frontiers | Research Topics

Insights
Frontiers in Built Environment

Edited by
Zhen Chen, Akira Matsumoto, Joan Ramon Casas, Vagelis Plevris, George Tsiatas, Hongling Guo, Yongkui Li and Sakdirat Kaewunruen

Published in
Frontiers in Built Environment

ISSN 1664-8714
ISBN 978-2-8325-4744-1
DOI 10.3389/978-2-8325-4744-1

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Insights: Frontiers in Built Environment

Topic editors

Zhen Chen — University of Strathclyde, United Kingdom
Akira Matsumoto — Nihon University, Japan
Joan Ramon Casas — Universitat Politecnica de Catalunya, Spain
Vagelis Plevris — Qatar University, Qatar
George Tsiatas — National Technical University of Athens, Greece
Hongling Guo — Tsinghua University, China
Yongkui Li — Tongji University, China
Sakdirat Kaewunruen — University of Birmingham, United Kingdom

Citation

Chen, Z., Matsumoto, A., Casas, J. R., Plevris, V., Tsiatas, G., Guo, H., Li, Y., Kaewunruen, S., eds. (2024). *Insights: Frontiers in Built Environment*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4744-1

Table of contents

frontiers in Built Environment

OPINION
published: 04 March 2022
doi: 10.3389/fbuil.2022.861018

The Involvement of Local Skilled Labour in Malaysia's Construction Industry

Mohd Ashraf Mohd Fateh[1]*, Mohamed Rizal Mohamed[2] and Siti Aisyah Omar[1]

[1]Centre of Studies for Construction, Faculty of Architecture, Planning and Surveying, Universiti Teknologi Mara, Shah Alam, Malaysia, [2]School of Architecture, Building and Design, Taylor's University, Subang Jaya, Malaysia

Keywords: local labour involvement, foreign labours, influx the labour force, skilled labour, construction

OPEN ACCESS

Edited by:
AbdulLateef Olanrewaju, Tunku Abdul Rahman University, Malaysia

Reviewed by:
Nurul Hafizah Yasin, Universiti Malaysia Kelantan, Malaysia

***Correspondence:**
Mohd Ashraf Mohd Fateh
mohdashraf@uitm.edu.my

Specialty section:
This article was submitted to Construction Management, a section of the journal Frontiers in Built Environment

Received: *24 January 2022*
Accepted: *03 February 2022*
Published: *04 March 2022*

Citation:
Mohd Fateh MA, Mohamed MR and Omar SA (2022) The Involvement of Local Skilled Labour in Malaysia's Construction Industry. Front. Built Environ. 8:861018. doi: 10.3389/fbuil.2022.861018

1 INTRODUCTION

In the 80s, Malaysia was one of the fastest developing countries with the introduction and development of mega-projects in the region. As reported by (Ibrahim et al., 2010), the construction industry plays an important role in the country's transition and its goal of becoming a developed country. It is expected that the Economic Transformation Program (ETP), the 11th Malaysia Plan (11 MP), and the 12th Malaysia Plan (12 MP), will transform Malaysia into one of the world's high-income nations. According to Olanrewaju et al. (2017), over the next 10 years, the country will need over one million construction workers. Olanrewaju et al. (2017) added that one of the Malaysian construction industry's biggest challenges is the lack of skilled labour. Mustafa Kamal et al. (2012) stated that the construction industry in Malaysia is still struggling with many problems and is associated with low quality, low productivity, unskilled labour, project delays, poor maintenance, non-conducting, and high on-site accident rates. The government aims to be a world-class, creative and knowledgeable solution provider for the Malaysian construction industry. To achieve the aim, the government, together with the CIDB, has made many efforts to upgrade the level of knowledge and skills among the construction players (Construction Industry Development Board Malaysia 2021). In the construction industry, 93% of registered foreign labourers with CIDB are unskilled labourers (Hisyam 2015). This was echoed by Hamzah et al. (2020) and Mohd Fateh et al. (2020), since most foreign labour comes from other countries, such as Indonesia, Bangladesh, and Myanmar, and is mostly unskilled labour. Many of them come from their county and only have basic construction knowledge. This issue has resulted directly in the lack of building expert labour in this country as work from these small nations has exposure to limited knowledge. The recruitment of low-salary foreign labour may ease the expenditure of the contractor, but it will not ensure the quality or even benefit of the Malaysian people.

The rapid development of the construction industry causes the demand for labour to increase, and the lack of local participation cannot meet the demand. Labour is one of the most constrained challenges faced by the Malaysian construction industry as the nation relies too heavily on a large foreign labour workforce (Najib et al., 2019). In the Malaysian construction industry, skilled labourers such as bricklayers, carpenters, painters, electricians, welders, plumbers, plant operators, among others, form a large part of the site labour force whose input determines, to a great extent, the quality of the industry's product. The common problems of low-skilled labour participation in construction have been traced to the unfair salary of labour, poor safety in construction sites, lack of clear-cut career paths, diminishing skilled labour training programs, and delays in the schedule of work on-site (Hussain, Xuetong, and Hussain 2020). Lingard (2013) reported that around 60–90% of the building work was carried out by foreigners and illegal foreign labourers, about half a million of whom came to Malaysia without a working permit or visa. However, Zaki et al. (2012) added that since the construction industry has a bad image for Malaysian labour, employers in the industry will have no choice but to bring in foreign labour to meet the demand for labour because these foreign labour will willingly accept the low salary and poor working conditions that construction sites offer.

Local skilled labour shortages in construction have become the Malaysian construction industry's biggest challenge. In addition, the participation of local labour is not highly motivated, and it is known that the needs of industry are not met by skilled labour created through vocational training. Some of them quit the construction industry after being educated by vocational education institutions. Malaysian construction has problems in obtaining the root of labour as well as retaining skills and needs to rely on foreign labour to meet the high demand for skilled labour due to rapid growth in Malaysia and poor local participation (Zaki, Mohamad, and Yusof 2012). The recruitment of foreign labours in Malaysia's construction sector is not a new issue. This issue starts since the British reign, Malaysia has begun taking foreign labours to fill the industry of need. Mustafa Kamal et al. (2012) highlighted that the critical issue in the industry is the dependency on a low salary, low skills, low overhead approach as price and affordability. Most of them cannot also learn and pass new technology into the building process. Whether foreign labours coming to Malaysia are skilled or not, it can also trigger several domestic cases that affect individuals or countries if not encourage the locals to involve in the construction industry then reduce the upcoming of foreign labour. Noh et al. (2016) reported that many of cases crimes are related to foreign and the people who involve in crime cases most of them are illegal immigrants and foreign labours. Foreign labour has caused several problems in Malaysian companies in the construction industry productivity. Firstly, when the organization decide to adopt more productive and modern methods of construction such as IBS. Most of the labours are not well trained and exposed to such a construction approach (Ismail et al., 2018). reported that some foreign labours have been unable to cope with a new working environment in large-scale projects because they do not have any experience handling big size projects. Some issues were also faced by the Malaysian construction organization when some of the foreign labours were absent during working hours and ran away after they reached Malaysia where the work on-site have begun halfway. Foreign labour also has a substantial indirectly economic effect on the nation. The remittances of foreign labours will be returned to their countries of origin, resulting in a currency outflow that causes our Malaysian ringgit to deteriorate.

In contrast, the inflation rate of foreign labours working in Malaysia is higher. This is because avoid paying a tax on salary, they buy goods and enjoy free benefits of charge. Not only that, their employers and other facilities guarantee foreign labour security without any related expenses (Malaysian Employers Federation 2014). Hiring foreign labour will also affect the wage structure of the industry. Employers are satisfied with the low salaries paid to foreign labours with the passage of time and the increase of foreign become highest than local labours decreasing due to employers could hire low skilled labour (Abdul-Rahman et al., 2012). The industry's reliance on foreign labours solely for the event with low value-added that require just lower level of expertise and low pay as emphasized by Del Carpio and Wagner (2015). Mahmood et al. (2021) conveyed that it has been stated that possible instability has such activities have been increased as violence, foreign worker's culture and illegal strikes. On the other hand, for the local labours, the bad sides of the construction industry make it more difficult to encourage locals to participate in the construction industry (Hamid, 2013; Mohd Najib et al., 2020). According to CIDB (2017) a survey in the Construction Industry Transformation Programme (CITP) shows that local people often see the construction industry as challenging to work in a negative sense, making them reluctant to take up jobs. In addition, jobs in the construction industry sector are on a contract or project basis (Construction Industry Development Board 2017). As the project was finished, their work contract also ended. This setting is one of the reasons locals decline to join the construction industry because the majority seek a permanent job that will be able to provide a secure and stable income. Labourers benefit from stability because it allows them to plan their cash flow and loan repayments.

Therefore, from all the issues highlighted, this paper is intended: 1) To investigate the current level of participation of local skilled labours in the Malaysian construction industry; 2) To determine the causes of the poor participation of local skilled labours in the Malaysian construction industry; and 3) To recommend initiatives to encourage the participation of local skilled labours in the construction industry.

2 LITERATURE REVIEW

2.1 The Construction Industry

Cambridge University Press (2019) defines construction as new development, modification, upgrading work, demolition, equipment and plant installation, and any remodeling from the original building design. While the construction industry is defined by the Project Management Institute (2013) as an application source for producing construction processes based on performance or achievement, scope, time, cost, and involvement of the temporary organisation in the project, like a construction firm, consultant firms like suppliers, professional services people, and a financial institution like a bank or government sector. Construction involves manpower, plant, construction materials, and management. The construction industry helps in developing and connecting infrastructure in the civil, structural, mechanical, and electrical aspects.

Malaysia's construction industry is one of the productive sectors that has made a significant contribution to the Malaysian economy as a growth catalyst for other industries. Malaysia's government plays a key role in the construction industry. One of the government's objectives is to allocate the nation's wealth to people to increase their standard of living. This is achieved by raising people's earnings and also creating employment opportunities. In addition, it is shown that the government can directly regulate the industry's demand. According to the CIDB, the labourers distinguish themselves as skilled labourers in the construction industry after obtaining some sort of accreditation from the relevant recognised bodies. This may include undergoing proper training or studying within the stipulated duration. In the context of obtaining recognition by the CIDB, the labourers will have to pass a test or courses that will be organised by some authorised bodies, such as Akademi Binaan Malaysia (ABM). The low participation of local skilled labour, in particular, is an issue that snowballs into a severe problem and may cause other problems, directly or indirectly, in the construction industry. The Malaysian construction industry is too dependent on foreign labour. Local skilled labourers are not attracted to working in the construction industry, especially if they are working on a construction site, because it is classified as undignified work and they demand higher pay. On the other hand, foreign workers are cheaper and can work long hours without any complaints.

As a result of this scenario, most employers preferred to hire foreign labour over local skilled labour. The International Trade and Industry Minister of Malaysia (MITI) stated that the local skilled labour shortage was a serious problem and needed immediate attention. Some of the construction organisations stated that they are willing to hire local skilled labour, but the labour leaves the organisation after receiving the necessary training. According to reports, some organisations invested a substantial amount of money on labour recruitment and training programmes, but only about 50% of the trainees served the companies, while others left (International Labour Organization 2016). It has also been observed that brain drain is one of the significant causes of the non-availability of local skilled labour in the industry (Jarmolowicz and Knapińska 2011). Most Malaysians prefer to work abroad, primarily in Singapore, Australia, the United States, and the United Kingdom, because these countries offer better opportunities and higher pay for their skills and experience (Jarmolowicz and Knapińska 2011).

2.2 Labour in the Construction Industry

Labour in the construction industry can be divided into two categories: unskilled and skilled labour. Work that requires no specific education or experience is often categorised as unskilled labour. The Department of Skill Development defines "skilled labour" as a worker who obtains level three certificates based on the Malaysia Skill Certificate (SKM) as a minimum qualification. Therefore, skilled labourers are people that have served an apprenticeship, practise the trade learned activity, and, because of their knowledge and vocational capacity, are given tasks that are particularly difficult and require a lot of experience that involve different trades of specialisation.

Labour is a key element in developing construction management processes. The lack of skilled labour has had a significant negative impact on construction costs, schedules, and construction performance. Skilled labour has received professional training; has at least 2 years of work experience, is knowledgeable about building materials, equipment, and safety, and is physically up-to-date (Abdul Hamid 2013). There are three possible avenues for training skilled labour, namely schools, vocational training centres, workshops, and on-site training directly (Sulaiman and Mohd Salleh 2016). Sulaiman and Mohd Salleh (2016) also added that the most commonly used skilled labour in the construction industry where their services are required most in construction projects are as follows:

- Carpenters: The job scope for carpenters is to construct, erect, install, or repair structures, fittings, or furniture made of wood; building frameworks, including partitions, joists, struts and rafters, wood staircases, window and door frames, and hardwood floors using their skills and tools (UNESCO 2021). The carpenters build the wood framing for houses, roofs, stairs, and decks, and construct formwork to support concrete work such as footings, columns, and stairs. Besides that, the carpenter is responsible for carpentry work such as installing cabinet siding, drywall rails, building cabinets and countertops, and working on drywall, wood flooring, metal jambs, and ceilings. Normally, carpenters are skilled at exterior and interior finish work.
- Bricklayers: Build walls, partitions, and other structures made of brick, and they also work with concrete blocks, bricks, tiles, marble, and terra cotta. The bricklayers can also do both construction and maintenance work. According to Wahab and Lawal (2011), many tools used by bricklayers include; trowels, brick hammers, chisels, levels, plumbs, and measuring squares, whereby unskilled workers will assist the bricklayer in doing the heavy work such as carrying materials and mixing mortars.
- Painters: According to Ali (2016), painters and decorators are often the last line of construction labour to finish a project before the tenants occupy the structure. The main duty of the painter is to provide not only aesthetic considerations for a bare structure, but also enhance the natural shape of a building and provide additional protection from the weather, wear and tear, and natural ageing process. Based on the findings by Božić-Štulić et al. (2019), painting is usually used on four elements of a building: interior walls, exterior walls, ceiling surfaces, wooden surfaces, and metal surfaces. Therefore, painters are highly skilled workers in the construction industry.
- Plasterer: It is one of the oldest and most required construction trades, which is responsible for applying stucco and plaster to building components for insulation, support, aesthetic, and smooth background on walls both internally and externally. To achieve this, the efficient services of skilled masonry work are required for the plastering and rendering finishes.
- Tiler: Tilers are responsible for installing hard tile and marble on floors, decks, and walls as specified in any construction project. There is also a roof tiler who can lay roof tiles on the roof frame. According to Lam and Fu (2019), tile fixers must have the training to set their tiles properly. Furthermore, qualified tile fixers must be able to be fast and efficient to ensure less risk of the project running behind schedule. However, tile fixers are susceptible to some injuries whereby constant straining and bending over a surface can lead to repetitive stress and injuries. For this reason, a trained tile fixer is needed to minimise the vulnerability involved in the work.

TABLE 1 | Numbers of foreign labour in Malaysia's construction industry.

Year	Numbers of foreign labour in Malaysia's construction industry
2015	428,469
2016	369,774
2017	339,712
2018 (until February)	315,614
Total	1,453,569

(Source: Surendran, 2021)

TABLE 2 | Statistic of registered personnel by CIDB.

Year	Local labour	Foreign labour
2016	208,087	14,820
2017	198,375	28,178
2018	109,156	41,736
Total	515,618	84,734

(Source: Construction Industry Development Board Malaysia, 2021).

2.3 Foreign Labour Forces in Malaysia's Construction Industry

According to current estimates, nearly RM5 billion is sent out of the country each year by foreign workers working in the construction sector, who send money back to their home countries. The Malaysian CIDB stated that there were currently 420,000 foreign workers in the sector who were registered with the Malaysian Immigration Department (The Star 2019). Malaysia, with a population of 32 million people, is heavily dependent on foreign labour. They are concentrated on blue-collar jobs. There are some 1.76 million foreign workers who work legally in Malaysia, and an estimated 3.9 million to 5.5 million more illegal immigrants, as highlighted in **Table 1**. This shows the numbers of registered unskilled foreign labour in the Malaysian construction industry until 2018, excluding illegal and non-registered labour. Even though the numbers are decreasing year to year, the majority of the foreign labour forces in Malaysia are unskilled labour, which is not preferable since the construction industry requires skilled labour.

2.4 The Pattern of Local Skilled Labour Participation in the Construction Industry

Malaysia seeks to overcome its reliance on low-skilled foreign labour as it attempts to move up the economic ladder. Nevertheless, it might impact other industries in the country. In the Malaysian construction industry, thousands of labourers are involved, but most of them are classified as unskilled labour. The country is experiencing a shortage of skilled and productive labour adequately trained for certain jobs. Skilled labour is the most important thing to improve the industry. The essential information of skilled labour in the construction sector must be kept up to date. If the current scenario is exposed, the authorities will find it easier to take concrete steps to improve the performance and efficiency of Malaysian labour (Najib et al., 2019). Furthermore, one of the national agendas is to reduce the number of overseas labourers by more than half in 5 years while getting companies to hire more high-skilled Malaysians to become a more developed economy. **Table 2** presents statistics of registered building personnel by CIDB from 2016 to 2018. The local labour force that registered with CIDB from 2017 to 2018 plunged by −45%, while the foreign labour force rocketed by +48%. This data is alarming as the increasing numbers of foreign labourers in this country concerns us because most of them are not skilled labourers who degrade work quality and also become a social issue (Manoharan et al., 2021).

2.5 The Needs of Local Skilled Labour in the Malaysian Construction Industry

Saieed (2016) highlights that after the pandemic, the economy is expected to grow, although the headwinds buffeting the Malaysian economy make it difficult to reach the target's upper band. Growing would mean delivering higher-value goods and services, which will increase productivity. An increase in productivity would mean a better salary for the labour force. This is the policymakers' basic argument as they speak about how human capital will help the economy. However, the reality is different. According to data from the Malaysian Productivity Corp, the average annual labour productivity growth between 2011 and 2015 was 1.8%, while the 11 MP has a target of 3.7% annual growth. Doubling the growth in labour productivity is needed to meet the New Economic Model's high-income target. Department of Statistics Malaysia, (2020) notes that the economy saw a 3.3% increase in labour productivity last year but claims that it will be difficult for labour productivity to grow in the coming years due to the lack of skilled labour.

The economy will struggle to move up the value chain without more skilled labour and will not be able to attract large investments in resources (Saieed, 2016). However, the target of 11 MP is well below the ratio of skilled labour to developing economies, where the ratio is at least half of the total workforce. The government's plans to raise the skill levels of Malaysian labourers have so far only shown mixed results, with a gap between the plans and their actual implementation (International Labour Organization 2018). In the context of the construction industry, it requires a large number of skilled workers to meet the demand for construction projects. Skilled labour may face difficulties in meeting the demand. The industry is in recovery and a lack of expertise could threaten to disrupt what could well be the most positive outlook for construction in a decade.

2.6 Causes of Poor Participation of Local Skilled Labour in Malaysia's Construction Industry

Formal education achievement cannot be used as a criterion or yardstick in the requirements for jobs in the context of skilled labour. Employers often use the rate and perceived quality of education as a fast-screening mechanism for hiring. Often, formal education does not perfectly align with one's ability to do the job. The great ability to do the work comes from experience. People learn from their own and other people's experiences. That is why training on-site operations is vital. When they are trained in real conditions, they gain real experience. Working with good quality results from good work skills, which cannot be obtained solely through classroom learning.

2.6.1 Poor Working Environment in the Construction Industry

Construction Industry Transformation Programme (CITP) findings by CIDB (2017) reveal that local people still view the construction industry with a negative perception as being difficult to work in, making local labour refuse to take a job in the construction industry. Not only does the negative image stem from the nature of construction, which some attribute to a dirty workplace, dangerous working conditions, and difficulty understanding the method and process. It is the main reason for hiring foreign labour. Building labour around the world has always been poor in terms of employment (International Labour Organization 2016). In addition, accidents on the construction site are common until the labourers believe they are inevitable. Malaysian labourers work without fully equipped types of machinery on the job site and their working conditions are more dangerous than in other developed countries. Local labour claims that there are numerous unreported incidents that worry them about the lack of procedure in safety management.

According to Ofori (2003), reducing construction accidents and fatalities will improve the industry's image and better suit the information society by attracting more Malaysian labour. Furthermore, local labourers have lost interest in construction work as a result of the filthy, difficult, and dengue issues that have long been associated with the industry, leading local people to refuse to be accepted by the industry, and some employers choose to hire foreign labour rather than improve working conditions to attract local people (Abdul-Rahman et al., 2012). The working environment was unable to attract local labour to meet the labour force's rapid demand (Del Carpio et al., 2015).

2.6.2 Low Assuredness of Salary

Local labour prefers to find the opportunity to work in a different country to earn better salaries than Malaysia (Abdul-Rahman et al., 2012). In addition, the Malaysian Trade Union Congress (MTUC) has failed to convince the government to stop the influx of foreign labour into the country, which depresses the wage structure and reduces opportunities to hire local labour (Najib et al., 2019). This was echoed by Trevena (2013), stating that foreign labourers are willing to accept relatively lower salaries compared to local labourers, thus undermining the wage structures. Local labour typically earns 40% more than foreign labour, which includes pension funds, medical and social benefits. Therefore, hiring foreign labour is a cheap option to keep the production costs low and reduce the bargaining power of the local labour force in the market (International Labour Office 2018). Due to the low salary, the foreign labour has a big opportunity to work in the Malaysian construction industry since the locals prefer to work in their neighbours' countries.

There is a small incentive offered to contractors for more productive, better quality, and safer technologies to be adopted and implemented. The condition also affects the willingness of contractors to hire highly skilled labour (Construction Industry Development Board Malaysia 2021). In particular, the construction industry offers many job opportunities to people who are not selective in their profession and who desperately need a job. Unfortunately, only foreign labourers have joined the bandwagon. However, it leads to poor workmanship, waste of materials, improper use of equipment, and permanent employment that is not enjoyed by labour. Labourers also did not have the benefits normally related to employment and welfare benefits. This arrangement does not attract local youth to participate because it implies that working in the construction sector is not valuable.

2.6.3 Higher Education Level Among Malaysians

Malaysia's education sector appears to be growing steadily. However, this progress means labour forces can only be occupied by low-skilled foreign employees. To destabilise an economy, the country needs to demonstrate strong resilience to face the crisis of globalisation where the economy was generated by the development of manpower in a few sectors in Malaysia. So, while manpower will generate the economy, Malaysia must ensure their employability, or the entire field may suffer as a result of the loss of various jobs in Malaysia. Malaysian educational standards have been continuously improving over the years, assisting the current generation to achieve higher levels of learning.

Malaysian will prefer to look for more lucrative jobs where they are educated; graduates will only focus on jobs that match their qualifications, causing them to become choosy and exacting in their job selection. Dom et al. (2012) emphasised this point, stating that the local youth are pursuing higher education in exchange for a more lucrative profession, and as a result, they are living differently and better than their parents. The industry has been reduced to low-value-added activities requiring only lower skill levels and offering low pay due to the reliance on foreign labour (Del Carpio and Wagner 2015). Working on a construction site always appears to be a low-status job, lacking prestige, class, and respectability. This was agreed upon by Mohamed (2015), stating that the stigma affects those who avoid the construction industry. This is why some of the local graduates prefer to remain unemployed rather than work in the construction industry and indirectly contribute to society (International Labour Organization 2016).

As for foreign labour, they are driven to improve the quality and value of their lifecycle, thus willing to work in any condition. They are motivated to provide a better future for their children (International Labour Organization 2016). Furthermore, the demand for foreign labour is lower in terms of supply and demand in the job market. It demonstrates that they are willing to work in a tough environment and for long hours, even if the salary is slightly lower. Mohamed (2015) reported that the local labour force is more choosy and aware of their rights when they inquire about and challenge their employers. It is evident when they dare to inquire about any unfair management practises directed toward them. It is the opposite with foreign labour, where they only obey the rules, when instructions are given, and they are more obedient where they accept the jobs without any serious complaint as long as they get the jobs and get paid. The employer prefers to hire foreign labourers who are resilient enough to agree to work in any tough environment, even if the pay is minimum.

2.6.4 Monopoly by Foreign Labour

Malaysia's dependence on foreign labour is very high, especially in the service, construction, agriculture, and manufacturing sectors. The arrangements benefit both parties, namely Malaysia and foreign workers from countries such as Indonesia, Nepal, Myanmar, Bangladesh, and others. Trevena (2013) reported that unskilled foreign labour fills the vacancies in the construction and other economic sectors as these jobs are physically demanding and labour-intensive. Due to this scenario, it may appear that foreign labour has monopolised the labour force, and leaving local labour as minorities. Since the number of graduates in Malaysia is rising, some fresh graduates may need to work in any sector at entry level before making the transition to their desired position or sectors. The monopolised state is making locals feel they need to compete with them as well. The tremendous monopolising of foreign labour into the country is also causing an outflow of money. As much as 80% of a foreign worker's pay is sent back to their respective home countries regularly, which usually means currency exchange for Malaysian Ringgit (MYR) is higher compared to their home countries (Fateh et al., 2020).

2.7 Initiatives to Improve Local Skilled Labour Participation in the Construction Industry

The lack of experienced and skilled construction labour has become a worldwide problem. Local participation in the workforce is not very encouraging, and it is recognised that skilled labour created through vocational training does not meet industry needs. It has also become a major challenge facing the Malaysian construction industry. Some of them left the construction sector after graduating from construction-related courses. There are problems with Malaysian construction in the ability to obtain the origin of work and maintain skills, and it has to rely on foreign labour to meet the high demand for skilled labour due to rapid growth in Malaysia and low participation by Malaysians (Zaki, Mohamad, and Yusof 2012).

Therefore, the government has identified the need to intensify the construction industry. Mechanization, new building methods and technology, and prefabrication have the effect of reducing reliance on labour and improving contractors' performance and productivity. The government has tightened the work permit requirements to limit foreign labour inflows and increased the levy on foreign labour based on their field of employment. The government also decided not to extend work permits to non-qualified foreign labourers who have been in the country for more than 5 years and has granted amnesty to illegal foreign labourers to return to their own countries without legal action. The CIDB has announced several initiatives to improve the image of the industry and raise awareness among local labour of the benefits of joining the industry, including improving incentive programs, technical and vocational education and training, raising wages to attract local labour, improving site conditions and safety practices, and raising awareness of the opportunity to get local participation and governance structure and implementing policies. Nevertheless, the changes might not happen overnight. The effects are very slow to take effect from the reforms initiated by the CIDB.

2.7.1 Improving Working Conditions and Safety Practice

The poor impression of working conditions and safety practises on the construction site is one of the reasons that most local labourers are turning away from work in the construction industry. French & Jones (2019) reported that most labourers nowadays have high-risk awareness. In this day and age, these issues need to be addressed. The introduction of standard safety equipment such as safety helmets and protective clothing to prevent accidents and injuries on site is one of the few ways to improve working conditions (Hammer et al., 2016). If the construction industry has a respectable reputation, more local people will be intrigued by taking up construction jobs, thus decreasing the dependence on foreign labour. Kalagatur et al. (2018) added that there are several ways of ensuring a decent working environment by reducing adverse environmental effects such as dust, noise, debris, and unhealthy emissions. This method is being implemented gradually in order to create a positive perception, even though it may be difficult for the construction player to implement it overnight.

Employers play a critical role in improving workplace safety practises, which will boost the local labour force's confidence in joining the construction industry (Manap, Noh, and Syahrom 2017). Although most construction-related activities are high-risk, employers are also responsible for ensuring that their employees are not exposed to risks that could jeopardise their safety and health. The contractor must have adequate safety and health policies in place to reduce workplace risk (Deros et al., 2014). The government is concerned about this. The CIDB has made it mandatory for all construction workers to complete the 'green card' course. The course will cover the fundamentals of security measures, safety and health issues, and risk management while on the job. This will provide some exposure for the labourers who will be working on the construction site.

2.7.2 Awareness of Opportunities

One way to instil the interest of the younger generations in taking part in the construction industry is to cultivate their interest while they are still acquiring university education. French & Jones (2019) suggested that apart from learning the importance of the theories and calculations in construction, the syllabus can also expose them to the new state-of-the art technology and innovation that is happening in the construction industry. The university can also invite industry players to give a positive “pep talk” to the students on the perception of working in the construction industry. In addition to that, throwing in some success stories will also increase the students’ interest in the construction industry. The platform can be used by construction industry participants to dispel any negative and misleading image of the construction industry. According to Couth et al. (2019), any university career fair can be a good platform to raise awareness for joining the construction industry. Career advisors should illustrate the potential for career enrichment and advancement in the construction industry, as agreed by Abdul-Rahman et al. (2012) and Jaafar et al. (2015).

2.7.3 Encourage Labour-Intensive Construction With an Industrial Building System

The CIDB described IBS as a construction approach using a mixture of construction components that are either manufactured on or off-site. The finished components were then mobilised and installed on-site like a “lego” system. In addition, Mohd Fateh et al. (2021) stated that advanced countries such as Australia, the UK, and Singapore have implemented the IBS construction approach, which is able to reduce unskilled labour and increase productivity. In the Malaysian context, IBS adoption might be able to reduce the large-scale number of foreign labourers on-site and attract local labourers to a more conducive and controlled environment (manufacturing yard or factory) (Mydin et al., 2014). Lui and Wen. (2020) and Fateh et al. (2019) detailed that the government should encourage labour-intensive sectors to utilise modern technology, especially IBS. This move will help the government achieve its goal of producing 35% of skilled labour by 2020, in line with its efforts to strengthen the country’s TVET.

2.7.4 Governance Structures and Recruitment Policies

Osborne & Hammoud (2017) and Abdul-Rahman et al. (2012) emphasise that improving the governance structures and policies could reduce the influx of foreign labour in the country. Del Carpio et al. (2015) highlighted that more impactful policies need to be introduced to reduce over-dependency on foreign labour without harming the economic sectors in Malaysia. Existing instruments such as quotas, dependence thresholds and levies can be changed to drive the market more while promoting performance in line with Malaysia’s economic goals. It should also be remembered that any proposed reforms need to be complemented by active oversight and implementation on the ground.

3 RESEARCH METHODOLOGY

Literature reviews and questionnaire surveys were used for the data collection. The questionnaires were distributed online. This is suitable for the new normal where close contact and discussion are reduced. The questionnaire comprises four sections. The first section focuses on the respondent’s demographics, which includes years of experience in the industry. Next, the second section dives into the current scenario of local skilled labour participation in Malaysia. In the third section, emphasis is placed on determining the causes of poor participation of local skilled labour in the Malaysian construction industry. Lastly, the last section recorded the recommendation to encourage higher participation of local skilled labour in the construction industry. A 5-point Likert scale was used for sections two, three, and four with a scale of 1-totally disagree to 5-totally agree.

TABLE 3 | Respondent’s position in the company.

Respondent’s position in the company	Frequency	Percentage (%)
Managing Director	6	5.66
Project Manager	64	60.38
Site Manager	30	28.30
Human Resource Manager	6	5.66
Total	106	100.00

Mohd Fateh et al. (2016) recommended running a pilot test on the questionnaire to check on typos, structures, and the clarity of the question itself. Therefore, the researcher has run a pilot test on five (5) respondents consisting of two (2) senior lecturers, one (1) quantity surveyor, and three (3) contractors. With all the comments and feedback from the pilot test, the questionnaire was amended accordingly. Simple random sampling was used based on the record on CIDB’s website. The questionnaire was distributed to 200 contractors (grade 7) in the Klang Valley. The contractor was chosen because they were dealing directly with the labour on and off-site. Thus, their opinions and feedback are valuable and highly related to the research context. A total of 106 respondents were received, which translates to a 53% response rate.

According to Takim & Adnan (2008) and Mohd Fateh et al. (2020) for a built environment research and self-administrated questionnaire, the acceptable response rate was 20–35%. Thus, based on the response rate, the data collection is reliable and acceptable. An internal reliability test was done using the Cronbach’s alpha analysis. The result will be discussed in the next section. All the data collected was analysed using the Statistical Package for the Social Sciences (SPSS). Variable statistical analysis was used, such as mean, frequencies, and percentage analysis.

4 FINDINGS AND DISCUSSION

4.1 Internal Reliability Test (Cronbach’s Analysis)

A reliability test was performed to assess the questionnaire survey’s internal consistency. The Cronbach’s alpha formula was used to compute the interval consistency level for twenty (20) items, and the result was 0.795, indicating that the items were interrelated and compatible with the study sample; hence, the interval consistency level is adequate (Pallant 2016). The closer the result to 1, the better the internal consistency it is. (Sekaran and Bougie 2016).

TABLE 4 | Respondents' work experience.

Year	Frequency	Percentage (%)
1 year to 5 years	10	9.43
6–10 years	25	23.58
11–20 years	33	31.14
21 years and above	38	35.85
Total	106	100.00

TABLE 5 | Respondents' academic qualifications.

Respondents' academic qualifications	Frequency	Percentage (%)
Diploma	8	7.55
Bachelor's Degree	62	58.49
Master's Degree	36	33.96
Total	106	100.00

4.2 Demographic

According to **Table 3**, the majority of respondents worked as project managers, with 64 respondents (60.38%). The second highest was the site manager, with 30 respondents (28.30%), followed by the managing director and the human resource manager, with 6 respondents (5.66%) each. All positions were directly engaged with labour on-site. Nevertheless, the human resources manager and the managing director were included as well because they were directly involved in the recruitment process, and their input will be useful and relevant to the research.

According to **Table 4**, respondents with 1–5 years of experience have the lowest response rate, with only 10 respondents (9.43%), followed by 5–10 years of experience, with 25 respondents (23.58%). Respondents with 10–20 years are in third place with 33 respondents (31.14%). The highest were the respondents with 20 or more years of experience, with 38 respondents (35.85%).

Table 5 shows that the majority of the respondents were bachelor's degree holders, with 62 respondents (58.49%). Next, 36 respondents (33.96) were Master's degree holders, and lastly, 8 respondents (7.55%) were diploma holders. The demographic shows that the respondents have vast experience in the industry and sit in prominent positions in the organisation, thus their input is reliable and resourceful.

4.3 The Current Scenario of Local Skilled Labour Participation in Malaysia

Table 6 illustrates the findings from respondents about the current scenario local skilled labour participation in the construction industry. The table exhibited the ranking and mean value of current issues related to the low participation of local skilled labour in the Malaysian construction industry. The first rank of the main current scenario participation of local skilled labours in Malaysian construction is that many local skilled labours join training education such as "Sijil Kecekapan Kemahiran (SKK)" but are not participating in the construction industry with a mean value of 3.88. Most of the respondents highlighted that after the locals completed the SKK provided by the government, many preferred to work in other sectors rather than stick to the construction industry. Construction-related training is administered by the CIDB, which is part of ABM (The Star 2009). Approximately 40,000 trainees were trained under the CIDB Youth Skills Training Program between 1999 and June 2007, and CIDB also trained 40,000 construction employees at the same time. Surprisingly, the vast majority of these trainees (80%) do not work in the construction industry. The main reason behind this is that the trainees were not exposed to building sites and were not equipped with the appropriate mentality for working in the real world.

The second highest is that local skilled labour prefers building work such as architectural, mechanical, and electrical work over structural work. This statement shows that most respondents agreed that local labourers do not prefer to work under harsh weather conditions because of the structural work exposed under hot conditions, instead of mechanical and electrical work that is not exposed directly. It can be seen that the average recorded a mean value of 3.86. Next, most respondents agreed that the local labour prefers to be a site leader, such as a site supervisor, rather than become normal labour with a mean value of 3.73. From their perspective, normal labour is required to handle heavy task work and expose themselves to harsh weather conditions, whereas a site supervisor is more focused on the monitoring of work and leading the labour on the arrangement of work that needs to be done.

Finally, respondents moderately agree that even with a higher salary, local labour participation may still be low, with a mean value of 3.64. The construction industry portrays an image of a 3D (dirty, dangerous, and demeaning) working environment. Some locals are not comfortable being in these settings. Some

TABLE 6 | The current scenario of the participation of local skilled labour in Malaysia.

Current scenario of the participation of local skilled labour in Malaysia	Mean value	Rank
Many local skilled labours join the training education such as 'Sijil Kecekapan Kemahiran (SKK)' but are not participating in the construction industry	3.88	1
Local skilled labours prefer to choose work in building such as architectural works and M&E works compare to structural work	3.86	2
Local skilled labours prefer to work as a leader such as site supervisors rather than being just labour	3.73	3
Even with a high salary, the participation of local skilled labours is still low	3.64	4

TABLE 7 | Causes of low local skilled labour participation in Malaysia's construction industry.

Causes of low local skilled labour participation in Malaysia's construction industry		Mean	Overall mean
Monopolize by foreign labours	Malaysia's dependence on foreign labours is very high	4.16	3.90
	Foreign labours are monopolizing sectors of employment in the country	3.90	
	An increase in the number of local skilled labours in Malaysia is causing many of them to be unemployed as there are insufficient jobs around	3.66	
Low Assuredness of Salary	Employers prefer to hire foreign labours because of lower salaries	3.94	3.87
	Local skilled labours prefer to find a job overseas to earn a better salary	3.81	
Academic qualification level	Local skilled labours would find work that matches their qualification/certification instead of just accepting any work from contractors	3.86	3.82
	With higher education levels, local skilled labours prefer more lucrative positions	3.81	
	Local skilled labours are more demanding and have more choice regarding the type of work they want to do	3.79	
Poor Working Environment	Local skilled labours believe that poor image from the nature of construction industry which is the dirty workplace, dangerous condition and difficult to understand the method of work	3.54	3.44
	Accidents on construction sites are common and these lead to local skilled labours thinking that accidents could also happen to them	3.47	
	Local skilled labours claim that there are many other unreported incidents and these make them worry about the lack of procedure in managing their safety	3.32	

respondents even highlighted that even if the salary is slightly lower but there is an opportunity to work in an air-conditioned office, they are willing to go. Nevertheless, our local labourers who are willing to work on-site prefer to go to our neighbouring country, Singapore. Higher salary prospects become more attractive than in Malaysia, and the salaries are paid in Singapore Dollars, where the conversion rate significantly multiplies the amount.

4.4 Causes of Low Local Skilled Labour Participation in Malaysia's Construction Industry

This section discussed the causes of the poor participation of local labour in the construction industry. **Table 7** summarises the findings and is arranged based on the overall mean value. The findings can be categorised into four (4) themes, namely (based on rank) migration development, the low assurance of salary, higher education level, and poor working environment.

Foreign labour monopoly ranks first among the causes for poor participation of local labourers, with a 3.90 overall mean value. The construction is driven by foreign labour. It demonstrates that Malaysia heavily relies on foreign labour. The migrants are monopolising the labour employment, which makes the locals look like minorities if they work as labour as well. This scenario could play out in other sectors in the country, such as service and manufacturing. The economic landscape and political governance may play a significant role in attracting foreign labour to come and work here in Malaysia. The MTUC failed to convince the government to stop the influx of foreign labour, which has the potential to depress the wage structures (New Straits Times 2020). This was echoed by Mohd Fateh et al. (2020), who pointed out that in Malaysia, it is a common perception that hiring foreign workers is less expensive than hiring local labour.

Next, is the low assurance of salary, with an overall mean value of 3.87. Most employers prefer to hire foreign labour because of their lower salaries and willingness to work long hours compared to local labour. Foreign labourers are willing to accept a low salary as long as they have a secure job (Abdul-Aziz, 2001; Narayanan & Lai, 2014). In the long run, it will have a bad impact on the construction industry because the overall average salary rate will deteriorate, and the rate of increment is very slow. Indirectly, it will affect the attractiveness of the local labour market (Del Carpio et al., 2015).

The third reason for the low participation of local labour in the Malaysian construction industry is the academic qualification level, with an overall mean value of 3.82. Most local labourers would find a position that matches their academic qualification level only, instead of accepting any position that is being offered. For example, diploma holders would like to be in a position that suits their qualifications and will not accept anything less. For a fresh graduate, it is better to gain experience first-hand rather than be too picky about the position offered. Vast experience is a quality that the construction industry values. Even if the labourer lacks formal qualifications, he has vast experience under his belt, which will carry significant weightage in his resume.

In 2021, the Construction Research Institute of Malaysia (CREAM) (2011) reported that the perception that working on a construction site is portrayed as a low-status job with a lack of prestige, class, and respectability. This perception is misleading and affects participation indirectly. The International Labour Office (2018) suggested that rather than being unemployed, which might lead to other problems, going to work in the construction industry is a decent choice to make. There are plenty of things that the applicants can learn, and they can be stepping stones for them to explore in the future. While it is the total opposite for foreign labour, Mohamed (2015) highlighted that they will grab any available jobs even if it means long hours and a harsh environment. The foreign labourers are more obedient as long as they get their payment as promised.

The poor participants of the local labours in the construction sector ranked the poor working environment on the construction site the lowest, with an overall mean value of 3.44. Mohd Fateh

TABLE 8 | Initiatives to encourage the participation of local labour in the construction industry.

Initiatives to encourage the participation of local labour in the construction industry		Mean	Overall mean
Provide Economic Welfare to the Labour	Provide comfortable housing or accommodation to local skilled labours	4.37	4.21
	Contractors should provide some activities as incentives whenever labours have completed some important milestones in the projects	4.05	
Encourage Labours Intensive with Industrial Building System (IBS)	Introduce to local skilled labours the IBS technology, thus improving participation	4.13	4.09
	With the IBS hiring, local skilled labours will increase as compared to hiring low-skilled foreign labours	4.09	
	IBS system can attract the local skilled labours to join the construction industry because the IBS method is simpler to construct	4.07	
Raising Salary to Attract the Local Labour	Give added benefits or rewards to local skilled labours who perform their jobs well	4.11	3.92
	Differentiate salary of local and foreign skilled labours	3.73	

et al. (2020) highlighted that the construction industry has been battered by chronic issues such as poor health and safety conditions, poor working conditions, and lack of quality, resulting in poor productivity and overall efficiency. Local labour believes that working on construction sites is a 3D (dirty, dangerous, and demeaning) environment that may lead to site injuries and accidents. Working on construction sites may portray less prestigious because it is claimed that it does not require a great deal of skill or intelligence (Rahim et al., 2018). However, the claim is deceptive. As you progress through the stages, the complexity of the work will increase as well. It requires creativity, critical thinking, and good communication skills to solve any issues that may arise.

Safety precautions on the construction site are a serious concern. Most organisations now have a competent safety officer who monitors and recommends the safety policy and procedure to ensure that the site is a safe place to work.

4.5 Initiatives to Increase Local Skilled Labour Participation in Construction

This section elaborates on the recommendation on the initiatives that relevant stakeholders can look into on how to encourage the local labour force to participate more in the construction industry. According to the findings, there are six (6) themes (based on the rank): provide economic welfare to labour, encouraging labours intensive with IBS, raising wages to attract local labour, improving working conditions and safety practices, and raising awareness of opportunities. **Table 8** summarises the overall mean and ranks the initiatives on how to encourage the local labour force to participate more in the construction industry.

The first initiative, with an overall mean value of 4.21, agreed upon by the respondents on how to encourage the local labour force to participate more in the construction industry is to provide economic welfare to the labour force, which includes providing comfortable accommodation and providing some activities as incentives whenever labour has completed some important milestones in the projects. The comfortable accommodation can attract local labour to join the construction industry. Usually, only temporary accommodation was given to the labourers, which led to overcrowding, inadequate sanitation, unregulated drainage of surface water, and improper disposal of waste.

Malaysians are reluctant to remain in that condition and feel that jobs in the construction sector are not dignified enough. cramped, dirty, and unhygienic conditions. This was agreed upon by Marhani et al. (2012), who reported that the accommodation is cramped and dirty. Indirectly, it might impact the labourers' health and the surrounding neighbourhood. The respondents added that the labourers have spent 8–9 h on-site and are looking forward to going back to decent accommodation to rest and recharge for the next day. Mohd Fateh et al. (2020) added that the accommodation needs to meet with the national legislation and international good practice, for example, a minimum amount of space for each worker, provision of sanitary, laundry and cooking facilities, and portable water.

Next, the second initiative that was agreed upon by the respondents is fully utilising the IBS, with an overall mean of 4.09. The respondents highlighted that most of the work being done for the IBS construction approach is off-site and in a controlled environment. CIDB (2015) described IBS as a construction system using a mixture of construction components that are either manufactured on or off-site and then installed and assembled into construction structures. It can be a factory or fabrication yard where the conditions are not too harsh compared to a typical construction site. In these settings, it can reduce the perceptions of the 3D workplace that have been portrayed. The increase of IBS in Malaysia can be a catalyst to reduce the dependency on foreign labour and start engaging with local talents. Mydin et al. (2014) concluded that since IBS construction relies more on state-of-the-art construction technology, reliance on manual labour can be reduced, and subsequently, foreign labour flows to Malaysia are reduced.

Low-skilled foreign labour fills vacancies in construction and other economic sectors because these jobs are physically demanding and labour intensive, and thus are often avoided by most locals (Kupets 2016). Many foreign workers are willing to accept lower wages in comparison to Malaysian workers, undermining wage structures. Kupets (2016) also exposed that local construction labour generally receives 40% higher pay than their foreign counterparts. Pension funds, medical and social benefits are among the legal requirements, and other employee benefits are relieved by employers upon hiring a foreign worker. Therefore, hiring a cheap foreign workforce keeps the production costs low and reduces the bargaining power of locals in the labour market. The tremendous influx of foreign labour into the country is also causing an outflow of money. Some studies have shown that as much as 80% of a foreign worker's pay is sent back to their respective home countries regularly.

The third initiative by the respondents was to increase the minimum salary and introduce performance benefits for the labour in the construction sectors, with an overall mean value of 3.92. This was agreed upon by Narayanan & Lai (2014), who highlighted that recruiting local labour to participate more in the construction industry increases the minimum salary. Pillai et al. (2016) reported that the MTUC had called on the government to create a minimum wage for the lowest-paid labourers to meet the cost of living that had arisen in the 2000s. It is also a prominent employee engagement programme that includes career advancement and learning and development opportunities. The CIDB claims that if Malaysia can build more participation, they can be paid higher by skilled local labour and industry, and the country will break foreign workers' reliance.

5 CONCLUSION

In conclusion, this is one of the construction industry's most pressing issues. The construction industry is a labour-intensive sector. Malaysia has been dependent on immigrant labour for many years to meet its demand for construction workers. However, Malaysia's construction industry is plagued by the low-salary and low-productivity traps resulting from heavy reliance on low-skilled foreign labour. Not only are the numbers of local workers small, but they also dramatically decline as a result of migration for better job opportunities in other industries. The findings from research might be able to improve the participation of local skilled labourers in the construction labour force. The related stakeholders can utilise the findings and assist in improving the subject matter.

AUTHOR CONTRIBUTIONS

MF and SO: conceptualization, resources, technique, formal analysis, discussion, visualisation, and writing—first drafts preparation and visualization. MF and MM are in charge of writing, evaluating, editing, and disseminating the paper. The published version of the manuscript has been read and approved by all authors.

ACKNOWLEDGMENTS

The publication is part of the author's ongoing research at Universiti Teknologi Mara (UiTM), Malaysia. The Construction Industry Development Board (CIDB) and UiTM deserve special recognition for their contributions to the research.

REFERENCES

Abdul-Aziz, A.-R. (2001). "Foreign Workers and Labour Segmentation in Malaysia's Construction Sector.". *Construction Manag. Econ.* 19 (December), 789–798. doi:10.1080/01446190110072022

Abdul-Rahman, H., Wang, C., Wang, C., Wood, L. C., and Low, S. F. (2012). Negative Impact Induced by Foreign Workers: Evidence in Malaysian Construction Sector. *Habitat Int.* 36 (4), 433–443. doi:10.1016/j.habitatint.2012.03.002

Ali, Z. A. (2016). Improving Skilled Workers' Performance in Construction Projects in Nigeria. *Johor: Universiti Tun Hussein Onn Malaysia*. Available at: https://core.ac.uk/download/pdf/141497896.pdf.

Božić-Štulić, D., Željko, M., and Gotovac, S. (2019). "Deep Learning Approach in Aerial Imagery for Supporting Land Search and Rescue Missions.". *Int. J. Comp. Vis.* 127 (9), 1256–1278.

Cambridge University Press (2019). *Cambridge Dictionary*. Cambridge University Press. 2019.

Carpio, X. D., Özden, Ç., Testaverde, M., Marouani, M., Nilsson, B., and Wagner, M. (2015). Foreign Workers in Malaysia: Labour Market and Firm Level Analysis. *Malaysian J. Econ. Stud.* 52 (1), 1–19.

Carpio, X. V. D., and Wagner, M. (2015). The Impact of Syrian Refugees on the Turkish Economy. *World Bank Pol. Res. Working Paper*. no. August. Available at: http://www.mdpi.com/2076-0760/6/4/129.

CIDB (2015). Government Circular on IBS. Available at: http://www.cidb.gov.my/cidbv4/index.php?option=com_content&view=article&id=595&Itemid=559&lang=en.

Construction Industry Development Board Malaysia (2021). Construction Industry Statistic (Open Data). *Construction Statistic* 2021. Available at: https://www.cidb.gov.my/en/construction-info/construction-statistic/construction-industry-statistic-open-data.

Construction Industry Development Board (2017).Malaysia Country Report. In 22nd Asia Construct Conference. Seoul. Available at: http://www.cidb.gov.my/images/content/international/Malaysia–-Country-Report-2017–22nd-Asia-Construct.pdf.

Construction Research Institute of Malaysia (CREAM) (2011). *The Current State of Industrialised Building System (IBS) Construction in Malaysia: Proceedings of 1st IBS Roundtable Workshop. Makmal Kerja Raya Malaysia*. 1st ed. Kuala Lumpur: Makmal Kerja Raya Malaysia. https://books.google.com.my/books/about/The_Current_State_of_Industrialised_Buil.html?id=CzGenQEACAAJ&redir_esc=y.

Department of Statistics Malaysia (2020). "Labour Market Review, First Quarter 2020." Kuala Lumpur. https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=100&bul_id=R09wdGZSektvNmw5T1VCeVphNXRqdz09&menu_id=TE5CRUZCblh4ZTZMODZIbmk2aWRRQT09.

Deros, B. M., Ahmad, R. I., Ghani, J. A., and Yusri, M. (2014). Conformity to Occupational Safety and Health Regulations in Malaysian Small and Medium Enterprises. *Am. J. Appl. Sci.* 11 (3), 499–504. doi:10.3844/ajassp.2014.499.504

Dom, N. M., Kasim, N., Shamsudin, A., and Shamsudin, A. (2012). Factors Influencing Human Resource Planning (HRP) for Local Workforce Supply in Malaysian Construction Industry. *J. Construction Eng. Project Manag.* 2 (1), 5–13. doi:10.6106/jcepm.2012.2.1.005

Fateh, M., Ashraf, M., Affendy Arshad, R., and Hamzah, F. H. (2020). "THE UNFAIR WELFARE PRACTICES TOWARDS FOREIGN WORKERS IN. *IUKL Res. J.* 8 (1), 23–46. Available at: https://iukl.edu.my/rmc/wp-content/uploads/2021/06/B5_IUKLRJ2020_final.pdf.

Fateh, M., Ashraf, M., and Mohammad, M. F. (2021). The Framework of Factors for the Improvement of the Significant Clauses in the Standard Form of Contract for the Ibs Construction Approach in Malaysia. *Int. J. Sust. Construction Eng. Tech.* 12 (1), 164–169. doi:10.30880/ijscet.2021.12.01.016

Fateh, M., Ashraf, M., Mohammad Mohammad, F., and Shukor, A. S. Abd. (2016). Review in Formulating the Standard Form of Contract for Industrialized Building System (IBS) Construction Approach in Malaysia. The 9th International Unimas Stem Engineering Conference (ENCON 2016)

"Innovative Solutions for Engineering and Technology Challenges 87. Kuching, MalaysiaMATEC Web Conf., 7. doi:10.1051/matecconf/20178701001

Fateh, M., Ashraf, M., and Nijar, N. N. (2019). Perspective Analysis on Ibs Provision in Standard Form of Contract in Malaysia. *Malaysian Construction Res. J.* 6 (1), 87–105.

Fateh, M., Ashraf, M., Zakariah, H., and Ezanee, S. E. (2020). Improvement for Significant Clauses in the Standard Form of Contract for Industrialized Building System Construction. In IOP Conference Series: Materials Science and Engineering 713(1). doi:10.1088/1757-899X/713/1/012037

French, J., and Jones, L. (2019). Positive You: a Self-Advocate's Arts-Based Approach for Building Self-Esteem. *Disabil. Soc.* 34 (2), 189–203. doi:10.1080/09687599.2018.1539649

Hamid, A., and Rahim, A. (2013). The Construction Labour Shortage in Johor Bahru, Malaysia. *Int. J. Res. Eng. Tech.* 02 (October), 508–512. doi:10.15623/ijret.2013.0210078

Hammer, L. B., Johnson, R. C., Crain, T. L., Bodner, T., KossekDavis, E. E., DavisKelly, K. D., et al. (2016). Intervention Effects on Safety Compliance and Citizenship Behaviors: Evidence from the Work, Family, and Health Study. *J. Appl. Psychol.* 101 (2), 190–208. doi:10.1037/apl0000047

Hamzah, I. S., Ridhwan Sarifin, M., Shamsinor Abdul Aziz, M., and Fakhrul Azri Abdullah, M. (2020). Malaysia as Attraction of International Foreign Workers. *J. Crit. Rev.* 7 (8), 797–804. doi:10.31838/jcr.07.08.171

Hisyam, K. (2015). A Nation Built by Foreigners, Brick by Brick. *Malaysian Trades Union Congress* 2015. Available at: http://www.mtuc.org.my/a-nation-built-by-foreigners-brick-by-brick/.

Hussain, S., Xuetong, W., and Hussain, T. (2020). Impact of Skilled and Unskilled Labor on Project Performance Using Structural Equation Modeling Approach. *SAGE Open* 10 (March), 215824402091459. doi:10.1177/2158244020914590

Ibrahim, A., Roy, M., Ahmed, Z., and Imtiaz, G. (2010). An Investigation of the Status of the Malaysian Construction Industry. *Benchmarking: Int. J.* 17 (April), 294–308. doi:10.1108/14635771011036357

International Labour Office (2018). *International Labour Organization (ILO)*. Geneva: ILO Publications. doi:10.18356/cfa20042-en

International Labour Organization (2016). *Non-Standard Employment Around the World. ILO Cataloguing in Publication Data*. 1st ed., 44. Geneva: International Labour Organization. Available at: https://www.ilo.org/wcmsp5/groups/public/–-dgreports/–-dcomm/–-publ/documents/publication/wcms_534326.pdf.

Ismail, R., Palliyaguru, R., Karunasena, G., and Othman, N. (2018). Health and Safety (H&S) Challenges Confronted by Foreign Workers in the Malaysian Construction Industry: A Background StudyIn 7th CIOB World Construction Symposium, 277. Sri Lanka: CEYLON INSTITUTE of BUILDERS(CIOB) SRI LANKA. Available at: https://www.researchgate.net/publication/326635913_Health_and_Safety_HS_Challenges_Confronted_by_Foreign_Workers_in_the_Malaysian_Construction_Industry_A_Background_Study.

Jaafar, M., Md Noor, S., and Rasoolimanesh, S. M. (2015). The Effects of a Campaign on Awareness and Participation Among Local Youth at the Lenggong Valley World Heritage Site, Malaysia. *Conservation Manag. Archaeological Sites* 17 (4), 302–314. doi:10.1080/13505033.2016.1175907

Jarmolowicz, W., and Knapińska, M. (2011). Labour Market Theories in Contemporary Economics. *Transformations Business Econ.* 10 (January), 268–280.

Kalagatur, N. K., Nirmal Ghosh, O. S., Sundararaj, N., and Mudili, V. (2018). Antifungal Activity of Chitosan Nanoparticles Encapsulated with Cymbopogon Martinii Essential Oil on Plant Pathogenic Fungi Fusarium Graminearum. *Front. Pharmacol.* 9, 610. doi:10.3389/fphar.2018.00610

Kamal, M., Ernawati, M., and Roger, F. (2012). "Understanding Absorptive Capacity in Malaysian Small and Medium Sized (SME) Construction Companies.". *J. Eng. Des. Tech.* 10 (July), 180–198. doi:10.1108/17260531211241176

Kupets, O. (2016). Education-job Mismatch in Ukraine: Too many People with Tertiary Education or Too many Jobs for Low-Skilled? *J. Comp. Econ.* 44 (1), 125–147. doi:10.1016/j.jce.2015.10.005

Lam, P. T. I., Franco, C., and Fu, S. (2019). "Exploratory Study on Current Status of Startups in the Hong Kong Built Environment Sector.". *J. Manag. Eng.* 35 (4), 5019005. doi:10.1061/(asce)me.1943-5479.0000696

Lingard, H. (2013). Occupational Health and Safety in the Construction Industry. *Construction Manag. Econ.*, 31. (June). doi:10.1080/01446193.2013.816435

Lui, J. W. (2020). *The Impact of Industrialised Building System (IBS) on Local Labour*. Nilai, Malaysia: INTI International University.

Mahmood, S., Awang, Z., Abdel Rahman Kmeil, F., and Jakada, A. H. (2021). Participation of Local Youth Workers in Construction Industry and Strategies for Improvement: A Study in Kelantan and Terengganu. *Int. J. Acad. Res. Business Soc. Sci.* 11 (2), 394–401. doi:10.6007/ijarbss/v11-i2/8225

Malaysian Employers Federation (2014). Practical Guidelines for Employers on the Recruitment, Placement, Employment and Repatriation of Foreign Workers in Malaysia. *Kuala Lumpur*. Available at: http://www.mef.org.my/Att_GeneralNews/MEFReport_PGERPERFWM.pdf.

Manap, N., Mohd Noh, N., and Syahrom, N. (2017). Recruitment Criteria and Attraction Strategies for Local Trained Labour in Malaysia's Construction Industry. In IOP Conference Series: Earth and Environmental Science, 109: 12011. IOP Publishing.

Manoharan, K., Dissanayake, P., Pathirana, C., Deegahawature, D., and Silva, R. (2021). Comparison of Skills between Sri Lankan and Foreign Construction Labour. doi:10.31705/wcs.2021.18

Marhani, M. A., Adnan, H., Einur Baharuddin, H., Reza Esa, M., and Hassan, A. A. (2012). Dependency of Foreign Workers in Malaysian Construction Industry. *Built Environ. J.* 9 (1), 39–50.

Mohamed, R. (2015). Manpower Needs in Construction (Foreign Labour Issues). Pahang: universiti malaysia pahang. Available at: Available at: http://umpir.ump.edu.my/id/eprint/13055/1/FKASA - RAZLINA BINTI MOHAMED - CD 9742.pdf.

Mydin, M. A. O., R Buyung, N., Md Sani, N., and F Abas, N. (2014). "Trends and Reliance on Foreign Labourers in Malaysia: Conventional Construction versus Industrialized Building System Construction.". *SHS Web of Conferences* 11, 1004. doi:10.1051/shsconf/20141101004

Najib, I., Mohd Nordin, R., Ahnuar, E., and Sukor, K. M. (2019). Malaysian as the Component of Labour Force for Construction Industry in Malaysia. *MATEC Web of Conferences* 266 (January), 1007. doi:10.1051/matecconf/201926601007

Mohd Najib, M., Zulkarnain, I., Mohd Nordin, R., and Ismail, Z. (2020). Analyzing the Participation Factors of Local Labour in the Malaysian Construction Industry. *Int. J. Sust. Construction Eng. Tech.* 11 (2), 234–241. doi:10.30880/ijscet.2020.11.02.029

Narayanan, S., and Lai, Y.-W. (2014). Immigrant Labor and Industrial Upgrading in Malaysia. *Asian Pac. Migration J.* 23 (3), 273–297. doi:10.1177/011719681402300302

New Straits Times (2020). MTUC to Home and HR Ministries: Address Foreign Workers Issue Together. *New Straits Times*. Available at: https://www.nst.com.my/news/nation/2020/06/602696/mtuc-home-and-hr-ministries-address-foreign-workers-issue-together.

Noh, N. A., Wahab, H. A., Bakar Ah, S. H. A., and Islam, M. R. (2016). Public Health Services for Foreign Workers in Malaysia. *Soc. Work Public Health* 31 (5), 419–430. doi:10.1080/19371918.2015.1125321

Ofori, G. (2003). Preparing Singapore's Construction Industry for the Knowledge-Based Economy: Practices, Procedures and Performance. *Construction Manag. Econ.* 21 (2), 113–125. doi:10.1080/0144619032000079680

Olanrewaju, A., Tan, S. Y., and Kwan, L. F. (2017). Roles of Communication on Performance of the Construction Sector. *Proced. Eng.* 196, 763–770. doi:10.1016/j.proeng.2017.08.005

Osborne, S., and Hammoud, M. S. (2017). "Effective Employee Engagement in the Workplace.". *Int. J. Appl. Manag. Tech.* 16 (1), 50–67. doi:10.5590/ijamt.2017.16.1.04

Pallant, J. (2016). SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS. McGraw-Hill Education. Available at: https://books.google.com.my/books?id=TSzOjwEACAAJ.6th ed.

Pillai, K. V., Rasiah, R., Williams, G., and Williams, G. (2016). The Impact of Trans Pacific Partnership (TPP) Agreement on US and Malaysian Business' Foreign Labour Practices. *Proced. - Soc. Behav. Sci.* 219, 589–597. doi:10.1016/j.sbspro.2016.05.038

Project Management Institute (2013). *Guide to the Project Management Body of Knowledge (PMBOK® Guide)*. 5th ed. Pennsylvania: Project Management Institute.

Rahim, A., Hamid, A., Nur Inayah Mohd Khazid, R. Y., Hazruwani, A. H., and Amall Raihan, A. R. (2018). The Emerging of Employment Gap in the Malaysian Construction Industry. *J. Phys. Conf. Ser.* 1049(1). doi:10.1088/1742-6596/1049/1/012033

Saieed, Zunaira. (2016). "Malaysia's Skilled Labour Shortage.". *The Star* 2016. https://www.thestar.com.my/business/business-news/2016/11/26/malaysias-skilled-labour-shortage/.

Sekaran, U., and Bougie, R. (2016). *Research Methods for Business: A Skill Building Approach*. Wiley. Available at: https://www.wiley.com/en-al/Research+Methods+For+Business%3A+A+Skill+Building+Approach%2C+7th+Edition-p-9781119266846.

International Labour Organization (2018). Situation and Gap Analysis on Malaysian Legislation, Policies and Programmes, and the ILO Forced Labour Convention and ProtocolUnited States of America. Available at: https://www.ilo.org/wcmsp5/groups/public/–-asia/–-ro-bangkok/documents/genericdocument/wcms_650658.pdf.

Sulaiman, N., and Salleh, K. M. (2016). The Development of Technical and Vocational Education and Training (Tvet) Profiling for Workforce Management in Malaysia: Ensuring the Validity and Reliability of Tvet Data. *Man. India* 96 (9), 2825–2835. Available at: https://www.researchgate.net/publication/309403104_The_development_of_technical_and_vocational_education_and_training_tvet_profiling_for_workforce_management_in_Malaysia_Ensuring_the_validity_and_reliability_of_tvet_data#fullTextFileContent.

Surendran, S. (2021). The State of the Nation: Malaysia in Foreign Worker Quandary. *The Edge Malaysia* 2021. https://www.theedgemarkets.com/article/state-nation-malaysia-foreign-worker-quandary.

Takim, R., and Adnan, H. (2008). Analysis of Effectiveness Measures of Construction Project Success in Malaysia. *Asian Soc. Sci.* 4 (7), 74–91. doi:10.5539/ass.v4n7p74

The Star (2019). Construction Industry Chief: Foreign Workers Remit RM5bil Out of the Country Annually. *Nation*. Available at: https://www.thestar.com.my/news/nation/2019/09/21/construction-industry-chief-foreign-workers-remit-rm5bil-out-of-the-country-annually.

The Star (2009). Sustainable Manpower and Skills Needed for Nation to Grow. *Business*. Available at: https://www.thestar.com.my/business/business-news/2009/02/03/sustainable-manpower-and-skills-needed-for-nation-to-grow.

Trevena, P. (2013). *Mobility in Transition*. Cambridge: Amsterdam University Press, 169–190. doi:10.1017/9789048515493.009Why Do Highly Educated Migrants Go for Low-Skilled Jobs?: A Case Study of Polish Graduates Working in London

UNESCO. (2021). "Technical Vocational and Education Training (TVET) 1. Sub-Education Policy Review Report: Technical and Vocational and Training (TVET). Jakarta. Available at: https://en.unesco.org/sites/default/files/tvet_final_-_january_2021.pdf."

Wahab, A. B., and F Lawal, A. (2011). "An Evaluation of Waste Control Measures in Construction Industry in Nigeria." *Afr. J. Environ. Sci. Tech.* 5 (3), 246–254.

Zaki, S., Mohamad, S., and Yusof, Z. (2012). Construction Skilled Labour Shortage – the Challenges in Malaysian Construction Sector.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Built Environment

ORIGINAL RESEARCH
published: 30 May 2022
doi: 10.3389/fbuil.2022.912326

Analysis of the Current Maintenance Management Process in School Buildings: Study Area of Primorje-Gorski Kotar County, Republic of Croatia

Ksenija Tijanić Štrok[1]*, *Saša Marenjak*[2] *and Diana Car-Pušić*[1]

[1]*Faculty of Civil Engineering, University of Rijeka, Rijeka, Croatia,* [2]*Faculty of Civil Engineering and Architecture Osijek, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia*

OPEN ACCESS

Edited by:
Ivan Zambon,
Other, Austria

Reviewed by:
Valentino Sangiorgio,
University of Studies G d'Annunzio Chieti and Pescara, Italy
Vít Hromádka,
Brno University of Technology, Czechia

***Correspondence:**
Ksenija Tijanić Štrok
ksenija.tijanic@uniri.hr

Specialty section:
This article was submitted to Construction Management, a section of the journal Frontiers in Built Environment

Received: *04 April 2022*
Accepted: *02 May 2022*
Published: *30 May 2022*

Citation:
Tijanić Štrok K, Marenjak S and Car-Pušić D (2022) Analysis of the Current Maintenance Management Process in School Buildings: Study Area of Primorje-Gorski Kotar County, Republic of Croatia. Front. Built Environ. 8:912326. doi: 10.3389/fbuil.2022.912326

Due to their educational purpose, impact on users and quality of education, school buildings are vital institutions in the community, and their adequate maintenance should be mandatory. Studies conducted around the world point to specific inadequacies that occur in the school maintenance segment, however studies that examine the entire maintenance management process in detail are rare. With such studies, it is possible to observe all components of maintenance and its management in the observed buildings, it is possible to get a detailed insight into the present inadequacies and problems, and give the best possible answer to them. Observing the territory of the Republic of Croatia, it was noticed that there are no adequate databases on school building maintenance. Given all the above, this paper aims to establish a methodological framework and analyze the entire maintenance management process in Croatian schools to detect potential inadequacies, reveal problems, and identify areas for action to improve maintenance performance and carry out maintenance more efficiently. The current maintenance management process of elementary school buildings in the Primorje-Gorski Kotar County, Republic of Croatia, was examined based on the developed methodological framework. Data on the management process were collected through a questionnaire survey among school principals. The questionnaire was compiled based on an extensive literature review. Maintenance performance through management functions was investigated, namely: maintenance planning, maintenance organization, maintenance directing, and maintenance controlling. Important management factors within each of the mentioned functions were observed. The data analysis revealed specific inadequacies and problems in the maintenance process, and it was shown that maintenance management is not entirely effective. Recommendations were given to improve the current situation. In this way, it is possible to quickly and easily examine the entire management process and identify measures that need to be introduced in schools to improve their business in the maintenance segment. The developed methodological framework can be applied to schools from other areas as well as to other types of buildings.

Keywords: maintenance, maintenance management, school buildings, elementary schools, Primorje-Gorski Kotar County, current state

1 INTRODUCTION

Concerning their educational purpose, school buildings are essential facilities in the community (Katić et al., 2021). Their function is to protect users and equipment from natural influences and provide internal space for learning and teaching (Palis and Misnan, 2018).

Education is crucial for human development, facilitates a healthy mindset and develops cognitive abilities (Sharma, 2020). Education through the development of individuals also affects society as a whole and is a direct driver of its economic, social, political (and all other) growth and development (Tijanić Štrok, 2021). Quality of education depends significantly on the state of built assets in which it is provided (Xaba, 2012; Yong and Silieman, 2015; Teixeira et al., 2017). For school buildings to be of high quality and to perform their required function, it is necessary to implement their maintenance continuously.

Building maintenance can be defined as a set of all activities undertaken to preserve, protect and improve buildings (following the standards in force) to serve the intended functions during the building exploitation phase. Maintenance management is a practice of Facility Management (FM) that establishes strategies, goals, maintenance responsibilities, ways of performing maintenance, with their implementation through functions of planning, organizing, directing and controlling (Tijanić Štrok, 2021).

No comprehensive research on school buildings maintenance has been conducted so far in the Republic of Croatia. There are no adequate maintenance databases, no records of the state of the maintenance management process, and there is no effective framework within which maintenance work would be performed, so maintenance is often neglected, disorganized, expensive, and does not provide value for money (Tijanić Štrok, 2021).

According to the conducted studies, inadequate state of schools is a pattern that appears everywhere in the world (Xaba, 2012; Albader and Kandil, 2013; ElSamadony et al., 2013; Izobo-Martins, 2014; Ropi and Tabasi, 2014; Vieira and Cardoso, 2014; Dickerson and Ackerman, 2016; Wuni et al., 2018). Studies that approach maintenance by observing the overall school maintenance management process are rare.

Considering the stated, this paper aims to establish a methodological framework for the analysis of the overall school maintenance management process. The framework will quickly and easily detect potential inadequacies and problems and identify the areas to be affected to improve the maintenance performance, i.e., to make the school maintenance more efficient.

Along with the Introduction, the paper is organized as follows: Chapter 2 provides a detailed literature review within which adequate maintenance management factors for public (educational) buildings have been identified; Chapter 3 presents the developed research methodological framework and describes the study area as well as the data collection method and the characteristics of the collected data; in Chapter 4, an analysis of the data collected and a discussion of the results was conducted; Chapter 5 provides concluding remarks and recommendations.

2 STATE OF THE ART

2.1 Previous Research on the School Buildings Maintenance

The school maintenance program is an organizational activity carried out by the school community to extend the lifespan of the school building with all its parts. School maintenance refers to the repair, replacement, and general maintenance of physical characteristics found in school buildings, terrain and safety systems (Izobo-Martins, 2014).

School buildings are specific and fail rapidly due to their age, overcapacity, extensive use (ElSamadony et al., 2013; Saraiva et al., 2018). A great challenge is the various building components with different repair requirements (ElSamadony et al., 2013). Schools are often located in buildings whose original purpose was not educational, and a number of them have been declared a cultural asset which makes maintenance even more difficult due to the special requirements to be met (Tijanić Štrok, 2021).

Different research on school buildings maintenance was conducted in Asia, primarily Malaysia. It was found that there is no adequate maintenance knowledge base; the attitude toward maintenance is mostly negative; the maintenance workforce is unqualified (Nah et al., 2012); the maintenance is neglected in the design phase (Ali et al., 2013); the planned maintenance is not applied; the funding is not sufficient to meet all the maintenance needs (Ropi and Tabasi, 2014); the users are dissatisfied with the buildings in which they are educated or work (Yong and Sulieman, 2015). In other countries, research outcomes show similar results.

Liu et al. (2021) point out that proper maintenance was neglected in schools in Indonesia. Izobo-Martins, (2014) showed in Nigeria that in secondary schools, users are not satisfied with the building's maintenance and that there is no maintenance policy that would define the procedure and method of maintenance implementation of the aforementioned institutions. Albader and Kandil, (2013) (USA) develop a framework for improving the overall school buildings state that will improve the inefficient use of maintenance funds and improve students' performance. Dickerson and Ackerman (2016) point out that a quarter of schools in the United States are in a weaker operational state and that they require major repairs. The inadequate condition also have school buildings in Egypt where Marzouk and Award (2016) developed a model for assessing the state of schools based on the Analytic Hierarchy Process (AHP). A survey conducted by Xaba (2012) in South Africa showed that school maintenance mostly comes down to emergency and corrective interventions. Wuni et al. (2018) research the school maintenance in the Ghana area and establish that the causes of inadequate management are the lack of professional maintenance managers, postponement of maintenance, budgetary constraints, deviations in planning, non-connection of the design phase with later project phases.

The public school building's funding is usually from the state budget, so it is limited. Insufficient financial resources and their

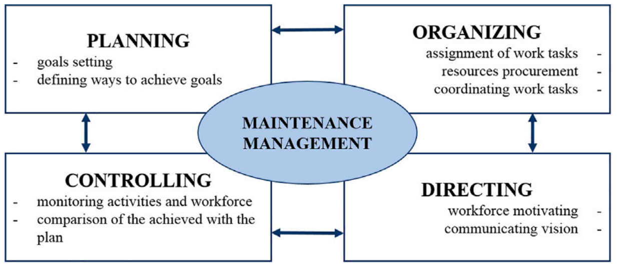


FIGURE 1 | Maintenance management model (adapted from Ogunoh et al., 2018).

inefficient spending are facts that flow through numerous research therefore certain researchers have developed cost estimation models based on methods such as regression analysis, neural networks, etc., to contribute to solving this problem (ElSamadony et al., 2013; Lee and Jeon, 2017; Kim et al., 2018).

Systematic research on school buildings maintenance in the Republic of Croatia has not been conducted so far. There are sporadic studies and literature reports that point to certain inadequacies in the current management process. According to the State Audit Office, (2018), Croatian schools operate in outdated facilities, as much as 25% of all schools were built more than 80 years ago, resulting in neglect and damage due to wear-out and inadequate maintenance. The lack of regular maintenance is mainly attributed to insufficient budget allocations from the public administration (Tijanić Štrok, 2021).

In some schools, health and ecologically unacceptable materials such as asbestos are still installed. School founders do not keep records of the condition and equipment of the schools they manage (State Audit Office, 2018; Tijanić Štrok, 2021). The Government of the Republic of Croatia (2012) points out that there is a great need at the level of the Republic of Croatia to invest in public infrastructure, including schools.

The research on health and environmental conditions in schools was also carried out by Capak et al. (2015) and found that there are problems with the moisture and mold in schools, as well as the state of sanitary nodes. According to this research, weaker construction and technical condition of the observed school buildings need to be solved by rehabilitation and reconstruction based on sustainable architectural solutions.

2.1.1 School Buildings Maintenance Management Factors

Building maintenance management implies integrating four primary management functions: maintenance planning, maintenance organization, maintenance directing, and maintenance controlling. Functions are consecutive but interconnect and sometimes overlap (Olanrewaju and Abdul-Aziz, 2014). The building maintenance management process model is shown in **Figure 1**.

Management activities (factors) within prominent functions are the main part of this model since their implementation fulfills the conditions for creating an adequate management process. It is important to identify all those factors that are appropriate for individual types of buildings to develop an adjusted, robust and comprehensive framework for their effective management.

Effective school maintenance management, according to Tijanić Štrok (2021), means proactive and high-quality implementation of maintenance activities while minimizing resource consumption, primarily maintenance costs, avoiding work interruptions, and increasing staff and student satisfaction by creating conditions that ensure their health and safety, facilitating teaching, learning and improving academic results. The school maintenance management factors should support the fulfillment of precisely these criteria.

To recognize the aforementioned factors, a detailed literature analysis was conducted and **Table 1** was created, summarizing the attitudes of the relevant newer researchers on the factors necessary for implementation into the maintenance management process. In order to obtain a more extensive overview of factors, research dealing with various types of public buildings has been elaborated, while **Table 1** includes those which, according to the authors of this paper, have a direct impact on school buildings. The implementation of these factors contributes to more efficient maintenance management. Factors are divided within the previously mentioned management functions.

According to Olanrewaju and Abdul-Aziz, (2014), recognizing and fulfilling users' expectations is an important part of effective maintenance management of educational institutions, especially since users' needs are becoming increasingly demanding. Efficient learning and teaching can only take place in a favorable environment for users therefore within the system of performance indicators, it is necessary to examine how satisfied the users are with this environment (Tijanić Štrok, 2021).

During the maintenance process, considering a large number of management factors, there is a huge amount of different information. For this reason, light is thrown on the characteristics of computer information systems and the way they handle information. Computer support has recently been recognized as an irreplaceable and indispensable factor for more efficient maintenance of buildings (Alzaben, 2015; Shohet and Nobili, 2017; Wong et al., 2018). The software can include management functions in one place as well as all stakeholders participating in the process and serves as maintenance data storage. The Computerized Maintenance Management System (CMMS) creates a decision-making platform and is a kind of maintenance management guide. CMMS application in schools is beneficial primarily because of the highlighted maintenance problems and the complexity of school buildings, the amount of data generated in them, limited resources, the number of users and their specific needs (Tijanić Štrok, 2021).

In order to create the basis for effective school maintenance management, some authors point out that it is important to consider decisions made before the phase of facility use, primarily the design phase is considered (Ali et al., 2013; Jandali and Sweis, 2017; Ogunoh et al., 2018). It can often be established that buildings are expensive for maintenance due to inappropriate priorities during project preparation. At this stage, the right decisions and a proper plan can save a lot,

TABLE 1 | Factors for school buildings effective maintenance management.

Function	Management Factor	Researchers Who Considered Listed Factors Important for an Efficient Maintenance Management Process of Public (School) Buildings
Planning	Defined maintenance policy	Ochien'g, (2012), Xaba, (2012), Izobo-Martins, (2014), Olanrewaju and Abdul-Aziz, (2014), Velmurugan and Dhingra, (2014), Vieira and Cardoso, (2014), Alzaben, (2015), Jandali and Sweis, (2017), Mong et al. (2018)
	Defined maintenance standard	Olanrewaju and Abdul-Aziz, (2014), Jandali and Sweis, (2017)
	Maintenance included in the mission and vision of the institution	Olanrewaju and Abdul-Aziz, (2014), Velmurugan and Dhingra, (2014)
	Defined maintenance goals	Harun et al. (2013), Olanrewaju and Abdul-Aziz, (2014), Vieira and Cardoso, (2014), Mong et al. (2018)
	Defined maintenance priorities	Nik-Mat et al. (2011), Wing et al. (2016)
	Maintenance plan developed	Ochien'g, (2012), Xaba, (2012), Harun et al. (2013), Izobo-Martins (2014), Velmurugan and Dhingra, (2014), Vieira and Cardoso, (2014), Mong et al. (2018), Ogunoh et al. (2018)
Organizing	Developed organizational structure of the institution	Xaba, (2012), Olanrewaju and Abdul-Aziz, (2014), Alzaben, (2015), Mong et al. (2018), Ogunoh et al. (2018)
	Defined lines of communication within the structure	Olanrewaju and Abdul-Aziz, (2014), Alzaben, (2015), Mong et al. (2018)
	Defined maintenance inventory management procedure	Harun et al. (2013), Olanrewaju and Abdul-Aziz, (2014), Velmurugan and Dhingra, (2014), Alzaben, (2015), Jandali and Sweis, (2017), Mong et al. (2018)
	Defined procedure for resolving maintenance requests	Olanrewaju and Abdul-Aziz, (2014), Alzaben, (2015), Alshehri, (2016), de Morais and Lordsleem Júnior, (2019)
	Defined procedure and method of documentation on maintenance	Vieira and Cardoso, (2014), Alzaben, (2015), Alshehri, (2016), Jandeli and Sweis, (2017), de Morais and Lordsleem Júnior, (2019)
Directing	Conducting staff training	Harun et al. (2013), Izobo-Martins, (2014), Olanrewaju and Abdul-Aziz, (2014), Alzaben, (2015), Aliyu et al. (2016), Alshehri, (2016), Jandali and Sweis, (2017), Mong et al. (2018)
	Staff motivation	Harun et al. (2013), Olanrewaju and Abdul-Aziz, (2014), Alzaben, (2015)
	Encouraging teamwork	Olanrewaju and Abdul-Aziz, (2014), Alzaben, (2015)
	Good communication	Harun et al. (2013), Olanrewaju and Abdul-Aziz, (2014)
Controlling	Carrying out maintenance supervision	Ochien'g, (2012), Harun et al. (2013), Izobo-Martins, (2014), Olanrewaju and Abdul-Aziz, (2014), Alzaben, (2015), Mong et al. (2018)
	Carrying out quality control over maintenance	Izobo-Martins, (2014), Alzaben, (2015), Alshehri, (2016), Jandali and Sweis, (2017), Mong et al. (2018)
	Measuring maintenance performance with performance indicators	Harun et al. (2013), El Shorafa, (2013), Olanrewaju and Abdul-Aziz, (2014), Velmurugan and Dhingra, (2014), Alzaben, (2015), Shohet and Nobili, (2017), Mong et al. (2018), de Morais and Lordsleem Júnior, (2019)

so maintenance experts should be consulted from the earliest project stages (Aliya et al., 2016). At the design stage, maintenance costs should already be planned in the form of a life costs analysis when different project variants can be developed and the most cost-effective ones can be selected (Tijanić Štrok, 2021). Buildings should be designed/built so that their maintenance requires minimum costs and time (Olanrewaju and Abdul-Aziz, 2014).

For all recognized management factors, their functioning in school buildings' current maintenance management process will be examined.

3 METHODOLOGY

After systematic consideration of the authors on the research problem, a methodological framework was developed based on which this research is conducted.

The research was conducted through several segments, which are in a meaningful way divided within the chapters of this paper. The methodological framework shows how the current maintenance management process in school buildings will be analyzed, as shown in **Figure 2**.

The research consists of two distinctive fields, theoretical examination and analysis of literature, and data collection and processing for the selected study area.

Within theoretical consideration, it was observed that problems in their maintenance are repeatedly present in school buildings. No efforts were made to create broader databases on maintenance inadequacies, which would provide valuable data to help identify the necessary measures to improve the current situation.

After the research aim (stated in the Introduction) has been defined, a detailed literature analysis has been conducted, identifying all important segments and functions within the school maintenance management process and identifying factors that can help create a more efficient management process.

The following hypotheses of research have been created:

- with the help of a defined methodological framework, it is possible to analyze the current maintenance management process in school buildings;
- current school maintenance management is not fully effective, i.e., there are inadequacies in maintenance management that can be affected through the introduction of adequate improvement measures.

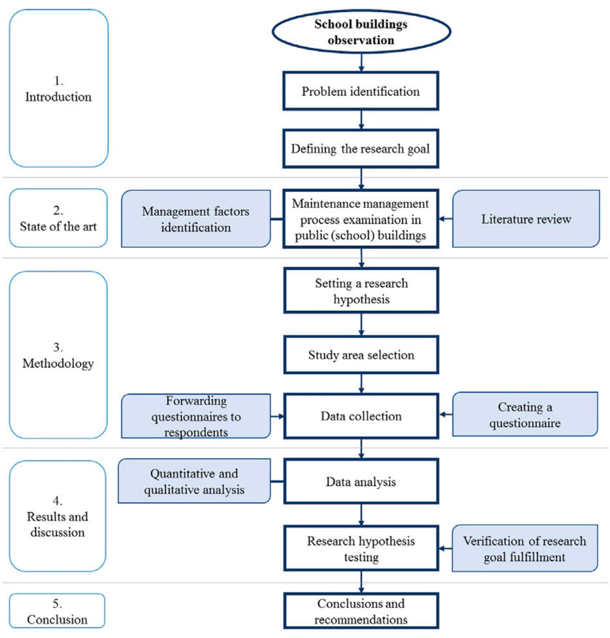


FIGURE 2 | Research methodological framework.

Hypotheses will be examined in the case study area covering the territory of Primorje-Gorski Kotar County in the Republic of Croatia. The study is limited to public elementary schools in the mentioned area.

The case study area will provide a comprehensive analysis of the school maintenance management process in the selected area. The aim is to develop a methodological framework and theories on a smaller data sample that will be applicable for the entire school population as well as with certain modifications for other types of facilities. The sample size in the research is not all that determines the scientific strength, but a consistent and systematic approach and suitability for purpose are also important.

Databases will be collected within the observed area using the questionnaire survey. The questionnaire was recognized as an adequate method of data collection due to the size of the study area and the number of schools in it, the amount of data needed and the duration of the study.

Once the data have been collected, they will be processed mainly by quantitative methods (statistical analysis), after which we will get an insight into the results, which will enable verification of the authenticity of the hypotheses set.

The last step in the research is to draw conclusions with certain recommendations.

3.1 Case Study Area

The area covered by the study is located in the western part of the Republic of Croatia and covers the area of Primorje-Gorski Kotar County. Primorje-Gorski Kotar County covers 3,587 km^2 of populated area, where about 6.9% of the population of Croatia resides with an average population density of 83 inhabitants/km^2. Its seat is Rijeka, Croatia's third-largest city. The County comprises 14 towns and 22 municipalities and 510 settlements in towns and municipalities (Primorje- Gorski Kotar County, 2020).

Out of 21 areas in the Republic of Croatia, the Primorje-Gorski Kotar County is the fourth in the number of elementary schools. Fifty-seven public elementary schools are operating by County area. The largest number of schools, thirty-one of them, is managed by Primorje-Gorski Kotar County, the City of Rijeka runs twenty-three schools, two schools are run by the town of Crikvenica, while the town of Opatija is the founder of one school.

Primorje-Gorski Kotar County was selected as an area of research due to the availability of data, development, size and sufficient number of elementary schools to create a representative data sample.

3.2 Survey Questionnaire

A survey was conducted among elementary school principals. The questionnaire consisted of three parts in which three data groups were requested.

The first group of data consists of general data on school buildings collected in order to obtain knowledge on the basic sample characteristics. Data on construction year, last renovation year, the total building inner surface, floors, number of students, departments, employees, shifts, the operational state of the school building; are requested.

The second part of the questionnaire examines the state of the current maintenance management process in school buildings. Questions were raised about the establishment of recognized management factors (**Table 1**) in the current school maintenance management process. The factors in this paper imply activities (criteria or subfunctions) within maintenance management functions that must be implemented for the whole maintenance process to function effectively. The questionnaire in this part of the survey is in the form of a 3-degree Likert scale, where the questions asked have answers in forms: no (1)—partially (2)—yes (3).

The third questionnaire part consists of the principal's views. In this part, the principals could leave their comments or point out, according to them, the biggest problems in the current maintenance management process.

The questionnaire is anonymous. To improve the questionnaire design and response rate, questions are as simple and shorter as possible. Where possible within the questionnaire, Cronbach's alpha test was used to measure reliability. The usual rule for surveys conducted in questionnaires form is that the results with Cronbach's alpha coefficient above 0.700 are considered acceptable (Kušljić, 2012). The questionnaire was changed several times, while in cooperation with two construction experts it was not estimated that a sufficiently valid and reliable version was obtained.

Questionnaires were sent to the examinees via e-mail. They were forwarded to a total of 57 addresses and 31 respondents replied to the questionnaire, representing a return rate of 54.38%. Taking into account that certain surveys on similar issues were carried out on samples of a smaller number and lower rates of return (Oyenuga et al., 2012; Xaba, 2012; ElSamadony et al., 2013; El Shorafa, 2013; Olanrewaju and Abdul-Aziz, 2014), the

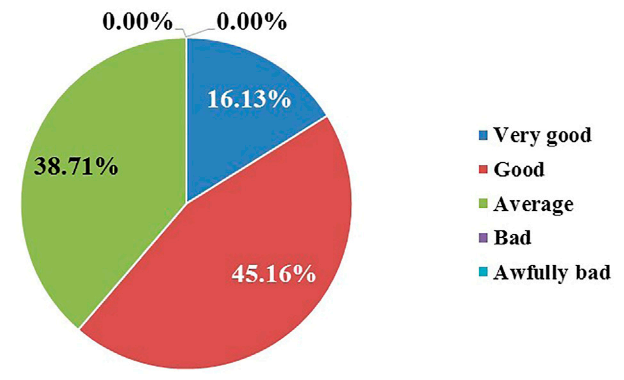


FIGURE 3 | School buildings condition.

questionnaire return rate is considered acceptable for the continuation of this study.

4 RESULTS AND DISCUSSION

Since data were collected for 31 schools (>30), a normal distribution is presumed for the data sample (Kušljić, 2012).

The collected data are statistically processed in the Microsoft Excel software and the results are presented below.

Descriptive data statistics on the general school characteristics are presented in **Table 2**.

The observed elementary schools are 69 years old on average. The oldest school was built in 1876 and the youngest in 2005. Most of the observed schools had a certain type of reconstruction or upgrade during their existence, and on average 18 years have elapsed since such greater intervention, compared to 2022. The school inner surface is around 2,600.00 m^2 with an average of 2.3 floors. The smallest facility in the sample has a surface area of 554.00 m^2, while the largest surface area is almost 6,000.00 m^2. On average, looking at the sample of schools, the number of students in them is 245.55. They are deployed in 14.48 departments. There are 44 employees in schools, taking into account non-teaching personnel. Schools mostly work in one shift. These data refer to the school year 2020/2021.

Within the first part of the questionnaire, principals assessed the operational (physical) state of schools considering the issue of damage to the building. Within the literature review, it has been established that there is no record of damage and condition of school buildings (Tijanić Štrok, 2021), therefore this situation has yet to be established. Knowledge about the condition and damages of school buildings is valuable data for creating a priorities system for the school buildings maintenance. The principals evaluated the status of their school with scores from 1 to 5, where grade 1 signifies awfully bad condition and 5 marks a very good condition. The results are shown in **Figure 3**.

According to principals, the largest number of schools, 14 of them, have a good (4) operational condition. There are 12 schools with average (3) condition (neither good nor bad). No principal rated the state of his school as bad (2) or awfully bad (1). In future research, it is recommended to examine the state of schools of all important elements of the structure to obtain a more detailed perception of the problem of damage on buildings. It is important to ensure that all students have equal conditions for work and study.

In the second part of the questionnaire, principals were directly asked whether, in their opinion, the current overall maintenance management process is efficient, positive response "yes" was given by 8, i.e., 25.81% respectively, while all others consider that it is not (22.58%) or that it is only partially effective (51.61%).

In order to determine efficiency within management functions, the principal's cognition of the functioning of certain management factors within the functions of planning, organizing, directing, and controlling maintenance, was examined. Responses within the questionnaire refer to whether these factors are implemented within the current school maintenance management process. The reliability control of this part of the questionnaire was performed, and Cronbach's alpha coefficient of 0.8654 was obtained, which is a satisfactory result. The questionnaire survey results are given in **Table 3**.

Given the scale used, it was determined that in order for the implementation of a certain function or factor in the school sample to be considered sufficiently satisfactory, the mean value of the data should be at least 2.50.

From **Table 3**, it can be seen that only a few factors have been implemented at a satisfactory level. The most positive responses were given to the factor concerning defined communication lines between maintenance personnel, 70.97%. A high proportion of the answers in part and extremely negative answers are concerns. A very high percentage of the answer "no" has a factor that concerns the education of the principal, as much as 80.65%. This trend is risky because all maintenance requests should have a starting point in the school principal. The principal is involved in building management and administrative issues. The principal's responsibility as head of the school is to initiate the procedure for the maintenance of the school building and to inform the founder of all problems (Tijanić Štrok, 2021). The principals often do not have any technical knowledge, they come from among the teachers, and therefore certain education would undoubtedly be helpful and compulsory.

The factor that also has a high share of negative responses concerns the measurement of maintenance performance by performance indicators. Performance indicators are used to monitor the performance of business processes and planned objectives, in this case, the maintenance management process, which is one of the fundamental activities of operations within public institutions. Implementation of performance indicators is very important because, in this way, an analysis of achievements in the field of maintenance can be performed. By reviewing achievements, adequate changes and improvements can be made in future planning, decision making, organization of works, etc. This process is the basis for systematic and continuous improvement of the overall school maintenance management process (Tijanić Štrok, 2021). It is particularly important to measure user satisfaction since they are the ones who spend most of their time in school buildings, and buildings

TABLE 2 | Descriptive statistics of general data on schools.

Examined Variables	Arithmetic Mean	Standard Deviation	Min Value	Max Value
Age	68.7419	34.5639	17	146
Years passed since the last renovation	18.4583	17.1235	2	50
Inner surface (m^2)	2,584.0968	1,404.9039	554.00	5,990.00
Number of floors	2.3226	0.7911	1	4
Number of students	245.5484	178.0056	19	765
Number of class departments	14.4839	7.7239	5	39
Number of employees	44	17.5499	19	87
Number of shifts	1.2258	0.4250	1	2

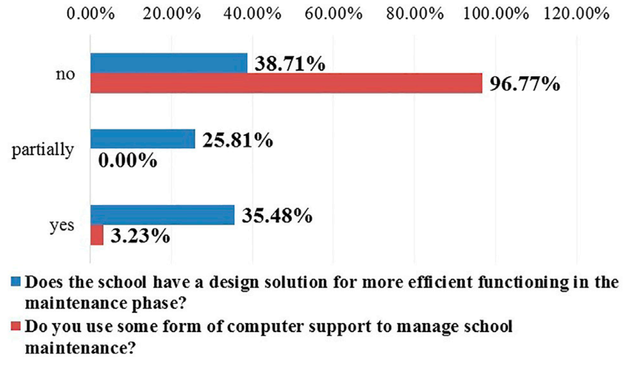


FIGURE 4 | Shares of response (assessment) of the condition of school buildings.

are there to meet their needs and create an environment suitable for learning and teaching. According to the results obtained from **Table 3**, measuring the satisfaction of users (students and staff) is a practice that is not common.

The maintenance policy and standard also received relatively few very positive answers. The maintenance policy determines the maintenance procedures and the way of making maintenance decisions. The policy helps to identify maintenance activities to be applied in a particular situation, therefore, it is one of the main aspects of successful building maintenance management. School institutions should have a defined standard (level) at which their buildings and individual building components should be maintained. Schools should implement a defined standard and should be guided by it in all other maintenance activities (Tijanić Štrok, 2021).

Looking at the results of maintenance performed according to maintenance management factors given in **Table 3**, it can be seen that there is room for improvement in all maintenance management functions and all management factors.

The questionnaire also examined the issue of school design solutions and CMMS application, and the results are shown in **Figure 4**.

According to the graph, the negative response prevails in terms of the school's design solution, which was done without considering the future of the building and its maintenance. Given the years of construction of the buildings in the sample, it is not surprising that most lack an adequate project solution.

What is very problematic nowadays is that almost no form of computer support is used to manage school maintenance. This factor, although a particularly important aspect of efficient management according to literature, received the most negative answers, 96.77% of them. This result shows that the greatest changes and improvements should go toward applying newer technologies and computerizing the entire schoolwork, together with maintenance management.

The connection with the variable on the current effectiveness of the overall school maintenance management process was examined for all collected data on the current implementation (state) of management factors. The aim was to examine, for example, how much and what the connection is between the current implementation of CMMS in schoolwork and the current effectiveness of the overall maintenance management process. Quantitative data from the first questionnaire part and the principal's assessments of the implementation of individual factors from the second questionnaire part, were used in the analysis. The degree of correlation was examined in the form of the Spearman correlation coefficient since most data are presented in the ordinal form.

When interpreting the correlation coefficient, the strength of the correlation between variables is interpreted as shown in **Table 4**.

To correctly interpret the correlation coefficient, it is necessary to calculate the significance of the correlation coefficient in the form of the p-value. In this case, the p-value was determined at a level of less than $\alpha = 0.05$, assuming that there is only a 5% probability that the results from the sample considered occurred accidentally. The null hypothesis is that there is no statistically significant relationship between the current efficiency of the maintenance management process and the current state of observed management factors. The results indicate that: if the p-value is $\leq \alpha$, the correlation is statistically significant, the null hypothesis is rejected; if the p-value is $> \alpha$, the correlation is not statistically significant, the null hypothesis is not ruled out. The results of the correlation analysis are presented in **Table 5**.

The results show a certain correlation between almost all observed variables and the current effectiveness of the overall maintenance management process. Given the amounts of p-values, in most cases, the correlation coefficients must not be interpreted, the null hypotheses cannot be rejected, the results

TABLE 3 | Are individual maintenance management factors implemented in the current school maintenance management process?

Function	Management Factor	Answer (%)			Mean	St. Dev
		No (1)	Partially (2)	Yes (3)		
Planning	Defined maintenance policy	41.94	45.16	12.90	1.7097	0.6925
	Defined maintenance standard	38.71	45.16	16.14	1.7742	0.7169
	Maintenance included in the mission and vision of the institution	12.90	32.26	54.84	2.4193	0.7199
	Defined maintenance goals	19.35	38.71	41.94	2.2258	0.7620
	Defined priorities and maintenance plans	12.90	35.48	51.61	2.3871	0.7154
	Average	25.16	39.35	35.48	2.1032	0.7213
Organizing	Developed organizational structure of the institution	25.81	35.48	38.71	2.1290	0.8059
	Defined lines of communication within the structure	6.45	22.58	70.97	2.6452	0.6082
	Defined maintenance inventory management procedure	12.90	35.48	51.61	2.3870	0.7154
	Defined procedure for resolving maintenance requests	19.35	35.48	45.16	2.2581	0.7732
	Defined procedure and method of documentation on maintenance	12.90	32.26	51.61	2.3871	0.7154
	Average	15.48	32.90	51.61	2.3613	0.7236
Directing	Conducting training of technical staff	25.81	35.48	38.71	2.1290	0.8059
	Conducting trainings for principals	80.65	12.90	6.45	1.2580	0.5755
	Providing guidance and support (motivation, teamwork, good communication)	9.68	25.81	64.52	2.5483	0.6752
	Average	38.71	24.73	36.56	1.9784	0.6855
Controlling	Carrying out supervision and quality control over maintenance	3.23	32.26	64.52	2.6129	0.5584
	Measuring maintenance performance with performance indicators	51.61	41.94	6.45	1.5483	0.6238
	Measuring student and staff satisfaction	32.26	41.94	25.81	1.9355	0.7718
	Average	29.03	38.71	32.26	2.0322	0.6513

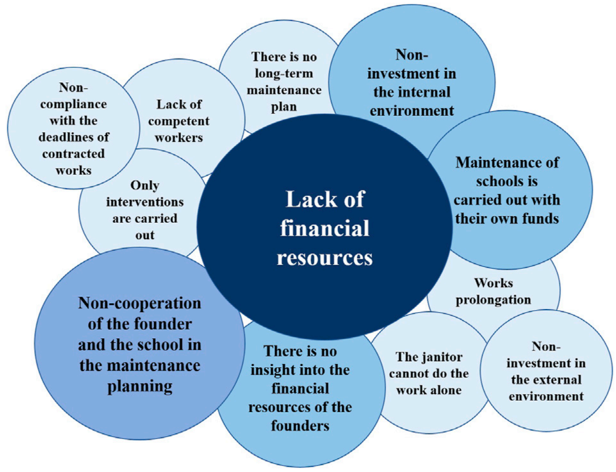


FIGURE 5 | Highlighted problems in the current schools' maintenance management process according to principals.

are not valid for the entire school population, and they should be examined on a larger data sample.

Correlation coefficients that have high significance are those between the current efficiency of the overall maintenance management process and the overall state of the school, design solution, defined maintenance standard, performance measurement with performance indicators. The most significant connection has proved to be with a current state of design solution therefore, it is recommended that during the renovation of old and construction of new school buildings, a maintenance expert should be included in the design phase, thus achieving savings in later stages of building maintenance and use and achieving more efficient maintenance management overall (Tijanić Štrok, 2021).

In the third part of the questionnaire, principals highlighted the current maintenance problems presented in **Figure 5**.

The largest number of principals stressed that adequate maintenance lacks financial resources. Therefore, the funds available must be spent more efficiently and purposefully, which can be achieved by proper planning, increasing the number of preventive works, better organization, supervision and control, etc. The problem highlighted by several principals was the lack of cooperation between the school founders and the principals in planning future maintenance, and the schools also are not familiar with the funds founders intend to invest. The persons in charge of school maintenance should work together, on the team, but everyone should have a role that contributes to the achievement of the goals set, and the founder and principal should be the ones who encourage work and cooperation (Tijanić Štrok, 2021).

To improve the current state of the school maintenance management process overall, according to Tijanić Štrok, (2021), it is recommended that a new maintenance management model be implemented in schools. This model is a kind of graphic and descriptive maintenance manual that directs school leaders and all other stakeholders on how to achieve the effective performance of maintenance management activities. In particular, the use of a CMMS is emphasized within the model. The model provides instructions and guidelines for actions within the planning, organization, directing, and controlling of maintenance activities in school buildings, as well as within other phases of the construction project.

TABLE 4 | Correlation coefficient values and strength of correlation between variables (adapted from Meghanathan, 2016).

Range of Correlation Coefficient Values	Level of Correlation
0.80 to 1.00, –1.00 to –0.80	very strong positive/negative
0.60 to 0.79, –0.79 to –0.60	strong positive/negative
0.40 to 0.59, –0.59 to –0.40	moderate positive/negative
0.20 to 0.39, –0.39 to –0.20	weak positive/negative
0.00 to 0.19, –0.19 to –0.01	very weak positive/negative

After the results are presented, the fulfillment of the research goal will be checked and the hypotheses will be tested.

The research established a methodological framework for analyzing the current school maintenance management process, thus achieving the primary objective of this paper and thus confirming the first research hypothesis. The analysis of the collected data showed that maintenance management of studied schools is not fully efficient, that is, there are certain inadequacies and problems identified within this chapter of the paper, thus proving the second research hypothesis. All identified inadequacies can be affected by certain measures, which need to be systematically considered and decisions made that will improve the current situation. One of the essential measures is undoubtedly the change of the overall business by introducing a new maintenance management model.

The following can be emphasized if a parallel is drawn between the theoretical background of the observed maintenance problem and the research conducted here. In the Republic of Croatia, there is no systematic research on the maintenance and condition of school buildings, and on the world stage, there are no studies that approach the schools' maintenance in a way that examines the functioning of the entire maintenance management process. The purpose of this paper was to close the gaps in the literature, suggest an approach, and conduct research that can be applied to larger areas and other buildings to identify shortcomings in the maintenance process and its management. Main problems in world research were also confirmed in a study area in which: older school buildings that are inefficiently maintained are in operation; maintenance struggles with a limited budget; a clear policy and standard of maintenance is lacking; sufficient attention is not paid to maintenance when designing buildings which according to the results of this research is a crucial factor for effective maintenance management. In addition to these inadequacies, shortcomings in measuring maintenance performance with performance indicators as well as measuring user satisfaction, and the lack of training for school principals, were also shown. The backwardness of Croatian schools is particularly evident in applying computer systems for maintenance management.

Both the literature review and this research have shown that additional efforts need to be made to improve the overall school maintenance process to create an environment in which students can reach their full potential. Measures proposal to overcome the identified deficiencies in maintenance management is given in the next chapter.

TABLE 5 | Correlation coefficients and *p*-values between the observed maintenance management factors.

Current Situation in Schools	Efficiency of Maintenance Management Process	
	Coefficient	*p*-value (<0.05)
Age	–0.1543	0.4071
Years passed since the last renovation	–0.1241	0.5632
Inner surface	0.1251	0.5023
Number of floors	–0.1642	0.3775
Number of students	0.0894	0.6326
Number of class departments	0.0116	0.9504
Number of employees	–0.1519	0.4145
Number of shifts	–0.0236	0.8997
School condition	0.4769	0.0067
Adequate design solution	0.7569	8.3E-07
Implemented CMMS	0.2569	0.1629
Defined maintenance policy	0.3134	0.0860
Defined maintenance standard	0.4769	0.0067
Maintenance included in the mission and vision of the institution	0.0820	0.6609
Defined maintenance goals	0.0031	0.9868
Defined priorities and maintenance plans	0.2439	0.1859
Developed organizational structure of the institution	–0.0795	0.6709
Defined lines of communication within the structure	0.3166	0.0830
Defined maintenance inventory management procedure	0.2370	0.1992
Defined procedure for resolving maintenance requests	0.0539	0.7735
Defined procedure and method of documentation on maintenance	0.0874	0.6459
Conducting training of technical staff	0.2238	0.2261
Conducting trainings for principals	–0.2381	0.1971
Providing guidance and support (motivation, teamwork, good communication)	0.1746	0.3476
Carrying out supervision and quality control over maintenance	0.2272	0.2190
Measuring maintenance performance with performance indicators	0.5575	0.0011
Measuring student and staff satisfaction	0.3154	0.0839

5 CONCLUSION

The research was conducted on the current state of the maintenance management process in elementary schools in Primorje-Gorski Kotar County. So far, no comprehensive research on school buildings maintenance and related issues has been conducted in the Republic of Croatia. The research aim has been achieved and hypotheses have been confirmed.

Main research results indicate that the observed school buildings are old and the creation of their design solutions did not consider the influence on subsequent maintenance, which, according to the size of the correlation coefficient, has a pronounced influence on the overall efficiency of the maintenance management process. There is no record of the operational state of school buildings based on which a particular system of school maintenance priorities could be developed. Schools do not have sufficient maintenance funds available that would meet all their maintenance needs, which is the responsibility of the state government. Still, since the increase in funds is not likely, it is essential to consider all ways to spend them more effectively. Since many stakeholders participate in school maintenance, it is necessary to achieve better cooperation and teamwork, which has not been the case so far. One of the distinguished stakeholders are principals, who head the school buildings and all maintenance requirements have a starting point in them, but they do not have any maintenance knowledge. The school maintenance policy and standard are not defined well enough, nor is there a system for measuring the overall maintenance performance. The use of CMMS has almost not been reported at all.

If the obtained results are compared with world studies, it is evident that school maintenance in the Republic of Croatia follows similar negative trends.

The observed inadequacies in Croatian schools contribute to the inefficiency of the current maintenance management process, so the following set of essential recommendations is given:

- the condition of school buildings should be recorded in detail, based on which an investment priority system will be developed, which will ensure the most critical and necessary work at the beginning of maintenance plans and enable more transparent planning of capital projects and maintenance work;
- the ways and procedure in which maintenance takes place in schools should be precisely defined in the form of a maintenance policy that will make it clear how maintenance decisions are made and who is involved in this process;
- the standard on which schools are maintained should be determined, which will achieve the satisfaction of the Basic Requirements for Construction Works (BRCW), functional conditions, aesthetic conditions, and not least important, users satisfaction;
- education and training should be conducted regularly for all persons involved in the school maintenance process, especially for principals who are in charge of schools and who are chronically lacking in technical knowledge;
- cooperation between schools and their founders should be improved, primarily during the planning and reallocation of financial resources for maintenance, so that founders have a better insight into the specific user needs and work of the school;
- a system of key performance indicators should be developed based on which maintenance performance will be measured and, given the results of measurements, improvements will be made in problematic maintenance areas; this is very important for the continuous improvement of the overall process of maintenance management in schools;
- it is especially necessary to measure the satisfaction of students and staff; effective learning and teaching can only take place in a user-friendly environment, so it is necessary to examine how satisfied they are with that environment and to identify areas where special attention should be paid according to them;
- CMMS should be implemented to facilitate maintenance management; the software will enable more efficient communication and faster and easier maintenance planning, directing staff to activities, planning resource consumption, monitoring maintenance costs and times, keeping maintenance records, documenting data on failures, damage, their causes, etc.; the software includes all functions and management factors as well as persons involved in maintenance, from technical staff to the founder;
- when designing school buildings, maintenance experts should be involved in this process to develop a solution that will have the most cost-effective performance during the building use phase.

All these recommendations, as well as improvements in other school maintenance segments, should be included through a new maintenance management model that will show the way to the effective functioning of the entire management process. Through this model, maintenance is taken care of from the earliest stages of the project, from design. Within the model, the use of CMMS is particularly emphasized. This model should instruct school leaders and other stakeholders on how to plan, organize, lead, and control maintenance activities. Future research should go in the direction of detailed elaboration and implementation of such a model in schoolwork.

The population studied in this paper is limited to elementary school buildings in Primorje-Gorski Kotar County in the Republic of Croatia, which is one of the limitations of the research, therefore the obtained results on the state of management process cannot be presumed for other schools because there is no statistical justification for this. The study is also limited by the scope of theoretical analysis based on which the studied management factors are determined. Although a comprehensive literature analysis has been carried out, there is always the possibility of omitting a certain source of literature that may be relevant for the research.

Regardless of limitations, the developed methodological framework, along with potential minor modifications, can be applied to schools from other areas as well as to other types of buildings, which can provide a good insight into the current state of their maintenance management process, based on which they can then improve their operations in the maintenance segment.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

KTŠ designed and conducted the research and wrote the manuscript. SM and DC-P provided guidance, read and revised drafts of the manuscript text and figures, and helped develop and advance the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work has been supported by the University of Rijeka under the project numbers uniri-tehnic-18-125 and uniri-pr-tehnic-19-18.

REFERENCES

Albader, H., and Kandil, A. (2013). An Agent-Based Framework of a Maintenance Budget Allocation System for Educational Facilities. *GEN* 62 (1), 1–9.

Ali, A. S., Keong, K. C., Zakaria, N., Zolkafli, U., and Akashah, F. (2013). The Effect of Design on Maintenance for School Buildings in Penang, Malaysia. *Struct. Surv.* 31 (3), 194–201. doi:10.1108/SS-10-2012-0030

Aliyu, A., Bello, A., Muhammad, S., Singhry, M., and Bukar, M. G. (2016). "An Assessment of Building Maintenance Management Practice for Abubakar Tafawa Balewa University Teaching Hospital, Bauchi," in Proceedings/Abstracts and Programmes of the Academic Conference on Transformation Assessment, 12–20.

Alshehri, A. R. (2016). *Quality Management System for Building Maintenance*. PhD Thesis. Edinburgh: Heriot-Watt University.

Alzaben, H. (2015). *Development of a Maintenance Management Framework to Facilitate the Delivery of Healthcare Provisions in the Kingdom of Saudi Arabia*. PhD Thesis. Nottingham: Nottingham Trent University.

Capak, K., Barišin, A., Brdarić, D., Jeličić, P., Janev, N., Poljak, V., et al. (2015). *Zdravstveno-ekološki Čimbenici U Osnovnim Školama U Republici Hrvatskoj*. Croatian Institute of Public Health. Available at: https://www.hzjz.hr/download/DIZAJN_BROSURE_HZJZ_WEB.pdf (Accessed May 10, 2022).

de Morais, G. A. T., and Lordsleem Júnior, A. C. (2019). Building Maintenance Management Activities in a Public Institution. *Ecam* 26 (1), 85–103. doi:10.1108/ECAM-01-2018-0024

Dickerson, D. E., and Ackerman, P. J. (2016). Risk-based Maintenance Management of U.S. Public School Facilities. *Procedia Eng.* 145, 685–692. doi:10.1016/j.proeng.2016.04.069

El Shorafa, F. E. (2013). *Key Performance Indicators for Maintenance in Hospitals Buildings in Gaza Strip*. PhD Thesis. Gaza: The Islamic University of Gaza.

ElSamadony, A., Hossny, O., ElHakeem, A., and Hussein, D. (2013). An Asset Management System for Maintenance and Repair of Educational Buildings. *Int. J. Sci. Eng. Res.* 4 (6), 2053–2064.

Gala Mong, S., Fikri Mohamed, S., and Mohd. Saidin Misnan, M. (2018). Maintenance Management Model: an Identification of Key Elements for Value-Based Maintenance Management by Local Authority. *Int. J. Eng. Tech.* 7 (325), 35–43. doi:10.14419/ijet.v7i3.25.17467

Harun, M. T., Salamudin, N., and Hushin, H. F. (2013). Appraisal of the Sport Facilities Maintenance Management Practices of Malaysian Stadium Corporations. *Asian Social Science* 9 (12), 93. doi:10.5539/ass.v9n12p93

Izobo-Martins, O. (2014). *Maintenance Strategies and Condition of Public Secondary School Buildings in Ado-Odo/Ota Local Government Area Ogun State, Nigeria*. PhD Thesis. Ota: Covenant University.

Jandali, D., and Sweis, R. (2019). Factors Affecting Maintenance Management in Hospital Buildings. *Int. J. Build. Patho. Adaptation* 37 (1), 6–21. doi:10.1108/IJBPA-12-2017-0064

Katić, D., Krstić, H., and Marenjak, S. (2021). Energy Performance of School Buildings by Construction Periods in Federation of Bosnia and Herzegovina. *Buildings* 11 (2), 42. doi:10.3390/buildings11020042

Kim, J.-M., Kim, T., Yu, Y.-J., and Son, K. (2018). Development of a Maintenance and Repair Cost Estimation Model for Educational Buildings Using Regression Analysis. *J. Asian Archit. Build. Eng.* 17 (2), 307–312. doi:10.3130/jaabe.17.307

Kušljić, D. (2012). *Determination of Criteria for Evaluating the Success of Public-Private Partnership Construction Projects*. PhD Thesis. Osijek: University of Josip Juraj Strossmayer in Osijek.

Lee, C-K., and Jeon, Y-I. (2017). A Maintenance Cost Prediction Model for Elementary Schools by Correcting FM Budget History and Performance Data. *Int. J. Appl. or Innovation Eng. Manag.* 6 (5), 084–094.

Liu, S.-S., and Faizal Ardhiansyah Arifin, M. (2021). Preventive Maintenance Model for National School Buildings in Indonesia Using a Constraint Programming Approach. *Sustainability* 13 (4), 1874. doi:10.3390/su13041874

Marzouk, M., and Awad, E. (2016). Establishing Multi-Level Performance Condition Indices for Public Schools Maintenance Program Using AHP and Fuzzy Logic. *Stud. Inf. Control* 25 (3), 343–352. doi:10.24846/v25i3y201608

Meghanathan, N. (2016). Assortativity Analysis of Real-World Network Graphs Based on Centrality Metrics. *Cis* 9 (3), 7–25. doi:10.5539/cis.v9n3p7

Nah, M. N. N., Abdullah, S., Razak, A. A., and Hanafi, M. H. (2012). Property Management Implementation Problems within Malaysian Public School. *Int. J. Soc. Sci.* 1 (1), 17–28.

Nik-Mat, N. E. M., Kamaruzzaman, S. N., and Pitt, M. (2011). Assessing the Maintenance Aspect of Facilities Management through a Performance Measurement System: A Malaysian Case Study. *Procedia Eng.* 20, 329–338. doi:10.1016/j.proeng.2011.11.174

Ochien'g, R. R. (2012). *Developing a Model Framework for Effective Maintenance of Public Office Buildings in Kenya: A Survey of Multi Storeyed Office Buildings in NairobiPhD Thesis*. Nairobi: Jomo Kenyatta University of Agriculture and Technology.

Ogunoh, P. E., Mbanusi, E. C., and Okoye, U. P. (2018). Effective Implementation of Maintenance Models in Building Maintenance Process. *J. Eng. Res. Rep.* 1–10. doi:10.9734/JERR/2018/43466

Olanrewaju, A. L. A., and Abdul-Aziz, A. R. (2014). *Building Maintenance Processes and Practices: The Case of a Fast Developing Country*. New York: Springer.

Oyenuga, S. O., Akinsola, O. E., Hussaini, P. O., and Fatokun, A. O. (2012). Maintenance of University Facilities in Developing Country: Case Study of Lagos State University Ojo Nigeria. *Mediterr. J. Soc. Sci.* 3 (11), 69–75. doi:10.5901/mjss.2012.v3n11p69

Palis, P., and Saidin Misnan, M. (2018). A Review of Key Factors that Affect University Building Maintenance Costs. *Int. J. Eng. Tech.* 7 (325), 32–34. doi:10.14419/ijet.v7i3.25.17466

Primorje-Gorski Kotar County (2020). O Županiji. Available at: https://www.pgz.hr/o-zupaniji/(Accessed January 24, 2022).

Ropi, R. M., and Tabassi, A. A. (2014). Study on Maintenance Practices for School Buildings in Terengganu and Kedah, Malaysia. In O. Mydin Proceedings of Building Surveying and Technology Undergraduate Conference. Langkawi, Malaysia, doi:10.1051/matecconf/20141003003

Saraiva, T., De Almeida, M., Bragança, L., and Barbosa, M. (2018). Environmental Comfort Indicators for School Buildings in Sustainability Assessment Tools. *Sustainability* 10 (6), 1849. doi:10.3390/su10061849

Sharma, D. (2020). The Importance of School Education in Child Development. Available at: https://www.educationworld.in/the-importance-of-school-education-in-child-development/(Accessed January 20, 2022).

Shohet, I. M., and Nobili, L. (2017). Application of Key Performance Indicators for Maintenance Management of Clinics Facilities. *Int. J. Strategic Prop. Manag.* 21 (1), 58–71. doi:10.3846/1648715X.2016.1245684

State Audit Office (2018). *Izvješće O Obavljenoj Reviziji Učinkovitosti Kapitalnih Ulaganja U Osnovne I Srednje Škole Na Području Republike Hrvatske*. Sisak.

Teixeira, J., Amoroso, J., and Gresham, J. (2017). Why Education Infrastructure Matters for Learning. Available at: https://blogs.worldbank.org/education/why-education-infrastructure-matters-learning (Accessed January 24, 2022).

The Government of the Republic of Croatia (2012). *Okvirni Program Izgradnje, Dogradnje I Rekonstrukcije Javnih Građevina Prema Ugovornom Obliku Javno Privatnog Partnerstva*. Available from: https://investcroatia.gov.hr/wp-content/uploads/2015/11/okvirni-program-jpp.pdf (Accessed April 25, 2022).

Tijanić Štrok, K. (2021). *Development of the Model for Efficient Maintenance Management of Public Educational Buildings*. PhD Thesis. Osijek: University of Josip Juraj Strossmayer in Osijek.

Velmurugan, R. S., and Dhingra, T. (2015). Maintenance Strategy Selection and its Impact in Maintenance Function. *Int. J. Operations Prod. Manag.* 35 (12), 1622–1661. doi:10.1108/IJOPM-01-2014-0028

Vieira, A. C. V., and Cardoso, A. J. M. (2014). "Maintenance Conceptual Models and Their Relevance in the Development of Maintenance Auditing Tools for School Buildings' Assets: an Overview," in Proceedings of Maintenance Performance Measurement and Management Conference, 3–10. doi:10.14195/978-972-8954-42-0_1

Wing, A. C. K., Mohammed, A. H., and Abdullah, M. N. (2016). A Review of Maintenance Priority Setting Methods. *Int. J. Real Estate Stud.* 10 (2), 1–9.

Wuni, I. Y., Agyeman-Yeboah, S., and Boafo, H. K. (2017). Poor Facility Management in the Public Schools of Ghana; Recent Empirical Discoveries. *J. Sustain. Dev. Stud.* 11 (1), 1–30.

Xaba, M. I. (2012). A Qualitative Analysis of Facilities Maintenance - a School Governance Function in South Africa. *Saje* 32 (2), 215–226. doi:10.15700/saje.v32n2a548

Yong, C. Y., and Sulieman, M. Z. (2015). Assessment of Building Maintenance Management Practice and Occupant Satisfaction of School Buildings in Perak, Malaysia. *J. Teknol.* 75 (5), 57–61. doi:10.11113/jt.v75.4995

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Built Environment

ORIGINAL RESEARCH
published: 22 June 2022
doi: 10.3389/fbuil.2022.909018

Revenue-Sharing Contract Design for Construction Onsite Equipment Sharing

*Zhi Liu[1,2], Hongwei Wang[3] and Yong Xie[1]**

[1]School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China, [2]China Construction Eighth Engineering Division Corp., Ltd., Shanghai, China, [3]School of Management, Huazhong University of Science and Technology, Wuhan, China

Equipment sharing between contractors can relieve equipment shortages and enhance construction productivity. Previous studies focused on how to allocate cooperative gains of resource sharing among contractors, and yet how to ensure the owner and contractors form a coalition that contributes to maximizing resource-sharing gains still needs to be studied. This article examines the contract design problem of motivating the contractors and owner to willingly cooperate to increase the amount of shared equipment to an optimal level to maximize the project's overall gain through equipment sharing. First, two trilateral equipment-sharing game models including a critical contractor, a non-critical contractor, and the owner are developed, which represent contractors' different leadership positions in equipment sharing. Then, a set of revenue-sharing contracts are devised to compensate the contractors' equipment-sharing cost and the owner's extra duration reward cost by reassigning the consequently increased operation income. Eventually, a numerical study demonstrates that the proposed contracts can prompt members to make decisions according to an optimal solution to maximize the overall gains of equipment sharing, and each member's gain and the construction duration compression are improved. This article contributes to onsite resource management by introducing a revenue-sharing contract to fairly compensate construction members' equipment-sharing costs to achieve optimal cooperative gains.

Keywords: equipment sharing, revenue-sharing contract, construction time-cost function, duration reward/penalty, construction time and cost

OPEN ACCESS

Edited by:
Yongkui Li,
Tongji University, China

Reviewed by:
Jiayu Chen,
City University of Hong Kong, Hong Kong SAR, China
Ge Wang,
Huazhong Agricultural University, China

****Correspondence:***
Yong Xie
xyhust@163.com

Specialty section:
This article was submitted to Construction Management, a section of the journal Frontiers in Built Environment

Received: *31 March 2022*
Accepted: *11 May 2022*
Published: *22 June 2022*

Citation:
Liu Z, Wang H and Xie Y (2022) Revenue-Sharing Contract Design for Construction Onsite Equipment Sharing. Front. Built Environ. 8:909018. doi: 10.3389/fbuil.2022.909018

1 INTRODUCTION

Critical construction equipment can be found in large-scale infrastructure construction projects; examples of such equipment include tower belt conveyors (which pour massive, highly intensified concrete in hydropower station projects) and deepwater riprap leveling barges (which are used in the seabed-immersed tube construction of the Hong Kong-Zhuhai-Macao bridge). Gransberg et al. (2006) defined construction machines that undertake construction tasks on a critical path as critical equipment. Since it is imperative to identify which construction equipment assigned to the project can directly affect the project completion time, the concept of critical equipment includes two key points:

(1) The first is to check the float available in each activity in the schedule and classify the equipment associated with each critical activity as critical equipment.

(2) The second point is to declare the equipment that physically or technically cannot be increased within a specific period as critical equipment, letting the non-critical ones rise and fall based on the needs of the project.

To be clear, critical equipment in the research is defined as construction machines that are limited and fixed in quantity during a certain period and undertake construction tasks on a critical path. For example, the tunnel boring machine is critical equipment compared to excavators in a metro construction project. Due to the complex and uncertain characteristics of the construction environment, critical equipment shortages occasionally happen, thus resulting in project postponements and extra costs (Zhang et al., 2020). Though equipment sharing can mitigate critical equipment shortages and improve equipment productivity, contractors make decisions independently and lack mutual trust. Consequently, one contractor may incur equipment idleness, while another is experiencing equipment shortages, thereby decreasing the performance of the project (Bendoly et al., 2010). Thus, the key problem of contractors' resource sharing is to design a contract to coordinate the conflicts of interests among contractors and the owner, which is grounded in the theory of construction supply chain management (CSCM) (Xue et al., 2007). CSCM focuses on cooperative relationships among members during construction engineering, and its objective is to optimize construction performance and improve the client's utility with fewer costs (Feldman and Tamir, 2012). In previous studies of CSCM coordination, these selfish behaviors can be coordinated through mechanisms such as auction and negotiation, when the owner supplies most equipment resources (Fink, 2006). In auctions, contractors bid for resources within a certain time period, and the use of the resources is appointed to the highest bidder by the owner (Lee et al., 2003). In the negotiation mechanism, multiple contractors negotiate with a mediator (the owner or a general contractor), who allocates the shared resources according to protocols of minimizing project costs or delays (Lau et al., 2006). Auction and negotiation help bridge contractors' onsite needs of resources and the owner's resource allocation, which are implemented when most resources belong to the owner (Xia et al., 2008). However, in large-scale projects, critical equipment resources are often dispersedly owned by multiple contractors, thus limiting the coordinating effect of the owner, and requiring a coordination mechanism representing the contractors' interests and decision-making.

To reflect the contractors' interests during resource sharing, current studies applied cooperative game theories that focus on assigning overall cooperative gain to contractors (Gkatzelis et al., 2016; Moradi et al., 2019; Akhbari, 2020). These studies contribute to determining how to verify the fairness and validity of cooperative gain allocation rules. Nevertheless, in construction practice, how to devise a cooperative mechanism that ensures the owner and contractors act in ways consistent with the optimal resource-sharing solution still needs to be researched. Therefore, the authors consider the contract design problem of how to optimize every participant's interest and to maximize the project gain of equipment sharing. The novelty of this research is proposing a cooperative contract based on revenue-sharing to motivate equipment-sharing members to comply with the equipment-sharing solution which optimizes the project's overall gain through modeling a trilateral game among the critical contractor, noncritical contractor, and owner.

In this study, equipment sharing among multiple contractors and the owner is as follows: when the critical contractor (the contractor who undertakes construction activity on a critical path) needs to expedite construction time, he will pay rent to a noncritical contractor (the contractor who undertakes construction activity that is not on a critical path) for extra equipment. Then, by reducing the critical path duration, the critical contractor gains more time reward, and the owner benefits from the project's running income due to early completion. The authors analyze and model the decision-making based on game theory: first, by modeling a trilateral game among the critical contractor, noncritical contractor, and owner, the individual equipment-sharing gains of these participants under a non-cooperative situation and construction system optimal equipment-sharing plan are determined. Second, concerning the effect of different equipment-sharing leaderships, several revenue-sharing contracts based on static transferable utility are devised to maximize all the participants' gains when the participants make decisions according to the system optimal equipment-sharing scheme.

The remainder of the article is organized as follows: in **Section 2**, the authors review previous construction resource-sharing research and explain how the proposed coordination mechanism can coordinate the interest conflicts of the owner and contractors and realize the construction system optimum. In **Section 3**, the authors describe the basic problem in equipment sharing. In **Section 4**, Stackelberg game models under two leading scenarios are developed. Then, in **Section 5**, the system optimal equipment-sharing scheme is determined, and a revenue-sharing contract is devised. In **Section 6**, a numerical analysis is conducted to demonstrate the effect of the contract. Finally, the article concludes and discusses the authors' research work.

2 LITERATURE REVIEW

2.1 Construction Resource-Sharing Concept

In previous literature on the resource-sharing problem, the notation of resource sharing in the construction project is introduced by Perera (1983) regarding the resource allocation problem of linear construction. Perera applied resource-hour as the basic unit of sharing constrained resources (labor, equipment, etc.) between construction activities. By scheduling intermittent construction of noncritical activities, Karra and Nasr (1986) presented resource sharing between noncritical activity and critical activity to save the cost of renting equipment for critical activities. Xu et al. (2013) defined resource sharing as the optimal allocation of multiple types of equipment to multiple construction sites in several stages to raise construction productivity and reduce cost. Considering the above literature, the authors conclude in this study that resource sharing consists of cooperative behaviors in which contractors share the usage of equipment, thereby significantly affecting the construction schedule so that both their individual gains and construction schedule can be optimized.

2.2 Coordination Mechanism in Resource-Sharing Problem

Previous studies classify onsite construction resource-sharing problems as resource-constrained project scheduling problems (RCPSPs) and aim mainly to minimize construction costs and improve the construction schedule (Hartmann and Briskorn, 2010; Xu and Zeng, 2011; Kucuksayacigil and Ulusoy, 2020). Unlike traditional RCPSP studies, which regard equipment sharing as a solution to optimizing construction engineering objectives, this study sees the equipment-sharing problem as a process that coordinates the interest and decision-making of every participant. Previous research on the coordination mechanism of resource sharing can be analyzed from three perspectives: combinatorial auction, negotiation, and game theory. In a combinatorial auction, contractors bid for the price of using several resources at a certain amount for a certain duration. The owner who acts as the auctioneer will distribute the combinatorial resources to the highest bidder and then start the next round of the auction until all the contractors' needs are satisfied (Confessore et al., 2007; Villahoz et al., 2010; Wang et al., 2014).

In a negotiation, the owner acts as a negotiator who receives conflicting resource demands from contractors and determines the resource allocation according to a series of negotiation rules that aim to achieve project objectives, such as minimizing the project cost or construction time (Homberger, 2007; Homberger, 2012). Compared to traditional resource allocation optimization methods, combinatorial auction and negotiation contribute to the timely reflection of resource demands from the contractors and can provide mutually beneficial contracts after multiple rounds of negotiations until contracts are accepted by both the owner and contractors. Additionally, the owner is assumed to own all the construction resources in the previous coordination mechanisms and is, therefore, responsible for determining the cooperative criterion in the coordination process and final determination of the resource allocation.

2.3 Game Theory Application

However, in large-scale construction projects, the owner's coordinating role can be challenged by contractors who own most of the construction equipment; thus, both independent decision-making and coalition forming by contractors should be examined. Game theory, including non-cooperative game and cooperative game, presents a suitable analytical framework for decentralized decision-making during resource sharing (Cachon and Zipkin, 1999).

Non-cooperative game theory can facilitate predicting the expected gains and behaviors of decision-makers who aim at maximizing individual interests. Samaddar et al. (2006) applied the Stackelberg game model to simulate joint knowledge resource sharing between corporations and determined the requirements of resource sharing between new and established firms. Moufid et al. (2017) modeled cheating behaviors during equipment sharing (such as hiding sharing time and lowering maintenance cost in production equipment sharing) among small-sized manufacturing enterprises and revealed the relationship between the probability of detecting cheating and detection cost. Concerning construction equipment-sharing research, Liu et al. (2018) suggested the motivative effect of a time incentive applied by the owner can benefit equipment sharing among contractors by developing the leader–follower game model. When there is no binding cooperative contract or authority to coordinate, non-cooperative game theory can help studies on resource sharing establish theoretical frameworks to investigate the influence of self-interested behaviors of resource-sharing members on their cooperation results.

Different from the non-cooperative game for maximizing individual interests, if cooperation can bring more income than competition, the players in an equipment-sharing game will choose to form a cooperative coalition and decide the fair and efficient allocation of the benefits of resource-sharing cooperation. The allocation methods of cooperative benefits are applied in construction resource-sharing scenarios such as cost savings/profit allocation among partnering firms, delay penalties and expedition rewards allocation among activities, and apportionment of construction costs for a public project (Eissa et al., 2021a). Considering the minimal equipment reserve for each contractor to carry out construction work, Asgari et al. (2014) developed a cooperative game model of sharing temporary and constant equipment among activities and allocated cooperative gains according to a Shapley value solution. By integrating the sharing cost of equipment operation, transportation, and installation, Hafezalkotob et al. (2018) proposed a cooperative game model of resource sharing between different construction stages to optimize construction system gain. Moradi et al. (2019) established a robust optimization model of sharing resources that are uncertain in availability during uncertain activity duration and compared subcontractors' satisfaction levels to different cooperative gain allocation rules. In the project-scheduling problem of sharing renewable resources between subprojects, Akhbari (2020) applied the Shapley–Shubik power index and DP (propensity to disrupt) value to evaluate the stability and fairness of allocation rules, including the Shapley value, nucleolus, and Nash–Harsanyi solution. Eissa et al. (2021b) developed a conceptual framework to allocate profit among construction joint ventures (CJVs) based on each firm's marginal contribution in order to reduce profit-share-related disagreements among CJV members. Li et al. (2021) proposed a multi-agent-based cooperative approach to mitigate resource competition of decentralized project managers and reach a negotiation protocol of resource allocation with a central decision-maker of all projects.

Previous studies on cooperative game theory shed light on how to allocate cooperative gains and how to verify the fairness and validity of allocation rules (Kemahlioglu-Ziya and Bartholdi, 2011; Guajardo and Rönnqvist, 2015). However, in practice, the

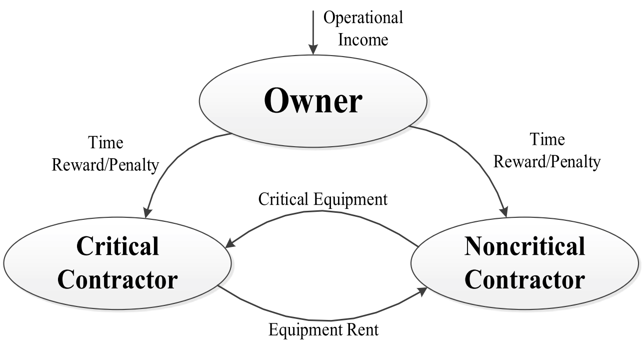


FIGURE 1 | Critical equipment-sharing model.

pursuit by contractors of their own self-interests may impede the implementation of gain allocation methods. Thus, the authors expand the cooperative gain allocation research on traditional cooperative resource leveling to coordination mechanism design, which focuses on how to prompt decision-makers to willingly achieve the system optimum.

3 PROBLEM DESCRIPTIONS

As in a BOT project, the concession period defines the time span in which the private investor has the right to commercially operate the infrastructure facility before it is transferred back to the government. Since the concession period is fixed, early completion of construction will result in a longer operation period which brings additional income (such as highway tolls and business advertisements). Considering construction contractors' equipment sharing can reduce the delay of the construction project, the financing model BOT (Build-Operate-Transfer) is used to describe the functional relationship between equipment sharing and the owner's interest. In a BOT (Build-Operate-Transfer) construction project, a critical contractor, a noncritical contractor, and the owner together constitute the basic equipment-sharing problem. Specifically, to prevent construction delay, the owner rewards contractors for early completion and penalizes them for delays. When additional equipment is needed to compress the construction duration to gain more time reward or lessen a penalty, a critical contractor will share equipment from other noncritical contractors. Additionally, the compression of the construction period can produce more operation income for the owner since the operation period will be extended in the fixed concession period of the BOT project (Zhang et al., 2016). Decision-making in equipment sharing involves two steps: first, the critical contractor (who needs equipment) proposes an equipment rent price to the noncritical contractor (who offers extra equipment). Then, taking the possible time penalty and construction cost into account, the noncritical contractor will determine the amount of shared equipment to maximize his or her own interest in responding to the critical contractor's request. The gains of both contractors will then reach equilibrium after the contractors' negotiation. The equipment-sharing process is modeled as a Stackelberg game in which the critical contractor acts as the leader and the noncritical contractor acts as the follower. The game model is presented in **Figure 1**.

During equipment sharing between contractors, bargaining power, which represents the right to decide equipment rent price, usually determines the leadership of the decision-making process. In practice, the construction time of critical contractors who undertake critical activities is directly associated with the project schedule. Thus, the critical contractor often possesses the bargaining power and acts as the leader of equipment sharing. However, if more than one critical contractor proposes identical equipment demands, if more critical machines break down, or if equipment demand surges due to design modification, the bargaining power will be handed over to the noncritical contractors, who have excessive equipment resources. In this scenario, equipment sharing can be modeled as a Stackelberg game in which the noncritical contractor acts as the leader and the critical contractor acts as the follower. In this article, the effects of two different leadership scenarios on the game equilibrium of the owner and contractors are analyzed.

In this article, the owner and contractors are assumed to be rational individuals who seek to maximize their own interests. Moreover, each player's information is assumed to be symmetric, i.e., the owner and contractors have full information on their income and payment.

3.1 Reward/Penalty Type for Construction Duration

Three important parameters are introduced in this section: reward/penalty type, construction cost, and amount of shared equipment. To be specific, the reward/penalty type reflects the owner's influence on the contractors' construction duration. Construction cost represents the relationship between the contractor's gain and construction duration. The amount of shared equipment shows the quantitative effect of equipment sharing on each contractor's construction duration. Based on these parameters, gains of the owner and contractors are associated, and how both participate in equipment sharing is revealed.

In the construction industry, the reward/penalty type for construction duration is usually implemented in linear form; therefore, the reward for early completion or punishment for the delay is based on a fixed reward/penalty coefficient (Jaraiedi et al., 1995; Shr and Chen, 2003). The function of reward/penalty is $W = I_o \cdot (d - d_0)$, where W represents the reward/penalty amount, I_o stands for the fixed reward/penalty coefficient, d is the contract construction duration, and d_0 is the actual construction duration. Thus, the reward/penalty amount of the critical contractor is expressed as follows:

$$W_a = I_o \cdot (d - d_0). \tag{1}$$

The reward/penalty amount of the noncritical contractor is expressed as follows:

$$W_b = I_o \cdot (d - d_0). \tag{2}$$

TABLE 1 | List of symbols and descriptions.

Symbol	Description
l_o	Reward/penalty coefficient of construction duration
P	Equipment rent price
a_1	Critical contractor's duration-cost coefficient
a_0	Critical contractor's fixed cost
M_t	Amount of shared equipment
b_1	Noncritical contractor's duration-cost coefficient
b_0	Noncritical contractor's fixed cost
d_{a1}	Construction duration of the critical contractor before equipment sharing
d_{a2}	Construction duration of the critical contractor after equipment sharing
d_{b1}	Construction duration of the noncritical contractor before equipment sharing
d_{b2}	Construction duration of the noncritical contractor after equipment sharing
R	Operation income in unit time

3.2 Construction Cost Function

Construction cost can be approximatively described as a convex function of construction duration (Callahan et al., 1992; Shr and Chen, 2003). In this study, the relationship between construction cost and duration is expressed as the quadratic function $C(d) = a_1 (d - d_0)^2 + a_0$, where $C(d)$ represents construction cost; a_1 represents the cost–duration coefficient, which is associated with the contractor's management ability; and a_0 represents the fixed construction cost. Thus, the construction cost–duration function of the critical contractor is expressed as follows:

$$C_a(d) = a_1 (d - d_{a1})^2 + a_0. \tag{3}$$

The construction cost–duration function of the noncritical contractor is expressed as follows:

$$C_b(d) = b_1 (d - d_{b1})^2 + b_0. \tag{4}$$

Considering critical activities will generally be allocated with more resources (manpower, equipment, and materials) than noncritical activities, the authors assume that duration changes for critical contractors consume more cost than those for noncritical contractors ($a_1 > b_1$).

3.3 Influence of Shared Equipment on Construction Duration

To describe the relationship between the amount of shared equipment and construction cost, the amount of shared equipment is represented by a machine team marked as M_t, which means one team of operators working on one piece of equipment in a unit period. The construction production of the machine team is marked as D_w. While performing construction tasks, the productivity of the same equipment is assumed to be identical for every contractor and match their estimated level. The amount of the critical contractor's own equipment is marked as m_a, and the contract duration before and after equipment sharing is marked as d_{a1} and d_{a2}, respectively. Since the total construction task equals the product of production per unit time and construction duration, the critical contractor's construction amount is expressed as $d_{a1} \cdot m_a \cdot D_w$. The construction amount is assumed to be unchanged during equipment sharing and is shown as $d_{a1} \cdot m_a \cdot D_w = d_{a2} \cdot m_a \cdot D_w + M_t \cdot D_w$. By solving this equation, the duration variation of the critical contractor is given as follows:

$$\sigma d_a = d_{a1} - d_{a2} = \frac{M_t}{m_a}. \tag{5}$$

Similarly, the duration variation of the noncritical contractor is given as follows:

$$\sigma d_b = d_{b2} - d_{b1} = \frac{M_t}{m_b}. \tag{6}$$

A list of mentioned symbols and descriptions is presented in **Table 1**.

4 TRILATERAL GAME MODEL OF EQUIPMENT SHARING

4.1 Critical Contractor Leads Equipment Sharing (Scenario I)

4.1.1 Critical Contractor's Gain Function

The critical contractor's gain consists of three parts: contract income, which includes a fixed fee (C_a) and a reward/penalty (W_a); the equipment rent payment, which is the product of the rent price (P) and shared equipment amount (M_t); and the construction cost ($C(d_a)$). The gain function before equipment sharing is as follows:

$$U_{a1} = C_a + W_a - C(d_{a1}). \tag{7}$$

The critical contractor's gain function after equipment sharing is as follows:

$$U_{a2} = C_a + W_a - P \cdot M_t - C(d_{a2}). \tag{8}$$

As the leader in the Stackelberg equipment-sharing game, the critical contractor will decide the optimal rent price P before the noncritical contractor makes a decision. The decision-making problem is as follows:

$$\begin{aligned} &\max_P U_{a2} = C_a + W_a - P \cdot M_t - C(d_{a2}) \\ &\text{s.t.}\ \ P > 0, d_{a1} > d_{a2} > 0 \end{aligned}. \tag{9}$$

4.1.2 Noncritical Contractor's Gain Function

The gain of the noncritical contractor consists of three parts: contract income, which includes a fixed fee (C_b) and a reward/penalty (W_b); equipment rent income, which is the product of the rent price (P) and shared equipment amount (M_t); and the construction cost ($C(d_b)$). The gain function before equipment sharing is as follows:

$$U_{b1} = C_b + W_b - C(d_{b1}). \tag{10}$$

The noncritical contractor's gain function after equipment sharing is as follows:

$$U_{b2} = C_b + W_b + P \cdot M_t - C(d_{b2}). \tag{11}$$

As the follower in the equipment-sharing game, the noncritical contractor's decision-making problem is to decide the optimal shared equipment amount M_t according to the critical contractor's optimal rent price to maximize the gain of the noncritical contractor. The decision-making problem is as follows:

$$\begin{aligned} &\max_{M_t} U_{b2} = C_b + W_b + P \cdot M_t - C(d_{b2}) \\ &\text{s.t.} \quad M_t > 0, d_{b2} > d_{b1} > 0 \end{aligned}. \tag{12}$$

4.1.3 Equilibrium of the Equipment-Sharing Game

Based on backward induction, the solution of the Stackelberg equipment-sharing game is obtained in three steps (Rasmusen, 2006). First, in response to any certain rent price P from the critical contractor, the noncritical contractor's optimal shared amount M_t is determined by solving the derivative of the gain function ($\frac{\partial U_b}{\partial M_t} = 0$). The optimal shared equipment amount is as follows:

$$M_t = \frac{1}{2} \frac{(P \cdot m_b - I_o) m_b}{b_1}. \tag{13}$$

Second, as the second derivative of the gain function $\frac{\partial^2 U_b}{\partial M_t^2} = -\frac{2b_1}{m_b^2} < 0$, the noncritical contractor's gain U_b is the convex function of M_t. By inserting optimal M_t (**Eq. 13**) in the critical contractor's gain function (**Eq. 8**), the critical contractor's equilibrium rent price P is calculated as follows:

$$P^* = \frac{I_o \left(b_1 m_a m_b + a_1 m_b^2 + b_1 m_b^2\right)}{m_b \left(a_1 m_b^2 + 2 b_1 m_b^2\right)}. \tag{14}$$

The final step is to substitute the equilibrium rent price P (**Eq. 14**) in **Eq. 13**; thus, the noncritical contractor's equilibrium shared equipment amount is as follows:

$$M_t^* = \frac{1}{2} \frac{I_o m_a (m_b - m_a) m_b}{a_1 m_b^2 + 2 b_1 m_b^2}. \tag{15}$$

Equation 15 shows that only when the noncritical contractor's original equipment amount is higher than that of the critical contractor ($m_b > m_a$), equipment sharing will occur (M_t*>0). This observation can be explained by comparing the duration variations of both contractors (**Eqs 5, 6**). Once the shared amount M_t is determined, the more original equipment there is, the less the duration variation and duration reward/penalty. In other words, only when the noncritical contractor's penalty for the delay is less than the critical contractor's reward can equipment sharing be profitable and applicable to both contractors.

