frontiers | Research Topics

Energy efficiency analysis and intelligent optimization of process industry

Edited by
Zhiqiang Geng, Xiang Zhang, Yongming Han and Xingxing Zhang

Published in
Frontiers in Energy Research

ISSN 1664-8714
ISBN 978-2-8325-3576-9
DOI 10.3389/978-2-8325-3576-9

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

Energy efficiency analysis and intelligent optimization of process industry

Topic editors

Zhiqiang Geng — Beijing University of Chemical Technology, China
Xiang Zhang — University of North Carolina at Charlotte, United States
Yongming Han — Beijing University of Chemical Technology, China
Xingxing Zhang — Dalarna University, Sweden

Citation

Geng, Z., Zhang, X., Han, Y., Zhang, X., eds. (2023). *Energy efficiency analysis and intelligent optimization of process industry*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-3576-9

Table of contents

frontiers | Frontiers in Energy Research

TYPE Editorial
PUBLISHED 13 September 2023
DOI 10.3389/fenrg.2023.1283021

OPEN ACCESS

EDITED AND REVIEWED BY
Ellen B. Stechel,
Arizona State University,
United States

*CORRESPONDENCE
Yongming Han,
hanym@mail.buct.edu.cn
Zhiqiang Geng,
gengzhiqiang@mail.buct.edu.cn

RECEIVED 25 August 2023
ACCEPTED 06 September 2023
PUBLISHED 13 September 2023

CITATION
Han Y, Wu P, Geng Z, Zhang X and Zhang X (2023), Editorial: Energy efficiency analysis and intelligent optimization of process industry. *Front. Energy Res.* 11:1283021. doi: 10.3389/fenrg.2023.1283021

Editorial: Energy efficiency analysis and intelligent optimization of process industry

Yongming Han[1,2]*, Peng Wu[1,2], Zhiqiang Geng[1,2]*, Xingxing Zhang[3] and Xiang Zhang[4]

[1]College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China, [2]Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing, China, [3]Department of Energy and Construction Engineering, Dalarna University, Falun, Sweden, [4]Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC, United States

KEYWORDS

energy systems, efficiency analysis, intelligent optimization, intelligent detection, process industry

Editorial on the Research Topic
Energy efficiency analysis and intelligent optimization of process industry

Introduction

Given the rapid economic development, energy-saving and reduction of carbon dioxide emissions are now recognized as primary global goals. One of the most effective approaches for achieving energy efficiency and emissions reduction is the utilization of energy efficiency analysis and intelligent optimization methods, which find wide application in the process industry.

Common methods include traditional mechanism methods based on momentum transport, energy and quality transport (TT), as well as reaction engineering (RG) (TT-RG), in addition to data-driven artificial intelligence techniques. However, mechanism methods require well-defined parameters to achieve accurate modeling, posing a significant challenge within the process industry. In line with the advancement of artificial intelligence (AI) and big data, data-driven AI methods have progressively emerged as widely employed modeling tools, without considering internal mechanisms and strong nonlinear approximation ability, which can provide theoretical guidance for energy saving and emission reduction of the process industry.

The objective of this Research Topic is to establish an energy efficiency analysis and optimization model that employs machine intelligence, deep learning, and other AI methods to achieve energy savings and reduce carbon dioxide emissions. Advanced AI-based optimization algorithms are necessary to address the challenge of local optima and inability to attain global optima in the optimization process. To tackle issues like strong coupling, nonlinearity, and missing data in the process industry, effective data analysis and processing techniques including dimensionality reduction, matrix completion, and feature extraction should be proposed. Furthermore, it is crucial to construct a robust and highly generalized prediction model that can surmount problems related to low precision and poor generalization in process industry modeling.

It is important to highlight that following a rigorous peer review process, 10 articles have been accepted for inclusion in this Research Topic, spanning the following categories.

Energy system efficiency analysis and optimization

A plethora of innovative solutions were proposed by researchers to tackle multifaceted challenges in energy systems. Addressing the weak continuity and volatility of solar power generation data, Liu et al. proposed a photovoltaic power generation prediction method, which amalgamated Radial Basis Function Neural Networks (RBFNNs), the Adaptive Black Widow Optimization algorithm (ABWO), Similar Day Analysis (SDA), and K-means Clustering. The outcome is an enhanced stability and power quality for the grid. In the context of the burgeoning integration of 5G base stations, Guo et al. delved into the operational framework of microgrids and the carbon-reducing potential of these 5G stations on the power system, and crafted a multi-objective optimal operational model for microgrids access with 5G base stations to achieve the dual goals of minimizing microgrid operation costs and carbon emissions. Wang and Liu put forth a multi-objective evolutionary algorithm including the NDWA-GA and Pareto optimal space PCA for the optimal capacity allocation problem of multi-energy complementary systems to minimize both the system's total investment cost and battery capacity, thereby amplifying the utilization of clean energy. Empirical evidence showcased the superior convergence and economic viability of the proposed method. To address the energy loss observed in high temperature heat pump system under extensive temperature elevations, Hao et al. formulated a thermodynamic model for a double flash combined cycle system. Combined with the multivariate simulated annealing algorithm, the COP of the system was taken as the optimization objective to complete the calculation of the steady-state thermodynamic parameters of this system. The double flash combined cycle system was validated to possess superior steam generation capabilities under pronounced temperature rises and elevated condensation temperatures. Collectively, these research endeavors fortify the technical foundation for energy systems, ensuring they operate efficiently, stably, and with a reduced carbon footprint.

Optimization of energy system management and trading mechanism

For enhanced financial budgeting and localized operations, Lu et al. integrated the Support Vector Machine—based Recursive Feature Elimination (SVM-RFE) technique with a variant of the Autoregressive and Moving Average (ARMA) model to predict energy consumption and operational costs, thereby refining management in the process industry. The precision of the proposed method was corroborated through case studies. In pursuit of bolstering the resilience of the natural gas market, Liu et al. devised a novel customer value portrait framework based on different types of behavioral characteristics and emerging trends in the natural gas market to identify industrial customer value. By harnessing varied behavioral data, it aptly encapsulates the value of natural gas industry clientele. Highlighting the economic and ecological potential of microgrids necessitates the creation of a proficient power trading mechanism. Traditional centralized power management models often grapple with issues of unreliability and confidentiality breaches during information exchanges. Therefore, Wang et al. introduced a blockchain-anchored distributed community energy trading mechanism, termed CE-SDT. And the thorough analysis affirmed its suitability. The introduction of various new technologies and strategies provided the possibility to achieve more efficient, economical and environmentally friendly energy utilization.

Energy system safety and fault detection

Within the realm of energy systems, the emphasis on technical safety and fault detection remains paramount. In view of the evident shortcomings in the standard interpretation of transformer fault detection and the inherent limitations of Adaptive Neuro Fuzzy Inference System (ANFIS), Equbal et al. introduced an online system for the early identification of transformer fault based on e-nose and ANFIS, which was demonstrated promising results upon testing. With the widespread deployment of the smart grid, as an open cyber physical system, it faces various security threats. Among them, False Data Injection Attack (FDIA) had become a major security risk, which bypassed the conventional detection of the system by constructing and injecting forged data. In order to cope with the diversity of this attack, Lin et al. unveiled a detection methodology for false data injection attacks, anchored in Deep Reinforcement Learning (DRL), which had notably enhanced detection efficacy. By integrating cutting-edge fault detection techniques with robust protection measures, a fortified safeguard for energy system was established.

Material science and environmental protection

The degradation of metals and alloys through corrosion can be effectively mitigated using corrosion inhibitors. However, traditional organic and inorganic inhibitors present issues related to toxicity, undesirable side effects, and environmental contamination. Recognizing these challenges, there has been a shift in research focus towards water-soluble polymer corrosion inhibitors, which offer environmentally benign, non-toxic, and minimal pollution attributes. Yihang conducted an extensive review on the action mechanisms of polymer-based inhibitors and the research status of natural polymer inhibitors and synthetic polymer inhibitors. This review aimed to furnish insights that could guide the advancement of eco-friendly metallic coatings.

In summary, this Research Topic encompasses a wide range of scholarly articles focusing on energy efficiency analysis and intelligent optimization in the process industry. The contributions within this compilation delve into various aspects of the field, exploring novel strategies and methodologies. A common thread among these works is the utilization of advanced

technological tools, including neural networks, the Internet, and blockchain, to enhance energy production, management, and utilization, as well as to improve the performance and reliability of energy systems.

The insights shared through this Research Topic will significantly contribute to the advancement of the field of energy efficiency analysis and intelligent optimization. This Research Topic of research exemplifies the potential of these innovations in propelling the energy sector forward, with the capacity to deliver greener, more efficient, and sustainable energy solutions. These solutions are instrumental in addressing the world's escalating energy demands while mitigating the impact on the environment.

Author contributions

YH: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Methodology, Project administration, Resources, Writing–original draft, Writing–review and editing. PW: Data curation, Formal Analysis, Methodology, Resources, Writing–original draft. ZG: Investigation, Project administration, Resources, Supervision, Writing–review and editing. XxZ: Conceptualization, Methodology, Resources, Validation, Writing–review and editing. XZ: Formal Analysis, Validation, 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 was funded by National Natural Science Foundation of China under Grant No. 62066005, U21A20464, Project of the Guangxi Science and Technology under Grant No. AD21196006, the National Natural Science Foundation of China (Grant Nos. 51876055, 51806060, and U1504524), the Natural Science Foundation of Henan Province (Grant Nos. 182300410233) and the Shanghai Heimdallr Energy Saving Technology Co., Ltd. (05N14030820), the National Natural Science Foundation of China (62273151, 61873096, and 62073145), Guangdong Basic and Applied Basic Research Foundation (2020A1515011057, 2021B1515420003), Guangdong Technology International Cooperation Project Application (2020A0505100024, 2021A0505060001), Fundamental Research Funds for the central Universities, SCUT (2020ZYGXZR034), the Horizon 2020 Framework Programme-Marie Skłodowska-Curie Individual Fellowships (891627), the National Natural Science Foundation of China under Grant Nos. 71804167, 72174188, the Research and Development in the key areas of Guangdong Province (2020B0101090001), and the Doctoral Fund Project of Chongqing Institute of Industry and Technology (Grant No. 2022G2YBS2K2-12). This work is supported by the National Natural Science Foundation of China (21978013) and the Fundamental Research Funds for the Central (XK1802-4).

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 12 September 2022
DOI 10.3389/fenrg.2022.990018

OPEN ACCESS

EDITED BY
Xingxing Zhang,
Dalarna University, Sweden

REVIEWED BY
Salim Heddam,
University of Skikda, Algeria
Vishnupriyan Jegadeesan,
Chennai Institute of Technology, India

*CORRESPONDENCE
Yongquan Zhou,
zhouyongquan@gxun.edu.cn

SPECIALTY SECTION
This article was submitted to Process and Energy Systems Engineering, a section of the journal Frontiers in Energy Research

RECEIVED 09 July 2022
ACCEPTED 11 August 2022
PUBLISHED 12 September 2022

CITATION
Liu H, Zhou Y, Luo Q, Huang H and Wei X (2022), Prediction of photovoltaic power output based on similar day analysis using RBF neural network with adaptive black widow optimization algorithm and K-means clustering. *Front. Energy Res.* 10:990018. doi: 10.3389/fenrg.2022.990018

Prediction of photovoltaic power output based on similar day analysis using RBF neural network with adaptive black widow optimization algorithm and K-means clustering

Hui Liu[1,2], Yongquan Zhou[1,2,3]*, Qifang Luo[1,2], Huajuan Huang[1,2] and Xiuxi Wei[1,2]

[1]College of Artificial Intelligence, Guangxi University for Nationalities, Nanning, China, [2]Guangxi Key Laboratories of Hybrid Computation and IC Design Analysis, Nanning, China, [3]Xiangsihu College of Guangxi University for Nationalities, Nanning, China

Solar photovoltaic power generation has become the focus of the world energy market. However, weak continuity and variability of solar power data severely increase grid operating pressure. Therefore, it is necessary to propose a new refined and targeted forecasting method to broaden the forecasting channels. In this paper, a hybrid model (KM-SDA-ABWO-RBF) based on radial basis function neural networks (RBFNNs), adaptive black widow optimization algorithm (ABWO), similar day analysis (SDA) and K-means clustering (KM) has been developed. The ABWO algorithm develops adaptive factors to optimize the parameters of RBFNNs and avoid getting trapped in local optima. SDA and K-means clustering determine the similarity days and the optimal similarity day through meteorological factors and historical datasets. Nine models compared forecast accuracy and stability over four seasons. Experiments show that compared with other well-known models on the four indicators, the proposed KM-SDA-ABWO-RBF model has the highest prediction accuracy and is more stable.

KEYWORDS

radial basis function neural networks, adaptive black widow optimization algorithm, similar day analysis, k-means clustering, photovoltaic power prediction, metaheuristic

1 Introduction

The traditional method of generating electricity is to use non-renewable fossil fuels such as coal and oil. However, due to rising electricity prices, the depletion of non-renewable energy sources and the consequent issues of climate change, solar photovoltaic (PV) is one of the most popular renewable energy sources in the utility grid. Figure 1 (Renewable Energy Policy) shows the global capacity and annual increase in PV power system from 2011 to 2021. The solar PV capacity has increased from 70 to 942 Gigawatts

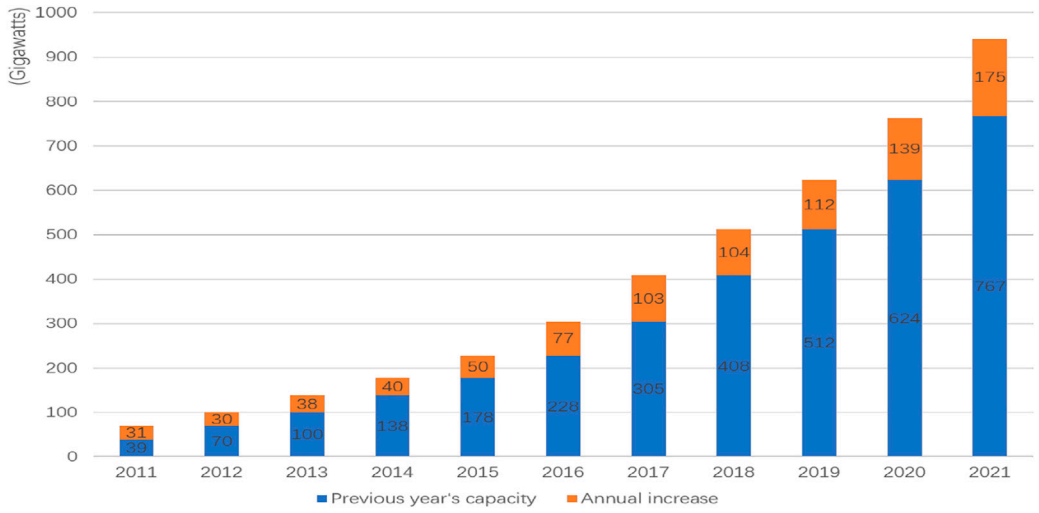


FIGURE 1
Solar PV global capacity and annual additions, 2011–2021.

in the last 10 years. However, the photovoltaic (PV) power generation industry and power system are faced with severe challenges (Mahmud et al., 2021). Due to the influence of the amount of solar radiation by weather factors (temperature, etc.), photovoltaic power generation has the defects of instability, uncertainty, difficulty in control, randomization and intermittency. During grid-connected operation, the stable operation and power quality of the power grid are disturbed. The stability and safety of photovoltaic power generation grid-connected operation not only has an important influence on social and economic development, but also is the focus of current social and economic concerns. Therefore, the method and system research of PV power generation prediction has important academic and application value. In the past, complex physical models and statistical models were used in the field of photovoltaic power generation forecasting. Reikard (Reikard, 2009) predicted solar irradiance using the Autoregressive Integrated Moving Average (ARIMA) methods. To take climatic factors into account, the ARMAX model is proposed, and the prediction accuracy is higher than that of the ARIMA model (Li et al., 2014). However, recent study (Sobri et al., 2018) has demonstrated that machine learning algorithms, such as neural networks (NNs) and support vector machines (SVMs), are superior to other forecasting methods for predicting photovoltaic power generation. A weighted SVM was proposed in Ref. (Xu et al., 2012). Chen et al, (2011) proposed an ANN-based weather classification method and utilized radial basis function neural networks (RBFNNs) for short-term photovoltaic power generation forecasting. Using appropriate algorithms to optimize the parameters of the model is one of the keys to improve the prediction effect (Du et al., 2018). Several studies have found that the combined approach of parameter optimization algorithms combined with NN or SVM shows higher prediction accuracy (Dolara et al., 2015). In Ref. (Ghimire et al., 2018), adaptive differential evolution extreme learning machine model outperformed all 9 benchmark models. Hossain and Mahmood, 2020) proposed a novel clustering method to classify irradiance values by weather type, improving prediction accuracy by 33%. Dong et al, (2020) proposed a convolutional neural network model based on genetic algorithm and particle swarm algorithm to reduce the complexity of irradiance prediction. Zhang et al, (2015) proposed SDD Engine to construct a Euclidean distance metric for historical data to obtain the similarity days of predicted days. Wang et al, (2018) designed a RBFNN based on the multi-objective dragonfly algorithm and obtained excellent prediction results. Liu et al, (2015) used a BPNN to forecast output power by researching the influence of aerosol index data etc. Combination forecasting method, that is, combining the advantages of different forecasting methods to forecast is a hot research topic at present (Lin et al., 2018). Akhter et al, (2019) demonstrated that the model combining machine learning methods (ANN, SVM, ELM) and metaheuristics achieved the combined advantages of two or more techniques in PV power generation forecasting. The performance of the hybrid method is determined by the performance of each method (Xu et al., 2012). Therefore, several proven methods with better performance are synthesized, which can maximize the performance of the hybrid model.

Compared with other neural networks, RBFNNs have stronger robustness, learning and memory ability. It not only has good classification ability, but also is the optimal network to

complete the mapping. However, the original RBFNNs are easy to get trapped in local optima and have poor global convergence ability (Han et al., 2017). Therefore, it is necessary to select an appropriate learning algorithm to improve the convergence performance of RBFNNs. The focus is on optimizing the parameters of RBFNNs (i.e. the center, width of the basis functions, and the weights from the hidden layer to the output layer), as well as determining the number of hidden layer neurons. Combined with common training algorithms, RBFNNs have been widely used in a large number of classification and regression problems. For example, Huang et al, (2004) designed a Growing and pruning RBF (GAP-RBF), which was improved and called GGAP-RBF (Huang et al., 2005). The goal of these two models is to make the network more compact to achieve less time consumption. Liu et al, (2005) tested on UCI classification datasets and seismic datasets using original RBFNNs and outperformed other models. However, what leaves these algorithms with little room for improvement is that they do not have a way to adjust parameters and network size simultaneously (Xie et al., 2012). Later, some studies introduced metaheuristic algorithms to simultaneously optimize network parameters and scale, which provided some unified and regular learning algorithms for researchers, broadening the space for improvement and optimization. Such as, Feng (Feng, 2006) proposed the SORBFNN based on particle swarm optimization (PSO) algorithm to solve three nonlinear problems such as approximation of nonlinear functions. Joe et al. (Qiao et al., 2019) combined an improved immune algorithm (IA) to optimize RBFNNs. Li et al. (Lee and Ko, 2009) developed the time-varying evolutionary PSO algorithm and compared it with other improved versions of the PSO algorithm. Duvvuri et al. (Duvvuri and Anmala, 2019) used the genetic algorithm to learn RBFNNs to predict fecal coliforms.

It is statistically known that the PSO algorithm and other improved PSO algorithms are used more for training RBFNNs than other metaheuristic algorithms. However, the PSO algorithm converges faster, but produce a low performance in exploitation search ability, and have low convergence precision (Parsopoulos and Vrahatis, 2002). To overcome this limitation, other metaheuristic algorithms that balance exploration and exploitation search capabilities can be considered, such as cuckoo search (CS) (Yang and Deb, 2009), symbiotic organisms search (SOS) (Cheng and Prayogo, 2014), gravitational search algorithm (GSA) (Rashedi et al., 2009), sine cosine algorithm (SCA) (Mirjalili, 2016), grey wolf optimizer (GWO) (Mirjalili et al., 2014), differential evolution (DE) (Storn and Price, 1997), etc. Although many metaheuristic algorithms have been proposed for many years, however, due to the "No Free Lunch (NFL) *Theorem*" (Wolpert and Macready, 1997), there is no one-size-fits-all algorithm that can reach the upper limit of the optimization capability, which is sufficient to solve all optimization problems. In terms of training neural network parameters, the black widow optimization algorithm proposed in the past two years showed excellent talent due to its powerful search and extraction capabilities. Since BWO was proposed, several papers have verified its superiority in optimizing neural networks, such as adaptive network-based fuzzy inference system (ANFIS). For example, Katooli et al. (Katooli and Koochaki, 2020) applied BWO in the training process of ANFIS to improve its accuracy and robustness, and combined the association rule mining technique (ARMT) to select the most necessary input data. Tightiz et al, (2020) demonstrated that BWO trains faster and has higher classification accuracy than gradient-based learning algorithms. Memar et al, (2021) used BWO to learn ANFIS and SVR and forecast maximum wave height. Panahi et al, (2021) proposed the ANFIS-BWO and SVM-BWO models for hydraulic engineering prediction. It is known from previous research that both BWO and RBFNN have excellent performance. Therefore, the combined method of BWO and RBFNN combines the advantages of both methods, and it is expected to have higher prediction performance than other methods.

In addition, it is necessary to ensure not only the efficiency of the predictive model, but also the accuracy of the input data. This requires the prediction model not only to consider the relationship between power generation and meteorological factors, but also to take historical power generation data as one of the influencing factors (Lin et al., 2018). Similar day analysis (SDA) is a mathematical model used to filter historical datasets, selecting similar days for training from a large amount of historical data (Zhou et al., 2020). Moreover, many literatures have also demonstrated that clustering algorithms can improve the prediction efficiency (Ghayekhloo et al., 2015; Lin et al., 2018; Hossain and Mahmood, 2020). Combining the advantages and disadvantages of previous research works, this paper proposes a combined prediction model based on *K*-means, SDA, ABWO and RBFNN, namely KM-SDA-ABWO-RBF, for real-life photovoltaic power generation prediction. The innovations and contributions of this paper are summarized as follows:

1) The ABWO algorithm is designed. The performance of the original BWO algorithm on optimizing RBFNNs needs to be improved. Numerous studies (Mirjalili et al., 2012; Gonzalez et al., 2015; Justus and Anuradha, 2022) have found that PSO algorithm, gravitational search algorithm (GSA), golden eagle optimizer (GEO) have better results than other algorithms in training neural network parameters. By analyzing the parameter design methods in these algorithms, it is known that they all have adaptive inertia weights that vary with the number of iterations. Therefore, the BWO algorithm is adaptively designed, and the parameters of the crossover stage are changed from random values to adaptively reduce with the increase of the number of iterations. Compared with many algorithms such as PSO, the accuracy of the adaptive black widow optimization algorithm is guaranteed.

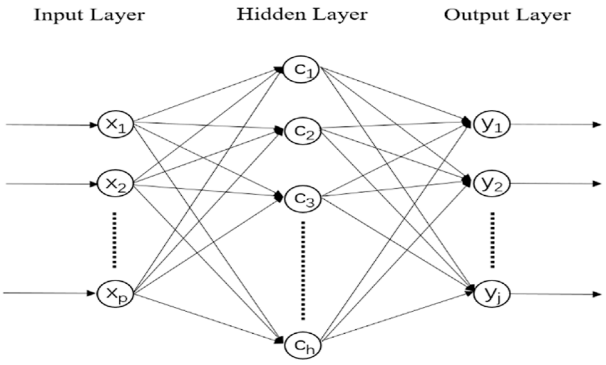


FIGURE 2
Multiple-input multiple-output RBFNN.

2) The goal of the metaheuristic algorithm is to find the optimal value of a function. Moreover, the goal of RBFNNs in the PV power generation problem is to minimize the prediction error, which boils down to the problem of solving the function minimum problem. Therefore, the metaheuristic algorithm is used to optimize RBFNNs to solve the PV power prediction problem.
3) In photovoltaic power generation forecasting, in order to avoid excessively long forecasting time caused by a large number of irrelevant abnormal historical data, and to synthesize the advantages of different forecasting methods, this paper adopts a combined forecasting method, namely the KM-SDA-ABWO-RBF model. The model uses the Pearson correlation coefficient to analyze the correlation between historical data weather factors and photovoltaic power generation, and then calculates the weighted Euclidean metric. Some days with smaller weighted Euclidean metric results are used as the similarity days useful for forecasting days. The influencing factors of power generation are not only temperature, humidity, irradiance, etc. Under different weather types (sunny, cloudy, rainy), the output power is different. Therefore, this paper conducts in-depth cluster analysis on the input data, extracts refined modalities, and performs classification prediction.
4) A limitation of most studies is that they do not take into account the stability of the predicted results. The model proposed in this paper not only has a small prediction error, but also has a relatively stable prediction result.

The remainder of this article is summarized as follows: The specific scheme of the KM-SDA-ABWO-RBF model is proposed in Section 2. The specific implementation process of PV power prediction is introduced and discussed in Section 3. Section 4 presents and discusses the experimental results. Finally, conclusions and future work in Section 5.

2 KM-SDA-adaptive black widow optimization-RBF model

2.1 RBF neural network

RBFNN is a special type of fully connected feedforward network (Er et al., 2002), (Chen et al., 1991), consisting of only three layers: input layer, output layer and hidden layer, each of which has an activation function. RBFNNs have achieved higher convergence accuracy, with more convergence speed for training, less convergence time (Mandal et al., 2012). The network topology of a multiple-input multiple-output RBFNN is drawn in Figure 2 below.

The hidden layer can directly map the input vector to the high dimension space without the need for weight connection. When the center point of the RBFNN is determined, this mapping relationship is also determined. The number of hidden layer neurons determines the network size. The activation function of the hidden layers neuron uses the Gaussian function, which is a nonlinear transformation function. Each hidden layer neuron of the RBFNN obtains the output of the hidden layer by calculating the activation function, and the final output of the output layer is the linear weighted sum of the outputs of the hidden layer. The output equation is described as follows:

$$y_j = \sum_{i=1}^{h} w_{i,j} e^{-\|\boldsymbol{x}-\boldsymbol{c}_i\|^2 / 2\sigma_i^2} \tag{1}$$

where $i = 1, 2, ..., h$, h is the number of hidden layer; $j = 1, 2, ..., n$, n is the number of output neurons; $w_{i,j}$ is the output weight between the ith hidden neuron and the jth output neuron. The input vector is expressed as $\boldsymbol{x} = [x_1, x_2, ..., x_p]^{\mathrm{T}}$, and $\boldsymbol{x} \in \Re^{p\times 1}$; p is the number of input neurons; $\boldsymbol{c}_i$ is the center vector of the ith hidden neuron; $\| \boldsymbol{x} - \boldsymbol{c}_i \|$ is the Euclidean distance between the input vector and the center vector; σ_i is the width of the ith hidden neuron. The proposed ABWO-based method simultaneously searches all parameters of RBFNNs.

2.2 BWO and adaptive black widow optimization algorithm

Metaheuristic algorithms can be divided into three broad categories: evolutionary algorithms, physicochemical-based algorithms, and swarm intelligence algorithms (Wang et al., 2020). Swarm intelligence algorithms are based on the natural biological proliferation of social connections and the behavior of natural animals and other people in their lives (Abualigah and Diabat, 2021). As a swarm intelligence algorithm, the BWO algorithm (Hayyolalam and Kazem, 2020) is based on the mating and reproduction behavior of black widow spiders. Each individual spider corresponds to a candidate solution to

the optimization problem, and survivability represents a fitness function. Getting the strongest spider is getting the optimal solution to the problem. The mating and reproduction behaviors of black widow spiders include crossover, cannibalism and mutation.

1) *Crossover:* The crossover phase is the exploration search phase. Firstly, combined with the parameter optimization problem of RBFNNs, the solution of root mean square error (*RMSE*) is used as the fitness function. Secondly, the populations are sorted according to their fitness. According to the fertility rate, spiders with high fitness values in the population were selected to participate in mating, and a pair of parents (male and female spiders) were randomly selected from them for mating and reproduction each time. Each black widow spider displays the value of the problem variable. The structure is treated as a one-dimensional array, where n is the dimension of the feature:

$$spider = [s_1, s_2, ..., s_n] \tag{2}$$

Each pair of parents simulates the reproductive process with the help of an α array:

$$\begin{cases} y_1 = \alpha.*s_1 + (1-\alpha).*s_2 \\ y_2 = \alpha.*s_2 + (1-\alpha).*s_1 \end{cases} \tag{3}$$

where $\alpha_i$representative is a random value in [0,1], s_1 and s_2 representative is parent, y_1 and y_2 representative is offspring, and each pair of parents mating is repeated $n/2$ times.

2) *Cannibalism:* The behavior of cannibalism refers to strong spiders eating weak spiders, that is, keep spiders with high fitness values and eliminating spiders with low fitness values. It includes sexual cannibalism and sibling cannibalism. Sexual cannibalism is when female spiders eat male spiders due to their smaller size, gaining nutrients that give offspring a higher chance of surviving. After mating, this behavior is achieved by destroying the mated parent. Corresponding to the optimization problem, this means that those with high fitness values are females, those with low fitness values are males, and females with higher fitness values are retained. Sibling cannibalism is when the offspring hatch and the stronger offspring eats the weaker siblings, thereby increasing the survival rate of the survivors. It achieves its goal by destroying a portion of offspring with weak fitness values according to the cannibalism rate. Each time through the crossover and cannibalistic stages, only one parent and one offspring are retained for each pair of parent spiders that mate.
3) *Mutation:* The mutation phase is the exploitation search phase. At this stage, the algorithm selects multiple spiders with higher fitness values according to the mutation rate, changes the attribute information of these spiders, and obtains new candidate solutions, so as to increase the diversity of the population. The mutation method randomly swaps the two eigenvalues in the array for each spider. e.g.,$[s_1, ..., s_i, ..., s_j, ..., s_n]$ mutates to $[s_1, ..., s_j, ..., s_i, ..., s_n]$.

The advantage of the BWO algorithm is that the equation is simple, easy to improve, and the principle is obvious, easy to understand, and even has a fast convergence speed. In order to increase the diversity and difference of the population and avoid getting trapped in local optima, only the equation of the crossover stage is improved, and the other stages remain unchanged. The parameter optimization of the improved algorithm on RBFNNs will be more accurate. Since the improved algorithm is adaptive, it is called adaptive black widow optimization (ABWO) algorithm.

The specific improvements are as follows: Eq. 3 shows that in the crossover stage, the α array is used to represent the process of generating offspring from the parent generation. In Eq. 4, the β array is introduced:$\beta = [\beta_1, \beta_2, ..., \beta_n]$. Each component of this array is not randomly selected in the range [0, 1] like the α array. It is randomly given in the range [0, 1] at the beginning, and then it adaptively decreases as *it* increases. The detailed description of the β array is as follows:

$$\beta_i = \frac{5 \times (Maxiter - it + 1)}{Maxiter} \tag{4}$$

Eq. 3 is changed to

$$\begin{aligned} y_1 &= \alpha.*s_1 + (1-\alpha).*s_2 \\ y_2 &= \beta.*s_2 + (1-\beta).*s_1 \end{aligned} \tag{5}$$

where s_1 and s_2 are parents, y_1 and y_2 are descendants, $Maxiter$ is the maximum number of iterations, and *it* is the current number of iterations.

2.3 Similar day analysis

In order to avoid training numerous irrelevant abnormal data, which leads to a long prediction time, this paper introduces the similar day analysis to obtain historical data that is instructive for the prediction results. The weather characteristics of certain days in the historical data are very similar to the weather characteristics of the forecasting days, and these days are called the similarity days (Zhou et al., 2020). The similarity days are obtained from the day the dataset starts to the day before the prediction day. There are many machine learning methods to obtain the similarity days: Euclidean metric (Mandal et al., 2007), PCC and cosine similarity, etc. When comparing the similarity of a certain meteorological factor between two days, more attention is paid to the numerical difference of weather characteristics between two days, rather than the similarity of change trends and the difference of numerical direction.

Therefore, a weighted Euclidean metric based on the influence of multiple meteorological factors was chosen instead of PCC and cosine similarity.

When predicting photovoltaic power generation, only the influence of meteorological factors on the output power needs to be considered. Different meteorological factors will lead to different output photovoltaic power. In order to obtain the influence weight of each meteorological factor on the output power, this paper uses the Pearson correlation coefficient (PCC) to conduct a correlation analysis on the relationship between weather factors and photovoltaic power generation. PCC (Bugała et al., 2018) is used to measure the degree of correlation between two variables, and its absolute value ranges from [0,1]. The closer the calculated PCC value is to 1 or -1, the stronger the correlation between the two variables. Conversely, the closer the value of PCC is to 0, the weaker the correlation. This study randomly selects a day from the dataset and calculates the PCC between the output power and each meteorological factor, as the weight coefficient of the weighted Euclidean metric. The greater the influence of the meteorological factor on the output power, the higher the weight coefficient of the meteorological factor. PCC is calculated as follows:

$$P(k) = \frac{\sum_{i=1}^{n}(u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^{n}(u_i - \bar{u})^2}\sqrt{\sum_{i=1}^{n}(v_i - \bar{v})^2}} \tag{6}$$

where n represents the number of power data of the day; k represents the number of weather factors of the day; u_i represents the kth weather factor at the ith hour of the day; v_i represents the photovoltaic power generation at the ith hour of the day. $P(k)$ is the PCC between the kth weather factor and the output power, that is, the degree of influence of the weather factor on the output power.

The Euclidean distance for each meteorological factor between two days is described as follows:

$$ED(k) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} \tag{7}$$

where X_i and Y_i represent the kth weather factor at the ith hour of the forecasting day and the history day, respectively; n represents the number of power data per day; k represents the number of weather factors per day. The smaller the value of $ED(k)$, the higher the similarity of the kth weather factor between two days.

The similarity between a historical day and a forecasting day is defined as the sum of the weighted Euclidean metric, described as follows:

$$S = \sum_{k=1}^{K} P(k)ED(k) \tag{8}$$

where $P(k)$ represents the weight coefficient of the kth weather factor; $ED(k)$ represents the Euclidean distance of the kth weather factor between two days. K represents the number of weather factors per day. The smaller the value of S, the higher the similarity between the historical day and the predicted day. The historical day corresponding to the minimum value of S is considered as the best similar day.

2.4 K-means clustering

In addition to meteorological factors, the type of weather can also affect the output power value. However, there is no information on weather types such as sunny, cloudy, rainy, etc. In the dataset used in this paper. In order to determine the weather type for each day of each season, the K-means algorithm is used to classify the input meteorological factors. Meteorological factors such as temperature and humidity in the same season are not very different. On the contrary, the irradiance of each day varies greatly. Therefore, the proposed clustering algorithm classifies the irradiance within the forecast horizon (similar days and the forecasting day). After the cluster center is obtained, the original irradiance value is replaced with the cluster center of the class to which the irradiance belongs. Different irradiance values represent different weather types. The flowchart of the clustering process is shown in Figure 3.

2.5 adaptive black widow optimization-RBF neural network

ABWO-RBF neural network can adjust all parameters during the learning process. Taking the structure of multiple-input and multiple-output as an example, in the widow initialization stage, the ith widow is described as

$$widow_i = \left[\boldsymbol{c}_{i,1}^{\mathrm{T}}, \sigma_{i,1}, \boldsymbol{w}_{i,1}^{\mathrm{T}}, \boldsymbol{c}_{i,2}^{\mathrm{T}}, \sigma_{i,2}, \boldsymbol{w}_{i,2}^{\mathrm{T}}, \ldots, \boldsymbol{c}_{i,h}^{\mathrm{T}}, \sigma_{i,h}, \boldsymbol{w}_{i,h}^{\mathrm{T}}\right] \tag{9}$$

where $\boldsymbol{c}_{i,h}$, $\sigma_{i,h}$, and $\boldsymbol{w}_{i,h}$ are the center vector, width, and output vector of the hth hidden neuron in the ith widow, respectively; h is the network size; D is the dimension of the widow satisfying $D = p + 1 + j$, with p being the number of input variable and j being the number of output variable. The fitness value of each widow represents the accuracy of the ABWO-RBF neural network. In order to optimize the parameters of RBFNNs, the fitness value of each widow is the RMSE. Taking the structure of multi-input and multi-output as an example, the specific definitions are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_j - f\left(x_j\right)\right)^2} \tag{10}$$

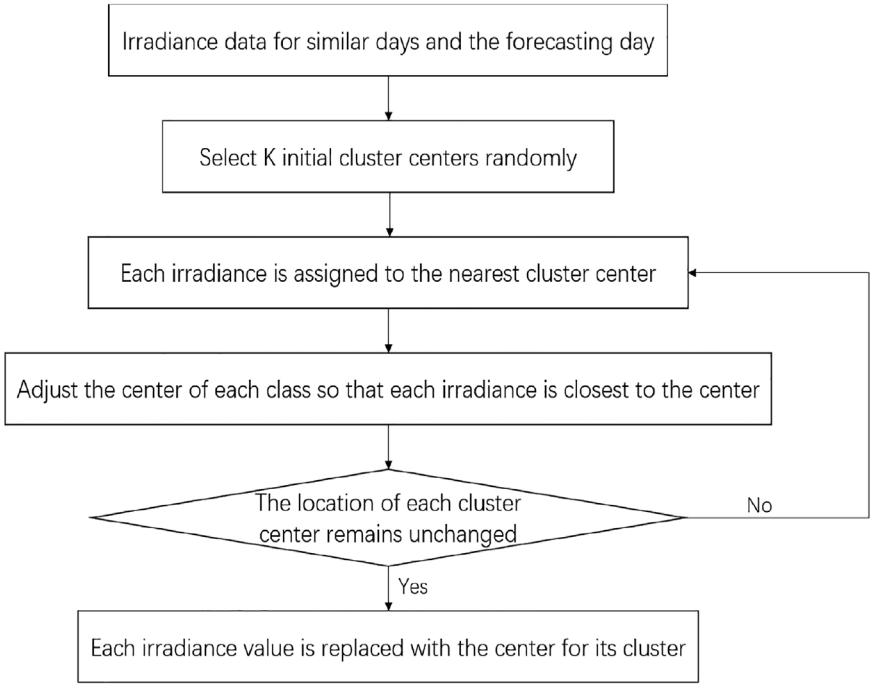


FIGURE 3
The process of the K-means clustering.

where n is the number of output neuron, y_j and $f(x_j)$ are the real output of the network and the desired output, respectively.

2.6 KM-SDA-adaptive black widow optimization-RBF model

The system architecture of the KM-SDA-ABWO-RBF model for PV power generation forecast is shown in Figure 4. The procedure is elaborated as follows:

Step 1. Obtain historical datasets, including attribute information such as several meteorological factors and photovoltaic power generation.

Step 2. The data is preprocessed for outliers and normalized.

Step 3. According to the equation of weighted Euclidean metric, several days with more similar meteorological characteristics to the predicted day are selected from the historical data as similar days, that is, days with smaller weighted Euclidean metric value. The day with the smallest weighted Euclidean metric is the optimal similarity day.

Step 4. The K-means algorithm classifies the daily average global horizontal radiation and the daily average diffuse horizontal radiation respectively, and replaces the original radiation value with the corresponding cluster center.

Step 5. Initialize the center, width and weights of the RBFNN using the ABWO algorithm.

Step 6. The data of similar days are input into RBFNN as training samples, including daily average weather temperature, daily average weather relative humidity, daily average global horizontal radiation, daily average diffuse horizontal radiation and active power; ABWO algorithm updates the parameters of RBFNN to obtain the optimal parameters of the KM-SDA-ABWO-RBF model.

Step 7. Input the data of the optimal similarity day and the forecasting day as the test sample into RBFNN to obtain the photovoltaic power generation on the forecasting day.

3 Specific implementation process of PV power prediction

All the schemes were programmed in MATLAB version 2019 and run on a normal personal computer with an Intel(R) Core(TM) i7-9700 CPU and 16.0 GB RAM, under a Microsoft Windows 10.0 environment.

3.1 Data description

Effective forecasting needs to meet two requirements, one is the validity of forecasting methods and software, and the other is the reliability of historical meteorological data. The dataset for

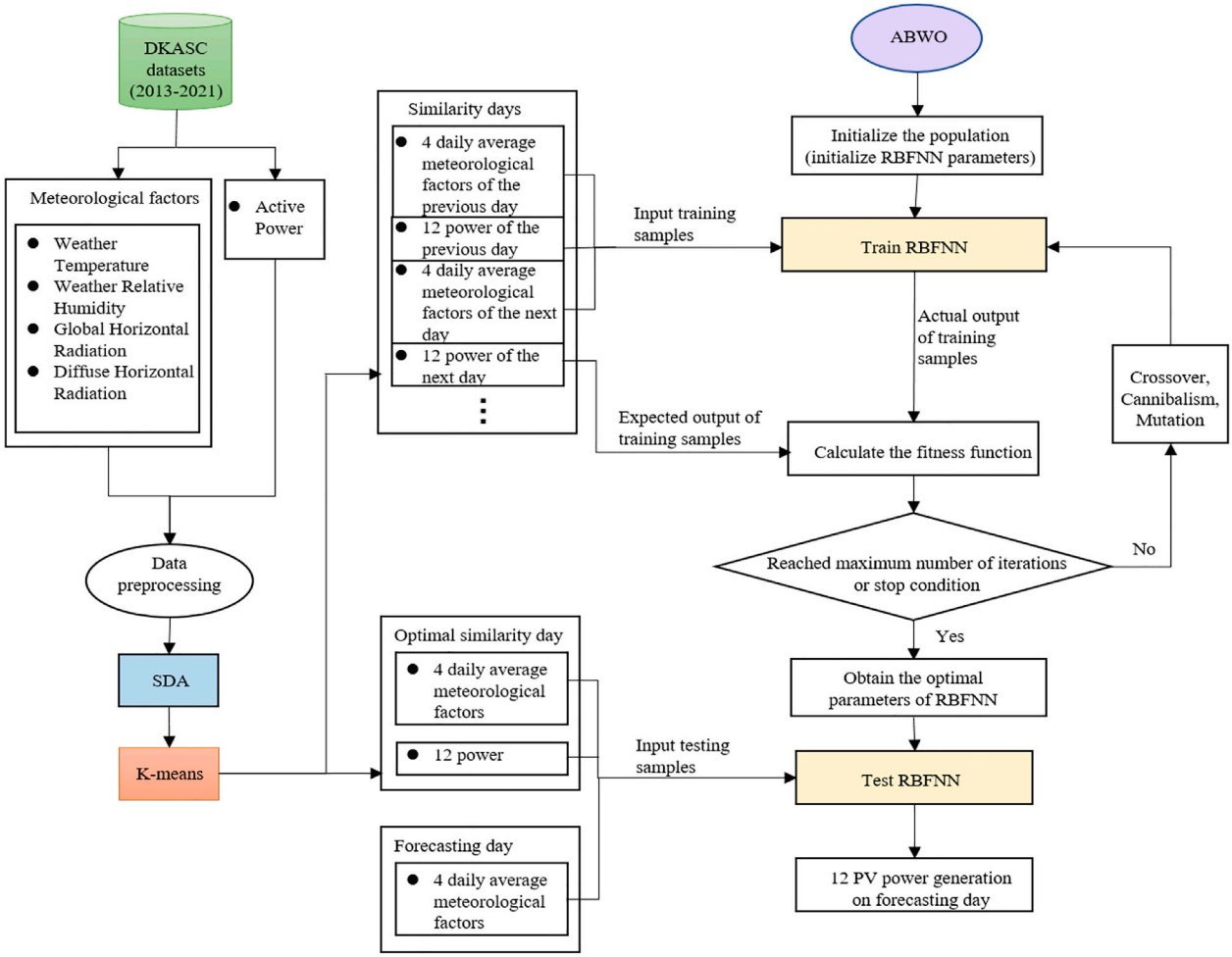


FIGURE 4
System architecture of the KM-SDA-ABWO-RBF model.

the experiment is from the Desert Knowledge Australia Solar Centre (DKASC), Alice Springs, (http://dkasolarcentre.com.au/download?location=alice-springs). The photovoltaic power plants use 60 monocrystalline panels and were installed in 2009 with an array rating of 10.5 kW. The dataset is from August 14, 2013 to November 28, 2021 with a time interval of 5 min. A large number of original datasets are inconvenient to organize, so this section extracted data from 7:30 to 18:30 with an interval of 1 h. The properties of the dataset include weather temperature, weather relative humidity, global horizontal radiation, diffuse horizontal radiation, and active power. The four seasons in Australia: spring days are from September to November; summer days are from December to February; autumn days are from March to May; winter days are from June to August. According to the time of the four seasons, the datasets from 2013 to 2021 are divided into four categories by season. Four days were arbitrarily selected from the four seasons in the dataset, that is, November 19, 2021, February 22, 2020, May 13, 2021 and June 29, 2021 as the forecasting day, respectively. The similar days for four forecasting days were selected from the day from the beginning of the four types of datasets to the day before the forecasting day, respectively.

Data mining methods are very common in classification and regression problems (Hossain and Mahmood, 2020). The data mining methods used in this section are dataset selection, outlier processing, data normalization and correlation analysis. There may be errors in the process of data collection and transmission, such as metering failure, system disconnection for cabling works, system outage for new array connection and so on. In a word, there are incomplete and abnormal data in a large number of original datasets, so the data needs to be preprocessed. So when the data at a certain sample is empty, the empty value is replaced with the data at the previous sample. When data missing is severe, consider removing missing data for a whole day. Data

TABLE 1 PCC of meteorological factors and active power.