By substituting **Eq. 15** in the critical contractor's duration variation (**Eq. 5**), the duration variation of the construction project is as follows:

$$\sigma d_a = \frac{1}{2} \frac{I_o (m_b - m_a) m_b}{a_1 m_b^2 + 2 b_1 m_a^2}. \tag{16}$$

By substituting the equilibrium value of P (**Eq. 14**) and M_t (**Eq. 15**) in the critical contractor's gain function (**Eq. 8**), we can obtain the critical contractor's optimal gain:

$$U_a = \frac{1}{4} \frac{I_o^2 m_b^2 - 2 I_a^2 m_a m_b + I_o^2 m_a^2 - 4 a_0 a_1 m_b^2 - 8 a_0 b_1 m_a^2}{a_1 m_b^2 + 2 b_1 m_a^2}. \tag{17}$$

By a similar calculation, the noncritical contractor's optimal gain is as follows:

$$U_b = \frac{1}{4} \frac{I_o^2 b_1 m_a^2 m_b^2 - 2 I_o^2 b_1 m_a^3 m_b + I_o^2 b_1 m_a^2 m_b^2 - 4 a_1^2 b_0 m_b^4 - 16 a_1 b_0 b_1 m_a^2 m_b^2 - 16 b_0 b_1^2 m_a^4}{\left(a_1 m_b^2 + 2 b_1 m_a^2\right)^2}. \tag{18}$$

4.1.4 The Owner's Gain Function

From the perspective of whole-life management, early completion of construction will result in additional income (such as highway tolls and business advertisements) in the project operation period (Chen and Ma, 2007; Lv et al., 2015). As in a BOT project, the concession period of the owner is certain and marked as d_c, and the expected unit operation income is marked as r. The operation period d_r will be prolonged when the construction period d_a is shortened, thus resulting in extra operation income r. Thus, operation income U_r becomes a function of the construction duration:

$$U_r = r \cdot d_r = r(d_c - d_a). \tag{19}$$

In this study, we assume the owner pays a fixed total price and duration reward/penalty to the contractors. The contract payment to both contractors is marked as P_{a+b}.

$$P_{a+b} = C_a + I_o(d_{a1} - d_{a2}) + C_b + I_o(d_{b1} - d_{b2}). \tag{20}$$

Since the owner's gain consists of contract payment P_{a+b} and operation income U_r, the objective of the owner is to determine the optimal reward/penalty coefficient I_o to maximize his or her gain.

$$\begin{aligned} &\max_{I_o} \mathrm{U}_o = \mathrm{U}_r - P_{a+b} \\ &= r(d_c - d_a) - (C_a + I_o(d_{a1} - d_{a2}) + C_b + I_o(d_{b1} - d_{b2})) \end{aligned}. \tag{21}$$

Proposition 1: $U_o(I_o)$ is convex with respect to I_o.

Proof: By substituting **Eqs 5, 6**, and **15** in **Eq. 21**, we obtain

$$U_o = \frac{-(m_a^2 - 2 m_a m_b + m_b^2) I_o^2 - (m_a m_b r - m_b^2 r) I_o - 2 a_1 m_b^2 r d_{a1} + 2 a_1 m_b^2 r dc - 4 b_1 m_a^2 r d_{a1} + 4 b_1 m_a^2 r d_c}{2\left(a_1 m_a^2 + 2 b_1 m_a^2\right)}.$$

By calculating and simplifying the second derivative of U_o, we obtain

TABLE 2 | Equilibrium values in Scenario I.

Parameter	Equilibrium value
P^*	$\frac{r(a_1 m_b^2+b_1 m_a^2+b_1 m_a m_b)}{2(m_b-m_a)(a_1 m_b^2+2b_1 m_a^2)}$
$M_t{}^*$	$\frac{m_a m_b^2 r}{4(a_1 m_b^2+2b_1 m_a^2)}$
$I_o{}^*$	$\frac{m_b r}{2(m_b-m_a)}$
$U_a{}^*$	$\frac{m_b^2 r^2}{16(a_1 m_b^2+2b_1 m_a^2)}-a_0$
$U_b{}^*$	$\frac{b_1 m_a^2 m_b^2 r^2}{16(a_1 m_b^2+2b_1 m_a^2)^2}-b_0$
$U_o{}^*$	$\frac{m_b^2 r^2}{8(a_1 m_b^2+2b_1 m_a^2)}+(d_c-d_{a1})r$

$$\frac{d^2\mathrm{U}_o}{dI_a^2} = -\left(\mathrm{m_a}-\mathrm{m_b}\right)\frac{}{a_1 m_b^2+2b_1 m_a^2<0.}$$

The optimal reward/penalty coefficient I_o is determined by solving the first derivative of the owner's gain function ($\frac{d\mathrm{U}_o}{dI_o}=0$) as follows:

$$I_o^* = \frac{m_b r}{2(m_b-m_a)}. \tag{22}$$

The owner's optimal gain is as follows:

$$U_o^* = \frac{m_b^a r^2}{8(a_1 m_b^2+2b_1 m_a^2)} + r(d_c-d_{a1}). \tag{23}$$

The results of **Eqs 14–18** show that when the contractors' original equipment amounts and duration-cost coefficients are fixed, the equilibrium values of the equipment-sharing game are determined by the owner's decision, which is the reward/penalty coefficient. In other words, the contractors' optimal equipment rent price, shared amount, and gains all depend on the optimal reward/penalty coefficient. Therefore, by substituting **Eq. 22** in **Eqs 14**, **15**, **17**, and **18**, the equilibrium values of the owner and two contractors are calculated, as shown in **Table.2**.

4.2 Non-critical Contractor Leads Equipment Sharing (Scenario II)

When demand for equipment resources surpasses supply, the noncritical contractor will gain bargaining power and take leadership of the equipment sharing. In this situation, the game sequence is in reverse: first, the noncritical contractor will offer equipment rent price P. Then, considering the rent payment and duration reward, the critical contractor will respond to the offered rent price by deciding its optimal shared equipment amount M_t.

As the leader in the equipment-sharing game, the noncritical contractor will decide the optimal rent price P to maximize his or her gain.

$$\begin{aligned}&\max_P U_{b2} = C_b - W_b + P\cdot M_t - C(d_{b2})\\&\text{s.t.}\quad P>0, d_{b2}>d_{b1}>0\end{aligned}. \tag{24}$$

In reaction to the noncritical contractor's rent price, the critical contractor will determine the optimal shared equipment amount M_t to maximize his or her gain. The critical contractor's decision-making is described as follows:

$$\begin{aligned}&\max_{M_t} U_{a2} = C_a + W_a - P\cdot M_t - C(d_{a2})\\&\text{s.t.}\quad M_t>0, d_{a1}>d_{a2}>0\end{aligned}. \tag{25}$$

The solution of the equilibrium values in Scenario II is similar to the solution in Scenario I, and the values are presented in **Table 3**.

4.3 Comparison of Scenarios I and II

By comparing the rent price and shared amount in both scenarios, we obtain

$$P^* - P^l = -\frac{(a_1^2 m_b^4 + a_1 b_1 m_a^2 m_b^2 + b_1^2 m_a^4)r}{2m_a(2a_1 m_b^2+b_1 m_a^2)(a_1 m_b^2+2b_1 m_a^2)}, \tag{26}$$

$$M_t^* - M_t^l = \frac{m_a m_b^2 r(a_1 m_b^2 - b_1 m_a^2)}{4(a_1 m_b^2+2b_1 m_a^2)(2a_1 m_b^2+b_1 m_a^2)}. \tag{27}$$

As **Eqs 26** and **27** show, the equipment rent price is lower in Scenario I than that in Scenario II. Furthermore, as $a_1 > b_1$ and $m_b > m_a$, the shared equipment amount is higher in Scenario I than that in Scenario II ($M_t{}^* > M_t{}^l$).

Proposition 2. When the critical contractor leads the equipment sharing (Scenario I), the shared equipment amount is higher and the equipment rent price is lower than in Scenario II.

By comparing the contractors' gains in both scenarios, we obtain

$$U_a^* - U_a^l = \frac{m_b^2 r^2(3a_1^2 m_b^4 + 2a_1 b_1 m_a^2 m_b^2 + b_1^2 m_a^4)}{16(a_1 m_b^2+2b_1 m_a^2)(2a_1 m_b^2+b_1 m_a^2)^2}, \tag{28}$$

$$U_b^* - U_b^l = -\frac{m_b^2 r^2(a_1^2 m_b^4 + 2a_1 b_1 m_a^2 m_b^2 + 3b_1^2 m_a^4)}{16(a_1 m_b^2+2b_1 m_a^2)^2(2a_1 m_b^2+b_1 m_a^2)}. \tag{29}$$

Equations 28 and **29** indicate that $U_a{}^* > U_a{}^l$ and $U_b{}^* < U_b{}^l$.

Proposition 3: When the non-critical contractor leads the equipment sharing, the critical contractor's gain will decrease, and the non-critical contractor's gain will increase.

The difference in the owner's gains in both scenarios is as below:

$$U_o^* - U_o^l = \frac{m_b^2 r^2(a_1 m_b^2 - b_1 m_a^2)}{8(a_1 m_b^2+2b_1 m_a^2)(2a_1 m_b^2+b_1 m_a^2)}. \tag{30}$$

TABLE 3 | Equilibrium values in Scenario II.

Parameter	Equilibrium value
P_l	$\frac{m_b r(a_1 m_a m_b+a_1 m_b^2+b_1 m_a^2)}{2m_a(m_b-m_a)(2a_1 m_b^2+b_1 m_a^2)}$
M_{tl}	$\frac{m_a m_b^2 r}{4(2a_1 m_b^2+b_1 m_a^2)}$
I_{ol}	$\frac{m_b r}{2(m_b-m_a)}$
U_{al}	$\frac{a_1 m_b^4 r^2}{16(2a_1 m_b^2+b_1 m_a^2)^2}-a_0$
U_{bl}	$\frac{m_b^2 r^2}{16(2a_1 m_b^2+b_1 m_a^2)}-b_0$
U_{ol}	$\frac{m_b^2 r^2}{8(2a_1 m_b^2+b_1 m_a^2)}+(d_c-d_{a1})r$

TABLE 4 | Comparison of equilibrium values in Scenarios I and II.

Parameter	Value
$P_{*} - P_{l}$	$-\frac{(a_1^2 m_b^4 + a_1 b_1 m_a^2 m_b^2 + b_1^2 m_a^4) r}{2m_a (2a_1 m_b^2 + b_1 m_a^2)(a_1 m_b^2 + 2b_1 m_a^2)}$
$M_{t^*} - M_{tl}$	$\frac{m_a m_b^2 r (a_1 m_b^2 - b_1 m_a^2)}{4 (a_1 m_b^2 + 2b_1 m_a^2)(2a_1 m_b^2 + b_1 m_a^2)}$
$I_{o^*} - I_{ol}$	0
$U_{a^*} - U_{al}$	$\frac{m_b^2 r^2 (3a_1^2 m_b^4 + 2a_1 b_1 m_a^2 m_b^2 + b_1^2 m_a^4)}{16 (a_1 m_b^2 + 2b_1 m_a^2)(2a_1 m_b^2 + b_1 m_a^2)^2}$
$U_{b^*} - U_{bl}$	$-\frac{m_b^2 r^2 (a_1^2 m_b^4 + 2a_1 b_1 m_a^2 m_b^2 + 3b_1^2 m_a^4)}{16 (a_1 m_b^2 + 2b_1 m_a^2)^2 (2a_1 m_b^2 + b_1 m_a^2)}$
$U_{o^*} - U_{ol}$	$\frac{m_b^2 r^2 (a_1 m_b^2 - b_1 m_a^2)}{8 (a_1 m_b^2 + 2b_1 m_a^2)(2a_1 m_b^2 + b_1 m_a^2)}$

TABLE 5 | Comparison of I_o and U_o in Scenarios I–III.

Parameters	Values
$I_{os^*} - I_{o^*}$	$\frac{m_b r}{(m_b - m_a)} \left(\frac{b_1 m_a^2}{a_1 m_b^2 + b_1 m_a^2} + \frac{1}{2}\right)$
$I_{osl} - I_{ol}$	$\frac{m_b r}{(m_b - m_a)} \left(\frac{a_1 m_b^2}{a_1 m_b^2 + b_1 m_a^2} + \frac{1}{2}\right)$
$U_{os^*} - U_{o^*}$	$-\frac{m_b^2 r^2 (a_1^2 m_b^4 + 2a_1 b_1 m_a^2 m_b^2 + 3b_1^2 m_a^4)}{16 (a_1 m_b^2 + 2b_1 m_a^2)^2 (2a_1 m_b^2 + b_1 m_a^2)}$
$U_{osl} - U_{ol}$	$-\frac{(9a_1^2 m_b^4 + 6a_1 b_1 m_a^2 m_b^2 + b_1^2 m_a^4) m_b^2 r^2}{8 (2a_1 m_b^2 + b_1 m_a^2)(a_1 m_b^2 + b_1 m_a^2)^2}$

Proposition 4. The owner's gain is higher in Scenario I than that in Scenario II.

From Proposition 2–4, it can be concluded that when the noncritical contractor leads the equipment sharing, the critical contractor has to pay more rent but will acquire fewer equipment. Since the noncritical contractor takes the privilege of setting the equipment rent price, he or she will select the price that benefits merely himself or herself. The critical contractor has no choice but to follow the noncritical contractor's decision. As a result, through offering equipment and receiving rent, the noncritical contractor actually obtains most of the construction duration reward belonging to the critical contractor. When the reward/penalty coefficient remains the same, fewer equipment will be shared in Scenario II. Consequently, the reduction of the construction duration becomes limited, and the owner's operation income falls.

A comparison of the values obtained in the two scenarios is presented in **Table 4**.

5 REVENUE-SHARING CONTRACT DESIGN

5.1 Construction Engineering System Optimal Solution (Scenario III)

In construction engineering, as owners and contractors are subjects who aim to optimize their self-interest, their resource allocation scheme is not always consistent with the utility of the construction engineering system (Meng and Gallagher, 2012; Kerkhove and Vanhoucke, 2016). In the construction engineering system, if there exists an optimal solution that maximizes the total gain of all equipment-sharing members, obtaining this solution will be the objective of the coordination mechanism (Kuipers et al., 2013). The gain of the system U_{sys} consists of the gains of the owner U_o, critical contractor U_a, and noncritical contractor U_b:

$$
\begin{aligned}
&\max_{M_t} U_{sys} = U_a + U_b + U_o \\
&= [C_a + W_a - P \cdot M_t - C(d_{a2})] + [C_b + W_b + P \cdot M_t - C(d_{b2})] \\
&+ [r(d_c - d_a) - (C_a + W_a + C_b + W_b)] \\
&= r(d_c - d_a) - C(d_{a2}) - C(d_{b2}) \\
&= r \cdot \frac{M_t}{m_a} - \left(\frac{a_1}{m_a^2} M_t^2 + a_0\right) - \left(\frac{b_1}{m_b^2} M_t^2 + b_0\right) \\
&= \left(-\frac{a_1}{m_a^2} - \frac{b_1}{m_b^2}\right) M_t^2 + \frac{r}{m_a} M_t + (d_c - d_{a1}) r - a_0 - b_0
\end{aligned}
. \tag{31}
$$

Equation 31 indicates that the system gain U_{sys} includes only the owner's operation income and the construction costs of both contractors. By substituting **Eqs 3–6** in **Eq. 31**, the simplified gain function of the system shows the system gain U_{sys} is a quadratic function of the shared equipment amount M_t.

Proposition 5: U_{sys} is convex with respect to M_t.

Proof: By calculating the second derivative of system gain, we obtain

$$\frac{d^2 U_{sys}}{d M_t^2} = -\frac{2a_1 m_b^2 + 2b_1 m_a^2}{m_a^2 m_b^2} < 0.$$

The systematic optimal amount of shared equipment is marked as M_t^s. By solving the equation of the first derivative of the system gain function ($\frac{dU_{sys}}{dM_t} = 0$), M_t^s is determined as follows:

$$M_t^s = \frac{m_a m_b^2 r}{2(a_1 m_b^2 + b_1 m_a^2)}. \tag{32}$$

By comparing the optimal shared equipment amounts of Scenarios I ($M_t^* = \frac{m_a m_b^2 r}{4(a_1 m_b^2 + 2b_1 m_a^2)}$), II ($M_t^l = \frac{m_a m_b^2 r}{4(2a_1 m_b^2 + b_1 m_a^2)}$), and III ($M_t^s = \frac{m_a m_b^2 r}{2(a_1 m_b^2 + b_1 m_a^2)}$), we obtain $M_t^s > M_t^*$, $M_t^s > M_t^l$. In other words, the shared equipment amounts in scenarios that maximize individual interest are less than those of the system optimum. Given the above analysis, when contractors reach the equilibrium of the Stackelberg game, the owner's reward/penalty coefficient I_o directly determines shared equipment amount M_t. Thus, the expected system optimal values of I_o in Scenarios I and II can be calculated as follows:

$$I_{os}^* = \frac{m_b (a_1 m_b^2 + 2b_1 m_a^2) r}{(a_1 m_b^2 + b_1 m_a^2)(m_b - m_a)}, \tag{33}$$

$$I_{os}^l = \frac{m_b (2a_1 m_b^2 + b_1 m_a^2) r}{(a_1 m_b^2 + b_1 m_a^2)(m_b - m_a)}. \tag{34}$$

The owner's decisions (I_{os}^* and I_{os}^l) will lead to the owner's gains when the owner selects I_o according to the system optimal solution in Scenarios I and II, respectively.

$$U_{os}^{*} = \left(d_c - d_{a1} - \frac{b_1 m_a^2 m_b^2 r}{2\left(a_1 m_b^2 + b_1 m_a^2\right)^2}\right) r, \tag{35}$$

$$U_{os}^{l} = \left(d_c - d_{a1} - \frac{a_1 m_b^4 r}{2\left(a_1 m_b^2 + b_1 m_a^2\right)^2}\right) r. \tag{36}$$

A comparison of the owner's decision variable I_o and gain U_o in Scenarios I–III is presented in **Table 5**.

Based on the results in **Table 5**, in Scenarios I and II where individual interests are maximized, the reward/penalty coefficients are less than those in Scenario III where the system gain is optimized ($I_{os}{}^{*} > I_o{}^{*}$ and $I_{os}{}^{l} > I_o{}^{l}$). Additionally, the owner's gains are higher in Scenarios I and II than those in Scenario III. Specifically, to achieve system optimum, the owner should raise the reward/penalty coefficient to a certain value ($I_{os}{}^{*}$ and $I_{os}{}^{l}$); therefore, the owner would pay more reward to reduce the same construction duration. In contrast, contractors will benefit from the owner's system optimal decision. As a consequence, the owner will retain his or her individual interest-maximizing decisions ($I_o{}^{*}$ and $I_o{}^{l}$) instead of the system optimal decisions ($I_{os}{}^{*}$ and $I_{os}{}^{l}$).

According to the above analysis, for the sake of the construction engineering system optimum, the owner and contractors should sign revenue-sharing contracts, which fairly distribute operational income brought by equipment sharing so that the owner can willingly raise the reward/penalty coefficient according to the system optimum.

5.2 Revenue-Sharing Contracts

To limit the members' selfish behaviors during equipment sharing, a coordination mechanism based on revenue-sharing contracting is designed to allocate cooperative system gain (Wang et al., 2004). The contracts must comply with two requirements:

(1) The members' gain must not be lower than the gain before the revenue-sharing contract is signed.
(2) The contracts must ensure only the system optimal solution can bring the members the highest gain so that the members will not deviate from the system optimal solution. The revenue-sharing contracts are presented below:

When the critical contractor leads the equipment sharing (Scenario I), the critical contractor's gain function under the contract is $\overline{U_a} = U_a + T_a$, the noncritical contractor's gain function under the contract is $\overline{U_b} = U_b + T_b$, and the owner's gain function under the contract is $\overline{U_o} = U_o + T_o$.

When the non-critical contractor leads the equipment sharing (Scenario II), the critical contractor's gain function under the contract is $\widetilde{U_a} = U_a + T_a$, the noncritical contractor's gain function under the contract is $\widetilde{U_b} = U_b + T_b$, and the owner's gain function under the contract is $\widetilde{U_o} = U_o + T_o$.

In both Scenarios I and II, the efficiency principle should be satisfied: $T_a + T_b + T_o = 0$; thus, the total gain of the system optimum is fully divided among the members.

The authors adopt a revenue-sharing contract based on a transferable subsidy (Cachon and Lariviere, 2005). The form of the contract is $T_i = k_i I_o + w_i$, $i = a, b, o$. The first part of the contract represents every member's extra subsidy (which is linear in the owner's decision variable I_o), and k_i represents the subsidy that each member has to pay when the owner raises/decreases a unit of the reward/penalty coefficient. Furthermore, k_i ensures the derivative of every member's gain function under contract will be 0 when the member makes a system optimal decision. For example, $\partial\overline{U_o}/\partial I_o|_{I_o=I_{os}^{*}} = 0$, and $\partial\overline{U_a}/\partial M_t|_{M_t=M_t^{s}} = 0$. Constant w_i allocates the share of system optimal gain for each member. Generally, w_i is determined by multi-round negotiations among contractors and the owner until all members recognize the allocation results.

In this study, to reflect the effect of different bargaining powers of the contractors and the owner on the equipment-sharing game, each member's gain under the contract is assigned a share proportional to the gain that the individual optimal gain represents over the system gain before contracting (Audy et al., 2012; Guajardo and Rönnqvist, 2015), as follows:

$$U_i = \frac{U_i}{\sum_{i \in a,b,o} U_i} \cdot U_{sys}\left(I_{os}^{*}\right), \tag{37}$$

$$U_i^{\sim} = \frac{U_i}{\sum_{i \in a,b,o} U_i} \cdot U_{sys}\left(I_{os}^{l}\right). \tag{38}$$

Proposition 6: When the critical contractor leads the equipment sharing (Scenario I), under the contracts $T_a = -k_o I_o + w_a$, $T_b = w_b$, and $T_o = k_o I_o - w_a - w_b$, the owner will select the system optimal reward/penalty coefficient I_{os}^{*} and $k_o = \frac{1}{2}\frac{(a_1 m_b^2 + 3b_1 m_a^2)(m_b - m_a)m_b r}{(a_1 m_b^2 + b_1 m_a^2)(a_1 m_b^2 + 2b_1 m_a^2)}$.

Proof: According to **Eq. 37**, the critical contractor's gain function under the contract is as follows:

$$\overline{U_a} = U_a + T_a = C_a + W_a - P \cdot M_t - C(d_{a2}) - k_o \cdot I_o + w_a.$$

The non-critical contractor's gain function under the contract is as follows:

$$\overline{U_b} = \mathrm{U}_b + T_b = C_b - W_b + P \cdot M_t - C(d_{b2}) + w_b.$$

By conducting a backward induction, after the contracts are signed, the critical contractor's optimal rent price will be

$$P^{*} = \frac{(b_1 m_a m_b + a_1 m_b^2 + b_1 m_a^2) I_o}{m_b\left(a_1 m_b^2 + 2b_1 m_a^2\right)}.$$

The noncritical contractor's optimal shared equipment amount will be

$$M_t^{*} = \frac{1}{2}\frac{m_a m_b (m_b - m_a) I_o}{a_1 m_b^2 + 2b_1 m_a^2}.$$

The owner's gain function under the contract is as follows:

$$\begin{aligned}\overline{U_o} &= \mathrm{U}_o + T_o \\ &= r(d_c - d_a) - (C_a + I_o(d_{a1} - d_{a2}) + C_b + I_o(d_{b1} - d_{b2})) - k_o \\ &\quad \cdot I_o - w_a - w_b.\end{aligned}$$

Since the second derivative of the gain function is negative ($\frac{d^2\overline{U_o}}{dI_o^2} = -\frac{(m_b - m_a)^2}{a_1 m_b^2 + 2b_1 m_a^2} < 0$), the owner has an optimal solution. By solving $\frac{d\overline{U_o}}{dI_o} = 0$, the owner's optimal reward/penalty coefficient is as follows:

$$I_o = \frac{1}{2}\frac{2a_1k_om_b^2 + 4b_1k_om_a^2 - m_am_br + m_b^2r}{(m_a - m_b)^2}.$$

According to the contract, as $k_o = \frac{1}{2}\frac{(a_1m_b^2+3b_1m_a^2)(m_b-m_a)m_br}{(a_1m_b^2+b_1m_a^2)(a_1m_b^2+2b_1m_a^2)}$, the optimal reward/penalty coefficient will be $I_o = \frac{m_b(a_1m_b^2+2b_1m_a^2)r}{(a_1m_b^2+b_1m_a^2)(m_b-m_a)}$, which is equal to the system optimal reward/penalty coefficient I_{os}^*.

Based on **Eq. 37**, constants w_a and w_b are determined as follows:

$$w_a = \frac{r^2m_b^2(2a_1^2m_b^4 + 11a_1b_1m_a^2m_b^2 + 15b_1^2m_a^4)}{2(3a_1m_b^2 + 7b_1m_a^2)(a_1m_b^2 + b_1m_a^2)^2}, \quad (39)$$

$$w_b = -\frac{m_a^2b_1m_b^2r^2(a_1m_b^2 + 3b_1m_a^2)}{2(3a_1m_b^2 + 7b_1m_a^2)(a_1m_b^2 + b_1m_a^2)^2}. \quad (40)$$

Proposition 7: When the noncritical contractor leads the equipment sharing (Scenario II), under the contracts $T_a = -j_oI_o + w_a$, $T_b = w_b$, and $T_o = j_oI_o - w_a - w_b$, the owner will select the system optimal reward/penalty coefficient I_{os}^l and $j_o = \frac{1}{2}\frac{r(a_1m_b^2+3b_1m_a^2)m_b(m_b-m_a)}{(a_1m_b^2+b_1m_a^2)(2a_1m_b^2+b_1m_a^2)}$.

The proof of Proposition 7 is similar to the proof of Proposition 6. Based on **Eq. 42**, constants w_a and w_b are determined as follows:

$$w_a = \frac{3m_b^2r^2(6a_1^2m_b^4 + 5a_1b_1m_a^2m_b^2 + b_1^2m_a^4)}{2(7a_1m_b^2 + 3b_1m_a^2)(a_1m_b^2 + b_1m_a^2)^2}, \quad (41)$$

$$w_b = -\frac{m_b^2r^2(6a_1^2m_b^4 + 5a_1b_1m_a^2m_b^2 + b_1^2m_a^4)}{2(7a_1m_b^2 + 3b_1m_a^2)(a_1m_b^2 + b_1m_a^2)^2}. \quad (42)$$

Constants w_a and w_b are applied to ensure the owner and contractors will receive the same gain proportion in system gain before contracting. In addition, the importance of subsidy coefficients k_o and j_o is that members obtain the maximized system gain share only when each member makes a decision that complies with the system optimum instead of the individual optimum.

6 NUMERICAL STUDY

For a better illustration of the effect of the developed revenue-sharing contracts, this article used a numerical example with two different scenarios. Consider a large construction project in which two contractors undertake construction tasks in critical and noncritical paths, respectively, and the owner applies a duration reward/penalty to both contractors to control the project schedule. During the construction period, the fixed contract price of a critical contractor is $50 million, and the contract construction duration of the project is 300 days. Additionally, the fixed contract price of a noncritical contractor is $10 million, and the contract construction duration of the project is 200 days. Before the contractors start equipment sharing, the critical contractor owns 10 pieces of a certain kind of heavy machine, and the noncritical contractor owns 22 pieces of an identical machine. In the construction phase of the projects, contractors are encouraged to spontaneously share heavy machines among themselves to save construction costs and to be rewarded for early completion (or to avoid a penalty). The values of other parameters in this section are based on the authors' investigations of two Chinese large-scale infrastructure projects (the Xiangjiaba hydropower station and the Hong Kong-Zhuhai-Macao bridge). The logical relationship of the parameters is based on the analysis of the functional relationship between construction cost and time (Callahan et al., 1992; Shr and Chen, 2006). The parameters are listed in **Table 6**.

TABLE 6 | Parameters of the numerical study.

Parameter	Value
a_1	4
a_0	50
b_1	4
b_0	10
d_c	2000
R	150

6.1 Revenue-Sharing Contracts in Scenario I

The contracts can be obtained according to **Section 5**:

$$T_a = -9.7I_o + 2134.6, T_b = -145.4, T_o = 9.7I_o - 1989.2.$$

Figure 2 shows the owner's gain functions before and after revenue-sharing contracts are signed. The owner will raise his reward/penalty coefficient from $I_o^* = 137.5$ (**Eq. 22**) before contracting to $I_{os}^* = 322.1$ (**Eq. 33**) after contracting. Moreover, the owner's gain rises from $U_o = 497.53$ (**Eq. 23**) before contracting to $\overline{U_o} = 740.86$ (**Eq. 37**) after contracting. **Figure 3** presents the critical contractor's gain functions before and after the contracts are signed. The critical contractor will increase his optimal equipment rent price from $P^* = 7.35$ (**Eqs 14**, **22**) before contracting to $P_s^* = 17.21$ (**Eqs 14**, **33**) after contracting. The critical contractor's gain rises from $U_a = 248.77$ (**Eq. 17**) before contracting to $\overline{U_a} = 370.43$ (**Eq. 37**) after contracting. **Figure 4** shows the noncritical contractor's gain functions before and after the contracts are signed. The noncritical contractor selects the shared equipment amount $M_t^* = 66.34$ (**Eqs 15**, **22**) and then chooses to raise the amount to $M_t^s = 155.39$ (**Eqs 15**, **33**) after the revenue-sharing contracts are signed. Additionally, his gain rises from $U_b = 36.37$ (**Eq. 18**) before contracting to $\overline{U_b} = 54.16$ (**Eq. 37**) after contracting.

Table 7 shows the comparisons of the construction system members' gains and the construction durations in Scenario I when the coordination mechanism is implemented. When revenue-sharing contracts are put into effect, the construction system gain increases from $U_{sys} = 782.67$ (**Eq. 31**) to $U_{sys} = 1,165.45$ (**Eq. 31**). In addition, the construction duration is reduced by 8.9 days. The proportions of members' gains to system gain remain the same before and after contracting. In other words, the system gain allocation rules

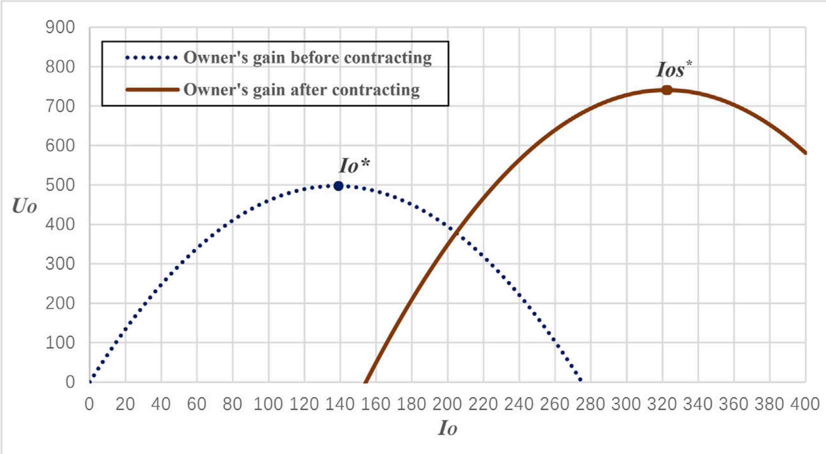


FIGURE 2 | The owner's gain function before and after contracting in Scenario I.

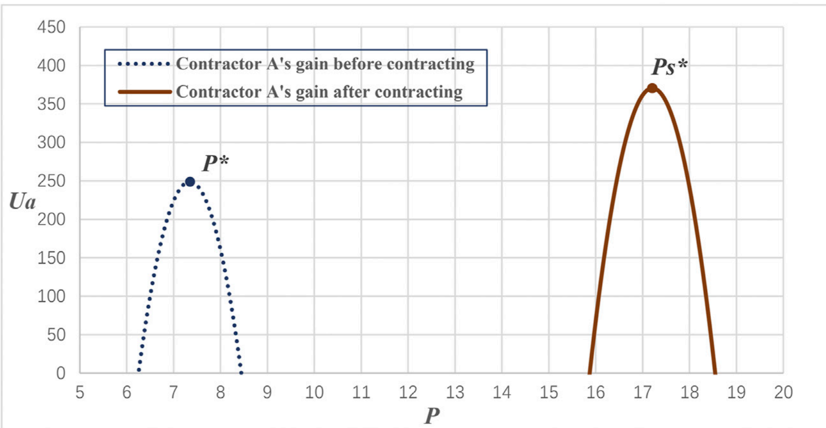


FIGURE 3 | The critical contractor's gain function before and after contracting in Scenario I.

of the contracts are based on the members' individual optimal gains as determined by the bargaining power of the members during equipment sharing.

6.2 Revenue-Sharing Contracts in Scenario II

The contracts are obtained according to the coordination mechanism design section:

$$T_a = -12.3I_o + 5381.5, T_b = -1793.8, T_o = 12.3I_o - 3587.7.$$

Figure 5 shows the owner's gain functions before and after revenue-sharing contracts are signed. The owner will raise his reward/penalty coefficient from I_o^l = 137.5 (**Table 3**) before contracting to I_{os}^l = 502.9 (**Eq. 34**) after contracting. The owner's gain rises from U_o = 318.64 (**Table 3**) before contracting to $\widetilde{U_o}$ = 675.0 (**Eq. 38**) after contracting. **Figure 6** presents the noncritical contractor's gain functions before and after the contracts are signed. The noncritical contractor will increase his optimal equipment rent price from P^l = 10.35 (**Table 3**) before contracting to P_s^l = 37.86 (**Table 3**, **Eq. 34**) after contracting. The noncritical contractor's gain rises from U_b = 159.32 (**Table 3**) before contracting to $\widetilde{U_b}$ = 337.5 (**Eq. 38**) after contracting. **Figure 7** shows the critical contractor's gain functions before and after the contracts are signed. The critical contractor selects the shared equipment amount M_t^l = 42.49 (**Table 3**) and then chooses to raise the amount to M_t^s = 155.39 (**Table 3**, **Eq. 34**) after the revenue-sharing contracts are signed. Additionally, his gain rises from U_a = 72.20 (**Table 3**) before contracting to $\widetilde{U_a}$ = 152.95 (**Eq. 38**) after contracting.

Table 8 shows the comparisons of the construction system members' gains and the construction durations in Scenario II. When revenue-sharing contracts are implemented, the construction system gain increases from U_{sys} = 550.16 (**Eq. 31**) to U_{sys} = 1,165.45 (**Eq. 31**). In addition, the construction duration is reduced by 11.29 days.

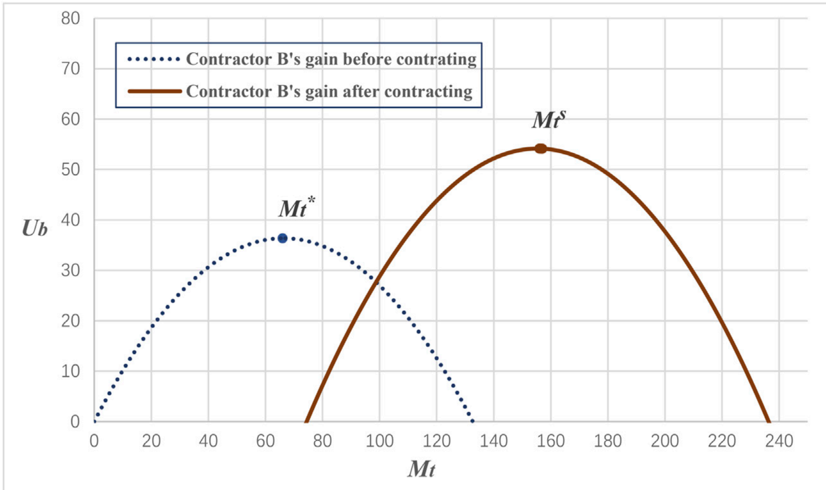


FIGURE 4 | The noncritical contractor's gain function before and after contracting in Scenario I.

TABLE 7 | Comparisons of gain functions and construction durations in Scenario I.

	Critical contractor's gain	Noncritical contractor's gain	Owner's gain	Construction system gain	Construction duration reduction
Before contracting	248.77 (31.78%)	36.37 (4.65%)	497.53 (63.57%)	782.67	6.63
After contracting	370.43 (31.78%)	54.16 (4.65%)	740.86 (63.57%)	1,165.45	15.54
Difference value	121.66	17.79	243.33	382.78	8.91

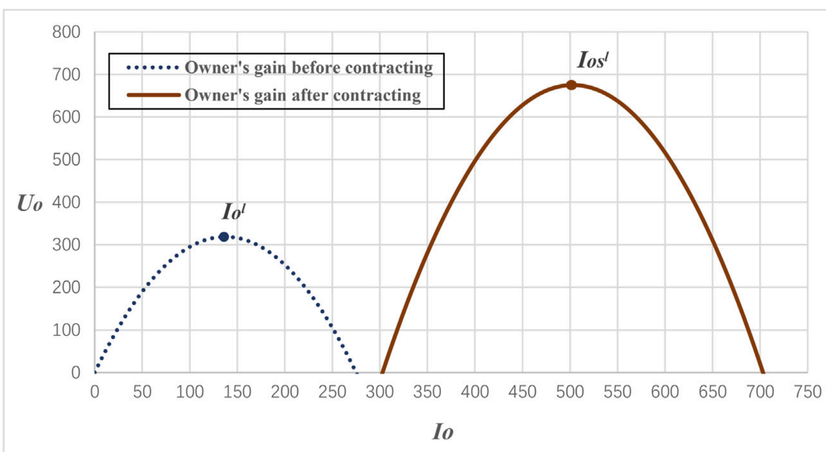


FIGURE 5 | The owner's gain function before and after contracting in Scenario II.

6.3 Conclusion of Numerical Study

By analyzing the results of **Section 6**, the authors can make three conclusions as follows:

Conclusion 1: In both Scenarios I and II, the revenue-sharing contracts can motivate the owner and contractors to willingly make decisions according to the system optimal solution.

The objective of the revenue-sharing contracts is to compensate the owner by reallocating system gain so that the owner can select the system optimal reward/penalty coefficient I_o, which is the key decision-making variable to achieve system optimum according to the coordination mechanism design section. Under the contracts, to make the owner raise the reward/penalty coefficient, both the

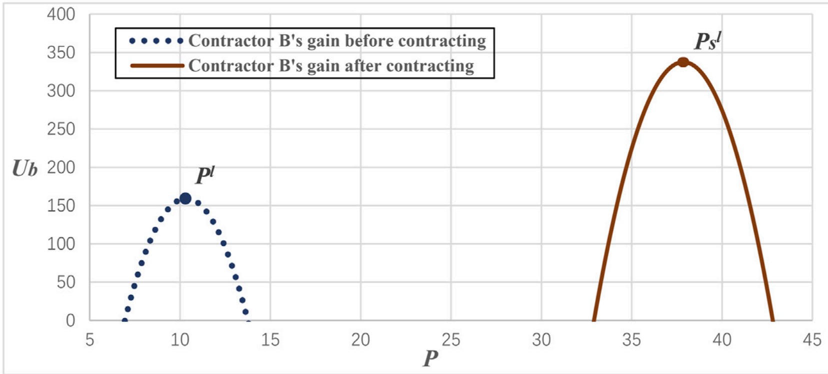


FIGURE 6 | The noncritical contractor's gain function before and after contracting in Scenario II.

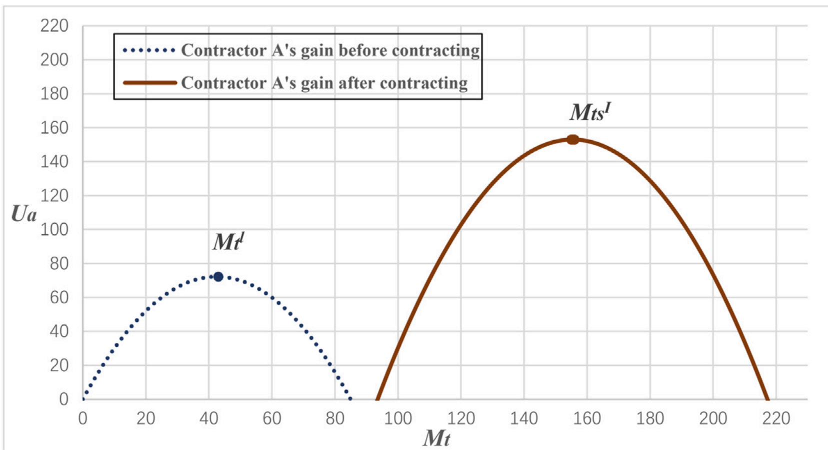


FIGURE 7 | The critical contractor's gain function before and after contracting in Scenario II.

TABLE 8 | Comparisons of the gain functions and construction durations in Scenario II.

	Critical contractor's gain	Noncritical contractor's gain	Owner's gain	Construction system gain	Duration compression
Before contracting	72.20 (13.12%)	159.32 (28.96%)	318.64 (57.92%)	550.16	4.25
After contracting	152.95 (13.12%)	337.50 (28.96%)	675.0 (57.92%)	1,165.45	15.54
Difference value	80.75	178.18	356.36	615.29	11.29

critical and noncritical contractors should share certain parts of their gains as subsidies with the owner.

Conclusion 2: In both Scenarios I and II, the gain of the equipment-sharing leader is higher than the gain of the follower.

According to the leader–follower game models of equipment sharing, the leader's gain is always higher than the follower's gain. The stronger bargaining power (the critical contractor's influence on duration and the noncritical contractor's equipment ownership), which is possessed by the leader, gives the leader a first-move advantage. The leader can always make a decision to benefit himself more than the follower, while the follower can react only after the leader has decided the rent price.

Conclusion 3: Before contracting, both the construction system gain and duration compression are higher in Scenario I than in Scenario II.

When the noncritical contractor (instead of the critical contractor) leads the equipment sharing, the critical contractor needs to pay more rent to acquire the same equipment amount. As a result, the cost of equipment sharing rises, and the amount of shared equipment drops. Furthermore, the construction duration decreases; this decrease will reduce the system income in the project operation period.

Based on the above analysis and conclusions, several managerial implications are provided for construction engineering practitioners and academics:

(1) Equipment is shared only when the critical contractor's equipment rent is sufficient, which fairly compensates the noncritical contractor's delay penalty and construction cost.
(2) To acquire sufficient equipment and income, the critical contractor needs to avoid sharing equipment with a noncritical contractor who has more bargaining power. Additionally, the critical contractor can cooperate with more than one noncritical contractor to strengthen his leadership.
(3) To ensure the sustainability of equipment sharing cooperation, the owner can raise the reward/penalty coefficient of the critical contractor (thereby creating more reward to rent equipment) or reduce the reward/penalty coefficient of the non-critical contractor (thereby lowering the rent price), since the owner can profit from the project's operation income with more shared-equipment.
(4) While designing the construction contracts, the keynote is to find the decision-making variable that directly determines the system gain by analyzing interactions among the contractors and the owner. Then, through subsidies to the key decision-making variables, such as time incentives and shared-equipment amount, the construction engineering members will comply with the system optimal solution.

CONCLUSION AND DISCUSSION

In this research, the problem of construction equipment sharing between contractors in leadership and non-leadership roles during the construction period is studied. Based on analyzing the effect of the duration reward/penalty on equipment-sharing productivity and operation income for the owner brought by duration compression, two trilateral equipment-sharing game models of the owner, critical contractor, and noncritical contractor are developed. To maximize system gain and reduce duration, revenue-sharing contracts based on static transferable utility are devised to motivate equipment-sharing members to make decisions according to the system optimal solution. The numerical study demonstrates that the revenue-sharing contracts can make the owner and contractors willingly select the system optimal solution rather than the individual optimal solution, no matter which contractor leads the equipment sharing. Under the contracts, the gains of all members will increase, and the construction duration is compressed.

This study contributes to the academic research on construction engineering for building a trilateral game model of the decision-making among the owner and contractors. By comparing the effects of different leadership roles and duration reward/penalty on equipment-sharing results, researchers can further study bilateral and trilateral games of construction engineering corporations. However, this article applies revenue-sharing contracting to the coordination problem of construction resource sharing. Concerning resource cooperation between contractors, unlike traditional studies that focus on cooperative gain allocation rules design, the authors focus on contracting, which can diminish members' selfish deviation from the system optimal solution. The contract design pattern in this article can help project management better utilize incentive contracting and select resource-sharing partners.

During contract designing, since equipment-sharing members are independent from each other and make decisions on their own, the composition of each member's gain function needs in-depth study; the elements needing in-depth study include construction cost, duration reward/penalty, operation income, and equipment rent. Furthermore, when the system optimal solution is determined, how to assign system gain fairly and efficiently to members is also important to the availability of the signed contracts. In this article, to better reflect the equipment-sharing leadership and bargaining power of each member, the authors adopt a proportional allocation method in which every member's proportionate share of system gain under the contracts stays the same as the share before contracting. However, in construction engineering practice, both the gain compositions and leaderships of contractors are more complex. Thus, system gain allocation rules based on the construction onsite situation that can promptly express the needs of construction firms should be studied in further research.

The authors consider the equipment-sharing problem while assuming the critical path of construction activities will not change during equipment sharing. However, in practice, equipment sharing may cause several changes to the critical path, such as a critical path turns to a noncritical path and one critical path grows to two or more critical paths. When the critical path changes, contractors who need extra equipment should not only carefully select equipment suppliers but also consider the following additional construction cost and potential equipment shortage.

As the objective of equipment sharing is to reduce project delay, when the noncritical path becomes a new critical one, more shared equipment will not reduce the construction time of the new critical path but prolong it and induce an unnecessary delay penalty. At this moment, equipment sharing will come to an end. It means that the contractors' game of equipment sharing will end before it reaches equilibrium. In other words, the amount of shared equipment and contractors' gains are determined by the construction schedule (the time difference between the critical path and noncritical path), instead of the leader–follower game equilibrium. As the total equipment sharing gain is fixed before equipment sharing, the bargaining game theory provides a systematic and analytical framework of how to allocate the gains between the contractors, which mainly focus on price negotiation, revenue and risk allocation, and concession period negotiation in the construction field (Nash, 1950; Rubinstein, 1982). In the next research, the authors will apply the bargaining game theory to the problem of equipment sharing gain allocation.

In addition, equipment-sharing models can be extended to more general forms, such as one critical contractor sharing with two or more non-critical contractors or multiple contractors sharing with each other to meet their specific construction requirements, thereby leading to more complicated game models. Finally, this study assumes that game information is symmetric to all members. While in an unsymmetrical information situation, contractors and the owner may reach other kinds of equilibria, such as a Bayesian Nash equilibrium. Then, the contracts in this article need to be redesigned.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

ZL conducted scientific research, performed numerical study, and wrote the paper. HW and YX reviewed and edited the manuscript drafts. All authors were involved in conceiving the study, data analysis, and structure of the paper.

FUNDING

This work has been supported by the National Natural Science Foundation of China (Nos. 71821001 and 71390524).

ACKNOWLEDGMENTS

The authors thank the editors and reviewers for their beneficial comments and suggestions, which have helped them improve this article.

REFERENCES

Akhbari, M. (2020). Contractors' Partnership in Project Resource Management Application of Cooperative Game Theory Approach. *Sci. Iran.* 27 (1), 469–480.

Asgari, S., Afshar, A., and Madani, K. (2014). Cooperative Game Theoretic Framework for Joint Resource Management in Construction. *J. Constr. Eng. Manage.* 140 (3), 04013066. doi:10.1061/(asce)co.1943-7862.0000818

Audy, J.-F., D'Amours, S., and Rönnqvist, M. (2012). An Empirical Study on Coalition Formation and Cost/Savings Allocation. *Int. J. Prod. Econ.* 136 (1), 13–27. doi:10.1016/j.ijpe.2011.08.027

Bendoly, E., Perry-Smith, J. E., and Bachrach, D. G. (2010). The Perception of Difficulty in Project-Work Planning and its Impact on Resource Sharing. *J. Operations Manag.* 28 (5), 385–397. doi:10.1016/j.jom.2009.11.011

Cachon, G. P., and Lariviere, M. A. (2005). Supply Chain Coordination with Revenue-Sharing Contracts: Strengths and Limitations. *Manag. Sci.* 51 (1), 30–44. doi:10.1287/mnsc.1040.0215

Cachon, G. P., and Zipkin, P. H. (1999). Competitive and Cooperative Inventory Policies in a Two-Stage Supply Chain. *Manag. Sci.* 45 (7), 936–953. doi:10.1287/mnsc.45.7.936

Callahan, M. T., Quackenbush, D. G., and Rowings, J. E. (1992). *Construction Project Scheduling*. New York: McGraw-Hill.

Chen, J. H., and Ma, S. H. (2007). A Revenue Incentive Model of Project Duration Coordination between Project Corporation and Contractor. *Chin. J. Manag. Sci.* 15 (3), 114–122.

Confessore, G., Giordani, S., and Rismondo, S. (2007). A Market-Based Multi-Agent System Model for Decentralized Multi-Project Scheduling. *Ann. Oper. Res.* 150 (1), 115–135. doi:10.1007/s10479-006-0158-9

Eissa, R., Eid, M. S., and Elbeltagi, E. (2021b). Conceptual Profit Allocation Framework for Construction Joint Ventures: Shapley Value Approach. *J. Manage. Eng.* 37 (3), 04021016. doi:10.1061/(asce)me.1943-5479.0000911

Eissa, R., Eid, M. S., and Elbeltagi, E. (2021a). Current Applications of Game Theory in Construction Engineering and Management Research: A Social Network Analysis Approach. *J. Constr. Eng. Manage.* 147 (7), 04021066. doi:10.1061/(asce)co.1943-7862.0002085

Feldman, M., and Tamir, T. (2012). Conflicting Congestion Effects in Resource Allocation Games. *Operations Res.* 60 (3), 529–540. doi:10.1287/opre.1120.1051

Fink, A. (2006). *Supply Chain Coordination by Means of Automated Negotiations between Autonomous agentsMultiagent Based Supply Chain Management*. Berlin: Springer, 351–372. doi:10.1007/978-3-540-33876-5_13

Gkatzelis, V., Kollias, K., and Roughgarden, T. (2016). Optimal Cost-Sharing in General Resource Selection Games. *Operations Res.* 64 (6), 1230–1238. doi:10.1287/opre.2016.1512

Gransberg, D. D., Popescu, C. M., and Ryan, R. C. (2006). *Construction Equipment Management for Engineers, Estimators, and Owners*. London: Taylor & Francis.

Guajardo, M., and Rönnqvist, M. (2015). Cost Allocation in Inventory Pools of Spare Parts with Service-Differentiated Demand Classes. *Int. J. Prod. Res.* 53 (1), 220–237. doi:10.1080/00207543.2014.948577

Hafezalkotob, A., Hosseinpour, E., Moradi, M., and Khalili-Damghani, K. (2018). Multi-resource Trade-Off Problem of the Project Contractors in a Cooperative Environment: Highway Construction Case Study. *Int. J. Manag. Sci. Eng. Manag.* 13 (2), 129–138. doi:10.1080/17509653.2017.1323240

Hartmann, S., and Briskorn, D. (2010). A Survey of Variants and Extensions of the Resource-Constrained Project Scheduling Problem. *Eur. J. Operational Res.* 207 (1), 1–14. doi:10.1016/j.ejor.2009.11.005

Homberger, J. (2007). A Multi-Agent System for the Decentralized Resource-Constrained Multi-Project Scheduling Problem. *Int. Trans. Oper. Res.* 14 (6), 565–589. doi:10.1111/j.1475-3995.2007.00614.x

Homberger, J. (2012). A (μ, λ)-coordination Mechanism for Agent-Based Multi-Project Scheduling. *OR Spectr.* 34 (1), 107–132. doi:10.1007/s00291-009-0178-3

Jaraiedi, M., Plummer, R. W., and Aber, M. S. (1995). Incentive/Disincentive Guidelines for Highway Construction Contracts. *J. Constr. Eng. Manag.* 121 (1), 112–120. doi:10.1061/(asce)0733-9364(1995)121:1(112)

Karaa, F. A., and Nasr, A. Y. (1986). Resource Management in Construction. *J. Constr. Eng. Manag.* 112 (3), 346–357. doi:10.1061/(asce)0733-9364(1986)112:3(346)

Kemahlioglu -Ziya, E., and Bartholdi, J. J. (2011). Centralizing Inventory in Supply Chains by Using Shapley Value to Allocate the Profits. *Manuf. Serv. Operations Manag.* 13 (2), 146–162.

Kerkhove, L. P., and Vanhoucke, M. (2016). Incentive Contract Design for Projects: The Owner's Perspective. *Omega* 62, 93–114. doi:10.1016/j.omega.2015.09.002

Kucuksayacigil, F., and Ulusoy, G. (2020). Hybrid Genetic Algorithm for Bi-objective Resource-Constrained Project Scheduling. *Front. Eng. Manag.* 7 (3), 426–446. doi:10.1007/s42524-020-0100-x

Kuipers, J., Mosquera, M. A., and Zarzuelo, J. M. (2013). Sharing Costs in Highways: A Game Theoretic Approach. *Eur. J. Operational Res.* 228 (1), 158–168. doi:10.1016/j.ejor.2013.01.018

Lau, J. S. K., Huang, G. Q., Mak, K. L., and Liang, L. (2006). Agent-based Modeling of Supply Chains for Distributed Scheduling. *IEEE Trans. Syst. Man. Cybern. A* 36 (5), 847–861. doi:10.1109/tsmca.2005.854231

Lee, Y.-H., Kumara, S. R. T., and Chatterjee, K. (2003). Multiagent Based Dynamic Resource Scheduling for Distributed Multiple Projects Using a Market Mechanism. *J. Intelligent Manuf.* 14 (5), 471–484. doi:10.1023/a:1025753309346

Li, F., Xu, Z., and Li, H. (2021). A Multi-Agent Based Cooperative Approach to Decentralized Multi-Project Scheduling and Resource Allocation. *Comput. Industrial Eng.* 151, 106961. doi:10.1016/j.cie.2020.106961

Liu, Z., Wang, H., and Li, H. (2018). Model of Equipment Sharing between Contractors on Construction Projects. *J. Constr. Eng. Manag.* 144 (6), 04018039. doi:10.1061/(asce)co.1943-7862.0001485

Lv, J. N., Ye, G., Liu, W., Shen, L. Y., and Wang, H. X. (2015). Alternative Model for Determining the Optimal Concession Period in Managing BOT Transportation Projects. *J. Manag. Eng.* 31 (4). doi:10.1061/(asce)me.1943-5479.0000291

Meng, X., and Gallagher, B. (2012). The Impact of Incentive Mechanisms on Project Performance. *Int. J. Proj. Manag.* 30 (3), 352–362. doi:10.1016/j.ijproman.2011.08.006

Moradi, M., Hafezalkotob, A., and Ghezavati, V. (2019). Robust Resource-Constrained Project Scheduling Problem of the Project's Subcontractors in a Cooperative Environment under Uncertainty: Social Complex Construction Case Study. *Comput. Industrial Eng.* 133, 19–28. doi:10.1016/j.cie.2019.04.046

Moufid, M. E., Roy, D., Hennequin, S., and Cortade, T. (2017). Game Theory Model of a Production Resource Sharing Problem: Study of Possible Cheatings. *IFAC-PapersOnline* 50 (1), 10532–10537. doi:10.1016/j.ifacol.2017.08.1301

Nash, J. F. (1950). The Bargaining Problem. *Econometrica* 18 (2), 155–162. doi:10.2307/1907266

Perera, S. (1983). Resource Sharing in Linear Construction. *J. Constr. Eng. Manag.* 109 (1), 102–111. doi:10.1061/(asce)0733-9364(1983)109:1(102)

Rasmusen, E. (2006). *Games and Information: An Introduction to Game Theory*. third edition. New York: Wiley-Blackwell.

Rubinstein, A. (1982). Perfect Equilibrium in a Bargaining Model. *Econometrica* 50 (1), 97–110. doi:10.2307/1912531

Samaddar, S., Nargundkar, S., and Daley, M. (2006). Inter-organizational Information Sharing: The Role of Supply Network Configuration and Partner Goal Congruence. *Eur. J. Operational Res.* 174 (2), 744–765. doi:10.1016/j.ejor.2005.01.059

Shr, J. F., and Chen, W. T. (2003). A Method to Determine Minimum Contract Bids for Incentive Highway Projects. *Int. J. Proj. Manag.* 21 (8), 601–615. doi:10.1016/s0263-7863(02)00056-x

Shr, J. F., and Chen, W. T. (2006). Functional Model of Cost and Time for Highway Construction Projects. *J. Mar. Sci. Technol.* 14 (3), 127–138. doi:10.51400/2709-6998.2066

Villahoz, J. J. L., del Olmo Martínez, R., and Arauzo, A. A. (2010). Combinatorial Auctions for Coordination and Control of Manufacturing MAS: Updating Prices Methods. *Proc. 5th Int. Workshop Soft Comput. Models Industrial Appl. Guimaraes, Portugal* 73, 27–30. doi:10.1007/978-3-642-13161-5_4

Wang, H., Guo, M., and Efstathiou, J. (2004). A Game-Theoretical Cooperative Mechanism Design for a Two-Echelon Decentralized Supply Chain. *Eur. J. Operational Res.* 157 (2), 372–388. doi:10.1016/s0377-2217(03)00233-9

Wang, L., Zhan, D. C., and Nie, L. S. (2014). Multi-project Decentralized Scheduling Problem Solving by Market Mechanism. *Jisuanji Jicheng Zhizao Xit. Integr. Manuf. Syst.* 20 (8), 1969–1979.

Xia, Y., Chen, B., and Kouvelis, P. (2008). Market-based Supply Chain Coordination by Matching Suppliers' Cost Structures with Buyers' Order Profiles. *Manag. Sci.* 54 (11), 1861–1875. doi:10.1287/mnsc.1080.0900

Xu, J., Meng, J., Zeng, Z., Wu, S., and Shen, M. (2013). Resource Sharing-Based Multiobjective Multistage Construction Equipment Allocation under Fuzzy Environment. *J. Constr. Eng. Manage.* 139 (2), 161–173. doi:10.1061/(asce)co.1943-7862.0000593

Xu, J., and Zeng, Z. (2011). Applying Optimal Control Model to Dynamic Equipment Allocation Problem: Case Study of Concrete-Faced Rockfill Dam Construction Project. *J. Constr. Eng. Manage.* 137 (7), 536–550. doi:10.1061/(asce)co.1943-7862.0000325

Xue, X., Wang, Y., Shen, Q., and Yu, X. (2007). Coordination Mechanisms for Construction Supply Chain Management in the Internet Environment. *Int. J. Proj. Manag.* 25 (2), 150–157. doi:10.1016/j.ijproman.2006.09.006

Zhang, J., Zheng, H., He, W., and Huang, W. (2020). West-East Gas Pipeline Project. *Front. Eng. Manag.* 7 (1), 163–167. doi:10.1007/s42524-019-0056-x

Zhang, X., Bao, H., Wang, H., and Skitmore, M. (2016). A Model for Determining the Optimal Project Life Span and Concession Period of BOT Projects. *Int. J. Proj. Manag.* 34 (3), 523–532. doi:10.1016/j.ijproman.2016.01.005

Conflict of Interest: Author ZL was employed by the company China Construction Eighth Engineering Division Corp., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Built Environment

TYPE Original Research
PUBLISHED 29 September 2022
DOI 10.3389/fbuil.2022.972796

Check for updates

OPEN ACCESS

EDITED BY
Joan Ramon Casas,
Universitat Politecnica de Catalunya,
Spain

REVIEWED BY
Pavlo Maruschak,
Ternopil Ivan Pului National Technical
University, Ukraine
Zhoujing Ye,
University of Science and Technology
Beijing, China

*CORRESPONDENCE
Yasutoshi Nomura,
y-nomura@fc.ritsumei.ac.jp

SPECIALTY SECTION
This article was submitted to Bridge
Engineering,
a section of the journal
Frontiers in Built Environment

RECEIVED 19 June 2022
ACCEPTED 12 September 2022
PUBLISHED 29 September 2022

CITATION
Nomura Y, Inoue M and Furuta H (2022),
Evaluation of crack propagation in
concrete bridges from vehicle-
mounted camera images using deep
learning and image processing.
Front. Built Environ. 8:972796.
doi: 10.3389/fbuil.2022.972796

Evaluation of crack propagation in concrete bridges from vehicle-mounted camera images using deep learning and image processing

Yasutoshi Nomura[1]*, Masaya Inoue[1] and Hitoshi Furuta[2]

[1]Department of Civil and Environmental Engineering, Ritsumeikan University, Kusatsu, Japan,
[2]Department of Civil Engineering, Osaka Metropolitan University, Osaka, Japan

In Japan, all bridges should be inspected every 5 years. Usually, the inspection has been performed through the visual evaluation of experienced engineers. However, it requires a lot of load and expense. In order to reduce the inspection work, an attempt is made in this paper to develop a new inspection method using deep learning and image processing technologies. While using the photos obtained by vehicle-mounted camera, the damage states of bridges can be evaluated manually, it still requires a lot of time and load. To save the time and load, deep learning, which is a method of artificial intelligence is introduced. For image processing, it is necessary to utilize such pre-processing techniques as binarization of pictures and morphology treatment. To illustrate the applicability of the method developed here, some experiments are conducted by using the photos of running surface of concrete bridges of a monorail took by vehicle-mounted camera.

KEYWORDS

crack detection, deep learning, crack propagation, concrete bridge, image processing

1 Introduction

In recent years, the structural integrity of social infrastructures, which have become unsafe due to aging, has become a great concern, and how to efficiently and accurately assess their structural integrity is important in developing maintenance plans. In 2015, the Road Act of Japan was revised to require close visual inspections once every 5 years for bridges with spans of 2 m or more. Currently, these bridges are visually inspected by professional engineers. However, in preparation for the shortage of engineers in the future due to the declining population, it is important to develop labor saving system for inspections while ensuring the inspection accuracy of skilled engineers.

In the past few decades, many traditional computer vision technologies, for example, threshold segmentation Cheng et al. (2003); Fujita et al. (2006), Otsu, 1979 (Otsu, 1979), edge detection (Behara et al. (2012); Martin et al. (2004), and seed region growth Huang and Xu, (2006); Li et al. (2011) have been explored in vison-based crack detection. However, the detection accuracy of these methods does not meet the engineering requirements for real applications. Also, as a representative study aiming at labor-saving visual inspection work, a method to create damage maps by detecting and quantifying cracks from images acquired by Unmanned Aerial Vehicles (UAVs) equipped with cameras has been proposed. Nishimura et al. (2012). Have developed a technique for quantifying crack opening width etc. From 3D models constructed by Structure from Motion (SfM) and from a large number of orthoimages through image processing (Nishimura et al., 2012a; Nishimura et al., 2012b). However, since SfM constructs a 3D model using the correlation of luminance distribution among images taken, it is difficult to evaluate the damage including cracks depending on the state of the object surface, such as a structure with uniform luminance distribution due to painting. On the other hand, the recent dramatic improvement in image/pattern recognition accuracy through artificial intelligence and deep learning has led to the development of inspection of social infrastructures. Numerous studies have been conducted on the detection of cracks in concrete for a variety of subjects. Chun and Igo et al. (2015) proposed a method for detecting cracks on concrete surfaces using Random Forest (Chun and Igo, 2015). Random Forest is an algorithm that performs classification according the results of multiple decision trees constructed from randomly selected training data and explanatory variables. Yoshida et al. (2020) proposed a method to detect crack initiation points using a deep learning model from panoramic images of riverbank revetments and to evaluate the actual size of crack detection from orthoimages (Yoshida et al., 2020). However, this method has been applied to only one type of design, and it has been shown that the accuracy of crack detection is greatly reduced for different structures. The uniformity of the structural part of the riverbank revetment also makes it easier for the learning model to recognize the feature patterns in the crack region. Cha et al. (2017) constructed a convolutional neural network that performs two-class classification: “cracked” and “intact” (Cha et al., 2017). The results show that cracks are detected with higher accuracy than the traditional Canny and Sobel edge detection methods Canny, (1986), Kanopoulos et al. (1988). Dung and Anh (2108) proposed an encoder-decoder FCN for segmenting an image of concrete crack into “crack” and “non-crack” pixels (Dung and Anh, 2018). Their proposed architecture was trained end-to-end on a subset of crack images of the same dataset and reached approximately 90% for both the max F1 and AP scores on training, validation, and test sets. However, the system is shown to be susceptible to ambient noise. Ju et al. (2019) developed a crack deep network based on the Faster Region Convolutional Neural Network (Fast-RCNN) (Ju et al., 2019). Results indicated that the developed system achieves the detection mean average precision of higher than 0.90, which outperforms both the original Faster-RCNN and SSD300, but requires intense computational resources. Liu et al. (2019) used U-Net network structure to build a deep learning model for crack detection (Liu et al., 2019). The proposed system was found to be accurate detections, but only accepts 3 × 512 × 512 image input. There is still a long way to go for engineering applications. Also, in recent years, pixel-level crack detection systems have been proposed. Bang et al. (2019) proposed a pixel-level crack detection system based on a deep convolutional encoder-detector network (Bang et al., 2019). Liu et al. (2020) proposed a two-step pavement crack detection and segmentation method based on CNN where the modified YOLOv3 (Redmon and Farhadi, 2018) and U-Net (Ronneberger et al., 2015) were, respectively, used for these two tasks. The results indicated that the precision, recall, and F1 score of proposed pavement crack detection and segmentation system are higher than other state-of the-art methods. Yamane et al. (2019) proposed a pixel-level crack detection method using Convolutional Neural Networks (CNN) and Mask R-CNN (He et al., 2018) which is one of semantic segmentation techniques. Meanwhile, numerous studies have been also conducted on damage detection in steel structures. Konovalenko et al. (2021) developed a classification system based on residual neural network for damage of metal surfaces such as scratches, scrapes and abrasions (Konovalenko et al., 2021). The results shows that the best model classifier is based on ResNet152 deep convolutional neural network. As for pixel-level defect detection for metal surfaces, Konovalenko et al. (2022) developed and researched 14 neural network models (Konovalenko et al., 2022). U-Net-like architectures with encoders such as ResNet (He et al., 2015), DenseNet (Huang et al., 2018) and so on were investigated as part of the problem, which consisted in detecting defects such as “scratch abrasion”. Results show that the highest recognition accuracy was attained using the U-Net model with a ResNet152 backbone. Most of the above-mentioned studies have been successful in detecting cracks or some kind of defects with high accuracy; however, they do not describe the aging or propagation of damage. In addition, Deep learning and Machine Learning based image classification systems, which have been applied in the above studies, require more time to detect damage in a single image than detection system because it is necessary to determine whether a crack is present or not. These are time-consuming processes when inspecting large structures, so there is room for improvement in terms of time efficiency. Moreover, deep learning methods require the generation of a large amount of data because accuracy is affected by the training data, which is a challenge in terms of labor-saving training data generation.

In this study, we attempt to develop a system that can detect damage on structural surfaces with high accuracy and high speed, and also a system that automatically outputs not only the damage initiation area but also progression area, by utilizing the visual inspection results accumulated by engineers. We also attempt to semi-automatically convert the accumulated visual inspection results into YOLO training data in order to save labor in the creation of training data. The proposed system is a multi-stage system that applies object recognition technology using convolutional neural networks (CNN) and a morphology algorithm that detects damage in pixel units based on the characteristics of image brightness values, in addition to detection technology using

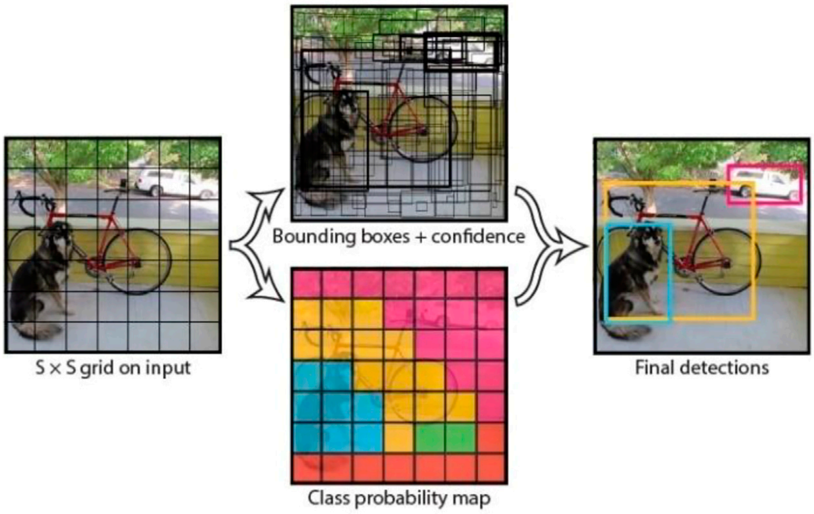


FIGURE 1
Object detection by YOLO (Redmon et al., 2016).

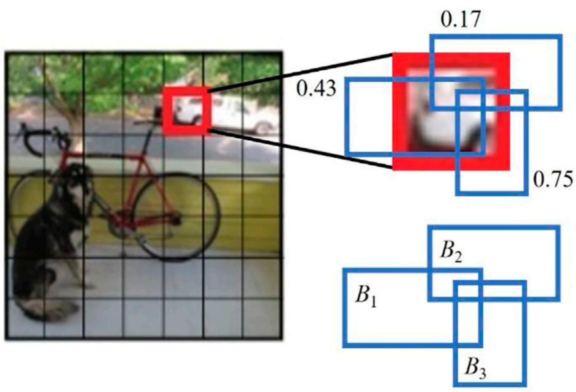


FIGURE 2
Bounding box and Confidence.

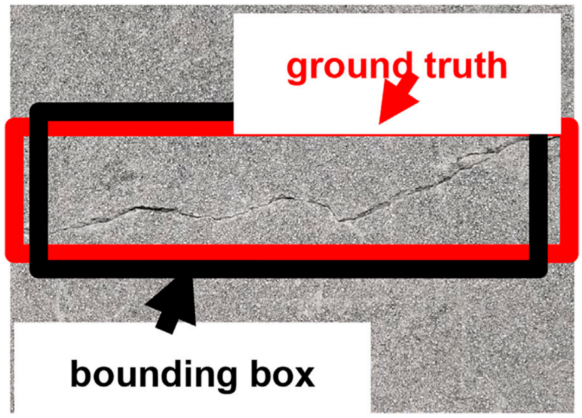


FIGURE 3
Ground_truth and bounding box in IOU.

YOLO, which can detect objects in moving images. To illustrate the applicability of the method developed here, some experiments are conducted by using the photos of running surface of concrete bridges of a monorail took by vehicle-mounted camera.

2 System flowchart

This study uses images taken from an on-board camera to inspect the concrete surface of a 40-km monorail bridge, the running surface. The main types of damage are cracks and free lime. The training data is created based on the images of inspection results that have been accumulated by the administrator. The system proposed in this study can be divided into five stages of processing. The first stage of processing uses YOLOv2 (Redmon and Farhadi, 2016), a common object detection algorithm, to detect damaged areas. The data applied in this study deals with a large set of images, which is about 100,000 images. Therefore, it is appropriate to use YOLOv2 for damage detection because of its high detection speed. The images handled in the second and subsequent stages are more limited than those detected in the first stage, and this data limitation can significantly reduce processing time.

In the second stage of processing, the area where damage was detected by YOLOv2 is first trimmed. Next, images of the damaged area are collected, and the damaged area is carefully inspected using the network model VGG16 (Simonyan and Zisserman, 2015), because damage detection by YOLOv2 alone may result in over-detection. The third stage of processing uses a morphological algorithm to extract pixel-by-pixel cracks from the damage area examined in the second stage of processing. Pixel-by-pixel damage detection allows for more accurate application to the fifth step, progression evaluation processing. Morphological image processing includes color correction, median filtering, line enhancement, and binarization. After morphological processing, the areas of damage are colored with a specified color. In the fourth stage of processing, the binarized damage areas are attached to the original image. This makes it possible for the inspection engineer to identify the location of damage in the original image. In the final stage of processing, the difference in the damaged area is extracted as an evaluation of the progress of damage. Here, the evaluation of the progress is carried out through the comparison with images evaluated in the above processing at the time of the previous inspection is used.

3 Damage detection based on deep learning

3.1 You only look once

The screening technique for structural surface damage in this study is based on general object detection using deep learning, and YOLO (Redmon et al., 2016) is used as a detection technique. The main feature of YOLO for object detection is that it simultaneously extracts a candidate region of a target object and calculates the class probability of that candidate region in a single guess. This makes it possible to perform object detection at very high speed and in real time, even for moving images. YOLO is one of the fastest general object detection systems proposed so far, but the same authors have also proposed YOLOv2 and YOLOv3, which enable object detection with higher accuracy. They, like YOLO, are provided as part of Darknet. In this study, we use YOLOv2, which is the fastest detection method. The inference flow of the YOLO system is shown below.

3.2 Inference procedure

Figure 1 shows the inference procedure for YOLO. The first step is to divide the input image into S × S grid cells. Each grid cell has several bounding boxes, as shown in Figure 2, and the confidence for each box is calculated. Confidence is calculated as follows.

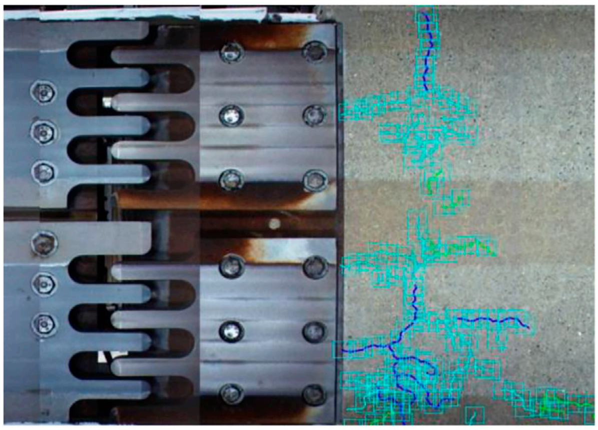


FIGURE 4
An example of annotation data.

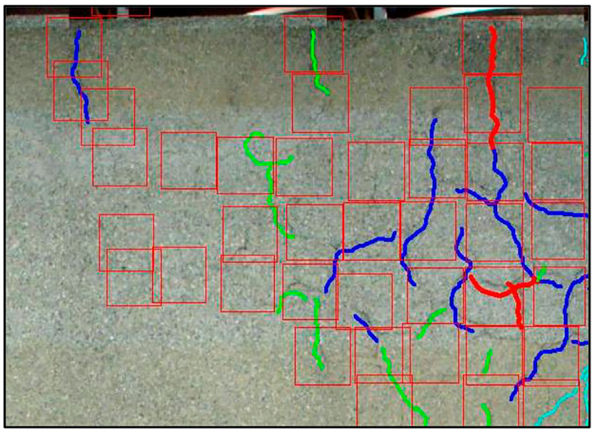


FIGURE 5
An example of detection results.

$$\boldsymbol{Confidence} = \boldsymbol{Pr}(\boldsymbol{Object}) \cdot \boldsymbol{IOU}^{truth}_{pred}$$

$$\boldsymbol{IOU}^{truth}_{pred} = \frac{|B_i \cap ground\_truth|}{|B_i \cup ground\_truth|}$$

where $\boldsymbol{Pr}(\boldsymbol{Object})$ is the probability that the bounding box contains some object, and $\boldsymbol{IOU}^{truth}_{pred}$ (Intersection Over Union) is the proportion of overlap between the bounding box (B_i) predicted from the grid cells and the object (ground_truth) in the image actually given as training data, which is shown in Figure 3. The confidence level is composed of the product of the probability $\boldsymbol{Pr}(\boldsymbol{Object})$ that a bounding box contains an object and the overlap ratio $\boldsymbol{IOU}$ between the bounding box and the ground truth. If there is no object to be detected in the grid cell of interest, then each bounding box has a confidence level of 0, since there is no object in that box. In addition to the confidence level, each bounding box keeps its own center coordinates (*x*, *y*) and size (*height*, *width*) *as* predicted in YOLO.

On the other hand, a grid cell has a conditional class probability ***Pr***(*Class*$_i$ | Object) whether the object to be detected is contained within its own grid cell or not. For example, the grid cell of interest in Figure 1 has ***Pr***(Class = Dog | Object) = 0.05, ***Pr***(Class = Car | Object) = 0.92, ***Pr***(Class = Bicycle | Object) = 0.03, etc. Then, each grid cell outputs two bounding boxes with high confidence, which are expressed in terms of thickness as the bounding box + confidence in Figure 1, depending on the magnitude of the confidence. Finally, only the bounding boxes with confidence exceeding the threshold are adopted, and the class of the bounding box is determined by combining the adopted bounding box and the class with the highest probability in the corresponding grid cell (final detection in Figure 1). As described above, YOLO's general object detection is structured in such a way that class classification is performed for each grid cell region and the bounding box is used to detect candidate object regions. YOLO randomly generates three bounding boxes with a certain aspect ratio in each grid cell.

3.3 How to learn you only look once

YOLO is trained with a convolutional neural network consisting of convolutional layers (conv. layers) and all connected layers (conn. layers), although the number of layers varies with each version. YOLO outputs the location information (x, y, height, width) and confidence level of the two bounding boxes and the category to be detected for each grid cell in the input image. Therefore, the number of units in the output layer is the sum of the number of classes to be detected, the location information of the two bounding boxes, and the confidence level [(x, y, height, width, confidence level) × 2], multiplied by the number of grids S × S. The loss functions for optimizing the filters and biases in the network are shown below.

$$
\begin{aligned}
\textbf{Loss function} = {} & \lambda_{\mathrm{cood}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\mathrm{obj}} \left[(x_i - \hat{x}_i)^2 \right. \\
& \left. + (y_i - \hat{y}_i)^2 \right] + \lambda_{\mathrm{cood}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\mathrm{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 \right. \\
& \left. + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] + \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} (C_i - \hat{c}_i)^2 \\
& + \lambda_{\mathrm{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\mathrm{noobj}} (C_i - \hat{c}_i)^2 \\
& + \sum_{i=0}^{S^2} I_i^{\mathrm{obj}} \sum_{c \in \mathrm{classes}} (p_i(c) - \hat{p}_i(c))^2
\end{aligned}
$$

whereI_{ij}^{obj} is a function that returns 1 if the center coordinate of the *j-th* bounding box is in the *i-th* grid cell and 0 otherwise, andI_i^{obj} returns 1 if the object is in grid cell *i* and 0 otherwise. Note thatλ_{cood} andλ_{noobj} are the program's default settings of 5 and 0.5, respectively, but these values can be set arbitrarily. In this study, the default settings are used. The first term of the loss function is the sum of the predicted center coordinates of the bounding box ($\hat{x}_i$, $\hat{y}_i$) and the center coordinates of the object's cutout region (ground_truth) given as the training data (x_i, y_i) of the object's cutout region (ground_truth) given as the training data. The second term represents the error between the predicted size of the bounding box (*height*, *weight*) = ($\hat{h}_i, \hat{w}_i$) and the size of the object's cutout area

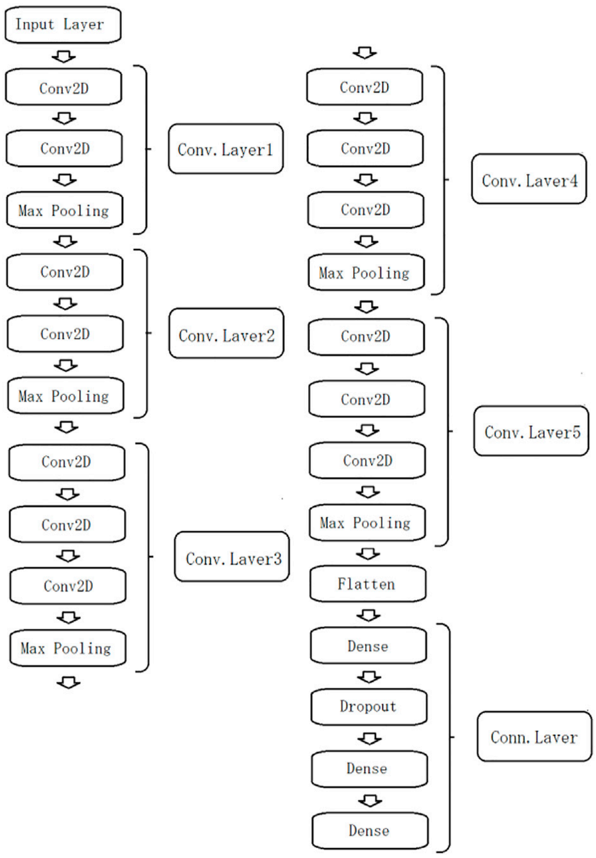


FIGURE 6
Structure of vgg16.

TABLE 1 Classified targets and Number of images.

	Class label	Training	Validation	Test
No.1	Crack	3062	171	171
No.2	Free lime	4517	251	251
No.3	Undamaged	3780	210	210
No.4	Joint	1784	100	100
No.5	Concrete formwork line	1083	61	61
No.6	Other	3062	171	171

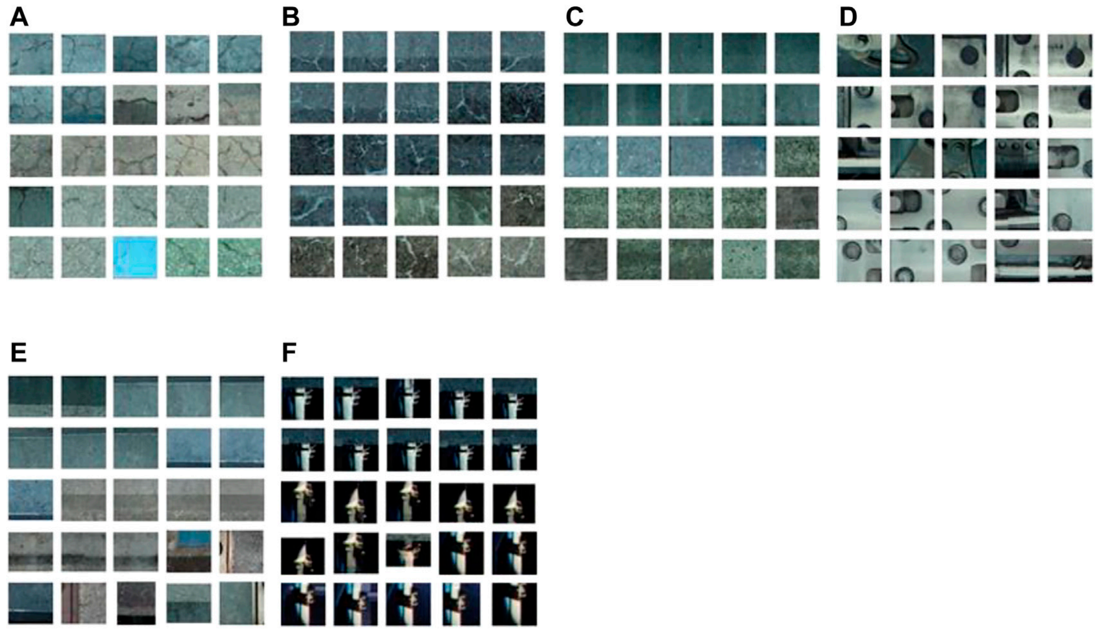


FIGURE 7
An example of Classified targets. **(A)** Crack. **(B)** Free lime. **(C)** Undamaged. **(D)** Joint. **(E)** Concrete formwork line. **(F)** Others.

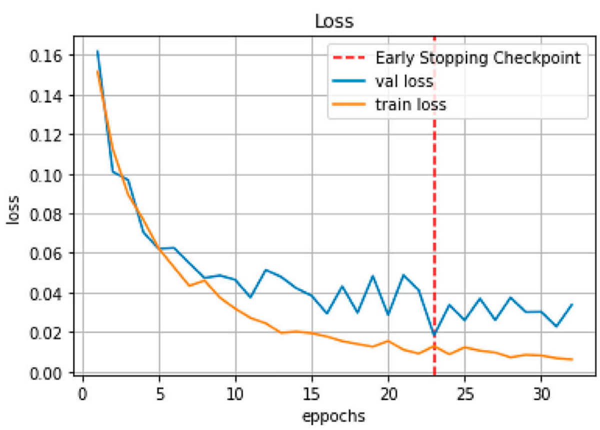


FIGURE 8
Time history of Loss values between training data and validation data.

(ground_truth) given in the training data $(height, weight) = (h_i, w_i)$ of the object given in the training data. The third term corresponds to the confidence level of the bounding box$\hat{c}_i$ and the confidence level of the training dataC_i (=1), which is calculated only for the grid cell with the center coordinates of the bounding box. The fourth term computes the prediction error of the unreliability of the bounding box,

TABLE 2 Classification results for each class.

Class label	Recall	Precision	F1score
Crack	0.98	0.94	0.96
Free lime	0.96	0.99	0.97
Undamaged	0.92	0.94	0.93
Joint	0.97	0.99	0.98
Concrete formwork line	0.98	0.98	0.96
Others	0.93	1.00	0.96

and is computed only for grid cells that do not have the center coordinates of the bounding box. The fifth term computes the error in the conditional class probability$\hat{p}_i(c)$ for the classification of a grid cell. The filters and biases inside the deep convolutional neural network are optimized so that each of the above five terms converges to zero.

3.4 Learning of supervisory data from crack images

When performing general object detection, training data requires a region cut out from the image data of the object to be detected, the location information such as the center

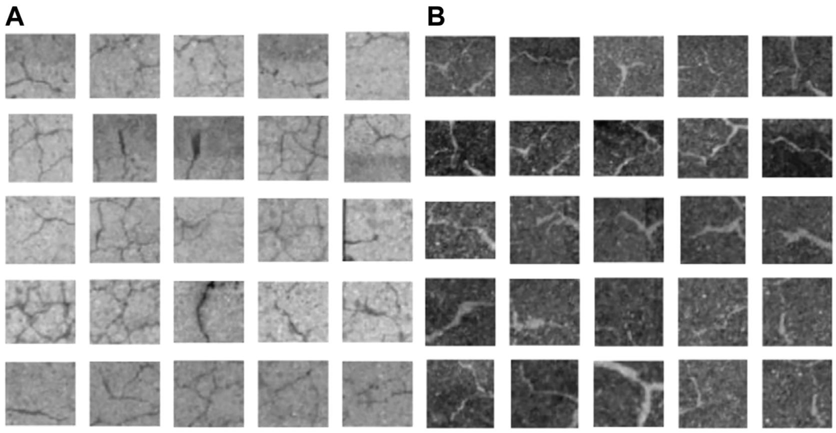


FIGURE 9
Images before color tone correction. **(A)** Crack. **(B)** Free lime.

coordinates and size of the region, and a labeled class. YOLO reads the location information of the object to be detected and automatically generates training data. The network weight coefficients are learned using the image of the object and the object's location. Image processing software is used to extract the image data and positional information from the image data, but LabelImg is used in this study. This method prepares an image showing cracks, specifies the cracked area from the image, and creates location information for YOLOv2 objects. Figure 4 shows an example of cutting out an object from an image. In this study, 25,000 images of cracks were prepared, from which the objects to be detected were extracted in detail and reserved as training data (ground_truth). The rectangular region in each image serves as the ground_truth during training. However, YOLOv2 performs random resizing, mirror image flipping, rotation, and HSV color space transformation on the teacher images during training to increase the number of apparently different image types (data augmentation). The convolutional neural network in this study uses the weights of a trained network for general object detection provided by the YOLOv2 developers as initial values. In one epoch, 64 images and their associated ground_truth were randomly selected for training, and training was completed in 100,000 epochs. The average IOU at the end of training was about 0.8. Considering our previous experience and the reliability of the objects, we judged that the given training data were adequately trained.

3.5 Detection result

Figure 5 shows an example of detection results. In Figure 5, the crack tracing results, which were visually confirmed by the engineer, are overlaid to make the detection results easier to understand. As shown in Figure 5, although some over-detection was observed, the recall of 50 images with cracks was about 88.9%. The recall in this study was calculated as the ratio of the total number of pixels of cracks traced by the engineer to the number of traced pixels in the predicted rectangular area.

4 Post-detection scrutiny by convolutional neural networks-based image recognition

In object detection using YOLOv2, increasing the confidence threshold for crack and other class classification results enables detection of only highly accurate areas, but it is likely to increase the lack of detection of small cracks. On the other hand, lowering the threshold reduces the number of the lack of detection of small cracks, but at the same time, dirt on the concrete surface may also be detected as cracks, resulting in more false detections. Therefore, we attempt to improve accuracy by incorporating image recognition technology using CNN after detection, and by conducting a close inspection after crack detection. Specifically, the threshold for the confidence value, which is output from YOLO, is set low, and all the rectangle determined to be a crack in YOLO is cropped as an image. All of the images are then input to the CNN and examined to determine if they are indeed cracks. The images that are not cracks are rejected, and only the images that are cracks are reflected. As described above, the system is designed to reduce both detection omissions and false positives by performing a close inspection using image recognition after detection at a low threshold value. In this study, the CNN used for the scrutiny was a fine-tuned version of VGG16 as a general-purpose model.

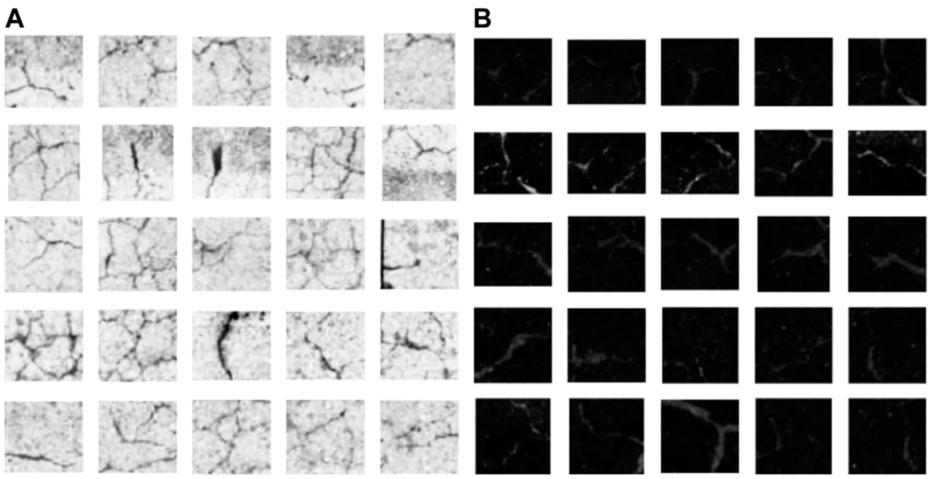


FIGURE 10
Images after color tone correction. **(A)** Crack. **(B)** Free lime.

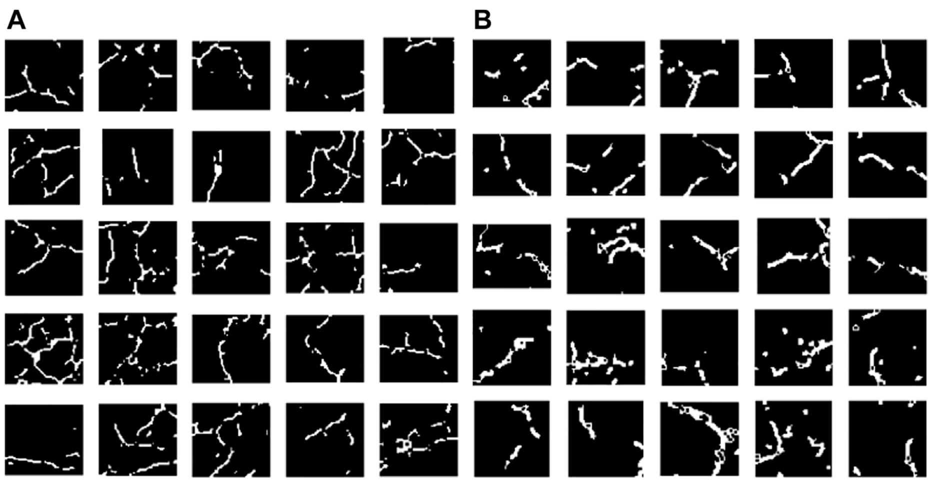


FIGURE 11
Results of binarization. **(A)** Crack. **(B)** Free lime.

VGG16 is a CNN model consisting of 16 layers trained on a large image dataset called ImageNet, and was developed by a research group called VGG at the University of Oxford in 2014. Because of its simple architecture, it is one of the learned models used in various studies. VGG16 consists of 13 convolutional layers and 3 all-connected layers, for a total of 16 layers, and its output layer is a neural network with 1,000 units and 1,000 classes to classify. The structure of the VGG16 model is shown in Figure 6. One of the features of this neural network is that it uses a 3 × 3 convolutional filter, which suppresses the increase in the number of parameters in the deeper layers and improves accuracy. The convolutional stride is fixed at 1 pixel, and there are 5 Maxpooling layers, where the size is halved. The stack of convolutional layers is followed by three all-combining layers. The last layer is the softmax layer, which uses ReLU as the activation function. When utilizing a trained network, feature extraction is processed in the convolutional layer within the network. When new data is applied to the network, a unique classifier is used in the network after the convolutional layer. The convolutional layers can be reused because their feature maps are generic and are likely to be useful in any computer vision application. Based on the results of the aforementioned damage detection, we further classified the data into six different classes.

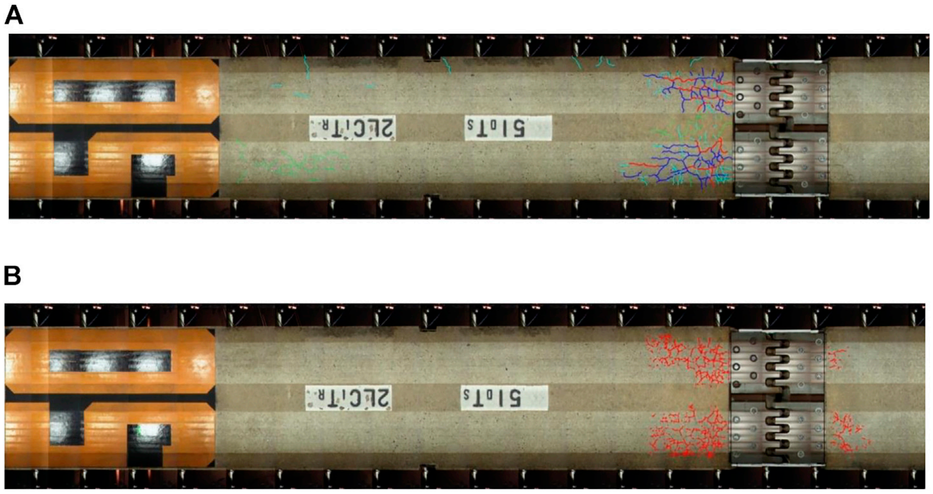


FIGURE 12
An example of the results for concrete girder of monorail bridge (surface A). **(A)** Damage detection results of manual tracing by engineers. **(B)** Damage detection results by the proposed system.

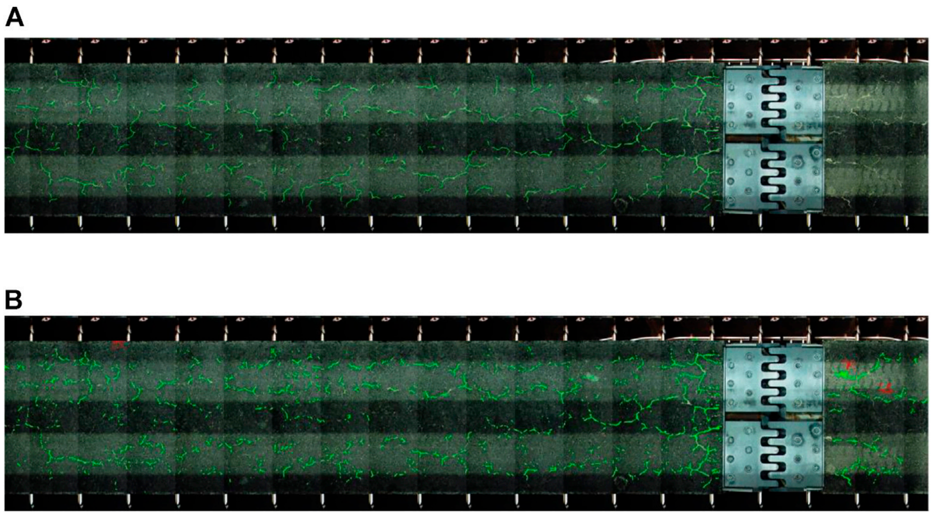


FIGURE 13
An example of the results for concrete girder of monorail bridge (surface B). **(A)** Damage detection results of manual tracing by engineers. **(B)** Damage detection results by the proposed system.

The six classes and the number of images for each class are listed in Table 1. The examples of images in the above six classes are shown in Figure 7. The most important classes are "crack" and "free lime". Therefore, images identified as "crack" or "free lime" are binarized through morphological processing. In this study, we set a function to terminate learning, which is called as early stopping, to avoid an overlearning. In addition, we employ the Cost-sensitive Learning approach, which defines a loss function that imposes a heavier penalty for misclassification of data with fewer labels due to the unbalanced amount of training data.

The time history of errors in the training and validation data is shown in Figure 8. As shown in Table 1, approximately 5% of the data in each class was allocated to the test data. The Precision, Recall, and F1Score for each class is shown in Table 2. It is found from the figure that the F1score is higher than 0.95 for the six classes.

TABLE 3 The performance of damage detection.

Image no.	Precision	Recall
Surface A 1	0.441	0.766
Surface A 2	0.279	0.717
Surface A 3	0.285	0.721
Surface B 1	0.283	0.980
Surface B 2	0.285	0.971
Surface B 3	0.355	0.984

TABLE 4 The range of α and coloring.

Gap (pixel)	Coloring
0px	White
$0px \lneqq \alpha \leq 3px$	Blue
$3px \lneqq \alpha \leq 5px$	Cyan
$5px \lneqq \alpha \leq 10px$	Green
$10px \lneqq \alpha \leq 20px$	Yellow
$20px \lneqq \alpha$	Red

5 Morphological processing for pixel-by-pixel damage detection

5.1 Color tone correction

In this study, morphological processing is used to enhance linear features from the image luminance value and to binarize cracks and free lime areas. In order to perform damage detection on a pixel-by-pixel basis, a color tone correction process is required as a preprocessing step. The color correction process in this study includes histogram flattening, gamma correction, and contrast correction.

First, histogram flattening reduces variations in the brightness and darkness of the image caused by factors such as illumination placement, shadows, and characteristics of the visible-light camera. Since the histogram flattening process corrects luminance values, it is necessary to correct the exposure level. In this study, gamma correction is used to correct the exposure. Gamma correction results in a luminance value close to black (0) for "cracks" and close to white (255) for "free lime". Contrast correction is then used to correct for differences in brightness and darkness in the image. In this study, the "cracks" are corrected to be darker than the background, and the "free lime" is corrected to be whiter than the background to emphasize the distinction between the target area and the background area. Figures 9, 10 show examples of images before and after color correction, respectively.

5.2 Median filtering, line enhancement and binarization processing

Contamination is apparent on concrete surfaces. In this study, a median filter is applied to remove elements other than "cracks" and "free lime. The median filter removes spike noise and produces a smoothed image that preserves smooth edges by giving the median value in the local area set by the filter size. Scale line enhancement is then applied to emphasize areas of "cracks" and "free lime". In this study, the Hesse matrix is applied to the line enhancement process. However, the output value of the line enhancement process has a brightness value of $0 \leq x \leq 255$ at each pixel. In order to evaluate the progress of damage, it is easier to evaluate the damage concisely if the luminance value at each pixel is 0 or 1. Therefore, in this study, the Canny edge detection algorithm of OpenCV is used to perform binarization detection on the output of the line enhancement process. Figure 11shows examples of the results by binarization for crack and free lime.

6 Experiment of damage detection for monorail bridge

In the proposed system, the engineer simply inputs an image and the system performs damage detection using YOLO, damage scrutiny using VGG16 for the detection results by YOLO, and pixel-by-pixel damage detection using binarization based on morphology processing. In other words, the engineer simply inputs an image to the proposed system, and the system outputs the image with the damaged areas colored.

Figures 12, 13 shows an example of the detection results by the proposed system to an image of a concrete girder of a monorail bridge. Green-colored areas indicate free lime, and red-colored areas indicate cracks. Figure 12 shows the detection results for the concrete surface where cracks are dominant as damage (surface A), and Figure 13 shows the detection results for the concrete surface where free lime is dominant as damage (surface B).

Figures 12, 13A are the damage detection results of manual tracing by engineers. Figures 12, 13B are the damage detection results by the proposed system. To quantify the detection accuracy, we evaluated the accuracy of the trace image and the detection result image in pixel units. The evaluation indices of detection accuracy applied in this study were Recall and Precision. Recall was defined as the probability of correctly predicting the correct pixel, and Precision as the probability of being the correct pixel among the predicted pixels. Table 3 shows the performance of damage detection for the surface A and surface B.

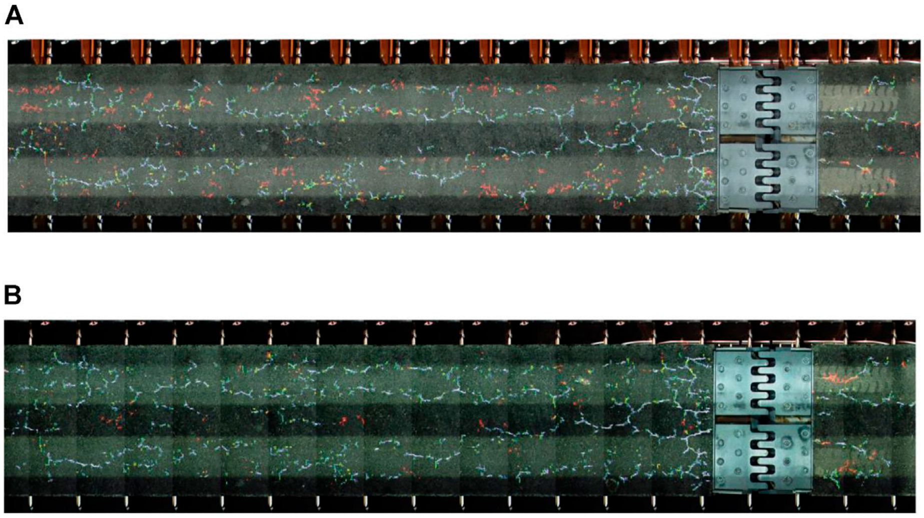


FIGURE 14
An example of Damage progression results. **(A)** Damage progression in running surface from 2014 to 2015. **(B)** Damage progression in running surface from 2015 to 2016.

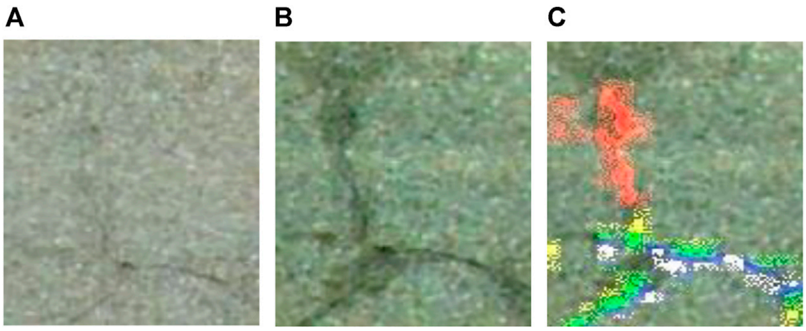


FIGURE 15
The detail of crack progression results. **(A)** Surface in 2015. **(B)** Surface in 2016. **(C)** Progression results of cracks.

As for the surface A, it is shown that Recall obtained a value exceeding 0.7. Meanwhile, as for the surface B, it is found that Recall obtained a value exceeding 0.9. These indicate that although there are some missing detections, the performance of the proposed system is adequate.

In particular, the reason for the high accuracy of the detection of free lime may be that "free lime" has thick line features due to the overflow of white substances such as efflorescence, and the brightness values were easy to distinguish from the those of background of the concrete surface.

On the other hand, Precision is lower than Recall, which indicates that there are many false positives. Thus, further improvement of the system is needed in the future. In addition, it is necessary for engineers to examine the tracing results more closely.

7 Damage progression assessment

In this study, the evaluation of the damage progression is limited to cases where the pixel gap between the two images is within approximately 20px, and is not applicable to cases where the difference between the two images is larger than 20px. In order to evaluate the progress of damage, it is essential to have yearly images taken at the same location. These images are obtained by applying a vehicle-mounted surveying system that continuously acquires three-dimensional coordinate data and visible light images of the road

10.3389/fbuil.2022.972796

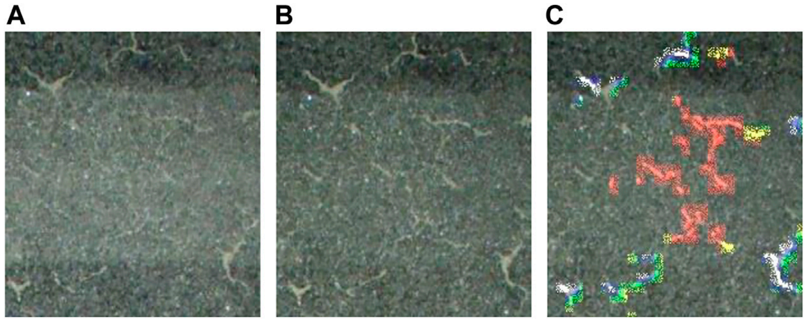


FIGURE 16
The detail of free lime progression results. **(A)** Surface in 2015. **(B)** Surface in 2016. **(C)** Progression results of free limes.

surface. We propose a method to evaluate the progress of the system by allowing for the occurrence of such misalignments.

Specifically, the latest inspection result (Image A) and the inspection result from one period ago (Image B) are prepared, and the damage result is examined pixel by pixel between the two images. The coordinate information (x, y) of damage X in image A is extracted, and the existence of damage X' corresponding to the coordinate information (x, y) in image B is confirmed; if there is no discrepancy of 1px and the damage occurrence area X and X' are identical, it is judged that there is no progress and colored white.

However, since the two images usually have a discrepancy in the angle of view, we attempt to check whether the damage exists within the coordinate information ($x \pm \alpha$, $y \pm \alpha$) in image B. Table 4 shows the range of α and the coloring. If X' is found in the range of $0 \lneqq \alpha \le 3$, it is colored blue, if X' is found in the range of $3 \lneqq \alpha \le 5$, it is colored cyan, if X' is found in the range of $5 \lneqq \alpha \le 10$, it is colored green, if X' is found in the range of 10 If X' is found in the range of $5 \lneqq \alpha \le 5$, it is colored green, and if X' is found in the range of $10 \lneqq \alpha \le 20$, it is colored yellow. If no damage X' is found in image B after 20px of circumference, it is considered highly likely to be a new damage and is colored red.

Figure 14 shows an example of the evolution of damages such as cracks and free limes in comparison with previous years. In particular, as progress from 2014 to 2015 shown in Figure 14A, while there is overlapping damages between previous years that is visualized in white, there are many areas colored in red. Although this might be due to a discrepancy in the angle of view, it is highly likely that minor cracks and free lime that existed in previous years have developed. On the other hand, as progress from 2015 to 2016 shown in Figure 14B, although red-colored areas are observed, they are fewer than those in Figure 14A. Although this is a qualitative assessment, it suggests that there was not much progression of cracks and free limes between 2015 and 2016.

Finally, examples of the details of crack and free-lime progression results are shown in Figures 15, 16, respectively. The left and center figures in Figures 15, 16 show the concrete surface of an area in 2015 and 2016, respectively, while the right figure shows the progression results using the proposed method. As shown in these figures, although it might be not always possible to detect only the areas of progression of damages due to the slip of the both images, the red-colored areas can be identified as newly formed cracks or free limes.

8 Conclusion

The objective of this study was to develop a system that can detect damage on structural surfaces with high accuracy and speed using the results of visual damage inspections, as well as a system that automatically outputs an evaluation of the progress of damage. In addition to the detection technology using YOLO, which is capable of high-speed processing of a large area and a large number of images, we aimed to improve the accuracy of damage detection by using image recognition technology based on CNN and the network model VGG16 to classify the detected rectangular images and introduce scrutiny to the object detection technology. In order to perform the evaluation of progression at the fine pixel level, morphology processing was applied, which is less dependent on the reliability of the trace image as a teacher image compared to deep learning methods. The morphology processing included color correction, median filtering, line enhancement, and binarization, and the final image was output as a pixel-by-pixel damage detection result attached to the original image. By automating these processes, we were able to realize a function in which an engineer inputs the original image and automatically outputs an image in which the damage occurrence areas and the damage development result are drawn on the original image. The findings obtained through this research and future works are summarized as follows.

YOLO was found to be able to detect difficult-to-see damage in images. The results also show that introducing image recognition processing learned by CNN after YOLO detection can improve the accuracy of YOLO. This shows the usefulness of CNNs in terms of performing a close inspection after general object detection.

Through morphological processing, it was found that the recall of damage detection at the pixel level was found to be 0.7 to 0.9.

In the evaluation of progression, a few pixel discrepancies were allowed in stages, and it was possible to identify the areas where progression was likely to be high for minor cracks and free lime that had existed in previous years by coloring the areas in stages.

In the future, the results of inspections by managers, which have been accumulated as training data for object detection and image classification, will be carefully examined, and the latest algorithms, including YOLOv5, which is fast and accurate, will be applied for practical use. In addition, we will attempt to detect cracks and free lime at the pixel level by using deep learning segmentation techniques for binarization of damage. Finally, the proposed method still has many cases of over-detection, so we need to examine the training data again to improve not only the recall but also the precision and so on. We also need to consider image registration techniques for evaluating progress of the crack, since large image shifts can cause problems.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bang, S., Park, S., Kim, H. H., and Kim, H. (2019). Encoder-decoder network for pixel-level road crack detection in black-box images. *Computer-Aided Civ. Infrastructure Eng.* 34 (8), 713–727. doi:10.1111/mice.12440

Behara, S., Mohanty, M. N., and Patnaik, S. (2012). "A comparative analysis on edge detection of colloid cyst: A medical imaging approach," in *Soft computing techniques in vision science*. Editors P. Srikanta and Y.-M. Yang (Berlin, Heidelberg: Springer), 63–85.

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-8, 679–698. doi:10.1109/tpami.1986.4767851

Cha, Y.-J., Chai, W., and Büyüköztürk, W. O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civ. Infrastructure Eng.* 32 (5), 361–378. doi:10.1111/mice.12263

Cheng, H. D., Glazier, X. J. C., and Glazier, C. (2003). Real-time image thresholding based on sample space reduction and interpolation approach. *J. Comput. Civ. Eng.* 17 (4), 264–272. doi:10.1061/(asce)0887-3801(2003)17:4(264)

Chun, P., and Igo, A. (2015). Crack detection from image using random forest. *J. JSCE* 71 (2), I_1–I_8. doi:10.2208/jscejcei.71.i_1

Dung, C. V., and Anh, L. D. (2018). Autonomous concrete crack detection using deep fully convolutional neural network. *Automation Constr.* 99, 52–58.

Fujita, Y., Mitani, Y., and Hashimoto, Y. (2006). "A method for crack detection on a concrete structure," in *18th International conference on pattern recognition (ICPR'06)* (IEEE), 3, 901–904. doi:10.1109/icpr.2006.98

He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2018). *Mask R-CNN*. arXiv:1703.06870v3.

He, K., Zhang, X., Ren, S., and Sun, J. *Deep residual learning for image recognition*, arXiv:1512.03385v1, (2015).

Huang, G., Liu, Z., Maaten, L., and Weinberger, K. Q. (2018). *Densely counnected convolutional networks*. arXiv:1608.06993v5.

Huang, Y., and Xu, B. (2006). Automatic inspection of pavement cracking distress. *J. Electron. Imaging* 15 (1), 013017. doi:10.1117/1.2177650

Ju, H., Li, W., Tighe, S., Zhai, J., and Chen, Y. (2019). Detection of scaled and unsealed cracks with complex backgrounds using deep convolutional neural network. *Automation Constr.* 107, 102946.

Kanopoulos, N., Baker, N. R. L., and Baker, R. L. (1988). Design of an image edge detection filter using the sobel operator. *IEEE J. Solid-State Circuits* 23 (2), 358–367. doi:10.1109/4.996

Konovalenko, I., Maruschak, P., Brezinová, J., Brezina, O. J., and Brezina, J. (2022). Research of U-Net-Based CNN architectures for metal surface defect detection. *Machines* 10 (5), 327. doi:10.3390/machines10050327

Konovalenko, I., Maruschak, P., Prentkovskis, V. O., and Prentkovskis, O. (2021). Recognition of scratches and abrasions on metal surfaces using a classifier based on a convolutional neural network. *Metals* 11 (4), 549. doi:10.3390/met11040549

Li, Q., Zou, Q., Mao, D. Q., and Mao, Q. (2011). Fosa: F* seed-growing approach for crack-line detection from pavement images. *Image Vis. Comput.* 29 (12), 861–872. doi:10.1016/j.imavis.2011.10.003

Liu, J., Yang, X., Lau, S., Wang, X., Luo, S., Ding, V. C. L., et al. (2020). Automated pavement crack detection and segmentation based on two-step convolutional neural network. *Computer-Aided Civ. Infrastructure Eng.* 35, 1291–1305. doi:10.1111/mice.12622

Liu, Z., Cao, Y., Wang, Y. W., and Wang, W. (2019). Computer vision-based concrete crack detection using u-net fully convolutional networks. *Automation Constr.* 104, 129–139. doi:10.1016/j.autcon.2019.04.005

Martin, D., Malik, C. J., and Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (5), 530–549. doi:10.1109/tpami.2004.1273918

Nishimura, S., Demizu, A., and Matsuda, H. (2012). The Proposal to the infrastructure research of the future as seen from survey and verification on gunkanjima-island -3D laser scanner photogrammetry UAV AR-. *J. JJSEM* 12 (3), 147–158.

Nishimura, S., Hara, K., Kimoto, K., and Matsuda, H. (2012). The measurement and Draw damaged plans at Gunkan-Island by Using 3D laser scanner and Digital Camera. *J. JSPRS* 51 (1), 46–53. doi:10.4287/jsprs.51.46

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man. Cybern.* 9 (1), 62–66. doi:10.1109/tsmc.1979.4310076

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). *You only look once: Unified, real-time, object detection*. arXiv preprint arXiv:1506.02640v5

Redmon, J., and Farhadi, A. (2016). *YOLO9000: Better, faster, stronger"*. arXiv preprint arXiv:1612.08242v1.

Redmon, J., and Farhadi, A. (2018). *YOLOv3: An incremental inprovement"*. arXiv:1804.02767v1.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "Unet: Convolutional networks for biomedical image segmentation," in *International conference on medical image computing and computer-assisted intervention* (Springer), 234241. doi:10.1007/978-3-319-24574-4_28

Simonyan, K., and Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition"*. arXiv: 1409.1556v6.

Yamane, T., Ueno, Y., Kanai, K., Izumi, S., and Chun, P. (2019). Reflection of crack location to 3D model of bridge using semantic segmentation. *J. Struct. Eng.* 65A, 130–138.

Yoshida, R., Fujii, J., Okubo, J., and Amakata, M. (2020). Creation of crack detector on revetments and measuring crack using orthophoto. *J. AI Data Sci.* 1 (1), 508–513.

TYPE Original Research
PUBLISHED 07 October 2022
DOI 10.3389/fbuil.2022.1009944

OPEN ACCESS

EDITED BY
Xianbo Zhao,
Central Queensland University, Australia

REVIEWED BY
Syahrul Fithry Senin,
Universiti Teknologi Teknologi MARA, Malaysia
Ke Chen,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Tomáš Hanák,
hanak.t@fce.vutbr.cz

SPECIALTY SECTION
This article was submitted to Construction Management, a section of the journal Frontiers in Built Environment

RECEIVED 02 August 2022
ACCEPTED 26 September 2022
PUBLISHED 07 October 2022

CITATION
Hanák T and Vítková E (2022), Causes and effects of contract management problems: Case study of road construction.
Front. Built Environ. 8:1009944.
doi: 10.3389/fbuil.2022.1009944

Causes and effects of contract management problems: Case study of road construction

Tomáš Hanák* and Eva Vítková

Institute of Construction Economics and Management, Faculty of Civil Engineering, Brno University of Technology, Brno, Czechia

Contract management is a crucial component of any project. As construction projects are complex and difficult to manage, adequate attention must be given to the related contract management issues. Inexpert management can bring about serious unfavourable consequences that can even result in a project failure. This study aims to explore a specific case study: A private road construction project implemented in the Czech Republic. The analysis consists in identifying potential problems and discussion of their implications. The problem is investigated on three levels: economic, technical and legal. The paper also considers contrasting attitudes of the contracting parties (buyer and supplier), e.g., from the point of view of the causes of problems. Several recommendations are formulated on the basis of the research findings. The results of this study have an ambition to improve contract management capabilities in the construction sector in order to prevent the occurrence of similar problems in future and contribute to our understanding of long-term effects of contract management problems throughout the life cycle.

KEYWORDS

construction project, contract management, road construction, project documentation, dispute

Introduction

Construction projects typically involve a high degree of complexity. In the theory, complexity is understood as the number and heterogeneity of different interrelated elements (Burke and Morley, 2016). Taking into consideration that construction projects 1) involve a high number of stakeholders, are 2) subject of frequent change orders, 3) of a long-term nature, and 4) affected by various influencing factors such as location, inflation, schedule targets, constrained budget, etc., it is difficult to understand, foresee and keep the under control (Vidal et al., 2011; Vařbuchta et al., 2017; Kermanshachi and Safapour, 2019). In order to cope with the project complexity, the project itself should be supported with good-quality technical, economic, and contractual documentation. Unfortunately, many projects suffer from disputes between the parties involved, leading up to adverse consequences such as significant delays, loss of quality in follow-up work, or lack of future cooperation (Lu et al., 2015). To imagine, National Construction Contracts and Law Survey reported that 30 percent of companies in the Great Britain had been involved in at least one dispute in the previous 12 months (RIBA Enterprises, 2013). Such disputes are often caused by a misunderstanding of the contractual terms and conditions as well as violation of obligations, no matter if real or only perceived by one of the parties. Given regular occurrence of disputes in construction projects, good negotiating skills are becoming crucial (Chow et al., 2012).

Research problem and aim of the paper

This paper deals with contract management problems and their influence on the construction and operation of a private road. Despite the fact that a wide body of knowledge on causes of claims, dispute resolution techniques, time and costs overruns, and project performance in general is available, there is a lack of studies that monitor these issues from a long-term perspective. In the case of the construction industry, this is essential because the duration of the life cycle of buildings is expected to be in the tens of years. At the same time, it is necessary to be aware that sooner or later the consequences of the decisions made in the early stages of the project will be felt. Although we know what the most common causes of disputes are and how they can be resolved, it is usually unclear to what extent the consequences of failing to resolve the dispute may increase in the future.

Accordingly, this study aims to investigate causes and effects of dispute in different phases of the life-cycle of a building facility. Therefore, the study addresses how contract management problems resulting from insufficient project preparation may affect not just the construction phase, but also the operation phase of the building facility's life cycle. The single-case study analysed in this research points out how a long-term dispute can negatively affect the operation of an owner's business; how it, for many years, has been tying contractor to a project which, from his point of view, had already been finished long ago; and especially how the continuously growing complexity of the problem significantly complicates its resolution.

Literature review

Project management literature provides a substantial body of knowledge concerning the assessment of a construction project's success. The main success criteria are time, cost, and quality, also known as the "iron triangle" (Ljevo et al., 2017). These criteria usually act as contradictive objectives and can be considered as interdependent parameters in a building project (Hu and He, 2014). This mutual interdependency was analysed by many researchers, e.g. in terms of time and cost for water supply systems (Zujo et al., 2017) and the balance of cost, time and quality for reinforced concrete (Hosseini et al., 2017). The aspect of project cash flow is also crucial as it allows the assessment of working capital requirements (Maravas and Pantouvakis, 2012).

The above-mentioned factors are crucial in construction disputes as well. Ilter (2012) deals with three categories of disputes, namely: extension of time, payments, and quality of works clearly corresponding to the iron triangle. Furthermore, study of Ilter (2012) identifies the relations between dispute categories with dispute factors such as late instructions by the employer, inadequate/incomplete specifications or unclear contractual terms. It should be emphasised that the contract is among the most important components of the construction project (Safa et al., 2017) and project documentation forms an integral part of the contract. This importance is documented by Leśniak et al. (2018), who point out that problems with project documentation are among the main causes of delays in both Poland and Slovakia. Disputes may arise for various reasons, in addition to usual ones between client and contractor such as cost and time overruns for ongoing projects (Nahidi et al., 2017; Johnson and Babu, 2020; Shoar et al., 2022; Subramanya et al., 2022), disputes may also be related to inadequate cost estimations for repairing damages caused by large scale natural disasters, such as floods or windstorms (Hanak and Korytarova, 2014).

Construction disputes may be destructive for projects (Ilter and Dikbas, 2009), therefore efforts should be made to prevent them in the first place. This issue is highly topical within the research community; e.g., Lee et al. (2016) found 140 articles dealing with dispute prevention. In this relation, Molenaar et al. (2000) developed a structural equation model for predicting of construction contract disputes. As presented by Naji et al. (2020), structural equation modelling enables to determine a regression model for the dispute occurrence as an output variable in terms of the dispute causes as real variable. However, since it is not possible to prevent all disputes, it is subsequently necessary to resolve them. In the case of systematic violations of one party's obligations, a judicial decision is typically needed (Yaskova and Zaitseva, 2017). This solution is commonplace in modern construction projects (Biering et al., 2016); however, it may take long (Dziadosz et al., 2015)—disputes can drag on for many years, which is why parties also consider alternative dispute resolution (hereinafter referred as ADR) methods, an option that arose as an alternative to lengthy and costly processes of arbitration and litigation (Cheung, 1999). The use of ADR (such as negotiation, arbitration or adjudication) depends on the nature of claims that are to be decided and also on the parties' perception of fairness and outcomes of win and losses in claims (Lee et al., 2016). The list of factors that affect the selection and use of ADR provided by Lee et al. (2016) is comprehensive, involving–apart from fairness and outcomes–e.g., bindingness, cost, confidentiality, control over the proceedings, lawyer's influence, perception of risk, and complexity of disputes. Regarding related costs, disputes resolution carries explicit costs such as lawyer's fees, court fees, consultant costs etc., however, other types of costs (so-called hidden costs) should also be considered, e.g., reduced project working efficiency, damaged reputation, or the aspect of future cooperation (Lu et al., 2015).

The frequency of disputes occurrence was analysed by Ilter and Dikbas (2009). They have revealed that *design and build* causes 8 precent more disputes in contrary to the *build* method and also that the frequency is positively correlated with the size of the contractor. According to Tazelaar and Snijders (2010) there is an 84 percent probability that at least some problem will occur within the project and 81 percent probability that the problem will be discussed successfully with the other party. However, if the problem involves a delayed payment or claim of damage, then there is 10 percent probability that the case will lead to arbitration or other legal steps.

The case study analysed in this paper deals with road construction. Road construction projects also involve the iron triangle; however, it should be noted that problems and risks are higher due to the duration and size of such projects, as well as unforeseen economic and other conditions such as material and energy prices. In the case of road construction, increased energy costs influence material costs and may slow down projects significantly (Hashem Mehany and Guggemos, 2015). An example of the procedure of contractor's claims management in a road works project is provided by Rybka et al. (2017). Mishmish and El-Sayegh (2018) have reported causes of claims in road construction projects in the UAE. Based on their findings, the most likely causes consist in variations initiated by the client or engineer, delays caused by the contractor (e.g., as a consequence of lack of resources and machinery on the construction site), and inadequate site investigation before bidding.

Providing sufficient background materials for each dispute is crucial. Typically, contractors lack documents such as adequate photographic evidence, time sheets, site diaries and revised drawings to pursue their claims (Vidogah and Ndekugri, 1998). Therefore, an experienced supervisor (Kongsong and Pooworakulchai, 2018) as well as the use of modern technical devices and solutions on the construction site (Banaszek et al., 2017; Acosta et al., 2019) are essential.

Materials and methods

The dispute analyzed in this case study was settled in court. Since authors of the paper participated in the preparation of the expert opinion for the court, they had at disposal complete documentation from the court file. Detailed documentation that has been available for the selected case study has been carefully studied. This documentation includes the following: contract for work, technical documents including initial road design, relevant documents provided by both parties to the dispute, and materials obtained during personal investigation of the site. All sensitive information and materials are published in an anonymised form (incl. location, time, name of the parties involved, the amount of costs, name of the technology used). The article contains only the facts that are relevant to the achievement of the paper's objective, i.e., to identify dispute factors and the related consequences.

After examining the textual and graphical documentation in the court file (analysis no. 1), an additional set of information was obtained during personal investigation of the site that took place in 2018 (analysis no. 2). Due to the long duration of the dispute, 6 years passed before the revision expert opinion was ordered and processed. On the other hand, it enables to explore evidentiary material covering the whole six-year period. This investigation primarily involved physical documentation of visible damages, the extent of which progressively evolved over time (when compared to previous evidence available in the court file). Subsequently, the disputed matters were discussed with both parties (investor and supplier) in order to obtain the widest possible material for the subsequent assessment of responsibility for the occurrence of defects. When all the defects have been identified, the causes of problems were investigated. Then, cause-effect relations were assigned and graphically presented in multi-dimensional diagram. In addition, the entire analysis has been performed in the context of the iron triangle (i.e., time, cost, and quality issues). The case study research methodology is graphically presented in Figure 1. Based on the findings, several lessons learnt are provided at the end of the manuscript.

Regarding scientific methods use, analytical-synthetic cognitive procedures have been applied to achieve an understanding of the qualitative data. Of these, a causal analysis was particularly important as it helps identify causes of phenomena. It is crucial to ensure objective and deep insight into the problem, especially in situations when observer just see certain process, but its manifestations are dependent on a very complex causal chain of causes and effects (Molnar et al., 2012, p. 22). For graphical interpretation for casual models, different types of casual diagrams are used, e.g., mind mapping and casual loop diagrams. Mind maps provides nonlinear graphical interpretation of data and casual loop diagrams are used to analyse qualitative data (Milen et al., 1997). Generally speaking, casual diagrams help to predict how the system would respond to hypothetical interventions (Pearl, 2000). For this reason, casual diagram was used as a graphical tool to present and understand causes and effects in analysed case study and, in the second stage, to facilitate the formulation of lessons learned. Project life cycle as well as the perspectives of both parties to the dispute in view of dispute categories are considered within the performed analysis.

Case study of a road construction project

The subject of the analysis consists in construction of a road on a privately-owned plot of land and its connection to the

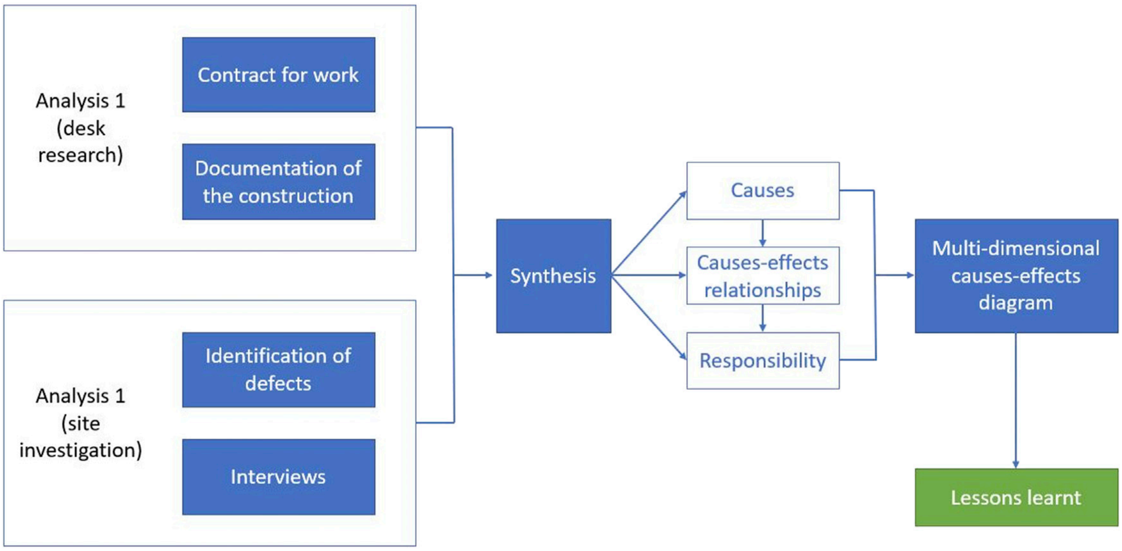


FIGURE 1
Case study research methodology.

TABLE 1 Basic characteristics of the analysed project.

Scope of the project	Bituminous road, 947 m^2
	Concrete interlocking pavement, 391 m^2
	Concrete drainage pavement, 233 m^2
Duration of the construction phase of the project	7 days (October/November 2012)
Structure of the bituminous road	asphalt concrete (50 mm)
	bituminous infiltration spraying (0.9 kg/m^2)
	base layer and subsoil stabilization (cement 55 kg/m^2)
	Subsoil - clay with medium plasticity (250 mm)

existing public asphalt road. Specifically, the road was to be used for transport to and from company premises. The road thus enables the entrance to the investor's premises, parking and the connection of the public road to the storage hall. The investor needed to bring the premises into operation quickly and, for this reason, it began to consider a change in the originally envisaged technological solution for the road. The investor and the contractor reached an agreement where the contractor proposed to use a non-standard construction process which promised to shorten the construction period and reduce price. The technology lacked binding rules in the Czech Republic and there was no demonstrable experience with its use in other projects in the country. The basic characteristics of the project are provided in Table 1 and a cross-section drawing is shown in Figure 2.

The investor and the contractor entered into a contract for work, but the contract unfortunately contained serious shortcomings consisting, e.g., in the absence of proper project documentation concerning the new technological solution attached to the contract as its integral part. All the above-mentioned facts subsequently led to disagreements and disputes between the contractual parties.

The road was built by the contractor, but defects and faults appeared during handover, resulting in the investor's refusal to accept the road and its request for a repair of these defects and faults (consisting, e.g., of incorrectly implemented drainage). Although the work was neither officially accepted nor paid for, the investor started using it for its business activities, which involved heavy freight traffic on the road which had not been designed for that purpose. Consequently, the defects and faults deteriorated further, manifesting as faults on the road near the drainage sites, ruts in the covering layer, cracks and unevenness in the road and area deformations. Figures 3–5 show examples of defects on the analysed road and view on its structure.

Since the parties were unable to agree on a solution to their dispute, they referred the matter to the court. As part of the

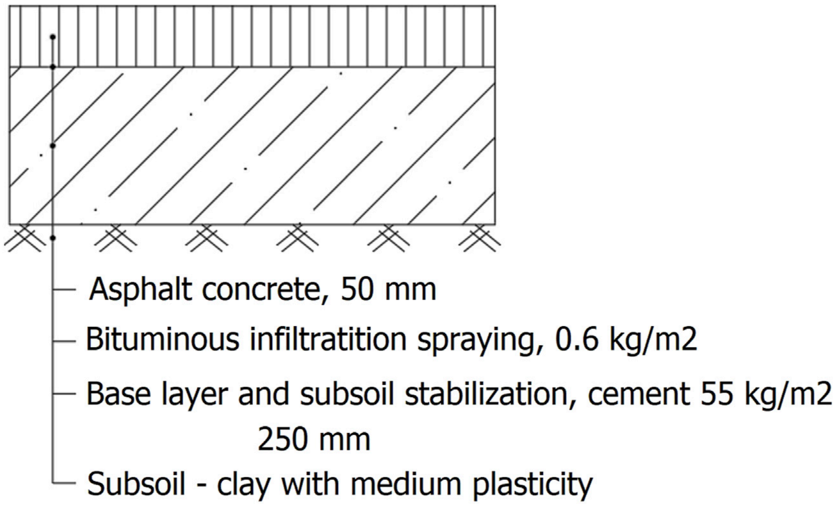


FIGURE 2
Cross section drawing of the road.

FIGURE 3
Cracks and deformations.

investigation of the causes of the defects and faults and the liability for them, the following factors were identified: inadequate preparation of the ground plane, incorrectly implemented drainage, uneven thickness of the asphalt layer, partially insufficient compaction of the underlying structure, insufficient binding between the covering and the supporting layer, and overloading the road by heavy freight traffic.

The situation was further worsened by the increasing number and scope of faults as time progressed, the fact that there was no building diary, as well as the absence of as-built documentation. Selected most serious faults and defects were provisionally repaired by the investor on the spot, yet documentation was again not made with respect to these repairs. As a result, the entire dispute became fairly complex, especially with regard to the assignment of liability for the individual defects and faults.

Figure 6 provides a lucid two-dimensional overview of the above facts. The individual factors at play and their consequences

FIGURE 4
Road surface irregularities.

FIGURE 5
View of the road structure.

are divided based on their time sequence into three life cycle stages, i.e., preparation, construction and operation. The second dimension consists in dispute categories (time, payment and quality). Colours indicate the decisions/activities/requirements of the parties in the dispute (blue = investor; orange = contractor; colour gradient = both).

The previous analysis showed that the project failure was caused by four main general factors: 1) selection of the contractor, 2) selection of the technology, 3) insufficient documentation, and 4) unsuitable use of the structure. The interaction of these factors made solving the dispute by agreement of the parties practically impossible.

The contractor selection process should generally be conducted in a way that results in implementation of the work by a construction company that has sufficient technical, managerial and economic qualifications to deliver the work in the required quality. Although the entire process of selecting the contractor was not part of the analysis, the authors note that a proper verification of the contractor's qualifications is crucial in this regard. Only qualified contractors should be selected to

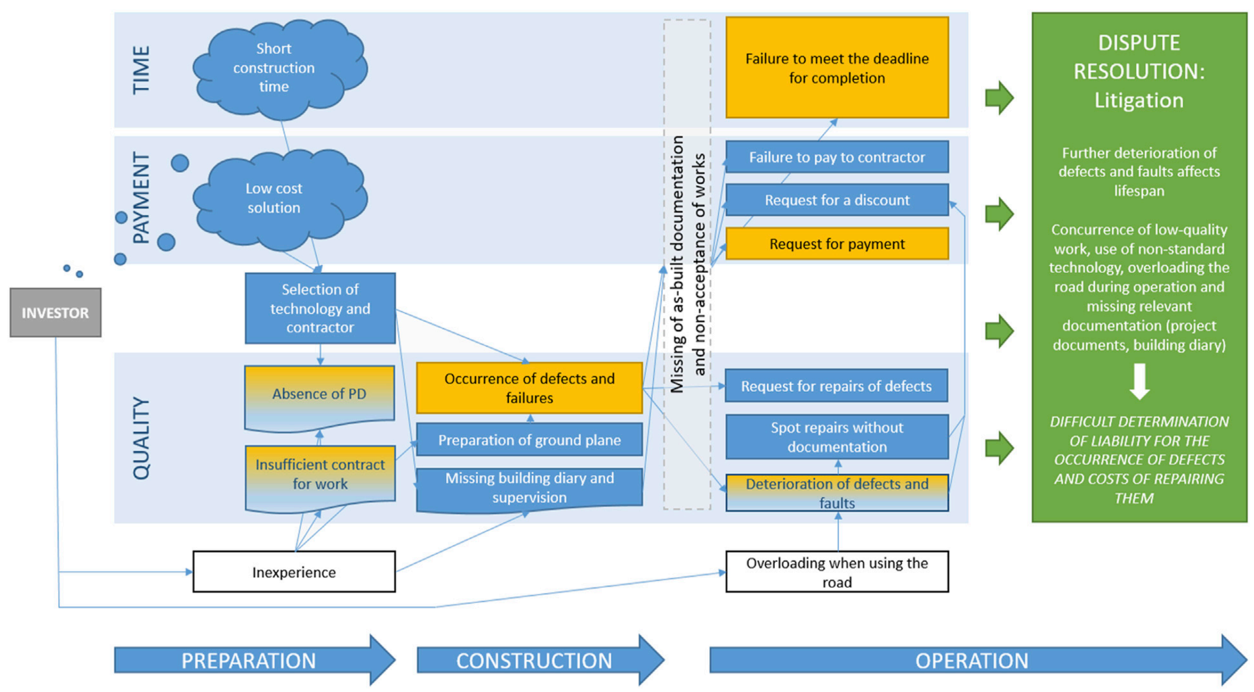


FIGURE 6
Causes and effects diagram.

TABLE 2 Recommendations resulting from the case study analysis.

Problem	Measure	Effect
Selection of an unsuitable solution	Clear specification of parameters of the work according to operational requirements	Elimination of adverse effects of operation on the quality of the work and its lifespan
Selection of the contractor	A proper assessment of qualifications incl. verification of references	Proper completion of the work in the desired quality
Contract for work	Consulting the contents of the contract for work with an expert in construction contract management	Clear definition of rights and obligations following from the contract and elimination of potential disputes
Handover of the construction site	Record of the handover of the construction site	Written confirmation that the investor handed over the construction site in a condition agreed in the contract for work
Supervision over the construction process	Keeping of building diary and the presence of a supervisor	Supervision over the quality of the construction and record-keeping for documenting the scope of the works carried out, defects and liability for potential future defects
Handover of the work	Provisions in the contract for work	Specification of potential defects and non-performances preventing proper use of the work, as well as other defects and non-performances

implement the contract, which means contractors who are competitive, competent and capable (Lam et al., 2010). A part of the qualifications should have consisted in documented (and verified by the investor) references proving that the contractor was competent and capable.

The above requirement for submission of references is also related to the selection of the technology to be used. The investor did not verify if other projects using this technology had been implemented in the Czech Republic. This fact is especially important since this construction technology was not tried and proven in the Czech Republic and lacked relevant generally binding standards.

The dispute escalated and was difficult to resolve particularly due to the absence of relevant documents. This means both design documents for construction and as-built documentation. Since the contract for work concluded in this particular case study did not contain the usual provisions concerning construction contracts and the investor was not experienced in this area, no building diary was made during the construction and the investor did not arrange for adequate supervision of the work. These factors significantly hindered the possibility for determining qualitative shortcomings of the construction, e.g., in relation to bad binding of the covering and supporting layers since these structures are not visible.

Finally, another important factor consisted in the fact that the investor started using the road and overloaded it with heavy freight vehicles for which the surface was not designed. This brought about a situation where both parties to the dispute bore some responsibility for the occurrence of certain defects and faults. For instance, as regards the road cracks, it is not clear to what extent they were caused by, e.g., insufficient drainage of the site as opposed to heavy traffic. Determining the exact degree of liability for the individual defects and faults is thus nearly technically impossible, taking into consideration the fact that the scope of the faults grew as the time progressed.

Finally, several lessons learnt from the analysis of the case study can be presented. Firstly, it is necessary to choose a technical solution suitable to the structure's operational requirements. In this relation, a detailed project documents must be drawn up.

Secondly, the contractor selection process should take place on the basis of sufficient qualifications. The qualification requirements should enable a selection from a sufficient number of bidders with the aim of ensuring sufficient competition within the selection procedure and, simultaneously, exclude from the procedure such potential contractors that lack the qualifications for successful completion of the project or who have previously developed reference projects that had quality issues.

Thirdly, contractual provisions in the contract for work should be consulted with experts in construction contract management in order to minimise the possibility of disputes in the course of the project's implementation and afterwards.

Fourthly, handover of the construction site to the contractor has to be properly recorded in order to ensure that the investor has met all its obligations as agreed in the contract for work.

Fifthly, during the course of the construction, the investor should arrange for oversight of the works by a supervisor; keeping a building diary including photographic documentation by the contractor is also necessary. These measures enable timely identification of low-quality work and achieving appropriate remedy; potentially, the same documents can serve as evidence for determination of the scope and liability for defects appearing during operation, which is especially important with regards to structures that are covered and cannot be visually inspected.

Sixthly, defects and shortcomings found during delivery of the completed work should be recorded with participation of both contractual parties and the supervisor. The defects and faults should be classified into two categories based on their severity, i.e., defects preventing the use of the work and defects not preventing the use of the work. This categorisation should be supported by the provisions of the contract for work; material defects preventing proper use of the work are generally considered to constitute grounds for non-acceptance of the work by the investor. A list of these lessons is provided in Table 2 showing the relationships between the problem, the measures adopted, and their desired effects.

As a result, the impact of this dispute is negative mainly for the investor, but also for the supplier. For investor, the effects consist mainly in the limited possibility to use built communication and in additional incurred costs for carrying out local repairs. Since the supplier consistently and completely rejects its liability for defects, its main burden consists of attending court hearings and the related costs of legal representation.

Conclusion

This study aimed to identify dispute factors and the related consequences on a case study involving the construction of a road to commercial premises. The subject of the analysis involved a critical evaluation of the interaction of individual factors both in terms of the dispute categories and the life cycle of the project. The mutual interactions are depicted in casual diagram (Figure 6), which shows the process in which a dispute becomes increasingly more complex and thus more difficult to resolve by reaching a mutually acceptable compromise solution.

Presented research is not without limitations. As this is a case study of one particular project, the observations and conclusions reached herein cannot be generalised to the entire construction sector. However, due to the high level of complexity of the analyzed case in the perspective of the life cycle, substantial implications can be derived not only for practice but also for theory.

This paper provides two main managerial implications. Firstly, the discussion of the case study brings several recommendations on how to avoid potential problems based on lessons learned from the situation that occurred in the studied project. Secondly, the conducted analysis shows that in case of protracted disputes arising from insufficient documentation, determining the party responsible for defects becomes extremely difficult.

From a theoretical perspective, performed analysis contributes to the current body of knowledge by highlighting potential undesirable growth in project complexity in relation to inexpert and controversial steps of the parties to the dispute. Furthermore, the findings presented extend our understanding of the long-term effects of contract management problems within individual stages of the project life cycle on both parties involved in the dispute and within iron triangle dimensions.

In order to reach generalisable findings in the area of transport construction projects, further project case studies should be analysed; in aggregate, they would have the potential to identify recurring negative scenarios, project inflection points and possibilities for general improvement in the area of contract management.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: business case of dispute. Requests to access these datasets should be directed to hanak.t@fce.vutbr.cz.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This paper has been written with the support of a research grant FAST-S-22-7970 "Economic and managerial processes in civil engineering".

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Acosta, L. M., Torres, A., Krejza, Z., Gibert, V., and Serrat, C. (2019). Exploring technological alternatives to the visual inspection method in the built environment. *IOP Conf. Ser. Earth Environ. Sci.* 222, 012027–12110. doi:10.1088/1755-1315/222/1/012027

Banaszek, A., Banaszek, S., and Cellmer, A. (2017). Possibilities of use of UAVS for technical inspection of buildings and constructions. *IOP Conf. Ser. Earth Environ. Sci.* 95 (3), 032001–6. doi:10.1088/1755-1315/95/3/032001

Biering, C., Hanus, J. P., and Verma, R. "Day in court - teaching contract disputes in construction management," in Proceedings of the ASEE Annual Conference and Exposition, New Orleans, USA, 2016-June, 1–11. paper ID 16249. doi:10.18260/p.26643

Burke, C. M., and Morley, M. J. (2016). On temporary organizations: A review, synthesis and research agenda. *Hum. Relat.* 69 (6), 1235–1258. doi:10.1177/0018726715610809

Cheung, S. (1999). Critical factors affecting the use of alternative dispute resolution processes in construction. *Int. J. Proj. Manag.* 17 (3), 189–194. doi:10.1016/S0263-7863(98)00027-1

Chow, P. T., Cheung, S. O., and Yiu, T. W. (2012). A cusp catastrophe model of withdrawal in construction project dispute negotiation. *Automation Constr.* 22, 597–604. doi:10.1016/j.autcon.2011.12.006

Dziadosz, A., Tomczyk, A., and Kapliński, O. (2015). Financial risk estimation in construction contracts. *Procedia Eng.* 122, 120–128. doi:10.1016/j.proeng.2015.10.015

Hanak, T., and Korytarova, J. (2014). 10.5937/jaes12-6098 = Risk zoning in the context of insurance: Comparison of flood, snow load, windstorm and hailstorm. *Istraz. i Proj. za privredu* 12 (2), 137–144. doi:10.5937/jaes12-6098

Hashem Mehany, M. S., and Guggemos, A. (2015). Risk management for asphalt road construction and maintenance under performance-based contracts. *Int. J. Constr. Educ. Res.* 11 (4), 292–315. doi:10.1080/15578771.2014.990121

Hosseini, S. A., Akbarpour, A., Ahmadi, H., and Aminnejad, B. (2017). Balance of cost, time, and quality related to construction projects regarding the reinforced concrete of underground structures using a meta-heuristic algorithm. *Archives Civ. Eng.* 63 (4), 103–121. doi:10.1515/ace-2017-0043

Hu, W., and He, X. (2014). An innovative time-cost-quality tradeoff modeling of building construction project based on resource allocation. *Sci. World J.* 2014, 1–10. doi:10.1155/2014/673248

Ilter, D., and Dikbas, A. "An investigation of the factors influencing dispute frequency in construction projects," in Proceedings of the COBRA 2009 - Construction and Building Research Conference of the Royal Institution of Chartered Surveyors, Cape Town, South Africa, 10-11 September 2009, 1496–1504.