Meteorological factors	PCC
Weather temperature	0.22523
Weather relative humidity	−0.13469
Global horizontal radiation	0.91518
Diffuse horizontal radiation	0.87048

normalization refers to scaling each meteorological attribute and output power respectively and mapping them to the interval [-1, 1]. Through outlier processing and normalization, the quality of the data is improved, and can be better adapted to algorithms and neural network models.

In order to study the correlation between the four weather factors and active power, an arbitrary day was selected from the dataset to calculate the PCC of weather factors and active power.

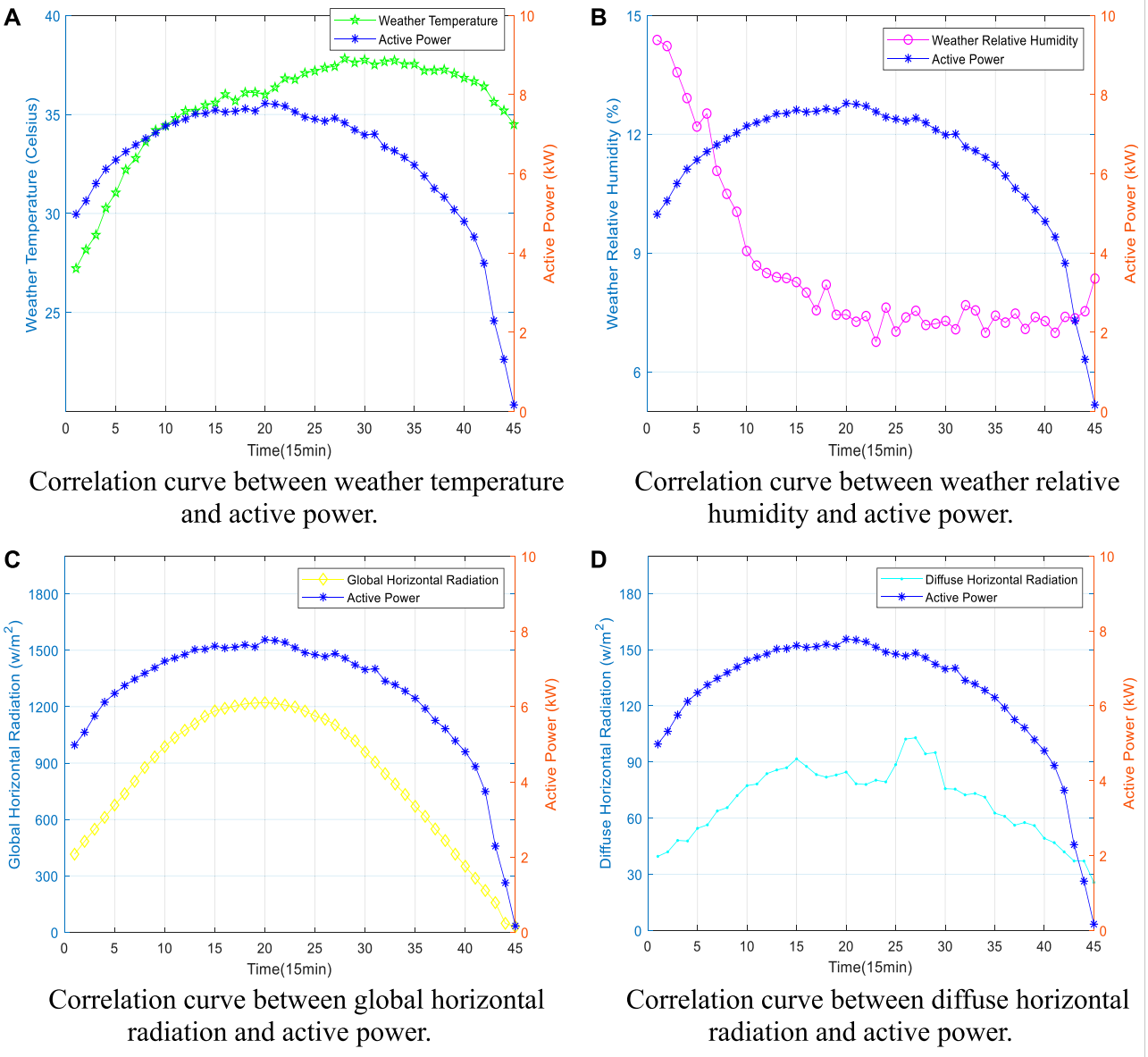


FIGURE 5
Correlation curve. **(A)** Between weather temperature and active power. **(B)** Between weather relative humidity and active power. **(C)** Between global horizontal radiation and active power; **(D)** Between diffuse horizontal radiation and active power.

TABLE 2 The optimal similarity day.

Season	Spring	summer	autumn	winter
Optimal similarity day	November 18, 2021	February 19, 2020	May 13, 2019	June 30, 2020

The specific calculation results are shown in Table 1. Moreover, Figures 5A–D shows the correlation curves of weather temperature, weather relative humidity, global horizontal radiation, diffuse horizontal radiation and photovoltaic power, respectively. The time point interval for this day is 15 min. From the correlation analysis, it is known that among the multivariate meteorological factors affecting photovoltaic power generation, the correlation coefficient of global horizontal radiation is the highest, while the correlation coefficient of weather temperature is the lowest. In descending order of influence on power generation are global horizontal radiation, diffuse horizontal radiation, weather relative humidity and weather temperature. Therefore, the predicted days and the similarity days have the most consistency in radiation intensity, followed by temperature and humidity.

After determining the correlation between meteorological factors and power generation, the SDA model is used to determine the similarity days and the optimal similarity day. For each forecasting day, the weighted Euclidean metric values are sorted in ascending order, and the first 56 days are selected as the similarity days and the first 1 day as the optimal similarity day. The optimal similarity days of the 4 forecasting days are shown in Table 2. The K-means clustering method sets three cluster centers for the two irradiances respectively. It takes temperature, humidity, sunny/cloudy/rainy, irradiance all into account to get complete weather information. Thereafter, taking the weather information of the similarity days as the input value, the KM-SDA-ABWO-RBF model was used to establish the functional relationship between the weather factors and the output power of photovoltaic power generation, and finally the output power value was obtained.

3.2 Experimental setup and evaluation metrics

This paper compares the proposed SDA-ABWO-RBF model with SDA-BWO-RBF, SDA-PSO-RBF, SDA-RBF, SDA-ELM, SDA-BPNN, RBF, ELM, BPNN. The population size of all metaheuristic algorithms is 50 and the number of iterations is 250. Acceleration constants of PSO is [2.1, 2.1], and inertia weights is [0.9, 0.6]. In addition, the number of hidden neurons is 4. The number of input neurons and output neurons are 16 and 12, respectively.

The performance of the SDA-ABWO-RBF model is evaluated by four evaluation indicators, namely RMSE, MAE, the standard deviation of error (SDE) (Eseye et al., 2018) and the coefficient of determination (R^2). Four metrics are expressed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{p,i} - y_{a,i}\right)^2} \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{p,i} - y_{a,i}\right| \tag{12}$$

$$SDE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{p,i} - y_{a,i} - \bar{e}\right)^2} \tag{13}$$

$$R^2 = \frac{\left(n\sum_{i=1}^{n} y_{p,i}y_{a,i} - \sum_{i=1}^{n} y_{p,i}\sum_{i=1}^{n} y_{a,i}\right)^2}{\left(n\sum_{i=1}^{n} y_{p,i}^2 - \left(\sum_{i=1}^{n} y_{p,i}\right)^2\right)\left(n\sum_{i=1}^{n} y_{a,i}^2 - \left(\sum_{i=1}^{n} y_{a,i}\right)^2\right)} \tag{14}$$

$$\bar{e} = \frac{1}{n}\sum_{i=1}^{n}\left(y_{p,i} - y_{a,i}\right) \tag{15}$$

where $y_{p,i}$ and $y_{a,i}$ are the predicted output and actual output at ith hour; n is the number of predicted samples, which is 12.

4 Prediction results and discussion

In order to verify the superiority of the proposed KM-SDA-ABWO-RBF model in PV power generation prediction, this paper compares the model with 8 prediction models (SDA-ABWO-RBF, SDA-BWO-RBF, SDA-RBF, SDA-ELM, SDA-BPNN, RBF, ELM and BPNN). The experiment selects 4 days, spring, summer, autumn and winter, and 12 forecast points per day for forecasting. The optimized number of neurons in the hidden layer is 4 in both ABWO-RBF and BWO-RBF in four seasons. The prediction results obtained from the experiments are discussed from two aspects. On the one hand, four metrics (RMSE, MAE, SDE and R^2) are used to evaluate the accuracy of the predicted values; on the other hand, each prediction model is run 10 times independently, and the stability of the prediction results of the 10 runs is compared on the four metrics sex.

4.1 The accuracy of the forecast results

Figures 6A–D shows the predicted power output curve of the 9 models in different seasons. It can be seen that the predicted values of the KM-SDA-ABWO-RBF model and the ABWO-RBF model are closer to the actual values, while the predicted outputs

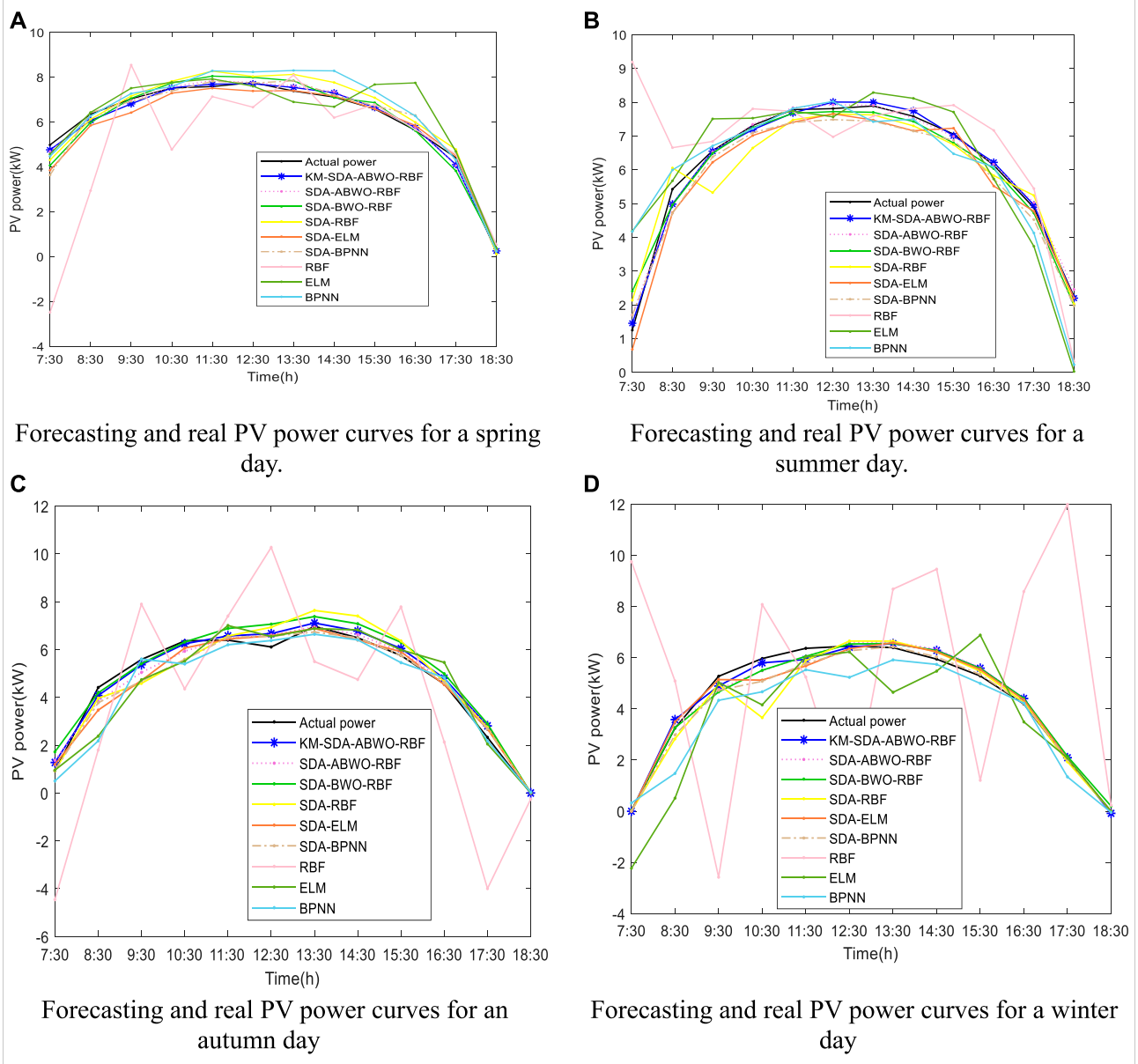


FIGURE 6
Forecasting and real PV power curves. **(A)** A spring day. **(B)** A summer day. **(C)** An autumn day. **(D)** A winter day.

of the other models are far from the actual values. Figures 7A–D shows the predicted error curve of 9 models in different seasons. Similarly, the error of the proposed model is less volatile around 0. Figure 8 shows the correlation between the predicted power of the proposed model and the actual power in four seasons. It is obvious from Figure 8 that R^2 of the proposed model in the four seasons is 0.9929, 0.9934, 0.9862 and 0.9881, respectively. Its value is very close to 1, showing the superiority of the model for prediction. summer in the selected dataset has the largest R^2 and the most accurate forecast result.

Figure 9A shows predicted RMSE of 9 models for 4 forecasting days. Figure 9B shows predicted MAE of 9 models for 4 forecasting days. Figure 9C shows predicted *SDE* of 9 models for 4 forecasting days. Figure 9D shows predicted R^2 of 9 models for 4 forecasting days. RMSE, *MAE*, *SDE* are several types of errors; the smaller the predicted error value, the higher the prediction accuracy of the model. R^2 represents the correlation between two variables; a larger value of R^2 indicates a more reliable model. From the figure it can be concluded that the proposed model achieved the smallest *RMSE*,

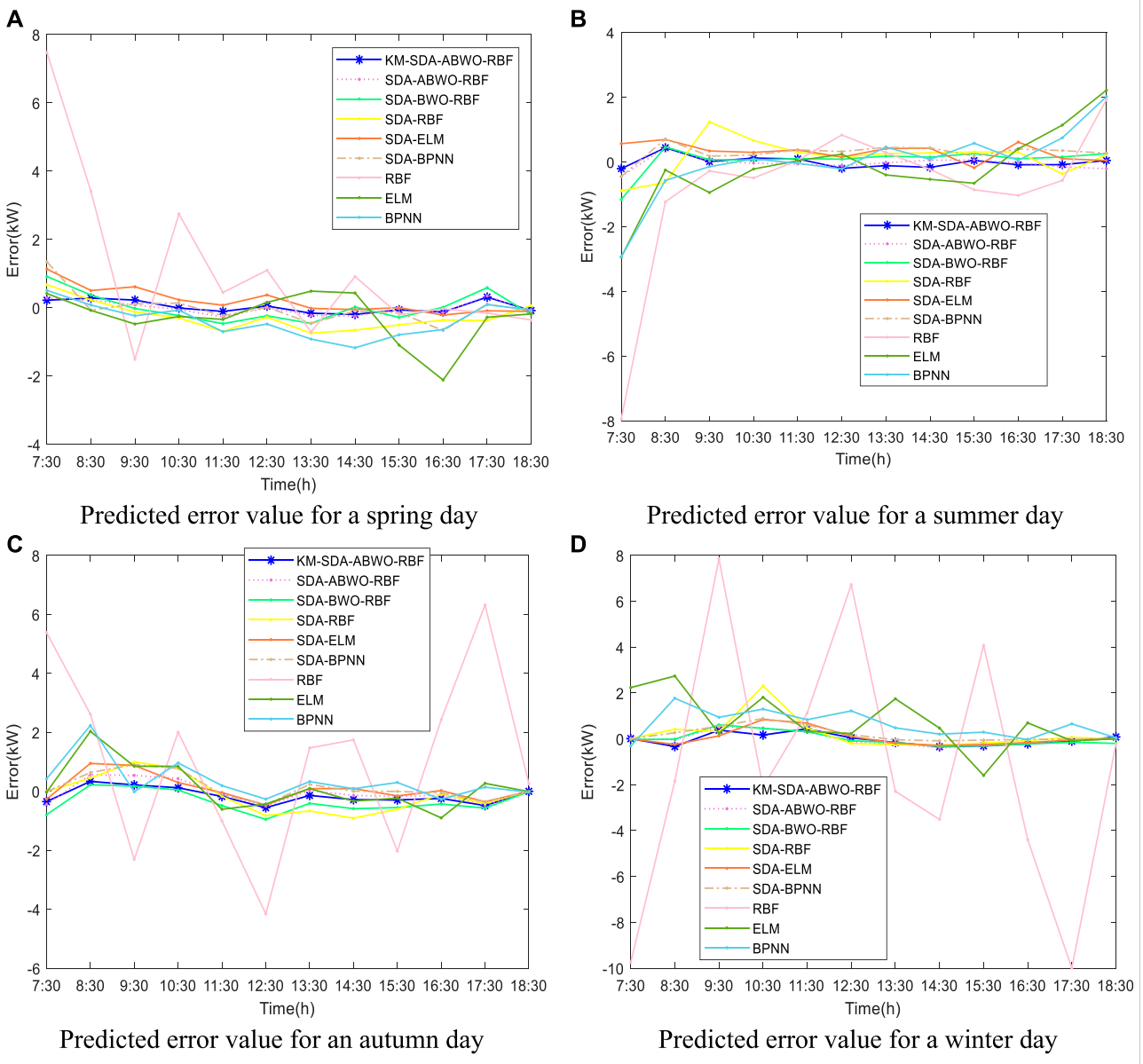


FIGURE 7
Predicted error value. **(A)** A spring day.

MAE, *SDE*, and the largest R^2 for each forecasting day. The prediction accuracy of SDA-ABWO-RBF and SDA-BWO-RBF is second only to the proposed model and higher than the other 6 traditional models that lack metaheuristic optimization. It proves that the metaheuristic algorithm to optimize neural network parameters greatly improves the prediction performance of traditional neural network models. In addition, SDA-ELM outperforms SDA-BPNN and SDA-RBF; SDA-RBF and RBF exhibit the worst performance. The reason for this result is that the parameters of the ELM are random so that the predicted output has a wide range. Therefore, it is possible to obtain more accurate output power values. Traditional BPNN and RBF have a high probability of leading to early convergence. Overall, it can be seen that on each evaluation metric, SDA-RBF performed better than RBF, SDA-ELM performed better than ELM, and SDA-BPNN performed better than BPNN. It can be shown that the SDA method greatly improves the prediction performance.

Figures 10A–D shows Taylor diagram of 9 models in different seasons. It visually presents the performance of 9 models on Standard Deviation, RMSE and R^2. KM-SDA-ABWO-RBF is closest to the observation in all four seasons,

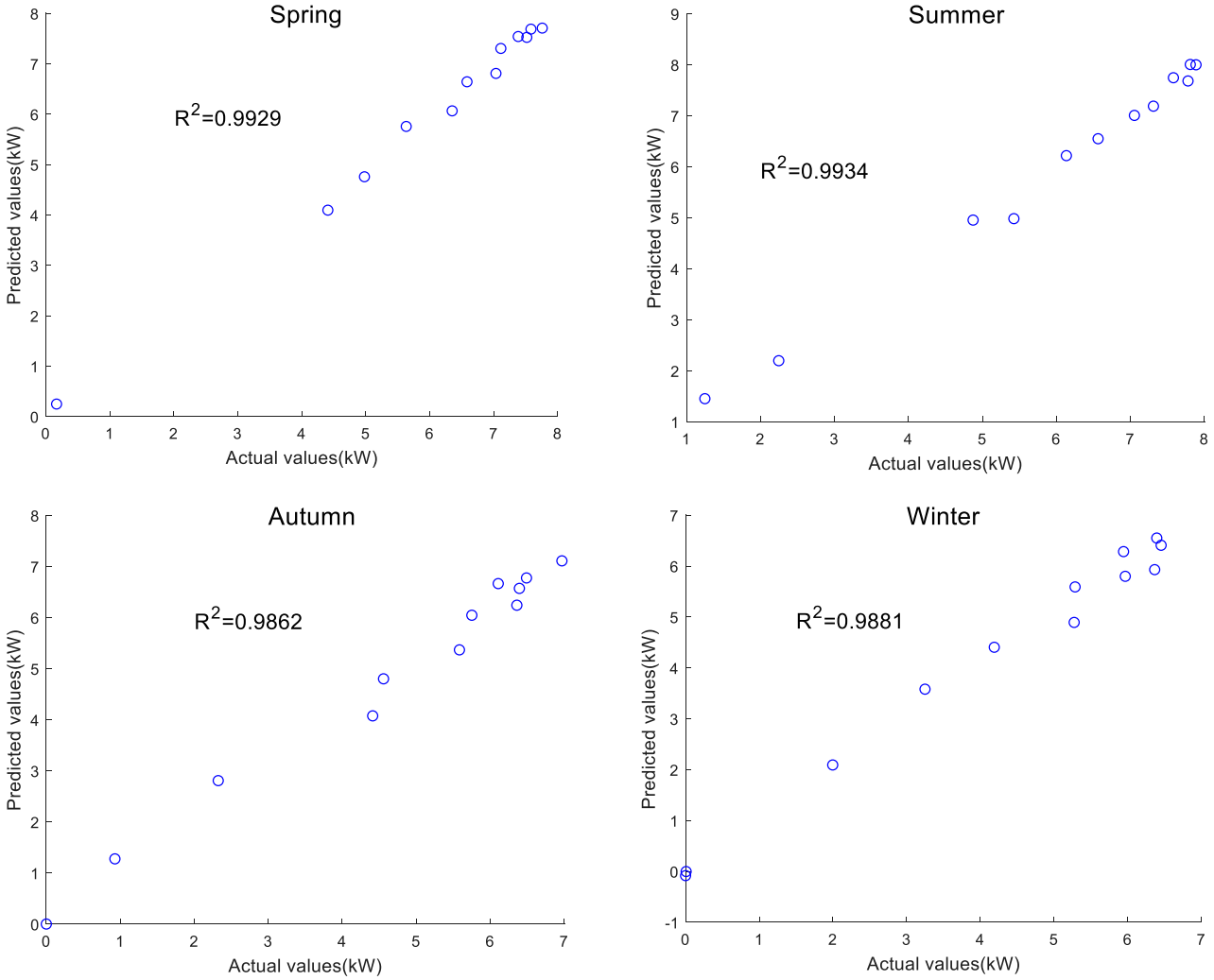


FIGURE 8
The correlation between the predicted power of KM-SDA-ABWO-RBF and the actual power.

while RBF is farthest from observation. Accordingly, it validates the accuracy of KM-SDA-ABWO-RBF model on three metrics.

Table 3 specifically describes the mean values of RMSE, MAE, SDE and R^2 for the four seasons. The mean values of RMSE, MAE, SDE and R^2 of the proposed model were 0.227,275, 0.19175, 0.215,725 and 0.99015, respectively. KM-SDA-ABWO-RBF has a 13%, 8% and 17% decrease in RMSE, MAE and SDE, and a 0.4% increase in R^2 compared to SDA-ABWO-RBF. It shows that adding K-means algorithm can further improve the prediction accuracy. In addition, the errors of KM-SDA-ABWO-RBF, SDA-ABWO-RBF and SDA-BWO-RBF were smaller than those of SDA-RBF, and the correlation with the actual value was greater. It illustrates the reliability of the combination of metaheuristics and machine learning methods. Again, the prediction results of KM-SDA-ABWO-RBF and SDA-ABWO-RBF were more accurate than SDA-BWO-RBF. The effectiveness of the improved BWO algorithm is proved. In a word, KM-SDA-ABWO-RBF achieves the best value for all forecasting days on all evaluation metrics. In the PV power prediction, the experimental results verified that the KM-SDA-ABWO-RBF model outperforms the other models in terms of robustness, effectiveness and accuracy.

4.2 The stability of the forecast results

Figures 11A,B show the RMSE, MAE, SDE, and R^2 values for 10 runs of 9 models for a spring day. Figures 12A,B show the RMSE, MAE, SDE, and R^2 values for 10 runs of 9 models for a summer day. Figures 13A–D show the RMSE, MAE, SDE, and R^2 values for 10 runs of 9 models for an autumn day. Figures 14A–D show the RMSE, MAE, SDE, and R^2 values for 10 runs of

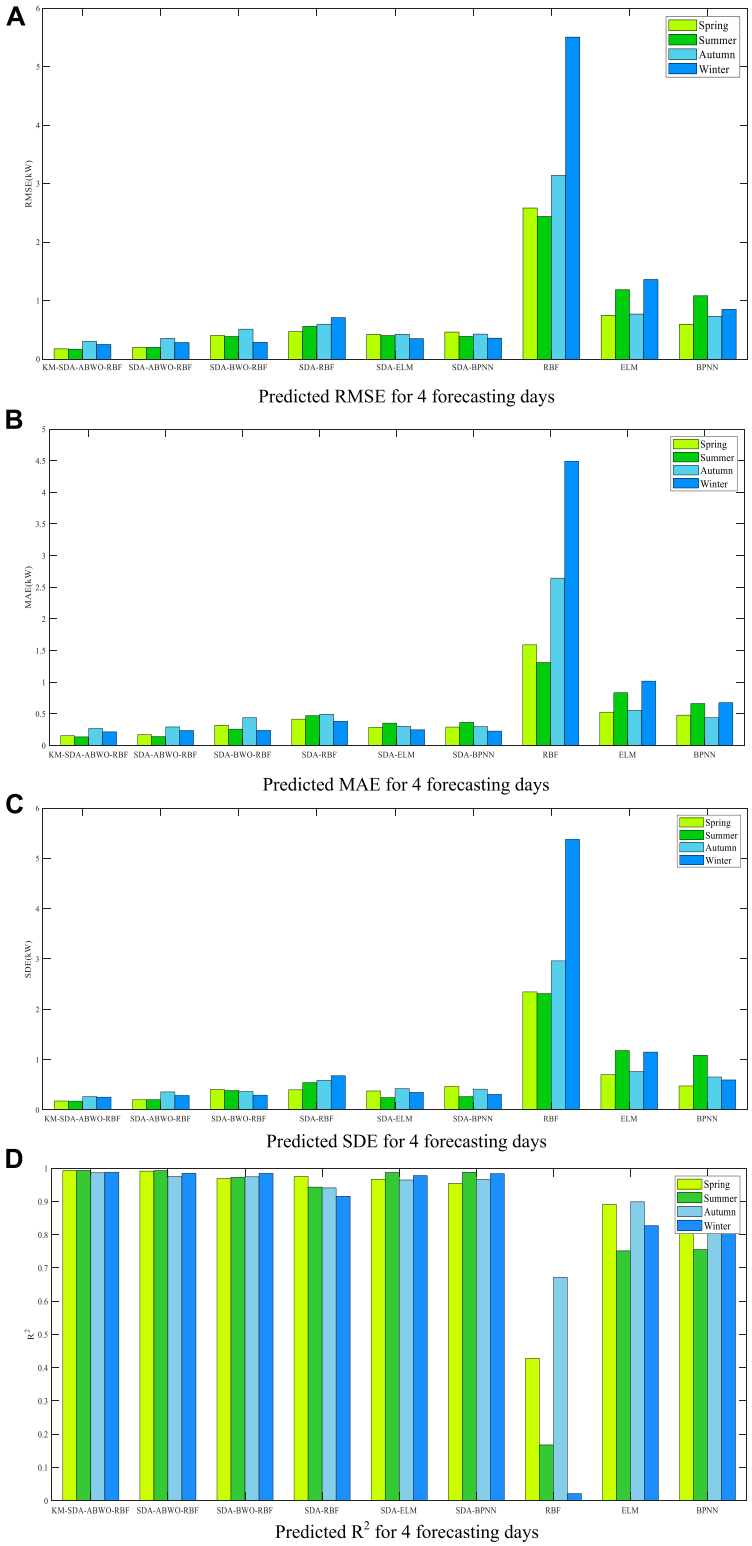


FIGURE 9
Predicted **(A)** RMSE. **(B)** MAE. **(C)** SDE. **(D)** R^2 for 4 forecasting days.

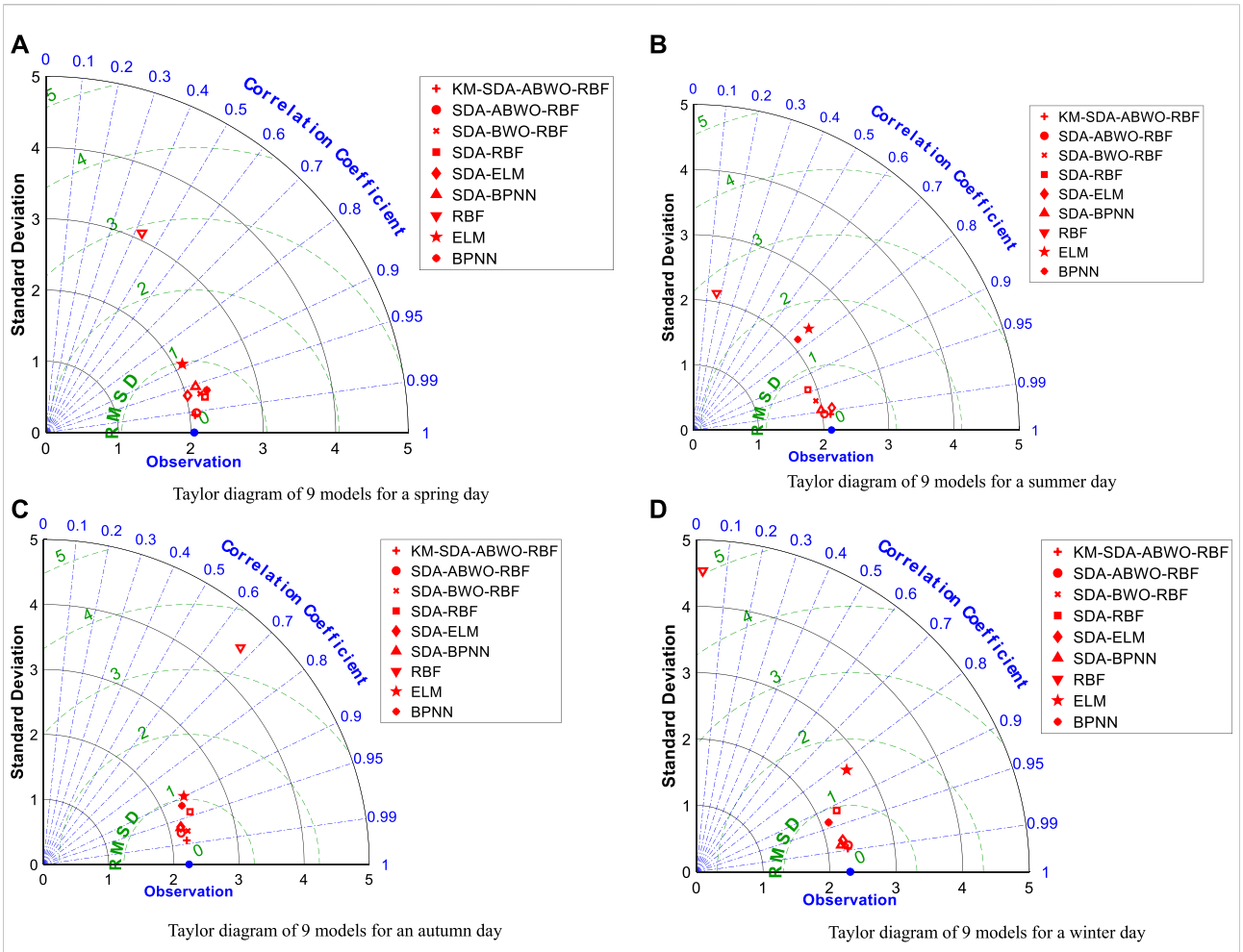


FIGURE 10
Taylor diagram of 9 models. **(A)** A spring day. **(B)** A summer day. **(C)** An autumn day. **(D)** A winter day.

TABLE *3 Average value of RMSE, MAE, SDE and R^2 in four seasons.*

Model	RMSE	MAE	SDE	R^2
KM-SDA-ABWO-RBF	0.227,275	0.19175	0.215,725	0.99015
SDA-ABWO-RBF	0.261,695	0.20815	0.2603	0.98595
SDA-BWO-RBF	0.400,383	0.312,025	0.361,625	0.975,025
SDA-RBF	0.586,443	0.43995	0.550,675	0.943,675
SDA-ELM	0.402,948	0.29595	0.346,525	0.974
SDA-BPNN	0.412,488	0.294,275	0.361,975	0.973
RBF	3.421	2.508,375	3.2487	0.32185
ELM	1.01976	0.73165	0.9475	0.841,875
BPNN	0.818,888	0.56285	0.701,475	0.894,425

9 models for a winter day. Taking a spring day as an example, it can be seen from Figures 26, 27 that the violin plots of SDA-ELM, SDA-BPNN, ELM and BPNN are narrower and longer, while KM-SDA-ABWO-RBF, SDA-ABWO-RBF and SDA-BWO-RBF are flatter and wider. It shows that the RMSE and MAE values of KM-SDA-ABWO-RBF, SDA-ABWO-RBF and SDA-BWO-RBF are closer and more stable. The SDA-RBF and RBF models have the highest stability, with the same values for 10 runs but the lowest accuracy. Roughly, the stability differences of KM-SDA-ABWO-RBF, SDA-ABWO-RBF and SDA-BWO-RBF are small, and the prediction results of the three models all have high stability and accuracy. However, SDA-BWO-RBF has higher stability than SDA-ABWO-RBF, which is more stable than KM-SDA-ABWO-RBF. Stability is reflected in the length of the box in Figures 11A,B. It can be clearly found from Figures 11A,B that SDA-ELM and SDA-BPNN are more stable than ELM and BPNN, respectively. It verifies that SDA can not only improve the prediction accuracy of PV generation power prediction, but also enhance the stability. Similarly, the same conclusion is reached for summer, autumn and winter.

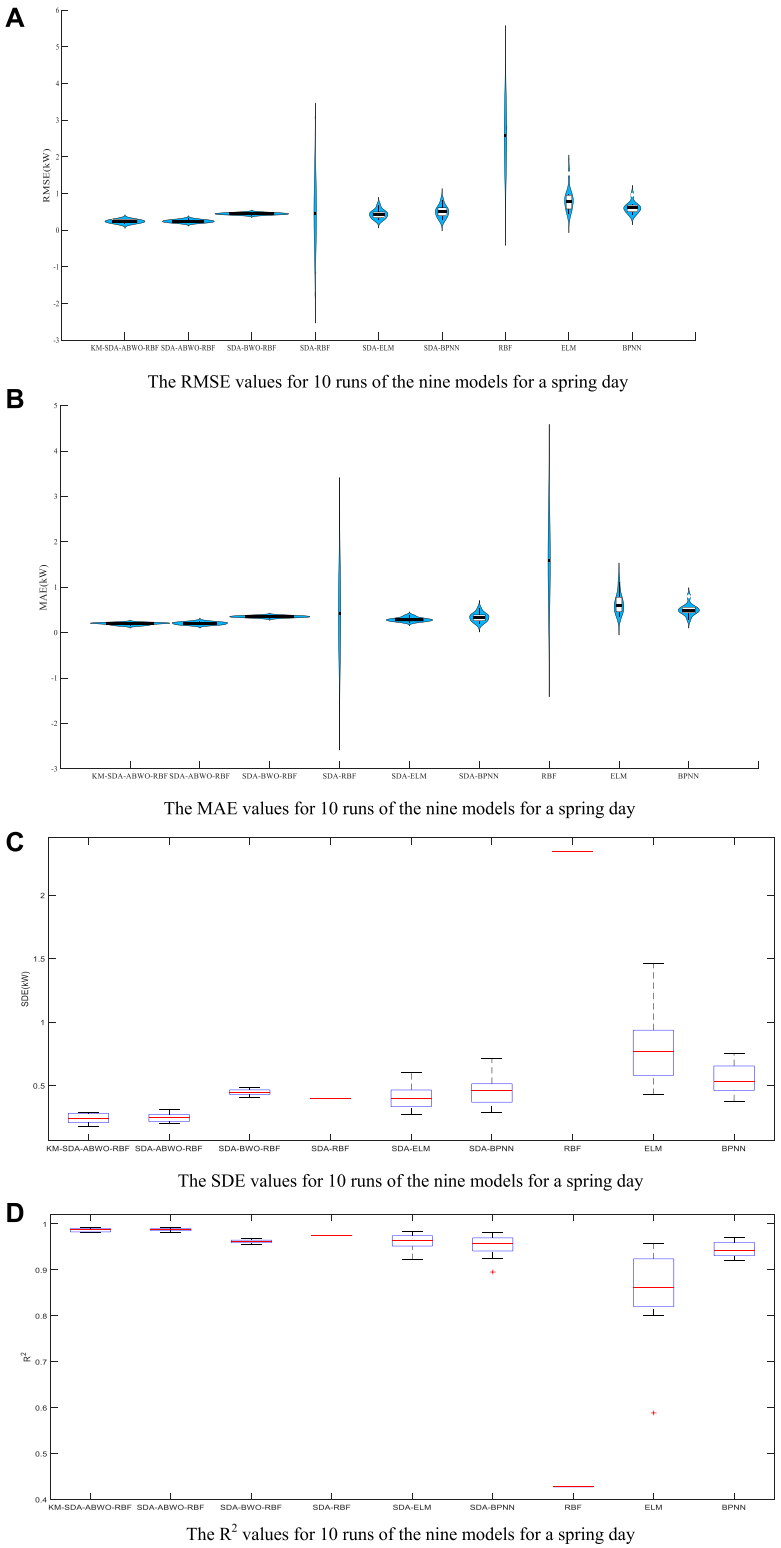


FIGURE 11
(A–D) shows the RMSE, MAE, SDE, and R^2 values for 10 runs of 9 models for a spring day.

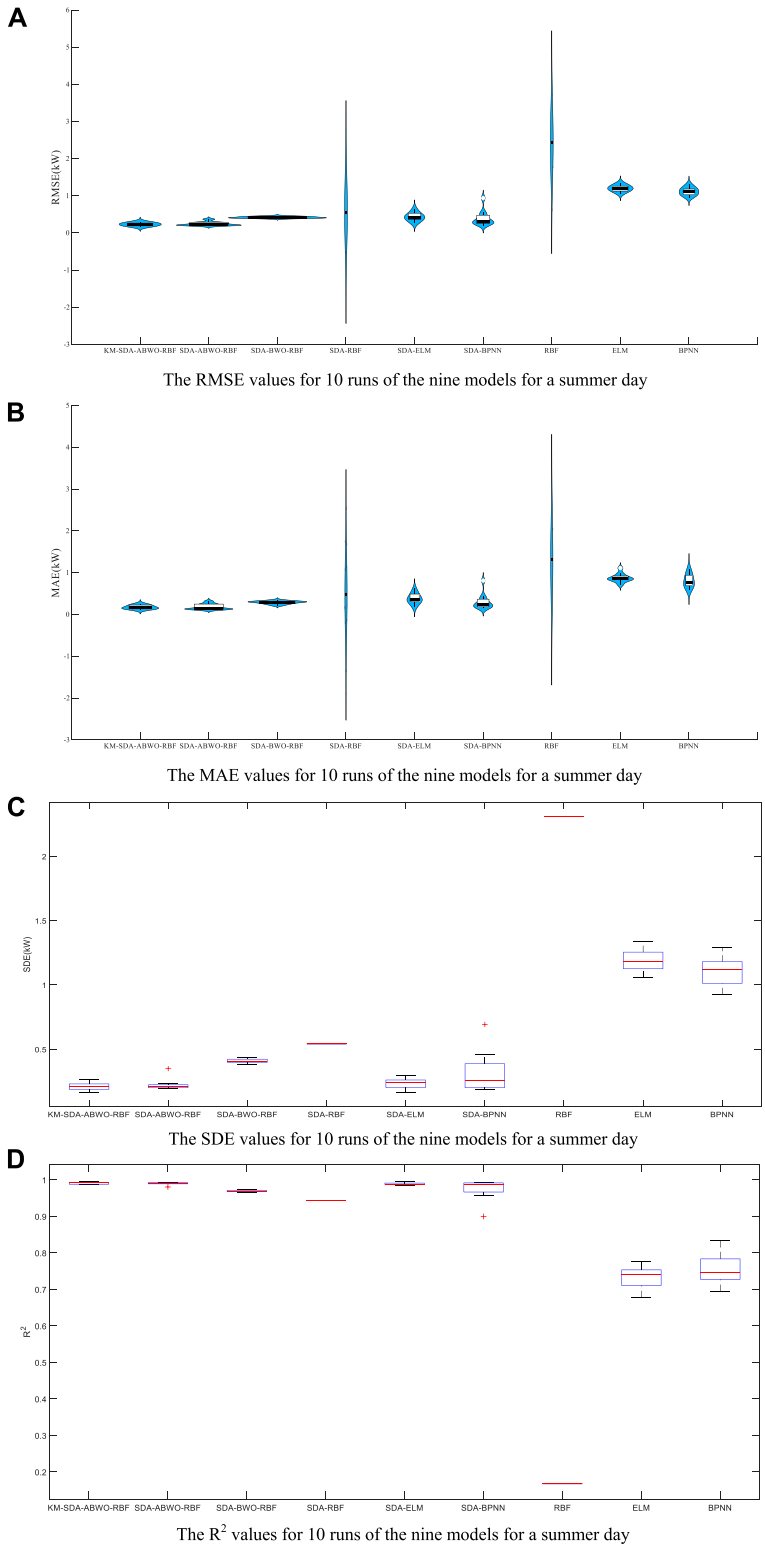


FIGURE 12
(A–D) show the RMSE, MAE, SDE, and R^2 values for 10 runs of 9 models for a summer day.

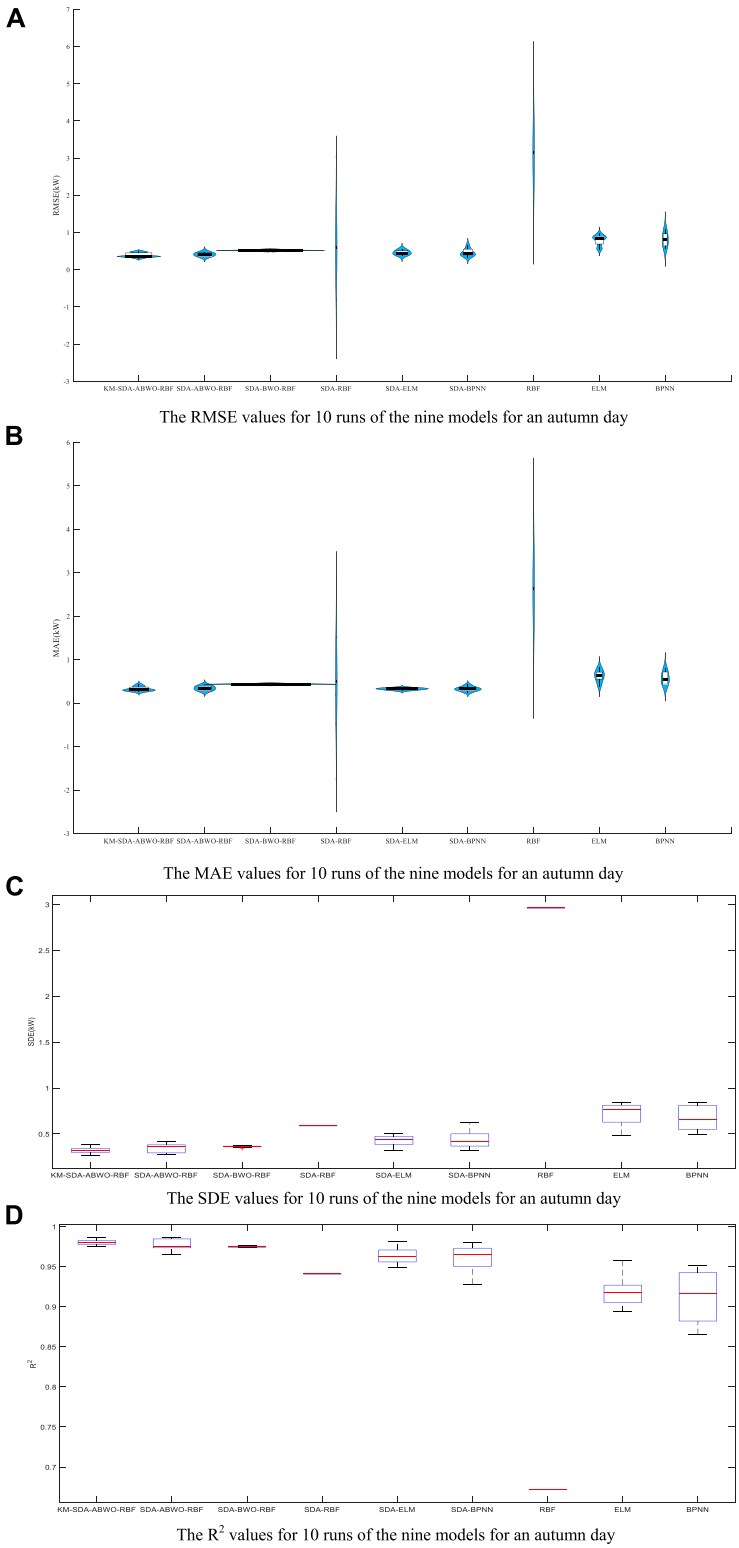


FIGURE 13
(A–D) show the RMSE, MAE, SDE, and R^2 values for 10 runs of 9 models for an autumn day.

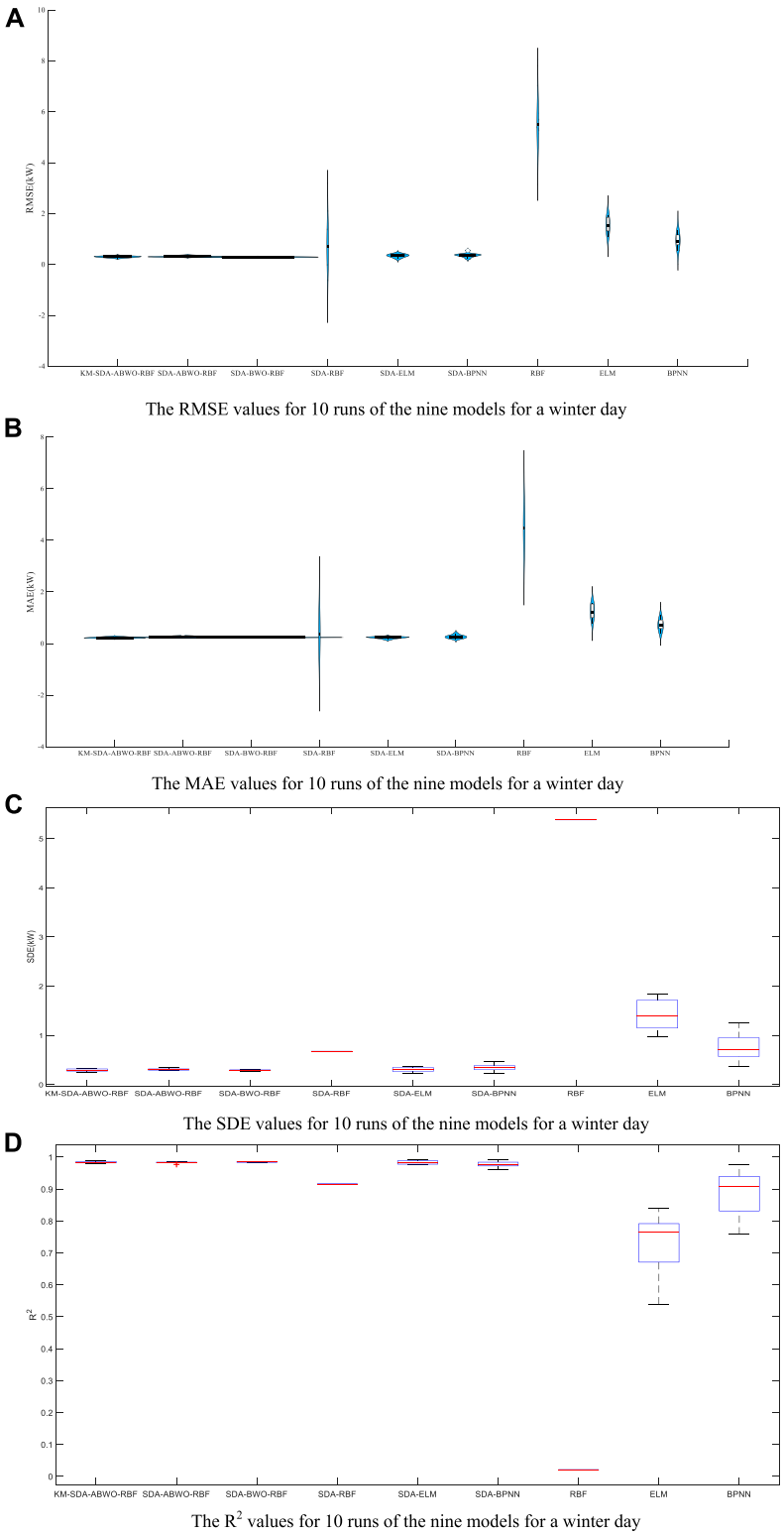


FIGURE 14
(A–D) show the RMSE, MAE, SDE, and R^2 values for 10 runs of 9 models for a winter day.

5 Conclusion and future work

This paper develops a KM-SDA-ABWO-RBF model, and utilizes the ABWO algorithm to train the parameters of RBFNNs to improve the generalization performance of RBFNNs and reduce the computational complexity. The ABWO algorithm enhances the search ability by introducing an adaptive factor, that is, the ability to search for optimal neural network parameters. Since the structure and parameters of the network determine the performance of RBFNNs, constructing the ABWO-RBF algorithm improves the performance and stability of the neural network. In addition, SDA and K-means methods are introduced to obtain historical data with strong correlation with the forecast day, which improves the overall accuracy of ABWO-RBF neural network in photovoltaic power prediction. The experimental results demonstrate the simplicity and high efficiency of the proposed model in the prediction of PV power generation. In the future, the hybrid model is still valid for forecast of the PV power output. Deep learning methods used in this field are expected to have higher prediction accuracy. The model in this paper can be used can be used in power plants to stably and efficiently regulate photovoltaic power generation in the future.

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.

Funding

This research is funded by National Natural Science Foundation of China under Grant No. 62066005, U21A20464, Project of the Guangxi Science and Technology under Grant No. AD21196006.

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

Abualigah, L., and Diabat, A. (2021). Advances in sine cosine algorithm: A comprehensive survey. *Artif. Intell. Rev.* 54 (4), 2567–2608. doi:10.1007/s10462-020-09909-3

Akhter, M. N., Mekhilef, S., Mokhlis, H., and Mohamed Shah, N. (2019). Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. *IET Renew. Power Gener.* 13 (7), 1009–1023. doi:10.1049/iet-rpg.2018.5649

Bugała, A., Zaborowicz, M., Boniecki, P., Janczak, D., Koszela, K., Czekala, W., et al. (2018). Short-term forecast of generation of electric energy in photovoltaic systems. *Renew. Sustain. Energy Rev.* 81, 306–312. doi:10.1016/j.rser.2017.07.032

Chen, C., Duan, S., Cai, T., and Liu, B. (2011). Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Sol. energy* 85 (11), 2856–2870. doi:10.1016/j.solener.2011.08.027

Chen, S., Cowan, C. F. N., and Grant, P. M. (1991). Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans. Neural Netw.* 2 (2), 302–309. doi:10.1109/72.80341

Cheng, M. Y., and Prayogo, D. (2014). Symbiotic organisms search: A new metaheuristic optimization algorithm. *Comput. Struct.* 139, 98–112. doi:10.1016/j.compstruc.2014.03.007

Dolara, A., Grimaccia, F., Leva, S., Mussetta, M., and Ogliari, E. (2015). A physical hybrid artificial neural network for short term forecasting of PV plant power output. *Energies* 8 (2), 1138–1153. doi:10.3390/en8021138

Dong, N., Chang, J. F., Wu, A. G., and Gao, Z. K. (2020). A novel convolutional neural network framework based solar irradiance prediction method. *Int. J. Electr. Power & Energy Syst.* 114, 105411. doi:10.1016/j.ijepes.2019.105411

Du, P., Wang, J., Yang, W., and Niu, T. (2018). Multi-step ahead forecasting in electrical power system using a hybrid forecasting system. *Renew. energy* 122, 533–550. doi:10.1016/j.renene.2018.01.113

Duvvuri, S. P., and Anmala, J. (2019). Fecal coliform predictive model using genetic algorithm-based radial basis function neural networks (GA-RBFNNs). *Neural comput. Appl.* 31 (12), 8393–8409. doi:10.1007/s00521-019-04520-2

Er, M. J., Wu, S., Lu, J., and Hock Lye Toh (2002). Face recognition with radial basis function (RBF) neural networks. *IEEE Trans. Neural Netw.* 13 (3), 697–710. doi:10.1109/tnn.2002.1000134

Eseye, A. T., Zhang, J., and Zheng, D. (2018). Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information. *Renew. energy* 118, 357–367. doi:10.1016/j.renene.2017.11.011

Feng, H. M. (2006). Self-generation RBFNs using evolutional PSO learning. *Neurocomputing* 70 (1-3), 241–251. doi:10.1016/j.neucom.2006.03.007

Ghayekhloo, M., Ghofrani, M., Menhaj, M. B., and Azimi, R. (2015). A novel clustering approach for short-term solar radiation forecasting. *Sol. Energy* 122, 1371–1383. doi:10.1016/j.solener.2015.10.053

Ghimire, S., Deo, R. C., Downs, N. J., and Raj, N. (2018). Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and Reanalysis atmospheric products in solar-rich cities. *Remote Sens. Environ.* 212, 176–198. doi:10.1016/j.rse.2018.05.003

Gonzalez, B., Valdez, F., Melin, P., and Prado-Arechiga, G. (2015). Fuzzy logic in the gravitational search algorithm for the optimization of modular neural networks in pattern recognition. *Expert Syst. Appl.* 42 (14), 5839–5847. doi:10.1016/j.eswa.2015.03.034

Han, H., Wu, X., Zhang, L., Tian, Y., and Qiao, J. (2017). Self-organizing RBF neural network using an adaptive gradient multiobjective particle swarm optimization. *IEEE Trans. Cybern.* 49 (1), 69–82. doi:10.1109/tcyb.2017.2764744

Hayyolalam, V., and Kazem, A. (2020). Black widow optimization algorithm: A novel meta-heuristic approach for solving engineering optimization problems. *Eng. Appl. Artif. Intell.* 87, 103249–103249.28. doi:10.1016/j.engappai.2019.103249

Hossain, M. S., and Mahmood, H. (2020). Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. *IEEE Access* 8, 172524–172533. doi:10.1109/access.2020.3024901

Huang, G. B., Saratchandran, P., and Sundararajan, N. (2005). A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation. *IEEE Trans. Neural Netw.* 16 (1), 57–67. doi:10.1109/tnn.2004.836241

Huang, G. B., Saratchandran, P., and Sundararajan, N. (2004). An efficient sequential learning algorithm for growing and pruning RBF (GAP-RBF) networks. *IEEE Trans. Syst. Man. Cybern. B* 34 (6), 2284–2292. doi:10.1109/tsmcb.2004.834428

Justus, J. J., and Anuradha, M. (2022). A golden eagle optimized hybrid multilayer perceptron convolutional neural network architecture-based three-stage mechanism for multiuser cognitive radio network. *Int. J. Commun.* 35 (4), e5054. doi:10.1002/dac.5054

Katooli, M. S., and Koochaki, A. (2020). Detection and classification of incipient faults in three-phase power transformer using DGA information and rule-based machine learning method. *J. Control Autom. Electr. Syst.* 31 (5), 1251–1266. doi:10.1007/s40313-020-00625-5

Lee, C. M., and Ko, C. N. (2009). Time series prediction using RBF neural networks with a nonlinear time-varying evolution PSO algorithm. *Neurocomputing* 73 (1-3), 449–460. doi:10.1016/j.neucom.2009.07.005

Li, Y., Su, Y., and Shu, L. (2014). An ARMAX model for forecasting the power output of a grid connected photovoltaic system. *Renew. Energy* 66, 78–89. doi:10.1016/j.renene.2013.11.067

Lin, P., Peng, Z., Lai, Y., Cheng, S., Chen, Z., and Wu, L. (2018). Short-term power prediction for photovoltaic power plants using a hybrid improved Kmeans-GRA-Elman model based on multivariate meteorological factors and historical power datasets. *Energy Convers. Manag.* 177, 704–717. doi:10.1016/j.enconman.2018.10.015

Liu, J., Fang, W., Zhang, X., and Yang, C. (2015). An improved photovoltaic power forecasting model with the assistance of aerosol index data. *IEEE Trans. Sustain. Energy* 6 (2), 434–442. doi:10.1109/tste.2014.2381224

Liu, Y., Li, Y., Li, G., Zhang, B., and Wu, G. (2005). "Constructive ensemble of RBF neural networks and its application to earthquake prediction," in International Symposium on Neural Networks, Chongqing, China, May 30–June 1, 2005 (Berlin, Heidelberg: Springer), 532–537.

Mahmud, K., Azam, S., Karim, A., Zobaed, S., Shanmugam, B., and Mathur, D. (2021). Machine learning based PV power generation forecasting in alice springs. *IEEE Access* 9, 46117–46128. doi:10.1109/access.2021.3066494

Mandal, P., Madhira, S. T. S., Meng, J., and Pineda, R. L. (2012). Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. *Procedia Comput. Sci.* 12, 332–337. doi:10.1016/j.procs.2012.09.080

Mandal, P., Senjyu, T., Urasaki, N., Funabashi, T., and Srivastava, A. K. (2007). A novel approach to forecast electricity price for PJM using neural network and similar days method. *IEEE Trans. Power Syst.* 22 (4), 2058–2065. doi:10.1109/tpwrs.2007.907386

Memar, S., Mahdavi-Meymand, A., and Sulisz, W. (2021). Prediction of seasonal maximum wave height for unevenly spaced time series by black widow optimization algorithm. *Mar. Struct.* 78 (10), 103005. doi:10.1016/j.marstruc.2021.103005

Mirjalili, S. A., Hashim, S. Z. M., and Sardroudi, H. M. (2012). Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm. *Appl. Math. Comput.* 218 (22), 11125–11137. doi:10.1016/j.amc.2012.04.069

Mirjalili, S., Mirjalili, S. M., and Lewis, A. (2014). Grey wolf optimizer. *Adv. Eng. Softw.* 69, 46–61. doi:10.1016/j.advengsoft.2013.12.007

Mirjalili, S. (2016). Sca: A sine cosine algorithm for solving optimization problems. *Knowledge-based Syst.* 96, 120–133. doi:10.1016/j.knosys.2015.12.022

Panahi, F., Ehteram, M., and Emami, M. (2021). Suspended sediment load prediction based on soft computing models and Black Widow Optimization Algorithm using an enhanced gamma test. *Environ. Sci. Pollut. Res.* 28 (35), 48253–48273. doi:10.1007/s11356-021-14065-4

Parsopoulos, K. E., and Vrahatis, M. N. (2002). Recent approaches to global optimization problems through Particle Swarm Optimization. *Nat. Comput.* 1 (2-3), 235–306. doi:10.1023/a:1016568309421

Qiao, J., Li, F., Yang, C., Li, W., and Gu, K. (2019). A self-organizing RBF neural network based on distance concentration immune algorithm. *IEEE/CAA J. Autom. Sin.* 7 (1), 276–291. doi:10.1109/jas.2019.1911852

Rashedi, E., Nezamabadi-Pour, H., and Saryazdi, S. (2009). Gsa: A gravitational search algorithm. *Inf. Sci.* 179 (13), 2232–2248. doi:10.1016/j.ins.2009.03.004

Reikard, G. (2009). Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Sol. energy* 83 (3), 342–349. doi:10.1016/j.solener.2008.08.007

Renewable Energy Policy (2017). Renewable energy policy network for the 21st century (REN21), renewables global status report (paris: Various editions). http://www.ren21.net/status-of-renewables/global-status-report/ (Accessed 27 January 2017).

Sobri, S., Koohi-Kamali, S., and Rahim, N. A. (2018). Solar photovoltaic generation forecasting methods: A review. *Energy Convers. Manag.* 156, 459–497. doi:10.1016/j.enconman.2017.11.019

Storn, R., and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* 11 (4), 341–359. doi:10.1023/a:1008202821328

Tightiz, L., Nasab, M. A., Yang, H., and Addeh, A. (2020). An intelligent system based on optimized ANFIS and association rules for power transformer fault diagnosis. *ISA Trans.* 103, 63–74. doi:10.1016/j.isatra.2020.03.022

Wang, J., Yang, W., Du, P., and Li, Y. (2018). Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system. *Energy* 148, 59–78. doi:10.1016/j.energy.2018.01.112

Wang, P., Zhou, Y., Luo, Q., Han, C., Niu, Y., and Lei, M. (2020). Complex-valued encoding metaheuristic optimization algorithm: A comprehensive survey. *Neurocomputing* 407, 313–342. doi:10.1016/j.neucom.2019.06.112

Wolpert, D. H., and Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* 1 (1), 67–82. doi:10.1109/4235.585893

Xie, T., Hao, Y., Hewlett, J., Rozycki, P., and Wilamowski, B. (2012). Fast and efficient second-order method for training radial basis function networks. *IEEE Trans. Neural Netw. Learn. Syst.* 23 (4), 609–619. doi:10.1109/tnnls.2012.2185059

Xu, R., Chen, H., and Sun, X. (2012). "Short-term photovoltaic power forecasting with weighted support vector machine," in 2012 IEEE International Conference on Automation and Logistics (IEEE), 248

Yang, X. S., and Deb, S. (2009).*Cuckoo search via Lévy flights. 2009 World congress on nature & biologically inspired computing (NaBIC)*. Ieee, 210

Zhang, Y., Beaudin, M., Taheri, R., Zareipour, H., and Wood, D. (2015). Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators. *IEEE Trans. Smart Grid* 6 (5), 2253–2262. doi:10.1109/tsg.2015.2397003

Zhou, Y., Zhou, N., Gong, L., and Jiang, M. (2020). Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine. *Energy* 204, 117894. doi:10.1016/j.energy.2020.117894

frontiers | Frontiers in Energy Research

TYPE Mini Review
PUBLISHED 20 September 2022
DOI 10.3389/fenrg.2022.997107

OPEN ACCESS

EDITED BY
Zhiqiang Geng,
Beijing University of Chemical Technology, China

REVIEWED BY
Lei Guo,
Tongren University, China

*CORRESPONDENCE
Zhang Yihang,
Luckyzhangyh@foxmail.com

SPECIALTY SECTION
This article was submitted to Process and Energy Systems Engineering, a section of the journal Frontiers in Energy Research

RECEIVED 18 July 2022
ACCEPTED 29 August 2022
PUBLISHED 20 September 2022

CITATION
Yihang Z (2022), Application of water-soluble polymer inhibitor in metal corrosion protection: Progress and challenges.
Front. Energy Res. 10:997107.
doi: 10.3389/fenrg.2022.997107

Application of water-soluble polymer inhibitor in metal corrosion protection: Progress and challenges

Zhang Yihang*

Chongqing Industry Polytechnic College, Chongqing, China

The use of corrosion inhibitors can effectively avoid the corrosion of metals and alloys, but the conventional organic/inorganic corrosion inhibitors have certain toxic and side effects and environmental pollution problems (including some expensive toxic reagents and catalysts), prompting researchers to turn their attention to the research of water-soluble polymer corrosion inhibitors with green environment friendly and low pollution. This paper reviews the action mechanism of polymer inhibitors, the research status of natural polymer inhibitors (chitosan-based inhibitors, cellulose and its derivatives inhibitors, other carbohydrate inhibitors, protein inhibitors) and synthetic polymer inhibitors (epoxy resin inhibitors, polyethylene glycol inhibitors, conductive polymer inhibitors), The purpose is to provide some reference for the development of eco-friendly metal coatings.

KEYWORDS

corrosion, inhibitor, polymer, mechanism, review

Introduction

Carbon steel, alloys (magnesium, aluminum alloys) and other metal materials are widely used in industry, agriculture, transportation, daily life and other fields because of their good plasticity, easy processability and excellent mechanical properties (Pludek, 1977; Bardal et al., 1993; Musa et al., 2012). However, under natural conditions, metal materials usually react with corrosive media such as water, Cl^-, SO_4^{2-}, O_2 in the surrounding environment, which changes their chemical properties, resulting in a large amount of metal components peeling off with the corrosive media. The essence of metal corrosion is the process of metal atoms losing electrons and being oxidized. Metal corrosion can be generally divided into chemical corrosion and electrochemical corrosion. Electrochemical corrosion is the most common and important type of metal corrosion. The corrosion of steel in humid air is the most prominent example of electrochemical corrosion. In the humid air, a thin water film will be adsorbed on the steel surface. If the water film is strongly acidic, hydrogen will be released from the electrons of H^+, and this electrochemical corrosion is called hydrogen evolution corrosion; If the water film is weakly acidic or neutral, more oxygen can be dissolved. At this time, O_2 gets electrons and precipitates OH^-. This electrochemical corrosion is called oxygen absorption corrosion

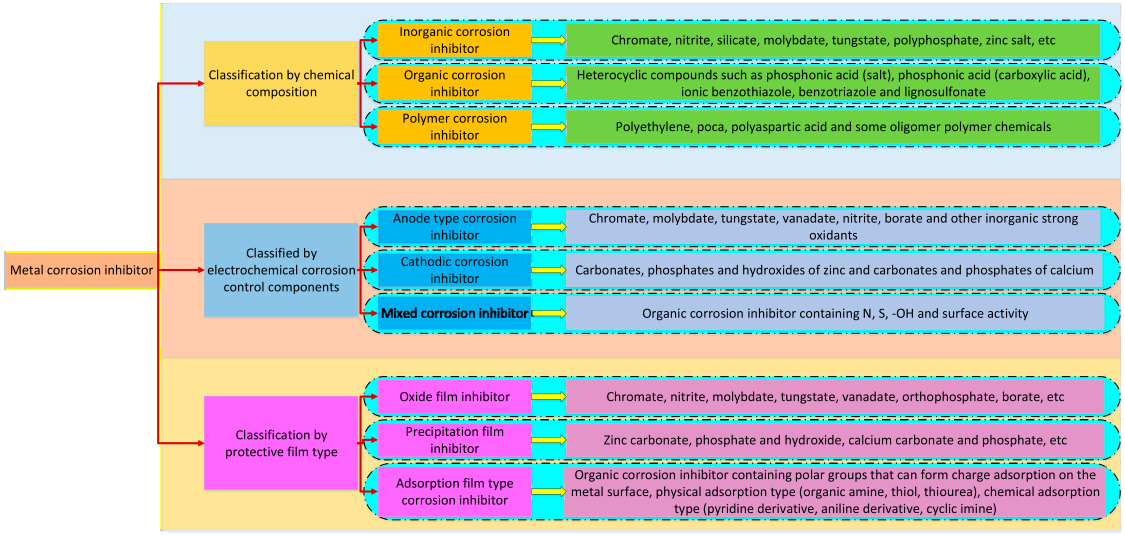


FIGURE 1
Type of metal corrosion inhibitor.

TABLE 1 Data obtained for water-soluble polymer as a corrosion inhibitor.

No.	Sub.	Env	Type	Adsor	Iso	Ref
1	Al	HCl + H_2SO_4	Anodic	Chemical–physical	Langmuir	Nwanonenyi et al. (2019)
2	Mg	HCl	Anodic	Chemical	Fruendlich	Hassan and Ibrahim (2021)
3	Cu	HCl	-	Physical	Langmuir	Sobhi and Eid (2018)
4	St	HCl	Cathodic	Chemical–physical	—	Wu et al. (2016)
5	St	NaCl	Anodic	Chemical	—	Ochoa et al. (2013)
6	St (C-Mn)	HCl	Cathodic	Chemical	Freundlich	Lahrour et al. (2019)
7	St	HCl	—	Chemical	Langmuir	Farag et al. (2018)
8	St	HCl	—	Chemical–physical	Langmuir	Rabizadeh and Asl (2019)
9	St	NaCl	—	Chemical	Langmuir	Shen et al. (2019)

and is the main cause of steel corrosion. According to the National Association of Corrosion Engineers (NACE), up to now, the global loss caused by corrosion is equivalent to about 3.4% of the world GDP, and safety accidents caused by corrosion continue to occur (Sheikh et al., 2021; Guo et al., 2020). In order to reduce the loss caused by metal corrosion, metal corrosion prevention methods such as surface modification, anodic oxidation, electroless plating or electroplating, and chemical coating have been developed in succession (Vijayan and Al-Maadeed, 2019; Grundmeier et al., 2000). The method of using heterocyclic compound organic coating to protect metal substrates has the advantages of low cost, fast construction, and excellent corrosion resistance. It is one of the fastest-growing anti-corrosion methods in recent decades, and is widely used in the field of metal protection. However, there are still many defects in the use of these compounds as corrosion inhibitors. The first is that these compounds are not environmentally friendly. Because most of the compound raw materials, solvents and catalysts are expensive and have toxic side effects, their application is greatly limited (Balgude and Sabnis, 2012; Guo et al., 2020; Tan et al., 2022). Polymer has the advantages of low cost, stable properties, non-toxic and so on. In recent years, it has gradually become an environmental protection inhibitor to replace the conventional heterocyclic inhibitor (Figure 1). Because the polymer contains various macromolecules and functional groups (–N=N–, R–NH_2, R–COOH, etc.), it can be used as both an anti-corrosion coating and a corrosion inhibitor (Table 1). Among them, amine, imine, amide polymers, vinyl and vinyl polymers account for 76.08% of polymer corrosion inhibitors (Fathima

Sabirneeza et al., 2015). In addition, there are heteroatoms in biopolymers or macromolecules formed by the degradation of active organisms (animals, plants, bacteria, microorganisms, etc.), which show a certain adsorption capacity, so as to achieve the purpose of preventing metal surface corrosion (Sushmitha and Rao, 2019). Therefore, this paper reviews the action mechanism of polymer inhibitors, the application status of biopolymer inhibitors (chitosan-based inhibitors, cellulose and its derivatives inhibitors, other carbohydrate inhibitors, protein inhibitors) and synthetic polymer inhibitors (epoxy resin inhibitors, polyethylene glycol inhibitors, conductive polymer inhibitors), The purpose is to provide some reference for the development of eco-friendly metal coatings.

Mechanism of polymer corrosion inhibitor

Metal corrosion is the result of galvanic cell reaction on the metal surface, which is the main cause of corrosion. Galvanic cell reaction usually includes anode reaction and cathode reaction. If corrosion inhibitor can inhibit either anode or cathode reaction, it will slow down the galvanic cell reaction process and weaken the corrosion rate of metal (Zeng and Chen, 2012; Jevremović et al., 2020). At present, the mechanism of corrosion inhibitor has not been fully understood, but the main corrosion inhibition principles are as follows:

(1) Physical/chemical adsorption at the metal/electrolyte interface;
(2) Electrostatic interaction of the protonated groups existing in corrosion inhibitor molecules and cations on surface of the metal matrix;
(3) Chemical interaction between unshared electrons (electrons on N, S, O atoms) and empty d orbitals in the electronic structure of metal atoms;
(4) By forming stable metal corrosion inhibitor complexes.

The action mechanism of polymer inhibitor is similar to heterocyclic inhibitor. It mainly forms a protective film for blocking aqueous medium by adsorbing on the surface of metal matrix. The stability and blocking effect of the protective film depend on the hydrophilicity and hydrophobicity of the polymer, which directly determines the adhesion ability of the polymer film. Polymer adsorption can be divided into physical adsorption, chemical adsorption and mixed adsorption. Among them, physical adsorption is mainly caused by the van der Waals force between the polymer inhibitor molecules and the surface of the metal matrix, while chemical adsorption uses the electron transfer and exchange between the polymer inhibitor and the surface atoms of the metal matrix to form the adsorption of chemical bonds. In addition, there are also mixed adsorption methods covering two adsorption characteristics (Hui et al., 2020; Verma et al., 2022). The adsorption process of polymer corrosion inhibitor can be divided into the following steps:

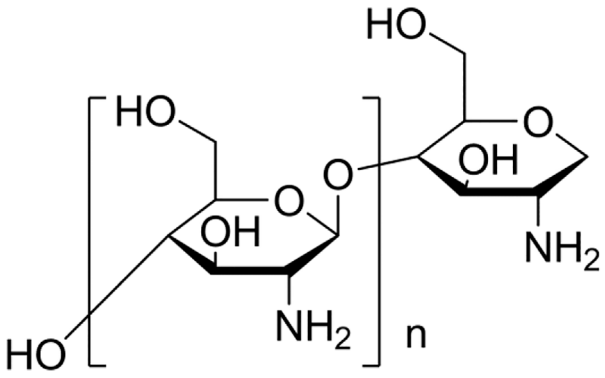


FIGURE 2
Molecular structure of chitosan.

(1) Polymer molecules replace acid ions and water molecules on the metal surface;
(2) Polymer molecules and acid ions form flocs;
(3) Polymer molecules use the hydroxyl, carboxyl and other organic functional groups contained in them to provide electrons to the empty orbit of the metal, form chelates with metal ions, and adsorb on the metal surface to form a hydrophobic protective film.

At the same time, in the process of polymer corrosion inhibitor, the metal corrosion efficiency will also be related to the chemical stability, coating uniformity and adsorption effect of the polymer. In addition, the density of donor electrons, electrolyte composition, pH value, apparent morphology of metal matrix, charge characteristics, etc. will also affect the effect of polymer corrosion inhibitor (Dariva and Galio, 2014).

Aqueous polymer inhibitor

Chitosan based corrosion inhibitor

As a linear polysaccharide, chitosan (Cht) contains various polar functional groups (amino, hydroxyl, acetyl, hydroxymethyl, etc.) on its linear polymer chain (Figure 2). These functional groups help to improve the solubility of chitosan in polar electrolytes and make chitosan easy to adsorb on metal surfaces. And because the acidity coefficient (pKa) value of amino group in chitosan is 6.5, it has a high degree of protonation in neutral and acidic environment. This property makes chitosan water-soluble and can form a good adhesion with negatively charged metal surfaces. In addition, chitosan can be

modified to improve its biological tolerance and corrosion resistance. It is because chitosan and its derivatives have the characteristics of non-toxic, environmental protection, non-sensitization, biocompatibility and degradability that they have been widely concerned in food, biomedicine, agriculture, environment, pharmaceutical and other industries.

Blanda et al. (2019) deposited a thin chitosan gel layer on the surface of 304 stainless steel sample through galvanic displacement reaction, and measured that the I_{corr} of steel before and after chitosan coating were 24.5 and 1.14 respectively μA cm^{-2}, while E_{corr} values were -0.161 and 0.0045 V respectively. The deposited chitosan coating successfully slowed down the corrosion rate of the matrix. It is found that the adhesion strength of chitosan coating is related to the type of solution, potential and other factors. For example, the adhesion of chitosan in the solution system is: methanol < isopropanol < ethanol solution. In addition, the thickness of chitosan coating is proportional to the potential. In order to improve the adhesion between chitosan and steel surface, polar functional groups can also be added to the main chain of chitosan for modification to strengthen its corrosion inhibition ability. Chauhan et al. (2019) used 4-amino-5-methyl-1,2,4-triazole-3-mercaptan for the first time to chemically modify chitosan corrosion inhibitor. Through the evaluation of its corrosion inhibition performance on the surface of carbon steel in 1 M HCl solution, it was found that the adsorption of the modified corrosion inhibitor conformed to the Langmuir isotherm, which showed physical adsorption and chemical adsorption. In addition, with the increase of inhibitor concentration, its polarization resistance gradually increases, which is a mixed inhibitor with cathode advantage, and the purpose of enhancing the corrosion inhibition effect can be achieved by improving the combination of Fe inhibitor system. Farhadian et al. (2020) used 1,3-propanesulfonic acid to sulfonate chitosan with different molecular weights to improve its solubility and overcome the incompatibility between hydrate inhibitor (HHIs) and inhibitor. The test shows that low molecular weight sulfonated modified chitosan (SCS) has obvious inhibition effect on metal corrosion in 2 M HCl solution, and the inhibition rate is close to 95.6 ± 0.1%. In addition, chitosan derivatives can also be used as metal corrosion inhibitors. Suyanto et al. (2015) synthesized two chitosan derivatives, carboxymethyl chitosan benzaldehyde (CMChi-B) and carboxymethyl chitosan urea glutaric acid (CMChi-UGLU), using chitosan as raw materials. By using NaCl 2% solution as the corrosion medium, the maximum corrosion inhibition rates of CMChi-B and CMChi-UGLU on steel are 80.8% and 80.6% respectively. Both inhibitors show strong corrosion inhibition ability.

Chitosan and its derivatives have a wide range of sources and low cost. By adjusting the parameters of the coating solution and introducing polar functional groups, the adhesion strength can be effectively improved, showing good application potential.

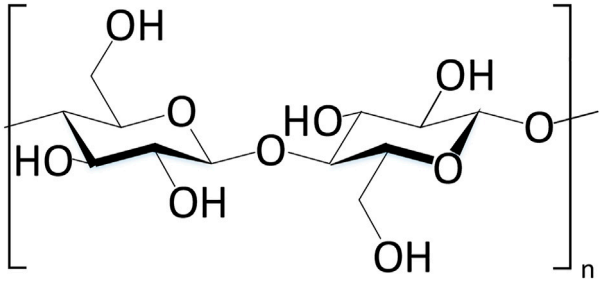


FIGURE 3
Molecular structure of cellulose.

Corrosion inhibitor of cellulose and its derivatives

Cellulose is a natural polymer rich in carbohydrates, which is widely distributed in cotton, wood, hemp and other animal and plant raw materials (Figure 3). Due to the environmental protection of its source, it is currently one of the main sources of compounds as eco-friendly metal corrosion inhibitors (Rajeswari et al., 2013).

Nwanonenyi et al. (2019) studied the adsorption and corrosion inhibition properties of hydroxypropyl cellulose (HPC) for aluminum in 0.5 M HCl and 2 M H_2SO_4 media. The results show that HPC has strong anodic effect in both corrosion solutions and is a mixed inhibitor. The sustained release rate increases with the increase of concentration, but its sustainability decreases with the increase of temperature. Hassan and Ibrahim. (2021) studied the corrosion inhibition effect of methylcellulose (MC) on magnesium (Mg) metal in hydrochloric acid. The results show that the corrosion inhibition rate of Mg metal surface increases with the increase of inhibitor concentration and decreases with the increase of temperature. Its adsorption conforms to the Langmuir and Freundlich isotherm model. This is because MC inhibitor can form OH^- bridging on the metal surface and adsorb on the Mg metal surface, thus delaying the corrosion of the metal surface. Sobhi and Eid (2018) studied the corrosion inhibition effect of methyl hydroxyethyl cellulose (MHEC) on metal copper in 1 m HCl. The results showed that the adsorption of MHEC on the copper surface followed the Langmuir isotherm, mainly physical adsorption. Its corrosion inhibition depends on the concentration of inhibitor. This inhibition is due to the adsorption of MHEC on the surface of copper, which hinders the charge transfer on the surface of copper, so as to protect the copper surface from aggressive ions. Manimaran et al. (2013) evaluated the corrosion inhibition rate of carboxymethyl cellulose (CMC)-Zn^{2+} system on carbon steel in formation water with or without Zn^{2+} by weight loss method. The results show that the corrosion inhibition rate of the formula composed of 250ppm CMC and 50 ppm Zn^{2+} is as high as 98%. It

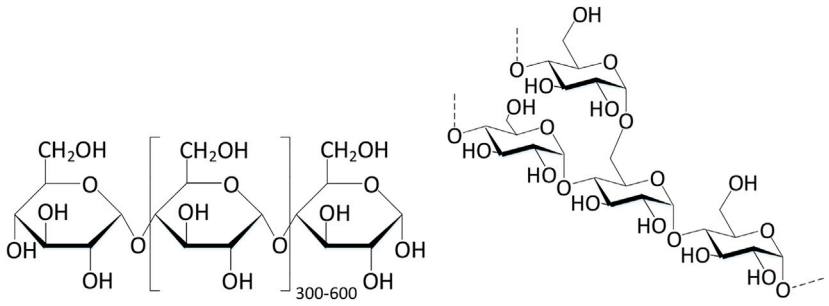


FIGURE 4
Molecular structure of starch (left: amylose; right: amylopectin).

is pointed out that the corrosion inhibition mechanism of CMC is due to the formation of protective film between Fe^{2+}-CMC complex and $Zn(OH)_2$, and CMC-Zn^{2+} system has synergistic effect. In addition, conventional cellulose can be modified or cellulose derivatives can be selected to avoid the limited solubility of conventional cellulose. Farhadian et al. (2021) developed a chemically modified hydroxyethyl cellulose (CHEC) and evaluated the anti-corrosion effect of CHEC in low carbon steel. The results show that CHEC can effectively maintain its anti-corrosion activity in the temperature range of 20–80°C, and the corrosion inhibition rate is as high as 93% at 80°C. In terms of mechanism analysis, CHEC inhibits the metal dissolution of the anode and the hydrogen evolution reaction of the cathode by generating chemical adsorption on the surface of the corroded metal. Its adsorption conforms to the Langmuir adsorption isotherm model.

Because cellulose has various electron rich polar functional groups, but its solubility in polar solvents is limited, it is often necessary to introduce polar substituents into the molecular structure of cellulose for chemical modification. These polar functional groups can improve the corrosion inhibition effect of inhibitors by increasing the molecular size and solubility of cellulose.

Starch corrosion inhibitor

As a natural renewable polymer carbohydrate, starch has the characteristics of low cost, non-toxic, degradable and so on. It is widely used in food industry, pharmaceutical, papermaking and other fields. It mainly exists in corn, potatoes, wheat and other grains, and is composed of various glucose units. However, natural starch is easy to stick together, there are problems of aging and biological corruption, and because of its poor corrosion inhibition ability and low water solubility, it is rarely used as a corrosion inhibitor directly. However, due to the existence of amylose (20–25%) and amylopectin (75–80%), there is a large space for modification (Figure 4).

Nwanonenyi et al. (2017) studied the corrosion inhibition performance of millet starch (MS) in 30 ± 1°C and 0.5 M HCl solution. The results show that MS has obvious inhibition effect on the corrosion of low carbon steel in acid solution, and the inhibition effect is proportional to the concentration of MS. Wu et al. (2016) studied the corrosion inhibition performance of carboxymethyl starch (CMS) on low carbon steel in hydrochloric acid solution. The results show that CMS has good corrosion inhibition effect in 5% HCl solution at 24–60°C. When the concentration of CMS increases to 2 g/L, the corrosion rate of low carbon steel decreases to 11.19 mm/year, and the corrosion inhibition rate is 80.61%. CMS acts as a cathodic inhibitor, and its corrosion inhibition rate increases with the increase of CMS concentration, and the corrosion inhibition effect at room temperature is the best. However, conventional starch has poor corrosion inhibition ability and low water solubility, so it cannot be directly used as a metal corrosion inhibitor, so it often needs to be modified. Ochoa et al. (2013) studied the corrosion inhibition effect of activated carboxymethylated cassava starch derivatives (CMS) on carbon steel XC35 in 200 mg L^{-1} NaCl solution. The results showed that activated starch was a mixed inhibitor, while carboxymethylated cassava starch mainly inhibited the anodic reaction. The slow-release rate of the two corrosion inhibitors increases with the increase of concentration. Activated starch can form a more stable and dense protective film than CMS, because the alkoxy and carboxylic groups in starch can form chelates with the surface of oxidized metal to block its active sites. Anyiam et al. (2020) modified sweet potato starch (NS) extract by alkali treatment to obtain alkali modified starch (AMS), and evaluated its corrosion inhibition performance under acidic conditions. The results show that AMS has high corrosion inhibition effect, and the corrosion inhibition effect tends to be stable with the extension of time. In addition, the synergistic use of potassium iodide (KI) can also improve the corrosion inhibition effect of AMS. Li et al. (2020) synthesized cassava starch sodium allyl sulfonate acrylamide graft copolymer (CS-SAS-AAGC) ternary graft copolymer by chemically modifying cassava starch (CS) and grafting with sodium allyl

FIGURE 5
Molecular structure of protein-conformational state (Burra et al., 2009)

sulfonate (SAS) and acrylamide (AA). The corrosion inhibition performance of CS-SAS-AAGC on cold rolled steel (CRS) in 1.0 M HCl solution was studied. The results show that the best corrosion inhibition rate of CS-SAS-AAGC at 50 mg L^{-1} is 97.2%, and its corrosion inhibition ability is stronger than CS, SAS and AA. Lahrour et al. (2019) used glycerol grafted starch as a metal corrosion inhibitor, grafted glycerol molecules on amylopectin, and then evaluated its corrosion inhibition effect in 1 m HCl solution and C-Mn steel. The results show that the inhibitor may be of cathode type. With the increase of inhibitor concentration, the corrosion current density decreases, and the inhibition effect of the inhibitor gradually increases. When the concentration of inhibitor is increased to 300 mg L^{-1}, the inhibition rate can reach 94.25%.

Starch has a large space for modification. Starch corrosion inhibitor based on certain activation and modification of starch raw materials is also an effective choice for metal corrosion prevention.

Protein inhibitor

As a kind of biological macromolecule (or biopolymer), protein is composed of one or more amino acid (AAS) residue polymerization chains. It plays a role in DNA replication and catalytic metabolism in organisms (Figure 5). Because of its wide source, easy degradation and low cost, it is widely used in industrial and biological fields. At present, studies have shown that proteins have specific amino acid functional sites. Through protonation, molecules can interact with metal surfaces to produce colloidal organometallic complexes. In addition, protein denaturation will also form a corrosion inhibition film on the metal surface, thereby delaying the corrosion process of metal (Talha et al., 2019).

Ma et al. (2021) studied the corrosion behavior of AZ31 magnesium alloy under the action of different concentrations of bovine serum albumin (BSA). The results showed that the corrosion resistance of BSA to AZ31 increased with the increase of its concentration in the range of 0–2.0 g/L. Among them, BSA mainly delays the process of metal corrosion by physical adsorption, but continue to increase the concentration of BSA to 5.0 g/L. BSA will chelate with dissolved Mg^{2+}, Ni^{2+} and other metal ions to form a soluble complex, which will increase the surface roughness of the metal and accelerate its corrosion process. Höhn et al. (2019) also confirmed the view that protein can be used as both a corrosion promoter and a corrosion inhibitor for magnesium alloys, which depends on the selected corrosion inhibitor system, protein concentration, protein size, protein surface interaction (surface charge), medium and protein-protein interaction. Farag et al. (2018) deproteinized the chitin in shrimp shell with acetic acid and sodium carbonate to obtain shrimp shell waste protein (RSWP) with low extraction cost and environmental friendliness, and explored the corrosion inhibition effect of RSWP as a carbon steel inhibitor in 1 M HCl solution. The results show that RSWP, as a mixed corrosion inhibitor, its corrosion inhibition depends on its adsorption on the steel surface, which follows the Langmuir adsorption isotherm and is a spontaneous chemical adsorption mode. The inhibition rate increased with the increase of RSWP concentration, and its inhibition effect was determined by the number of donor atoms in amino acids in protein. Rabizadeh and Asl (2019) studied the effect of bovine cheese protein on the corrosion behavior of low carbon steel in 0.1 M HCl, and found that casein can act on the metal surface by physical or chemical adsorption, and its selective adsorption conforms to the Langmuir adsorption isotherm. In addition, this process is accompanied by the increase of activation energy and activation enthalpy of corrosion process, and the increase of casein concentration will also lead to the decrease of corrosion current density.

In addition to the corrosion inhibition performance of animal protein extracts, researchers have also conducted relevant research on the corrosion inhibition effect of plant proteins. Zunita et al. (2022) separated proteins and their amino acids (Glycine, methionine, proline, threonine, tryptophan, etc.) from tofu pulp, and evaluated their corrosion inhibition in brackish water medium. The results showed that the corrosion inhibition rate of the separated protein was 92% at the concentration of 80 ppm and 25°C, and its adsorption process followed the Langmuir isotherm curve ΔG_{ads} is −33.5 kJ/mol. Zhang et al. (2019) studied the

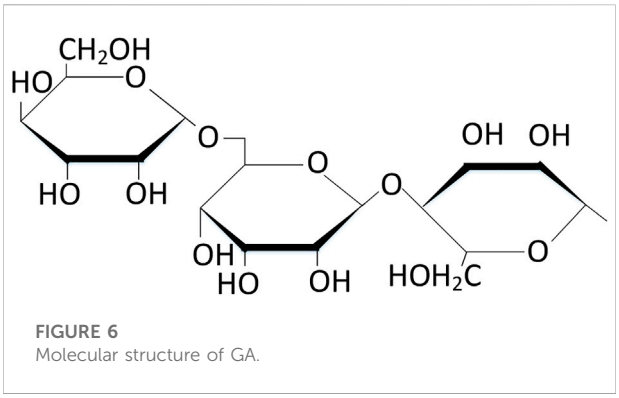


FIGURE 6
Molecular structure of GA.

corrosion inhibition effect of corn gluten powder extract on reinforcement in 3.0 wt% NaCl solution. The results show that the amide bond structure of the main components in the extract is conducive to the adhesion of the steel surface. The corrosion inhibition rate of the extract to the steel is 62.7%–88.1%. Its adsorption process follows the Langmuir isotherm curve and is mainly controlled by physical adsorption.

Animal/plant proteins are supplied from a wider range of sources. By changing protein concentration, protein size, protein surface interaction (surface charge), medium and protein-protein interaction, the corrosion rate of metals can be effectively reduced. However, at present, the mechanism of protein corrosion inhibition needs to be further studied.

Arabic gum inhibitor

Arabic gum (GA) is a common natural gum. Because of its low cost, non-toxic and good biocompatibility, it is widely used in the food and coating industry as a stabilizer, film-forming agent, emulsifier and viscosity control agent. GA is a complex mixture of arabinogalactan, oligosaccharides, polysaccharides and glycoproteins. It has a highly branched complex polysaccharide structure, which makes it have a relatively small hydrodynamic volume, thus allowing more functional groups to be adsorbed on the metal surface (Figure 6), showing a certain corrosion inhibition potential (Ali et al., 2009).