Ilter, D. (2012). Identification of the relations between dispute factors and dispute categories in construction projects. *Int. J. Law Built Env.* 4 (1), 45–59. doi:10.1108/17561451211211732

Johnson, R. M., and Babu, R. I. I. (2020). Time and cost overruns in the UAE construction industry: A critical analysis. *Int. J. Constr. Manag.* 20 (5), 402–411. doi:10.1080/15623599.2018.1484864

Kermanshachi, S., and Safapour, E. (2019). Identification and quantification of project complexity from perspective of primary stakeholders in us construction projects. *J. Civ. Eng. Manag.* 25 (4), 380–398. doi:10.3846/jcem.2019.8633

Kongsong, W., and Pooworakulchai, C. (2018). Improving for construction contract management of government construction projects. *Int. J. Civ. Eng. Technol.* 9 (7), 253–260.

Lam, K. C., Tao, R., and Lam, M. C. (2010). A material supplier selection model for property developers using Fuzzy Principal Component Analysis. *Automation Constr.* 19 (5), 608–618. doi:10.1016/j.autcon.2010.02.007

Lee, C. K., Yiu, T. W., and Cheung, S. O. (2016). Selection and use of alternative dispute resolution (ADR) in construction projects - past and future research. *Int. J. Proj. Manag.* 34 (3), 494–507. doi:10.1016/j.ijproman.2015.12.008

Leśniak, A., Piskorz, G., Spišáková, M., and Mačková, D. (2018). Causes of delays in construction works resulting from the provisions of the contract in Poland and Slovakia. *Sci. Rev. Eng. Environ. Sci.* 27 (1), 71–81. doi:10.22630/PNIKS.2018.27.1.7

Ljevo, Ž., Vukomanović, M., and Rustempašić, N. (2017). Analysing significance of key quality factors for management of construction projects. *Gradjevinar* 69 (5), 359–366. doi:10.14256/JCE.1723.2016

Lu, W., Zhang, L., and Pan, J. (2015). Identification and analyses of hidden transaction costs in project dispute resolutions. *Int. J. Proj. Manag.* 33 (3), 711–718. doi:10.1016/j.ijproman.2014.08.009

Maravas, A., and Pantouvakis, J. P. (2012). Project cash flow analysis in the presence of uncertainty in activity duration and cost. *Int. J. Proj. Manag.* 30 (3), 374–384. doi:10.1016/j.ijproman.2011.08.005

Milen, D. R., Schriefer, A., Lehder, D. Z., and Dray, S. M. "Mind maps and causal models: Using graphical representations of field research data," in CHI EA '97 CHI '97 Extended Abstracts on Human Factors in Computing Systems, Atlanta, Georgia, March 1997 (ACM), 265–266.

Mishmish, M., and El-Sayegh, S. M. (2018). Causes of claims in road construction projects in the UAE. *Int. J. Constr. Manag.* 18 (1), 26–33. doi:10.1080/15623599.2016.1230959

Molenaar, K., Washington, S., and Diekmann, J. (2000). Structural equation model of construction contract dispute potential. *J. Constr. Eng. Manag.* 126 (4), 268–277. doi:10.1061/(ASCE)0733-9364(2000)126:4(268)

Molnar, Z., Midleova, S., Rezankova, H., Brixi, R., and Kalina, J. (2012). *Pokrocile metody vedecke prace (Advanced methos of scientific work)*. Czech Republic: VŠE. 978-80-7259-064-3.

Nahidi, S., Fountas, G., Sarvani, S. P., Sarwar, M. T., and Anastasopoulos, P. C. (2017). Project discrepancies in roadway construction and preservation: A statistical analysis of public–private partnership contract types in the US. *Front. Built Environ.* 3. doi:10.3389/fbuil.2017.00015

Naji, K. K., Mansour, M. M., and Gunduz, M. (2020). Methods for modeling and evaluating construction disputes: A critical review. *IEEE Access* 8, 45641–45652. doi:10.1109/ACCESS.2020.2976109

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.

RIBA Enterprises, (2013). National construction contracts and Law Survey 2013. Available at: https://www.thenbs.com/knowledge/nbs-national-construction-contracts-and-law-survey-2013.

Rybka, I., Bondar-Nowakowska, E., Pawluk, K., and Połoński, M. (2017). Risk of contractors' claims on the example of road works. *IOP Conf. Ser. Mat. Sci. Eng.* 245 (7), 072009–9. doi:10.1088/1757-899X/245/7/072009

Safa, M., Shahi, A., Haas, C. T., and Hipel, K. W. (2017). Construction contract management using value packaging systems. *Int. J. Constr. Manag.* 17 (1), 50–64. doi:10.1080/15623599.2016.1167369

Shoar, S., Chileshe, N., and Edwards, J. D. (2022). Machine learning-aided engineering services' cost overruns prediction in high-rise residential building projects: Application of random forest regression. *J. Build. Eng.* 50, 104102. doi:10.1016/j.jobe.2022.104102

Subramanya, K., Kermanshachi, S., and Pamidimukkala, A. (2022). Evaluation of E-Ticketing technology in construction of Highway projects: A systematic review of adoption levels, Benefits, Limitations and Strategies. *Front. Built Environ.* 8. doi:10.3389/fbuil.2022.890024

Tazelaar, F., and Snijders, C. (2010). Dispute resolution and litigation in the construction industry. Evidence on conflicts and conflict resolution in The Netherlands and Germany. *J. Purch. Supply Manag.* 16 (4), 221–229. doi:10.1016/j.pursup.2010.08.003

Vařbuchta, P., Kovářová, H., Hromádka, V., and Vítková, E. (2017). Risk variables in evaluation of transport projects. *IOP Conf. Ser. Mat. Sci. Eng.* 236 (1), 012112–012117. doi:10.1088/1757-899X/236/1/012112

Vidal, L., Marle, F., and Bocquet, J. (2011). Measuring project complexity using the analytic hierarchy process. *Int. J. Proj. Manag.* 29 (6), 718–727. doi:10.1016/j.ijproman.2010.07.005

Vidogah, W., and Ndekugri, I. (1998). Improving the management of claims on construction contracts: Consultant's perspective. *Constr. Manag. Econ.* 16 (3), 363–372. doi:10.1080/014461998372385

Yaskova, N., and Zaitseva, L. (2017). Application of alternative dispute resolution in the field of construction projects. *IOP Conf. Ser. Earth Environ. Sci.* 90 (1), 012182–012184. doi:10.1088/1755-1315/90/1/012182

Zujo, V., Car-Pušić, D., Zileska-Pančovska, V., and Cecez, M. (2017). Time and cost interdependence in water supply system construction projects. *Technol. Econ. Dev. Econ.* 23 (6), 895–914. doi:10.3846/20294913.2015.1071292

frontiers | Frontiers in Built Environment

TYPE Original Research
PUBLISHED 08 November 2022
DOI 10.3389/fbuil.2022.956066

New trends in assessing the prestress loss in post-tensioned concrete bridges

Mariano Angelo Zanini[1]*, Flora Faleschini[1,2] and Carlo Pellegrino[1]

[1]Department of Civil, Environmental, and Architectural Engineering, University of Padova, Padova, Italy,
[2]Department of Industrial Engineering, University of Padova, Padova, Italy

OPEN ACCESS

EDITED BY
Joan Ramon Casas,
Universitat Politecnica de Catalunya, Spain

REVIEWED BY
Michele Fabio Granata,
University of Palermo, Italy
Yuguang Yang,
Delft University of Technology, Netherlands

*CORRESPONDENCE
Mariano Angelo Zanini,
marianoangelo.zanini@unipd.it

SPECIALTY SECTION
This article was submitted to Bridge Engineering,
a section of the journal
Frontiers in Built Environment

RECEIVED 29 May 2022
ACCEPTED 20 October 2022
PUBLISHED 08 November 2022

CITATION
Zanini MA, Faleschini F and Pellegrino C (2022), New trends in assessing the prestress loss in post-tensioned concrete bridges.
Front. Built Environ. 8:956066.
doi: 10.3389/fbuil.2022.956066

The evaluation of the current prestress force represents often a challenging issue during the assessment of existing post-tensioned concrete (PTC) members. In this contribution, two case studies are presented to show the efficiency of some experimental methods applied both to laboratory beams and to an in-service PTC bridge. First, the outcomes of an experimental campaign carried out on three PTC beams with a straight post-tensioned, non-grouted tendon are discussed. For each beam, three-point bending tests (3PBTs) were carried out applying increasing load steps until failure and varying the prestress force; at each load step, non-destructive tests (NDTs) were performed, namely, dynamic free vibration and ultrasonic tests. The variation of non-destructive parameters was evaluated as a function of the prestress force and of the structural damage. The second case study deals with a 50-year-old PTC bridge with grouted ducts. In this case, other than evaluating the existing geometry, structural details, and material properties, the following NDTs were performed: endoscopies, tests based on stress release, i.e., both saw-cut and wire-cut tests, and lastly, X-ray diffractometries (XRDs). The results indicate the high potential of this latter technique, which resulted more reliable and less invasive than the previous ones.

KEYWORDS

dynamic identification, non-destructive tests, post-tensioned concrete, prestress loss, structural assessment, x-ray diffraction, bridges

1 Introduction

In the field of civil infrastructures, the proper management of stocks of existing bridges and other civil structures has become an extremely complex issue that involves technical, social, economic, and political considerations. In this context, several research projects were financed aimed at identifying the broadly recognized procedures and strategies that can be applied by public and private authorities for the management of infrastructural assets (COST345, 2004; SAMARIS, 2005; COST TU1404, 2014). However, while this issue is gaining increasing social attention with time, as a consequence of structural failure events that had a great media impact (Faleschini et al., 2018; Calvi et al., 2019), there is still the need for more research efforts on this topic. Indeed, the current practice on the performance assessment of bridges, and more generally infrastructures, suffers from several deficiencies, the most well known of that being based on visual observations of structural damage only, thus being limited on qualitative data (Zanini et al., 2017a; Quirk et al., 2018). In-depth investigations are instead needed when safety and serviceability are evaluated during structures service life, and for them, it is necessary to improve and/or establish reliable procedures and techniques to carefully assess the current health condition of structures. In the field of post-tensioned concrete (PTC) structures, this becomes even more fundamental because the variation of prestress force within the element is an additional parameter of uncertainty (Osborn et al., 2012), other than the effects of deterioration (Tu et al., 2019), variation of the external loads (Crespo-Minguillón and Casas 1997), time-dependent phenomena (Oh and Yang 2000) involving material properties variation (Badalassi et al., 2017), e.g., for concrete, creep and shrinkage, and steel relaxation, natural hazard occurrence (Zanini et al., 2017b), etc.

The knowledge of the residual prestress force in existing PTC bridges is considered one of the fundamental keys for understanding the actual health condition of the whole structure. Tendons represent in fact the main load-carrying elements in this type of structural elements. Different deterioration phenomena may accelerate the prestress loss, which are often linked to corrosion processes acting on the tendons. Other than the well-known carbonation and chloride-induced corrosion (Angst et al., 2020; Berto et al., 2020), the stress-corrosion cracking represents another harmful phenomenon because the interaction between tensile stresses acting on a tendon and the corrosion favors the speed of the reaction (Calabrese and Proverbio, 2020). Furthermore, hydrogen embrittlement is also considered one possible source for the deterioration of metals in general, and thus, hazardous for steel tendons (Aliofkhazraei et al., 2021). Additionally, it should be recalled that an insufficient design of the drainage system, poor material properties (low-strength concrete with many voids leading to high diffusion of carbon and chlorides in the material), mistakes during the construction process (e.g., for grouted ducts, the well-known problem of the partial injection of the ducts with mortar, its segregation and the availability of free water to come into contact with the tendons inside the duct), and lastly insufficient concrete covers, make the PTC bridges high-vulnerable structures. Brittle collapses of these kind of structures were recorded and described in the literature (Moravčík and Bujňáková, 2017; Anania et al., 2018), demonstrating that PTC bridges can suddenly fail due to the reduction of the load-carrying capacity provided by the sole tendons. Furthermore, if a strand fails, the expected redistribution of stresses to the others might not occur, especially when a large number of voids are present in the grout (Martin et al., 2001).

A substantial difference between design and current prestress force might lead to serviceability and safety impairments (Saiidi et al., 1994), linked to the increase of tensile stresses in concrete and to the arise of possible cracking phenomena, with consequent effects on the exposure of reinforcement, and lastly leading to potential structural failure. Several studies highlighted discrepancies when experimental residual prestress force has been compared with the theoretical one, as predicted applying different codes, for prestressed members that were extracted from structures in service for more than 25 years (Azizinamini et al., 1996; Bagge et al., 2017). Such deviations (typically underestimating prestress loss) might be significant even if most of the codes are generally conservative in the estimation of prestress losses (Garber et al., 2015), which occur within the first months after the element casting and that are relatively easy to be predicted. Hence, such higher losses are associated to the aforementioned phenomena linked to material degradation, long-term effects, construction defects, or structural damage, that may lead to a progressive deterioration of the prestressing level along the PTC member.

In this context, the recent Italian Guidelines for the Risk Classification and Management of Bridges (Ministry of Italian Infrastructure, 2020) have classified the PTC bridges as high-vulnerable structures, for which specific assessment is often necessary. One of the key parameters that need to be investigated properly is the prestress loss, which should be estimated for a sufficient number of tendons with instrumental techniques operated *in situ*. However, there is still an open question about which test method is more suitable to obtain a reliable evaluation of this measure.

This work aims to describe some new insights about estimating the prestress loss in PTC members, by providing an additional contribution to the state-of-the-art of the scientific literature devoted to identify the most proper investigation techniques to be implemented for the quantification of prestress losses in PRC elements based on recent authors experiences carried out both in laboratory and *in-field* tests. First, a review of the existing available methods proposed in the literature is presented, with the aim to give an overview of current methods that can be implemented singularly or in combination with each other for testing PTC members. Then, the results of a first series of experimental tests on PTC laboratory beams are discussed: beams have 6 m length and a straight post-tensioned tendon, and were subject to varying prestress forces and damage. Lastly, *in situ* tests on a 50-year PTC bridge with grouted ducts were carried out. In all the cases, complementary tests were executed, aiming to compare the different techniques and the obtained results, which are critically discussed.

TABLE 1 Current experimental methods applied to PTC bridges for investigating prestress loss.

Method	Advantage	Disadvantage
Visual inspection	Low cost. Not invasive	Few relevant as PTC bridges rarely show distress and relevant cracking pattern before failure
Manual drilling of tendon duct with visual inspection using endoscopy	Medium cost. It provides direct evidence of the state of conservation of the tendons inside the duct	It may be dangerous if tendons are cut during the drilling operations. The statistical relevance is limited and depends on the number of tests. It can be applied only in some parts of the section (generally, at PTC beam bottom, depending on tendons longitudinal profiles)
Acoustoelastic methods (based on wave velocity)	Medium-low cost; not invasive; it provides useful results about voids presence when used to realize UPV (ultrasonic pulse velocity) maps with a tomographic input/output scheme	The results may be not conclusive when large amount of reinforcement is present, as it may shield the presence of voids
Impact-echo	Medium cost; not invasive. If a sufficiently high-frequency is used with respect to the defect, it generally provides good results to detect voids in the ducts	The presence of inhomogeneities in the grout may affect the results
Ground-penetrating radar (GPR)	Medium cost. Not invasive. Effective to detect voids in plastic ducts (present in recent bridges). Useful to localize metallic ducts	The main limitation is that it cannot be used to detect voids in steel ducts (only in non-metallic ducts)
Strain-based methods (on concrete or steel)	It provides an evaluation of strain and stress acting in the analyzed section. It may be applied directly to a single strand (exposed) or to a concrete portion. If applied to concrete, is considered an NDT	Medium–high cost; it is necessary to monitor also the temperature with thermocouples; strain variations in the strand may be in the same order of those due to temperature variations. Indirect methods
Radiography	It may be able to detect both the corrosion of wires or strands and the grouting consistency in ducts	High costs. Generally, cannot be carried out with open traffic on the bridge. Radiation risk for the operators (highly skilled)
Vibration-based tests (Modal testing with Forced Vibration or Ambient Vibration)	Not invasive; evaluate the overall behavior of the PTC bridge; often applied for damage detection. It may be useful if more tests in time are available	High cost; it may show insignificant changes of the main frequencies and modal shapes until relatively high prestress losses
Load test	Evaluate the overall behavior of the PTC bridge	High cost; it may be dangerous if the PTC bridge is already damaged; failure may occur before any meaningful deflection response is obtained

2 Review on current experimental methods to evaluate the prestress loss

The direct access to prestressing tendons is often a hard task, especially for those grouted. Thus, the direct evaluation of the effective strain, and thus, stress acting on each strand inside a duct becomes impossible in the practice, unless a monitoring system is conceived since the design of the PTC element, which should be applied before concrete casting. Hence, a series of indirect experimental methods have been proposed in the past and are currently used in the practice by bridge inspectors and laboratories to evaluate the current stress acting on the tendons. Table 1 lists a comparison of the most common methods with the aim to help a reader to select the most proper testing method or a combination of test methods in relation to different specific conditions, aims of the experimental campaign, field conditions, etc., to achieve a specific goal when testing post-tensioned concrete members.

Other than these tests methods, some more recent experimental procedures were proposed and applied, worth mentioning that based on the X-ray diffraction technique (Morelli et al., 2021), which is an innovative NDT (non-destructive test) applied to PTC bridges. Such technique is commonly diffused in material science, as it is often used to analyze the microstructure of crystalline materials and it is based on the well-known Bragg equation:

$$n \cdot \lambda = 2 \cdot d \cdot \sin\theta, \quad (1)$$

which states that when the X-ray (with wavelength λ) is incident onto a crystal surface, its angle of incidence θ will reflect back with the same angle of scattering θ. When the path difference d (which is the distance between the planes of the crystalline lattice) is equal to the whole number n of wavelength λ, a constructive interference will occur. Such technique has been also applied to measure the so-called residual stresses in different applications, since more than 25 years (Prevéý, 1996). If the crystalline lattice is subject to a stress state, the interatomic distance modifies and this makes in turn a variation of the angle of scattering θ. Knowing the material composition (i.e., the steel type used for the strand realization) and its elastic properties (i.e., the elastic modulus E and the Poisson coefficient υ), it is possible to directly

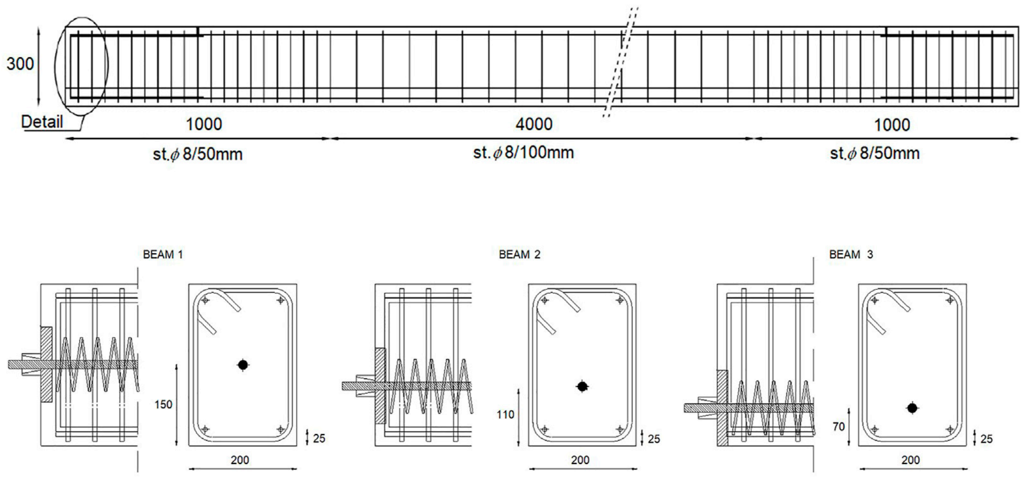


FIGURE 1
Laboratory PTC beams geometry and reinforcement arrangement.

reconstruct the stress and strain state applied to the material. In Section 4 of the present work, more details on this technique are provided.

3 Laboratory tests on post-tensioned concrete beams with straight unbonded tendons

In this section, the results of an experimental campaign on three PTC beams with non-adherent (unbonded) strands, inserted in plastic ducts and placed at different eccentricity, are shown. The aim of this experimental campaign is to evaluate the potential of applying two NDT techniques, that is, a vibration-based and the UPV tomography, for damage assessment and the evaluation of prestress loss. The complete test report of this experimental campaign is provided in (Frizzarin, 2019).

3.1 Post-tensioned concrete beam geometry and material properties

All the beams are characterized by the same geometry and steel reinforcement but they differ from the position of the tendon, which have a different eccentricity and is non-grouted (see Figure 1). A rectangular cross section and a constant eccentricity e is selected to represent an idealized configuration of the PTC member and to remove potential secondary effects linked to the tendons layout, that can make understanding the results more difficult. The prestressing force was applied to each beam tensioning the strand with a hydraulic hollow jack, positioned at one end of the beam. The strand was locked with wedges, and at the interface between the hydraulic jack and the concrete surface 20 mm thick plates were positioned, in order to distribute the concentrated load on a 200 × 200 mm^2 area. The amount of the prestress force is varied during the test through a hydraulic jack. Table 2 summarizes the main features of these beams.

Concerning material properties, the concrete cubic compressive strength is $f_{c,cube}$ = 64.7 ± 2.5 MPa, evaluated at the time of testing (about 55 days after concrete casting). For reinforcement, ordinary B450C steel was used, with experimental average yield strength f_y = 512 MPa at about ε_y = 0.2%, tensile strength f_t = 612 MPa, and average elongation = 9%. The prestressing steel was also experimentally tested, obtaining in average a nominal strength at 1% of total elongation $f_{p(1)k}$ = 1703 MPa, tensile strength f_{pt} = 1925 MPa, and average elongation = 6.4%.

3.2 Test setup

PTC beams were subjected to 3PBTs, with a test program that includes a series of load–unload cycles up to failure. Between each loading cycle, NDTs were carried out, at different values of the prestress force (regulated by the hydraulic jack). Figure 2A shows the loading scheme: note that the vertical load application point is asymmetric. In the same figure, the instrumentation adopted is indicated. Other than the vertical load and the prestress force, the static parameters were continuously also acquired during the test : vertical displacements along the length of the element through five LVDTs (linear variable differential transducers) with a precision of ± 0.01 mm;

TABLE 2 Main features of the laboratory PTC beams.

ID	L (m)	L_s (m)	A_c (mm^2)	A_p (mm^2)	F_p (kN)	σ_c (MPa)	e (mm)	e/h (-)
Beam 1	6.00	5.60	60000	139	0–160	0–2.67	0	0
Beam 2	6.00	5.60	60000	139	0–160	0–2.67	40	0.134
Beam 3	6.00	5.60	60000	139	0–160	0–2.67	80	0.267

L = length of the element; L_s = clear span between the supports; A_c = concrete section with (b x h) = (200 × 300) mm^2; A_p = prestressing steel area, made with a single 7-wires single strand, equivalent diameter 6/10"; F_p = range of prestressing force applied to the strand; σ_c = range of prestress acting on the concrete; e = strand eccentricity; e/h = relative eccentricity compared to the section height.

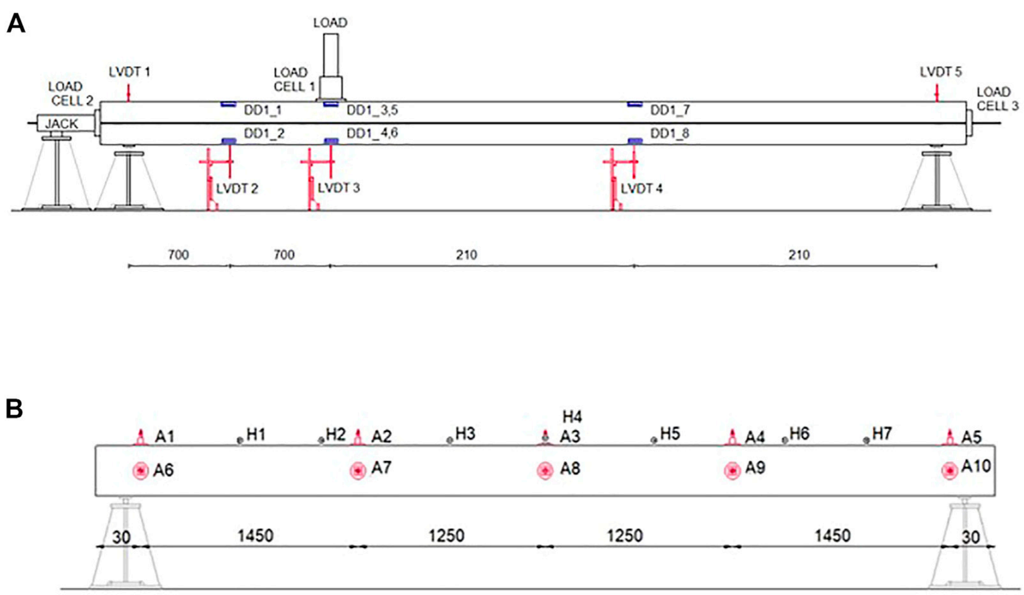


FIGURE 2
Test setup and instrumentation disposition: **(A)** 3PBT test and **(B)** NDT1 test (A = accelerometers; H = instrumented hammer).

concrete strains at relevant locations, adopting four couples of DD1s (distortion detectors) with ± 0.001 mm precision, placed closed to the top and bottom of the sections, to reconstruct the strain profile at each analyzed section.

The NDTs applied in this experimental campaign are 1) dynamic free vibrations measurements and 2) ultrasonic tests. The first test method (NDT1) was carried out using 10 accelerometers, located at relevant positions of the element and an instrumented hammer that triggers the impulsive force. The acquisition system was a NI PXI-1042Q, which allowed to record the signals that were further analyzed in the frequency domain through frequency domain decomposition (FDD) to derive the modal parameter of interest. Figure 2B shows the locations of the accelerometers and those where the impulsive force was applied. The second test method (NDT2) deals with the analysis of the ultrasonic pulse waves, adopting a tomography configuration for the instrumentation. To achieve such configuration, the distances defining the net between each emitter and receiver shear wave transducers are 30 cm in length and 20 cm in width. Each test consists of a triplet of ultrasonic measurements with probe frequency equal to 54 kHz, being one direct and two inclined, with the aim of covering the majority of the element surface. The signal acquisition was performed using the *PunditLab* + acquisition system with the *PunditLink* software.

The NDTs were carried out always on the unloaded beams (after flexural loading, at varying prestress forces). The following test procedure and loading history were adopted in this experimental campaign:

1. NDTs on beams with no flexural damage, carried out varying the prestress force from 0% to 100% in steps

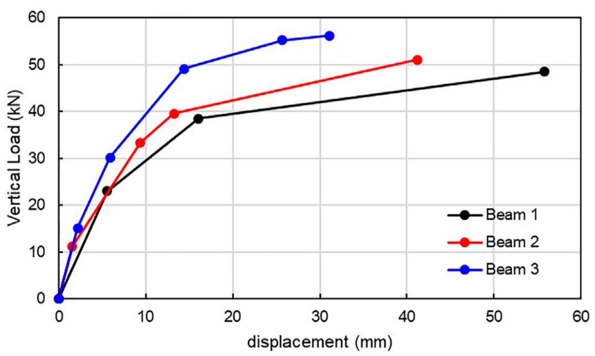


FIGURE 3
Vertical load vs. displacement (at LVDT3) curves for the laboratory PTC beams.

2. NDTs on beams with slight–moderate flexural damage (loaded before cracking and then unloaded), carried out varying the prestress force from 100% to 0% in steps
3. NDTs on beams with high flexural damage (loaded above 60% of the expected ultimate load and then unloaded), carried out varying the prestress force from 100% to 0% in steps
4. NDTs on beams with a complete flexural damage (loaded until failure P_{max} and then unloaded), carried out varying the prestress force from 100% to 0% in steps

3.3 Results

Figure 3 shows the results obtained in terms of load-displacement curve for the PTC beams, where the displacement is evaluated at the load application section (LVDT3). As expected, Beam 1 (with $e = 0$) showed a lower failure load of 48.51 kN and a higher deformability with respect to Beam 3 (with the highest e value), failed at 56.21 kN and a 45% reduced vertical displacement value, whereas Beam 2 showed intermediate outcomes, with a failure load of 51.09 kN.

Regarding dynamic tests (NDT1), fundamental frequencies were calculated from the recorded signals. Then, modal shapes of the tested beams were also evaluated, particularly analyzing the first two vertical and two transverse modes, with the hypothesis of adopting two perfect hinges at the supports.

The first result is that the first fundamental frequency is affected by the flexural damage, as already well known from the literature (Casas and Aparicio, 1994; Gentile and Saisi, 2007; Venanzi et al., 2020; Pisani et al., 2021). Even if the fundamental frequency is not always highly sensible to damage occurrence, here, Figure 4 shows clearly the variation of this parameter (in the y-axis) with the applied flexural load (in the x-axis), at varying prestress force values. The reported results in terms of frequencies are the average from multiple readings; the coefficient of variation is very low in all the cases, about 1%. The lightest-colored lines identify the behavior of the beams with full prestress: from their curves, a decreasing trend of the fundamental frequency with the increase of the applied flexural load is visible. The maximum reduction ranges from 12 to 30% when the beams are fully damaged, depending on the strand disposition in the section. In the same figure, the darker-colored lines indicate the behavior of the beams with reducing prestress force while the fundamental frequency value is almost the same in the undamaged condition regardless of the prestress force, there is a non-negligible effect after the cracking condition. The reduction in the fundamental frequency occurs as a consequence of crack formation when the vertical load increases: the contribution of increasing the prestress force is to restrain the cracks opening, thus limiting the reduction of this parameter. Such results underline how a single dynamic test cannot give additional knowledge about the circumstance of a prestress loss, if the PTC member has no relevant cracks, i.e., detectable with a visual survey of the element surfaces. On the other side, if relevant cracks can be visually detectable, it may be not necessary to perform a dynamic test to conclude that a prestress loss occurred, thus concluding how a single dynamic

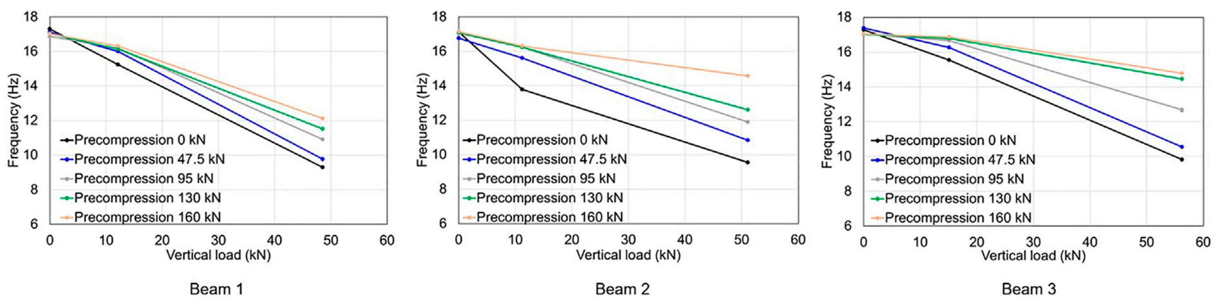


FIGURE 4
Average fundamental frequency as a function of the applied flexural load and prestressing force.

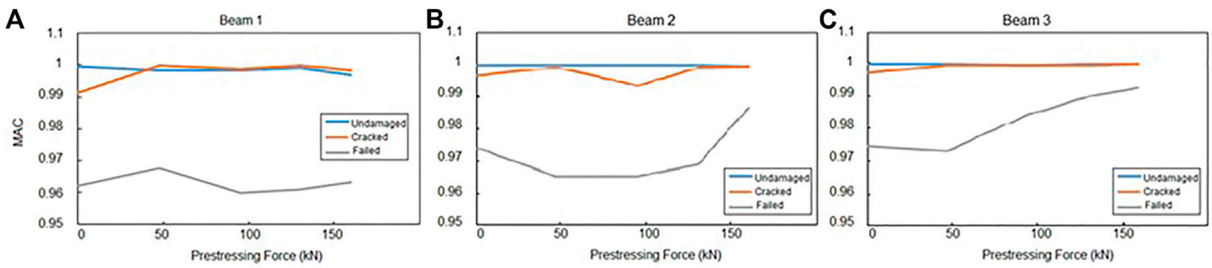


FIGURE 5
MAC index as a function of the applied flexural load (undamaged - cracked - failed) and prestressing force for the first vertical modal shape.

test, without any other additional tests or information, cannot help to understand the condition state of PTC members in ordinary conditions (when they are in operation and do not show relevant crack patterns).

Modal shape changes were quantified by the Modal Assurance Criterion (MAC), which provides a measure of the correlation between two sets of modal shapes, evaluating their least square deviation (Ewins, 2009). A MAC value close to 1 means that modes differ little, while a value close to 0 means that the modes are totally different. Figure 5 shows MAC value variation for the first vertical mode for each of the analyzed PTC beams, resulting in a similar trend to those already observed for natural frequencies, even if characterized by less sensitivity to damage presence. The maximum reduction ranges from 3 to4% in the presence of a full damage imposed by the external flexural load; the influence of the prestress force is almost negligible. Such result seems to underline how in the case of PTC members continuously monitored with a SHM system, such performance indicator may not be the best choice to highlight the occurrence of non-negligible crack patterns.

Regarding the ultrasonic tests (NDT2), first maps of the velocity were drawn from the interpolation of the collected values by developing an algorithm in the Matlab® environment (for more details see Frizzarin 2019). The results obtained from this technique were not encouraging: even if it was possible to identify a general decreasing trend of the UPV with the increase of damage (see Figure 6 for Beam 1, as an example, together with its cracking pattern at failure), for the highest levels of flexural applied load, the ultrasonic signal suffers for a great signal disturbance. Accordingly, it was not possible to calculate the signal velocity at each location of the beam, revealing a limitation of this method. Hence, similar conclusions can be drawn to those discussed for the few utility of dynamic tests for the quantification of potential prestress losses.

Then, the attenuation of the ultrasonic signal was also evaluated through the computation of the area underlying the power spectral density (PSD) of the ultrasonic signal, measured in time. This calculation ensures to find a result for those signals that also suffer from disturb , where the time of flight of the signal (and therefore velocity) are difficult to be assessed clearly. Results of such analysis are presented in Figure 7 in the shape of tomographic contour plot. There, attenuation plots are presented for the case of Beam 1 in the undamaged condition (a) and after complete damage (b), using the data from the y-direction. This means that Figure 7 shows the attenuation across the beam height (0.3 m), over about 1.5 m length of the element. The main result is that moving from the null to full prestress acting in the element, the attenuation decreases, even if results are not always clear in average. The role of prestressing in the closure of cracks is to reduce the signal attenuation. Results show some correlation with flexural damage too, comparing the maps of Beam 1 when subject to complete damage to the undamaged condition, the UPV attenuation values are, in average, higher. Furthermore, these plots give a quite detailed view not only on the presence of the damage but also of the location of the cracks (see for comparison, Figure 6C).

4 Tests on an in-service 50-year post-tensioned concrete bridge

In this section, the results of an experimental campaign on an in-service 50-year PTC bridge, built in 1968 in Padova district, Italy, are shown. The case study is a three-span bridge, with a single carriage and one vehicle and one pedestrian lane. The deck is made with four simply-supported double-T beams, with an average length of 32.10 m and height of 1.60 m (Figure 8). The distance between each web is 2.50 m, whereas the collaborating slab has 0.2 m width. Overall, the deck width is 10 m; the beams are transversally connected through intermediate elements, each placed at 8 m from the other. Concerning the prestressing steel, from the original design, it was possible to know that eight prestressing tendons are present in each beam, realized with a different number of single wires (with diameter ϕ = 7 mm), depending on their position. Furthermore, their tensile strength was f_{ptk} = 1650 MPa and their stress during exercise condition (after losses) was design to be about f_{pe} = 889 MPa. The state of condition of the PTC beams is the objective of this study:

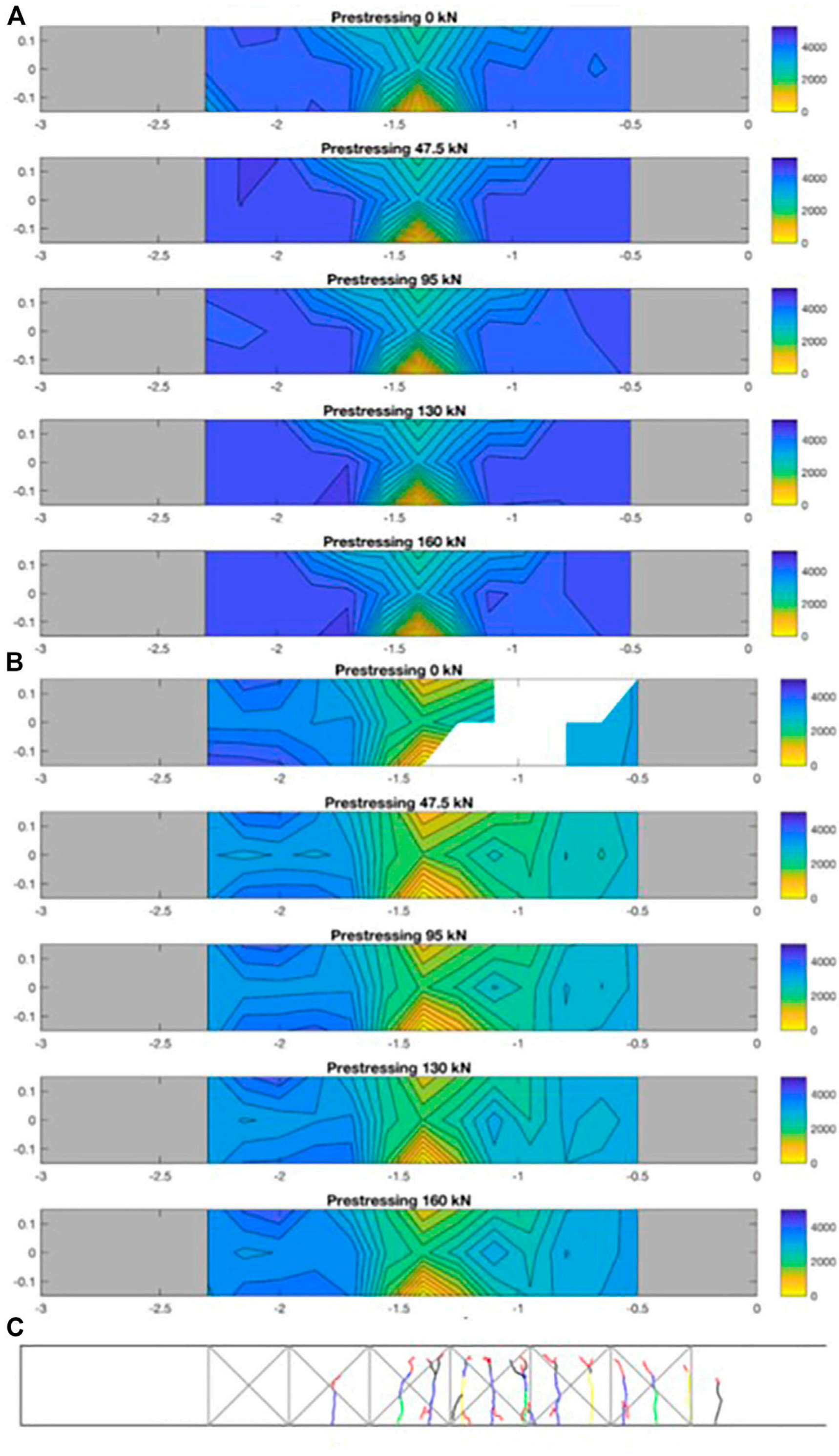


FIGURE 6
UPV tomography (m/s) in the vertical direction (PTC cross-section height = 0.3 m) of Beam 1 in **(A)** undamaged condition, **(B)** after complete damage, and **(C)** cracking pattern at failure.

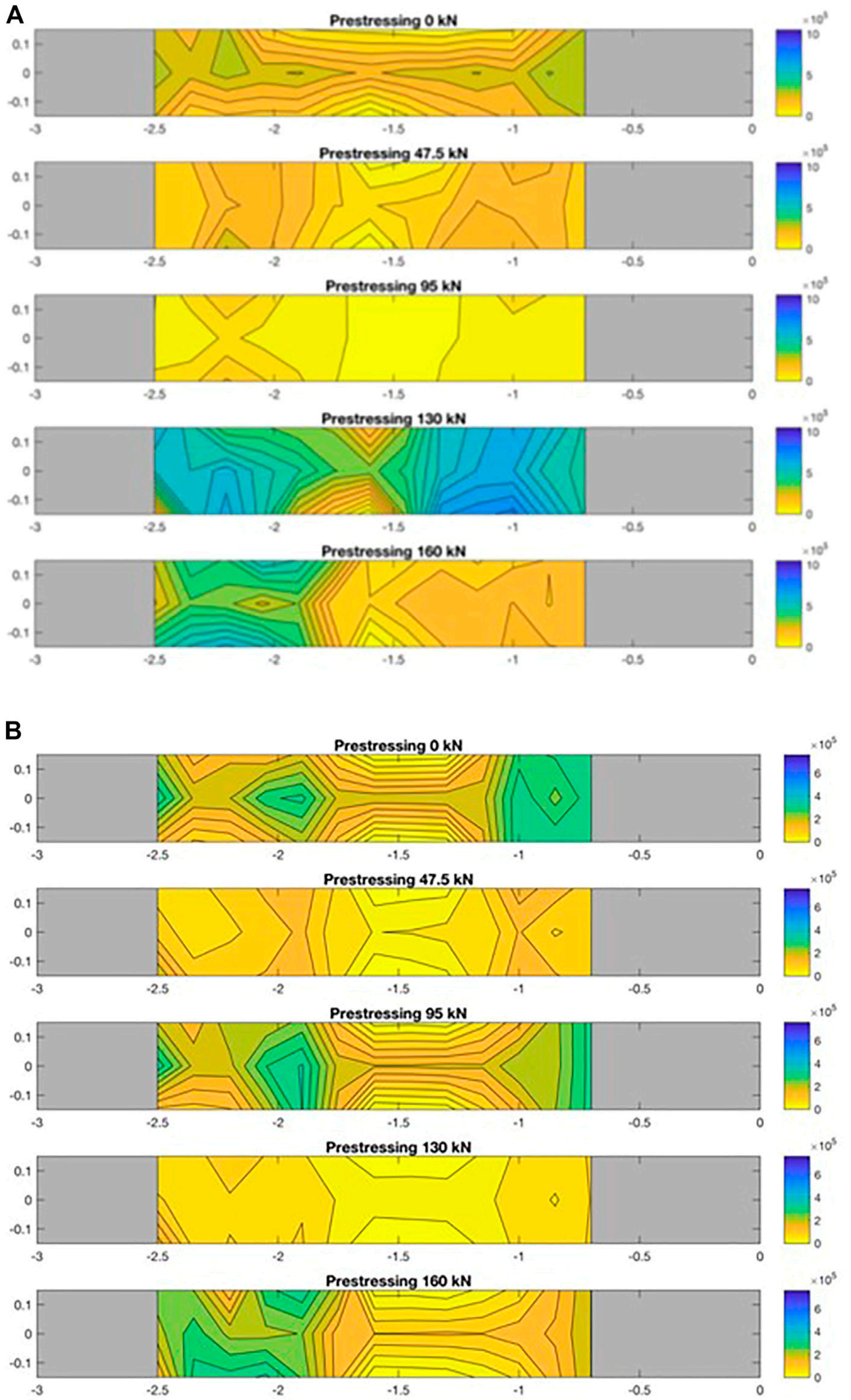


FIGURE 7
UPV attenuation (m/s) in the vertical direction (PTC cross-section height = 0.3 m) of Beam 1 in **(A)** undamaged condition and **(B)** after complete damage.

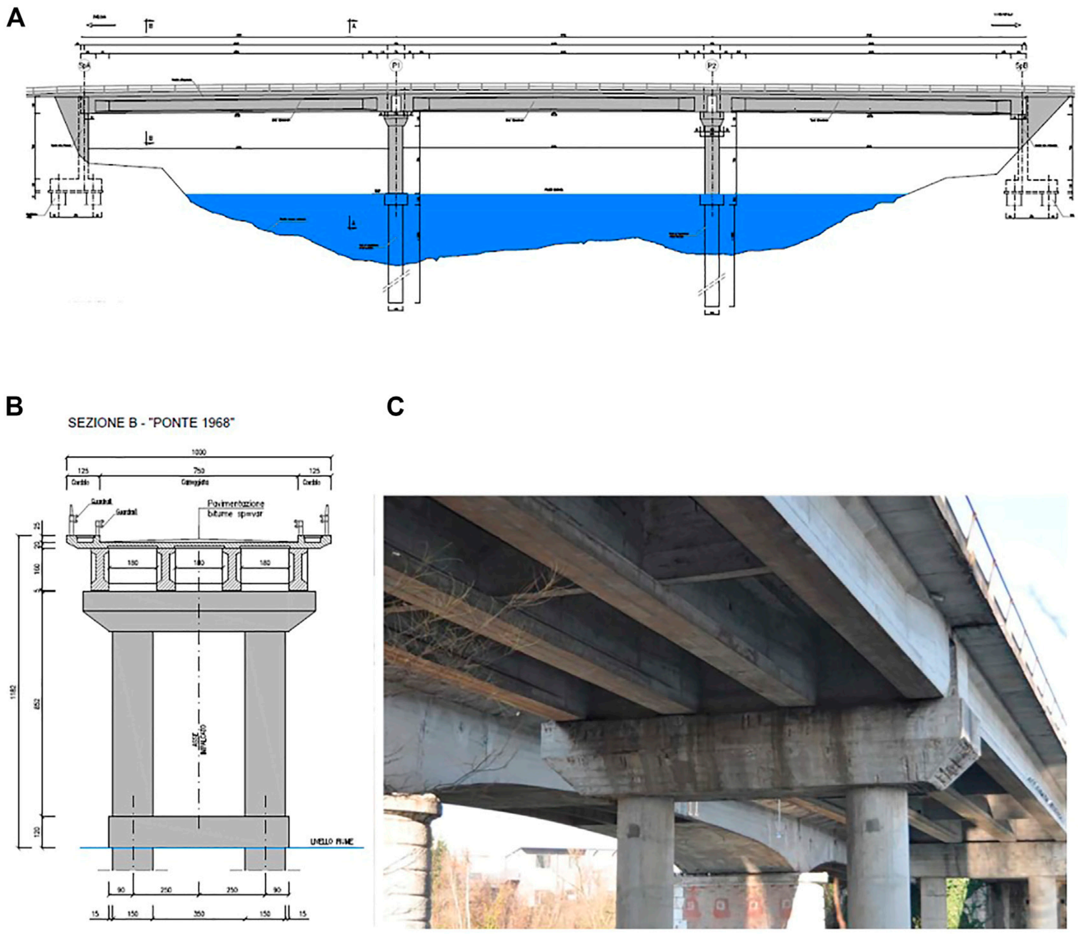


FIGURE 8
Case study: four PTC beams of a 50-year bridge with post-tensioned tendons and grouted ducts: **(A)** longitudinal view, **(B)** transversal cross-section, and **(C)** detail of the current condition state.

particularly, here, different NDTs (endoscopies, stress-release methods, and XRD) were applied to investigate the residual prestress force acting on the grouted tendons. Preliminary, visual inspections, the assessment of material properties and the current tendons layout along the elements have been evaluated.

4.1 Preliminary tests: Visual inspection, material properties, and tendon disposition

A preliminary visual inspection was carried out to evaluate the main deterioration phenomena active on the structural elements. The state of conservation of the deck was found to be generally good with regard to the main prestressed beams, with no visible damage, the same applies for the slab intrados. This result was explained by the high quality of the concrete, the absence of obvious defects in drainage system, and almost ordinary environmental conditions (limited freezing/thawing cycles, no relevant sources for chlorides, limited air pollution), which are all factors that contributed to maintain sufficiently well the material properties. Instead, the piers cap and edge elements display some diffused deterioration, with concrete cover expulsion, steel bars oxidation, and some corrosion phenomena.

The results of two successive experimental campaigns performed in 2019 and 2021 were collected together, aiming at improving the knowledge on material properties. The main results deal with concrete cores and steel bars extraction and testing, other than some NDTs such as *in situ* SONREB tests (coupling ultrasonic and rebound tests with Schmidt hammer) on concrete and Brinell hardness tests on steel bars. The tests were carried out on the double-T beams, transverse beams, slab, pile caps, piles, and abutments.

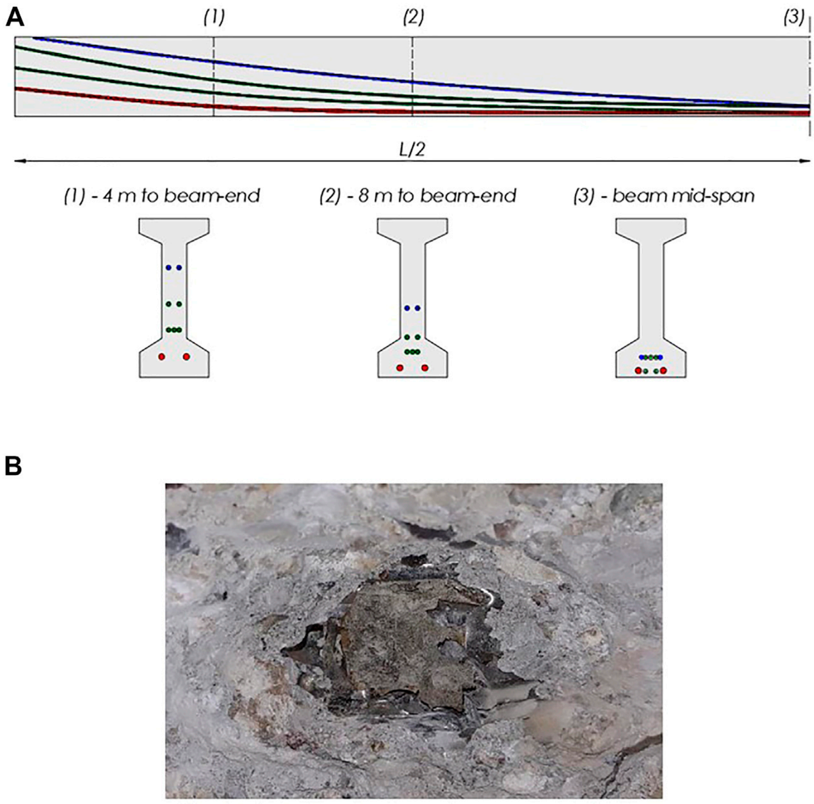


FIGURE 9
(A) Longitudinal profile of prestressing tendons along the beam length (half-span) and cross-section geometry of the double-T beams in three relevant sections and **(B)** condition state of the metallic duct and internal grout.

Specifically, for the double-T beams, the average *in situ* concrete compressive strength f_c = 43.8 MPa, whereas the ordinary steel is characterized by yielding strength f_y = 531 MPa. Concerning the slab, instead, material properties were characterized by f_c = 57.0 MPa and f_y = 459 MPa, for concrete and steel, respectively.

The cross-section, ordinary steel and prestressing tendon disposition in the PTC elements were investigated by means of visual inspection and electromagnetic method (i.e., using a covermeter for some details and GPR for both ordinary rebars, steel ducts localization, and prestressing tendons identification, adopting different acquisition setups depending on the type of reinforcement analyzed). According to the GPR results, it was possible to confirm the information available from the design of the bridge and to reconstruct the tendons layout. Specifically, Figure 9A shows the longitudinal profile of the tendons along the beams' length, and three relevant sections of the beams with the details about the number of prestressing wires on each tendon. Note that from this technique no information about the possible presence if voids inside the tendons can be obtained. Indeed, the steel ducts reflect the waves and do not allow any penetration inside.

4.2 Non-destructive tests for prestress loss assessment

The following NDTs were carried out to obtain useful information about the current stress state of the prestressing steel:

- Micro-coring and endoscopies on two positions for each beam in two over three spans of the bridge (i.e., 16 point of measures)
- Tests based on stress release on both concrete and directly on the prestressing steel (one position for each beam)
- XRD tests (in the same position of stress release tests)

Particularly, the micro-cores were carried out to evaluate if there are voids or signs of corrosion in relevant positions that are the intrados of the PTC beams at the mid-span. In fact, in the absence of visible defects on, e.g., drainage system, which may be responsible for inducing damage in one specific section, the most vulnerable section is considered the mid-span. There, given the parabolic profiles of the prestressing tendons, they have the minimum concrete cover. Furthermore, due to their

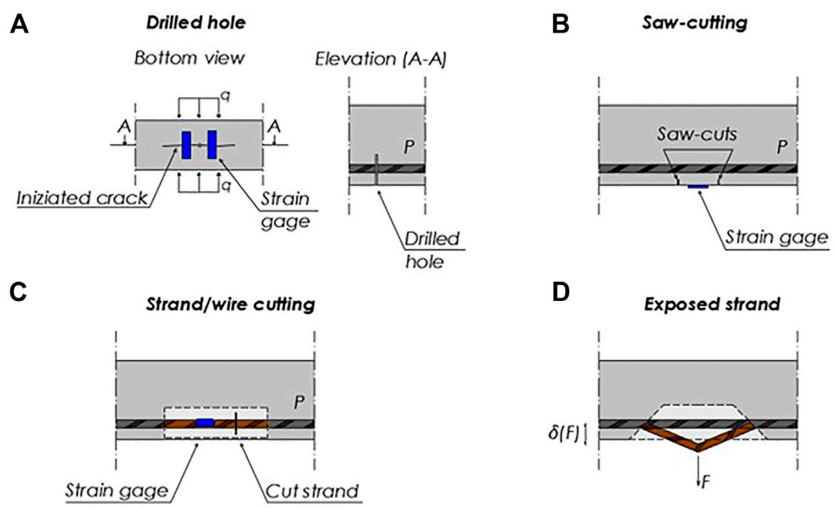


FIGURE 10
Indirect methods for prestress loss evaluation.

longitudinal profiles, if there were any segregation or bleeding phenomena in the grouted ducts during their realization, water would migrate to this position. It should also be recalled that, in the practice, the mid-span section is also the easiest to be accessed by the operators. However, it should be mentioned that this choice does not mean that any damage can be present in other positions.

In those positions, the cores were drilled until reaching the metallic ducts. At this stage, the cover was reported, and in one single case external corrosion of the duct was observed. In one case, indeed, the concrete cover was almost absent and signs of poor compaction were identified (honeycombs). Then, the ducts were cut and the grout was inspected to find any void, signs of poor compaction, bubbles, or presence of free water. None of these phenomena were observed and all wires seemed immersed in a good quality grout. Figure 9B shows an example of one analyzed duct.

Regarding the stress release methods, used also in this work, these techniques are generally based on the principle of estimating the current prestress force based on the stress released in a concrete member, measuring a strain value, as recently reviewed by (Kraľovanec et al., 2021a; Kraľovanec and Prokop, 2021b). These methods can be considered partially non-destructive, that is, they typically induce a local damage on the structure, which can however be promptly repaired with no long-term effects (neither for the statics nor for the durability) on the overall structure. Among them, the saw-cut method, stress-release coring (or drilling method), strand-cutting method (or wire-cutting), and exposed strand methods are worth to be cited (Kraľovanec et al., 2021c). A summarizing view of these methods is shown in Figure 10.

Here, the saw-cut method is applied, which is very similar to the drilling method: the technique is based on observing the changes in strain and stress in the area adjacent to the hole (or saw-cut isolated concrete block, in the companion method). The concrete block can be considered to be fully isolated when, increasing the cut depth, no further changes of the measured strain are read. Practically, the following procedure was followed: first, concrete surface was cleaned to guarantee a good adhesion of the electrical strain gages to the support (for their location, refer to Figure 11A); strain gages application and start of signal recording; cuts execution on the concrete surface with pyramid trunk shape in this case (see Figure 11B); final records after concrete de-heating. The concrete section was immediately repaired with a tyxotropic mortar.

The results in terms of concrete stress obtained from this test can be, thus used to quantify the residual prestress force acting in the tendons, using, for example, finite element (FE) numerical models, or more simply, Navier's formulation. In this last case, all the factors contributing to the strains at the monitoring point should be considered, that is, the prestress force in the member, including the influence of eccentricity in prestressing element positions; the restraint forces due, for instance, to eccentricity in the prestressing element positions in members that are not free to deform; the dead load of the member and the external applied load (Bagge et al., 2017). The results obtained here were reliable only for two points of measures over four: in half the cases, in fact, results could not be used due to a poor quality of the electrical signals and also because of the complexity of the methods. These tests have been rarely applied to real case structures with confirmed results but they found their main applications to relatively simple members (in terms of support

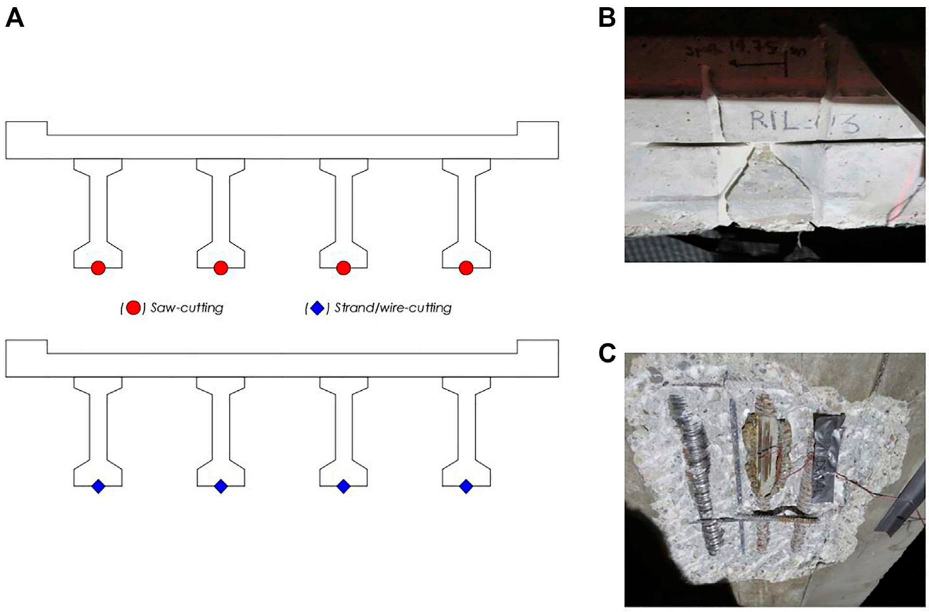


FIGURE 11
Stress-release method with saw-cut and wire-cut: location of test points **(A)**, saw-cuts **(B)**, and wire-cuts **(C)**.

conditions, member geometry, and prestressing reinforcement) in controlled environments. The thermal strains due to the cutting operations may be difficult to be quantified and subtracted from the overall reading. Overall, however, it was possible to obtain an average σ_c = 2.55 MPa at the beam intrados, which corresponds to about f_{se} = 810 MPa, under current service loads.

The second stress release method is the wire-cutting: one of the wires belonging to a tendon is exposed for a minimum length of 15 cm in the present case, then a strain gauge is installed and used to measure the strains that develop when it is cut. Then, corresponding prestress force in the wire can be determined easily. Also in this case, the section was then immediately restored using a tyxotropic mortar. Compared to other test methods, this approach is slightly destructive because one single wire is cut; however, the stresses can be re-distributed in other wires of the tendons if the grout has sufficiently good quality. Figure 11A shows the location of the tests, whereas Figure 11C provides a detail of the application of the strain gage on one wire.

The results of this last test method were quite non-conclusive: the high variability of the readings made the results non-reliable. In three cases over four, the prestress loss obtained was sensibly high, indicating almost a complete loss of the prestress force acting in the tendon. In another case, no variation of the prestress force was recorded. Such results are in line with the observations made by Halsey and Miller (1996), who tested two 40-year-old PTC beams extracted from a bridge, applied the wire-cutting method, and compared to other destructive and non-destructive testing. They obtained that the wire-cutting method was responsible for the highest prestress loss results, and they indicated, as possible reasons, the following: the removal of surrounding concrete reduces the local confinement, and thus a higher prestress loss; the instantaneous cut of the wire releases high-energy, which may affect the electric signal of the strain gage; lastly, the reading from a single wire may be not accurate to describe the behavior of a whole tendon. The authors would also add another possible reason related to the thermal strains induced by the cutting operation that may affect the reading: in this case, they are even higher than in the saw-cut method due to the higher coefficient of thermal expansion of high-alloy alloy (up to $1.7 \cdot 10^{-5}$ (°C^{-1})) steel than concrete ($1.2 \cdot 10^{-5}$ (°C^{-1})).

The last test method is the XRD, which has been briefly discussed in Section 2. The test was carried out on the same positions for the wire-cut but on more wires. Specifically, two-three wires were investigated for each measure point, in most cases repeating the reading to validate the result obtained. In detail, here the XRD-$sin^2(\psi)$ technique was adopted (Prevey, 1986), originally developed from the theories of crystallography and solid mechanics and applied for the surface residual stress evaluation mainly on thin films, coatings, and metals (Matejicek et al., 1998; Ma et al., 2002; Yazdanpanah et al., 2021a; Yazdanpanah et al., 2021b). In fact, since X-rays have a limited penetration in solid surfaces, the method allows measuring the stresses in a depth of a few micrometers as maximum.

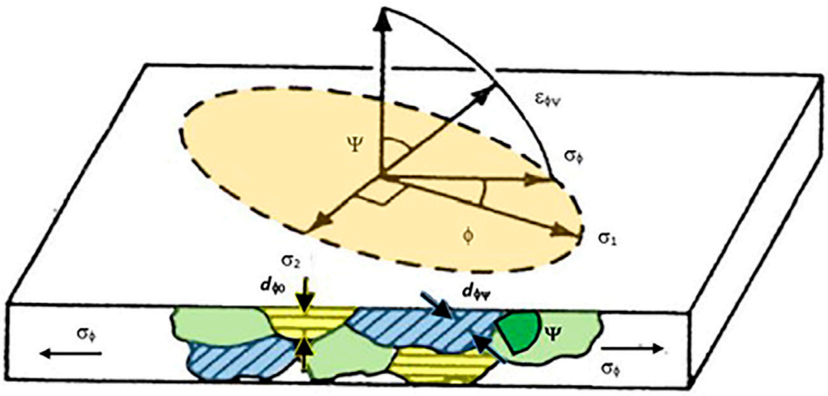


FIGURE 12
Plane-stress elastic model (adapted from Prevéy, 1986).

The main assumption at the base of the method is that a plane stress-state applies, see Figure 12, which shows the in-plane stress σ_ϕ with respect to the two principal stress components σ_1 and σ_2. When a X-ray beam hits the surface of the test sample at an incident angle Ω, the (*hkl*) lattice planes meeting the Bragg diffraction condition and having an off-axis angle ψ with respect to the sample surface normal, emit a diffraction X-ray beam at a diffraction angle 2θ (Luo and Jones, 2010). Accordingly, the spacing $d_{\phi\psi}$ of the (*hkl*) lattice plane is measured. Knowing the elastic properties of the materials (E and v) in the normal direction to the (*hkl*) orientation of the material and the initial lattice spacing in the unloaded (null stress) condition d_0, the following equation can be used:

$$d_{\phi\psi} = \left[\left(\frac{1+v}{E}\right)_{(hkl)} \cdot \sigma_\phi d_0 sin^2 \psi\right] - \left[\left(\frac{v}{E}\right)_{(hkl)} \cdot d_0 (\sigma_1 + \sigma_2) + d_0\right] \quad (2)$$

Thanks to the almost complete knowledge of the steel alloys used to realize construction materials, and particularly both ordinary and prestressing reinforcement, material properties and d_0 can be easily known. From Eq. (2), it is possible to obtain, under certain assumption, the surface stress acting on the material, according to the following equation:

$$\sigma_\phi = \left(\frac{E}{1+v}\right)_{(hkl)} \frac{1}{d_{\phi 0}} \left(\frac{\partial d_{\phi\psi}}{\partial sin^2 \psi}\right) \quad (3)$$

For a complete review of this technique, the work by Prevey (1986) is referred. Note that in the practice, the methods require that for a single determination of a $d_{\phi\psi}$ lattice spacing, to carry out multiple readings (in this case, nine executed; six after outliers removal), varying the angle ψ (the selected range is ± 35°, with counting time equal to 60 s), and the stress can be obtained by a least square minimization of the regression equation obtained. In this experimental campaign, a portable Spider TM X GNR diffractometer, working at 30 kV and 90 µA was used for measurements with a radiation source of chromium (Yazdanpanah et al., 2022).

FIGURE 13
XRD tests execution on the bridge PTC beams.

The portable XRD was mounted on a support to anchor the instrumentation at the deck intrados, allowing a minimum distance of the emitters/detectors to the whole length (about 50 cm) of the wires analyzed. Figure 13 shows an example of the tests carried out. Preliminary to carry out the readings, the wires were prepared: particularly, their surface was manually cleaned and then a chemical pickling was made, adopting a solution with sulfuric and hydrochloric acid, and demineralized water. Overall, for each measure point, 15 min are necessary to obtain one stress value. Furthermore, surface preparation time should be considered.

The results found through this method have a good repeatability and show consistency among themselves: for instance, repeating the reading two times on the same wire, results differ by 50–60 MPa. In fact, for Beam 1 and Beam 2, respectively, the results of successive readings on single wires were: 666 ± 20 MPa and 657 ± 21 MPa; 778 ± 63 MPa and 713 ± 44 MPa. The variability among the results is considered sufficiently limited, overall. Furthermore, the values obtained from this test are associated to higher prestress loss than those obtained from the saw-cut method, this probably being due to the direct exposition of the wires and the reduction of local confinement.

5 Conclusion

This work discussed pros and cons of different test methods to evaluate the actual prestress state in the prestressing steel reinforcement of post-tensioned concrete members. The failure of some PTC structures has raised the attention on this peculiar structural type, particularly for those elements having grouted ducts with non-accessible tendons. The assessment of current prestress in these members still represents a challenging objective because direct methods cannot be applied, unless tendons were instrumented at the time of realization, or they provide non-reliable results. An in-depth review of the current methods proposed in the literature and in practical applications was first performed. Then, recent authors experience both in laboratory tests and *in situ* applications were illustrated in detail with the aim to provide an additional contribution to the state-of-the-art of the scientific literature devoted to identify the most proper investigation techniques to be implemented for the quantification of prestress losses in PRC elements.

In detail, through two experimental campaigns, carried out both at the laboratory scale and in-field on an existing in-service PTC bridge, we have shown which results can be obtained with different techniques. Based on the authors' experience, vibration-based methods can be applied satisfactorily to identify existing damage; however, when the beams are still not cracked, the variation of the main dynamic parameters is too small to identify any significant prestress variation. In addition, the mutual connection of beams with other deck elements like slab, pavements, and bearings can significantly impact on the dynamic properties of the simply-supported beams individually considered. This issue implies additional efforts to carefully quantify the contribution of each component to the overall stiffness of a bridge deck that can rarely lead to a clear interpretation of the condition state of the prestressed beams. On the other hand, a promising technique is the evaluation of the UPV attenuation signal: when the test is carried out in the tomography configuration and the whole signal is recorded and analyzed, attenuation maps can be developed, showing a correlation with damage patterns. However, some criticalities have been observed for higher damage levels, where the ultrasonic signal suffers for a great signal disturbance.

Concerning indirect methods, according to the authors' opinion, an integrated approach should be preferred coupling different test methods, starting from visual inspections, endoscopies, and lastly, point measure such as saw-cut tests and XRD ones. In this regard we presented the results on a real-scale bridge, where some of these methods were applied, combining the use of the aforementioned different NDTs and comparing their results. In summary, based on such experience, the XRD method, even very recent in the field of bridge engineering, has shown very promising results, as also recently suggested by Morelli et al. (2021).

Lastly, as final remark, authors want to underline how a direct comparison between experimental results coming from laboratory tests and *in situ* measurements can be strongly flawed for the two following aspects that characterize the existing bridge decks as compared to the experimental layouts: 1) deterioration phenomena that can affect with unknown patterns the capacity of the beams and 2) the internal redundancy of real bridge decks—even if statically determined if considered as a whole—that can in turn give place to a beneficial redistribution of internal actions in case of prestress losses.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material; further inquiries can be directed to the corresponding author.

Author contributions

MZ, FF, and CP contributed to the conception and design of the study and performed the analysis of the results. FF wrote the first draft of the manuscript. MZ and CP wrote sections of the manuscript. MZ, FF, and CP reviewed the final version of the manuscript. All authors contributed to manuscript revision and read and approved the submitted version.

Acknowledgments

The authors would like to acknowledge all the colleagues that contributed to this work: Manuele Dabalà, Claudio Gennari, and Luca Pezzato from the University of Padova for their expertise on the XRD residual stress technique; Michele Frizzarin, Paolo Franchetti (from the Franchetti

SpA), and Eng. Filippo Andreose (from the University of Padova) for their work on the PTC laboratory beams and the elaboration of data (fully published in Frizzarin PhD thesis); and Eng. Silvano Vernizzi and Eng. Alessandro Zago (from Veneto Strade SpA) for the availability in carrying out the tests on the case study. Veneto Strade SpA is particularly thanked for the financial support and for giving the possibility to carry out and use the test results on the existing bridge. All the private partners (laboratories and engineers) who participated in the experimental tests of the case study are gratefully acknowledged. Finally, this work was developed within the Italy–Croatia INTERREG Project FIRESPILL—Fostering Improved Reaction of crossborder Emergency Services and Prevention Increasing safety LeveL.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aliofkhazraei, M., Dukeman, D., and Alexander, C. L. (2021). "The susceptibility of prestressing strands within grouted post-tensioned tendons to hydrogen embrittlement: A review," in *Corrosion 2021 (NACE)* (Houston, TX: Four Point Publishing LCC).

Anania, L., Badalà, A., and D'Agata, G. (2018). Damage and collapse mode of existing post tensioned precast concrete bridge: The case of Petrulla viaduct. *Eng. Struct.* 162, 226–244.

Angst, U., Moro, F., Geiker, M., Kessler, S., Beushausen, H., Andrade, C., et al. (2020). Corrosion of steel in carbonated concrete: Mechanisms, practical experience, and research priorities–a critical review by RILEM TC 281-CCC. *RILEM Tech. Lett.* 5, 85–100. doi:10.21809/rilemtechlett.2020.127

Azizinamini, A., Keeler, B., Rohde, J., and Mehrabi, A. B. (1996). Application of a new nondestructive evaluation technique to a 25-year-old prestressed concrete girder. *pcij.* 41, 82–95. doi:10.15554/pcij.05011996.82.95

Badalassi, M., Braconi, A., Cajot, L. G., Caprili, S., Degee, H., Gündel, M., et al. (2017). Influence of variability of material mechanical properties on seismic performance of steel and steel–concrete composite structures. *Bull. Earthq. Eng.* 15 (4), 1559–1607. doi:10.1007/s10518-016-0033-2

Bagge, N., Nilimaa, J., and Elfgren, L. (2017). *In-situ* methods to determine residual prestress forces in concrete bridges. *Eng. Struct.* 135, 41–52. doi:10.1016/j.engstruct.2016.12.059

Berto, L., Caprili, S., Saetta, A., Salvatore, W., and Talledo, D. (2020). Corrosion effects on the seismic response of existing rc frames designed according to different building codes. *Eng. Struct.* 216, 110397. doi:10.1016/j.engstruct.2020.110397

Calabrese, L., and Proverbio, E. (2020). A review on the applications of acoustic emission technique in the study of stress corrosion cracking. *Corros. Mater. Degrad.* 2 (1), 1–33. doi:10.3390/cmd2010001

Calvi, G. M., Moratti, M., O'Reilly, G. J., Scattarreggia, N., Monteiro, R., Malomo, D., et al. (2019). Once upon a time in Italy: The tale of the morandi bridge. *Struct. Eng. Int.* 29 (2), 198–217. doi:10.1080/10168664.2018.1558033

Casas, J. R., and Aparicio, A. C. (1994). Structural damage identification from dynamic-test data. *J. Struct. Eng. (N. Y. N. Y).* 120 (8), 2437–2450. doi:10.1061/(asce)0733-9445(1994)120:8(2437)

COST TU1406 (2014). *Quality specifications for roadway bridges, standardization at a European level (Bridgespec)*. Brussels: European Cooperation in the field of Scientific and Technical Research.

COST345 (2004). "Procedures required for the assessment of highway infrastructures," in *European research project under the framework of European cooperation in the field of scientific and technical research* (Brussels: EU Commission-Directorate General Transport and Energy).

Crespo-Minguillon, C., and Casas, J. R. (1997). A comprehensive traffic load model for bridge safety checking. *Struct. Saf.* 19 (4), 339–359. doi:10.1016/s0167-4730(97)00016-7

Ewins, D. J. (2009). *Modal testing: Theory, practice and application*. John Wiley & Sons.

Faleschini, F., Zanini, M. A., and Pellegrino, C. (2018). "Quality control, infrastructure management systems and their implementation in medium-size highway networks," in *Proc eBook for the 8th Workshop meeting Barcelona EUROSTRUCT* (Guimaraes, Portugal: EUROSTRUCT Association).

Frizzarin, M. (2019). An experimental study on the correlation between residual prestress force in pre-tensioned and post-tensioned concrete beams and nondestructive tests. Ph.D. Thesis. Padova (Italy): University of Padova. Available At: https://www.research.unipd.it/handle/11577/3423298.

Garber, D. B., Gallardo, J. M., Deschenes, D. J., and Bayrak, O. (2015). Experimental investigation of prestress losses in full-scale bridge girders. *ACI Struct. J.* 112 (5), 7909. doi:10.14359/51687909

Gentile, C., and Saisi, A. (2007). Ambient vibration testing of historic masonry towers for structural identification and damage assessment. *Constr. Build. Mater.* 21 (6), 1311–1321. doi:10.1016/j.conbuildmat.2006.01.007

Halsey, J. T., and Miller, R. (1996). Destructive testing of two forty-year-old prestressed concrete bridge beams. *pcij.* 41 (5), 84–93. doi:10.15554/pcij.09011996.84.93

Kraľovanec, J., Bahleda, F., Prokop, J., Moravčík, M., and Neslušan, M. (2021a). Verification of actual prestressing in existing pre-tensioned members. *Appl. Sci.* 11 (13), 5971. doi:10.3390/app11135971

Kraľovanec, J., Moravčík, M., Bujňáková, P., and Jošt, J. (2021c). Indirect determination of residual prestressing force in post-tensioned concrete beam. *Materials* 14 (6), 1338. doi:10.3390/ma14061338

Kraľovanec, J., and Prokop, J. (2021b). Indirect methods for determining the state of prestressing. *Transp. Res. Procedia* 55, 1236–1243. doi:10.1016/j.trpro.2021.07.105

Luo, Q., and Jones, A. H. (2010). High-precision determination of residual stress of polycrystalline coatings using optimised XRD-sin2ψ technique. *Surf. Coatings Technol.* 205 (5), 1403–1408. doi:10.1016/j.surfcoat.2010.07.108

Ma, C. H., Huang, J. H., and Chen, H. (2002). Residual stress measurement in textured thin film by grazing-incidence X-ray diffraction. *Thin solid films* 418 (2), 73–78. doi:10.1016/s0040-6090(02)00680-6

Martin, J., Broughton, K. J., Giannopolous, A., Hardy, M. S. A., and Forde, M. C. (2001). Ultrasonic tomography of grouted duct post-tensioned reinforced concrete bridge beams. *Ndt E Int.* 34 (2), 107–113. doi:10.1016/s0963-8695(00)00035-9

Matejicek, J., Sampath, S., and Dubsky, J. (1998). X-ray residual stress measurement in metallic and ceramic plasma sprayed coatings. *J. Therm. spray Technol.* 7 (4), 489–496. doi:10.1361/105996398770350701

Ministry of Italian Infrastructure (2020). *Guidelines on Risk classification and management, safety assessment and monitoring of existing bridges*. Rome, Italy: Consiglio Superiore dei Lavori Pubblici.

Moravčík, M., and Bujňáková, P. (2017). Analysis of brittle failure of prestressed bridge in the north of Slovakia. *Rocz. Inżynierii Bud.*, 27–34.

Morelli, F., Panzera, I., Piscini, A., Salvatore, W., Chichi, F., Marconi, G., et al. (2021). X-ray measure of tensile force in post-tensioned steel cables. *Constr. Build. Mater.* 305, 124743. doi:10.1016/j.conbuildmat.2021.124743

Oh, B. H., and Yang, I. H. (2000). Sensitivity analysis of time-dependent behavior in PSC box girder bridges. *J. Struct. Eng. (N. Y. N. Y).* 126 (2), 171–179. doi:10.1061/(asce)0733-9445(2000)126:2(171)

Osborn, G. P., Barr, P. J., Petty, D. A., Halling, M. W., and Brackus, T. R. (2012). Residual prestress forces and shear capacity of salvaged prestressed concrete bridge girders. *J. Bridge Eng.* 17 (2), 302–309. doi:10.1061/(asce)be.1943-5592.0000212

Pisani, M. A., Limongelli, M. P., Giordano, P. F., and Palermo, M. (2021). On the effectiveness of vibration-based monitoring for integrity management of prestressed structures. *Infrastructures* 6 (12), 171. doi:10.3390/infrastructures6120171

Prevéy, P. S. (1996). Current applications of X-ray diffraction residual stress measurement. *Dev. Mater. Charact. Technol.* 103.

Prevey, P. S. (1986). X-ray diffraction residual stress techniques. *ASM Int. ASM Handb.* 10, 380–392.

Quirk, L., Matos, J., Murphy, J., and Pakrashi, V. (2018). Visual inspection and bridge management. *Struct. Infrastructure Eng.* 14 (3), 320–332. doi:10.1080/15732479.2017.1352000

Saiidi, M., Douglas, B., and Feng, S. (1994). Prestress force effect on vibration frequency of concrete bridges. *J. Struct. Eng. (N. Y. N. Y).* 120 (7), 2233–2241. doi:10.1061/(asce)0733-9445(1994)120:7(2233)

SAMARIS (2005). “Sustainable and advanced materials for road infrastructure,” in *VI framework programme* (Brussels: SAMARIS Consortium).

Tu, B., Dong, Y., and Fang, Z. (2019). Time-dependent reliability and redundancy of corroded prestressed concrete bridges at material, component, and system levels. *J. Bridge Eng.* 24 (9), 04019085. doi:10.1061/(asce)be.1943-5592.0001461

Venanzi, I., Kita, A., Cavalagli, N., Ierimonti, L., and Ubertini, F. (2020). Earthquake-induced damage localization in an historic masonry tower through long-term dynamic monitoring and FE model calibration. *Bull. Earthq. Eng.* 18 (5), 2247–2274. doi:10.1007/s10518-019-00780-4

Yazdanpanah, A., Biglari, F. R., Fallahi Arezoodar, A., and Dabalà, M. (2021a). Role of grinding induced surface residual stress on probability of stress corrosion cracks initiation in 316L austenitic stainless steel in 3.5% sodium chloride aqueous solution. *Corros. Eng. Sci. Technol.* 56 (1), 81–92. doi:10.1080/1478422x.2020.1812818

Yazdanpanah, A., Franceschi, M., Bergamo, G., Bonesso, M., and Dabalà, M. (2022). On the exceptional stress corrosion cracking susceptibility of selective laser melted 316L stainless steel under the individual effect of surface residual stresses. *Eng. Fail. Anal.* 136, 106192. doi:10.1016/j.engfailanal.2022.106192

Yazdanpanah, A., Lago, M., Gennari, C., and Dabalà, M. (2021b). Stress corrosion cracking probability of selective laser melted 316l austenitic stainless steel under the effect of grinding induced residual stresses. *Metals* 11 (2), 327. doi:10.3390/met11020327

Zanini, M. A., Faleschini, F., and Pellegrino, C. (2017a). Bridge residual service-life prediction through Bayesian visual inspection and data updating. *Struct. Infrastructure Eng.* 13 (7), 906–917. doi:10.1080/15732479.2016.1225311

Zanini, M. A., Faleschini, F., and Pellegrino, C. (2017b). Probabilistic seismic risk forecasting of aging bridge networks. *Eng. Struct.* 136, 219–232. doi:10.1016/j.engstruct.2017.01.029

frontiers | Frontiers in Built Environment

TYPE Original Research
PUBLISHED 17 November 2022
DOI 10.3389/fbuil.2022.1008973

Check for updates

OPEN ACCESS

EDITED BY
Vagelis Plevris,
Qatar University, Qatar

REVIEWED BY
Maurizio Orlando,
Maurizio Orlando, Italy
Tadesse Gemeda Wakjira,
University of British Columbia,
Okanagan Campus, Canada

*CORRESPONDENCE
Themistoklis Tsalkatidis,
themistoklis.tsalkatidis@nmbu.no

SPECIALTY SECTION
This article was submitted to Computational Methods in Structural Engineering,
a section of the journal
Frontiers in Built Environment

RECEIVED 01 August 2022
ACCEPTED 02 November 2022
PUBLISHED 17 November 2022

CITATION
Vogiatzis T, Tsalkatidis T and Efthymiou E (2022), The wall–frame interaction effect in CLT-steel hybrid systems.
Front. Built Environ. 8:1008973.
doi: 10.3389/fbuil.2022.1008973

The wall–frame interaction effect in CLT-steel hybrid systems

Tzanetis Vogiatzis[1], Themistoklis Tsalkatidis[2]* and Evangelos Efthymiou[1]

[1]Department of Civil Engineering, Institute of Metal Structures, Aristotle University of Thessaloniki, Thessaloniki, Greece, [2]Faculty of Science and Technology, Norwegian University of Life Sciences, Ås, Norway

Behaviour and capacity of cross-laminated timber (CLT) infills built inside steel frames have been given increasing research attention in recent years. It is widely accepted that when the CLT wall panel is built in tight contact with the bounding steel frame to participate in the load sharing, its inherently large in-plane stiffness will attract additional forces to the frame area and change the behaviour of the hybrid system. If not designed properly, the structural integrity of both the infill and the frame will be compromised. It is thus crucial to accurately evaluate the contribution of the infill CLT wall panel to the stiffness and strength of the hybrid system. To that end, a finite element study was performed to investigate the frame-wall interaction effect on the behaviour of hybrid systems. The lateral stiffness, lateral load capacities and hysteretic characteristics of the hybrid systems with frictional and connected interfaces were investigated. The load-sharing effect between the CLT wall and the steel frame was studied. The numerical results showed that the connected models were very effective as the infill absorbed a substantial part of the lateral load, during the initial stages of loading.

KEYWORDS

steel–timber hybrid structure, shear wall, cross-laminated timber, steel moment resisting frame, seismic behavior, numerical analysis, ductility

1 Introduction

Traditionally, hybrid wall systems involve the use of reinforced concrete (RC) walls or composite steel-concrete (CSC) walls to provide a lateral-load resisting system for earthquake ground motions (Wallace and Wada, 2000). One example of CSC is the composite steel frame with RC walls system (SRCW) (Hajjar, 2002), classified according to the frame of Eurocode 8 as Type 1 composite structural system (European Commision, 2008). The main advantage of a code-designed SRCW system is that it has both higher flexural resistance and stiffness in comparison to a conventional RC wall of the same cross-sectional geometry (Fardis et al., 2005). These composite structures are also referred to as infilled frames and have increased lateral stiffness than structures that have bare frames. In the past six decades, extensive experimental and numerical research has been conducted to understand the combined behaviour of infilled frames, but most of these

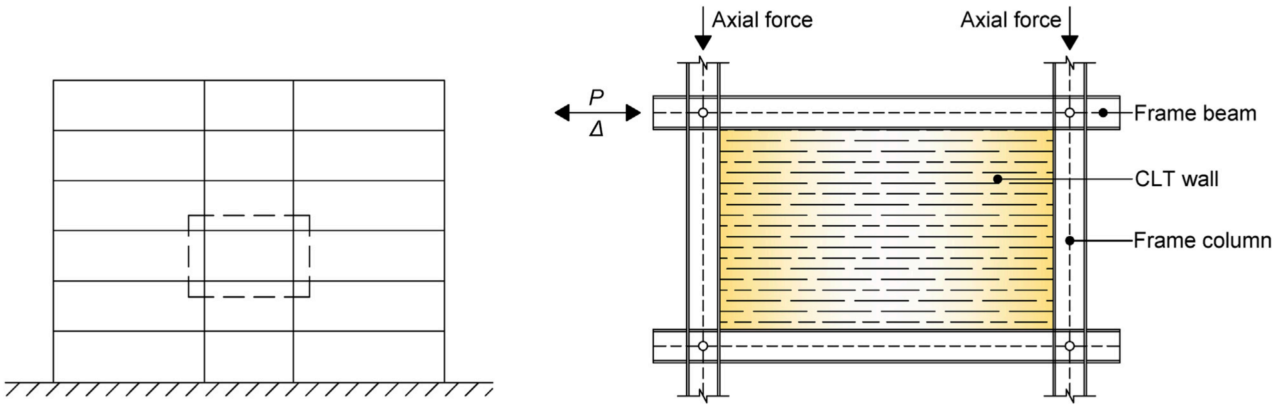


FIGURE 1
Typical elements of a steel frame with CLT shear wall hybrid system, STSW.

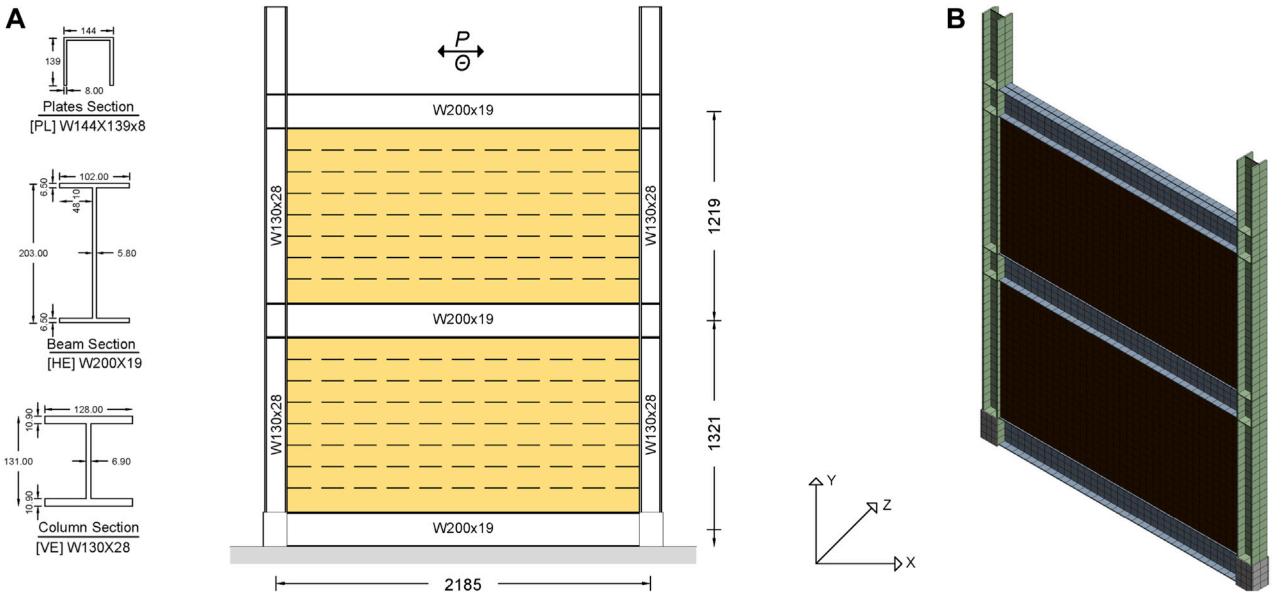


FIGURE 2
STSW Hybrid System **(A)** Specimen configuration, **(B)** Finite element model.

TABLE 1 Specifications of STSW models.

FE-model	Frame-infill interaction	CLT layers (No)	Infill thickness t_w, (mm)
HW1-3PLY81T	Frictional—0.40	3	81
HW1-3PLY102T	Frictional—0.40	3	102
HW1-5PLY120T	Frictional—0.40	5	120
HW2-3PLY81T	Fully connected	3	81
HW2-3PLY102T	Fully connected	3	102
HW2-5PLY120T	Fully connected	5	120

TABLE 2 Mechanical properties of steel members used in the FE-models.

Element	E_s (GPa)	f_y (MPa)	f_u (MPa)	ε_{st} (%)	ε_u (%)
Frame columns	200	312	496	1.69	11
Frame beams	200	353.50	540	2.38	12

scholars investigated the behaviour and capacity of masonry infills built inside concrete or steel frames (Mehrabi et al., 1996; Chen and Liu, 2016; Margiacchi et al., 2016; Repapis and Zeris, 2019). Only a few researchers investigated the response of steel frames connected with infill RC walls (Makino et al., 1980; Saari et al., 2004; INNO-HYCO, 2015; Vogiatzis and Avdelas, 2018; Pallarés et al., 2020).

Today, as the construction must comply with the requirements of sustainable developments, timber has been gaining increased popularity in both residential and non-residential projects (Wang et al., 2015), as a renewable building material with a low carbon footprint, environmental friendliness, and easy assembly. Cross-laminated timber (CLT) is a prefabricated multi-layered engineer wood product, manufactured using at least three pieces of parallel boards by gluing their surfaces together using adhesives (FPInnovations Institute, 2013; Pečnik et al., 2021). Due to its cross-lamination technique, this type exhibits substantially better dimensional stability in comparison to solid wood, and offers high in-plane stiffness and strength when placed as a vertical wall element in mid-rise buildings (Carrero et al., 2021). CLT systems provide not only a high level of prefabrication and flexibility in planning, which reduces hence total cost of projects [FPInnovations (Institute), 2013]. Countries such as Japan have a long history of timber buildings that can effectively resist earthquakes. More recently, Canada, New Zealand, Italy and Greece have been developing structural systems for either larger buildings or semi-open sports facilities that can withstand earthquakes with minimal damage.

The CLT building material has opened new dimensions in steel-timber hybrid systems and allows researchers to focus on new solutions for structures with sustainable design considerations. Moreover, the benefits of CLT-Steel hybrid systems include more efficient use of materials and improved seismic resistance. Increased efficiency can be obtained because the strength, stiffness and weight of some materials can be utilised where it is most beneficial within a building. Seismic performance can similarly be enhanced by combining materials to optimise ductility, strength, and stiffness, all of which contribute to the dynamic structural behaviour of buildings (Quintana Gallo and Carradine, 2018). The most popular concepts of steel-timber hybrid systems are the steel frame with CLT shear walls (Dickof et al., 2014; Tesfamariam et al., 2014; Vogiatzis et al., 2019) and the timber frame with steel plate shear walls, (Conrad and Phillips, 2019; Igbal et al., 2020), the former concept is the focus of this work. Please note that in this as well as the following sections, unless otherwise stated, the abbreviation STSW refers to the concept of steel frame with CLT shear walls, as shown in Figure 1.

One of the first works on hybrid systems composed of steel frames and CLT shear walls was conducted by (Tesfamariam et al., 2014). The results of this work showed that the peak inter-story drift was less than two per cent for multistorey buildings. To overcome this challenge, a hybrid system composed of steel columns, timber beams, and oriented strand board panels with good in-plane strength and ductility was proposed by (Trutalli et al., 2017). Moreover, a hybrid wall system consisting of a light-frame wood shear wall and steel frame was tested by (He et al., 2014). The aforementioned tested hybrid system was numerically investigated by (Li et al., 2014), and further developed with a post-tensioned steel frame and light frame wood shear wall by (Cui et al., 2020). Recently, a hybrid system composed of steel frames with semi-rigid connections and CLT shear walls was designed and analyzed by the authors (Vogiatzis et al., 2019), although the post-elastic phase was not considered in that work.

2 Methodology section

2.1 Design consideration

The STSW system considered for this research represents the first two stories of a multi-story mixed building designed for core-wall hybrid lateral load-resisting systems (Shahrooz et al., 1996). The specimen geometry, material, and boundary conditions reflected the physical model of the SRCW system tested by (Tong et al., 2001), as an idealized representation (one-third scale) of the bottom two stories of a six-story building, although the frame and the infill were modeled with cross-laminated timber shear walls and rigid connections, respectively. The columns consist of W130X28 wide flange

TABLE 3 Mechanical properties of timber members used in the FE models.

$E_{0,m}$	$E_{90,m}$	G_m	$f_{t,0,k}$	$f_{t,90,k}$	$f_{c,0,k}$	$f_{c,90,k}$	$f_{v,090,k}$	$f_{v,9090,k}$
(MPa)	(MPa)	(MPa)	(MPa)	(MPa)	(MPa)	(MPa)	(MPa)	(MPa)
11,600	370	690	14	0.12	21	2.5	4.0	0.8

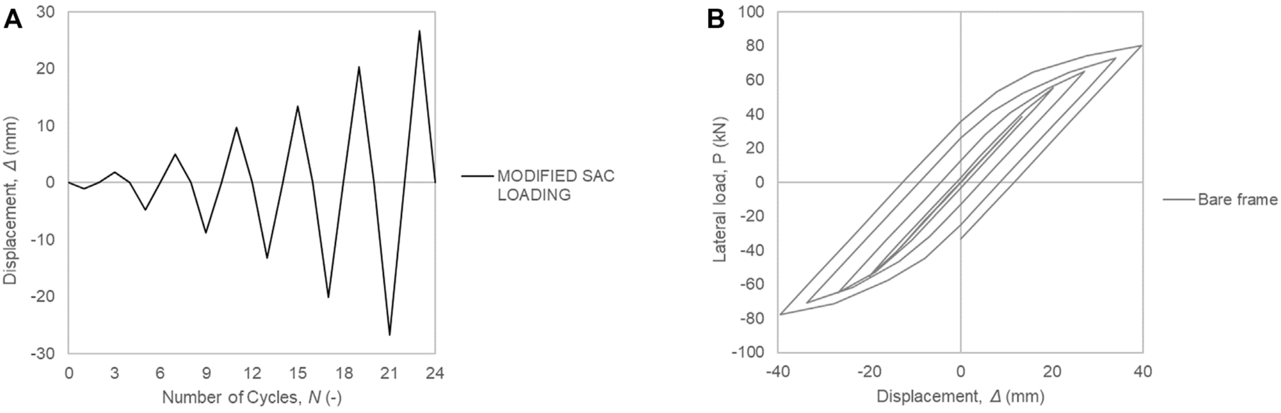


FIGURE 3
(A) Loading history, **(B)** Hysteretic curve of bare frame.

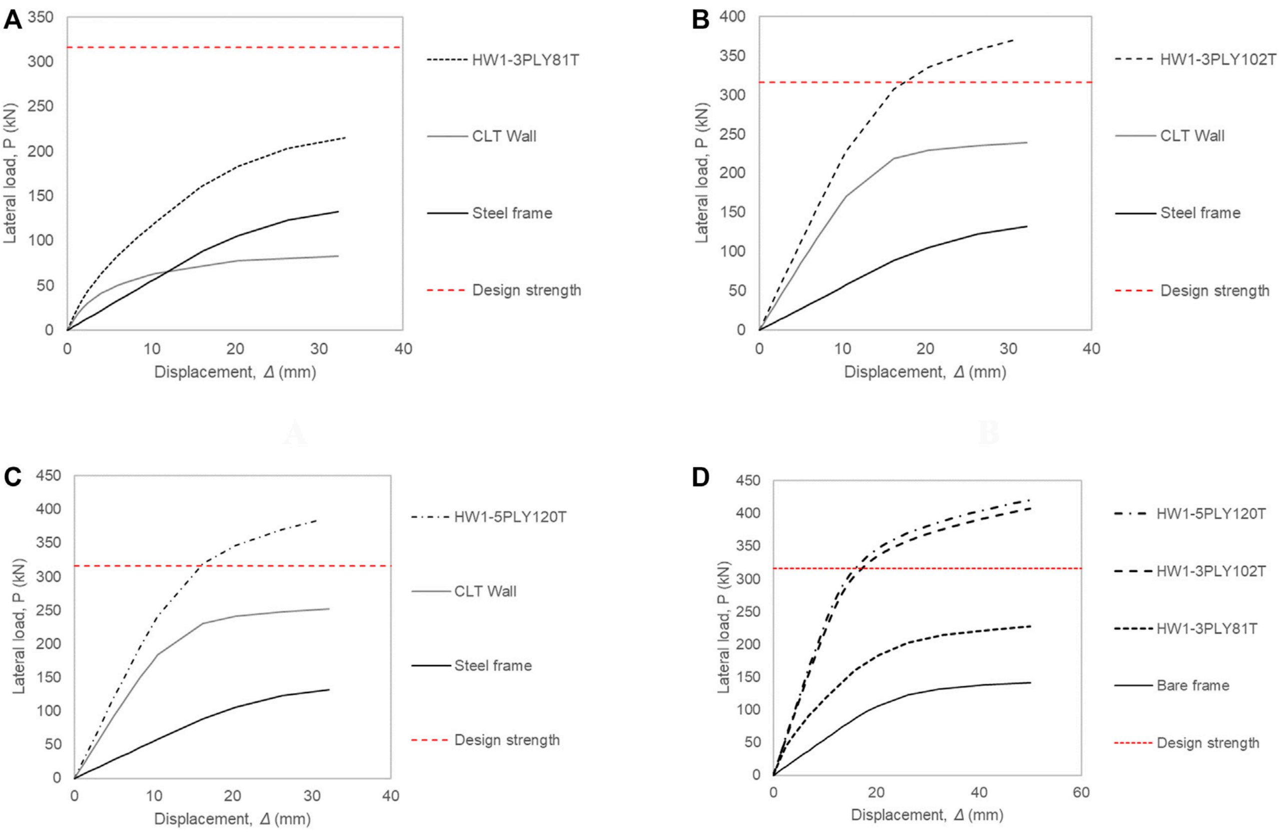


FIGURE 4
Behaviour of STSW specimens with frictional infill-frame interface **(A)** 3-ply (81 mm thick), **(B)** 3-ply (102 mm thick), **(C)** 5-ply (120 mm thick), **(D)** Comparative diagram.

steel sections and the beams consist of W200X19 wide flange steel sections. The infill was assumed to transfer 100% of the seismic story shear. The building was designed according to the NEHRP provision, using an acceleration and velocity coefficient of 0.4 and a force reduction factor R, of 5.5, (larger than the Eurocodes' requirement for q due to overstrength) (Vogiatzis,

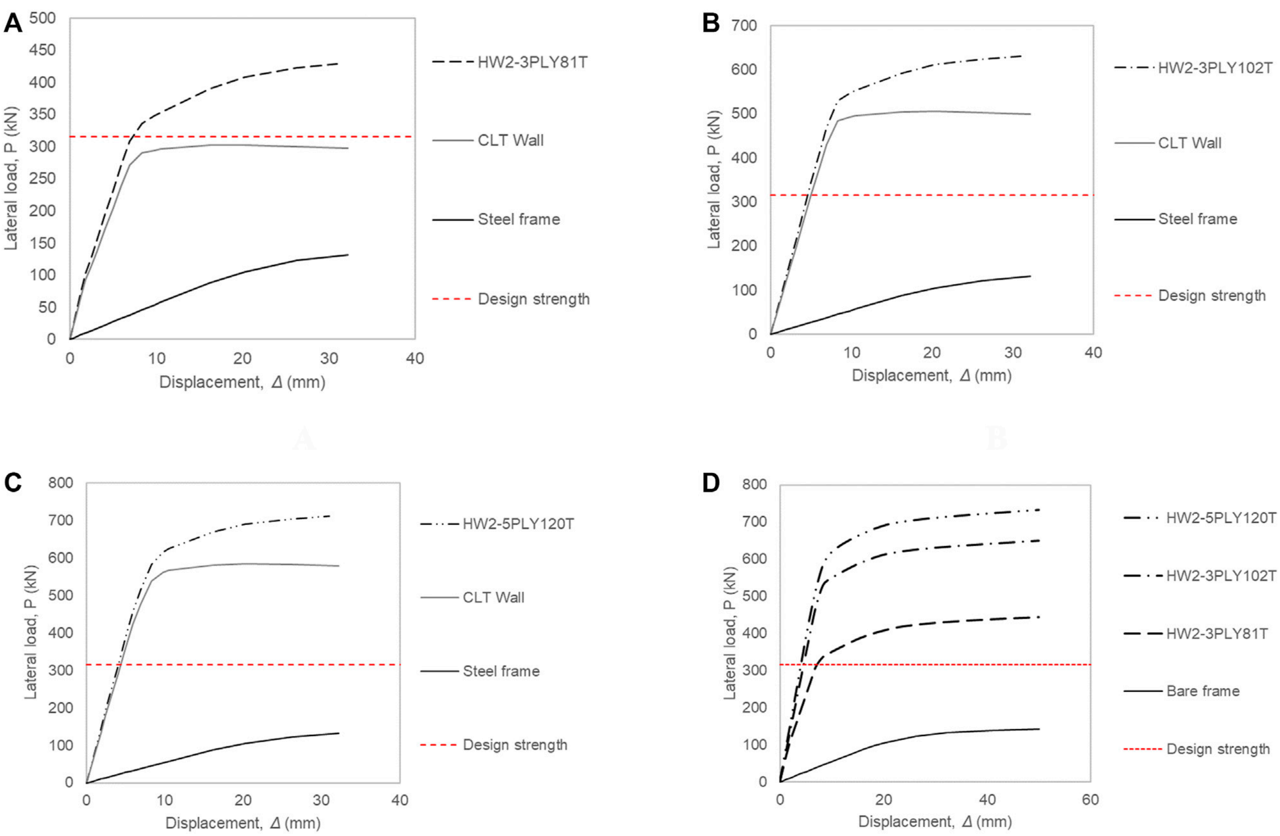


FIGURE 5
Behaviour of STSW specimens with connected infill-frame interface **(A)** 3-ply (81 mm thick), **(B)** 3-ply (102 mm thick), **(C)** 5-ply (120 mm thick), **(D)** Comparative diagram.

2019). Figure 2A shows a schematic view of the STSW system considered for this work. Table 1 summarises the FE models developed in this study, while a list of abbreviations for different parameters is also included.

There is a clear lack of studies exploring the wall–frame interaction effect on the overall behaviour of STSW systems under monotonic and cyclic loading conditions. Hence, in the current, work numerous STSW models were developed as part of a parametric study, that aims to understand the influence of the wall–frame interaction effect on the response of STSW systems under different loading conditions. The parametric study was performed on a frame that had fixed-ended columns and rigid beam-to-column connections, as shown in Figure 2A. The dimensions of the STSW models are 2,184 mm (length) × 2,540 mm (height), measured from center to center of the steel sections. The beam and column dimensions are also shown in Figure 2A. The steel parameters are listed in Table 2. The STSW systems, had an infill CLT panel with length set to 2,054 mm and a height set to 1,016 mm, by using outside adhesion, between the infill CLT panels and the boundary steel frame. The material properties of the CLT shear wall under consideration are given in Table 3.

2.2 Finite element analysis

2.2.1 Finite element model

To accurately simulate the behaviour of the STSW system, all the components of the specimen must be included in the simulation. These components, shown in Figure 3A, are the infill CLT panels and the boundary frame members (beams, columns). Furthermore, the interaction between components is critical. Both H-shaped steel frame and infill CLT wall are modelled in ANSYS R1 (ANSYS Inc, 2019) with the twenty-node structural brick element (SOLID186) as presented in Figure 2B. This element has been commonly used for three-dimensional finite element modelling of steel-timber hybrid structures (Rahmzadeh and Iqbal, 2021). Specifications of the analyzed models are shown in Table 1.

2.2.2 Steel material model

Multi-linear forms are usually used to define the steel stress-strain relationship, giving acceptable results under monotonic loading. However, Shi et al. (2011) showed that in the cases of cyclic loading, it was difficult for their results to meet the calculation accuracy. Therefore, the constitutive model suggested by Chaboche (Chaboche, 1986; Chaboche, 1989) is adopted which is parameterized in ANSYS R1 (ANSYS Inc, 2019). The material properties of each steel element were adopted from (Vogiatzis et al., 2019) and are given in Table 2.

2.2.3 Timber material model

The complexity of the mechanical response of CLT origins in the orthogonal grain direction and the overall anisotropy of wood as a material. The elastic mechanical properties have different values along the three principal axes. There is the grain (axial), the circumferential and the radial directions. The stiffness and strength in the axial direction have greater values than those in the other two directions (Furtmüller et al., 2018). Regarding the failure modes of timber, the three most common failure mechanisms are: 1) failure occurring due to compression parallel to grain (ductile failure mode), 2) failure occurring due to compression perpendicular to the grain (ductile failure mode), and 3) failure caused by shear parallel to the grain that it is accompanied by tension perpendicular to the grain (brittle failure mode) (Xu et al., 2014).

The theory of plasticity employs a set of constitutive equations to describe the complex multiaxial stress state. This is achieved by using three basic parameters: a yield criterion, a flow rule and a hardening rule. To capture the material's non-linearity with accuracy, Hill's yield criterion (Hill, 1948; Hill, 1998) has been selected to describe the anisotropic plastic behaviour of the cross-laminated timber. The accuracy of Hill's model for capturing the non-linear behaviour and failure mode of timber has been previously investigated by (Xu et al., 2014; He et al., 2018; Vogiatzis et al., 2020).

The validity of Hill's criterion, as implemented in ANSYS, has been investigated for timber structural systems under cyclic loading conditions by (Rahmzadeh and Iqbal, 2021). In that work the form of the quadratic Hill yield criterion was presented as shown in Eq. 1. Where the stresses σ_{ij} are the normal yield stresses according to the principal directions of anisotropy, and the constants F, G, H, L, M and N, can be defined either experimentally or by Eqs 2–5. More information on the background of Hill's criterion is given by (Imaoka, 2008).

$$F\left(\sigma_{yy}-\sigma_{zz}\right)^2+G\left(\sigma_{zz}-\sigma_{xx}\right)^2+H\left(\sigma_{xx}-\sigma_{yy}\right)^2+2L\tau_{yz}^2+2M\tau_{zx}^2+2N\tau_{xy}^2 \tag{1}$$

$$F=\frac{1}{2}+\left(\frac{1}{\left(\sigma_{yy}^{y}\right)^2}+\frac{1}{\left(\sigma_{zz}^{y}\right)^2}-\frac{1}{\left(\sigma_{xx}^{y}\right)^2}\right) \tag{2}$$

$$G=\frac{1}{2}+\left(\frac{1}{\left(\sigma_{zz}^{y}\right)^2}+\frac{1}{\left(\sigma_{xx}^{y}\right)^2}-\frac{1}{\left(\sigma_{yy}^{y}\right)^2}\right) \tag{3}$$

$$H=\frac{1}{2}+\left(\frac{1}{\left(\sigma_{xx}^{y}\right)^2}+\frac{1}{\left(\sigma_{yy}^{y}\right)^2}-\frac{1}{\left(\sigma_{zz}^{y}\right)^2}\right) \tag{4}$$

$$L=\frac{1}{2\left(\tau_{yz}^{y}\right)^2}\,M=\frac{1}{2\left(\tau_{zx}^{y}\right)^2}\,N=\frac{1}{2\left(\tau_{xy}^{y}\right)^2} \tag{5}$$

where σ_{ij}^{y} $(i = j)$ are yield stresses in the principal axes of anisotropy and τ_{ij}^{y} $(i \neq j)$ are yield stresses in shear concerning the principal axis of anisotropy. Based on Hill's criterion, in the case of an anisotropy with three mutually orthogonal planes of symmetry, yielding under a multiaxial stress state occurs when Eq. 1 is equal to one. The matrix must be positive-definite so that negative strain energy is avoided. This thermodynamic requirement can be satisfied using the following equations (Lempriere, 1968):

$$\left(1-\nu_{yz}\nu_{zy}\right),\left(1-\nu_{xz}\nu_{zx}\right),\left(1-\nu_{xy}\nu_{yx}\right)\rangle 0 \tag{6}$$

$$1-\nu_{xy}\nu_{yx}-\nu_{yz}\nu_{zy}-2\nu_{xy}\nu_{yz}\nu_{zx}>0 \tag{7}$$

The values of the wood elastic properties, for both the longitudinal and the transverse layers of the CLT have been obtained from (Stazi et al., 2019), corresponding to red spruce C24 boards. Table 3 summarises these parameters: parallel-to-grain ($E_{0,m}$) and perpendicular-to-grain ($E_{90,m}$) moduli of elasticity, shear moduli values (G_m), parallel-to-grain ($f_{t,0,k}$) and perpendicular-to-grain ($f_{t,90,k}$) tensile strength values, parallel-to-grain ($f_{c,0,k}$) and perpendicular-to-grain ($f_{c,90,k}$) compressive strength values, parallel-to-grain ($f_{v,090,k}$) and perpendicular-to-grain ($f_{v,9090,k}$) characteristic shear strength values.

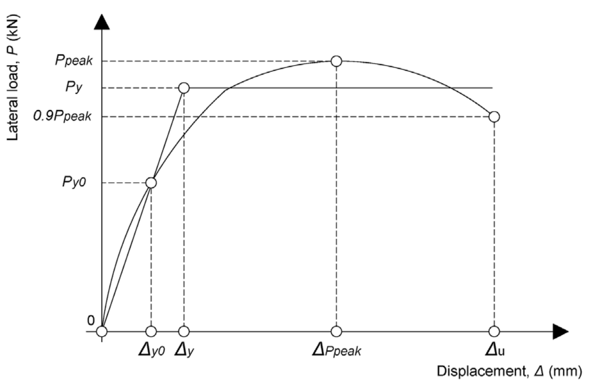


FIGURE 6
Actual response curve and equivalent idealized elastic-plastic curve.

2.2.4 Contact modelling

The contact elements for the STSW models were simulated with CONTA174 elements and target elements were modeled with TARGE170. It was proven by (He et al., 2018) that the CLT neighboring layers could be bonded in finite element model (He et al., 2018), which could not affect the produced results. Hence, the interaction at the interface of different layers of the CLT assembly were simulated using paired surface-to-surface contact elements (CONTA174 and TARGE170). This indicates that the relative movement between two contacting surfaces was not

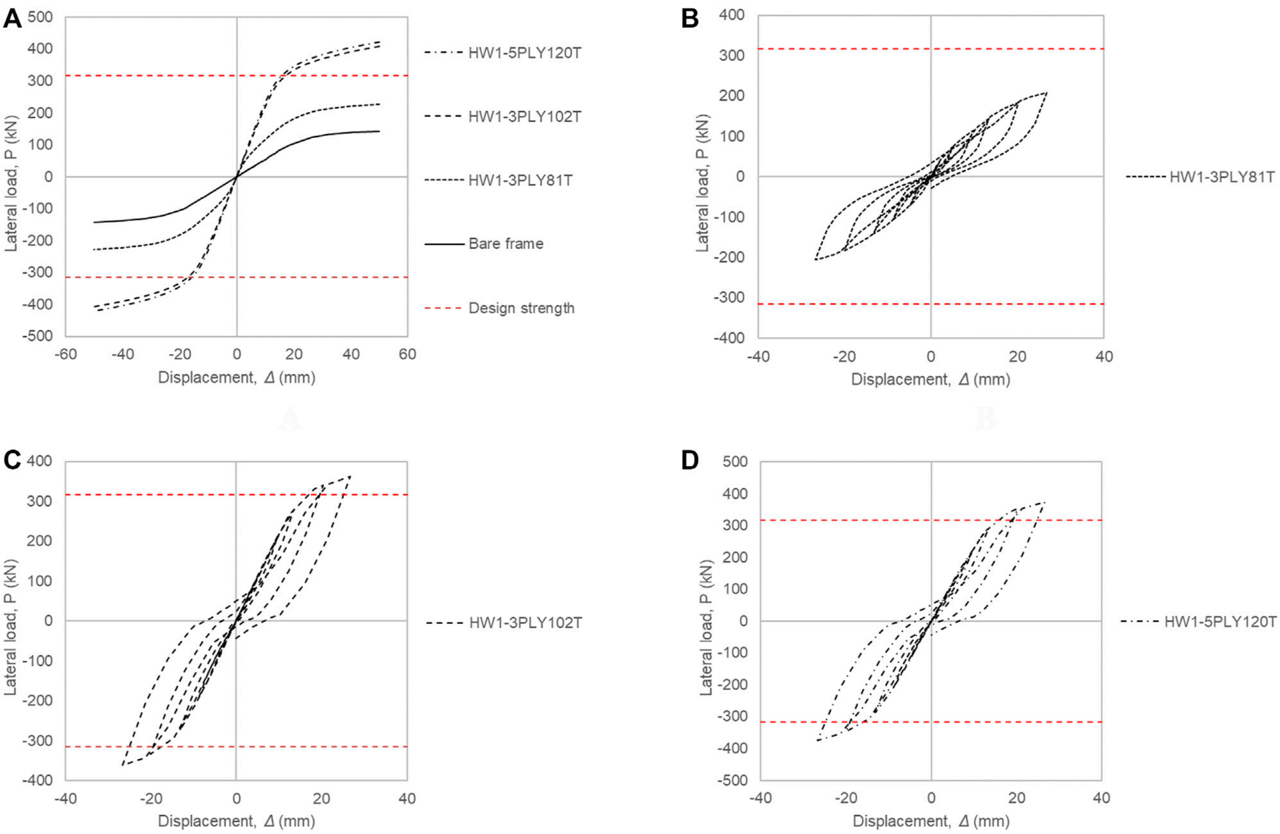


FIGURE 7
Hysteretic and backbone curves of the specimens with frictional infill-frame interface **(A)** Comparative diagram, **(B)** 3-ply (81 mm thick), **(C)** 3-ply (102 mm thick), **(D)** 5-ply (120 mm thick).

permitted. Those interface elements were also used for the contact along the perimeter between the steel boundary frame members and the CLT shear wall. This contact surface, along the interface of the steel sections and the CLT panel, was defined as a frictional contact for the HW1 models and bonded for the HW2 models. Coefficient of friction for all frictional contact bodies was set at 0.4.

2.2.5 Boundary conditions, loading procedure and validation

During the numerical analysis, lateral displacements are applied to the upper flange of the top beam as shown in Figure 2A. In the finite element model, the out-of-plane degrees of freedom along the height of the beam centerline were constrained to simulate the lateral support. All the degrees of freedom for the bottom of the columns and the lower flange of the ground beam were completely constrained to simulate the fixed support. Firstly, the gravity loads are applied. Secondly, the lateral displacement Δ is imposed to simulate the monotonic or cyclic loading conditions. At the first step of loading, the gravity loads were implemented and kept constant throughout the subsequent step. Next step was to apply cyclic loading, using displacement control mode, at the top flange of the systems upper beam. The imposed cyclic loading pattern is shown in Figure 3A, as it was modified according to SAC protocol for testing of steel beam-column connections and other steel elements (Tong et al., 2001).

The numerical model could not be validated against experimental results since there is lack of available experimental data of the same hybrid system in the international literature. Therefore, the authors used as a guide a numerical model of a similar hybrid system that they had built in a recent study (Vogiatzis et al., 2020) which had been validated using experimental data found in (Conrad and Phillips, 2019). Material parameters and techniques used in that model were employed in the current numerical simulation to ensure the validity and accuracy of the model. Moreover, the validation of a numerical model based on previous numerical simulations of similar systems as a technique was also used within the classification validation types conducted by (Archambeault and Connor, 2008). To sum up, the numerical model used to simulate the bare steel frame, as shown in Figure 3B, had been

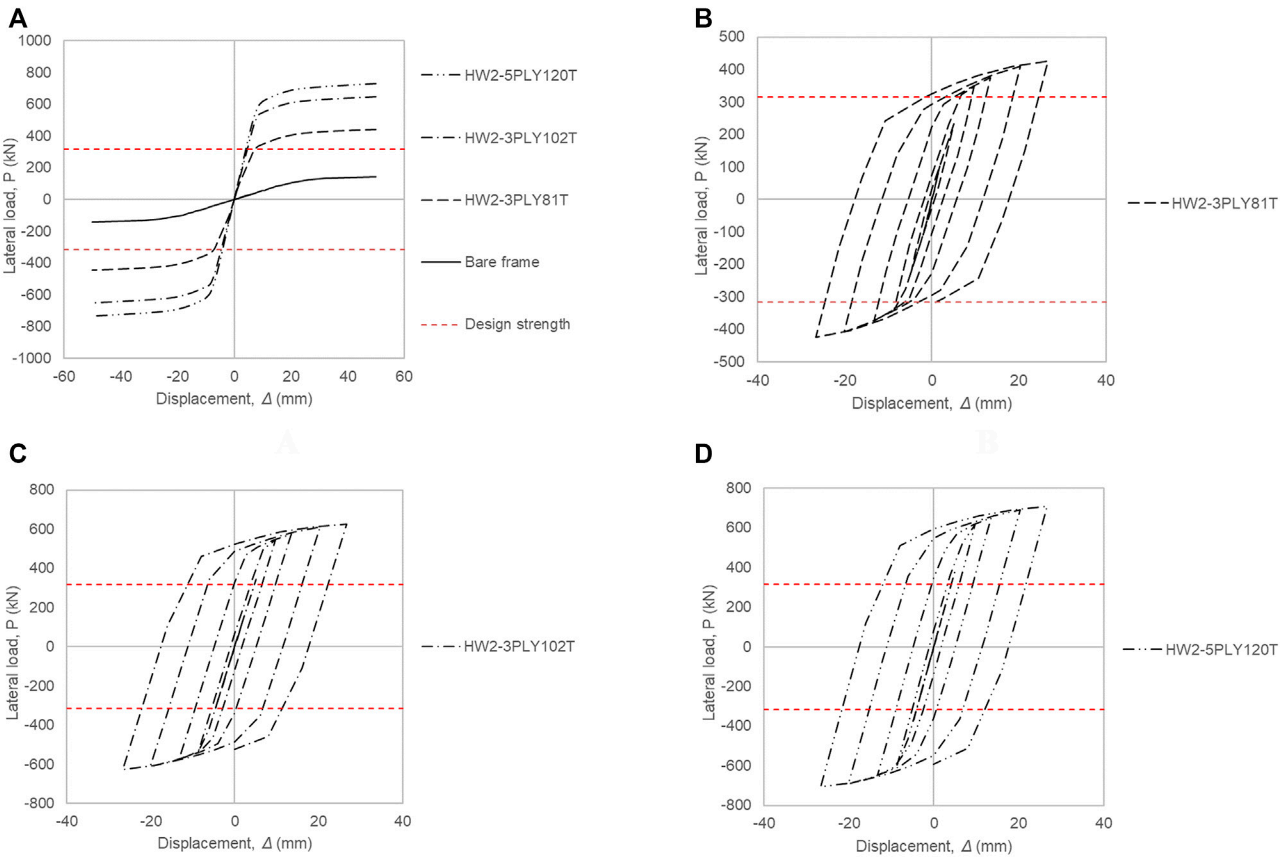


FIGURE 8
Hysteretic and backbone curves of specimens with connected infill-frame interface **(A)** Comparative diagram, **(B)** 3-ply (81 mm thick), **(C)** 3-ply (102 mm thick), **(D)** 5-ply (120 mm thick).

already validated against experimental results by (Vogiatzis, 2019). The CLT shear walls were modelled based on the methodology developed by (Vogiatzis et al., 2020) for a similar steel-timber hybrid system under monotonic loading which had been validated using experimental data. It has been shown in the literature that the Hill's criterion, as implemented in ANSYS, once validated for monotonic loading then it has the ability to be used for the investigation of the same specimen under cyclic loading conditions without further calibration on new inputs required (Rahmzadeh and Iqbal, 2021).

3 Results

The effect of the two main wall–frame connection types, on the response of CLT-Steel hybrid wall systems under monotonic and cyclic loading conditions is explored. To this end, several finite element models with thicknesses of the infill CLT walls t_w, equal to 81, 102 and 120 mm, respectively, are evaluated. The lateral load P versus displacement Δ curves obtained from this study are shown in Figures 4A–D for the frictional STSW models and in Figures 5A–D for the connected STSW models. Where, the horizontal line marks the designed shear strength of the prototype building.

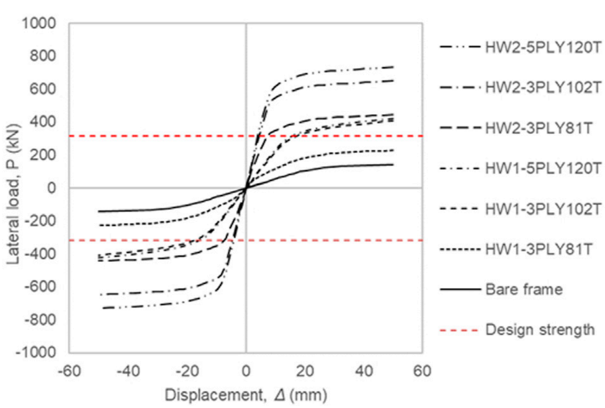


FIGURE 9
Backbone curves for the developed CLT-Steel hybrid systems.

TABLE 4 Summary of numerical results.

FE-model	K_e	Δ_y	$0.9P_{peak}$	G'	G_{bf}	μ	Ratio	
	(N/mm)	(mm)	(kN)	(N/mm)	(N/mm)	(-)	G'/G_{bf}	P_u/V_d
HW1-3PLY81T	13.61	13.43	182.78	174.96	70.54	1.96	2.48	0.58
HW1-3PLY102W	23.40	13.77	322.17	300.83	74.49	1.91	4.04	1.02
HW1-5PLY120W	24.78	13.46	333.45	318.65	74.55	1.95	4.27	1.06
HW2-3PLY81T	46.24	8.24	380.82	594.59	74.96	3.19	7.93	1.21
HW2-3PLY102W	66.85	8.42	562.55	859.55	74.97	3.12	11.46	1.78
HW2-5PLY120W	71.94	8.82	634.47	924.94	75.00	3.09	12.33	2.01

Ductility (μ) is a sign of displacement induced in plastic region without strength degradation. The ductility factor can be defined from Eq. 8. It is assumed that the P-Δ backbone curve is elastic and perfectly plastic according to the procedure explained and used by (Kennedy-Kuiper et al., 2022).

$$\mu = \frac{\Delta_u}{\Delta_y} \tag{8}$$

The yield displacement (Δ_{yield}) was measured through the concept of equal plastic energy, so that the area enclosed by the idealized elasto-plastic curve was equal to that of the actual pushover curve, as depicted in Figure 6. Figures 7A, 8A, 9 compare the backbone curves of the CLT-steel hybrid systems. The capacity parameters of the developed specimens are presented in Table 4. Where the stiffness (K_e) is computed by Eq. 9, the secant shear modulus of the CLT wall (G') is given by Eq. 10, and the secant shear modulus of the bare steel frame (G_{bf}) is produced by Eq. 11.

$$K_e = 0.4\frac{P_{peak}}{\Delta_{0.4P_{peak}}} \tag{9}$$

$$G' = \frac{0.4P_{peak}}{\Delta_{0.4P_{peak}}} \times \frac{H}{L} \tag{10}$$

$$G_{bf} = \frac{P_{bf}}{\Delta_{bf}} \times \frac{H}{L} \tag{11}$$

The findings of this study show that an ascent on the infill plate thickness from 81 to 102 mm can increase the shear capacity up to 40% for frictional STSW models, and up to 70% for frictional STSW models. Moreover, as presented in Figures 7, 8, the frictional STSW specimens fail to reach the design strength by 40% when the thickness is 80 mm but they provide adequate strength when the thickness of the infill is either 102 or 120 mm. From the same results it is obvious that this is not the case for the connected STSW specimens as their strength ratio is ranging from 1.06 to 2.01, meaning they provide overstrength more than 70% Comparison with the bare frame results, as presented in Figures 4, 5 and Table 4, show that the incorporation of the CLT infill wall can significantly increase the initial lateral stiffness of the bare steel frame, up to 3 times for the frictional STSWs and up to 11 times for the connected STSW systems. Further, it is observed that stronger infills led to higher lateral load capacity, initial lateral stiffness, and yield load. However, the increase of the ductility was limited.

4 Conclusion

Based on the findings of this research study, the following conclusions can be drawn within the limitation of the current research:

- For the connected hybrid steel-timber wall systems the lateral stiffness is decreased with the horizontal displacement increasing, showing a strong nonlinear feature, as shown before in Figures 4, 5.
- As expected, an increase in the number of the plies of the panels positively influences the load-carrying capacity of the wall. However, the contribution of their thickness (width) is more prominent since it affects to a greater extent the structural performance of the wall systems.
- An increase in the number of the plies of the panels does not necessarily reduces the displacement of the hybrid wall. This has been concluded by comparison of 3-ply and 5-ply CLT panels. On the contrary, a reduction in the thickness of the CLT infill results in the smaller displacement of the wall system.
- Among the models of the hybrid steel-timber wall systems examined, the models with the presence of friction have showed higher values of lateral drift capacities but in smaller lateral loads. This indicates the effect of friction on these types of wall systems.

- The connected STSW models have shown higher energy dissipation capacities than the frictional STSW models of the same wall systems. The increase in the energy absorption has been more significant in the case of the 5-ply CLT infilled hybrid wall.

Overall, hybrid CLT-Steel wall systems have shown great potential and they can be regarded as reliable alternative systems to traditional shear walls. The effects of parameters, such as different geometries and mechanical properties, that can affect the structural performance of STSW systems have not been examined as they have not been within the scope of the paper. However, these studies can produce interesting findings in addition to the results from the current study.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

TV, TT, and EE contributed conception and design of the study. All authors contributed to manuscript revision, read, and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

ANSYS Inc (2019). *Ansys®Academic Research Release*.

Archambeault, B., and Connor, S. (2008). "Proper model validation is important for all EMI/EMC applications," in IEEE International Symposium on Electromagnetic Compatibility 2008-Janua, August 18–August 22, 2008, Detroit Michigan, United States. doi:10.1109/ISEMC.2008.4652152

Carrero, T., Montaño, J., Berwart, S., Santa María, H., and Guindos, P. (2021). Seismic behavior of innovative hybrid CLT-steel shear wall for mid-rise buildings. *Bull. Earthq. Eng.* 19, 5917–5951. doi:10.1007/s10518-021-01204-y

Chaboche, J. L. (1989). Constitutive equations for cyclic plasticity and cyclic viscoplasticity. *Int. J. Plast.* 5, 247–302. doi:10.1016/0749-6419(89)90015-6

Chaboche, J. L. (1986). Time-independent constitutive theories for cyclic plasticity. *Int. J. Plast.* 2, 149–188. doi:10.1016/0749-6419(86)90010-0

Chen, X., and Liu, Y. (2016). A finite element study of the effect of vertical loading on the in-plane behavior of concrete masonry infills bounded by steel frames. *Eng. Struct.* 117, 118–129. doi:10.1016/j.engstruct.2016.03.010

Conrad, K., and Phillips, A. R. (2019). Full scale testing and development of wood-steel composite shear walls. *Structures* 20, 268–278. doi:10.1016/j.istruc.2019.04.010

Cui, Y., Chen, F., Li, Z., and Qian, X. (2020). Self-centering steel-timber hybrid shear wall: Experimental test and parametric analysis. *Materials* 13, 2518. doi:10.3390/ma13112518

Dickof, C., Stiemer, S. F., Bezabeh, M. A., and Tesfamariam, S. (2014). CLT-steel hybrid System: ductility and overstrength values based on static pushover analysis. *J. Perform. Constr. Facil.* 28, A4014012. doi:10.1061/(asce)cf.1943-5509.0000614

European Commision (2008). *Eurocode-8:– Design of structures for earthquake resistance – Part 1: General rules, seismic actions and rules for buildings*. Brussels.

Fardis, M., Carvalho, E., Elnashai, A., Faccioli, E., Pinto, P., and Plumier, A. (2005). *Designers guide to EN 1998-1 and EN 1998-5, Eurocode 8: Design of structures for earthquake resistance*. London, U.K: Thomas Telford Ltd.

FPInnovations Institute (2013). *CLT handbook: Cross-laminated timber*. US edition: Quebec, Canada.

Furtmüller, T., Giger, B., and Adam, C. (2018). General shell section properties and failure model for cross-laminated timber obtained by numerical homogenization. *Eng. Struct.* 163, 77–92. doi:10.1016/j.engstruct.2018.02.017

Hajjar, J. F. (2002). Composite steel and concrete structural systems for seismic engineering. *J. Constr. Steel Res.* 58, 703–723. doi:10.1016/S0143-974X(01)00093-1

He, M., Li, Z., Lam, F., Ma, R., and Ma, Z. (2014). Experimental investigation on lateral performance of timber-steel hybrid shear wall systems. *J. Struct. Eng. (N. Y. N. Y).* 140, 04014029. doi:10.1061/(asce)st.1943-541x.0000855

He, M., Sun, X., and Li, Z. (2018). Bending and compressive properties of cross-laminated timber (CLT) panels made from Canadian hemlock. *Constr. Build. Mat.* 185, 175–183. doi:10.1016/j.conbuildmat.2018.07.072

Hill, R. (1948). A theory of the yielding and plastic flow of anisotropic metals. *Proc. R. Soc. Lond A Math. Phys. Sci.* 193, 281–297. doi:10.1098/rspa.1948.0045

Hill, R. (1998). *The mathematical theory of plasticity*. New York, United States: Clarendon Press.

Igbal, A., Todorov, B., and Billah, A. (2020). "Combination of steel plate shear walls and timber moment frames for improved seismic performance," in 17th world conference on earthquake engineering (Sendai, Japan), 2.

Imaoka, S. (2008). *Hill's potential*. ansys.net Tips and Tricks.

INNO-HYCO (2015). *Innovative hybrid and composite steel-concrete structural solutions for building in seismic area, Final Report*. Brussels: EUR 26932 EN. doi:10.2777/85404

Kennedy-Kuiper, R. C. S., Wakjira, T. G., and Alam, M. S. (2022). Repair and retrofit of RC bridge piers with steel-reinforced grout jackets: An experimental investigation. *J. Bridge Eng.* 27, 1–11. doi:10.1061/(asce)be.1943-5592.0001903

Lempriere, B. M. (1968). Poisson's ratio in orthotropic materials. *AIAA J.* 6, 2226–2227. doi:10.2514/3.4974

Li, Z., He, M., Lam, F., Li, M., Ma, R., and Ma, Z. (2014). Finite element modeling and parametric analysis of timber-steel hybrid structures. *Struct. Des. Tall Spec. Build.* 23, 1045–1063. doi:10.1002/tal.1107

Makino, M., Kawano, A., Kurobane, Y., Saisho, M., and Yoshinaga, K. (1980). "An investigation for the design of framed structures with infill walls," in *The seventh world conference on earthquake engineering* (Turkey, Istanbul: Kelaynak), 369.

Margiacchi, F., Salvatori, L., Orlando, M., de Stefano, M., and Spinelli, P. (2016). Seismic response of masonry-infilled steel frames via multi-scale finite-element analyses. *Bull. Earthq. Eng.* 14, 3529–3546. doi:10.1007/s10518-016-0012-7

Mehrabi, A. B., Benson Shing, P., Schuller, M. P., and Noland, J. L. (1996). Experimental evaluation of masonry-infilled RC frames. *J. Struct. Eng. (N. Y. N. Y).* 122, 228–237. doi:10.1061/(asce)0733-9445(1996)122:3(228)

Pallarés, L., Agüero Ramon-Llin, A., Martí-Vargas, J. R., and Pallarés, F. J. (2020). Behaviour of headed studs subjected to cyclic shear in steel frames with reinforced concrete infill walls. *Constr. Build. Mat.* 262, 120018. doi:10.1016/j.conbuildmat.2020.120018

Pečnik, J. G., Gavrić, I., Sebera, V., Kržan, M., Kwiecień, A., Zając, B., et al. (2021). Mechanical performance of timber connections made of thick flexible polyurethane adhesives. *Eng. Struct.* 247, 113125. doi:10.1016/j.engstruct.2021.113125

Quintana Gallo, P., and Carradine, D. (2018). *State of the art of timber- based hybrid seismic- resistant structures*. Porirua, New Zealand: BRANZ research report SR400.

Rahmzadeh, A., and Iqbal, A. (2021). Numerical modelling and analysis of post-tensioned timber beam-column joint. *Eng. Struct.* 245, 112762. doi:10.1016/j.engstruct.2021.112762

Repapis, C., and Zeris, C. A. (2019). Seismic assessment of non-conforming infilled RC buildings using IDA procedures. *Front. Built Environ.* 4, 88. doi:10.3389/fbuil.2018.00088

Saari, W. K., Hajjar, J. F., Schultz, A. E., and Shield, C. K. (2004). Behavior of shear studs in steel frames with reinforced concrete infill walls. *J. Constr. Steel Res.* 60, 1453–1480. doi:10.1016/j.jcsr.2004.03.003

Shahrooz, B. M., Gong, B., and Arnold, J. (1996). *Seismic behavior and design of composite coupling beam: Internal Report*. Cincinnati, U.S.A: University of Minnesota.

Shi, Y., Wang, M., and Wang, Y. (2011). Experimental and constitutive model study of structural steel under cyclic loading. *J. Constr. Steel Res.* 67, 1185–1197. doi:10.1016/j.jcsr.2011.02.011

Stazi, F., Serpilli, M., Maracchini, G., and Pavone, A. (2019). An experimental and numerical study on CLT panels used as infill shear walls for RC buildings retrofit. *Constr. Build. Mat.* 211, 605–616. doi:10.1016/j.conbuildmat.2019.03.196

Tesfamariam, S., Stiemer, S. F., Dickof, C., and Bezabeh, M. A. (2014). Seismic vulnerability assessment of hybrid steel-timber structure: Steel moment-resisting frames with CLT infill. *J. Earthq. Eng.* 18, 929–944. doi:10.1080/13632469.2014.916240

Tong, X., Schultz, A. E., Hajjar, J. F., and Shield, C. K. (2001). *Seismic behavior of composite steel frame-reinforced concrete infill wall structural system: A report from the national science foundation U.S.-Japan cooperative research program phase 5, composite and hybrid structuresMineapolis, Minnesota 55455-0116*. United States: University of Minnesota.

Trutalli, D., Marchi, L., Scotta, R., Pozza, L., and Stefani, L. de (2017). Seismic response of a platform-frame system with steel columns. *Buildings* 7, 33. doi:10.3390/buildings7020033

Vogiatzis, T., and Avdelas, A. (2018). Study of composite steel frame with reinforced-concrete infill. *Proc. Institution Civ. Eng. - Struct. Build.* 171, 178–192. doi:10.1680/jstbu.16.00192

Vogiatzis, T. (2019). *Nonlinear numerical study on the behaviour of seismic resistant composite structural systems of steel moment frames with reinforced concrete infill walls*. Thessaloniki, Greece: Aristotle University of Thessaloniki. doi:10.12681/eadd/45230

Vogiatzis, T., Tsalkatidis, T., and Avdelas, A. (2019). Steel framed structures with cross laminated timber infill shear walls and semi-rigid connections. *Int. J. Eng. Technol.* 8, 433. doi:10.14419/ijet.v8i4.29742

Vogiatzis, T., Tsalkatidis, T., and Avdelas, A. (2020). Wood-steel composite shear walls with openings. *Int. J. Eng. Technol.* 10, 14. doi:10.14419/ijet.v10i1.31255

Wallace, J., and Wada, A. (2000). "Hybrid wall systems: US-Japan research," in Proceedings of 12th World Conference on Earthquake Engineering, January 30–February 04, 2000, Auckland, New Zeland.

Wang, Z., Gong, M., and Chui, Y. H. (2015). Mechanical properties of laminated strand lumber and hybrid cross-laminated timber. *Constr. Build. Mat.* 101, 622–627. doi:10.1016/J.CONBUILDMAT.2015.10.035

Xu, B.-H., Bouchaïr, A., and Racher, P. (2014). Appropriate wood constitutive law for simulation of nonlinear behavior of timber joints. *J. Mat. Civ. Eng.* 26, 04014004. doi:10.1061/(asce)mt.1943-5533.0000905

frontiers | Frontiers in Built Environment

TYPE Original Research
PUBLISHED 06 December 2022
DOI 10.3389/fbuil.2022.1049121

Check for updates

An analysis of the harassments and challenges faced by the public transport users in a developing country of South Asia

Samnun Yasir, Tanzim Ahmad and Annesha Enam*

Department of Civil Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

OPEN ACCESS

EDITED BY
Sakdirat Kaewunruen,
University of Birmingham,
United Kingdom

REVIEWED BY
Deepthi Mary Dilip,
Birla Institute of Technology and Science, United Arab Emirates
Hourakhsh Ahmad Nia,
Alanya Hamdullah Emin Pasa University, Turkey

*CORRESPONDENCE
Annesha Enam,
annesha@ce.buet.ac.bd

SPECIALTY SECTION
This article was submitted to Transportation and Transit Systems, a section of the journal Frontiers in Built Environment

RECEIVED 20 September 2022
ACCEPTED 21 November 2022
PUBLISHED 06 December 2022

CITATION
Yasir S, Ahmad T and Enam A (2022), An analysis of the harassments and challenges faced by the public transport users in a developing country of South Asia.
Front. Built Environ. 8:1049121.
doi: 10.3389/fbuil.2022.1049121

Bangladesh—located in south Asia and home to almost 167 M people, is one of the most densely populated countries in the world. Despite high population density, the country is yet to have a well-coordinated mass transit system. However, most people rely on public transportation due to comparatively low motorization. Public transport users, especially vulnerable cohorts (e.g., women), face enormous challenges, including molestations, assaults, and rape during their daily commute. Few studies that analyzed the state of harassment exclusively focused on women, which might not be comprehensive enough to understand the state of the problem and devise effective policies. Therefore, the current study explores the state of harassment in the three cities of Bangladesh, namely, Dhaka—the capital; Rajshahi; and Mymensingh across all genders. The study also investigates people's perception of women's mobility, the experience of using public transport, and the desirable safety precautions to understand the gender differences and the variability across the three cities. A detailed questionnaire survey was conducted to collect data from the three cities. The study found considerable gender differences regarding the perception of women's mobility hindrances, the experience of using public transportation, and desirable safety precautions. The differences across the three cities were also noticeable. Contrary to the general belief, the study found that people from smaller towns like Mymensingh were more likely to get harassed than those from the bigger cities like Dhaka and Rajshahi—insinuating the lack of reporting from the smaller towns of the country. The findings from the study could be helpful for the transit and city planners in creating a conducive transit ambiance in Bangladesh. Based on the desirable safety precautions, female-only rides, especially at night, would be beneficial. City planners could also plan small and medium business activities around transit stops to attract the crowd and reduce the possibility of getting harassed while walking alone to access transit.

KEYWORDS
public transport, harassments, challenges, developing country, gender

1 Introduction

Bangladesh is one of the most densely populated countries in the world—the population density of the country is almost three times that of the neighboring country, India[1]. Dhaka—the capital of Bangladesh, is one of the fastest-growing megacities. The country's population primarily relies on public transportation (e.g., public bus, taxi, CNG) for daily commutes. According to an inner cordon survey of the Dhaka metropolitan area, the modal share for public bus account for more than sixty percent[2]. However, a glaringly high percentage of the population, especially women, report abuse (87% of women reported violence against them[3]) in public transportation. Persistent crime could discourage the use of the public mode and push people towards individualized transportation modes which could be deadly for the country's traffic congestion.

Over the last few years, Bangladesh has become a vulnerable region to accelerating gender-based violence, especially in public places, which has restricted women's mobility (Mazumder and Pokharel, 2019; Closing the gaps for gender, 2016). It is postulated that many women in Bangladesh will endure sexual harassment, abuse, and assault at some point in their lives (Khairuzzaman, 2019). According to a national survey on violence against women, 21% of women consider public transportation the most likely place for sexual assault (Hossein). A recent study interviewing 2,500 women from every district of Bangladesh found that 90% of women and girls have suffered sexual harassment while on public transport (Ferdous and Dipu, 2019). Additionally, bus stops, footpaths, and stations are frequently reported as places of harassment in Bangladesh (Rahman, 2010).

Though most studies show that women are more likely to be victimized than men, violence against men on public transportation is not unheard of (Gordon and Roger, 1989). A study conducted on 5,000 participants in New Delhi, India, reported harassment of both men and women. However, sexual harassment is seen as the gravest threat to women's safety in Delhi, according to a study published a few years ago (Strategic_Framework.pdf, 2022). In developing countries such as India, women travel fewer distances within restricted geographical areas for multipurpose trips (e.g., shopping, child escort, healthcare) and consequently rely more on public transport than men (Need to make public transport, 2017). Furthermore, women are more likely than men to travel at a slower speed and spend a more significant percentage of their time on transportation, exposing them to gender-based violence and several forms of transit assault (Strategic_Framework.pdf, 2022). A survey on public transportation in Chennai, India, found that 30% of respondents had experienced all forms of abuse, with inappropriate touching (37%), stalking (37%), being asked for sexual favors (17%), and comments on physical attributes (14%) (Valan, 2020). According to another study conducted in Lucknow, India, 45% of respondents believed that young women are the most common targets of harassment on public transportation. However, there was no agreement on whether any attribute, such as traveling alone or during rush hours, making them particularly vulnerable to sexual harassment on public transportation (Tripathi et al., 2017).

Other developing countries such as Nepal, Pakistan, Indonesia, Turkey, and Bogota face similar challenges regarding harassment on public transportation. According to a study conducted among 280 female students in the Kathmandu valley, Nepal, 219 students faced harassment before (Gautam et al., 2019). As per another study in Nepal, 97% of the people surveyed had experienced at least one or more incidents of sexual harassment on public transport in Kathmandu. This shows that sexual harassment is a ubiquitous experience for consumers of public transportation in Nepal (Neupane and Chesney-Lind, 2014). According to research on public transit harassment in Lahore, Pakistan, 77% of those surveyed have been harassed. 14% of them said it happened to them regularly (Awan, 2020).

Similarly, Kirchhoff discovered in his research that 39% of the females surveyed in Jakarta, Indonesia, had been sexually harassed on public transportation and that the harassment incidents were a frequent occurrence for them (Kirchhoff et al., 2007). A study in Istanbul, Turkey, asked participants to share their experiences utilizing public transportation. All participants reported ample negative experiences, including discrimination, assault, or harassment. The reported incidents varied from minor harassment (e.g., finger-pointing) to physical assault (Shakibaei and Vorobjovas-Pinta, 2022). Bogota, the capital of Colombia, was named the most dangerous place to travel alone at night, especially for women (EXCLUSIVE-, 2014). According to a survey on public transportation in Bogotá, 84% of women have been harassed at some point. Interestingly, only 10% of women who have experienced harassment have reported it—demonstrating severe underreporting (Quinones, 2020).