Shen et al. (2019) studied the corrosion inhibition performance of GA on 1018 carbon steel in brine. The results show that GA surface mainly improves the corrosion inhibition ability of metals by chemical adsorption, and the corrosion inhibition rate increases with the increase of GA concentration. Its adsorption follows the Langmuir isotherm, and the corrosion inhibition rate at room temperature is up to 94%. Palumbo et al. (2020) studied the corrosion inhibition effect of GA on N80 carbon steel in high-pressure CO_2 salt water environment. The results show that the corrosion inhibition rate of N80 carbon steel increases with the increase of inhibitor concentration and CO_2 partial pressure. After soaking for 24 h, the maximum corrosion inhibition rate when PCO_2 = 4 MPa is 84.53%. The partial pressure was maintained for 168h, and the corrosion inhibition rate of GA was 74.41%. In addition to investigating the corrosion inhibition effect of GA on different types of corrosive media and materials, halide ions are also one of the important parameters affecting the corrosion inhibition performance of GA. Djellab et al. (2018) explored the corrosion inhibition effect of halide ions (KI, KCl, and KBr) on the surface of API5L X70 pipeline steel in 30°C and 0.5 m H_2SO_4 solution. The results show that the synergistic effect of 2 $g{\cdot}L^{-1}$ GA and 0.5 M KI can obtain significant corrosion inhibition effect (the inhibition rate is 99%), and its adsorption follows the Langmuir adsorption isotherm. In addition, the order of inhibition gain effect between GA and halide is: I > Cl > Br. In addition to halides, nanoparticles are also an important component that affects the corrosion inhibition performance of GA. Asaad et al. (2021) studied the corrosion inhibition effect of nanoparticles Arabic gum (GA-NPs) on reinforced concrete exposed to CO_2 for 180 days. The results show that in the concrete mixed with 3% GA-NPs, the corrosion inhibitor molecules form a protective layer by adsorption on the surface of the reinforcement, the corrosion inhibition rate is increased to 94.5%, and the corrosion rate is reduced to 0.57×10^{-3}mm/year. In addition, due to the consumption of $Ca(OH)_2$, GA-NPs reduced Ca/Si to 3.72% and 0.69% respectively. Nawaz et al. (2021) added about 30.0wt% GA to cerium oxide nanoparticles (CONPs) anti-corrosion coatings, and developed an environment-friendly anti-corrosion additive for coated steel in marine environment. The inhibitor combination can effectively reduce the corrosion active area by releasing Ga from CONPs and adsorbing it on the steel surface, thus playing a corrosion inhibition role.

Similarly, as a biological corrosion inhibitor, the corrosion inhibition performance of Arabic gum is relatively insufficient. At present, the research on Arabic gum inhibitor mainly focuses on auxiliary corrosion inhibition, but the complex polysaccharide structure of Arabic gum provides the basis for its modification and enhancement.

Other plant corrosion inhibitors

At present, the selection of low-cost and environment-friendly biological corrosion inhibitors has become the current mainstream development trend. Many plant extracts contain heteroatoms such as N and P, conjugated double bonds and polar functional groups (Tan et al., 2020). Their electronegativity makes this kind of corrosion inhibitor can form a barrier diaphragm to isolate corrosive media by adsorption on the metal surface.

Guo et al. (2021) found that banana leaves contain a large number of organic compounds containing heteroatoms (N, O, etc.), which can form coordination bonds with the three-dimensional empty orbital of iron atoms, thus forming stable

chemical adsorption on the surface of X70 steel. Hu et al. (2020) found that durian pulp skin/core extract (DPSE) contains rich terminal hydroxyl, amino, carboxylic acid and amide groups. Through the corrosion inhibition performance test of DPSE on copper in 0.5 mol/L H_2SO_4 solution, DPSE can effectively adsorb on the copper surface to slow down corrosion, and the corrosion inhibition rate can reach 92.6%, which is in line with Langmuir isotherm. Chen et al. (2022) explored the corrosion inhibition performance of Ginkgo biloba extract on X70 Steel in 1mol/L HCl solution, and found that the main components of the extract form a dense and stable adsorption film on the surface of × 70 steel, which can effectively prevent the attack of corrosive species on the steel substrate. Tan et al. (2022) explored the corrosion inhibition effect of areca leaf extract (BLE) on Q235 steel and found that BLE could inhibit the cathodic and anodic reactions of Q235 steel. When BLE concentration is 400 mg/l, the anti-corrosion efficiency is close to 94%.

Recent studies have proved the effectiveness of plant extracts as metal corrosion inhibitors, but it is also necessary to ensure that the extraction solvent used in the extraction process should meet the basic requirements of environmental friendliness, low price, good solubility and so on. At the same time, the final extract must be non bioaccumulative, non-toxic and biodegradable.

Synthetic polymer corrosion inhibitor

Synthetic polymers usually refer to polymers synthesized under laboratory conditions. Because of their high molecular weight, they can still maintain good coating effect even at low concentration, thus playing a certain anti-corrosion effect. Therefore, the application status of resin-based corrosion inhibitor, polyethylene glycol corrosion inhibitor and conductive polymer corrosion inhibitor which are widely used at present are introduced respectively.

Resin based corrosion inhibitor

Epoxy resin, also known as epoxy polymer, has the advantages of many varieties, large curing/modification space, strong adhesion of peripheral polar substituents, low curing shrinkage, high mechanical strength, strong acid and alkali resistance, etc. Due to the polymer characteristics of epoxy resin, it can form excellent surface coverage and strong adsorption capacity on the metal surface, and heteroatoms such as O, N and s in epoxy resin are easy to protonate, which is conducive to the dissolution of epoxy polymer in polar electrolyte (Jin et al., 2015). In addition, epoxy ring can also be protonated in strong acidic solution and undergo ring opening reaction to form -OH substituents with strong polarity and interaction with water (Figure 7). However, due to the high

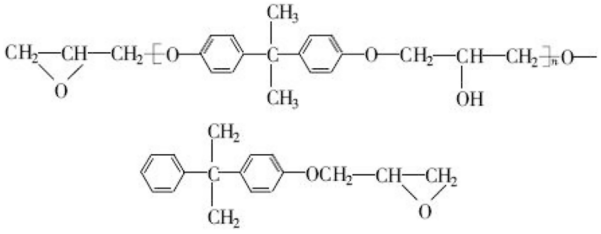


FIGURE 7
Molecular structure of resin.

molecular weight of most epoxy resins, the solubility in polar electrolytes is still limited, so at present, waterborne epoxy resins are mainly used as metal corrosion inhibitors.

Hsissou et al. (2020) studied the corrosion inhibition effect of diglycidyl aminobenzene (DGAB) epoxy prepolymer on acid corrosion of carbon steel. The results show that DGAB has good metal adsorption and corrosion barrier ability. Its adsorption process conforms to the Langmuir adsorption isotherm, and the corrosion inhibition rate is as high as 95.9% at the concentration of 10^{-3} M. Dagdag et al. (2020) used methylene diphenylamine (MDA) as curing agent to build a macromolecular epoxy coating (DGEDDS-MDA) based on bisphenol s diglycidyl ether (DGEDDS). It was found that the electron rich center of DGEDDS-MDA showed strong adsorption capacity with Fe (110) metal surface, and its adsorption energy was-300.67 kJ/mol. In addition to exploring the influence of the type of epoxy resin on metal corrosion, researchers also modified the resin-based materials to obtain better corrosion inhibition effect.

In addition to using different types of epoxy resins for metal corrosion inhibition experiments, researchers also improved the corrosion inhibition effect of conventional resin materials by doping silicon and carbon nanoparticles or using chemical reagents. Joseph et al. (2020) synthesized ceria zirconia solid solution ($CexZr_{1-x}O_2$ nanoparticles) with different doping concentrations by co deposition method, and prepared $CexZr_{1-x}O_2$ nanoparticles waterborne epoxy resin coating on low carbon steel substrate by spin coating technology. The results showed that $CexZr_{1-x}O_2$ nanoparticles effectively improved the corrosion resistance of waterborne resin. In CE rich Ce–Zr solid solution ($0.6 \leq \times \leq 0.8$), with the increase of Zr doping concentration, the corrosion resistance of $CexZr_{1-x}O_2$ nanoparticles waterborne epoxy resin is enhanced, because a large amount of oxygen vacancies improves the redox performance of the resin-based system. Sharma et al. (2022) used carbon nanotubes/reduced graphene oxide (CNT/rGO) to modify the epoxy resin corrosion inhibition coating, and found that CNT/rGO can effectively improve the mechanical strength of the coating matrix, and can effectively improve the dispersion and bonding characteristics of nano fillers by cooperating with silane agents. The corrosion results show that CNT/rGO modified epoxy coating

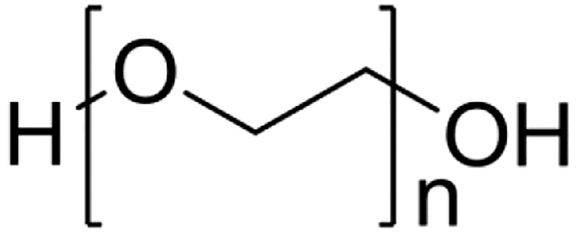


FIGURE 8
Molecular structure of PEG.

can form a dense, impermeable and durable coating on the surface of reinforcement. This is because the functional groups on the rGO sheet provide a dense network structure, and the rGO sheet also has a certain hydrophobicity, which can effectively reduce the infiltration ability of corrosive media. Karaxi et al. (2019) studied the modification effect of liquid rubber polymer on epoxy resin. The experimental results show that the corrosion inhibition coating can form a stable chelate on the surface of galvanized steel, has strong barrier performance, and the coating shows a certain self-reaction and self-healing effect on corrosion damage. Rahman et al. (2019) found that the use of indole butyric acid (IBA) modified gypsum can effectively improve the corrosion resistance of epoxy coating to reinforcement, and the coating has no adverse effect on the bond strength between embedded reinforcement and concrete. The reason is related to the healing effect of metal surface passivation in corrosive environment; On the other hand, it is attributed to the adhesion and good isolation ability of the resin. Dagdag et al. (2019) explored the corrosion inhibition effect of zinc phosphate on the surface of aa7075-t6 aluminum plate substrate with bisphenol epoxy resin coating in NaCl solution (3wt%). The results showed that the addition of zinc phosphate promoted the formation of the coating barrier layer and effectively delayed the penetration rate of corrosive ions on the surface of the coating in the epoxy resin polyamide anticorrosive system.

Resin based corrosion inhibitor is the mainstream metal corrosion inhibitor at present because of its good acid, alkali and salt resistance. And the epoxy resin has a large modification space, which can further improve its corrosion inhibition performance by doping various functional nano materials.

Polyethylene glycol corrosion inhibitor

Polyethylene glycol (PEG) as a high molecular polymer, its molecular weight ranges from 200 to 10 of thousands (Figure 8). Due to its good dispersion, solubility, adhesion and low raw material cost, peg is widely used in chemical, pharmaceutical, food and other fields (Chen et al., 2005). In recent years, PEG and its derivatives have been used as metal corrosion inhibitors because of its good thermal stability, chemical stability and recyclability.

Deyab and Abd El-Rehim (2013) used peg with different average molecular weights (mw = 1200, 4,000 and 6,000) as corrosion inhibitor to explore the corrosion inhibition effect of PEG on carbon steel in 1.0 M butyric acid medium. The results show that the corrosion rate of carbon steel in butyric acid decreases with the increase of PEG concentration and molecular weight, and intensifies with the increase of temperature. Peg retards the corrosion of carbon steel by physical adsorption, which follows the Temkin adsorption isotherm. Awad (2013) studied the correlation between the molecular structure of PEG and the corresponding corrosion inhibition rate. Quantum chemical calculation shows that there is a good correlation between the quantum chemical parameters of the studied compound and the corrosion inhibition rate of its corrosion process. The function $\triangle E$ of its reaction activity is related to the corrosion inhibition rate of the inhibitor. With the decrease of $\triangle E$, the reactivity of the inhibitor to the metal surface increases. Similarly, the higher the adsorption energy, the higher the stability of the formed complex, and the higher the corrosion inhibition rate.

Like other corrosion inhibitors, the corrosion inhibition effect of PEG is also affected by some common surfactants, catalysts, emulsifiers and other additives. Mobin and Khan (2014) studied the adsorption and corrosion inhibition behavior of PEG with sodium dodecylbenzene sulfonate (SDBS) and cetyltrimethylammonium bromide (CTAB) on low carbon steel in 0.1 M H_2SO_4 solution and 30–60°C. The results show that the addition of a small amount of surfactant can synergistically enhance the corrosion inhibition of PEG. The adsorption process is completely spontaneous and conforms to the Langmuir adsorption isotherm. Hamlaoui et al. (2013) studied the effect of PEG and cerium nitrate solution on galvanized steel. The results show that adding peg to cerium nitrate solution can effectively weaken the hydrogen reduction reaction, reduce the thickness of the coating, reduce the formation of cracks, and then improve the corrosion resistance of the coating. Umoren et al. (2009) studied the effect of halides (KCl, KBr and KI) on the corrosion inhibition of PEG. The results show that PEG can slow down the corrosion of aluminum in alkaline medium, and its corrosion inhibition rate increases with the increase of inhibitor concentration. The adsorption of PEG on aluminum surface conforms to Flory Huggins and Temkin adsorption isotherms. The addition of halides has obvious synergistic effect on the corrosion inhibition rate of PEG, and the order is: $I^- > Br^- > Cl^-$. El-Lateef and Hany (2016) studied the corrosion inhibition of carbon steel in 0.1 M H_2SO_4 solution when two kinds of PEG with different molecular weight cooperated with rare earth Ce^{4+}. The results show that the synergistic use of PEG and Ce^{4+} ions can significantly improve the corrosion inhibition of carbon steel surface when used alone.

Polyethylene glycol has poor performance when used as a metal corrosion inhibitor. It is often used together with other chemical reagents in the form of additives, and can improve the corrosion resistance of the coating by weakening the hydrogen reduction reaction in the process of metal corrosion.

Conductive polymer inhibitor

Conductive polymer (CPS) is a special type of polymer. Because its main chain has extensive π electron delocalization, it shows good conductivity during oxidation and reduction. Because of the electronic properties of semiconductors and the ease of processing of conventional polymers, they are widely used in daily production and life. In addition, because of its large structural units and good coating effect, conductive polymers are also efficient corrosion inhibitors for metal corrosion prevention. The commonly used conductive polymer corrosion inhibitors include polyaniline (PANI), polypyrrole (PPy), polyindole (PIN), polythiophene (PTh) and polyaniline (PPD). Because the conductive polymer coating has good barrier performance, it can effectively inhibit the corrosion diffusion of corrosion medium and firmly adhere to the metal surface (Advincula, 2015; Namsheer and Rout, 2021).

Danaee et al. (2020) studied 2,2′-(2-hydroxyethylimino) bis [N-(α-Corrosion inhibition effect of dimethyl phenylethyl) -n-methylacetamide] on low carbon steel in 1 m HCl solution. The results show that the corrosion inhibitor is mainly adsorption, and its adsorption process conforms to the Langmuir adsorption isotherm. It is a mixed corrosion inhibitor, and its slow-release rate is directly proportional to the corrosion concentration. Ali Fathima Sabirneeza and Subhashini (2013) prepared a new water-soluble conductive polymer poly (vinyl alcohol histidine) composite with persulfate as raw material by free radical condensation method, and explored the corrosion inhibition effect of the material on the corrosion behavior of low carbon steel in 1 M HCl. The results show that when the mass fraction of the composite is 0.6%, the corrosion inhibition rate can reach more than 95%. As a mixed inhibitor, the adsorption of the inhibitor conforms to the Temkin adsorption isotherm. Duan et al. (2022) obtained a new polymer poly [n-alkyl—(4-hydroxybutyl-2-alkynyl) pyridine bromide] by radical polymerization of n-alkyl—(4-hydroxybutyl-2-alkynyl) pyridine bromide under the catalysis of Pd-C catalyst, and then verified its corrosion inhibition effect on × 70 Steel in HCl solution. The results show that the optimum corrosion inhibition concentration of the polymer is 60 ppm, and the increase of acid concentration and soaking time can enhance the adsorption and barrier effect of the polymer.

In addition, Awuzie (2017) found that conductive polymers can improve their conductivity, mechanical flexibility and high thermal stability by doping. Aly et al. (2019a) studied the effect of sodium bentonite on the solubility, thermal stability and corrosion inhibition of CoPm conductive polymer coating. The results show that the corrosion inhibition rate of the blend polymer to low carbon steel in acidic environment is 97%, which is much higher than that of the same copolymer. Shukla et al. (2008) synthesized water-soluble self-doped poly (o-aminobenzoic acid) corrosion inhibitor with different oxidants at different polymerization times, and studied the corrosion inhibition effect of the inhibitor on low carbon steel in 0.5 M HCl. The results show that the inhibitor is mainly in the form of adsorption inhibition. When the concentration of inhibitor is 60 mg/l, the maximum inhibition rate is 92%.

In the process of metal corrosion, some CPS charges will be consumed due to corrosion, and anion exchange will occur to form a passive oxide layer, reducing the adhesion between CPS and metal. Therefore, researchers proposed to use nano inorganic materials in conjunction with CPS to improve the corrosion inhibition ability of metal materials. Aly et al. (2019b) prepared cycloalkanone nano clay CPS composites by *in-situ* polymerization, and compared them with CPS polymers without nano clay. The results show that CPS modified by cyclohexanone has better shielding and protection ability, and its corrosion inhibition rate is as high as 95.9%, indicating that increasing the molecular weight of the polymer (the larger the molecular weight is, the larger the metal area can be covered) or using modified nano clay can effectively enhance the corrosion inhibition ability of CPS. Elhalawany et al. (2014) prepared a kind of conductive polymer nanoparticles (CPNs) by using micro lotion polymerization technology, and added the prepared CPNs and polymethylaniline (PAns) - polyaniline (PTol) copolymer (CCPNs) dispersed in aqueous medium to the aqueous coating for corrosion performance evaluation. The results show that the composite inhibitor system has stronger barrier effect and is easier to participate in the formation of oxide film. Through the dual redox catalysis of CCPNs, the corrosion resistance of metal surface is significantly improved.

There are many kinds of conductive polymer corrosion inhibitors, which have strong corrosion resistance to corrosive media, good adhesion, and can be used together with similar conductive polymers, but too many doping methods have brought challenges to its mechanism research.

Composite polymer inhibitor

In order to obtain better anticorrosion performance of metal materials, at present, the combination of inhibitors is often used to balance the advantages and disadvantages of each inhibitor and maximize the anticorrosion effect of the coating. Abd El-Ghaffar et al. (2020) used ammonium persulfate as initiator, used toluidine and phenylenediamine monomers to carry out *in-situ* chemical oxidation polymerization in the presence of barium sulfate to prepare a new intelligent and eco-friendly polymer conductive polymer inhibitor. Based on the synergistic epoxy polyester coating formula, they carried out a salt spray test of

soaking in 3.5% NaCl solution for 1,500 h, and obtained good corrosion resistance. Alva et al. (2018) explored the corrosion inhibition effect of GA cooperating with cellulose acetate on the surface of aluminum plate. It was found that in high concentration of strong acid, the water solubility of GA and cellulose acetate was enhanced, which was due to the increase of solution polarity by the polysaccharide components in the two biological corrosion inhibitors. In addition, the two polymers can also hinder the chemical reaction between aluminum and other substituents by forming an organic coating, and the N, O, S, etc. in the two polymers play an important role in reducing the formation of aluminum surface by-products. Chauhan et al. (2018) explored the effect of PEG crosslinked chitosan (Cht-PEG) on the corrosion performance of carbon steel in 1 M sulfamic acid solution. The results show that the adsorption of Cht-PEG on the surface of low carbon steel conforms to the Langmuir isotherm, and the corrosion inhibition rate can reach 93.9% at the optimal concentration of 200ppm. As a mixed corrosion inhibitor with cathode advantages, it can effectively inhibit the anodic and cathodic corrosion processes. Kong et al. (2019) explored the effect of polyaniline (PANI)/chitosan (CTS) on the corrosion properties of Q235 steel in 0.5 M HCl medium. The results show that polyaniline/CTS is essentially a mixed corrosion inhibitor because it can provide N, O, −C = N and other groups. In the corrosion test, the increase of its concentration can effectively delay the cathodic reaction, resulting in the increase of charge transfer resistance, thus delaying the corrosion process of Q235 steel. Sabirneeza et al. (2013) synthesized a new corrosion inhibitor polyvinyl alcohol leucine composite inhibitor (PVAL), and studied the corrosion inhibition effect of PVAL on low carbon steel in 1 M HCl solution. The results show that PVAL is a mixed inhibitor, which exhibits spontaneous adsorption on the surface of low carbon steel. It inhibits the corrosion of low carbon steel by blocking the active sites of metals. When its concentration increases to 0.6%, the inhibition rate can reach more than 95%. Yang et al. (2018) synthesized a new clam shaped adhesive polymer (PHEA-DOPA) containing 3,4-dihydroxyphenylalanine (DOPA) functional group based on polyasparagine derivatives, and explored the corrosion inhibition performance of the polymer on waterborne epoxy coatings. The results show that PHEA-DOPA can effectively delay the corrosion process of the metal solution interface by forming an insoluble DOPA-Fe complex passivation layer on the surface of carbon steel to improve the cross-linking density of the coating.

Conclusion

Through the systematic development of aqueous polymer corrosion inhibitors, it can be seen that in water-soluble polymers, such as chitosan, cellulose, animal and plant proteins and their derivatives have the characteristics of wide source, easy degradation and low cost, but these materials are often used in large amounts and have poor solubility, and will be affected by the pH value of the solution, and the understanding of the mechanism is still not clear; In synthetic polymers, epoxy resin, polyethylene glycol, conductive polymers, etc. have the advantages of low dosage, containing multiple adsorption sites, high corrosion inhibition rate, long action time limit, high temperature resistance, large modification space, etc., but the disadvantages are that it is difficult to analyze the corrosion inhibition mechanism of synthetic polymers, containing certain biological toxicity, and difficult to control the molecular weight; The composite polymer inhibitor effectively takes into account the balance between natural polymers and synthetic polymers, which can not only play a synergistic corrosion inhibition effect, but also improve other properties required by the metal matrix (such as wear, lubrication, hydrophilic or hydrophobic properties, etc.). However, the composition of the composite polymer is miscellaneous, and the analysis of its corrosion inhibition mechanism is more difficult, but this uncertainty provides infinite possibilities for future exploration. It is believed that with the continuous advancement of relevant research work and the development of technology, the corrosion inhibition mechanism of various polymers will become more and more clear. By reasonably controlling the cost of polymer corrosion inhibitor, it will become the mainstream development trend of metal corrosion protection in the future.

Author contributions

ZY investigated and processed the data, pointed out the framework of the paper, made a summary and analysis, and completed the writing of the paper. ZY reviewed the paper and supported the project.

Author Contributions

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

Funding

This work was supported by The Doctoral Fund Project of Chongqing Institute of Industry and Technology (Grant No. 2022G2YBS2K2-12).

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

Abd El-Ghaffar, M., Abdelwahab, N., Fekry, A., Sanad, M., Sabaa, M., and Soliman, S. (2020). Polyester-epoxy resin/conducting polymer/barium sulfate hybrid composite as a smart eco-friendly anti-corrosive powder coating. *Prog. Org. Coatings* 144 (7), 105664–105711. doi:10.1016/j.porgcoat.2020.105664

Advincula, R. (2015). "Conducting polymers with superhydrophobic effects as anticorrosion coating," in *Intelligent coatings for corrosion control* (Oxford: Butterworth-Heinemann), 409–430. doi:10.1016/B978-0-12-411467-8.00011-8

Ali, B., Ziada, A., and Blunden, G. (2009). Biological effects of gum Arabic: A review of some recent research. *Food Chem. Toxicol.* 47 (1), 1–8. doi:10.1016/j.fct.2008.07.001

Ali Fathima Sabirneeza, A., and Subhashini, S. (2013). A novel water-soluble, conducting polymer composite for mild steel acid corrosion inhibition. *J. Appl. Polym. Sci.* 127 (4), 3084–3092. doi:10.1002/app.37661

Alva, S., Sundari, R., Rahmatullah, A., and Wahyudi, H. (2018). Inhibition effect of Arabic gum and cellulose acetate coatings on aluminium in acid/base media//IOP Conference Series: Materials Science and Engineering. *IOP Conf. Ser. Mat. Sci. Eng.* 324 (1), 012070–8. doi:10.1088/1757-899X/324/1/012070

Aly, K., Younis, O., Mahross, M., Orabi, E. A., Abdel-Hakim, M., Tsutsumi, O., et al. (2019a). Conducting copolymers nanocomposite coatings with aggregation-controlled luminescence and efficient corrosion inhibition properties. *Prog. Org. Coatings* 135 (10), 525–535. doi:10.1016/j.porgcoat.2019.06.001

Aly, K., Younis, O., Mahross, M., Tsutsumi, O., Mohamed, M. G., and Sayed, M. M. (2019b). Novel conducting polymeric nanocomposites embedded with nanoclay: synthesis, photoluminescence, and corrosion protection performance. *Polym. J.* 51 (1), 77–90. doi:10.1038/s41428-018-0119-6

Anyiam, C., Ogbobe, O., Oguzie, E., and Madufor, I. C. (2020). Synergistic study of modified sweet potato starch and KI for corrosion protection of mild steel in acidic media. *J. Bio. Tribocorros.* 6 (3), 70–10. doi:10.1007/s40735-020-00362-x

Asaad, M., Huseien, G., Baghban, M., Raja, P. B., Fediuk, R., Faridmehr, I., et al. (2021). Gum Arabic nanoparticles as green corrosion inhibitor for reinforced concrete exposed to carbon dioxide environment. *Materials* 14 (24), 7867–7925. doi:10.3390/ma14247867

Awad, M. K. (2013). Quantum chemical studies and molecular modeling of the effect of polyethylene glycol as corrosion inhibitors of an aluminum surface. *Can. J. Chem.* 91 (4), 283–291. doi:10.1139/cjc-2012-0354

Awuzie, C. (2017). Conducting polymers. *Mater. Today Proc.* 4 (4), 5721–5726. doi:10.1016/j.matpr.2017.06.036

Balgude, D., and Sabnis, A. (2012). Sol–gel derived hybrid coatings as an environment friendly surface treatment for corrosion protection of metals and their alloys. *J. Solgel. Sci. Technol.* 64 (1), 124–134. doi:10.1007/s10971-012-2838-z

Bardal, E., Drugli, J., and Gartland, P. (1993). The behaviour of corrosion resistant steels in seawater. A review. *Corros. Sci.* 35 (1-4), 257–267. doi:10.1016/0010-938X(93)90157-C

Blanda, G., Brucato, V., Carfì, F., Conoscenti, G., La Carrubba, V., Piazza, S., et al. (2019). Chitosan-coating deposition via galvanic coupling. *ACS Biomater. Sci. Eng.* 5 (4), 1715–1724. doi:10.1021/acsbiomaterials.8b01548

Burra, P., Zhang, Y., Godzik, A., and Stec, B. (2009). Global distribution of conformational states derived from redundant models in the PDB points to non-uniqueness of the protein structure. *Proc. Natl. Acad. Sci. U. S. A.* 106 (26), 10505–10510. doi:10.1073/pnas.0812152106

Chauhan, D., Quraishi, M., Sorour, A., Saha, S. K., and Banerjee, P. (2019). Triazole-modified chitosan: A biomacromolecule as a new environmentally benign corrosion inhibitor for carbon steel in a hydrochloric acid solution. *RSC Adv.* 9 (26), 14990–15003. doi:10.1039/C9RA00986H

Chauhan, D., Srivastava, V., Joshi, P., and Quraishi, M. A. (2018). PEG cross-linked chitosan: A biomacromolecule as corrosion inhibitor for sugar industry. *Int. J. Ind. Chem.* 9 (4), 363–377. doi:10.1007/s40090-018-0165-0

Chen, J., Spear, S., Huddleston, J., and Rogers, R. D. (2005). Polyethylene glycol and solutions of polyethylene glycol as green reaction media. *Green Chem.* 7 (2), 64–82. doi:10.1039/B413546F

Chen, L., Shi, Y., and Li, H. (2022). Research progress of plant extract-based green corrosion inhibitors in aggressive media. *J. Chongqing Univ.* 25 (4), 1–25. doi:10.11835/j.issn.1000-582X.2021.114

Dagdag, O., Berisha, A., Safi, Z., Dagdag, S., Berrani, M., Jodeh, S., et al. (2020). Highly durable macromolecular epoxy resin as anticorrosive coating material for carbon steel in 3% NaCl: Computational supported experimental studies. *J. Appl. Polym. Sci.* 137 (34), 49003–49012. doi:10.1002/app.49003

Dagdag, O., El Gana, L., Hamed, O., Jodeh, S., and El Harfi, A. (2019). Anticorrosive formulation based of the epoxy resin–polyaminoamide containing zinc phosphate inhibitive pigment applied on sulfo-tartaric anodized AA 7075-T6 in NaCl medium. *J. Bio. Tribocorros.* 5 (1), 25–29. doi:10.1007/s40735-019-0218-8

Danaee, I., RameshKumar, S., RashvandAvei, M., and Vijayan, M. (2020). Electrochemical and quantum chemical studies on corrosion inhibition performance of 2, 2'-(2-hydroxyethylimino) bis [N-(alphaalpha-dimethylphenethyl)-N-methylacetamide] on mild steel corrosion in 1M HCl solution. *Mat. Res.* 23 (2), 1–16. doi:10.1590/1980-5373-MR-2018-0610

Dariva, C., and Galio, A. (2014). Corrosion inhibitors–principles, mechanisms and applications. *Dev. Corros. Prot.* 16 (2), 365–378. doi:10.5772/57255

Deyab, M., and Abd El-Rehim, S. (2013). Influence of polyethylene glycols on the corrosion inhibition of carbon steel in butyric acid solution: weight loss, EIS and theoretical studies. *Int. J. Electrochem. Sci.* 8 (12), 12613–12627. doi:10.1007/s10008-013-2231-z

Djellab, M., Bentrah, H., Chala, A., and Taoui, H. (2018). Synergistic effect of halide ions and gum Arabic for the corrosion inhibition of API5L X70 pipeline steel in H_2SO_4. *Mater. Corros.* 70 (1), 149–160. doi:10.1002/maco.201810203

Duan, M., Wu, G., Gu, T., Jiang, X., Zhao, L., Zhou, L., et al. (2022). Conductive and anticorrosive poly (N-alkyl-2-(4-hydroxybut-2-ynyl) pyridinium bromides): Synthesis and characterization. *J. Polym. Sci.* 60 (12), 1468–1479. doi:10.1002/pol.20210672

El-Lateef, A., and Hany, M. (2016). Synergistic effect of polyethylene glycols and rare Earth Ce^{4+} on the corrosion inhibition of carbon steel in sulfuric acid solution: electrochemical, computational, and surface morphology studies. *Res. Chem. Intermed.* 42 (4), 3219–3240. doi:10.1007/s11164-015-2207-y

Elhalawany, N., Mossad, M., and Zahran, M. (2014). Novel water-based coatings containing some conducting polymers nanoparticles (CPNs) as corrosion inhibitors. *Prog. Org. Coatings* 77 (3), 725–732. doi:10.1016/j.porgcoat.2013.12.017

Farag, A., Ismail, A., and Migahed, M. (2018). Environmental-friendly shrimp waste protein corrosion inhibitor for carbon steel in 1 M HCl solution. *Egypt. J. petroleum* 27 (4), 1187–1194. doi:10.1016/j.ejpe.2018.05.001

Farhadian, A., Kashani, S., Rahimi, A., Oguzie, E. E., Javidparvar, A. A., Nwanonenyi, S. C., et al. (2021). Modified hydroxyethyl cellulose as a highly efficient eco-friendly inhibitor for suppression of mild steel corrosion in a 15% HCl solution at elevated temperatures. *J. Mol. Liq.* 338 (9), 116607–116616. doi:10.1016/j.molliq.2021.116607

Farhadian, A., Varfolomeev, M., Shaabani, A., Nasiri, S., Vakhitov, I., Zaripova, Y. F., et al. (2020). Sulfonated chitosan as green and high cloud point kinetic methane hydrate and corrosion inhibitor: Experimental and theoretical studies. *Carbohydr. Polym.* 236 (5), 116035–116112. doi:10.1016/j.carbpol.2020.116035

Fathima Sabirneeza, A., Geethanjali, R., and Subhashini, S. (2015). Polymeric corrosion inhibitors for iron and its alloys: A review. *Chem. Eng. Commun.* 202 (2), 232–244. doi:10.1080/00986445.2014.934448

Grundmeier, G., Schmidt, W., and Stratmann, M. (2000). Corrosion protection by organic coatings: electrochemical mechanism and novel methods of investigation. *Electrochimica Acta* 45 (15-16), 2515–2533. doi:10.1016/s0013-4686(00)00348-0

Guo, L., Tan, B., Li, W., Li, Q., Zheng, X., and Obot, I. B. (2021). Banana leaves water extracts as inhibitor for X70 steel corrosion in HCl medium. *J. Mol. Liq.* 327 (11), 114828–114911. doi:10.1016/j.molliq.2020.114828

Guo, L., Tan, J., Kaya, S., Leng, S., Li, Q., and Zhang, F. (2020). Multidimensional insights into the corrosion inhibition of 3, 3-dithiodipropionic acid on Q235 steel in H_2SO_4 medium: A combined experimental and *in silico* investigation. *J. colloid interface Sci.* 570 (6), 116–124. doi:10.1016/j.jcis.2020.03.001

Hamlaoui, Y., Tifouti, L., and Pedraza, F. (2013). Corrosion protection of electro-galvanized steel by ceria-based coatings: effect of polyethylene glycol (PEG) addition. *J. Mat. Eng. Perform.* 22 (9), 2706–2715. doi:10.1007/s11665-013-0574-3

Hassan, R., and Ibrahim, S. (2021). Performance and efficiency of methyl-cellulose polysaccharide as a green promising inhibitor for inhibition of corrosion of magnesium in acidic solutions. *J. Mol. Struct.* 1246 (101), 131180–131187. doi:10.1016/j.molstruc.2021.131180

Höhn, S., Virtanen, S., and Boccaccini, A. (2019). Protein adsorption on magnesium and its alloys: A review. *Appl. Surf. Sci.* 464 (1), 212–219. doi:10.1016/j.apsusc.2018.08.173

Hsissou, R., Benhiba, F., Dagdag, O., El Bouchti, M., Nouneh, K., Assouag, M., et al. (2020). Development and potential performance of prepolymer in corrosion inhibition for carbon steel in 1.0 M HCl: outlooks from experimental and computational investigations. *J. colloid interface Sci.* 574 (8), 43–60. doi:10.1016/j.jcis.2020.04.022

Hu, Y., Fan, B., and Liu, H. (2020). Corrosion inhibition mechanism of durian peel and stone extract for copper in Sulphuric solution. *Corros. Prot.* 41 (7), 17–24. doi:10.11973/fsyfh-202007003

Hui, X., Guo, N., and Liu, T. (2020). Slow release mechanism and application progress of cellulose based corrosion inhibitors. *Equip. Environ. Eng.* 17 (12), 47–53. doi:10.7643/issn.1672-9242.2020.12.008

Jevremović, I., Chen, Y., and Altin, A. (2020). "Mechanisms of inhibitor action," in *Corrosion inhibitors in the oil and gas industry* (New Jersey: Wiley).

Jin, F., Li, X., and Park, S. (2015). Synthesis and application of epoxy resins: A review. *J. Industrial Eng. Chem.* 29 (9), 1–11. doi:10.1016/j.jiec.2015.03.026

Joseph, A., John Mathew, K., and Vandana, S. (2020). Zirconium-doped ceria nanoparticles as anticorrosion pigments in waterborne epoxy–polymer coatings. *ACS Appl. Nano Mat.* 4 (1), 834–849. doi:10.1021/acsanm.0c03162

Karaxi, E., Kartsonakis, I., and Charitidis, C. (2019). Assessment of self-healing epoxy-based coatings containing microcapsules applied on hot dipped galvanized steel. *Front. Mat.* 222 (9), 1–17. doi:10.3389/fmats.2019.00222

Kong, P., Feng, H., Chen, N., Lu, Y., Li, S., and Wang, P. (2019). Polyaniline/chitosan as a corrosion inhibitor for mild steel in acidic medium. *RSC Adv.* 9 (16), 9211–9217. doi:10.1039/C9RA00029A

Lahrour, S., Benmoussat, A., Bouras, B., Mansri, A., Tannouga, L., and Marzorati, S. (2019). Glycerin-grafted starch as corrosion inhibitor of C-Mn steel in 1 M HCl solution. *Appl. Sci.* 9 (21), 4684–4718. doi:10.3390/app9214684

Li, X., Deng, S., Lin, T., Xie, X., and Du, G. (2020). Cassava starch ternary graft copolymer as a corrosion inhibitor for steel in HCl solution. *J. Mater. Res. Technol.* 9 (2), 2196–2207. doi:10.1016/j.jmrt.2019.12.050

Ma, Y., Talha, M., Wang, Q., Li, Z., and Lin, Y. (2021). Evaluation of the corrosion behavior of AZ31 magnesium alloy with different protein concentrations. *Anti-Corrosion Methods Mater.* 69 (1), 47–54. doi:10.1108/ACMM-08-2021-2524

Manimaran, N., Rajendran, S., Manivannan, M., et al. (2013). Corrosion inhibition by carboxymethyl cellulose. *Eur. Chem. Bull.* 2 (7), 494–498. doi:10.1016/S0956-5663(01)00286-X

Mobin, M., and Khan, M. (2014). Adsorption and corrosion inhibition behavior of polyethylene glycol and surfactants additives on mild steel in H_2SO_4. *J. Mat. Eng. Perform.* 23 (1), 222–229. doi:10.1007/s11665-013-0767-9

Musa, A., Mohamad, A., Al-Amiery, A., and Tien, L. T. (2012). Galvanic corrosion of aluminum alloy (Al_{2024}) and copper in 1.0M hydrochloric acid solution. *Korean J. Chem. Eng.* 29 (6), 818–822. doi:10.1007/s11814-011-0233-z

Namsheer, K., and Rout, C. (2021). Conducting polymers: A comprehensive review on recent advances in synthesis, properties and applications. *RSC Adv.* 11 (10), 5659–5697. doi:10.1039/d0ra07800j

Nawaz, M., Shakoor, R., Kahraman, R., and Montemor, M. (2021). Cerium oxide loaded with gum Arabic as environmentally friendly anti-corrosion additive for protection of coated steel. *Mater. Des.* 198 (1), 109361–109412. doi:10.1016/j.matdes.2020.109361

Nwanonenyi, S., Arukalam, I., Obasi, H., Ezeamaku, U. L., Eze, I. O., Chukwujike, I. C., et al. (2017). Corrosion inhibitive behavior and adsorption of millet (Panicum miliaceum) starch on mild steel in hydrochloric acid environment. *J. Bio. Tribocorros.* 3 (4), 54–11. doi:10.1007/s40735-017-0115-y

Nwanonenyi, S., Obasi, H., and Eze, I. O. (2019). Hydroxypropyl cellulose as an efficient corrosion inhibitor for aluminium in acidic environments: Experimental and theoretical approach. *Chem. Afr.* 2 (3), 471–482. doi:10.1007/s42250-019-00062-1

Ochoa, N., Bello, M., Sancristóbal, J., Balsamo, V., Albornoz, A., and Brito, J. L. (2013). Modified cassava starches as potential corrosion inhibitors for sustainable development. *Mat. Res.* 16, 1209–1219. doi:10.1590/S1516-14392013005000126

Palumbo, G., Kollbek, K., Wirecka, R., Bernasik, A., and Gorny, M. (2020). Effect of CO_2 partial pressure on the corrosion inhibition of N80 carbon steel by gum Arabic in a CO_2-water saline environment for shale oil and gas industry. *Materials* 13 (19), 4245–4324. doi:10.3390/ma13194245

Pludek, V. (1977). *Design and corrosion control*. London: The Macmillan Press. ISBN-13: 978-1349027897.

Rabizadeh, T., and Asl, S. (2019). Casein as a natural protein to inhibit the corrosion of mild steel in HCl solution. *J. Mol. Liq.* 276 (2), 694–704. doi:10.1016/j.molliq.2018.11.162

Rahman, M., Islam, M., Khan, M., Ong, H. R., and Uddin, M. T. (2019). IBA-modified gypsum-containing epoxy resin coating for rebar: corrosion performance and bonding characteristics. *Int. J. Plast. Technol.* 23 (1), 20–28. doi:10.1007/s12588-019-09238-3

Rajeswari, V., Kesavan, D., Gopiraman, M., and Viswanathamurthi, P. (2013). Physicochemical studies of glucose, gellan gum, and hydroxypropyl cellulose—inhibition of cast iron corrosion. *Carbohydr. Polym.* 95 (1), 288–294. doi:10.1016/j.carbpol.2013.02.069

Sabirneeza, A., Subhashini, S., and Rajalakshmi, R. (2013). Water soluble conducting polymer composite of polyvinyl alcohol and leucine: An effective acid corrosion inhibitor for mild steel. *Mater. Corros.* 64 (1), 74–82. doi:10.1002/maco.201106096

Sharma, N., Sharma, S., Sharma, S. K., Mahajan, R. L., and Mehta, R. (2022). Evaluation of corrosion inhibition capability of graphene modified epoxy coatings on reinforcing bars in concrete. *Constr. Build. Mater.* 322 (3), 126495–126521. doi:10.1016/j.conbuildmat.2022.126495

Sheikh, M., Kamal, K., Rafique, F., Sabir, S., Zaheer, H., and Khan, K. (2021). Corrosion detection and severity level prediction using acoustic emission and machine learning based approach. *Ain Shams Eng. J.* 12 (4), 3891–3903. doi:10.1016/j.asej.2021.03.024

Shen, C., Alvarez, V., Koenig, J., and Luo, J. L. (2019). Gum Arabic as corrosion inhibitor in the oil industry: experimental and theoretical studies. *Corros. Eng. Sci. Technol.* 2019 (4), 444–454. doi:10.1080/1478422X.2019.1613780

Shukla, S., Quraishi, M., and Prakash, R. (2008). A self-doped conducting polymer "polyanthranilic acid": An efficient corrosion inhibitor for mild steel in acidic solution. *Corros. Sci.* 50 (10), 2867–2872. doi:10.1016/j.corsci.2008.07.025

Sobhi, M., and Eid, S. (2018). Chemical, electrochemical and morphology studies on methyl hydroxyethyl cellulose as green inhibitor for corrosion of copper in hydrochloric acid solutions. *Prot. Met. Phys. Chem. Surf.* 54 (5), 893–898. doi:10.1134/S2070205118050210

Sushmitha, Y., and Rao, P. (2019). Material conservation and surface coating enhancement with starch-pectin biopolymer blend: A way towards green. *Surfaces Interfaces* 16 (9), 67–75. doi:10.1016/j.surfin.2019.04.011

Suyanto, H., Ratih, R., and Leo, S. (2015). Application chitosan derivatives as inhibitor corrosion on steel with fluidization method. *J. Chem. Pharm. Res.* 7 (2), 260–267. Available at: http://repository.unair.ac.id/id/eprint/49764.

Talha, M., Ma, Y., Kumar, P., Lin, Y., and Singh, A. (2019). Role of protein adsorption in the bio corrosion of metallic implants–A review. *Colloids Surfaces B Biointerfaces* 176 (4), 494–506. doi:10.1016/j.colsurfb.2019.01.038

Tan, B., Zhang, S., Cao, X., Fu, A., Guo, L., Marzouki, R., et al. (2022). Insight into the anti-corrosion performance of two food flavors as eco-friendly and ultra-high performance inhibitors for copper in sulfuric acid medium. *J. Colloid Interface Sci.* 609 (11), 838–851. doi:10.1016/j.jcis.2021.11.085

Tan, B., Zhang, S., Qiang, Y., Li, W., Li, H., Feng, L., et al. (2020). Experimental and theoretical studies on the inhibition properties of three diphenyl disulfide derivatives on copper corrosion in acid medium. *J. Mol. Liq.* 298 (1), 111975–112013. doi:10.1016/j.molliq.2019.111975

Umoren, S., Ebenso, E., and Ogbobe, O. (2009). Synergistic effect of halide ions and polyethylene glycol on the corrosion inhibition of aluminium in alkaline medium. *J. Appl. Polym. Sci.* 113 (6), 3533–3543. doi:10.1002/app.30258

Verma, C., Quraishi, M., and Rhee, K. (2022). Aqueous phase polymeric corrosion inhibitors: Recent advancements and future opportunities. *J. Mol. Liq.* 348 (2), 118387–118422. doi:10.1016/j.molliq.2021.118387

Vijayan, P., and Al-Maadeed, M. (2019). Self-repairing composites for corrosion protection: A review on recent strategies and evaluation methods. *Materials* 12 (17), 2754–2817. doi:10.3390/ma12172754

Wu, W., Chen, T., Du, H., Li, D., and Liu, J. (2016). Carboxymethyl starch as corrosion inhibitor for mild steel. *Emerg. Mater. Res.* 5 (2), 277–283. doi:10.1680/jemmr.15.00037

Yang, M., Wu, J., Fang, D., Li, B., and Yang, Y. (2018). Corrosion protection of waterborne epoxy coatings containing mussel-inspired adhesive polymers based on polyaspartamide derivatives on carbon steel. *J. Mat. Sci. Technol.* 34 (12), 2464–2471. doi:10.1016/j.jmst.2018.05.009

Zeng, J., Chen, H., Li, S. L., and Wang, Y. Z. (2012). Chitin whiskers: An overview. *Biomacromolecules* 025 (03), 1–11. doi:10.1021/bm201564a

Zhang, Z., Ba, H., and Wu, Z. (2019). Sustainable corrosion inhibitor for steel in simulated concrete pore solution by maize gluten meal extract: Electrochemical and adsorption behavior studies. *Constr. Build. Mater.* 227 (12), 117080. doi:10.1016/j.conbuildmat.2019.117080

Zunita, M., Wahyuningrum, D., Wenten, I. G., and Boopathy, R. (2022). Carbon steel corrosion inhibition activity of tofu associated proteins. *SSRN J.* 17 (2), 1–7. doi:10.2139/ssrn.3970448

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 11 October 2022
DOI 10.3389/fenrg.2022.1020040

OPEN ACCESS

EDITED BY
Yongming Han,
Beijing University of Chemical Technology, China

REVIEWED BY
Yunkai Wu,
Jiangsu University of Science and Technology, China
Hasmat Malik,
University of Technology Malaysia, Malaysia

*CORRESPONDENCE
Tahir Khurshaid,
tahir@ynu.ac.kr
Salah Kamel,
skamel@aswu.edu.eg

SPECIALTY SECTION
This article was submitted to Process and Energy Systems Engineering, a section of the journal Frontiers in Energy Research

RECEIVED 15 August 2022
ACCEPTED 15 September 2022
PUBLISHED 11 October 2022

CITATION
Equbal MD, Nezami MM, Hashem H, Bajaj M, Khurshaid T, Ghoneim SSM and Kamel S (2022), IOT based classification of transformer faults using emerging techniques of E-nose and ANFIS. *Front. Energy Res.* 10:1020040. doi: 10.3389/fenrg.2022.1020040

IOT based classification of transformer faults using emerging techniques of E-nose and ANFIS

Md. Danish Equbal[1], Md. Manzar Nezami[2], Hythem Hashem[3], Mohit Bajaj[4], Tahir Khurshaid[5]*, Sherif S. M. Ghoneim[6] and Salah Kamel[7]*

[1]Department of Electrical Engineering, Galgotias College of Engineering and Technology, Greater Noida, India, [2]Department of Electronics and Communication Engineering, GLA University, Mathura, India, [3]Department of Computer Science, Faculty of Science and Arts at Belqarn, University of Bisha, Bisha, Saudi Arabia, [4]Department of Electrical Engineering, Graphic Era (Deemed to be University), Dehradun, India, [5]Electrical Engineering Department, Yeungnam University, Gyeongsan, South Korea, [6]Department of Electrical Engineering, College of Engineering, Taif University, Taif, Saudi Arabia, [7]Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan, Egypt

E-Nose finds its use in a wide range of applications such as quality assessment in food processing to toxic gas identification in chemical industry either in the offline or online mode. Their usage can be extended to transformer condition monitoring in the online mode. Considering the importance of transformers in power system and the impact it could create if faults in them are unidentified or left unattended, their functioning should be monitored on a real time basis. This work, describes the realization of a prospective E-Nose for online transformer incipient fault identification. The resistive gas sensor array has been simulated in real time using variable resistances forming one arm of a Wheatstone bridges. Separate variable resistances have been calibrated using characteristics of different fault gas sensors. The sensor array of the E-Nose helps to identify the transformer fault gases resulting from an incipient fault condition at the nascent stage itself and prompts for the necessary corrective action well before a catastrophic situation arises. Furthermore, ANFIS model of the Duval's Triangle (DT) method have been developed to facilitate the online classification of incipient faults. The ANFIS models of other popularly used incipient fault interpretation methods, reported in earlier works, have also been used for a comparative analysis on their diagnostic capabilities. The developed model has been tested using the fault cases of IEC-TC10 fault database and the results thus obtained have been found to be very promising.

KEYWORDS

E-nose, anfis, duval triangle method, incipient faults, power transformer

Introduction

Power transformers are essential and costly power system assets which are used primarily for changing voltage levels in the power system. Transformers operate with their insulation stressed both under normal stable operation and even during fault where the magnitude of stress is augmented by thermal or electrical fault conditions in the insulation. A timely diagnosis of these slow developing anomalies in the transformer could save the transformer and the system as a whole from large scale damages (Duval, 1989). Diagnosis of faults at the incipient stage can prevent large scale component outages. Incipient fault identification methodologies have been designed for the diagnosis of winding faults in squirrel cage induction motors (Wu et al., 2020) and for transformers as well. Dissolved Gas Analysis (DGA) is one such diagnostic tool that aids to predict such eventualities before their occurrences. However, DGA based incipient fault diagnosis has failed to give consistent reliable diagnosis at all times. Hence, the fault interpretation standards have apparent shortcomings in the form of no diagnosis or unresolved diagnosis. This calls for use of other tools that would aid the standards in the diagnosis process by removing these apparent fallacies (Duval, 2003; Duval, 2002; Singh and Bandyopadhyay, 2010a; Wani et al., 2015; Khan et al., 2020; Benmahamed et al., 2021; Zhang et al., 2022).

E-Nose system is essentially an array of sensors which can respond to a wide range of inputs. The need for E-Nose for transformer health monitoring arises from the requirement of nonintrusive diagnostic method (Gardner and Bartlett, 1994; Wilson and Baietto, 2009).

A host of soft computing methodologies have been implemented using the diagnostic standards of IEEE and IEC (Interpretation of the analysis of, 1978; Rogers, 1978; Hooshmand and Banejad, 2008; Author anonymous, 2009). The possibility of transformer incipient fault diagnosis and transformer insulation health monitoring has been explored using Artificial Neural Network (ANN) and Machine Learning (ML) (Guardado et al., 2001; Equbal et al., 2018; Nezami et al., 2021a; Ghoneim et al., 2021; Kherif et al., 2021; Taha et al., 2021), Fuzzy Logic (FL) system (Dukarm, 1993; Dhote and Helonde, 2012; Huang and Sun, 2013; Noori et al., 2017) and Adaptive Neuro Fuzzy Inference System (ANFIS) (Hooshmand et al., 2012; Khan et al., 2014; Vani and Murthy, 2014; Khan et al., 2015; Nezami et al., 2021b). However, these methods have their own limitations which needs to be addressed for their effective use in an online system.

In this work an online system of incipient fault identification of transformer is showcased through the use of an E-Nose and ANFIS based fault interpretation model. The E-Nose is realized by using an array of variable resistances simulating gas sensors instead of directly employing the sensors for identifying the evolved gas concentrations. These variable resistances simulate various resistive fault gas sensors. Variable resistance values are converted into gas concentrations by calibrating them using real sensor characteristics. Furthermore, one of the popular methods of incipient fault classification, the Duval's Triangle Method (Duval, 1989; Duval, 2003; Duval, 2002; Singh and Bandyopadhyay, 2010a; Singh and Bandyopadhyay, 2010b), has been implemented using ANFIS. The E-Nose is interfaced with the online incipient fault detection model using a microcontroller. The gas concentrations obtained from the E-Nose are used as inputs in the ANFIS model which in turn helps diagnosis the fault developing in the transformer.

This paper proposes an online non-destructive procedure of transformer incipient fault identification without the use of sophisticated instruments and without interrupting the transformer operation. It aims to identify the gases evolved as a result of transformer oil decomposition and relay this information to an online system hosting the ANFIS model for interpretation of fault type. The conventional method involves a destructive procedure requiring the extraction of transformer oil and determination of dissolved gases in oil using gas chromatography and then interpreting the fault gases to converge upon the fault type. Here, the objective is to obtain a model that facilitates non-intrusive online detection of incipient fault with a high degree of accuracy. The testing of the developed system has been done using fault cases reported in IEC-TC10 fault database (Duval and DePablo, 2001; Gouda et al., 2021). The outcome of the diagnostic tests has been found to be reliable and suggests that the developed model can be regarded as a prospective online incipient fault identification system.

Development of an ANFIS model of the duval's triangle method for online transformer incipient fault classification

Dissolved gas analysis

Transformer insulating oils, under normal aging as well as under increased thermal and electrical stress resulting from incipient faults, decomposes to release fault gases. The type and concentration of gases is fault dependent such that aging of oil will result in the production of the low molecular weight gases while incipient fault-related gases include hydrocarbons, oxides of carbon and some non-fault gas types as well. Sustained faults in transformers that are allowed to persist for longer duration may eventually lead to catastrophic failure of the transformer. The different types of gases that evolve due to transformer oil decomposition are: 1) Hydrocarbon and hydrogen-a. Methane (CH_4) b. Ethane (C_2H_6) c. Ethylene (C_2H_4) d. Acetylene (C_2H_2) e. Hydrogen (H_2) 2) Carbon oxides-a. Carbon monoxide (CO) b. Carbon dioxide (CO_2) and 3) Non fault gases a. Nitrogen (N_2) b. Oxygen (O_2). Further, the classification may also be done on the basis of

combustible gases (H_2, CO, CH_4, C_2H_6, C_2H_4 and C_2H_2) and non-combustible gases (CO_2, N_2 and O_2). The gases released in the process of oil decomposition get dissolved in the liquid insulation. The identification of these gas types and their concentration would indicate the possible fault type (Singh and Bandyopadhyay, 2010a).

DGA is one of the most commonly used tools for online health monitoring of transformers in service. DGA has gained immense acceptability in maintenance procedures due to its ability to detect incipient faults in addition to providing early warnings of slow developing faults and the rate of the fault build up.

A number of ratio and concentration-based transformer incipient fault diagnosis methods have been universally adopted for determining the existing fault type in the transformer oil insulation using the DGA data. Various indicators such as Total Dissolved Combustible Gaz (TDCG), Total Dissolved Hydrogen and Hydrocarbon Gas (TDHG), the primary gases- Methane, Ethane and Ethylene (MEA) and thermal fault gases-Total Heat Gases (THG) use combination of gas concentrations for performing health assessment of transformer insulation. The TDCG method is most commonly used method among them and has been elaborated in the IEEE C57.104-2008 standard (Author anonymous, 2009). The ratio-based methods include the Roger's Ratio Method, Doernenberg Ratio Method, and IEC methods while the Duval's triangle method is a graphical method that uses relative gas concentration for performing diagnosis functions (Duval, 2002; Singh and Bandyopadhyay, 2010a; Zhang et al., 2022).

Duval's triangle method of incipient fault identification

The Duval's Triangle method is a transformer incipient fault diagnosis method that has gained world-wide acceptability due to its simplicity in application and interpretation, as well as its effectiveness in diagnosing fault conditions (Duval, 2002). It is a graphical method in which the region within the triangle is demarcated into seven zones, each representing a fault condition. This method uses the relative concentration of three fault gases (CH_4, C_2H_4 and C_2H_2) to identify seven possible fault conditions namely, Partial Discharge (PD), Discharges of Low Energy (D1), Discharges of High Energy (D2), Thermal Fault, T< 300°C (T1), Thermal Fault, 300°C < T < 700°C (T2), Thermal Fault, T > 700°C (T3) and an intermediate fault zone comprising of a mix of electrical and thermal fault (DT). Figure 1 shows the Duval's triangle along with the fault regions represented as polygons within the triangle. The relative concentration values of the three gases are calculated and plotted on their respective sides of the triangle using Eqs 1–3. The zone in which the point of intersection of the projection of the relative gas concentration from each side lies indicates the incipient fault type.

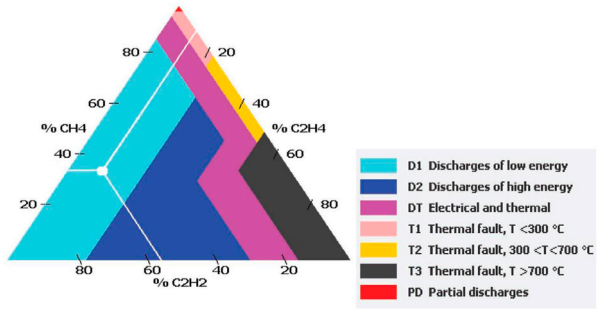


FIGURE 1
Duval's triangle zones showing the identifiable fault types.

This method provides a fairly reliable diagnosis. The formulation of this graphical method is easy but the interpretation of the results requires precision. In addition, for transformer incipient fault diagnosis, the Duval's triangle method mandates a fault condition to exist as it does not have the provision to diagnose a non-fault condition. It assigns a fault type to the no-fault cases, thus giving a wrong diagnosis (Duval, 2003) (Khan et al., 2020) (Singh and Bandyopadhyay, 2010b).

$$\%CH_4 = \frac{Conc.of\ CH_4 in\ ppm}{Total\ Conc.of\ (CH_4 + C_2H_4 + C_2H_2)} \times 100, \quad (1)$$

$$\%C_2H_4 = \frac{Conc.of\ C_2H_4 in\ ppm}{Total\ Conc.of\ (CH_4 + C_2H_4 + C_2H_2)} \times 100, \quad (2)$$

$$\%C_2H_2 = \frac{Conc.of\ C_2H_2 in\ ppm}{Total\ Conc.of\ (CH_4 + C_2H_4 + C_2H_2)} \times 100, \quad (3)$$

Development of ANFIS model of the duval's triangle method

ANFIS is a hybrid approach to input-output mapping that makes use of the advantages of neural networks' capacity for learning within the context of fuzzy inference systems. The generalised structure of ANFIS, which consists of nodes connected to one another through directed linkages, is depicted in Figure 2. ANFIS contains five layers, each of which has a fixed or adaptable node. The output of the adaptive square nodes depends on the input parameters. While the circular nodes are stationary and their output is dependent on the output of the previous layer. The directed linkages offer a conduit for the signal's flow without changing its weight (Khan et al., 2015; Nezami et al., 2021b).

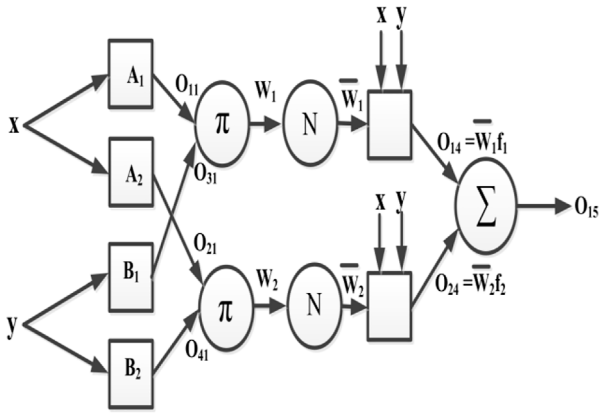


FIGURE 2
Generalized ANFIS structure.

Here, A_1, A_2, B_1, and B_2 are the four nodes that make up layer 1, the top layer. Each nodes represent a fuzzy set that converts the input to the output based on how closely it resembles the fuzzy sets. The output of node k in the first layer, O_{k1}, is determined by Eqs 4–7.

$$O_{11} = \mu_{A1}(x) \tag{4}$$
$$O_{21} = \mu_{A2}(x) \tag{5}$$
$$O_{31} = \mu_{B1}(y) \tag{6}$$
$$O_{41} = \mu_{B2}(y) \tag{7}$$

The circular shape of the layer two nodes indicates its fixed nature. These nodes combine the input signals to produce an output that is represented by Eqs 8, 9.

$$w_1 = \mu_{A1}(x) \times \mu_{B1}(y) \tag{8}$$
$$w_2 = \mu_{A2}(x) \times \mu_{B2}(y) \tag{9}$$

The third layer's nodes carry out the layer 2's output's normalizing function. Eqs 10, 11 can be used to express this.

$$\bar{w}_1 = w_1/(w_1 + w_2) \tag{10}$$
$$\bar{w}_2 = w_2/(w_1 + w_2) \tag{11}$$

The outputs of layer four nodes are of the adaptive kind, and they depend on the nodes' ensuing parameters. Because of this, the outputs of this layer mix the output of layer 3 with the parameters chosen by the fuzzy sugeno system. Eqs 12, 13 serve as representations of this form.

$$O_{14} = \bar{w}_1 f_1 \tag{12}$$
$$O_{24} = \bar{w}_2 f_2 \tag{13}$$

Where $\bar{w}_1$ and $\bar{w}_2$ stand for the layer three outputs that have been normalised, and f_1 and f_2 are the fuzzy Sugeno system parameters. Eqs 14, 15 are used to determine these parameters.

$$f_1 = p_1 x + q_1 y + r_1 \tag{14}$$
$$f_2 = p_2 x + q_2 y + r_2 \tag{15}$$

where the system parameter set is composed of p, q, and r. In order to produce the final output, layer 5, the last layer, summarizes the incoming signals O_{14} and O_{24} to estimate O_{15}. Eq. 16 is used to calculate this layers output.

$$O_{15} = (w_1 f_1 + w_2 f_2)/(w_1 + w_2) \tag{16}$$

The ANFIS model for the implementation of Duval's triangle method has been developed in the MATLAB environment. The model has been trained based on the confinement of each fault type in a given zone. Separate ANFIS models have been trained suitably to identify their respective fault type using a training dataset comprising of 1821 simulated data points. Each of the ANFIS model has been trained to keep the error within a tolerance of 0.001. The triangular membership function when selected as the input MF gave the most accurate diagnosis. Hence, triangular MF has been chosen to represent the inputs while linear MF has been selected as the output MF in the developed ANFIS model. Seven ANFIS models have been developed, one for each of the seven faults that can be detected by the Duval's Triangle method. The input gas concentrations are simultaneously applied to all the seven ANFIS model and the fault corresponding to the highest value in the output vector is identified as the prevalent fault type. The input-output pattern which forms the basis for the development of the ANFIS models is given in Table 1. The input values of P1, P2 and P3 are the contributions in terms of relative percentage of the three fault gases forming the three sides of the triangle. The input codes are taken as the coordinates of the vertices of the polygon representing the respective zones formed inside the triangle. Additional care has been taken to handle overlapping points by preventing replication of data points in two fault zones up to four places of decimals.

Faults lying on the boundaries of two fault zones has been treated to be of the more severe type among them so as to prompt for immediate corrective action. This generalization for the ANFIS model has been made in view of the fact that if a given fault condition is sustained for larger time duration will lead to a more severe fault condition.

The MATLAB based GUI of the ANFIS model for Duval's triangle method has been shown in Figure 3. This GUI can be utilized for online detection of the prevalent incipient fault in the transformer.

Development of E-NOSE based online transformer incipient fault identification system

Realization of E-Nose for transformer incipient fault gas identification

The objective here is to make a prototype of an E-Nose that would work in conjunction with an ANFIS based fault interpretation model

TABLE 1 Input-output codes for the ANFIS model of the Duval's triangle.

No.	Type of fault	Input coordinates			Output codes						
		P1	P2	P3	F1	F2	F3	F4	F5	F6	F7
1	Partial Discharge (PD)	0.7999	0.199	0.0000	1	0	0	0	0	0	0
		1.0000	0.0000	0.0000							
		0.9799	0.0000	0.0199							
2	Discharge of Low Energy (D1)	0.0000	0.0000	1.0000	0	1	0	0	0	0	0
		0.0000	0.2299	0.7699							
		0.6399	0.2299	0.1299							
		0.8699	0.0000	0.1299							
3	Discharge of High Energy (D2)	0.0000	0.2300	0.7700	0	0	1	0	0	0	0
		0.0000	0.7099	0.2899							
		0.3099	0.3999	0.2899							
		0.4699	0.3999	0.1299							
		0.6400	0.2300	0.1300							
4	Thermal Fault, T<300°C (T1)	0.7599	0.1999	0.0399	0	0	0	1	0	0	0
		0.7999	0.1999	0.0000							
		0.9800	0.0200	0.0000							
		0.9800	0.0000	0.0200							
		0.9599	0.0000	0.0399							
5	Thermal Fault, 300°C < T<700°C (T2)	0.4599	0.4999	0.0399	0	0	0	0	1	0	0
		0.4999	0.4999	0.0000							
		0.8000	0.2000	0.0000							
		0.7600	0.2000	0.0400							
6	Thermal Fault, T>700°C (T3)	0.0000	0.8499	0.1499	0	0	0	0	0	1	0
		0.0000	1.0000	0.0000							
		0.5000	0.5000	0.0000							
		0.3499	0.4999	0.1499							
7	Mix of Both Electrical and Thermal Faults (DT)	0.0000	0.7100	0.2900	0	0	0	0	0	0	1
		0.0000	0.8500	0.1500							
		0.3500	0.5000	0.1500							
		0.4600	0.5000	0.0400							
		0.9600	0.0000	0.0400							
		0.8700	0.0000	0.1300							
		0.4700	0.4000	0.1300							
		0.3100	0.4000	0.2900							

for assessing the condition of the transformer and identifying the type of incipient fault existing in the transformer, if any. However, the E-Nose is not designed using an array of actual gas sensors but instead adopts the characteristics of these standard gas sensors to calibrate variable resistances, one for each fault gas type. One arm of the Wheatstone bridge is a variable resistance to account for the resistance change in accordance with change in gas concentration following the sensor characteristics. Output of the Wheatstone bridges have been calibrated using four gas sensors characteristics. The sensors based on which the system has been calibrated are MQ-2 for Hydrogen, MQ-4 for Methane, Figaro TGS2620 for ethylene and a lab fabricated sensor reported in (Lin et al., 2015) for acetylene. The schematic of the experimental setup showing the arrangement of the Wheatstone bridges and the interfacing equipment in the form of an Arduino Mega 2,560 microcontroller for online readings is shown in Figure 4. The essential specifications of the four sensors are summarized in Table 2.

As stated earlier, the feasibility of usage of an E-Nose for transformer incipient fault identification is studied through an arrangement of variable resistances simulating the actual characteristics of the sensors. The variable resistance is varied imitating the resistance variation of the sensing material due to

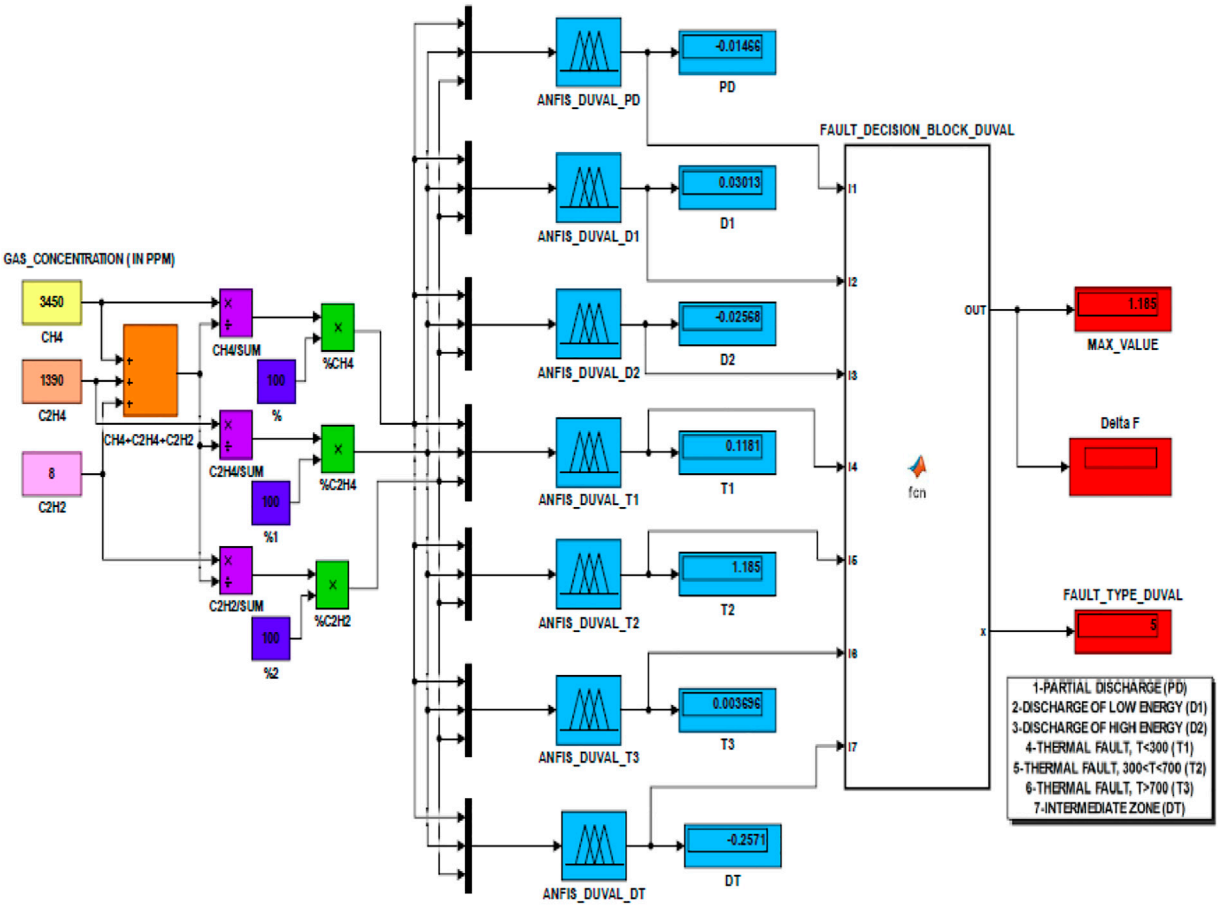


FIGURE 3
MATLAB GUI for ANFIS based implementation of Duval's triangle method.

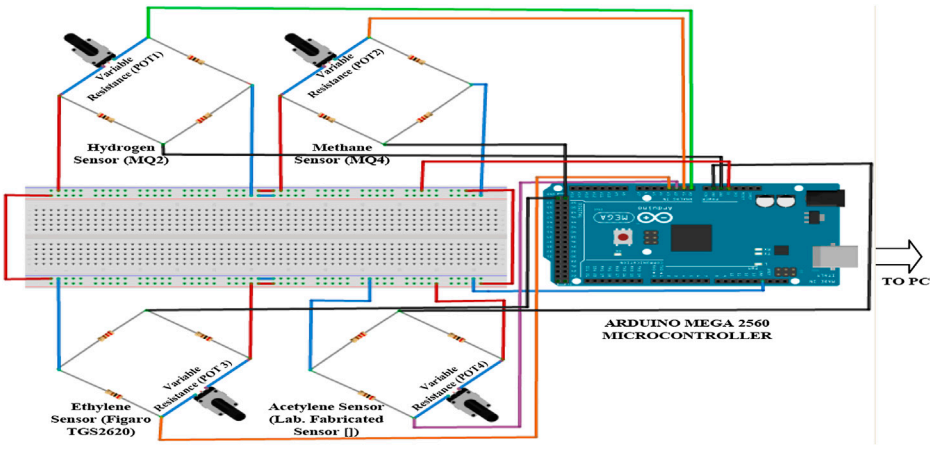


FIGURE 4
Schematic arrangement of a simulated E-Nose for online transformer incipient fault identification.

TABLE 2 Specifications of gas sensors.

S. No	Sensor	Identifiable gas	Sensing material	Concentration range	Type
1	MQ-2	H_2	SnO_2	300–10000 ppm	Resistive
2	MQ-4	CH_4	SnO_2	200–10000 ppm	Resistive
3	Figaro TGS2620	C_2H_4	MOS	0–100 ppm	Resistive
4	Lab. Fabricated Sensor of (Lin et al., 2015)	C_2H_2	NiO/SnO_2 Composite	0–4,000 ppm	Resistive

the change in gas concentration. The resistance values corresponding to a given concentration is determined from the equations established using the sensitivity characteristics of the sensors.

In the balanced state, the voltage across the output terminals of the bridge remains zero. As the concentrations of the fault gases change, the balance of the respective Wheatstone bridges is disturbed due to the change in the sensor resistance and a voltage appears across the detector terminals of the Wheatstone bridge. This voltage representing the output voltage is detected and displayed on a monitor using a microcontroller-based interface. In order to obtain a diagnosis of the prevailing fault condition in the transformer the output voltage needs to be converted into gas concentrations in ppm. Calibration has been carried out using the voltage-concentration relation derived for each sensor.

The response characteristics (resistance-gas concentration relationship) and the voltage-gas concentration relationship for the MQ-2, MQ-4, Figaro TGS2620 and the laboratory fabricated acetylene sensor of (Lin et al., 2015) are given in Figure 5 (*i*)–(*iv*), respectively. The resistance-gas concentration relationships are obtained by extracting data points from the response characteristics given in the data sheet of the respective sensor while the voltage-gas concentration plot is obtained by observing the voltage across the detector for the for the whole range of sensor resistance extracted from the response characteristics. In the figures Rs denotes the sensor resistance while Ro is the resistance of sensor at a particular concentration (say 1,000 ppm) in fresh air. Its value can be determined by calibrating the actual sensors in fresh air. In this experimental study, Ro has been designated a value of 1 KΩ for all the sensors. It can be observed from the plots of all the characteristics that the R-squared values i.e. the regression is very close to 1, thus, indicating a good fit.

The relation between gas concentration in ppm and the sensor resistance within the detectable concentration range and the relation between the output voltage and gas concentration are given by Eqs 17, 18, respectively for the MQ-2 hydrogen sensor, Eqs 19, 20, respectively for MQ-4 methane sensor, Eqs 21, 22, respectively for ethylene sensor and Eqs 23, 24, respectively for acetylene sensor.

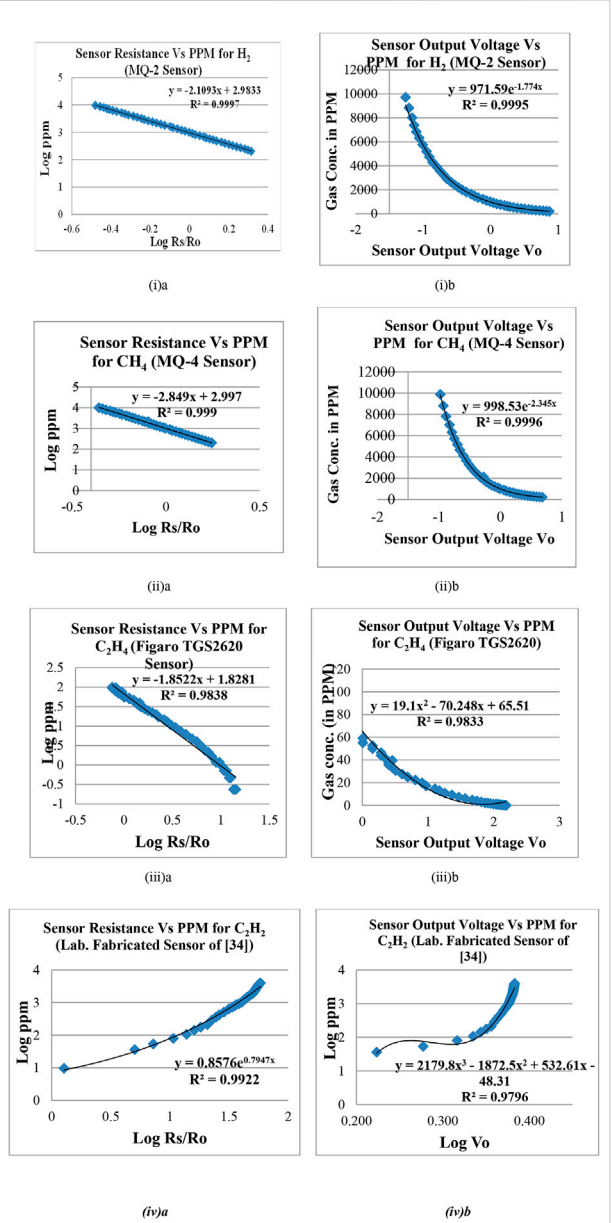


FIGURE 5
(A) Sensitivity characteristics plot and **(B)** output Voltage-PPM plot- (i) MQ-2 sensor for H_2 (ii) MQ-4 sensor for CH_4 (iii) Figaro TGS2620 for C_2H_4 (iv) lab. Fabricated sensor for C_2H_2 (Lin et al., 2015)

$$\text{Logppm} = -2.109 \times (\log R_s/R_o) + 2.983, \quad (17)$$

$$\text{ppm} = 971.5e^{-1.77Vo} \quad (18)$$

$$\text{logppm} = -2.849 \times \log(R_s/R_o) + 2.997 \quad (19)$$

$$\text{ppm} = 998.5e^{-2.34Vo} \quad (20)$$

$$\text{logppm} = -1.852 \times \log(R_s/R_o) + 1.828 \quad (21)$$

$$\text{ppm} = 19.1 \times Vo^2 - 70.24 \times Vo + 65.51 \quad (22)$$

$$\text{logppm} = 0.857e^{0.794 \times \log(R_s/R_o)} \quad (23)$$

$$\text{logppm} = 2179 \times (\log Vo)^3 - 1872(\log Vo)^2 + 532.6 \times (\log Vo) - 48.31 \quad (24)$$

Interface for online transformer incipient fault study

The setup for the simulated E-Nose based online transformer incipient fault identification is given in Figure 6. It can be observed that the sensor array output voltages samples are acquired to host computer using an Arduino Mega 2,560 microcontroller interface. The Arduino Mega 2,560 has an ATmega2,560 microcontroller housed on board that provides adequate clock speed, RAM and analog input channels for the application. The objective of acquiring the data on a computer is to enable a host fault diagnosis algorithm to identify the fault type after calibration of the sensor output voltage.

In the setup the variable resistances are simulating the actual gas sensors of the E-Nose and are calibrated using their characteristics. The change in gas concentration leads to change in sensor resistance which results in an output voltage appearing across the output terminals of the Wheatstone bridge. The output terminals of the bridges are wired to the analog input of Arduino Mega 2,560. The Simulink model on the host computer needs to acquire this voltage available with the hardware so that after suitable conversion into gas concentration the diagnostic functions can be performed online.

In order to acquire the voltage data from the Arduino hardware a communication interface needs to be setup. Simulink model has been executed in the normal mode as normal mode execution of the simulation model ensures its validation as well as accelerates the process since no code generation or model deployment is required. The Simulink IO that comes with the Simulink support package for Arduino hardware creates a communication link that allows the Simulink model and the IO server on the hardware to communicate among themselves. Figure 7 gives an insight on the interaction between the simulation model and the Arduino hardware in the normal mode with Simulink IO.

Arduino being a microcontroller-based device are capable of performing the Analog to Digital Conversion (ADC) when the inputs are given to the analog ports on the board of an Arduino Mega 2,560. The output of the ADC is a ratiometric values, proportional to the system voltage. For a 10-bit microcontroller the output of an ADC ranges from 0 to 1,023, where 0 corresponds to an output voltage of 0V while and an ADC output reading of 1,023 would imply 5V output. Hence, a conversion of the ADC readings needs to be done for obtaining output voltages in Volts using a simple relation given in Eq. 25.

$$\frac{Resolution\,of\ ADC\ (1023\ for\ Arduino\ Mega)}{System\ Input\ Voltage\ (Here\ 5V)} = \frac{ADC\ Reading}{Output\ Analog\ Voltage} \quad (25)$$

Based on this relation the voltage data acquired in the simulation model from Arduino Mega 2,560 is a 10-bit voltage value with a resolution of 0.0048875V.

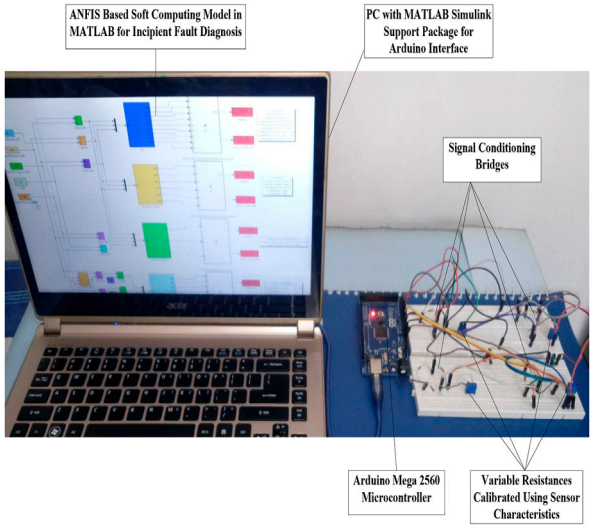


FIGURE 6
Setup for online transformer incipient fault diagnosis using an E-Nose prototype.

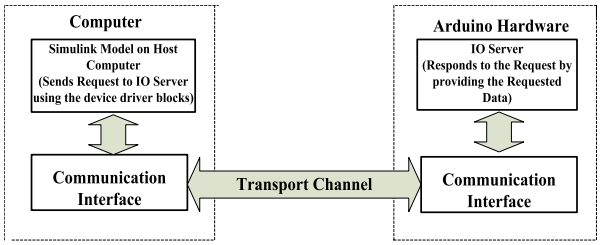


FIGURE 7
Communication between host computer and Arduino hardware.

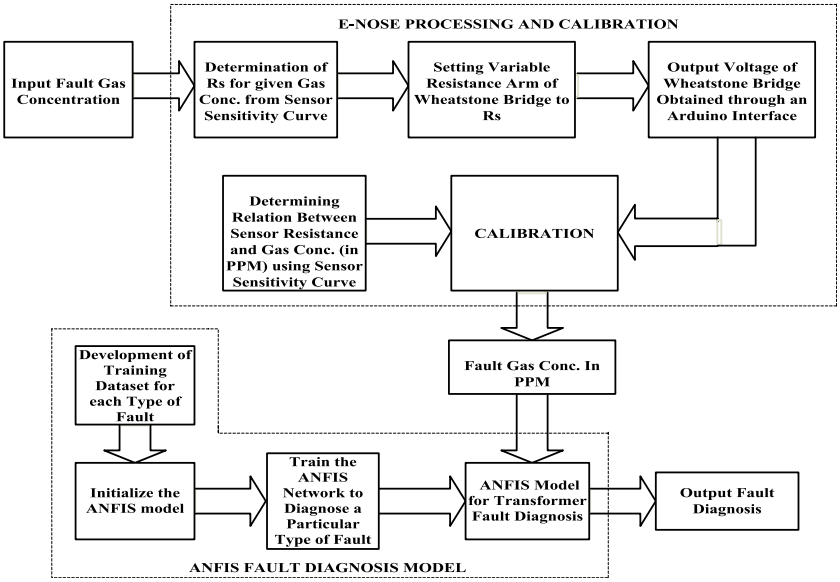


FIGURE 8
Flow diagram of the fault diagnosis process.

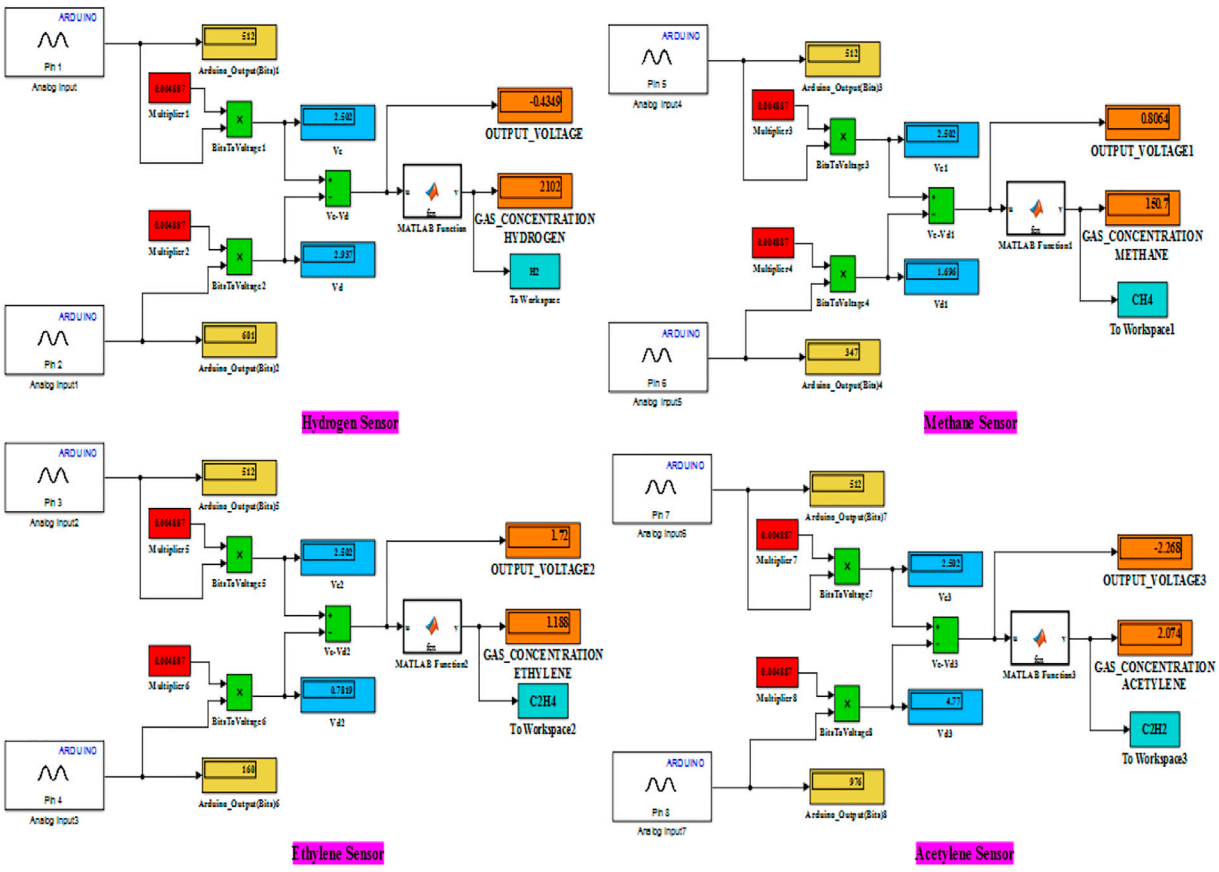


FIGURE 9
MATLAB subsystem for acquiring output voltages and conversion to gas concentration.

Online transformer incipient fault detection and classification model

The output from the Arduino Mega after suitable conversion to voltages is used to obtain the gas concentrations in ppm. The conversions to ppm follows the relations obtained from the voltage-ppm curve for each of the sensors. These gas concentrations are used as input to the fault identification algorithm for diagnosing the existing incipient fault. The complete process of the online fault identification model is shown as a flow diagram in Figure 8.

The output of the E-Nose in the form of fault gas concentrations obtained after suitable conversion of the output voltages resulting from variation in sensor resistances into gas concentrations in ppm serves as the inputs for the ANFIS based fault diagnosis model. The MATLAB Simulink model to acquire the output voltages and their subsequent conversions into gas concentrations is shown in Figure 9. The fault interpretation model is a composite model comprising ANFIS model of the DT method developed above.

And the ANFIS models based on IEC-599 standard, Roger's ratio method, Doernenberg ratio method as reported in (Khan et al., 2015). The standard fault identification methods, in general, depend on the concentration of five primary gases to recognize the fault condition in the transformer. These gases are namely, hydrogen, methane, ethane, ethylene and acetylene. However, in the prototype for E-Nose based online transformer incipient fault detection method only four sensors have been considered, one each for hydrogen, methane, ethylene and acetylene. The ethane gas concentration has been manually fed in the fault identification process. In the event that an ethane gas sensor with sensing capabilities in the desired gas concentration range is obtained/ fabricated, their incorporation would be in a manner similar to the other sensors.

The ANFIS based fault interpreting Simulink model is given in Figure 10.

Result and discussion

The ANFIS model based on the Duval's triangle method has been tested to identify its diagnostic capability using incipient fault cases from field data of faulty transformers reported in (Duval and DePablo, 2001). The accuracy

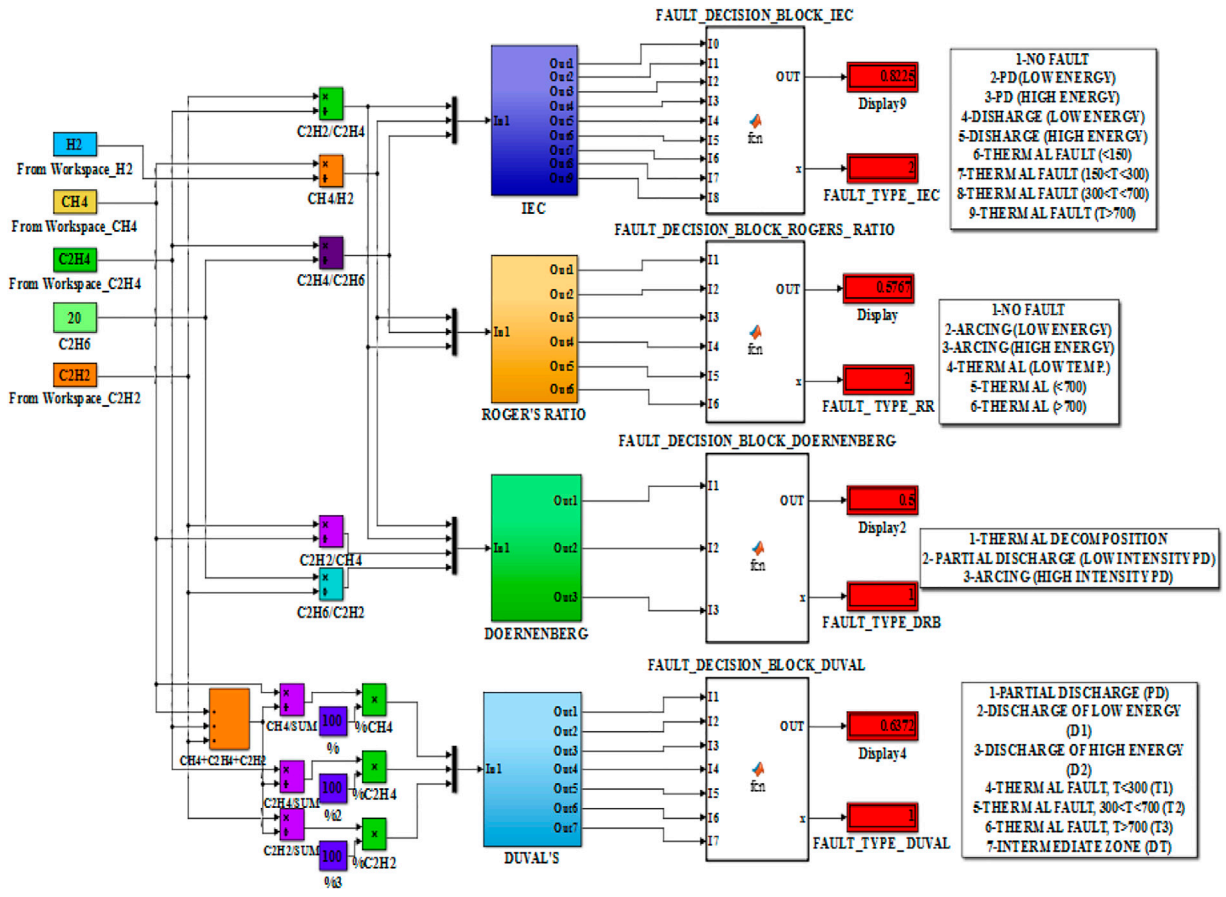


FIGURE 10
ANFIS model for online transformer incipient fault diagnosis.

of the ANFIS model for DT method was determined to be 97%.