The harassment scenario on public transportation is not very different in the more developed countries either. The most susceptible cohort to such harassment in Japan is middle-aged employed women and school and university students (Burgess and Horii, 2012). In Seoul, South Korea, sexual harassment is pervasive in packed vehicles as passenger mobility is limited, resulting in unwanted physical contact (Kim et al., 2020). The number of transport assaults has also climbed drastically in recent years in Hong Kong. Chui and Ong conducted a study on public transit in Hong Kong, finding 125 cases of sexual assault over 10 months in 2006 (Chui and Ong, 2008). The most

---

1 https://ourworldindata.org/most-densely-populated-countries - Accessed on October 2022.

2 ESCAP (2018). Final report on Sustainable Urban Transport Index for Dhaka, Bangladesh.

3 https://www.ucanews.com/news/most-bangladeshi-women-suffer-abuse-on-public-transport/98579 - Accessed on October 2022.

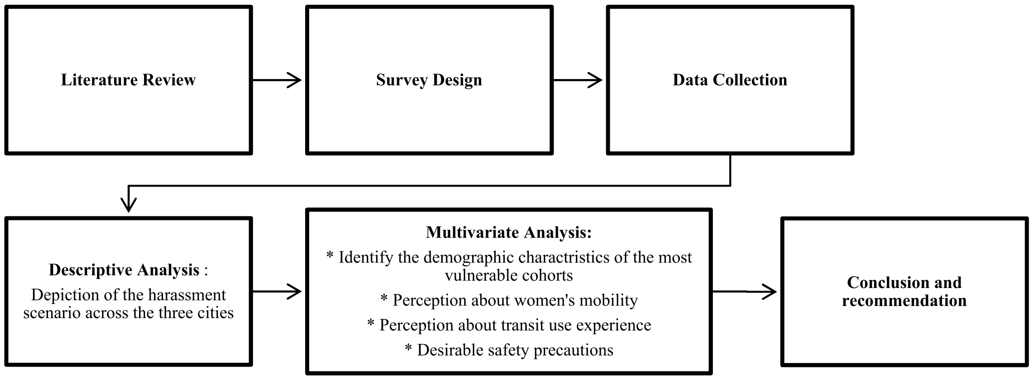


FIGURE 1
Methodology of the current study.

prevalent forms of harassment on public transportation in the United States include groping, unwanted touch, stalking, and accosting (Hsu, 2011). According to an international survey conducted in 2014, in London, 32% of women are being harassed regularly, with 19% having experienced direct physical assault on public transportation (Foundation, 2022).

From the literature review above, it can be inferred that harassment inside public transportation is common in Bangladesh and other developing and developed nations. Patrolling public transportation could be one of the ways to reduce the occurrence of such incidents. However, continuous policing is not feasible, especially in a high-density country like Bangladesh. Instead, educating harassers and enabling victims to undertake desirable safety measures could be more effective means to safeguard travelers. Such measures would require a better understanding of individuals' perceptions of the challenges and the acceptable safety precautions. There is a dearth of systematic studies, especially in the context of developing nations, that explored the root cause of the spreading of violence inside the transportation system. Very few researchers that have investigated these issues have often only analyzed women's opinions on the subject (Tokey and Shioma, 2017)—which is not comprehensive enough to understand the depth of the problem and determine the causes of prejudice against victims. Therefore, this study attempts to understand the barriers to safe mobility inside transportation by investigating people's perceptions and desirable safety precautions. The study also investigates the harassment scenario across all genders in three different cities of Bangladesh.

The current study has four primary objectives, namely (i) depict the harassment scenario in the three major cities of Bangladesh—Dhaka, Rajshahi, and Mymensingh; investigate the gender difference in the (ii) perception of women's mobility; (iii) experience of using public transportation; and (iv) desirable safety precautions. The variability across the three cities is also investigated. To this end, the study has undertaken a primary data collection effort—a survey is designed to solicit information from different demographic groups residing in Dhaka, Rajshahi, and Mymensingh. The collected information is analyzed using the descriptive analysis technique, followed by the development of multivariate models which control for different demographic factors while investigating the effect of gender and geography. Specifically, a binary logit model is developed to identify the most vulnerable cohorts of the three cities. To understand the perception of different demographic cohorts on women's mobility, public transportation experience, and desirable safety precautions, multiple ordered probit models are developed. To the best of the authors' knowledge, this is one of the first studies to investigate the heterogeneity in the perception towards women's mobility, the experience of riding public transportation, and desirable safety precautions to devise policies against harassment inside public transportation. The findings from the study will be helpful for city and transportation planners to promote a safe and secure transit system in Bangladesh and other countries with similar socio-economic backgrounds.

The rest of the paper is organized as follows. The next section provides a brief description of the survey. The harassment scenario of the three cities is also depicted in this section through descriptive analysis. The section briefly elaborates on the two statistical model forms used in the analysis—the binary logit and the ordered probit models. The result section presents the findings of the different statistical models. The paper ends with a succinct description of the research findings. The policy implications of the findings are also presented in the discussion section.

2 Materials and methods

The research started with a literature review. Next, a survey was designed to understand the respondents' perceptions about women's mobility, the experience of using public transportation, and the desirable safety precautions of transit users. The survey intended to collect information about the harassment incidents encountered by the survey participants. A detailed descriptive analysis was conducted to understand the harassment scenario across the three cities. Particularly the descriptive analysis focused on the spatial and temporal distribution and the modal distribution of the harassment incidents. The descriptive analysis was followed by developing a binary logit model to capture the demographic characteristics of the most vulnerable population. Additionally, multivariate analysis is conducted to understand the gender and geographic differences in the perception of women's mobility, public transportation experience, and the desirable safety precautions. A flow chart depicting the significant steps of the study is provided in Figure 1.

2.1 Survey

The study conducted a detailed questionnaire survey in the three cities of Bangladesh, namely, Dhaka, the capital city; Rajshahi; and Mymensingh, between September and November 2021. The survey was distributed online and restricted to people over 16 years old. The data collection followed the convenience sampling technique since we used our social media network to augment the number of respondents. The survey was distributed among all genders since we intended to depict the scenario of harassment and understand the perception of harassment among both genders.

The survey included multiple sections on demographics (e.g., age, gender, household income, education level, family structure, marital history, and so on), daily travel behavior (e.g., household car ownership, frequency of trips executed for a variety of purposes by different modes of transportation), perception about women's mobility, experience of using public transportation, desirable safety precautions to reduce harassment, and personal and family harassment history. The answers to the perception questions (towards women's mobility, public transportation, and safety measures) were collected on a five-point Likert scale ranging from strongly agree to disagree strongly. To capture participants' perceptions of the barriers to women's mobility, the respondents were asked to indicate their (dis) agreement about multiple contributing factors such as fear, discrimination in rights, income, employment, and tendency to rely on others. The experience within public transport was recorded on several dimensions, such as space and congestion inside the vehicles, the driver/conductor's behavior, and the quantity and condition of seats. To understand the safety precautions, the questionnaire asked to indicate different actions the respondents exercise, such as avoiding rush hours, traveling alone, traveling at night, and sitting closer to drivers. The last section of the survey collected information about respondents' and their family-friends' harassment history, including the location, time, and the severity of the incident if they ever faced any. There was also an open-ended question asking respondents to provide additional detail about their harassment incident.

2.2 Descriptive data analysis

Table 1 summarizes the distribution of the important socio-economic variables. The survey collected information from 452 respondents. Almost 2/3rd respondents are female, and the rest are male. About 50% of the respondents belong to the 16–25 years age group—the high representation of this age group might be due to the tech savviness of this generation (as mentioned earlier, the survey was conducted online). Only 2% of the survey respondents are above 55 years old. Due to the high representation of the young age group, 60% of the survey respondents are students, and only 34% of the respondents are employed. Slightly less than half of the respondents are married. 42% of the sample completed undergraduate education, and 20% of the respondents completed higher secondary certificate level education. Slightly more than 70% of the respondents' household consists of more than three members. Around 40% of the respondents have a household income of more than 50,000 BDT per month[4].

2.2.1 Spatial and temporal distribution of the harassment incidents

Figure 2 presents the proportion of respondents who faced harassment at least once in their lifetime. The harassment proportion of female respondents in each of the three cities is much more significant than that of male respondents. Also, Mymensingh has the highest rate of harassment towards women than the other two cities.

Figure 3 presents the spatial and temporal distribution of the harassment incidents faced by women in the three cities. According to the figure, for most women (around 1/3rd of the sample), the harassment happened inside public transportation—this trend is very similar across the three cities. Similarly, the percentage of women harassed while getting into vehicles is higher in the major cities—Dhaka and Rajshahi than in Mymensingh. It should be mentioned that due to extremely high population density and poor transit system, public transportation is usually overcrowded in the country. Hence, harassing women while boarding transit vehicles is a widespread occurrence. The lower reporting of such occurrences in Mymensingh might be due to the relatively less crowded

4 1 US Dollar = 101.81 BDT as f 26 Nov, 2022.

TABLE 1 Distribution of the socio-economic variables of the sample.

Sample size		452			
Socio-demographic characteristics		% (count) (All 3 cities)	% (count) (Dhaka)	% (count) (Rajshahi)	% (count) (Mymensingh)
Gender	Male	42.1% (172)	50.9% (85)	33.8% (48)	39.39% (39)
	Female	57.7% (236)	49.1% (82)	66.19% (94)	60.6% (60)
Age	16–25	47.7% (195)	62.87% (105)	41.26% (59)	31.31% (31)
	26–35	22% (90)	15.57% (26)	27.27% (39)	25.25% (25)
	36–45	18.1% (74)	14.97% (25)	15.38% (22)	27.27% (27)
	46–55	10.5% (43)	5.99% (10)	13.98% (20)	13.13% (13)
	55+	1.7% (7)	0.598% (1)	2.09% (3)	3.03% (3)
Marital status	Married	40.8% (167)	25.75% (43)	48.25% (69)	55.56% (55)
	Unmarried	55.8% (228)	70.66% (118)	48.25% (69)	41.41% (41)
	Divorced	1.7% (7)	2.99% (5)	0.69% (1)	1.01% (1)
	Widow/Widower	1.7% (7)	0.59% (1)	2.89% (4)	2.02% (2)
Student	Yes	60.6% (243)	72.12% (119)	55% (75)	49.5% (49)
	No	39.4% (158)	27.88% (46)	45% (62)	50.5% (50)
Highest level of education	I have not received any formal education	0.49% (2)	0.60% (1)	0.69% (1)	—
	Secondary school (up to class eight)	0.98% (4)	1.2% (2)	—	2.04% (2)
	Secondary School Certificate (SSC)	2.7% (11)	2.41% (4)	4.19% (6)	1.02% (1)
	Higher Secondary Certificate (HSC)	20.64% (84)	26.51% (44)	16.78% (24)	16.33% (16)
	College Degree	12.04% (49)	11.45% (19)	12.59% (18)	12.24% (12)
	Undergraduate (4-year university degree)	42.01% (171)	42.77% (71)	46.85% (67)	33.67% (33)
	Postgraduate - Master's degree	13.02% (53)	10.24% (17)	13.98% (20)	16.33% (16)
	Postgraduate - Ph.D	4.91% (20)	4.21% (7)	4.89% (7)	6.12% (6)
	Technical Education	1.23% (5)	—	—	5.1% (5)
	Diploma	1.96% (8)	0.60% (1)	—	7.14 (7)
Household Members	1	0.7% (3)	1.19% (2)	—	1.01% (1)
	2	4.2% (17)	3.59% (6)	4.89% (7)	4.04% (4)
	3	27.4% (112)	23.95% (40)	24.48% (35)	37.37% (37)
	4	45.2% (185)	44.31% (74)	55.24% (79)	32.32% (32)
	More than 4	22.5% (92)	26.95% (45)	15.38% (22)	25.25% (25)
Household Members (Below 16-year-old)	0	53.8% (220)	58.08% (97)	55.24% (79)	44.44% (44)
	1	32.8% (134)	29.34% (49)	31.46% (45)	40.40% (40)
	2	10% (41)	7.78% (13)	11.89% (17)	11.11% (11)
	3+	3.4% (14)	4.79% (8)	1.39% (2)	4.04% (4)
Household's monthly income (in Bangladeshi Taka or BDT)	Less than 5,000	0.24% (1)	0.59% (1)	—	—
	5,000–10,000	1.7% (7)	1.79% (3)	1.39% (2)	2.02% (2)
	11,000–20,000	3.4% (14)	5.39% (9)	2.09% (3)	2.02% (2)
	21,000–30,000	10.5% (43)	8.38% (14)	13.98% (20)	9.09% (9)
	31,000–40,000	17.8% (73)	11.38% (19)	19.58% (28)	26.26% (26)
	41,000–50,000	19.6% (80)	17.37% (29)	20.97% (30)	21.21% (21)
	More than 50,000	40.1% (164)	46.71 (78)	36.36% (52)	34.34% (34)
	I prefer not to say	6.6% (27)	8.38% (14)	5.59% (8)	5.05% (5)

(Continued on following page)

 10.3389/fbuil.2022.1049121

TABLE 1 (*Continued*) Distribution of the socio-economic variables of the sample.

Sample size		**452**			
Socio-demographic characteristics		**% (count) (All 3 cities)**	**% (count) (Dhaka)**	**% (count) (Rajshahi)**	**% (count) (Mymensingh)**
Employment	Employed	34% (139)	28.74% (48)	32.17% (46)	45.45% (45)
	No - not looking for employment	48.2% (197)	53.89% (90)	50.35% (72)	35.35% (35)
	No - looking for employment	17.8% (73)	17.36% (29)	17.48% (25)	19.19% (19)

condition of the transit vehicles in Mymensingh than in the other two cities. The footpaths in Rajshahi are less safe than that in the other two cities.

Interestingly, in terms of temporal distribution, in Dhaka—the capital city, the harassment is almost equally likely to happen between 11 a.m. and 11 p.m. In Rajshahi, the highest harassment occurs during the evening rush hours. In general, the rush hours facilitate more public contact and hence more opportunities for misconduct. Dhaka is a more densely populated city than Rajshahi and remains crowded for most of the day, spreading the harassment incidents over a longer time window. On the other hand, in Mymensingh, the highest amount of harassment happen between 7 p.m. and 11 p.m. Mymensingh being a smaller town becomes quieter earlier than the other two cities—this might create an opportunity for the harassers to indulge in misconduct during this time window.

According to Figure 3, every form of the harasser is approximately equally involved in the harassment incidence, indicating a lack of education and respect towards people. Harassment by standing co-passengers is significantly higher in Rajshahi, while it is lower by the driver and helpers than in the other two cities. The narration of the harassment incidents by some of our respondents is provided below.

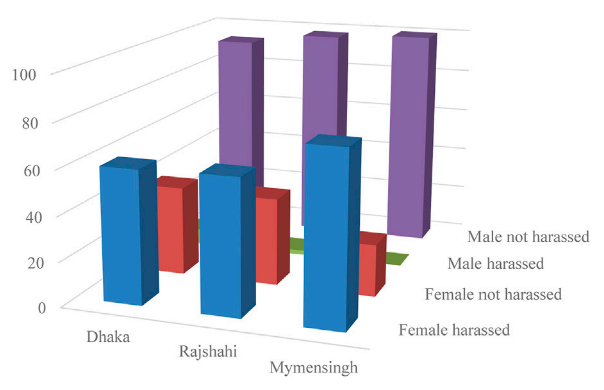


FIGURE 2
Distribution of harassed respondents by gender.

1) "I was trying to get to my seat, my co-passenger stretched his leg, and I fell. He also tried to touch my private parts. I could do nothing but leave the bus."
2) "Helper tried to touch my back while I was getting off the bus."
3) "A middle-aged guy was saying double-meaning words towards me. That was so embarrassing."
4) "My sister was going to meet one of her friends *via* Auto Rickshaw. A co-male passenger, after a while, joined the ride and tried to sit close to her though she had notified him to maintain a distance. However, the male passenger did not listen to my sister. My sister was a little afraid and stopped the driver and left the ride immediately."

2.2.2 Modal distribution of the harassment incidents

According to Figure 4, most harassment occurs on public buses, the predominant type of public transportation in Bangladesh. Interestingly, harassments are highly likely to occur inside CNG–one form of auto-rickshaw driven by Compressed Natural Gas, hence the name CNG—in Mymensingh than in the other two cities. This could be due to the differences in the operational characteristics of this mode across the cities. In Mymensingh, unrelated travelers share trips in CNG—while in Dhaka and Rajshahi, only friends and families travel as a group in CNG, reducing the chances of harassment. The percentage of harassment in Lagoona (human haulers) is higher in Dhaka than in Rajshahi and Mymensingh. This might be due to a higher share of Lagonna (human haulers) operating in the capital than in the other cities. According to recent statistics, 18,025 registered human haulers are operating around the country—of which almost one-third are operating in the capital city (Human haulers a threat in, 2022).

2.3 Statistical model

The descriptive analysis conducted above helps understand the impact of the single independent variable on the endogenous variable—for example, the probability

10.3389/fbuil.2022.1049121

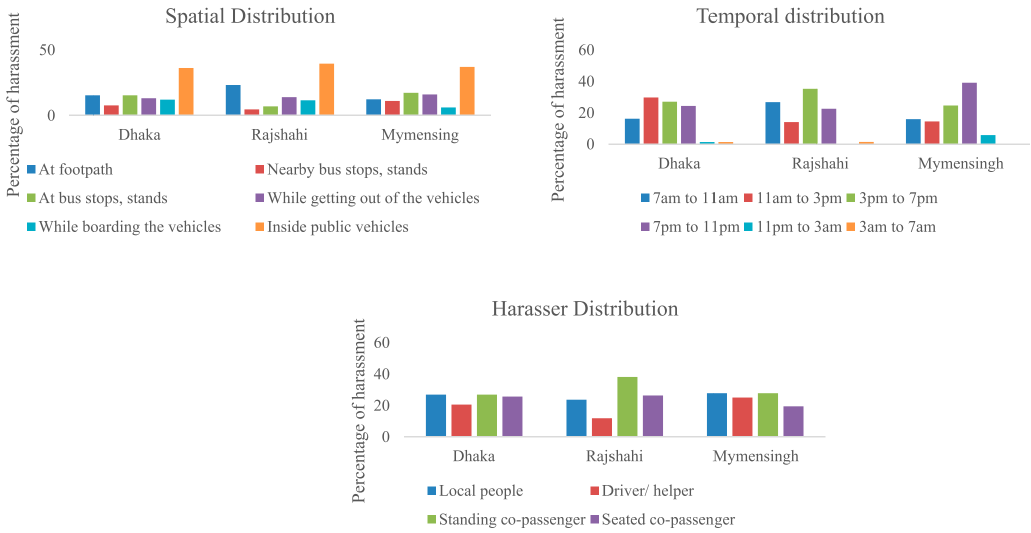


FIGURE 3
Spatial, temporal, and harasser distribution of the harassment incidents.

of getting harassed. Multivariate analysis is necessary to capture the relative influence of multiple exogenous variables on the variable(s) of interest. In this study, the heterogeneity in the probability of getting harassed is captured using the binary logit model. The heterogeneity in the responses to different statements capturing perception towards mobility challenges, riding experience in public transportation, and desirable safety precautions are captured using the ordered probit model. It can be noted that the ordered probit model captures the effect of different demographic characteristics. However, primarily the models plan to address the variation across gender and geography in Bangladesh. Below is a brief description of the two model forms—binary logit and ordered probit.

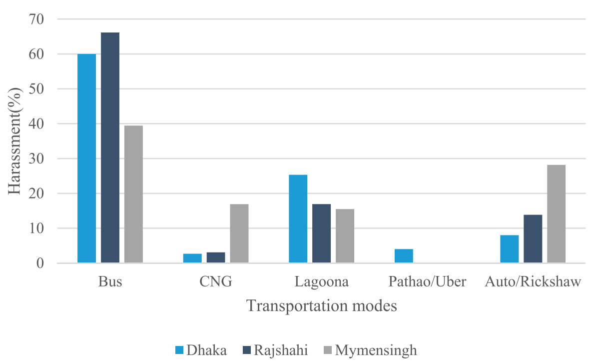


FIGURE 4
Harassment in various modes.

2.3.1 Binary logit model

The probability of an individual getting harassed can be expressed using equation (Mazumder and Pokharel, 2019) below.

$$P_{ih} = \frac{\exp\ (U_{ih})}{\exp(U_{ih}) + \exp(U_{inh})} \tag{1}$$

In Equation 1, P_{ih} is the probability of an individual, i, getting harassed, and U_{ih} and U_{inh} represent the "(dis)utility" of getting harassed and not harassed, respectively. U_{ih} is a function of multiple exogenous variables, including demography (e.g., location. gender, income, education) and travel characteristics (e.g., trip purpose). U_{ih} and U_{inh} can be expressed using Eq. 2 below, where β is the vector of the parameter to be estimated, X is the vector of exogenous variables, and ε is the vector of random error assumed to be Gumbel distributed. In Eq. 2, the coefficient vector, β, is estimated concerning the harassed "alternative" or, in other words, not harassed alternative serves as the base alternative in the model.

$$U_{ih} = \beta' X_i + \varepsilon_{ih} \tag{2a}$$

$$U_{inh} = \varepsilon_{inh} \tag{2b}$$

2.3.2 Ordered probit model

As mentioned earlier, the (dis)agreement to multiple statements captures the perception of women's mobility, the experience of using public transportation, and desirable safety precautions. The responses to these statements were collected on a Likert scale. Hence, the responses are treated as categorical variables. The latent propensity that derives the responses to the statements could be captured using Equation 3 below.

TABLE 2 Binary logit models results: who are more likely to get harassed.

Independent variables	Who are more likely to get harassed	
	Estimate	Robust t-ratio
Constant (harassed)	−3.503	−7.65
Female Indicator	3.703	9.45
Location Indicator (Dhaka)	−0.450	−1.24
Location Indicator (Rajshahi)	−0.285	−0.87
Trip Purpose Indicator by Bus (Shopping)	0.760	0.92
Trip Purpose Indicator by Bus (Recreation)	1.109	1.26
Trip Purpose Indicator by Bus (Work)	1.331	1.52
Student Indicator	0.596	1.87
Employment Indicator (Employed)	1.064	3.25
Household Vehicle Indicator (Private Car)	−0.121	−0.39

$$U_{ik}^{*} = \alpha_k^{'} X_i + \varepsilon_{ik} \quad (3)$$

Where U^* is the latent propensity of individual i towards statement k, α is the vector of parameters, and X is the vector of exogenous variables, including the individual's demographic characteristics and travel conditions. ε is the vector of random error that is assumed to be independently, identically, and logistically distributed across the individuals and the statements.

The latent propensity could be related to the responses of the individual using Equation 4 below with the help of the threshold parameters.

$$U_{ik} = j \, if \, \theta_k^j < U_{ik}^{*} < \theta_k^{j+1} \quad (4)$$

In Equation 4, j is the response category of the individual i to the k^{th} statement, and θ is the corresponding threshold parameter.

The binary logit and the ordered probit models are estimated using the maximum likelihood estimation technique. The Apollo package in R[5] is used to estimate the models.

3 Results

3.1 Likelihood of being harassed

The binary logit model results on the most likely cohort to get harassed are provided in Table 2. It can be noted that the primary objective of the model is to identify the gender and geographic difference in the propensity to get harassed; however, the model controls for other demographic attributes and trip characteristics. According to the model results, people from Dhaka and Rajshahi are less likely to get harassed than those from Mymensingh. One reason could be that in the bigger cities, harassers are often afraid to engage in widespread misconduct due to the availability of closed-circuit (CC) cameras. Interestingly, this finding contradicts the widespread belief that people from bigger cities are more prone to harassment than those from smaller towns. Low media coverage in the smaller towns could be the reason behind such intuition. The study reveals that being a student or a female increases the likelihood of getting harassed. Being employed also increases the probability of getting harassed significantly. The findings are similar to those obtained from other developing cities. For example, 55% of the women surveyed in Karachi, Pakistan, reported being sexually harassed on public transportation. In comparison, only 16% of working women and 18% of students said they had never been accosted on public transit (Rapid Assessment of Sexual Harassment, 2014). Employed people are more likely to travel alone, increasing their chances of getting harassed. People traveling to work by bus are very likely to get harassed. The study found that owning household vehicles decreases the possibility of getting harassed. However, this finding is not statistically significant.

3.2 Perception of Women's mobility

Table 3 presents the results on the perception of women's mobility. Four statements were used to capture the perception of women's mobility. The results of the four ordered probit models corresponding to four statements are discussed below.

3.2.1 Statement: Fear of possible victimization hinders women's mobility

Previous research has posited that fear of potential violence influences women's travel, consequently limiting their mobility (Mukerjee, 2019). However, whether different cohorts of society brace such a notion is yet to be verified. According to the current

5 Hess, S., and D. Palma. 2019. "Apollo: A Flexible, Powerful and Customisable Freeware Package for Choice Model Estimation and Application." Journal of Choice Modelling 32: 100170.

10.3389/fbuil.2022.1049121

TABLE 3 Ordered probit model results on perception of women's mobility.

Independent variables	Women's mobility hindrance			
	Fear of possible victimization	Income/employment status	Gender Discrimination	Tendency to rely on Others
	Estimate (Robust t-ratio)	Estimate (Robust t-ratio)	Estimate (Robust t-ratio)	Estimate (Robust t-ratio)
Female Indicator	0.396 (2.57)	0.404 (2.79)	0.589 (3.65)	0.164 (1.07)
Harassment Indicator (Individual)	—	−0.129 (−0.92)	0.229 (1.71)	0.214 (1.49)
Harassment Indicator (Friends & Family)	0.207 (1.67)	—	—	—
Location Indicator (Dhaka)	−0.447 (−2.63)	−0.263 (−1.99)	0.218 (1.45)	−0.104 (−0.73)
Location Indicator (Rajshahi)	−0.969 (−5.97)	−0.132 (−0.99)	0.495 (3.38)	0.211 (1.44)
Age Indicator (26–35)	—	—	−0.291 (−1.85)	0.207 (1.49)
Age Indicator (36–45)	0.272 (1.65)	—	−0.574 (−3.08)	0.296 (2.22)
Age Indicator (More Than 45)	—	—	−0.642 (−3.11)	—
Married Female Indicator	−0.223 (−1.39)	—	0.213 (1.22)	—
Employment Indicator (Employed)	0.319 (2.23)	0.055 (0.47)	—	−0.185 (−1.45)
Employment Indicator (Looking for Employment)	0.302 (1.66)	—	—	0.228 (1.33)
Household Vehicle Indicator (Motorcycle)	−0.378 (−3.01)	0.146 (1.31)	0.269 (2.33)	0.250 (1.99)
Household Vehicle Indicator (Private Car)	−0.195 (−1.28)	0.225 (1.73)	0.080 (0.60)	0.172 (1.27)
Dependent Children Indicator (One or More than one)	—	−0.131 (−1.19)	—	—

study, irrespective of geography, women are more likely (than men) to agree that fear of victimization is a likely cause of their mobility impairment. On the other hand, respondents from Dhaka and Rajshahi are less likely than Mymensing to agree that fear impairs women's mobility. Dhaka and Rajshahi being more urbanized, have fostered women's empowerment to a greater extent. Consequently, these two cities have higher women's employment rates than Mymensingh. Even though Dhaka is rated seventh in the world for violence against women (Report, 2017)- women continue to travel for jobs and survival in the city. Hence, people are less likely to agree that women being fearful of possible victimization is one of the major causes of their impaired mobility.

Interestingly, married women are less likely to agree that the fear of being harassed restricts their mobility. Married women in Bangladesh are more likely than unmarried women to be accompanied while traveling. This might help married women to be less fearful about possible victimization. Compared to those not working, employed respondents and those seeking employment regard fear of victimization on public transportation as a substantial mobility impediment. The trip-making frequency is higher among the employed cohort—which might make them more dreadful about the harassment situation. Interestingly, respondents with household-owned vehicles and motorcycles are less likely to agree that fear of victimization is a likely cause of women's mobility impairment. Respondents with access to household vehicles might be less dependent on public transportation resulting in such a perception.

3.2.2 Statement: Income and employment status hinders women's mobility

Women consider unfavorable income and employment status impair women's mobility. People from Dhaka and Rajshahi are less likely than Mymensingh to agree that the poor income and employment condition of women is a likely reason behind their lack of mobility. Dhaka and Rajshahi being more progressive might be the reason for not supporting the claim as strongly as Mymensingh. Employed individuals agree with the perception more strongly than unemployed individuals. People owning household cars and motorcycles also agree that dissimilar income and employment create additional challenges to women's mobility.

3.2.3 Statement: Discrimination in gender hinders women's mobility

Compared to males, females are more agreeable regarding gender discrimination as a barrier to mobility. Interestingly, unlike the other two statements, Dhaka and Rajshahi respondents agree more with this statement than Mymensingh's. Notably, people from bigger cities feel gender discrimination more strongly than those from smaller cities. People in the middle age group agree with the statement less vigorously than those in the younger and older cohorts. Unlike

TABLE 4 Ordered probit model results on the experience of using public transportation.

Independent variables	Bus conductor and driver misbehave with women	Inside of public vehicles are overcrowded	The condition of the seat is very poor	Space inside public vehicles is not sufficient
	Estimate (Robust t-ratio)	Estimate (Robust t-ratio)	Estimate (Robust t-ratio)	Estimate (Robust t-ratio)
Female Indicator	0.549 (3.21)	0.099 (0.65)	0.151 (1.03)	0.624 (4.00)
Harassment Indicator (Individual)	0.101 (0.69)	0.154 (1.06)	0.202 (1.68)	0.252 (1.71)
Harassment Indicator (Friends & Family)	—	0.249 (1.97)	—	0.139 (1.15)
Location Indicator (Dhaka)	0.266 (1.62)	0.295 (1.94)	−0.067 (−0.47)	−0.239 (−1.70)
Location Indicator (Rajshahi)	0.306 (1.98)	0.348 (2.32)	0.305 (2.29)	−0.268 (−1.89)
Age Indicator (26–35)	0.311 (1.88)	−0.554 (−3.32)	0.198 (1.36)	0.249 (1.80)
Age Indicator (36–45)	0.219 (1.16)	−0.769 (−5.47)	0.219 (1.54)	0.299 (2.41)
Age Indicator (More Than 45)	0.325 (1.57)	−0.577 (−3.14)	—	0.263 (1.53)
Married Female Indicator	−0.337 (−0.01)	—	—	—
Employment Indicator (Employed)	0.143 (1.03)	—	—	—
Household Vehicle Indicator (Motorcycle)	—	—	—	0.083 (0.74)
Household Vehicle Indicator (Private Car)	0.226 (1.60)	—	—	0.137 (1.02)

the fear of victimization statement, married women do consider gender discrimination to be a reason for mobility hindrance. This might reveal the social norm of the country, where anecdotal evidence suggests that married women's lives are more closely monitored (most often by the spouse and in-laws) than those of unmarried women. Respondents from households owning motorcycles agree with the statement strongly. The country's predominant motorcycle owners/drivers are young males (Bray and Holyoak, 2015). Therefore, having a motorcycle might make the household members more aware of gender discrimination.

3.2.4 Statement: Relying on others hinders women's mobility

Gender difference is least pronounced in this statement than in the last three. Individuals from Rajshahi support this claim more than those from Dhaka and Mymensingh. Unlike the previous statement, the middle age group agrees with this statement more than the younger and older age groups. Employed people are less likely to believe in this statement. Having household-owned cars and motorcycles also make people agree with this statement more strongly.

3.3 Experience of traveling in public transportation

Table 4 provides the results of the respondents' experiences while using public transportation. The experience of using public transportation was captured using four statements. The results of the four ordered probit models corresponding to the four statements are discussed below.

3.3.1 Statement: Bus conductor and driver misbehave with women

Females agree more strongly with this statement than males. Moreover, people with friends and families who have experienced harassment before agreed with this statement more strongly. Respondents from Dhaka and Rajshahi are more likely to support this statement than those from Mymensingh. This might indicate that public bus operators from bigger cities are more likely to harass passengers than those from smaller towns like Mymensingh. Individuals in the middle age group agree with the statement more strongly than the younger and older cohorts. It is interesting to note that married women are less likely to agree with this statement. Married women, most often being accompanied by fellow travelers, might face such harassment less than their unmarried counterparts. Employed people agree with the statement slightly more strongly than those who are not employed. Frequent travel might be the reason behind the comparatively more bitter experience of the employed cohort than those who are not employed. Also, people belonging to households with access to private cars agree with the statement more intensely than people from households without private vehicles.

3.3.2 Statement: Inside public vehicles are overcrowded

Unlike the previous statement, people from opposite genders are less likely to argue about this statement—indicated by the non-significant coefficient to the female indicator. Having harassment

TABLE 5 Ordered probit model results on safety measures taken by individuals.

Independent variable	Avoid rush hour	Avoid night travel	Avoid walking alone	Sit close to the driver
	Estimate (Robust t-ratio)	Estimate (Robust t-ratio)	Estimate (Robust t-ratio)	Estimate (Robust t-ratio)
Female Indicator	0.187 (1.32)	1.081 (8.844)	0.619 (3.86)	0.189 (1.22)
Harassment Indicator (Individual)	0.123 (0.93)	—	0.212 (1.46)	0.189 (1.25)
Harassment Indicator (Friends & Family)	—	0.146 (1.18)		—
Location Indicator (Dhaka)	0.043 (0.28)	−0.256 (−1.17)	−0.209 (−1.52)	0.226 (1.64)
Location Indicator (Rajshahi)	−0.184 (−1.31)	−0.193 (−1.29)	−0.224 (−1.63)	0.062 (0.46)
Age Indicator (26–35)	—	—	0.286 (1.99)	0.489 (3.18)
Age Indicator (36–45)	−0.194 (−1.29)	—	0.313 (2.04)	0.442 (2.67)
Age Indicator (More Than 45)	−0.331 (−2.02)	—	0.252 (1.39)	0.229 (1.24)
Employment Indicator (Employed)	0.146 (1.09)	—	−0.052 (−0.38)	−0.277 (−1.95)
Employment Indicator (Looking for employment)	0.343 (2.05)		−0.07 (−0.44)	−0.217 (−1.34)
Household Vehicle Indicator (Motorcycle)	—	—	0.099 (0.87)	—
Household Vehicle Indicator (Private car)	0.094 (0.68)	−0.07 (−0.56)	0.148 (1.11)	—

experience either first-hand (i.e., experienced by the respondents) or second-hand (experienced by their friends and family) make respondents more likely to agree with this statement. This might indicate that victims of harassment consciously or subconsciously associate harassment with the overcrowding of public vehicles. Respondents from Dhaka and Rajshahi are more likely to agree with the statement than people from Mymensing. Dhaka and Rajshahi's overcrowded public transportation system could be the reason behind more robust agreement towards this statement from these two cities. Interestingly, people from the middle age group less strongly agree with this statement than those from the younger and older cohort. This is non-intuitive since this age group travels ravenously for various purposes. It might happen that frequent travel has made the middle age group less sensitive to the crowded condition of public vehicles.

3.3.3 Statement: The condition of the seat is very poor

The (dis)agreement to this statement by various societal cohorts is very similar to the previous statement (overcrowding). One significant difference is that respondents from Dhaka agree with this statement less likely than those from the other two cities. Though this observation is only marginally significant, this might indicate that the condition of the public vehicles in the capital city is comparatively better than that of the other two cities.

3.3.4 Statement: Space inside public vehicles is not sufficient

The gender difference is very significant in this statement. Like the previous statements, earlier harassment experiences (either personally or with family and friends) make people agree with this statement more strongly. This might indicate that people with a harassment history associate space insufficiency inside public transportation with harassment incidents. Respondents from Dhaka and Rajshahi are less likely to agree with this statement than Mymensingh. This might indicate that spaces inside the public transportation in the bigger cities are comparatively more conducive than those in Mymensingh. People from households owning private vehicles agree with the statement more strongly than those not having household vehicles. Having alternative transportation options might make people more sensitive to public transportation conditions, resulting in a more positive coefficient.

3.4 Desirable safety precautions for traveling *via* public transportation

Table 5 shows the outcomes of individual safety measures adopted during travel. Four statements were used to capture the desirable safety measure. The results of the four ordered probit models corresponding to the four statements are discussed below.

3.4.1 Statement: Try to avoid rush hours while traveling in public transport

Compared to men, women are more likely to contemplate avoiding rush hour. This finding is like previous research. For example, in a study conducted in the United Kingdom, girls and young women reported preferring empty vehicles to prevent harassment (Finsgate S +44300 777 9777, 2018). People who experienced harassment have slightly higher chances of avoiding rush hours than those who did not. Respondents from Rajshahi are less likely to agree with the statement—comparatively less congestion and low bus frequency in Rajshahi could have contributed to this perception. Employed individuals are more likely to agree with this statement than others. Employed individuals are more likely to travel during rush hours—which might have persuaded them to be extra sensitive towards the inconvenience of rush hour travel. People in the middle age group are less likely to agree—this age group usually takes care of various family, and work responsibilities and hence must travel frequently. All-consuming travel needs might have made this cohort less sensitive toward rush hours.

3.4.2 Statement: Try to avoid traveling at night in public transport

Compared to males, females more strongly agree with the need to avoid nighttime travel to ensure safety. This is supported by previous research—according to World Economic Forum Annual Meeting in 2020, women are more likely than men to opt for costly transportation alternatives (e.g., taxis, privately owned vehicles) at night to ensure safety (There are differences in how, 2022). People who encounter harassment among family and friends are more likely to avoid night travel than others. Compared to Mymensingh, respondents from Dhaka and Rajshahi are less likely to agree with this statement. Mymensingh is a suburban town that gets quiet earlier than the two major cities—Dhaka and Rajshahi. Avoiding night travel might not be feasible for respondents from the two major cities.

3.4.3 Statement: Try to avoid walking alone

Females, unlike men, are more apprehensive about safety while traveling—hence, they avoid walking alone to reduce risk. Individuals who have been harassed prefer not to go alone, indicating they associate walking alone with getting harassed. This finding has been corroborated by earlier research—according to the research by Logan and Walker in the USA; women are more likely than men to avoid traveling alone (Logan and Walker, 2021). Residents of Dhaka and Rajshahi disagree with the statement more than those from Mymensingh. Walking alone might deem riskier in Mymensingh because of the low population density. Older adults are more likely to agree with the statement than younger adults. It might be easier for older adults to find a company than younger adults, making it feasible to practice this safety measure.

3.4.4 Statement: Try to sit close to the driver

Females agree with this statement more than males. People who got harassed previously agree with this statement more than those who did not. Dhaka's Respondents agree more with this statement than those from Rajshahi and Mymensingh. People of the middle to high age group (more than 26 years old) agree with this statement more than those of the younger age group. Interestingly employed individuals agree with this statement less than those who are unemployed. Frequent need for travel might make it impractical for the employed individuals to find a seat near drivers, which might have caused the insignificant estimate of this parameter.

4 Conclusion

Harassments on public transportation have become common, particularly in underdeveloped countries. However, such incidents appear infrequent due to insufficient reporting. Furthermore, the very little research conducted in this realm in the context of developing nations does not investigate the public perception of the situation—which is an essential precursor for developing successful mitigation strategies. The current study was designed to understand the perspective of different demographic cohorts on various issues related to the use of public transportation in the context of Bangladesh—a densely populated, developing country in South Asia. The study starts by depicting the harassment scenario on public transportation in three major cities of the country—including the identification of demographic differences in the occurrence of harassment incidents and the spatial and temporal variation in the incidents across the three cities. Next, the study explores the perception of different demographic groups on women's (the most vulnerable group, according to previous literature) mobility, experience of using public transportation, and desirable safety precautions to avoid harassment.

The study's primary contribution lies in identifying the gender and geographic differences in the perceptions towards mobility challenges faced inside public transportation and the desirable safety precautions. According to current (and previous) studies, women are the predominant victims of harassment in public transport. Therefore, by changing the established attitudes and mentalities towards women, it will be possible to mitigate the problem largely. Moral principles of the younger generations must be strengthened at home and school through educational curricula such as gender discrimination awareness, gender equality, and female empowerment. While analyzing the barriers to women's mobility, the study found that men are not as likely as women to agree about the challenges women face in transportation (for example, gender discrimination, unequal social status in terms of income and employment, and fear of victimization). Therefore, special care should be taken to educate the boy child.

It is also interesting to note that no significant gender difference was noted in some of the riding experience indicators—such as crowding situations and seating conditions of public transportation. However, in terms of misconduct of the transport operators, women are much more likely to agree to the indicator than men. Therefore, special care should be taken to educate the transportation operators, such as ticket takers, drivers, and other support personnel. Mandatory training can educate transportation operators who frequently interact with passengers since most do not receive formal education. According to the study, this might be more important for the smaller towns such as Mymensingh—since people from bigger cities such as Dhaka and Rajshahi are less likely to experience such misconduct. Newspapers, television, and social media might play a role in this realm.

The notable gender difference was found in terms of the desirable safety precautions. Women are significantly more likely to avoid night travel and traveling alone than men to ensure safety. This could seriously limit women's mobility and discourage women's empowerment. Particular attention should be given for ensuring women's safety at night. City and transport planners should think of launching female-only services to ensure safe commutes for women who need to travel alone. Exceptional security could also be provided in selected services at night to safeguard lone female travelers. According to the study, small towns such as Mymensigh are much more in need of such services than the bigger cities such as Dhaka and Rajshahi.

In general, since the propensity of females traveling alone is increasing rapidly because of women's empowerment, city and transport planners should reserve seats for women in public transportation. This is not a long-term solution, but it will assist in lessening harassment for the time being. A follow-up survey could be conducted to determine the desired number of seats to be reserved. In the beginning, women could be allocated at least one-fourth of all seats available in a public vehicle.

Peak-hour traffic is generally crowded, and our research found that the rate of harassment is exceptionally high at these times. Due to the more significant number of female passengers than the number of reserved seats during periods of heavy traffic, females are compelled to stand while traveling on public buses and consequently become exposed to harassment. Therefore, more female-friendly transit facilities need to be supplied during peak hours.

Unfortunately, despite the several rules and regulations, the harassment situation on public transportation in Bangladesh is exacerbating continuously. Therefore, adequate enforcement of the regulations and proper punishment of the harassers must be ensured. On the other hand, victims of harassment must not feel ashamed and raise their voices to draw the attention of the concerned authorities.

Finally, city and transportation planners need to recognize the impact of harassment on the utilization of public transit—a predominant mode of transport in developing countries like Bangladesh. A proper understanding of the challenges would help transit planners build efficient transit systems and foster a safety culture across the country. The recommendations outlined in this article could be helpful in other countries with similar socio-economic backgrounds facing similar challenges in public transportation.

Due to the COVID-19 situation, the survey was conducted online, which might have limited the participation of the urban poor due to a lack of internet access. Future studies should aim to collect a portion of the survey data in person to capture the perception of the underprivileged population of Dhaka, Rajshahi, and Mymensingh.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Study concept and design: SY, TA, AE; Data collection: SY, TA; analysis and interpretation of results: SY, TA, AE; draft manuscript preparation: SY, TA, AE.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. 3rd ed. Hoboken, N.J: Wiley-Interscience, 721. (Wiley series in probability and statistics).

Awan, S. Z. (2020). Stud. Public Transp. Lahore An Analysis Harass. *Women Passengers* 28 (1), 30.

Bray, D., and Holyoak, N. (2015). "Motorcycles in developing asian cities. A Case Study of Hanoi," in *The 37th Australasian Transport Research Forum (ATRF)* (Sydney, NSW, Australia: A Case Study of Hanoi).

Burgess, A., and Horii, M. (2012). Constructing sexual Risk:'Chikan', collapsing male authority and the emergence of women-only train carriages in Japan. *Health Risk Soc - Health RISK SOC*, 14.

Chui, W., and Ong, R. (2008). Indecent assault on the public transport in Hong Kong. *Int. J. Law Crime. Justice* 36, 2–14. doi:10.1016/j.ijsl.2007.06.001

Closing the gaps for gender equality: The role of the Bangladesh Police in fighting Gender Based Violence [Internet].The Daily Star. 2016 [cited 2022 Jun 16]. Available from: https://www.thedailystar.net/supplements/unfpasupplement/closing-the-gaps-gender-equality-the-role-the-bangladesh-police.

Exclusive-, P. O. L. L. (2014). *Latin American cities have the most dangerous transport for women, NYC best*. Reuters [Internet]https://www.reuters.com/article/women-poll-idUKL6N0S32MQ20141029.[cited 2022 Jun 17]; Available from:

Ferdous, N., and Dipu, T. A study of Women Harassment in public transports in Bangladesh.*Int. J. Sci. Bus*, 2019;3:215–222.

Finsgate, S. +44300 777 9777 PIU, London 5-7 Cranwood Street, Uk E 9lh. *It is not OK* [Internet]. Plan International UK. 2018 [cited 2022 Jun 17]. Available from: https://plan-uk.org/street-harassment/its-not-ok.

Foundation, T. R. *Thomson Reuters Foundation* [Internet]. [cited 2022 Jun 17]. Available from:http://www.trust.org/home/index.html.

Gautam, N., Sapakota, N., Shrestha, S., and Regmi, D. (2019). <p>Sexual harassment in public transportation among female student in Kathmandu valley</p>. *Risk Manag. Healthc. Policy* 12, 105–113. doi:10.2147/rmhp.s196230

Gordon, M. T., and Roger, S. (1989). *The female fear*. 1st edition. New York : London: Free Press, 230.

Hossein, A. Measuring gender-based violence: *Results of the violence against women (VAW) survey in Bangladesh*, 35.2015

Hsu, H. P. How does fear of sexual harassment on transit affect women's use of transit? Proceedings of the Transportation Research Board ConferenceProceedings [Internet]. 2011 [cited 2022 Jun 17]. Available from:https://trid.trb.org/view/1101830.

Human haulers a threat in Bangladesh capital [Internet]. *New Age | The Most Popular Outspoken English Daily in Bangladesh*. [cited 2022 Jun 17]. Available from: https://www.newagebd.net/article/46513/human-haulers-a-threat-in-bangladesh-capital/article/index.php.

Khairuzzaman, Z. A. M. *Women's safety in public transport: A case for the would-be city father* [Internet]. The Daily Star. 2019 [cited 2022 Jun 17].Available from: https://www.thedailystar.net/opinion/society/news/womens-safety-public-transport-case-the-would-be-city-father-1703476.

Kim, T., Paek, S., and Lee, J. (2020). An examination of subway sex offense modus operandi: A case of seoul, South Korea. *Int. J. Environ. Res. Public Health* 17, 8979. doi:10.3390/ijerph17238979

Kirchhoff, G., Morosawa, H., and Barkhuizen, J. The asian passengers' safety study of sexual molestation on trains and buses: The Indonesian pilot study. 2007; 20:1.

Logan, T. K., and Walker, R. (2021). The gender safety gap: Examining the impact of victimization history, perceived risk, and personal control. *J. Interpers. Violence* 36 (1–2), 603–631. doi:10.1177/0886260517729405

Mazumder, H., and Pokharel, B. (2019). Sexual violence on public transportation: A threat to women's mobility in Bangladesh. *J. Aggress. Maltreat. Trauma* 28 (8), 1017–1019. doi:10.1080/10926771.2018.1491487

Mukerjee, N. (2019). Investigating women's safety in New Delhi's urban transport systems. *Int. J. Soc. Sci. Econ. Res.* 4 (05), 3876–3886.

Need to make public transport in Delhi women-friendly: Study [Internet]. Hindustan Times. 2017 [cited 2022 Jun 17]. Available from: https://www.hindustantimes.com/delhi-news/need-to-make-public-transport-in-delhi-women-friendly-study/story-Eq8h997zRiq8XTdIr7dQ0H.html.

Neupane, G., and Chesney-Lind, M. (2014). Violence against women on public transport in Nepal: Sexual harassment and the spatial expression of male privilege. *Int. J. Comp. Appl. Crim. Justice* 38 (1), 23–38. doi:10.1080/01924036.2013.794556

Quinones, L. M. (2020). Sexual harassment in public transport in Bogotá. *Transp. Res. Part A Policy Pract.* 139, 54–69. doi:10.1016/j.tra.2020.06.018

Rahman, M. (2010). Bus Serv. "Women Only" *Dhaka City An Investigation*3, 17–32.

Rapid assessment of sexual harassment in public transport and connected spaces in Karachi - Final Report [Internet]. Asian Development Bank. 2014 [cited 2022 Jun 17]. Available from: https://www.adb.org/projects/documents/rapid-assessment-sexual-harassment-public-transport-and-connected-spaces-karachi.

Report, S. O. *Dhaka 7th worst megacity for women: Poll* [Internet]. The Daily Star. 2017 [cited 2022 Jun 17]. Available from: https://www.thedailystar.net/city/dhaka-7th-worst-megacity-women-thomson-reuters-foundation-poll-london-cairo-sexual-violence-culture-healthcare-finance-1477684.

Shakibaei, S., and Vorobjovas-Pinta, O. (2022). Access to urban leisure: Investigating mobility justice for transgender and gender diverse people on public transport. *Leis. Sci.*,6 1–19. doi:10.1080/01490400.2021.2023372

Strategic_Framework.pdf [Internet]. [cited 2022 Jun 17]. Available from: http://www.jagori.org/wp-content/uploads/2006/01/Strategic_Framework.pdf.

There are differences in how men and women see mobility - this is why that is important [Internet]. *World economic Forum.* [cited 2022 Jun 17]. Available from: https://www.weforum.org/agenda/2020/01/mobility-in-2020-a-female-perspective/.

Tokey, A. I., and Shioma, S. A. (2017). *Women safety in public transports.*

Tripathi, K., Barron, H., and Belur, J. (2017). Sexual harassment of students on public transport: An exploratory study in Lucknow, India. *Crime. Prev. Community Saf.* 19, 240–250. doi:10.1057/s41300-017-0029-0

Valan, M. L. (2020). Victimology of sexual harassment on public transportation: Evidence from India. *J. Vict. Victim. Justice* 3 (1), 24–37. doi:10.1177/2516606920927303

frontiers | Frontiers in Built Environment

TYPE Perspective
PUBLISHED 06 January 2023
DOI 10.3389/fbuil.2022.942651

Check for updates

OPEN ACCESS

EDITED BY
Yifei Sun,
Hohai University, China

REVIEWED BY
Serdar Dindar,
Izmir Kâtip Çelebi University, Türkiye
Sakdirat Kaewunruen,
University of Birmingham, United Kingdom

*CORRESPONDENCE
Utpal Deka,
p20190905@dubai.bits-pilani.ac.in

SPECIALTY SECTION
This article was submitted to Transportation and Transit Systems, a section of the journal Frontiers in Built Environment

RECEIVED 12 May 2022
ACCEPTED 21 December 2022
PUBLISHED 06 January 2023

CITATION
Deka U, Varshini V and Dilip DM (2023), The journey of demand responsive transportation: Towards sustainable services.
Front. Built Environ. 8:942651.
doi: 10.3389/fbuil.2022.942651

The journey of demand responsive transportation: Towards sustainable services

Utpal Deka*, Vetha Varshini and Deepthi Mary Dilip

Department of Civil Engineering, BITS Pilani Dubai Campus, Dubai, United Arab Emirates

The concept of Demand Responsive Transport (DRT) describes a technology-enabled shared mobility service that has a flexible operating schedule and/or provides virtual stops and/or flexible routes. While these on-demand services are not new, the developments in communication and tracking technology (vehicle positioning systems) have revived these services in the past decade. The benefits of adopting demand responsive transport services for intra-community and FLM travel needs are widely accepted, but there is still a cautious approach towards their implementation due to the failure of many promising demand responsive transport schemes in the past. This article 1) creates an overview of the various on-demand services introduced across the world, to understand the factors that may have contributed to the failure of these services in the past 2) identifies the progress made towards sustainable demand responsive transport ventures through analysis of global case studies 3) provides an overview of the flexibility of vehicle and deployment technologies in the demand responsive transport sphere. A bibliometric analysis, where the top keywords were further categorised using VOSviewer's default clustering algorithm, highlighted the importance of sustainability in demand responsive transport ventures. By the progress made towards sustainable demand responsive transport ventures, it can be concluded that environmentally sustainable demand responsive transport ventures can be achieved through the adoption of electric and autonomous vehicles for demand responsive transport services, by reducing mileage of the vehicle and/or adjusting the length of route. The study concludes by reviewing existing research gaps regarding performance expectation, and recommending policy and practice implications, based on the case study of the Bus-on-Demand in Dubai, UAE.

KEYWORDS

demand responsive transport (DRT), sustainability, public private partnership (PPP), public transportation, integrated public transport

1 Introduction

The world is experiencing rapid urbanisation today. Over 66% of the world's population will reside in cities by 2050, creating 80% of global wealth, and consuming 60% of the world's energy (Gerardo 2018). As urban mobility is the lifeline of modern cities, one of the most pressing issues for cities across the globe is the planning and development of smart cities that provide an effective, equitable and sustainable mechanism for moving people. Over the past few decades, both public transportation (including buses and railway/metro) and personalised vehicles have facilitated urban mobility. Although conventional public transportation is more affordable and environmentally sustainable than personal vehicles, the latter trumps mass transportation when it comes to comfort and convenience. To this end, governments across the world have tried to improve accessibility, service quality, frequency, network expansion, financial incentives, and several other methods to attract citizens towards public transportation. However, only a handful of public transportation systems globally are profitable (Canales et al., 2017). The rest are only making enough to keep the system going or more often sustain losses and still function. With urban areas becoming increasingly denser, the provision of mass transit to urban dwellers has become an expensive proposition to the city authorities. Moreover, with the growing earning/car ownership levels and the ensuing dispersal of activity centres and ride styles (Khattak & Yim, 2004), conventional public transportation (including buses and railway/metro) which operate with fixed stops, routes, and schedules, can no longer meet the travelling needs of large sections of society.

Innovations in mobility, such as ride-hailing (Feigon and Murphy, 2016; Erhardt et al., 2019) have paved the way for the concept of shared mobility, which combines the advantages of public transport and personal transport (Marković et al., 2016). Shared mobility can be achieved either through vehicle-sharing (car sharing or peer to peer car sharing), ride-sharing (taxi sharing or carpooling) (Balcombe et al., 2004), or flexible-transit (micro-transit or Bus-on-Demand) (Feigon and Murphy, 2016). The use of mobility-on-demand services has been shown to improve transportation efficiency, user satisfaction, and the environment (Greenblatt and Shaheen, 2015; Beiker, 2016; Djavadian and Chow, 2017). With recent developments in communication and sensor-based ICT technology, as well as the necessity to integrate shared mobility with other public transportation modes (inter-modal journeys), demand responsive transportation (DRT) systems have been developed (Marković et al., 2016). These systems are characterised by flexible routes and small vehicles that operate in a shared-ride mode along routes consistent with passenger needs (Ambrosino et al., 2003).

A well-designed DRT system strikes the right balance between the dependability of conventional public transportation and the flexibility of private vehicles (Kamargianni et al., 2016), and at prices far below that of taxis (Rodier et al., 1998). Therefore, demand responsive public transportation is becoming the preferred solution for today's fast growing cities. It complements the city's mass transport networks and provides new travel options that reduce congestion air pollution, while also increasing ridership and customer happiness (Bürstlein et al., 2021).

While the benefits of adopting DRT services for intra-community and FLM (first-mile last-mile) travel requirements have been well documented (Shu et al., 2021), there still remains a cautiousness towards the implementation of DRT services, with many examples of promising DRT schemes that have failed (Pettersson, 2019). The steep fares for DRT services in comparison with fixed-route public transport was cited as the main reason for the failure of these systems (Enoch et al., 2004). Currie and Fournier (2020), reported the failure of Bridj which shut down its U.S. operations in 2017 (Marshall, 2017; Schmitt, 2018) and Chariot (Marshall, 2019), and attributed the "success" of the DRT ventures to the "media hype," and the lack of documentation relating DRT failures to the high costs. In spite of this, the fact that the pilot projects have been deployed in over 900 cities globally over the last four to 5 years (Foljanty, 2020) is a testimony that DRT services may well be able to bridge the gap between an efficient public transportation system and convenient individual mobility.

In this paper, we examine the various on-demand services introduced across the globe to learn from past initiatives and understand the major roadblocks that led to their failure. While the resurgence of interest in these services can be seen from the publication trend over the years, a bibliometric analysis of the keywords indicated a keen interest by researchers on sustainability aspects. Therefore, an assessment of the progress made towards sustainable DRT ventures is carried out, focusing specifically on the DRT ventures in the past decade. Through the lens of sustainability, DRT services can be beneficial for internal community transport as well as first-mile last-mile (FLM) requirements.

2 Towards sustainable DRT services

DRT was first experimented with in Atlantic City in 1916 when jitneys were used to respond to public requests for rides (O'Leary, 1982; Strobel, 1987). However, as these services provided stiff competition to the existing trolley systems, they were throttled by the conventional transit authorities and regulated to operate on fixed routes, thereby reducing their relevance by the 1920s (Eckert and Hilton, 1972). The DRT's made a come-back much later in the 1960s, with the spread of the low-density areas in the United States (Cole, 1968). The CARS project of the Massachusetts Institute of Technology, aimed at the development of a "many-to many" origins and destinations algorithm to efficiently assign demand, reducing the workforce necessary to provide the offer, making a door-to-door transit system affordable (Wilson et al., 1969).

The high operating costs associated with DRT prior to 1980 (Wilson and Simpson, 1975), the lower speeds (Murphy et al., 1975) and the higher ticket prices (Enoch et al., 2004) led to the failure of many of these ventures. While the 1980s and 90s saw the improvement of different technologies that enhanced communication and data collection, most paratransit services were characterised as either low-tech or high-tech. As internet technology developed in the early 2000s (Lasdon et al., 2000), this concept did not meet much acceptance as a replacement for regular transportation, and it was deemed inefficient (Davison et al., 2012; Mulley et al., 2012), and not economically viable (Davison et al., 2014).

The DRT services have evolved over the decades, from the 'traditional' paratransit service, paving the way for the demand-responsive modes today, being launched in urban/metropolitan areas for the general public (Pettersson, 2019). The development of communication technologies (smartphones) and tracking technologies (global positioning systems) has greatly increased the popularity of these services in recent years. From a review of Scopus-indexed publications (Pirola et al., 2020), 169 documents addressing demand responsive transportation have been identified since the first papers on the topic in 1973, following the elimination of irrelevant documents. Only 22 publications were published between 1973 and 2003, while 24 publications were published in 2020 itself. This trend has been reflected on the ground, with on-demand transportation growing 3 times globally from 2009 to 2017 (Goldman Sachs Annual Report, 2017) and many of the world's cities either exploring or in the initial stage of running of pilot DRT projects, to understand the possibility of using a suitable on-demand technology that can complement or replace traditional public transport services (Barrett et al., 2019).

The high failure rate of previous DRT services (before 2000) underscores the importance of learning from the past deployments to ensure the success of the DRT services. Based on the approach suggested by Petticrews' Practical Guide to Systematic Reviews in the Social Sciences (2006), a systematic review of the DRT development over the years have been carried out.

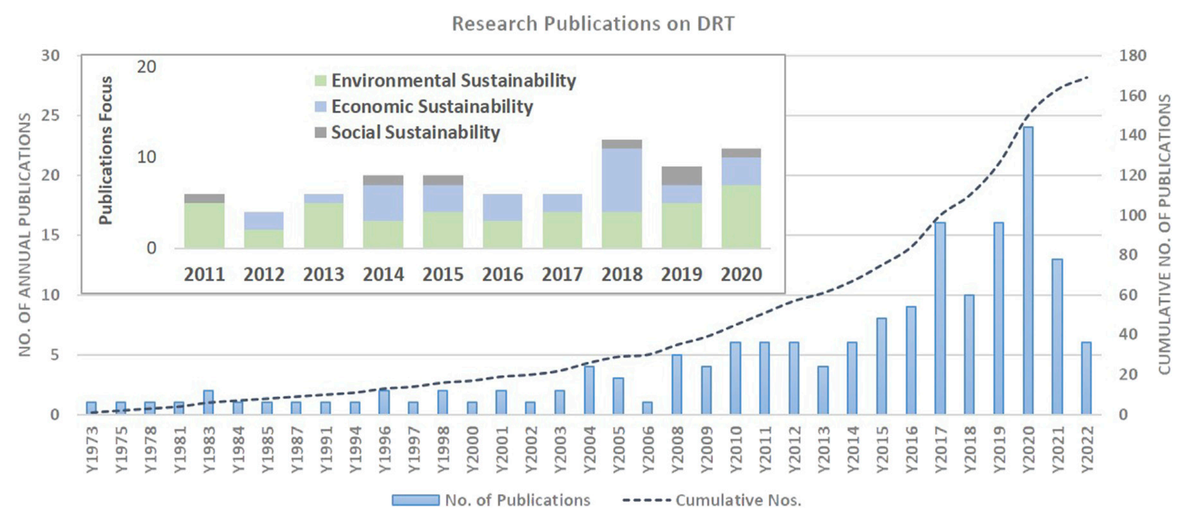


FIGURE 1
DRT research publications trend (as of April 2022).

DRT systems contribute to social sustainability by merging the advantages of public transportation with the conveniences of private vehicles (Brake et al., 2004). According to a review of journal publications (Scopus) over the last decade, there has been an increasing interest in sustainable DRT ventures, as illustrated in Figure 1 above. The focus during 2011 was on environmental sustainability. However, the research focus was also equally shifted to economic sustainability from 2012 onwards and during 2018, majority of the sustainability related publications were focused on economic aspects of DRT development.

The research focus on economic sustainability could potentially be attributed to the practical implementation of DRT pilot programs all across the world cities indicating importance of financial sustainability of such deployments.

2.1 Commercial sustainability

In the period between 2000 and 2010, nearly all the wider proposals for DRT adoption were premised on receiving public monetary support (Mulley and Nelson, 2009). Researchers (Tsubouchi et al., 2010; Davison et al., 2014) have suggested that getting aid from local authorities or even the government to start-up the project would aid in resolving the first big obstacle any DRT project would face. Stating that DRT projects may fail, especially when they are not realistically priced or designed with a full understanding of the market they are to serve, Gomes et al. (2015) proposed an approach integrating simulation and optimization to balance the envisaged costs and service efficiency.

As developing a DRT system with adequate software and labour does not come cheap, and researchers (Furuhata et al., 2014; Ryley et al., 2014) claim that it is very difficult to earn back the investments made and to make the system financially stable. Another main challenge faced is how the operating costs and profits are to be shared fairly among the passengers.

As driver costs and scheduling constraints are the main reasons why introduction of large-scale DRT may not be feasible (Bösch et al., 2018; Lioris et al., 2018; Winter et al., 2018), DRT services with autonomous vehicles (AVs) were proposed as a solution as they are claimed to be economical, as they minimise or absolutely deduct the cost of labour for transportation.

Experiences across the world have indicated that a DRT system can be long lasting only if it is economical for the passenger when compared to a private or other public transports. A study by Ryley et al. (2014) explained that changing the competitor makes it easier for DRT to succeed. Competing against railway or airline modes of transportation will make it easy for a DRT system to be preferred over the former.

3 Case study: Bus on demand in Dubai, UAE

Since February 2020, Dubai Bus on Demand (DBOD) has been operating in the United Arab Emirates (UAE) as a technology-enabled shared bus service. A third-party service provider operates the services under a 3-year Public Private Partnerships (PPP) Contract. For a service fee based on service kilometres, the service provider brings the standard-sized (12–14 seat capacity wheelchair accessible vehicles), the drivers, the technology, and the marketing of the services. Figure 2 illustrates a month by month comparison of the capacity utilisation i.e., riders per vehicle per hour in 2020 and 2021 as well as the average hourly cost recovery ratio (CRR) of the DRT operation. During 2021, the average monthly utilisation level increased by 23.5% from 2020. During August 2020 (peak COVID-19), the capacity utilisation had reached its peak, requiring the deployment of additional vehicles. The system's hourly utilisation reaches 8+ making the Dubai BOD one of the most efficient micro-transit deployments of its kind. These numbers are comparable with the best cities that has provision of FLM & intra-community DRT services with similar capacity vehicles internationally that varies between 1.69–5.5 riders/vehicle/hour.

The data set used for case study analysis is collected and re-organised with approval from Roads and Transport Authority, Government of Dubai which are based on actual operational data of DBOD over three consecutive months during 2021 that includes week days and week ends in one of the operational area (Al Barsha).

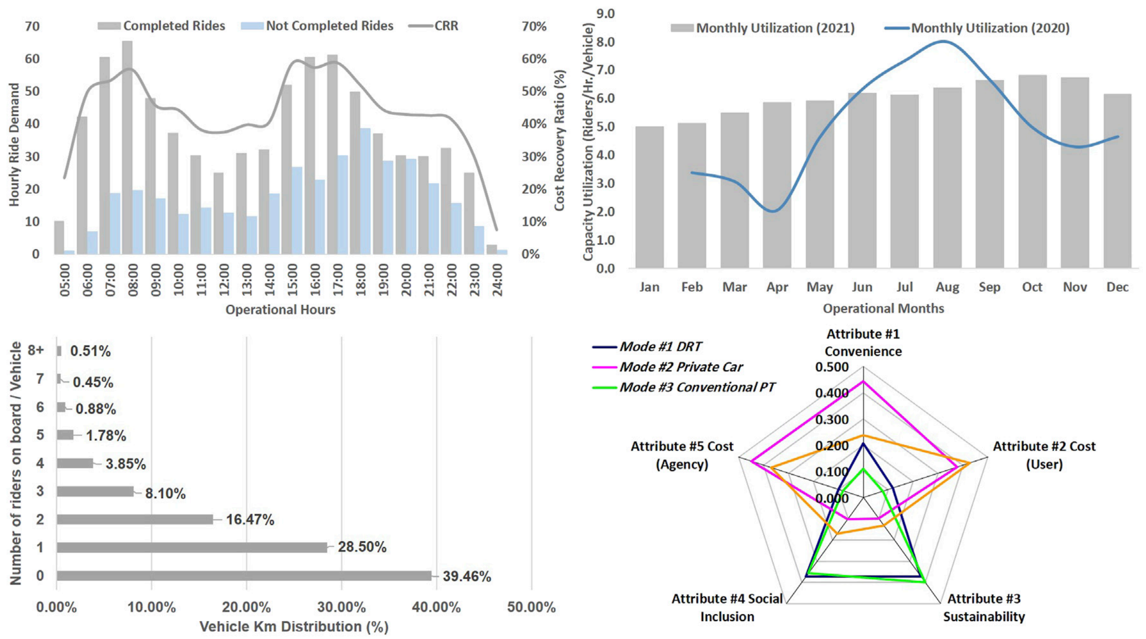


FIGURE 2
Operational parameters of DBOD (Source: DBOD operational data).

3.1 Key factors for DRT commercial sustainability

From the first author's experience, the authors' recommend the following factors that need to be considered for sustainable DRT ventures:

3.1.1 DRT mode choice

From a basic Analytic Hierarchy Process (AHP) model (see Figure 2), based on the results of stated preference survey of users of DBOD, sustainability and social inclusion are seen to be the two main attributes to be considered for DRT. Despite the launch of these on-demand services in many cities, there are still many misgivings regarding the commercial sustainability of these ventures.

3.1.2 Integrated planning for cost recovery

Cost Recovery Ratio (CRR), which is the percentage of operating costs covered by fares, often depends on the route (feeder, urban, and intercity). Normally the intercity routes with more frequent service and faster journey times tend to carry more passengers with higher fare level thereby having a higher return on investment, while other routes (feeder or urban) servicing first mile/last mile or serving the outskirts of the city or newly developing communities often carry fewer passengers and therefore have lesser CRR. Lesser cost recovery inevitably leads to significantly higher public sector budgetary subsidies. Moreover, longer distances covered during off-peak hours without optimal passenger on-board can significantly impact the mileage and reduce cost recovery.

From the experience of BDOD, while planning for the vehicle deployment during the operational hours, there is significant additional capacity required during the morning and evening peak hours in order to meet the key performance criteria (ETA, pick-up walk distance) of the service deployment. As the number of riders increases during the peak hours, Vehicle-Kilometres travelled with riders on-board also proportionately increases. Another important aspect that can be observed from Figure 2 is the distribution of Vehicle-Kilometres with no riders on board even during the peak hours of operation. This is due to the nature of the on-demand services, but it is also apparent that aggregation of riders plays a crucial role in determining the commercial rigour of the services. The CRR of the service operation is the single most indicator for commercial sustainability of DRT, and the DBOD operations highlights the need for optimal fleet deployment for higher cost recovery.

A balance between the fare level that can be levied on the riders versus the cost of operation covering both revenue and non-revenue kilometres need careful consideration by keeping in mind the public transport users economic profile.

From Dubai's Bus on Demand experience (refer Figure 2), approximately 40% of the vehicle mileage consists of vehicle travel without any passenger on board and the balance 60% distance travelled with passengers on board, 28.5% boarded km distribution consists of one rider, 16% has two riders, and 4% has three riders on-board, indicating the need for additional aggregation of riders for achieving the desired commercial rigour with optimization of service planning. Accordingly, a successful DRT venture should be flexible enough so as to minimise the mileage with no passengers on board, thus minimising the non-revenue mileage (Bellini et al., 2003).

3.1.3 Service efficiency

A key lesson from the pilot run of DBOD was that the larger vehicles (Mercedes Sprinters) had difficulty making U-turns. By introducing smaller vehicles (Toyota Hiaces), both Expected Time of Arrival (ETA) and mileage have improved. The operational data averaged over a month for both large vehicles (Mercedes Sprinter) and small vehicles (Toyota Hiace) indicates a reduction of 22.5% in case of pickup ETA time with an average reduction of 14% travel duration for each passenger. Accordingly, accommodating smaller vehicle sizes skews the scale in favour of optimal vehicle utilisation compared to traditional public transport buses. This result is in accordance with the guidelines suggested by Wright (2013), who concluded that the product of the demand and the average trip distance provides a very strong indicator of the suitable vehicle type.

4 Discussion and recommendations

Despite the potential barriers, the popularity of DRT services has been increasing exponentially since 2017, with the highest number of on-demand shared mobility pilot services till date, launched in Europe, followed by North America (Foljanty, 2020). One of the likely reasons behind these initiatives were due to the early experiences of long-existing DRT services in various countries in Europe, e.g., Germany, France or the United Kingdom. During 2019 and 2020, there have been multiple deployments all across Asian cities like Singapore, Japan, India, Indonesia, Korea, Israel, and UAE, closely following the market dynamics of Europe and North America. Thus, there are many DRT operators and technology platform providers working in silos in many parts of the world. However, a comprehensive approach for the commercial sustainability of a citywide DRT System, that addresses the urban mobility requirements of a city in an integrated manner with a multi-modal approach, is lacking.

5 Conclusion

The main purpose of Demand Responsive Transport is to address the citizens' travel demands at the most basic level by creating an integrated platform for journey options and modal choices. While the system can run independently, taking over the traditional modes, there is also a possibility of integrating DRT with the traditional public transit modes in the city. FLM solutions including DRT foster end-to-end "door-to-door" mobility, enabled by new technology, which is critical to provide alternatives to individual private transport. Making the services customer friendly with flexible timings play a factor in a successful DRT service though there is significant additional capacity required during the peak hours necessitating extra capacity that can meet the key performance criteria and customer satisfaction. However, the cost recovery ratio of the service operation is the single most indicator for commercial sustainability determination of a DRT system. A balance between the passenger fare and the cost of operation considering both capital investment and operational expenses (vehicle type, fuel type, seating capacity) covering both revenue and non-revenue kilometres requires careful policy considerations by keeping in mind the public transport users economic profile.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

UD—conceived the study, drafted manuscript, visualization of key word matrix and conducted assessment of case study. VV and DD—were involved in data collection, provided key information and revised the manuscript.

Acknowledgments

The authors acknowledge the Roads and Transport Authority, Government of Dubai for providing necessary permission to access the operational data of the Dubai Bus on Demand for carrying out the case study analysis of this academic research project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbuil.2022.942651/full#supplementary-material

References

Ambrosino, G., Nelson, J., and Romanazzo, M. (2003). *Demand Responsive Transport Services: Towards the Flexible Mobility Agency*. Rome, Italy: Italian National Agency for New Technologies, Energy and the Environment.

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., et al. (2004). *The demand for public transport: A practical guide*. Berlin, Germany: Researchgate.

Barrett, S., Santha, N., and Khanna, A. (2019). *On-demand public transport: Key learnings from global pilots*. New York, NY, USA: LEK.

Beiker, S. (2016). "Implementation of an automated mobility-on-demand system," in *Autonomous driving* (Berlin, Heidelberg: Springer), 277–295.

Bellini, C., Dellepiane, G., and Quaglierini, C. (2003). The demand responsive transport services: Italian approach. *WIT Trans. Built Environ.* 64, 10. doi:10.2495/UT030071

Bösch, P. M., Becker, F., Becker, H., and Axhausen, K. W. (2018). Cost-based analysis of autonomous mobility services. *Transp. Policy* 64, 76–91. doi:10.1016/j.tranpol.2017.09.005

Brake, J., Nelson, J. D., and Wright, S. (2004). Demand responsive transport: Towards the emergence of a new market segment. *J. Transp. Geogr.* 12 (4), 323–337. doi:10.1016/j.jtrangeo.2004.08.011

Bürstlein, J., López, D., and Farooq, B. (2021). Exploring first-mile on-demand transit solutions for North American suburbia: A case study of markham, Canada. *Transp. Res. Part A Policy Pract.* 153, 261–283. doi:10.1016/j.tra.2021.08.018

Canales, D., Bouton, S., Trimble, E., Thayne, J., Da Silva, L., Shastry, S., et al. (2017). *Connected urban growth: Public-private collaborations for transforming urban mobility*. London and Washington DC: Coalition for Urban Transitions c/o World Resources Institute, 39–45.

L. M. Cole (Editor) (1968). *Tomorrow's transportation: New systems for the urban future (vol. 62)* (Washington, D.C: US Government Printing Office).

Currie, G., and Fournier, N. (2020). Why most DRT/Micro-Transits fail–What the survivors tell us about progress. *Res. Transp. Econ.* 83, 100895. doi:10.1016/j.retrec.2020.100895

Davison, L., Enoch, M., Ryley, T., Quddus, M., and Wang, C. (2014). A survey of demand responsive transport in Great Britain. *Transp. Policy* 31, 47–54. doi:10.1016/j.tranpol.2013.11.004

Davison, L., Enoch, M., Ryley, T., Quddus, M., and Wang, C. (2012). Identifying potential market niches for demand responsive transport. *Res. Transp. Bus. Manag.* 3, 50–61. doi:10.1016/j.rtbm.2012.04.007

Djavadian, S., and Chow, J. Y. (2017). An agent-based day-to-day adjustment process for modeling 'Mobility as a Service' with a two-sided flexible transport market. *Transp. Res. part B Methodol.* 104, 36–57. doi:10.1016/j.trb.2017.06.015

Eckert, R. D., and Hilton, G. W. (1972). The jitneys. *J. Law Econ.* 15 (2), 293–325. doi:10.1086/466738

Enoch, M., Potter, S., Parkhurst, G., and Smith, M. (2004). *Inter-mode: Innovations in demand responsive transport, Report for department for transport and greater manchester passenger transport executive, final report*. London: Department for Transport. Final report.

Erhardt, G. D., Roy, S., Cooper, D., Sana, B., Chen, M., and Castiglione, J. (2019). Do transportation network companies decrease or increase congestion? *Sci. Adv.* 5 (5), eaau2670. doi:10.1126/sciadv.aau2670

Feigon, S., and Murphy, C. (2016). *Shared mobility and the transformation of public transit (No. Project J-11, Task 21).* Washington, D.C., United States: National Academies Press.

Foljanty, L. (2020). Mapping the global on-demand ridepooling market. Medium. Available at: https://lukas-foljanty.medium.com/mapping-the-global-on-demand-ridepoolingmarket-f8318de1c030.

Furuhata, M., Daniel, K., Koenig, S., Ordonez, F., Dessouky, M., Brunet, M. E., et al. (2014). Online cost-sharing mechanism design for demand-responsive transport systems. *IEEE Trans. Intelligent Transp. Syst.* 16 (2), 1–16. doi:10.1109/tits.2014.2336212

Gerardo, P. G. (2018). 68% of the world population projected to live in urban areas by 2050, says UN. Available at: https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html.

Goldman Sachs Annual Report (2017). *Goldman Sachs annual report*. New York, United States: Goldmansachs.

Gomes, R., de Sousa, J. P., and Dias, T. G. (2015). Sustainable Demand Responsive Transportation systems in a context of austerity: The case of a Portuguese city. *Res. Transp. Econ.* 51, 94–103. doi:10.1016/j.retrec.2015.07.011

Greenblatt, J. B., and Shaheen, S. (2015). Automated vehicles, on-demand mobility, and environmental impacts. *Curr. sustainable/renewable energy Rep.* 2 (3), 74–81. doi:10.1007/s40518-015-0038-5

Kamargianni, M., Li, W., Matyas, M., and Schäfer, A. (2016). A critical review of new mobility services for urban transport. *Transp. Res. Procedia* 14, 3294–3303. doi:10.1016/j.trpro.2016.05.277

Khattak, A. J., and Yim, Y. (2004). Traveler response to innovative personalized demand-responsive transit in the san francisco bay area. *J. urban Plan. Dev.* 130 (1), 42–55. doi:10.1061/(asce)0733-9488(2004)130:1(42)

Lasdon, L. S., Mahmassani, H. S., and Dutt, S. (2000). *Transit scheduling data integration: Paratransit operations review and analysis (No. 0-1884-1).* Austin, United States: University of Texas at Austin. Center for Transportation Research.

Lioris, J., Hai-Salem, H., Seidowsky, R., and Lebacque, J. P. (2018). "Dynamic strategies optimizing benefits of fully autonomous shared vehicle fleets," in 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 04-07 November 2018 (IEEE), 2771–2776.

Marković, N., Kim, M. E., and Schonfeld, P. (2016). Statistical and machine learning approach for planning dial-a-ride systems. *Transp. Res. Part A Policy Pract.* 89, 41–55. doi:10.1016/j.tra.2016.05.006

Marshall, A. (2019). *Ford axes its Chariot shuttles, proves mobility is hard*. San Francisco, California: WIRED.

Marshall, A. (2017). *How a failed experiment could still be the future of public transit*. San Francisco, California: WIRED.

Mulley, C., and Nelson, J. D. (2009). Flexible transport services: A new market opportunity for public transport. *Res. Transp. Econ.* 25 (1), 39–45. doi:10.1016/j.retrec.2009.08.008

Mulley, C., Nelson, J., Teal, R., Wright, S., and Daniels, R. (2012). Barriers to implementing flexible transport services: An international comparison of the experiences in Australia, Europe and USA. *Res. Transp. Bus. Manag.* 3, 3–11. doi:10.1016/j.rtbm.2012.04.001

Murphy, R. E., Paisley, P. L., and Siersema, J. N. (1975). in *Transportation research board special Report* (Washington, D.C: Transportation Research Board). 154.*Research and development in demand-responsive transportation*

O'Leary, K. (1982). *Planning for new and integrated demand-responsive systems*. Washington, D.C: National Academies Press. No. HS-037 987.

Pettersson, F. (2019). *An international review of experiences from on-demand public transport services*. Berlin, Germany: Researchgate. The Swedish Knowledge Centre for Public Transport.

Pirola, F., Boucher, X., Wiesner, S., and Pezzotta, G. (2020). Digital technologies in product-service systems: A literature review and a research agenda. *Comput. Industry* 123, 103301. doi:10.1016/j.compind.2020.103301

Rodier, C. J., Johnston, R. A., and Shabazian, D. R. (1998). Evaluation of advanced transit alternatives using consumer welfare. *Transp. Res. Part C Emerg. Technol.* 6 (1-2), 141–156. doi:10.1016/s0968-090x(98)00013-8

Ryley, T. J., Stanley, P. A., Enoch, M. P., Zanni, A. M., and Quddus, M. A. (2014). Investigating the contribution of Demand Responsive Transport to a sustainable local public transport system. *Res. Transp. Econ.* 48, 364–372. doi:10.1016/j.retrec.2014.09.064

Schmitt, A. (2018). *The story of 'micro Transit'Is consistent*. USA: Dismal Failure.

Shu, P., Sun, Y., Xie, B., Xu, S. X., and Xu, G. (2021). Data-driven shuttle service design for sustainable last mile transportation. *Adv. Eng. Inf.* 49, 101344. doi:10.1016/j.aei.2021.101344

Strobel, H. (1987). "Computer controlled urban transportation: A survey of concepts, methods, and experiences," in *Control in transportation systems 1986* (Oxford: Pergamon Press), 25–45.

Tsubouchi, K., Yamato, H., and Hiekata, K. (2010). Innovative on-demand bus system in Japan. *IET Intell. Transp. Syst.* 4 (4), 270–279. doi:10.1049/iet-its.2009.0113

Wilson, N. H. M., Sussman, J. M., Goodman, L. A., and Higonnet, B. T. (1969). *Simulation of a computer aided routing system (CARS).* New Jersey: IEEE.

Wilson, R. J., and Simpson, A. U. (1975). "Planning for new and integrated demand-responsive systems," in *Transportation research board special Report* (Washington, D.C: Transportation Research Board). 154.

Winter, K., Cats, O., Correia, G., and van Arem, B. (2018). Performance analysis and fleet requirements of automated demand-responsive transport systems as an urban public transport service. *Int. J. Transp. Sci. Technol.* 7 (2), 151–167. doi:10.1016/j.ijtst.2018.04.004

Wright, S. (2013). Designing flexible transport services: Guidelines for choosing the vehicle type. *Transp. Plan. Technol.* 36 (1), 76–92. doi:10.1080/03081060.2012.745757

frontiers | Frontiers in Built Environment

TYPE Original Research
PUBLISHED 16 February 2023
DOI 10.3389/fbuil.2023.1143829

Check for updates

OPEN ACCESS

EDITED BY
Argaw Gurmu,
Deakin University, Australia

REVIEWED BY
Xin Hu,
Deakin University, Australia
Arka Ghosh,
Deakin University, Australia

*CORRESPONDENCE
Ramin Asadi,
raminasadi2013@yahoo.com

SPECIALTY SECTION
This article was submitted to Construction Management, a section of the journal Frontiers in Built Environment

RECEIVED 13 January 2023
ACCEPTED 01 February 2023
PUBLISHED 16 February 2023

CITATION
Asadi R, Rotimi JOB and Wilkinson S (2023), Rework causes classification model with liable parties of the contract in construction projects.
Front. Built Environ. 9:1143829.
doi: 10.3389/fbuil.2023.1143829

Rework causes classification model with liable parties of the contract in construction projects

Ramin Asadi*, James Olabode Bamidele Rotimi and Suzanne Wilkinson

School of Built Environment, Massey University, Auckland, New Zealand

Rework is an interesting topic in the contract management of construction projects. An effective way of improving performance on construction projects is to manage rework. However, managing rework is challenging because of the dynamic nature of construction activities. A rational starting point is the identification of the root causes of rework then a framework for its management can be developed for improving construction performance. This paper reviews rework-related studies in the construction industry through a critical review of literature to investigate the main causes of rework in the construction contracts. A content analysis of the previously proposed classification methods for rework revealed that all rework causes could be grouped under various project stages. The causes of rework also could be ascribed to different project parties. However, previous studies have suggested various categories of rework, there is no commonly used classification model for rework causes in construction contracts. The current study proposes a model in three levels to address this gap. The developed model categorised rework causes in five constructs linked to three main stages of the project under two liable contract parties. The study findings show that the procurement stage has fewer categories of rework causes than design and construction. The result also reveals that the involvement level of contract parties in rework occurrence can be investigated in contract documents. Thus, this paper suggests further research in procurement stage to address rework causes in the contract conditions.

KEYWORDS

rework, construction, contract, defect, classify

1 Introduction

In construction projects, the combination of stakeholders works together to execute a complex and unique product. Rationally, each stakeholder prefers to save more benefits, so motivation to work collaboratively is low (Liu et al., 2020). Such an attitude has steered projects to split the main work into smaller packages to be accomplished by distinct organizations individually. Thus, the stages of design, procurement, and construction of projects have been performed separately for a long period. This breakup of the project's main stages has degraded construction success indicators such as performance, productivity, and competitiveness (Ye et al., 2015). Novel contracting strategies have emerged over the last few years to coordinate the relationship between contractors and clients to achieve better performance. The cost of construction projects is effectively reduced if the main stages of the project life cycle are taken into consideration (Bao et al., 2018) under rework events. Rework is an issue that appears across various stages of the project and has cross-functional

TABLE 1 An overview on the classification methods used in the reviewed papers in this study.

Number of occurence		18	2	10	8	1	2	21	2	5	5	9	7	5	4	1	1	2	1	6	7
[35]	Nigeria	~	~	~	~	~	~	√	~	~	~	~	~	~	~	~	~	~	~	~	~
[34]	Saudi Arabia	~	~	~	~	~	~	~	~	√	~	~	~	~	~	~	~	~	~	~	~
[33]	China	√	~	~	√	~	~	√	√	~	√	~	~	~	√	~	~	~	~	√	~
[32]	USA	~	~	~	~	~	√	~	~	~	√	√	~	~	~	~	~	~	~	~	~
[31]	Singapore	√	~	√	~	√	~	√	~	~	~	~	~	~	√	~	~	~	~	~	~
[30]	Ukrine	√	~	√	~	~	~	√	~	~	~	~	~	~	~	~	~	~	~	~	~
[29]	Nigeria	√	~	√	~	~	~	√	~	~	~	~	~	~	~	~	~	~	~	√	~
[28]	Nigeria	√	√	√	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
[27]	Malaysia	~	~	~	~	~	~	√	~	~	√	√	~	√	~	~	~	~	~	~	~
[26]	Nigeria	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	√
[25]	South Africa	~	~	~	~	~	~	√	~	~	~	~	√	~	~	~	~	√	~	~	~
[24]	Palestine	√	~	√	~	~	~	√	~	√	~	√	~	√	~	~	~	~	~	√	~
[23]	Nigeria	~	~	~	~	~	~	√	~	~	~	~	~	~	~	~	~	~	~	~	~
[22]	Nigeria	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	√
[21]	India	√	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
[20]	India	~	~	√	~	~	~	√	~	√	~	√	~	√	~	~	~	~	~	√	~
[19]	Palestine	√	√	√	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	√	~
[18]	Nigeria	√	~	√	~	~	~	√	~	~	~	~	~	~	~	~	~	~	~	~	~
[17]	Nigeria	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	√
[16]	Iran	√	~	~	√	~	~	√	~	~	~	~	~	~	~	~	~	~	~	~	~
[15]	China	√	~	√	√	~	~	√	√	~	√	~	~	~	√	~	~	~	~	√	~
[14]	Spain	~	~	~	~	~	√	~	~	~	√	√	~	~	~	~	~	~	~	~	~
[13]	Singapore	√	~	~	~	~	~		~	~	~	~	~	~	~	~	~	~	~	~	~
[12]	Hong Kong	√	~	~	√	~	~	√	~	~	~	~	√	~	~	~	~	~	~	~	~
[11]	Nigeria	√	~	√	~	~	~	√	~	~	~	~	~	~	~	~	~	~	~	~	~
[10]	Canada	~	~	~	~	~	~	~	~	~	~	√	~	~	√	~	~	~	~	~	~
[9]	Australai	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	√

(Continued on following page)

TABLE 1 (*Continued*) An overview on the classification methods used in the reviewed papers in this study.

Number of occurence		18	2	10	8	1	2	21	2	5	5	9	7	5	4	1	1	2	1	6	7
[8]	Nigera	~	~	~	~	~	~	~	~	~	~	√	~	~	~	√	~	~	√	~	~
[7]	Australia	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	√
[6]	Hong Kong	√	~	~	√	~	~	√	~	~	~	~	√	~	~	~	~	~	~	~	√
[5]	Hong Kong	√	~	~	√	~	~	√	~	~	~	~	√	~	~	~	~	~	~	~	√
[4]	Hong Kong	~	~	~	√	~	~	√	~	√	~	~	√	~	~	~	~	~	~	~	~
[3]	Canada	~	~	~	~	~	~	√	~	~	~	√	~	√	~	~	√	√	~	~	~
[2]	Australia	√	~	~	√	~	~	√	~	√	~	~	√	~	~	~	~	~	~	~	~
[1]	Sweden	√	~	~	~	~	~	√	~	~	~	√	√	√	~	~	~	~	~	~	~
References	Country Category	Client	Consultant	Contractor	Subcontractor	Vendor/Supplier	Organizational factors	Design/Engineering review	Contract Management	Construction	Project	Workmanship/People/Human Resources	Site Management/Productopn/Operation	Material and equipment/Machine	Process/Change/Error/Ommision	Technical	Leadership and Communications	Planning and Scheduling	Quality Management	External/Environmental aspects	No classification/General
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

[1] Josephson et al. (2002); [2] Love and Smith (2003); [3] Robinson et al. (2004); [4] Palaneeswaran et al. (2005); [5] Palaneeswaran (2006); [6] Palaneeswaran et al. (2008); [7] Love et. (2009); [8] Oyewobi and Ogunsemi (2010); [9] Love et. (2010b); [10] Zhang et al. (2012); [11] Aiyetan (2013); [12] Palaneeswaran et al. (2014); [13] Hwang et al. (2014); [14] Forcada et al. (2014); [15] Ye et al. (2015); [16] Miri and Khaksefidi (2015); [17] Ajayi and Oyeyipo (2015); [18] Aiyetan and Das (2015); [19] Mahamid (2016); [20] Shah et al. (2016); [21] Ahmed and Naik (2016); [22] Oyewobi et al. (2016); [23] Wilson and Odesola (2017); [24] Enshassi et al. (2017); [25] Ndwandwa et al. (2017); [26] Ajayi (2017); [27] Yap et al. (2017); [28] Eze et al. (2018b); [29] Eze et al. (2018a); [30] Trach et al. (2019); [31] Hwang et al. (2019); [32] Safapour et al. (2019); [33] Liu et al. (2020); [34] Mahamid (2020); [35] Salihu and Babarinde (2020).

10.3389/fbuil.2023.1143829

impacts on the construction industry (Zhang et al., 2018). Identifying rework causes at the main stages of a project is required to facilitate construction projects' success and to provide solutions for managing rework. Therefore, implementing a life cycle philosophy for rework management is critical due to the integrated activities of the construction contracts.

In the study area of construction rework, available literature mostly has focused on general trends of rework topics. Identified causes of rework from previous studies have generally been classified in various methods based on research scopes (Love and Smith, 2003; Robinson et al., 2004; Liu et al., 2020). Literature on this topic lacks systematic analysis of rework causes from a perspective that covers various stages of projects and liable contract parties (Taggart et al., 2014). The various categories of rework causes in the literature show that there is no unify used classification method for categorizing causes of rework in construction contracts (Love and Li, 2000; Zhang et al., 2012; Safapour and Kermanshachi, 2019). The variety of uses of different classification methods is presented in Table 1. The lack of systematic literature review on rework causes will require further investigation to unify all identified causes in a model that refers to the contract parties under various project stages (Ye et al., 2015; Asadi et al., 2021b). As such, a classification model is needed to be designed and proposed to address rework causes under the contract party's liability. Furthermore, there is no detailed analysis and systematic review of the identified rework causes link to the contract documents, while the other aspects such as change orders, payments, safety issues, conflicts and disputes have been studied in detail (El-adaway et al., 2016; El-adaway et al.,2017; Abdul Nabi et al., 2020; Saseendran et al., 2020). Therefore, the lack of addressing rework causes in the construction contracts under the project's main stages is considered a knowledge gap. Rework can be addressed in the contract documents, and to find an appropriate guideline for mapping rework in the contract (Asadi et al., 2023), firstly, causes of rework need to be identified and classified (Hwang et al., 2009) with the liable contract parties.

Several definitions for rework can be found in the literature (Josephson et al., 2002; Hwang et al., 2019; Liu et al., 2020), and under each definition, rework prevention or reduction requires both contract parties' attention and effort. In this study, rework is considered an activity that needs to be redone due to non-compliance with the contract. Since the impact of rework does not vary significantly with different procurement methods used (Love, 2002), this paper has reviewed all rework causes regardless of various procurement routes to avoid missing any items related to different contract types. Procurement is the key element of a project that plays a vital role in the success of the construction industry (Aiyetan, 2013). Contract as the main output of the procurement stage has been studied under different situations (Mendis et al., 2015; El-adaway et al., 2016). Non-etheless, the assessment of rework in contract documents needs further investigation as literature shows the lack of addressing rework in the contracts, which results in contractual claims and disputes (Wang et al., 2019; Asadi et al., 2021a). Thus, identifying the main causes of rework in construction contracts is considered as the key objective for this study followed by proposing a classification model.

2 Literature review

2.1 List of rework root causes under projects' main stages

The investigation into the sources of rework has taken more attention than the other study area of rework such as rework impacts, proposed strategies, implemented theories, reduction models and other solutions (Asadi et al., 2021b). Identifying the root causes of rework has always been an essential part of research around rework as it is the first step towards rework management either through reduction or prevention (Hwang et al., 2009; Ye et al., 2015; Ndwandwa et al., 2017). Rework identification is a necessity in construction projects as the consequences of rework lead to cost overruns and delay, poor organizational performance, and contractual claims (Love and Smith, 2003; Kim and Skibniewski, 2020). Cost and the other effects of rework are reduced when the causal structure of rework causes is understood (Love et al., 2010a). Various research have attempted to identify causes of rework in different types of construction projects including residential and commercial buildings, (Oyewobi and Ogunsemi, 2010; Hwang ans Yang, 2014; Ajayi and Oyeyipo, 2015; Yap et al., 2017; Liu et al., 2020; Mahamid, 2020), civil and infrastructure projects (Palaneeswaran et al., 2008; Love et al., 2010b; Zhang et al., 2012; Palaneeswaran et al., 2014; Enshassi et al., 2017; Hwang et al., 2019), and industrial plants.

The primary sources of rework are deviation from quality or non-conformance, changes, failures, defects, damages, errors, and omissions (Palaneeswaran et al., 2005). Numerous rework causes have been identified in different research by implementing various methods of benchmarking, cause and effect, regression, system dynamic, artificial neural networks, and fuzzy set theory based (Love and Smith, 2003; Robinson et al., 2004; Palaneeswaran et al., 2008; Love et al., 2010b; Aiyetan and Das, 2015; Hwang et al., 2019). The first list of rework causes with six categories (Josephson et al., 2002) was presented as a benchmark rework cost in the Swedish construction industry. In 2004, a full list of rework root causes under five categories was developed in Canada to measure and classify construction field rework using a fishbone diagram of cause and effects (Robinson et al., 2004). The identified causes of rework through the literature have been used for different purposes, for example, providing a framework to monitor rework in building projects (Palaneeswaran et al., 2005) or to determine the cost of rework in civil and infrastructure projects (Love and Li, 2000; Love et al., 2010b; Miri and Khaksefidi, 2015; Mahamid, 2016).

The causal structure of rework causes must be identified to understand the overall influence of rework on project performance (Love et al., 2010a). Thus, the list of rework root causes has evolved through empirical research and case studies worldwide. In some studies, the list of rework causes was identified and tested over construction projects, and in some others, it has been used for proposing a classification model or developing a framework of rework management (Love and Edwards, 2004; Love et al., 2009; Forcada et al., 2014; Aiyetan and Das, 2015; Zhang et al., 2018). Part of the reviewed papers focused on the identification of rework causes from the perspective of their impacts on project performance through measuring contractual claims, time, cost, quality and safety indicators (Palaneeswaran et al., 2008). Previous studies also have provided or used the list of rework root causes for various purposes such as investigating client-related factors (Hwang et al., 2014), minimizing the effects of rework to achieve more effective construction (Shah et al., 2016), reducing design-related causes (Wilson and Odesola, 2017; Salihu and Babarinde, 2020), examining the perception of professionals on the factors that trigger the occurrence of rework (Eze et al., 2018a), prioritizing rework indicators (Balouchi et al., 2019; Safapour et al., 2019), and evaluating the relationship between material waste and rework (Mahamid, 2020). Overall, it is required to be noted that rework occurs at various stages of a project. Non-etheless, when the emphasis is given to managing rework at the early stages, the impacts of rework will be reduced in the following steps (Ma et al., 2019). Thus, rework management at the design and procurement stages may result in fewer numbers of rework in downstream phases such as construction (Hossain and Chua, 2013).

Despite all efforts to investigate the causes of rework in different projects worldwide (Ye et al., 2015), the literature lacks a list of rework root causes to cover the main stages of a project under two parties of the contract. In the literature review, identified rework causes were largely from one and, to a lesser extent, from two project stages. Rework generally originates from the design stage, while it may occur in the other project stages (Love et al., 2010a). The focus of this paper is to identify and classify causes that result in rework in construction projects. Thus, it covers all causes that may appear in the process, starting from design to construction completion. Rework may also occur after construction work during the operation stage of the project (Kakitahi et al., 2013). Rework after practical completion, including the defect liability period till the end of the operation stage, is called failure and latent defect (Love and Smith, 2018) or defective work (Yap et al., 2017). Rework at this project stage is the result of causes initiated from previous project stages such as ineffective communication, inadequate design information, non-complaint building material, and insufficient works supervision (Kakitahi et al., 2013). The literature review showed no classified rework causes under the operation stage as there is no practical work after construction. Therefore, all the identified and classified causes are analysed under three main stages of design, procurement and construction; however, they may result in rework after the practical completion and during the defect liability period.

2.2 Background of the rework classification methods

The classification of rework causes is an essential part of the rework management process (Robinson et al., 2004; Zhang et al., 2012). The root causes of rework are traceable when a classifying method is employed. The classification provides referencing for future rework management and strategies (Love and Edwards, 2013). The literature review shows that most previous studies had classified the causes of rework based on their research aims. For example, in the study (Love and Li, 2000), the adopted classification system comprised two design and construction categories with five causes, including changes, errors, omissions, and damage. Literature indicates that the root causes of rework are categorized into various groups, including a) stakeholders, b) project stages, c) organizational factors and activities, and d) other miscellaneous aspects. Most of studies have classified rework causes based on the initial sources of rework from stakeholders "client, contractor, consultant and subcontractors" (Aiyetan, 2013; Palaneeswaran et al., 2014). Classification based on the occurrence of rework at different stages of the project (design, contracting management and construction) was the second-most used method in the literature (Zhang et al., 2012; Wilson and Odesola, 2017). The rest of the papers have added some other categories or used a combination of stakeholders and projects stages in their classified methods. A lesser proportion of the papers have not adopted a specific classification method and simply listed the causes of rework (Oyewobi et al., 2016). Table 1 presents the rework classification methods used in previous studies. All category names of rework causes used in the classification methods are listed in Table 1. The most used category names among the classification methods are design (21 studies), client (18 studies) and contractor (10 studies).

Detailed analysis of previous classification methods show that the list of rework causes under each category varies among different studies. The lists of rework root causes in most of the studies were adopted based on the research needs, project types and stakeholders' requirements. Therefore, the existing classified groups do not cover a wide range of rework causes liable to the contract parties. Furthermore, the same identified causes are found in different categories from different levels. For example, lack of communication is identified in the category of client (Mahamid, 2016; Eze and Idiake, 2018; Trach et al., 2019), contractor (Hwang et al., 2019), and subcontractor (Ye et al., 2015; Liu et al., 2020). Lack of communication also is found under various project stages such as design (Wilson and Odesola, 2017; Liu et al., 2020), construction (Mahamid, 2020), or other categories (Oyewobi et al., 2016; Enshassi et al., 2017). A single cause may occur in each project stages liable to different stakeholders, thus available categories in the literature may unable to show the accurate roots of rework causes. Lack of knowledge and training is another example that can occur in the design or construction stages (Eze et al., 2018b; Eze and Idiake, 2018; Hwang et al., 2019; Trach et al., 2019) and can be linked to both sides of the contract, either client or contractor. It also had been classified in the grouping of human resources in some other studies (Zhang et al., 2012; Shah et al., 2016).

By referring to above evidence and despite all the efforts to identify rework causes, there is a lack of a standard classification model to cover three project stages in conjunction with liable contract parties. A structured classification model including a list of rework root causes removes such limitations and summarizes the long list of identified causes from previous studies. Such a structured classification model effectively facilitates long-term solutions for eliminating or reducing rework issues (Taggart et al., 2014). The list of rework causes can be used and modified by project parties at initiation steps of the construction projects and assists practitioners to manage impacts of rework in later steps. As a result, the provided list is used to further investigate rework issues in the contracts by addressing rework causes in the contract terms and conditions. Addressing rework causes in the contract conditions results in fewer claims and disputes (Ndwandwa et al., 2017; Kim and Skibniewski, 2020; Asadi et al., 2022). Thus, the identified and classified rework causes in this study contributes to the body of knowledge in construction contracts.

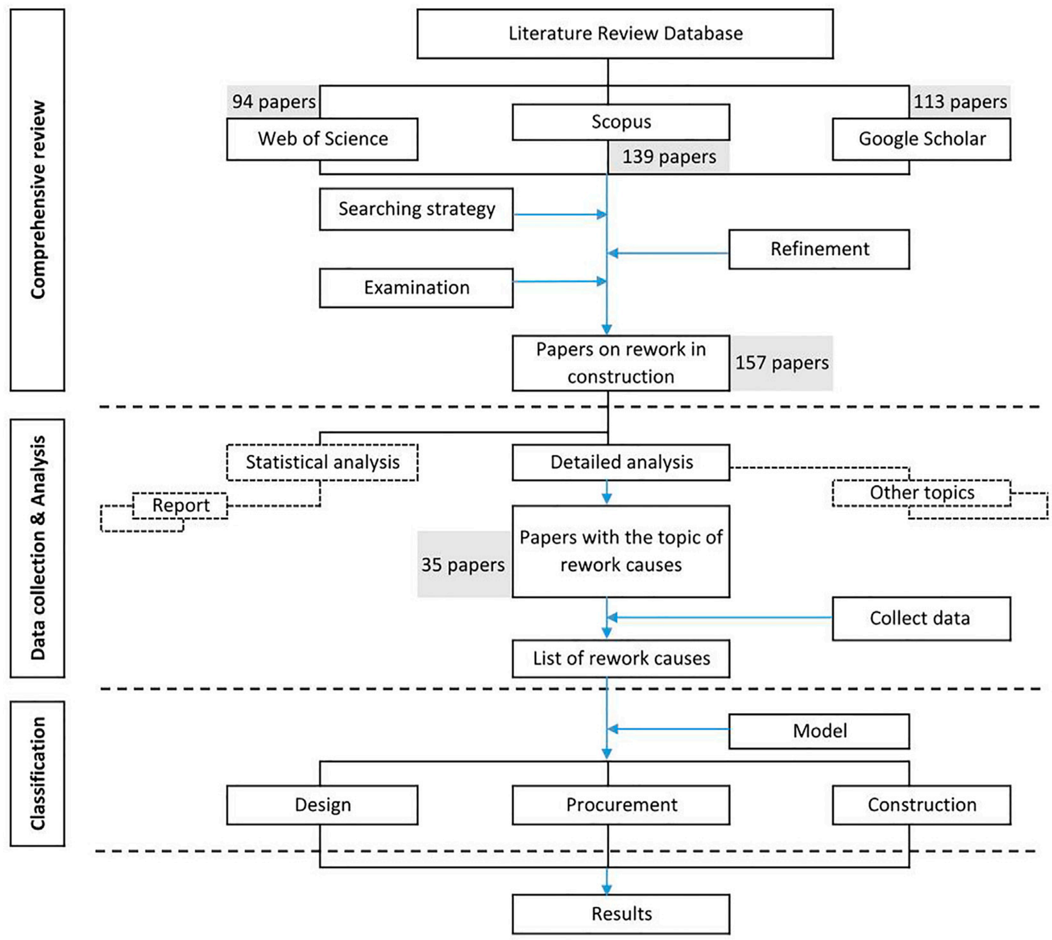


FIGURE 1
The process of research methodology.

3 Methodology

This paper is part of a research study that explores rework management in construction contracts. Poor contract management and scope ambiguity in the contract documents creates contractual claims and change orders and leads to rework occurrence (Palaneeswaran, 2006; Al-jababi et al., 2020). Thus, identifying the causes of rework is the key priority of this study, as previous research has recommended further investigation of the contract documents under rework circumstances (Forcada et al., 2014). One way to provide a list of identified rework root causes is through searching the published articles on rework in the construction sector. An effective way to identify knowledge gaps and pave the way for future studies is through literature review. Published papers on the topic of rework from 1990 to 2021 were obtained through the following systematic literature review steps. The reason for the selected period is that rework research has started expanding after the first official definition of rework by Construction Industry Development Agency in 1995 (Love et al., 2018). A systematic literature review as a scientific activity is more often used for management practices. An extensive literature review on rework causes identifies gaps in each stage of the project and provides the opportunity for further recommendations to address rework issues. The sequences of implementing this comprehensive review consist of identifying research by generating a strategy, selecting criteria to include and exclude studies, collecting relevant articles, quality assessment of the collected documents, data extraction, and result summarization (Park and Tucker, 2017). This review was initiated to answer this question of what the root causes of rework in construction projects are? Rework is a term widely used in many research fields and is interchangeable with several other words (Love and Smith, 2003). Such similar words may cause interferences in the search result and since the paper aimed to identify the causes of rework in contracts, the scope of searching publications was confined only to the keyword of "rework" in the paper's title considering all stages of the project. The search was further limited only to construction projects with the document type of article and conference papers. Other published papers under the broad categories of editorial, book review, forum, discussions/closures, letters to the editor, articles in the press, index, foreword, introduction, conference/seminar report, briefing sheet, and comment were excluded from the search analysis.

Following the commonly used method in previous studies, suitable search engines of Scopus, Web of Science "WoS," and

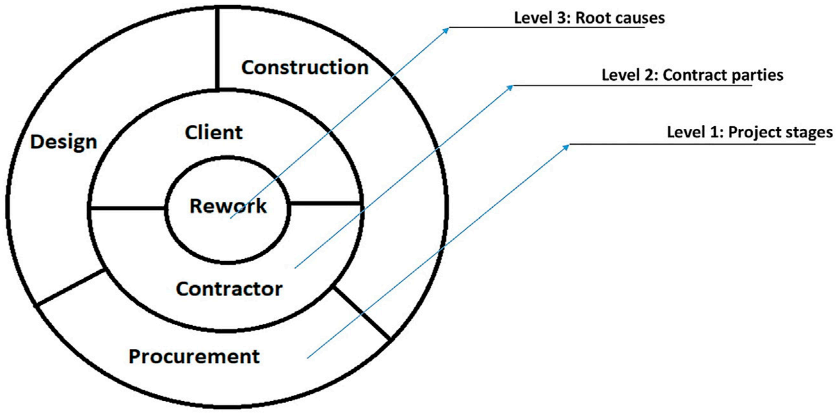


FIGURE 2
Conceptual structure for rework root causes classification.

Google Scholar were employed for searching (Bao et al., 2018; Habibi and Kermanshachi, 2018; Wang et al., 2020). The study began by a) searching for the rework topics in the Scopus followed by two complementary searching engines, b) selecting relevant papers and c) clustering them based on the subtopics, and then d) detailed analysis of rework causes in three stages of design, procurement and construction. In the first search round, Scopus was utilized as it is widely used for reviewing construction literature (Park and Tucker, 2017). The search was then completed using the Web of Science and Google Scholar search engines (Chan et al., 2020). Using the second and third search engines ensured that no relevant document is missed. The supplementary search engines cover the potential limitations of each other. The following algorithm has been considered to meet the systematic literature review criteria, as described in Figure 1.

3.1 Comprehensive literature review

The critical part of each research finding can be acquired and shared through publications for further advancement in a specific area (Bao et al., 2018). The publications provide a wider perspective of the topics and accordingly researchers can build a new idea by continuing the work of others. As a result, a comprehensive review of published papers identifies the current status and would help researchers to investigate future required research (Ke et al., 2009). The review process of finding relevant journal and conference papers is started when the research requirements are adopted based on the scope of the study (Safapour et al., 2019; Malek and Desai, 2020). The afore-mentioned requirements then are divided into three main steps: defining a search strategy, refining documents for inclusion in and exclusion from the papers, and examining of the documents to ensure that only selected papers are in the study's scope and are not duplicated. The search for rework in three search engines was conducted in June 2021 with the following codes:

- (TITLE (rework) AND ALL (construction)) AND DOCTYPE (ar OR cp) AND PUBYEAR > 1989 AND (LIMIT-TO (LANGUAGE , "English")) in the Scopus search engine,
- (TITLE: (rework) AND TOPIC: (construction), Timespan: 1990-2021. Databases: WOS, BIOABS, CABI, CCC, FSTA, KJD, MEDLINE, RSCI, SCIELO. Search language=English) in the Web of Science search engine and,
- (with all of the words (rework) AND (construction) in the title of the article, Return articles dated between, 1990–2021, excluding citations AND search English pages)) in the Google Scholar search engine.

The search then resulted in 139 papers in Scopus, 94 papers in Web of Science, and 113 papers in Goggle Scholar. A total of 346 papers were then reviewed to remove the irrelevant documents and duplicated papers among three search engines. The selection of more relevant papers will lead to more accurate results. Thus, initially, all paper titles were reviewed carefully to check whether they meet the research scope or not. The papers with titles out of the research scope were removed. Following that, the abstract of all remaining documents was scanned to segregate the relevant papers according to research scope. Besides, all papers with the same titles were identified to avoid duplications. The outcome of this process resulted in 93 papers in Scopus, 12 papers in Web of Science, and 52 papers in Google Scholar. In total, 157 papers remained for further analysis.

3.2 Data collection and analysis

All 157 papers were studied to take out the required data around the causes of rework and provide a list of rework root causes that covers both liable parties of the contract under three project stages. The following strategy was developed to generate a report on the extracted data. The main criteria for selecting papers were mainly based on discussing the causes and sources of rework. Thus, all papers that contain the list of rework causes were clustered in a separate group for further analysis and providing a report. Of 157 publications, the total number of 35 papers had provided a list of rework causes. Since the paper aimed to unify the identified causes and classify them, all identified lists, including their classification methods, were transferred to an Excel file for further analysis. Different categories appeared from the extracted data of the collected classification methods. The Excel file also revealed that the number of cases in each study varied and ranged between a minimum of seven items (Hwang et al., 2014) and a maximum of 77 items (Oyewobi and Ogunsemi, 2010). The Findings and Discussion sections of the paper contain more details about the analysis of rework causes and their categories.

3.3 Classifying the results

Despite the various points of view on classifying rework root causes, having a model for managing rework consequences is essential when rework causes are identified (Hwang et al., 2009). As shown in Table 1, various rework classification methods have been used in the construction industry for categorizing identified rework causes and measuring their impacts. The identified rework causes from the literature can be categorized into different groups based on their similarity and differences. A flexible model for various construction projects can cover all identified rework causes and provide possible solutions or recommendations. Since this research aims to assess rework in contracts across the main stages of projects, the model needs to contain design, procurement, and construction with liability to the contract parties. The classified causes of rework will be used for further studying of the contract (Al-Jababi et al., 2020), so the liable contract parties would need to be included in the classification model. For this purpose, a classification system at three levels in Figure 2 is adopted to address all reviewed and analyzed rework root causes. The novelty of this research is to propose a structure to address rework causes under project stages with their liable contract parties.

3.4 Proposing a rework classification model

Providing a systematic approach for classifying the causes of rework will improve the rework management consequences (Miri and Khaksefidi, 2015). Rework causes can be grouped based on their sources, impacts, occurrence in project stages, and involved parties. However, rework is found at many project stages, such as inspection, construction, after handover, or even during defect liability period and project maintenance, the causes of rework are generally investigated and linked to the operational stages (Taggart et al., 2014). The primary purpose of this research is to investigate rework issues in construction contracts. The standard form of contract in the construction industry is involved two parties overall. Thus, this study has focused on these two main parts—The client and the contractor. The client-side of the contract generally encompasses consultants and the contractor side covers subcontractors and their suppliers. Depending on the contract types, both parties are involved from the initiation stage of a project. Therefore, both client and contractor may contribute to rework occurrences at each project stage. The stage of project shows when rework occurs, during the contract period. The contract's liable parties indicate where the causes of rework originate. Lastly, root causes of rework may sit under each of the stages and liable contract parties to explain how they may be initiated. The proposed model in this study covers rework causes either from various sources or at different project stages. The proposed model has incorporated contract parties and various project stages at two separate levels to tackle these contradictions. As a result, six clusters appear in the model, namely, design-client (cluster 1), design-contractor (cluster 2), procurement-client (cluster 3), procurement-contractor (cluster 4), construction-client (cluster 5), and construction-contractor (cluster 6). Assessing rework causes under these six clusters is the novelty of the research, which has not yet been examined in the literature. The proposed model in three levels describes when, where, and how rework occurs. These three questions are the critical points to address rework in the contract documents.

The employed concepts and terminologies in Table 1 are used as the basis of the proposed model in this paper. The concept and criteria applied in previous studies have been considered in the proposed model to avoid missing or duplicating any of the identified causes. The identified rework root causes from the literature then are grouped into five constructs under each of the clusters. Grouping the similar rework root causes is necessary to understand the interdependency of causes and avoid duplications (Siraj and Robinson, 2019). The selected five categories in this paper are align with five group of rework causes in previous studies (Asadi et al., 2021a; Asadi et al., 2022). The category names in the proposed model were adjusted precisely to cover all identified categories in Table 1. In the case where a category name is found with other similar terms, the more inclusive name has been chosen. For example, the category name of "Contract Management" is used to cover "Procurement" and the category name of "Construction" covers both "Site Management" and "Construction". Furthermore, "Process" is selected over "Change/Error/Omission" as it covers all the interchangeable names, and "Human Resources" is selected over the other terms of "People and Workmanship" (Yap et al., 2017). Overall, the proposed classification model clearly shows the various groups of rework causes with their liable contract parties in three main stages of a project as described below.

3.4.1 Level 1: Project stages

According to PMI "Project Management Institute," the life cycle of a project consists of developed sequences from initiation to closing. This definition in the construction industry in overall, consists of three main stages of design, procurement and construction (Habibi et al., 2018), which are sometimes called EPC in megaprojects. The design stage covers items from previously identified and classified rework causes in Table 1 under design/engineering reviews category. The procurement stage covers rework causes from previous contracting management categories (Coleman et al., 2020), and the construction stage covers previous categories of site management and construction (Palaneeswaran et al., 2005). Rework causes in the category of "Project" are distributed among all stages depending on the nature of the items.

3.4.2 Level 2: Contract liable parties

Since contract documents in construction projects are signed off between the client and contractor, this paper has classified rework causes into these two main groups as described in introduction section. The client parts cover two previous categories of client and consultants. The contractor parts cover three categories of contractor, subcontractor and supplier from previous classification methods (Assaad et al., 2020). Rework causes under "Organizational factor" are distributed among both parties depending on the nature of the items. The study conducted by Taggart et al., (2014) showed that rework root causes are identified when a more collaborative approach is adopted among the supply chain stakeholders. Rework may also contribute to other parties such as consultants, suppliers and subcontractors. This study investigates rework issues between two main sides of the contract as the client and contractor and that is another limitation to the study which can be covered by further research considering other parts involved in the project.

3.4.3 Level 3: Root causes

Continuance to the afore-mentioned levels; this paper has categorized the root causes of rework under five distinct constructs: process, human resources, material and equipment, technical, and general/external factors. The group of process-related factors covers the items from the previously classified category with the same title (Zhang et al., 2012). Human resources-related factors cover the previous categories under the names of workmanship, people, and human resources (Robinson et al., 2004; Forcada et al., 2014; Yap et al., 2017). The material and equipment related factors cover the same category in previous models plus the machine's category (Josephson et al., 2002; Enshassi et al., 2017). The technical-related factors cover the following categories of planning and scheduling, leadership and communication, technical and quality management (Robinson et al., 2004; Oyewobi and Ogunsemi, 2010; Ndwandwa et al., 2017). The last group of general/external factors covers the previous categories of environmental/external, and other related factors (Ye et al., 2015; Mahamid, 2016; Eze et al., 2018b; Liu et al., 2020). In this way, all identified causes of rework throughout the literature are allocated to the proposed model.

Figure 3 presents a diagram of the proposed model in which the horizontal line has divided it into two similar patterns that may occur on both sides of the contract. The upper section addresses the client and the lower section is for the contractor. This line also has been considered as the border between the two and can be treated as the contract. The contract documents are signed at the end of the procurement stage, while before and after that point are known as pre-contract and post-contract, respectively (Love et al., 2009; Kakitahi et al., 2014). It also shows how a project evolves from design on the left side of the diagram towards the end of construction on the right side. Each of the five constructs has been placed on both sides and three stages to check the probability of rework occurrence in the entire process. Each box in the diagram contains numbers of rework root causes, as described in Tables 2–7. Three abbreviations based on the three-level classification system have been allocated for coding of each boxes. The first character of the code shows the project's stage, and includes "D" for design, "*P*" for procurement, and "C" for construction. The second character shows the liable contract parties in which "E" is used for the client and "S" for the contractor. The third character shows the relevant category of root causes as "*P*" for the process, "H" for human resources, "M" for material and equipment, "T" for technical, and "G" for general/external related factors. For example, DST means technical related factors that originated by contractor at the design stage and may comprise more than one root cause that will be demonstrated as DST1, DST2, and more. Each identified rework cause places in one of these boxes. The diagram has been developed based on the proposed model, and it is part of the study's novelty. All identified rework causes were then placed in the adopted classification model to provide a list of rework causes that can be utilized in construction contracts throughout various project stages. Allocating rework root causes to the diagram facilitates the preparation of the list of rework causes from various studies. It contributes to the body of knowledge by generating a matrix that addresses rework management gaps in construction projects, as shown in Table 8.

4 Findings

4.1 Rework causes classification and ranking in the project stages

Several research efforts have attempted to identify and classify the causes of rework and quantify its overall extent (Aiyetan, 2013), and each study's outcomes show different results. Engineering and reviews such as unsuitable or faulty design and lack of coordination in the design stage were known as the most significant causes of rework at the design stage (Josephson et al., 2002; Robinson et al., 2004). According to Love et al. (2000), the main source of rework is documentation that is basically used for construction activities. In other studies, human resources capability, and rework factors related to the contractors were prioritized as the major rework categories that affect the projects' performance (Enshassi et al., 2017; Eze and Idiake, 2018). The top five rework factors were identified as error and omission, labour skills, inadequate supervision, scope changes, and NCR to the specifications and requirements of the project (Mahamid, 2020). The importance and frequency of these causes vary from one study to another depending on the project type and locations.

Distribution of the identified rework root causes within the proposed classification model is presented in Tables 2–7. Each table presents a list of classified rework causes in a project stage that is linked to one side of the contract. The identified causes under each study are analyzed separately to bring together the same items under each table. The frequency of repeating an item over the studies is indicated in the last column of the tables. This indicator is for sorting the causes and does not make any priority in terms of severity or probability of rework occurrences. It only reflects the high interest of previous researchers on the identified cause and explains how it was commonly used among previous studies. The severity or probability of occurrence is generally used to measure the impacts (Robinson et al., 2004; Love et al., 2011). However, such measurement is not in the scope of this research as this study focuses on identifying the causes of rework regardless of their severity or occurrence probability. Thus, the number of citations in previous studies is considered an indicator for sorting the identified causes. For making the list of causes in each project stage, the following process take place to maximize the accuracy of referencing from the literature (Chan et al., 2020).

- Provide a list of rework causes under each author name to define the interrelations between causes

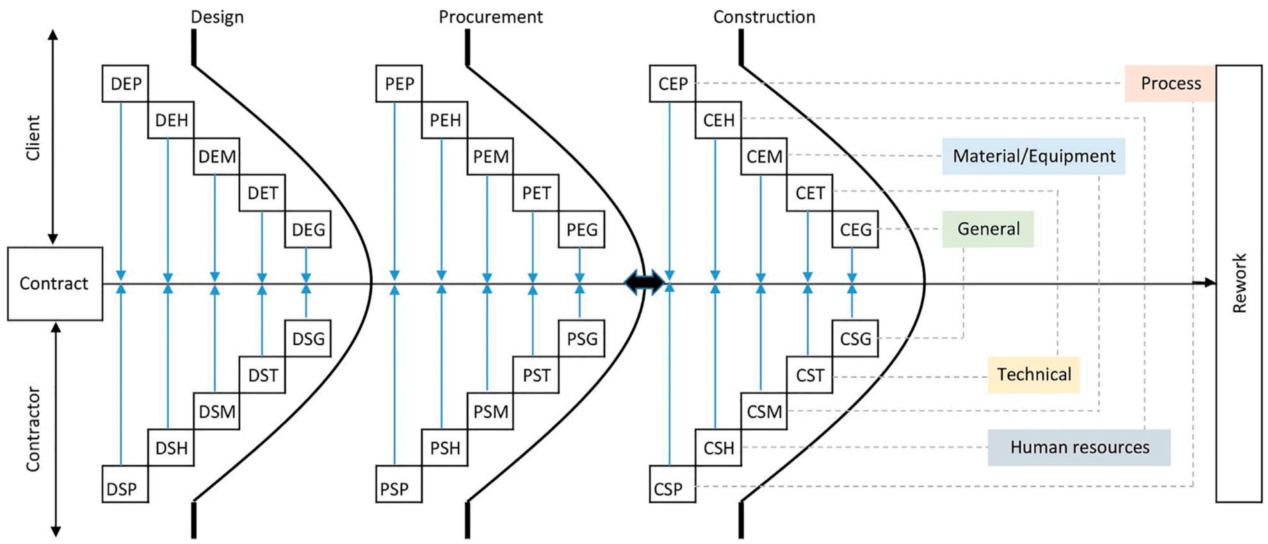


FIGURE 3
Classification model for rework root causes in life cycle of project with liable parties of the contract.

- Review the identified cause regarding the relevant paper's title as precisely as possible
- Ensure that categorizations of causes are uniform by comparing the provided lists
- Match the retrieved information with the details of the project stage
- Optimize the results through merging the same items in each stage separately

This processing confirms that all causes have been grouped in the most relevant and correct places. It also reduces the variety of views in providing more general causes that will incorporate easily into the more overall frameworks. Moreover, it would help the identification of more effective response strategies that can be allocated to the appropriate contract party. It finally generates a list of rework causes that can be adopted for different projects with two liable contract parts. Below are details of rework causes under three main project stages.

4.1.1 Design stage

The design stage consists of two main parts: pre-design/conceptual design and detailed design (Yeo and Ning, 2002). Error and incomplete information in the design stage generate rework (Love and Li, 2000). The development of this stage has been identified as the most important activity in a project's life cycle (Habibi et al., 2018) as the initial concept of the client's requirement consolidates into reality within this stage. During the design process, naturally, errors occur and sometimes lead to change and, consequently, rework (Love et al., 2010a). All items under this category from previous research studies were carefully reviewed to combine altogether and make a single list of causes in the design stage. Table 2 comprises 24 root causes of rework from the client-side of the contract concerning their code number, references and numbers of citations over various studies.

The findings suggest that rework causes in the design stage of a project have appeared in all five factors. However, the distribution of root causes among the categories is not in balance for both sides of the contract. The most frequent category in the design stage is human resources related factors, on both sides following the technical related factors. The low number of design errors brings more reputation to the firm and presents a better image of the designers (Love et al., 2000). Overall, the total number of rework root causes at this stage of the project is more likely to slope towards the client rather than the contractor with 24 items against 21. Table 3 comprises 21 root causes of rework from the contractor side of the contract concerning their code number, references and numbers of citations over various studies.

The most cited cause in both contract sides is a poor communication system with a frequency of more than 20 times from 35 publications. This particular cause was reported as one of the most severe causes of rework (Mahamid, 2016) in residential buildings that frequently trigger the need for rework management (Eze et al., 2018a). An acceptable logic for that is because communication provides channels for exchanging information between involved parties (Ye et al., 2015). Design error is another common cause of rework that can be reduced when participants follow integrated mechanisms to manage documentation in the design process (Love et al., 2000).

4.1.2 Procurement stage

The procurement stage plays a critical role in the success of construction projects as the project's cost and time are affected by different procurement strategies (Love et al., 2000). The errors that appear in the design stage affect the procurement process and reduce the project performance (Love et al., 2010a). Procurement is a stage between design and construction and generally overlaps with both stages. The overlapping of procurement with design and construction stages creates higher uncertainties during contract preparation (Yeo and Ning, 2002). Since procurement follows the

TABLE 2 Rework causes classification in design stage—liable to the client-side of the contract.

	Rework root causes	Code	Covered by references	Sum
1	Poor communication system	DET1	[2,5,6,7,9,10,11,12,15,16,17,18,19,21,24,27,28,29,30,33]	20
2	Changes, modification, and revisions	DEP1	[1,2,3,4,7,8,9,10,11,13,15,18,19,22,24,25,30,31,33]	19
3	Financial issues and lack of funding	DEG1	[2,5,6,7,9,11,12,13,16,17,18,19,21,24,28,29,30,31]	18
4	Conflicting and incomplete information	DET2	[1,3,5,6,9,11,14,15,16,17,18,19,22,23,24,26]	16
5	Lack of client involvement	DEG2	[2,5,6,7,8,9,12,16,17,19,21,25,26,28,29,30]	16
6	Lack of experiences and personal expertise	DEH1	[2,5,6,7,8,9,12,14,16,18,19,21,28,29,30]	15
7	Lack of education and poor knowledge of team	DEH2	[2,5,6,7,8,9,10,12,16,19,21,28,29,30,32]	15
8	Time pressure to complete design tasks	DEG3	[2,7,8,9,11,12,14,18,19,21,23,28,29,30]	14
9	Inefficient management	DET3	[8,10,11,13,17,18,24,31]	8
10	Inadequate planning and poor scheduling of workload	DET4	[7,8,9,10,11,14,18,19]	8
11	Lack of skill	DEH3	[3,8,10,14,17,21,24]	7
12	Defective material, non-adherence to material spec	DEM1	[1,3,10,13,14,20]	6
13	Ineffective use of quality management practices and deviation or failure due to poor monitoring	DET5	[7,8,9,11,14,18]	6
14	Incomplete design, omission in design, drawings, spec	DEP2	[3,7,8,9,10,19]	6
15	Lack of documents control	DEP3	[3,10,18,23,32]	5
16	Inadequate supervision staff	DEH4	[3,9,14,19,27]	5
17	Error in design, drawings and specifications	DEP4	[7,8,9,15,19]	5
18	Poor technology application and lack of IT use	DET6	[7,8,9,17,18]	5
19	Conflict of interest	DEH5	[11,20,24,31]	4
20	Inappropriate personal attitude	DEH6	[8,10,24]	3
21	Lack of employee motivation and rewards	DEH7	[8,10,24]	3
22	Unclear line of authority	DEG4	[20,24]	2
23	The absence of job security and other safety rules	DEH8	[20,24]	2
24	Unanticipated consequences of change	DEG5	[8]	1

design stage of the project, it sometimes is called post design. In some other studies, procurement has been referenced as the pre-construction stage (Love and Sing, 2013). During the procurement stage, documents from design stage are collated to perform tenders till contract awards. In the construction industry, procuring the required equipment and material is part of the contractor's main responsibility (Habibi et al., 2018). Thus, the procurement stage in some projects involves supplying materials and equipment. All relevant items from previous research studies were carefully reviewed and analysed, then placed in a single list. Table 4 comprises seven root causes of rework from the client-side of the contract concerning their code number, references and numbers of citations over various studies.

The procurement method is the primary element of the project that is influenced by rework causes (Love et al., 2010a). The evidence of rework root causes at this stage revealed that only the three categories of process, technical, and general/external related factors appeared in both sides of the contract, and there is no evidence of studying the other two categories in previous research. An investigation into the categories of human resources and material/equipment from the procurement perspective in the future would benefit the projects. As such, further studying of the following items will contribute to the body of knowledge: lack of manpower, staff motivation, inadequate knowledge and experience from the category of human resources factors, lack of information technology use and lack of attention to quality from the category of technical factors in addition to political effects from the category of general/external related factors. Similar to the design stage, the total number of rework root causes from the client-side of the contract is higher than the contractor side. Table 5 comprises five root causes of rework from the contractor side concerning their code number, references and numbers of citations over various studies.

Poor contract documentation and omission of items in the contract are the highly-cited causes of rework under the client-

TABLE 3 Rework causes classification in design stage—liable to the contractor-side of the contract.

	Rework root causes	Code	Covered by references	Sum
1	Poor communication system	DST1	[1,2,5,6,7,8,9,10,11,14,15,18,22,23,24,25,26,28,29,30,31,32,33,35]	24
2	Inadequate planning and poor scheduling of design resources	DST2	[1,2,3,5,6,7,8,9,10,12,14,16,20,23,25,28,29,35]	18
3	Design changes in any form	DSP1	[3,4,7,8,9,10,11,15,18,20,22,23,25,28,29,30,31]	17
4	Time pressure to complete design tasks	DSG1	[3,5,6,7,8,9,11,12,14,16,23,24,27,28,29,35]	16
5	Poor technology application and lack of information technology use	DST3	[2,5,6,7,8,9,11,12,16,17,18,23,28,29,30,35]	16
6	Ineffective use of quality management practices and deviation or failure due to poor monitoring	DST4	[2,5,6,7,8,9,10,11,12,14,16,23,27,29,35]	15
7	Design errors in any form	DSP2	[3,7,8,9,10,11,15,22,23,24,25,28,29,31]	14
8	Inefficient management and ineffective coordination	DST5	[8,10,12,15,16,17,24,27,28,29,32]	11
9	Omission in design process and incomplete design information	DSP3	[1,3,8,10,11,15,22,23,24,25,31]	11
10	Lack of experiences and personal expertise	DSH1	[7,8,9,11,14,17,18,23,25,26,32]	11
11	Lack of skill	DSH2	[3,8,9,10,23,24,26,28,29,30,32]	11
12	Lack of education and poor knowledge of team	DSH3	[7,8,9,10,14,17,20,23,24,26]	10
13	Inadequate manpower to complete the task	DSH4	[2,5,6,7,8,9,12,16,25,35]	10
14	Labor reallocation, alteration and staff turnover	DSH5	[2,5,6,7,8,9,12,16,35]	9
15	Lack of document control	DSP4	[3,10,14,22,23,32]	6
16	Poor project documentations	DST6	[10,20,22,23,24]	5
17	Financial issues and lack of funding, cost pressure	DSG2	[17,20,24,25,26]	5
18	Lack of employee motivation and rewards	DSH6	[8,10,20,24,32]	5
19	The absence of job security and other safety rules	DSH7	[8,20,22,24,32]	5
20	Non-attention to constructability problems raised at early stages	DSG3	[5,6,15,23]	4
21	Inappropriate personal attitude	DSH8	[8,10,24]	3

TABLE 4 Rework causes classification in procurement stage—liable to the client-side of the contract.

	Rework root causes	Code	Covered by references	Sum
1	Poor contract documentations, missing documents	PET1	[2,4,7,8,9,11,12,16,17,18,19,23,25,26,27,28,29,30,31,33]	20
2	Time pressure, insufficient time to prepare contract documentation	PEG1	[2,5,6,7,8,9,12,17,19,25,26,28,29,30,35]	15
3	Incomplete design at the time of tender	PEP1	[2,5,6,7,9,12,19,25,26,28,29,30]	12
4	Financial issues such as low fees for preparing contract documents	PEG2	[2,9,12,21,28,29,30]	7
5	Improper contractor selection	PEP2	[8,11,12,21,22]	5
6	Errors made in the contract documentation	PEP3	[2,4,7,9,30]	5
7	Incomplete information at the time of award	PET2	[5,6,8,11]	4

procurement cluster with the frequency of 20 times over publications. Under the cluster of contractor-procurement, two other causes, namely "ambiguity in contract due to conflicting and incomplete information" and "improper subcontractor selection," were prevalent as the most studied items over 10 publications. Ambiguous scope in the contract documents results in poor contract management and leads to rework (Ye et al., 2015). Employing qualified contractors has been pointed out as the key contributing factor of the industry's success (Chan et al., 2020).

4.1.3 Construction stage

However, many research studies have emphasized the importance of the design stage in the process of construction

TABLE 5 Rework causes classification in procurement stage—liable to the contractor-side of the contract.

	Rework root causes	Code	Covered by references	Sum
1	Conflicting and incomplete information, ambiguity of items from contract documentation	PST1	[5,6,14,15,17,28,29,30,33,35]	10
2	Improper subcontractor selection	PSP1	[4,5,6,8,11,12,19,22,25,34]	10
3	Poor project documents	PST2	[5,6,15,23,28,29,30,33]	8
4	Inadequate procurement method, poor contract execution	PSP2	[8,11,15,22,23,24,31]	7
5	Financial issues such as low contract fees	PSG1	[7,9,15,33]	4

TABLE 6 Rework causes classification in construction stage—liable to the client-side of the contract.

	Rework root causes	Code	Covered by references	Sum
1	Lack of knowledge of construction	CEH1	[2,5,6,7,8,9,10,12,14,16,19,20,21,24,26,28,29,30]	18
2	Lack of experience and personal expertise	CEH2	[2,5,6,7,8,9,10,12,14,16,17,19,21,26,28,29,30]	17
3	Financial issues and cost pressure	CEG1	[2,3,7,8,9,11,12,13,15,18,24,26,33]	13
4	Changes or modification in the construction process or after completed work	CEP1	[1,3,4,7,8,9,11,13,18,22,24,34]	12
5	Lack of client involvement	CEG2	[2,5,6,7,8,9,11,17,25,26,28,29]	12
6	Lack of constructability	CEG3	[3,5,6,15,17,20,24,26,28,29,31]	11
7	Inadequate construction planning and poor planning of workload	CET1	[3,7,8,9,10,11,12,14,17,18,24]	11
8	Ineffective management practice	CET2	[8,11,13,14,18,22,24,31]	8
9	Conflicting and incomplete information	CET3	[1,8,11,14,15,18,22,24]	8
10	Defective materials	CEM1	[1,8,10,13,14,20,32]	7
11	Ineffective use of quality management practices and deviation or failure due to poor monitoring	CET4	[7,8,9,11,14,18,24]	7
12	Poor technology application and lack of IT use	CET5	[7,8,9,12,17,26]	6
13	Changes in government regulations, laws, and policies	CEG4	[15,20,24,31]	4
14	Unpredictable factors from different sources	CEG5	[8,15]	2

projects, yet most problems such as delay, and cost overruns were raised from rework within the construction stage. Construction commencement before design completion affects the project's performance as the overlapping of these two stages transfers the impact of design errors to the job site thus, increasing the time and cost of the project (Habibi et al., 2018). The previous classification systems of rework under construction group included contractor-related factors, site management, and subcontractor factors (Palaneeswaran et al., 2005). In this study, all previously identified items under these categories and other related factors were compared to provide the following single list of rework root causes at the construction stage. Table 6 consists of 14 root causes from the client side of the contract concerning their code number, references and numbers of citations over various studies.

The findings of rework root causes in the construction stage of a project revealed that, as with the design stage, all categories of causes are in existence in both contract sides. The highest number of rework causes in this stage has fallen in the technical category following by the general/external related factors. The other most frequent cause of rework in this stage is human resources in the contractor side; however, the number of causes in this category from the client-side is deficient. In contrast to the pre-mentioned stages, the total number of rework root causes from the contractor side is higher than the client-side. This evidence implies that previous studies have mostly focused on the contractor side as they have been more involved in this project stage. Reducing the causes of rework in the construction stage improves the performance of construction firms by cost saving (Love and Li, 2000). Table 7 comprises 27 root causes of rework from the contractor side of the contract concerning their code number, references and numbers of citations over various studies.

Under the cluster of client-construction, lack of knowledge was the most identified cause by 18 publications. Lack of client knowledge has been introduced as the root causes of many problems in construction projects (Trach et al., 2019). The used of skilled and experienced professionals by both contract parties throughout different project stages also had been recommended to achieve free rework construction (Eze et al., 2018b). Ineffective use of quality management with 25 citations is listed first under the cluster of construction-contractor as well as the highest frequency

TABLE 7 Rework causes classification in construction stage—liable to the contractor-side of the contract.

	Rework root causes	Code	Covered by references	Sum
1	Ineffective use of quality management practices and deviation or failure due to poor monitoring	CST1	[2,3,4,5,6,7,8,9,10,11,12,14,15,18,19,20,22,24,25,27,28,29,31,32,33]	25
2	Inadequate construction planning and poor planning of workload	CST2	[1,2,3,4,5,6,7,8,9,10,12,14,17,20,22,24,25,26,28,29,30,33]	22
3	Lack of skills in both labour and supervisory levels	CSH1	[2,3,4,5,6,8,10,12,16,17,19,20,22,24,25,27,28,29,30,32,34]	21
4	Damage, defect or deviation of products due to carelessness and poor safety consideration	CSG1	[2,4,5,6,7,8,9,11,12,15,16,17,18,19,22,26,28,29,31,34]	20
5	Use of poor-quality material and substandard products	CSM1	[1,2,4,5,6,8,9,11,12,15,16,17,18,19,20,25,26,31,32,33]	20
6	Construction error	CSP1	[2,4,5,6,7,8,9,11,12,15,16,18,19,22,25,28,29,31,34]	19
7	Ineffective management practice and poor site management	CST3	[3,8,11,14,15,16,17,18,19,22,24,25,26,27,28,29,33,34]	18
8	Changes or modification in the construction process or after completed work	CSP2	[2,4,5,6,7,8,9,11,15,16,18,22,28,29,30,31,33,34]	18
9	Poor site condition, environmental conditions	CSG2	[5,6,8,15,17,19,20,22,24,26,27,28,29,31,32,33,34]	17
10	Poor workmanship approach and inappropriate personal attitude	CSH2	[1,5,6,8,10,16,17,18,19,20,24,26,28,29,30,31,34]	17
11	Lack of knowledge, unqualified technical staff due to lack of training	CSH3	[7,8,9,10,14,17,20,22,24,26,28,29,30,31,32]	15
12	Poor communication system	CST4	[7,8,9,10,14,15,17,19,22,24,25,31,32,33,34]	15
13	Labour reallocation, alteration and staff turnover	CSH4	[2,4,5,6,7,8,9,12,17,19,22,26,28,29,34]	15
14	Schedule acceleration, time pressures	CSG3	[3,7,8,9,14,15,17,20,22,24,25,26,28,29]	14
15	The omission of some tasks during construction	CSP3	[2,4,5,6,7,8,9,11,17,22,26,30,31,34]	14
16	Poor technology application and lack of information technology use	CST5	[2,4,5,6,7,8,9,12,15,17,24,26,27]	13
17	Defective or damaged materials	CSM2	[1,3,8,10,11,15,18,20,24,25,27,31]	12
18	Inexperienced personnel	CSH5	[7,8,9,14,17,22,25,26,28,29,32]	11
19	Replacement or misplacement of material	CSM3	[1,10,14,15,20,24,28,29,30,31]	10
20	Use of inefficient equipment	CSM4	[1,10,15,19,20,24,25,27,33,34]	10
21	Financial weakness such as inadequate funding, cost pressure	CSG4	[3,8,15,20,22,24,25,26,32,33]	10
22	Conflicting and incomplete information	CST6	[3,8,14,20,22,25,26,28,29,33]	10
23	Lack of motivation and care, Carelessness	CSH6	[8,10,19,20,22,24,32,34]	8
24	Untimely deliveries of material and equipment	CSM5	[1,3,8,10,20,24,25]	7
25	Lack of manpower to complete the tasks	CSH7	[7,8,9,17,26,31]	6
26	Poor project documents	CST7	[8,20,22,24]	4
27	Unpredictable factors from different sources	CSG5	[8,15]	2

among all identified rework causes throughout the project. This is aligning with the previous research directions that rework is due to the lack of quality focus (Love et al., 1999). However, quality management does not correspond to previous studies' outcomes as none of them have highlighted quality related factors in the list of top rework causes.

5 Discussion

Classification of the identified rework causes is the first required step of rework management (Palaneeswaran et al., 2005). This study presents an extensive literature review of rework causes related to both sides of the contract and the three project stages. Most previously generated classification methods are comparable as they almost follow the same pattern. Thus, achieving a comprehensive categorization scheme for classifying rework causes is implicitly possible. Researchers have used different phrases to present the same rework causes in the literature. This study carefully surveyed each identified item from different sources to bring together the most interrelated concepts of the causes. For example, excessive overtime (Robinson et al., 2004), fixed time for a task (Palaneeswaran, 2006; Miri and Khaksefidi, 2015), time boxing (Love et al., 2010b), schedule pressure (Enshassi et al., 2017), accelerating or shortening the schedule (Ye et al., 2015), and insufficient time for activities (Wilson and Odesola, 2017) are

TABLE 8 List of rework causes in three stages of the project with liable contract parties.

Project stages			Design		Procurement		Construction	
Contract parties			Client	Contractor	Client	Contractor	Client	Contractor
Group factor	Rework root causes	Code	DE	DS	PE	PS	CE	CS
Process	Changes, modification and revisions in design/ construction changes		P1	P1	—	—	P1	P2
	Error in design, drawings and specifications/ construction error		P4	P2	P3	—	—	P1
	Incomplete design, any omission in the design or construction process		P2	P3	P1	—	—	P3
	Inadequate procurement methods/poor contract execution		—	—	—	P2	—	—
	Improper contractor and subcontractor selection		—	—	P2	P1	—	—
	Lack of document control		P3	P4	—	—	—	—
Human Resources	Lack of experience and personal expertise in design and construction		H1	H1	—	—	H2	H5
	Inadequate supervision staff		H4	—	—	—	—	—
	Inadequate manpower to complete the task		—	H4	—	—	—	H7
	Lack of skills		H3	H2	—	—	—	H1
	Poor knowledge of team member, lack of education and training		H2	H3	—	—	H1	H3
	Lack of employee motivation and rewards		H7	H6	—	—	—	H6
	Poor workmanship approach and inappropriate personal attitude		H6	H8	—	—	—	H2
	The absence of job security and other safety rules		H8	H7	-	-	-	-
	Labor reallocation, alteration and staff turnover		—	H5	—	—	—	H4
	Conflict of interests		H5	—	—	—	—	—
Material/Equipment	Defective materials, Non-adherence to material specifications		M1	—	—	—	M1	M2
	Poor-quality material or substandard products/ Prefabrication errors		—	—	—	—	—	M1
	Replacement or misplacement of material		—	—	—	—	—	M3
	Inefficient equipment use		—	—	—	—	—	M4
	Untimely deliveries of material and equipment		—	—	—	—	—	M5
Technical	Ineffective use of quality management practices/ deviation due to poor monitoring		T5	T4	—	—	T4	T1
	Poor technology application and lack of information technology use		T6	T3	—	—	T5	T5
	Poor communication system for coordinating between members		T1	T1	-	-	-	T4
	Inefficient management process, poor site management practice		T3	T5	-	-	T2	T3
	Poor project documents, such as poor contract documents		—	T6	T1	T2	—	T7
	Conflicting and incomplete information		T2	—	T2	T1	T3	T6
	Inadequate planning and poor scheduling of workload		T4	T2	—	—	T1	T2

(Continued on following page)

TABLE 8 (*Continued*) **List of rework causes in three stages of the project with liable contract parties.**

Project stages			Design		Procurement		Construction	
Contract parties			Client	Contractor	Client	Contractor	Client	Contractor
Group factor	Rework root causes	Code	DE	DS	PE	PS	CE	CS
General/External	Financial issues such as lack of funding, low contract or payment fee, delay in payment and cost pressure		G1	G2	G2	G1	G1	G4
	Lack of client involvement		G2	-	-	-	G2	-
	Unclear line of authority		G4		—	—	—	—
	Time pressure, schedule acceleration to finish the task, insufficient time to prepare contract documentation		G3	G1	G1	—	—	G3
	Lack of constructability		—	G3	—	—	G3	—
	Damage/defects/deviations in the product due to poor handling and safety considerations		—	—	—	—	—	G1
	Governmental regulations and policies		—	—	—	—	G4	—
	Environmental conditions		—	—	—	—	—	G2
	Unpredictable factors from different sources		G5	—	—	—	G5	G5

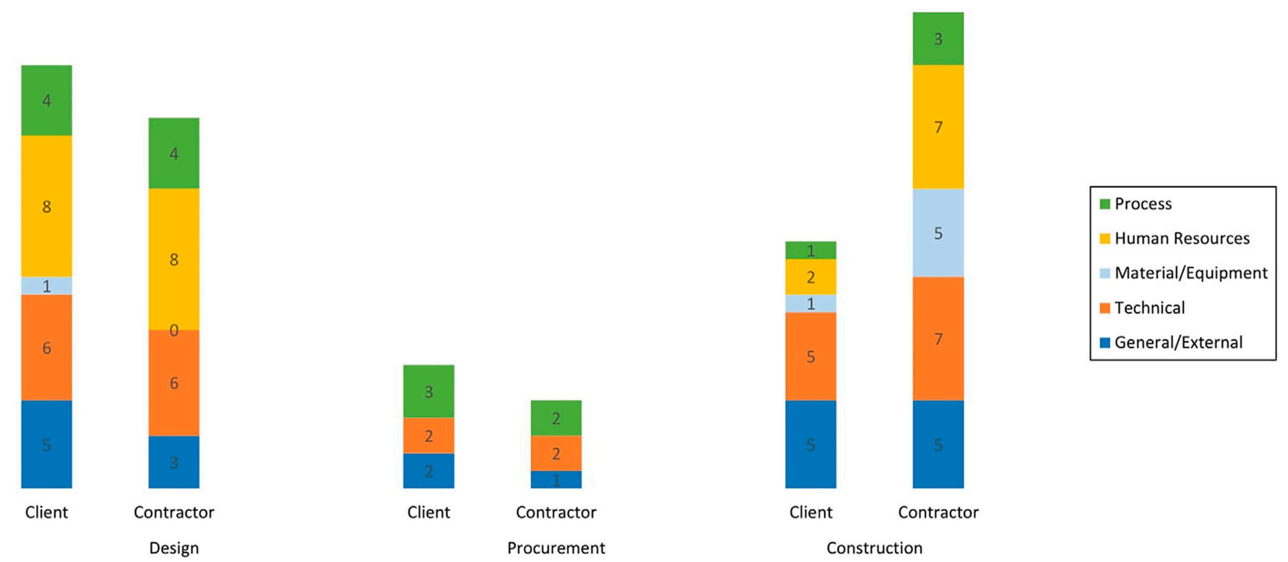


FIGURE 4
Distribution of rework factors in three stages of project.

recorded as one identified cause under the time related concerns. Tables 2–7 show classified rework causes based on collected data from previous studies, and Figure 4 has summarized the number of causes based on the proposed classification model. As shown in Figure 4, interest for studying causes of rework under the construction stage of a project was slightly higher than at the design stage. Both construction and design were constantly studied more than the procurement stage.