The developed setup, simulating an E-Nose, applied for the identification of transformer incipient faults has been also tested using the fault cases reported in (Duval and DePablo, 2001). Since the sensors are limited to operate within a specified gas concentration range, only those fault cases which satisfy these ranges have been selected for testing. These diagnosable cases along with their respective gas concentrations are given in Table 3. These fault cases were diagnosed to check the diagnostic efficiency of the developed online E-Nose system. The sensor resistance for a particular sensor is obtained using the sensitivity characteristics of that sensor by projecting the given gas concentration on the ppm-sensor resistance curve or by using the equation derived from the characteristics.

The combined diagnosis results for all the fault cases under study are given in Table 4. It is observed that the ANFIS models based on the IEC-599, Roger's ratio method, Doernenberg method and the Duval's triangle method have an accuracy of 40%, 44%, 72% and 80%, respectively. Unlike the other fault interpretation models the Duval's triangle method requires only threes gas inputs i.e. gas concentrations of CH_4, C_2H_4 and C_2H_2 and they are all acquired from the output of the sensor array prototype.

In order to showcase the diagnostic capability of the proposed method a comparison of diagnosis of sample fault cases reported in published works has been shown in Table 5. The cases have been carefully selected to satisfy the ppm limitations of the CH_4, C_2H_2, C_2H_4 sensors of the E-Nose while the limitations of other gases are ignored as they are insignificant in incipient fault diagnosis by Duval's method. It can be observed that the ANFIS model of the Duval Triangle method successfully diagnoses all cases while the E-Nose based diagnosis incorrectly diagnoses the high thermal fault in Case 3 as a mix of Thermal Electrical fault.

Furthermore, the proposed E-Nose based transformer incipient fault model has been tested for the impact in diagnosis due to errors introduced in the measurement of gases, errors in the interfacing devices or due to the conditions in the transformers. Table 6 shows the sample transformer cases of Table 3 being subjected to ±5%, ±10%, ±15% and ±20% errors. The conclusion drawn

TABLE 3 Transformer fault cases for diagnostic testing of E-Nose model.

S. No.	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	Fault identified by inspection
1	8,266	1,061	0	0	22	PD
2	1,000	500	500	400	1	D1
3	60	10	4	4	4	D1
4	95	10	39	11	0	D1
5	35	6	482	26	3	D1
6	210	22	7	6	6	D1
7	385	60	159	53	8	D1
8	595	80	244	89	9	D1
9	120	25	40	8	1	D1
10	8	0	101	43	0	D1
11	1,330	10	182	66	20	D1
12	4	1	52	7	2	D1
13	57	24	30	27	2	D1
14	120	31	94	66	0	D2
15	13	3	6	3	1	D2
16	137	67	104	53	7	D2
17	34	21	56	49	4	D2
18	150	130	30	55	9	D2
19	75	15	26	14	7	D2
20	90	28	32	31	8	D2
21	48	610	0	10	29	TH1
22	12	18	0	4	4	TH1
23	2031	149	0	3	20	TH1
24	150	22	11	60	9	TH2
25	1	8	6	100	8	TH2

PD- Partial Discharge; D1- Discharge of Low Energy; D2- Discharge of High Energy; TH1- Thermal Fault Less than 700°C and TH2- Thermal Fault in excess of 700°C

TABLE 4 Transformer fault diagnosis results using the E-Nose setup.

S. No.	IEC-599 standard		Rogers ratio method (RRM) diagnosis		Doernenberg ratio method (DRM) diagnosis		Duval's triangle method (DTM) diagnosis	
	IEC-599 diagnosis	E-Nose based Diagnosis	RRM Diagnosis	E-Nose based Diagnosis	DRM Diagnosis	E-Nose based diagnosis	DTM Diagnosis	E-Nose based diagnosis
1	F0	**F2**	F0	F0	NV	**F1**	F1	F1
2	ND	**F0**	ND	**F0**	NV	F3	F3	F3
3	F3	**F0**	ND	**F0**	NV	**F2**	F2	**F7**
4	F3	F3	ND	F2	NV	F3	F2	F2
5	ND	**F0**	ND	**F0**	F3	**F1**	F2	F2
6	F2	F2	ND	**F0**	NV	F3	F2	F2
7	F3	F3	F2	F2	F3	F3	F2	**F3**
8	F3	F3	F2	F2	F3	F3	F2	F2
19	F3	**F2**	ND	**F0**	F3	F3	F2	F2
10	F4	**F0**	ND	**F0**	NV	F3	F3	F3
11	F3	**F4**	ND	F1	F3	F3	F3	F3
12	F3	F3	ND	**F0**	NV	**F2**	F2	F2
13	F4	**F0**	F2	**F0**	NV	F3	F3	F3
14	F4	**F0**	F2	**F0**	F3	F3	F3	F3
15	F3	**F4**	ND	F2	NV	F3	F3	F3
16	F4	F4	F2	F2	F3	F3	F3	F3
17	F4	**F8**	F2	**F1**	NV	F3	F3	F3
18	ND	**F3**	F2	**F3**	NV	**F2**	F3	**F7**
19	F3	F3	ND	F2	NV	F3	F3	F3
20	F4	F4	F2	F2	NV	F3	F3	F3
21	F6	F6	ND	**F2**	NV	F1	F4	**F1**
22	F7	**F3**	F4	**F2**	NV	F1	F4	**F7**
23	F1	**F2**	F1	**F0**	F2	**F1**	F1	F1
24	F4	**F3**	F2	**F3**	NV	**F2**	F6	F6
25	F8	F8	F5	F5	NV	F1	F6	F6

ND: Non-Diagnosable; NV: Not Valid Ratio

IEC- F0: No Fault; F1: Partial Discharge with low energy density; F2: Partial Discharge with high energy density; F3: Discharge (arc) with low energy; F4: Discharge (arc) with high energy; F5: Thermal faults of temperatures <150°C; F6: Thermal faults of temperatures between 150°C and 300°C; F7:Thermal faults of temperatures between 300°C and 700°C; F8:Thermal faults of temperatures >700°C.
RRM- F0: No Fault; F1: Low-energy density arcing-PD; F2: Arcing-High-energy discharge; F3: Low temperature thermal; F4: Thermal fault < 700°C; F5: Thermal fault >700°C.
DRM- F0: No Fault; F1: Thermal Decomposition; F2: Partial Discharge(Low intensity PD); F3: Arcing (High intensity PD).
Bold values indicate deviation in diagnosis from expected or correct diagnosis.

from the diagnosis is that the even though noises does influence the performance of the model, the E-Nose system along with the ANFIS is able to effectively classify most cases particularly the low ppm fault cases. The faults causing release of higher concentration of fault gases have shown some deviation in their diagnosis.

Conclusion

In this work, an online system for transformer incipient fault identification using an E-Nose has been implemented. Actual sensor characteristics have been used to realize the E-Nose by suitably calibrating variable resistances simulating sensors. The Wheatstone bridge is used as a signal conditioning circuit for the sensors. The outputs of the bridges were voltages, equivalent to the sensor output voltages. The prototype model utilized these output voltages to obtain the gas concentration online through an Arduino Mega 2,560 interface. An ANFIS model using the Duval's Triangle method was developed and used along with ANFIS models of popular interpretation standards were used to arrive upon a fault diagnosis online. The offline testing of the AFIS model based on DT method gave an accuracy of 97%. The selected fault cases in compliance with the detectable concentration range of the sensors, from a fault larger transformer incipient fault database were used to test this online fault diagnosis system. The Duval's triangle

TABLE 5 Comparison of Transformer fault diagnosis with reported results.

S. No	Ref. No.	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	Method used	Diagnosed fault type	Proposed model diagnosis	
									ANFIS model of DT method	E-Nose based diagnosis
1	Li et al. (2016)	48	610	0	10	29	Genetic Algorithm based ratio method	Low and Medium Thermal	F4 (T<300°C)	F4 (T<300°C)
2	Mang-Hui Wang, (2003)	18	262	0	28	41	Extension Method based on the matter-element model and extended relation functions	150–300°C Thermal Fault	F4 (T<300°C)	F4 (T<300°C)
3	Zhang et al. (1999)	56	286	7	28	96	Fuzzy Equivalent Matrix	High Temp. Overheating	F6 (T<700°C)	F7 (DT)
4	Gouda et al. (2018)	180	652.9	4	50	75	Three Ratio Technique	Thermal<150°C	F4 (T<300°C)	F4 (T<300°C)
5	Gouda et al. (2018)	56	334.1	31	32	75	Three Ratio Technique	Low Energy Discharge	F2 (D1)	F2 (D1)
6	Gouda et al. (2018)	33,046	619	0	2	58	Three Ratio Technique	Low Energy Corona Partial Discharge	F1 (PD)	F2 (D1)
7	Gouda et al. (2018)	40,280	1,069	1	1	1,060	Three Ratio Technique	High Energy Corona Partial Discharge	F3 (D2)	F3 (D2)
8	Gouda et al. (2018)	9,340	995	7	6	60	Three Ratio Technique	High Energy Corona Partial Discharge	F3 (D2)	F3 (D2)

TABLE 6 Impact of noise on incipient fault diagnosis.

Transformer Case No. (Table 3)	Noise	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	Expected fault type as per DTM	E-Nose diagnosis
1	+5%	8,679.3	1,114.05	0	0	23.1	PD	PD
	−5%	7,852.7	1,007.95	0	0	20.9	PD	PD
24	+5%	157.5	23.1	11.55	63	9.45	T1	**T3**
	−5%	142.5	20.9	10.45	57	8.55	T1	**T3**
16	+10%	150.7	73.7	114.4	58.3	7.7	D2	D2
	−10%	123.3	60.3	93.6	47.7	6.3	D2	D2
21	+10%	52.8	671	0	11	31.9	T1	**PD**
	−10%	43.2	549	0	9	26.1	T1	**PD**
2	+15%	1,150	575	575	460	1.15	D1	**D2**
	−15%	850	425	425	340	0.85	D1	**D2**
6	+15%	241.5	25.3	8.05	6.9	6.9	D1	D1
	−15%	178.5	18.7	5.95	5.1	5.1	D1	D1
20	+20%	108	33.6	38.4	37.2	9.6	D2	D2
	−20%	72	22.4	25.6	24.8	6.4	D2	D2
22	+20%	14.4	21.6	0	4.8	4.8	T2	T2
	−20%	9.6	14.4	0	3.2	3.2	T2	T2

Bold values indicate deviation in diagnosis from expected or correct diagnosis.

method using three senor output gas concentrations as its input in terms of relative gas concentration, gave an encouraging diagnosis trend with an accuracy of 80% with ample scope to improve upon the accuracy with the elimination of errors. Some errors were introduced in the process of obtaining the gas concentrations due to approximations and using equipment of lesser precision. The accuracy of the setup could certainly be enhanced by using sensors of higher resolution and interfacing equipment lesser prone to errors. Other methods such as IEC-599, Roger's ratio method and the Doernenberg ratio method had diagnostic efficiencies lower than the Duval's triangle method. Their performance would also be augmented with the removal of inaccuracies.

This preliminary work does make pathways for a reliable online transformer incipient fault detection model using an actual gas-based sensor array as an E-Nose that could be deployed in the hostile transformer environment. The improvement in the incipient fault diagnosis of transformers can be explored by implementing the incipient fault interpretation methods using machine learning and deep neural network methodologies. The hydrocarbon gas sensors used in the E-Nose can be made more sensitive to a wide range of gas concentration ranging from sub-ppm levels to thousands of ppm. An ethane gas sensor having resolution for an appropriate gas range needs to be fabricated and incorporated into the E-Nose so that gas ratios of all interpretation methods can be obtained from a common setup. Furthermore, better precision is required in interfacing the E-Nose with the online fault identification system so as to obtain gas concentration values closer to the true value and more accurate incipient fault diagnosis can be performed.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

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

Author anonymous (2009). *IEEE guide for the interpretation of gases generated in oil-immersed transformers, IEEE std. C57.104-2008 (Revision of IEEE std. C57.104-1991)*. Piscataway: IEEE, 1–36.

Benmahamed, Y., Kherif, O., Teguar, M., Boubakeur, A., and Ghoneim, S. S. M. (2021). Accuracy improvement of transformer faults diagnostic based on DGA data using SVM-BA classifier. *Energies* 14, 2970. doi:10.3390/en14102970

Dhote, N. K., and Helonde, J. B. (2012). Diagnosis of power transformer faults based on five fuzzy ratio method. *WSEAS Trans. Power Syst.* 7 (3), 114–125.

Dukarm, J. J. J. (1993). "Transformer oil diagnosis using fuzzy logic and neural networks," in *Proceedings of the Canadian conference on electrical and computer engineering*, 14-17 September 1993, Vancouver: Canada.

Duval, M. (2002). A review of faults detectable by gas-in-oil analysis in transformers. *IEEE Electr. Insul. Mag.* 3, 154–196. doi:10.1109/MEI.2002.1014963

Duval, M., and DePablo, A. (2001). Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases. *IEEE Electr. Insul. Mag.* 17, 31–41. doi:10.1109/57.917529

Duval, M. (1989). Dissolved gas analysis: It can save your transformer. *IEEE Electr. Insul. Mag.* 5, 22–27. doi:10.1109/57.44605

Duval, M. (2003). New techniques for dissolved gas-in-oil analysis. *IEEE Electr. Insul. Mag.* 19, 06–15. doi:10.1109/mei.2003.1192031

Equbal, M. D., Khan, S. A., and Islam, T. (2018). Transformer incipient fault diagnosis on the basis of energy-weighted DGA using an artificial neural network. *Turk. J. Elec. Eng. Comp. Sci.* 26, 77–88. doi:10.3906/elk-1704-229

Gardner, J. W., and Bartlett, P. N. (1994). A brief history of electronic noses. *Sensors Actuators B Chem.* 18, 210–211. doi:10.1016/0925-4005(94)87085-3

Ghoneim, S. S. M., Farrag, T. A., Rashed, A. A., El-Kenawy, E. M., and Ibrahim, A. (2021). Adaptive dynamic meta-heuristics for feature selection and classification in diagnostic accuracy of transformer faults. *IEEE Access* 9, 78324–78340. doi:10.1109/access.2021.3083593

Gouda, O. E., El-Hoshy, S. H., and E. L.-Tamaly, H. H. (2018). Proposed three ratios technique for the interpretation of mineral oil transformers based dissolved gas analysis. *IET Gener. Transm. &amp. Distrib.* 12, 2650–2661. doi:10.1049/iet-gtd.2017.1927

Gouda, O. E., El-Hoshy, S. H., and Ghoneim, S. S. M. (2021). Enhancing the diagnostic accuracy of DGA techniques based on IEC-TC10 and related databases. *IEEE Access* 9, 118031–118041. doi:10.1109/access.2021.3107332

Guardado, J. L., Naredo, J. L., Moreno, P., and Fuerte, C. R. (2001). A comparative study of neural network efficiency in power transformers diagnosis using dissolved gas analysis. *IEEE Trans. Power Deliv.* 16, 643–647. doi:10.1109/61.956751

Hooshmand, R. A., Parastegari, M., and Forghani, Z. (2012). Adaptive neuro-fuzzy inference system approach for simultaneous diagnosis of the type and location of faults in power transformers. *IEEE Electr. Insul. Mag.* 28, 32–42. doi:10.1109/mei.2012.6268440

Hooshmand, R., and Banejad, M. (2008). Application of fuzzy logic in fault diagnosis in transformers using dissolved gas based on different standards. *World Acad. Sci. Engg. Techn* 2 (5), 1172–1176. doi:10.5281/zenodo.1076352

Huang, Y., and Sun, H. (2013). Dissolved gas analysis of mineral oil for power transformer fault diagnosis using fuzzy logic. *IEEE Trans. Dielectr. Electr. Insul.* 20, 974–981. doi:10.1109/tdei.2013.6518967

Interpretation of the analysis of gases in transformers and other oil- filled electrical equipment in service, *IEC Publ.* 599, 1978.

Khan, S. A., Equbal, M. D., and Islam, T. (2015). A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and anfis models. *IEEE Trans. Dielectr. Electr. Insul.* 22, 590–596. doi:10.1109/tdei.2014.004478

Khan, S. A., Equbal, M. D., and Islam, T. (2014). "ANFIS based identification and location of paper insulation faults of an oil immersed transformer," in *Proceedings of the IEEE power India international conference*.05-07 December 2014, New Delhi, India.

Khan, S. A., Khan, M. A., Wani, S. A., and Equbal, M. D. (2020). Performance enhancement and extension of DGA based transformer fault diagnosis methods using soft-computing techniques. Soft Computing Condition Monitoring and Diagnostics of Electrical and Mechanical Systems. *Adv. Intelligent Syst. Comput.* 1096, 287–324.

Kherif, O., Benmahamed, Y., Teguar, M., Boubakeur, A., and Ghoneim, S. S. M. (2021). Accuracy improvement of power transformer faults diagnostic using KNN classifier with decision tree principle. *IEEE Access* 9, 81693–81701. doi:10.1109/access.2021.3086135

Li, J., Zhang, Q., Wang, K., Wang, J., Zhou, T., and Zhang, Y. (2016). Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine. *IEEE Trans. Dielectr. Electr. Insul.* 23, 1198–1206. doi:10.1109/tdei.2015.005277

Lin, Y., Li, C., Wei, W., Li, Y., Wen, S., Sun, D., et al. (2015). A new type of acetylene gas sensor based on a hollow heterostructure. *RSC Adv.* 5, 61521–61527. doi:10.1039/c5ra10327d

Mang-Hui Wang, M.-H. (2003). A novel extension method for transformer fault diagnosis. *IEEE Trans. Power Deliv.* 18 (1), 164–169. doi:10.1109/tpwrd.2002.803838

Nezami, M. M., Equbal, M. D., Khan, S. A., and Sohail, S. (2021). An ANFIS based comprehensive correlation between diagnostic and destructive parameters of transformer's paper insulation. *Arab. J. Sci. Eng.* 46, 1541–1547. doi:10.1007/s13369-020-05180-4

Nezami, M. M., Equbal, M. D., Khan, S. A., Sohail, S., and Ghoneim, S. S. M. (2021). Classification of cellulosic insulation state based on smart life prediction approach (SLPA). *Processes* 9, 981. doi:10.3390/pr9060981

Noori, M., Effatnejad, R., and Hajihosseini, P. (2017). Using dissolved gas analysis results to detect and isolate the internal faults of power transformers by applying a fuzzy logic method. *IET Gener. Transm. &amp. Distrib.* 11, 2721–2729. doi:10.1049/iet-gtd.2017.0028

Rogers, R. R. (1978). IEEE and IEC codes to interpret incipient faults in transformers, using gas in oil analysis. *IEEE Trans. Elect. Insul.* EI-13, 349–354. doi:10.1109/tei.1978.298141

Singh, S., and Bandyopadhyay, M. N. (2010). Duval triangle: A noble technique for DGA in power transformers. *Int. J. Electr. Power Eng.* 4, 193–197. doi:10.3923/ijepe.2010.193.197

Singh, S., and Bandyopadhyay, M. N. (2010). Dissolved gas analysis technique for Incipient Fault diagnosis in power transformers: A bibliographic survey. *IEEE Electr. Insul. Mag.* 26, 41–46. doi:10.1109/mei.2010.5599978

Taha, I. B. M., Ibrahim, S., and Mansour, D. -E. A. (2021). Power transformer fault diagnosis based on DGA using a convolutional neural network with noise in measurements. *IEEE Access* 9, 111162–111170. doi:10.1109/access.2021.3102415

Vani, A., and Murthy, P. S. C. R. (2014). An adaptive neuro fuzzy inference system for fault detection in transformers by analyzing dissolved gases. *IOSR J. Electr. Electron. Engg* 9, 07–14. doi:10.9790/1676-09460714

Wani, S. A., Farooque, M. U., Khan, S. A., Gupta, D., and Khan, M. A. (2015). "Fault severity determination in transformers using dissolved gas analysis (DGA)," in *Proceedings of the 12th IEEE conference electronics, Energy, environment, communication*.New Delhi, India, 17-20 December 2015

Wilson, A. D., and Baietto, M. (2009). Applications and advances in electronic-nose technologies. *Sensors* 9, 5099–5148. doi:10.3390/s90705099

Wu, Y., Jiang, B., and Wang, Y. (2020). Incipient winding fault detection and diagnosis for squirrel-cage induction motors equipped on CRH trains. *ISA Trans.* 99, 488–495. doi:10.1016/j.isatra.2019.09.020

Zhang, G., Yasuoka, K., and Ishii, S. (1999). "Application of fuzzy equivalent matrix for fault diagnosis of oil-immersed insulation," in 13th International Conference on Dielectric Liquids (ICDL'99), Nara, Japan, July 25, 1999, 400–403.

Zhang, Y., Feng, J., and Wang, S. (2022). .A new method for transformer fault diagnosis by using improved clustering methodCSEE Journal of Power and Energy Systems. *Early Access*, 1–11. doi:10.17775/CSEEJPES.2021.03880

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 14 November 2022
DOI 10.3389/fenrg.2022.1032993

OPEN ACCESS

EDITED BY
Yongming Han,
Beijing University of Chemical Technology, China

REVIEWED BY
Yiming Ke,
Jinan University, China
Gejirifu De,
Technical University Dortmund, Germany
Yue Zhou,
Cardiff University, United Kingdom

*CORRESPONDENCE
Weishang Guo,
18800161023@163.com

SPECIALTY SECTION
This article was submitted to Process and Energy Systems Engineering, a section of the journal Frontiers in Energy Research

RECEIVED 31 August 2022
ACCEPTED 21 October 2022
PUBLISHED 14 November 2022

CITATION
Guo W, Mao Y, Chen Y and Zhang X (2022), Multi-objective optimization model of micro-grid access to 5G base station under the background of China's carbon peak shaving and carbon neutrality targets.
Front. Energy Res. 10:1032993.
doi: 10.3389/fenrg.2022.1032993

Multi-objective optimization model of micro-grid access to 5G base station under the background of China's carbon peak shaving and carbon neutrality targets

Weishang Guo[1,2,3]*, Yihua Mao[1], Yufen Chen[3] and Xuexing Zhang[3]

[1]College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, Zhejiang, China, [2]School of Civil Engineering and Architecture, Taizhou University, Taizhou, Zhejiang, China, [3]Fangyuan Construction Group Co.,Ltd., Taizhou, Zhejiang, China

Abstract: a large number of 5G base station are connected, which provides a new possibility for the future low-carbon development of power systems. By encouraging 5G base station to participate in demand response and incorporating it into the Microgrid, it can reduce the power consumption cost of 5G base stations and promote the efficient utilization of renewable energy. Based on the microgrid operation structure, 5G base station and multi-objective problem algorithm, a multi-objective optimization operation model of microgrid access to 5G base station is built. Considering the physical constraints of Microgrid, the objective is to minimize the operating cost and carbon emission. Through the joint dispatching of distributed clean energy generation, micro gas turbine, energy storage system and 5G base station in Microgrid, the comprehensive optimization of system economy and low-carbon benefits can be achieved. In this paper, a microgrid in Beijing is taken as the research object, and the Whale Optimization Algorithm algorithm is used to solve the multiobjective problem. The analysis results show that 5G base station can flexibly respond to microgrid scheduling, which helps microgrid to improve the consumption and utilization efficiency of renewable energy, thus bringing higher economic benefits and low-carbon benefits, and helping China to achieve the goal of carbon peak shaving and carbon neutrality at an early date.

KEYWORDS

microgrid, multi objective optimization, 5G base station, carbon peak shaving and carbon neutrality, whale optimization algorithm

1 Introduction

1.1 Literature review

China's "14 Five-Year Plan" for electric power development has been made clear. China will no longer build new coal-fired power generation projects, and arrange a certain scale to ensure the electric power supply according to demand, so as to vigorously promote the consumption of clean energy such as wind energy and photovoltaic energy. By 2030, the installed capacity of wind power and photovoltaic in China is expected to reach over 1.8 billion GW, accounting for about 25% of the total power generation. In 2060, the installed capacity of wind power and photovoltaic power generation is expected to reach over 6 billion GW, accounting for about 75% of the total power generation. With the increasing popularity of new energy, the peak-valley difference of net load fluctuation after the superposition of new energy output and load demand will increase year by year.

China's microgrid construction is still in the exploratory stage. With the deepening of energy consumption revolution and power system reform, and the continuous maturity of grid-connected technology of distributed generation, the interactive operation with external large power grid will remain a research focus. Microgrid is usually integrated by distributed generation, energy storage system, demand side and energy management system, and becomes a small power generation and distribution system which can not only be used as backup grid connected operation, but also independent of external power grid. At the executive meeting of China's State Council in April 2020, seven major new infrastructure projects were established. They include 5G base station construction, UHV, intercity high-speed railway and urban rail transit, charging piles for new energy vehicle, big data center, artificial intelligence and industrial internet. Therefore, the construction of 5G base station connected to microgrid is one of the inevitable directions of new power system construction in China. Espina et al. (2020) summarized the development of distributed cooperative control system, they thought applied the cooperative algorithm of the system, including the minimum cost operation. Agalgaonkar and Robinson (2021) discussed relationship between communication system and power system of Microgrid, and the latest development of Microgrid is described. Hirsch et al. (2018) considered that the application of Microgrid has entered into commercial application from experiment, which improves the reliability of power supply, manages distributed energy. Chen et al. (2020a) considered communication and functional integration comprehensively to realize the optimal and sustainable operation of Microgrid. Parhizi et al. (2015) discussed the different design schemes in islanding state and connected with large power grid, and reviewed the technologies of distributed generation, protection and communication. Ke et al. (2022a); Ke et al. (2022b) studied the investment scheme selection of urban comprehensive energy system and photovoltaic poverty alleviation project, and analyzed the economic and environmental benefits of these two new power systems from the perspectives of economic cost and energy utilization.

Distributed generation technology is undoubtedly an important generation side component of microgrid. System planning and energy policy making are of great significance for the coordinated control of distributed generation. Lopes et al. (2007); Gupta et al. (2020); Zhao et al. (2020) improved the power quality of wind energy and photovoltaic in distribution network and reduced the harmonics through the adaptive signal extraction method and the design of T-shaped filter. Zhang et al. (2011) simulated a new grid connected inverter for distributed wind power generation in MATLAB. Liu et al. (2006) discussed the safe operation state of grid connected control. Lee et al. (2021); Yuvaraj et al. (2021) used different algorithms to find the optimal capacity and location of distributed generation. Naval and Yusta (2021) optimized the hourly scheduling of pumping equipment proves that distributed photovoltaic warfare reduces the energy cost of pumping station. Lotfi (2021) proved that using demand response to change user habits can achieve optimization effect when the joint operation of distributed generation and reactive power compensation capacitor combined. Katiraei and Iravani (2006) mainly studied the active and reactive power management strategy of distributed generation unit. Alabdali and Nahhas (2021) discussed the scope and development prospect of photovoltaic grid connection in Arab region. Onojo et al. (2021) compared the differences between two different systems including distributed generation unit and diesel generating unit.

The optimization of microgrid has always been the focus of attention of experts and scholars, which can be roughly divided into two aspects: one is the optimization of scheduling operation, and the other is the optimization of capacity allocation. In terms of scheduling, (Weishang et al., 2022) used the PSO algorithm to construct and solve the optimal scheduling of the gas heat virtual power plant. Marzband et al. (2018a) considered a large number of distributed energy suppliers in the energy market firstly. A heuristic algorithm is adopted to maximize the economic benefits of all subjects. Marzband et al. (2018b) put forward the main body of demand response. The market mechanism is used to change the consumption mode of consumers, so as to improve the operation optimization mechanism of power market. Tavakoli et al. (2018) optimized the cost strategy of Microgrid for commercial buildings by using linear programming model after quantifying the cost of distributed generation in bad weather by conditional value at risk theory. Zia et al. (2018) compared several communication technologies and analyzed the optimal cost management strategy. Ali et al. (2021) used the bi level programming model based on meta heuristic to solve the randomness of uncertain generation and electric vehicle random charge and discharge. Junejo et al. (2021) used the improved

MILP method to optimize the Microgrid access scheme in rural areas can reduce the total net present value and carbon emissions. Tan et al. (2020) established a model based on multi task learning and least squares support to predict the electric, heating, cooling and gas loads in the integrated energy system, which is helpful to improve the operation efficiency of the system. Zhou et al. (2021) used the robust model to ensure the high robustness and stability of wind power inverter in Microgrid, which can not only save energy but also ensure the fast convergence. Zandrazavi et al. (2022) proposed a stochastic multi-objective optimization model for Microgrid to minimize the total operating cost and voltage deviation. by ε-Constraint method and fuzzy satisfaction method. Chen et al. (2020b) proposed an unconscious islanding transition control strategy of three-phase/single-phase MMG based on artificial emotion reinforcement learning to reduce the time required for voltage recovery and improve the power supply reliability of Microgrid. Aa et al. (2021) combined Particle swarm optimization and bat algorithm to optimize the location and capacity of distributed solar power generation in Microgrid. Guo et al. (2021) used Quantum behaved particle swarm optimization (QPSO) to solve the capacity allocation optimization problem of isolated Microgrid and realized power supply reliability. Cheng et al. (2021) used the NSGA-II algorithm to solve the optimization model of grid cost-benefit composition, so as to ensure the safety of power consumption of important users in Microgrid (Wang et al., 2021) used the genetic particle swarm optimization algorithm to solve the capacity allocation problem of photovoltaic, wind power and energy storage equipment in Microgrid.

In recent years, with the rapid popularization and promotion of 5G communication technology in China, 5G base station has gradually developed into an important emerging load. According to the statistics of the Ministry of industry and information technology of China, by the end of 2021, 1.425, million 5G base stations have been built and opened in China. The total number of 5G base stations accounts for more than 60% of the world. The 5G network has covered all prefecture level cities, More than 98% of the counties and cities and 80% of the towns and townships. By the end of 2020, the number of base stations per person has increased by nearly 10.1 million. In the operation process, through scientific means to dispatch and manage the power supply and power consumption equipment in 5G base station, the interactive response potential of 5G base station can be brought into full play, and considerable flexibility support can be provided for the power grid, At the same time, it can improve the clean energy consumption and carbon emission reduction. The massive access of 5G base stations provides a new possibility for the development of low-carbon power system in the future, (ishchenko et al., 2022) discussed the MIMO antenna array in the fifth generation millimeter wave UWB data transmission network. Chandran and Abdulla (2021) compared the infrastructure update of 5G communication technology and 4G communication technology in the Internet of things. Hao et al. (2020); Wen et al. (2021) discussed the nano materials, antennas and other equipment needed for 5G communication technology. Yong et al. (2021) introduced the basic equipment and power consumption characteristics of 5G base station, and studied its demand response potential. Asia et al. (2022) introduced the changes of demand growth of copper clad laminate for 5G base stations in Europe and the United States and the changes of 5G base stations. Weissberger (2020) introduced the development of China's 5G communication technology, the technical impact of the US.-China trade war and the export control promulgated by the US. government on the development of 5G communication.

1.2 Innovation and contribution

According to the comparison of literature research, the innovations and contributions of this paper are as follows.

1) Give full play to the low-carbon energy potential of 5G base station in microgrid operation, deeply consider its operation characteristics, and establish an optimized operation model of 5G base station operation adjustable characteristics under the coupling of multiple energies of microgrid connected to external network.
2) Most of the existing researches adopt system cost minimization or system operating income maximization. After deeply analyzing the role of microgrid operating framework and 5G base station in carbon emission reduction of power system, based on the two dimensions of economic benefit and environmental benefit, this paper comprehensively considers the physical constraints of microgrid units, and constructs the double objectives of microgrid operating cost and carbon emission minimization under the access of 5G base station.
3) 5G base station participating in microgrid dispatching can effectively optimize the system load curve, and the fluctuation of system load is obviously optimized.

2 Multi objective optimization model of microgrid access to 5G base station

2.1 Basic structure of microgrid access to 5G base station

The basic structure of Microgrid can be divided into four parts: generation side, user side, energy storage system and energy management system in Microgrid. After 5G base station is connected, the specific composition of Microgrid system is shown in Figure 1.

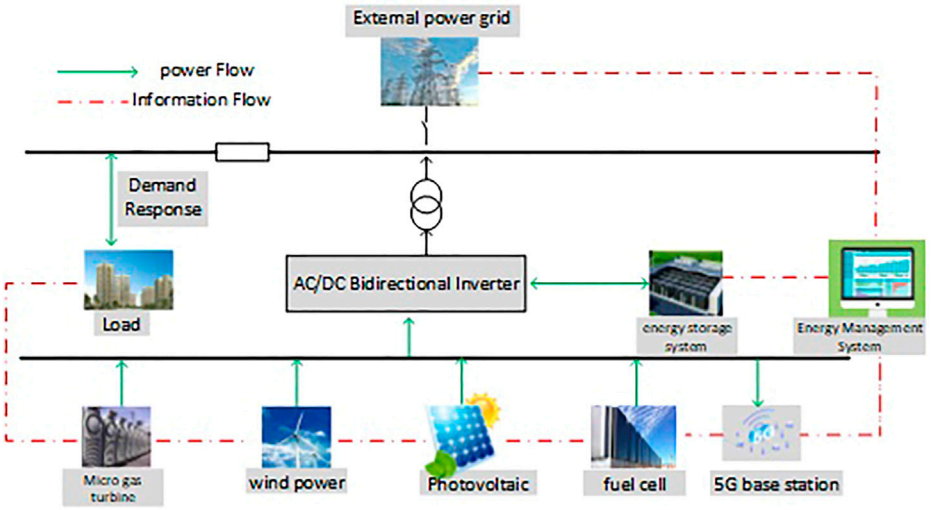


FIGURE 1
Schematic diagram of Microgrid system structure.

As can be seen from Figure 1, the power generation side of the system mainly includes controllable power sources, such as micro turbine (MT) and fuel cell (FC), and distributed power generation, such as wind turbine (WT) and photovoltaics (PV), which can be connected to 5G base station, thus further embodying the characteristics of reducing the abandonment rate of wind and photovoltaics. In this paper, the distributed energy consumption maximization principle is adopted, that is, wind power and photovoltaic power are preferentially absorbed, and controllable power supply mainly plays a role of adjustment and supplement. Meanwhile, the microgrid designed in this paper is connected to the power grid through the point of common coupling (PCC).

2.2 Simplified mathematical model of microgrid

2.2.1 Simplified photovoltaic output model

Assuming that the photovoltaic output is only related to photovoltaic intensity and ambient temperature, the power generation of photovoltaic array can be expressed by the following formula:

$$P_{PV} = f_{PV} P_{PV,rate} \left(\frac{I_T}{I_S} \right) \left[1 + \alpha_P \left(T_{cell} - T_{cell,S} \right) \right] \tag{1}$$

P_{PV} is the actual output power of photovoltaic; f_{PV} is the derating factor of photovoltaic array, that is, the ratio of actual power generation power to rated power generation power of photovoltaic array, which is used to display the loss of photovoltaic panel itself, usually taken as 0.9. $P_{PV,rate}$ is the rated capacity of photovoltaic array, kW; I_T and I_S are the actual photovoltaic intensity and the photovoltaic intensity under the standard test environment, kW/M^2; α_P is power temperature coefficient,%/°C; T_{cell} and $T_{cell,S}$ are the actual temperature of the PV panel and the temperature under the standard test conditions respectively.

2.2.2 Wind turbine output model

When the wind speed v is lower than v_{in} (cut-in wind speed), the output of wind turbine is 0. When the wind speed v is higher than the cut in wind speed v_{in}, wth the increase of wind speed, the output power increases continuously. When the wind speed reaches the rated wind speed v_r, the fan maintains a constant output power P_r. When the wind speed exceeds the cut-out wind speed v_{out}, the fan stops running. Therefore, the power generation model of wind turbine can be approximately expressed by the following formula:

$$P_{WT} = \begin{cases} 0 & v \le v_{in} \\ \dfrac{v^3 - v_{in}^3}{v_r^3 - v_{in}^3} P_{WT,rate} & v_{in} \le v \le v_r \\ P_r & v_r \le v \le v_{out} \\ 0 & v \ge v_{out} \end{cases} \tag{2}$$

P_{WT} is the actual output of wind turbine; $P_{WT,rate}$ is the rated power of the fan.

2.2.3 Power generation model of micro gas turbine

Micro gas turbine is a kind of ultra-small gas turbine. Usually, methane and natural gas are used as fuel with a power of less than 1 million kilowatts. Micro-gas turbine can generate electricity through various fuels, which is not only environmentally friendly, efficient, simple in operation and maintenance, and mature in technology, but also has better controllability compared with new energy generation, and can also be used for rotary standby, so it is widely used in the field of distributed power generation. The output power of the micro gas turbine can be increased or decreased by controlling the fuel supply according to the load demand, but it has the upper limit and lower limits, and the change speed, that is, climbing speed, is limited by this range.

Therefore, the power generation model of micro gas turbine is as follows:

$$P_{MT} = q^{*}\eta_{MT}{}^{*}m_f \tag{3}$$

P_{MT} is the output power of micro gas turbine; q is the calorific value of the fuel; η_{MT} is the power generation efficiency of micro gas turbine; m_f is the flow rate of the fuel.

Furthermore, the power generation cost (fuel cost) of micro gas turbine can be calculated by the following formula:

$$C_{MT.fuel} = c^{*}\frac{1}{LHV}{}^{*}\sum\frac{P_{MT}}{\eta_{MT}} \tag{4}$$

Among them, $C_{MT.fuel}$ is the power generation cost of micro gas turbine, c is Unit price of natural gas, LHV is the low calorific value of natural gas.

Distributed micro sources such as micro gas turbine and fuel cell will generate emission cost during power generation. The calculation formula is as follows:

$$C_{MT.en} = \sum_{i=1}^{N}\alpha_i{}^{*}EF_{MT.i}{}^{*}P_{MT} \tag{5}$$

$C_{MT.en}$ is the environmental cost of micro gas turbine; N is the type of pollutant discharged; α_i is the conversion coefficient of pollutants, yuan/kg; $EF_{MT.i}$ is the micro gas turbine i Unit emission of pollutants, kg/kW.

2.2.4 Fuel cell power generation model

The fuel cell converts the chemical energy in fuel (usually hydrogen or natural gas) and oxidant into electric energy through the positive and negative electrodes and electrolyte materials in the device, mainly emits CO_2 and water, and the power generation efficiency is between 50% and 70%. The fuel cell converts the chemical energy in fuel (usually hydrogen or natural gas) and oxidant into electric energy through the positive and negative electrodes and electrolyte materials in the device, mainly emits CO_2 and water, and the power generation efficiency is between 50% and 70%.

Similar to the power generation cost of micro gas turbine, the calculation formula of power generation cost under normal operation of fuel cell is as follows:

$$C_{FC.fuel} = c^{*}\frac{1}{LHV}{}^{*}\sum\frac{P_{FC}}{\eta_{FC}} \tag{6}$$

$C_{FC.fuel}$ is the power generation cost of fuel cell.

Similar to micro gas turbine, the environmental cost of fuel cell is calculated by the following formula.

$$C_{FC.en} = \sum_{i=1}^{N}\alpha_i{}^{*}EF_{FC.i}{}^{*}P_{FC} \tag{7}$$

$C_{FC.en}$ is the environmental cost of a fuel cell; $EF_{MT.i}$ is that the fuel cell i Unit emission of pollutants, kg/kW.

2.2.5 Operation model of energy storage system

The remaining charge of the battery at that time and its state of charge. In this paper, the state of charge (SOC) is used to reflect the remaining capacity of the battery, which is expressed by the ratio of the remaining capacity of the battery to the total capacity.

When the battery is charged:

$$\mathrm{SOC}_{ESS}^{t+1} = \mathrm{SOC}_{ESS}^{t} + \frac{\eta_{ch}P_{ESS,ch}^{t}}{E_{ESS}} \tag{8}$$

When the battery is discharged:

$$\mathrm{SOC}_{ESS}^{t+1} = \mathrm{SOC}_{ESS}^{t} - \frac{P_{ESS,dis}^{t}}{\eta_{dis}E_{ESS}} \tag{9}$$

SOC_{ESS}^{t} and SOC_{ESS}^{t+1} are the state of charge of the battery during the period t and $t+1$, η_{ch} and η_{dis} are the charging and discharging efficiency of the battery, $P_{ESS,ch}^{t}$ and $P_{ESS,dis}^{t}$ are the charging and discharging power of the battery during the period t, E_{ESS} is the rated capacity of the battery.

2.2.6 5G base station energy consumption model

The energy consumption of 5G base station includes static power consumption and dynamic power consumption. Static power consumption refers to the energy demand independent of service load and output transmission power, which is mainly composed of fixed losses of power supply system, BBU baseband unit signal processing and cooling system Dynamic power consumption refers to the energy requirement related to 5G service load. It is a function of 5G base station output transmission power. To sum up, the total power consumption of a single 5G base station can be expressed as follows.

$$\begin{gathered}P_{i,t}^{BS} = \sum_{j=1}^{N_{i,t}^{TR}}P_{i,j,t}^{TR} + P_{i,t}^{A} \\ i \in \Omega^{BS}, j \in N_{i,t}^{TR}, t \in \Omega^{T}\end{gathered} \tag{10}$$

$P_{i,t}^{BS}$ is total power consumption in time period t of base station i. $N_{i,t}^{TR}$ is the number of transceivers, which is

related to the base station carrier configuration; $P_{i,j,t}^{TR}$ is the Internal power consumption of Transceiver j in Base stations i in time period t; $P_{i,t}^{A}$ is power consumption of cooling system. Ω^{BS} is the set of 5G base stations; Ω^{T} is the time period set. According to reference (Hao et al., 2020) the power consumption of a single transceiver can be calculated as follows:

$$P_{i,j,t}^{TR} = \delta_{i,j,t}\left(P_{i,t}^{0} + \Delta\beta_i P_{i,j,t}^{\text{out}}\right) + \left(1 - \delta_{i,j,t}\right)P_i^{\text{sleep}} \tag{11}$$

$\delta_{i,j,t}$ is a 0–1 variable representing the running state of the transceiver, which is 1 when it is in the working state and 0 when it is in dormant state; $P_{i,t}^{0}$ is no-load power consumption; $\Delta\beta_i$ is the slope related to the dynamic power consumption of a single transceiver; P_i^{sleep} is the power consumption of a single transceiver in sleep state; $P_{i,j,t}^{\text{out}}$ is the output transmission power of base station transceiver, which is a function of signaling power and user data power.

$$P_{i,j,t}^{\text{out}} = \left(1 - \beta^{OH}\right)\frac{P_i{}^{\max}}{r_i^{bd}} r_{i,j,t}^{tr} + \beta^{OH} P_i{}^{\max} k \tag{12}$$

β^{OH} is the proportion of fixed signaling signal in the transmission power; $P_i{}^{\max}$ is the maximum transmission power of single transceiver of each base station; r_i^{bd} is the maximum bandwidth utilization of a single transceiver of each base station, $r_{i,j,t}^{tr}$ is the number of bandwidth used for data transmission; k is a weighting factor, which is the level of signaling power transmitted by the base station under different operating conditions. Its value is related to the operation status of transceiver and the number of data transmission bandwidth in the current period.

During the operation scheduling period, the relationship between the energy storage capacity of the energy storage equipment in 5G base station in the current period, the state of the former sequence electricity and the charging and discharging power is shown as follows:

$$\begin{gathered} E_{i,t} = E_{i,t-1} + \eta^{\text{CH}} P_{i,t}^{\text{ch}} \Delta t - P_{i,t}^{\text{dis}} \Delta t / \eta^{\text{DIS}} \\ i \in \Omega^{\text{BS}}, t \in \Omega^{\text{T}} \end{gathered} \tag{13}$$

$E_{i,t}$ is the energy storage capacity of energy storage battery of node i base station in period T; $\eta^{\text{CH}}, \eta^{\text{DIS}}$ Charging and discharging efficiency of energy storage device.

2.3 Planning objectives

2.3.1 Mathematical model of objective function

2.3.1.1 System operation cost minimization

Total cost of Microgrid generation and operation C_g involves the fuel cost, sewage cost, daily operation and maintenance cost, power purchase cost from external power grid and energy storage loss cost of 5G base station.

$$\begin{aligned} \min C_g = \sum_{t=1}^{T}\Bigg[\sum_{j=1}^{N}\left(C_{j,fuel}\left(P_j^t\right) + C_{j,op}\cdot P_j^t + C_{j,en}\cdot P_j^t\right) + \sum_{m=1}^{M} C_m P_m^t \\ + p_{grid}^t P_{grid}^t\Bigg] \sum_{t\in\Omega^{\text{T}}}\sum_{i\in\Omega^{\text{BS}}}\left[\sigma^{\text{ch}}\left(P_{i,t}^{\text{ch}}\right)^2 + \sigma^{\text{dis}}\left(P_{i,t}^{\text{dis}}\right)^2\right] \end{aligned} \tag{14}$$

N is the number of types of power generation equipment in microgrid, including micro gas turbine, fuel cell, photovoltaic generator set and wind turbine set. $C_{j,fuel}(P_j^t)$ is the fuel cost of various power generation equipment, the specific expression and the value of relevant parameters are given in reference (He et al., 2012); $C_{j,op}, C_{j,en}$ are the environmental cost and fuel cost of distributed photovoltaic power generation and wind power generation, which are not considered in this paper; P_j^t is the actual power of the period t. σ^{ch}、 σ^{dis} are the unit depreciation cost corresponding to the charging and discharging of the energy storage battery in 5G base station. The energy storage loss cost is the quadratic function of the charging power $P_{i,t}^{\text{ch}}$ and discharge power $P_{i,t}^{\text{dis}}$ (He et al., 2012).

2.3.1.2 System carbon emission minimization

In order to comprehensively reflect the benefits of carbon emission reduction brought by 5G base station participating in demand response for microgrid, this paper takes the minimum carbon emission of the system as another objective of optimizing the operation model. Among them, the carbon emission of microgrid is directly related to the fossil fuel consumption of micro gas turbine. As the fossil fuel consumption of micro gas turbine is very small compared with the external power grid, it can be neglected. Therefore, the carbon emission of the system is equal to the electricity purchased from the external power grid multiplied by the coal consumption coefficient of unit power generation and then multiplied by the carbon emission coefficient corresponding to unit coal consumption power generation, as follows:

$$min\, C^{\text{CA}} = \varepsilon f \sum\nolimits_{t\in\Omega^{\text{T}}} P_t^{\text{grid}} \Delta t \tag{15}$$

ε is the carbon emission coefficient corresponding to unit coal consumption power generation; f is coal consumption coefficient corresponding to unit power generation in external power grid.

2.3.1.3 After weighting these two objective functions, the final multi-objective optimization function can be obtained as follows:

$$\text{minF} = \lambda_1 C_{\text{g}} + \lambda_2 C^{\text{CA}} \tag{16}$$

λ_1 and λ_2 respectively represent the weights corresponding to each sub-objective function. $\lambda_1 + \lambda_2 = 1$. The values of λ_1 and λ_2 can be selected subjectively according to the actual situation of

scheduling strategy. In this paper, the values of λ_1 and λ_2 are both set to 0.5.

2.3.2 Physical constraints of microgrid

2.3.2.1 System power balance constraints

$$P_{PV}^{t} + P_{WT}^{t} + P_{MT}^{t} + P_{FC}^{t} + P_{ESS}^{t} + P_{grid}^{t} = P_{af}^{t} + P_{loss}^{t} \tag{17}$$

P_{PV}^{t} is the photovoltaic power of the period t, P_{WT}^{t} is wind power generation of period t, P_{loss}^{t} is operation loss of the period t.

2.3.2.2 Energy storage system constraints

1) Power constraint of energy storage system

$$P_{ESS}^{\min} \le P_{ESS}^{t} \le P_{ESS}^{\max} \tag{18}$$

$P_{ESS}^{\max}$ and $P_{ESS}^{\min}$ are the upper and lower limits of the operating power of the energy storage system, and the lower limit is negative, indicating the state of charge.

2) State of charge constraints of energy storage system

$$\mathrm{SOC}_{ESS,\min} \le \mathrm{SOC}_{ESS}^{t} \le \mathrm{SOC}_{ESS,\max} \tag{19}$$

SOC_{ESS}^{t} is the state of charge of the energy storage system. $\mathrm{SOC}_{ESS,\max}$ and $\mathrm{SOC}_{ESS,\min}$ are the upper and lower limits of the state of charge of the energy storage system.

If the optimal scheduling period of Microgrid is continuous, then the state of charge of the energy storage system should be consistent at the beginning and the end of each scheduling period.

$$\mathrm{SOC}_{ESS}^{0} = \mathrm{SOC}_{ESS}^{T} \tag{20}$$

2.3.2.3 Internal load balancing constraints

$$P_{be}^{t} = P_{un}^{t} + P_{m}^{t} \tag{21}$$

$$\sum_{t=1}^{T} P_{be}^{t} = \sum_{t=1}^{T} P_{af}^{t} + \sum_{t=1}^{T} P_{cut}^{t} \tag{22}$$

P_{un}^{t} is the uncontrollable load of the period t.

2.3.2.4 Controllable load capacity constraint

$$0 \le P_{m}^{t} \le P_{m,\max} \tag{23}$$

$P_{m,\max}$ is the maximum capacity of controllable load.

2.3.3 Constraints on generation side of microgrid

The constraints on the generation side of Microgrid mainly include climbing power constraint, maximum minimum output constraint and power exchange constraint between Microgrid and external power grid.

2.3.3.1 Interactive power constraints between power grids

$$P_{grid,\min} \le P_{grid}^{t} \le P_{grid,\max} \tag{24}$$

$P_{grid,\max}$ and $P_{grid,\min}$ are the upper and lower limits of power for interaction between Microgrid and external power grid. The lower limit is negative, and the upper limit is positive. The absolute values of the two are the same.

2.3.3.2 Unit output constraints

$$P_{j,\min} \le P_{j}^{t} \le P_{j,\max} \tag{25}$$

$P_{j,\max}$ and $P_{j,\min}$ are the upper and lower output limits of the generator sets j.

2.3.3.3 Unit climbing constraints

$$\Delta_{j}^{-} \le P_{j}^{t} - P_{j}^{t-1} \le \Delta_{j}^{+} \tag{26}$$

Δ_{j}^{+}、 Δ_{j}^{-} are the upper and lower ramp rate limits of the distributed units j.

2.3.4 5G base station constraints

During the operation, the maximum number of transceivers in 5G base station should meet the following constraints due to the limitation of equipment configuration capacity:

$$0 \le N_{i,t}^{TR} \le N_{i,t}^{\max} \tag{27}$$

$N_{i,t}^{\max}$ is the number of transceivers configured for the base station.

In addition, since 5G operators need to meet the requirements of transmitting signals within the allocated bandwidth and meet the transmission equipment standards, the data transmission bandwidth of transceiver in 5G base station should also meet the following constraints:

$$0 \le r_{i,j,t}^{tr} \le r_{i}^{bd} \tag{28}$$

3 Solving algorithm

3.1 Whale optimization algorithm

The whale optimization algorithm is based on the behavior of whales rounding up their prey. Whales are social mammals, and

they will cooperate with each other to drive and round up their prey when hunting.

There are three ways to find the best prey, that is, to find the best prey by soaking net and to find whale prey by soaking net.

3.1.1 Surround the prey

During hunting, whales can find out the coordinates of their prey and surround it. Before solving the problem, the position of prey in the solution space is unknown to whales. Therefore, in the WOA algorithm, the position of the best whale in the current population is assumed to be the position of the prey, and other whales in the population are surrounded by the best whale. The mathematical model is described as follows:

$$D = \left|C \cdot X^{best}(t) - X\right| \tag{29}$$

$$X(t+1) = X^{best}(t) - A \cdot D \tag{30}$$

t is the current number of iterations, X is the current position of the whale, X^{best} is the position of the best whale in the current whale group. A and C are coefficient vectors. The mathematical model is described as follows:

$$A = 2a \cdot r - a \tag{31}$$

$$C = 2 \cdot r \tag{32}$$

a is a convergence factor, with the predation iteration of the whale group, the convergence factor is determined. a is the value decreases linearly from 2 to 0, r is random numbers between $[0, 1]$.

3.1.2 Bubble net attack mode

We choose the helix update position mechanism to calculate the distance (X, Y) and (X^{best}, Y^{best}). A spiral equation is used to simulate the spiral motion related to humpback whales. The mathematical model is described as follows:

$$X(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \tag{33}$$

$$D' = \left|X^*(t) - X(t)\right| \tag{34}$$

D' is the distance between the optimal whale individual and the current whale individual in the next iteration, b is the constant of the logarithmic spiral equation, l is a random number between $[0, 1]$.

Humpback whales swim in a short circle around their prey along a spiral path. In order to get this behavior optimized in the process_50% of the probability is between the encirclement mechanism and the spiral update position mechanism to update the position of the individual whale. The mathematical model is described as follows.

$$X(t+1) = \begin{cases} X^*(t) - A \cdot D & p < 0.5 \\ D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) & p \geq 0.5 \end{cases} \tag{35}$$

p is a random number between $[0, 1]$. After the bubble net attack, humpback whales begin to search for prey at random.

3.1.3 Looking for prey

In the process of hunting prey, we can use the method that changes with the iteration process vector a. In fact, humpback whales will randomly explore the space according to each other's position. Therefore, when the random value of A is greater than 1 or less than 1, it is used to update the whale's position in the global exploration stage to stay away from the current individual. Compared with the local development stage, the global exploration stage. The position of a single whales is upgraded on the basis of randomly selected individuals, instead of the best individual found so far. It is assumed that in the optimization process, there is a 50% probability of randomly selecting the location of the updated individual whale between the circling mechanism and the spiral updating location mechanism. The mathematical model is described below.

$$D = |C \cdot X_{rand} - X| \tag{36}$$

$$X(t+1) = X_{rand} - A \cdot D \tag{37}$$

X_{rand} is the position of randomly selected individual whales in the current whale population.

3.2 Solution steps

1) Input variables of parameter setting of microgrid power supply including wind turbine capacity, wind turbine cut in wind speed and other initial parameter into the algorithms;
2) Set the whale number n and the maximum iteration number tmax of the algorithm, and initialize the position information;
3) Calculate the fitness of each whale, find the current optimal whale position and keep it;
4) Calculate parameters a, p and coefficient vectors A, C. Judge whether the probability p is less than 0.5, if so, go directly to step (5), otherwise, adopt the bubble net predation mechanism: update the position by formula (38 formula -formula 39);
5) Judge whether the absolute value of coefficient vector A is less than 1, if so, surround the prey: update the position according to formula (35); Otherwise, globally randomly search for prey: update the position according to formula (41 formula -formula 42);
6) After the position update, calculate the fitness of each whale, and compare it with the previously reserved optimal whale position. If it is better than that, replace it with a new optimal solution. Judging whether the current calculation reaches the maximum number of iterations, if so, obtaining the optimal solution, and finishing the calculation, otherwise, entering the next iteration, and returning to step (4);
7) Output the final calculation result.

GAMS-BARON solver is used to solve the mixed integer nonlinear programming model of optimal scheduling of micro-

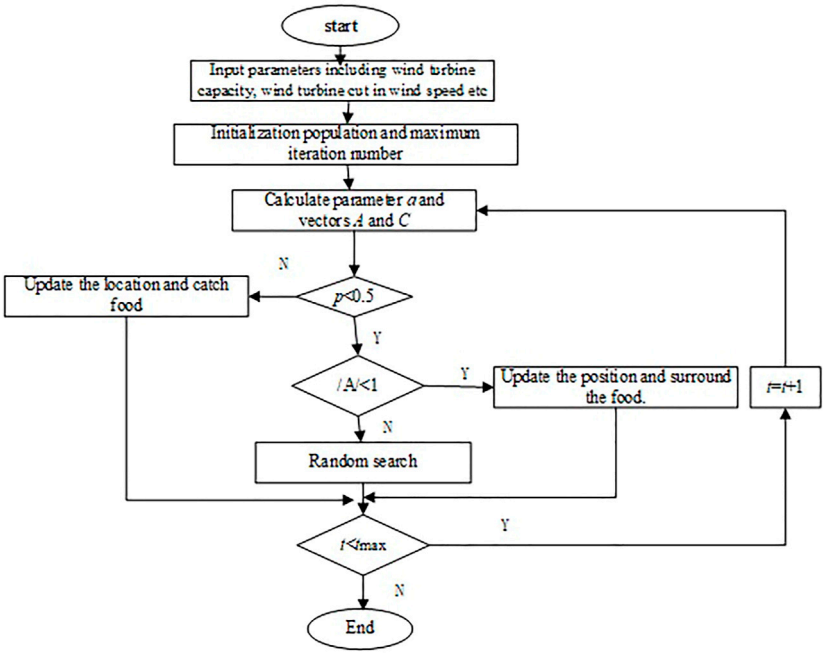


FIGURE 2
Solution flow.

TABLE 1 Parameter setting of Microgrid power supply (Yu and Zheng, 2020).

Wind turbine capacity	20 kW
Wind turbine cut in wind speed v_{in}	4.2 m/s
Rated wind speed of wind turbine v_r	15.5 m/s
Wind turbine cut-out wind speed v_{out}	20 m/s
Operation and maintenance cost of wind turbine	0.030 yuan/kW
Photovoltaic capacity	25 kW
Operation and maintenance cost of photovoltaic unit	0.012 yuan/kW
Battery charge discharge efficiency	94%
Battery capacity	180 kWh
Battery storage efficiency	[10%,100%]
Maximum SOC	0.9
Maximum SOC	0.2
Initial SOC	0.3
Battery operation and maintenance cost	0.028 yuan/kW
Power range of gas turbine	[15,100]/kW
Operation and maintenance cost of gas turbine	0.041/kW
Gas turbine ramp rate	15 kW/h
Power range of external power grid	[-50,100]/kW

energy grid operation configuration considering the characteristics of electric load. The specific process is shown in Figure 2.

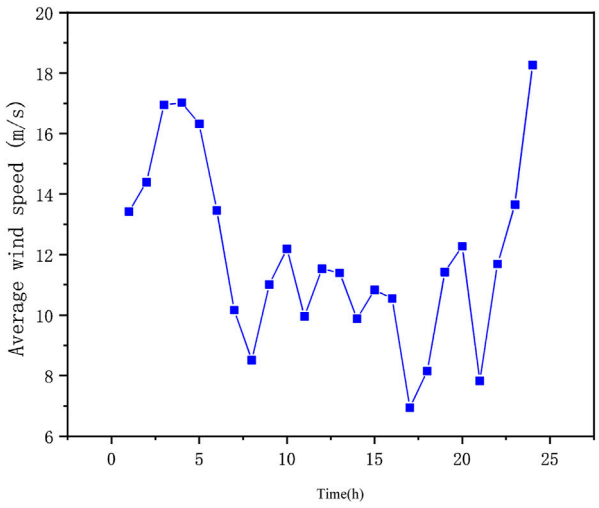


FIGURE 3
Average wind speed forecast in the next 24 h.

4 Simulation example

In this paper, a microgrid in Beijing is selected as the study object. The dispatching cycle of the Microgrid is 1 day, which is divided into 24 dispatching periods, that is, the

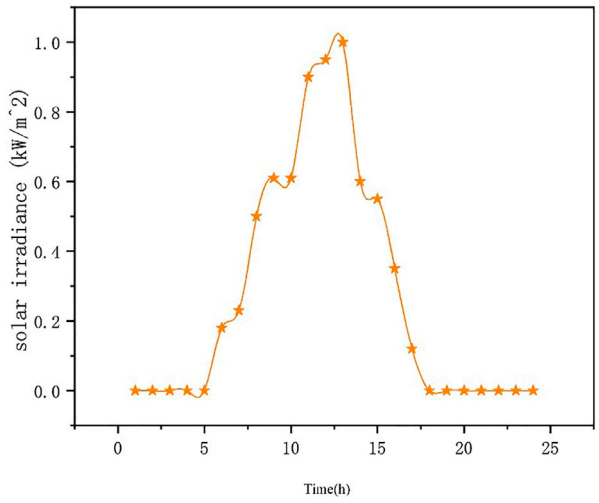


FIGURE 4
Forecast of solar irradiance in the next 24 h.

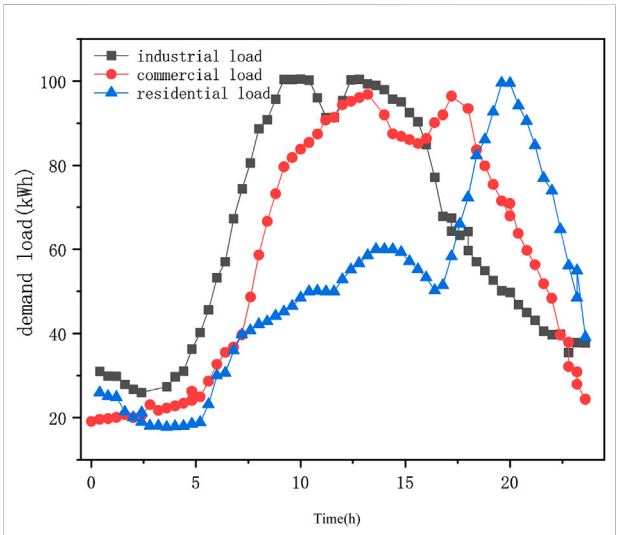


FIGURE 5
Daily load curve of different users.

dispatching interval is 1 h. Refer to the literature (Wen et al., 2021) for the setting of parameters such as the emission coefficient and converted cost of the micro-combustion turbine and fuel cell.

4.1 Basic data

In the selected Microgrid, the parameter settings of various power sources in the Microgrid are shown in Table 1.

See Figures 3, 4 for the forecast of average wind speed and solar irradiance in the next 24 h.

Based on the mathematical relationship between Figures 3, 4 and Formula (1-2), calculate the power of photovoltaic and wind turbines. When calculating the output power of WT and PV, it is assumed that the power factor of these power sources is 1, and no reactive power is generated.

The typical daily load curves of high industrial load, total commercial load and smart home load can be predicted according to historical data, as shown in Figure 5.

Time-of-use electricity prices of external power grid are shown in Table 2.

4.2 Simulation

In order to compare the absorption and efficient utilization of renewable energy in microgrid system by 5G base station, and consider whether to access 5G base station or not, the model established in this paper is configured with two operating modes, as follows (Table 3).

In order to reveal the economic and environmental benefits of 5G base station participating in microgrid, this section makes a comparative analysis of the scheduling schemes of 5G base station participating in microgrid and not participating in it. Scenario 1 without 5G access means that 5G base station do not participate in grid interaction when dispatching power in microgrid, and only manage the power consumption of electrical equipment from the perspective of their own energy consumption. WOA algorithm is used to determine the best solution set.

4.3 Analysis

The wind curtailment power of the system of the two scenarios is shown in Figure 6.

As can be seen from Figure 6, the flexible interaction of 5G base stations significantly reduces wind power, and the amount of wind power connected to the grid greatly increases during peak wind power generation. The main reasons are as follows. In the peak period of power load, on the premise of meeting the communication service quality, the base station controls the number of transceivers, which significantly reduces the power consumption of 5G base stations. In addition, due to the discharge of the energy storage system, the total energy consumption of each 5G base station group is greatly reduced. Inversely proportional to the total load curve of the system, thus reducing the power purchase of the main network and improving the operation efficiency.

Figure 7 shows the operation state of 5G base station before and after participating in demand response.

It can be seen from the figure that 5G base station participating in microgrid dispatching can effectively

TABLE 2 Time-of-use electricity prices.

Time period classification	Price (yuan/kWh)	
	Purchase	Sell
Peak hours (10:00–15:00, 18:00–22:00)	0.92	0.74
Normal period (07:00–10:00, 15:00–18:00, 22:00–24:00)	0.48	0.39
Trough period (00:00–07:00)	0.17	0.12

TABLE 3 Operation modes.

Scenario	Operating mode
Scenario 1	5G base station is not connected to the microgrid
Scenario 2	5G base station access to microgrid

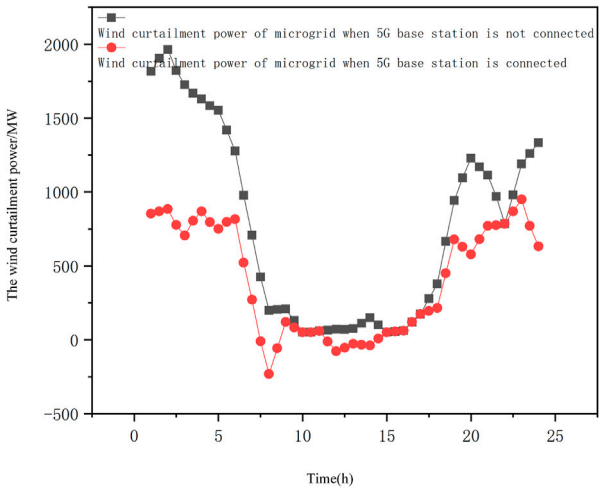


FIGURE 6
The wind curtailment power of the system.

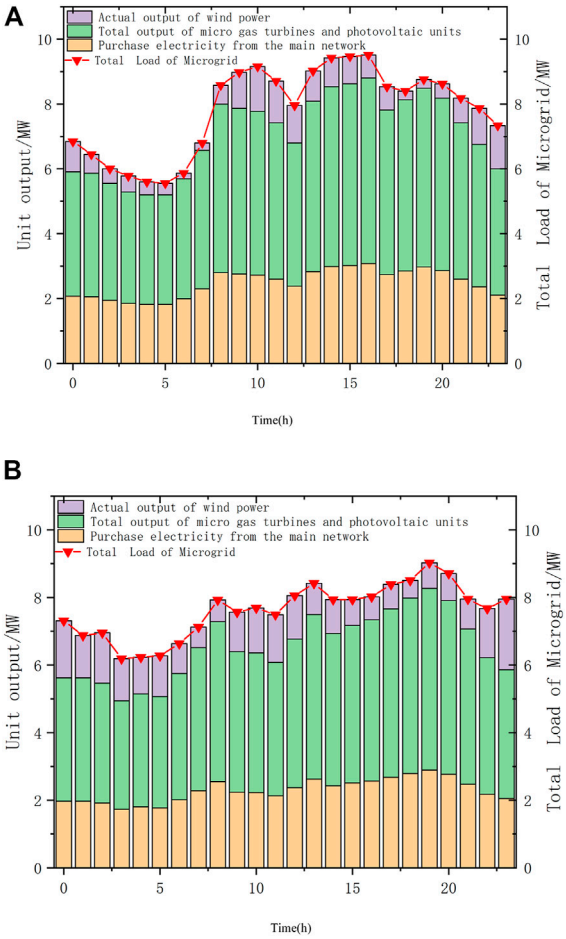


FIGURE 7
(A) 5G base station is not connected to the microgrid.**(B)**. 5G base station access to microgrid.

optimize the system load curve and obviously optimize the system load fluctuation. In Scenario 1, the peak value of power load is 9.5111 mw, and the valley value is 5.5556 MW. In the Scenario 2, the peak load is 9.0292 mw and the valley value is 6.1912 MW. It can obviously cut off the peak and fill in the valley, and improve the contradiction between power supply and demand during peak and low load periods.

When 5G base station does not participate in demand response, it can be regarded as a normal power load. After the 5G energy storage system is charged, it is necessary to increase the energy consumption of the base station. In the low-power generation period of wind power generation, energy storage system or grid-connected transmission can reduce the purchase of electricity from the external grid by microgrid, thus reducing the carbon emission on the power generation side. Because 5G base station can control its energy consumption by changing its own communication

TABLE 4 Comparison of system operation benefits in different scenarios.

Index	System operation cost (yuan)	System carbon emission (t)	Electricity purchase cost (yuan)	Wind power consumption (kWh)
Scenario 1	190,485.32	336.55	156,241.81	17,892.37
Scenario 2	188,542.32	241.39	136,204.3	26,541.30

equipment, reduce its energy consumption during peak power load, and use energy storage of base station to discharge at peak power load and charge at low power load, thus improving wind power consumption and reduce power purchase of main grid.

As can be seen from Table 4, when 5G base stations participates in the optimal dispatch of microgrid, although the operating cost of the system has increased, the wind power consumption of the system has increased by 8,648.93 kW h, indirectly reduces the electricity purchased by the microgrid from the external grid, and the electricity purchase cost of the main grid has decreased by 20,037.51 yuan. As a result, the total carbon emission of the system decreased by 95.16 tons.

5 Conclusion

In order to promote the realization of China's dual-carbon goal, and make use of new power system to vigorously promote the efficient consumption of clean energy, and realize the carbon emission reduction of power system, this paper takes the interactive response of 5G base station connected to microgrid as the research direction, and puts forward a multi-objective optimization model of 5G base stations connected to microgrid as the breakthrough point, taking the improvement of wind power utilization efficiency as the breakthrough point. Compared with existing research. In this paper, the characteristics of 5G microgrid, the low-carbon operation characteristics of microgrid and the optimal operation of microgrid are studied.

1) 5G base station can reduce its own power demand and improve the energy efficiency of the system by technical management means such as base station sleeping, transceiver turning off, downlink power control, *etc.*, thus indirectly reduce the carbon emission on the power generation side.
2) 5G base station participating in microgrid dispatching can effectively optimize the system load curve, and the fluctuation of system load is obviously optimized. Under the 5G disconnected scenario, the peak value of power load is 9.5111 MW and the valley value is 5.5556 MW. Under the 5G connection scenario, the peak load is 9.0292 MW, and the valley value is 6.1912 MW. It can obviously cut off the peak and fill in the valley, and improve the contradiction between power supply and demand during peak and low load periods.
3) The multi-objective model constructed in this paper can take into account the economic and environmental benefits of microgrid operation at the same time. When 5G base stations participates in the optimal dispatch of microgrid, although the operating cost of the system has increased, the consumption of wind power in the system has increased by 8,648.93 kWh, indirectly reduces the purchase of electricity from the main grid, and the purchase cost of electricity from the main grid has decreased by 20,037.51 uan.

The flexible access of 5G base station to microgrid can improve the utilization efficiency of electric equipment, balance the power flow of lines and reduce some network loss. According to the load situation, adjust the tap-changer of the main transformer at any time, so as to ensure that the grid voltage is within the fluctuation range specified in the regulations. Avoid the power quality degradation caused by the low voltage level in the grid during the peak load, and also increase the voltage at the end of the line, reduce the line current, and achieve the purpose of loss reduction. 5G base station carries out power management and flexible charging and discharging of energy storage battery, which can significantly change the power flow/voltage distribution in the distribution network, not only reduce the network loss, but also improve transmission efficiency. It can also effectively alleviate line congestion and improve the grid-connected capability of clean energy generation microgrid. However, there is no mathematical model for this discovery in this paper, which is the limitation of this paper and the next direction of our research.

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

Authors GW, ZX, and CY was employed by Fangyuan Construction Group Co., Ltd.

The remaining author declares 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

Aa, A., Htr, B., Mak, C., Sk, D., and Yn, E. (2021). A hybrid algorithm (BAPSO) for capacity configuration optimization in a distributed solar PV based Microgrid. *Energy Rep.* 7, 7906–7912. doi:10.1016/j.egyr.2021.01.034

Agalgaonkar, A., and Robinson, D. (2021). Cooperative control of microgrids: A review of theoretical frameworks, applications and recent developments. *Energies* 14, 8026. doi:10.3390/en14238026

Alabdali, Q. A., and Nahhas, A. M. (2021). Simulation study of grid connected photovoltaic system using PVsyst software: Analytical study for yanbu and rabigh regions in Saudi arabia. *Am. J. Energy Res.* 9, 30–44. doi:10.12691/ajer-9-1-4

Ali, A., Mahmoud, K., and Lehtonen, M. (2021). Optimal planning of inverter-based renewable energy sources towards autonomous Microgrids accommodating electric vehicle charging stations. *IET Generation Trans. Dist.* 16, 219–232. doi:10.1049/gtd2.12268

Asia, J. L., Taipei Willis, Ke, and Digitimes, n. d. (2022). *CCL demand for 5G base stations may gain momentum in Europe, US in*. [WWW Document]. DIGITIMES. URL Availabe at: https://www.digitimes.com/news/a20210621PD203.html (accessed 8 9, 22).

Chandran, T. L., and Abdulla, M. S. (2021). ALS-ORAN 5G: Adapting to lowpower-WAN & satellite - open radio access network in 5G. *Work. Pap.* 2021, 480.

Chen, B., Wang, J., Lu, X., Chen, C., and Zhao, S. (2020). Networked microgrids for grid resilience, robustness, and efficiency: A review. *IEEE Trans. Smart Grid* 12, 18–32. doi:10.1109/TSG.2020.3010570

Chen, B., Wang, J., Lu, X., Chen, C., and Zhao, S. (2020). Networked microgrids for grid resilience, robustness, and efficiency: A review. *IEEE Trans. Smart Grid* 12, 18–32. doi:10.1109/TSG.2020.3010570

Cheng, Z., Zeng, W., Shu, Z., Liao, W., and Lin, Y. (2021). Grid-connected micro-grid power supply and distribution transformer capacity optimization system design in the electricity sales environment. *J. Phys. Conf. Ser.* 1852, 022050. doi:10.1088/1742-6596/1852/2/022050

Espina, E., Llanos, J., Burgos-Mellado, C., Cárdenas-Dobson, R., Martínez-Gómez, M., and Sáez, D. (2020). Distributed control strategies for microgrids: An overview. *IEEE Access* 8, 193412–193448. doi:10.1109/ACCESS.2020.3032378

Guo, Y., Lei, X., and Wang, Q. (2021). Capacity coordination planning of isolated Microgrid and battery swapping station based on the quantum behavior particle swarm optimization algorithm. *Int. Trans. Electr. Energy Syst.* 31. doi:10.1002/2050-7038.12804

Gupta, T. N., Murshid, S., and Singh, B. (2020). Improving power quality of single phase utility grid connected to wind-PV system using multilayer-frequency adaptive fundamental signal extractor. *IET Renew. Power Gener.* 14, 2126–2134. doi:10.1049/iet-rpg.2019.0928

Hao, H., Hui, D., and Lau, D. (2020). Material advancement in technological development for the 5G wireless communications. *Nanotechnol. Rev.* 9, 683–699. doi:10.1515/ntrev-2020-0054

He, Y. F., Venkatesh, B., and Uuan, L. (2012). Optimal scheduling for charging and discharging of electric vehicles. *IEEE Trans. Smart Grid* 3 (3), 1095–1105. doi:10.1109/TSG.2011.2173507

Hirsch, A., Parag, Y., and Guerrero, J. (2018). Microgrids: A review of technologies, key drivers, and outstanding issues. *Renew. Sustain. Energy Rev.* 90, 402–411. doi:10.1016/j.rser.2018.03.040

ishchenko, E. A., Badaev, A. S., Volodko, A. V., and Fedorov, S. M. (2022). 5G base station antenna integrated into solar panel. *IOP Conf. Ser. Earth Environ. Sci.* 979, 012104. doi:10.1088/1755-1315/979/1/012104

Junejo, A. K., Elmorshedy, M. F., Elmorshedy, M. F., Wang, S., Buker, M. S., and Akindeji, K. T, et al.,(2021). Optimal planning and EMS design of PV based standalone rural microgrids. *IEEE Access* 9, 32908–32930. doi:10.1109/ACCESS.2021.3060031

Katiraei, F., and Iravani, M. R. (2006). Power management strategies for a microgrid with multiple distributed generation units. *IEEE Trans. Power Syst.* 21, 1821–1831. doi:10.1109/TPWRS.2006.879260

Ke, Y., Liu, J., Meng, J., Fang, S., and Zhuang, S. (2022a). Comprehensive evaluation for plan selection of urban integrated energy systems: A novel multi-criteria decision-making framework. *Sustain. Cities Soc.* 81, 103837. doi:10.1016/j.scs.2022.103837

Ke, Y., Tang, H., Liu, M., and Qi, X. (2022b). A hybrid decision-making framework for photovoltaic poverty alleviation project site selection under intuitionistic fuzzy environment. *Energy Rep.* 8, 8844–8856. doi:10.1016/j.egyr.2022.07.012

Lee, D., Son, S., and Kim, I. (2021). Optimal allocation of large-capacity distributed generation with the volt/var control capability using particle swarm optimization. *Energies* 14, 3112. doi:10.3390/en14113112

Liu, Q. H., He, Y. K., and Zhang, J. H. (2006). Investigation of control for AC-exited VSCF wind power generation system connected to grid. *Proc. CSEE* 26, 109–114. doi:10.1016/S0379-4172(06)60085-1

Lopes, J. P., Hatziargyriou, N., Mutale, J., Djapic, P., and Jenkins, N. (2007). Integrating distributed generation into electric power systems: A review of drivers, challenges and opportunities. *Electr. Power Syst. Res.* 77, 1189–1203. doi:10.1016/j.epsr.2006.08.016

Lotfi, H. (2021). Optimal sizing of distributed generation units and shunt capacitors in the distribution system considering uncertainty resources by the modified evolutionary algorithm. *J. Ambient. Intell. Humaniz. Comput.* 13, 4739–4758. doi:10.1007/s12652-021-03194-w

Marzband, M., Azarinejadian, F., Savaghebi, M., Pouresmaeil, E., Guerrero, J. M., and Photovoltaicbody, G. (2018). Smart transactive energy framework in grid-connected multiple home Microgrids under independent and coalition operations. *Renew. Energy* 126, 95–106. doi:10.1016/j.renene.2018.03.021

Marzband, M., Azarinejadian, F., Savaghebi, M., Pouresmaeil, E., Guerrero, J. M., and Photovoltaicbody, G. (2018). Smart transactive energy framework in grid-connected multiple home Microgrids under independent and coalition operations. *Renew. Energy* 126, 95–106. doi:10.1016/j.renene.2018.03.021

Naval, N., and Yusta, J. M. (2021). Optimal short-term water-energy dispatch for pumping stations with grid-connected photovoltaic self-generation. *J. Clean. Prod.* 316, 128386. doi:10.1016/j.jclepro.2021.128386

Onojo, J. O., Okozi, S. O., Okafor, K. C., Uzoechi, L. O., and Nkwachukwu, C. (2021). Towards smart grid economics using distributed generation systems-hybrid biomass, solar and fossil fuel plants. *Int. J. Mechatronics Manuf. Syst.* 11, 5019.

Parhizi, S., Lotfi, H., Khodaei, A., and Bahramirad, S. (2015). State of the art in research on microgrids: A review. *IEEE Access* 3, 890–925. doi:10.1109/ACCESS.2015.2443119

Tan, Z., De, G., Li, M., Lin, H., Yang, S., Huang, L., et al. (2020). Combined electricity-heat-cooling-gas load forecasting model for integrated energy system based on multi-task learning and least square support vector machine. *J. Clean. Prod.* 248, 119252. doi:10.1016/j.jclepro.2019.119252

Tavakoli, M., Shokridehaki, F., Akorede, M. F., Marzband, M., Vechiu, I., and Pouresmaeil, E. (2018). CVaR-based energy management scheme for optimal resilience and operational cost in commercial building Microgrids. *Int. J. Electr. Power & Energy Syst.* 100, 1–9. doi:10.1016/j.ijepes.2018.02.022

Wang, X., Zhu, Q., and Wang, Y. (2021). Optimal allocation of wind-solar storage capacity of Microgrid considering carbon emission reduction benefits. *IOP Conf. Ser. Earth Environ. Sci.* 804, 032015. doi:10.1088/1755-1315/804/3/032015

Weishang, G., Yihua, M., Xuexing, Z., and Huan, Y. (2022). Internal benefit optimization model of gas-thermal power virtual power plant under China's carbon neutral target. *Energy Sci. Eng.* 10, 1227–1239. doi:10.1002/ese3.1097

Weissberger, A. (2020). 5G base station deployments. Open-RAN Competition & HUGE 5G BS Power Problem [WWW Document]. Technology Blog. URL Availabe at: https://techblog.comsoc.org/2020/08/07/5g-base-station-deployments-open-ran-competition-huge-5g-bs-power-problem/(accessed 8 9, 22).

Wen, S., Xu, Y., and Dong, Y. (2021). Low-profile wideband omnidirectional antenna for 4G/5G indoor base station application based on multiple resonances. *IEEE Antennas Wirel. Propag. Lett.* 20, 488–492. doi:10.1109/LAWP.2021.3054682

Yong, P., Zhang, N., Ci, S., and Kang, C. (2021). 5G communication base stations participating in demand response: Key technologies and prospects. *Zhongguo Dianji Gongcheng Xuebao/Proceedings Chin. Soc. Electr. Eng.* 41, 5540–5551. doi:10.13334/j.0258-8013.pcsee.210183

Yu, X., and Zheng, D. (2020). Cross-regional integrated energy system scheduling optimization model considering conditional value at risk. *Int. J. Energy Res.* 44, 5564–5581. doi:10.1002/er.5307

Yuvaraj, T., Devabalaji, K. R., Prabaharan, N., Alhelou, H. H., Siano, P., Pal, P., et al. (2021). Optimal integration of capacitor and distributed generation in distribution system considering load variation using bat optimization algorithm. *Energies* 14, 3548. doi:10.3390/en14123548

Zandrazavi, S. F., Guzman, C. P., Pozos, A. T., Quiros-Tortos, J., and Franco, J. F. (2022). Stochastic multi-objective optimal energy management of grid-connected unbalanced Microgrids with renewable energy generation and plug-in electric vehicles. *Energy* 241, 122884. doi:10.1016/j.energy.2021.122884

Zhang, H. Q., Wei, P. F., and Li, W. W. (2011). Study on simulation of three-phase Z-source inverter for grid-connected wind power generation. *Adv. Mat. Res.* 383–390, 2018–2024. doi:10.4028/www.scientific.net/amr.383-390.2018

Zhao, Z., Wang, H., and Zhao, H. (2020). Research on filter parameter optimization method of grid-connected photovoltaic system. *IOP Conf. Ser. Earth Environ. Sci.* 619, 012020. doi:10.1088/1755-1315/619/1/012020

Zhou, M., Cheng, S., Xu, L., Wang, L., Guo, Q., and Cai, W. (2021). Robust switching gain-based fractional-order sliding mode control for wind-powered microgrids. *Complexity* 2021, 1–12. doi:10.1155/2021/6697792

Zia, M. F., Elbouchikhi, E., and Benbouzid, M. (2018). Microgrids energy management systems: A critical review on methods, solutions, and prospects. *Appl. Energy* 222, 1033–1055. doi:10.1016/j.apenergy.2018.04.103

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fenrg.2022.1043087

OPEN ACCESS

EDITED BY
Yongming Han,
Beijing University of Chemical Technology, China

REVIEWED BY
Jiangping Chen,
Shanghai Jiao Tong University, China
Jian Zhang,
University of Wisconsin Green Bay, United States

*CORRESPONDENCE
Zhang Yanting,
ytzhang@upc.edu.cn
Wang Lin,
wlhaust@163.com

SPECIALTY SECTION
This article was submitted to Process and Energy Systems Engineering, a section of the journal Frontiers in Energy Research

RECEIVED 13 September 2022
ACCEPTED 29 November 2022
PUBLISHED 04 January 2023

CITATION
Hao Z, Yanting Z, Lin W, Jingyu X, Zheng H, Guangzhi Z and Yunlong Z (2023), System optimization of steam generation in the dual-flash compound circulation system.
Front. Energy Res. 10:1043087.
doi: 10.3389/fenrg.2022.1043087

System optimization of steam generation in the dual-flash compound circulation system

Zhang Hao[1], Zhang Yanting[1]*, Wang Lin[2]*, Xu Jingyu[3], Huang Zheng[1], Zhang Guangzhi[1] and Zheng Yunlong[1]

[1]Mechanical and Electrical Engineering, China University of Petroleum (East China), Qingdao, Shandong, China, [2]School of Civil Engineering, Henan University of Science and Technology, Luoyang, China, [3]Shanghai Heimdallr Energy Saving Technology Co., Ltd., Shanghai, China

Concentrating on the problem of massive energy loss in the compressor, expansion valve, and the other components present in the high-temperature heat pump system under extensive temperature lift, the dual-flash compound circulation system is proposed and the thermodynamic model of the dual-flash compound circulation system was established. The article combines the multivariate simulated annealing algorithm, utilizes the system COP as the optimization goal, and completes the calculation of the thermodynamic parameters in the steady-state of the system that is based on satisfying the conditions of the system process. Using R245fa as the refrigerant, the condensation temperature is set within the range of 110°C–140°C for the model calculation. The results show that, compared to the traditional two-stage compression system under the same environment, the COP of the dual-flash compound circulation system can be increased up to 5.71%–12.13%, and the exergy efficiency can be increased by 5.11%–10.71%, respectively. Besides, steam production per unit refrigerant is also increased by 3.79%–5.14%. Finally, the feasibility of the theoretical model is verified by simulation, and it is concluded that the dual-flash compound circulation system has better steam production performance at the extensive temperature lift and the elevated condensation temperature.

KEYWORDS

high-temperature heat pump steam-system, multiple annealing optimization, energy analysis, the dual-flash compound circulation system, large temperature lift conditions

Introduction

With the rapid development of the world economy, the serious pollution caused by industrial production can not be ignored. According to the 2019 China Statistical Yearbook, from 1990 to 2018, China's total energy consumption increased from 987.03 million tons of standard coal to 4.64 billion tons, a nearly five-fold increase in energy consumption (China National Bureau of Statistics, 2019). At present, the main types of energy consumption are still coal and oil, but with the development of green energy technology, the proportion of fossil energy consumption is gradually declining. As shown in Figure 1A, China's total energy consumption in 2017 was 4.485 billion tons of

standard coal. The energy consumption of each industry and its proportion statistics are shown in Figure 1B. Industrial and manufacturing industries consume the highest energy, accounting for 41.11% and 34.22%, respectively. In that year, industrial sulfur dioxide, industrial nitrogen oxide and other harmful gas emissions reached 2.3782 million tons.

In order to reduce the emission of harmful gases, many scholars turn their attention to the field of high temperature heat pumps (HTHP) (Chamoun et al., 2014; Wu et al., 2016). It is hoped to use HTHP technology to replace part of the traditional calcination boiler technology to achieve 100–130°C steam production (Meggers and Leibundgut, 2011; Xing et al., 2014).

The HTHP system usually has the characteristics of high condensation temperature and the temperature lift is huge, which also makes the system has higher requirements for its circulation structure and working refrigerant. HTHP technology is divided into the compression HTHP technology and the absorption HTHP technology. Gao et al. (2021) proposed an absorbing compound HTHP with large temperature lift condition by using air source as heat source. The coefficient of performance was between 1.2 and 1.7. Liu et al. (2022) proposed to use the waste heat of low-temperature flue gas at 170°C to produce steam at 150°C, and verified by the experimental prototype that the steam heat output reached 70.15 kW. KOBELCO and CRIEPI (Japan Electric Power Company) (KOBELCO and TEPCO, 2011; Kaida et al., 2015) first proposed the use of HTHP technology for steam production, and in 2011 designed the first commercial high-temperature heat pump steam system, model SGH120. The energy efficiency coefficient (COP) of the system at 120°C was 3.5. Gong YL et al. (Liu et al., 2015) analyzed and compared five high-temperature refrigerants (R123, R141b, R245ca, R245fa and R114) based on the SGH120 model. The results show that R245fa has better comprehensive performance when the temperature lift of the system is in the range of 60–80°C, and the COP of the system is 3.65% higher than that of R114. Subsequently, Gong YL's team (Liu et al., 2017) optimized the model of the HTHP system. By introducing the air supplement parameter (air supplement rate B) and the sub-cooler parameters (heat source water subcooling rate A), it is pointed out that the parameter-B and the parameter-A also have significant influence on the thermodynamic performance of the system. When the typical values of B and A are .4 and .2, respectively, the COP of the system is 13.5% higher than that of SGH120. Bamigbetan et al. (2018) pointed out that HTHP systems using butane as refrigerant have the ability to replace traditional boilers for steam production. To test this theoretical hypothesis, the team (Bamigbetan et al., 2019) designed a butane compressor and built an experimental platform with a HTHP system. D. H et al. (KANG et al., 2017) first analyzed the difference between SGH120 and SGH165. Summarized the influence of the flash tank and MVR device on the steam system of HTHP. In 2019, the team (Hoon et al., 2019) developed a HTHP system for steam generation, including an internal heat exchanger (IHX) based on theoretical research. The final steam produced by the system can reach temperatures of 130°C. Dai et al. (2020) adopted the structure of dual-pressure condenser to carry out piecewise heat transfer for thermal products, effectively reducing the exergic loss caused by excessive heat exchange temperature difference. Zhang YT et al. (Hao et al., 2022) conducted relevant research on the influence of IHX in the steam system of a complex HTHP, and the results showed that the COP after IHX positioning was increased by 4.87% compared with the conventional cycle. Xu et al. (2021) pointed out that the compressed heat pump had higher COP but lower exergy efficiency, indicating that exergy loss generated by the cycling process should be reduced if the compressed HTHP system was used to produce steam.

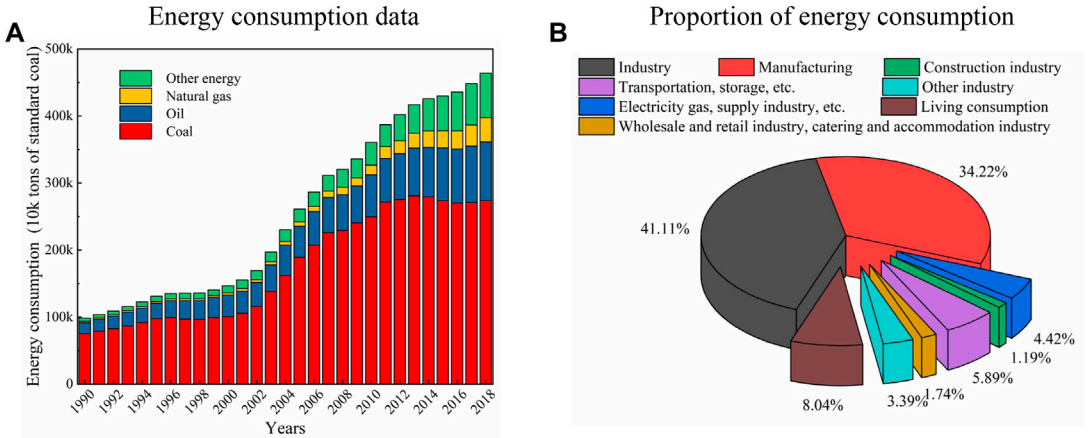


FIGURE 1
Energy consumption in different years and various fields in China. (A) Energy consumption data. (B) Proportion of energy consumption.