Following the proposed classification model, all identified causes of rework are clustered throughout the project stages. An overview of all identified rework causes in each project stage has been displayed in Table 8. This table comprises a list of 37 collected rework causes which are categorized into five primary constructs. The process related factor is involved with the projects' main activities. The human resources related factor deals with the human attributes. The material/equipment related factor supplies the project requirement. The technical related factor contributes to the supporting project activities, and the general/external related factor covers outsource items or causes that cannot be categorized under the four previously stated constructs. Referring to the nature of each construct, six causes were classified under process, 10 under human resources, five under material/equipment, seven under technical, and nine under the general/external category. Each category's underlying causes in the three project stages were placed under liable contract parties with the same reference code from Tables 2 to 7. Therefore, coding numbers in each cause does not show an in-order pattern. For example, in the process category, changes are listed first under the cluster of design-client with reference code of DEP1, while under the cluster of construction-contractor, errors are listed first with reference code of CSP1. The listing of rework causes in each category is based on the citation frequencies, and simultaneously it is assumed as an index for sorting the identified causes. Obviously, the higher number of citations does not indicate higher occurrence probability but depicts higher research interests.

In the design stage, the number of identified causes of rework in the client side of the contract is higher than contractor side and it is not surprising because the client involvement during design has been known as the primary contributor (Love et al., 2010b; Forcada et al., 2016; Mahamid, 2016; Yap et al., 2017; Eze et al., 2018b). The distribution of rework causes in both sides of the contract under this stage in Table 8 shows a total number of 28 out of 37 identified causes. Nearly all five categories under the cluster of client-design were cited as highly relevant, while the material/equipment related factor under the cluster of contractor-design seemed to be absent in previous studies.

Procurement is the other important stage that affects project performance thereby contributing to rework occurrence (Forcada et al., 2017). In the procurement stage, two categories of human resources and material/equipment were undeniably missed in previous studies. The reasons for lack of research in this area needs to be investigated, whether due to the selected procurement strategy by the projects or redundancy. Research interests for assessing rework causes under the procurement stage in Table 8 have identified eight items out of 37 causes. Most of them deal with contract management in different ways (Ye et al., 2015). Who accepts the responsibility of rework is defined fundamentally in the contract (Love et al., 2006). The lower number of studies on rework causes in this stage is probably linked to the defined delivery system of projects that mostly merges two stages of design and construction together. This was aggravated by improper implementation of procurement strategies that have not been adopted widely at the design stage (Salihu and Babarinde, 2020). Even though the procurement stage is listed with low citation frequencies, the underlying causes in both clusters under procurement are very critical in rework management (Al-jababi et al., 2020). More collaboration between contract parties results in early identification of rework causes and provides better solutions to prevent reoccurrence of rework (Taggart et al., 2014) at this stage.

Construction as the other stage of a project presents different results. However, it encompasses 30 identified causes over 37 listed items; distribution of the causes in both contract sides is not equally balanced. While the contractor has more contribution to the causes of rework, the other side takes half of the causes in terms of quantity. It is because the contractor plays the main role at this stage. However, the three categories of process, human resources, and material/equipment, present a small number of causes under the client's side of the contract, they become more predominant at the other side.

6 Conclusion

The study aimed to identify and classify rework root causes in the main stages of a project to facilitate a better understanding of rework with liable parties of the construction contracts. To achieve the study's aim, a comprehensive review of the literature on the sources of rework was conducted. Reduction of the rework impacts has received extensive attention over the years but rework nonetheless continues to exist (Hwang et al., 2009; Love et al., 2015) probably because the effects of rework have not been integrated in the main stages of a project (Xue et al., 2010). Although interest in reviewing the topic of rework has increased in recent years, a systematic review on rework root causes under three main stages of a project with liable sides of the contract remained undiscovered. Based on the structured method used in this study, a total of 157 papers were identified directly relevant to rework in construction projects. The aggregated publications then were systematically reviewed. The causes of rework were identified through a detailed analysis of selected publications from academic journals and conference. This paper presented a list of 37 rework root causes that have been integrated into a structured classification model. The proposed classification model was applicable to three stages of a project and two liable contract parties.

The summarized result in Table 8 is the main contribution of this study and it illustrates a diversity of rework root causes in a wider perspective. The proposed model shows a full picture of rework root causes in six clusters generated from three project stages and two contract sides. Thus, it alleviates the combination of different levels used in previous classified methods. The proposed model adds value to the existing knowledge as it shows the area of rework studies which has not been explored yet. Compared with previous methods, this model shows rework causes under each project stage linked to the liable contract parties. The result contributes to the theory by proposing a model and classified list of causes. This list will be used to investigate rework in the construction contracts that bridges the existing knowledge gap by addressing rework causes in the contract conditions. Thus, finding of this study contributes to the contract management body of knowledge. The classified rework causes from this study will be used as a platform for further investigation of the contract to improve contractual terms and conditions by addressing rework issues. Improving contract conditions of construction projects prevents rework issues and will result in fewer contractual claims and disputes. The classified causes of rework will also be used as a guideline for construction projects to regulate and adopt more reliable strategies on rework management. It displays the distribution of causes in both sides of the contract and offers opportunities for further practical research.

The topmost frequently cited rework causes in the reviewed papers include, but are not limited to, ineffective use of quality management practices, poor communication systems, inadequate construction planning, insufficient skills in both labour and supervisory levels, damage, defect or deviation of products due to carelessness and poor safety consideration, use of poor-quality material, and poor contract documentations. The comparison between categories reflected that all stages of a project had not been covered thoroughly with the identified causes. Rework at the construction stage includes the most frequent causes, whereas procurement shows fewer identified causes. However, the result of critical analysis has adequate evidence in each project stages, more focus on studying of the procurement stage is recommended. Since the proposed model needs validation in various contexts, the next step of research will investigate this purpose through a questionnaire survey that examines the effects of rework on contractual claims. Contract documents as the main output of the procurement stage define parties' authorities. Incorrect contract information is presumed to lead rework occurrence (Kakitahi et al., 2014). Searching the relationship between contract clauses and rework within various project stages would benefit future models of rework management by developing a framework that is capable to evaluate contracts in terms of identified rework causes (Mendis et al., 2015; Asadi et al., 2022). Identifying rework from a contract party's perspective will result in higher awareness of client and contractor by recognizing the root causes of rework at the time of contract preparation.

In this paper, the review is limited to the identified causes of rework and does not cover the other aspects of measuring rework impacts. Future research on the identified gaps is recommended to provide a framework for more practical experiments and develop rework management trends. The result inspires further investigations specifically on the categories that have received less attention previously, such as the following areas of study:

- The material/equipment related factor of both sides of the contract at the design stage and the client-side of the contract at the construction stage.
- Human resources and material/equipment related factors of both sides of the contract at the procurement stage.

This paper has some other limitations as the approach used was on the matters related to the causes of rework to be investigated in the contracts. Contracts are the outcome of procurement, and various procurement options in projects may result in different types of contract. Since the causes of rework appear in construction projects regardless of their procurement routes, the result of this study is not subject to a specific type of contract. Therefore, further investigation of rework causes under different procurement approaches is advised in future studies. This paper is also limited to the project's design, procurement, and construction stages. Investigating the other stages such as operation, maintenance, and demolishing, is another area for future studies to complete the study of rework in the project life cycle. The approach used for sorting the identified causes under each stage of the project was based on the number of references from previous studies to show the level of interest from previous researchers. This approach is considered as another limitation to this research but based on the research needs; it can be justified by the inapplicability of measuring the severity of causes in this study. Lastly, even though the result of this paper cannot be generalized for various types of projects; nevertheless, it supports the next stage of the research to investigate addressing rework issues in the construction contracts.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdul Nabi, M., El-adaway, I. H., Fayek, S., Howell, C., and Gambatese, J. (2020). Contractual guidelines for construction safety–related issues under design–build standard forms of contract. *J. Constr. Eng. Manag.* 146 (7), 04020074. doi:10.1061/(asce)co.1943-7862.0001855

Ahmed, S. A., and Naik, B. H. (2016). Rework management in construction projects and comparison with time and cost. *Int. J. Innovative Res. Sci. Eng. Technol.* 5 (10), 17973–17980.

Aiyetan, O. A. (2013). Causes of rework on building construction projects in Nigeria. *Interim Interdiscip. J.* 12 (3), 1–15.

Aiyetan, O. A., and Das, D. (2015). Using system dynamics modelling principles to resolve problems of rework in construction projects in Nigeria. *J. Constr. Proj. Manag. Innovation* 5 (2), 1266–1295.

Ajayi, O. M. (2017). Occurrence of rework on components of building project in Lagos state, Nigeria. *Int. J. Sustain. Constr. Eng. Technol.* 8 (1), 218084–324293.

Ajayi, O., and Oyeyipo, O. (2015). Effect of rework on project performance in building project in Nigeria. *Int. J. Eng. Res. Technol.* 4 (2), 294–300.

Al-Jabâbi, A. M., Abdel-Monem, M. S., and El-dash, K. M. (2020). Factors causing rework and their impact on projects' performance in Egypt. *J. Civ. Eng. Manag.* 26 (7), 666–689. doi:10.3846/jcem.2020.12916

Asadi, R., Rotimi, J. O. B., and Wilkinson, S. (2022). Investigating the relationship between reworks and contractual claims: The salience of contract conditions. *J. Leg. Aff. Dispute Resolut. Eng. Constr.* 14 (1), 04521046. doi:10.1061/(asce)la.1943-4170.0000519

Asadi, R., Wilkinson, S., and Rotimi, J. O. B. (2023). Guidelines for mapping rework in the contract conditions of construction projects journal of legal affairs and dispute resolution in engineering and construction. *J. Leg. Aff. Dispute Resolut. Eng. Constr.* 15 (1), 04522041. doi:10.1061/%28ASCE%29LA.1943-4170.0000563

Asadi, R., Wilkinson, S., and Rotimi, J. O. B. (2021a). The common causes of rework in construction contracts: A diagnostic approach. *J. Eng. Des. Technol.* doi:10.1108/JEDT-04-2021-0215

Asadi, R., Wilkinson, S., and Rotimi, J. O. B. (2021b). Towards contracting strategy usage for rework in construction projects: A comprehensive review. *Constr. Manag. Econ.* 39 (12), 953–971. doi:10.1080/01446193.2021.2004609

Assaad, R., Elsayegh, A., Ali, G., Abdul-Nabi, M., and El-adaway, I. H. (2020). Back-to-Back relationship under standard subcontract agreements: Comparative study. *J. Leg. Aff. Dispute Resolut. Eng. Constr.* 12 (3), 04520020. doi:10.1061/%28ASCE%29LA.1943-4170.0000406

Balouchi, M., Gholhaki, M., and Niousha, A. (2019). Reworks causes and related costs in construction: Case of parand mass housing project in Iran. *Int. J. Qual. Reliab. Manag.* 36 (8), 1392–1408. doi:10.1108/ijqrm-06-2018-0155

Bao, F., Chan, A. P. C., Chen, C., and Darko, A. (2018). Review of public-private partnership literature from a project lifecycle perspective. *J. Infrastructure Syst.* 24 (3), 04018008. doi:10.1061/(asce)is.1943-555x.0000424

Chan, A. P. C., Tetteh, M. O., and Nani, G. (2020). Drivers for international construction joint ventures adoption: A systematic literature review. *Int. J. Constr. Manag.* 22, 1571–1583. doi:10.1080/15623599.2020.1734417

Coleman, E., Nooni, I. K., Fianko, S. K., Dadzie, L., Neequaye, E. N., Owusu-Agyemang, J., et al. (2020). Assessing contract management as a strategic tool for achieving quality of work in Ghanaian construction industry. *J. Financial Manag. Prop. Constr.* 25 (1), 41–60. doi:10.1108/jfmpc-08-2018-0046

El-adaway, I., Fawzy, S., Allard, T., and Runnels, A. (2016). Change order provisions under national and international standard forms of contract. *J. Leg. Aff. Dispute Resolut. Eng. Constr.* 8 (3), 03716001. doi:10.1061/(asce)la.1943-4170.0000187

El-adaway, I. H., Fawzy, S., Burrell, H., and Akroush, N. (2017). Studying payment provisions under national and international standard forms of contracts. *J. Leg. Aff. Dispute Resolut. Eng. Constr.* 9, 33. doi:10.1061/(ASCE)LA.1943-4170.0000200

Enshassi, A., Sundermeier, M., and Abo Zeiter, M. (2017). Factors contributing to rework and their impact on construction projects performance. *J. Leg. Aff. Dispute Resolut. Eng. Constr. J. Sustain. Constr. Eng. Technol.* 8 12.

Eze, E. C., and Idiake, J. E. (2018). Analysis of cost of rework on time and cost performance of building construction projects in Abuja, Nigeria. *Int. J. Built Environ. Sustain.* 5 (1), 56–67.

Eze, E. C., Idiake, J. E., and Ganiyu, B. (2018a). Analysis of rework risk triggers in the Nigerian construction industry. *Organ. Technol. Manag. Constr.* 10, 1778–1793. doi:10.2478/otmcj-2018-0008

Eze, E. C., Idiake, J. E., and Ganiyu, B. O. (2018b). Rework risks triggers in the Nigerian construction industry a view of built environment professionals. *Indep. J. Manag. Prod.* 9 (2), 448–472. doi:10.14807/ijmp.v9i2.729

Forcada, N., Alvarez, A. P., Love, P. E. D., and Edwards, D. J. (2016). Rework in urban renewal projects in Colombia. *J. infrastructure Syst.* 2016, 1–19. doi:10.1061/%28ASCE%29IS.1943-555X.0000332

Forcada, N., Gangolells, M., Casals, M., and Macarulla, M. (2017). Factors affecting rework costs in construction. *J. Constr. Eng. Manag.* 143 (8), 04017032. doi:10.1061/(asce)co.1943-7862.0001324

Forcada, N., Rusinol, G., Macarulla, M., and Love, P. E. D. (2014). Rework in highway projects. *J. Civ. Eng. Manag.* 20 (4), 445–465. doi:10.3846/13923730.2014.893917

Habibi, M., and Kermanshachi, S. (2018). Phase-based analysis of key cost and schedule performance causes and preventive strategies, Research trends and implications. *Eng. Constr. Archit. Manag.* 25 (8), 1009–1033. doi:10.1108/ecam-10-2017-0219

Habibi, M., Kermanshachi, S., and Safapour, E. (2018). Engineering, procurement, and construction cost and schedule performance, leading indicators: State-of-the-Art review. *Proc. Constr. Res. Congr.* 2018, 378–388. doi:10.1061/9780784481271.037

Hossain, M. A., and Chua, D. K. H. (2013). Overlapping design and construction activities and an optimization approach to minimize rework. *Int. J. Proj. Manag.* 32, 983–994. doi:10.1016/j.ijproman.2013.10.019

Hwang, B. G., Thomas, S. R., Haas, C. T., and Caldas, C. H. (2009). Measuring the impact of rework on construction cost performance. *J. Constr. Eng. Manag.* 135 (3), 187–198. doi:10.1061/(asce)0733-9364(2009)135:3(187)

Hwang, B. G., and Yang, S. (2014). Rework and schedule performance. *Constr. Archit. Manag.* 21 (2), 190–205. doi:10.1108/ecam-10-2012-0101

Hwang, B. G., Zhao, X., and Goh, K. J. (2014). Investigating the client-related rework in building projects; the case of Singapore. *Int. J. Proj. Manag.* 32, 698–708. doi:10.1016/j.ijproman.2013.08.009

Hwang, B. G., Zhao, X., and Yang, K. W. (2019). Effect of BIM on rework in construction projects in Singapore status Quo, magnitude, impact, and strategies. *J. Constr. Eng. Manag.* 145 (2), 04018125. doi:10.1061/(asce)co.1943-7862.0001600

Josephson, P. E., Larsson, B., and Li, H. (2002). Illustrative benchmarking rework and rework costs in Swedish construction industry. *J. Manag. Eng.* 18 (2), 76–83. doi:10.1061/(asce)0742-597x(2002)18:2(76)

Kakitahi, J. M., Alinaitwe, H. M., Landin, A., and Rodrigue, M. J. (2014). A Comparison of construction related rework in Uganda and Mozambique. *J. Constr. Proj. Manag. Innovation* 4 (1), 770–781.

Kakitahi, J. M., Landin, a., and Alinaitwe, H. M. (2013). An exploratory study of rework causality in Uganda. *Constr. Innov.* 13 (3), 266–280. doi:10.1108/ci-nov-2011-0051

Ke, Y., Wang, S., Chan, A. P. C., and Cheung, E. (2009). Research trend of public-private partnership in construction journals. *J. Constr. Eng. Manag.* 135 (10), 1076–1086. doi:10.1061/(asce)0733-9364(2009)135:10(1076)

Kim, Y. J., and Skibniewski, M. J. (2020). Cash and claim: Data-based inverse relationships between liquidity and claims in the construction industry. *J. Leg. Aff. Dispute Resolut. Eng. Constr.* 12 (3), 0452002. doi:10.1061/(asce)la.1943-4170.0000402

Liu, Q., Ye, G., Feng, Y., Wang, C., and Peng, Y. (2020). Case based insights into rework costs of residential building projects in China. *Int. J. Constr. Manag.* 20 (4), 347–355. doi:10.1080/15623599.2018.1484856

Love, P. D. E., Ackermann, F., Teo, P., and Morrison, J. (2015). From individual to collective learning: A conceptual learning framework for enacting rework prevention. *J. Constr. Eng. Manag.* 141 (11), 05015009. doi:10.1061/(asce)co.1943-7862.0001013

Love, P. D. E., and Smith, J. (2018). Unpacking the ambiguity of rework in construction: Making sense of the literature. *Civ. Eng. Environ. Syst.* 35 (NOS 1-4), 180–203. doi:10.1080/10286608.2019.1577396

Love, P. E. D., and Edwards, D. J. (2013). Curbing rework in offshore projects systemic classification of risks with dialogue and narratives. *Struct. Infrastructure Eng.* 9 (11), 1118–1135. doi:10.1080/15732479.2012.667419

Love, P. E. D., and Edwards, D. J. (2004). Determinants of rework in building construction projects. *Eng. Constr. Archit. Manag.* 11 (4), 259–274. doi:10.1108/09699980410547612

Love, P. E. D., Edwards, D. J., Irani, Z., and Goh, Y. M. (2011). Dynamics of rework in complex offshore hydrocarbon projects. *J. Constr. Eng. Manag.* 137, 1060–1070. doi:10.1061/(asce)co.1943-7862.0000377

Love, P. E. D., Edwards, D. J., and Smith, J. (2006). Contract documentation and the incidence of rework in projects. *Archit. Eng. Des. Manag.* 1, 247–259. doi:10.1080/17452007.2005.9684596

Love, P. E. D., Edwards, D. J., Smith, J., and Walker, D. H. T. (2009). Divergence or congruence; A path model of rework for building and civil engineering projects. *J. Perform. Constr. Facil.* 23 (6), 480–488. doi:10.1061/(asce)cf.1943-5509.0000054

Love, P. E. D., Edwards, D. J., Watson, H., and Davis, P. (2010b). Rework in civil infrastructure projects determination of cost predictors. *J. Constr. Eng. Manag.* 136 (3), 275–282. doi:10.1061/(asce)co.1943-7862.0000136

Love, P. E. D., Holt, G. D., Shenb, L. Y., Li, H., and Irani, Z. (2002). Using systems dynamics to better understand change and rework in construction project management systems. *Int. J. Proj. Manag.* 20, 425–436. doi:10.1016/s0263-7863(01)00039-4

Love, P. E. D. (2002). Influence of project type and procurement method on rework costs in building construction projects. *J. Constr. Eng. Manag.* 128, 18–29. doi:10.1061/(asce)0733-9364(2002)128:1(18)

Love, P. E. D., Li, H., and Mandal, P. (1999). Rework a symptom of a dysfunctional supply chain. *Eur. J. Purch. Supply Manag.* 5, 1–11. doi:10.1016/s0969-7012(98)00017-3

Love, P. E. D., and Li, H. (2000). Quantifying the causes and costs of rework in construction. *Constr. Manag. Econ.* 18, 479–490. doi:10.1080/01446190050024897

Love, P. E. D., Mandal, P., and Li, H. (2010a). Determining the causal structure of rework influences in construction. *Constr. Manag. Econ.* 17 (4), 505–517. doi:10.1080/014461999371420

Love, P. E. D., Mandal, P., Smith, J., and Li, H. (2000). Modelling the dynamics of design error induced rework in construction. *Constr. Manag. Econ.* 18, 567–574. doi:10.1080/014461900407374

Love, P. E. D., and Sing, C. P. (2013). Determining the probability distribution of rework costs in construction and engineering projects. *Struct. Infrastructure Eng.* 9 (11), 1136–1148. doi:10.1080/15732479.2012.667420

Love, P. E. D., Smith, J., Ackermann, F., and Irani, Z. (2018). The praxis of stupidity an explanation to understand the barriers mitigating rework in construction. *Prod. Plan. Control Manag. Operations* 29 (13), 1112–1125. doi:10.1080/09537287.2018.1518551

Love, P. E. D., and Smith, J. (2003). Benchmarking, bench-action, and bench-learning rework mitigation in projects. *J. Manag. Eng.* 19 (4), 147–159. doi:10.1061/(asce)0742-597x(2003)19:4(147)

Ma, G., Jia, J., Zhu, T., and Jiang, S. (2019). A critical design structure method for project schedule development under rework risks. *Sustainability* 11, 7229. doi:10.3390/su11247229

Mahamid, I. (2016). Analysis of rework in residential building projects in Palestine. *Jordan J. Civ. Eng.* 10 (2), 197–208. doi:10.14525/jjce.10.1.3536

Mahamid, I. (2020). Impact of rework on material waste in building construction projects. *Int. J. Constr. Manag.* 22, 1500–1507. doi:10.1080/15623599.2020.1728607

Malek, J., and Desai, T. N. (2020). A systematic literature review to map literature focus of sustainable manufacturing. *J. clear Prod.* 256, 120345. doi:10.1016/j.jclepro.2020.120345

Mendis, D., Hewage, K. N., and Wrzesniewski, J. (2015). Contractual obligations analysis for construction waste management in Canada. *J. Civ. Eng. Manag.* 21 (7), 866–880. doi:10.3846/13923730.2014.893907

Miri, M., and Khaksefidi, M. (2015). Cost management in construction projects rework and its effects. *Mediterr. J. Soc. Sci.* 6 (6S6), 209–215. doi:10.5901/mjss.2015.v6n6s6p209

Ndwandwa, S., Simpeh, E. K., and Smallwood, J. J. (2017). "Factors influencing the occurrence of rework in construction," in Proceedings of international research conference 2017, sharping tomorrow's built environment, University of Salford, UK, 757–770.

Oyewobi, L. O., Falemu, O. A., and Ibironke, O. T. (2016). The impact of rework and organisational culture on project delivery. *J. Eng. Des. Technol.* 14 (2), 214–237. doi:10.1108/jedt-05-2013-0038

Oyewobi, L. O., and Ogunsemi, D. R. (2010). Factors influencing rework occurrence in construction; A study of selected building projects in Nigeria. *J. Build. Perform.* 1 (1), 1–20.

Palaneeswaran, E., Kumaraswamy, M. M., and Ng, S. T. (2005). A framework for monitoring rework in building projects. *Proc. Tall Build. Eng. Sustain.* XIII, 710–715.

Palaneeswaran, E., Love, P. E. D., Kumaraswamy, M. M., and Ng, T. S. T. (2008). Mapping rework causes and effects using artificial neural networks. *Build. Res. Inf.* 36 (5), 450–465. doi:10.1080/09613210802128269

Palaneeswaran, E., Love, P. E. D., Kumaraswamy, M. M., and S.T. Ng, T. (2014). Causal ascription of rework in building and civil engineering projects; A multivariate exploration. *Eng. Constr. Archit. Manag.* 21 (1), 111–126. doi:10.1108/ecam-04-2010-0029

Palaneeswaran, E. (2006). "Reducing rework to enhance project performance levels," in Proceedings of the One day Seminar on Recent Developments in Project Management, Hong Kong, 12 May 2006, 1–10.

Park, J., and Tucker, R. (2017). Overcoming barriers to the reuse of construction waste material in Australia: A review of the literature. *Int. J. Constr. Manag.* 17 (3), 228–237. doi:10.1080/15623599.2016.1192248

Robinson, F. A., Dissanayake, M., and Campero, O. (2004). Developing a standard methodology for measuring and classifying construction field rework. *Can. J. Civ. Eng.* 31, 1077–1089. doi:10.1139/l04-068

Safapour, E., and Kermanshachi, S. (2019). Identifying early indicators of manageable rework causes and selecting mitigating best practices for construction. *J. Manag. Eng.* 35 (2), 04018060. doi:10.1061/(asce)me.1943-5479.0000669

Safapour, E., Kermanshachi, S., and Taneja, P. (2019). "Investigation and analysis of the rework leading indicators in construction projects state-of-the-art review," in ASCE International Conference on Computing in Civil Engineering 2019, Atlanta, USA, June 17-19, 1–10.

Salihu, C., and Babarinde, S. A. (2020). Strategies for reduction of design-related rework in the Nigerian construction industry. *PM World J.* IX (II), 1–11.

Saseendran, A., Bigelow, B. F., Rybkowski, Z. K., and Jourdan, D. E. (2020). Disputes in construction: Evaluation of contractual effects of ConsensusDOCS. *J. Leg. Aff. Dispute Resolut. Eng. Constr.* 12 (2), 04520008. doi:10.1061/(asce)la.1943-4170.0000377

Shah, K. K., Shah, R. A., and Sharma, N. D. (2016). Development of field rework index to minimize the impact of rework for effective construction work. *Int. J. Sci. Dev. Res.* 1 (5), 370–375.

Siraj, N. B., and Robinson, F. A. (2019). Risk identification and common risks in construction: Literature review and content analysis. *J. Constr. Eng. Manag.* 145 (9), 03119004. doi:10.1061/(asce)co.1943-7862.0001685

Taggart, M., Koskela, L., and Rooke, J. (2014). The role of the supply chain in the elimination and reduction of construction rework and defects an action research approach. *Constr. Manag. Econ.* 32 (7–8), 829–842. doi:10.1080/01446193.2014.904965

Trach, R., Pawluk, K., and Lendo-Siwicka, M. (2019). Causes of rework in construction projects in Ukraine. *Archives Civ. Eng.* LXV (3), 61–74. doi:10.2478/ace-2019-0034

Wang, C., Yap, J. P. H., Wood, L. C., and Abdul-Rahman, H. (2019). Knowledge modelling for contract disputes and change control. *Prod. Plan. Control* 30 (8), 650–664. doi:10.1080/09537287.2019.1572247

Wang, T., Chan, A. P. C., He, Q., and Xu, J. (2020). Identifying the gaps in construction megaproject management research: A bibliographic analysis. *Int. J. Constr. Manag.* 15623599, 1585–1596. doi:10.1080/15623599.2020.1735610

Wilson, J. U., and Odesola, I. A. (2017). Design-related causes of rework and the performance of oil and gas projects in Nigeria. *Int. J. Sustain. Constr. Eng. Technol.* 8 (1), 60–76.

Xue, X., Shen, Q., and Ren, Z. (2010). Critical review of collaborative working in construction projects: Business environment and human behaviors. *J. Manag. Eng.* 26 (4), 196–208. doi:10.1061/(asce)me.1943-5479.0000025

Yap, J. B. H., Low, P. L., and Wang, C. (2017). Rework in Malaysian building construction impacts causes and potential solutions. *J. Eng. Des. Technol.* 15 (5), 591–618. doi:10.1108/jedt-01-2017-0002

Ye, G., Jin, Z., and Xia, B. (2015). Analysing causes for reworks in construction projects in China. *J. Manag. Eng.* 6, 04014097.

Yeo, K. T., and Ning, J. H. (2002). Integrating supply chain and critical chain concepts in engineer-procure-construct (EPC) projects. *Int. J. Proj. Manag.* 20, 253–262. doi:10.1016/s0263-7863(01)00021-7

Zhang, D., Haas, C. T., Goodrum, P. M., Caldas, C. H., and Granger, R. (2012). Construction small-projects rework reduction for capital facilities. *J. Constr. Eng. Manag.* 138 (12), 1377–1385. doi:10.1061/(asce)co.1943-7862.0000552

Zhang, S., Duan, H., Zhao, X., Xia, B., Feng, Y., and Galvin, S. (2018). Learning on rework management of construction projects a case study. *Int. J. Constr. Manag.* 1562-3599, 246–260. doi:10.1080/15623599.2018.1521361

TYPE Original Research
PUBLISHED 02 March 2023
DOI 10.3389/fbuil.2023.1119696

OPEN ACCESS

EDITED BY
Eugenia Gasparri,
The University of Sydney, Australia

REVIEWED BY
Marina Bonomolo,
University of Palermo, Italy
Piercarlo Romagnoni,
Università Iuav di Venezia, Italy

*CORRESPONDENCE
Nimish Biloria,
nimish.biloria@uts.edu.au

SPECIALTY SECTION
This article was submitted to Sustainable Design and Construction, a section of the journal Frontiers in Built Environment

RECEIVED 09 December 2022
ACCEPTED 02 February 2023
PUBLISHED 02 March 2023

CITATION
Biloria N, Makki M and Abdollahzadeh N (2023), Multi-performative façade systems: The case of real-time adaptive BIPV shading systems to enhance energy generation potential and visual comfort. *Front. Built Environ.* 9:1119696.
doi: 10.3389/fbuil.2023.1119696

Multi-performative façade systems: The case of real-time adaptive BIPV shading systems to enhance energy generation potential and visual comfort

Nimish Biloria[1,2]*, Mohammed Makki[1] and Nastaran Abdollahzadeh[1]

[1]Faculty of Design Architecture and Building, School of Architecture, University of Technology Sydney, Sydney, NSW, Australia, [2]Department of Urban and Regional Planning, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa

Building envelopes invariably tend to be static systems that encounter various performance limitations such as inefficient illuminance admittance, and heat and moisture transmission owing to their non-responsiveness towards environmental fluctuations. In contrast to such façade solutions, responsive façade systems with embedded sensing, actuation, and control systems have been proven to perform with up to 65% higher efficiency by being able to adapt their physical characters, such as orientation, and material property in real-time as a response to fluctuating environmental conditions (visual and thermal) and user preferences. Advancements in artificial intelligence and machine learning processes further aid such responsive façade systems to optimize multiple parameters such as illuminance level and the associated lighting energy, visual discomfort caused by solar glare, solar heat gain, thermal resistance (heating energy and comfort level), and natural ventilation simultaneously. This research investigates the case of a real-time adaptive Building Integrated Photo Voltaic (BIPV) shading system and its ability (in comparison with traditional static building integrated photo voltaic façade systems) to perform as regards visual comfort and energy generation potential simultaneously within the humid subtropical climate of Sydney, Australia. A simulated case scenario wherein a real-time adaptive building integrated photo voltaic shading systems is deployed on a typical multistorey building façade in Sydney, Australia is accordingly presented. The conducted simulation considers the responsive building integrated photo voltaic system as a double-skin façade system and uses multi-objective evolutionary computing principles to decipher its integrability potential. A comparative analysis between traditional static mounted Photo Voltaic (PV) systems as opposed to multi-objective optimization driven real-time adaptive building integrated photo voltaic shading configurations is subsequently presented. The ability to maximize generated energy, while simultaneously maintaining visual comfort is thus a unique proposition of this research.

KEYWORDS

real-time adaptive BIPV systems, double-skin facade, architecture, climate change, multi-performative design, multi-objective evolutionary optimization, resource efficiency, energy generation

1 Introduction

The building and the construction sector are responsible for 32% of the global energy consumption and 37% of greenhouse gas emissions, respectively. Residential and commercial buildings alone account for 30% of the final energy consumption (International Energy Agency and the United Nations Environment Program, 20121). Simultaneous, population growth and the exponential rate of urbanization further increases the demand for energy consumption, resulting in adverse environmental impacts (Yang et al., 2020). Dense urban patterns incorporating high-rise buildings as a means to cater to this population growth further increases the amount of energy consumed for heating and cooling purposes (Abdollahzadeh and Biloria, 2022). It is expected that the energy use and associated environmental pollutant emissions that impact public health are set to double or even tripled by the mid of the 21st century (International Energy Agency, 2013). Planetary health and environmental concerns have thus undoubtedly become common concerns globally (Yang et al., 2020), resulting in an urgency around the development of preventative mitigation measures to be considered for countering human-induced climate change (Secretariat, 2015). A paradigm shift embracing the large-scale implementation of renewable energy sources is thus quintessential.

In this context, apart from building new, energy efficient buildings, the potential for energy and Carbon dioxide (CO_2) mitigation (up to 50%–90%) of existing, non-energy efficient building stock (Lucon and Ürge-Vorsatz, 2014; Biloria and Abdullahzadeh, 2022) by adopting clever means of retrofitting cannot be undermined. In Australia, commercial buildings consume 10% of the national energy out of which 25% is consumed by office buildings (Council of Australian Governments, 2012; Ernst and Young, 2016). According to the International Energy Agency (IEA), global renewable power capacity shall increase by 50% in 2024 (since 2019) (IEA, 2019). Solar PV systems will constitute 60% (equivalent to 1200 Giga Watt) of this increased renewable power generation capacity (Poon et al., 2020). Energy retrofitting of existing buildings is thus deemed a potential solution to address this energy demand and BIPVs can be specifically used as a net-positive energy tool in this regard.

A buildings' façade is a crucial element that can impact energy consumption, presenting itself as an excellent element for building retrofitting. The façade acts as a connecting element between the indoor and the outdoor environment, thus affecting heating, cooling, and lighting conditions significantly. This in-turn influences the amount of energy consumed for satisfying comfort conditions in the interiors of buildings. Thermal fluxes through a building's envelope can further influence heating and cooling loads. One mitigation measure for such façade traits is a double skin façade system (DSF) with building integrated photo voltaic systems (BIPV). A double skin façade (DSF) can greatly improve the performance of building facades with respect to maintaining optimal thermal comfort and energy savings simultaneously. Such facade systems have the capability to control the amount of daylight received in interiors, solar heat gain, ventilation, and thermal exchange through a building's envelope (Lee et al., 2002).

Typical DSFs consist of an external glazed skin and a ventilated cavity with operable vents between the main envelope and the exterior skin facade. A high rate of ventilation through the cavity (open vents) in summer and minimum ventilation in winter (closed vents) increases thermal comfort and enhances a building's environmental performance. This ventilation it helps BIPVs in thermal washing: cooling of building elements in contact with the cavity thus increasing the efficiency of PV panels (Agathokleous and Kalogirou, 2016; Yang and Athienitis, 2016; Prieto et al., 2019). According to Agathokleous and Kalogirou (2016), DSFs should be able to provide a comfortable indoor climate, sound protection, use of daylighting while minimizing energy demand simultaneously. In this configuration, BIPV facade systems, not only generate extra electricity; but also protect the building's envelope against excessive heat transfer. PV installation on roofs and facades have the potential to exceed the local non-baseload demand and fulfill 50%–75% requirements of the total energy use (Karava et al., 2012).

The following sections further elaborates upon the role of BIPVs for climate conscious retrofitting.

2 Building integrated photovoltaics (BIPVs) as a climate conscious retrofitting solution

BIPV facade systems can be used as a retrofitting solution (Eicker et al., 2015; Evola and Margani, 2016; Hachem et al., 2014) to enhance the appearance and performance of existing buildings, resulting in material and energy savings simultaneously (Martín-Chivelet et al., 2018). According to Martin-Chivelet et al. (2018), PV panels can be easily integrated into the building envelope at a competitive cost (Delponte et al., 2015), and their characteristics can be coherent with the overall building's design in terms of architectural composition, colors, and textures (Martín-Chivelet et al., 2018). Additionally, their lightweight and flexible nature allow for easy integration into building facades (Yu et al., 2021). Also, developments in thin-film BIPV technology in terms of their efficiency and cost effectiveness have made their integration into novel architectural designs with complex geometry feasible (Jelle et al., 2012; Kaelin et al., 2004; Kushiya, 2014; Wilson, 2015). It can thus be argued that "high electrical efficiency, low cost, and the ease of installation are key to the wide acceptance and adoption of BIPV" systems (Yu et al., 2021), for addressing growing energy demands (Defaix et al., 2012; Raugei and Frankl, 2009). Local electricity generation using eliminate losses incurred during energy transportation, thus proving to be more cost-effective. BIPV as a double skin façade also offers acoustic benefits. Improving aesthetics and ease of installation are the other features that can be considered as advantages that BIPV elements offer, especially for retrofitting and renovation purposes (Biyik et al., 2017).

BIPV facade systems have been studied in different climatic contexts, including South Korea (Joe et al., 2013), Hong Kong (Peng et al., 2013; Peng et al., 2016), etc., however, very few studies have been conducted in the southern hemisphere, and subtropical or temperate climates. Within these limited studies, a study by Joe et al. (2013) suggests that a BIPV facade system reduces heating and cooling energy use by 16% and 7%, respectively (Joe et al., 2013). However, a later study, also indicates a 51% solar heat gain in

TABLE 1 Thermal properties of "naturally-ventilated BIPV/T-DSF" system (Australian Building Codes Board, 2016).

Parameters	Naturally ventilated BIPV/T-DSF	Requirement of the local regulation
U-value of the external wall (W/m2K)	0.51	Climate zone 1, 2, and 3: U ≥ 0.3
		Climate zone 4, 5, 6, and 7: U ≥ 0.36
		Climate zone 8: U ≥ 0.26
U-value of external roof (W/m2K)	0.316	Climate zone 1, 2, 3, 4, 5, and 6: U ≥ 0.31
		Climate zone 7: U ≥ 0.27
		Climate zone 8: U ≥ 0.21
Floor—a slab on the ground (boundary condition)	Adiabatic	N/A
U-value of internal window (W/m2K)	5.68	Climate zone 1, 2, and 3: 5.0 ≤ U ≤ 7.9
		Climate zone 4 and 5: 3.5 ≤ U ≤ 7.0
		Climate zone 6, 7, and 8: 3.0 ≤ U ≤ 6.0
Solar transmittance of internal window	83%	N/A
Solar reflectance of internal window	7.5%	N/A

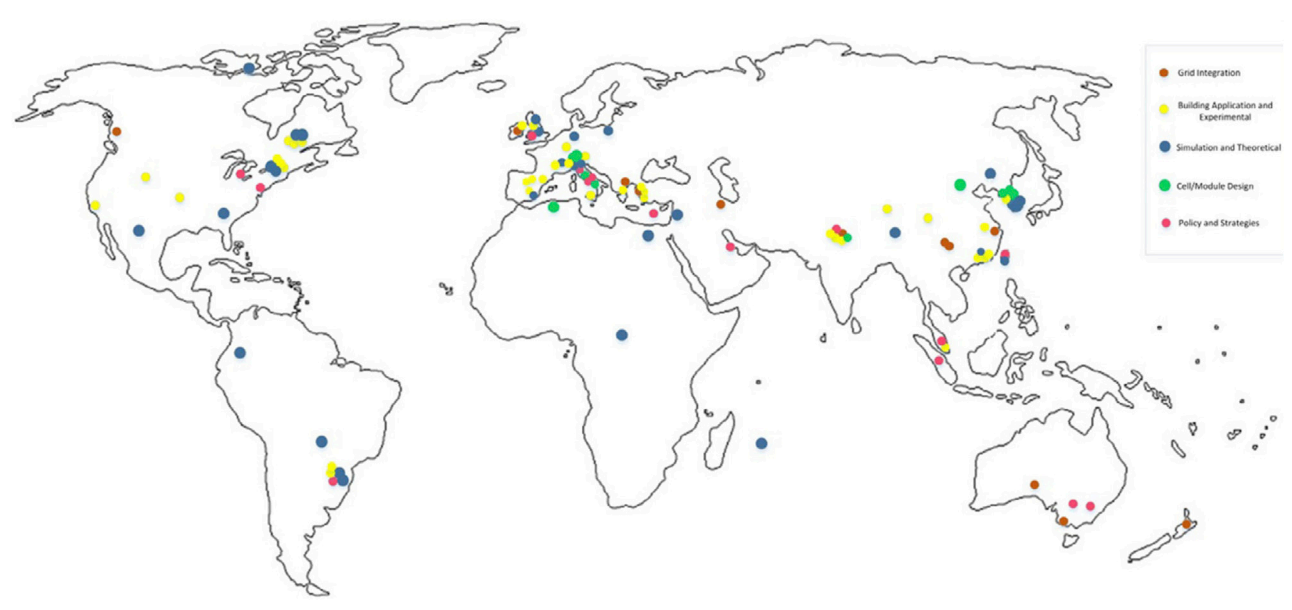


FIGURE 1
Worldwide studies on BIPV systems (Biyik et al., 2017).

summer and 32% heat loss in winters through such systems (Peng et al., 2013). Fossa et al. (2008) also investigated the thermal behavior of the BIPV/T-DSF system in Sydney (Australia) in a climatic context like this study (Fossa et al., 2008). Nevertheless, Fossa et al.'s study only reports on indoor environment tests (Yang et al., 2020). Australian building codes board also offers certain thermal properties for external envelope construction (naturally ventilated BIPV/T-DSF) to control heat loss and gain, which directly affects the energy demand of buildings (Australian Building Codes Board, 2016). These values are summarized in Table 1.

Furthermore, a comparative study on double skin facades with integrated BIPV technologies by Heidari Matin and Eydgahi, indicate that such efficient facade systems typically offer 30% improvement in visual comfort, and a 50% and 20% reduction in energy use values and carbon emissions, respectively. Moreover, an 80% reduction in thermal loads and a 25% reduction in cooling loads are reported (Heidari Matin and Eydgahi, 2019), which results in 15%–20% (cooling) cost efficiency respectively (Kolarevic and Parlac, 2015). Figure 1 indicates the studies conducted on BIPV worldwide (till 2017), according to a review conducted by Biyik et al., 2017. Figure 1 also categorizes the nature of conducted studies into five categories: Grid integration, Building application and experimental, Simulation and theoretical, Cell/Module design, and Policy and Strategies.

According to a review conducted by Agathokleous and Kalogirou (2016), to keep the BIPV temperature at a low level through natural ventilation, an optimal air gap of 10–15 cm between the PV panels and the building envelope (Agathokleous and Kalogirou, 2016) should be maintained. However, they also suggest that depending on the climatic condition and ventilation rate, the width of this air gap and the associated efficiency of the modules can be increased. In a study conducted by Martín-Chivelet et al., 2018, facade ventilation caused an annual 2.5% increase in the efficiency of PV panels as compared to a non-ventilated facade. An increase in temperature could however also result in a 3% decrease in the obtained power (Martín-Chivelet et al., 2018). A recent study by Goncalves et al. (2020) also employs the sensitivity analysis (SA) method along with computational simulation to explore the performance of naturally ventilated BIPV facade systems. The study reported exterior convective heat transfer coefficient to be the most influential factor impacting the performance of BIPV systems. The study also indicates that cavity ventilation can become more important when the exterior convective heat transfer decreases. Furthermore, solar irradiance is also found to be a vital factor affecting BIPV power outputs (Gonçalves et al., 2020).

2.1 The multifunctional use-case of BIPV façade systems

BIPV systems can be more cost-effective when they offer more than one function besides power generation. Some such functions include BIPVs acting as shading systems and promoting active solar heating and lighting (Agathokleous and Kalogirou, 2016; Farkas et al., 2013; Bonomo et al., 2015; Agathokleous and Kalogirou, 2016). According to Agathokleous and Kalogirou, thermal buffer zones, solar preheating of ventilated air, sound protection, wind and pollutant protection, night cooling and space for energy collection devices like PV cells are some of the use cases that can be considered when designing double skin facades as a cost-effective material substitute for buildings envelopes (Agathokleous and Kalogirou, 2016). Besides this, BIPVs can also act an insulation layer, a weather barrier, or a sun shading system (Gonçalves et al., 2020; Quesada et al., 2012).

Such multi-functionality of BIPV's is thus ideally suited for improving power generation as well as enhancing thermal comfort, natural lighting, heating, cooling, etc., (Yoo, 2011). They are also ideal for acting as an element of a double façade building envelope - serving as insulator during the night besides generating energy during the day. This use-case scenario is tested in a study by Yoo and Manz (2011) using SOLCEL, a photovoltaic system analysis program (Yoo and Manz, 2011). This computational method is also used in a later study of theirs with a focus on the position of the PV panels as shading devices to maximize power generation in the climatic context of the Suwon area, Korea. The findings of this study indicate that south/east (at 50°, building azimuth 130°) and south/west (at 50°, building azimuth 230°) offered a higher energy generation capacity compared to the exact south. A considerable impact of the PV's angle, building azimuth, and albedo on the power generation potential of BIPVs is thus reported in an evidence-based manner (Yoo, 2011). The use-case of BIPV window system design is, however, more sensitive as a minimum amount of transparency is required in windows, and PV films offer acceptable SHG value for this purpose. According to a study by Cannavale et al. (2017) in southern Italy, BIPV windows with shading systems have the potential to reduce the overall annual energy use by 18% compared to standard clear glass windows (Cannavale et al., 2017).

A study in the Australian contexts investigating the performance of building-integrated photovoltaic/thermal double-skin facade (BIPV/T-DSF) reported a total annual energy savings of 34.1%, 86%, and 106% annually could be attained in Darwin, Sydney, and Canberra, when compared with conventional technologies. This study explored different design parameters such as cavity depth of a double skin facade, rate, and mode of ventilation (natural vs. mechanical), thermal transmittance, and the opening ratio of ventilation louvers, to decipher impacts on heating and cooling loads and associated energy consumption of buildings through a sensitivity analysis. Although an additional energy saving is predicted for mechanically ventilated DSFs (lowest cooling energy use), non-ventilated DSFs offered the lowest heating energy use in the subtropical climate zone of Australia (Yang et al., 2020). A review conducted by Agathokleous and Kalogirou, suggest further research is vital to extrapolate advantages and effective usage of BIPVs (Agathokleous and Kalogirou, 2016).

A vital use-case for BIPV systems takes up the form of shading devices. The following section examines the application of BIPV's as facade shading systems when they are deployed in fixed, tilted, and adaptive configurations.

2.2 BIPV as shading systems

A study by Jayathissa et al. (2017) classifies previous research on BIPV shading systems in two categories: The effects of BIPV on building energy performance, and the integration of building energy performance simulation with shading systems (Jayathissa et al., 2017). PV integrated louvers used for shading result in significant energy savings, especially when used for cooling in hot climatic conditions (Palmero-Marrero and Oliveira, 2010). A tilted BIPV louver can typically generate 20%–40% more electricity than a flat vertical one (Freitas and Brito, 2015), and can reduce cooling demand by up to 51.6% (Sun et al., 2012).

Another shading system design in Denmark indicates the efficiency of dynamic shading systems over static/fixed ones (Nielsen et al., 2011). An adaptive shading system can control both solar heat gain and lighting conditions, thus improving a buildings' performance in terms of energy use, and thermal and visual comfort simultaneously. Such BIPV shading system can thus increase visual comfort, and prevent excessive solar radiation, thus reducing the need for excessive energy spending for cooling during summer while providing sufficient solar access during winter for passive heating (Yu et al., 2021). On a sunny winter and summer day, adaptive solar facades (ASFs) can compensate for 62% and 270% of the energy demand of buildings in the climatic context of Zurich, and Switzerland (Jayathissa et al., 2017). While in Los Angeles, the annual HVAC energy reduction is reported to be 30% using thin-film PV on glazed surfaces (Chae et al., 2014).

Such adaptive systems can also affect the thermal and optical condition of interiors. ASF's as dynamic systems continuously adapt to solar geometry to control direct and indirect radiation penetration into the buildings and thus reduce net HVAC loss (caused by less solar heat gain in colder climates (Chae et al., 2014), while increasing occupant comfort simultaneously (Loonen et al., 2013).

BIPV adaptive shading systems can thus supply the electricity required for heating, cooling, and lighting simultaneously (Jayathissa et al., 2017). A recent study by Yu et al. (2021) investigates three different categories of 'outdoor PV blinds, indoor PV blinds and middle PV blinds' (named based on the position of blinds relative to the windows). The study concluded that BIPV blinds are easier to adjust compared to BIPV windows and are thus more efficient in terms of energy generation and solar heat gain. Additionally, double-glazed BIPV shading blinds could perform better in winters compared to double glazing (semi-transparent façade) due to a higher solar heat gain coefficient (SHGC) and lower U-value (Yu et al., 2021). Another study on movable BIPV sun-shading systems installed on windows indicates that the thermal load of buildings can be reduced by up to 16% while electricity generation can be increased by 70% as compared to a fixed BIPV system as a secondary skin facade (Paydar, 2020).

Different studies are conducted on outdoor (Bahr, 2014; Park et al., 2016; Vadiee et al., 2016; Gao et al., 2018; Taveres-Cachat et al., 2019; Paydar, 2020), middle (Kang et al., 2012; Luo et al., 2018; Luo et al., 2017; Hong et al., 2017; Jeong et al., 2017; Koo et al., 2017), and indoor (Davidsson et al., 2010; Davidsson et al., 2012) BIPV shading blinds. Results of these studies suggested that middle PV blinds have the potential of energy savings up to 12.2% and 25.6% compared to traditional double-glazed windows or windows without blinds (Luo et al., 2017). The application of outdoor shading blinds is however limited due to their high costs. However, where external shadings are appropriate to install, outdoor PV blinds would also exhibit potential benefits (Yu et al., 2021). Jayathissa et al. (2017), conducted a similar study on dynamic photovoltaic systems for adaptive shading purposes to optimize the orientation of PV shading panels for maximum energy generation and minimizing heating, cooling, and lighting demand in office interiors. The finding indicates a 20%–80% energy saving (compared to static PV shadings), and that a 90° and 15°–45° (to the vertical plane) angle are required for the adaptive solar facade (ASF) to perform most efficiently. Moreover, PV generated energy supply can compensate for the annual energy demand by up to 95%, if an efficient HVAC system is installed simultaneously (Coefficient of performance: 6) (Jayathissa et al., 2017).

Considering such scientific evidence, and increasing climate emergencies, this research further elaborates upon a case study undertaken by the authors. Dewidar et al claim that responsive façade systems have been proven to perform with up to 65% higher efficiency (Dewidar et al., 2010) by being able to adapt their physical characters, such as orientation, and material property in real-time as a response to fluctuating environmental conditions (visual and thermal) and user preferences. The study pertains to a real-time adaptive BIPV shading system installed as a double skin facade on an educational institute building in Sydney, Australia. The study explores the real-time adaptive nature of a BIPV double skin set-up as a plausible energy retrofit solution within the subtropical climate of Sydney. The retrofit solution is seen from a multi-performative perspective wherein its ability to maximize generated energy, while simultaneously increasing visual comfort is put to the test. The real-time adaptive nature of the proposed system implies the BIPV panels to augment their physical position in real-time based on the sun angle throughout the day. A multi-objective optimization driven computational process is deployed for simulating such adaptivity of the BIPV panels. A comparison between a static BIPV panel system vs. the proposed real-time adaptive panel system is subsequently conducted to reveal the advantages of the proposed system, thus adding a novel unexplored dimension of real-time adaptive BIPV systems to the knowledge base of BIPV façades.

3 Methodology

3.1 Multi-objective evolutionary algorithm for real-time adaptive BIPV systems

Multi-objective evolutionary algorithms mimic the natural cycle of an evolutionary process in which a base phenotype (the geometry that the algorithm will act on) is developed through a defined set of variables (genes). A MOEA incrementally "mutates" (varies) the genes that define the phenotype to create a generation of solutions; at the end of each generation, the algorithm evaluates the evolved solutions and retains the "successful" solutions and discards the "unsuccessful" solutions; this process forms the fundamental workflow of all evolutionary algorithms (De Jong, 2006). In the algorithmic process, "success" is defined by a numeric fitness function that each solution is tested against; if the applied genetic mutation results in an improvement of a solution's fitness function, it is retained, and if the mutation results in a worsening of the solution's fitness function, it is discarded. The genes of the retained solutions cross over with one another to create the next-generation; as more and more "fit" solutions are selected and cross over with each other, each subsequent generation evolved by the algorithm is comprised from solutions with higher fitness values.

One of the main advantages of an MOEA is that the designer can integrate multiple conflicting fitness functions to evaluate each solution simultaneously, thus allowing the algorithm to evolve a population of solutions that have been independently optimized to the different fitness functions (through the continuous minor improvement of solutions through incremental mutations), consequently generating a varied population of optimized phenotypes. MOEAs have been used extensively across multiple disciplines since the mid-20th century; with a sharp increase in their use within Architecture and Design in the past decade due to the proliferation of various MOEA tools within mainstream 3d modeling software (Showkatbakhsh and Mohammed, 2022).

Considering the aforementioned advantages, a MOEA is thus employed in this study to optimize the BIPV configurations in response to simultaneously satisfy three objectives: To maximize irradiance value on the BIPV, minimize internal illuminance values above 3000 lux, and minimize internal illuminance levels below 300 lux. The use of an MOEA allows for the optimization of conflicting objectives, which in the case study presented herein is

FIGURE 2
UTS Tower (Wpcpey, 2017), located in the city of Sydney is used as the base model for the presented experiment.

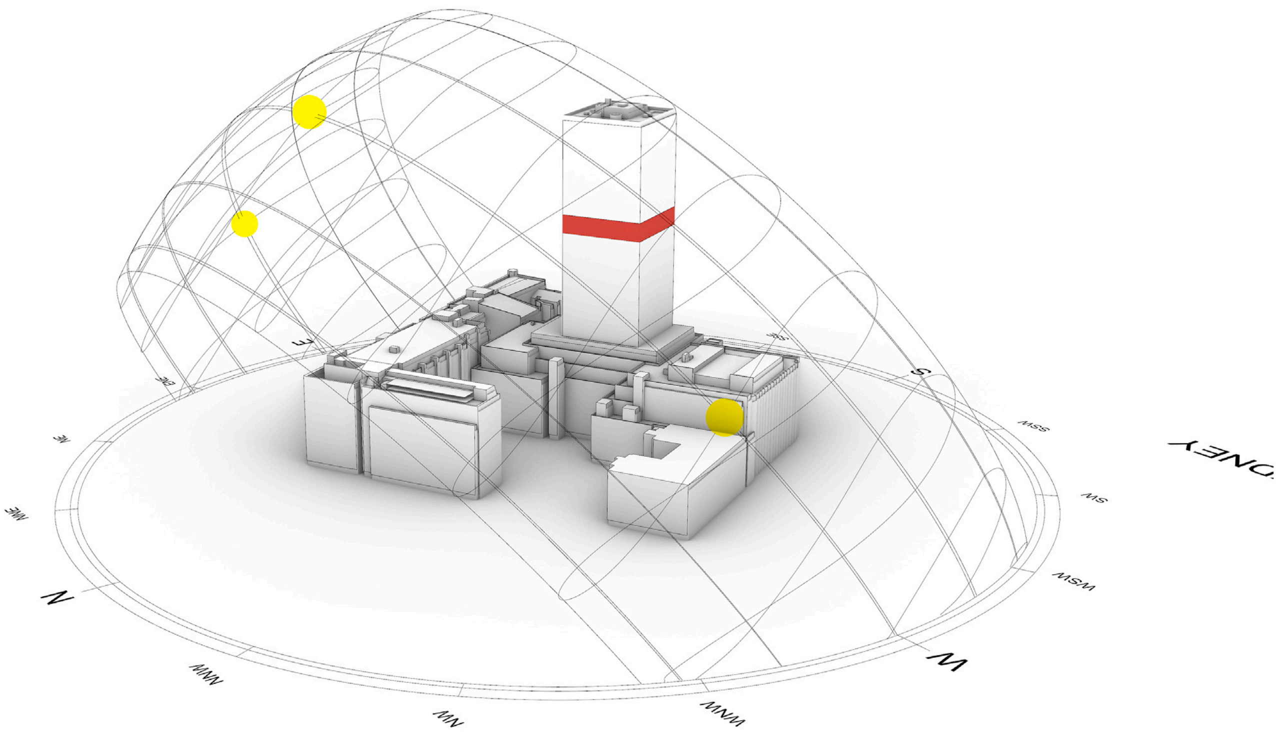


FIGURE 3
Solar analysis conducted for three different times of the day: March equinox (21st of March): 8:00am, 12:00pm and 4:00pm.

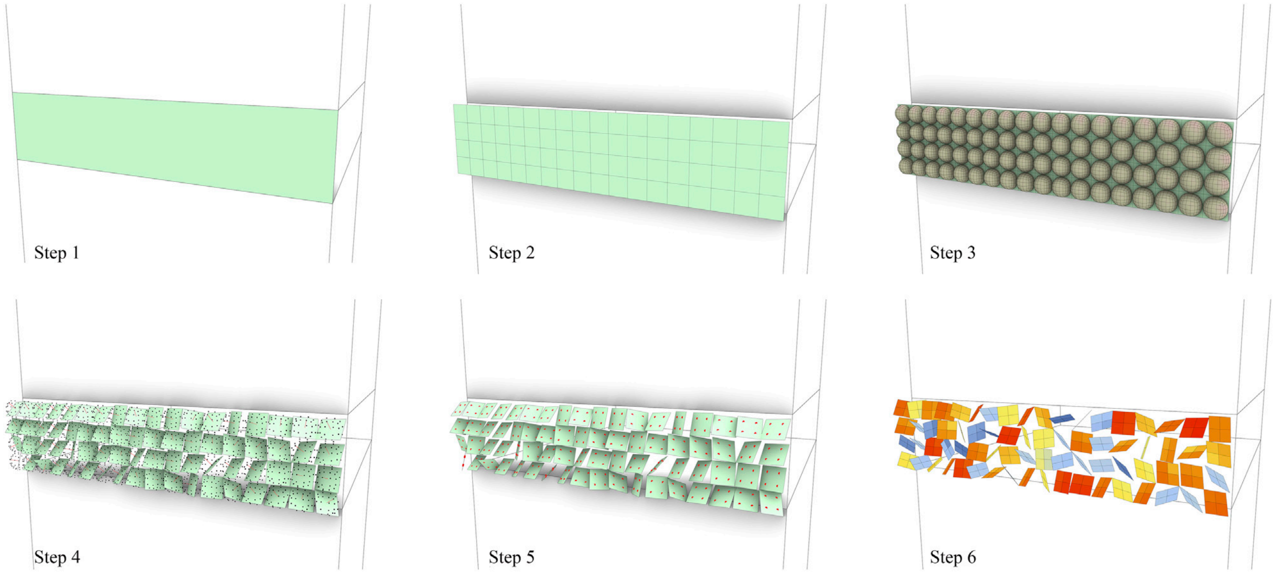


FIGURE 4
The step-by-step construction of the photovoltaic panels in the primitive phenotype.

critical to ensure the efficiency of the BIPVs is maximized with minimal intervention on the visual discomfort of the internal spaces, since the BIPV system is supposed to act as an energy generator and a shading device simultaneously. Moreover, due to the exponential number of possible configurations of the BIPV, utilizing an MOEA avoids the need to identify, model, and evaluate every possible configuration manually. This simultaneous multi-optimization of three objectives in an automated manner to develop real-time adaptive BIPV systems is thus seen as a novel contribution of this research.

3.2 Case study setup

The presented experiment selects a typical high-rise building in the city of Sydney: in this case, the University of Technology Sydney, Australia's Tower (Figure 2), and extracts two levels from the tower to run the MOEA. Selecting a section of the tower rather than the entire tower was primarily owing to the excessive runtime of the algorithm and the computational load associated with calculating irradiance. For the purpose of the presented study, conducting the analysis on a section of the tower is deemed sufficient as it allows for a comparative analysis between the algorithmic results and current approaches. The study presents three simulations, each optimizing the adaptive system for a different time of day during the March equinox (21st of March); 8:00 a.m., 12:00 p.m., and 4:00 p.m. (Figure 3).

The base phenotype for the experiment consists of 72 photovoltaic panels, each one approximately 2 m × 2 m in size, distributed across two levels of the tower's northern facade. A single chromosome, comprising 72 unique genes, controls the orientation of each panel, with a freedom of movement of 180° (Figure 4). The phenotype is optimized against three fitness functions; the first maximizes the irradiance value on the photovoltaic cells, where each cell is divided into 4 sample points for the irradiance calculation. The other two objectives evaluate interior illuminance levels out of Useful Daylight Illuminance (UDI Autonomous) bonds (300–3000 lux), as this range provides desirable visual conditions (A. Chi et al., 2018): The second fitness function minimizes the percentage of the internal space (in the two selected levels) with an illuminance value above 3000 lux (as a lux value above this level is not recommended for internal spaces and increases glare probability), the third fitness function minimizes the percentage of internal space with an illuminance value below 300 lux. For the second and third fitness functions, the internal space (Figure 5A) is divided into a 1.2m grid (Figure 5B) for calculating the illuminance levels (Figure 5C).

The computational process involves two steps: First, all three simulations are conducted using the software Wallacei (Makki et al., 2018) inside the Rhino 3D and Grasshopper 3D ecosystem, which employs a non-dominated sorting genetic algorithm (NSGA-2) algorithm (Deb et al., 2000), Second, solar analyses of irradiance and illuminance utilizes the software Honeybee, also situated within the same ecosystem (Roudsari and Mackey, 2012). The simulation settings and algorithm properties are detailed in Table 2.

4 Results

Each simulation evolved 4900 solutions across 140 generations. As can be observed in the charts in Figure 6, which present the algorithm's progress in the optimization process for each fitness function separately, the 8 a.m., simulation was successful in optimizing for both maximizing irradiance on panels and minimizing illuminance values of the internal space below 300 lux. Due to the position of the solar vector at this time of the day, it was not possible for illuminance values to go over 3000 lux and thus that fitness function remained at 0. The results for the 12 p.m. simulation and 4 p.m. simulation share similar patterns of behavior, both of which differ from the 8 a.m. simulation. In these two simulations, the fitness functions for maximizing irradiance on panels and minimizing illuminance above 3000 lux showed clear

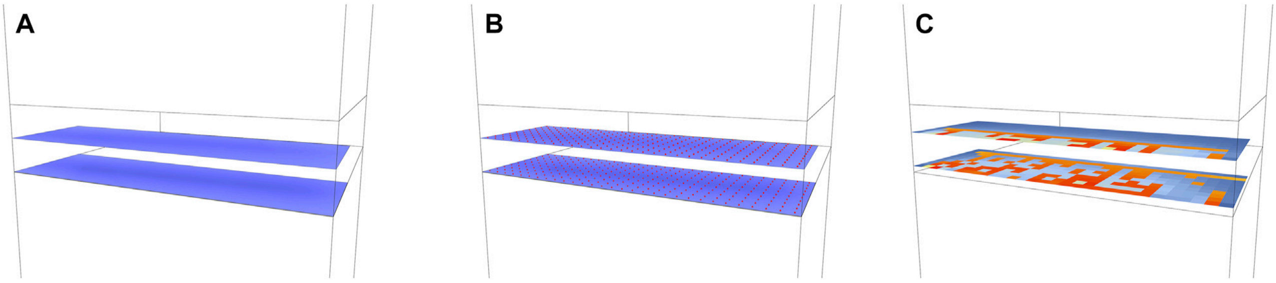


FIGURE 5
The step-by-step construction of the internal space in the primitive phenotype.

TABLE 2 The simulation and algorithm settings of the MOEA.

Weather data		Sydney (33.8688°S, 151.2093°E) EPW file	
Simulation tool		Ladybug, Honeybee V. 1.5	
Analysis tool		Wallacei	
Simulation time		A typical summer day (21.06) at 8 a.m., 12 p.m., 4 p.m.	
Analysis grid		1.2*1.2 m^2	
Material reflectance and transmittance		Facade: 0.35, PV: 0.05, glazing: 0.88	
Simulation size		Algorithm settings	
Generation size/count	35/140	Mutation rate	1/n (*n* = no. of var.)
Population size	4900	Crossover probability	0.9
No. of chromosomes	1	Mutation distribution index	20
No. of variables	6264	Crossover distribution index	20
Size of search space	4.4×10^{139}	Algorithm runtime	112 h., 20 min, 50 s

indication of convergence towards an optimal solution, (with the 12 p.m. simulation demonstrating better convergence than the 4 p.m. simulation), however due to the conflicting nature of the fitness functions, in both simulations, the algorithm struggled to converge the third fitness function that minimizes illuminance below 300 lux.

This conflict indicates that the algorithm has found it more efficient to optimize for irradiance on solar panels and minimize illuminance above 3000 lux than to minimize illuminance below 300 lux. It is important to note that despite the algorithm's inability to converge the third fitness function towards an optimal solution, a further analysis of the results showcase successful results for this fitness function.

Due to the conflicting nature of the fitness functions being optimized for in each simulation, there is no single optimal solution that is generated by the algorithm, as what is optimal for one fitness function, may not be for another. Instead of outputting a single optimal solution, the algorithm outputs a solution set that forms part of the "Pareto Front." The solutions in the Pareto Front are the "best" solutions evolved by the algorithm, in which no solution can be improved without making another solution on the Pareto Front worse. In each of the three simulations (8 a.m., 12 p.m., 4 p.m.), the Pareto Front consisted of 14 solutions, 46 solutions, and 85 solutions respectively. To capture the variation of solutions across the Pareto Front, two types of solutions were selected from each simulation for further analysis; the outlier solutions, which are the most optimal solutions for each fitness function separately; and the "Utopia" solution, which is the solution that is closest to the 'ideal' solution (defined as the solution that is impossible to achieve due to the conflict between the fitness functions) (Showkatbakhsh and Mohammed, 2022) (Figure 7).

5 Discussion

This research revolves through three main objectives using a BIPV shading system:

- Minimum over lit spots with an illuminance above 3000 lux.
- Minimum underlit spots with an illuminance under 300 lux to increase visual comfort.
- Maximum irradiation (W/m2) on PV cells to increase energy generation potentials.

Accordingly, a MOEA is employed to represent the most efficient design considering either one of the objectives (the outlier solutions) or all together (Utopia solution). Figure 8

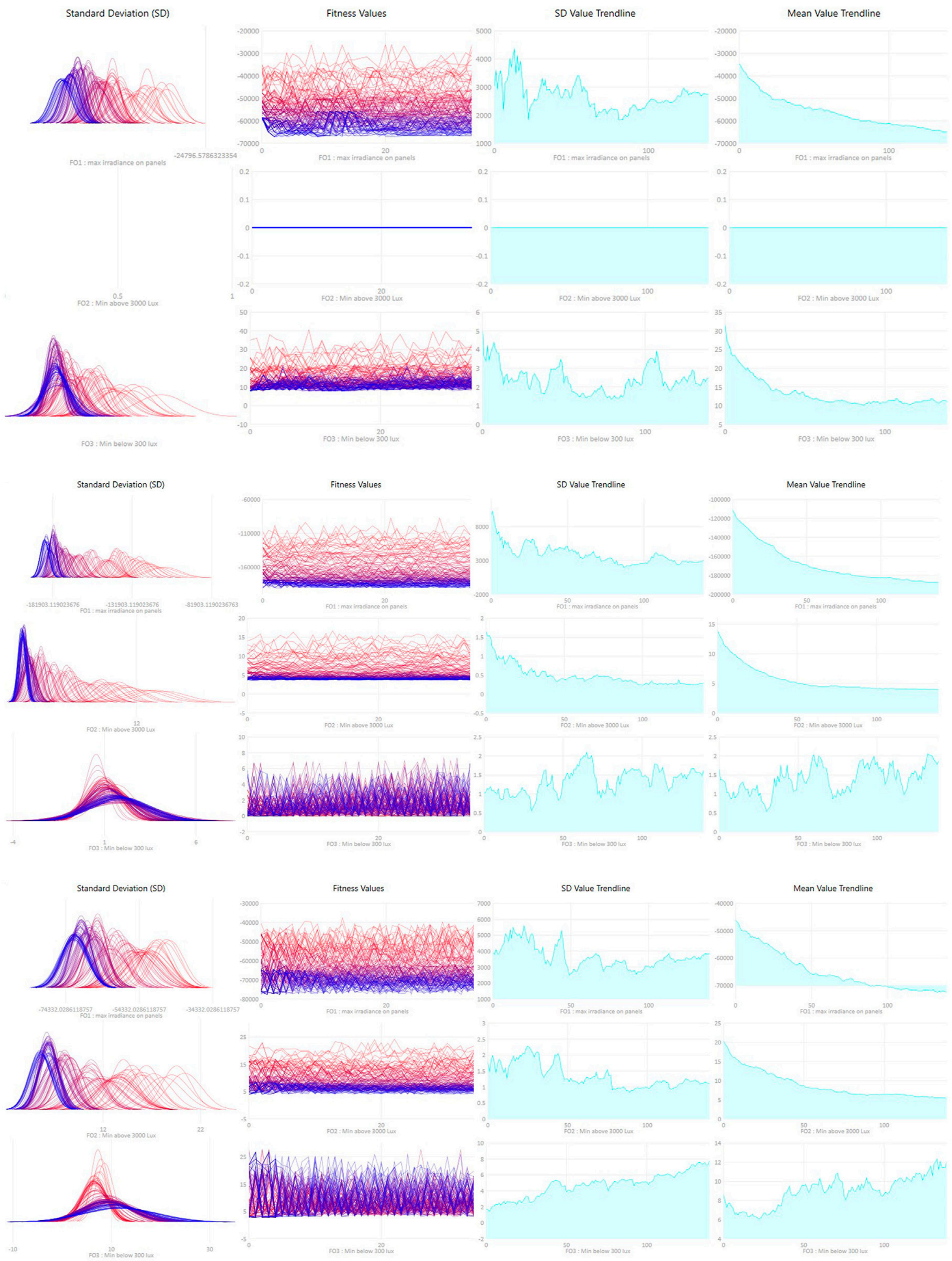


FIGURE 6
The results of the MOEA. Each set of charts corresponds to a different simulation: from top to bottom, 8 a.m., 12 p.m., and 4 p.m. The charts present the fitness values for every solution in the population through different graphical analyses. From left to right: Standard Deviation chart presents the variation and mean of each generation (the red to blue color scheme represents the first to last generation). The second chart is the fitness values chart, which presents each solution evolved by the population; each line on this chart represents a generation and each point on the line represents a solution in that generation. Finally, the third and fourth charts represent the variation and average of each generation respectively across all 140 generations.

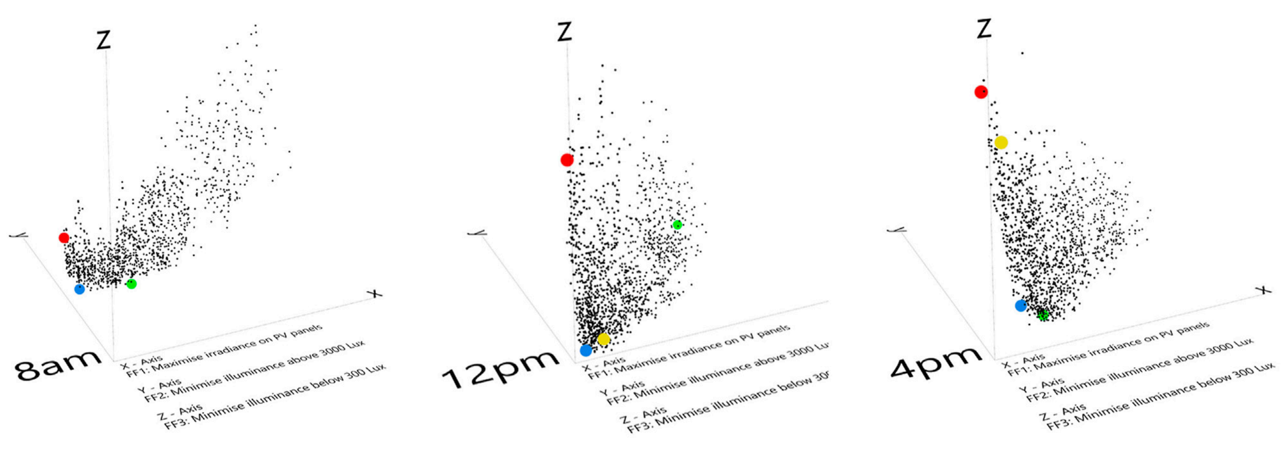


FIGURE 7
The solutions of each simulation distributed in the objective space to present the relationship between solutions and spread of fitness values across the entire population. The outlier solutions are highlighted (red highlights most optimal solution for irradiance on PV panels, green highlights the most optimal solution for minimizing lux below 300, and yellow highlights the most optimal solution for minimizing lux above 3000) as well as the Utopia solutions (the solution closest to the 'ideal' point, which in the presented charts is located on the 0,0,0 point of the graph. The Utopia point is highlighted in blue).

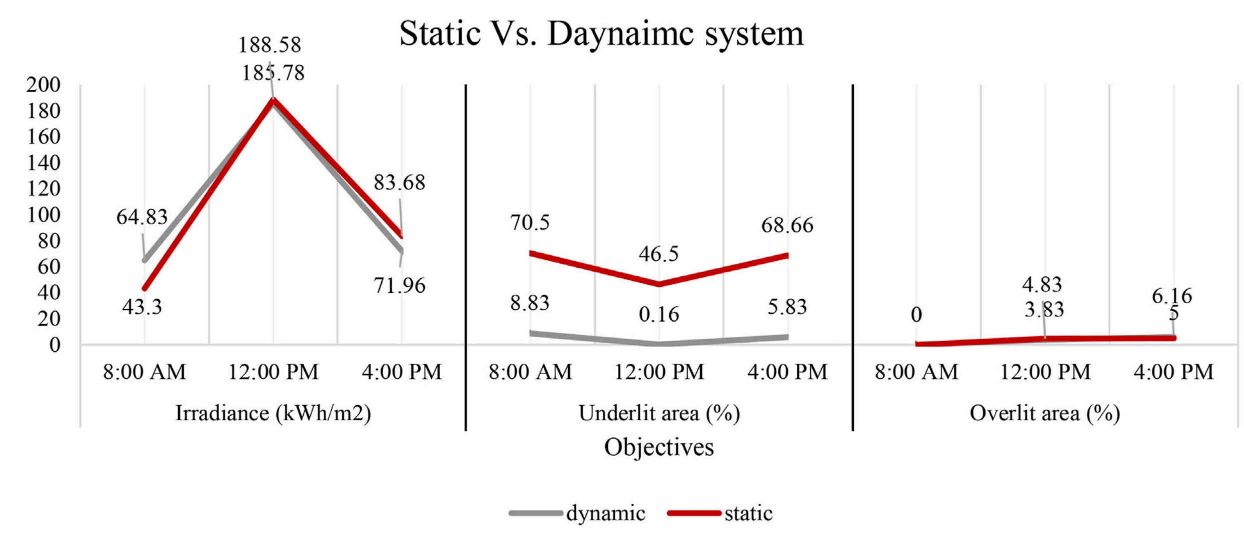
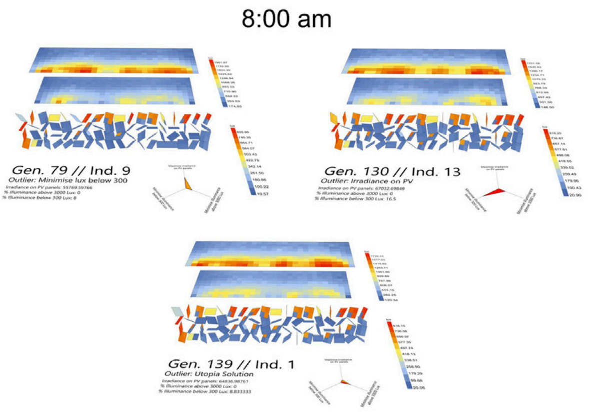


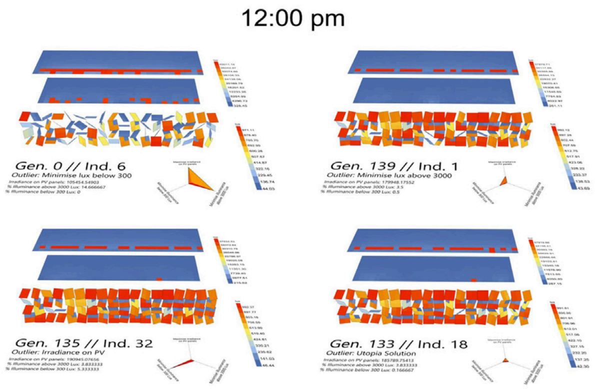
FIGURE 8
A comparison of static and Dynamic BIPV systems' performance.

presents the selected phenotypes from each simulation (the outliers and Utopia solution) with each solution's fitness values and the relationship between the fitness values as visualized through the diamond chart under each phenotype. As can be observed (Table 3), the Utopia solution identifies the solution that most equally optimizes for all fitness functions, while the outlier solutions prioritize the optimization of their respective fitness function, thus highlighting the conflict to the remaining fitness functions. The comparative analysis between the outliers and Utopia solution is vital in the presented study; although the natural progression towards selection is to choose the Utopia solution (as its the solution that most evenly optimizes for all fitness functions), the problem at hand presents a strong argument to utilize all the outlier solutions as they present valid configurations for the photovoltaics that respond to a specific time of day and a specific activity taking place inside the space being analyzed. In the case of the presented results, in each of the three simulations, the outlier that favors irradiance on the PVs is the most suitable panel configuration in the hours that the internal space is not inhabited (for example, before or after working hours or on weekends), as there is minimal need to ensure specific lux values are maintained during these uninhabited hours. Whereas in instances where the space is inhabited, the Utopia solution is more suitable as it evenly balances between irradiance on PVs as well as lux values in the interior spaces. This adaptability and reconfiguration of PVs throughout the day, an affordance provided by an alternative approach to static PV panels as facade systems,

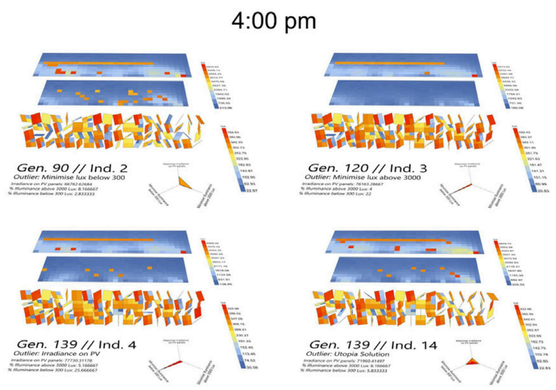
TABLE 3 The Outlier and Utopia solution and their associated fitness values for each simulation.



8:00 am	Outlier 1 (Best Solution for lux below 300)	
	Solution Location	Gen 79 / Ind 9
	Irradiance	55.76 kWh/m2
	% below 300 lux	8.00
	% above 3000 lux	0.00
	Outlier 2 (Best Solution for Irradiance)	
	Solution Location	Gen 130 / Ind 13
	Irradiance	67.03 kWh/m2
	% below 300 lux	16.50
	% above 3000 lux	0.00
	Utopia Solution	
	Solution Location	Gen 139 / Ind 1
	Irradiance	64.83 kWh/m2
	% below 300 lux	8.83
	% above 3000 lux	0.00



12:00 pm	Outlier 1 (Best Solution for lux below 300)	
	Solution Location	Gen 0 / Ind 6
	Irradiance	105.45 kWh/m2
	% below 300 lux	0.00
	% above 3000 lux	14.67
	Outlier 2 (Best Solution for lux above 3000)	
	Solution Location	Gen 139 / Ind 1
	Irradiance	179.94 kWh/m2
	% below 300 lux	0.50
	% above 3000 lux	3.50
	Outlier 3 (Best Solution for Irradiance)	
	Solution Location	Gen 135 / Ind 32
	Irradiance	190.94 kWh/m2
	% below 300 lux	5.33
	% above 3000 lux	3.83
	Utopia Solution	
	Solution Location	Gen 133 / Ind 18
	Irradiance	185.78 kWh/m2
	% below 300 lux	0.17
	% above 3000 lux	3.83



4:00 pm	Outlier 1 (Best Solution for lux below 300)	
	Solution Location	Gen 90/ Ind 2
	Irradiance	66.76 kWh/m2
	% below 300 lux	2.83
	% above 3000 lux	8.17
	Outlier 2 (Best Solution for lux above 3000)	
	Solution Location	Gen 120 / Ind 3
	Irradiance	76.16 kWh/m2
	% below 300 lux	22.00
	% above 3000 lux	4.00
	Outlier 3 (Best Solution for Irradiance)	
	Solution Location	Gen 139 / Ind 4
	Irradiance	77.73 kWh/m2
	% below 300 lux	25.67
	% above 3000 lux	5.17
	Utopia Solution	
	Solution Location	Gen 139 / Ind 14
	Irradiance	71.96 kWh/m2
	% below 300 lux	5.83
	% above 3000 lux	6.17

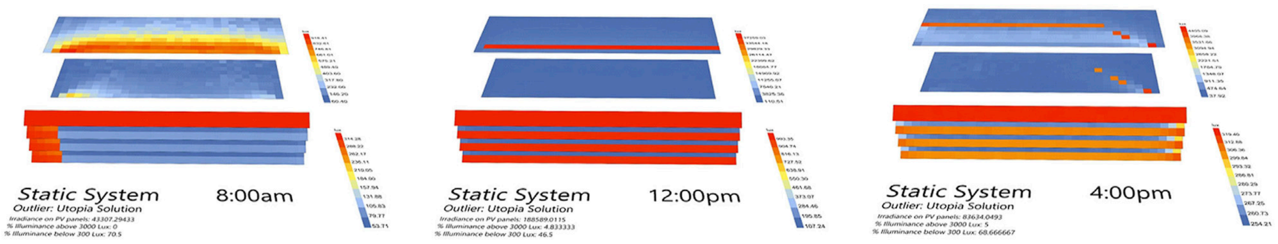


FIGURE 9
Static BIPV shading system performance (indoor illuminance and outdoor irradiation).

coupled with multi-objective optimization algorithms, is critical to increasing efficiency of solar exposure throughout different times of the day.

Meanwhile, a 60-degree tilted static system showed a weak performance, especially with regards to reducing underlit spots with an illuminance lower than 300 lux. In other words, as shown in Figure 8, an average daily of 56.94% difference in the minimum threshold level of illuminance is observed between static and dynamic BIPV shading system. This is while energy generation potentials in dynamic system reaches to its maximum (increasing 21.53%) in the morning (at 8 a.m.) with an average daily improvement percentage of 2.35%, compared to the static one. The spatial percentage of area with an illuminance over 3000 lux, which increases visual discomfort and glare probability is also alleviated by up to 1.16% in the dynamic system.

The spatial distribution of underlit and overlit interior spots and irradiation on fixed PV panels for different times of the day (8 a.m., 12 p.m., 4 p.m.) is shown in Figure 9.

Fixed static shading systems, not only have shown a lower performance regarding both energy generation potentials and visual comfort but also can block outward views, which directly affects occupants' health and wellbeing. Although dynamic systems have a higher installation and maintenance cost, they offer numerous advantages to users and investors. For example, another benefit of dynamic systems is that they can control direct and indirect radiations, which balances solar heat gain and loss in extreme seasons (winter and summer) (Jayathissa et al., 2017), and eventually brings comfort to the occupants and reduces the annual energy demand of buildings.

6 Conclusion

This research aims to show the importance of adopting a multi-functional real-time adaptive BIPV system as a response to climate change and associated energy use augments. To support this suggestion, a real-time adaptive BIPV shading system is designed and evaluated with the use of multi-objective evolutionary computation methods. This multi-objective study operated on extracting optimal BIPV façade configurations while addressing three fitness functions simultaneously: To maximize irradiance value on the BIPV, minimize internal illuminance values above 3000 lux, and minimize internal illuminance levels below 300 lux, thus increasing visual comfort in building interiors, minimizing underlit and overlit spots, and maximizing energy generation potential of the BIPV façade. The performance comparison of the real-time adaptive BIPV façade solution with typical static BIPV panel installations revealed the benefits of the proposed adaptive system considering its attainment of optimum visual comfort conditions and energy generation potential irrespective of the time of the day. This aspect is seen as a substantial achievement considering the limitations of a static BIPV system with regards to its exposure to solar irradiation and limited adaptive ability towards visual comfort.

The proposed multi-objective evolutionary computing method and the resulting real-time adaptive BIPV façade solution can certainly be used for proposed new-built scenarios, and for existing building stock, in the form of a retrofitting initiative. The adoption of such retrofit processes will be beneficial to occupants, investors, and the climatic context alike resulting in economic, health, and environmental benefits. Moreover, there is an opportunity to establish an interactive model between the inhabitants and the BIPV system (similar to an inhabitant's interaction with a HVAC system) where the inhabitants can control internal illuminance comfort levels through choosing from a pre-defined set of BIPV configurations at different times throughout the day. The research findings thus profess an integrated approach wherein computational tools and techniques can aid in the multi-performative deployment of BIPV systems, thus contributing towards on-going efforts to mitigate climate change and maximizing renewable energy use potential.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

Methodology, NB, MM, NA; Case Study, MM; Analysis, MM, NA; Writing-original draft preparation, NB, MM, NA; Writing-review and editing, NB, MM, NA.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Agathokleous, R. A., and Kalogirou, S. A. (2016). Double skin facades (DSF) and building integrated photovoltaics (BIPV): A review of configurations and heat transfer characteristics. *Renew. Energy* 89, 743–756. doi:10.1016/j.renene.2015.12.043

Abdollahzadeh, N., and Biloria, N. (2022). Urban microclimate and energy consumption: A multi-objective parametric urban design approach for dense subtropical cities. *Front. Archit. Res.*, 1–13. doi:10.1016/j.foar.2022.02.001

Australian Building Codes Board (2016). *National construction code volume one. Building code of Australia, class 2 to class 9 buildings*. Canberra: Australian Building Codes Board.

Bahr, W. (2014). A comprehensive assessment methodology of the building integrated photovoltaic blind system. *Energy Build.* 82, 703–708. doi:10.1016/j.enbuild.2014.07.065

Biloria, N., and Abdollahzadeh, N. (2022). "Energy-efficient retrofit measures to achieve nearly zero energy buildings." In *Nearly zero energy building (NZEB) - materials, design and new approaches*, London, UK: IntechOpen. doi:10.5772/intechopen.101845

Biyik, E., Araz, M., Hepbasli, A., Shahrestani, M., Yao, R., Shao, L., et al. (2017). A key review of building integrated photovoltaic (BIPV) systems. *Eng. Sci. Technol. Int. J.* 20, 833–858. doi:10.1016/j.jestch.2017.01.009

Bonomo, P., Chatzipanagi, A., and Frontini, F. (2015). Overview and analysis of current BIPV products: New criteria for supporting the technological transfer in the building sector. *Vitr. J. Archit. Technol. Sustain.* 2015, 67–85. doi:10.4995/vitruvio-ijats.2015.4476

Cannavale, A., Ierardi, L., Hörantner, M., Eperon, G. E., Snaith, H. J., Ayr, U., et al. (2017). Improving energy and visual performance in offices using building integrated perovskite-based solar cells: A case study in southern Italy. *Appl. Energy* 205, 834–846. doi:10.1016/j.apenergy.2017.08.112

Chae, Y. T., Kim, J., Park, H., and Shin, B. (2014). Building energy performance evaluation of building integrated photovoltaic (BIPV) window with semi-transparent solar cells. *Appl. Energy* 129, 217–227. doi:10.1016/j.apenergy.2014.04.106

Chi, A., Rangel, D., and Navarro, J. (2018). Correlating daylight availability metric with lighting, heating, and cooling energy consumptions. *Build. Environ.* 132, 170–180. doi:10.1016/j.buildenv.2018.01.048

Council of Australian Governments (COAG) (2012). Baseline energy consumption and greenhouse gas emissions.

Davidsson, H., Perers, B., and Karlsson, B. (2010). Performance of a multifunctional PV/T hybrid solar window. *Sol. Energy* 84, 365–372. doi:10.1016/j.solener.2009.11.006

Davidsson, H., Perers, B., and Karlsson, B. (2012). System analysis of a multifunctional PV/T hybrid solar window. *Sol. Energy* 86, 903–910. doi:10.1016/j.solener.2011.12.020

Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T. (2000). "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II," in *International conference on parallel problem solving from nature*: Springer, 849–858.

De Jong, K. A. (2006). *Evolutionary computation: A unified approach*. Massachusetts, United States: MIT Press.

Defaix, P. R., Van Sark, W., Worrell, E., and de Visser, E. (2012). Technical potential for photovoltaics on buildings in the EU-27. *Sol. Energy* 86, 2644–2653. doi:10.1016/j.solener.2012.06.007

Delponte, E., Marchi, F., Frontini, F., Polo, C., Fath, K., and Batey, M. (2015). "BIPV in EU28, from niche to mass market: An assessment of current projects and the potential for growth through product innovation," in Proceedings of the 31st European Photovoltaic Solar Energy Conference and Exhibition, Hamburg, Germany, 14-18 September 2015, 3046–3050.

Dewidar, K., Mahmoud, A. H., Magdy, N., and Ahmed, S. (2010). "The role of intelligent façades in energy conservation Khaled," in Proceeding Int. Conf. Sustain. Futur. Futur. Intermed. sus- tainable cities, Egypt.

Eicker, U., Demir, E., and Gürlich, D. (2015). Strategies for cost efficient refurbishment and solar energy integration in European Case Study buildings. *Energy Build.* 102, 237–249. doi:10.1016/j.enbuild.2015.05.032

Ernstand Young (2016). *The Upside of Disruption: Megatrends shaping 2016 and beyond*.

Evola, G., and Margani, G. (2016). Renovation of apartment blocks with BIPV: Energy and economic evaluation in temperate climate. *Energy Build.* 130, 794–810. doi:10.1016/j.enbuild.2016.08.085

Farkas, K., Frontini, F., Maturi, L., Munari Probst, M. C., Roecker, C., and Scognamiglio, A. (2013). *Designing photovoltaic systems for architectural integration*. France: Farkas, Klaudia pour International Energy Agency Solar Heating and Cooling.

Fossa, M., Ménézo, C., and Leonardi, E. (2008). Experimental natural convection on vertical surfaces for building integrated photovoltaic (BIPV) applications. *Exp. Therm. fluid Sci.* 32, 980–990. doi:10.1016/j.expthermflusci.2007.11.004

Freitas, S., and Brito, M. C. (2015). "Maximizing the solar photovoltaic yield in different building facade layouts," in Proceedings of the European Photovoltaic Solar Energy Conference and Exhibition, Hamburg, Germany, 14-18 September 2015, 14–18.

Gao, Y., Dong, J., Isabella, O., Santbergen, R., Tan, H., Zeman, M., et al. (2018). A photovoltaic window with sun-tracking shading elements towards maximum power generation and non-glare daylighting. *Appl. Energy* 228, 1454–1472. doi:10.1016/j.apenergy.2018.07.015

Gonçalves, J. E., van Hooff, T., and Saelens, D. (2020). Understanding the behaviour of naturally-ventilated BIPV modules: A sensitivity analysis. *Renew. Energy* 161, 133–148. doi:10.1016/j.renene.2020.06.086

Hachem, C., Athienitis, A., and Fazio, P. (2014). Energy performance enhancement in multistory residential buildings. *Appl. Energy* 116, 9–19. doi:10.1016/j.apenergy.2013.11.018

Heidari Matin, N., and Eydgahi, A. (2019). Technologies used in responsive facade systems: A comparative study. *Intell. Build. Int.* 2019, 54–73. doi:10.1080/17508975.2019.1577213

Hong, T., Koo, C., Oh, J., and Jeong, K. (2017). Nonlinearity analysis of the shading effect on the technical–economic performance of the building-integrated photovoltaic blind. *Appl. Energy* 194, 467–480. doi:10.1016/j.apenergy.2016.05.027

IEA (2019). *Renewables 2019: Analysis and forecasts to 2024*. Paris: Int. Energy Agency.

International Energy Agency (2013). *Renewable energy medium-term market report – market trends and projections to 2018*. Paris, France: International Energy Agency.

Jayathissa, P., Luzzatto, M., Schmidli, J., Hofer, J., Nagy, Z., and Schlueter, A. (2017). Optimising building net energy demand with dynamic BIPV shading. *Appl. Energy* 202, 726–735. doi:10.1016/j.apenergy.2017.05.083

Jelle, B. P., Breivik, C., and Røkenes, H. D. (2012). Building integrated photovoltaic products: A state-of-the-art review and future research opportunities. *Sol. Energy Mat. Sol. Cells* 100, 69–96. doi:10.1016/j.solmat.2011.12.016

Jeong, K., Hong, T., Koo, C., Oh, J., Lee, M., and Kim, J. (2017). A prototype design and development of the smart photovoltaic system blind considering the photovoltaic panel, tracking system, and monitoring system. *Appl. Sci.* 7, 1077. doi:10.3390/app7101077

Joe, J., Choi, W., Kwon, H., and Huh, J.-H. (2013). Load characteristics and operation strategies of building integrated with multi-story double skin facade. *Energy Build.* 60, 185–198. doi:10.1016/j.enbuild.2013.01.015

Kaelin, M., Rudmann, D., and Tiwari, A. N. (2004). Low-cost processing of CIGS thin film solar cells. *Sol. Energy* 77, 749–756. doi:10.1016/j.solener.2004.08.015

Kang, S., Hwang, T., and Kim, J. T. (2012). Theoretical analysis of the blinds integrated photovoltaic modules. *Energy Build.* 46, 86–91. doi:10.1016/j.enbuild.2011.10.042

Karava, P., Jubayer, C. M., Savory, E., and Li, S. (2012). Effect of incident flow conditions on convective heat transfer from the inclined windward roof of a low-rise building with application to photovoltaic-thermal systems. *J. Wind Eng. Ind. Aerodyn.* 104, 428–438. doi:10.1016/j.jweia.2012.03.026

Kolarevic, B., and Parlac, V. (2015). Building dynamics: Exploring architecture of change. *Build. Dyn. Explor. Archit. Chang.* 2015, 1–298. doi:10.4324/9781315763279

Koo, C., Hong, T., Jeong, K., Ban, C., and Oh, J. (2017). Development of the smart photovoltaic system blind and its impact on net-zero energy solar buildings using technical-economic-political analyses. *Energy* 124, 382–396. doi:10.1016/j.energy.2017.02.088

Kushiya, K. (2014). CIS-based thin-film PV technology in solar frontier KK. *Sol. Energy Mat. Sol. Cells* 122, 309–313. doi:10.1016/j.solmat.2013.09.014

Lee, E., Selkowitz, S., Bazjanac, V., Inkarojrit, V., and Kohler, C., 2002. High-performance commercial building façades.Report LBNL-50502

Loonen, R. C. G. M., Trčka, M., Cóstola, D., and Hensen, J. L. M. (2013). Climate adaptive building shells: State-of-the-art and future challenges. *Renew. Sustain. energy Rev.* 25, 483–493. doi:10.1016/j.rser.2013.04.016

Lucon, O., and Ürge-Vorsatz, D. (2014). Fifth assessment report, mitigation of climate change. *Intergov. Panel Clim. Chang.*, 674–738.

Luo, Y., Zhang, L., Liu, Z., Xie, L., Wang, X., and Wu, J. (2018). Experimental study and performance evaluation of a PV-blind embedded double skin façade in winter season. *Energy* 165, 326–342. doi:10.1016/j.energy.2018.09.175

Luo, Y., Zhang, L., Wang, X., Xie, L., Liu, Z., Wu, J., et al. (2017). A comparative study on thermal performance evaluation of a new double skin façade system integrated with photovoltaic blinds. *Appl. Energy* 199, 281–293. doi:10.1016/j.apenergy.2017.05.026

Makki, M., Showkatbakhsh, M., and Song, Y. (2018). *Wallacei, An evolutionary and analytic engine for Grasshopper 3D, Primer 2.0.*

Martín-Chivelet, N., Gutiérrez, J. C., Alonso-Abella, M., Chenlo, F., and Cuenca, J. (2018). Building retrofit with photovoltaics: Construction and performance of a BIPV ventilated façade. *Energies* 11, 1719. doi:10.3390/en11071719

Nielsen, M. V., Svendsen, S., and Jensen, L. B. (2011). Quantifying the potential of automated dynamic solar shading in office buildings through integrated simulations of energy and daylight. *Sol. Energy* 85, 757–768. doi:10.1016/j.solener.2011.01.010

Palmero-Marrero, A. I., and Oliveira, A. C. (2010). Effect of louver shading devices on building energy requirements. *Appl. Energy* 87, 2040–2049. doi:10.1016/j.apenergy.2009.11.020

Park, H. S., Koo, C., Hong, T., Oh, J., and Jeong, K. (2016). A finite element model for estimating the techno-economic performance of the building-integrated photovoltaic blind. *Appl. Energy* 179, 211–227. doi:10.1016/j.apenergy.2016.06.137

Paydar, M. A. (2020). Optimum design of building integrated PV module as a movable shading device. *Sustain. Cities Soc.* 62, 102368. doi:10.1016/j.scs.2020.102368

Peng, J., Curcija, D. C., Lu, L., Selkowitz, S. E., Yang, H., and Zhang, W. (2016). Numerical investigation of the energy saving potential of a semi-transparent photovoltaic double-skin facade in a cool-summer Mediterranean climate. *Appl. Energy* 165, 345–356. doi:10.1016/j.apenergy.2015.12.074

Peng, J., Lu, L., Yang, H., and Han, J. (2013). Investigation on the annual thermal performance of a photovoltaic wall mounted on a multi-layer façade. *Appl. Energy* 112, 646–656. doi:10.1016/j.apenergy.2012.12.026

Poon, K. H., Kämpf, J. H., Tay, S. E. R., Wong, N. H., and Reindl, T. G. (2020). Parametric study of URBAN morphology on building solar energy potential in Singapore context. *Urban Clim.* 33, 100624. doi:10.1016/j.uclim.2020.100624

Prieto, A., Knaack, U., Auer, T., and Klein, T. (2019). Coolfacade: State-of-the-art review and evaluation of solar cooling technologies on their potential for façade integration. *Renew. Sustain. Energy Rev.* 101, 395–414. doi:10.1016/j.rser.2018.11.015

Quesada, G., Rousse, D., Dutil, Y., Badache, M., and Hallé, S. (2012). A comprehensive review of solar facades. Transparent and translucent solar facades. *Renew. Sustain. Energy Rev.* 16, 2643–2651. doi:10.1016/j.rser.2012.02.059

Raugei, M., and Frankl, P. (2009). Life cycle impacts and costs of photovoltaic systems: Current state of the art and future outlooks. *Energy* 34, 392–399. doi:10.1016/j.energy.2009.01.001

Roudsari, M. S., and Mackey, C. (2012). Ladybug Tools (1.5) [WWW Document]. Ladybug Tools. Available at: https://www.ladybug.tools/.

Secretariat, U. (2015). "United Nations framework convention on climate change," in Report of the Conference of the Parties on its twenty-first session, Paris, 30 November to 13 December 2015.

Showkatbakhsh, M., and Mohammed, M. (2022). Multi-objective optimisation of urban form: A framework for selecting the optimal solution. *Build. (Basel).* 12 (9), 1473. doi:10.3390/buildings12091473

Sun, L., Lu, L., and Yang, H. (2012). Optimum design of shading-type building-integrated photovoltaic claddings with different surface azimuth angles. *Appl. Energy* 90, 233–240. doi:10.1016/j.apenergy.2011.01.062

Taveres-Cachat, E., Lobaccaro, G., Goia, F., and Chaudhary, G. (2019). A methodology to improve the performance of PV integrated shading devices using multi-objective optimization. *Appl. Energy* 247, 731–744. doi:10.1016/j.apenergy.2019.04.033

Vadiee, A., Yaghoubi, M., Martin, V., and Bazargan-Lari, Y. (2016). Energy analysis of solar blind system concept using energy system modelling. *Sol. Energy* 139, 297–308. doi:10.1016/j.solener.2016.09.039

Wpcpey (2017). UTS tower building 201708 [WWW Document]. Available at: https://commons.wikimedia.org/wiki/File:UTS_Tower_Building_201708.jpg (Accessed Febuary 23 2023).

Wilson, G. (2015). Cell efficiency records. Availabe at:http://www.nrel.gov/ncpv/.

Yang, S., Cannavale, A., Di Carlo, A., Prasad, D., Sproul, A., and Fiorito, F. (2020). Performance assessment of BIPV/T double-skin façade for various climate zones in Australia: Effects on energy consumption. *Sol. Energy* 199, 377–399. doi:10.1016/j.solener.2020.02.044

Yang, T., and Athienitis, A. K. (2016). A review of research and developments of building-integrated photovoltaic/thermal (BIPV/T) systems. *Renew. Sustain. Energy Rev.* 66, 886–912. doi:10.1016/j.rser.2016.07.011

Yoo, S. H., and Manz, H. (2011). Available remodeling simulation for a BIPV as a shading device. *Sol. Energy Mat. Sol. Cells* 95, 394–397. doi:10.1016/j.solmat.2010.02.015

Yoo, S. H. (2011). Simulation for an optimal application of BIPV through parameter variation. *Sol. Energy* 85, 1291–1301. doi:10.1016/j.solener.2011.03.004

Yu, G., Yang, H., Luo, D., Cheng, X., and Ansah, M. K. (2021). A review on developments and researches of building integrated photovoltaic (BIPV) windows and shading blinds. *Renew. Sustain. Energy Rev.* 149, 111355. doi:10.1016/j.rser.2021.111355

Frontiers | Frontiers in Built Environment

TYPE Review
PUBLISHED 07 April 2023
DOI 10.3389/fbuil.2023.1176621

Check for updates

OPEN ACCESS

EDITED BY
Giovanni Castellazzi,
School of Engineering, University of Bologna, Italy

REVIEWED BY
Annalaura Vuoto,
University of Minho, Portugal
Marco Funari,
University of Surrey, United Kingdom

*CORRESPONDENCE
Alejandro Jiménez Rios,
alejand@oslomet.no

SPECIALTY SECTION
This article was submitted to Computational Methods in Structural Engineering, a section of the journal Frontiers in Built Environment

RECEIVED 28 February 2023
ACCEPTED 28 March 2023
PUBLISHED 07 April 2023

CITATION
Jiménez Rios A, Plevris V and Nogal M (2023), Bridge management through digital twin-based anomaly detection systems: A systematic review.
Front. Built Environ. 9:1176621.
doi: 10.3389/fbuil.2023.1176621

Bridge management through digital twin-based anomaly detection systems: A systematic review

Alejandro Jiménez Rios[1]*, Vagelis Plevris[2] and Maria Nogal[3]

[1]Department of Built Environment, Oslo Metropolitan University (OsloMet), Oslo, Norway, [2]Department of Civil and Architectural Engineering, Qatar University, Doha, Qatar, [3]Department of Materials, Mechanics, Management and Design, Delft University of Technology, Delft, Netherlands

Bridge infrastructure has great economic, social, and cultural value. Nevertheless, many of the infrastructural assets are in poor conservation condition as has been recently evidenced by the collapse of several bridges worldwide. The objective of this systematic review is to collect and synthesize state-of-the-art knowledge and information about how bridge information modeling, finite element modeling, and bridge health monitoring are combined and used in the creation of digital twins (DT) of bridges, and how these models could generate damage scenarios to be used by anomaly detection algorithms for damage detection on bridges, especially in bridges with cultural heritage value. A total of 76 relevant studies from 2017 up to 2022 have been taken into account in this review. The synthesis results show a consensus toward the future adoption of DT for bridge design, management, and operation among the scientific community and bridge practitioners. The main gaps identified are related to the lack of software interoperability, the required improvement of the performance of anomaly-detection algorithms, and the approach definition to be adopted for the integration of DT at the macro scale. Other potential developments are related to the implementation of Industry 5.0 concepts and ideas within DT frameworks.

KEYWORDS

bridges, digital twins, anomaly detection algorithms, finite element method, cultural heritage conservation, bridge information modeling, bridge health monitoring

1 Introduction

In 2018 the Morandi bridge collapsed in Genova, Italy, killing 43 people, forcing the displacement of 200 families living below the bridge, causing damages of EUR 422 million and yearly losses of EUR 784 million to the industry sector in the region (Xuequan, 2018). During the last 2 decades, the collapse of more than 120 bridges worldwide has caused major economic losses and casualties (Wang et al., 2022). A total of 9 661 structures representing 12.4% of all bridges and tunnels in Canada are reported to be in poor/very poor condition (Canada Infrastructure, 2019), whereas 46 ,154 bridges, equivalent to 7.5% of this kind of asset in the United States are considered structurally deficient (ASCE, 2021). In comparison, the percentages of deficient bridges in European countries such as France, Germany, and the United Kingdom are even higher, 39%, 30%, and 37%, respectively (European Commission et al., 2019). Besides, many old bridges are considered to have a Cultural Heritage (CH) value (Jiménez Rios and O'Dwyer, 2019) and some of them are even included in the UNESCO World Heritage List (World Heritage Centre, 2023) thanks to their outstanding universal cultural value. In addition to human and economic losses, the damage or collapse of a historical bridge also entails the painful loss of a cultural asset.

Because of the large number of existing bridges and the limited availability of human and economic resources (PIARC, 2023), it is not feasible to continuously inspect and assess the structural condition of every bridge using conventional methods. In the current practice, bridge inspections are performed on a code-prescriptive fixed-scheduled periodic basis varying between two to 6 years (EuroStruct, 2020). However, these periodic revisions have proven to be ineffective, as damage could appear after a periodic inspection and not be detected until the next one, leading to further deterioration of the bridge and increased cost of its eventual repair or replacement, if not to its collapse. In addition to the particular condition of a bridge, other factors can be considered in scheduling and performing bridge inspections. Most approaches consider the current and future usage of the bridge, its role in the transportation network, as well as other environmental, political, and social factors. It is of paramount importance to integrate CH values with bridge management methodologies, in agreement with international principles of conservation (Petzet, 2004), otherwise irreplaceable parts of our built environment may be lost forever.

A theoretical way to tackle the issue of insufficient resources at a network level, while adequately considering the CH value of a bridge, is to adopt a novel Digital Twin (DT) paradigm (Shabani et al., 2022). A DT of a bridge contains a virtual replica of a real-world bridge and a connectivity module that allows both the physical and virtual assets to be synchronized along the life cycle stages of the asset. The 3D geometry of the bridge can be created through a Bridge Information Modelling (BrIM) approach, whereas a structural twin can be constructed in Finite Element (FE) software. Sensors installed during a Bridge Health Monitoring (BHM) process can provide data about the environmental conditions, loads and response of the structure to those loads, either at local-element or global bridge scale. A series of damage and decay scenarios can be simulated on the virtual asset, which will reproduce the structural response of its physical counterpart through a series of FE models. This digital approach allows testing the bridge and generating the required data under several “normal” and “damaged” scenarios necessary for training Artificial Intelligence (AI) data-driven models such as Anomaly Detection Algorithms (ADAs) capable to detect damage in quasi-real time. The bridge management team or other stakeholders use the generated information to make an informed decision, thus optimizing the resources they have at their disposal. Therefore, the employment of a DT methodology leads to improved bridge performance and CH conservation, an increase in the bridge service life, and an eventual reduction of the maintenance and operation costs of the bridge network.

This systematic review aims to collect and synthesize state-of-the-art knowledge and information about how BrIM, FE, and BHM are combined and used in the creation of DTs of bridges and how the use of these models could generate damage scenarios to be employed by AI ADAs for damage detection on bridges (especially for bridges with high CH value). To this end, the proposed systematic review answers the following questions: i) what are the available ways to build bridge DTs based on BrIM, FEs, and BHM?; and ii) what are the available ADAs that could be used in the damage detection of conventional and CH bridges?

The value of this study lies in the need of having a comprehensive perspective of the current state of the art as the keystone for further research and development. The rest of this paper is organized as follows: Section 2 presents the methodology applied for the search strategy, bibliometric analysis, and synthesis of the found information, Section 3 contains the bibliometric results, and Section 4, the narrative synthesis. Finally, in Section 5 some conclusions are drawn, highlighting the gaps and further research suggestions derived from the systematic review work.

2 Methodology

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 methodology (Page et al., 2021a) was adopted. Thus, this systematic review has followed the checklist provided by PRISMA and a protocol was developed in accordance with the guidelines of the PRISMA-P Explanation and Elaboration (Page et al., 2021b). Following the guidelines, our systematic review protocol was registered in the Open Science Framework (OSF) Registries with registration number sh9b2 (Jiménez Rios et al., 2023b). The protocol of this systematic review can be consulted in Jiménez Rios et al. (2023d).

2.1 Search strategy

The quality of systematic reviews heavily relies on the search strategy implemented for the information retrieval process. Nevertheless, search strategies are commonly not adequately reported. This systematic review has adopted a search strategy methodology based on the PRISMA-S checklist (Rethlefsen et al., 2021) and it can be consulted in Jiménez Rios et al. (2023e).

The search strategy implemented was performed in Scopus because of its wide coverage of the literature, its high-quality content and its advanced data extraction capabilities (Elsevier, 2023). Initially, seven main keywords of interest were selected, namely, *“bridge,” “digital twin,” “bridge information modeling,” “finite element methods,” “bridge health monitoring,” “anomaly detection algorithms”* and *“cultural heritage”*). These keywords (and similar terms such as *“bridge”* and *“bridges”*) were combined to obtain six search queries in which every search combined a keyword with the *“bridge”* keyword. Thus, the queries obtained were.

- bridge* AND “digital twin*”
- bridge* AND (BrIM OR “bridge information model*“)
- bridge* AND (FEM OR FEA OR “finite element method*” OR “finite element analy*“)
- bridge* AND (“bridge health monitoring” OR “structural health monitoring”)
- bridge* AND (ADA OR “anomaly detection algorithm*“)
- bridge* AND (“cultural heritage” OR “monument* bridge*” OR “old bridge*” OR “ancient bridge*” OR “historic* bridge*“)

where * represents the wild character, AND and OR are Boolean operators, “·” are used to group individual words into

TABLE 1 Full queries used for the search and the respective number of records found.

#	Query	# Of records found
1	TITLE-ABS-KEY (bridge* AND "digital twin*") AND (LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017)) AND (LIMIT-TO (SUBJAREA, "ENGI")) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ch"))	178
2	TITLE-ABS-KEY (bridge* AND (brim OR "bridge information model*")) AND (LIMIT-TO (SUBJAREA, "ENGI")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ch")) AND (LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017)) AND (LIMIT-TO (LANGUAGE, "English"))	56
3	TITLE-ABS-KEY (bridge* AND (fem OR fea OR "finite element method*" OR "finite element analy*")) AND (LIMIT-TO (SUBJAREA, "ENGI")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ch")) AND (LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017)) AND (LIMIT-TO (LANGUAGE, "English"))	5137
4	TITLE-ABS-KEY (bridge* AND ("bridge health monitoring" OR "structural health monitoring")) AND (LIMIT-TO (SUBJAREA, "ENGI")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ch")) AND (LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017)) AND (LIMIT-TO (LANGUAGE, "English"))	2941
5	TITLE-ABS-KEY (bridge* AND (ada OR "anomaly detection algorithm*")) AND (LIMIT-TO (SUBJAREA, "ENGI")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ch")) AND (LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017)) AND (LIMIT-TO (LANGUAGE, "English"))	10
6	TITLE-ABS-KEY (bridge* AND ("cultural heritage" OR "monument* bridge*" OR "old bridge*" OR "ancient bridge*" OR "historic* bridge*")) AND (LIMIT-TO (SUBJAREA, "ENGI")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ch")) AND (LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017)) AND (LIMIT-TO (LANGUAGE, "English"))	351

multi-word keywords and (·) are used to group several similar terms. The six searches were limited to journal articles, conference papers, reviews, and book chapters, written in English, that were published after 2017, on the subject of *Engineering*. The searches were performed within the fields of *title*, *abstract*, and *keywords*. Table 1 presents the full queries used in the search, which was performed on 10/12/2022, and the respective number of records found for each one of them.

A total of 8 673 records were found. Detailed bibliometric information about all these records was downloaded from Scopus both in. ris and. csv format and has been made available in the open-source database Jiménez Rios et al. (2023a). Deduplication, filtering, screening, and eligibility assessment of all those records were carried out following PRISMA flow chart (Page et al., 2021a) (see Figure 1).

All duplicated records (based on the DOI number) and those records without a DOI number were removed. In total, 1 838 records were discarded after this first filtering. As a second filter, a search combination was performed. The initial six searches were combined using the AND operator (resulting in 15 new searches) to obtain relevant records dealing with at least three of the initially selected keywords (as the *"bridge"* keyword was used in all original six searches). Thus, 6 446 records were excluded and only 389 records remained, including the top-ten most cited papers from each one of the original searches (60 papers in total). The most cited papers were deemed to be of paramount importance to the state-of-the-art of the field due to their major impact on all related publications.

The selection phase (according to PRISMA 2020 item checklist #8) started at this point by manually screening the title and abstract of the remaining 389 records based on the authors' criteria and previous knowledge of the field. Those records that did not fully fit within the scope of the review were excluded. Thus, 108 works remained and were subjected to full paper examination to assess their eligibility. From this list of 108 works, 2 were removed as they were duplicates, another 2 were excluded as they only dealt with the construction of new bridges, 12 more were not considered as they did not deal with DTs, and lastly, 16 papers were rejected as they were not related to bridges. As a result, a total of 76 studies were finally included in this systematic review.

2.2 Bibliometric analysis methodology

A bibliometric analysis represents a quantitative methodology by which meaningful insights can be obtained from large quantities of data (Broadus, 1987). The main outcomes of a bibliometric analysis are the identification of emerging research trends in a field, collaboration, and publication patterns, and exploration of literature structure (Koutsantonis et al., 2022). The approaches of a bibliometric analysis could be categorized into two main groups: performance analysis and science mapping (Solorzano and Plevris, 2022).

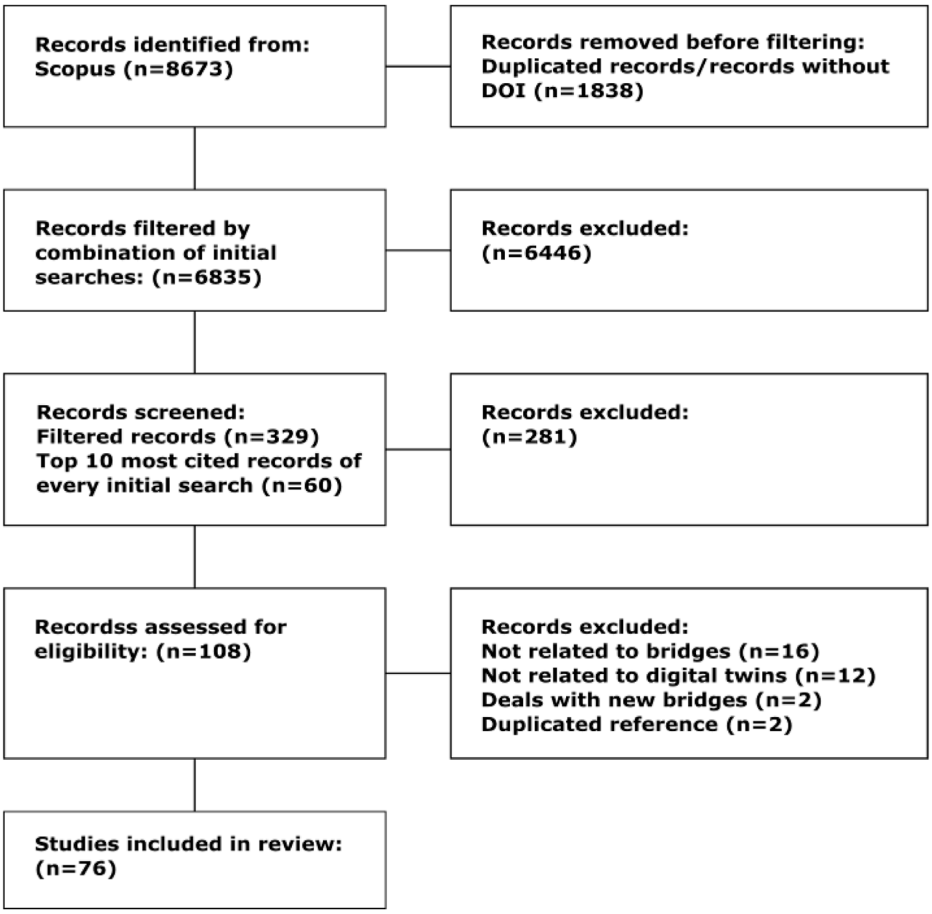


FIGURE 1
PRISMA flow diagram.

In this systematic review, the performance analysis was carried out by querying, filtering, and sorting the bibliographic database obtained from the search strategy, namely, using the 6835 records obtained after the first filtering phase, whereas the science mapping was performed using the VOSviewer v1.6.18 software (https://www.vosviewer.com/). Performance analysis is presented in terms of publications per year, most cited authors, most cited records, documents per country, keyword occurrence, and most used source for publication. On the other hand, science mapping focuses on analyzing the co-authorship relationships in terms of authors and countries, as well as the co-occurrence relationships between keywords (both author and index keywords). Keywords mapping allows visualizing the interconnections of core concepts and topics within a certain research area. For further insights into how the maps are created interested readers can consult (van Eck and Waltman, 2014) and the software manual (van Eck and Waltman, 2022).

2.3 Synthesis methodology

The information of the studies included in this systematic review has been qualitatively summarized in a narrative synthesis as the findings are characterized by heterogeneity. Data has been analyzed and classified within 5 major themes, namely,: i) Bridge DTs; ii) BrIM and FE modeling; iii) BHM, AI and ADAs; iv) Unmanned Aerial Vehicles (UAVs), satellite monitoring, and other DT-related emerging technologies for bridge inspection; and v) historical and CH bridges. Based on this classification, the findings of the systematic review are presented, the strengths and limitations of the studies are highlighted, their influence on practice and research is discussed, and future research recommendations are suggested.

3 Bibliometric analysis results and discussion

3.1 Performance analysis

Regarding the number of publications per year, Figure 2 shows that over 1 000 papers containing the keywords of interest of this systematic review were constantly published per year between 2017 and 2020. The trend though shows an increase in the number of publications from the last 2 years, with 25% and 50% increments on the number of yearly publications for the years 2021 and 2022, respectively.

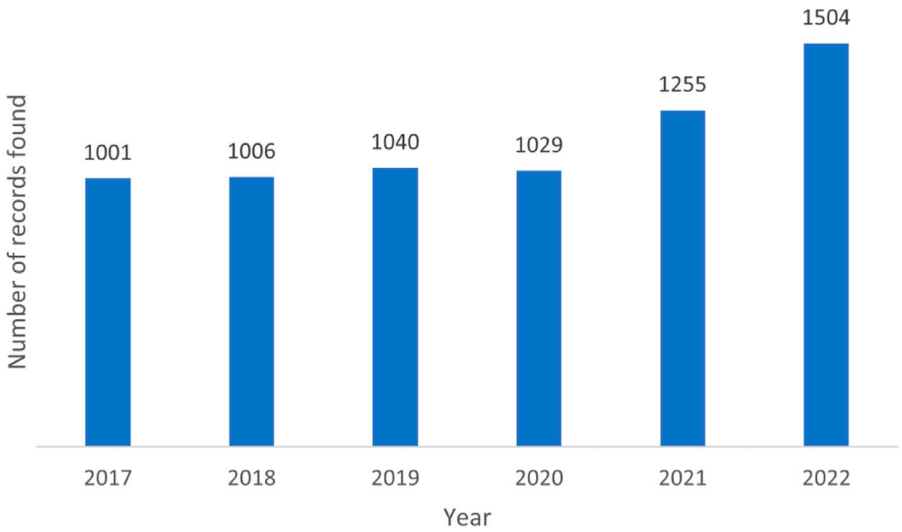


FIGURE 2
Publications per year.

Table 2 presents the top 20 most cited works (as of 10/12/2022, the day the search was conducted). The paper with the most citations is "Shaping the DT for design and production engineering" (Schleich et al., 2017) with a total of 644 citations. Nevertheless, after the filtering process shown in Figure 1, this paper was not included in this systematic review as it is not directly related to bridges. The paper included in this systematic review with the most citations is the second in the list, "Structural Health Monitoring Using Wireless Sensor Networks: A Comprehensive Survey" (Noel et al., 2017) with a total number of 273 citations in Scopus.

Another interesting metric related to citations is that of the most cited authors. This parameter considers the accumulated number of citations for all papers of an author. Thus, Wang, H., Li, H. and Bao, Y. are the most cited authors with 1308, 1276, and 1222 citations, respectively (see Figure 3).

Research is normally fostered at a national level by the National Research Council of each country. Figure 4 presents the countries with at least 100 publications in the field over the past 6 years. China is the country with the most publications (2599) followed by the United States and the United Kingdom with 1282 and 420 publications, respectively. The last country on the list is Turkey (out of the top 16 countries with more publications), with 105 publications. Note that the number of publications per country is based on the country of the authors' affiliations, not on the nationality of the authors.

In terms of keywords occurrence, it is not surprising to find out that *"bridges," "FEM"* and *"SHM"* are among the most frequently used keywords (based on the graphical information presented in Figure 5), as they were explicitly included in the search queries. On the other hand, the absence of terms such as *"digital twins"* and *"bridge information modeling"* may be explained by their relatively new adoption in the field, whereas the absence of keywords related to *"cultural heritage"* or *"conservation"* is directly tracked to the generalized lack of attention towards these topics by the engineering research community.

Most of the research considered in this systematic review has been published in three main scientific journals, namely, *Engineering Structures*, *Journal of Bridge Engineering*, and *Lecture Notes in Civil Engineering*, 445, 254, and 182 works in each one, respectively (see Figure 6). The total number of works concentrated in only these three main sources of publication represents 12.9% of the total number of records after deduplication found from the initial searches of this systematic review.

3.2 Science mapping

Co-author relationships are qualitatively analyzed using a network visualization map. Each circle in Figure 7 represents one of the top 100 authors with the most publications, as found after performing the search strategy described previously. The size of each circle depicts its strength or weight within the network, where larger circles correspond to authors with a larger number of publications. Moreover, the lines that are observed in this figure represent co-authorship links, in other words, who works mostly with whom. Analogously to the size of the items, the thickness of the links represents their strength, i.e., the strength of the co-authorship links of a given researcher with other researchers.

The items in Figure 7A are color-coded into nine different clusters based on network connectivity. Furthermore, Figure 7B shows an overlay visualization of the co-author relationships color-coded in terms of average publication year based on the scores assigned to each item of the network. It can be observed in Figure 7A that Liu Y. (green, 121 publications), Li J. (orange, 100 publications), Li Y. (red, 90 publications), Zhang Y. (pink, 90 publications) and Wang H. (yellow, 88 publications) are the centroids of the five more prominent clusters identified in the network. From these five networks, it can be seen in Figure 7B that the research group

TABLE 2 Most cited records.

#	Title	References	# Citations
1	Shaping the digital twin for design and production engineering	Schleich et al. (2017)	644
2	Structural Health Monitoring Using Wireless Sensor Networks: A Comprehensive Survey	Noel et al. (2017)	273
3	A Digital Twin-Based Approach for Designing and Multi-Objective Optimization of Hollow Glass Production Line	Zhang et al. (2017)	266
4	Computer vision and deep learning–based data anomaly detection method for structural health monitoring	Bao et al. (2019b)	228
5	Building Information Modeling (BIM) for transportation infrastructure—Literature review, applications, challenges, and recommendations	Costin et al. (2018)	216
6	Digital twin-driven rapid individualised designing of automated flow-shop manufacturing system	Liu et al. (2019)	193
7	Experimental validation of cost-effective vision-based structural health monitoring	Feng and Feng (2017)	190
8	The State of the Art of Data Science and Engineering in Structural Health Monitoring	Bao et al. (2019a)	173
9	A review of the piezoelectric electromechanical impedance based structural health monitoring technique for engineering structures	Na and Baek (2018)	170
10	Review of Bridge Structural Health Monitoring Aided by Big Data and Artificial Intelligence: From Condition Assessment to Damage Detection	Sun et al. (2020)	170
11	Autonomous UAVs for Structural Health Monitoring Using Deep Learning and an Ultrasonic Beacon System with Geo-Tagging	Kang and Cha (2018)	156
12	Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring	Tang et al. (2019)	154
13	Environmental effects on natural frequencies of the San Pietro bell tower in Perugia, Italy, and their removal for structural performance assessment	Ubertini et al. (2017)	135
14	Digital twin in smart manufacturing	Li et al. (2022)	128
15	A review on deep learning-based structural health monitoring of civil infrastructures	Ye et al. (2019)	128
16	Structural Displacement Measurement Using an Unmanned Aerial System	Yoon et al. (2018)	126
17	A state of the art review of modal-based damage detection in bridges: Development, challenges, and solutions	Moughty and Casas (2017)	125
18	Structural health monitoring of bridges: a model-free ANN-based approach to damage detection	Neves et al. (2017)	124
19	Investigation of dynamic properties of long-span cable-stayed bridges based on 1-year monitoring data under normal operating condition	Mao et al. (2018)	118
20	Recent progress and future trends on damage identification methods for bridge structures	An et al. (2019)	114

spear-headed by Zhang Y. is the one with the most recent average year of publication (2020.3).

Another interesting co-authorship relationship, now in terms of countries, is showcased in Figure 8. In this instance, three main clusters can be observed from Figure 8A where the strongest items are; China (pink, 2599 publications), United States (red, 1282 publications), followed by the United Kingdom and Italy (green, 420, and 415 publications, respectively). Among them, Italy is the country with the most recent average publication year (2020.09, see Figure 8B).

The co-occurrence relationships between keywords have similarly been analyzed through network and overlay visualization maps as displayed in Figure 9. The *"bridges"* keyword plays a predominant role in this network, which is not surprising because it is the main topic of interest in this systematic review. It is closely related to *"SHM"* and *"Damage detection"* as they belong to the same cluster (red) and have thick link lines (see Figure 9A). Regardless of its relatively small strength, *"AI"* has one of the most recent average publication years (2020.36), which shows its relatively new adoption in the field of bridge engineering (see Figure 9B).

4 Narrative synthesis and discussion

4.1 Bridge digital twins

The life-cycle stages of a bridge include: i) Planning and design; ii) Construction; iii) Inspection and maintenance; iv) Rehabilitation or replacement; and v) Demolition or decommissioning. Accounting for the entire life-cycle of a bridge within the DT paradigm requires the parallel evolution of both the digital and physical assets from the planning and design phase (inspection and maintenance for existing bridges) until the final demolition or decommissioning of the structure. For such purposes, deterioration models that can predict the progressive decay of the structural performance of the physical asset are of paramount importance (Cervenka et al., 2020; Jiang et al., 2021). Thus, Giorgadze et al. (2022) suggest an ontological modeling approach that includes not only components related to the structural elements of the bridge itself but also resources, processes, and risks related to the management and operation activities along the life of a bridge.

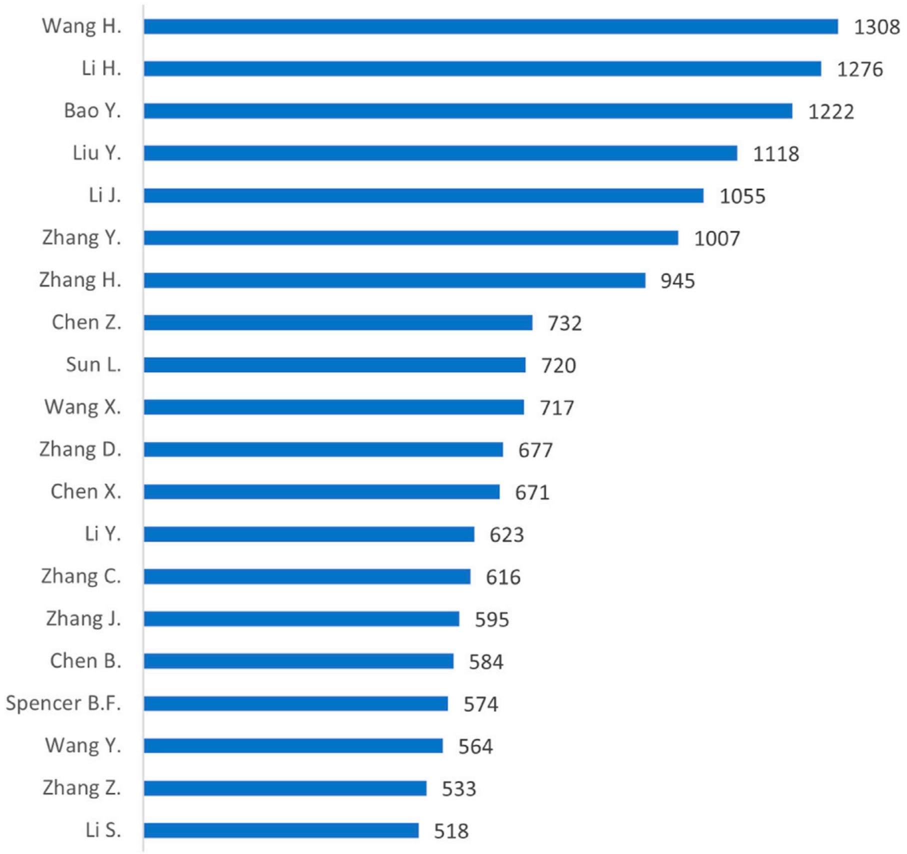


FIGURE 3
Most cited authors.

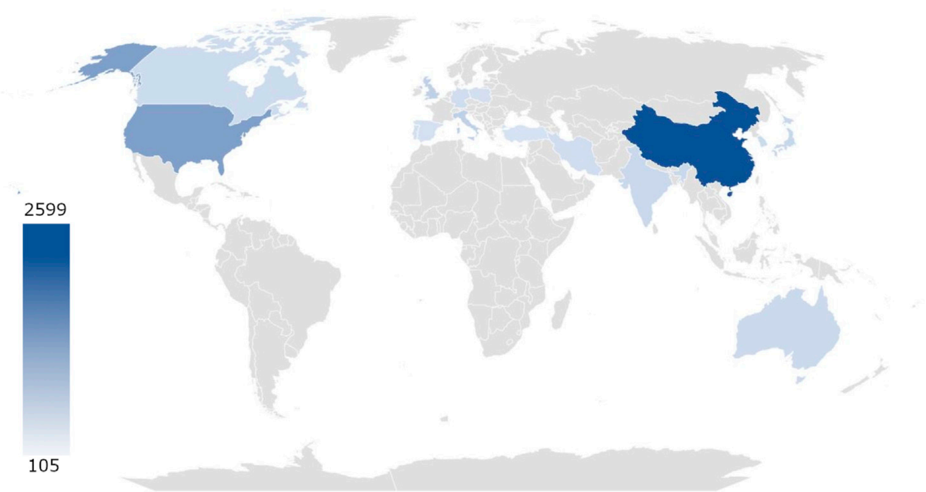


FIGURE 4
Documents per country.

In terms of maturity, Shim et al. (2019) group DTs into three progressive categories based on their Level of Detail (LOD): i) partial DT (LOD 200–300, used during conceptual and detailed design/analysis); ii) clone DT (LOD 400, which provides construction information); and iii) augmented DT (LOD 500, capable of assisting during operation and management stages). Analogously,

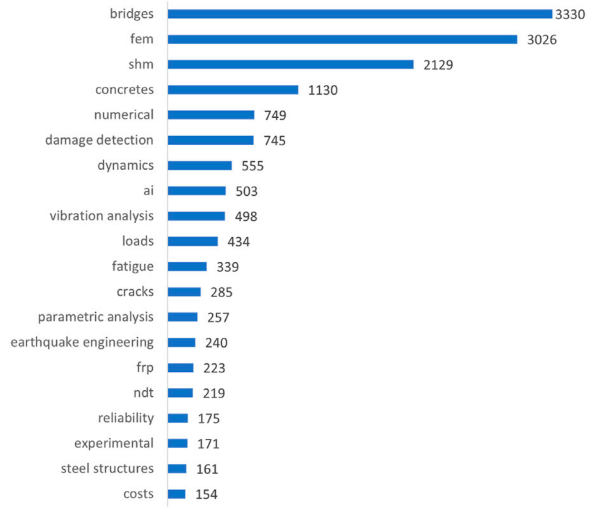


FIGURE 5
Keyword occurrence.

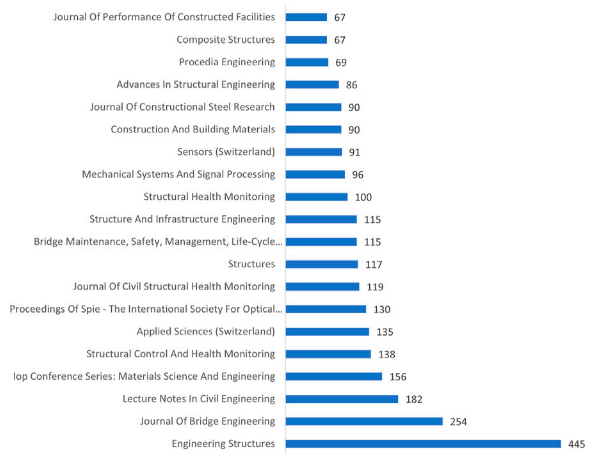


FIGURE 6
Most used sources for publication.

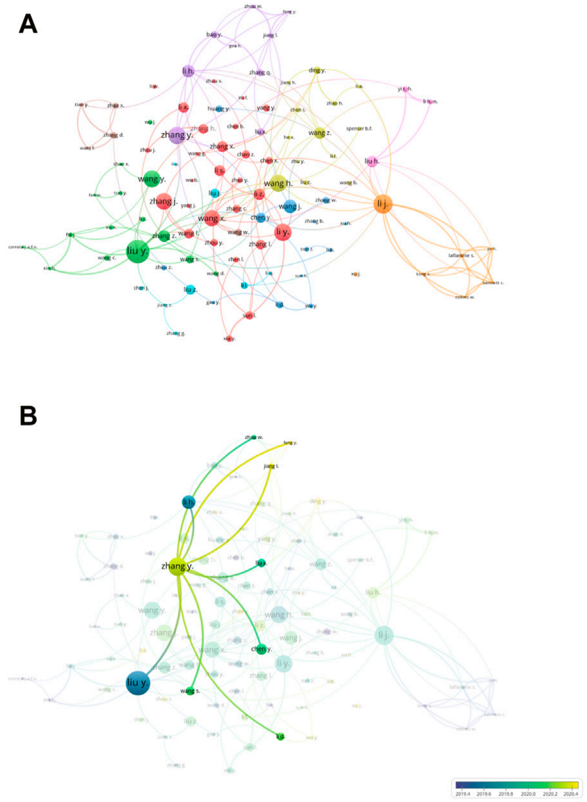


FIGURE 7
Co-authorship relationships in terms of authors: **(A)** Authors clusters and **(B)** Most recent publication averaged year per author, Zhang Y.

Kang et al. (2021) classify DT maturity into three progressive levels of complexity: functional, connected, and intelligent. Yet another classification based on the DT features and scopes is identified by Saback de Freitas Bello et al. (2022). In this threefold classification, a digital model replicates a physical asset but lacks data connectivity between the two, a digital shadow possesses automated one-way data connectivity between the physical and digital counterparts, and finally, in a digital twin the real-time data connectivity is granted in both directions and the digital asset evolves along with the physical one through its service life.

The multi-scale nature of DTs is explored in the work presented by Lu et al. (2020), where they develop a hierarchical architecture to build a DT at both city and building levels. According to the authors' vision, the DT of a bridge could as well be integrated within the DT of a city, and this city DT would eventually form part of a DT at a national level. Although this vision makes sense for urban bridges, the integration of most bridges as part of transportation networks in rural or natural areas would perhaps be more appropriate if a DT is created at the transportation network level (including DTs of roads, tunnels, etc.) which could additionally be benefited from traffic data sharing tools, as explored by Dan et al. (2022). This suggests that the direction for the creation of macro-DTs that integrate DT of individual infrastructure assets is still not clearly defined and it needs to be determined whether a geographical, systemic, or another kind of ontological integration approach would be more favorable for grouping bridge DTs into the macro-DT of an entire transportation network, country, or continent.

A key component for the successful implementation of the DT paradigm in bridge monitoring is the integration of Cloud Computing (CC) within the adopted framework. Jeong et al. (2019) build on top of the OpenBrIM schema proposed by Jeong et al. (2017) and develop an Infrastructure as a Service (IaaS)/Platform as a Service (PaaS) CC environment in the Microsoft Azure cloud platform where an open-source distributed NoSQL database (Apache Cassandra) was employed to ensure scalability, flexibility, fault-tolerance, and high-performing data management. IaaS was offered in the form of Virtual Machines (VMs) that can be scaled either vertically (increasing the computational capabilities of the VM) or horizontally (by adding extra VMs). Furthermore, Software as a Service (SaaS) is provided through an online platform from

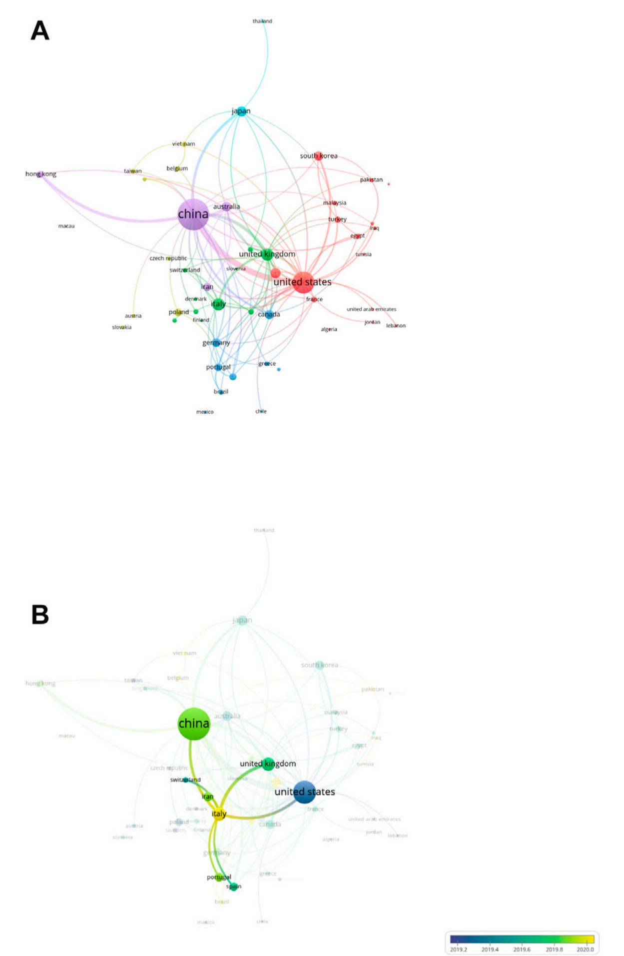


FIGURE 8
Co-authorship relationships in terms of countries: **(A)** Countries clusters and **(B)** country with the most recent average publication, Italy.

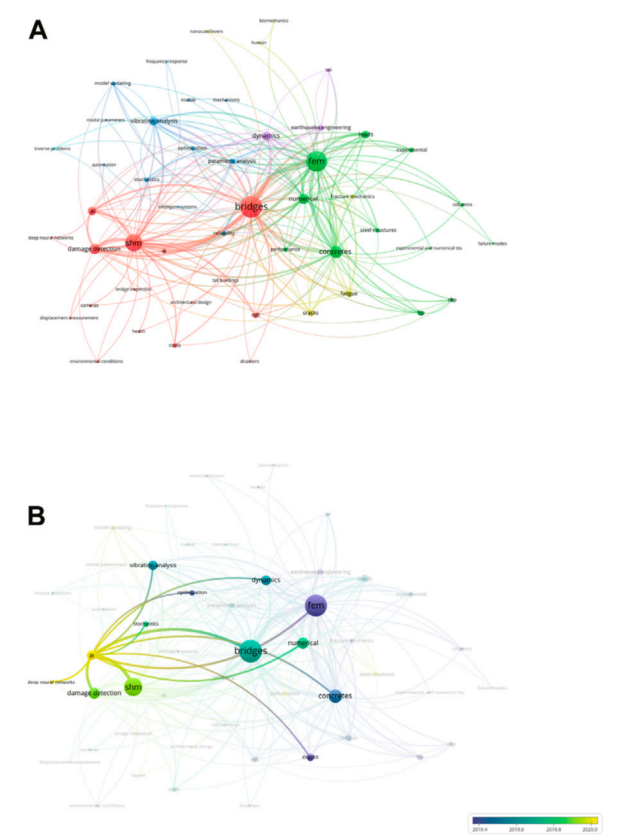


FIGURE 9
Co-occurrence relationships between keywords: **(A)** Keywords clusters and **(B)** AI links.

where the user can query the information of interest and download the model that can be regenerated in structural engineering software such as CSIBridge.