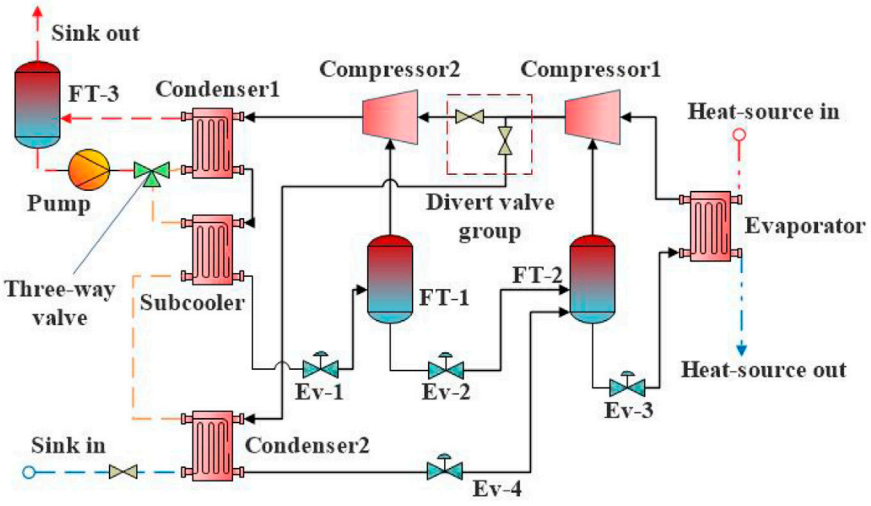


FIGURE 2
The structure of the dual-flash compound circulation system.

In summary, from the perspective of steam production process, the global optimization of the system is often easily neglected. The single optimization of high temperature heat pump system will easily lead to a large temperature difference between the initial temperature of thermal products and the condensing temperature of refrigerant, which will increase the exergy loss of the system. Therefore, the high temperature heat pump system is optimized globally in this paper. On the one hand, the COP and exergy efficiency of the system were improved through double flash gas injection process. On the other hand, the method of piecewise heat transfer is adopted to reduce the exergy loss caused by large heat transfer temperature difference. Finally, a set of the Dual-Flash Compound Circulation System was put forward. Based on this system, the multiple simulated annealing algorithm is used to optimize the thermodynamic parameters of the system, and a multi-objective parameter optimization algorithm for complex heat pump system is proposed, so as to improve the steam production performance of HTHP.

System model

Cycle structure and principle

The research result of Zhang YT et al. (Hao et al., 2022) showed that avoiding part of refrigerant from entering the expansion valve (Ev) was an effective way to improve the exergy efficiency of heat pump system without affecting the gas-liquid ratio of FT. Meanwhile, in order to reduce the loss of exergy generated by condensation, the dual-flash compound circulation system combined multiple vapor injection process with segmental heat transfer process. The design structure of the system is shown in Figure 2:

As shown in the figure, the dual-flash compound circulation is composed of two thermal cycles. Circulation-1 is a dual-flash vapor injection circulation consisting of two compressors and two flash tanks. The refrigerant enters into compressor-1 from the evaporator for quasi-primary compression. It is then mixed with the saturated vapor refrigerant of FT-2 and compressed again by compressor-1. This process is called the vapor injection compression process. Part of the refrigerant at the outlet of compressor-1 flows into compressor-2 through the shunt valve assembly to realize the second vapor injection compression. After this part of refrigerant completes all the compression process, it enters condenser-1 and the subcooler to transfer heat to the thermal product for steam preparation. After heat exchange, the refrigerant passes through Ev-1, Ev-2 and Ev-3 in turn, and finally enters the evaporator to form the dual-flash vapor injection circulation structure. Circulation-2 is the traditional FT quasi-two-stage compression cycle structure. The two circulation structures are coupled by sharing FT-2, Ev-3, and evaporator, and the refrigerant is shunted through the shunt valve set. Condenser-1, subcooler and condenser-2 were used to realize the segmented heat transfer process for heat products.

The pressure enthalpy characteristics of the dual-flash compound circulation system are shown in Figure 3 below:

It can be seen from the pressure enthalpy characteristics of the system in Figure 3 that there are five pressures in the HTHP system, namely condensing pressure pcond1 of the condenser-1, flash pressure p_{FT1} of the FT-1, condensing pressure p_{cond2} of the condenser-2 (Intermediate pressure p_{mid}), flash pressure p_{FT2} of

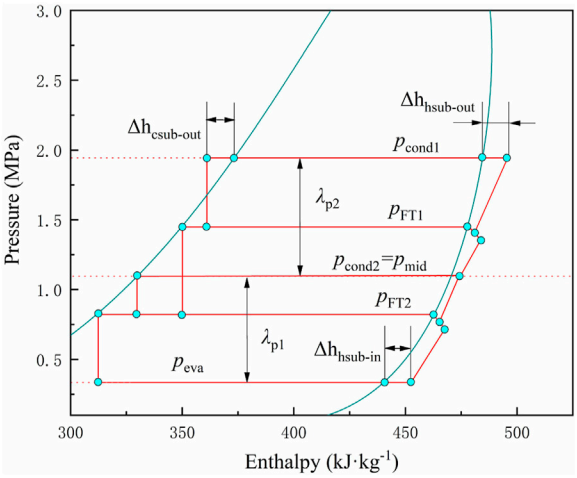


FIGURE 3
The pressure and enthalpy characteristics of the dual-flash compound circulation system.

the FT-2 and evaporating pressure p_{eva} of the evaporator. In order to facilitate the subsequent optimization calculation, the intermediate pressure optimization experience of the traditional two-stage compression system is used, which sets the pressure p_{mid} is set to a fixed value, denoted as Eq. 1.

$$p_{mid} = p_{cond2} = \sqrt{p_{eva} \times p_{cond1}} \tag{1}$$

The relationship between the five pressures in the system is shown in Eq. 2:

$$p_{cond1} > p_{FT1} > p_{mid} = p_{cond2} > p_{FT2} > p_{eva} \tag{2}$$

Therefore, when the p_{cond2} and p_{eva} in the system are determined, the intermediate pressure of the system will become a fixed value. The two flash pressures of the system will be used as variable parameters to optimize the system.

Thermodynamic model

According to the dual-flash compound circulation system, the thermodynamic models of heat source, compound circulation and steam generation were established respectively.

The heat exchange of the heat source thermodynamic model is the main heat source of steam produced by the system. When the mass flow rate of the heat source is m_c, Q_c is used to represent the heat supply of the heat source:

$$Q_c = m_c \times (h_{c1} - h_{c2}) \tag{3}$$

where h_{c1} is the specific enthalpy of the heat source before heating, kJ/kg; h_{c2} is the specific enthalpy of the heat source after heating, kJ/kg; m_c is the mass flow of the quantitative heat source, kg/s.

The HTHP compound circulation is the hub of heat transfer in the system, and its advancement determines the thermodynamic performance of the heat pump system. Meanwhile, the shunt coefficient x of the compound circulation has a significant influence on the system performance. The shunt coefficient x is the proportion of the refrigerant flow from the shunt valve to compressor-2. The mass flow rate of the refrigerant at the outlet of compressor-1 is m_1, and that at the outlet of compressor-2 is m_2. The relationship between m_1, m_2 and x can be expressed as follows:

$$m_2 = m_1 \cdot x \tag{4}$$

They are assuming that the gas ratios of FT-1 and FT-2 are denoted as a and b, respectively. According to the law of mass conservation, the expressions of a and b are shown in Eq. 5.

$$\begin{cases} a = (h_{co-out} - h_{FT1-l}) / (h_{FT1-v} - h_{FT1-l}) \\ b = [(1-x) \times h_{cond2} + x \cdot h_{FT1-l} - h_{FT2-l}] / (h_{FT2-v} - h_{FT2-l}) \end{cases} \tag{5}$$

Among them, $h_{co\text{-}out}$ represents the specific enthalpy of the refrigerant discharged from the sub-cooler, kJ/kg; $h_{FT1\text{-}l}$, $h_{FT1\text{-}v}$, $h_{FT2\text{-}l}$, and $h_{FT2\text{-}v}$ represent the specific enthalpy of the saturated liquid refrigerant and the saturated gaseous refrigerant in the FT-1, respectively. The specific enthalpy and the specific enthalpy of the saturated liquid refrigerant and the specific enthalpy of the saturated gaseous refrigerant in the FT-2, kJ/kg; h_{cond2} represents the specific enthalpy of the condenser-2, in the saturated liquid state of the refrigerant, kJ/kg.

Combining Eqs 4, 5, the expression of m_1 is determined from the heat supply of the heat source, which is Eq. 6.

$$m_1 = \frac{Q_c}{(h_{com1-in} - h_{Ev3-out}) \cdot (1-b)} \tag{6}$$

where $h_{com1\text{-}in}$ represents the specific enthalpy of the refrigerant at the inlet of the compressor-1, kJ/kg; $h_{Ev3\text{-}out}$ represents the specific enthalpy of the refrigerant at the outlet of the Ev-3, kJ/kg.

According to the process, the heat released by Condenser-1, sub-cooler, and the Condenser-2 in the high-temperature heat pump subsystem is defined as Q_1, Q_2, and Q_3, respectively, and expressed as:

According to the steam production process of the system, the heat released by condenser-1, subcooler and condenser-2 in the HTHP compound cycle is defined as Q_1, Q_2, and Q_3, as shown in Eq. 7:

$$\begin{cases} Q_1 = m_2 \cdot (h_{com2-out} - h_{cond1}) \\ Q_2 = m_2 \cdot (h_{cond1} - h_{co-out}) \\ Q_3 = m_1 \cdot (h_{com1-out} - h_{cond2}) \times (1-x) \end{cases} \tag{7}$$

Among them, $h_{com1\text{-}out}$ and $h_{com2\text{-}out}$ represent the specific enthalpy of the refrigerant discharged from the compressor-1 and the compressor-2, kJ/kg; h_{cond1} represents the specific

enthalpy of the refrigerant in the saturated liquid state after heat exchange in Condenser-1, kJ/kg.

The isentropic efficiency of the air-injected compressor is calculated according to the total pressure ratio of the suction and the exhaust, referring to the isentropic efficiency fitting formula that is $\eta_{is} = .9343–.04478 \times \lambda_p$ proposed by Zheng (2017), the Compressor-1 and the isentropic efficiency of Compressor-2 is expressed as:

By referring to the isentropic efficiency fitting formula of $\eta_{is} = .9343–.04478 \times \lambda_p$ proposed by Zheng (2017), the isentropic efficiency of the vapor-injected compressor is calculated according to the pressure ratio of suction and exhaust., then the isentropic efficiency of compressor-1 and compressor-2 can be expressed as:

$$\begin{cases} \eta_{is-1} = 0.9343 - 0.04478 \times (p_{mid}/p_{eva}) \\ \eta_{is-2} = 0.9343 - 0.04478 \times (p_{cond1}/p_{mid}) \end{cases} \tag{8}$$

The power consumption of compressor-1 and compressor-2 is expressed as W_1 and W_2:

$$\begin{cases} W_1 = m_1 \cdot [(1-b) \times (h_{com1-1} - h_{com1-in}) + (h_{com1-out} - h_{mix1})] \\ W_2 = m_2 \cdot [(1-a) \times (h_{com2-1} - h_{com1-out}) + (h_{com2-out} - h_{mix2})] \end{cases} \tag{9}$$

where h_{com1-1} and $h_{com2-1,}$ respectively, represent the specific enthalpy of the quasi-stage compression of the refrigerant by the compressor-1 and compressor-2, kJ/kg; $h_{com1\text{-}in}$ and $h_{com2\text{-}in}$ represent the specific enthalpy of refrigerant at the suction port of compressor-1 and compressor-2, kJ/kg; h_{mix1} and $h_{mix2,}$ respectively represent the specific enthalpy after the vapor injection refrigerant and the quasi-stage compression refrigerant in the compressor-1 and compressor-2 are mixed, kJ/kg.

The COP of the dual-flash compound circulation system can be expressed as Equation 10.

$$COP = \frac{Q_1 + Q_2 + Q_3}{W_1 + W_2} \tag{10}$$

It is assumed that the initial temperature of the inlet hot water at the steam generating side is T_{sink0}, and the heat loss in the heat transfer process is ignored. The mass flow rate of the hot water at the steam generating side is m_{sink}, which is the steam output obtained by the system, as shown in Equation 11.

$$m_{sink} = \frac{Q_1 + Q_2 + Q_3}{h_{sink} - h_{sink0}} \tag{11}$$

h_{sink} is the specific enthalpy of the final saturated steam, kJ/kg; h_{sink0} is the inlet hot water at the steam generating side, kJ/kg.

Exergy analysis

When the ambient temperature is T_{Amb}, the specific enthalpy and specific entropy of the corresponding environment are expressed as h_{Amb} and s_{Amb}. According to the definition of exergy, exergy loss expressions of the dual-flash compound circulation system equipment in each process are shown in Table 1 below:

Through the expression of exergy loss of equipment in each process of the system, the total exergy loss of the system can be obtained, which is expressed as Eq. 12.

$$E_{total} = E_{cond1} + E_{cond2} + E_{subcooler} + E_{Eva} + E_{Ev1} + E_{Ev2} + E_{Ev3} + E_{Ev4} + E_{FT1} + E_{FT2} + E_{com1} + E_{com2} \tag{12}$$

The exergy efficiency of the system is the ratio of exergy output and exergy obtained by the system, which can reflect the advancement of a circulatory system to a large extent, and can be expressed by Eq. 13.

$$E_{ee} = \frac{1 - E_{total}}{W_1 + W_2} \tag{13}$$

Process constraints

Exergy loss of the system could be reduced and thermodynamic performance of the system could be improved in the hierarchical heat transfer process, but it had certain restrictions on the heat transfer relationship of the system. When the initial temperature of the supplementary hot water is $T_{sink0,}$ and the mass flow rate is m_{sink}, assuming that the pinch temperature in the heat transfer process is 5 °C, according to the second law of thermodynamics, the water temperature reached after the absorption of Q_3 should not be higher than T_{cond2} _5°C. Set this limit temperature as $T_{sink}' = T_{cond2}$ _5°C, then the specific enthalpy of the hot water added at this temperature is h_{sink}'. When the hot water absorbs Q_2 and Q_1, the final production steam temperature should not be higher than T_{cond1} _5, denoted as T_{sink}. The h_{sinkv} is the specific enthalpy of saturated steam and h_{sinkl} is the specific enthalpy of saturated hot water. Due to the conservation of energy and mass during the cycle operation of the system, two heat specific requirements related to the phase transition of hot water were derived, denoted $R_{condition1}$ and $R_{condition2}$ respectively. The relevant expression is as follows (14).

$$\begin{cases} R_{condition1} = \dfrac{h_{sinkv} - h_{sinkl}}{h_{sinkl} - h_{sink0}} \\ R_{condition2} = \dfrac{h_{sinkl} - h_{sink'}}{h_{sink'} - h_{sink0}} \end{cases} \tag{14}$$

Based on this, when the system satisfies Eq. 15 at the same time, the long-term steady operation of the system can be realized.

TABLE 1 Exergy loss model of each component.

System equipment	Exergy loss expression	
Evaporator		$E_{Eva} = T_{Amb} \times [m_1 \times (1-b) \cdot (s_{com1-in} - s_{Ev3-out}) - m_c \cdot (s_{c1} - s_{c2})]$
Condenser-1		$E_{cond1} = T_{Amb} \times [m_{sink} \cdot (s_{sinkl} - s_{sink2}) - m_2 \cdot (s_{com2-out} - s_{cond1})]$
Condenser-2		$E_{cond2} = T_{Amb} \times [m_{sink} \cdot (s_{sink1} - s_{sink0}) - m_1 \times (1-x) \cdot (s_{com1-out} - s_{cond2})]$
Subcooler		$E_{subcooler} = T_{Amb} \times [m_{sink} \cdot (s_{sink2} - s_{sink1}) - m_2 \cdot (s_{cond1} - s_{co-out})]$
Ev-1		$E_{Ev1} = T_{Amb} \cdot m_2 \times (s_{Ev1-out} - s_{co-out})$
Ev-2		$E_{Ev2} = T_{Amb} \cdot m_2 \cdot (1-a) \times (s_{Ev2-out} - s_{FT1-l})$
Ev-3		$E_{Ev3} = T_{Amb} \cdot m_1 \cdot (1-b) \times (s_{Ev3-out} - s_{FT2-l})$
Ev-4		$E_{Ev4} = T_{Amb} \cdot m_1 \cdot (1-x) \times (s_{Ev4-out} - s_{cond2})$
FT-1		$E_{FT1} = T_{Amb} \cdot m_2 \cdot [a \cdot s_{FT1-v} + (1-a) \cdot s_{FT1-l} - s_{Ev1-out}]$
FT-2		$E_{FT2} = T_{Amb} \cdot m_1 \cdot [b \cdot s_{FT2-v} + (1-b) \cdot s_{FT2-l} - s_{Ev2-out} \cdot x - s_{Ev4-out} \cdot (1-x)]$
Compressor-1	E_{com1-1}	$E_{com1-1} = T_{Amb} \cdot m_1 \cdot (1-b) \times (s_{com1-1} - s_{com1-in})$
	E_{com1-2}	$E_{com1-2} = T_{Amb} \cdot m_1 \times (s_{com1-out} - s_{mix1})$
	$E_{com1-mix}$	$E_{com1-mix} = T_{Amb} \cdot m_1 \times [s_{mix1} - (1-b) \cdot s_{com1-1} - b \cdot s_{FT2-v}]$
Compressor-2	E_{com2-1}	$E_{com2-1} = T_{Amb} \cdot m_2 \cdot (1-a) \times (s_{com2-1} - s_{com1-out})$
	E_{com2-2}	$E_{com2-2} = T_{Amb} \cdot m_2 \times (s_{com2-out} - s_{mix2})$
	$E_{com2-mix}$	$E_{com2-mix} = T_{Amb} \cdot m_2 \times [s_{mix2} - (1-a) \cdot s_{com2-1} - a \cdot s_{FT1-v}]$

$$\begin{cases} R_{condition1} \leq \dfrac{Q_1}{Q_2 + Q_3} \\ R_{condition2} \leq \dfrac{Q_2}{Q_3} \end{cases} \tag{15}$$

It is assumed that the evaporation temperature of the system is 50°C, the evaporation superheat is 5°C, the condensation temperature is 120°C, and the condensation supercooling degree is 5°C. When the mass flow rate of heat source is 1 kg/s, and the initial temperature is 60°C. The above parameters were quantified, the shunt coefficient x and p_{FT1} and p_{FT2} were set as variable parameters, and the calculation system was optimized to maximize COP.

The relationship between $R_{condition1}$, $R_{condition2}$, and Q_1, Q_2, and Q_3 are shown in Figure 4.

It can be seen that, with the change of x in the dual-flash compound circulation system, the heat transfer ratio of $R_{condition1}$ was constant 8.11, while $R_{condition2}$ was always equal to 1.3947. It follows that when the evaporation temperature is set, the value of $R_{condition1}$ depends only on the maximum condensing temperature of the system, while $R_{condition2}$ is not only affected by the T_{cond1}, but also related to the T_{cond2} of the system. Combined $R_{condition1}$ and $R_{condition2}$, the shunt coefficient x needs to be greater than .958 and .952 at the same time. Therefore, the shunt coefficient x greater than .958 under this condition is the technical condition limitation of the system.

Methodology

Due to the complex cycle process of the system, the parameters in the system have coupling and variability. Therefore, it is necessary to optimize each parameter in the system by appropriate optimization algorithm to ensure the heating performance of the system. The multivariable simulated annealing algorithm has the advantages of simple optimization process and strong robustness, which is suitable for the multi-objective parameter optimization process of the Dual-Flash Compound Circulation system (Kryukova et al., 2019; Zhang et al., 2022).

Simulated annealing algorithm (SA) is an optimization algorithm which is often used to find the global nadir through the simulated annealing process. For the parameter optimization of the high point bit, the optimization process can be realized by adding symbols or changing the operation direction. The schematic diagram of the algorithm is shown in Figure 5 (Bai and Liu, 2016).

It can be seen from the annealing calculation that there are multiple annealing points during the whole annealing process. Under the action of multiple influential variables, the highest annealing point of the whole process is finally obtained as the unique solution of the system optimization. In order to maintain the stability of the system, some unnecessary variables need to be reduced. Thermodynamic model based on the system, the shunt coefficient of x, system too cold quantity $\Delta T_{sub\text{-}co}$, the FT-1 flash

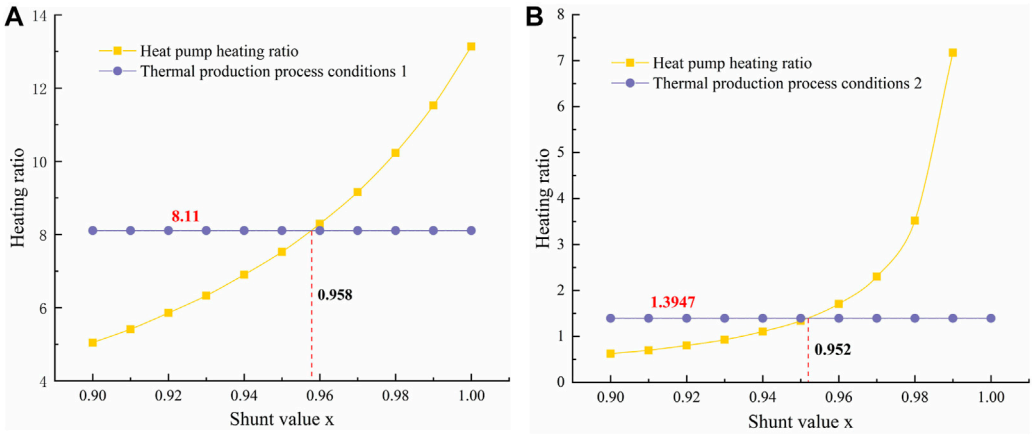


FIGURE 4
The influence of the shunt coefficient *x* on the process limitation under certain parameters. **(A)** $R_{condition1}$ relationship. **(B)** $R_{condition2}$ relationship.

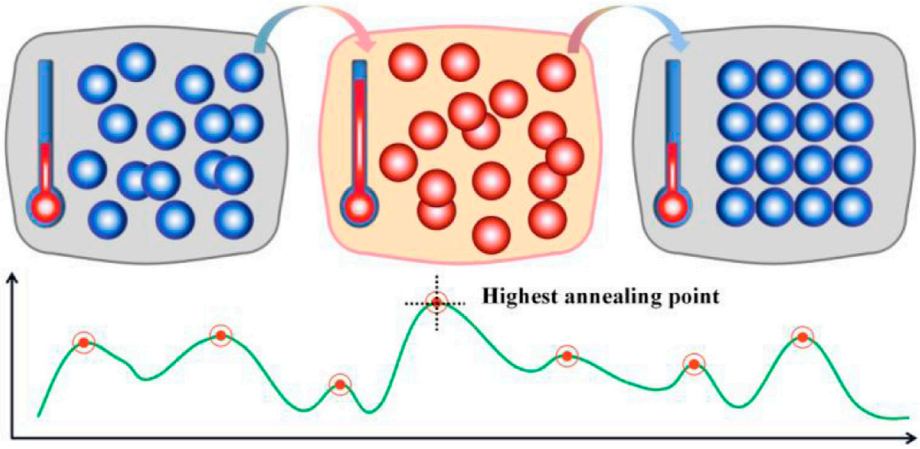


FIGURE 5
Principle of the simulated annealing algorithm.

pressure p_{FT}, FT-1 flash p_{FT} COP of the system have a significant impact, including x and $\Delta T_{sub\text{-}co}$ has also changed the constraints in the process of system. Therefore, the above four parameters as variables, determine multivariate calculation process of simulated annealing algorithm, as shown in Figure 6.

Due to the system variables is more, the algorithm parameters must be large enough to satisfy the movement step length, step length is set to 100,000. Simulated annealing temperature gradient reduction should be small enough, the period is set to .1[30]. The thermodynamic model of the system was introduced into the algorithm flow, as shown in Figure 6, and the maximum COP was taken as the optimization condition of the judicial system. When the judgment condition and termination condition are satisfied at the same time, the system optimization parameters obtained are substituted into the system thermodynamic model to obtain the system conditions and other thermodynamic parameters.

Results and discussions

Optimization and analysis of parameters

The working condition of the system is formulated. The evaporation temperature is 50°C, the condensation temperature tends to change within the range of 110–140°C, the pinch

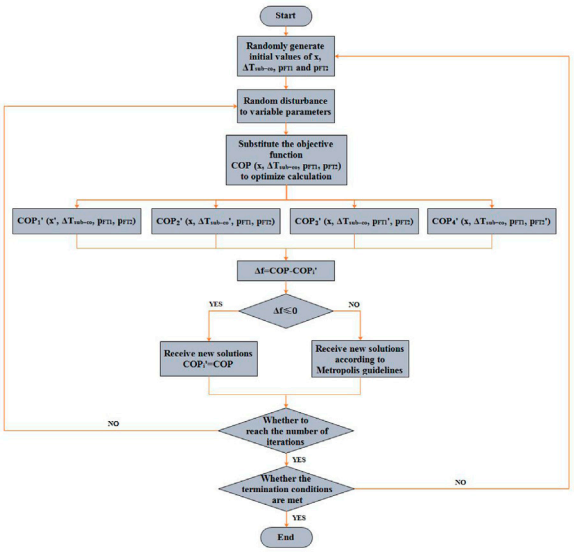


FIGURE 6
System multivariate simulated annealing optimization algorithm flow.

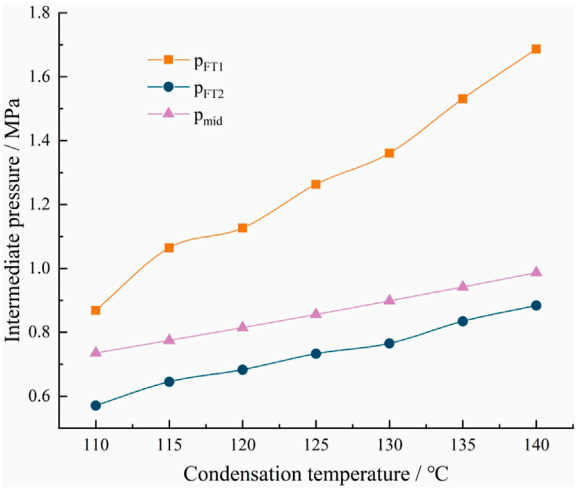


FIGURE 7
Variable relationship of intermediate pressure at different condensation temperatures.

temperature of heat transfer in the system is 5°C, and the intermediate pressure is calculated through Eq. 1 to reduce the calculation time of the optimization model. The other intermediate parameters were optimized according to the multivariate simulated annealing algorithm. As the condensation temperature of the system changes, the relationship between the intermediate pressure corresponding to the maximum COP of the system is shown in Figure 7 below.

The p_{mid} is a variable related to the condensing pressure and evaporation pressure. According to Eq. 1, when the evaporation pressure does not change, p_{mid} shows a linear change with a slow increase as the condensation temperature increases. The flash pressure p_{FT1} and the p_{FT2} fluctuated with the increase of the condensation temperature. This is due to the parameter fluctuation caused by considering the global optimal solution in the multi-objective optimization process of the system. From the fluctuation range, with the increase of condensation temperature, the change of pressure variables in the system is more significant. However, the variation range of the other two variables parameters is relatively small, and the variation trend is shown in Table 2.

From the optimized parameters, it can be concluded that the split coefficient x is in the limit position as the condensation temperature changes. Therefore can be concluded that under the conditions, the smaller x leads to the higher the COP of the system. When the condensing temperature was higher than 125°C, the minimum value of the shunt coefficient x was reduced from .96 to .95, which verified that $R_{condition1}$ and $R_{condition2}$ were affected by the condensing temperature and gradually decreased with the increase of the condensing temperature. At the same time, the optimal subcooling degree of the system increases first and then decreases with the increase of the condensation temperature, and the cut-off point is the same as the condensation temperature of 125°C. Therefore, the optimal value of the subcooling degree has a certain correlation with the condensation temperature T_{cond1} and the shunt coefficient x. To analyze the influence of x on the subcooling degree, the working condition in Table 2 is defined as working condition-1, and the x is set as working condition-2. In this case, the variation of the parameters of the subcooling degree corresponding to the optimal COP of the system is shown in Figure 8.

Figure 8 shows that the system subcooling under the two working conditions shows a trend of first increasing and then decreasing. However, the difference is that the variation of the value in working condition-1 is relatively smooth, while the value in working condition-2 drops abruptly when the condensation temperature is higher than 125°C. The reason for this trend is that the temperature lift of the system causes a change in the intermediate pressure, which indirectly changes the $R_{condition2}$ required by the process through the condensing temperature T_{cond2}. When the final condensation temperature T_{cond1} of the system is higher than 125°C, The SA algorithm will reduce the subcooling to meet the requirements of $R_{condition2}$.

Thermal performance analysis

The dual-flash compound circulation system is mainly used to convert hot water into steam. In addition to the COP of the system, the steam production and cooling capacity of the unit

TABLE 2 x **and** $\Delta T_{sub\text{-}co}$ **change trend with condensation temperature.**

Condensation temperature/°C	110	115	120	125	130	135	140
$\Delta T_{sub\text{-}co}$/°C	4.7	4.9	5.2	5.3	4.6	4.4	4.1
X	.96	.96	.96	.96	.95	.95	.95

TABLE 3 Technical parameters of the expansion valve to be used.

Parameter	Constant coefficient/-	Valve diameter/mm	Poppet angle/°	Open limit/mm
Value	.75	10	45	5

TABLE 4 Technical parameters of the proposed compressor.

Parameter	Empirical parameters/-	Coefficient of variation/-	Inspiratory volume/L	
Symbol	a_{Kim}	b_{Kim}	m_n	V_{Ins}
Value	1.025	.085	1.5	.2

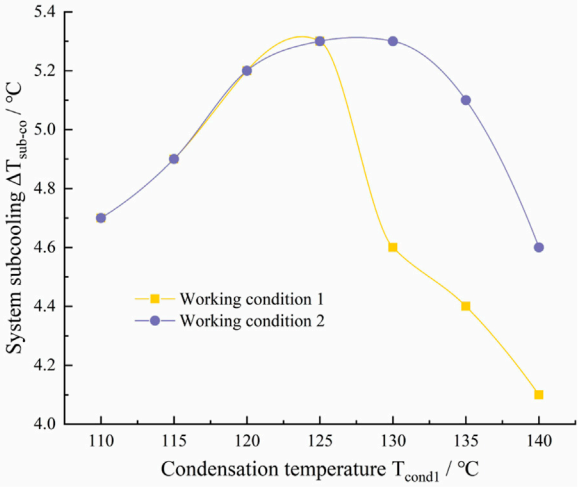


FIGURE 8
System subcooling changes under two working conditions.

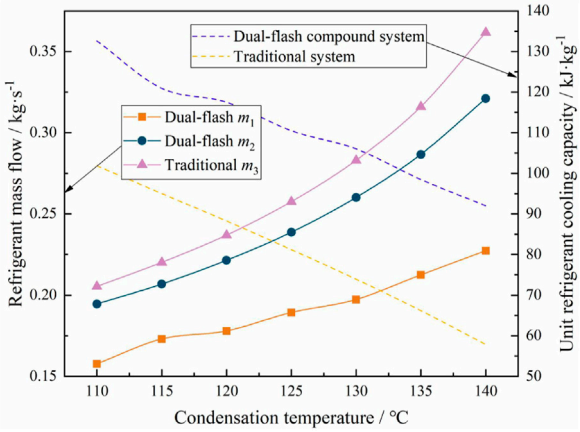


FIGURE 9
Refrigerant changes with condensing temperature under rated cooling capacity.

refrigerant will also be used as the standard parameters to evaluate the thermodynamic performance of the system.

To simplify the calculation, the influence of system refrigerant flow rate on the dual-flash compound circulation system was analyzed under the specified cooling capacity, and the mass flow relationship of refrigerant in each stage of compound circulation was analyzed is show in Figure 9.

When the cooling capacity and evaporation temperature are constant, the mass flow of refrigerant through the evaporator in the dual-flash compound circulation system and the traditional two-stage compression system are both m_1, and the mass flow of refrigerant discharged from the high-pressure compressor are m_2 and m_3, respectively. As shown in Figure 8, with the same cooling capacity, m_2 increases from .195 kg/s to 0.321 kg/s, and m_3 increases from .205 kg/s to .362 kg/s with the increase of condensation temperature T_{cond1}. The m_3 is 5.13%–12.8% higher than m_2. Under the same cooling capacity, the increase in refrigerant demand leads to an increase in cost, and the circulation process may cause a greater resistance loss. In addition, according to the mass flow relationship of refrigerant, under the action of equivalent circulating

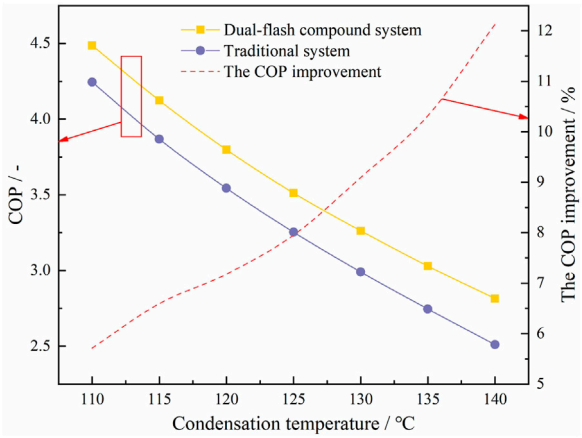


FIGURE 10
System COP changes at different condensation temperatures.

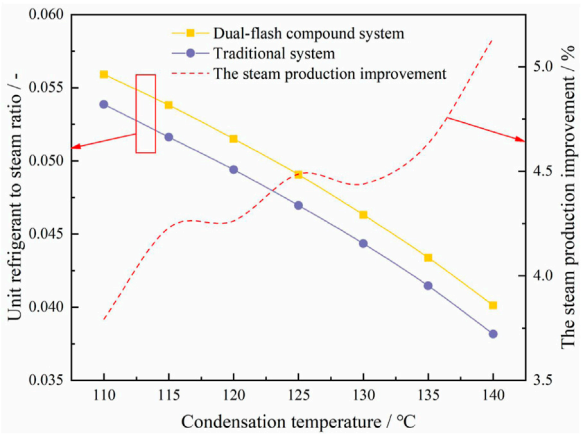


FIGURE 11
The change of unit steam volume of the system at different condensing temperatures.

refrigerant, the unit refrigerant refrigerating capacity of the traditional two-stage compression system and the dual-flash compound circulation system showed a downward trend with the increase of the condensation temperature. However, the unit refrigerating capacity of the dual-flash compound circulation system is 21.42%–37.17% higher than that of the traditional two-stage compression system. Therefore, facing the same production scale, the dual-flash compound circulation system can more fully use of the heat source of heat.

The thermodynamic models of the traditional two-stage compression system and the dual-flash compound circulation system were calculated. Under the same environment, the maximum COP of the two systems at different condensing temperatures could be obtained through parameter optimization by SA algorithm, as shown in Figure 10.

As shown in Figure 10, with the increase of the condensation temperature, the maximum COP of both the traditional two-stage compression system and the dual-flash compound circulation system showed a downward trend. Among them, the COP of the dual-flash compound circulation system decreased from 4.49 to 2.81, which was 5.71% ~ 12.13% higher than that of the traditional two-stage compression system in the same environment. This trend shows that the higher the condensation temperature required by the system, the more advantages the dual-flash compound circulation system has in heating.

To verify this advantage, the vapor production per unit refrigerant was compared between the two systems under different condensing temperature requirements, as shown in Figure 11.

Figure 11 shows the variation of steam production for the two systems with the same mass flow rate of refrigerant. When the condensing temperature of the system was within the range of 110–140°C, the mass flow rate of steam generated by the dual-flash compound circulation system was .0401–.0559 kg/s. Compared with the conventional two-stage compression system, the steam production is increased by 3.79%–5.14%. In addition, in the process of steam production, there are two fluctuation points when the system condensing temperature is 120°C and 130°C. This phenomenon is exact because the two fluctuation points of m_1 in Figure 10 correspond to the change of the system steam production caused by the change of m_1, which does not affect the optimization theory.

Exergy analysis

In the thermodynamic analysis of heat pump system, the COP can usually reflect the heating performance of the system at the economic benefits, but it cannot reflect the advanced nature of the thermal cycle of the system. The advancement of thermodynamic cycle of a system is usually expressed by thermodynamic perfection or exergy efficiency. Therefore, exergy model of the dual-flash compound circulation system was established, exergy changes of each component in the compound circulation of the system were analyzed, and the pros and cons of the system were further judged. As the change of condensation temperature, describes the exergy loss of each component in the system as shown in Figure 12.

Under the condition of the large temperature lift, the total exergy loss of the system increases from 2.646kW to 4.649 kW with the increase of condensing temperature, and the exergy loss of the compressor accounts for 50.12%–54.35%. It can be concluded that the compressor is the main component causing exergy loss in the system under the condition of high condensing temperature and the large temperature lift. As the

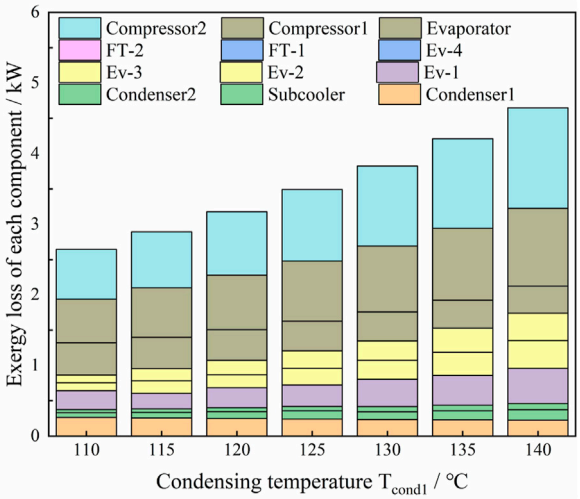


FIGURE 12
The exergy loss of each component of the system with the condensing temperature under the same heating capacity.

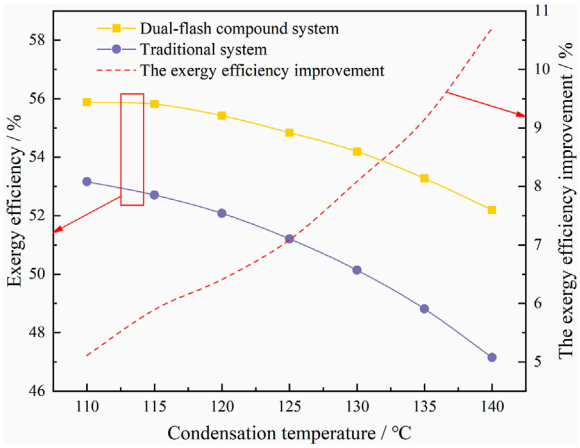


FIGURE 13
The trend of exergy efficiency of the system under the same cooling capacity.

condensing temperature increased in the system, the percentage of exergy loss of compressor fluctuated slightly, while the percentage of exergy loss of condensers -1, evaporator, Ev-2 and Ev-3 changed significantly.

When the condensing temperature of the system increased from 110°C to 140°C, the proportion of exergy loss of condenser-1 increased from 9.98% to 17.16%, and that of evaporator from 4.83% to 8.22%. The exergy loss of Ev-2 decreased from 4.23% to 4.11%, the Ev-3 decreased from 8.49% to 8.33%. The result showed that under the same heating capacity, with the increase of temperature T_{cond1}, the exergy loss generation process of the system gradually changed from pressure transformation process to heat transfer process. It is worth noting that the exergy loss ratio of Ev-1 did not fluctuate. Therefore, it can be concluded that the main reason for the change of exergy loss of Ev-2 and Ev-3 is the vapor injection process. In order to further understand the advancement of thermal cycle of the system, exergy efficiency of the system was calculated and compared with that of traditional two-stage compression cycle.

According to the exergy model of the system, the variation trend of exergy efficiency of the system at different condensing temperatures is obtained, as shown in Figure 13.

As can be seen, when the condensing temperature increased from 110°C to 140°C, the exergy efficiency of the dual-flash compound circulation system decreased from .5588 to .522. On the one hand, the overall exergy loss of the system increases with the increase of condensing temperature. On the other hand, the increase of pressure ratio increases the compression power of the compressor. However, the exergy efficiency of the traditional two-stage compression system decreased from .5316 to .4715 in the same environment, which was 5.11%–10.71% lower than that of the dual-flash compound circulation system. Thus, from the perspective of the advanced nature of thermal cycle, the dual-flash compound circulation system has obvious advantages.

Simulation verification

To verify the feasibility of the theoretical model, the system can be modeled and simulated by using Aspen Plus V9 software. Due to the large number of parameter variables and constraints in the theoretical model, in order to simplify the calculation process, the heat transfer iteration process of the simulation model is replaced by energy value through the relationship between the output heat in $R_{condition1}$ and $R_{condition2}$, and the system model is shown in Figure 14 below.

The simulation environment of the simulation system takes the evaporation temperature of 50 °C and the condensation temperature of 120 °C as an example. The physical property data of REFPROP are substituted into Aspen Plus, and QOUT1, QOUT2 and QOUT3 are set as the cold utilities. The boundary conditions are set respectively according to the theoretical calculation values. The pressure of thermal products is set as a fixed value of .169 MPa (saturated steam 115°C), and the inlet temperature of QOUT3 is 50°C. The outlet vapor fraction of QOUT1 is 1. QIN is the hot utilities, and the initial value of heat refers to the theoretical calculation of heat supply from the heat source.

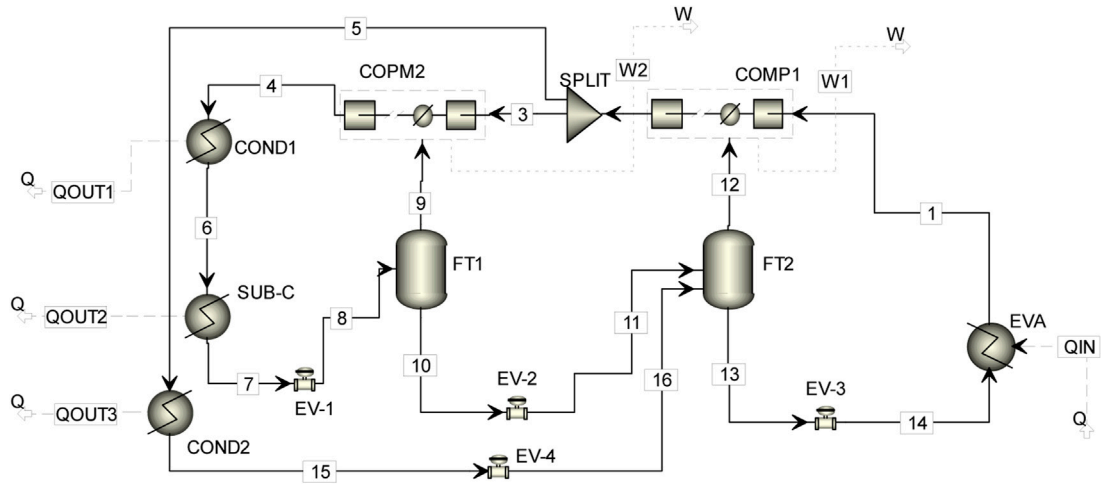


FIGURE 14
Simulation verification model.

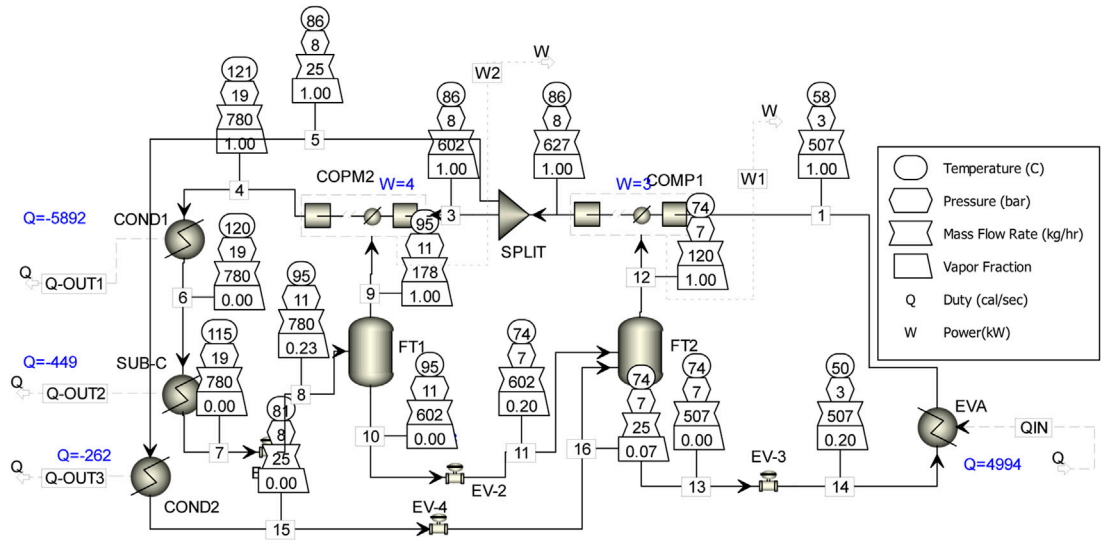


FIGURE 15
The state of each node of the model after the steady-state.

The technical parameters of the compressor and expansion valve in circulation are shown in the following Table 3 and Table 4.

The resistance loss along the pipeline is not considered. The theoretical optimized system parameters were used as the initial values of each pressure in the simulation model, where p_{cond1} = 1.9275 MPa, p_{FT1} = 1.126 MPa, p_{mid} = .8145 MPa, p_{FT2} = .6828 MPa, p_{eva} = .3442 MPa. The compressor is selected as a compressor module with double trays