To further improve the performance of a DT, Dang et al. (2022) propose the implementation of an intermediate level defined as Fog Computing (FC, computing done in the data generation device itself), which is capable of filtering the great amount of data generated by BHM systems before transferring only the relevant data to the CC layer. They also recognize the need for having several sub-models as part of the digital replica of a DT, each suitable for particular tasks, namely, analytical models based on mechanics and probability theory capable of providing exact and fast results in terms of structural response, reliability and safety for relatively simple idealized structures, physics-based numerical models (i.e., FE) which can replicate the structural response of complex systems for undamaged/damaged scenarios, be used for prognosis purposes and to generate synthetically augmented data. These data, along with the ones collected from the BHM of the physical asset, can be exploited by a third type of data-driven model, capable of performing real-time damage detection.

Along with FC intermediate data filtering, the implementation of enhanced data acquisition techniques such as compressive sampling, suitable for sparse data signals (Bao et al., 2019a), can drastically reduce the amount of data that would be stored and analyzed in the DT. In terms of visual acquisition data, the amount of information required for processing could be reduced if appropriate compression techniques and image quality percentages are adequately determined as done, for example, by Ri et al. (2020). By using a BrIM model in combination with Genetic Algorithms (GA) and Discrete Event Simulation (DES), Nili et al. (2021) propose a simulation-based framework to optimize bridge intervention (maintenance, rehabilitation and replacement) considering crew limitations. The framework is developed using Microsoft Visual Studio environment, Microsoft Access for the data management and data query, Autodesk Navisworks Manage as the BrIM application software, GA engine for the planning and sequencing modules, and a DES engine of Symphony core service, with a customized. Net programming language code. Nevertheless, this framework lacks consideration for CH conservation philosophy and methodologies when applied to bridges with CH value.

4.2 BrIM and FE modeling

The concept of BrIM is the adaptation of Building Information Modeling (BIM) methodologies applied to bridges (McKenna et al., 2017), whereas Historical BIM (HBIM), has been developed having CH buildings in mind (Pepe et al., 2020). Based on the life-cycle stage of the bridge in which the BrIM model is built, it could be classified as either as-designed, if the model is produced since the planning and design stage, or as-built, if it is created after the construction phase, or as-is, if the model has been effectively interconnected with the physical asset and is capable of updating its status along the further life-cycle stages of the structure (Hosamo and Hosamo, 2022).

Therefore, by following a multi-level and multi-modal approach as suggested by Xiao et al. (2017), an augmented As-Is Historical Bridge Information Modeling (AI-HBrIM), would be the adequate tool to implement the BIM methodology for existing historical bridges within the context of the DT paradigm. It is estimated that the adoption of this approach could result in up to a 30% reduction in traffic-related costs and a 10% reduction in the overall management and operation activities along the entire life of a bridge (Saju, 2022).

A suitable methodology to keep AI-HBrIM digital models interconnected to the physical asset is through FE model updating. Using this approach, the FE model is informed by the actual measured data coming from the physical asset (Yu et al., 2022). Ramancha et al. (2020) implement an advanced Bayesian inference approach using Sequential Monte Carlo (SMC) simulations to update the material and damping model parameters of a full-scale reinforced-concrete column under dynamic loading based on the heterogeneous data collected by accelerometers, strain gauges, GPS displacements, and potentiometers. Similarly, by applying a Bayesian inference approach, Ghahari et al. (2022) successfully update an FE model including soil-structure interaction effects. This was possible thanks to the motion identification at foundation level based on the acceleration measurement data obtained from the BHM.

While nowadays there are several data formats (specific protocols for data storing and retrieving) and schema (organization and structure such as XML, STEP, etc.) proposed for achieving AI-HBrIM interoperability, the OpenBrIM Platform (OpenBrIM, 2023) seems to be the most up-to-date option, whereas Industry Foundation Classes (IFC) (IFC, 2023) development team is currently preparing a new standard (IFC5) including data definitions required for both buildings and bridges over their life cycle. Both OpenBrim and IFC are XML schema-based. In this regard, Jeong et al. (2017) expand the OpenBrIM standard by enriching it with libraries for structural elements (e.g., mesh, constraints, and coordinate systems), load and analysis conditions (e.g., vehicle loads, modal, static, and multi-step) and sensors (e.g., accelerometers, strain gauges, and thermistors). The input data is organized and stored in a NoSQL database and Python is used to create the interface between the database and the analysis software (CSI Bridge) by parsing the XML objects. On the other hand, Park et al. (2018) propose using the functional meaning of bridge components (i.e., column, beam, etc.) to improve the usability of IFC applied to bridges by exploiting IFC's basic modular structure and its framework for the sharing of information between various areas of the construction industry.

Another practicality that has received attention from researchers is the initial geometry modeling process of the AI-HBrIM model. Lu and Brilakis (2019) propose an automatic geometry modeling method to advance the creation of HBrIM models characterized by a slicing-based object-fitting approach. They recreated the geometry of an existing concrete bridge using 3D solid elements in IFC format based on a pre-processed labeled point cluster, a work previously presented by the same authors in Lu et al. (2019). Although their work was limited to a LOD level of 250 and only four general bridge elements, namely, slab, pier cap, pier, and girder, they achieved an impressive time reduction in comparison with manual geometric modeling techniques currently in practice.

Also in this subject, McKenna et al. (2017) present a case study where 3D laser scanning was undertaken to capture as-is geometry and condition data using a Leica P20 pulse-based Terrestrial Laser Scanner (TLS). Scans are colored using imagery obtained from a Nikon D200 camera mounted on a Nodal Ninja bracket to create high-resolution 360° panoramic images and then processed using Leica Cyclone proprietary software to create a 3D solid Autocad model of the structure. Two approaches have been followed to transform the CAD model into an HBrIM one. Leica CloudWorx for Revit is used first and then Autodesk ReCap software. Most of the modeling work is done manually, though.

As an alternative to conventional geometry data capture of existing bridges necessary to build a DT, Rashidi and Karan (2018) propose a low-cost, automatic, videogrammetry methodology. It consists of videotaping the bridge from several views and directions to reduce occlusions, transforming the 2D images captured into a 3D points cloud through the use of a Patch-based Multi-View Stereo (PMVS) algorithm, applying computer vision algorithms to identify the bridge components and exporting those elements to an XML format compatible with major BrIM software (RM Bridge, LEAP Bridge Enterprise, AutoCAD Civil 3D, Revit Structure, and Tekla Structures).

Although limited to presenting the applications, challenges, and recommendations of BIM applied to transportation infrastructure (without integration within the DT methodology), Costin et al. (2018) present a comprehensive review of BIM. They highlight the lack of interoperability within the different tools and methodologies currently in practice (Del Rio et al., 2020; Bouzas et al., 2022; Polania et al., 2022) as one of the main needs to be addressed to facilitate the implementation of BIM on the field of transportation infrastructure. Other significant challenges are the assurance of data quality, methodology cost reduction, inherent limitations, and institutional barriers as well as resistance to change by the industry agents.

4.3 BHM, AI, and ADAs

BHM aims to improve asset performance by measuring and learning from in-service structural behavior (Ye et al., 2022). Moreover, in earthquake-prone countries, BHM supports emergency management actions (Limongelli et al., 2019) and it can even be used to provide real-time traffic information (Burrello et al., 2020). BHM systems are usually designed based on the structural response observed on an a-priori FE model (Ye et al., 2020). Although model-based BHM approaches (Gonen and Soyoz, 2021; Gonen et al., 2023) can predict future bridges' structural response under idealized load scenarios, their use results unfeasible for real-time damage detection applications (due to the high computational resources and the relatively long simulation periods required). With the rapid surge and adoption of AI, a new BHM and damage detection paradigm have recently gained importance: the model-free, also known as data-driven, paradigm. Data-driven methodologies can provide quasi-real-time results when damage occurs, on the other hand, they require large data to be trained and it is difficult to assign physical meaning to the detected damage. Moreover, databases containing information from real damaged bridges are scarce, as highlighted by Kim et al. (2021).

Neves et al. (2017) present a data-driven damage detection approach based on Machine Learning (ML). They test their methodology and overcome the lack of large data by creating a synthetic database with the help of an FE model. The data set consists of accelerations from 300 simulations of healthy and two damage scenarios of a bridge, of which 150 are used for training of an unsupervised Artificial Neural Network (ANN) and the remaining 150 for validation and verification purposes. Even though their approach is effective, the authors list a series of necessary improvements before it could be put into practice, such as considering the effect of environmental and operational conditions, including multiple damage scenarios, extending it for damage location and dealing with factors such as minimum reliability levels (the CH value of the bridge must as well be considered) for the determination of the threshold value. In that regard, Kostic and Gül (2017) try to include environmental and operational effects in their proposed ANN damage detection methodology by implementing a time series analysis, which allows for the successful detection of damage under low levels of temperature-induced noise ($< 3\%$).

By leveraging the mutual advantages of model-based and data-driven approaches, Zhang and Sun (2021) develop a physics-guided ML monitoring strategy. Their methodology consists of training an ANN using a baseline undamaged condition from observations of a bridge and enriched with damage scenarios data synthetically generated through a FE model. To detect damage, it uses the Normalized Frequency Change Ratios (NFCR) and the change of the first several mode shapes of the bridge, combined in a novel cross-entropy loss function. According to the authors, this mixed approach is not only capable of detecting damage, but also of locating and quantifying it.

While some authors have focused on the development of damage-detection data-driven methodologies, others have tried to improve the BHM, which is traditionally based on bridge instrumentation and results economically unfeasible for short and medium-span bridges. Sreevallabhan et al. (2017) present a comprehensive literature review of Structural Health Monitoring (SHM) using Wireless Sensor Networks (WSNs), which are a low-cost alternative to the wired sensor networks commonly used nowadays. Wang et al. (2022) explore the installation and operation of novel piezoelectric transducers, which use a Coda Wave Interferometry (CWI) technique, to assess the condition of existing concrete bridges based on waves generated by the passing vehicles.

On the other hand, OBrien et al. (2017) propose an indirect bridge monitoring approach based on the instrumentation of the vehicles driving through the bridge. This so-called drive-by monitoring provides acceleration data that can be decomposed into three main components; vehicle frequency, bridge natural frequency, and pseudo frequency associated with vehicle speed. These three components are obtained through the means of Empirical Mode Decomposition (EMD). Drive-by monitoring approaches have proved effective not only in damage detection but also in damage location. A research gap identified by them that needs to be addressed to improve the effectiveness of drive-by monitoring is the effect that road roughness has on indirect monitoring. More recently, Locke et al. (2020) present a drive-by monitoring approach capable of not only considering road roughness, which is modeled based on power spectral density functions (International Organization for Standardization, 2016), but also variable environmental and operational conditions.

Another alternative proposed for BHM cost reduction consists of the use of non-contact vision-based displacement sensors to measure bridge displacements, which is a parameter directly related to the stiffness of the structure. These approaches exploit a series of available template matching/registration techniques such as Up-sampled Cross Correlation (UCC), pattern matching, edge detection, Orientation Code Matching (OCM), Digital Image Correlation (DIC), Hough transforms, and RANSAC (Feng and Feng, 2017). More recently, Shao et al. (2020, 2021) propose a holographic visual sensor coupled with computer-vision-based algorithms in a non-contact displacement and vibration measurement system, capable of capturing bridge full-field displacement and vibrations. Nevertheless, vision-based displacement sensors' efficiency highly relies on image quality, which is commonly affected by illumination variation, partial target occlusion, partial shading, and background disturbance, factors usually present in normal bridge operational conditions. In this regard, Shao et al. (2020) suggest the use of denoising and contractive auto encoders to reduce low image-quality errors and improve visual-based monitoring effectiveness.

The problem of damage detection in bridge monitoring may be presented like a simple classification problem, i.e., identifying whether there is or there is not any damage in the bridge. However, the individual and highly complex nature of bridges may result in different dynamic responses, which adds complexity to the task. Conventional classification approaches are rarely successful due to the important imbalance between normal and anomalous cases, resulting in too many false negatives. An excessive number of false negatives may hinder the detection of actual damages or substantial decay, ultimately affecting the performance of a bridge and, in critical cases, leading to its collapse. Conversely, a large number of false positives would lead to unnecessary spending of resources. By contrast, an acceptable number of false positives may be even desirable for damage detection on CH bridges, which could be obtained with the application of a fine-tuned ADA. Table 3 presents a compilation of the diverse ADAs methodologies found in this systematic review.

A comprehensive review of supervised learning, unsupervised learning, novelty detection, and deep neural network methodologies used for generalized damage detection is provided by Sun et al. (2020). Furthermore, Ye et al. (2019) list a series of deep learning techniques specifically used for crack detection, damage detection, loosened bolt detection, and damage state classification of bridges. Finally, a comprehensive list of damage detection methods classified either as model-based (FE model updating) or data-driven (ML and statistical methods) can be found in Vagnoli et al. (2018).

TABLE 3 Various ADAs found on the works included in this systematic review.

#	References	Description	Type
1	Lu et al. (2020)	Cumulative Sum Charts (CUSUM) to automatically detect vibration deviations on a pump.	Vibration-based
2	Shim et al. (2019)	Edge detection algorithms combined with fuzzy logic to automatically analyse images captured *via* UVAs.	Visual-based
3	Perry et al. (2020)	Black Hat Transform and Canny Edge Detector damage detection algorithm in conjunction with a module to automatically track the change of a defect over time based on an Affine Transform. Damage mapping technique to relate the defects on 2D images to the 3D point-cloud by applying camera, intrinsic and extrinsic matrix multiplications	Visual-based
4	Neves et al. (2017)	ANN to predict the expected accelerations of a bridge based on accelerations at previous instant in time.	Vibration-based
5	OBrien et al. (2017)	Implementation of Intrinsic Mode Functions (IMFs) and pseudo frequency component obtained from indirect drive-by monitoring of a bridge.	Vibration-based
6	Kang and Cha (2018)	Deep Convolutional Neural Network (CNN) to analyze the images captured by an UAV and effectively detect concrete cracks.	Visual-based
7	Bao et al. (2019b) and Tang et al. (2019)	Two-steps computer vision and deep learning-based data-driven damage detection method: Transformed registries of time series signals into gray-scale image vectors which were subsequently labeled and used to train a deep neural network (DNN) capable of classifying data pattern anomalies.	Mixed-visual-vibration-based
8	García-Macías and Ubertini (2020)	Automated ADA based upon the pruned exact linear time (PELT) method.	Vibration-based
9	Al-Ghalib (2022)	Pipe-lined methodology combining Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).	Vibration-based
10	Meixedo et al. (2022b)	PCA combined with autoregressive exogenous input (ARX) and clustering algorithms.	Vibration-based
11	Meixedo et al. (2022a)	Continuous Wavelet Transform (CWT) combined with ARX and clustering algorithms.	Vibration-based
12	Febrianto et al. (2022)	Statistical FE modeling and confidence intervals.	Strain-based
13	Weinstein et al. (2018)	Bootstrapping scheme for the training of an ANN.	Strain-based
14	Soman et al. (2018)	Multi-metric data fusion combined with a flexibility index approach.	Mixed-strain-vibration-based[a]
15	Döhler et al. (2018)	Subspace-based residual function and a χ^2-test for a hypothesis testing.	Vibration-based

[a]The use of heterogeneous data fusion particularly in combination with denoising techniques has resulted in better data quality obtained from BHM, as reported by Ravizza et al. (2020)c, from which the different proposed damage detection techniques could benefit.

4.4 UAVs, satellite monitoring, and other DT-related emerging technologies for bridge inspection

One recurrent topic found in this systematic review is the use of UAVs both to capture bridge geometry and generate HBrIM models and to automatically detect damage and decay mechanisms based on visual computing techniques (Mongelli et al., 2017; Roselli et al., 2018). Over the past few years, the use of UAVs has increased thanks to the reduction of their costs, improvement of stability and maneuverability, as well as the development of more efficient visual computing techniques. They provide more advantages than manual inspections in terms of time, accuracy, safety, and costs (Albeaino et al., 2019). Furthermore, the GPS signal loss suffered by UAVs below bridge decks during inspections, has been overcome by the implementation of an array of navigation sensors such as optical, infrared, and ultrasonic sensors as proved by Kang and Cha (2018). The ongoing development of automatizing UAV flights would further boost the use of UAVs for bridge monitoring as it would result in operational time reductions and path re-usability.

For example, Perry et al. (2020) report UAVs as a key element of an automatic streamlined bridge inspection system capable of identifying and locating bridge surface defects, and generating as-built BrIM models for the storage and visualization of damage information. Their methodology includes photogrammetry software (Meshroom) for the creation of 3D point clouds and photorealistic models, Gaussian Mixture Model and Agglomerative Clustering using Python Scikit-learn for element identification along with the use of Revit and Dynamo for the creation of the AI-HBrIM model containing 3D geometry and damage cubes. However, their approach lacks dealing with the integration of FE and structural analysis tools.

Yoon et al. (2022) assess bridge condition based on images captured with a UAV from which damage is automatically detected through a mask Region-based Convolutional Neural Network (R-CNN) algorithm. The methodology proposed by these authors also included a FE model updating module based on a linear stiffness reduction corresponding to the level of bridge condition assessment, as per the damage grades defined in South Korean accepted codes.

The main limitation of visual-based damage detection techniques is that they can not identify sub-surface damages such as reinforcement corrosion and concrete delamination. These techniques need to be complemented with the application of remote sensing technologies such as Ground-Penetrating Radar (GPR), Infrared Thermography (IR) (Xu and Turkan, 2020), or the use of piezoelectric electromechanical sensors that can detect internal damage in a relatively inexpensive way (Na and Baek, 2018). Furthermore, in a comparative study between UAV photogrammetry scanning capabilities against a conventional TLS, Mohammadi et al. (2021) conclude that TLS provides more accurate results and is more suitable for the complex implementation of creating an AI-HBrIM model within the DT paradigm.

Another monitoring technology that has grown in importance over the past few years thanks to its capability for real-time remote monitoring of displacements in bridges is the Interferometry Synthetic Aperture Radar (InSAR). This technology has benefited from the increased number of available satellites and their specialized tools capable of performing millimeter-accuracy measurements. Alani et al. (2020) use InSAR in combination with GPR to assess the integrity of a historical masonry bridge and the effects that local floods have on its displacement seasonal trends. More recently, by taking advantage of the improvements in data processing techniques and the availability of larger SAR databases; Gagliardi et al. (2022) manage to detect the seasonal deformation components of a historical masonry bridge based on an enhanced Multi-Temporal InSAR (MT-InSAR) methodology. One more bridge satellite monitoring case, coupled with hydraulic monitoring of river conditions, is reported by Bianchi et al. (2022).

Regardless of the impressive advancements in UAVs, satellites, and AI applications experienced over the past few years, it is evident that the human component can not be entirely removed from any DT framework. In this regard, Karaaslan et al. (2022) develop a human-centered approach using Mixed Reality (MR) to improve the quality and effectiveness of conventional bridge inspections. This is achieved through the use of Hololens (https://www.microsoft.com/en-us/hololens), which provides the bridge inspector with visual information in real-time about the bridge condition and defects.

4.5 Historical and CH bridges

Historical bridges with CH value require an extra layer of care and special considerations from the part of bridge managers and operators as they not only play a key role in transportation networks but also hold important social, cultural, and artistic values (Pachón et al., 2018). Any intervention performed in this type of bridge must abide by the principles of evidence-based, minimum and incremental intervention, removable and distinguishable measures, and material compatibility established in the Venice Charter (ICOMOS, 1964) and strive to preserve the bridge's authenticity (ICOMOS, 1994). Furthermore, guidelines and recommendations found in the ISCARSAH documents (ICOMOS-ISCARSAH, 2003a; ICOMOS-ISCARSAH, 2003b) and in the Annex I of the ISO 13822 standard (International Organization for Standardization, 2010) must be followed to ensure the correct conservation of such valuable assets.

Interventions on CH bridges must be performed by a multidisciplinary team as shown by the work done by Conde et al. (2017) and Bautista-De Castro et al. (2018). Conde et al. (2017) carry out a comprehensive field survey fully based on non-destructive testing techniques, followed by accurate and detailed 3D FE simulations calibrated using the results obtained from a dynamic identification campaign based on an operational modal analysis approach. Bautista-De Castro et al. (2018) perform TLS, ambient vibration tests, and minor destructive tests. These works result in the detailed assessment of the corresponding bridges' structural condition and the determination of their acceptable safety level. The full adoption of a DT approach is desirable during interventions of CH bridges. A DT would have the ability to monitor in real-time the structure and detect any possible damage induced by the intervention procedure itself as validated by Andersen et al. (2019) in the case of the Henry Hudson I89 Bridge in New York, thus complying with the observational approach suggested by conservation guidance.

Perhaps one of the most advanced tools for damage detection of bridges (in which the CH value is also considered), is the one presented by García-Macías and Ubertini (2020). Their MOVA/MOSS software is capable of automatically performing Operational Modal Analysis (OMA) and system identification through four different techniques: Enhanced Frequency Domain Decomposition (EFDD) and Polyreference Least Squares Complex Frequency Domain method (p-LSCF), both frequency-domain-based, Covariance driven Stochastic Subspace Identification (COV-SSI) and DATA-driven Stochastic Subspace Identification (DATA-SSI), these last two being time-domain-based. Subsequently, it executes frequency tracking and detects changes in the dynamic properties of the structure by applying statistical process control tools, namely, Hotelling, Multivariate Cumulative Sum (MCUSUM), and Multivariate Exponentially Weighted Moving Average (MEWMA). Finally, automatic damage detection is done through the implementation of the Pruned Exact Linear Time (PELT) Method. Their tool, unfortunately, is not part of any DT framework and compatibility issues may arise during integration with other modules of available frameworks.

Along with the conservation of CH value, a sustainable DT framework must as well account for robustness and resilience. Structural robustness is the capability of a structure to sustain a certain amount of damage without suffering full collapse, whereas the resilience of a structure refers to its ability to resist and withstand sudden shocks and sustained stresses, as well as its ability to recover from damage and continue to function effectively (Hajdin et al., 2018). Assessing both the robustness and resilience of a bridge implies dealing with a series of uncertainties related to material resistances and external loads. An adequate assessment methodology could be, as suggested by Futai et al. (2022), the implementation of reliability-based and risk-based performance indicators. It has been proved that modeling uncertainties can greatly be reduced by the adoption of a DT framework (Rojas-Mercedes et al., 2022).

5 Conclusion

Although the scope of not all research works found in this systematic review encompass the digital twin paradigm, and regardless of all the challenges and limitations still in place for its full deployment and implementation in real practice, there seems to be only one school of thought and a consensus toward the future adoption of digital twins for bridge design, management, and operation among the scientific community and bridge practitioners.

A suitable digital twin framework capable of accounting for the cultural heritage of existing bridges would be primarily based on the creation of an as-is historical bridge information model with fully inter-operable data, geometry, finite element, and data-driven modules. The as-is historical bridge information model would be kept interlinked to its physical asset counterpart through the implementation of a multi-metric bridge health monitoring system that constantly generates data about the structural, environmental, and operational conditions of the bridge. That data would be effectively generated by optimized sampling methodologies and would pass through an intermediate fog computing layer before its final processing at a cloud computing service.

Current research gaps in the practical development and implementation of digital twins are mainly related to i) the lack of interoperability among the different proprietary and open-source software used along the digital twin model generation pipeline; ii) performance improvement of currently available anomaly detection algorithms; and iii) the direction for the creation of macro-digital twins that integrate digital twins of individual infrastructure assets. The latter needs to be determined and the benefits/drawbacks of whether it is done at geographical, systemic, or another kind of ontological integration approach, needs to be assessed.

The digital twin paradigm was born within the Industry 4.0 era. Future potential developments in the field are related to the implementation of Industry 5.0 concepts and ideas within digital twin frameworks such as sustainability, human-centrism, and resilience (European Commission et al., 2022).

We are moving to a digital age where physical assets will increasingly have digital representations. New opportunities arise. Looking towards the future, it appears that digital twins in the built environment have the potential to play a significant role, in a wide range of applications throughout their entire life cycle. This is evidenced by the fast-growing occurrence of digital twin-related articles in the scientific literature during recent years. This study concludes that although some of the technologies discussed are relatively new and there are certainly several challenges to tackle, it has great potential to become an extremely positive force of change in the architecture, engineering, and construction industry. As engineers, it is our responsibility to facilitate the digital transformation of the architecture, engineering, and construction industry and to make it ready for the challenges and opportunities of the future, and digital twins are bound to play a pivotal role in this transformation.

Author contributions

All authors contributed to the conception and design of the study. AJ wrote the protocol and search strategy, carried out the search, organized the data and created the database, performed the bibliometric analysis, did the synthesis of the literature, and wrote the first draft of the manuscript. VP and MN supervised the work. All authors contributed to the manuscript revision, and read, and approved the submitted version. All authors have as well contributed to obtaining the necessary funding to carry out this work.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 101066739. The APC was funded by Oslo Metropolitan University.

Acknowledgments

The content of the manuscript has previously appeared online in the form of a preprint Jiménez Rios et al. (2023c).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alani, A. M., Tosti, F., Ciampoli, L. B., Gagliardi, V., and Benedetto, A. (2020). An integrated investigative approach in health monitoring of masonry arch bridges using gpr and insar technologies. *NDT E Int.* 115, 102288. doi:10.1016/j.ndteint.2020.102288

Albeaino, G., Gheisari, M., and Franz, B. W. (2019). A systematic review of unmanned aerial vehicle application areas and technologies in the aec domain. *J. Inf. Technol. Constr.* 24, 381–405. doi:10.36680/j.itcon.2019.020

Al-Ghalib, A. A. (2022). Structural damage detection of old ada steel truss bridge using vibration data. *Struct. Control Health Monit.* 29. doi:10.1002/stc.3098

An, Y., Chatzi, E., Sim, S.-H., Laflamme, S., Blachowski, B., and Ou, J. (2019). Recent progress and future trends on damage identification methods for bridge structures. *Struct. Control Health Monit.* 26. doi:10.1002/stc.2416

Andersen, J. E., Rex, S., and Gala, A., Hardesty & Hanover (2019). *Structural health monitoring of henry hudson i89*. Lyngby, Denmark: International Association for Bridge and Structural Engineering IABSE, 2121–2131. COWI.

ASCE (2021). *2021 report card for America's infrastructure*. American Society of Civil Engineers.

Bao, Y., Chen, Z., Wei, S., Xu, Y., Tang, Z., and Li, H. (2019a). The state of the art of data science and engineering in structural health monitoring. *Engineering* 5, 234–242. doi:10.1016/j.eng.2018.11.027

Bao, Y., Tang, Z., Li, H., and Zhang, Y. (2019b). Computer vision and deep learning-based data anomaly detection method for structural health monitoring. *Struct. Health Monit.* 18, 401–421. doi:10.1177/1475921718757405

Bautista-De Castro, A., Sánchez-Aparicio, L. J., Ramos, L. F., Sena-Cruz, J., and González-Aguilera, D. (2018). Integrating geomatic approaches, operational modal analysis, advanced numerical and updating methods to evaluate the current safety conditions of the historical bôco bridge. *Constr. Build. Mater.* 158, 961–984. doi:10.1016/j.conbuildmat.2017.10.084

Bianchi, S., Biondini, F., D'Angelo, M., Ballio, F., Anghileri, M., Rosati, G., et al. (2022). "Satellite-based structural and hydraulic monitoring of a 50-year-old bridge over the oglio river in Italy," in *Lecture Notes in Civil engineering*, 200, 380–389. doi:10.1007/978-3-030-91877-4_44

Bouzas, O., Cabaleiro, M., Conde, B., Cruz, Y., and Riveiro, B. (2022). Structural health control of historical steel structures using hbim. *Autom. Constr.* 140, 104308. doi:10.1016/j.autcon.2022.104308

Broadus, R. N. (1987). Toward a definition of "bibliometrics". *Scientometrics* 12, 373–379. doi:10.1007/BF02016680

Burrello, A., Brunelli, D., Malavisi, M., and Benini, L. (2020). "Enhancing structural health monitoring with vehicle identification and tracking," in 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) (Zurich, Switzerland: DEI, University of Bologna IIS, ETH, Institute of Electrical and Electronics Engineers Inc.), 9128641. doi:10.1109/I2MTC43012.2020.9128641

Canada Infrastructure (2019). *Canadian infrastructure report card*. Available at: http://canadianinfrastructure.ca/en/index.html.

Cervenka, J., Jendele, L., Zalsky, J., Pukl, R., Novak, D., Zhao, B., et al. (2020). *Digital twin approach for durability and reliability assessment of bridges*. Brno, Czech Republic: Dep. of Civil Engineering, Technical University of Brno, International Federation for Structural Concrete, 1840–1848.

Conde, B., Ramos, L. F., Oliveira, D. V., Riveiro, B., and Solla, M. (2017). Structural assessment of masonry arch bridges by combination of non-destructive testing techniques and three-dimensional numerical modelling: Application to vilanova bridge. *Eng. Struct.* 148, 621–638. doi:10.1016/j.engstruct.2017.07.011

Costin, A., Adibfar, A., Hu, H., and Chen, S. S. (2018). Building information modeling (bim) for transportation infrastructure - literature review, applications, challenges, and recommendations. *Autom. Constr.* 94, 257–281. doi:10.1016/j.autcon.2018.07.001

Dan, D., Ying, Y., and Ge, L. (2022). Digital twin system of bridges group based on machine vision fusion monitoring of bridge traffic load. *IEEE Trans. Intell. Transp. Syst.* 23, 22190–22205. doi:10.1109/TITS.2021.3130025

Dang, H., Tatipamula, M., and Nguyen, H. X. (2022). Cloud-based digital twinning for structural health monitoring using deep learning. *IEEE Trans. Indust. Inf.* 18, 3820–3830. doi:10.1109/TII.2021.3115119

Del Rio, I., Cabaleiro, M., Conde, B., Riveiro, B., Caamaño, J. C., El Naggar, H., et al. (2020). *Hbim application to historical steel structures: The case study of lapela bridge*. Vigo, CP 36208, Spain: Department of Materials Engineering Applied Mechanics and Construction, School of Industrial Engineering, University of Vigo, Avestia Publishing, 160–1–160–7. doi:10.11159/icsect20.160

Döhler, M., Hille, F., and Mevel, L. (2018). Vibration-based monitoring of civil structures with subspace-based damage detection. *Intelligent Syst. Control Autom. Sci. Eng.* 92, 307–326. doi:10.1007/978-3-319-68646-2_14

Elsevier (2023). *Scopus*. Elsevier.

European Commission; Joint Research CentreGkoumas, K., Balen, M., Grosso, M., Pekár, F., et al. (2019). *Research and innovation in bridge maintenance, inspection and monitoring: A European perspective based on the transport research and innovation monitoring and information system (TRIMIS)*. Publications Office. doi:10.2760/16174

European Commission; Directorate-General for Research and InnovationRenda, A., Schwaag Serger, S., Tataj, D., Morlet, A., et al. (2022). *Industry 5.0, a transformative vision for europe: Governing systemic transformations towards a sustainable industry*. Publications Office of the European Union. doi:10.2777/17322

EuroStruct (2020). *Risk-based monitoring and assessment of bridges*. YouTube.

Febrianto, E., Butler, L., Girolami, M., and Cirak, F. (2022). Digital twinning of self-sensing structures using the statistical finite element method. *Data-Centric Eng.* 3, e31. doi:10.1017/dce.2022.28

Feng, D., and Feng, M. Q. (2017). Experimental validation of cost-effective vision-based structural health monitoring. *Mech. Syst. Signal Process.* 88, 199–211. doi:10.1016/j.ymssp.2016.11.021

Futai, M. M., Bittencourt, T. N., Carvalho, H., and Ribeiro, D. M. (2022). Challenges in the application of digital transformation to inspection and maintenance of bridges. *Struct. Infrastruct. Eng.* 18, 1581–1600. doi:10.1080/15732479.2022.2063908

Gagliardi, V., Ciampoli, L. B., D'Amico, F., Alani, A. M., Tosti, F., Benedetto, A., et al. (2022). "Remote sensing measurements for the structural monitoring of historical masonry bridges," in *Lecture Notes in Civil engineering*, 200, 632–641. doi:10.1007/978-3-030-91877-4_72

García-Macías, E., and Ubertini, F. (2020). Mova/moss: Two integrated software solutions for comprehensive structural health monitoring of structures. *Mech. Syst. Signal Process.* 143, 106830. doi:10.1016/j.ymssp.2020.106830

Ghahari, F., Malekghaini, N., Ebrahimian, H., and Taciroglu, E. (2022). Bridge digital twinning using an output-only bayesian model updating method and recorded seismic measurements. *Sensors* 22, 1278. doi:10.3390/s22031278

Giorgadze, I. M., Vahdatikhaki, F., and Voordijk, J. H. (2022). "Conceptual modeling of lifecycle digital twin architecture for bridges: A data structure approach," in 39th International Symposium on Automation and Robotics in Construction (ISARC 2022) (Netherlands: Department of Construction Management Engineering, University of Twente, International Association for Automation and Robotics in Construction IAARC), 199–206.

Gonen, S., and Soyoz, S. (2021). "Dynamic identification of masonry arch bridges using multiple methodologies," in *Special topics in structural dynamics & experimental techniques*. Editor D. S. Epp (Cham: Springer International Publishing), Vol. 5, 37–47.

Gonen, S., Demirlioglu, K., and Erduran, E. (2023). "Modal identification of a railway bridge under train crossings: A comparative study," in *Dynamics of Civil structures*. Editors H. Y. Noh, M. Whelan, and P. S. Harvey (Cham: Springer International Publishing), Vol. 2, 33–40.

Hajdin, R., Powers, N., Frangopol, D. M., Al-Mahaidi, R., and Caprani, C. (2018). *Managing existing bridges - on the brink of an exciting future*. Faculty of Civil Engineering, University of Belgrade, CRC Press/Balkema, 63–87.

Hosamo, H. H., and Hosamo, M. H. (2022). Digital twin technology for bridge maintenance using 3d laser scanning: A review. *Adv. Civ. Eng.* 2022, 1–15. doi:10.1155/2022/2194949

ICOMOS (1964). *International charter for the conservation and restoration of monuments and sites*. London: ICOMOS.

ICOMOS (1994). "The nara document on authenticity," in Ratified at Nara Conference on Authenticity in Relation to the World Heritage Convention, held at Nara, Japan, from 1-6 November, 1–6.

ICOMOS-ISCARSAH (2003a). "Principles for the analysis, conservation and structural restoration of architectural heritage," in Ratified by the ICOMOS 14th General Assembly Victoria Falls, Zimbabwe.

ICOMOS-ISCARSAH (2003b). *Recommendations for the analysis, conservation and structural restoration of architectural heritage*. Available at: https://iscarsah.files.wordpress.com/2023/01/part-ii-e28093-guidelines.pdf.

IFC (2023). *Ifc bridge design to construction information exchange*. IFC4 documentation.

International Organization for Standardization (2010). *Iso 13822:2010 bases for design of structures - assessment of existing structures*. Available at: https://www.iso.org/standard/46556.html.

International Organization for Standardization (2016). *Iso 8608:2016 mechanical vibration: Road surface profiles: Reporting of measured data*. Available at: https://www.iso.org/standard/71202.html.

Jeong, S., Hou, R., Lynch, J. P., Sohn, H., and Law, K. H. (2017). An information modeling framework for bridge monitoring. *Adv. Eng. Softw.* 114, 11–31. doi:10.1016/j.advengsoft.2017.05.009

Jeong, S., Hou, R., Lynch, J. P., Sohn, H., and Law, K. H. (2019). A scalable cloud-based cyberinfrastructure platform for bridge monitoring. *Struct. Infrastructure Eng.* 15, 82–102. doi:10.1080/15732479.2018.1500617

Jiang, F., Ding, Y., Song, Y., Geng, F., and Wang, Z. (2021). Digital twin-driven framework for fatigue life prediction of steel bridges using a probabilistic multiscale model: Application to segmental orthotropic steel deck specimen. *Eng. Struct.* 241, 112461. doi:10.1016/j.engstruct.2021.112461

Jiménez Rios, A., and O'Dwyer, D. (2019). "External post-tensioning system for the strengthening of historical stone masonry bridges," in *Structural analysis of historical constructions*. Editors R. Aguilar, D. Torrealva, S. Moreira, M. A. Pando, and L. F. Ramos (Cham: Springer International Publishing), 1566–1574.

Jiménez Rios, A., Plevris, V., and Nogal, M. (2023a). *Bibliographic data from the digital twin anomaly detection decision-making for bridge management systematic review*. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska- Curie grant agreement No 101066739. doi:10.5281/zenodo.7548017

Jiménez Rios, A., Plevris, V., and Nogal, M. (2023b). *Digital twin anomaly detection decision-making for bridge management (dtadd)*. doi:10.17605/OSF.IO/SH9B2

Jiménez Rios, A., Plevris, V., and Nogal, M. (2023c). *Dtadd systematic review preprint*. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska- Curie grant agreement No 101066739. doi:10.5281/zenodo.7673718

Jiménez Rios, A., Plevris, V., and Nogal, M. (2023d). *Dtadd systematic review protocol (amended)*. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska- Curie grant agreement No 101066739. doi:10.5281/zenodo.7546576

Jiménez Rios, A., Plevris, V., and Nogal, M. (2023e). *Dtadd systematic review search strategy (amended)*. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska- Curie grant agreement No 101066739. doi:10.5281/zenodo.7546557

Kang, D., and Cha, Y.-J. (2018). Autonomous uavs for structural health monitoring using deep learning and an ultrasonic beacon system with geo-tagging. *Computer-Aided Civ. Infrastruct. Eng.* 33, 885–902. doi:10.1111/mice.12375

Kang, J.-S., Chung, K., and Hong, E. J. (2021). Multimedia knowledge-based bridge health monitoring using digital twin. *Multimedia Tools Appl.* 80, 34609–34624. doi:10.1007/s11042-021-10649-x

Karaaslan, E., Zakaria, M., and Catbas, F. N. (2022). *Mixed reality-assisted smart bridge inspection for future smart cities*. Orlando, FL, United States: Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Elsevier, 261–280. doi:10.1016/B978-0-12-817784-6.00002-3

Kim, C.-W., Zhang, F.-L., Chang, K.-C., McGetrick, P. J., and Goi, Y. (2021). Ambient and vehicle-induced vibration data of a steel truss bridge subject to artificial damage. *J. Bridge Eng.* 26. doi:10.1061/(ASCE)BE.1943-5592.0001730

Kostic, B., and Gül, M. (2017). Vibration-based damage detection of bridges under varying temperature effects using time-series analysis and artificial neural networks. *J. Bridge Eng.* 22. doi:10.1061/(ASCE)BE.1943-5592.0001085

Koutsantonis, D., Koutsantonis, K., Bakas, N. P., Plevris, V., Langousis, A., and Chatzichristofis, S. A. (2022). Bibliometric literature review of adaptive learning systems. *Sustainability* 14, 12684. doi:10.3390/su141912684

Li, L., Lei, B., and Mao, C. (2022). Digital twin in smart manufacturing. *J. Indust. Inf. Integr.* 26, 100289. doi:10.1016/j.jii.2021.100289

Limongelli, M. P., Dolce, M., Spina, D., Guéguen, P., Langlais, M., Wolinieck, D., et al. (2019). "S 2 hm in some European countries," in *Springer tracts in Civil engineering*, 303–343. doi:10.1007/978-3-030-13976-6_13

Liu, Q., Zhang, H., Leng, J., and Chen, X. (2019). Digital twin-driven rapid individualised designing of automated flow-shop manufacturing system. *Int. J. Prod. Res.* 57, 3903–3919. doi:10.1080/00207543.2018.1471243

Locke, W., Sybrandt, J., Redmond, L., Safro, I., and Atamturktur, S. (2020). Using drive-by health monitoring to detect bridge damage considering environmental and operational effects. *J. Sound Vib.* 468, 115088. doi:10.1016/j.jsv.2019.115088

Lu, R., and Brilakis, I. (2019). Digital twinning of existing reinforced concrete bridges from labelled point clusters. *Autom. Constr.* 105, 102837. doi:10.1016/j.autcon.2019.102837

Lu, R., Brilakis, I., and Middleton, C. R. (2019). Detection of structural components in point clouds of existing rc bridges. *Computer-Aided Civ. Infrastruct. Eng.* 34, 191–212. doi:10.1111/mice.12407

Lu, Q., Parlikad, A. K., Woodall, P., Don Ranasinghe, G., Xie, X., Liang, Z., et al. (2020). Developing a digital twin at building and city levels: Case study of west cambridge campus. *J. Manag. Eng.* 36. doi:10.1061/(ASCE)ME.1943-5479.0000763

Mao, J.-X., Wang, H., Feng, D.-M., Tao, T.-Y., and Zheng, W.-Z. (2018). Investigation of dynamic properties of long-span cable-stayed bridges based on one-year monitoring data under normal operating condition. *Struct. Control Health Monit.* 25, e2146. doi:10.1002/stc.2146

McKenna, T., Minehane, M., O'Keeffe, B., O'Sullivan, G., and Ruane, K. (2017). Bridge information modelling (brim) for a listed viaduct. *Proc. Institution Civ. Eng. Bridge Eng.* 170, 192–203. doi:10.1680/jbren.16.00007

Meixedo, A., Ribeiro, D., Santos, J., Calçada, R., and Todd, M. D. (2022a). Real-time unsupervised detection of early damage in railway bridges using traffic-induced responses. *Struct. Integr.* 21, 117–142. doi:10.1007/978-3-030-81716-9_6

Meixedo, A., Santos, J., Ribeiro, D., Calçada, R., and Todd, M. D. (2022b). Online unsupervised detection of structural changes using train-induced dynamic responses. *Mech. Syst. Signal Process.* 165, 108268. doi:10.1016/j.ymssp.2021.108268

Mohammadi, M., Rashidi, M., Mousavi, V., Karami, A., Yu, Y., and Samali, B. (2021). "Case study on accuracy comparison of digital twins developed for a heritage bridge via uav photogrammetry and terrestrial laser scanning," in 10th International Conference on Structural Health Monitoring of Intelligent Infrastructure, SHMII 10 (Italy: Department of Information Engineering and Computer Science, University of Trento, International Society for Structural Health Monitoring of Intelligent Infrastructure, ISHMII), 1713–1720.

Mongelli, M., De Canio, G., Roselli, I., Malena, M., Nacuzi, A., De Felice, G., et al. (2017). 3D photogrammetric reconstruction by drone scanning for FE analysis and crack pattern mapping of the "bridge of the towers", spoleto. *Key Eng. Mat.*747, 423–430. doi:10.4028/www.scientific.net/kem.747.423

Moughty, J. J., and Casas, J. R. (2017). A state of the art review of modal-based damage detection in bridges: Development, challenges, and solutions. *Appl. Sci. Switz.* 7, 510. doi:10.3390/app7050510

Na, W. S., and Baek, J. (2018). A review of the piezoelectric electromechanical impedance based structural health monitoring technique for engineering structures. *Sensors Switz.* 18, 1307. doi:10.3390/s18051307

Neves, A. C., González, I., Leander, J., and Karoumi, R. (2017). Structural health monitoring of bridges: A model-free ann-based approach to damage detection. *J. Civ. Struct. Health Monit.* 7, 689–702. doi:10.1007/s13349-017-0252-5

Nili, M. H., Taghaddos, H., and Zahraie, B. (2021). Integrating discrete event simulation and genetic algorithm optimization for bridge maintenance planning. *Automation Constr.* 122, 103513. doi:10.1016/j.autcon.2020.103513

Noel, A. B., Abdaoui, A., Elfouly, T., Ahmed, M. H., Badawy, A., and Shehata, M. S. (2017). Structural health monitoring using wireless sensor networks: A comprehensive survey. *IEEE Commun. Surv. Tutorials* 19, 1403–1423. doi:10.1109/COMST.2017.2691551

Obrien, E. J., Malekjafarian, A., and González, A. (2017). Application of empirical mode decomposition to drive-by bridge damage detection. *Eur. J. Mech. A/Solids* 61, 151–163. doi:10.1016/j.euromechsol.2016.09.009

OpenBrIM (2023). "Openbrim platform where bridge engineering happens," in *OpenBrIM platform - engineering on cloud*.

Pachón, P., Castro, R., García-Macías, E., Compan, V., and Puertas, E. (2018). E. torroja's bridge: Tailored experimental setup for shm of a historical bridge with a reduced number of sensors. *Eng. Struct.* 162, 11–21. doi:10.1016/j.engstruct.2018.02.035

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021a). The prisma 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372, n71. doi:10.1136/bmj.n71

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021b). Prisma 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ* 372, n160. doi:10.1136/bmj.n160

Park, S. I., Park, J., Kim, B.-G., and Lee, S.-H. (2018). Improving applicability for information model of an ifc-based steel bridge in the design phase using functional meanings of bridge components. *Appl. Sci. Switz.* 8, 2531. doi:10.3390/app8122531

Pepe, M., Costantino, D., and Garofalo, A. R. (2020). An efficient pipeline to obtain 3d model for hbim and structural analysis purposes from 3d point clouds. *Appl. Sci. Switz.* 10, 1235. doi:10.3390/app10041235

Perry, B. J., Guo, Y., Atadero, R., and van de Lindt, J. W. (2020). Streamlined bridge inspection system utilizing unmanned aerial vehicles (uavs) and machine learning. *Meas. J. Int. Meas. Confed.* 164, 108048. doi:10.1016/j.measurement.2020.108048

Petzet, M. (2004). *International Charters for Conservation and restoration = Chartes Internationales sur la conservation et la restauration = Cartas internacionales sobre la conservacion y la restauracion*. ICOMOS.

PIARC (2023). *Piarc asset management manual*. World Road Association PIARC. Asset Management Manual — Asset Management ManualA guide for practitioners.

Polania, D. R., Tondolo, F., Osello, A., Fonsati, A., De Gaetani, C., Trincianti, C., et al. (2022). "Digitalization processes and bridge information modeling for existing bridges," in *Lecture Notes in Civil engineering*, 200, 944–953. doi:10.1007/978-3-030-91877-4_108

Ramancha, M. K., Astroza, R., Conte, J. P., Restrepo, J. I., Todd, M. D., and Mao, Z. (2020). *Bayesian nonlinear finite element model updating of a full-scale bridge-column using sequential Monte Carlo*. Santiago, Chile: Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Springer, 389–397. doi:10.1007/978-3-030-47638-0_43

Rashidi, A., and Karan, E. (2018). Video to brim: Automated 3d as-built documentation of bridges. *J. Perform. Constr. Facil.* 32. doi:10.1061/(ASCE)CF.1943-5509.0001163

Ravizza, G., Ferrari, R., Rizzi, E., Dertimanis, V., Chatzi, E. N., Papadrakakis, M., et al. (2020). "An integrated monitoring strategy for current condition assessment of historic bridges," in XI International Conference on Structural Dynamics (Stefano-Franscini-Platz 5, Zürich, CH-8093, Switzerland: ETH Zürich, Institute of Structural Engineering, Department of Civil, Environmental and Geomatic Engineering, European Association for Structural Dynamics), 1373–1387.

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., et al. (2021). Prisma-s: An extension to the prisma statement for reporting literature searches in systematic reviews. *Syst. Rev.* 10, 39. doi:10.1186/s13643-020-01542-z

Ri, S., Tsuda, H., Chang, K., Hsu, S., Lo, F., and Lee, T. (2020). Dynamic deformation measurement by the sampling moiré method from video recording and its application to bridge engineering. *Exp. Tech.* 44, 313–327. doi:10.1007/s40799-019-00358-4

Rojas-Mercedes, N., Erazo, K., and Di Sarno, L. (2022). Seismic fragility curves for a concrete bridge using structural health monitoring and digital twins. *Earthq. Struct.* 22, 503–515. doi:10.12989/eas.2022.22.5.503

Roselli, I., Malena, M., Mongelli, M., Cavalagli, N., Gioffrè, M., De Canio, G., et al. (2018). Health assessment and ambient vibration testing of the "ponte delle torri" of spoleto during the 2016-2017 central ItItalyeismic sequence. *J. Civ. Struct. Health Monit.* 8, 199–216. doi:10.1007/s13349-018-0268-5

Saback de Freitas Bello, V., Popescu, C., Blanksvärd, T., Täljsten, B., Pellegrino, C., Faleschini, F., et al. (2022). "Framework for bridge management systems (bms) using digital twins," in *Lecture Notes in Civil engineering*, 200, 687–694. doi:10.1007/978-3-030-91877-4_78

Saju, S. (2022). "Smart bridges based on bridge information modelling," in 2022 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES) (Trivandrum, India: Bridges design unit, chief design office, Kerala Public Works Department, Institute of Electrical and Electronics Engineers Inc.), 279–284. doi:10.1109/SPICES52834.2022.9774179

Schleich, B., Anwer, N., Mathieu, L., and Wartzack, S. (2017). Shaping the digital twin for design and production engineering. *CIRP Ann. - Manuf. Technol.* 66, 141–144. doi:10.1016/j.cirp.2017.04.040

Shabani, A., Skamantzari, M., Tapinaki, S., Georgopoulos, A., Plevris, V., and Kioumarsi, M. (2022). 3d simulation models for developing digital twins of heritage structures: Challenges and strategies. *Procedia Struct. Integr.* 37, 314–320. doi:10.1016/j.prostr.2022.01.090

Shao, S., Zhou, Z., Deng, G., Du, P., Jian, C., and Yu, Z. (2020). Experiment of structural geometric morphology monitoring for bridges using holographic visual sensor. *Sensors Switz.* 20, 1187. doi:10.3390/s20041187

Shao, S., Deng, G., Zhou, Z., Rizzo, P., and Milazzo, A. (2021). Structural geometric morphology monitoring for bridges using holographic visual sensor. *Lect. Notes Civ. Eng.* 128, 3–13. doi:10.1007/978-3-030-64908-1_1

Shim, C.-S., Dang, N.-S., Lon, S., and Jeon, C.-H. (2019). Development of a bridge maintenance system for prestressed concrete bridges using 3d digital twin model. *Struct. Infrastructure Eng.* 15, 1319–1332. doi:10.1080/15732479.2019.1620789

Solorzano, G., and Plevris, V. (2022). Computational intelligence methods in simulation and modeling of structures: A state-of-the-art review using bibliometric maps. *Front. Built Environ.* 8. doi:10.3389/fbuil.2022.1049616

Soman, R., Kyriakides, M., Onoufriou, T., and Ostachowicz, W. (2018). Numerical evaluation of multi-metric data fusion based structural health monitoring of long span bridge structures. *Struct. Infrastructure Eng.* 14, 673–684. doi:10.1080/15732479.2017.1350984

Sreevallabhan, K., Chand, B. N., Ramasamy, S., Vijayaraghavan, R., Senthilnathan, K., Saral, A. M., et al. (2017). "Structural health monitoring using wireless sensor networks," in IEEE Communications Surveys & Tutorials (Vellore-, Tamil Nadu, 632014, India: School of Electrical Engineering, VIT University, Institute of Physics Publishing). doi:10.1088/1757-899X/263/5/052015

Sun, L., Shang, Z., Xia, Y., Bhowmick, S., and Nagarajaiah, S. (2020). Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection. *J. Struct. Eng. (United States)* 146. doi:10.1061/(ASCE)ST.1943-541X.0002535

Tang, Z., Chen, Z., Bao, Y., and Li, H. (2019). Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring. *Struct. Control Health Monit.* 26, e2296. doi:10.1002/stc.2296

Ubertini, F., Comanducci, G., Cavalagli, N., Laura Pisello, A., Luigi Materazzi, A., and Cotana, F. (2017). Environmental effects on natural frequencies of the san pietro bell tower in perugia, Italy, and their removal for structural performance assessment. *Mech. Syst. Signal Process.* 82, 307–322. doi:10.1016/j.ymssp.2016.05.025

Vagnoli, M., Remenyte-Prescott, R., and Andrews, J. (2018). Railway bridge structural health monitoring and fault detection: State-of-the-art methods and future challenges. *Struct. Health Monit.* 17, 971–1007. doi:10.1177/1475921717721137

van Eck, N. J., and Waltman, L. (2014). *Visualizing bibliometric networks*. Cham: Springer International Publishing, 285–320. doi:10.1007/978-3-319-10377-8_13

van Eck, N. J., and Waltman, L. (2022). *Vosviewer manual*. Manual for VOSviewer version 1.6.18. 1.

Wang, X., Niederleithinger, E., and Hindersmann, I. (2022). The installation of embedded ultrasonic transducers inside a bridge to monitor temperature and load influence using coda wave interferometry technique. *Struct. Health Monit.* 21, 913–927. doi:10.1177/14759217211014430

Weinstein, J. C., Sanayei, M., and Brenner, B. R. (2018). Bridge damage identification using artificial neural networks. *J. Bridge Eng.* 23. doi:10.1061/(ASCE)BE.1943-5592.0001302

World Heritage Centre (2023). *World heritage list*. UNESCO World Heritage Centre.

Xiao, R., Lian, Y., Sun, B., Zhao, X., Liu, Z., Tang, P., et al. (2017). "Method of bridge structural analysis based on bridge information modeling," in ASCE International Workshop on Computing in Civil Engineering 2017 (United States: School of Sustainable Engineering and the Built Environment, Arizona State Univ., American Society of Civil Engineers ASCE), 92–100. P.O. Box 873005, Tempe, AZ 85287-3005. doi:10.1061/9780784480823.012

Xu, Y., and Turkan, Y. (2020). Brim and uas for bridge inspections and management. *Eng. Constr. Archit. Manag.* 27, 785–807. doi:10.1108/ECAM-12-2018-0556

Xuequan, M. (2018). *Morandi bridge absence costs Italy "784 mln euros a year in lost gdp"*. Report. Xinhua.

Ye, X. W., Jin, T., and Yun, C. B. (2019). A review on deep learning-based structural health monitoring of civil infrastructures. *Smart Struct. Syst.* 24, 567–585. doi:10.12989/sss.2019.24.5.567

Ye, S., Lai, X., Bartoli, I., and Aktan, A. E. (2020). Technology for condition and performance evaluation of highway bridges. *J. Civ. Struct. Health Monit.* 10, 573–594. doi:10.1007/s13349-020-00403-6

Ye, C., Kuok, S.-C., Butler, L. J., and Middleton, C. R. (2022). Implementing bridge model updating for operation and maintenance purposes: Examination based on UK practitioners' views. *Struct. Infrastructure Eng.* 18, 1638–1657. doi:10.1080/15732479.2021.1914115

Yoon, H., Shin, J., and Spencer, B. F. (2018). Structural displacement measurement using an unmanned aerial system. *Computer-Aided Civ. Infrastruct. Eng.* 33, 183–192. doi:10.1111/mice.12338

Yoon, S., Lee, S., Kye, S., Kim, I.-H., Jung, H.-J., and Spencer, B. F. (2022). Seismic fragility analysis of deteriorated bridge structures employing a uav inspection-based updated digital twin. *Struct. Multidiscip. Optim.* 65, 346. doi:10.1007/s00158-022-03445-0

Yu, S., Li, D., and Ou, J. (2022). Digital twin-based structure health hybrid monitoring and fatigue evaluation of orthotropic steel deck in cable-stayed bridge. *Struct. Control Health Monit.* 29. doi:10.1002/stc.2976

Zhang, Z., and Sun, C. (2021). Structural damage identification via physics-guided machine learning: A methodology integrating pattern recognition with finite element model updating. *Struct. Health Monit.* 20, 1675–1688. doi:10.1177/1475921720927488

Zhang, H., Liu, Q., Chen, X., Zhang, D., and Leng, J. (2017). A digital twin-based approach for designing and multi-objective optimization of hollow glass production line. *IEEE Access* 5, 26901–26911. doi:10.1109/ACCESS.2017.2766453

Glossary

ADA Anomaly Detection Algorithm

AI Artificial Intelligence

ANN Artificial Neural Network

AI-HBrIM As-Is Historical Bridge Information Modeling

BIM Building Information Modeling

BHM Bridge Health Monitoring

BrIM Bridge Information Modelling

CH Cultural Heritage

CC Cloud Computing

CWI Coda Wave Interferometry

COV-SSI Covariance driven Stochastic Subspace Identification

DATA-SSI DATA-driven Stochastic Subspace Identification

DIC Digital Image Correlation

DES Discrete Event Simulation

DT Digital Twin

EMD Empirical Mode Decomposition

EFDD Enhanced Frequency Domain Decomposition

FC Fog Computing

FE Finite Element

GA Genetic Algorithms

GPR Ground-Penetrating Radar

HBIM Historical BIM

IFC Industry Foundation Classes

IR Infrared Thermography

IaaS Infrastructure as a Service

InSAR Interferometry Synthetic Aperture Radar

LOD Level of Detail

ML Machine Learning

MR Mixed Reality

MT-InSAR Multi-Temporal InSAR

MCUSUM Multivariate Cumulative Sum

MEWMA Multivariate Exponentially Weighted Moving Average

NFCR Normalized Frequency Change Ratios

OSF Open Science Framework

OMA Operational Modal Analysis

OCM Orientation Code Matching

PMVS Patch-based Multi-View Stereo

PaaS Platform as a Service

p-LSCF Polyreference Least Squares Complex Frequency Domain

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PELT Pruned Exact Linear Time

R-CNN Region-based Convolutional Neural Network

SMC Sequential Monte Carlo

SaaS Software as a Service

SHM Structural Health Monitoring

TLS Terrestrial Laser Scanner

UAVs Unmanned Aerial Vehicles

UCC Up-sampled Cross Correlation

VM Virtual Machine

WSN Wireless Sensor Network.

frontiers | Frontiers in Built Environment

TYPE Original Research
PUBLISHED 08 June 2023
DOI 10.3389/fbuil.2023.1205235

Check for updates

OPEN ACCESS

EDITED BY
Sakdirat Kaewunruen,
University of Birmingham,
United Kingdom

REVIEWED BY
Filippo Pratico',
Mediterranea University of Reggio Calabria, Italy
Di Wang,
Aalto University, Finland

*CORRESPONDENCE
Muhammad Amjad Afridi,
afridi@kth.se

RECEIVED 13 April 2023
ACCEPTED 25 May 2023
PUBLISHED 08 June 2023

CITATION
Afridi MA, Erlingsson S and Sjögren L (2023), Municipal street maintenance challenges and management practices in Sweden.
Front. Built Environ. 9:1205235.
doi: 10.3389/fbuil.2023.1205235

Municipal street maintenance challenges and management practices in Sweden

Muhammad Amjad Afridi[1,2]*, Sigurdur Erlingsson[2,3,4] and Leif Sjögren[3]

[1]Skellefteå Municipality, Skellefteå, Sweden, [2]Department of Building Materials, KTH Royal Institute of Technology, Stockholm, Sweden, [3]Swedish National Road and Transport Research Institute (VTI), Linköping, Sweden, [4]Faculty of Civil and Environmental Engineering, University of Iceland, Reykjavik, Iceland

The municipal street network acts as a multifunctional asset by providing people, vehicles and public services with a well-functioning infrastructure. To keep it in good condition, optimal maintenance measures are required which would result in an efficient use of taxpayers' money. This paper investigates the street network deterioration processes and the management practices that the municipal administrations have applied in Sweden. The study is based on a survey with Swedish municipalities using questionnaires and complementary interviews. The answers provide insight into a wide range of common pavement distresses and deterioration factors, along with pavement management practices. The study identifies that potholes, surface unevenness and alligator cracking are the most cited challenges, while pavement ageing, heavy traffic and patches are the most noted causes. Similarly, the cold climate and population density are influential factors in pavement deterioration. Allocation of the maintenance and rehabilitation and reconstruction budget is higher in the northern part of the country as well as in densely populated municipalities. Condition data collection and use of commercial Pavement Management Systems (PMS) are limited. Addressing the challenges effectively may be possible through the enhancement of the budget, feasible/clear guidelines from municipal councils/politicians, and reducing the gap between street network administrations and utility service providers.

KEYWORDS

pavement management systems, road maintenance, municipalities, budget allocation, questionnaire, pavement deterioration, cold climate

1 Introduction

Streets or urban roads are an essential asset in any community due to their multifunctioning nature, as they provide not only movement to traffic and people but also utility services. Streets in good condition are therefore vital to the enhancement of socioeconomic value and urban liveability (International Road Federation, 2007). Demographic and environmental changes are the main factors that constantly put unanticipated pressure on street pavements. Consequently, timely maintenance is imperative to keep the serviceability of the street network at the predefined performance level (PIARC, 2019).

It is a constant challenge to maintain street pavements in good condition within the limited maintenance budget, particularly for small municipalities. The shortfall in funding for maintenance is a common issue in Europe (European Union Road Federation, 2018), resulting in delayed maintenance and rehabilitation (M&R) activities. Delaying maintenance treatment can compromise the condition of the pavement network significantly (Chang et al., 2018), with a remarkable increase in the long-term pavement maintenance cost as a consequence (Tavakoli et al., 1992), while road users and environmental costs are additional burdens. Furthermore, limited funding restricts the wide range of treatment measures (Hafez et al., 2019). Relatively expensive construction, operations and maintenance of cold region pavements highlight the scarcity of budget and resources among pavement administrations (Doré and Zubeck, 2009).

For the pavement maintenance staff, ensuring sustainable maintenance solutions in pavement management provides further complications (NAS, 2011). Ramani et al. (2009) noted that only 30% of the US state transportation agencies address the sustainability concept (economic, social and environmental) in their pavement maintenance decision-making. However, this requires up-to-date pavement condition data and prediction models to evaluate the impact of maintenance strategies on the socio-economic and environmental factors in achieving sustainable pavement management (Flintsch and Bryce, 2014).

Keeping streets in good condition may be achieved through implementing an effective pavement management strategy and better cooperation among stakeholders. Pavement Management System (PMS) is a useful tool in the decision process for prioritizing maintenance strategies to reach optimal use of the maintenance budget by predicting the future pavement condition (Tavakoli et al., 1992; Shahin, 2005; Uddin et al., 2013). In light of this, PMS has been in place at the network and project level for several decades (Haas et al., 2015) to maintain the pavement network systematically and strategically intact (Loprencipe and Pantuso, 2017; Saha et al., 2017). It has been developed by many transport agencies (Pérez-Acebo et al., 2018) for various network levels, e.g., state roads (Flintsch and McGhee, 2009; Zimmerman, 2017; Peshkin and Duncan, 2021), local roads and municipal streets (Douglas, 2011; Wolters et al., 2011; He et al., 2017).

The use of pavement data in a PMS is common in selecting treatment alternatives among US state transportation agencies, although the use of environmental sustainability data is lacking in it (Zimmerman, 2017). In another study, it was found that traffic volume, maintenance history and structural data are important factors in building PMS among Colorado local agencies (Hafez et al., 2019). In Iowa State, US, collection of pavement distress data is lacking in 70% (of 74 respondents) of the local agencies, and therefore subjective maintenance practice prevails (Abdelaty et al., 2017).

The condition of the pavement network can be assessed through manual (visual survey/windshield) and sensor-based surveys i.e., automated and semi-automated (Sholevar et al., 2022). The cost of the manual method is low compared to the sensor methods but is a subjective approach (Staniek, 2021). Sensor methods require different data collection devices, namely,: cameras, different road sensor devices, friction, and deflectometer (Coenen and Golroo, 2017). Survey vehicles equipped with multi-devices can be used to measure a wide range of pavement distresses and other pavement performance parameters (Coenen and Golroo, 2017; Benmhahe and Chentoufi, 2021). Sensor-based data collection is comparatively useful in data-driven pavement management, provided that effective condition data quality management is in place throughout the survey (Underwood et al., 2011).

The pavement maintenance management practices at the local level (i.e., urban roads/streets) in the UK, maintenance funding, safety concern and classification of roads are the most decisive factors respectively, while average daily traffic was the least decisive factor in prioritizing the pavement maintenance decisions (Alfar, 2016). A study revealed that the local maintenance authority in England and Wales has a maintenance backlog of nearly £12.64 billion, which reflects the poor condition of local roads and lack of funding (Asphalt Industry Alliance, 2022). Further, it was reported that pavement ageing and pothole repair are the main challenges that the local authorities are facing since the average age of pavements is 70 years and total spending on pothole repair reached £1 billion in 2022 (Asphalt Industry Alliance, 2022).

The classification of road networks is important in managing traffic and budget utilization in the Toronto province (City of Toronto, 2017). A survey revealed that rutting and fatigue crackings are the most commonly collected distresses among Canadian cities and municipalities, followed by ravelling, transverse cracking and edge cracking; however, the frequency of collecting pavement distress data is subject to the availability of funding (Farashah and Tighe, 2014). According to a study, almost 40% of municipal roads are either at a fair or significant deterioration stage among the Canadian municipalities that responded (CIRC team, 2019), which could be the reason that pavement rehabilitation and reconstruction are more predominant than pavement preservation among the municipalities in Alberta (Newstead et al., 2018). Moreover, it was reported that there is significant variation in the implementation of infrastructure management plans between small and large municipalities (CIRC Team, 2019). Another study revealed that the engineering judgement approach prevailed among Canadian municipalities, particularly in small municipalities (Hajek et al., 2004).

Municipalities in Sweden are politically administrated organisations that maintain their street networks through municipal tax (Alm et al., 2021). However, the estimated backlog of municipal streets amounted to 12 billion Swedish crowns (SEK), while 8% of the municipal road network needs immediate rehabilitation or reconstruction (Ekdahl et al., 2016). Furthermore, only 50% of pavements had over 7 years of remaining service life (Ekdahl et al., 2016). The choice of pavement condition data assessment varies among municipalities due to street network size and the extent of the maintenance budget (Wilén, 2016).

To highlight the challenges of pavement management at the municipality level in Sweden, a questionnaire was formulated and disseminated to all municipalities across the country. The main purpose of the survey was to get first-hand information about the present practices of maintaining the streets at the municipal level. Furthermore, some of the respondents were selected for in-depth interviews to discuss pavement maintenance approaches in their respective municipalities. This study presents the results of the questionnaire regarding pavement distresses and their causes, along with the municipal transport street/road administration's approach to meeting these management challenges. The aim of the study is to provide better information on the status of street maintenance among municipalities in Sweden, which can then be used as a foundation for improving pavement management practices in municipalities.

2 Municipal pavements network management

2.1 Pavement structural design and deterioration

Flexible pavements are generally designed for 20 years and are based on a multi-layered elastic approach to withstand both traffic and environmental loads. In the design process it is usually assumed that pavement materials are horizontally infinite while the subgrade is vertically infinite and traffic is represented as equivalent single axel loads (ESALs) (Huang, 2004; Garber and Hoel, 2009; Zhang, Mills-Beale, and You, 2011; Mallick and El-Korchi, 2013). Typically, flexible pavements are supposed to diminish the induced stresses (compression, tensile and thermal) due to traffic loading and environmental changes to prevent the occurrence of pavement distress. The core of pavement design is to withstand traffic loading or environmentally induced excessive deformations (Sun, 2016), which can be achieved by selecting suitable materials for pavement structures in the design, construction and maintenance stages. To meet the pavement design requirement, the design approach aims to control fatigue cracking and rutting by addressing the tensile stress at the underside of the bound layer (asphalt) and the vertical, bottom-up compressive stress at the top of the subgrade (White et al., 2002; Sun, 2016). The thickness of the pavement varies in order to meet the expected functionality and serviceability performance during its designed life period.

Municipality street networks generally have a relatively low traffic volume compared to highways, which may vary from one street to another due to the functionality of the streets. The speed limit on asphalt-paved streets in municipalities usually varies from 30 km/h to 70 km/h, which is relatively low compared to highways. Similarly, streets do not usually include side ditches but cycle paths, parking places, street light poles, traffic sign poles, green strips, drainage inlets and utility inlets are commonly adjacent to the pavement structure. This implies the need for a different strategy for the construction and maintenance of street pavements.

Generally, the performance of flexible pavements is influenced by several factors: traffic volume, axle loads and their configurations, environmental and climate conditions, construction and maintenance strategies (both summer and winter), drainage, and functionality of the road/street or a combination thereof (Mallick and El-Korchi, 2013; Sun, 2016). The environment and serviceability level of the pavement govern the influence of the aforementioned factors, e.g., the type and severity of distress may be different compared to urban roads or highways (Lavin, 2003). Consequently, understanding the occurrence of pavement distresses in an urban context is complex due to uneven traffic conditions and the diverse functionality of pavements (Sadeghi et al., 2017).

Asphalt pavement distresses/defects can be categorized into four groups: cracks, deformations (including surface abrasion and unevenness), surface defects and edge defects (Fwa, 2006). Pavement cracks are common defects and are herein limited to longitudinal (incl. alligator/fatigue cracking), transversal and frost heave/thaw cracking. The mechanisms involved in the occurrence of these cracks are mainly due to fatigue-induced stresses and temperature, along with construction joints/laps. Longitudinal cracks occur in the direction of the pavement, which may be induced either due to traffic loading or non-traffic loading (Lavin, 2003). Longitudinal cracks along the wheel paths are load-driven (Lavin, 2003), which if left unattended will result in alligator cracking and eventually potholes (Pearson, 2012). However, non-load-induced longitudinal cracks occur outside the loading area or wheel paths due to several reasons, e.g., differential frost heaving (Doré and Zubeck, 2009; Churilin et al., 2018), construction joints/laps etc (Lavin, 2003). Alligator cracking in flexible pavement, i.e., bottom-up cracking due to repeated application of traffic loading, may prematurely occur in cold regions due to poor pavement drainage as a result of frost thaw (FT) cycles (Doré and Zubeck, 2009). Large fluctuations in the pavement temperature may even result in longitudinal and transverse cracks (Huang, 2004). Transverse cracks develop perpendicular to the travel direction, mainly due to low temperature, an abrupt change in the pavement temperature or aged binder (Lavin, 2003; Mallick and El-Korchi, 2013; Sun, 2016). The initiation of low-temperature cracks in pavements generally takes place in the asphalt-bound layers, but it may also begin in the underlying frozen layers or subgrade due to the development of ice lenses (Doré and Zubeck, 2009). Furthermore, tree roots in the vicinity of the street may also produce cracks in the pavements.

Pavement deformation commonly includes rutting, corrugation, shoving and frost heaving. Rutting distress is a permanent longitudinal depression of asphalt pavements in the wheel paths as a consequence of repeated traffic loading (Doré and Zubeck, 2009; Erlingsson, 2012; Alaswadko and Hassan, 2018). It may occur due to one or more reasons in all layers, both asphalt-bound layers and unbound layers, including subgrade (Huang, 2004; Pearson, 2012). Rutting in the unbound layers or subgrade may occur as a result of the thaw period that significantly reduces the bearing capacity due to the presence of the excess amount of melted water in the unbound layers of the pavement (Doré and Zubeck, 2009; Salour and Erlingsson, 2013). Pavement deformation in the form of surface wear, i.e., rutting, is common due to the usage of studded tyres in cold regions (Lundberg et al., 2019). However, wear on the surface course is more prominent on free-flow roads compared to residential streets (Arrojo, 2000; Snilsberg et al., 2016). Similarly, in the urban environment asphalt pavement wears down commonly at road intersections and bus stops, due to braking, slow movement and acceleration of vehicles (Transit, 2000; Hajj et al., 2007; Al-Qadi et al., 2009; Ali et al., 2009; Li et al., 2013). Moreover, both raised and unraised pedestrian crossings are observed as places that are vulnerable to surface wear in an urban environment. The surface resistance of asphalt pavement to abrasion can be improved at the design stage by opting for a higher aggregate size of coarse aggregate in the dense-graded or stone-rich mastic asphalt concrete, which on the other hand might compromise the noise level (Snilsberg et al., 2016).

Corrugation and shoving distress are plastic deformations of wearing course that appear in the shape of ripples (perpendicular to the pavement) and horizontal displacement (in the direction of pavement), due to the combined effect of acceleration or deceleration of vehicles and inadequate asphalt mix, poor bonding between layers or insufficient stability (Lavin, 2003; Fwa, 2006; Mallick and El-Korchi, 2013). Similarly, in cold regions, the uneven upward expansion or heaving of pavement structure due to the formation of ice lenses in the pavement results in pavement deformations (Doré and Zubeck, 2009; Pearson, 2012). Pavement surface unevenness occurs both in longitudinal and transverse directions due to corrugation, shoving and frost heaving (Wågberg, 2003).

Flexible pavements can incur surface defects, which are herein referred to as bleeding, ravelling, potholes and patches that can significantly affect the serviceability and skid resistance of the pavement (Fwa, 2006). Bleeding is a construction defect which refers to the appearance of asphalt binder on the pavement surface that results in a soft pavement surface due to the presence of an excessive amount of bitumen in the asphalt mixture (Garber and Hoel, 2009). Repeated traffic loading or over-compaction of asphalt, typically in hot weather conditions, results in the upward movement of asphalt-rich binder to the surface (Lavin, 2003; Fwa, 2006; Pearson, 2012). On the other hand, ravelling is common in cold and wet conditions where it refers to the progressive disintegration of both binder and aggregates from pavement wearing course due to poor compaction, inadequate asphalt binder or pavement ageing (Lavin, 2003; Fwa, 2006).

A pothole is a formation of localized depression or cavities in the pavement surface that may be developed in both wet and dry conditions (Wilson and Romine, 2001; Doré and Zubeck, 2009). Potholes may be formed due to poor base support, the presence of moisture in the pavement, FT cycles and repetitive traffic loading, or a combination of these factors (Jassal, 1998; Wilson and Romine, 2001). The seepage of moisture into the pavement is due to the presence of low-temperature or alligator cracking (Han et al., 2019) which is either left unaddressed or is addressed but not early enough (Marasteanu, 2018). The occurrence of potholes is more common in the case of fluctuating weather conditions or FT cycles—FT cycles was cited as the major cause of potholes in Canadian provinces, followed by traffic load, poor drainage and pavement age (Biswas et al., 2018). Further factors which contribute to pothole formations include type and thickness of asphalt, oxidation, rate and amount of precipitation and winter maintenance strategy, together with the method and material used for pothole repair (Biswas et al., 2018).

Patches result from repairing potholes, low-scale improvements of pavement condition and regular cuts for utility networks. Regular cuts and backfilling for utility networks on street pavements result in premature pavement deterioration (Yapp et al., 2001; Wilde et al., 2018). In cold regions, patches may be vulnerable to frost heaving and FT cycles due to the use of non-homogenous material relative to the surrounding pavement or poorly backfilled material around stormwater inlets or other water supply installations (Wågberg, 2003). The impact of potholes, regular utility cuts and patching not only deteriorate the pavement but also has a huge impact on traffic delays, user cost and comfort, safety, air quality, local businesses, and public perception of the effectiveness of municipal maintenance departments (Arudi et al., 2000; Wilde et al., 2018).

Flexible pavements are also subject to edge defects in the form of edge deformation as a result of insufficient pavement width, poor edge/shoulder support and pavement resurfacing (Fwa, 2006). A weak unbound layer due to frost action may also result in edge deformation (Ahmed, 2017).

Ageing is a well-recognized distress in asphalt pavements (Huang and Di Benedetto, 2015) which has been reported in England and Wales (Asphalt Industry Alliance, 2022), the US and Canada (Kim, 2014) due to a lack of funding and resources. Asphalt ageing has negative effects on pavement performance due to the degradation of bitumen in asphalt over time (Chai et al., 2014; Jing et al., 2021). It is influenced by several mechanisms; however, oxidation is the most significant mechanism that affects the durability of the bitumen in asphalt (O'Nions and Preston, 2015; Speight, 2016). In addition to this, the characteristics of asphalt, aggregates and particle size distribution, as well as the void ratio of the asphalt mix, also affect asphalt ageing (Speight, 2016). Asphalt undergoes both short-term and long-term oxidations that result in the hardening of asphalt over time (Speight, 2016; Sirin et al., 2018). Short-term or first-phase oxidation takes place during the production, storage, transportation and laying of asphalt, while long-term or second-phase oxidation takes place during the in-service life of asphalt (Speight, 2016; Sirin et al., 2018). However, several aspects of ageing are not sufficiently well understood (Doré and Zubeck, 2009; Underwood, 2015). Initial significant ageing happens in the first phase and during the first few months of the second phase. This later gradually slows down but keeps continuing until the end of its design life (Doré and Zubeck, 2009; O'Nions and Preston, 2015; Sirin et al., 2018). Oxidation of the asphalt has a direct relation to the exposing temperature, period of exposure and bitumen film thickness in asphalt (O'Nions and Preston, 2015). Aged asphalt pavements are more vulnerable to cracks, moisture-related damage, pothole formation and surface wear (Doré and Zubeck, 2009; Roque et al., 2015; Underwood, 2015).

2.2 Swedish municipality perspective

Sweden is located in Northern Europe with a population of approx. 10.4 million and has a total of 290 municipalities (SCB, 2020). There are approximately 623 thousand km of roads in the country. Municipalities own approx. 7% of these (mainly streets) while the Swedish Transport Administration owns approx 16% of roads (highways) (SKR, 2021). Due to safety in severe winter conditions, winter tyres (studded and specially designed winter tyres) are mandatory (December to March) but the use of studded tyres is relatively higher than that of other winter tyres in the country. Overall, the estimated use of studded tyres in Sweden is 55%; however, in the northern part of Sweden it is around 95% due to a prolonged winter season (Trafikverket, 2021). Similarly, the use of studded tyres is common in cold climate regions, e.g., the Nordic countries (excl. Denmark), Alaska and Canada. As a result, pavement abrasion wear is a common distress in these places (Abaza et al., 2021).

Generally, municipalities in Sweden classify their street networks on the basis of their provided functions, namely, main or arterials, collectors, residential and industrial streets. Municipalities have the prerogative to decide the load-carrying

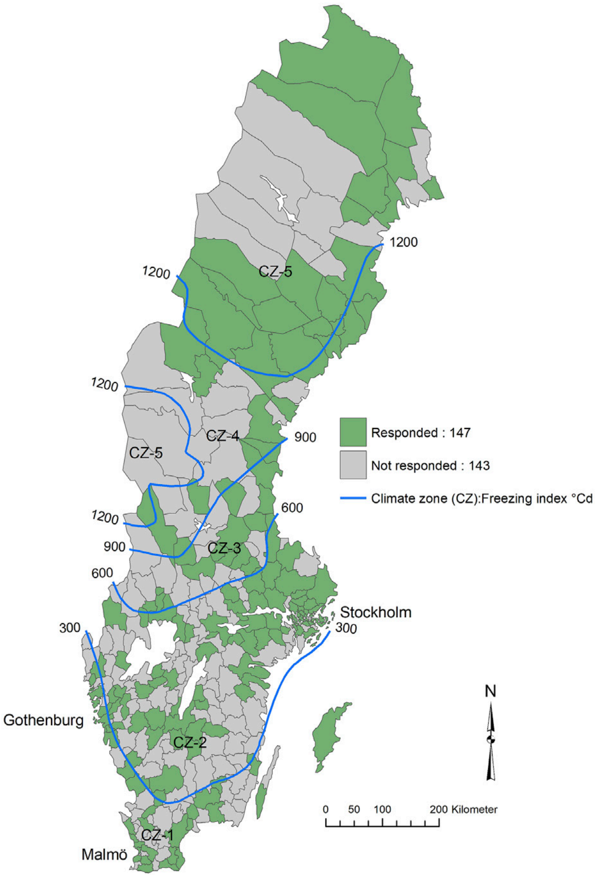


FIGURE 1
Map of Sweden showing the distribution of respondent municipalities and the five CZs based on the FI. The three biggest cities (Stockholm, Gothenburg, and Malmö) are also shown.

capacity group on their street network. Each municipality is responsible for maintaining the street network in their jurisdiction. Sweden is climatically divided into five climate zones (CZs) based on the average number of freezing degree days (freezing index, FI), i.e., from below 300 to above 1200 C ° d (Svensson, 1993; Trafikverket, 2011; Erlingsson and Saliko, 2020), namely, CZ 1 to CZ 5 (Figure 1).

The thickness of municipal pavement structure varies mainly with increased FI. Similarly, roads and streets need comparatively higher investments to build and maintain with increasing FI due to the severe climate and less competitive construction and maintenance industry. Typical pavement sections for residential streets in Swedish municipalities (Danderyd Kommun, 2011; Sandviken Kommun, 2021; Östersund Kommun, 2021; Helsingborg Kommun, 2022; Skellefteå Kommun, 2022) and state roads (Trafikverket, 2011) situated in different CZs are given in Table 1.

The thickness of the bound layers in Table 1 varies due to traffic conditions while the thickness of the subbase and frost protection layer increases with increasing FI. A geotextile on top of the subgrade is mandatory.

Pavement condition assessment is a significant step towards the effective utilization of the pavement budget. In Sweden, municipalities generally follow the locally developed guidelines of pavement condition assessment, which is useful for the visual assessment of pavement conditions (Wågberg, 2003). Similar types of guidelines have been developed in the US (Miller and Bellinger, 2014) and other European countries (Ragnoli et al., 2018). Similarly, there are locally developed guidelines to select the possible treatment alternatives for pavement distress (Wågberg, 2001). Generally, municipalities in Sweden outsource pavement condition assessment to private contractors or consultants. In fact, private vendors do not use the same methods of condition assessment, which makes it difficult for the municipalities to switch from one vendor to another (Wilén, 2016).

3 Method—Survey description

The survey presented here was composed of a questionnaire and in-depth interviews. The web-based questionnaire was sent to the customer service desk of all 290 Swedish municipalities. Thereafter, 14 in-depth interviews took place after the completion of the questionnaire. The criteria of the selection of municipalities for the interviews were solely based on the geographical location, the population size of the biggest urban area of the municipality, and the pavement network size (length per inhabitant) for streets. A total of 147 municipalities answered the survey, which yielded a 51% response rate. The geographical distribution of the municipalities is shown in Figure 1. In addition to this, 63 more municipalities had started the survey but neither finished nor submitted the survey in due time. Nevertheless, their completed parts of the survey are included in the results. Therefore, the response rate varies slightly from question to question relative to the 51% response rate. Population (SCB, 2020) and road network statistics of municipalities are presented in Table 2. As shown, the responding municipalities represent 70% of the total population in Sweden and 64% of the total municipal street network. The response rate is almost equal for all CZs, while the population density in the responding municipalities is somewhat higher than the national average.

Figure 2 highlights the research study and the results presented in this paper. Each step is described in the relevant sections.

3.1 Questionnaire content

The questionnaire was composed of 36 questions to get vital information related to the maintenance of streets owned by municipalities. The survey included: *i)* general information about municipality-owned pavement networks (8 questions), *ii)* flexible pavement condition data assessment and PMS (14 questions), *iii)* pavement distresses and their respective causes (4 questions), *iv)* budget and resources (10 questions). The question related to distresses in the questionnaire was formulated as "*How common are the following distresses on asphalt concrete street pavements*?", and was followed by a set of well-known pavement distresses (see results section). Similarly, in the case of distress causes, the question was formulated as "*How common are the following causes of distress on asphalt concrete street pavements*?". The choice of alternatives in both queries was based on a five-point Likert scale, i.e., "*None*",

TABLE 1 Typical pavement cross-sections of municipal streets and state roads in Sweden.

	Municipal street (residential pavement cross-sections)			Low-volume state road (<5 × $10^{\wedge 5}$ ESAL)
	CZ 1–5			CZ 1–5
	Thickness (mm)	Material	Aggregate size (mm)	Thickness (mm)
Wearing course	25–45	ABT[a]	11 or 16	45
Binder course or road base	35–60	AG[b]	16	-
Unbound base course	80–120	Crushed rock	0–32	80
Subbase	350–500	Crushed rock	0–80	420
			0–90	
			0–100	
Frost protection layer	50–650	Crushed rock	0–150	50–650
Geotextile	✓			✓
Subgrade (None to highly frost-susceptible soils)	✓			

[a]ABT, Dense Graded Asphalt Concrete pen 160/220.
[b]AG, Asphalt Gravel-Bitumen Bound pen 160/220.

"*Infrequently*", "*Quite frequently*", "*Very frequently*" and "*No info/No opinion*". Additionally, a place for any "*Comments*" was provided to strengthen the selected answers if needed.

4 Results

The results of pavement distress, distress causes, pavement management and budget are presented as follows. Figures containing the notation N = i represent the number of responses for each option.

4.1 Pavement distresses among responding municipalities

4.1.1 Share of distress distribution

Municipalities chose the most appropriate answers to different pavement distresses among the five alternatives (Figure 3) as per the formulated question "*How common are the following distresses on asphalt concrete street pavements*?". As shown in the aforementioned figure, potholes were the most commonly cited pavement distress (80%), followed closely by surface unevenness (75%), which refers to the occurrence of pavement unevenness in both longitudinal and transverse directions due to frost heaving, corrugation and shoving. Alligator cracking, rutting and longitudinal cracking were also considered as very frequent or quite frequent distress types by a large number of survey respondents. Fewer respondents observed transverse cracking, edge deformation or ravelling, while bleeding was the least experienced pavement distress.

4.1.2 Distresses with respect to climate zones

The above data (Figure 3) was analysed further to get information about the impact of cold climate on the pavement deterioration of street networks in different CZs. In this regard, the answer alternatives "Quite frequently" and "Very frequently" to the question "*How common are the following distress on asphalt concrete street pavements?*" of the respective distress in the respective CZ were summed up and plotted against the CZs (Figure 4). It can be seen that the pothole distress occurrence increases by 75%–100% from CZs 1–5, excluding CZ 3. Similarly, surface unevenness also shows a similar trend from south to north, which is probably linked to increased frost-heave action and a higher share of studded tyre usage in the northern part of the country. The trend in the occurrence of alligator cracking is unclear, while rutting occurrence is quite similar in all CZs except for a slight drop in CZs 4–5. On the whole, the occurrence of longitudinal and transverse cracks increases from south to north, reflecting the environmental impacts on the serviceability of the pavements. Edge deformation and ravelling occurrences are infrequent, while bleeding is almost negligible distress.

4.1.3 Distresses with respect to population and road network size

The distress data (Figure 3) was further analysed to identify the impact of the frequency distribution of most cited distresses (except edge deformation, ravelling and bleeding) in relation to the population density on the road network of each municipality. In this regard, the population of each municipality was divided by its street network (km). This ratio was used to classify the population density into 5 intervals, i.e., <120, 120–160, 160–200, 200–240 and above 240 persons per km street. The results are shown in Figure 5.

The frequency occurrence of potholes and surface unevenness are very similar and it seems that both distresses increase with increasing population density. The cited trend of rutting seems also to increase with increasing population density. The occurrence

10.3389/fbuil.2023.1205235

TABLE 2 Population, street network and survey statistics in the respondent municipalities.

Climate zone (CZ)	Total municipalities (no)	Total population (million)	Total street network[a] (thousand km)	Total population density on streets (persons/km)	Respondent municipalities (no)	Survey response rate (%)	Survey response rate relative to 51 per cent response (%)	Population in responded municipalities (million)	Street network in the respondent municipalities (thousand km)	Respondent municipality population density on roads (persons/km)
1	65	2.95	11.73	252	32	49	22	2.13	7.69	277
2	145	5.93	22.69	261	72	50	49	4.10	14.12	291
3	29	0.52	2.77	187	14	48	10	0.31	1.66	184
4	21	0.51	2.15	236	11	52	7	0.38	1.55	246
5	30	0.47	3.23	146	18	60	12	0.36	2.36	155
Total	290	10.38	42.57	244	147	51	100	7.28	27.37	266

[a]Partially collected from National Road Database, Trafikverket (NVDB, 2021).

frequency of alligator cracking is quite similar in all intervals while longitudinal cracking shows a trend, similar to a bell curve, that is difficult to interpret with regard to population density. On the other hand, transverse cracks show an almost uniform frequency occurrence except for the 20% decline in the most densely populated municipalities.

4.2 Share of cause of distress among the responding municipalities

In the case of common causes of pavement deterioration, the municipalities chose the most appropriate alternative among the five options to the question "*How common are the following causes of distress on asphalt concrete street pavements?*". The results are shown in Figure 6. Ageing (second-phase oxidation) was considered the most commonly cited cause of pavement distress (88%), followed closely by heavy vehicles (80%). Patching, high traffic flow and FT cycles were also indicated by a large number of survey respondents (more than 50%) as the main causes of distress. Fewer respondents observed roots and vegetation, frost heave and studded tyres among the common causes. High temperature was considered as the least frequent source of pavement distress.

4.2.1 Cause of distress with respect to climate zones

To highlight the possible impact of cold climate on the causes of pavement deterioration, the causes of distress data (Figure 6) were further analysed. In this regard, the answers ("Quite frequently" and "Very frequently") to the question "*How common are the following causes of distress on asphalt concrete street pavements*?" in the respective CZs were summed up and plotted against the CZs. The results are shown in Figure 7. Ageing was indicated as the most frequent source of pavement distress in all CZs except CZ 3. The trend suggests that ageing is almost the most significant source of deterioration in 127 municipalities located in CZs 1–2 and CZs 4–5. The low citation about ageing in CZ 3 is difficult to interpret. Heavy traffic as a source of deterioration seems to be more significant in CZs 1–2 relative to CZs 3–5. Frost heave and FT cycles are the most correlated frequent source of pavement distresses. Patching was cited in some way as an equally frequent cause of pavement deterioration. It could be associated with the abundance or presence of utility cuts, pothole fixing and localized pavement surface treatments. Studded tyres, roots and vegetation, and poor edge stability were cited as relatively low in the aforementioned sources of deterioration.

4.2.2 Cause of distress with respect to population and road network size

To highlight the possible correlation of the most cited causes of pavement deterioration to the population density on the road network, the causes of distress data (Figure 6) were further analysed. The results can be seen in Figure 8.

Ageing of pavement was cited as the most frequent source of pavement deterioration, followed by heavy vehicles and patching. Apart from a slight drop at the highest population density interval,

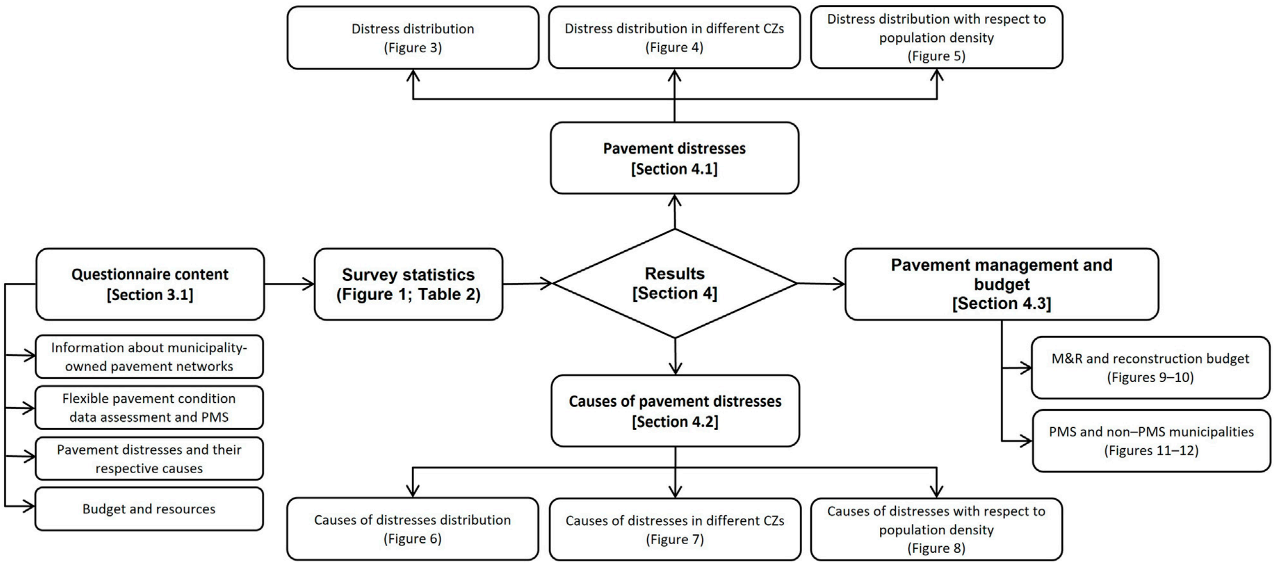


FIGURE 2
Flow chart of the study.

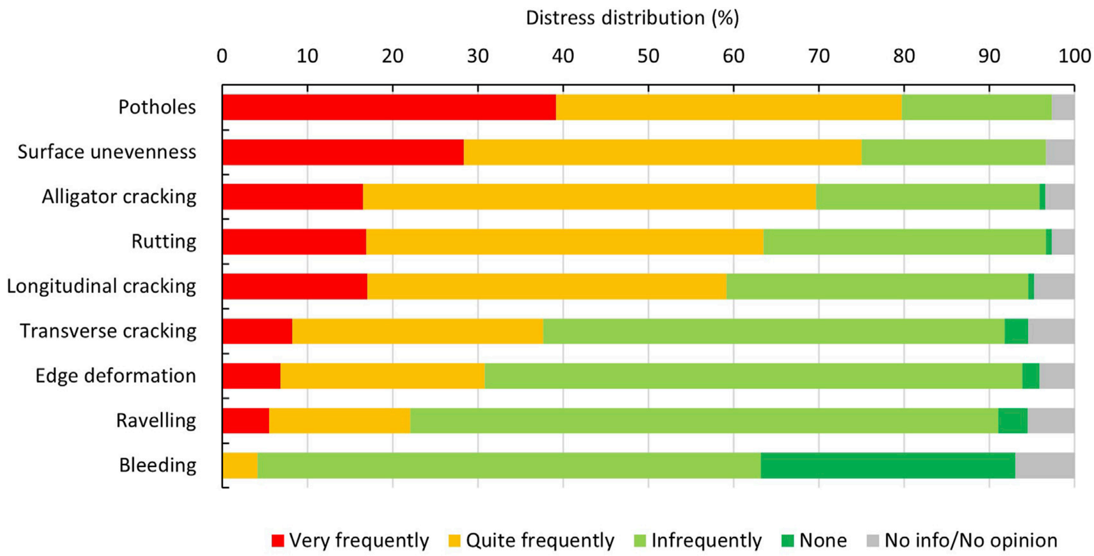


FIGURE 3
Distress distribution share among the responding municipalities.

there is an increasing trend in ageing and patching with population density intervals. Ageing, heavy vehicles, patching and high traffic flow are almost equally cited as the most frequent source of deterioration in most densely populated areas. Heavy vehicles and high traffic flow have slightly similar patterns with regard to an increase in population density intervals. The rest of the pavement deterioration causes (FT cycles, roots and vegetation, frost heave, studded tyres, poor edge stability) were cited with unclear patterns, which may be correlated with the climate rather than the population density intervals.

4.3 Pavement management and budget among responding municipalities

4.3.1 Maintenance and rehabilitation budget

The survey provides insight into the annual M&R budget allocation for the fiscal year 2020. A total of 118 municipalities reported that a total of 940 million SEK were allocated to the M&R budget to maintain about 23 thousand km of the street network (about 54% of the total street network in the country).

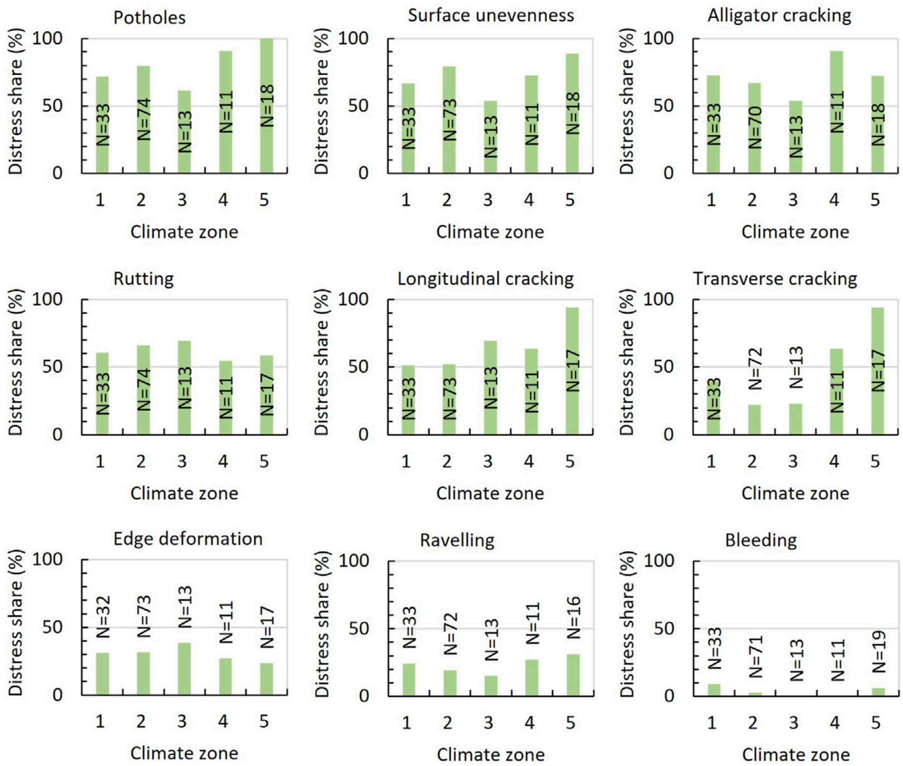


FIGURE 4
Pavement distress distribution in different CZs.

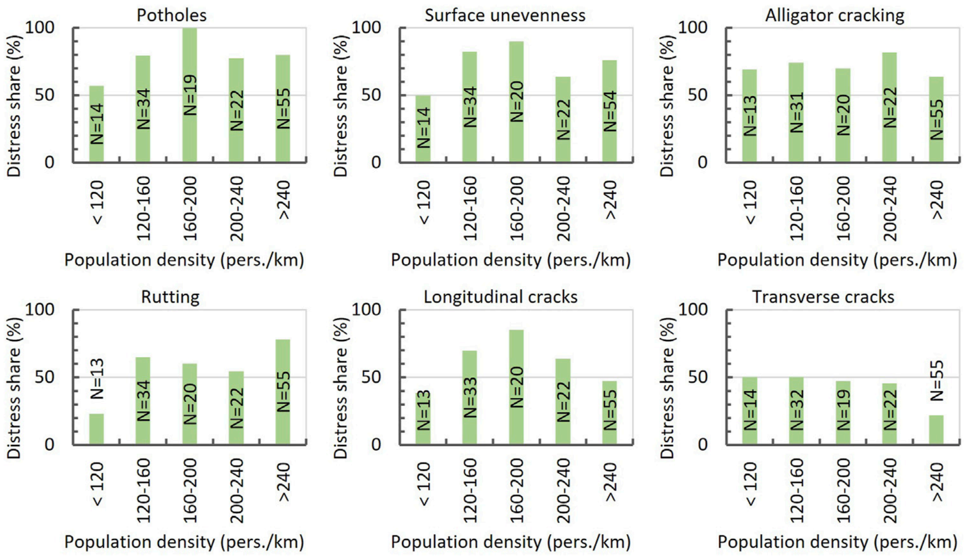


FIGURE 5
Occurrence of different pavement distresses for municipalities with respect to different population densities.

The cited M&R budget was analysed with regard to CZs and population density intervals to get an insight into the utilization of taxpayers' money on the management of the street network. In this regard, the budget of each municipality was divided by its population and the street network (km). The outcome of the taxpayers' contribution to street network management with

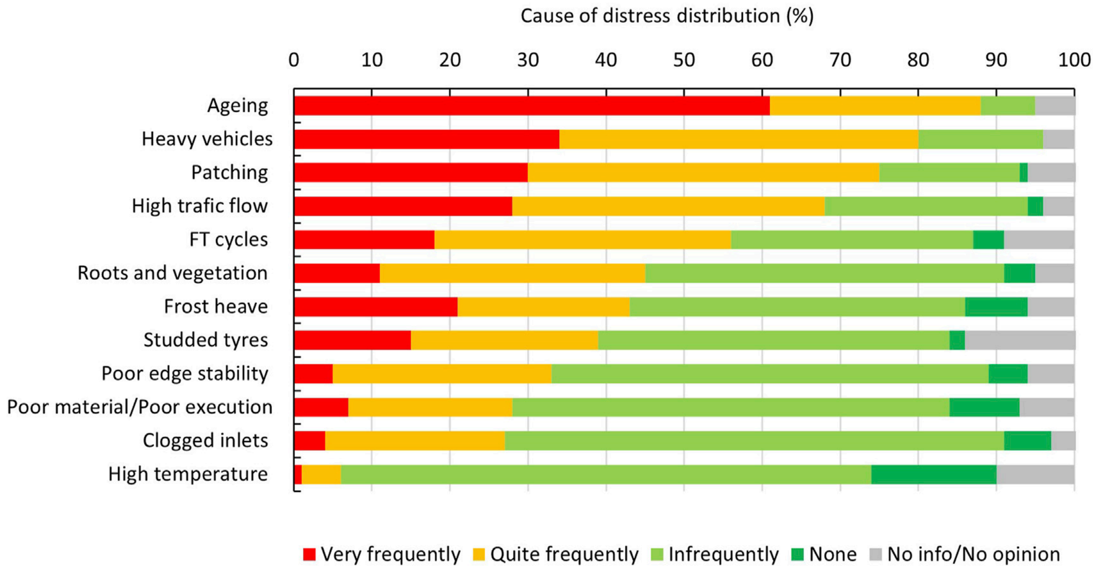


FIGURE 6
Cause of distress distribution share among the responding municipalities.

respect to CZ and population density is shown in Figure 9. It can be seen that the allocated budget for M&R activities tends to increase with an increase in FI while decreasing with an increase in population density.

The spending of taxpayers' money is highest in CZs 3–4 but lowest in CZ 2 (12 times less). On the other hand, the cited spending on M&R activities is highest in the most sparsely populated municipalities compared with densely populated municipalities. The spending trend with regard to population density seems to be strongly correlated when compared to CZ.

4.3.2 Reconstruction budget

For the fiscal year 2020, 78 municipalities cited that a total of SEK 749 million was allocated to the reconstruction budget to maintain about 16 thousand km of the street network (about 54% of the respondent municipalities). To highlight the utilization of taxpayers' money in the management of the street network, the cited spending on reconstruction activities is analysed with regard to CZs and population density. The outcome is presented in Figure 10. It is noticeable that the cited allocated expenditure on reconstruction is lower in CZs 1–2 and comparatively higher in CZs 3–5. The higher budget for reconstruction in CZ 4 (i.e., 21 times more than the CZ 2 spending) is difficult to interpret. On the contrary, the allocated budget for reconstruction decreases with an increase in population density. The most sparsely populated municipalities spent almost 14 times more than the most densely populated municipalities. Apparently, the taxpayers' spending is more associated with population density than with CZs.

4.3.3 Maintenance backlog

In response to the budget-related question, 70% of the responding municipalities cited dissatisfaction with catching up with the maintenance backlog due to the low maintenance budget. However, 25% of municipalities cited that they are satisfied in relation to their capacity and resources, while 5% were undecided. Similarly, to reduce the maintenance backlog, 34% of the municipalities indicated that they have received a budget increase over the last 5 fiscal years but it is still not enough. However,22% of municipalities cited that their maintenance budget shrank over time. One municipality cited that the maintenance backlog is huge and it is difficult to meet the performance goals even if there is enough budget, due to a lack of staff and a short summer season.

4.3.4 Pavement management

The majority of municipalities (94 out of 147) have no proper PMS in place due to a lack of resources and maintenance staff. Some municipalities even cited that the need for PMS is not required, due to the implementation cost of commercial PMS and their small pavement network. Most municipalities with no proper PMS have paper-based maintenance approaches or use spreadsheets to manage their network due to budget constraints. The budget (M&R and reconstruction) data in Figures 9, 10 have been broken down to get an insight into the role of PMS in the maintenance of the street network. The outcome is presented in Figure 11.

It can be seen in Figure 11 that the allocated M&R and reconstruction budget tends to decrease with an increase in population density, both in PMS and non-PMS municipalities. Moreover, the allocated budget for maintaining the street network is almost the same in PMS and non-PMS municipalities. It is therefore difficult to interpret or draw a conclusion about whether the cost of maintaining a street network decreases with a PMS in place or *vice versa*. Figure 12 shows the PMS and non-PMS municipalities among the responding municipalities.

It can be seen in Figure 12 that more than one-third of the respondent municipalities have a commercial PMS in place to manage 52% of the street network of respondent municipalities.

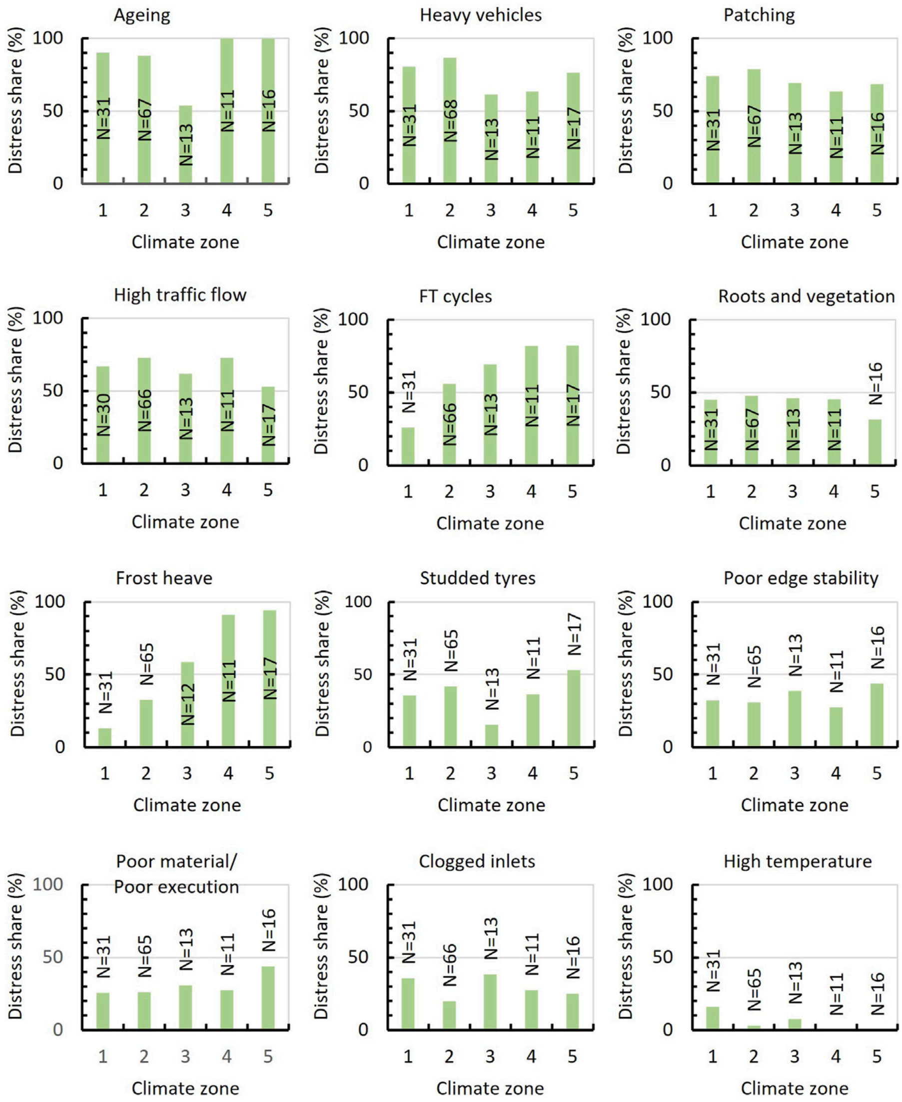


FIGURE 7
Cause of pavement distress distribution in different CZs.

The manual/windshield assessment method of condition assessment is predominant.

An annual assessment of the pavement network is most common among responding municipalities (64 out of 134) but storing such assessments is not practised. The vast majority of the PMS municipalities cited that they collected surface distress data every 4 or 5 years for the whole network. Even among PMS municipalities, storing maintenance history has not yet matured; this is a common problem among municipalities or small pavement administrations elsewhere as well.

In condition assessments, the vast majority of responding municipalities have collected the segment, type, severity and position of distress of the pavement area, as well as the functionality of the street. Nearly one-third of the responding municipalities (43 out of 147) collected data about traffic volume. However, the international roughness index (IRI) and data collection on traffic lanes, drainage inlets and side drains have never been part of the PMS among the responding municipalities. All municipalities cited that the selection of pavement segment for treatment depends on multiple factors, e.g., the type and severity of distress, the functionality of the street, traffic volume, preservation/preventive maintenance, available budget, complaints from road users, and suggestions from the municipal council.

In response to the distress analysis question, 129 out of 142 municipalities cited that they analysed the distresses through engineering judgement, even at the project level. The remaining follow locally developed guidelines and PMS recommendations. Onsite investigation or material testing happens rarely. Common pavement treatment measures that were identified include thin overlay asphalt and resurfacing after milling the deteriorated asphalt surface. Further, it was identified that pavement preservation is infrequent and is limited to cracking sealing.

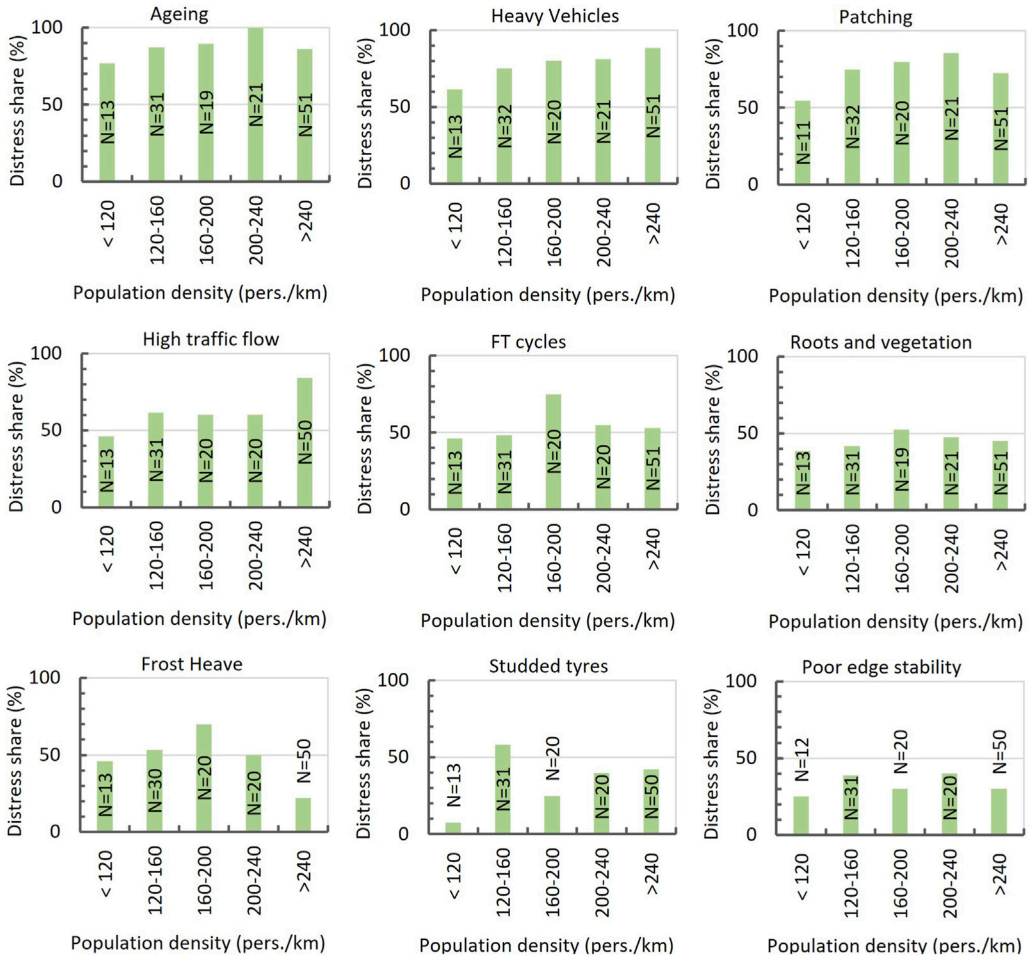


FIGURE 8
Causes of different pavement distress for municipalities with respect to different population densities.

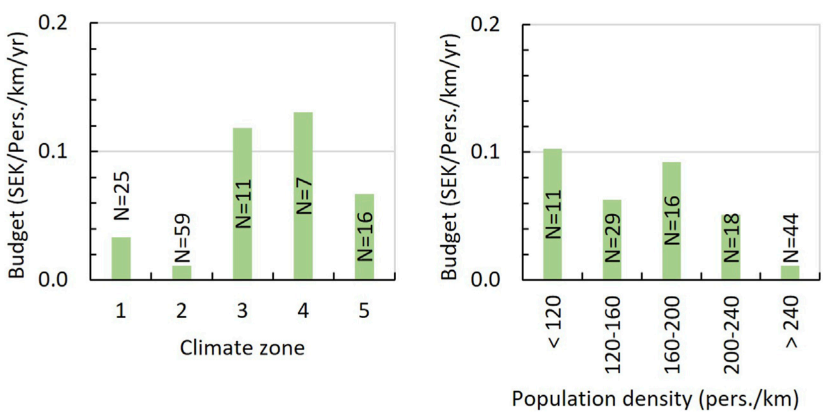


FIGURE 9
Comparison of taxpayers' contributions to streets M&R (CZ and population density).

The municipalities were asked to rank factors that could improve the current maintenance practices and street network conditions. A total of 114 municipalities indicated that the focus needs to be on the following top four areas: *i*) maintenance budget, 46% of municipalities cited that enhancement of the maintenance budget would be decisive in the matter for catching up with the

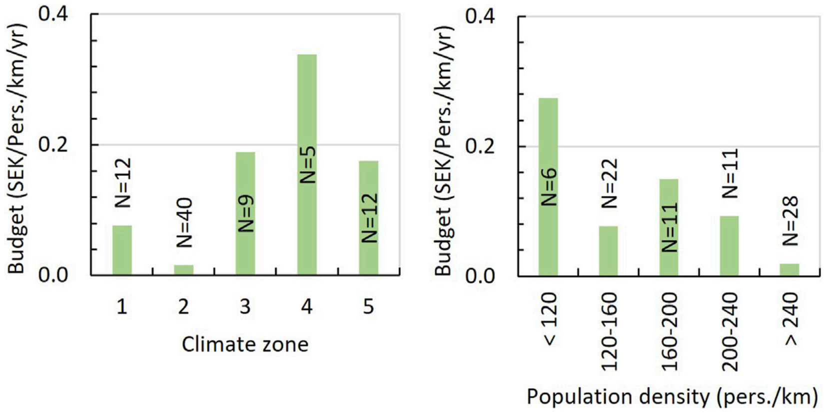


FIGURE 10
Comparison of taxpayers' contributions to street reconstruction (CZ and population density).

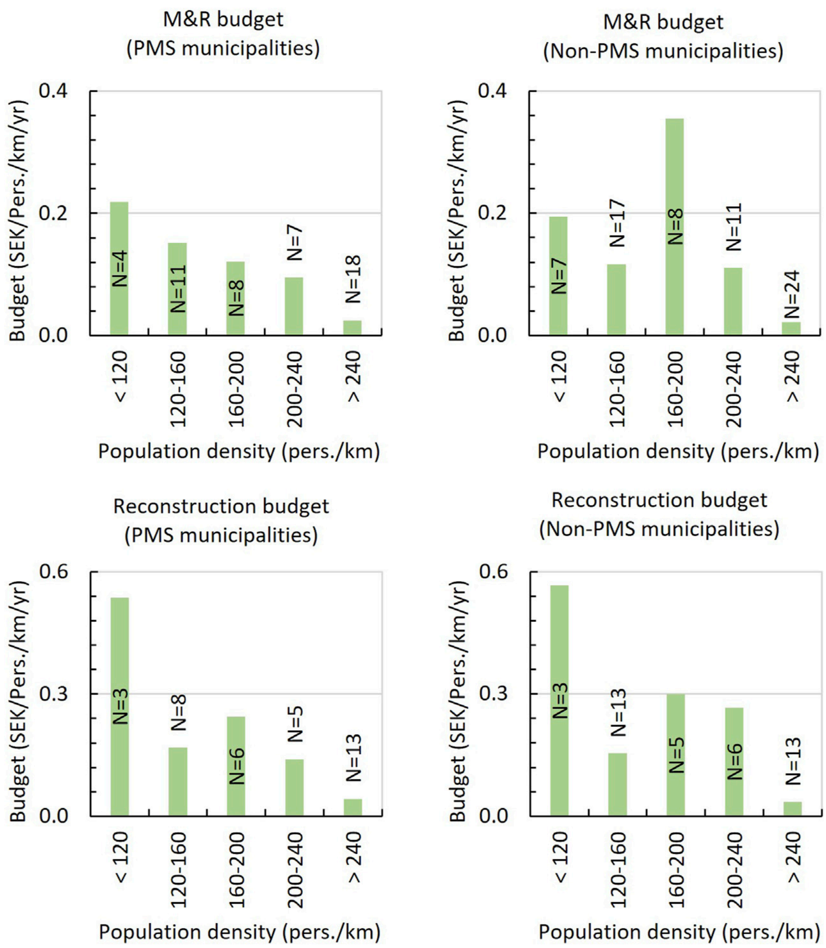


FIGURE 11
Comparison of budget in PMS and non-PMS municipalities with respect to different population densities.

maintenance backlog; *ii*) cooperation, 14% of municipalities cited that better cooperation between municipal street network administrations and utility service providers is needed; *iii*) policy guidelines, 8% of municipalities identified that better maintenance policy/guidelines from municipal councils/politicians to the municipal street network administration would lead to greater

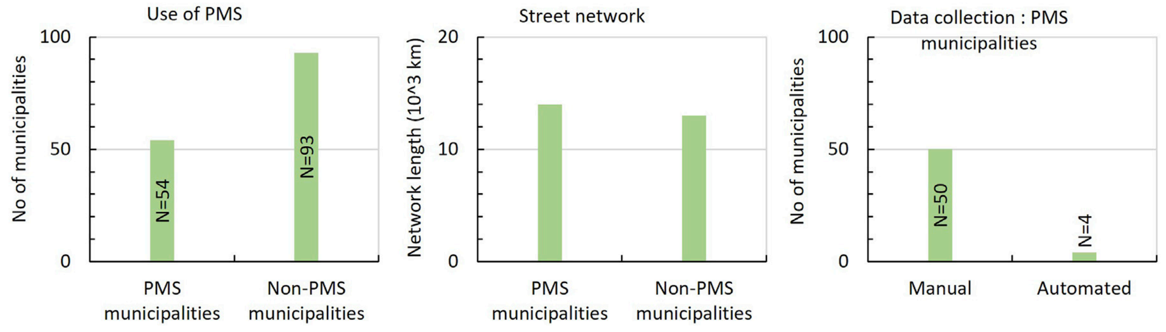


FIGURE 12
Pavement management approaches among PMS and non-PMS municipalities.

improvement; *iv*) maintenance staff, 7% of municipalities indicated that the street network condition would improve with an increase in the number of maintenance staff.

5 Discussion

5.1 Frequent pavement distress and causes

The questionnaire revealed that aged pavement as a source of pavement deterioration is the most serious issue throughout the responding municipalities (Figures 7, 8), particularly in CZs 4–5 and densely populated municipalities. This could be because of the low average frequency of resurfacing of municipal roads, since the vast majority of the municipal networks in the country were built in the 1970s. It seems that lack of resources (i.e., proper PMS usage and low frequency of pavement condition data collection) and maintenance budget play a vital role in the abundance of aged networks across the municipalities. The issue of aged pavement needs to be addressed on a priority basis since it contributes to the occurrence of several distresses.

Pavement unevenness is the second most frequent pavement distress after potholes which could be due to the combined effect of cold climate and population density on roads (Figures 4, 5). On the other hand, the occurrence of longitudinal and transverse cracks are strongly correlated to frost action or cold climate (Figure 4).

Patching is further one of the most common sources of deterioration throughout the country (Figures 7, 8). The high frequency of patching and utility cuts could be due to insufficient coordination and cooperation among street network administrations and utility service providers. The choice of material and method to repair potholes and backfill utility cuts is also decisive in keeping the functionality of the roads. In this regard, street network administrations are supposed to improve the quality assessment of backfilled materials and minimize the use of cold mix in winter due to its low durability. Additionally, to keep the street network in a serviceable condition, the street network administrators and utility service providers need to improve the share of information with each other. An integrated system or database would be useful in the timely execution of M&R activities, both for management of street networks and for utility services.

The reported occurrence of alligator cracks is surprisingly quite similar in terms of population density (Figure 5), as heavy vehicles and high traffic are cited as significant sources of distress (Figure 8). This is probably not due to insufficient asphalt-bound layer thickness (Table 1), but rather due to the abundance of aged pavements and pothole formation.

The frequency trend of rutting (Figures 4, 5) is reasonable since it reflects the significance of both cold climates (excess use of studded tyres and/or FT cycles) and the dense population density on roads. Similarly, the lowest frequency occurrence of rutting in the colder zones (Figure 4) can be related to the use of a winter road maintenance strategy, i.e., allowing snow to accumulate on the road surfaces up to a certain level. In northern Sweden, the residential streets in winter are frequently covered with snow/ice, which contributes to reducing the negative effects of studded tyres. Another reason for relatively less occurrence of rutting on municipal streets could be due to the large number of residential streets compared to main, arterial and collector streets. Residential streets are low-speed and low-trafficked streets that are mostly covered with snow, resulting in the low-frequency occurrence of abrasion wear.

5.2 Infrequent pavement distress and causes

The low-cited frequency of bleeding (Figure 4) could be related to the short summer season and the low- cited frequency of high temperatures, while the low-cited frequency of ravelling (Figure 4) might be an oversight due to the presence of issues like aged pavement, pothole formation and patches. The situation might even be the same for the roots and vegetation as a source of deterioration (Figures 7, 8). However, the low frequency of edge deformation (Figure 4) on streets could be due to the presence of pedestrian paths and houses along streets. In other words, edge deformation might be more frequent in the presence of non-functioning side drains or lack of sufficient distance between the side drain and the pavement edge. On the other hand, the reason for the low frequency of poor practices and clogged inlets (Figure 7) is difficult to interpret, since the quality standards are not the same throughout the country. Furthermore, data collection about the functionality of pavement drainage is rare.

5.3 The capacity of municipalities—strategies, resources and budget

Municipal councils, through relevant committees, decide the level of pavement performance goals and the street management budget in municipalities. Municipal pavement management staff make strategies within the available resources and budget to meet pavement performance goals. The representatives of responding municipalities realized a need for much better guidelines from the political administrations regarding the maintenance of street networks. Improvements in policies/guidelines may be achieved by presenting the current and predicted condition of the pavement network. Such assessments require the implantation of PMS.

The sophistication of the commercial PMS among the responding municipalities is low due to a lack of deterioration models, collection of roughness data, and maintenance decision rules/trees. Furthermore, a lack of maintenance history and outdated pavement data management are widespread among the responding municipalities. Consequently, it is difficult to track the quality of maintenance alternatives and maintenance practices. Annual pavement condition assessment is common but storing such data is not practised due to constraints on funding, time and resources. Subjective condition assessment and a subjective treatment selection approach are additional issues which need to be shifted to automated data collection and data-driven decision-making over time. Municipalities are generally satisfied with the current maintenance approach in relation to their competence and available budget. However, they agreed that there is room for improvement.

The street management budget reflects the approach for the utilization of taxpayers' money. The spending trend of taxpayers' money on both M&R and reconstruction activities, as per the fiscal year 2020, is similar in terms of both CZ and population density on road length (Figures 9, 10). However, the spending is 2–3 times more on reconstruction activities compared to M&R. Ideally, the spending should be more on M&R activities, though it needs a preventive maintenance approach. It can be said that the taxpayers are contributing more to the M&R and reconstruction of the municipal road network located in the northern region and sparsely populated municipalities. Timely M&R and reconstruction interventions are important in the effective utilization of the available budget, since fixing the distress in the early stage is relatively less costly.

6 Conclusion

The paper presents the results of a survey on street network deterioration and pavement management practices among Swedish municipalities. Fifty-one percent of survey responses were received out of a total of 290 municipalities across the country. The study highlights both the significant/frequent and insignificant/infrequent pavement distresses on the street network among the responding municipalities, along with the possible causes. Moreover, the study highlights the maintenance practices and the availability of budget to the municipalities to maintain the functionality and serviceability of street networks. The main conclusion of the study can be summarised as follows:

- The most cited frequently occurring distresses are potholes, surface unevenness, alligator cracking and rutting respectively.
- On the other hand, ageing, heavy vehicles, patching and high traffic flow are reported as the most frequent sources of pavement distresses.
- The use of commercial PMS is limited while visual assessment of pavements is predominant, due to budget and resource constraints.
- The M&R and reconstruction challenges varied from south to north and from sparsely to densely populated municipalities.

The share distribution of potholes, longitudinal and transverse cracking, and surface unevenness distresses tends to increase from south to north as per CZs. However, no such pattern is visible with increasing population density over road length.

Pavement age or ageing seems to be the main cause of pavement distress irrespective of CZ and population density, which reflects the high share distribution of pavement patches. This might indicate the delayed maintenance approach, possibly due to insufficient cooperation between street network and utility service administrators, a lack of resources and sophisticated PMS. Heavy vehicles and high traffic flow are almost equally dominant in terms of both population density and CZs, but the frequent occurrence of alligator cracks might be due to aged pavements. Frost heave and the number of FT cycles increases from south to north as a source of pavement deterioration. The downsides of studded tyres on residential streets are reduced due to low-speed limits and snow accumulation on street pavements.

A quarter of municipalities are satisfied with the current budget allocation for M&R and reconstruction to catch up with the maintenance backlog. Due to the short summer season and insufficient maintenance staffing, many municipalities would not be able to effectively reduce their maintenance backlog even if they have the needed budget. For the year 2020, the allocated budget both for M&R and rehabilitation decreases with increasing population density over road length, while it increases CZ-wise from southern to northern municipalities.

The challenges that the municipalities are facing are beyond pavement deterioration and cold climate. Additional challenges include getting the required pavement maintenance budget in both the short- and long-term and adapting to unclear or non-feasible maintenance policies/guidelines from municipal councils/politicians. Further challenges include developing better working procedures to improve coordination and cooperation with utility service providers, and lastly increasing the number of maintenance staff and enhancing their competence. In other words, a more national view with insights into different climate zones and capabilities of municipalities could be a good initiative in improving the safety and rideability standards on street networks.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization: SE, LS, and MA; Methodology: SE, LS, and MA; Investigation: MA, SE, and LS; Writing–original draft preparation: MA; Writing–review and editing: SE and LS; Visualization: MA, SE, and LS; Supervision: SE and LS; Project administration: SE and LS; Funding acquisition: SE and LS. All authors contributed to the article and approved the submitted version.

Funding

The project was funded by Mistra Infra Maint Project 1.8 (DIA 2016/28) and Skellefteå municipality. The APC was funded by the library of the KTH Royal Institute of Technology.

Acknowledgments

The authors gratefully acknowledge those who have taken part in the formulation of questionnaires at VTI and those who responded to the questionnaire and interviews.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abaza, O. A., Arafat, M., and Uddin, M. S. (2021). Physical and economic impacts of studded tyre use on pavement structures in cold climates. *Transp. Saf. Environ.* 3, 1–15. doi:10.1093/tse/tdab022

Abdelaty, A., Jeong, H. D., and Smadi, O. (2017). Practical approach for pavement treatment decisions for local agencies. *J. Constr. Eng. Proj. Manag.* 7, 30–36. doi:10.6106/JCEPM.2017.3.30.030

Ahmed, A. (2017). Effects of climatic loading in flexible pavement subgrades in Texas. Available at: https://rc.library.uta.edu/uta-ir/bitstream/handle/10106/27171/AHMED-DISSERTATION-2017.pdf?sequence=1&isAllowed=y.

Al-Qadi, I. L., Yoo, P. J., Elseifi, M. A., and Nelson, S. (2009). Creep behavior of hot-mix asphalt due to heavy vehicular tire loading. *J. Eng. Mech.* 135, 1265–1273. doi:10.1061/(ASCE)0733-9399(2009)135:11(1265)

Alaswadko, N., and Hassan, R. (2018). Rutting progression models for light duty pavements. *Int. J. Pavement Eng.* 19, 37–47. doi:10.1080/10298436.2016.1155123

Alfar, E. (2016). GIS based pavement maintenance management model for local roads in the UK. Available at: https://usir.salford.ac.uk/id/eprint/39679/1/F__Final%20Version%20of%20Emad%20Alfar%20PhD%20Thesis_Emad%20Alfar%20PhD%20Thesis%201.pdf.

Ali, B., Sadek, M., and Shahrour, I. (2009). Finite-element model for urban pavement rutting: Analysis of pavement rehabilitation methods. *J. Transp. Eng.* 135, 235–239. doi:10.1061/(ASCE)0733-947X(2009)135:4(235)

Alm, J., Paulsson, A., and Jonsson, R. (2021). Capacity in municipalities: Infrastructures, maintenance debts and ways of overcoming a run-to-failure mentality. *Local Econ.* 36, 81–97. doi:10.1177/02690942211030475

Arrojo, G. (2000). *Pavement wear caused by the use of studded tyres*. Linköping: Swedish National Road and Transport Research Institute.

Arudi, R., Pickering, B., and Flading, J. (2000). Planning and implementation of a management system for utility cuts. *Transp. Res. Rec.* 1699, 42–48. doi:10.3141/1699-06

Asphalt Industry Alliance (2022). Annual local authority road maintenance (ALARM) survey 2022. Wrington, England. Available at: https://www.asphaltuk.org/wp-content/uploads/ALARM-survey-2022-FINAL.pdf (Accessed November 14, 2022).

Benmhahe, B., and Chentoufi, J. A. (2021). Automated pavement distress detection, classification and measurement: A review. *IJACSA* 12, 1–11. doi:10.14569/IJACSA.2021.0120882

Biswas, S., Hashemian, L., and Bayat, A. (2018). Investigation of pothole severity and maintenance methods in Canada through questionnaire survey. *J. Cold Regions Eng.* 32, 1–9. doi:10.1061/(ASCE)CR.1943-5495.0000161

Chai, G., van Staden, R., Chowdhury, S. H., and Loo, Y. C. (2014). A study of the effects of pavement ageing on binder deterioration. *Int. J. Pavement Eng.* 15, 1–8. doi:10.1080/10298436.2012.752823

Chang, C., Rodriguez, E., Yapp, M., and Pierce, L. (2018). *Consequences of delayed maintenance of pavement networks*. Pretoria, South Africa: Jukwaa Media. Available at: https://repository.up.ac.za/bitstream/handle/2263/69516/1A_Chang_Consequences_2018.pdf?sequence=1&isAllowed=y.

Churilin, V., Efimenko, S., Matvienko, O., and Bazuev, V. (2018). Simulation of stresses in asphalt-concrete pavement with frost heaving. *MATEC Web Conf.* 216, 01011–01019. doi:10.1051/matecconf/201821601011

CIRC team (2019). Canada infrastructure report card 2019: Monitoring the state of Canada's core public infrastructure. Canadian Infrastructure. Available at: http://canadainfrastructure.ca/downloads/canadian-infrastructure-report-card-2019.pdf (Accessed January 2, 2023).

City of Toronto (2017). About the road classification system. City of Toronto. Available at: https://www.toronto.ca/services-payments/streets-parking-transportation/traffic-management/road-classification-system/about-the-road-classification-system/ (Accessed February 8, 2023).

Coenen, T. B. J., and Golroo, A. (2017). A review on automated pavement distress detection methods. *Cogent Eng.* 4, 1374822–1374823. doi:10.1080/23311916.2017.1374822

Danderyd Kommun (2011). Teknisk handbok (in Swedish). Available at: https://www.danderyd.se/contentassets/e24ebabb04bc4330a6ad0c3826f7aa16/teknisk-handbok.pdf (Accessed January 24, 2023).

Doré, G., and Zubeck, H. K. (2009). *Cold regions pavement engineering*. Reston, VA: New York: ASCE Press ; McGraw-Hill.

Douglas, R. A. (2011). Appropriate pavement maintenance and rehabilitation management system for local governments. *Transp. Res. Rec.* 2205, 206–212. doi:10.3141/2205-26

Ekdahl, P., Hellman, M., Henrikson, T., Svedben, P., and Silfwerbrand, J. (2016). *Skulden till underhåll. Det kommunala underhållsbehovet för gator, broar och belysning (in Swedish)*. Stockholm: Sveriges Kommuner och Landsting. Available at: https://webbutik.skr.se/bilder/artiklar/pdf/7585-446-5.pdf (Accessed September 29, 2021).

Erlingsson, S. (2012). Rutting development in a flexible pavement structure. *Road Mater. Pavement Des.* 13, 218–234. doi:10.1080/14680629.2012.682383

Erlingsson, S., and Saliko, D. (2020). "Correlating air freezing index and frost penetration depth—a case study for Sweden," in *Proceedings of the 9th international Conference on Maintenance and Rehabilitation of pavements—mairepav9 lecture notes in civil engineering*. Editor C. Raab (Cham: Springer International Publishing), 847–857. doi:10.1007/978-3-030-48679-2_79

European Union Road Federation (2018). *Road Asset Management: An ERF position paper for maintaining and improving a sustainable and efficient road network*. Brussels, Belgium: European Road Federation. Available at: https://erf.be/wp-content/uploads/2018/07/Road-Asset-Management-for-web-site.pdf.

Farashah, M. K., and Tighe, S. L. (2014). "Development practices for municipal pavement management systems application," in *Asset management: Reinventing organizations for the next 100 Years session* (Montreal, Quebec), 1–20.

Flintsch, G., and Bryce, J. (2014). "Sustainable pavement management," in *Climate change, energy, Sustainability and pavements green energy and Technology*. Editors K. Gopalakrishnan, W. J. Steyn, and J. Harvey (Berlin, Heidelberg: Springer), 373–392. doi:10.1007/978-3-662-44719-2_13

Flintsch, G. W., and McGhee, K. K. (2009). *Quality management of pavement condition data collection. Transportation research board*. Washington, D.C.: National Cooperative Highway Research Program Synthesis. doi:10.17226/14325

T. F. Fwa (Editor) (2006). *The handbook of highway engineering* (Boca Raton: Taylor & Francis).

Garber, N. J., and Hoel, L. A. (2009). *Traffic and highway engineering*. 4th ed. Australia; United States: Cengage Learning.

Haas, R., Hudson, W. R., and Falls, L. C. (2015). *Pavement asset management*. 1st edition. Salem, Massachusetts: Hoboken, New Jersey: Wiley-Scrivener.

Hafez, M., Ksaibati, K., and Atadero, R. (2019). Best practices to support and improve pavement management systems for low-volume paved roads. *Int. J. Pavement Eng.* 20, 592–599. doi:10.1080/10298436.2017.1316648

Hajek, J., Hein, D., and Olidis, C. (2004). "Decision making for maintenance and rehabilitation of municipal pavements," in *Annual Conference of the Transportation Association of Canada* (Québec, Canada), 1–9.

Hajj, E. Y., Siddharthan, R. V., Sebaaly, P. E., and Weitzel, D. (2007). Hot-mix asphalt mixtures for Nevada's intersections. *Transp. Res. Rec.* 2001, 73–83. doi:10.3141/2001-09

Han, S., Liu, M., Shang, W., Qi, X., Zhang, Z., and Dong, S. (2019). Timely and durable polymer modified patching materials for pothole repairs in low temperature and wet conditions. *Applied Sciences* 9, 1–13. doi:10.3390/app9091949

He, Z., Qin, X., Wang, H., and Comes, C. (2017). Implementing practical pavement management systems for small communities: A south Dakota case study. *Public Works Manag. Policy* 22, 378–391. doi:10.1177/1087724X17721714

Helsingborg Kommun (2022). Överbyggnadstabell—Teknisk handbok (in Swedish). Available at: https://tekniskhandbok.helsingborg.se/wp-content/uploads/sites/64/2020/03/overbyggnadstabell-tekniskhandbok-200330.pdf (Accessed January 24, 2023).

S.-C. Huang and H. Di Benedetto (Editors) (2015). *Advances in asphalt materials: Road and pavement construction*. 1st edition (Boston, MA: Woodhead Publishing).

Huang, Y. H. (2004). *Pavement analysis and design*. 2nd ed. Upper Saddle River, NJ: Pearson/Prentice Hall.

International Road Federation (2007). *The socio-economic benefits of roads in Europe*. Brussels: European Union Road Federation. Available at: https://www.aecarretera.com/cdc/ERF_Los%20beneficios%20socioeconomicos%20de%20las%20carreteras%20en%20Europa.pdf (Accessed January 2, 2023).

Jassal, K. S. (1998). Development of potholes from cracks in flexible pavements. Available at: https://www.collectionscanada.gc.ca/obj/s4/f2/dsk2/ftp01/MQ39100.pdf (Accessed March 9, 2022).

Jing, R., Varveri, A., Liu, X., Scarpas, A., and Erkens, S. (2021). Differences in the ageing behavior of asphalt pavements with porous and stone mastic asphalt mixtures. *Transp. Res. Rec.* 2675, 1138–1149. doi:10.1177/03611981211032218

Y. R. Kim (Editor) (2014). *Asphalt pavements* (London, UK: CRC Press/Balkema is an imprint of the Taylor & Francis Group, an Informa business).

Lavin, P. (2003). *Asphalt pavements: A practical guide to design, production and maintenance for engineers and architects*. London: Spon.

Li, L., Huang, X., Wang, L., and Li, C. (2013). Integrated experimental and numerical study on permanent deformation of asphalt pavement at intersections. *J. Mater. Civ. Eng.* 25, 907–912. doi:10.1061/(ASCE)MT.1943-5533.0000745

Loprencipe, G., and Pantuso, A. (2017). A specified procedure for distress identification and assessment for urban road surfaces based on PCI. *Coatings* 7, 65–26. doi:10.3390/coatings7050065

Lundberg, J., Janhäll, S., Gustafsson, M., and Erlingsson, S. (2019). Calibration of the Swedish studded tyre abrasion wear prediction model with implication for the NORTRIP road dust emission model. *Int. J. Pavement Eng.* 22, 432–446. doi:10.1080/10298436.2019.1614585

R. B. Mallick and T. El-Korchi (Editors) (2013). *Pavement engineering: Principles and practice*. 2nd edn. (Boca Raton: CRC Press).

Marasteanu, M. (2018). *Pothole prevention and innovative repair. Pillsbury drive*. Minneapolis, MN: Department of Civil, Environmental, and Geo- Engineering University of Minnesota. Available at: http://mndot.gov/research/reports/2018/201814.pdf.

Miller, J. S., and Bellinger, W. Y. (2014). *Distress identification manual for the long-term pavement performance project*. McLean,VA: Federal Highway Administration, US. Available at: https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/13092/13092.pdf (Accessed January 2, 2023).

NAS (2011). *Sustainable pavement maintenance practices*. Washington, DC: The National Academies Press.

Newstead, B., Hashemian, L., and Bayat, A. (2018). *A study on pavement network condition and reporting in the province of Alberta through a questionnaire survey*. Ottawa, Ontario: Transportation Association of Canada.

NVDB (2021). *Vägdata: Kommunala asfalterade bilvägar. Nationella Vägdatabasen (in Swedish).*

O'Nions, L., and Preston, N. (2015). "Chapter 20: Durability of bitumens and asphalts," in *The shell Bitumen handbook (westminster* (London: Published for Shell Bitumen by ICE Publishing), 573–589.

Östersund Kommun (2021). Tekniska handbok, sektioner körbana (in Swedish). Available at: https://tekniskhandbok.ostersund.se/vagutformning/korbana/ (Accessed January 24, 2023).

Pearson, D. (2012). *Deterioration and maintenance of pavements*. First edn. London: ICE Publishing.

Pérez-Acebo, H., Linares-Unamunzaga, A., Abejón, R., and Rojí, E. (2018). Research trends in pavement management during the first years of the 21st century: a bibliometric analysis during the 2000–2013 period. *Appl. Sci.* 8, 1–23. doi:10.3390/app8071041

Peshkin, D. G., and Duncan, G. M. (2021). *Maintenance and surface preparation activities prior to pavement preservation treatments*. Washington, DC: The National Academies Press.

PIARC (2019). *Innovative-innovative-approaches-asset-management. PIARC technical committee D.1 - management of road infrastructure assets. La defense cedex*. France: World Road asspciation. Available at: https://www.piarc.org/ressources/publications/11/8427e86-31264-2019R19EN-Innovative-Innovative-Approaches-Asset-Management.pdf (Accessed May 8, 2022).

Ragnoli, A., De Blasiis, M. R., and Di Benedetto, A. (2018). Pavement distress detection methods: A review. *Infrastructures* 3, 58–19. doi:10.3390/infrastructures3040058

Ramani, T., Zietsman, J., Eisele, W., Rosa, D., and Spillane, D. (2009). *Developing sustainable transportation performance measures for TxDOT's strategic plan: Technical report*. Austin, Texas: United States: Federal Highway Administration. Available at: https://static.tti.tamu.edu/tti.tamu.edu/documents/0-5541-1.pdf.

Roque, R., Zou, J., and Hernando, D. (2015). "5 - cracking mechanisms in asphalt mixtures," in *Advances in asphalt materials woodhead publishing series in civil and structural engineering*. Editors S.-C. Huang and H. Di Benedetto (Oxford: Woodhead Publishing), 135–167. doi:10.1016/B978-0-08-100269-8.00005-2

Sadeghi, J., Najafabadi, E. R., and Kaboli, M. E. (2017). Development of degradation model for urban asphalt pavement. *Int. J. Pavement Eng.* 18, 659–667. doi:10.1080/10298436.2015.1095912

Saha, P., Ksaibati, K., and Atadero, R. (2017). Developing pavement distress deterioration models for pavement management system using markovian probabilistic process. *Adv. Civ. Eng.* 2017, 1–9. doi:10.1155/2017/8292056

Salour, F., and Erlingsson, S. (2013). Investigation of a pavement structural behaviour during spring thaw using falling weight deflectometer. *Road Mater. Pavement Des.* 14, 141–158. doi:10.1080/14680629.2012.754600

Sandviken Kommun (2021). Grävbestämmelser sandvikens kommun (in Swedish). Available at: https://sandviken.se/download/18.69cc24a417ffd93f9a4bfe39/1650963101798/Sandviken%20kommuns%20gr%C3%A4vbest%C3%A4mmelser.pdf (Accessed December 13, 2022).

SCB (2020). *Folkmängden efter region, civilstånd, ålder och kön. År 1968 - 2021 (in Swedish)*. Statistikdatabasen. Available at: http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101A/BefolkningNy/ (Accessed January 26, 2023).

Shahin, M. Y. (2005). *Pavement management for airports, roads and parking lots*. 2nd ed. New York, NY: Springer.

Sholevar, N., Golroo, A., and Esfahani, S. R. (2022). Machine learning techniques for pavement condition evaluation. *Automation Constr.* 136, 1–17. doi:10.1016/j.autcon.2022.104190

Sirin, O., Paul, D. K., and Kassem, E. (2018). State of the art study on aging of asphalt mixtures and use of antioxidant additives. *Adv. Civ. Eng.* 2018, 1–18. doi:10.1155/2018/3428961

Skellefteå Kommun (2022). Ritningar och andra handlingar - skellefteå kommun (in Swedish). Available at: https://skelleftea.se/invanare/startsida/bo-trafik-och-miljo/bygga-nytt-andra-eller-riva/bygglov-och-bygganmalan/ritningar-och-andra-handlingar (Accessed January 24, 2023).

SKR (2021). Drift och underhåll: Fakta om svenska vägnätet (in Swedish). *Sveriges Kommuner och Reg.* Available at: https://skr.se/skr/samhallsplaneringinfrastruktur/trafikinfrastruktur/driftochunderhall.291.html (Accessed July 14, 2021).

Snilsberg, B., Saba, R. G., and Uthus, N. (2016). "Asphalt pavement wear by studded tires – effects of aggregate grading and amount of coarse aggregate," in *Proceedings of 6th eurasphalt & eurobitume congress* (Prague: Czech Technical University). doi:10.14311/EE.2016.072

Speight, J. G. (2016). *Asphalt materials science and Technology*. 1st ed. Oxford: Butterworth-Heinemann:Elsevier.

Staniek, M. (2021). Road pavement condition diagnostics using smartphone-based data crowdsourcing in smart cities. *J. Traffic Transp. Eng. Engl. Ed.* 8, 554–567. doi:10.1016/j.jtte.2020.09.004

Sun, L. (2016). in *Structural Behavior of asphalt pavements*. Editor L. Sun (Oxford: Butterworth-Heinemann:Elsevier).

Svensson, J. (1993). *VTT Tjäldataarkiv (in Swedish)*. Statens Väg-och Trafikinstitut., VTI meddelande 727. Available at: https://www.diva-portal.org/smash/get/diva2:672187/FULLTEXT01.pdf (Accessed February 5, 2021).

Tavakoli, A., Lapin, M. S., and Figueroa, J. L. (1992). Pmsc: Pavement management system for small communities. *J. Transp. Eng.* 118, 270–280. doi:10.1061/(ASCE)0733-947X(1992)118:2(270)

Trafikverket (2011). Trafikverkets tekniska krav Vägkonstruktion (in Swedish). Available at: https://trafikverket.ineko.se/Files/sv-SE/10750/RelatedFiles/2011_072_TRVK_vag_2.pdf (Accessed July 13, 2020).

Trafikverket (2021). Undersökning av däcktyp i Sverige – vintern 2021 (januari–Mars) (in Swedish). Available at: https://www.dackinfo.se/wp-content/uploads/2012/04/Dackundersokning_PV_vinter_TRV_publ_2021-215.pdf (Accessed May 12, 2022).

Transit (2000). Asset management manual. Available at: https://www.nzta.govt.nz/assets/resources/state-highway-asset-management-manual/docs/SM020-manual.pdf (Accessed December 5, 2022).

Uddin, W., Hudson, W. R., and Haas, R. (2013). *Public infrastructure asset management*. New York, US: McGraw-Hill Education.

Underwood, B. S. (2015). "9 - multiscale modeling approach for asphalt concrete and its implications on oxidative aging," in *Advances in asphalt materials woodhead publishing series in civil and structural engineering*. Editors S.-C. Huang and H. Di Benedetto (Oxford: Woodhead Publishing), 273–302. doi:10.1016/B978-0-08-100269-8.00009-X

Underwood, B. S., Kim, Y. R., and Corley-Lay, J. (2011). Assessment of use of automated distress survey methods for network-level pavement management. *J. Perform. Constr. Facil.* 25, 250–258. doi:10.1061/(ASCE)CF.1943-5509.0000158

Wågberg, L.-G. (2003). *Bära eller brista: Handbok i tillståndsbedömning av belagda gator och vägar (in Swedish)*. Stockholm: Svenska kommunförbundet.

Wågberg, L.-G. (2001). *I valet & kvalet: Handbok för val av beläggningsåtgärd (in Swedish)*. Stockholm: Svenska kommunförbundet.

White, T. D., Haddock, J. E., Hand, A. J., and Fang, H. (2002). *Contributions of pavement structural layers to rutting of hot mix asphalt pavements*. Washington, D.C: National Academy Press.

Wilde, J., Grant, C., and Nelson, P. (2018). *Manual for controlling and reducing the frequency of pavement utility cuts*. Washington, D.C: Federal Highway Administration, US. Available at: https://www.fhwa.dot.gov/utilities/utilitycuts/manual.pdf (Accessed January 18, 2023).

Wilén, Ni. (2016). Framtagning av standard för skadesammanställning på kommunala gator (in Swedish). Available at: http://lup.lub.lu.se/student-papers/record/8876577.

Wilson, T. P., and Romine, A. R. (2001). *Materials and procedures for repair of potholes in asphalt-surfaced pavements-manual of practice*. Washington, DC: Federal Highway Administration. Available at: https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/99168/99168.pdf (Accessed November 9, 2022).

Wolters, A., Zimmerman, K., Schattler, D. K., and Rietgraf, A. (2011). *Implementing pavement management systems for local agencies: Implementation guide*. Illinois: Illinois Center for Transportation. Available at: https://idot.illinois.gov/Assets/uploads/files/Transportation-System/Manuals-Guides-&-Handbooks/T2/P052.pdf (Accessed September 18, 2020).

Yapp, M., Lundy, J., and Chu, S. (2001). Impact of utility cuts on pavement performance in the city of seattle. Available at: https://trid.trb.org/view/793554 (Accessed September 23, 2021).

Zhang, Q., Mills-Beale, J., and You, Z. (2011) "Pavement engineering I: Flexible pavements," in *Handbook of transportation engineering, volume II: Applications and technologies*. Editor M. Kutz Second Edition (New York, Chicago, San Francisco, Lisbon, London, Madrid, Mexico City, Milan, New Delhi, San Juan, Seoul, Singapore, Sydney, Toronto: McGraw-Hill Education). Available at: http://www.accessengineeringlibrary.com/content/book/9780071614771/chapter/chapter11 (Accessed June 4, 2021).

Zimmerman, K. A. (2017). *Pavement management systems: Putting data to work*. Washington, DC: National Academy of Sciences.

frontiers | Frontiers in Built Environment

TYPE Review
PUBLISHED 10 October 2023
DOI 10.3389/fbuil.2023.1271889

Check for updates

OPEN ACCESS

EDITED BY
Vagelis Plevris,
Qatar University, Qatar

REVIEWED BY
Nimish Biloria,
University of Technology Sydney, Australia
Shenjun Yao,
East China Normal University, China

*CORRESPONDENCE
Yang Ye,
yeyang@hit.edu.cn

RECEIVED 03 August 2023
ACCEPTED 26 September 2023
PUBLISHED 10 October 2023

CITATION
Ye Y, Huang Y and Li S (2023), A multidisciplinary perspective on the relationship between sustainable built environment and user perception: a bibliometric analysis.
Front. Built Environ. 9:1271889.
doi: 10.3389/fbuil.2023.1271889

A multidisciplinary perspective on the relationship between sustainable built environment and user perception: a bibliometric analysis

Yang Ye*, Yi Huang and Shuqi Li

School of Architecture, Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Harbin Institute of Technology, Ministry of Industry and Information Technology, Harbin, China

The study of the relationship between sustainable built environment and user perception has often taken a single perspective, displaying neither a holistic view of the relationship nor a systematic and refined grasp of the research content previously. This has resulted in a biased understanding of the two research objects and their respective measurement methods and made it difficult to develop synergies. In this context, this paper summarises the current research hotspots and trends in the relationship between sustainable built environment and user perception through CiteSpace quantitative analyses such as keyword co-occurrence networks, emergent word detection, and disciplinary re-clustering, and takes a multidisciplinary perspective to focus on relevant research in public health, environmental science, and architecture and urban design, such as the progress of research between sustainable built environment and users in physical activity, environmental cognition, and image perception. The results show that in the environmental science field, environmental cognition is the core, based on the study of users' direct perception of sustainable built environment and its externalisation in the expression of behaviours. The public health field and the architecture and urban design field are dominated by the study of perceptual outcomes. There is a trend towards big data as a measurement tool for research subjects in all fields with a multidisciplinary perspective and the inclusion of more disciplines can produce more meaningful research results. The study provides a framework for research into the relationship between the two from a broader perspective and provides guidance for future multidisciplinary research, with implications for the construction of high-quality human-centred urban spatial environments.

KEYWORDS

multidisciplinary vision, sustainable built environment, user perception, Citespace, health behaviour, urban landscape

1 Introduction

The built environment is often defined as the human environment, such as communities, workplaces, and road environments (JF, 2006). Studies have shown that the built environment has a strong correlation with the health of the population (Gold, 1980; Barton, 2009), but the mechanisms by which it affects health are complex. This is due to the combined effects of specific social and economic conditions and individual usage patterns, along with environmental influences, including physical activity (Hur et al., 2010) and environmental pollution (Diez Roux and Mair, 2010). The rapid development of modern industrial cities has led to a series of public health problems caused by the reduced quality of the built environment. Such rapid development has gradually worsened the problems of resource scarcity, traffic congestion, and air pollution, and given rise to many disorders, such as homogeneity, chaos, and inhuman scale, (Chen et al., 2020), which, in turn, have led to the compression of social space as well as health problems, such as obesity and chronic diseases (Lu and Tan, 2015). Based on numerous empirical studies in Western countries, urban built environment factors such as housing conditions, building density, green space (Evans et al., 2003; Nielsen and Hansen, 2007; Van den Berg et al., 2010), neighbourhood support, and the degree of chaos in the neighbourhood (Kim, 2010; Qin et al., 2018) have a significant impact on residents' mental health. The design of the built environment has a strong guiding effect on people's daily lives and behaviour patterns (Liu and Guo, 2006). It is, therefore, necessary to examine whether the built environment meets the behavioural and psychological needs of residents, and how more refined, humanistic planning and design could improve the quality of urban space to enhance their sense of wellbeing and access. This has become one of the main directions for the improvement of built environment (Cao and Ettema, 2014). This is also a main issue that needs to be addressed in the process of urban physical examination and development of new cities.

Environmental perception refers to a person's awareness or feeling of the environment and the act of understanding the environment through the senses (Wang L. et al., 2020). Environmental perception of the built environment by people is the most direct result of the relationship between the two, such as the visual quality of streets and satisfaction with urban parks and green spaces (Altman and Zube, 2012). Individual perception of the built environment is influenced by a variety of factors, such as individual economic attributes, social environment, and spatial and psychological characteristics (Chen et al., 2017). Measuring residents' perceptions of the built environment can help analyse the relationship between their health and the various components of the built environment (Dong et al., 2020). Early studies on the impact of the built environment on resident health focused on the impact on physical activity and its health effects: the former was mostly in the fields of environmental science and public health (Kearns and Moon, 2002); the latter was in the fields of recreation and healthcare; and there was a lack of systematic and diverse studies on the built environment impact factors and pathways on resident health (Andrews and Evans, 2008). Existing studies on the perception of the built environment are very rich in connotations, covering various aspects, such as sensory perceptions of sight, hearing, smell, and touch (Xi et al., 2019); perceptions of community facilities and spatial forms (Zhu and Zhai, 2019); evaluation dimensions, such as the sense of safety and comfort in the environment (Chen et al., 2017); as well as high-level environmental experiences such as satisfaction (Carr et al., 2010), happiness (Dong and Qin, 2017), community identity, and emotional attachment (Wang, 2018), etc. Nowadays, the impact of the sustainable built environment on resident health is increasingly receiving focused attention from the fields of urban and rural planning, sociology, medicine, and other disciplines.

Some scholars have attempted to define 'multidisciplinarity' and 'interdisciplinarity'. They argue that multidisciplinarity draws on knowledge from different disciplines but stays within their respective boundaries. Generally speaking, collating the views of researchers from different disciplines on the same issue from their respective disciplines is called 'multidisciplinary research' and is often found in project applications. Interdisciplinary research is more concerned with methodological and theoretical aspects (Alvargonzález, 2011). Multidisciplinary and interdisciplinary research are often combined (Wu, 2021). The National Academy of Sciences defines 'interdisciplinarity' as a mode of research in a discipline in which a research team or individual integrates information, data, working perspectives, concepts, or theories from two or more disciplines or fields of expertise to solve or understand a problem that could not be solved by a single discipline or area of research practice alone and to break free from the constraints of a single disciplinary perspective (Committee on Facilitating Interdisciplinary Research and National Academy of Sciences, National Academy of Engineering, 1900). 'Sustainability' is multiple things at once and navigates interesting territory—it is a goal, an ideal, an umbrella, and a sub-discipline of multiple disciplines (Reyers et al., 2010). Clarifying the relationship between sustainable built environment and user perception is an important part of alleviating urban problems, improving the quality of the living environment, and enhancing urban refinement. This paper takes a multidisciplinary approach to the analysis of the design of human-centred spaces in the context of public health, environmental science, architectural engineering, transport, and urban planning. The multidisciplinary nature of conservation science has long been recognised, but seldom achieved. Multidisciplinarity features several academic disciplines in a theme-based investigation with multiple goals—essentially, studies "co-exist in a context" (Petts et al., 2008). If subdivided, multidisciplinary research focuses on collecting and comparing research results without attempting to cross boundaries (Tress et al., 2005). It is seen as a step up from multidisciplinary research, focusing on synthesising knowledge and generating new and different perspectives, rather than simply reproducing them (Attwater et al., 2005; Petts et al., 2008). It is a process of answering, solving, or posing a question that is more involved and complex than a single discipline can handle, drawing on various perspectives and integrating the results of various disciplines with the aim of developing a more integrated understanding and expanded perception.

Thus, the paper analyses the relationship between sustainable built environment and user perception from a multidisciplinary perspective, using a high-level view to make a clear link within the huge network of relationships, adopts an interdisciplinary approach to describe the research and its different aspects in more detail, synthesises and analyses the results of the research in order to generate new knowledge helpful for further research, outlines the current state of research using CiteSpace, understands the distribution characteristics of multidisciplinary fields, and summarises them into three major research areas. It is hoped that this will provide a reference for subsequent academic research and built environment-related design.

2 Research methodology

2.1 Study design

This paper begins with the design of a research framework (Figure 1). Through a preliminary reading of theory and literature,

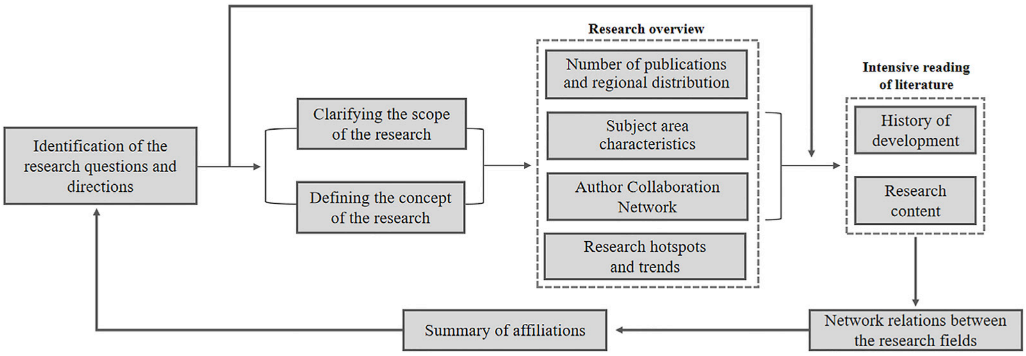


FIGURE 1
Research design framework diagram.

the research questions and directions were identified: the study of the relationship between sustainable built environment and user perception. The research scope was then clarified and the research concepts defined. This paper uses the literature information visualisation software CiteSpace (Chen et al., 2015) as the main analysis tool to 1) understand the total number of publications on the relationship between sustainable built environment and user perception and the regional distribution characteristics of the research objects in the literature; 2) explore the disciplinary distribution characteristics based on Web of Science (WoS) categories and related research literature; 3) use the keyword co-occurrence network and frequency analysis to study the relationship between sustainable built environment and user perception; 4) use the co-authorship network and frequency of keyword citation articles to analyse the core figures and research results of the relationship between sustainable built environment and user perception and combine with the clustering table to summarise the three main research areas; 5) based on the results of the inductive analysis, explore the future research directions of the relationship between sustainable built environment and user perception from three aspects: public health, environmental science, and architecture and urban design, in order to provide reference and guidance for subsequent multidisciplinary cross-research and cooperation, as well as subsequent optimisation of sustainable built environment.

2.2 Data acquisition

With the advent of the era of big data, the visualisation of relationships between categories of knowledge is increasingly valued by scholars. In order to ensure the comprehensiveness and accuracy of data, this paper uses the WoS core database as the data collection platform and CiteSpace software to draw and analyse the corresponding map. User perception also includes psychological needs and perceptions of sustainable built environment, so in addition to the general term 'user perception', specific psychological perception phrases are added, including 'safety', 'comfort', 'enjoyment', 'wellbeing', 'perceptibility', and 'playfulness'. Weakly related keywords such as 'satisfaction' and 'walkability' are excluded to ensure the comprehensiveness of the search results, and therefore, the search has shortcomings. After running several searches and filters to ensure that the search terms were as core as possible, 'sense of pleasure' and 'palpable' were removed; 'safety' was used instead of 'security'; and the advanced search formula TS=(sustainable built environment and (perception of user or perception of safety or comfortable sensation or happiness or playability)) was used. 2,465 documents were obtained. 1,870 journal articles were selected as the data source for this study by manually removing the papers that were double counted, with a data download date of 03 December 2022.

3 Review of literature on the relationship between sustainable built environment and user perception

The research on the relationship between sustainable built environment and user perception has a long history and was first started in Europe and the United States, which includes a great variety of contents, disciplines and research methods, but there are connections and overlaps in the internal structure. The analysis of the number of publications per year and regional distribution characteristics can show the hotspot time and region for the research on the relationship between sustainable built environment and user perception. The analysis of the Author collaboration network can show the key researchers and important papers for the research on the relationship between the two, in order to clearly define the research line and development direction. The analysis of the characteristics of the subject areas can clearly show the variety and complexity of the research content, which can clarify the main subject areas and essentially capture the core relationship between user perception of the sustainable built environment, and analyse the future development direction with the research hotspots' map.

3.1 Number of publications per year and regional distribution characteristics

The change in the number of publications reflects the level of attention and research on the topic. This paper presents a statistical analysis of the year-on-year changes in the number of publications

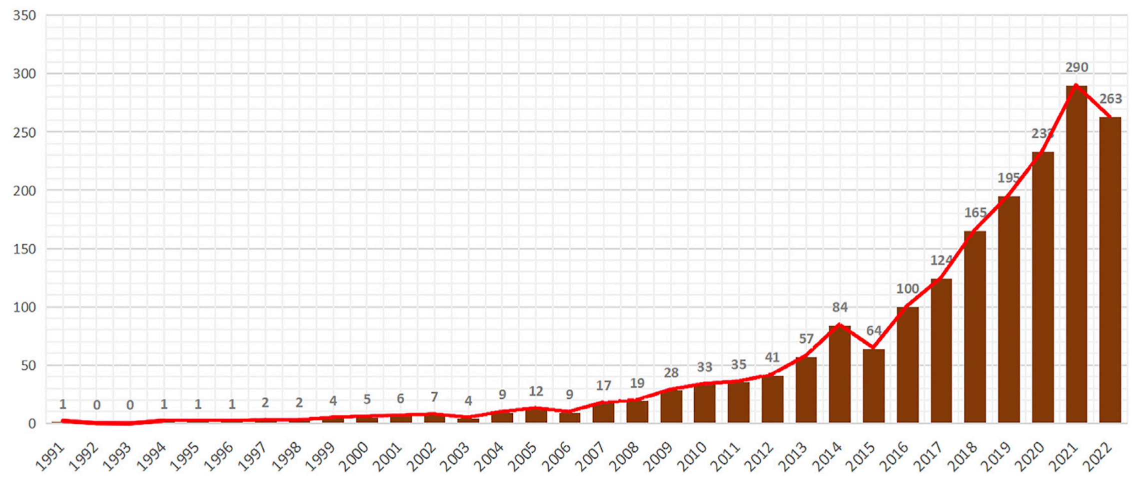


FIGURE 2
Trend of the number of studies on the relationship between sustainable built environment and user perception.

FIGURE 3
Knowledge map of the regional distribution of literature on the relationship between sustainable built environment and user perception.

on sustainable built environment and user perception (Figure 2). It shows that the annual number of publications on this topic has generally shown an upward trend since 2003 and exceeded 200 in 2020, indicating that scholars at home and abroad have gradually paid more attention to research on the relationship between the two in recent years. The research history can be divided into three phases: the first phase before 2003, when there was basically no research in this field, indicating that the relationship between sustainable built environment and user perception had not yet attracted the attention of scholars at the time; the second phase of slow development, 2004–2011, when people gradually paid more attention to their built environment as their material living standards improved; the third stage of rapid growth, 2012–2022, with a significant increase in the number of annual publications and active research dynamics, proving that user perception of built environment has become a popular aspect of the relationship between built environment and users, and also reflecting the importance and pursuit of people's spiritual feelings. Theoretical guidance was needed to keep pace with the growing material culture.

The regional distribution of the literature can visually reflect the popular countries and regions for research on the relationship between sustainable built environment and user perception. As can be seen in Figure 3, most of the scholars studying the relationship between the two are concentrated in the US, China, the United Kingdom, Australia, and Canada. The regional distribution of the literature also shows that the US is the earliest researcher on the topic and has a greater influence on other countries. As one of the most popular countries for research in this area, most of the research on the relationship between the two in China occurred after 2015, which is highly consistent with the third stage. This is closely related to China's urbanisation process, which reached a new level of domestic construction after two stages of expansion and coordinated development between 1993 and 2012.

3.2 Author collaboration network

The network mapping of co-authorship shows the relationships between scholars in the field, with larger nodes indicating greater influence of the scholar. Using CiteSpace to analyse the downloaded literature for co-authorship (Figure 4) yielded 654 nodes and 737 links, with a density of 0.0035. The analysis, combined with the regional distribution of the literature (Figure 3), shows that relevant research in the United Kingdom and the US began long ago, with clear collaborative relationships between authors, with Ilse De Bourdeaudhuij, James F. Sallis, and Benedicte Deforche being the most influential in the field and laying the foundation for research, while research in China started a little later but since gained great momentum, with Zhang Yin, Wang Zhe, and Liu Hong being more influential. Jacqueline Kerr, James F. Sallis, and others (Kerr et al., 2016), examining the strength and shape of the relationship between

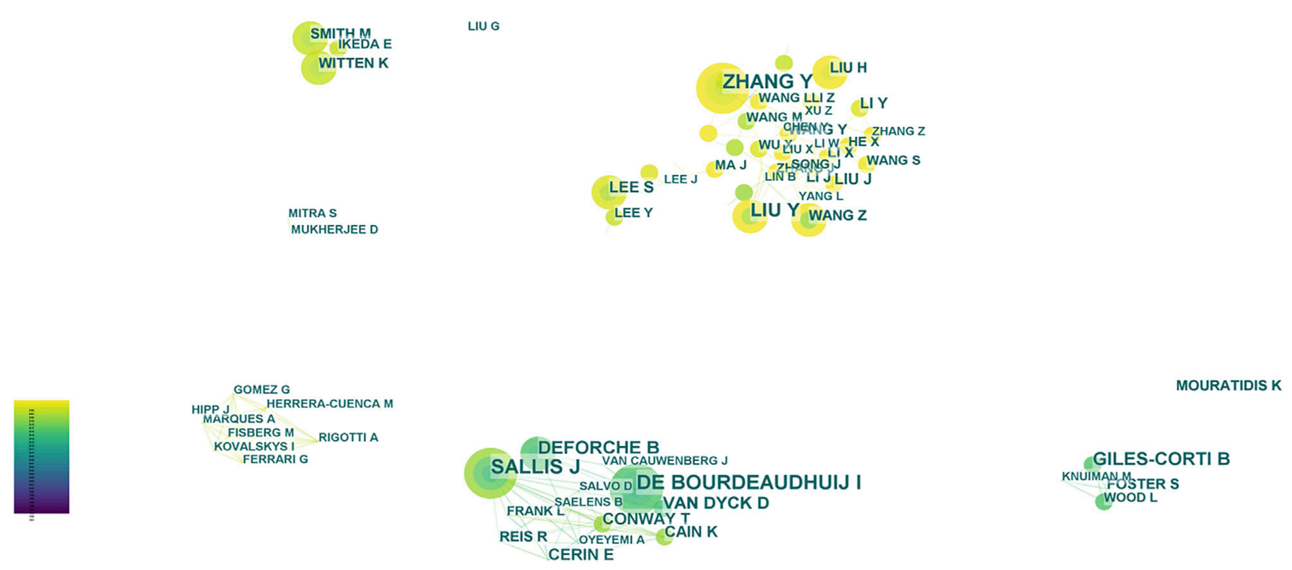


FIGURE 4
Authors' collaborative knowledge map for research of the relationship between sustainable built environment and user perception.

adults' perceptions of community and walking and cycling for transport in different environments, point out that many perceived environmental attributes support walking and cycling, that in a highly walkable environment, people are less likely to cycle as they choose to walk, and that these findings could guide the implementation of global health strategies. Ilse De Bourdeaudhuij and Benedicte Deforche together with Jelle Van Cauwenberg and other researchers (Van Cauwenberg et al., 2012) qualitatively studied the influence of the physical environment on older people's choice to walk, identifying influential elements such as functional facilities, walking facilities, traffic safety, familiarity, crime safety, social contact, aesthetics, and weather, and specifically stated that in order to encourage walking, the community should have more services and social spaces, and the perception of the environment should focus on crime prevention and a sense of security. Ester Cerin and Billie Giles-Corti work closely with the above researchers James F. Sallis and Ilse De Bourdeaudhuij. The IPEN (International Physical Activity and Environment Network) Recognition Project sought to aggregate and analyse the perceived community environment through the Neighbourhood Environment Walkability Scale (NEWS) and its intermittent version (NEWS-a) measure, using physical activity data from 12 countries and proposing country-specific modifications to the original NEWS/ NEWS-a scoring scheme to improve comparability between countries, although some differences between countries remain and need to be taken into account when interpreting the results for each country (Cerin et al., 2013).

3.3 Subject area characteristics

In this paper, a preliminary statistical analysis of the resulting 1815 documents according to WoS categories reveals that the research presents a characteristic of concentration in certain types of subject areas (Figure 5). The top fifteen categories are extracted for comparison with the others, with a ratio of about 5:1. The histogram of the top 15 categories shows that 'public, environmental & occupational health' is the largest with 358 articles, indicating that the relationship between built environment and users is of great interest to the public health sector. The categories 'environmental studies' and 'environmental sciences' account for more than 20% of the total number of articles, mainly due to their inherent properties related to sustainable built environment and user perception of the environment. The category 'architectural engineering', as an externalised manifestation of the relationship between the two, also receives a higher number of articles.

The co-occurrence mapping of subject areas derived from CiteSpace shows that 'public, environmental & occupational health', 'environmental studies', 'construction & building technology', 'civil engineering', 'transport', and 'environmental sciences' have a high citation intensity, while other clear nodes are 'green & sustainable science & technology', 'urban science & technology', 'urban studies', and other environmental disciplines (Figure 6).

The size of the nodes in the graph indicates the level of citation intensity, the line between the nodes represents the year when the literature between the two disciplines was first cited together, and the change in the brightness of the colour indicates how early research on the relationship between sustainable built environment and user perception emerged in the field, with the bluer the colour, the later the emergence. It can be seen that early research focused on the discipline of 'public, environmental & occupational health', but after 2009, the field of research on this topic began to expand with 'environmental studies', 'construction & building technology', 'engineering, civil', and 'geography'. After 2015, the original field of study expanded further, with 'urban studies', 'regional & urban planning', and other disciplines focusing on the relationship between sustainable built environment and user perception. The categories 'urban studies', 'regional & urban planning', etc. have all addressed this issue.

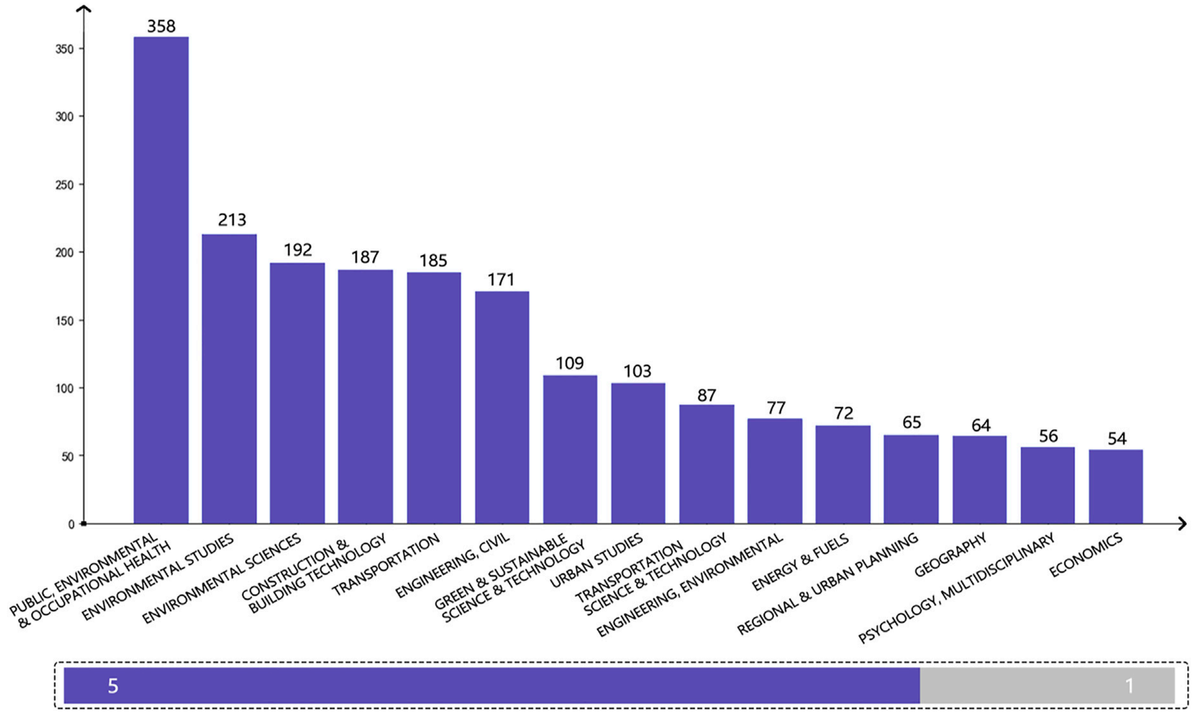


FIGURE 5
Discipline distribution map based on WoS categories.

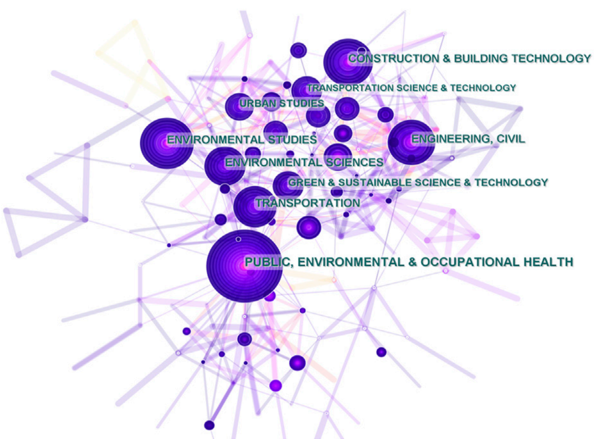


FIGURE 6
A disciplinary co-occurrence network for the study of the relationship between sustainable built environment and user perception.

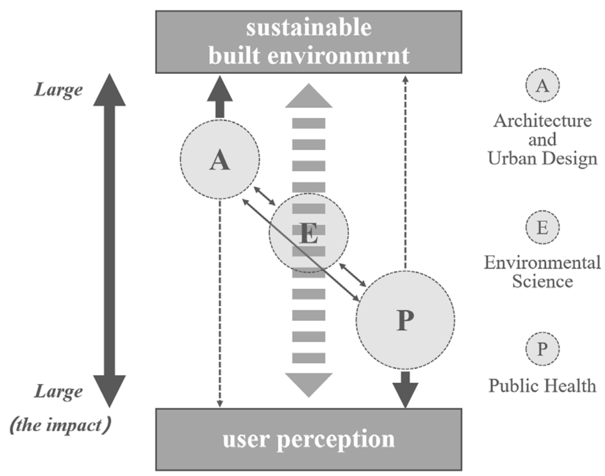


FIGURE 7
Diagrammatic representation of the relationship between the three subject fields.

In general, with the gradual expansion and deepening of research on the relationship between sustainable built environment and user perception, research on the topic has now formed an obvious 'one centre + two nodes + a number of branches' research structure layout, with the field of environmental science as the centre, and the field of public health and the field of architecture and urban design as the two nodes. In terms of research content, the field of environmental science serves as the core field of relationship between the two. In relationship research, the field of architecture and urban design is more inclined to the analysis and design of sustainable built environment, while the field of public health pays more attention to the perception and behavioural characteristics of users. The field of research has recently crossed over into the disciplines of psychology and computer science, reflecting the multidisciplinary and cross-study nature of this thematic research (Figure 7).

FIGURE 8
Keywords co-occurrence network diagram for the study of the relationship between sustainable built environment and user perception.

3.4 Research hotspots and trend analysis

In this paper, we used the CiteSpace software to analyse the keyword co-occurrence network and frequency of 1815 documents on the relationship between sustainable built environment and user perception from 1991 to 2022, with a 1-year time slice, node type ‘keyword’, pruning settings ‘pathfinder’ and ‘pruning sliced networks’, and other parameters as default. A network density of 0.0114 was obtained for 688 document nodes with 2,699 lines (Figure 8). The data was further counted and the top 20 high-frequency keywords were obtained (Table 1). As this study searched for ‘built environment’ and user perception’ as topics, the two keywords ‘built environment’ and ‘perception’ were excluded. The remaining high-frequency keywords in this area were ‘physical activity’, ‘health’, ‘walking’, ‘environment’, ‘impact’, ‘behaviour’, ‘perception of behaviour’, ‘design’, etc. As can be seen, the high-frequency keywords corresponded well to the characteristics of the disciplinary distribution mentioned above. The keywords ‘physical activity’, ‘health’, and ‘walking’ were more often studied in the context of public health, especially the topic ‘physical activity’. The keywords ‘environment’, ‘thermal comfort’, ‘happiness’, ‘satisfaction’, and ‘safety’ were more often associated with the field of environmental science. For example, Honf and Chen (Hong and Chen, 2014) studied the relationship between built environment, crime perception, and people’s walking behaviour. The study showed that people’s perceptions of the built environment affect their travel and use. The keywords ‘urban design’, ‘quality’, ‘virtual reality’, and ‘urban form’ are mostly based on architecture and urban design, using people’s perceptions of the built environment to identify and perceive urban images or landscapes.

Based on this, this paper used CiteSpace to cluster the literature nodes based on keywords, drew words from them by ‘log-likelihood rate’ to identify clusters, and obtained a total of 12 sub-clusters (Figure 9) with ‘modularity Q = 0.5232 > 0.3’ and ‘silhouette S = 0.7621 > 0.7’, which had a significant group structure and sufficient confidence (Chen et al., 2010). Sub-clusters ranged from ‘thermal comfort’ and ‘physical activity’ to ‘quality’ and ‘neighbourhood environment’, whereas ‘cancer causation’ was weakly related to the study topic and therefore excluded.

Combining Figure 10, it can be seen that the initial health-related terms ‘walking’, ‘physical activity’, and ‘environmental perception’ have matured and the research hotspots are gradually moving towards ‘urban scale’ and ‘emerging technologies’, with terms such as ‘neighbourhood design’, ‘city’, ‘virtual reality’, and ‘machine learning’ showing recent mutations.

TABLE 1 Keywords ranking list for the study of the relationship between sustainable built environment and user perception.

Serial number	Frequency	Keywords	Earliest time/ Year
1	508	built environment	2005
2	377	perception	2001
3	321	physical activity	2005
4	224	health	2002
5	197	walking	2005
6	188	environment	2003
7	147	impact	1999
8	128	behavior	2001
9	118	design	2005
10	117	model	2001
11	111	safety	2006
12	110	association	2008
13	97	thermal comfort	2002
14	89	happiness	2006
15	87	satisfaction	2010
16	85	quality	1999
17	80	performance	2012
18	67	neighborhood	2003
19	65	children	2007
20	65	obesity	2008

4 Trends in research on the relationship between built environment and user perception from a multidisciplinary perspective

The top ten references were filtered according to the frequency of citations. When analysing the journal sources of the cited literature, journals dealing with health and perception and the design of built environment at different scales dominated (Table 2), and almost half of them focused directly on the

 10.3389/fbuil.2023.1271889

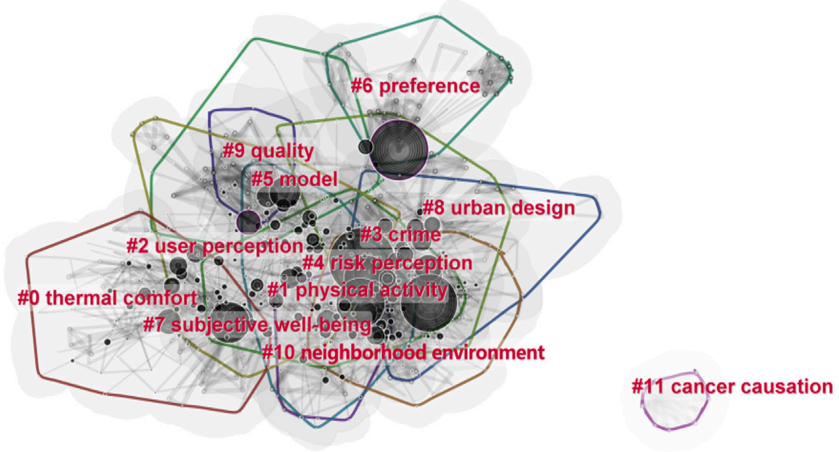


FIGURE 9
Keywords co-occurrence clustering map for the study of the relationship between sustainable built environment and user perception.

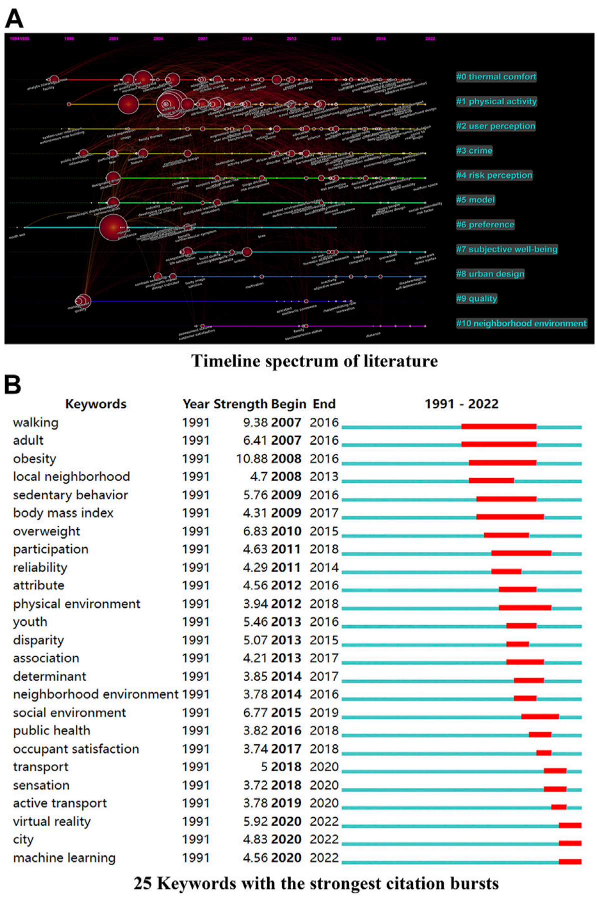


FIGURE 10
the Hotpots' tendency for the study of the relationship between sustainable built environment and user perception. **(A)** Timeline spectrum of literature **(B)** 25 Keywords with the strongest citation bursts.

relationship between environment and perception. The authors of the highly cited literature were mostly from three disciplinary backgrounds, namely, health, environment, and architecture and urban planning, reflecting to some extent the disciplines' health benefits and value as sources for promoting research on improving built environment.

Based on Table 2, the results are integrated according to the CiteSpace cluster analysis map to obtain Table 3, which shows that the research on the relationship between sustainable built environment and user perception, based on a multidisciplinary perspective, is divided into three main areas of concern: public health, environmental science, and architecture and urban design, with the corresponding research focus on 'health behaviour', 'environmental cognition', and 'image perception'. The three fields differ in their respective points of focus on the relationship: environmental science focuses directly on the mechanisms underlying the relationship, with an emphasis on quantifying and analysing user perception of sustainable built environment, whereas public health as well as architecture and urban design focus on applied research, with an emphasis on analysing, filtering, and applying perceptions within their own fields of study. The respective research trends are elaborated in all three areas.

4.1 External representation of the relationship between sustainable built environment and user perception from a public health perspective

From a public health perspective, sustainable built environment has a profound impact on the health of their users through the externalisation of user perception of the environment and their self-selection. Studies have shown that built environment has a strong correlation with user health (Barton, 2009). Environmental categories that could influence people's behaviour patterns, such as food resources,

TABLE 2 Top 10 cited literature for the study of the relationship between sustainable built environment and user perception.

Serial number	Frequency of citarions	Name of the journal article	Year of publication	Journal name
1	552	Literature survey on how different factors influence human comfort in indoor environments	2011	**Building And Environment**
2	452	The acceptance and use of a virtual learning environment in China	2008	**Computers & Education**
3	425	The relationship between nature connectedness and happiness: a meta-analysis	2016	**Frontiers In Psychology**
4	410	Multilevel modelling of built environment characteristics related to neighbourhood walking activity in older adults	2005	**Journal Of Epidemiology And Community Health**
5	390	User acceptance enablers in individual decision making about technology: Toward an integrated model	2002	**Decision Sciences**
6	304	Cycling and the built environment, a US perspective	2005	**Transportation Research Part D-Transport And Environment**
7	290	Benefits and wellbeing perceived by people visiting green spaces in periods of heat stress	2009	**Urban Forestry & Urban Greening**
8	272	Residents' perceptions of walkability attributes in objectively different neighbourhoods: a pilot study	2005	**Health & Place**
9	228	City structure, obesity, and environmental justice: An integrated analysis of physical and social barriers to walkable streets and park access	2009	**Social Science & Medicine**
10	211	The Collaborative Image of The City: Mapping the Inequality of Urban Perception	2013	**Plos One**

TABLE 3 Keywords clustering analysis for the study of the relationship between sustainable built environment and user perception.

Cluster number	Cluster name	Number of sub-cluster	Key areas of research
#1	**physical activity**	113	**Public Health area - Health behaviour -**
#5	**model**	47	
#0	**thermal comfort**	117	**Environmental Science area - Environmental awareness -**
#2	**user perception**	80	
#3	**crime**	75	
#4	**risk perception**	51	
#6	**preference**	41	
#7	**subjective wellbeing**	34	
#9	**quality**	21	
#8	**urban design**	28	**Architecture and Urban Design area - Image perception -**
#10	**neighborhood environment**	15	

information patterns, public facilities, and medical environments, have all been addressed in the study of built environment from a public health perspective (Ma and Cai, 2016). Thus, the relationship between sustainable built environment and user perception has become more of a health-oriented behavioural profile, and more disciplines are now focusing on the role of the environment in influencing population health, such as urban and rural planning, health geography, sociology, medicine, etc. Sallis et al. (Sallis et al., 2006) put forward the famous social ecology theory based on the intersection of multidisciplinary fields at the time, summarised a more complete active living ecological model based on a large number of health behaviour theories, and sorted out the elements of environmental interventions to promote health behaviour, which expanded the scope of research to explore the relationship between sustainable built environment and user perception based on health behaviour research. In *A New Perspective on the Health of Canadians*, Lalonde (1974) identified lifestyle, environment, medical facilities, and human biology as the four most important factors influencing health, with non-communicable diseases overtaking communicable diseases as the main threat to human life (except in specific emergencies). The 'environment' is the external factor affecting human health, while 'human biology' is broadly defined as the body's own characteristics, especially certain hereditary diseases such as

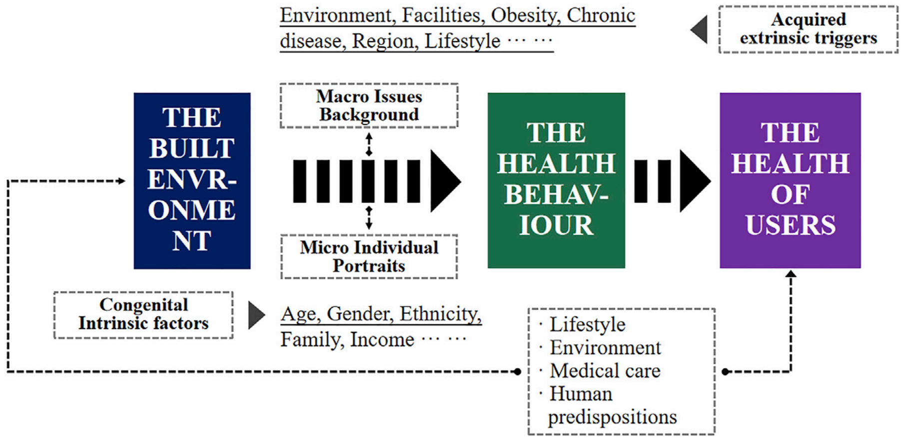


FIGURE 11
Graphical representation of the relationship between sustainable built environment and health behaviour.

obesity, hypertension, and myopia. The four types of elements interact with each other, with health-related behaviours being an important area of public health research into the relationship between the sustainable built environment and user perception, as well as an important way of influencing human health, of which physical activity is a particularly important part. The relationship between behaviour and health under the three main acquired factors of environmental quality, medical facilities, and obesity, as well as the relationship between human biology and physical activity are analysed below (Figure 11).

4.1.1 Research of behavioural patterns based on a multi-genre context of the times

In different contexts, user perception of various aspects of sustainable built environment influences corresponding behavioural patterns. In a multi-dimensional context, with the development of built environment and social awareness, the behavioural patterns of residents are becoming more complex, with significant health implications. The planning and design of sustainable built environment should, therefore, use the interdisciplinary research results and take into account environmental, social, and economic factors in order to promote the adoption of healthier and more sustainable behavioural patterns. Long-term indoor stay could lead to problems such as sedentariness and lack of exercise; crowding of activity and interaction spaces could lead to physical health problems such as obesity, respiratory diseases, and chronic illnesses (Lu and Tan, 2015); highly homogeneous urban environments and increasing social pressures could lead to mental health problems such as anxiety, loneliness, and depression (Qin et al., 2018). These suggest that sustainable built environment should not only focus on the physical aspects of sustainability, but also consider how to create more socially interactive and emotionally resonant spaces to improve user mental health.

In the context of the growing problem of environmental pollution in urban areas, air pollution (e.g., PM2.5) has become one of the major factors threatening the health of residents (Gee and Payne-Sturges, 2004). Due to the low quality of the environment, people are reluctant to spend too much time outdoors. Also, activities such as walking and cycling can only take place in certain places, making travel behaviour increasingly homogeneous. Studies of residents' travel behaviour have analysed the impact of factors such as mode of transport and commute time on their health status (Macmillan et al., 2013). For example, Yang and French (Yang and French, 2013) studied the relationship between transport use and body mass index (BMI). The analysis of air pollution exposure based on individual behavioural research paradigms in residential travel patterns and micro-behavioural settings would be one of the frontier topics.

In the context of equalisation of services, research on access behaviour could provide a basis for optimal allocation of healthcare resources, with accessibility and individual factors being the most critical. Andersen's model of healthcare behaviour is a widely used classic model that categorises the factors that influence an individual's decision to seek medical care as 'predisposing factors', 'enabling factors', and 'need factors' (Kehrer et al., 1972). In sustainable built environment, this model also needs to be integrated with environmental accessibility and user perception to ensure that healthcare resources are allocated appropriately and people's needs are met.

In the context of the rapidly growing obesity problem, the density of housing in built environment (Rodríguez et al., 2009), the number of commercial establishments (Nagel et al., 2008), the land mix, the walkability of neighbourhoods (Li et al., 2009), the density of public transport (Bird et al., 2009) etc. influence people's physical activity and thus their levels of obesity, but even more relevant is the food service environment, which could also lead to obesity if it does not offer healthier foods and there is an imbalance between energy intake and expenditure (Brug et al., 2006). Researchers such as Morland (Morland et al., 2006) examined the location of supermarkets and other food outlets and found that adults living in areas with supermarkets had lower rates of obesity than those without. Alter and Eny (Alter and Eny, 2005) found that the density of fast food outlets in Canada was associated with cardiovascular disease. This requires consideration of how the eating environment can be improved to encourage healthier food choices and provide convenient healthy eating options, and occupants can be encouraged to make healthy choices and avoid unhealthy behaviours. Therefore, the relationship between sustainable built environment and user perception based on dietary behaviour also needs to be studied in depth in order to promote health and sustainable development goals, which is the focus of current academic research.

4.1.2 Research on physical activity based on a diverse population profile

From a public health perspective, the relationship between sustainable built environment and physical activity depends not only on the characteristics of the physical environment, but also on user perception of and interactions with the built environment. Related research needs to consider the interaction between these two factors more comprehensively to better understand how built environments affect physical activity and health. Research on the relationship between built environment and physical activity in public health has reached a certain stage in many highly urbanised countries. Epidemiological studies were the main focus until 1996, and the second stage, 1997–2003, saw a steady annual publication of physical activity research involving built environment and affirmation of the health-promoting effects of physical activity (Ma et al., 2019). In September 2003, both the *American Journal of Public Health* and the *American Journal of Health Promotion* published special issues on the relationship between built environment, physical activity, and health, leading to a surge of literature in the third phase that has had a significant impact on the field of public health research (JF, 2006; Sallis et al., 2013). The field now divides physical activity into four categories: 1) work-related physical activity; 2) household physical activity; 3) transport physical activity; and 4) leisure or recreational physical activity (Lee and Moudon, 2004). In addition to the growing body of research on built environment, there is a growing body of environmental indicators. Canadian researchers Gavin R Mc Cormack and Alan Shiel examined the relationship between built environment and physical activity in adults using 20 cross-sectional studies and 13 quasi-experimental studies to establish a causal relationship between adult activity and built environment in communities across geographical areas, and identified environmental factors that had a significant impact on behaviour (Liberati et al., 2009).

As the field develops, more and more studies have begun to consider the relationship between sustainable built environment and user perception. The concept of a sustainable built environment includes not only the improvement of urban infrastructure, but also the sustainability of buildings, transport systems, and public spaces. In terms of user perception, users' views and experiences of the environment have also begun to receive attention, as these can influence their behaviours and health choices. For example, a user's willingness to walk or cycle may depend on their perception of the urban environment, including road safety, landscape aesthetics, and accessibility. However, differences in individual needs are often overlooked, resulting in a lack of thoughtful design of environments for different groups and a lack of public facilities related to health services. For older people who are less physically active, participation in physical activity is low, and loneliness and lack of moderate activity are two major barriers to healthy living: the prevalence of chronic disease is 3.2 times higher in older people than in the general population, with 64% of those aged 65 and over suffering from at least one chronic disease and more than a third suffering from varying degrees of mental illness (Jonasson et al., 2017). For older people, a short, flat, and directional road network (Burton, 2012) and high quality parks and public open spaces (Nielsen and Hansen, 2007) contribute to higher travel rates and healthier travel choices. The number of social spaces in a community (Maas et al., 2009) and access to services such as healthcare (Chiou and Chen, 2009) could increase their sense of belonging and community. Studies have shown that regular participation in leisure-time physical activity delays ageing, reduces the incidence of non-communicable diseases, and improves quality of life (Lahti et al., 2014).

Physical activity is particularly important for the development of adolescents. Encouraging them to be physically active and exercise by optimising the quality of built environment is a new concept to promote healthy living in the context of 'Healthy China'. However, the current highly informed and intelligent lifestyle has led to a lack of physical activity among adolescents (Monda et al., 2007). The prerequisite for effective interventions in this phenomenon is an in-depth study of the influencing factors and the identification of those elements of built environment that have a significant impact on adolescents' physical activity and health and are easy to change (Trapp et al., 2012). The link between built environment and adolescents' health activities is studied from a public health perspective. On the one hand, by optimising the quality of built environment, it is possible to break with the original controlled interventions based on the behavioural activities of adolescents, their families, or society, and to intervene actively, opening up new avenues of intervention. On the other, the built environment, as an important interface for health interventions at the macro level, could be planned and designed better, addressing adolescent health from the urban level.

In addition to age (adolescents, middle-aged adults, older adults), people can also be categorised by income level (low, middle, high) (Wallace et al., 2000; Xian Yu et al., 2011) or by gender. Female older adults are more susceptible to disease than male, but regular physical activity could reduce mortality in women (Rockhill et al., 2001), reduce joint pain (Bruce et al., 2002), increase flexibility, balance, speed, and reaction time (Liu-Ambrose et al., 2004; Tinetti and Kumar, 2010), and reduce the range of chronic diseases caused by obesity. However, the effects of built environment on leisure-time physical activity and body mass index in older women have been reported in some reviews and empirical studies (Chen et al., 2018), but not specifically from the perspective of older women, and follow-up studies are needed to fill this gap. Overall, in the field of public health, the research focus is on physical activity, which creates an important connection between people and their environments. Only by creating a well-perceived built environment can it fundamentally contribute to the quality of life of the inhabitants, which, in turn, feeds back into the improvement of the built environment.

4.2 Intrinsic mechanisms of the relationship between sustainable built environment and user perception from an environmental science perspective

People perceive information as a process of acquiring awareness or understanding perceptual information. In the field of environmental science, the perception of sustainable built environment has received more and more attention in recent years. User awareness of the built environment and the relationship between the user and the built environment are created from the perception of the environment through multisensory cells (Bell, 2012). The perception of the environment is the basis of spatial creation, so clarifying the structured relationship between the physical environment and the subject's perception is one of the inevitable directions to realise the 'return to humanism' in the current urban transformation process in China. Environmental science provides a quantitative and scientific basis for studying the relationship between sustainable built environment and user perception. The physical characteristics of the built environment influence its perceived characteristics. User perception of the built environment can influence their psychological state and behavioural patterns, for example, an environment perceived as safe and comfortable will enhance users' satisfaction and wellbeing, and at the same time motivate them to participate more actively in community activities, environmental protection, and other positive behaviours. Gold (Gold, 1980) suggests that the perception of the environment is the psychological environment formed by the user after receiving and processing information in the physical environment, and that this could guide external behaviour. Sustainable built environment and user perception are mutually influential. Several studies have shown that the daily lives of residents are closely related to many factors of the urban built environment, such as 'land use', 'functional layout', 'transport links', 'green space', 'environmental safety', 'accessibility', and other spatial attributes that have become the most important aspects of built environment. However, there are fewer perceptual elements based on a human perspective, especially objective spatial combination elements and subjective psychological perception elements. It is also the basis for the formation of a link between perception and behaviour.

4.2.1 Trends in perceptual content research: microcosm and diversity

While we now have mature evaluation methods for objective environmental measurements, such as environmental form, density, scale, greenery, sky openness, etc., there is no comprehensive evaluation framework or indicator system for subjective user perception and experience of sustainable built environment. Residents' daily lives and behaviours, and as such, their internal perception, are always influenced by the micro built environment. Therefore, clarifying the relationship between the two can positively improve it, which is conducive to the optimisation of the urban environmental space and the improvement of the healthy life of the residents. While the starting point for urban sustainable built environment and user perception is mostly macroscopic, lacking the matching of active perceptions and indicators of environmental elements at the microscopic scale, the human scale is mainly divided into perceptions of thermal comfort and quality of the street environment, such as safety, satisfaction, enjoyment, playability, etc.

The Chinese Healthy Building Standard, introduced in 2017, and the US WELL® Healthy Building Evaluation Standard (Institute, 2014) specifically address the evaluation of the indoor environment with the aim of promoting mental health. The sources of stimulation in the indoor environment of buildings specifically include two categories of 'interior design factors' (such as interior landscape, window views, interior interfaces, and spatial division) and 'physical environment factors' (such as thermal environment, light environment, sound environment, and indoor air quality) (Cooper and Marshall, 1976; Edwards and Burnard, 2003). The effect of preventing mental health problems of occupants by regulating the elements of the indoor environment is one of the key directions of built environment construction. In addition to the basic elements of traditional architectural design, the study of light environment has formed a more complete framework and system. With climate change in recent years, the indoor thermal environment has gradually attracted attention because it directly affects the physical and mental health of people and their willingness to move. Thermal comfort is the degree to which people perceive the thermal environment of space as good or bad, and its evaluation criteria have changed from being determined by temperature alone to incorporating multiple factors such as humidity, airflow, and radiation, and the evaluation model as a whole is divided into two types of models: mechanistic and empirical (de Freitas and Grigorieva, 2015). The study of the perception of thermal comfort has become an important topic of research.

The spatial environment of the street has both transport and public space attributes, and it is important to analyse the factors that influence the vitality of the street space from a human-centred scale and to redesign it to improve the spatial quality of the whole city and activate the vitality of the neighbourhood. The relationship between sustainable built environment and user perception is not just a theoretical link, but a practical necessity to achieve the goal of sustainable urban development by enhancing the perceptual experience of citizens through design and advanced technologies. In previous studies, the objective physical elements had mostly been matched with data on the vitality of the population to determine the framework of influence, without the intervention of subjective perceptions from a micro perspective. The process of human mental processing of the environment in environmental psychology divides the spatial perception framework into four generic layers: 'sensory', 'perceptual', 'cognitive', and 'behavioural' (Wang R. et al., 2020). Among them, with emphasis on the youth population and the aggravation of the ageing problem in society, the research on pedestrians' perception of safety in the street walking environment has become a hot research topic. Other related research theories are more abundant abroad, such as 'street eye', 'broken window theory', 'situational crime prevention theory', etc. Baran et al. (Baran et al., 2018) explored the sense of security in residential parks through immersive virtual environment experiments and showed that spaces with low and medium levels of enclosure provide a greater sense of security. Kuo (Kuo et al., 1998) found in a related study that the density of tree planting and the level of lawn maintenance in settlements had a greater impact on the sense of security and showed a positive correlation.

4.2.2 Trends in perceptual quantification: objectification and informatisation

The study of the relationship between the sustainable built environment and user perception in the field of environmental science focuses not only on specific perceptions, but also on how such perceptions can be quantified in order to provide a database for subsequent analysis of the relationship between the two.

Establishing correlations between spatial environmental indicators and user perception is a necessary part of combining a large amount of theory and practice, and ensuring that the relevant information is measured accurately and comprehensively is the foundation of this process. As the population has become more diverse, the research has become more sophisticated and the methodology has changed from subjective questionnaires to objective measurement techniques such as GIS, GPS, and pedometers, which are widely used in many watches and mobile phones as part of 'health monitoring functions'. Due to the difficulty of obtaining micro-data at street level and the lack of emphasis on micro-environments, the richness of micro-environments is often represented by coarse-grained studies of community plots (Long, 2016), leaving most of the current research under qualitative studies and lacking rich quantitative studies.

The two main types of measurement methods are qualitative analysis of survey indicators based on offline research and quantitative analysis based on geo-information data systems and digital technology tools. The former is the direct perception of the environmental climate by respondents in the field measured through questionnaires, interviews, or physiological assessments and behavioural observations to obtain comprehensive and intuitive information about the environment (Ewing and Handy, 2009). It can be divided into participant subjective assessment and researcher observational assessment. Participant subjective evaluation generally focuses on analysing residents' perceptions of built environment through their self-reported data and residents' environmental rating data, usually using paper questionnaires, electronic questionnaires, telephone surveys, face-to-face interviews, cognitive mapping, image recognition, and so on. Among many scales, the WHO–5 scale has been used in various foreign countries and groups and has been shown to have high reliability and validity in numerous mental health studies (Barton, 2009). Among these, researcher-monitored assessment is mainly based on observing and identifying elemental indicators and population information at a specific location and time, and then using statistical analysis models to match them, but it is time-consuming and labour-intensive. Many current monitoring tools, such as systematic walking and cycling, environmental scan (SPAC-ES), physical activity resource assessment (PARA) tool, Bedimo-Rung assessment tool - direct observation (BRAT-DO), etc. have improved in efficiency and accuracy.

For the latter, digital measurement tools are gradually replacing most of the traditional measurement methods. GIS technology and digital means can be used to collect and analyse the spatial characteristics of the built environment to provide quantitative spatial information for perception studies. Smartphone applications and social media platforms can collect realtime feedback and evaluation of the built environment by users, providing a large amount of perceptual data. Currently, machine learning is one of the major research trends and hotspots. For the collection and analysis of objective built environment elements, it mainly includes GIS technology and remote sensing image technology, and can use geographic information systems (GIS), open street map (OSM), Tencent map, and street image recognition means. Among these, WalkScore is the most representative for measuring the walkability of the physical environment, based on the type and spatial arrangement of everyday facilities, while introducing factors such as walking distance decay and intersection density to improve accuracy. Subjective measures of user perception can be obtained through smartphone signalling or through intelligent measurements from physiological sensors, represented by eye-tracking metrics (ETM), salivary cortisol, electroencephalogram (EEG), MRIs, hormone tests, etc. With the maturation and mobility of biosensors (Tang and Long, 2017; Zhang et al., 2018), these technologies offer new ways to intelligently measure the quality of built environment and also to explore the relationship between sustainable built environment and user perception through virtual reality and augmented reality techniques. These two aspects are, therefore, discussed below.

4.3 External representation of the relationship between sustainable built environment and user perception from an architecture and urban design perspective

Architecture and urban design are based on spatial perception, which is a real, direct, and comprehensive experience of human perception of the environment (Merleau-Ponty, 2004). In exploring the relationship between sustainable built environment and user perception, the field of architecture and urban design favours the study of the physical space of the built environment, which is also, in this field, the external expression of user perception. Architecture and urban design professionals should not only create built environments, including buildings and public spaces, and perform urban planning, but also create spatial sensations based on the results of environmental science research on user perception and experience. In the past, the design of built environment focused more on functional arrangement, planning and theoretical techniques, but less on the perception and behavioural characteristics of citizens in urban space. Given the richness of microscopic research perspectives, the guiding ideology of urban planning and design should shift from 'material-oriented' to 'human-centred', and gradually pay attention to the daily lives of residents and the distinctive culture of the city. This is an effective way to improve the quality of built environment at both the meso-community and macro-city scales through high quality planning and distinctive cultural design to enhance the comfort of residents, enhancing the experience and quality of life for residents and create a city's calling card.

4.3.1 Community environmental landscape study

The sustainable built environment is closely related to the perception of its users. The community environment is the one that is closest to people and felt the most. One of the most common and distinctive urban spatial phenomena in China is the unitary community, which is heavily influenced by history. However, the current design of community environments lacks humanistic thinking and rarely considers the relationship between sustainable built environment and user perception, with a lack of landscape facilities and insufficient emergency measures and capabilities. Previously, research on community environment focused on 'landscape elements in residential areas from the perspective of elderly health' and 'mechanisms of the effect of residential landscape environment on the health of the elderly'. A good residential outdoor environment could improve the physical performance of older people, improve sleep quality, reduce stress, and maintain a positive emotional state (Rantanen et al., 2012; Skalicky and Čerpes, 2019). In the keyword knowledge network and the clustering map, the words 'community', 'neighbourhood environment', 'landscape' occupy high frequencies. Especially under the influence of the global outbreak of the severe public health event, the New Hall pneumonia epidemic, people have begun to pay special attention to landscape planning and design in community environments, gradually focusing on the influence of mental health and community environment for all age groups. Hur (Hur et al., 2010) found that community building density and community vegetation coverage were related to residents' life satisfaction, but Dong H and Qin B (Dong and Qin, 2017) showed that only community green space significantly influenced residents' mental health, and it is a new research hotspot to clarify the relationship between the two in light of new demands. This also suggests that we need to pay more attention to creating and protecting green spaces in the design of community environments, especially in the design of unit communities, and improving resident mental health through landscape planning.

4.3.2 Urban cultural landscape identification

The built environment of a city is vast and various, but it always affects the internal perception and the external performance of its inhabitants. In the context of globalisation, urban form and architectural design are converging and becoming more universal. Cities are gradually losing their unique spiritual qualities. This 'urban homogenisation' reflects not only the unique influence of modern technology and materials, but also the loss of regional architectural culture. In this scenario, there is an urgent need to solve the problem of creating and inheriting urban characteristics and culture in the context of new urban development. This problem can be effectively solved in a point-to-point manner, taking the built environment and the user perception or experience, and using as a starting point the results of research, quantification, and analysis of the relationship between the two in the fields of public health and environmental science. People mainly understand space through four aspects of perception: visual, tactile, olfactory, and auditory. Urban colours are perceived through vision, from perception to imagery, and then to recognition and identification of urban culture (Figure 12). Big data in the context of information technology has expanded the spatial and temporal dimensions of urban space, bringing an opportunity to refine urban planning and management, gradually moving from top-down control and adjustment of indicators to perceiving the city from the users' perspective, using open maps, streetscape images, virtual reality technology, and other means (Long and Tang, 2019). Won, Lee, and Park (Won et al., 2020) investigated the effect of colour combinations between shop signs and buildings on colour harmony and legibility; Zhong et al. (Zhong et al., 2021) identified and extracted the dominant colours of urban buildings from street scene images and guided optimisation strategies. The study takes 'people' as the object of study, uses various measurement methods, uses 'subjective perceptions' as an intermediate medium and digitises them, uses statistics and analysis to investigate the correlation between the objective built environment and the subjective perceptions of users, and then translates the subjective perceptions into actionable aspects of the physical. The general idea behind this research is to use machine learning to improve the translation of the results into digital technology. This will be an important research trend in the future.

4.4 Compilation of multidisciplinary research

Based on a comprehensive study of disciplinary background, research content, and research methodology, this study finds that the research boundaries of various disciplines are not clear, and that the *status quo* of multidisciplinary intersection has become the norm. Therefore, it is useful to combine and integrate the results of research on the relationship between sustainable built environment and user perception from the fields of public health, environmental science, and architecture and urban design to understand their lineages, value, and significance as well as to identify the characteristics, problems, and future research directions (Table 4).

5 Review and prospects of multidisciplinary relationships in the study of the relationship between sustainable built environment and user perception

Multidisciplinarity, in a narrow sense, refers to the interface between the natural sciences and the humanities, but also, more generally, to broader disciplines that involve the intersection of different disciplines. In general, research on the relationship between sustainable built environment and user perception has given rise to a network of disciplines, such as public, environmental, and occupational health, environmental science, building science and engineering, transport engineering, geography, and psychology, which lays down a disciplinary vein structure and provides a rich knowledge base and profound technical support for the field, in terms of both research structure and content. The interaction of the disciplines with the field and their complementarity create a distinctly multidisciplinary character.

5.1 Network relationships of research structures

The relationship between sustainable built environment and user perception is structured as one centre + two nodes + several branches (Figure 13A). The field of public health is the one centre, which initially studies the relationship between built environment and physical activity based on lifestyle from a public health perspective, followed by research hotspots such as travel behaviour and research models; the field of environmental

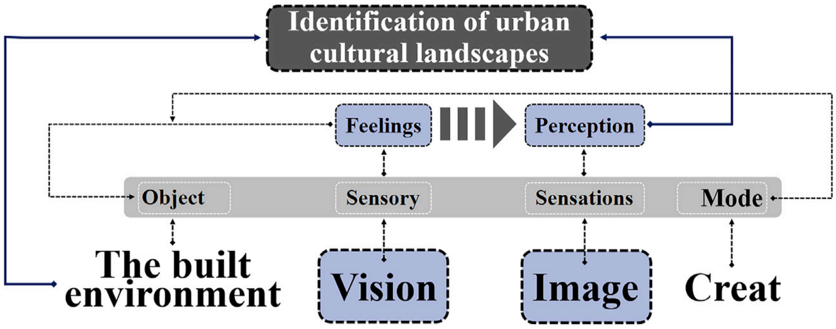


FIGURE 12
Flow chart of urban cultural landscape identification based on visual perception.

TABLE 4 Research results from various disciplines on the relationship between sustainable built environment and user perception.

Research field	Research dimension	Research content	Research methodology	Academic background	Research level
Public Health	physiological feature (extrinsic manifestation)	Physiological impacts of sustainable built environments on users in the context of psychological perceptions; external manifestations of the perceptual relationship between the two	Instrumental Measurement Social Observation	Sociological humanities medicine environmental hygiene	Applied Research
	Behavioural activities (extrinsic manifestation)	The direct impact of a sustainable built environment on the daily behaviour and life patterns of its users; the study of the characteristics of human behaviour in the environment	Social Observation Analysis of the phenomenon Behavioural analysis	Philosophy Sociological humanities medicine environmental hygiene	Theoretical research Applied Research
Environmental Science	psychological perception (intrinsic characteristics)	Intrinsic characteristics of the relationship between the sustainable built environment and user perceptions; pathways of influence in the formation of this relationship	Survey interview Case Study Psychological experiment	environmental science Sociological humanities	Theoretical research Fundamental research Interpretation tool
	quantitative analysis (intrinsic characteristics)	Direct/indirect measurement of sustainable built environment and user perceptions; analysis of the relationship between qualitative and quantitative measures	Instrumental Measurement public engagement	environmental science Sociological humanities Statistics	Applied Research Interpretation tool
Architecture and Urban Design	geographic planning (extrinsic manifestation)	Spatial Structure and Functional Distribution of Sustainable Built Environment; Focus on Urban Character and Humanistic Care	spatial analysis Case Study public engagement	Architecture Geography Urban and Rural Planning Landscape Architecture	Theoretical research Applied Research
	physical space (extrinsic manifestation)	Spatial design and sense of place creation under the core idea of human-centredness; targeted creation of different space types	Behavioural analysis spatial analysis public engagement Case Study	Architecture Urban and Rural Planning Landscape Architecture	Theoretical research Applied Research Interpretation tool

science and the field of architecture and urban design are the two nodes, with perceptual contents such as safety, comfort, walkability, indoor thermal comfort, and urban research as research hotspots. The addition of subfields such as psychology, computer science, and sociology enriches the network density and hierarchy.

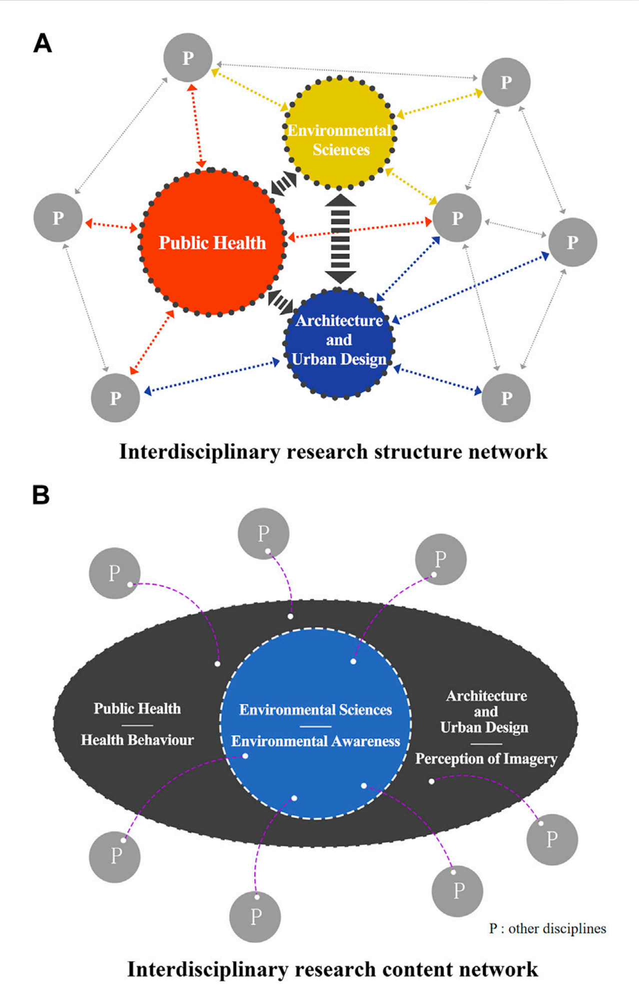


FIGURE 13
Interdisciplinary research network for the study of the relationship between sustainable built environment and user perception. **(A)** Interdisciplinary research structure network **(B)** Interdisciplinary research content network.

5.2 Network relationships of research content

The relationship between sustainable built environment and user perception is expressed in terms of research content in the form of concentric circles: core layer + outer circle (Figure 13B). Among the three main fields, the field of environmental science—environmental cognition—is the core, based on the study of users' direct perception of built environment and its externalisation in the expression of activities such as 'physical activity', 'urban design'. Of the various research fields in the outer circle, the fields of public health and architecture and urban design are not concerned with the study of perceptual content, but with the direct coupling of the relationship between the independent variable (built environment) and the dependent variable (externalised expression), using the results to speculate on the causal elements, while the other fields are more often added to the overall research network as research tools and grounded theory. The specific ways in which they affect each other are as follows.

5.2.1 Contribution of the public health to the study of perceptual relationships in the other two major fields

Public health is concerned with how the built environment affects the physical and mental health of its inhabitants. The findings from this field can provide environmental scientists with clues as to which environmental parameters—for example, air quality, water quality, or noise levels—may have a negative impact on health.

5.2.2 Contribution of the environmental science to the study of perceptual relationships in the other two major fields

Environmental science is responsible for measuring and monitoring key parameters of the built environment, which can provide a basis for public health research and improvement of indoor environments. The findings from this field can influence architecture and urban design. For example, research on indoor air quality can inspire architects to design better ventilation and filtration systems for the built environment that have a direct impact on the health and quality of life of its inhabitants.

5.2.3 Contribution of the architecture and urban design to the study of perceptual relationships in the other two major fields

Architecture shapes indoor environments by designing the layout, structure, and materials of buildings. Urban design improves urban environments by creating green spaces and walking/cycling paths and reducing traffic pollution, as well as by engaging the community and influencing behaviour. Public health can be influenced through community engagement, public space design, and urban planning that promotes healthy behaviours. Thoughtful design can inspire community residents to adopt healthier lifestyles.

5.3 Interdisciplinary collaboration and perspectives

Different research areas have different focuses and relative shortcomings. For example, the field of public health is more concerned with disease prevention, epidemiology, and the health and behaviour of users, and may lack a deeper understanding of the specific environmental parameters in architecture and urban design and how public health can be improved through design.

Interdisciplinary collaboration should capitalise on strengths in research content and pay more attention to the way in which the collaboration itself takes place: 1) Establish interdisciplinary research teams comprising public health experts, environmental scientists, and architects and urban designers to study the relationship between sustainable built environment and user perception. Such teams can work together to develop research plans, design experiments, collect data, and analyse results. 2) Share data and research tools across disciplines to better understand the links between sustainable built environment and user health and perception. 3) Jointly develop and apply integrated assessment methodologies that take into account the interactions between environmental parameters, social factors, and individual health, which may include the use of quantitative and qualitative methods to analyse, in an integrated manner, the impact of sustainable built environment on user perception. 4) Collaborate on field trials and case studies to test theories and drive innovation, for example, develop integrated construction projects in specific urban areas to test the impact of new urban planning and building design concepts on user health and perception.

6 Conclusion

Overall, the disciplines have different emphases in studying the relationship between sustainable built environment and user perception, and different network relationships in terms of research structure and content. In terms of research structure, the field of public health, which is researched earlier and published more papers, is taken as the centre, and the field of environmental science and the field of architecture and urban design are taken as the two nodes for subsequent development, forming a structure of 'one centre + two nodes'. In terms of research content, environmental cognition in the field of environmental science is at the core, based on the study of users' direct perception of the built environment and its reflection in behavioural performance, while the fields of public health and architecture and urban design are dominated by the study of perceptual outcomes. Currently, there is a trend towards close collaboration between the different disciplines working on this topic, and interdisciplinary work has an important role to play in the field of research on sustainable built environment and user perception, which requires the synthesis of knowledge and methods from different fields to study and solve problems in depth, and which can bring together people and ideas from different disciplines to construct problems, agree on methodologies, and analyse data, with the possibility of drawing new synthesised conclusions [E]. Methods from other fields can often solve the confusion of a single discipline, and the current means of measuring the object of study in various fields have a tendency to develop into big data. With the continuous deepening of the research and application of computer technology, such as the increasing improvement of web-based GIS technology, interdisciplinary and cross-sectoral research and technological synergies will become closer and closer, and the results of research on this topic will provide a steady stream of inspiration for the establishment of motivated lifestyle and behavioural habits in the future.

In the future, in order to improve the research network on sustainable built environment and user perception, it is necessary to further deepen the interdisciplinary cooperation and research content, build a more scientific and mature environmental spatial evaluation system, and then use quantitative analysis to study the inner mechanism of sustainable built environment on psychological perception, and further study the environmental psychological behaviour mechanism between the two, which will help enrich the civic activities, improve the street environment, create the community atmosphere, and form the spiritual image of the city. By visualising the information in the 1815 documents obtained from the search, the following conclusions can be drawn.

1. In terms of annual publication volume, research on the relationship between sustainable built environment and user perception has shown an increasing trend over the past 26 years, with the period from 2012 to the point of search showing a rapid increase in the number of articles on this topic, indicating that the dynamics of built environment–user relationship research will continue to be active in the future, and that there is a promising future for the analysis of built environment–user relationship from a perception perspective;
2. In terms of regional distribution, research in the US started early and developed to more maturity, followed by China, the United Kingdom, Australia, and Canada, but there are clear differences in the volume of research, with most of the Chinese studies occurring after 2015. In terms of author collaboration networks, the United Kingdom and US authors are well-established and their networks clear, with Ilse De Bourdeaudhuij, James F. Sallis, and Benedicte Deforche being the most influential in the field, and the trend of cross-discipline research leading to increasingly intensive collaborations between authors in the future;
3. At a disciplinary level, research on the relationship between the sustainable built environment and user perceptions has focused on 'public, environmental & occupational health', and since 2009, gradually moved towards 'environmental studies', 'construction & building technology', 'engineering, civil', and 'geography', with public health, environmental, and building engineering disciplines dominating the study, and a trend towards cross-disciplinarity that will lead to further integration in the future;
4. In terms of keyword co-occurrence, the keywords 'health', 'walking', 'environment', 'impact, 'behaviour', 'perception of behaviour', and 'design' are core for this topic. The hotspots of research in this field are often related to people's physical and mental health, which is highly compatible with the distribution of the subject areas. The keywords form a total of 12 sub-clusters—11 of which are under discussion—ranging from 'thermal comfort' and 'physical activity' to 'quality' and 'neighbourhood environment'. The keyword emergence and timeline spectrum suggest a gradual convergence of research hotspots towards urban-scale, emerging technologies such as 'neighbourhood design', 'city', 'virtual reality', and 'machine learning', progressively using new tools and methods of research.

Author contributions

YY: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing–review and editing. YH: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing–original draft. SL: Conceptualization, Data curation, Methodology, Software, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research has been supported by National Natural Science Foundation of China (No.52008128).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alter, D. A., and Eny, K. (2005). The relationship between the supply of fast-food chains and cardiovascular outcomes. *Can. J. Public Health* 96, 173–177. doi:10.1007/BF03403684

Altman, I., and Zube, E. H. (2012). *Public places and spaces*. Springer Science & Business Media.

Alvargonzález, D. (2011). Multidisciplinarity, interdisciplinarity, transdisciplinarity, and the sciences. *Int. Stud. philosophy Sci.* 25 (4), 387–403. doi:10.1080/02698595.2011.623366

Andrews, G. J., and Evans, J. (2008). Understanding the reproduction of health care: towards geographies in health care work. *Prog. Hum. Geogr.* 32 (6), 759–780. doi:10.1177/0309132508089826

Attwater, R., Booth, S., and Guthrie, A. (2005). The role of contestable concepts in transdisciplinary management of water in the landscape. *Syst. Res. Behav. Sci.* 22 (3), 185–192. The Official Journal of the International Federation for Systems Research. doi:10.1002/sres.668

Baran, P. K., Tabrizian, P., Zhai, Y., Smith, J. W., and Floyd, M. F. (2018). An exploratory study of perceived safety in a neighborhood park using immersive virtual environments. *Urban For. Urban Green.* 35, 72–81. doi:10.1016/j.ufug.2018.08.009

Barton, H. (2009). Land use planning and health and well-being. *Land use policy* 26, S115–S123. doi:10.1016/j.landusepol.2009.09.008

Bell, S. (2012). *Landscape: Pattern, perception and process*. Routledge.

Bird, S., Kurowski, W., Feldman, S., Browning, C., Lau, R., Radermacher, H., et al. (2009). The influence of the built environment and other factors on the physical activity of older women from different ethnic communities. *J. Women & Aging* 21 (1), 33–47. doi:10.1080/08952840802633669

Bruce, D. G., Devine, A., and Prince, R. L. (2002). Recreational physical activity levels in healthy older women: the importance of fear of falling. *J. Am. geriatrics Soc.* 50 (1), 84–89. doi:10.1046/j.1532-5415.2002.50012.x

Brug, J., van Lenthe, F. J., and Kremers, S. P. (2006). Revisiting kurt lewin: how to gain insight into environmental correlates of obesogenic behaviors. *Am. J. Prev. Med.* 31 (6), 525–529. doi:10.1016/j.amepre.2006.08.016

Burton, E. (2012). Streets ahead? The role of the built environment in healthy ageing. *Perspect. public health* 132 (4), 161–162. doi:10.1177/1757913912449570

Cao, X. J., and Ettema, D. F. (2014). Satisfaction with travel and residential self-selection: how do preferences moderate the impact of the hiawatha light rail transit line? *J. Transp. Land Use* 7 (3), 93–108. doi:10.5198/jtlu.v7i3.485

Carr, L. J., Dunsiger, S. I., and Marcus, B. H. (2010). Walk score™ as a global estimate of neighborhood walkability. *Am. J. Prev. Med.* 39 (5), 460–463. doi:10.1016/j.amepre.2010.07.007

Cerin, E., Conway, T. L., Cain, K. L., Kerr, J., De Bourdeaudhuij, I., Owen, N., et al. (2013). Sharing good NEWS across the world: developing comparable scores across 12 countries for the neighborhood environment walkability scale (NEWS). *BMC public health* 13 (1), 309–314. doi:10.1186/1471-2458-13-309

Chen, C., Ibekwe-SanJuan, F., and Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *J. Am. Soc. Inf. Sci. Technol.* 61 (7), 1386–1409. doi:10.1002/asi.21309

Chen, Y., Chen, C., Liu, Z., Hu, Z., Wang, X., Liu, Y., et al. (2015). Eight SNVs in NF-κB pathway genes and their different performances between subclinical mastitis and mixed Chinese Holstein cows. *Stud. Sci. Sci.* 33 (02), 242–249. doi:10.1016/j.gene.2014.11.011

Chen, Q., Fang, L., and Deng, Y. (2017). A study on neighborhood compound environment and walking behavior from thePerspective of heterogeneity:taking hangzhou as an example. *City Plan. Rev.* (9), 48–57. doi:10.11819/cpr20170907a

Chen, C., Tajnur, H., and Chen, Y. (2018). Built environment's influence on obesity of the older women and its planning responses. *Hum. Geogr.* 4, 76–81. doi:10.13959/j.issn.1003-2398.2018.04.011

Chen, C., Chen, J., Jia, Z., Deng, Z., Ma, S., Long, Y., et al. (2020). Large-scale measurement of the quality of urban street space based on physical disorder theory: A case study of area within the second ring of hefei city. *South Archit.* 25 (2), 10–18. doi:10.1186/s40001-020-00407-4

Chiou, S.-T., and Chen, L.-K. (2009). Towards age-friendly hospitals and health services. *Archives gerontology geriatrics* 49, S3–S6. doi:10.1016/S0167-4943(09)70004-4

Committee on Facilitating Interdisciplinary Research, National Academy of Sciences, National Academy of Engineering (1900). *Facilitating interdisciplinary research*. National Academies Press.

Cooper, C. L., and Marshall, J. (1976). Occupational sources of stress: A review of the literature relating to coronary heart disease and mental ill health. *J. Occup. Psychol.* 49 (1), 11–28. doi:10.1111/j.2044-8325.1976.tb00325.x

de Freitas, C. R., and Grigorieva, E. A. (2015). A comprehensive catalogue and classification of human thermal climate indices. *Int. J. biometeorology* 59, 109–120. doi:10.1007/s00484-014-0819-3

Diez Roux, A. V., and Mair, C. (2010). Neighborhoods and health. *Ann. N. Y. Acad. Sci.* 1186 (1), 125–145. doi:10.1111/j.1749-6632.2009.05333.x

Dong, H., and Qin, B. (2017). Exploring the link between neighborhood environment and mental wellbeing: A case study in Beijing, China. *Landsc. Urban Plan.* 164, 71–80. doi:10.1016/j.landurbplan.2017.04.005

Dong, W., Liu, Y., and Dong, Y. (2020). Measurement methods of urban residents' perception of built environment from a health perspective: A review. *Sci. Technol. Rev.* 38 (7), 61–68. doi:10.3981/j.issn.1000-7857

Edwards, D., and Burnard, P. (2003). A systematic review of stress and stress management interventions for mental health nurses. *J. Adv. Nurs.* 42 (2), 169–200. doi:10.1046/j.1365-2648.2003.02600.x

Evans, G. W., Wells, N. M., and Moch, A. (2003). Housing and mental health: A review of the evidence and a methodological and conceptual critique. *J. Soc. issues* 59 (3), 475–500. doi:10.1111/1540-4560.00074

Ewing, R., and Handy, S. (2009). Measuring the unmeasurable: urban design qualities related to walkability. *J. Urban Des.* 14 (1), 65–84. doi:10.1080/13574800802451155

Gee, G. C., and Payne-Sturges, D. C. (2004). Environmental health disparities: A framework integrating psychosocial and environmental concepts. *Environ. health Perspect.* 112 (17), 1645–1653. doi:10.1289/ehp.7074

Gold, J. R. (1980). *An introduction to behavioural geography*. USA: Oxford University Press.

Hong, J., and Chen, C. (2014). The role of the built environment on perceived safety from crime and walking: examining direct and indirect impacts. *Transportation* 41, 1171–1185. doi:10.1007/s11116-014-9535-4

Hur, M., Nasar, J. L., and Chun, B. (2010). Neighborhood satisfaction, physical and perceived naturalness and openness. *J. Environ. Psychol.* 30 (1), 52–59. doi:10.1016/j.jenvp.2009.05.005

Institute, I. W. B. (2014). *The well building Standard®*. Delos Living LLC.

Jf, S. (2006). Physical activity and the built environment, president's council on physical fitness and sports research digest. http://www.fitness.gov/publications/digests/december2006digest.pdf.18.

Jonasson, L. S., Nyberg, L., Kramer, A. F., Lundquist, A., Riklund, K., and Boraxbekk, C.-J. (2017). Aerobic exercise intervention, cognitive performance, and brain structure: results from the physical influences on brain in aging (PHIBRA) study. *Front. Aging Neurosci.* 8, 336. doi:10.3389/fnagi.2016.00336

Kearns, R., and Moon, G. (2002). From medical to health geography: novelty, place and theory after a decade of change. *Prog. Hum. Geogr.* 26 (5), 605–625. doi:10.1191/0309132502ph389oa

Kehrer, B. H., Andersen, R., and Glaser, W. A. (1972). A behavioral model of families' use of health services. *J. Hum. Resour.* 7 (1), 125. doi:10.2307/145064

Kerr, J., Emond, J. A., Badland, H., Reis, R., Sarmiento, O., Carlson, J., et al. (2016). Perceived neighborhood environmental attributes associated with walking and cycling for transport among adult residents of 17 cities in 12 countries: the IPEN study. *Environ. health Perspect.* 124 (3), 290–298. doi:10.1289/ehp.1409466

Kim, J. (2010). Neighborhood disadvantage and mental health: the role of neighborhood disorder and social relationships. *Soc. Sci. Res.* 39 (2), 260–271. doi:10.1016/j.ssresearch.2009.08.007

Kuo, F. E., Bacaicoa, M., and Sullivan, W. C. (1998). Transforming inner-city landscapes: trees, sense of safety, and preference. *Environ. Behav.* 30 (1), 28–59. doi:10.1177/0013916598301002

Lahti, J., Holstila, A., Lahelma, E., and Rahkonen, O. (2014). Leisure-time physical activity and all-cause mortality. *PLoS One* 9 (7), e101548. doi:10.1371/journal.pone.0101548

Lee, C., and Moudon, A. V. (2004). Physical activity and environment research in the health field: implications for urban and transportation planning practice and research. *J. Plan. literature* 19 (2), 147–181. doi:10.1177/0885412204267680

Li, F., Harmer, P., Cardinal, B. J., Bosworth, M., Johnson-Shelton, D., Moore, J. M., et al. (2009). Built environment and 1-year change in weight and waist circumference in middle-aged and older adults: portland neighborhood environment and health study. *Am. J. Epidemiol.* 169 (4), 401–408. doi:10.1093/aje/kwn398

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann. Intern. Med.* 151, W-65–W94. doi:10.7326/0003-4819-151-4-200908180-00136

Liu, B., and Guo, C. (2006). Promoting health through design: A brief ntroduction of"Active living byDesign",An American national program. *Urban Plan. Int.* 21 (2), 60–65. CNKI:SUN:GWCG.0.2006-02-012.

Liu-Ambrose, T., Khan, K. M., Eng, J. J., Janssen, P. A., Lord, S. R., and Mckay, H. A. (2004). Resistance and agility training reduce fall risk in women aged 75 to 85 with low bone mass: A 6-month randomized, controlled trial. *J. Am. Geriatrics Soc.* 52 (5), 657–665. doi:10.1111/j.1532-5415.2004.52200.x

Long, Y., and Tang, J. (2019). Large-scale quantitative measurement of the quality of urban street space: the research progress. *City Plan. Rev.* 43, 107–114. doi:10.11819/cpr20190616a

Long, Y. (2016). Street urbanism: A new perspective for urban studies and city planning in the new data environment. *Time+ Archit.* 2, 128–132. doi:10.13717/j.cnki.ta.2016.02.019

Lu, F., Tan, S., Wang, Y. H., Ma, H. D., Tang, W., Yang, J. B., et al. (2015). Built environment's influence on physical activity: review and thought. *Urban Plan. Int.* 30 (2), 62–72. CNKI:SUN:GWCG.0.2015-02-011. doi:10.1016/j.jaut.2015.05.008

Ma, M., Cai, Z., Michaelides, A., and Aeppli, G. (2016). Fast diffusion of water nanodroplets on graphene. *Chin. Landsc. Archit.* 32 (11), 66–71. CNKI:SUN:ZGYL.0.2016-11-014. doi:10.1038/nmat4449

Ma, M., zhou, J., Cai, z., Liu, J., Liang, Y., Zhang, C. D., et al. (2019). ZEB1-AS1 initiates a miRNA-mediated ceRNA network to facilitate gastric cancer progression. *J. Hum. Settlements West China* 34 (4), 27–34. doi:10.1186/s12935-019-0742-0

Maas, J., Van Dillen, S. M., Verheij, R. A., and Groenewegen, P. P. (2009). Social contacts as a possible mechanism behind the relation between green space and health. *Health & place* 15 (2), 586–595. doi:10.1016/j.healthplace.2008.09.006

Macmillan, A., Hosking, J., Connor, J., Bullen, C., and Ameratunga, S. (2013). A cochrane systematic review of the effectiveness of organisational travel plans: improving the evidence base for transport decisions. *Transp. policy* 29, 249–256. doi:10.1016/j.tranpol.2012.06.019

Merleau-Ponty, M. (2004). *The world of perception*. Routledge.

Monda, K. L., Gordon-Larsen, P., Stevens, J., and Popkin, B. M. (2007). China's transition: the effect of rapid urbanization on adult occupational physical activity. *Soc. Sci. Med.* 64 (4), 858–870. doi:10.1016/j.socscimed.2006.10.019

Morland, K., Roux, A. V. D., and Wing, S. (2006). Supermarkets, other food stores, and obesity: the atherosclerosis risk in communities study. *Am. J. Prev. Med.* 30 (4), 333–339. doi:10.1016/j.amepre.2005.11.003

Nagel, C. L., Carlson, N. E., Bosworth, M., and Michael, Y. L. (2008). The relation between neighborhood built environment and walking activity among older adults. *Am. J. Epidemiol.* 168 (4), 461–468. doi:10.1093/aje/kwn158

Nielsen, T. S., and Hansen, K. B. (2007). Do green areas affect health? Results from a Danish survey on the use of green areas and health indicators. *Health & place* 13 (4), 839–850. doi:10.1016/j.healthplace.2007.02.001

Petts, J., Owens, S., and Bulkeley, H. (2008). Crossing boundaries: interdisciplinarity in the context of urban environments. *Geoforum* 39 (2), 593–601. doi:10.1016/j.geoforum.2006.02.008

Qin, B., Zhu, W., and Dong, H. (2018). The influence of neighborhood environment and transportation on residents' mental wellbeing: based on questionnaire of 16 communities in beijing. *Urban Rural. Plan.* 3, 38–46. doi:10.12049/j.urp.201803005

Rantanen, T., Portegijs, E., Viljanen, A., Eronen, J., Saajanaho, M., Tsai, L.-T., et al. (2012). Individual and environmental factors underlying life space of older people–study protocol and design of a cohort study on life-space mobility in old age (LISPE). *BMC Public Health* 12, 1–17. doi:10.1186/1471-2458-12-1018

Reyers, B., Roux, D. J., Cowling, R. M., Ginsburg, A. E., Nel, J. L., and Farrell, P. O. (2010). Conservation planning as a transdisciplinary process. *Conserv. Biol.* 24 (4), 957–965. doi:10.1111/j.1523-1739.2010.01497.x

Rockhill, B., Willett, W. C., Manson, J. E., Leitzmann, M. F., Stampfer, M. J., Hunter, D. J., et al. (2001). Physical activity and mortality: A prospective study among women. *Am. J. Public Health* 91 (4), 578–583. doi:10.2105/ajph.91.4.578

Rodríguez, D. A., Evenson, K. R., Roux, A. V. D., and Brines, S. J. (2009). Land use, residential density, and walking: the multi-ethnic study of atherosclerosis. *Am. J. Prev. Med.* 37 (5), 397–404. doi:10.1016/j.amepre.2009.07.008

Sallis, J. F., Cervero, R. B., Ascher, W., Henderson, K. A., Kraft, M. K., and Kerr, J. (2006). An ecological approach to creating active living communities. *Annu. Rev. Public Health* 27, 297–322. doi:10.1146/annurev.publhealth.27.021405.102100

Sallis, J. F., Carlson, J. A., Mignano, A. M., Lemes, A., and Wagner, N. (2013). Trends in presentations of environmental and policy studies related to physical activity, nutrition, and obesity at society of behavioral medicine, 1995–2010: A commentary to accompany the active living research supplement to annals of behavioral medicine. *Ann. Behav. Med.* 45 (1), S14–S17. doi:10.1007/s12160-012-9415-0

Skalicky, V., and Čerpes, I. (2019). Comprehensive assessment methodology for liveable residential environment. *Cities* 94, 44–54. doi:10.1016/j.cities.2019.05.020

Tang, J., and Long, Y. (2017). Metropolitan street space quality evaluation: second and third ring of Beijing, inner ring of shanghai. *Planners* 33 (2), 68–73. doi:10.3969/j.issn.1006-0022.2017.02.011

Tinetti, M. E., and Kumar, C. (2010). The patient who falls:"It's always a trade-off". *Jama* 303 (3), 258–266. doi:10.1001/jama.2009.2024

Trapp, G. S., Giles-Corti, B., Christian, H. E., Bulsara, M., Timperio, A. F., McCormack, G. R., et al. (2012). Increasing children's physical activity: individual, social, and environmental factors associated with walking to and from school. *Health Educ. Behav.* 39 (2), 172–182. doi:10.1177/1090198111423272

Tress, B., Tress, G., and Fry, G. (2005). Researchers' experiences, positive and negative, in integrative landscape projects. *Environ. Manag.* 36, 792–807. doi:10.1007/s00267-005-0038-0

Van Cauwenberg, J., Van Holle, V., Simons, D., Deridder, R., Clarys, P., Goubert, L., et al. (2012). Environmental factors influencing older adults' walking for transportation: A study using walk-along interviews. *Int. J. Behav. Nutr. Phys. activity* 9, 85–11. doi:10.1186/1479-5868-9-85

Van den Berg, A. E., Maas, J., Verheij, R. A., and Groenewegen, P. P. (2010). Green space as a buffer between stressful life events and health. *Soc. Sci. Med.* 70 (8), 1203–1210. doi:10.1016/j.socscimed.2010.01.002

Wallace, B., Barnes, J., and Rutherford, G. S. (2000). Evaluating the effects of traveler and trip characteristics on trip chaining, with implications for transportation demand management strategies. *Transp. Res. Rec.* 1718 (1), 97–106. doi:10.3141/1718-13

Wang, L., Sun, W., and Wu, Y. (2020a). Study on the impact of subjective perception of urban environment on residents' health: based on a large sample survey of 60 counties and cities in China. *Hum. Geogr.* 35 (2), 55–64. doi:10.13959/j.issn.1003-2398.2020.02.007

Wang, R., Niu, Q., Gao, H., Huang, S., and Xu, Z. (2020b). Quantitative research and evaluation on street lnterface of historic district A case of sanhe ancient town in hefei. *Urbanism Archit.* 17 (28), 15–19+56. doi:10.19892/j.cnki.csjz.2020.28.03

Wang, M., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., et al. (2018). Research on community ldentity,Emotional structure of the Environment,and urban morphogenetic mechanism:based on the measurement and evaluation of caoyang xincun in shanghai. *City Plan. Rev.* 42 (12), 43–49. doi:10.1038/s41586-018-0063-9

Won, S., Lee, Y., and Park, Y. (2020). The impact of signboard-building color combinations on color harmony and legibility. *Color Res. Appl.* 45 (3), 527–541. doi:10.1002/col.22480

Wu, J. (2021). Three forms of multidisciplinary research. *Bull. Natl. Nat. Sci. Found. China* 35 (2), 5. doi:10.16262/j.cnki.1000-8217.2021.02.002

Xi, T., Kuang, X., Zhu, Y., and Fu, X. (2019). An exploration of the street renewal design based on human perception:A case study of the beautiful district renovation program in pengpu Town,Jing'an District,Shanghai. *Urban Plan. Forum.* doi:10.16361/j.upf.201907020

Xian Yu, J. C., Juan, Z. C., Gao, L. J., Ni, A. N., Zhang, W., and Wu, B. (2011). Empirical analysis of commuters' nonwork stop-making behavior in Beijing, China. *J. Transp. Eng.* 137 (5), 360–369. doi:10.1061/(ASCE)TE.1943-5436.0000228

Yang, J., and French, S. (2013). The travel—obesity connection: discerning the impacts of commuting trips with the perspective of individual energy expenditure and time use. *Environ. Plan. B Plan. Des.* 40 (4), 617–629. doi:10.1068/b38076

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., et al. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landsc. Urban Plan.* 180, 148–160. doi:10.1016/j.landurbplan.2018.08.020

Zhong, T., Ye, C., Wang, Z., Tang, G., Zhang, W., and Ye, Y. (2021). City-scale mapping of urban façade color using street-view imagery. *Remote Sens.* 13 (8), 1591. doi:10.3390/rs13081591

Zhu, W., and Zhai, B. (2019). Lnfluence of urban built environment cognition on children's outdoor activities. *City Plan. Rev.* 43 (12), 45–54. doi:10.11819/cpr20191207a

frontiers | Frontiers in Built Environment

TYPE Original Research
PUBLISHED 29 January 2024
DOI 10.3389/fbuil.2023.1300843

Check for updates

OPEN ACCESS

EDITED BY
Yongkui Li,
Tongji University, China

REVIEWED BY
Hasim Altan,
Prince Mohammad bin Fahd University, Saudi Arabia
Sheila Walbe Ornstein,
University of São Paulo, Brazil

*CORRESPONDENCE
Pirouz Nourian,
p.nourian@utwente.nl

RECEIVED 23 September 2023
ACCEPTED 22 December 2023
PUBLISHED 29 January 2024

CITATION
Azadi S, Bai N and Nourian P (2024), Ergonomics of spatial configurations: a voxel-based modelling framework for accessibility and visibility simulations.
Front. Built Environ. 9:1300843.
doi: 10.3389/fbuil.2023.1300843

Ergonomics of spatial configurations: a voxel-based modelling framework for accessibility and visibility simulations

Shervin Azadi [1], Nan Bai [2] and Pirouz Nourian [3]*

[1]Eindhoven University of Technology, Eindhoven, Netherlands, [2]Delft University of Technology, Delft, Netherlands, [3]University of Twente, Enschede, Netherlands

How can we assess the ergonomic comfort of a sizeable spatial configuration such as the indoor space of a complex building or an urban landscape when we design, plan, and manage the space? Is there a fundamental difference between indoor [architectural] spatial configurations and outdoor [urban] spatial configurations with respect to ergonomics? Can we have a unified approach to the computational study of spatial ergonomics? This paper addresses these fundamental questions while providing a brief taxonomic review of the scholarly literature on these matters from a mathematical point of view, including a brief introduction to the modelling-based approaches to the computational ways of studying the fundamental effects of spatial configuration on human behaviours. Furthermore, the paper proposes a computational approach for ergonomic assessment of spatial configurations that explicitly allows for combined accessibility and visibility analyses in the built environment. The gist of this approach is the conceptualisation of spatial configurations as rasterised (voxelated) 2D manifold walkable terrains whose voxels have 3D vistas, unifying the simulations and analyses of accessibility and visibility. The paper elaborates on how such a representation of space can provide for conducting various sorts of computational queries, analyses, and simulation experiments for research in spatial ergonomics. The paper concludes with a mapping of the computational modelling approaches pertinent to the study and assessment of spatial ergonomics; and marks avenues of future research on various categories of exploratory, generative, and associative models for ex-ante and ex-post assessment of ergonomic matters at spatial scales.

KEYWORDS

spatial ergonomics, accessibility, visibility, architectural morphology, urban morphology, simulation models

1 Introduction

The importance of ergonomics and human factors in spatial design and interventions can be understood, unfortunately, by examination of the disastrous results of numerous contemporary building developments and interventions in cities around the planet that disregard ergonomics and human factors. Such interventions have resulted in the so-called ghost cities (Jin et al., 2017; DeLyser, 1999), dangerous crime neighbourhoods (Angel, 1968; Newman, 1996; Cozens et al., 2005; Jacobs, 2016), and depressive social environments

(Wang R. et al., 2019; Galea et al., 2005). Systematically studying ergonomics and human factors based on the spatial configuration of the built environment can be beneficial for three primary purposes, in line with the three main activities of design processes as defined by Lawson (Lawson, 2006), all of which need to address fundamental questions regarding the relationship between environment and human behaviours:

- **Evaluation**: quality assurance of built environment, governance, standardization, certification through ex-post assessment of safety, operational efficiency, occupational comfort, and security;
- **Analysis**: spatial, social, behavioural sciences, and environmental psychology;
- **Synthesis**: scientific and evidence-based design methodology, design didactics, spatial-decision-support, and design research.

One of the pioneers of ergonomics research, Lillian Moller Gilbreth, formulated some of the oldest spatial questions in ergonomics about the *comfort* of kitchens and assembly lines for work and the relations between [spatial] movement, work efficiency, fatigue, and comfort (Perloff and Naman, 2012; Gilbreth, 1914). While the scale of a kitchen or the functional space for a stationary worker in an assembly line is too small to have non-trivial navigation challenges, the repetition of the work at scale (such as in hospitals or subway complexes) raises the importance of comfort of frequent human movements.

In an indoor environment, it is easy to observe that the configuration or layout of the environment can directly affect the spatial ergonomics of the building. However, the counterpart of configuration in urban land-use distributions, i.e., the functionalities assigned to spaces, has a rather subtle and yet substantial influence on the spatial ergonomics on the urban scale. The so-called *cost of transport* can show, for instance, that the *expected walking travel time* is dependent on the spatial configuration since the assignment of different functionalities to spaces is what initially motivates the movement of people between two spaces (q.v. (Nourian et al., 2021; Çubukçuoğlu et al., 2021)).

Notwithstanding the importance of morphological factors such as heights of ceilings, proportions of rooms, and alike, this article only addresses the ergonomic questions with regards to the syntactic relations between spaces rather than the morphological attributes of the spaces in isolation. We will address the discussion on the complex relationship between spatial configuration and spatial ergonomics, focusing on the potential comfort of pedestrians during way-finding and walking through the built environment. Our perspective is based on the assumption that accessibility and visibility are the main factors that are not only directly related to the holistic notion of human comfort and ergonomics in space, but also precisely in the purview of architectural designers and spatial decision-makers. This is because both of them are directly influenced by the design decisions concerning the configuration and shape of built environments.

Our proposed spatial modelling framework provides for combined accessibility and visibility simulations that can show the potential ergonomic comfort or discomfort of an environment with respect to walkability. This, however, does not mean that we are proposing a framework for simulations of actual walking behaviour. The framework, rather, paves the way for spatial ergonomics researchers to build models, test hypotheses, and collate their results in a standardised framework that is suitable for various approaches in modelling, ranging from mathematical models to models based on artificial intelligence. The key connective element, and thus the main advantage of the proposed framework, will be that it introduces *graphs* as natural representations of spaces, whether for accessibility analyses or visibility analyses. Furthermore, the fact that both kinds of graphs can be constructed out of the same superset of nodes (corresponding to voxels) means that the results of analyses obtained from the two types of studies can be easily collated and combined as desired. Moreover, the generality of the proposed approach is such that the distinction between indoor and outdoor environments will be rendered unnecessary, hence providing for a unified view towards analyses of architectural and urban configurations. Therefore, arguably, this approach opens up new avenues for rigorous research in spatial ergonomics.

2 Theoretical context

2.1 Design science, science of artificial, and generative science

A substantial part of Spatial Ergonomics as an interdisciplinary field of research is concerned with understanding the complex relations of pedestrians' behaviour within complex spatial environments. Such systematic understandings could inform and underpin design activities by introducing a scientific approach to evaluating the impact of spatial designs and interventions on human comfort. Moreover, the paper argues that a knowledge-based approach for informing decision-making in design activities is necessary for guaranteeing human qualities usually referred to as the live-ability of buildings and cities. In this sense, Spatial Ergonomics can be seen as a design science as called for by H.A. Simon in (Simon, 1970). This is related to, yet different from the studies of environmental psychology and behaviour, which attempts to inductively understand the bi-directional relation of humans and their environment from a cognitive and physiological point of view (Golledge et al., 1993; McAndrew, 1993; Golledge, 1999). However, we also argue that the *Problem-Oriented* approach of environmental psychology (Proshansky, 1978) and the *Design-Oriented* approach of spatial ergonomics are approaching the essence of such a relation from a different angle.

This difference in approach allows us to consider spatial ergonomics as a subject matter of the Sciences of Artificial (Simon, 1970), i.e., the decision sciences aimed at the study of practices to improve the current state of systems towards better ones (non-verbatim, ibid). The main goal in the sciences of artificial is to extend the natural sciences by going beyond understanding phenomena and discovering the relation of human choices and their consequences, (Schelling, 1984). Within the scale of the city, understanding the complex relation of the urban environment and human behaviour requires explicit simulation models (Batty, 2007). To illustrate the specifics of scientific studies and their application in spatial ergonomics research, we find it important to distinguish three types of models as to their purposes:

- **Exploratory Models** are also known as the exploratory model analysis (EMA). These sorts of models do not aim to validate or refute a specific hypothesis, nor do they intend to predict a quantity precisely. Rather, they focus on the uncertainty in the open systems, exploring the implication of varied assumptions and hypotheses based on a series of computational experiments (Bankes, 1993; Agusdinata, 2008). Exploratory models benefit greatly from improved computational power in recent years, providing new knowledge for understanding the phenomena or system behaviours through the information gained from many computational experiments, acting as the base for further hypothesis building. Exploratory models could be driven by data, questions, or even models (Bankes, 1993). Network centrality analyses inspired by Social Network Analysis research are typical applications of exploratory models, as often seen in Space Syntax research (various exemplary reference provided further on).
- **Associative Models** are statistical models or neural networks that can be fitted/trained using data describing the dependent and independent variables. Such dependent variables can be numerical or categorical, making the problem either regression or classification. A predefined (classical) statistical model (such as multi-linear/logistic regression or a gravity model) can be fitted to the independent inputs to predict the dependant outputs of interest. Similarly, Artificial Neural Networks can be trained to find non-trivial (most likely non-linear) relations between independent input and dependent output variables to capture a complex association (see e.g., (Wang Z. et al., 2019)).
- **Generative Models** are also known as simulation models, which are *mechanism replicators* made based on a mathematical or computational understanding formalised in a theory or a simulation model capable of explaining "how things work." Archetypal examples include an understanding or assumption of the local behaviour of a spatial agent (a virtual pedestrian) based on its surrounding environment and other agents in its vicinity; or an Agent-Based Model revealing the emergence of behavioural patterns from predefined situated patterns of behaviour (Crooks et al., 2015). Another example of such models is the simulation of mobility/accessibility through Random Walks (Nourian, 2016). Such models directly link to the idea of ergonomics and the so-called First Law of Geography, as formulated by Waldo Tobler: Everything is connected to everything else, but near things are more connected than distant things, (Tobler, 1970). Thus the easier it is to get somewhere, the higher the probability of going there: measuring distance in time, through stochastic paths (random walks) (Nourian, 2016).

To be able to understand and trace the effect of the spatial form (i.e., the geometry and topology of a built environment) on behavioural patterns, we need to move toward models that can explain and justify such complex relations. We should start from the fact that relationships between form factors and human behaviours are fundamentally complex not only on an individual/psychological level but also on a collective/social level and that they need to be studied through simulation models. If our only goal in studying ergonomics were to predict comfort or discomfort, then it would suffice to study the association of some spatial factors with various comfort criteria describing human perception or human behaviour within built environments. However, if we are to understand and explain the mechanisms underlying these [complex] associations, then we need to study the phenomena directly through simulation models (Epstein and Axtell, 1996; Epstein, 2008). Therefore, our perspective in spatial ergonomics is also to promote the notion of Generative Science which is the pursuit of explainable and reproducible knowledge through devising and experimenting with complex simulation models; introduced by Joshua Epstein (Epstein, 2012; Epstein, 1999). Thus, we suggest that we need to synthesise simulation models that can replicate the complex relations of causes/stimuli and effects/behaviours to understand the relation of the design decisions on the shape and configuration of built environments (choices) and social behaviours and human perceptions (consequences).

The relations between the form of the environment and human behaviour (e.g., movement) can be identified to have at least two types of complexities, one being the network complexity of navigable/walkable spatial environments and their effects on human behaviour; the other being the process complexity of the influences of the other actors on the behaviour of individual actors in an environment. In the face of such complexities, if the aim is to understand the mechanisms resulting in the emergence of specific behavioural patterns, it will be necessary to utilise generative simulation models that are capable of replicating the mechanisms of stimuli and responses and their inter-relations of multiple actors within a complex spatial environment (Epstein, 2012). Examples of such models are stochastic simulation models, Agent-Based Models, Cellular Automata, Game Theoretical Models, or network-based models such as Markov Chains (a.k.a. Random Walks). Furthermore, the simulation-driven approach has been widely applied to acquire insights and incorporate the complexity of the un-modelled uncertainties in the study of environment-agent interactions (DeLaurentis and Callaway, 2004; Macal and North, 2009; Heppenstall et al., 2011); a particular example of which is to assess how visibility influences path-finding (Gath-Morad et al., 2021).

However, developing such simulation models will introduce some complexities, as it is inevitable to step away from the principle of *Ceteris Paribus* (Schlicht, 2012). Considering this inherent complexity of the relations between the inputs and outputs of such morphology-behaviour models, we can distinguish the application of each category of models in the study of these inter-relations from two directions:

- **Ex-Ante Assessment** of environmental design decisions: given a design as a proposed new spatial configuration or intervention in an existing spatial configuration, ex-ante assessments focus on what will be the projected behaviours induced by the spatial configuration? Such questions can be best answered by building exploratory or generative models. These models can be validated and calibrated using exemplary data from the past.
- **Ex-Post Assessment** of environmental design decisions: given a building or a built environment configuration from the past and observed behavioural data (e.g., of pedestrian movement), ex-post assessments focus on how can the correlation between

spatial configuration and human behaviour be made explicit through statistical models or neural networks? Considering the complexity of these associations, such questions can be best answered by building associative models through model-fitting or model-training using exemplary data (evidence) from the past.

Considering that all such models require behavioural data and spatial configurations from the past to be either trained, validated, or calibrated, it is clear that for moving towards an "evidence-based design" discipline, not only data needs to be gathered from the past, but it should also be structured properly by explicitly considering such analytic prospects. One key challenge or question of importance will be about the representation of a *spatial configuration*. It is important to have such a representation of a building or a built environment to assess and further analyse on, otherwise the notion of assessment becomes ill-defined.

2.2 Spatial morphology and space syntax

The terms morphology and syntax are historically borrowed from Linguistics, respectively referring to the study of the shape of the words in isolation and the grammatical structures connecting them. However, in the contemporary academic circles of architectural and urban studies, the term *morphological* colloquially refers to both morphological and syntactic studies on the configuration of space in built environments (Hillier, 1998; Marshall, 2004) and its relation to human behaviour; especially focusing on *pedestrian movement in space* (Hillier and Hanson, 1989; Hillier, 1999). Unequivocally, we can define spatial configurations as distributions of human activities (land-uses in geospatial scales and designated spatial functions in architectural scales) in non-trivial/non-Euclidean network spaces. In this sense, provided a network abstraction of the underlying space is available, a particular spatial configuration would be a coloured/labelled graph as an abstraction of the space in question.

Multiple scholars have proposed different lenses for the study of spatial morphology and, in particular, the "spatial configuration". Conzen offered a categorisation of different elements of the urban landscape based on the time scale of their change: 1) the city layout, 2) building form, and 3) land use. Each of these elements represents different degrees of changes, i.e., land use changing most often, buildings' physical form lasting longer, and the city layouts being the most permanent elements (Conzen, 1960). Vance ties the changes in the urban environment to human activity and claims that they can not be reduced to alteration of the physical structure (Vance, 1977). Frey relates the emergence of cities to the interaction of navigation, communication, and land values (Frey, 1999). Despite different formulations, what most of these lenses have in common is the emphasis on the relationship between the form of urban environments and the social structures that emerge within them (Stephenson, 2008); this perspective is widely known in the scientific literature as the morphological approach that is mainly concerned with the study of spatial form that can be defined as the geometry and topology (also referred to as the configuration or layout) of an environment and their effects on human behaviour. Based on this relationship, an urban form can be evaluated via activity patterns (primarily through the so-called pedestrian or natural movement (Hillier et al., 1993)), particularly human congregation that emerges within it, e.g., some studies have used activity/presence/movement density, and intensity as proxies for urban functionality (Chiu et al., 2020).

Hillier (Hillier and Hanson, 1989; Hillier et al., 1996) introduced the clear framework of Space Syntax to structure, measure, and understand the relation between the form of spatial structures and people's behaviour within them. Space syntax offers an exploratory methodology to investigate the urban spatial structure and uncover the spatial qualities of urban areas and their correlations with various network centrality indicators as *proxies* for spatial activities related to pedestrian movement through spaces. These methods aim to quantify the configurational properties of the built environment through the study of the structural properties of spatial networks, mainly through network centrality indicators. At the architectural scale, syntactic measures have been shown to correlate with behavioural patterns in a variety of settings, including malls (Okamoto et al., 2014; Omer and Goldblatt, 2016; Koohsari et al., 2013); subway stations (van der Hoeven and van Nes, 2014), museums (Turner and Penn, 1999), and hospitals (Haq and Luo, 2012). At an urban scale, these centrality indicators (such as integration/closeness or choice/betweenness) have been investigated based on their potential for revealing various relations between urban structural elements and human activity patterns (Peponis et al., 1997a); the influence of spatial structure (as a network) on staying behaviour at certain Points of Interests (referred to as *movement to* spaces in the Space Syntax jargon), as well as the correlation of the frequency of pedestrian *movement through* spaces with the presence of socio-economic activities (such as retail).

While urban form (Pacione, 2009) is assumed to be the conductor of the pedestrian movement patterns (Hillier et al., 1993), the effect of land-use or the distribution of designated functions in urban or architectural scales cannot be ignored in shaping the pedestrian movement patterns; this is because the reasons people walk in an environment have to do with their social and economic needs (Algers et al., 2005; Meyer et al., 2001). However, human movement is arguably a focal point of attention for ergonomics research at a spatial level, especially because it is directly affected by the design decisions on the shape and layout of an environment. Consequently, these decisions on a larger scale can function as an attractor or a driving force of developments (Johnson and Fisher, 2013).

2.3 Visibility and accessibility analysis

A comprehensive understanding of the relation of the behavioural patterns and built environments (either in architectural scale or urban scale) pertains to two main ergonomic aspects: Accessibility and Visibility in space.

Within Hillier's approach, the relation between urban structural elements and human activity patterns is not limited to the accessibility provided by the street network; It also includes concepts to denote the visual relations of spatial elements such as axial graph (Bentley, 1985; Moughtin, 2007; De Smith et al., 2018) and Turner-Penn's topological model of space is derived from visibility graph (Turner, 2001; Penn, 2003). This model can be constructed, e.g., based on the intersections of visibility polygons (a.k.a. Isovists (Benedikt, 1979)) to quantify the inter-visibility and proximity of spatial elements (Turner, 2003; Turner et al., 2001). Penn argues that this approach (e.g., the use of axial lines of sight) contributes significantly to the understanding of environmental cognition by putting forward a topological space representing our cognitive model of space (Penn, 2003).

In his seminal book, "The Image of the City" (Lynch and for Urban Studies, 1960), Lynch argues that legibility facilitates the perception and understanding of the spatial structure, and consequently, there is a correlation of the sense of identity in the urban space and legibility. He specifies three elements that comprise legibility: clarity, visibility, and coherence (Lynch and for Urban Studies, 1960). Legibility is influential in way-finding (Weisman, 1981; Koseoglu and Onder, 2011), and this is not only in perceiving the space and constructing a cognitive map from it but also in utilising the cognitive map in way-finding (Belir and Onder, 2013). The two most important factors affecting legibility are *simplicity of spatial configuration* (2D) and *saliency of landmarks* (3D), which respectively correspond to accessibility and visibility of urban spaces (Lynch and for Urban Studies, 1960). Lynch ties the construction of the cognitive map of a city to five elements: nodes (2D), paths (2D), districts (2D), landmarks (3D), and edges (3D) (Lynch and for Urban Studies, 1960). He maintains that the 2D elements function as anchors of locations. Thus their role in the cognitive map of the user is topological, whereas the 3D elements, on the other hand, provide a sense of distance and direction for the user. Thus their role is more geometrical (Dalton et al., 2003). On a tangent line and from a descriptive point of view, O'Neill ties the representation of cognitive map to the concept of legibility by defining it as the potential of a space to facilitate the formation of a successful cognitive map (O'Neill, 1991).

People experience and understand space primarily through vision, mapping the relation of the visible features of the space, and identifying the navigation opportunities (Li et al., 2016). Landmarks or salient objects in particular play an important role in wayfinding (Daniel and Denis, 1998; Denis et al., 1999; Lovelace et al., 1999; Schroder et al., 2011). Sorrows et al. attribute the distinctiveness of landmarks to their visual (colour, shape, etc.), semantic (function, type, etc.), and structural salience (location as to other elements) (Sorrows et al., 1999). Golledge (Golledge et al., 1993) puts a relative perspective on the matter and ties the quality of being a landmark to the distinctiveness of an element compared to other elements in a local or global neighbourhood, a relative view towards salience that is fundamental to our decision-support perspective towards the computational study of ergonomics. Although characterisation and detection of potential elements as landmarks can be done without topological models (see an example in (Nothegger et al., 2004; Steiniger et al., 2008)), structural salience depends on the position of a landmark along a route (Klippel et al., 2005; Elias et al., 2023), thus requiring an implicit or explicit topological model. Specifically, visibility is proved to be influential for way-finding and navigation in complex space, even when such as space is a multi-level uncertain environment that has vertical dimension (Gath-Morad et al., 2021).

Although it is indisputable that the local and global structural properties of an urban environment (or the indoor space of a building) influence the cognitive representations of it (FREUNDSCHUH and EGENHOFER, 1997), the fact that legibility is a more subjective measure of the space rather than an objective property of it, has led to the complication of understanding of humans' spatial perception processes in built environments (Zmudzinska-Nowak, 2003). To untangle this subjective-objective dichotomy, a few attempts have been oriented towards comparing the quantitative representation of a given built environment with humans' perception of it. (Stamps, 2005; Gjerde, 2011; Long and Baran, 2011). In particular, the correlation of legibility (subjective) and intelligibility (objective) (Long and Baran, 2011), and modelling permeability using Isovists (Stamps, 2005). It is worth noting that there is a parallel field of research focused on way-finding and path-planning for robots, called Robot Motion Planning (cf. (Patle et al., 2019)) that is especially relevant to the discussion because of its history in developing environment maps as accessibility and visibility graphs for navigation.

Alternative models for studying accessibility from a cognitive/psychological point of view (and thus in relation to visibility) have the potential to take cognitive factors into account, e.g., by considering the simplicity or ease of navigation in path-finding (Turner, 2001; Duckham and Kulik, 2003; Nourian et al., 2015), or by utilising Random-Walk, Markov-Chains, or Spectral Models in analysing accessibility from a stochastic perspective as opposed to the inherent deterministic perspective associated with using centrality indicators based on geodesics or shortest paths (Nourian, 2016; Blanchard and Volchenkov, 2008; Nourian et al., 2016a). Moreover, the notions of accessibility and visibility lead to an interesting phenomenon: somewhere is considered accessible when it is *eventually* accessible under some conditions (a maximum time or distance before reaching the place), but somewhere is considered visible only if it is immediately *visible*. This semantic difference also becomes a challenge for unifying the analyses of accessibility and visibility.

In summary, studying accessibility from a cognitive and ergonomic point of view would eventually require studying visibility, especially for assessing the legibility of an environment for way-finding. This entails a need for a mathematical model of the relation between accessibility and visibility that can support exploratory and associative approaches, as well as generative simulations.

3 Proposed framework

3.1 Intertwined geometrical and topological models of space

Purely topological models of space are arguably inspired by the social network models studied in computational social science since the 1950/60s (Nourian, 2016; Bonacich, 1987). They are often considered for studying the configurational properties of the urban space (e.g., (Hillier, 1998)). Due to the historical prevalence of [urban] spatial network models consisting of lines (axial lines of sight in Space Syntax-oriented models or the street centre line in Transport Planning oriented models), much attention has been paid to the adequate representation of line-based networks for morphological studies. In retrospect, it can be seen that the orientation towards line-based networks has been mainly a pragmatic choice for keeping computational costs low. Multiple

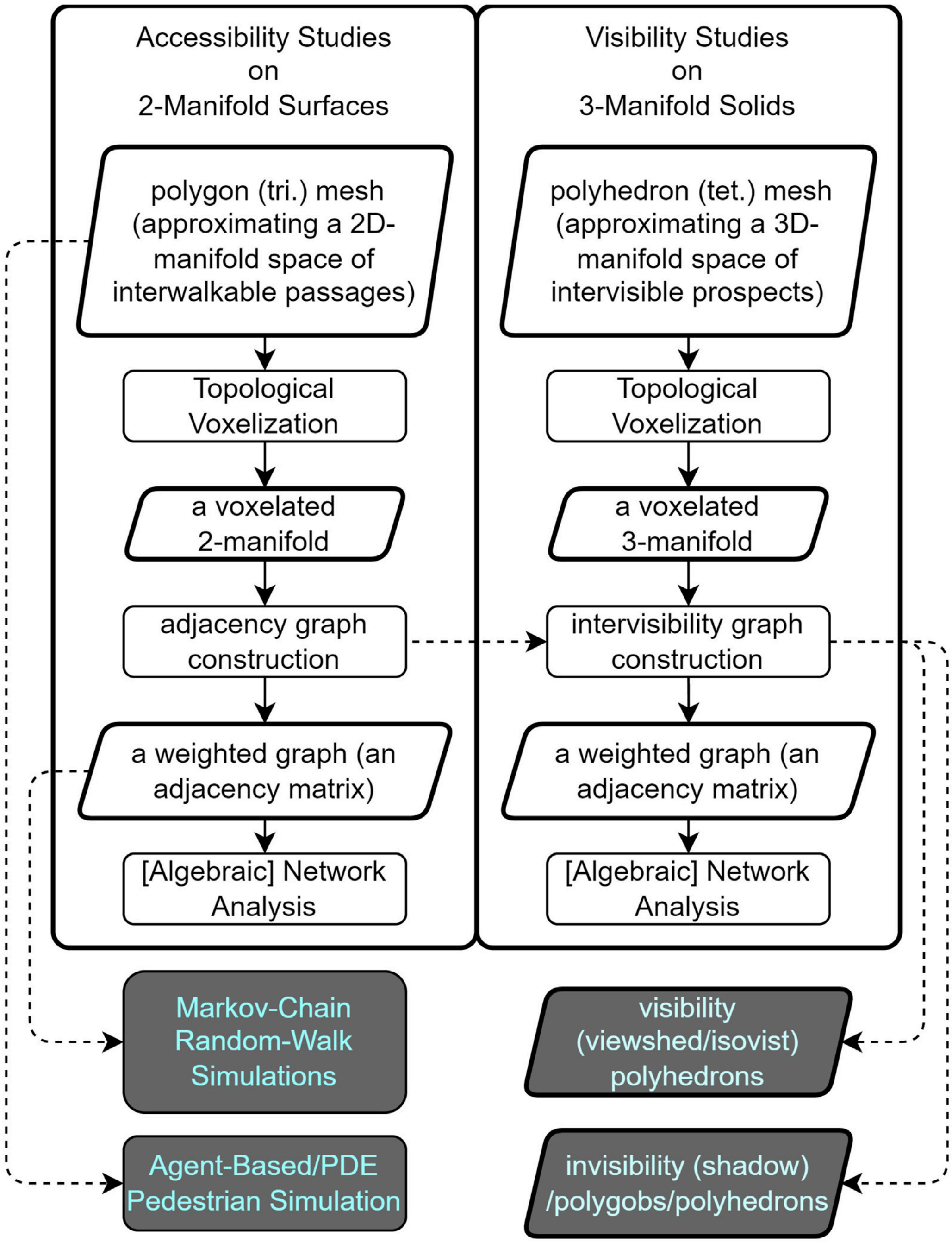


FIGURE 1
A schematic view of the proposed raster-based analytic workflows.

critiques of this approach have pointed to either the counter-examples for spaces that cannot be reduced to a line network or the failure of topological models in integrating the metric properties (Turner et al., 2001; Peponis et al., 1997b; Batty, 2001; Ratti, 2004a; Ratti, 2004b; Montello, 2007; Kostakos et al., 2010; Netto, 2015; Pafka et al., 2018). However, without dwelling much on the futile debates on the advantages or disadvantages of geometric or topological models of space, given that topological and geometric measures are not independent, we argue that models need to be constructed based on a close relation of geometry and topology rather than a supplementary attribution of metric weights.

We argue that the reduction of walkable or visible space into a few lines is overly simplistic and problematic. Conventionally, i.e., in Space Syntax studies, both accessibility and visibility are modelled through topological models made up of "lines of sight" dubbed Axial Lines (Penn, 2003). However, even apart from the reported problems in defining such axial-line networks ((Ratti, 2004b)), the connection (similarities and differences) between the two notions could be more explicit. Mathematically speaking, a walkable surface is essentially a 2D-Manifold and a visible prospect is fundamentally a 3D-Manifold. Studying accessibility requires working with geodesic/optimal-path algorithms on [1D]

TABLE 1 Glossary.

Term	Definition
Voxel	A volumetric 3D pixel, a voxel is a small box-like spatial unit, i.e., a bounding box in a spatial domain as a part of a 3D regular and orthogonal tiling of space that is indexed with three [integer] coordinates in the form of (i, j, k). A voxelated domain practically becomes a 3D raster (akin to a bitmap) image
Manifold	A spatial domain that is possibly not linear (Euclidean) in the global picture, but everywhere locally similar to a Euclidean space of a low dimension like a 1D, 2D, or 3D Euclidean space (a line, a plane, or a hyperplane). A k-Manifold is a space that is locally homeomorphic (topologically similar) to a Euclidean space of dimension k, i.e., $\mathbb{R}^k$
Topological Voxelation	A process for obtaining a voxelated representation of a continuous spatial domain such as a curve, a surface, or a solid with a minimal set of voxels that are necessary and sufficient to preserve the topological properties of the original 'smooth' geometries in the raster domain, such as connectedness, closure, number of genera (holes) and alike
Graph	An ordered pair of Vertices and Edges (also referred to as Nodes and Links), typically encoding relations of different sorts such as inter-connectivity or inter-visibility, either as binary data or weighted float variables: $G = (V, E)$, $V = v_i \| i \in [0, n)$, $E \subset V \times V$, the connectivity information of a graph can also be represented as a [sparse] matrix
Neighbourhood	A function showing the proximity, similarity, or connectivity of the data points/vertices/nodes in space as to which a graph can be constructed, possibly with a binary or real range. The outputs of such functions are used for constructing [possibly weighted] graphs

line-networks or [as we propose] on [2D] surface-meshes representing 2D manifold-surfaces of passages, possibly with many holes (mathematically known as surfaces of high genus/genera [holes]) and that studying visibility requires working with intersection algorithms within 3D manifold-solids of prospects (mathematically known as solids of high genus/genera [holes]). It suffices to say that from a mathematical point of view, a manifold surface or a manifold solid cannot be reduced into a line-network without losing the metric properties of space unless a relatively high density of graph vertices is produced to "sample" the space. Though accessibility and visibility analyses have different inherent characteristics, and thus modelling requirements, they have a similar constituent element, i.e., the spatial units. To make the connection more explicit for exploratory, associative, and generative modelling studies, we argue that the models be made separately by respecting the inherent differences of the two notions while using similar constituent elements to allow for collation and connection of their results.

We propose here to derive the topological model from a regular discretisation of the geometry of walkable surfaces and visible environments to provide for comprehensive ergonomic research (see Figure 1 for a complete picture). We argue that a regular spatial subdivision based on voxels can provide a uniform way of modelling 2-Manifold walkable spaces and 3-Manifold visible spaces for studying accessibility and visibility respectively. Space Syntax models were also later generalised by the models made by late Alasdair Turner for pixel-based visibility and accessibility analysis (called the Visibility Graph Analysis) integrated into Depth Map (Turner et al., 2001; Turner, 2007), which could be seen as consistent with the framework proposed in this paper. A few of the crucial concepts introduced in this section are explained in more detail in Table 1. The following sections elaborate on the steps introduced in Figure 1.

3.2 Spatial modelling based on topological voxelation

We propose to use the regular discretisation of spatial structures through voxels to construct a topological voxelization of 2D manifold surfaces and 3D manifold solids for accessibility and visibility analyses. Analysing both qualities on voxel models can be done by using graph/network representations constructed based on the topological relations between voxels, which includes (but is not necessarily limited to) the face-to-face connectivity between the voxels.

Both the walkable passages needed by accessibility analyses and visible prospect needed by visibility analyses could be modelled through topological voxelation on their original manifold. Afterwards, the voxelated model could be uniformly used to construct accessibility graphs and visibility graphs can be respectively constructed based on definitions of isochronal accessible neighbourhoods and isovist visible neighbourhoods. These local definitions of connectivity effectively result in a graph that encapsulates the connectivity structure of a spatial configuration in terms of immediate accessibility or immediate visibility. Figures 2, 3 show how the accessibility and visibility graphs could be constructed stepwise in a hypothetical voxelated space referenced with a 3D Morton curve, on a horizontal and vertical section, respectively. Both examples show the local neighbours of a separate voxel unit, demonstrating the edges and all connected nodes to this specific node in the graph to be constructed. After going through this process over all the nodes in the set, the corresponding graphs could be constructed (in the form of sparse matrices). In practice, interestingly enough, both such spatial models can be built by using regular volumetric cells or voxels with the help of software packages such as topoGenesis in a straightforward manner (Azadi and Nourian, 2021). It is also noteworthy that the visibility graph can be constructed within the raster through Fast Voxel Traversal algorithm (Amanatides and Woo, 1987).

It is worth mentioning that the scale of the graphs generated from walkable surfaces and visible solids are usually not the same due to the difference in dimensionality. The latter would have far more nodes than the former, while in a graph with binary weights on edges, the former is usually a subset of the latter, as well, since all the [immediately] accessible locations are [immediately] visible for people in specific perspectives from the corresponding voxel. As shown with a rightward arrow in Figure 1, the entire voxel space of the accessibility graph can be elevated to an eyesight level and

10.3389/fbuil.2023.1300843

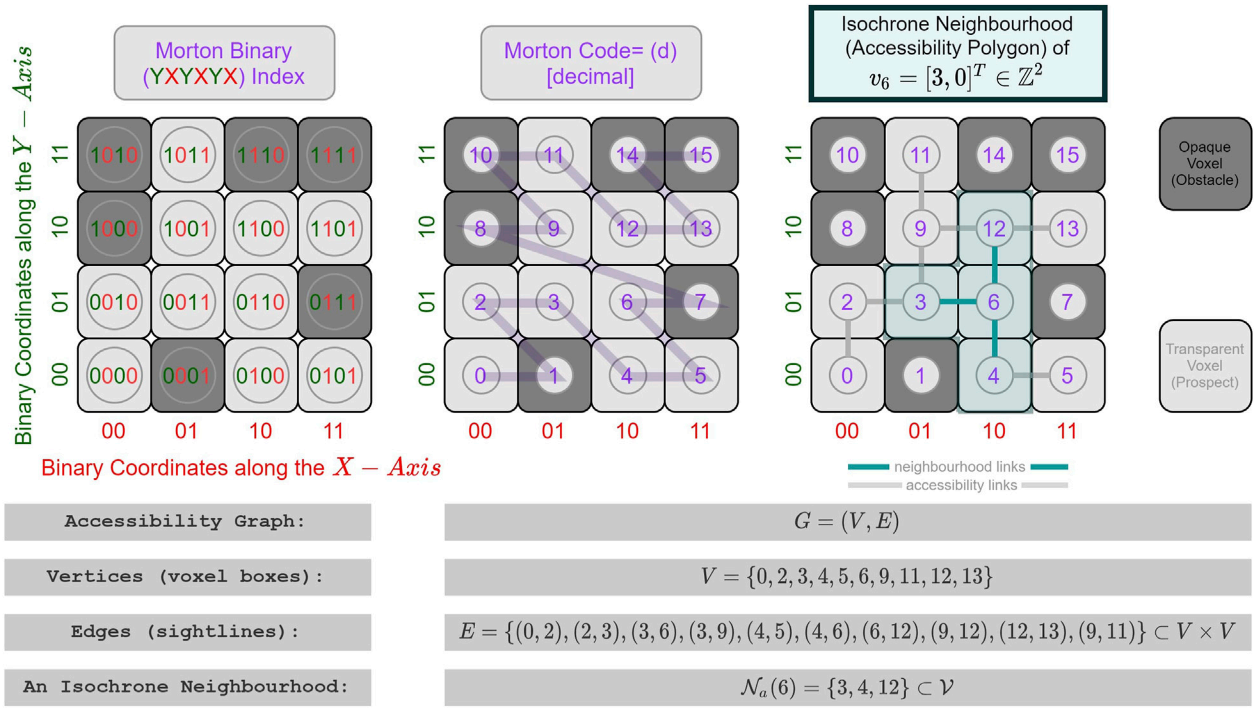


FIGURE 2
Accessibility graph construction on a hypothetical horizontal section of a voxelated passage space referenced with a 3D Morton curve.

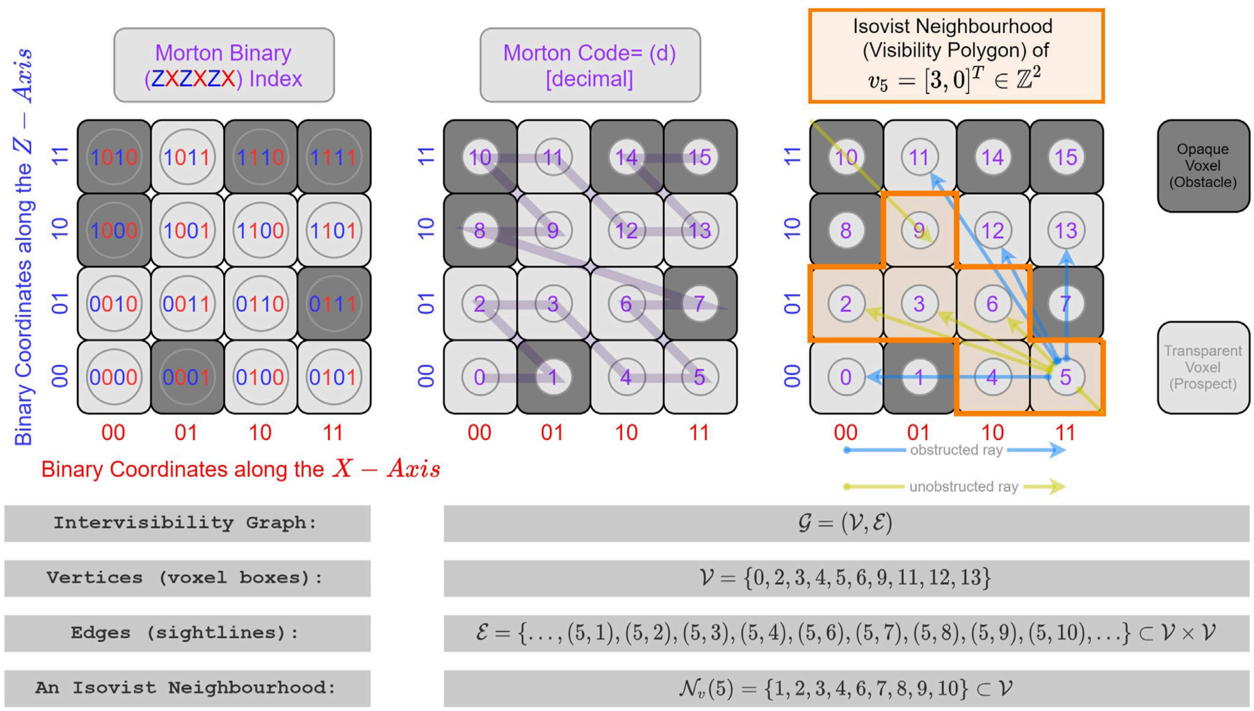


FIGURE 3
Visibility graph construction on a hypothetical vertical section of a voxelated prospect space referenced with a 3D Morton curve. Original mesh. Continuous 2D manifold. Accessibility Graph (Katz Centrality).

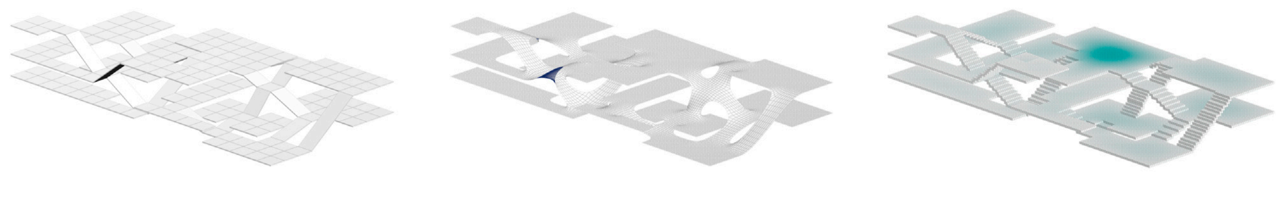


FIGURE 4
The original mesh is representing a 2-manifold walkable surface of a hypothetical architectural configuration (Nourian et al., 2015). After voxelization, the voxelated model has been used as the basis of the accessibility graph (16,548 nodes and 31,958 links). The result of Katz centrality has been visualized on the accessibility graph using Houdini 18.5 (https://www.sidefx.com/). Katz centrality was performed using NetworkX-NumPy (2.6.3) (Hagberg et al., 2008) implementation with the attenuation factor set to 99.9% of the reciprocal of the largest eigenvalue of the adjacency matrix. The procedure took 43 s on a 6-Core Intel Core i9.

incorporated into an intervisibility graph. In this way, a bipartite graph of inter-visibility can be formed to distinguish the visibility stances and visibility targets. Note that the examples shown in Figures 2, 3 are based on *immediate binary* connections. We want to stress the semantic difference of accessibility and visibility mentioned in Section 2.3. In the framework proposed by this paper, the two approaches are further unified: *direct/immediate* spatial or visual connections (visible and accessible neighbourhoods of a certain location) could be modelled with the edges in the graph, while the locations that are *eventually* accessible or *eventually* visible by people could be represented with a path in the graph traversing the walkable nodes. As a result, a global picture of [eventual] accessible or [eventual] visible spaces can be constructed from a "connected-component" (an island in the graph, in a manner of speaking). Furthermore, as described in Table 1, the relations represented by the edges in a graph can be either binary for marking if a pair of nodes are interconnected/inter-visible or not, or in the real-number range to get the relations weighted with some Gaussian normalisation, based on the different proximity/similarity/connectivity of the nodes. Furthermore, as described in Table 1, the relations represented by the edges in a graph can be either binary for marking if a pair of nodes are interconnected/inter-visible or not, or in the real-number range to get the relations weighted with some Gaussian normalisation, based on the different proximity/similarity/connectivity of the nodes. For example, for accessibility, the existence of slopes would make the comfort (ease) of walking in different directions vary (Nourian, 2016; Nourian et al., 2015), and places farther than a certain area would be homogeneously considered as "too far" for being accessible by people; thus changing the concept of accessibility to a *linguistic variable* in the sense of Fuzzy Logics (Nourian et al., 2016a). Similarly, for visibility, the angles of viewing can be assumed to be different in comfort and ease of viewing and finding a salient object (e.g., in the context of way-finding). Evidently, all these aspects could be modelled with adjustable parameters in the voxel-based modelling framework.

Our proposed approach makes the distinction between the topology and geometry of the environment unnecessary (Morello and Ratti, 2009; Fisher-Gewirtzman et al., 2013; Nourian et al., 2016b; Gorte et al., 2019), as the metric distance between locations is preserved even after graph construction and can be inferred from the integer coordinates of voxels and could be easily demonstrated on the walkable manifold (Nourian, 2016; Gorte et al., 2019), which would be otherwise lost in purely topological models (Ratti, 2004b).

Additionally, it provides a uniform way of studying spatial ergonomic matters such as way-finding in architectural (indoor) and urban (outdoor) environments as it removes the need to distinguish streets and corridors from other walkable areas such as squares and rooms (as seen in Figure 4). Both indoor (Gorte et al., 2019) and outdoor (Nourian et al., 2016b; Morello and Ratti, 2009) spaces can be modelled in the same way either as voxelated approximations of walkable 2-manifold surfaces or as voxelated approximations of visible 3-manifold solids. Specifically, Topological Voxelization at a geographical scale (Nourian et al., 2016b) preserves the most important properties of spatial forms and provides a uniform level of resolution while removing noise and unnecessary details. In scenarios where a combination of models are needed, voxel-indices or vertex indices in the graph can be globally unique indices referring to a space-filling curve, such as the Morton curve (see Figures 2, 3), thus making it possible to adjoin graphs if necessary without confusion. In addition, the globally unique indexing will facilitate querying data on both the graph and voxelated model. Specifically, the binary coordinates of the voxelated model referenced with a Morton curve can show the clustering information of the nodes in the space, since it effectively forms a Quadtree (2D) or Octree (3D) in space.

In summary, the voxelated framework provides a uniform way of constructing the base models on which both types of analyses/simulations (accessibility and visibility) can be performed.

3.3 Generality of voxel-based models

The accessibility and visibility graphs described above can facilitate spatial analysis in multiple ways: from providing for straightforward ways of converting questions to standard queries and property checks on graphs, at the simplest level, to facilitating problem-solving by utilizing graph traversal algorithms, centrality analyses on networks, and more importantly enabling the utilization of constructs from Spectral Graph Theory for simulating Random-Walks/Markov-Chains or solving Partial Differential Equations. Similarly, the same graphs can be used as "environments" for making Agent-Based Simulation Models. Understanding the value of graph data models is easier when they reveal patterns that could not be revealed [so easily] otherwise. Even the simplest uses of such graphs such as running the most routine queries, e.g., the neighbours of a visibility vertex will reveal unintuitive properties of the modelled space such as, for instance, all the points in a neighbourhood that can see the tip of a tower. The so-called "viewshed" of a heritage building can similarly be made by finding the union of the visibility neighbourhoods of the voxels associated with it. The advantage of such straightforward procedures must be clear for standardizing and streamlining assessment procedures, e.g., with respect to municipal regulations for view protection concerning architectural heritage (for planning heritage buffer zones). Furthermore, such graphs can also be used to answer generative design questions, e.g., what is the shape of the locus of points with a certain level of minimum view towards greenery in a neighbourhood or a hospital? Where should we place a sign with maximum view probability for pedestrians in a metro station? These questions refer back to the central question of ergonomics study in space: the comfort and discomfort of people in space and how a design or an intervention can change such comfort levels. In summary, there are various types of models (as previously mentioned in Section 2.1) that can be constructed based on the voxel-graph models to study spatial ergonomics questions as application scenarios:

- **Exploratory Models**: similar to social network analysis, the centrality-based analysis could be performed on the voxelated models. They are typically exploratory in the sense that a network centrality indicator is first computed and assumed to be an indicator of the *spatial potential* (or lack thereof) for human activities (e.g., pedestrian movement).
- **Associative Models**: we could collect data concerning urban spatial ergonomics, such as the real-world time needed for navigation, the felt comfort or discomfort and/or physiological parameters of comfort during walking, the pattern of cognitive maps formed in the space, the emotions, memories and other senses in space, the efficiency, effectiveness, and vitality of space (Nourian et al., 2021), and use the voxelated model as the common ground, attribute the aforementioned data to nodes or edges of the graph and perform associative analysis and predictions.
- **Generative Models**: we could also conduct various simulations on the voxelated model with different parametric configurations to come to an interpretable and explainable conclusion. For example, simulation models of pedestrian behaviour could be made by discretisation of a Partial Differential Equation describing universal rules of the pedestrian behaviour in a spatial domain (e.g., (Hoogendoorn et al., 2014)) or as an Agent-Based Simulation-Model (e.g., (Gath-Morad et al., 2021; Turner, 2007)).

4 Demonstration

To illustrate the potentials of the proposed framework, we have created a simple exploratory model of accessibility and visibility based on a portion of urban environment of Rotterdam, the Netherlands (see Figure 5). The Zomerhofkwartier (ZoHo) in Rotterdam, situated near Central Station, is undergoing redevelopment. The area is predominantly residential but it also includes some office spaces. On the boarder of the area lies "de Hofbogen," a former elevated rail track from the 19th century that is not in use any more. Historically facing socio-economic challenges, including poverty and perceived insecurity, the predominantly migrant-populated area has seen recent initiatives like the Waterplein Benthemplein project to enhance social resilience. In 2020, the City of Rotterdam announced new development plans for ZoHo aiming to create new homes, work spaces, and rooftop green landscapes that will be connected to the future Hofbogenpark. The planning process involves consultation with the ZoHo district-level council and groups such as ZoHo Citizens, representing local entrepreneurs and organizations (Peinhardt, 2021).

The surface mesh represents a portion of the ZoHo neighbourhood in Rotterdam. After voxelization, the voxelated model is used as the basis of the accessibility graph (1,631 nodes and 2,741 links) and the visibility graph (4,477 nodes and 960,523 links). The result of Katz centrality has been visualized on both graphs using Houdini 18.5. Katz centrality was performed using NetworkX-NumPy (2.6.3) (Hagberg et al., 2008) implementation with the attenuation factor set to 95.0% of the reciprocal of the largest eigenvalue of the adjacency matrix. The procedure took 0.1 and 4.5 s on a 6-Core Intel Core i9 for accessibility and visibility graph accordingly. The codes and models are open source and available at [redacted].

The models reveal the usefulness of accessibility and visibility graphs by visualizing a Katz Centrality indicator on both of them, which is a particular variant of Eigenvector Centrality proposed by Leo Katz in 1953 for revealing the relative importance of nodes in social networks as to the 'status' of actors (Katz, 1953). Eigenvector centrality indicators shows the relative centrality of the nodes in a network with respect to the relative centrality of their neighbours recursively, which can also be adjusted to be the final result of a Random-Walk Simulation. Intuitively, a node of high eigenvector centrality can be said to have a high influential position for diffusions or dispersion of information along the links of the network. As explained in (Nourian, 2016) (pp.228-236), such eigenvector centrality indicators are used in geographical analysis, social network analysis and webpage sorting (PageRank (Page et al., 1999)). The particular advantage of Katz centrality, as compared to PageRank and the 'normal' eigenvector centrality, is that it works for directed graphs and that it is not influenced by the degree regularity of the nodes that is a characteristic of accessibility graphs on voxelated manifolds. Katz centrality indicator has an attenuation parameter that defines the relative importance of longer walks as compared to short walks, which effectively allows for mitigating the adverse effects of the extreme degree-regularity of accessibility graphs. In particular, the results presented in Figure 5 highlight the utility of accessibility and visibility graphs by revealing the relative importance of some locations in a voxelated network. Interestingly, obtained with the same Katz centrality indicator, both results reveal meaningful patterns of influence with regards to accessibility and visibility. Theoretically, a place of high Katz centrality in the accessibility graph can be said to be highly accessible even for random walkers, thus beneficial for planning influential urban functions that need to be easily accessible; and a place of high Katz centrality in the visibility graph can be said to be highly visible

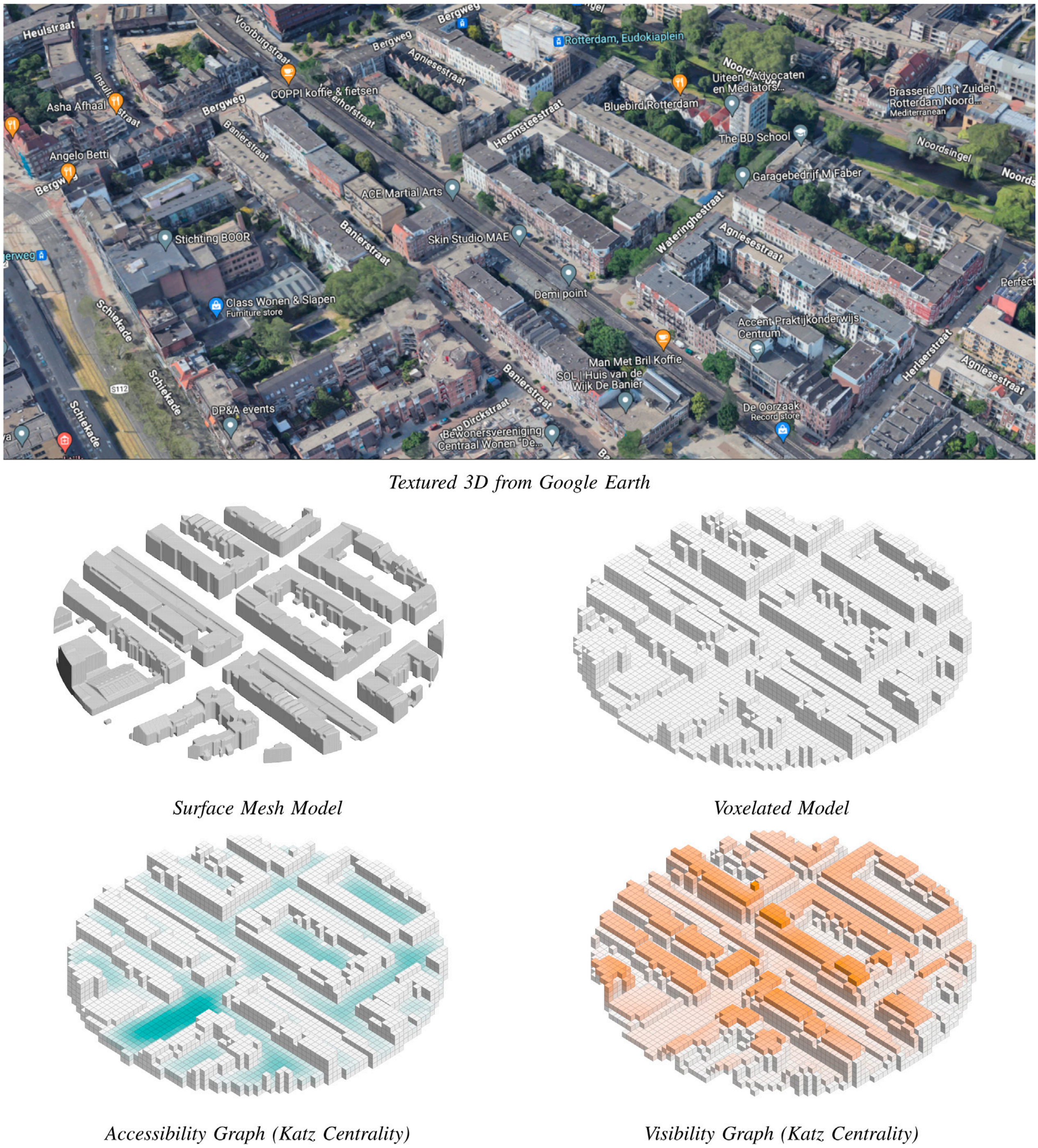


FIGURE 5
A portion of the ZoHo neighbourhood in Rotterdam: the surface mesh and voxelated models; Katz centrality for accessibility and visibility graphs; along with the view of the textured 3D model of the area from Google Earth to illustrate the urban context.

from the highly visible places, thus of importance with respect to visibility, proper for, e.g., the construction of a new landmark. Note the difference between the two voxel models used for accessibility and visibility analyses: the base map in voxels used for visibility analysis represents a one-layer-thick voxel envelope around the urban objects while the base map in voxels used for accessibility analysis represents the walkable ground surfaces. For example, note that the spots of high altitude in the visibility graph are highlighted in the centrality analysis and that the highlighted spots in the accessibility graph are obviously connected to layers of well-accessible voxels within the scope of the model. Such indicators would also be useful in creating safe spaces, e.g., following the normative ideas of the theory of Defensible Space Theory of Newman (Newman, 1996) and the arguments of Jane Jacobs (Jacobs, 2016) (see, e.g., (Byon et al., 2010; Rosser et al., 2017)). Therefore, we could argue that such exploratory computation is beneficial for both ex-ante and ex-post assessment of built environments as mentioned in Section 2.1.

Another rationale for demonstrating preliminary results from Katz centrality is out of pragmatic considerations concerning the computation costs. Accessibility graphs are utilized nowadays on a daily basis on navigation platforms such as Google Maps and thus they are already commonplace. However, visibility graphs and their value in the study of ergonomics are less known, perhaps due to the prohibitive time and memory costs of making visibility graphs of substantial size. As can be seen in the provided examples from such graphs in an urban context, the visibility graphs constructed as such can be quite large for urban environments (both due to involving many nodes and many more links), orders of magnitude bigger than their accessibility graph counterparts. Even though conventional centrality analyses (such as closeness centrality and the notoriously expensive betweenness centrality) will be prohibitively expensive to run on such humongous graphs, stochastic analyses such as Katz centrality and PageRank analyses can be performed successfully on such graphs, as shown in the examples.

We have only demonstrated the potential of exploratory models in these exemplary settings. The application of associative and generative models will be the future focus of our work. It is important to note that we only discussed the construction of unweighted and undirected binary graphs but more interesting questions can be answered by weighting the adjacency links in directed graphs (both for accessibility and visibility graphs) according to the purpose of the model, which will be supported by the Katz centrality indicator as introduced, without loss of generality.

5 Conclusion and discussion

Accessibility and visibility analyses could be considered the two most crucial problems concerned within urban spatial ergonomics to discuss the relations between the built environment and human behaviours. Considering the advantages of the voxel models in the unification of geometry and topology, they facilitate the synthesis of an overarching representation model for spatial configurations (at both architectural and urban scales). This is essentially different from the kind of ergonomics discussed in the context of industrial design as that field operates on the object-scale and focuses more on the geometrical and physical properties of the human body, whereas in the spatial-scale of the built environment, the discussion of ergonomics focuses on cognitive and locomotive capabilities of humans that govern their behaviour within the space.

In a voxelated space, a spatial configuration can be defined as a coloured graph, dual to a voxelated domain, in which the voxel colours/labels indicate the designated use (functionality) of spaces. Such a model of a spatial configuration would encompass all the aforementioned aspects of importance previously considered in architectural/urban morphological research (geometry, topology, and land-use/function-allocation). However, as compared to irregular networks, voxelated networks can be rather dense and thus heavy for computational simulations of accessibility and visibility (as seen in Figure 5). Whilst this can be a pragmatic barrier in the computational aspects of research in spatial ergonomics, the uniformity arising out of the simplicity of the voxel-graph models makes them unrivalled for spatial ergonomics models, especially for developing simulation models. In addition to urban environments, this framework and methodology is also applicable to the case of an indoor space with pieces of furniture that is voxelated (with appropriate size) to analyze the visibility and accessibility within indoor environments. Particularly, the added value of this framework is in having an integrated approach to how indoor and outdoor environment can be modelled and analysed.

One point of attention that was not discussed throughout the article is the temporal dimension of simulation models: in the study of visibility and accessibility, time can be a factor that can be inherently present in simulation models (think of daylight and its effects on visibility or temporal patterns of movement in a city/building), but exploratory models or associative models are relatively indifferent to time. On the other hand, in setting impact-assessment models for decision-support, one is often interested in the long-term and average temporal effects of decisions on an environment. This entails that the results of the simulations often need to be aggregated at least in time to provide such a long-term perspective. It might also be desirable to simplify or generalise the spatial structure of a building or a city to a few nodes or even one node for assessment purposes, hence another reason to aggregate/integrate the simulation results in space for assessment purposes, which is also not deeply touched by this paper.

Providing a uniform representation model such as the one proposed here enables the discipline to evaluate a given spatial scenario as to spatial ergonomics. Furthermore, the discretisation provides a clear and navigable structure for a solution space and provides for adopting a decision-making approach to urban and architectural design and developing explainable and explicit knowledge about ergonomics and human factors of architectural and urban design. We sincerely hope that the theoretical framework proposed in the paper could initiate some discussions.

Long Abstract (600 words) How can we assess the ergonomic comfort of a sizeable spatial configuration such as the indoor space of a complex building or an urban landscape when we design, plan, and manage the space? Is there a fundamental difference between indoor [architectural] spatial configurations and outdoor [urban] spatial configurations? Can we have a unified approach to the computational study of spatial ergonomics? This paper addresses these fundamental questions by providing a brief taxonomic review of the scholarly literature on these matters from a mathematical point of view. It presents a brief introduction to the modelling-based approaches to the study of spatial complexity and the computational ways of understanding the fundamental effects of spatial configuration on human behaviours. The paper argues that a substantial part of Urban Spatial Ergonomics (USE) as an interdisciplinary field of scholarly research is concerned with understanding the complex relations of pedestrians' behaviour with the configuration of complex spatial environments. Such systematic understandings could inform and underpin design activities by introducing a scientific approach to evaluating the impact of spatial designs and interventions on human comfort. Moreover, the paper argues that a knowledge-based approach for informing decision-making in design activities is necessary for guaranteeing human qualities usually referred to as the live-ability of buildings and cities, particularly by introducing ex-ante and ex-post assessment frameworks. In this sense, USE could be closely related to the notions of Design Sciences, the Sciences of Artificial, and Generative Sciences. The paper further introduces three types of models that could be beneficial for USE, namely, exploratory models, associative models, and generative models, suitable for application in the ex-ante and ex-post assessments of design decisions. Furthermore, the paper proposes a computational approach for ergonomic assessments of spatial configurations that explicitly allows for combined accessibility and visibility analyses in the built environment. The gist of this approach is the conceptualisation of spatial configurations as rasterised (voxelated) 2D manifold walkable terrains whose voxels have 3D vistas, unifying the simulations and analyses of accessibility and visibility with the mathematical representation of graphs. The paper elaborates on how such a representation of spatial configurations can provide for conducting various sorts of computational queries, analyses, and simulation experiments for research in spatial ergonomics. Specifically, a voxelated model of space (whether the original model of space is a 2D manifold surface or a 3D manifold solid) will have multiple advantageous properties, including the straightforwardness of constructing a topological model without losing metric distance information, and the unified representation on both indoor and outdoor spaces in both architectural and geographical scales. Additionally, the paper includes a few demonstrative examples illustrating how accessibility and visibility graphs could be constructed in both a hypothetical spatial configuration and a real-world urban district in Rotterdam, the Netherlands. Preliminary results of exploratory models using Katz centrality, a variant of eigenvector centrality originated in Social Network Analysis, in both graphs are presented, which show the potential of voxelated models of accessibility and visibility analyses in USE studies. The paper concludes with a mapping of the computational modelling approaches pertinent to the study and assessment of spatial ergonomics; and marks avenues of future research on various categories of exploratory, generative, and associative models for ex-ante and ex-post assessment of ergonomic matters at spatial scales.

Data availability statement

Original datasets are available in a publicly accessible repository: https//github.com/shervinazadi/Urban_Ergonomics

Author contributions

SA: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing–original draft, Writing–review and editing. NB: Conceptualization, Formal Analysis, Investigation, Methodology, Resources, Writing–original draft, Writing–review and editing. PN: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Supervision, Visualization, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Agusdinata, D. B. (2008). *Exploratory modeling and analysis: a promising method to deal with deep uncertainty*. Delft: TU Delft Open. \MakeLowercase []{ISBN} 9789071382260.

Algers, S., Eliasson, J., and Mattsson, L.-G. (2005). Is it time to use activity-based urban transport models? a discussion of planning needs and modelling possibilities. *Ann. Regional Sci.* 39 (4), 767–789. doi:10.1007/s00168-005-0016-8

Amanatides, J., and Woo, A. (1987). A fast voxel traversal algorithm for ray tracing. *Eurographics* 87 (3), 3–10. doi:10.2312/EGTP.19871000

Angel, S. (1968). *Discouraging crime through city planning*. University of California Institute of Urban and Regional Development.

Azadi, S., and Nourian, P. (2021). *topoGenesis*. original-date: 2020-05-31T13:29:16Z, Aug. 2021. Available: https://github.com/shervinazadi/topoGenesis(visitedon09/12/2021.

Bankes, S. (1993). Exploratory modeling for policy analysis. *Operations Res.* 41 (3), 435–449. \MakeLowercase []{ISSN} 0030-364X. doi:10.1287/opre.41.3.435

Batty, M. (2007). *Cities and complexity: understanding cities with cellular Automata, agent-based models, and fractals*. MIT Press. \MakeLowercase []{ISBN} 9780262524797.

Batty, M. (2001). Exploring isovist fields: space and shape in architectural and urban morphology. *Environ. Plan. B Plan. Des.* 28 (1), 123–150. 10.1068/b2725. Available: doi:10.1068/b2725

Belir, O., and Onder, D. (2013). *Accessibility in public spaces: spatial legibility for visually impaired people*.

Benedikt, M. L. (1979). To take hold of space: isovists and isovist fields. *Environ. Plan. B Plan. Des.* 6 (1), 47–65. doi:10.1068/b060047

Bentley, I. (1985). Responsive environments: a Manual for designers *(butterworth architecture)*. London: Architectural Press. \MakeLowercase []{ISBN} 9780750605663.

Blanchard, P., and Volchenkov, D. (2008). *Mathematical analysis of urban spatial networks*. Springer Science and Business Media.

Bonacich, P. (1987). Power and centrality: a family of measures. *Am. J. Sociol.* 92 (5), 1170–1182. \MakeLowercase []{ISSN} 00029602, 15375390. doi:10.1086/228631

Byon, Y.-J., Abdulhai, B., and Shalaby, A. (2010). Incorporating scenic view, slope, and crime rate into route choices: emphasis on three-dimensional geographic information systems with digital elevation models and crime rate geospatial data. *Transp. Res. Rec.* 2183 (1), 94–102. doi:10.3141/2183-11

Chiu, Y.-H., Chuang, I.-T., and Tsai, C.-Y. (2020). Analyzing correlation of urban functionality and spatial configuration. *Environ. Plan. B Urban Anal. City Sci.* 48 (5), 1296–1313. doi:10.1177/2399808320924673

Conzen, M. R. G. (1960). Alnwick, northumberland: a study in town-plan analysis. *Trans. Pap.* 27, iii. doi:10.2307/621094

Cozens, P. M., Saville, G., and Hillier, D. (2005). *Crime prevention through environmental design (cpted): a review and modern bibliography*. Leeds: Property management.

Crooks, A., Croitoru, A., Lu, X., Wise, S., Irvine, J., and Stefanidis, A. (2015). Walk this way: improving pedestrian agent-based models through scene activity analysis. *ISPRS Int. J. Geo-Information* 4 (3), 1627–1656. doi:10.3390/ijgi4031627

Çubukçuoğlu, C., Nourian, P., Tasgetiren, M., Sariyildiz, I., and Azadi, S. (2021). Hospital layout design renovation as a Quadratic Assignment Problem with geodesic distances. *J. Build. Eng.* 44, 102952–952. doi:10.1016/j.jobe.2021.102952

Dalton, R., Bafna, S., Conroy Dalton, R., and Bafna, S. (2003). "The syntactical image of the city: a reciprocal definition of spatial elements and spatial syntaxes," in *4th international space syntax symposium* (London, UK), 17–19.

Daniel, M.-P., and Denis, M. (1998). Spatial descriptions as navigational aids: a cognitive analysis of route directions. *Kognitionswissenschaft* 7 (1), 45–52. 10.1007/bf03354963. Available:. doi:10.1007/bf03354963

DeLaurentis, D., and Callaway, R. K. (2004). A system-of-systems perspective for public policy decisions. *Rev. Policy Res.* 21 (6), 829–837. doi:10.1111/j.1541-1338.2004.00111.x

DeLyser, D. (1999). Authenticity on the ground: engaging the past in a California ghost town. *Ann. Assoc. Am. Geogr.* 89 (4), 602–632. doi:10.1111/0004-5608.00164

Denis, M., Pazzaglia, F., Cornoldi, C., and Bertolo, L. (1999). Spatial discourse and navigation: an analysis of route directions in the city of venice. *Appl. Cogn. Psychol.* 13 (2), 145–174. doi:10.1002/(sici)1099-0720(199904)13:2<145::aid-acp550>3.0.co;2-4

De Smith, M., Goodchild, M., Longley, P., and Longley, A. (2018). *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*. London: Drumlin Security. \MakeLowercase []{ISBN} 9781912556045.

Duckham, M., and Kulik, L. (2003). ""simplest" paths: automated route selection for navigation," in *International conference on spatial information theory* (Springer), 169–185.

Elias, B., Brenner, C., and Fisher, P., 2023 "Automatic generation and application of landmarks in navigation data sets. 469–480. doi:10.1007/3-540-26772-7_36

Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity* 4 (5), 41–60. doi:10.1002/(sici)1099-0526(199905/06)4:5<41::aid-cplx9>3.3.co;2-6

Epstein, J. M. (2008). Why model? *J. Artif. Soc. Soc. Simul.* 11 (4), 12.

Epstein, J. M. (2012). *Generative social science*. Princeton University Press.

Epstein, J. M., and Axtell, R. (1996). *Growing artificial societies: social science from the bottom up*. Cambridge, Massachusetts: Brookings Institution Press.

Fisher-Gewirtzman, D., Shashkov, A., and Doytsher, Y. (2013). Voxel based volumetric visibility analysis of urban environments. *Surv. Rev.* 45 (333), 451–461. \MakeLowercase []{ISSN} 00396265. doi:10.1179/1752270613Y.0000000059

Freundschuh, S. M., and Egenhofer, M. J. (1997). Human conceptions of spaces: implications for GIS. *Trans. GIS* 2 (4), 361–375. doi:10.1111/j.1467-9671.1997.tb00063.x

Frey, H. (1999). Designing the city: Towards a more sustainable urban form *(designing the city: towards a more sustainable urban form)*. London: E and FN Spon. \MakeLowercase []{ISBN} 9780419221104.

Galea, S., Ahern, J., Rudenstine, S., Wallace, Z., and Vlahov, D. (2005). Urban built environment and depression: a multilevel analysis. *J. Epidemiol. Community Health* 59 (10), 822–827. doi:10.1136/jech.2005.033084

Gath-Morad, M., Thrash, T., Schicker, J., Hölscher, C., Helbing, D., and Aguilar Melgar, L. E. (2021). Visibility matters during wayfinding in the vertical. *Sci. Rep.* 11 (1), 1–15. doi:10.1038/s41598-021-98439-1

Gilbreth, L. M. (1914). *The psychology of management: the function of the mind in determining, teaching and installing methods of least waste*. New York: Sturgis and Walton Company.

Gjerde, M. (2011). Visual evaluation of urban streetscapes: how do public preferences reconcile with those held by experts? *URBAN Des. Int.* 16 (3), 153–161. doi:10.1057/udi.2011.10

Golledge, R. (1999). *Wayfinding behavior: cognitive mapping and other spatial processes*. Baltimore, London: Johns Hopkins University Press. \MakeLowercase []{ISBN} 9780801859939. Available: https://books.google.nl/books?id=TjzxpAWiamUC.

Golledge, R. G. (1993). "Chapter 2 geographical perspectives on spatial cognition," in *Behavior and environment, ser. Advances in psychology*, T. Gärling and R. G. Golledge 96, North-Holland, 16–46. doi:10.1016/S0166-4115(08)60038-2

Gorte, B., Zlatanova, S., and Fadli, F. (2019). *Navigation in indoor voxel models*.

Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). "Exploring network structure, dynamics, and function using networkx," in *Proceedings of the 7th Python in science conference*. Editors G. Varoquaux, T. Vaught, and J. Millman (Pasadena, CA USA), 11–15.

Haq, S., and Luo, Y. (2012). Space syntax in healthcare facilities research: a review. *HERD Health Environ. Res. Des. J.* 5 (4), 98–117. 10.1177/193758671200500409. Available:. doi:10.1177/193758671200500409

Heppenstall, A. J., Crooks, A. T., See, L. M., and Batty, M. (2011). *Agent-based models of geographical systems*. Springer Science and Business Media.

Hillier, B. (1998). *Space is the machine: a configurational theory of architecture*. Cambridge University Press. \MakeLowercase []{ISBN} 9780521645287. Available: https://books.google.nl/books?id=8CygHAAACAAJ.

Hillier, B. (1999). The common language of space:a way of looking at the social,economic and environmental functioning of cities on a common basis. *en* 11 (3), 344–349. Available: http://www.jesc.ac.cn/jesc_en/ch/reader/view_abstract.aspx?file_no=19990316.

Hillier, B., and Hanson, J. (1989). *The social logic of space*. Cambridge University Press. \MakeLowercase []{ISBN} 9781139935685.

Hillier, B., Major, M., Desyllas, J., Karimi, K., Campos, B., and Stonor, T. (1996). *Tate gallery, millbank: a study of the existing layout and new masterplan proposal*.

Hillier, B., Penn, A., Hanson, J., Grajewski, T., and Xu, J. (1993). Natural movement: or, configuration and attraction in urban pedestrian movement. *Environ. Plan. B Plan. Des.* 20 (1), 29–66. doi:10.1068/b200029

Hoogendoorn, S. P., van Wageningen-Kessels, F. L., Daamen, W., and Duives, D. C. (2014). Continuum modelling of pedestrian flows: from microscopic principles to self-organised macroscopic phenomena. *Phys. A Stat. Mech. its Appl.* 416, 684–694. doi:10.1016/j.physa.2014.07.050

Jacobs, J. (2016). *The death and life of great American cities*. London: Vintage.

Jin, X., Long, Y., Sun, W., Lu, Y., Yang, X., and Tang, J. (2017). Evaluating cities' vitality and identifying ghost cities in China with emerging geographical data. *Cities* 63, 98–109. doi:10.1016/j.cities.2017.01.002

Johnson, J., and Fisher, W. (2013). Urban geography: an introductory analysis *(commonwealth and international library of science, technolog)*. Elsevier Science. \MakeLowercase []{ISBN} 9781483181547.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika* 18 (1), 39–43. doi:10.1007/bf02289026

Klippel, A., and Winter, S. (2005). "Structural salience of landmarks for route directions," in *Spatial information theory*. Editors A. G. Cohn and D. M. Mark (Berlin, Heidelberg: Springer Berlin Heidelberg), 347–362.

Koohsari, M. J., Kaczynski, A. T., Giles-Corti, B., and Karakiewicz, J. A. (2013). Effects of access to public open spaces on walking: is proximity enough? *Landsc. Urban Plan.* 117, 92–99. doi:10.1016/j.landurbplan.2013.04.020

Koseoglu, E., and Onder, D. E. (2011). Subjective and objective dimensions of spatial legibility. *Procedia - Soc. Behav. Sci.* 30, 1191–1195. doi:10.1016/j.sbspro.2011.10.231

Kostakos, V. (2010). "Space syntax and pervasive systems," in *Geospatial analysis and modelling of urban structure and dynamics*. Editors B. Jiang and X. Yao (Dordrecht: Springer Netherlands), 31–52. doi:10.1007/978-90-481-8572-6_3

Lawson, B. (2006). *How designers think: the design process demystified*. London: Routledge.

Li, X., Hijazi, I., Koenig, R., Lv, Z., Zhong, C., and Schmitt, G. (2016). Assessing essential qualities of urban space with emotional and visual data based on GIS technique. *ISPRS Int. J. Geo-Information* 5 (11), 218. 10.3390/ijgi5110218. Available: . doi:10.3390/ijgi5110218

Long, Y., and Baran, P. K. (2011). Does intelligibility affect place legibility? understanding the relationship between objective and subjective evaluations of the urban environment. *Environ. Behav.* 44 (5), 616–640. 10.1177/0013916511402059. Available:. doi:10.1177/0013916511402059

Lovelace, K. L., Hegarty, M., and Montello, D. R. (1999). "Elements of good route directions in familiar and unfamiliar environments," in *Spatial information theory. Cognitive and computational foundations of geographic information science*. Editors C. Freksa and D. M. Mark (Berlin, Heidelberg: Springer Berlin Heidelberg), 65–82.

Lynch, K., and for Urban Studies, J. C. (1960). The Image of the city *(Harvard-MIT joint center for urban studies series)*. Harvard University Press. \MakeLowercase []{ISBN} 9780262620017.

Macal, C. M., and North, M. J. (2009). "Agent-based modeling and simulation," in *Proceedings of the 2009 winter simulation conference (WSC)* (IEEE), 86–98.

Marshall, S. (2004). *Streets and patterns*. London: Routledge.

McAndrew, F. (1993). *Environmental psychology*. Pacific Grove: Brooks/Cole. \MakeLowercase []{ISBN} 9780534193089. Available: https://books.google.nl/books?id=ip99AAAAMAAJ.

Meyer, M., Meyer, M., and Miller, E. (2001). *Urban transportation planning: a decision-oriented approach (civil engineering series)*. McGraw-Hill. \MakeLowercase []{ISBN} 9780072423327.

Montello, D. R. (2007). "The contribution of space syntax to a comprehensive theory of environmental psychology," in *Proceedings of the 6th international space syntax symposium* (Istanbul: Citeseer), 1–12.

Morello, E., and Ratti, C. (2009). A digital image of the city: 3D isovists in Lynch's urban analysis. *Environ. Plan. B Plan. Des.* 36 (5), 837–853. \MakeLowercase []{ISSN} 14723417. doi:10.1068/b34144t

Moughtin, C. (2007). *Urban design: street and square*. London: Taylor and Francis. \MakeLowercase []{ISBN} 9781136350344.

Netto, V. M. (2015). “what is space syntax not?” reflections on space syntax as sociospatial theory. *URBAN Des. Int.* 21 (1), 25–40. doi:10.1057/udi.2015.21

Newman, O. (1996). *Creating defensible space*. Darby: Diane Publishing.

Nothegger, C., Winter, S., and Martin, R. (2004). Selection of salient features for route directions. *Spatial Cognition Comput.* 4, 113–136. doi:10.1207/s15427633scc0402_1

Nourian, P. (2016). *Configraphics: graph theoretical methods for design and analysis of spatial configurations*. Delft: TU Delft. \MakeLowercase []{ISBN} 9789461867209.

Nourian, P., and Bai, N. (2021). “Reuse: configurational assessment of functionality of buildings before and after a change,” in Mastermind Crash: a method to reveal the impact of architectural redesign*, ser. Mastermind Crash: a method to reveal the impact of architectural redesign*. Editor A. Pereira Roders 3rd ed. (Delft: TU Delft Open), 23–73. Available: http://pure.tudelft.nl/ws/portalfiles/portal/88291501/CRASH_mastermind_2021_v3.pdf.

Nourian, P., Gonçalves, R., Zlatanova, S., Ohori, K. A., and Vo, A. V. (2016b). Voxelization algorithms for geospatial applications: computational methods for voxelating spatial datasets of 3d city models containing 3d surface, curve and point data models. *MethodsX* 3, 69–86. doi:10.1016/j.mex.2016.01.001

Nourian, P., Rezvani, S., Sariyildiz, I., and van der Hoeven, F. (2016a). “Spectral modelling for spatial network analysis,” in *Proceedings of the symposium on simulation for architecture and urban design (simAUD 2016)* (Delft: SimAUD).

Nourian, P., Van Der Hoeven, F., Rezvani, S., and Sariylidiz, S. (2015). *Easiest paths for walking and cycling*. London: 10TH INTERNATIONAL SPACE SYNTAX SYMPOSIUM.

Okamoto, K., Kaneda, T., Ota, A., and Meziani, R. (2014). *Correlation analyses between underground spatial configuration and pedestrian flows by space syntax measures: a case study of underground mall complex in nagoya station*. Tokyo: National Library of Australia Cataloguing-in-Publication entry, 116.

Omer, I., and Goldblatt, R. (2016). Using space syntax and q-analysis for investigating movement patterns in buildings: the case of shopping malls. *Environ. Plan. B Urban Anal. City Sci.* 44 (3), 504–530. 10.1177/0265813516647061. Available:. doi:10.1177/0265813516647061

ONeill, M. J. (1991). Evaluation of a conceptual model of architectural legibility. *Environ. Behav.* 23 (3), 259–284. 10.1177/0013916591233001. Available:. doi:10.1177/0013916591233001

Pacione, M. (2009). *Urban geography: a global perspective*. London: Taylor and Francis. \MakeLowercase []{ISBN} 9781134043095.

Pafka, E., Dovey, K., and Aschwanden, G. D. (2018). Limits of space syntax for urban design: axiality, scale and sinuosity. *Environ. Plan. B Urban Anal. City Sci.* 47 (3), 508–522. 10.1177/2399808318786512. Available:. doi:10.1177/2399808318786512

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: bringing order to the web. *Stanf. InfoLab, Tech. Rep*.

Patle, B. K., Pandey, A., Parhi, D. R. K., and Jagadeesh, A. J. D. T. (2019). A review: on path planning strategies for navigation of mobile robot. *Def. Technol.* 15 (4), 582–606. doi:10.1016/j.dt.2019.04.011

Peinhardt, K. (2021). *Resilience through placemaking: public spaces in rotterdam's climate adaptation approach*. Bonn: eng. Discussion Paper 1/2021, 2021. 10. 23661/dp1.2021. Available: http://hdl.handle.net/10419/228834.

Penn, A. (2003). Space syntax and spatial cognition. *Environ. Behav.* 35 (1), 30–65. 10.1177/0013916502238864. Available:. doi:10.1177/0013916502238864

Peponis, J., Ross, C., and Rashid, M. (1997a). The structure of urban space, movement and co-presence: the case of atlanta. *Geoforum* 28 (3-4), 341–358. doi:10.1016/s0016-7185(97)00016-x

Peponis, J., Wineman, J., Rashid, M., Kim, S. H., and Bafna, S. (1997b). On the description of shape and spatial configuration inside buildings: convex partitions and their local properties. *Environ. Plan. B Plan. Des.* 24 (5), 761–781. doi:10.1068/b240761

Perloff, R., and Naman, J. L. (2012). Lillian gilbreth: tireless advocate for a general psychology. *Portraits Pioneers Psychol. Volume II* 2, 107.

Proshansky, H. M. (1978). The city and self-identity. *Environ. Behav.* 10 (2), 147–169. 10 . 1177/0013916578102002. Available:. doi:10.1177/0013916578102002

Ratti, C. (2004a). Rejoinder to hillier and penn. *Environ. Plan. B Plan. Des.* 31 (4), 513–516. doi:10.1068/b3019b

Ratti, C. (2004b). Space syntax: some inconsistencies. *Environ. Plan. B Plan. Des.* 31 (4), 487–499. doi:10.1068/b3019

Rosser, G., Davies, T., Bowers, K. J., Johnson, S. D., and Cheng, T. (2017). Predictive crime mapping: arbitrary grids or street networks? *J. quantitative Criminol.* 33 (3), 569–594. doi:10.1007/s10940-016-9321-x

Schelling, T. (1984). *Choice and consequence*. Harvard University Press. \MakeLowercase []{ISBN} 9780674127715. Available: https://books.google.nl/books?id=MF08nRe6jQoC.

Schlicht, E. (2012). *Isolation and aggregation in economics*. Springer Berlin Heidelberg. \MakeLowercase []{ISBN} 9783642702983. Available: https://books.google.nl/books?id=o6%5C_tCAAAQBAJ.

Schroder, C. J., Mackaness, W. A., and Gittings, B. M. (2011). Giving the “right” route directions: the requirements for pedestrian navigation systems. *Trans. GIS* 15 (3), 419–438. doi:10.1111/j.1467-9671.2011.01266.x

Simon, H. (1970). *The sciences of the artificial (karl taylor compton lectures)*. Cambridge: Tun Huang Book Company. \MakeLowercase []{ISBN} 9780262690232. Available at: https://books.google.nl/books?id=mjZgAAAAMAAJ.

Sorrows, M. E., and Hirtle, S. C. (1999). “The nature of landmarks for real and electronic spaces,” in *Spatial information theory. Cognitive and computational foundations of geographic information science*. Editors C. Freksa and D. M. Mark (Berlin, Heidelberg: Springer Berlin Heidelberg), 37–50.

Stamps, A. E. (2005). Isovists, enclosure, and permeability theory. *Environ. Plan. B Plan. Des.* 32 (5), 735–762. 10.1068/b31138. Available:. doi:10.1068/b31138

Steiniger, S., Lange, T., Burghardt, D., and Weibel, R. (2008). An approach for the classification of urban building structures based on discriminant analysis techniques. *Trans. GIS* 12 (1), 31–59. doi:10.1111/j.1467-9671.2008.01085.x

Stephenson, J. (2008). The cultural values model: an integrated approach to values in landscapes. *Landsc. Urban Plan.* 84, 127–139. doi:10.1016/j.landurbplan.2007.07.003

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Econ. Geogr.* 46 (1), 234–240. doi:10.2307/143141

Turner, A. (2001). “Angular analysis,” in *Proceedings of the 3rd international symposium on space syntax* (Atlanta, GA: Georgia Institute of Technology), 30–31.

Turner, A. (2007). *Ucl depthmap 7: from isovist analysis to generic spatial network analysis*. Istanbul: New Developments in Space Syntax Software.

Turner, A., Doxa, M., O'Sullivan, D., and Penn, A. (2001). From isovists to visibility graphs: a methodology for the analysis of architectural space. *Environ. Plan. B Plan. Des.* 28 (1), 103–121. 10.1068/b2684. Available:. doi:10.1068/b2684

Turner, A. (2003). Analysing the visual dynamics of spatial morphology. *Environ. Plan. B Plan. Des.* 30 (5), 657–676. 10 . 1068/b12962. Available:. doi:10.1068/b12962

Turner, A., and Penn, A. (1999). “Making isovists syntactic: isovist integration analysis,” in *Proceedings of of second international symposium on space syntax*, 103–121.

Vance, J. (1977). This Scene of man: the Role and Structure of the City in the Geography of western civilization *(harper and row series in geography)*. Harper's College Press. \MakeLowercase []{ISBN} 9780061674075.

van der Hoeven, F., and van Nes, A. (2014). Improving the design of urban underground space in metro stations using the space syntax methodology. *Tunn. Undergr. Space Technol.* 40, 64–74. doi:10.1016/j.tust.2013.09.007

Wang, R., Liu, Y., Xue, D., and Helbich, M. (2019a). Depressive symptoms among Chinese residents: how are the natural, built, and social environments correlated? *BMC public health* 19 (1), 887–888. doi:10.1186/s12889-019-7171-9

Wang, Z., Liang, Q., Duarte, F., Zhang, F., Charron, L., Johnsen, L., et al. (2019b). Quantifying legibility of indoor spaces using deep convolutional neural networks: case studies in train stations. *Build. Environ.* 160, 106099–099. doi:10.1016/j.buildenv.2019.04.035

Weisman, J. (1981). Evaluating architectural legibility. *Environ. Behav.* 13 (2), 189–204. 10 . 1177/0013916581132004. Available:. doi:10.1177/0013916581132004

Zmudzinska-Nowak, M. (2003). Searching for legible city form: kevin lynch's theory in contemporary perspective. *J. Urban Technol.* 10 (3), 19–39. doi:10.1080/1063073032000175390

frontiers | Frontiers in Built Environment

TYPE Original Research
PUBLISHED 29 February 2024
DOI 10.3389/fbuil.2024.1301049

Check for updates

OPEN ACCESS

EDITED BY
Vagelis Plevris,
Qatar University, Qatar

REVIEWED BY
Vasant Annasaheb Matsagar,
Indian Institute of Technology Delhi, India
Georgios A. Drosopoulos,
University of KwaZulu-Natal, South Africa
Feifei Wang,
Hunan City University, China

*CORRESPONDENCE
Qahtan Al-Shami,
6122000034@tju.edu.cn
Mugahed Amran,
m.amran@psau.edu.sa
Yaser Gamil,
yaser.gamil@ltu.se

RECEIVED 24 September 2023
ACCEPTED 19 February 2024
PUBLISHED 29 February 2024

CITATION
Al-Shami Q, Huang J, Amran M, Mugahed S, Alluqmani AE, Al-Haaj M, Gamil Y and Abdelgader HS (2024), Efficient numerical simulations on the forest barrier for seismic wave attenuation: engineering safe constructions.
Front. Built Environ. 10:1301049.
doi: 10.3389/fbuil.2024.1301049

Efficient numerical simulations on the forest barrier for seismic wave attenuation: engineering safe constructions

Qahtan Al-Shami[1,2,3]*, Jiankun Huang[1], Mugahed Amran[4,3]*, Saleh Mugahed[3], Ayed Eid Alluqmani[5], Mohammed Al-Haaj[2], Yaser Gamil[6,7]* and Hakim S. Abdelgader[8,9]

[1]School of Soil and Water Conservation, Beijing Forestry University, Beijing, China, [2]Department of Civil Engineering, Tianjin University, Tianjin, China, [3]Department of Civil Engineering, Faculty of Engineering and IT, Amran University, Amran, Yemen, [4]Department of Civil Engineering, College of Engineering, Prince Sattam Bin Abdulaziz University, Alkharj, Saudi Arabia, [5]Department of Civil Engineering, Islamic University of Madinah, Madinah, Saudi Arabia, [6]Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, Luleå, Sweden, [7]Department of Civil Engineering, School of Engineering, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, Selangor, Malaysia, [8]Department of Civil Engineering, Faculty of Engineering, University of Tripoli, Tripoli, Libya, [9]Faculty of Civil Engineering and Architecture, Lublin University of Technology, Lublin, Poland

This paper aims to elucidate the clear visibility of attenuating seismic waves (SWs) with forest trees as natural metamaterials known as forest metamaterials (FMs) arranged in a periodic pattern around the protected area. In analyzing the changeability of the FM models, five distinct cases of “metawall” configurations were considered. Numerical simulations were conducted to study the characteristics of bandgaps (BGs) and vibration modes for each model. The finite element method (FEM) was used to illustrate the generation of BGs in low frequency ranges. The commercial finite element code COMSOL Multiphysics 5.4a was adopted to carry out the numerical analysis, utilizing the sound cone method and the strain energy method. Wide BGs were generated for the Bragg scattering BGs and local resonance BGs owing to the gradual variations in tree height and the addition of a vertical load in the form of mass to simulate the tree foliage. The results were promising and confirmed the applicability of FEM based on the parametric design language ANSYS 17.2 software to apply the boundary conditions of the proposed models at frequencies below 100 Hz. The effects of the mechanical properties of the six layers of soil and the geometric parameters of FMs were studied intensively. Unit cell layouts and an engineered configuration for arranging FMs based on periodic theory to achieve significant results in controlling ground vibrations, which are valuable for protecting a large number of structures or an entire city, are recommended. Prior to construction, protecting a region and exerting control over FM characteristics are advantageous. The results exhibited the effect of the ‘trees’ upper portion (e.g., leaves, crown, and lateral bulky branches) and the gradual change in tree height on the width and position of BGs, which refers to the attenuation mechanism. Low frequency ranges of less than 100 Hz were particularly well

Abbreviations: SW, Seismic wave; FM, Forest metamaterial; FEM, Finite element method; BG, Bandgap; EW, Elastic wave; LR, Local resonance; 3D, Three-dimension model; IBZ, Irreducible Brillouin Zone; SCM, Sound cone method; EDM, Energy density method; ARF, Amplitude reduction factor; FRF, Frequency response function.

suited for attenuating SWs with FMs. However, an engineering method for a safe city construction should be proposed on the basis of the arrangement of urban trees to allow for the shielding of SWs in specific frequency ranges.

KEYWORDS

forest metamaterial, seismic waves, periodic structure, low-frequency wide BGs, FRF curves

1 Introduction

Seismic waves (SWs) are one of the physical phenomena that affect people and their property, and because the occurrence of this phenomenon is difficult to prevent, it must be minimized (Thompson et al., 2008). Thus, comprehending the causes of its danger, how to deal with it, and the possibility of reducing its risk is critical. Even though elastic waves (EWs) can cause significant structural damage (Marazzani et al., 2021; Qahtan et al., 2022), there must be techniques to reduce their destructive harms to protect constructions that can resist the coercion of a disaster when it occurs. The complete understanding of EW propagation is the first step toward developing attenuation techniques (Du et al., 2017). Pu et al. (Pu and Shi, 2018) investigated the effectiveness of periodic barriers and developed a better method for reducing elastic surface wave propagation and damping in soil. New methods for dramatically reducing the propagation of EWs (Sang et al., 2018; Mandal and Somala, 2020; Huang et al., 2021; Li G. et al., 2022), mitigating SWs (Jia and Shi, 2010; Colombi et al., 2020; Daradkeh et al., 2022), controlling ground vibrations (Dijckmans et al., 2015; Gao et al., 2015; Jiang et al., 2015; Schevenels and Lombaert, 2017; Thompson et al., 2019; An, 2022), utilizing seismic clocks (Sens-Schönfelder, 2008; Aravantinos-Zafiris and Sigalas, 2015; Gahlmann and Tassin, 2022), using sound sensors (Dong, 2021; Yang et al., 2021), and guiding waves (Prati, 2006; Li et al., 2021; Thomes et al., 2022) based on periodic structure theory have emerged. For attenuation purposes, locally resonant structures have been used as periodic structures in the frequency of interest to mitigate EWs (Wang M. Y. and Wang X., 2013). Huang et al. (Huang and Shi, 2013) investigated a periodic array of piles with their dispersion relation for vibration attenuation. Wood et al. (Woods et al., 1974) came up with the idea of using a regular pattern of vertical holes to deliberate down surface waves.

In natural disasters, most damage is caused by elastic surface waves (Uhlemann et al., 2016). Protecting civil structures from EWs can be accomplished by converting surface waves into shear waves (Wei et al., 2015). Engineered metawall barriers use local resonance (LR) to transform surface waves into bulk waves that permeate far below the surface (Mir et al., 2018). Kim et al. (Kim, 2012) shed light on a number of exciting design challenges, such as the effect of urban trees on the attenuation of EWs and their applications for SW hazard. Over a specific frequency range, periodically arranged metamaterials show significant attenuation of surface waves better than those arranged in a random way (Huang et al., 2019). Metamaterials are synthetic composites with desirable properties that are more valuable than the individual components (Kaushik, 2023). In essence, the expression "metamaterial" is a novel concept in the field of material design (Seive, 2019). To realize appropriate material operations, this concept overcomes the apparent natural rule of order structure on a critical performance scale (Barbuto, 2015). The main purpose of using seismic metamaterials is to protect building structures from wave hazards (Mu et al., 2020). However, the use of inexpensive materials and the realization of ecological construction are desirable. Urban trees are becoming increasingly popular because they greatly attenuate surface waves in cities (Liu et al., 2019). Forest metamaterials (FMs) are organized in a periodic way as seismic metamaterials (Colombi et al., 2016a). In comparison with the normal state's disorganized arrangement, FMs have distinct bandgaps (BGs) (Muhammad et al., 2020). As a result, they are an efficient, economical, and environment-friendly way to reduce the effects of EWs and control ground vibrations. Recent studies have addressed a number of construction problems, which include the role of forest trees in EW attenuation (Colombi et al., 2016a; Colombi et al., 2016b; Lott et al., 2020; Muhammad et al., 2020; Qahtan et al., 2022; He et al., 2023). Because of indirect interactions between plants and ground reflection, low-frequency EWs are attenuated (Lim, 2021).

Natural available metamaterials of forest trees can be used at the subwavelength scale of low-frequency BGs (Colombi et al., 2016a; Miniaci et al., 2016; Liu et al., 2019; Muhammad et al., 2020; Lim, 2021). Maurel et al. (Maurel et al., 2018) reported the coupling between the LR modes and impedance mismatch of converting elastic Love waves into bulk waves due to planting trees in the ground. Although BGs appear generally in the frequency range of interest (low frequencies bellow 100 Hz), many studies have been conducted on the results of experiments and simulations for the LR and Bragg scattering (BS) characters (Wang Y.-F. and Wang Y.-S., 2013; Brûlé et al., 2014; Lim, 2019; Sánchez et al., 2021; Wang et al., 2021; Zeng et al., 2022). The arrangement of trees and their resonance eigenvalues are coupled with the induced elastic surface waves (Muhammad et al., 2020). The mechanism of attenuating EWs and obtaining wide BGs in low frequency ranges is still obscure. The influence of various factors of FMs, such as the geometric properties of natural trees and the mechanical properties of soil, is not clear enough to achieve a distinctive model of BGs with the same order at least as other industrial materials, such as steel sections (Zhang et al., 2021). In the realm of planting and urbanity, trees have the potential to contribute to ecological conception, economic effects, and civilization and environmental values (Roux et al., 2018). Related works show an arrangement for protecting functionality by using rows of forest trees on the upper layer of substrate soil (Muhammad et al., 2020; Li S. et al., 2022). Trees in the forest have been used as seismic shields given their properties as natural metamaterials (Colquitt et al., 2017; Sewar et al., 2022). The numerical calculation of (2D) models, which ignores information about the propagation in the third direction, may lead to the loss of an essential calculation point (Audusse et al., 2021). Through a three dimension (3D) simulation model, which is a type of engineering, numerical analysis can be used to represent the visibility of an urban forest in a way that effectively protects against low-frequency EWs (Huang et al., 2019). The use of vegetation as an environmental solution to ground vibration problems has gained popularity (Lombi and Susini, 2009). Environmental solutions for noise and ground vibration reduction in the low frequency range are achieved through the interplay between periodic structures and ground reflections, which occurs in a roundabout way (Garg, 2022).

The increasing interest in natural metamaterials opens new paths to protect threatened areas of EW propagation. A qualitative investigation has been carried out to study the effects of various geometrical characteristics, including green belt configurations, resonator height, radius, lateral side branches, multilayers of soil, water content, and mechanical properties (i.e., elastic modulus, density, and Poisson's ratio), on the elimination and control of BG characteristics (Azzouz, 2021). The search for sustainable and eco-friendly solutions to such problems has become a necessity. Forest trees can be manipulated and engineered to provide protection and safety against ground vibration and noise (Cook and Van Haverbeke, 1974). FM-on substrate soil provide an effective offense in EWs' attenuation mechanism (Maurel et al., 2018). However, the mechanism and modeling of vibration reduction should be clearly emphasized. Incident elastic surface waves can be transformed into bulk shear waves or reflected, resulting in a "seismic rainbow" effect similar to the electromagnetic surface optical rainbow (Colombi et al., 2016a). Although investigations have shown essential and promising development in the use of FMs as vibratory EW attenuators, considerable knowledge gaps remain in this research field. For example, several proposed models have been implemented to confirm observed simulation results (Colombi et al., 2016a; Zhang and He, 2021). While a surface wave propagates in forested areas and then gradually increases or decreases, it is reflected back or converted into a body wave and propagated downward (Colombi et al., 2017). There is a decent understanding of the effectiveness of resonators/FMs to produce a BG at their resonant frequency (Huang et al., 2019). Although, the ability to produce a wide BG at low frequencies (<100 Hz) for practical applications is desirable, it is still unclear for vibration reduction applications.

Accordingly, in this paper, 3D simulations of FMs, including their upper parts, are numerically highlighted and modeled in an engineering way. The aim is to identify the role of urban trees in the attenuation of EWs, as well as the effect of different heights of periodic structures upward and downward on creating wide BGs in the frequencies of interest less than 100 Hz. Research methods have shifted to metamaterials, rearranging periodic structures in soil in a specific geometric way. FMs' extensive range of applications toward urban areas includes the capacity to generate BS and LR BGs. The increasing interest in metamaterials is opening up new paths for future study toward defending major cities and buildings that are in danger from the propagation of EWs. Within the context of the significant interest shown in research activities, an opportunity has arisen to investigate the benefits offered by urban trees in the formation of environmental systems that are free from the effects of noise pollution and secure against the dangers posed by earthquakes and other forms of natural calamity.

This paper introduces a metawall composed of a forest, in which each unit has a unique configuration in terms of height, width, lattice, and distance from the source of EW energy. The proposed model investigates the role of the upper part of forest trees in attenuating SWs. It demonstrates the effectiveness of a 3D simulation of metawall with various configurations to assess the effect of foliage to create BGs on dispersion curves of wave propagation. Furthermore, this study examines how the width and position of BGs are influenced by multiple layers of soil, each with its own mechanical properties. The innovation of this work centers around the use of forest trees as natural metamaterials, termed FMs, for attenuating SWs. This paper introduces an innovative concept: using forest trees, arranged in a periodic pattern around a protected area, as a method for seismic wave suppression. It explores how different configurations of these forest metamaterials, including variations in tree height and the addition of mass to simulate foliage, can create band gaps in low-frequency ranges. This approach is novel in its integration of natural elements into engineering solutions for seismic wave attenuation, providing a sustainable and ecologically friendly method to protect urban areas from seismic activity. The main objective of this study is to determine the optimal distribution pattern of forest trees around designated conservation areas for maximum protection and effectiveness.

2 Numerical analysis and design methods

This work advances numerically in simulating SWs to address specific problems. It focuses on the effectiveness of FMs under SW excitation at low frequencies less than 100 Hz. The metawall of FMs represents that resonators interact with SWs in substrate soil; thus, behavior toward attenuating SWs in mechanism is called BG in the field of metamaterials, in which FMM is used as a natural material and the low frequency range is extended to 100 Hz. Through swapping waves via the first irreducible Brillouin zone, the features of BG are realized to attenuate SW propagation.

2.1 Metawall calculation model

This work is limited to the analysis of SWs by using a 3D model. The velocities of primary and secondary waves differ depending on how particles are distributed along the longitudinal and transverse axes. To avoid complexity, when the band structures of the FM resonator are arranged in a periodic pattern, EWs are attenuated. The lattice is considered a function of the wavelength of EW in the soil substrate. Herein, it equals to 50% of λ_ω, where λ_ω is the wavelength of EW in the first BG of BS. The commercial code of COMSOL Multiphysics 5.4a software is valuable to solving the model through the solid mechanics study module. COMSOL Multiphysics 5.4a is also adequate to calculate the corresponding eigenvalues of the unit cell frequency modes of the first irreducible Brillouin zone and swap the periodic arrangement in ΓXMΓ as shown in Figure 1B. In this study, the real growth of a tree species resembling a unit cell is shaped as a circular truncated cone, expressing the radius of the truck at the top equal to the half

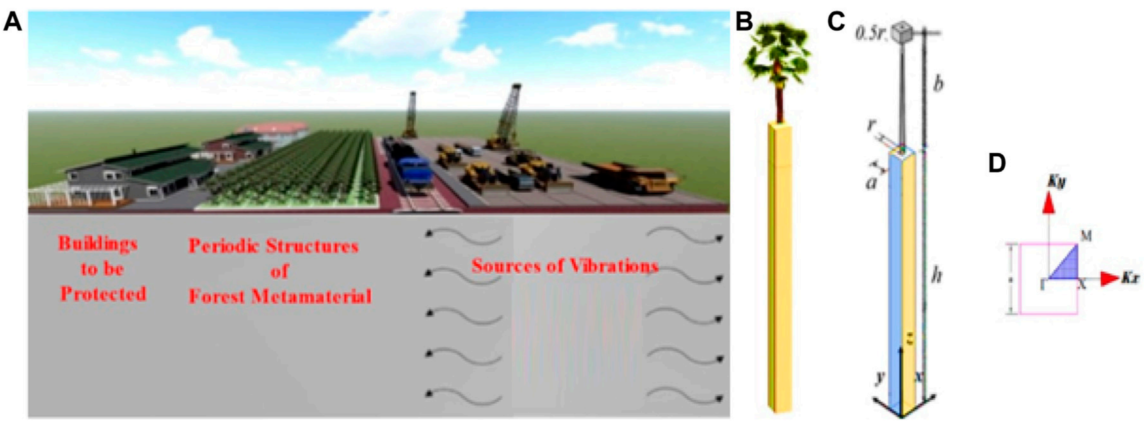


FIGURE 1
Seismic waves and periodic FM barriers, **(B)** Comparison of FM in Nature, **(C)** Tree-soil unit cell, and **(D)** first BZs (hatched blue areas) and first irreducible Brillouin zones (solid line).

TABLE 1 Mechanical properties for the proposed FM (Maurel et al., 2018; Muhammad et al., 2020).

Model	ρ (kg/m³)	E (MPa)	υ
Foliage	200	23.8	0.4
FMM	1,000	2000	0.3
Soil	2,000	200	0.35

value of the bottom radius (r), i.e., r = a/10. A mass-effect model is used to make simulations easy and represent the branches and crowns, measured in meters. The dimensions are considered a percentage of wavelength λ_ω, as demonstrated in Figure 1.

A periodic structure has a similar effect to the atomic periodic potential field on electrons. On the basis of this similarity, this study directly refers to the lattice in physics to describe the periodicity of structures. The concept of this phenomenon extends periodic structures to nanoscale. A periodic structure is represented as "metawall," which can be simplified into a cantilever bending wall on soil as locally vertical resonators, FMs. The effects of tree branches and crowns are considered when establishing the numerical model. The influence of roots on scattering is negligible, on the basis that the roots act as a connecting tool between the soil and the trees. Different configurations of a unit cell are demonstrated to represent the unit cell of urban trees in real images. The mechanical properties of unit cell configurations include the following: ρ is the density of the material, E is the modulus of elasticity, and υ is the Poisson's ratio for the corresponding material, as presented in Table 1.

An inhomogeneous mechanism is used by replacing an equivalent "metawall" filled with an isotropic medium of soil (Marigo and Maurel, 2017). The conforming proposed model of different configurations allows for an analysis to study the influence of many factors, such as the mechanical properties of the unit cell, changing height of trees, lattice shape, distance from wave sources, number of resonator rows, and multilayered soil. A user-controlled adaptive mesh for the unit cell is utilized, ensuring that the maximum element size is less than 1/10 of the wavelength of an ancient wave. Mesh size convergence is performed, and a portion of this analysis is achieved. A maximum element size of 0.2 m and a corresponding triangular finer mesh size are determined as adequate. Convergence plots are provided in Section 3.3 of this study to visually represent the stability and accuracy of the results as the mesh is refined.

2.2 Governing wave equation

The mathematically governing equation for the propagation of waves in a medium is considered the same as that in a linear elastic homogeneous medium (Mei et al., 2003), ignoring the body forces, regardless of whether the damping is expressed by the displacement vector, i.e.,:

$$\frac{E}{2\,(1+v)}\nabla^2 u + \frac{E}{2\,(1+v)\,(1-2v)}\nabla(\nabla.u) = -\rho\omega^2 u \quad (1)$$

The proposed FMs should satisfy periodic boundary conditions based on Bloch theory as:

$$u(r+a) = e^{ik.a}u(r) \quad (2)$$

where r = (r_x, r_y, r_z) refers to the components of location vector matching on the FM boundary, u(r) is the displacement vector of the "metawall" nodes, and a is the lattice constant. Then, through substituting Eq. 1 into Eq. 2, the equation of calculation eigenvalues can be expressed as:

$$(\Omega(K) - \omega^2 M(K)).u = 0 \quad (3)$$

where ω is the angular frequency; K is the wave vector; Ω and M are the stiffness and the mass matrix of the unit cell, respectively, which are functions of the wave vector, K. The wave propagation follows the equilibrium equation in the time domain for an inhomogeneous elastic medium, and the body force term is not included because it has no significant effect (Irschik et al., 1994). Several mathematical schemes and techniques could be used to calculate the swapping

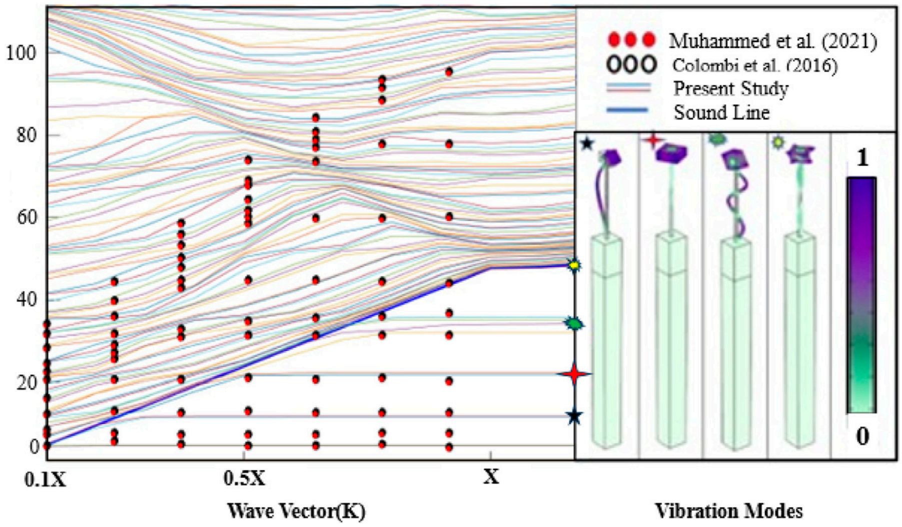


FIGURE 2
Validation of the dispersion curve through case verification with (Colombi et al., 2016a; Muhammad et al., 2020).

vector of elastic surface waves propagating in a given frequency range. To eliminate numerical calculations, the finite element method is used to estimate the solution inside a particular elastic domain as a compilation of basic solutions created by a collection of virtual sources located outside the domain. When the amplitude factor (K) is known, the systems of Eq. 2 can be established to solve the eigenvalue problems adopted via COMSOL Multiphysics 5.4a directly in a finite medium by sweeping the K-vector along the transition (front) sides of the irreducible Brillouin zone under complex boundary conditions. A free surface must be imposed on this situation (to the resonator), whereas a fixed boundary must be imposed on the bottom (u = 0). As a result of sweeping the wave vector, dispersion curves of the total dispersion relation are produced.

2.3 Comparison cases

To verify the reliability of the calculation methods in this paper with those applied in the electromagnetic field and postprocessing methods, two previous research models are selected; both are carried out on a nanoscale and are similar to the model studied in this work. Figure 2 compares the dispersion curve and the calculation model of a similar technique that was reported. The dispersion curve and calculation model depict the band structure of resonant frequencies applied on the top of a semi-infinite substrate soil for model validation by Colombi et al. (Colombi et al., 2016a), of hollow black origin. The sound cone method (SCM) processes the dispersion curve of periodic structures, and the calculated results have good correlation, which can be used to derive the surface wave eigenmodes from the bulk wave modes with excellent agreement. The substrate is composed of sedimentary soil, and the scatterers on the surface are trees. It is also used SCM to process the dispersion curve of their proposed model. The blue solid line in the figure is the sound line, and the light solid lines are the surface wave dispersion curves screened by SCM. The red dots in the figure represent the surface wave mode screened by Muhammad et al. (Muhammad et al., 2020), and the space areas represent the BGs. The figure shows that the computed BG result is compatible with past study results, confirming the accuracy of the current method. The calculation results are compared in the figure with all those obtained from the existing calculation methods. The double line is the difference between the 3D analysis and the 2D ones presented in previous studies.

3 Numerical simulation

3.1 Model shapes under analysis

In this section, the focus is on analyzing wave propagation and mode shapes in the context of EWs interacting with forest metamaterials. By employing numerical simulations, the study examines the combined effects of bulk and surface waves, which are fundamental to EW formation due to subsurface complexities. This section delves into the dispersion characteristics of these waves, highlighting their interactions with the metamaterials. EWs' BG characteristics are predicted via numerical simulations. Various bulk and surface waves combine to form EWs as a result of the subsurface formation's complexity. Seismic surveys can produce EWs when the shear wave velocity of one layer is substantially lower than the velocity of the layer above it. The mode shape analysis is consistent with periodicity, as shown in Figure 3. The metawall unit cell of FMM in square lattice with foliage for compression ignores the foliage effect and includes multilayers of soil.

Bulk and surface wave modes are included in the dispersion curves of the 3D model. SCM is used to differentiate between pure wave modes because of the peculiarities of EW propagation. The surface wave's dispersion curve is shown under the sound line, and the BGs visible in the blue hatched area are its dispersion gaps. Seismic surface wave BGs and bulk wave modes with extremely high frequencies can be found outside the sound cone. Moreover, the

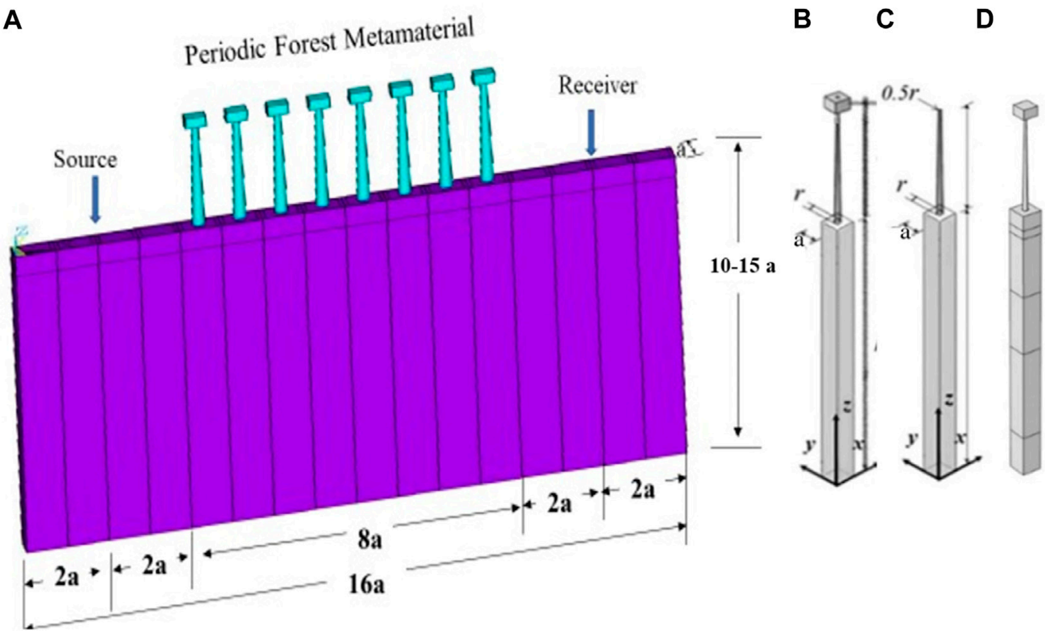


FIGURE 3
Various model shapes **(A)** 3D metawall, **(B)** with foliage, **(C)** without foliage, **(D)** with layered soil.

dispersion properties of elastic waves (EWs) and their interactions with forest metamaterials are explored. The mode shapes relevant to the analysis capture both surface and bulk wave dynamics. Emphasis is placed on modeling the upper sections of forest trees, representing tree branches and biomass as vertical loads to simulate foliage effects. This approach aligns the simulation outcomes with real-world forest configurations, ensuring a realistic representation of wave interactions within varied forest densities.

3.2 Postprocessing physics methods

Bulk waves travel down into deeper soil layers, while surface waves travel along the soil surface (Matthews et al., 1996). Surface waves concentrate their energy primarily on the ground within one or two Rayleigh wavelengths (Pu and Shi, 2017). To achieve surface waves, SCM is used, in which pure surface waves clearly appear below the acoustic line caused by the peculiarities of wave propagation on the surface (Khelif et al., 2010). The sound line constrains the sound cone when it corresponds to the smallest value of the soil phase velocity concept (Huang et al., 2017). Software such as COMSOL Multiphysics 5.4a offers predefined equations to simulate various physical phenomena (Huang and Shi, 2013). COMSOL Multiphysics 5.4a is used to calculate the dispersion curves of periodic systems because it can integrate the formula of eigenvalue and deal with boundary conditions. In this study, the dispersion curves for surface and bulk SWs are calculated using 3D periodic modes with no clear distinction; a locally resonant system is well suited for SCM. In terms of LR and BS, no surface wave mode occurs outside the sound line diagram. Pure surface SW modes are obtained inside the sound line (Khelif et al., 2010). Sound cone lines are calculated in various directions to plot the sound cone diagram, as shown as follows:

$$w = K \times Cs \tag{4}$$

where C_s is the shear wave velocity of the upper soil layer, which can be calculated as:

$$C_s = \sqrt{\frac{\mu}{\rho}} \tag{5}$$

where μ is the shear factor of the upper soil layer, and ρ refers to the soil's bulk density. The infinite periodic structure in the first irreducible Brillouin zone can be calculated using only a unit cell through the similarity of BZs. The EW vector (K) can be demarcated in the first irreducible Brillouin zone only. The unit cell must meet the periodic boundary conditions. Then, COMSOL Multiphysics 5.4a software with complex boundaries is used to solve the eigenvalues for calculating the dispersion curves of the 3D model. This work proposes the use of the energy density method (EDM) to comprehensively study high-frequency surface modes. In accordance with the propagation characteristics of surface EWs in the upper layer of the ground soil and bulk wave propagation in lower layers, the energy of surface waves should be mainly concentrated in 1 or 2 times of the wavelength for EWs on the ground (Huang and Shi, 2015). Hence, the energy dispersal in the 3D model can be observed, and the dispersion curves are verified by the EDM, as presented in the following equation:

$$DOE = \frac{\iiint_{h_t} \frac{1}{2} T_{ij} S_{ij}{}^* dxdydz}{\iiint_{H} \frac{1}{2} T_{ij} S_{ij}{}^* dxdydz} \tag{6}$$

The distribution of energy in Eq. 6 is defined as the energy of EW distribution factor, where T_{ij} represents the tensors of stresses, and

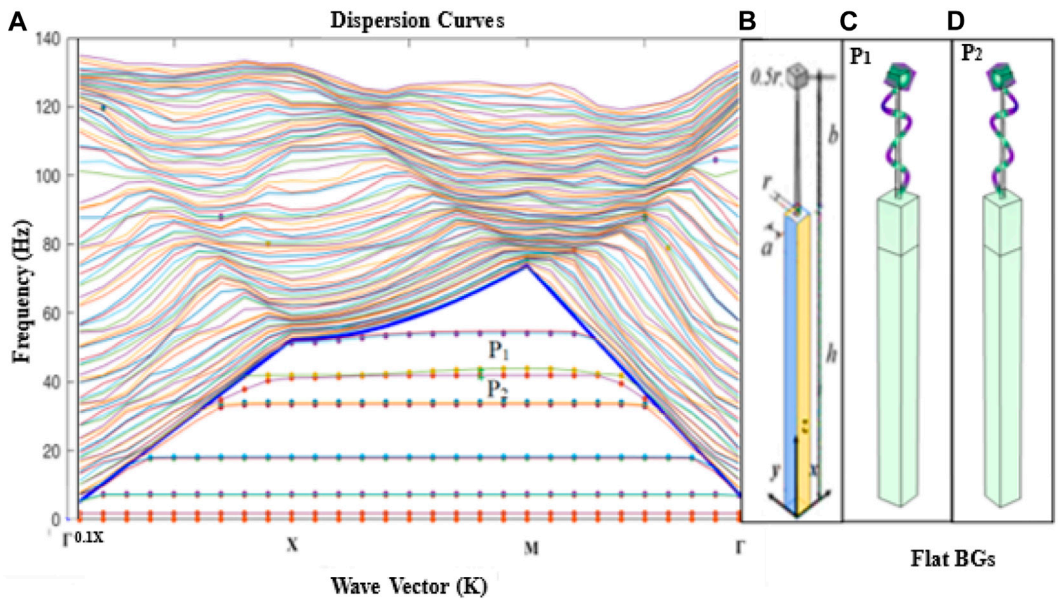


FIGURE 4
(A) dispersion curves in ΓXMΓ, **(B)** unit cell, and **(C, D)** flat bands vibration modes in the same frequency.

S^*_{ij} refers to the conjugate tensors of the strain. The integral limits on the numerators range of sedimentary soil from 0 to h_1 = 2a, measured from the ground surface in the z-axis, and the integral limits on the denominators range of bottom soil from h_1 to h_s, measured through the bottom layers of soil where, h_1 is the sedimentary soil, and h_s is the total height of soil layers; strain energy density is integrated to achieve elastic strain energy (Huang et al., 2017). Therefore, the value of energy distribution factor must change between 0 and 1. A high value of distribution of energy means that the energy on the soil surface (2a) is high. The strain energy generated by vibration is concentrated on the soil surface. Thus, the critical value of distribution of energy can be defined in small values, such as 0.6–0.9. The recognized surface wave mode is produced by the energy concentrated near the ground surface. In accordance with the ratio of the wavelength in BG frequencies to FM dimensions, as well as the lattice constant, BGs can be divided into BS and LR. Generally, these two types of BG formation are the result of the collective properties of periodic structures. No obvious difference exists between LR and BS systems. The BGs produced by these two actions are likely to exist in the proposed system. Although the "metawall" system mainly reflects the resonance characteristics, BGs are still produced by the apparatus of BS in the system. The periodicity of scatterer configurations plays a key role in determining BS BGs; for LR, the greatest role is played by the resonance characteristics of single FM. The SCM approach is mostly applicable to systems with BGs that are obtained by the apparatus of LR; however, it is blind to surface wave modes that are not shear waves. The dispersion curve of the surface wave can be seen below the sound cone; the blue line is the sound line, and the surface wave modes are inside the sound cone. Outside of the sound cone is a continuous area that represents entirely probable bulk wave modes and high frequencies of surface wave BGs. This region could be assumed as an extension of the sound cone. A representative unit cell is used in the calculation of dispersion curves and presented beside the dispersion curve in Figure 4, and a perfect wave-forbidding function is found in an infinite periodic FM. By contrast, practical uses of periodic trees require that they are structured with a limited number of unit cells. The BG generation mechanism is illustrated by calculating low and high BG vibration modes only at boundary of BGs.

According to the results of this calculation, wave mode coupling occurs at the resonance attachment point between the longitudinal patterns of FMs and the vertical motion of incoming waves. This coupling causes SWs to move in phase, and the periodic arrangement of trees acts constructively to provide a large BG. The upper section of FM depicts supplementary branches and/or foliage that appear to improve the coupling strength even further. It is demonstrated a significant degree of vibration reduction within the BG frequency by establishing the boundary conditions of the 3D model. The vibration modes in Figure 4 (a) include points (P_1) and (P_2). Owing to the bending of FMs within SCM, not all FM resonances can produce LR BGs. The vibration mode only differs in the direction of vibration and has the same vibration deformation. Several straight bands can be seen in the dispersion curves of Figure 4. No BG exists between these frequency bands. In the study of metamaterials, such bands are called flat bands. These flat bands are composed of two curves of frequency bands that are located for observation, and the vibration modes of points in the upper and lower frequency bands are selected.

3.3 Transmission spectrum analysis

To absorb propagating EWs, a symmetric load of boundary condition is added to the left and right sides of the model, which is perpendicular to the *y* direction, to simulate the actual mode of FMs. The "metawall" has a width of one unit representing the periodically planted forest on the ANSYS 17.2 finite element software. This model assumes that the forest is planted in a periodic manner, and the model volume is reduced to simplify the calculation. Transmission pertains to the ratio of the transmitted wave energy and the incident wave energy in sound cone theory. The mathematical expression of transmission is:

$$T = 20 \times \log\left(\frac{w_t}{w_i}\right) \tag{7}$$

where w_t and w_i are the transmitted wave energy and incident wave energy, respectively. The amplitude reduction factor (ARF) offers a transmission help in visualizing the attenuation mechanism. ARF is used to evaluate the transmission efficiency of barriers for SW attenuation and defined as follows:

$$ARF = \frac{Amplitude\ of\ substrate\ soil\ within\ barriers}{Amplitude\ of\ substrate\ soil\ throughout\ barriers} \tag{8}$$

With regard to physical properties, such as acceleration and displacement in the z-axis, maximum values are known as amplitudes. Over a certain range, ARF exhibits some degree of freedom. FRF, also known as the transfer function, is defined as the ratio of the complex output amplitude to the complex input amplitude. The frequency response function (FRF) represents the output per unit sinusoidal input at a specific frequency for a sinusoidal input. Measuring the frequency response typically entails stimulating the system with an input signal, determining the resulting output signal, calculating the frequency spectra of the two signals, and comparing the spectra to isolate the system's effect. In this case, the sources are turned on sequentially. It is also found that when the first source excites the FM, FRF can be calculated by dividing the FM response signal by the model without the FM signal. The average value of FRF is also used to describe the efficiency of the attenuation mechanism, and it can be calculated using the following formula:

$$FRF = \frac{1}{S}\int_S 20\log_{10}(ARF)ds \tag{9}$$

where s is the vibration-sensitive range located behind the barriers. The EW in the BG is forbidden from propagation, and the transmission factor will decay in parallel to the normal case of propagation. As a result, metawall arrangement is an excellent indicator of the location and width of the BG that has been created. Furthermore, the transmission changeability indicates the physical mechanism of the BG structure. The plot of dispersion curves and the FRF in the same plot are the best way to find the BGs. To ensure that resonance occurs at the expected frequencies, an evaluation of each configuration is carried out in each mode of the metaforest. At frequencies in which the transmission ratio is less than 1, the total amount of displacement is plotted against frequency to assess how it changes over time. After the meshing of the model divided by a small fine value, such as 0.2 m, in each direction by using ANSYS 17.2, a spring is added in the x, y, and z directions of each node. The viscoelastic spring boundary conditions are added, as shown in Figure 5. Afterward, the absorbed EW is measured smoothly, ensuring the simulation accuracy and reliability.

The analysis of the dynamic responses of "metawall" models is conducted using the ANSYS17.2 finite element software. In this study, the source of the excitation wave is set up at (2a) away from the closest side surface, and the observation (receiver) line is located at (2a) away from the other side. Viscoelastic boundaries are mainly interpreted as artificial boundaries. A damper for absorbing the energy of the wave propagating toward the boundary is utilized, in which the elastic modulus of the spring and the damping constant of the damper are determined by chosen material parameters of the periodic FMs. A comprehensive mesh convergence study was conducted, employing a user-controlled adaptive mesh for the unit cell. This study ensured that the maximum element size was maintained at less than one-tenth of the wavelength of an incident wave, adhering to rigorous accuracy standards. To scientifically quantify mesh convergence, it is incorporated specific metrics, including the percentage of allowable error. The results showed convergence with a maximum element size of 0.2 m and a finer mesh, which was determined to be adequate for our simulations. Through the ANSYS 17.2 parametric design language, when the viscoelastic boundary conditions are applied, the periodic FMs are set perpendicular to the x-axis and the bottom with the z-axis. A frequency-domain analysis is used instead of a time-domain analysis in COMSOL Multiphysics 5.4a because of the enormous computing time required. The x-direction end face of the wall receives a predetermined displacement. The other end face is subjected to a traction-free boundary condition, and the wall's lateral faces are subjected to periodic boundary conditions.

4 Results and discussions

Numerical computation of the incident SW values is performed using extensions of the eigenmode method via COMSOL Multiphysics 5.4a. The 3D model, with the mechanical properties presented in Section 2.1, is adopted. In this case, the effective surface wave velocity of the model is 190.4 m/s. The first BS BG center frequency (f_{m}) is 41.5 Hz. The corresponding wavelength λ_w is:

$$\lambda_w = \frac{C_s}{f_m} = \frac{190.4}{41.5} = 4.57m \tag{10}$$

This wavelength is almost twice the lattice constant, or λ_w = 2a. Accordingly, a = 2.3 m. The geometric parameters used in this case are expressed as a ratio of wavelength, as mentioned in Sec. 2.1. The tree space is a = 2.3 m; thus, the height of trunk h_t = 11.5 m, and the radius of the bottom of the trunk is r = 0.23 m. The soil depth should be sufficiently large compared to the tree spacing (Oudich and Badreddine Assouar, 2012; Xu and Peng, 2015). The accuracy of these calculations has been corroborated by the findings of Liu et al. (Liu et al., 2019). Hence, a unit cell with a height of 30 m (i.e., 10–15 times 'a') was considered reasonable for band gap calculation (Miniaci et al., 2016; Du et al., 2017; Du et al., 2018; Lim and Reddy, 2019). This method allows for an accurate determination of the stress-strain relationship when subjected to seismic wave excitations, offering a deeper understanding of the wave attenuation properties of the metamaterials. The soil, h_{S2} = 25.4 m, for ground height and, h_{S1} = 4.6 m, for the guiding layer. For the starting frequency represented by P_1, the tree and soil exhibit some degree of deformation. The bands are not fully opened, and only BGs in the "TX" region are determined. The best approach to explain the wave reflections induced by FMs is creating a model and calculating the finite size of "metawall".

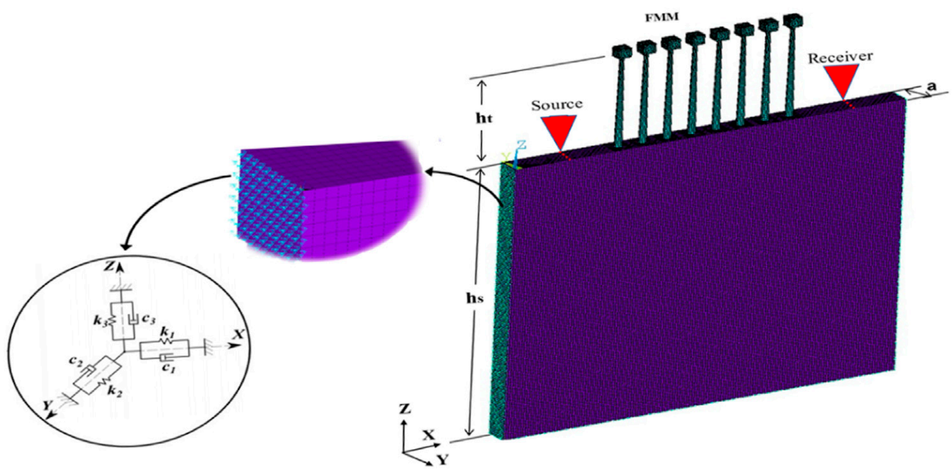


FIGURE 5
Mesh size and viscoelastic boundary conditions for 3D metawall with periodic FMM.

In this study, the FEM is employed to simulate the interaction between seismic waves and forest metamaterials. The FEM allows for detailed analysis of the complex geometries and material properties inherent in the proposed model. It is utilized a discretized model of the physical system, subdividing the forest metamaterials into smaller, manageable finite elements. The results from FEM simulations are instrumental in understanding and optimizing the design and arrangement of forest metamaterials for effective seismic wave attenuation. The simulation modeling is realized via ANSYS 17.2. The viscoelastic boundary elements are accurately modeled to a sufficient level. Boundary finite elements are added to the surfaces parallel to the x direction and the bottom face. To simplify the computation model, periodic boundary conditions are also added to the surface that is orthogonal to the *y* direction. Furthermore, the vibration modes of FMs/resonators are computed at the edges of the BGs, as highlighted in the gray area in Figure 6. The patterns through the points in the scattering curve of the upper and lower edges of the paired BGs show the patterns of the edge bands in all BGs. If the wave primarily propagates along the ground surface and is outside the BG, the EW would pass through the entire model without causing any disruption. Frequencies of 16 and 48 Hz are used as an example of permitting surface waves to pass outside BGs. However, the BG suppresses the EW's ability to propagate, such as at frequencies of 47, 50, and 73 Hz, inside BG frequencies, and both points inside and outside of the BGs. Figure 6A depicts the FRF curves. The FRF's attenuation region closely matches the BGs that could be solved within the sound cone, as observed in the dispersion curves in Figure 6B. The vibration patterns of maximum and minimum low-order longitudinal resonances of the resonator FMs are indicated by dark and light points, respectively, which are highlighted in the vibration modes presented in Figure 6C. Figure 6D displays the displacement distribution for simulating the surface wave feature in the metawall model. The outcomes demonstrate the BG's wave-forbidding property in FMs; at the same time, the earth surface's soil movement is nearly constant. Because the FMs are positioned as a cantilever column in the soil, when the soil shakes on the earth's surface, it indicates the movement of the stand; the FMs vibrate in accordance with this movement, as depicted in Figure 6D. The vibration of trees is a clear curvature movement during the arrival of the wave to the soil's surface and its vibration. The LR BGs below the sound line in the dispersion curves are observed. In consideration of the geometric parameters, the band structure shows the presence of wide LR in the frequency ranges of 30.61–36.12 Hz and 38.47–46.81 Hz separated at a passband around 37.25 Hz as gray highlighted regions. This solution is considered to elucidate the mechanism of BG generation in the BS eigenmodes beyond the sound line (black solid line), indicating highlighted BS (shaded gray regions) from 48.17 Hz to 53.61 Hz and from 68.11 Hz to 74.45 Hz. These gaps dampen the shear wave by about 40%. In general, the symmetric and antisymmetric eigenmodes' characteristics, as well as those of the substrate where the resonators are integrated, cause the longitudinal and curvature modes, which depend on these characteristics to produce BGs.

In the past few years, numerical simulations have been extensively used to determine the effectiveness of periodic structures for the attenuation of EWs and earthquakes in general and to study the factors affecting the attenuation mechanism in particular. Designing periodic structures based on the results of modular analysis is usually a fast, economical, and high-precision method. Therefore, numerical simulations are conducted to examine a group of models of natural metamaterials, arrange them as periodic structures in soil, know the effectiveness of 3D models of these resonators for mitigating the propagation of SWs, and prove the most effective models for attenuating SWs and protecting cities from the dangers of ground vibrations. The mechanism of the attenuation for SW hazard by periodic FMs is applied in this section in consideration of the effects of several factors, including multilayered soil with different properties, "tree–soil" unit cell with various configurations and lattice shapes, FMs' upper part, bulky side branches, and foliage with crown, and metawall with gradually changing trunk height. The numerical simulation of these factors is performed on the basis of the finite element analysis of the 3D model in five cases.

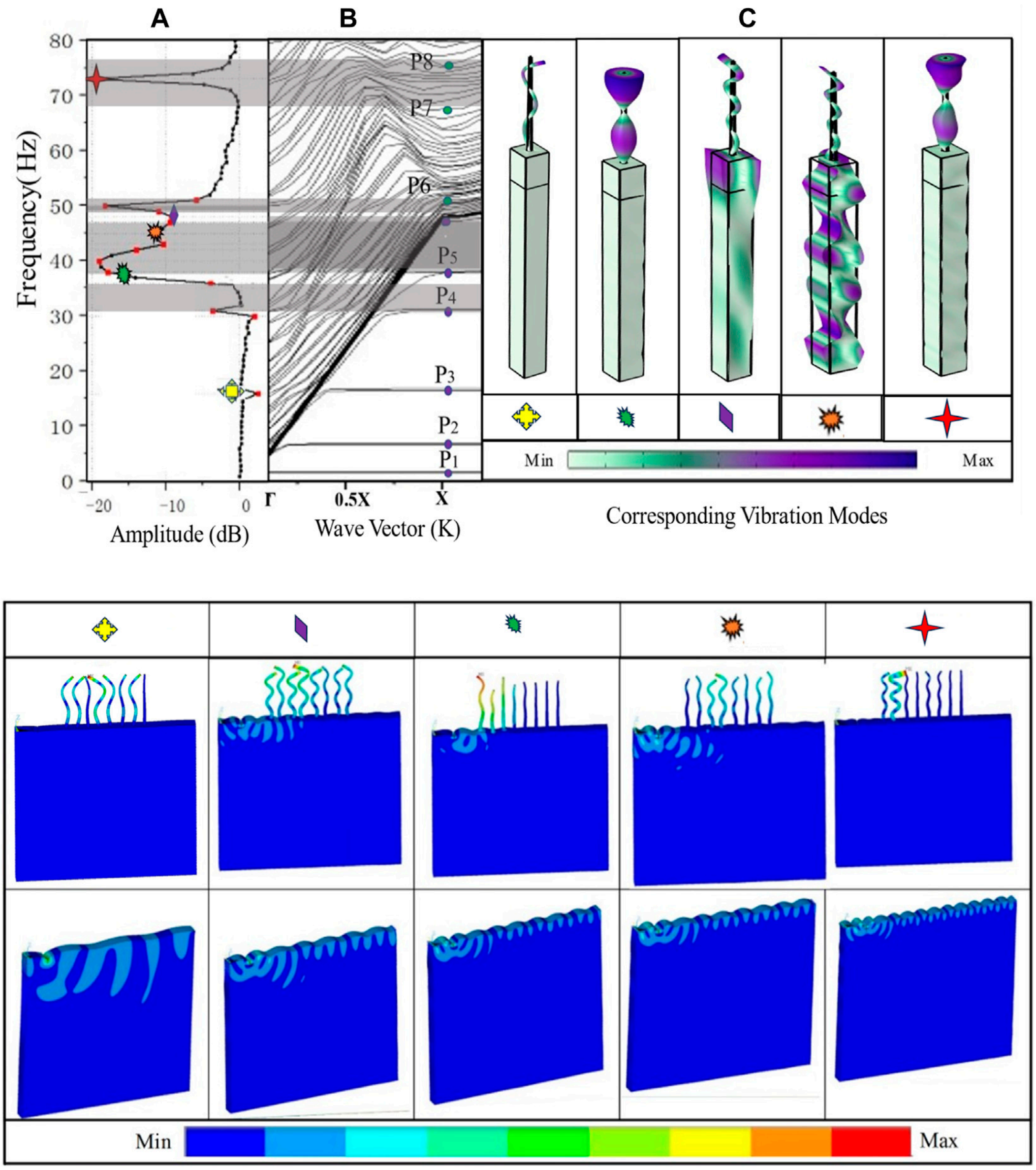


FIGURE 6
3D band structure of FMMs: **(A)** FRF curve, **(B)** dispersion curves in "ΓX," **(C)** vibration modes, **(D)** displacement distribution diagram in the corresponding frequencies.

4.1 Layered soil medium

The simulations in the model of this case are repeated as a transitional step, using soil parameters, via COMSOL Multiphysics 5.4a. Medium dense, dry, uniform sand is chosen as the type of soil because it exhibits the most linear behavior in reality. The assigned soil is still considered as a linear elastic material. On the basis of the assumption that studying and modeling soil layers with inhomogeneity and different consolidated soil characteristics, as well as the universal properties of many layers, are difficult, there may be considerable attenuation in ground vibrations. This condition validates the study's findings and concludes that resonant trees reduce vibrations better than single homogeneous layers of sedimentary soil, confirming the effectiveness of FMs in attenuating SWs in other stratified ground conditions. To make this study universal, layered soils are considered in the model of Figure 7A. The heterogeneity in the properties obtained from *in situ* experiments is accounted for to suit the properties of the soil layers (Pu and Shi, 2018). The dispersion available in multilayer soils comprises different modulus of elasticity, density, and Poisson's ratio compared with that in single-layer soils. The dispersion curves of SWs propagating in various layers of soil return unlike the dispersion curves with different rigidity and density in one-layer soil (Lim and Reddy, 2019). Gao et al. (Gao et al., 2015) conducted cone penetration experiments in the coastal plain of Shanghai City, China and discovered that soils may be classified into six strata on the basis of their mechanical properties (i.e., Young's modulus, Poisson's ratio, and density). The properties of the layers for

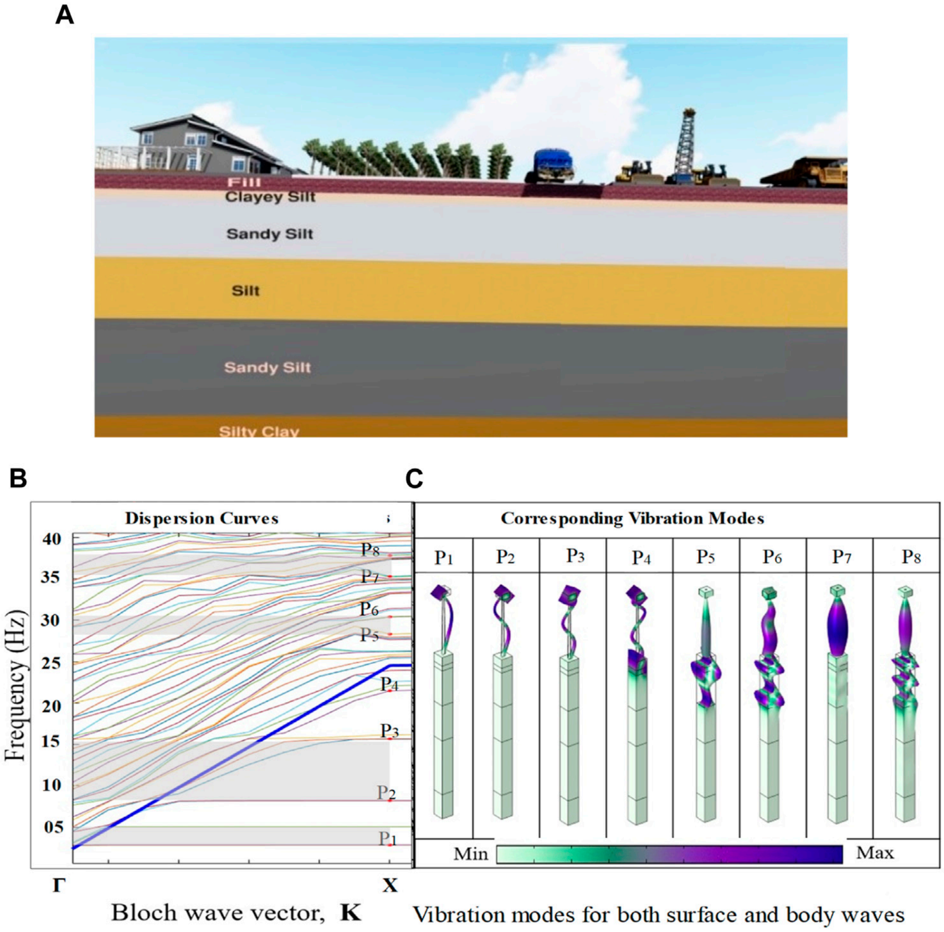


FIGURE 7
FMs on multilayers of soil strata: **(A)** proposed model, **(B)** dispersion curves, and **(C)** vibration modes.

TABLE 2 Properties of layers for simulating a multilayer model (Gao et al., 2015; Lim and Reddy, 2019).

Soil type	ρ (kg/m^3)	E (MPa)	Depth (m)	Height (m)	υ
Fill	1830	37	0–2	2	0.4
Clayey silt	1939	44	2–4	2	0.4
Sandy silt	1888	153	4–9.0	5	0.4
Silt	1898	141	9.0–15.5	6.5	0.45
Sandy silt	1837	90	15.5–25.5	10	0.45
Silty clay	1847	155	25.5–30	4.5	0.45

simulation in this case are indicated in Table 2. The properties of the material effects for single- and multilayer profiles are explored in relation to many geotechnical tests performed *in situ*. Previous studies (Atala et al., 2013; Davis and Hussein, 2014; Dai et al., 2017) have applied perfectly matched layers (PML) at the bottom and both ends of the finite domain model to ensure an infinite boundary and to avoid the back reflection of incident waves. It brief, it is proposed another effective approach by applying the concept of a viscous-spring boundary condition. The fixed boundary is applied at the bottom of the model in COMSOL Multiphysics 5.4a.

The dispersion curves in the direction of ΓX are presented in Figure 7B. The BGs are governed by the coupling of FMs with SWs that travel on the layered soils, as shown in Figure 7B; the vibration modes for surface waves move in this hypothetical periodic system, as illustrated in Figure 7C. The geometric parameters used in this case are expressed as follows: The tree space is a = 2 m, the height of the trunk is h_t = 11.5 m, the radius of the bottom of the trunk is r = 0.3 m, and the dimension of the foliage volume is 1.5 m × 1.5 m × 1 m. The properties of the foliage and the forest metamaterials (FMs) are consistent with those outlined in Table 1, Section 2.1.

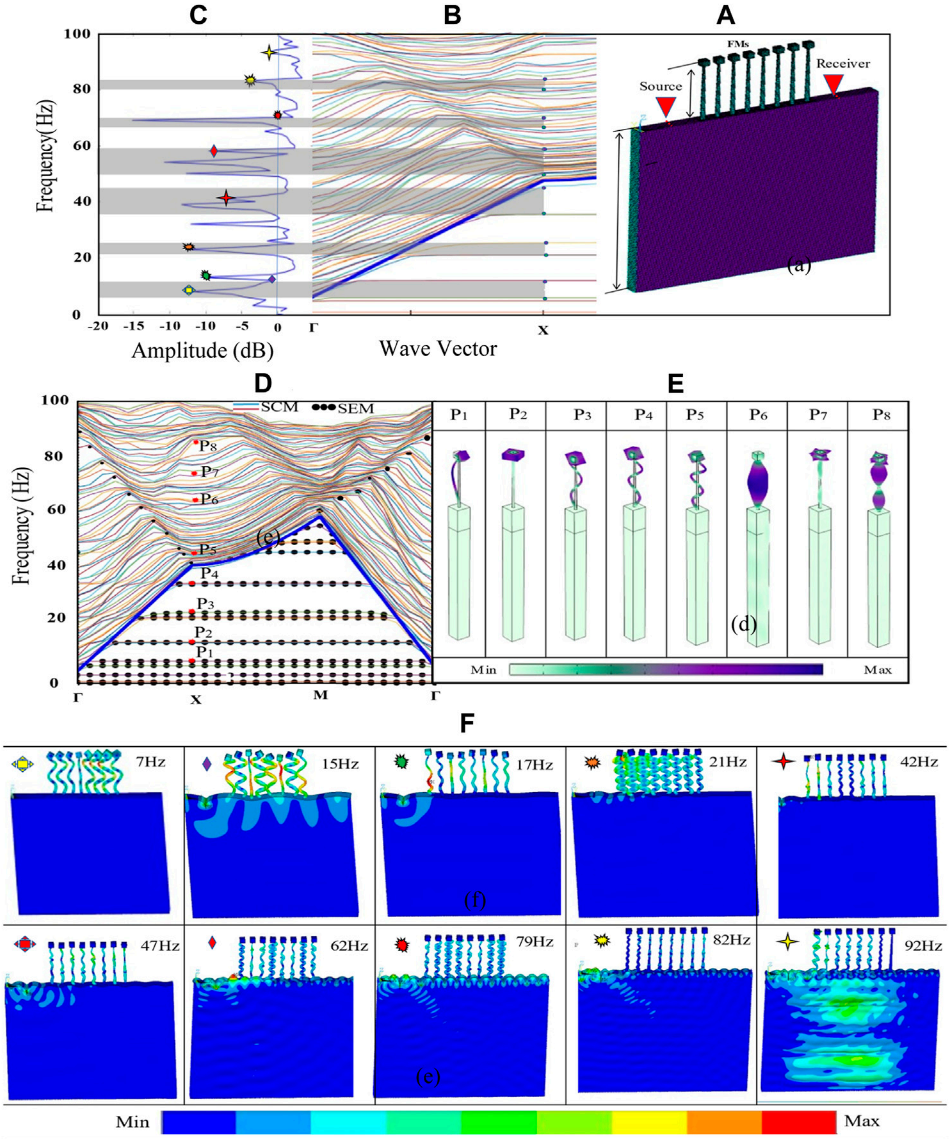


FIGURE 8
(A) Proposed model, **(B)** dispersion curve, **(C)** FRF plot, **(D)** dispersion curve, **(E)** vibration modes, and **(F)** distribution displacement.

Table 2 primarily focuses on detailing the characteristics of the soil layers.

The resonant modes of the infinitely periodic resonator act in a way that helps produce low-frequency BGs when periodicity conditions are applied. When the soil medium is replaced with layered soil and the geometric characteristics of this instance are maintained, the resonance of LR depends on the resistance misalliance and coupling of SW modes between the FMs and soil. The increase in the width of BG is attributed to the interaction with the resonant FMs. Because the velocity patterns of SWs change in each layer of a homogeneous medium, surface impulses travel at amplitudes and phase velocities. The ground effect results in deductive and eradicative intervention, causing a rise and minimizing the BG width and having an effective influence on the attenuation mechanism. For acoustically rigid and stiff surfaces, the changes in the magnitude of elastic modulus and in relative bulk density led to a decay in the attenuation ratio, thus improving the effectiveness of periodic barriers. For spongy soil surfaces, such as clay, sandy, forestay sand, and gray sandy clay, ground actions can result in relevant low frequencies. At high frequencies, soft ground can absorb SWs, so the resonance's amplitude is changed. The attenuation mechanism is determined clearly in the vibration modes of these models, as shown in Figure 7C. The medium's dispersion relation classifies waves depending on their component speeds. The sound cone technique can differentiate between surface and body waves because the surface wave speed is lower than the body wave speed. SCM is used to depict the BG for plane waves, and BG positions are marked as gray regions. According to Figure 7, the model's dispersion spectra exhibit three locally resonant BGs at frequencies of 2.55–5.04 Hz, 7.194–15.07 Hz, and 16.68–22.61 Hz.

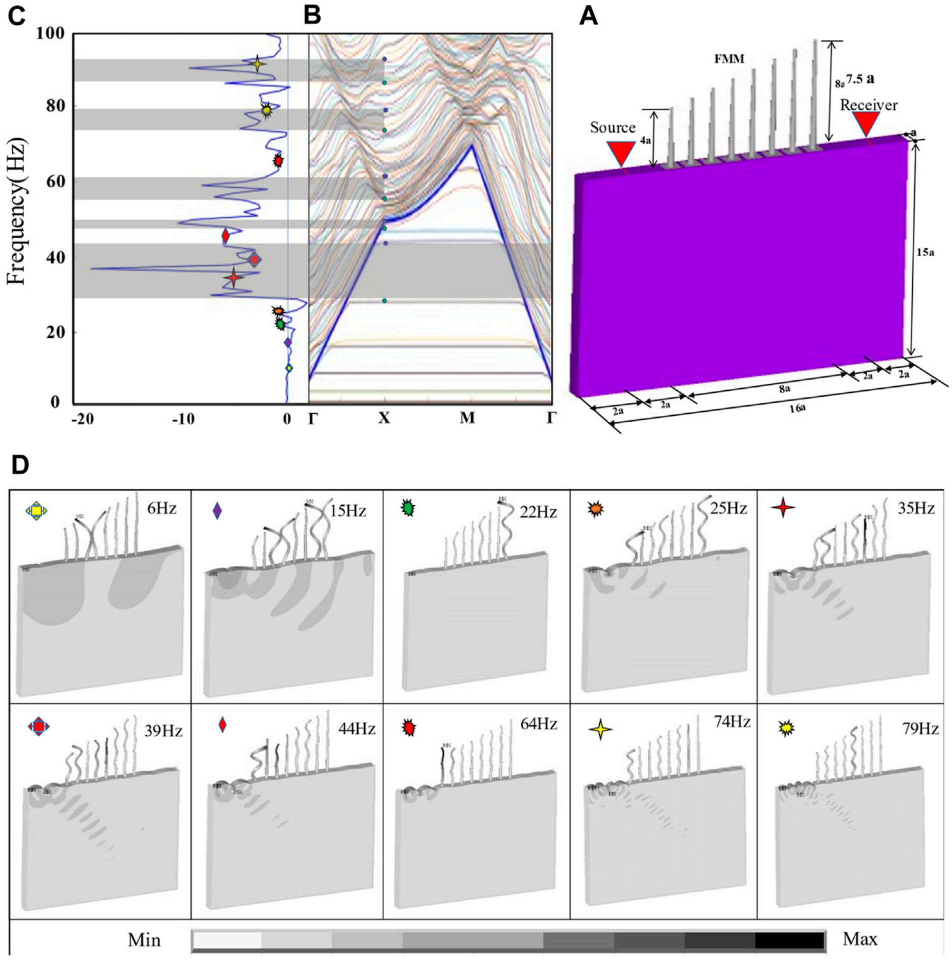


FIGURE 9
(A) Proposed model, **(B)** dispersion curve in the ΓX direction, **(C)** FRF curve, **(D)** distribution displacement corresponding.

Wide low-frequency BGs are achieved as a result of LR BGs developing because of the increased impedance mismatch between soil strata and built-up FMs. Meanwhile, the layered soil model shows other BGs on the outside of the sound line; those bands are BS BGs in the range of frequencies of 28.04–30.92 Hz and 36.08–37.54 Hz. BGs using SCM when applying the boundary condition refer to the radiation of the slowest bulk wave.

4.2 Characteristic of the upper part of trees

Owing to the unstable, irregular movement of trees' leaves caused by wind and the disparity in the shapes and sizes of foliage and secondary branches, they differ from one species of tree to another (Hong et al., 2018). Thus, realizing a geometric simulation model with actual shapes and sizes in the engineering software used for analysis is a very complex matter. In this study, the consideration of a cube block shape with mass added at the top of the resonator FMs is valuable to investigating the real configuration of tree foliage to study their contribution to the attenuation mechanism and verify the changes in BG width and position (Maurel et al., 2018). FMs interact with EWs, resulting in a wide BG in the low frequencies of interest (Liu et al., 2019). The movement of the lateral branches of trees during the occurrence of vibrations is in different directions, which in turn positively affects the ability to damp the vibrations (Lim, 2021). Hence, the effect of side branches connected to the trunk of trees should be considered to fully understand this mechanism. This case is evaluated using a simulation approach with the model including vertical load instead of the upper part of foliage on the top of the trunk (i.e., equal to 5 times the lattice constant in this study). The FMs are arranged at a distance (2a) from the source of vibration and the receiver at the same distance (2a) from the latest periodic FM. The lattice transformation constant in this case is 2.3 m center to center of FMs, as shown in Figure 8A. Dispersion relations of moving EWs in the X direction, as well as FRF curves, clearly demonstrate the FM BGs in the LR and BS BGs, as depicted in Figure 8B, C, respectively.

The LR BGs are highlighted under the sound line. The exact ranges are 7.32–13.71 Hz, 21.18–26.32 Hz, and 37.17–46.62 Hz. In the actual process using the COMSOL Multiphysics 5.4a software,

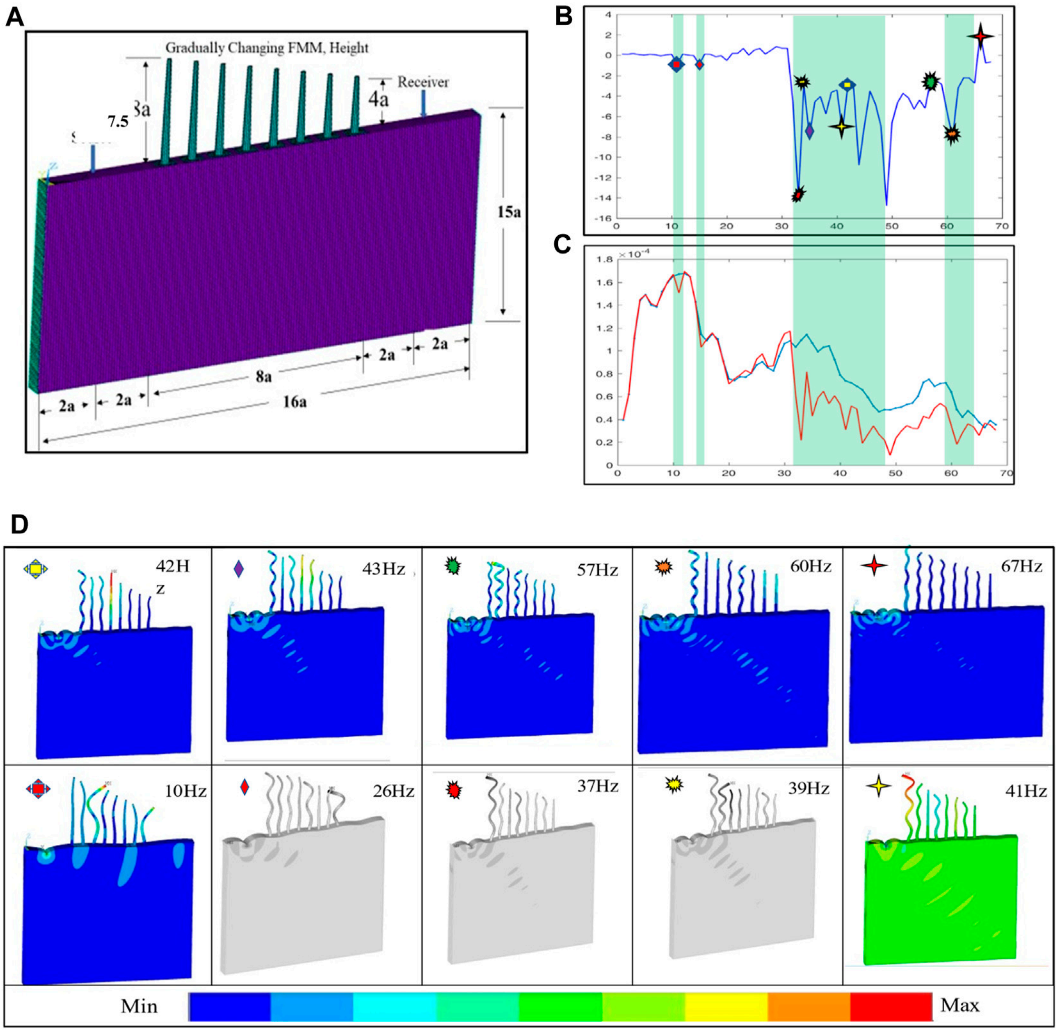


FIGURE 10
(A) Proposed model, **(B)** FRF curve, **(C)** comparison of the displacement distribution plot for the "metawall" model with (red line) and without (blue line) FMs, and **(D)** displacement distribution plot.

the vibration modes plotted in Figure 8D are integrated into EDM directly to achieve the related strain energy black dots in Figure 8E. The dispersion curves for moving surface waves in ΓXMΓ direction in the FMs with foliage in the internal curves of SCM are shown in the same figure, which are completely consistent with the elastic surface wave BGs determined in the figure. Some of the high-frequency surface wave modes are highlighted in gray regions. The simulation model also identifies some BS BGs for surface waves beyond the sound line in the ranges of 51.12–58.60 Hz, 74.22–79.24 Hz, and 80.06–84.24 Hz. The calculation methods in this case involve surface and bulk wave modes, revealing the changes in wave propagation in the upper part of forest trees. Therefore, through the presented intensive analysis and the calculation, it is determined the maximum percentage of the surface wave attenuation on soil and save the first 100 frequency bands. Furthermore, as shown in the simulation models, the effect of foliage appears not only for low bands of tens of hertz but also for the BGs at frequencies higher than 50 Hz.

The modes of vibration for unit cells indicate noticeable changes in vibration modes at points from P_1 to P_8 after damping the surface propagation of EW through the upper part of FMs. In the case of a foliage model, the increase in dimensions of foliage BG width declines. This case presents a detailed investigation into the effect of tree foliage and branches on ground vibration attenuation, particularly at low frequencies below 20 Hz. It demonstrates that these natural elements significantly contribute to the creation of Rayleigh wave band gaps, which are instrumental in reducing ground-borne vibrations. This finding emphasizes the potential of incorporating natural structures into seismic metamaterial design, offering ecological and effective solutions for vibration control. The inclusion of foliage, in particular, introduces vital band gaps at lower frequencies, enhancing the overall efficiency of vibration mitigation strategies. That is, the coupling of FMs and soil deteriorates with a negative effect, which means that the foliage can induce further noise when the incident wave interacts with the structure, especially in wind.

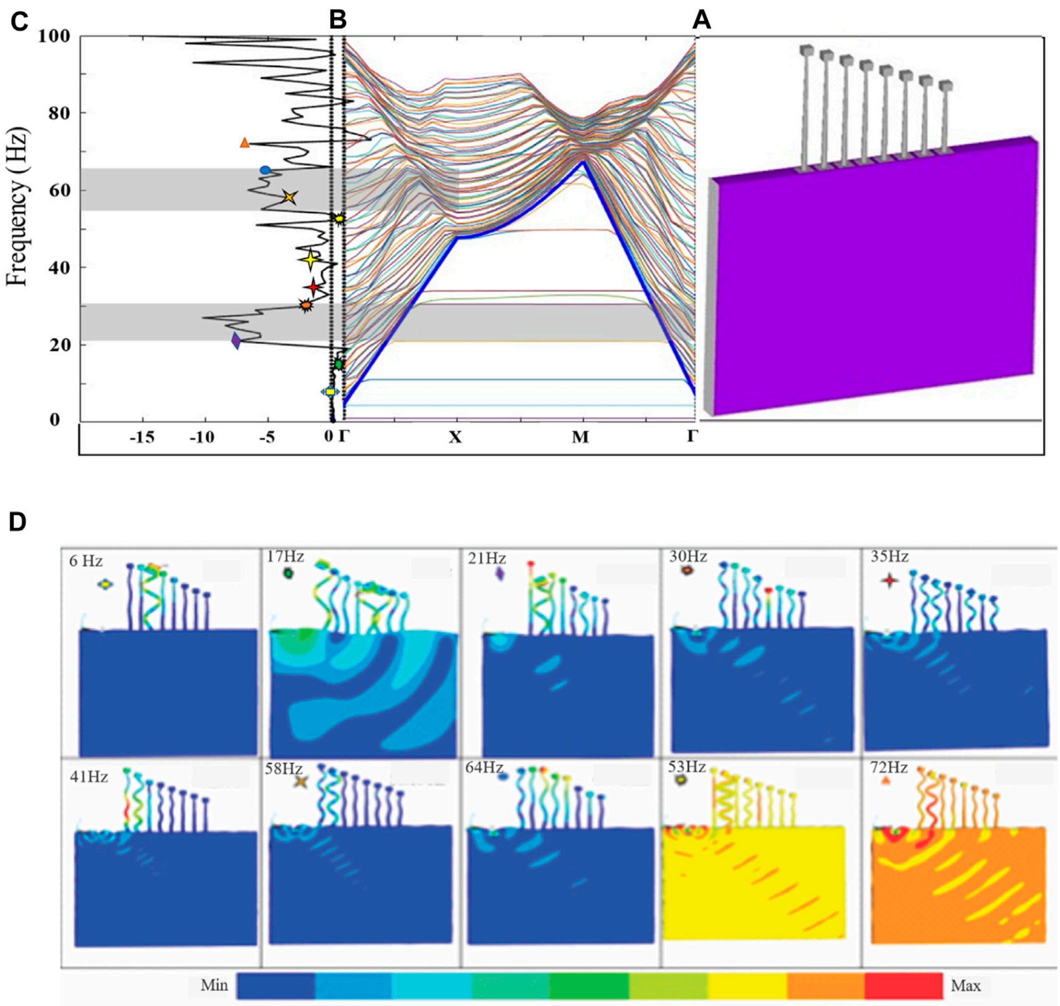


FIGURE 11
(A) Metawall model, **(B)** dispersion curves, **(C)** FRF plots, **(D)** total displacement distribution plots inside and outside BGs in the z direction.

4.3 Gradually increasing FM height

In addition to the considerable influence of layered soil and treetops, another interesting feature appears, which may significantly affect the propagation of directed waves through forestation of FMs by increasing tree heights. In this case, FMs are arranged in a special periodic model, with an increase in FMs' height from 4 times to 8 times the lattice constant. The FM parameters are as follows: the lattice is a = 2 m, the trunk radius at bottom is r = 0.3 m, the trunk radius at top is 0.5 r, and height changes from 8 m beside the wave sources to 15 m beside the recording point. The soil contains two layers with 4 and 26 m heights from up to down with the same parameters provided on Table 1. The metawall model includes eight rows of FMs, as shown in Figure 9A. The mechanical properties of the materials are listed in Section 2.1.

As the height of FM increases, it interacts with the inducing surface wave, generating extremely wide BGs in the low frequency range of interest below 100 Hz. To determine the variant restrictions of the BGs in the 3D models of FMs, the dispersion curves based on SCM and EDM methods typically establish the BGs' characteristics. The gradual change in FMs' height effects results in deductive and eradicative intervention, causing a rise and minimizing the SW attenuation. The sound line divides the dispersion relations of FMs for medium heights into LR and BS BGs. The unit cell motion shows that the vibration is related to FM oscillations. The FMs show a large sinusoidal distortion in LR BGs in the ranges of 24.15–43.86 Hz and 47.18–51.62 Hz and a strong surface distortion, almost as strong as the ground movement of the surface wave below the sound line. On the contrary, such distortion is noticeably attenuated beyond the sound line to generate some of the BS BGs of periodic FMs in many ranges, such as the ranges of 56.21–61.46 and 64.4–79.12 Hz. The analysis finding in the case of gradually increasing FM height results in extremely wide BGs for LR and BS BGs. Although BGs do not appear at low frequencies below 20 Hz, the BGs produced by BS correspond to surface waves converting into shear waves in the deep layers of soil.

4.4 Gradually decreasing tree height

This case includes a numerical analysis of a model with varying heights of FMs and considers the cutoff frequency for the propagating wave with decreasing FM height from $h_t = 15$ m toward shorter trees, $h_t = 8$ m. The other parameters used in this case are the same as those presented in case (III). The analysis in this case is suitable to illustrate surface wave propagation along FMs with decreasing stem height, which becomes increasingly adapted to being converted into shear wave in deeper soil layers. The calculation methods here are limited to explaining the effect of gradually reducing FM height. The results show that when a tree reaches a certain height, coupled with the resonant frequency of the vector, it is converted into a common wave, not adapted to pass the cutoff frequency and thus simply reflected back to the lower layers of soil. This strongly asymmetric behavior of surface love waves was shown in a previous study by Maurel et al. (Maurel et al., 2018). Surface wave scattering and shifting are also observed in the "metawall" model. It is considered a point source outside a forest of eight trees, which generates a love spell in the surface layer of soil. FRF clearly shows the BGs of the proposed models to verify the effect of tree height change on the BGs, as shown in Figure 10B. The displacement distribution diagram in Figure 10D at points selected from the FRF shows the wave transformation inside and outside of BGs. The incident wave causes this coupling, and the periodic formation of trees acts constructively to produce a wide BG. The results show a strong attenuation of ground vibration because the first extremally large BG starts from 30.25 Hz to 47.5 Hz, expressed as LR BGs. The establishment of the bulk wave damping in the BS BGs is visible in the frequency range of 57.14–65.43 Hz.

4.5 Gradually decreasing foliage height

In this case, the main concern is understanding how the attenuation effectiveness changes as surface waves move in the ΓXMΓ direction. As shown in Figure 11, on the top of the trees, the model includes a foliage effect as a vertical block load with dimensions of 1.5 m × 1.5 m × 1 m. Eight rows of forest trees are arranged, in which the height of FMs changes gradually from 15 m to 8 m. The other parameters of FMs are the same as in case (IV). Engineering FMs with altered stem height and foliage at the same time as actual tree planting has another effect. While foliage has a negative effect on increasing noise and low-frequency vibrations, it plays an important role in converting surface waves into bulk waves and generating wide BGs for LR and BS. When the foliage is presented with high trunk FMs, the reflection of Rayleigh wave remains almost constant throughout the considered high values. The coupling of strong SW modes creates great impedance through matching the resonance between FMs and the foliage fixed in the soil. As a result of this matching, BGs are generated at 22–31 Hz for LR BGs and 56–64 Hz for BS BGs. The negative effect of foliage disappears in the range of frequencies lower than 80 Hz, as depicted in Figure 11. The change in the mechanical properties of soil and FMs with foliage indicates the effects on the strength of wave mode coupling. Similarly, the band structure for foliage models achieves a strong response to absorb wave energy.

5 Conclusion

This work concluded that FMs acted as a harvesting device for SWs through wide BGs, as evidenced by the hybridization phenomenon in numerical simulation. To this end, surface wave attenuation using FMs has been studied intensively on the basis of periodicity theory. The effects of various parameters, such as layered soil, foliage, and gradually changing tree height, have been comprehensively discussed. Conclusions are summarized as follows:

- The surface waves at low frequencies below 100 Hz are refracted and deflected in the unit cell and converted into shear waves, as proven through numerical simulations of different configurations of FMs in this work.
- The changes in BGs' characteristics due to foliage are discernible. The results of additional characteristics are revealed in the case of the response of foliage in the low-frequency range of SWs, in which the first BGs start with small frequencies less than 10 Hz and show strong attenuation of the propagation of SWs. However, their effect becomes negative during wind waves, which causes further noise.
- According to the findings of this study, the wave propagation in layered soil results in refracted seismic surface waves at frequencies less than 50 Hz. Thus, the layered soil is believed to be the optimal damper median for surface waves because the waves are reflected multiple times as they travel in the various layers of soil.
- The gradual change in FMs' height results in the rising the starting point of BGs, but it also generates wide BGs for LR and BS BGs.
- The local amplification of ground vibration control by FMs can be directly applied to the improvement of urban ecosystem functions and has extremely important application potential in urban ecological construction.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

QA-S: Conceptualization, Methodology, Writing–review and editing, Writing–original draft, Investigation. JH: Conceptualization, Methodology, Data curation, Writing–review and editing, Supervision, Project administration. MA: Writing–review and editing, Formal Analysis, Resources, Supervision, Project administration, Validation, Funding acquisition. SM: Formal Analysis, Resources, Writing–review and editing, Validation. AA: Conceptualization, Methodology, Formal Analysis, Writing–review and editing, Validation, MA-H: Writing–review and editing, Resources. YG: Writing–review and

editing, Funding acquisition, Formal Analysis, Resources. HA: Formal Analysis, Writing–review and editing, Software, Visualization, Validation.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).

Acknowledgments

Authors gratefully acknowledge the Beijing Municipal Education Commission for their financial support through the Innovative Transdisciplinary Program "Ecological Restoration Engineering" and Peiyang Future Scholar Scholarship. The authors also appreciate the financial support given by the Islamic University of Madinah, Saudi Arabia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

An, N. M. N. (2022). Mitigation of ground vibration induced by high-speed trains using a periodic pile system 2022. *Geotextiles and Geomembranes*, 46,

Aravantinos-Zafiris, N., and Sigalas, M. M. (2015). Large scale phononic metamaterials for seismic isolation. *J. Appl. Phys.* 118, 64901. doi:10.1063/1.4928405

Atala, M., Aidelsburger, M., Barreiro, J. T., Abanin, D., Kitagawa, T., Demler, E., et al. (2013). Direct measurement of the zak phase in topological Bloch bands. *Nat. Phys.* 9, 795–800. doi:10.1038/nphys2790

Audusse, E., Steinstraesser, J. G. C., Emerald, L., Heinrich, P., Paris, A., and Parisot, M. (2021). Comparison of models for the simulation of landslide generated tsunamis. *ESAIM Proc. Surv.* 70, 14–30. doi:10.1051/proc/202107002

Azzouz, L. (2021). *Development and characterisation of novel biocomposites fabricated using natural fibres and rapid prototyping Technology*. University of Hertfordshire, Hatfield, UK,

Barbuto, M. (2015). *Design and implementation of metamaterial-inspired transmissive and radiating microwave components*. University of Hertfordshire, Hatfield, UK,

Brûlé, S., Javelaud, E. H., Enoch, S., and Guenneau, S. (2014). Experiments on seismic metamaterials: molding surface waves. *Phys. Rev. Lett.* 112, 133901. doi:10.1103/physrevlett.112.133901

Colombi, A., Ageeva, V., Smith, R. J., Clare, A., Patel, R., Clark, M., et al. (2017). Enhanced sensing and conversion of ultrasonic Rayleigh waves by elastic metasurfaces. *Sci. Rep.* 7, 6750–6759. doi:10.1038/s41598-017-07151-6

Colombi, A., Colquitt, D., Roux, P., Guenneau, S., and Craster, R. V. (2016b). A seismic metamaterial: the resonant metawedge. *Sci. Rep.* 6, 27717–27726. doi:10.1038/srep27717

Colombi, A., Roux, P., Guenneau, S., Gueguen, P., and Craster, R. V. (2016a). Forests as a natural seismic metamaterial: Rayleigh wave bandgaps induced by local resonances. *Sci. Rep.* 6, 19238–19247. doi:10.1038/srep19238

Colombi, A., Zaccherini, R., Aguzzi, G., Palermo, A., and Chatzi, E. (2020). Mitigation of seismic waves: metabarriers and metafoundations bench tested. *J. Sound. Vib.* 485, 115537. doi:10.1016/j.jsv.2020.115537

Colquitt, D. J., Colombi, A., Craster, R. V., Roux, P., and Guenneau, S. R. L. (2017). Seismic metasurfaces: sub-wavelength resonators and Rayleigh wave interaction. *J. Mech. Phys. Solids* 99, 379–393. doi:10.1016/j.jmps.2016.12.004

Cook, D. I., and Van Haverbeke, D. F. (1974). Tree-covered land-forms for noise control; forest service. *U. S. Dep. Agric.* 263.

Dai, H., Liu, T., Jiao, J., Xia, B., and Yu, D. (2017). Double Dirac cone in two-dimensional phononic crystals beyond circular cells. *J. Appl. Phys.* 121. doi:10.1063/1.4979852

Daradkeh, A. M., Hojat Jalali, H., and Seylabi, E. (2022). Mitigation of seismic waves using graded broadband metamaterial. *J. Appl. Phys.* 132, 54902. doi:10.1063/5.0089242

Davis, B. L., and Hussein, M. I. (2014). Nanophononic metamaterial: thermal conductivity reduction by local resonance. *Phys. Rev. Lett.* 112, 055505. doi:10.1103/physrevlett.112.055505

Dijckmans, A., Coulier, P., Jiang, J., Toward, M. G. R., Thompson, D. J., Degrande, G., et al. (2015). Mitigation of railway induced ground vibration by heavy masses next to the track. *Soil Dyn. Earthq. Eng.* 75, 158–170. doi:10.1016/j.soildyn.2015.04.003

Dong, Q. (2021). *Acoustic-structural interaction: understanding and application in sensor development and metamaterials*. Tokyo, Japan: Temple University libraries. doi:10.34944/dspace/6823

Du, Q., Zeng, Y., Huang, G., and Yang, H. (2017). Elastic metamaterial-based seismic shield for both lamb and surface waves. *AIP Adv.* 7, 75015. doi:10.1063/1.4996716

Du, Q., Zeng, Y., Xu, Y., Yang, H., and Zeng, Z. (2018). H-fractal seismic metamaterial with broadband low-frequency bandgaps. *J. Phys. D. Appl. Phys.* 51, 105104. doi:10.1088/1361-6463/aaaac0

Gahlmann, T., and Tassin, P. (2022). Deep neural networks for the prediction of the optical properties and the free-form inverse design of metamaterials. https://arxiv.org/abs/2201.10387,

Gao, G., Li, N., and Gu, X. (2015). Field experiment and numerical study on active vibration isolation by horizontal blocks in layered ground under vertical loading. *Soil Dyn. Earthq. Eng.* 69, 251–261. doi:10.1016/j.soildyn.2014.11.006

Garg, N. (2022). "Environmental control strategies," in *Environmental noise control* (Springer), Berlin, Germany, 277–344.

He, C., Zhou, S., Li, X., Di, H., and Zhang, X. (2023). Forest trees as a natural metamaterial for surface wave attenuation in stratified soils. *Constr. Build. Mater.* 363, 129769. doi:10.1016/j.conbuildmat.2022.129769

Hong, S.-W., Zhao, L., and Zhu, H. (2018). CFD simulation of airflow inside tree canopies discharged from air-assisted sprayers. *Comput. Electron. Agric.* 149, 121–132. doi:10.1016/j.compag.2017.07.011

Huang, J., Liu, W., and Shi, Z. (2017). Surface-wave attenuation zone of layered periodic structures and feasible application in ground vibration reduction. *Constr. Build. Mater.* 141, 1–11. doi:10.1016/j.conbuildmat.2017.02.153

Huang, J., Liu, Y., and Li, Y. (2019). Trees as large-scale natural phononic crystals: simulation and experimental verification. *Int. Soil Water Conserv. Res.* 7, 196–202. doi:10.1016/j.iswcr.2019.03.004

Huang, J., and Shi, Z. (2013). Application of periodic theory to rows of piles for horizontal vibration attenuation. *Int. J. Geomech.* 10, 132–142. doi:10.1061/(asce)gm.1943-5622.0000193

Huang, J., and Shi, Z. (2015). Vibration reduction of plane waves using periodic in-filled pile barriers. *J. Geotech. Geoenvironmental Eng.* 141, 4015018. doi:10.1061/(asce)gt.1943-5606.0001301

Huang, T. T., Ren, X., Zeng, Y., Zhang, Y., Luo, C., Zhang, X. Y., et al. (2021). Based on auxetic foam: a novel type of seismic metamaterial for lamb waves. *Eng. Struct.* 246, 112976. doi:10.1016/j.engstruct.2021.112976

Irschik, H., Bedford, A., and Drumheller, D. S. (1994). *Introduction to elastic wave propagation*. Hoboken, NJ, USA: John Wiley and Sons.

Jia, G., and Shi, Z. (2010). A new seismic isolation system and its feasibility study. *Earthq. Eng. Eng. Vib.* 9, 75–82. doi:10.1007/s11803-010-8159-8

Jiang, J., Toward, M. G. R., Dijckmans, A., Thompson, D. J., Degrande, G., and Lombaert, G., (2015). "Reducing railway induced ground-borne vibration by using trenches and buried soft barriers," in *Noise and vibration mitigation for rail transportation systems* (Springer), Berlin, Germany, 555–562.

Kaushik, M. V. (2023). Negative index materials: metamaterials. *Review Feature*, 9,

Khelif, A., Achaoui, Y., Benchabane, S., Laude, V., and Aoubiza, B. (2010). Locally resonant surface acoustic wave band gaps in a two-dimensional phononic crystal of pillars on a surface. *Phys. Rev. B* 81, 214303. doi:10.1103/physrevb.81.214303

Kim, S.-H. (2012). Seismic wave attenuator made of acoustic metamaterials. *J. Acoust. Soc. Am.* 131, 3292. doi:10.1121/1.4708309

Li, B., Zhang, C., Peng, F., Wang, W., Vogt, B. D., and Tan, K. T. (2021). 4D printed shape memory metamaterial for vibration bandgap switching and active elastic-wave guiding. *J. Mater. Chem. C* 9, 1164–1173. doi:10.1039/d0tc04999a

Li, G., Chen, Y., Chen, W., Liu, J., and He, H. (2022a). Local resonance–helmholtz lattices with simultaneous solid-borne elastic waves and air-borne sound waves attenuation performance. *Appl. Acoust.* 186, 108450. doi:10.1016/j.apacoust.2021.108450

Li, S., Qiu, C., Huang, J., Guo, X., Hu, Y., Mugahed, A.-S. Q., et al. (2022b). Stability analysis of a high-steep dump slope under different rainfall conditions. *Sustainability* 14, 11148. doi:10.3390/su141811148

Lim, C. W. (2019). Elastic waves propagation in thin plate metamaterials and evidence of low frequency pseudo and local resonance bandgaps. *Phys. Lett. A* 383, 2789–2796. doi:10.1016/j.physleta.2019.05.039

Lim, C. W. (2021). Natural seismic metamaterials: the role of tree branches in the birth of Rayleigh wave bandgap for ground born vibration attenuation. *Trees* 35, 1299–1315. doi:10.1007/s00468-021-02117-8

Lim, C. W., and Reddy, J. N. (2019). Built-up structural steel sections as seismic metamaterials for surface wave attenuation with low frequency wide bandgap in layered soil medium. *Eng. Struct.* 188, 440–451. doi:10.1016/j.engstruct.2019.03.046

Liu, Y., Huang, J., Li, Y., and Shi, Z. (2019). Trees as large-scale natural metamaterials for low-frequency vibration reduction. *Constr. Build. Mater.* 199, 737–745. doi:10.1016/j.conbuildmat.2018.12.062

Lombi, E., and Susini, J. (2009). Synchrotron-based techniques for plant and soil science: opportunities, challenges and future perspectives. *Plant Soil* 320, 1–35. doi:10.1007/s11104-008-9876-x

Lott, M., Roux, P., Garambois, S., Guéguen, P., and Colombi, A. (2020). Evidence of metamaterial physics at the geophysics scale: the METAFORET experiment. *Geophys. J. Int.* 220, 1330–1339. doi:10.1093/gji/ggz528

Mandal, P., and Somala, S. N. (2020). Periodic pile-soil system as a barrier for seismic surface waves. *SN Appl. Sci.* 2, 1184–1188. doi:10.1007/s42452-020-2969-8

Marazzani, J., Cavalagli, N., and Gusella, V. (2021). Elastic properties estimation of masonry walls through the propagation of elastic waves: an experimental investigation. *Appl. Sci.* 11, 9091. doi:10.3390/app11199091

Marigo, J.-J., and Maurel, A. (2017). Second order homogenization of subwavelength stratified media including finite size effect. *SIAM J. Appl. Math.* 77, 721–743. doi:10.1137/16m1070542

Matthews, M. C., Hope, V. S., and Clayton, C. R. I. (1996). Rayleigh the use of surface waves in the determination of ground stiffness profiles. *Proc. Inst. Civ. Eng. Eng.* 119, 84–95. doi:10.1680/igeng.1996.28168

Maurel, A., Marigo, J.-J., Pham, K., and Guenneau, S. (2018). Conversion of love waves in a forest of trees. *Phys. Rev. B* 98, 134311. doi:10.1103/physrevb.98.134311

Mei, J., Liu, Z., Shi, J., and Tian, D. (2003). Theory for elastic wave scattering by a two-dimensional periodical array of cylinders: an ideal approach for band-structure calculations. *Phys. Rev. B* 67, 245107. doi:10.1103/physrevb.67.245107

Miniaci, M., Krushynska, A., Bosia, F., and Pugno, N. M. (2016). Large scale mechanical metamaterials as seismic shields. *New J. Phys.* 18, 083041. doi:10.1088/1367-2630/18/8/083041

Mir, F., Saadatzi, M. S., Ahmed, R. U., and Banerjee, S. (2018). The possibility of harvesting electrical energy from industrial noise barriers using meta-wall bricks. *Proc. Sensors Smart Struct. Technol. Civ. Mech. Aerosp. Syst.* 10598, 679–683. doi:10.1117/12.2296422

Mu, D., Shu, H., Zhao, L., and An, S. (2020). A review of research on seismic metamaterials. *Adv. Eng. Mater.* 22, 1901148. doi:10.1002/adem.201901148

Wu, T., and Lim, C. W. (2020). Forest trees as naturally available seismic metamaterials: low frequency Rayleigh wave with extremely wide bandgaps. *Int. J. Struct. Stab. Dyn.* 20, 2043014. doi:10.1142/s0219455420430142

Oudich, M., and Badreddine Assouar, M. (2012). Surface acoustic wave band gaps in a diamond-based two-dimensional locally resonant phononic crystal for high frequency applications. *J. Appl. Phys.* 111. doi:10.1063/1.3673874

Prati, E. (2006). Microwave propagation in round guiding structures based on double negative metamaterials. *Int. J. Infrared Millim. Waves* 27, 1227–1239. doi:10.1007/s10762-006-9134-3

Pu, X., and Shi, Z. (2017). A novel method for identifying surface waves in periodic structures. *Soil Dyn. Earthq. Eng.* 98, 67–71. doi:10.1016/j.soildyn.2017.04.011

Pu, X., and Shi, Z. (2018). Surface-wave attenuation by periodic pile barriers in layered soils. *Constr. Build. Mater.* 180, 177–187. doi:10.1016/j.conbuildmat.2018.05.264

Qahtan, A. S., Huang, J., Amran, M., Qader, D. N., Fediuk, R., and Wael, A. D. (2022). Seismic composite metamaterial: a review. *J. Compos. Sci.* 6, 348. doi:10.3390/jcs6110348

Roux, P., Bindi, D., Boxberger, T., Colombi, A., Cotton, F., Douste-Bacque, I., et al. (2018). Toward seismic metamaterials: the METAFORET project. *Seismol. Res. Lett.* 89, 582–593. doi:10.1785/0220170196

Sánchez, P., Durda, D. D., Devaud, G., Fischer, A., Scheeres, D. J., and Dissly, R. (2021). Laboratory experiments with self-cohesive powders: application to the morphology of regolith on small asteroids. *Planet. Space Sci.* 207, 105321. doi:10.1016/j.pss.2021.105321

Sang, S., Sandgren, E., and Wang, Z. (2018). Wave attenuation and negative refraction of elastic waves in a single-phase elastic metamaterial. *Acta Mech.* 229, 2561–2569. doi:10.1007/s00707-018-2127-1

Schevenels, M., and Lombaert, G. (2017). Double wall barriers for the reduction of ground vibration transmission. *Soil Dyn. Earthq. Eng.* 97, 1–13. doi:10.1016/j.soildyn.2017.02.006

Seive, L. (2019). Coding metamaterials, digital metamaterials and programmable metamaterials. *Light Sci Appl* 3, doi:10.1038/lsa.2014.99

Sens-Schönfelder, C. (2008). Synchronizing seismic networks with ambient noise. *Geophys. J. Int.* 174, 966–970. doi:10.1111/j.1365-246x.2008.03842.x

Sewar, Y. Y., Zhang, Z., Meng, X., Wahan, M. Y., Qi, H., Al-Shami, Q. M., et al. (2022). Mechanical properties and constitutive relationship of the high-durable parallel strand bamboo. *J. Renew. Mater.* 10, 219–235. doi:10.32604/jrm.2021.016013

Thomes, R. L., Beli, D., and Junior, C. D. M. (2022). Space–time wave localization in electromechanical metamaterial beams with programmable defects. *Mech. Syst. Signal Process.* 167, 108550. doi:10.1016/j.ymssp.2021.108550

Thompson, D. J., Kouroussis, G., and Ntotsios, E. (2019). Modelling, simulation and evaluation of ground vibration caused by rail vehicles. *Veh. Syst. Dyn.* 57, 936–983. doi:10.1080/00423114.2019.1602274

Thompson, R. B., Margetan, F. J., Haldipur, P., Yu, L., Li, A., Panetta, P., et al. (2008). Scattering of elastic waves in simple and complex polycrystals. *Wave Motion* 45, 655–674. doi:10.1016/j.wavemoti.2007.09.008

Uhlemann, S., Hagedorn, S., Dashwood, B., Maurer, H., Gunn, D., Dijkstra, T., et al. (2016). Landslide characterization using P-and S-wave seismic refraction tomography—the importance of elastic moduli. *J. Appl. Geophys.* 134, 64–76. doi:10.1016/j.jappgeo.2016.08.014

Wang, M. Y., and Wang, X. (2013a). Frequency band structure of locally resonant periodic flexural beams suspended with force–moment resonators. *J. Phys. D. Appl. Phys.* 46, 255502. doi:10.1088/0022-3727/46/25/255502

Wang, Y., Zhang, C., Chen, W., Li, Z., Golub, M. V., and Fomenko, S. I. (2021). Precise and target-oriented control of the low-frequency lamb wave bandgaps. *J. Sound. Vib.* 511, 116367. doi:10.1016/j.jsv.2021.116367

Wang, Y.-F., and Wang, Y.-S. (2013b). Multiple wide complete bandgaps of two-dimensional phononic crystal slabs with cross-like holes. *J. Sound. Vib.* 332, 2019–2037. doi:10.1016/j.jsv.2012.11.031

Wei, W., Zhao, D., Xu, J., Wei, F., and Liu, G. (2015). P and S Wave tomography and anisotropy in northwest pacific and east asia: constraints on stagnant slab and intraplate volcanism. *J. Geophys. Res. Solid Earth* 120, 1642–1666. doi:10.1002/2014jb011254

Woods, R. D., Barnett, N. E., and Sagesser, R. (1974). Holography—a new tool for soil dynamics. *J. Geotech. Eng. Div.* 100, 1231–1247. doi:10.1061/ajgeb6.0000121

Xu, Y., and Peng, P. (2015). High quality broadband spatial reflections of slow Rayleigh surface acoustic waves modulated by a graded grooved surface. *J. Appl. Phys.* 117. doi:10.1063/1.4905948

Yang, J., Qi, L., Li, B., Wu, L., Shi, D., Uqaili, J. A., et al. (2021). A terahertz metamaterial sensor used for distinguishing glucose concentration. *Results Phys.* 26, 104332. doi:10.1016/j.rinp.2021.104332

Zeng, Y., Cao, L., Wan, S., Guo, T., Wang, Y.-F., Du, Q.-J., et al. (2022). Seismic metamaterials: generating low-frequency bandgaps induced by inertial amplification. *Int. J. Mech. Sci.* 221, 107224. doi:10.1016/j.ijmecsci.2022.107224

Zhang, G., and He, B.-J. (2021). Towards green roof implementation: drivers, motivations, barriers and recommendations. *Urban For. urban Green.* 58, 126992. doi:10.1016/j.ufug.2021.126992

Zhang, K., Luo, J., Hong, F., and Deng, Z. (2021). Seismic metamaterials with cross-like and square steel sections for low-frequency wide band gaps. *Eng. Struct.* 232, 111870. doi:10.1016/j.engstruct.2021.111870

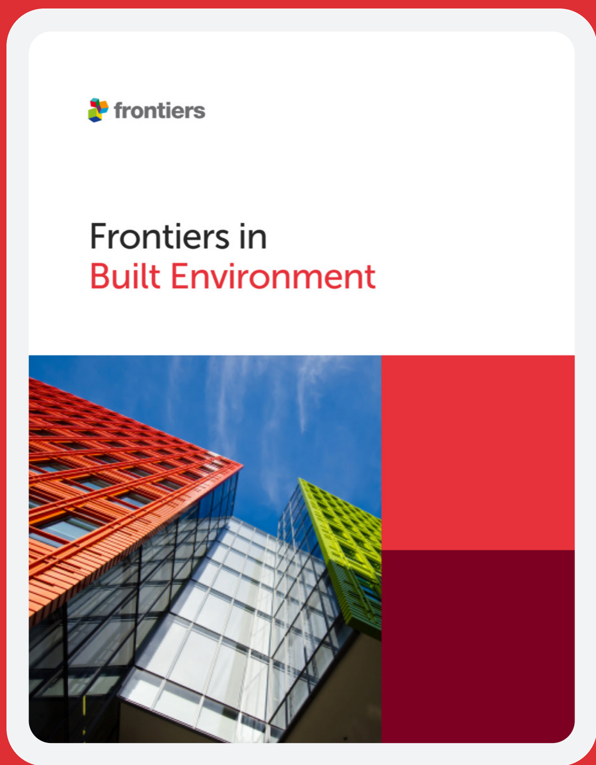
Frontiers in
Built Environment

Innovations in the engineering of sustainable buildings, cities, and urban spaces

An innovative journal that advances our knowledge of civil engineering. It focuses on the development of sustainable methodologies for the design and management of resilient buildings and infrastructure.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact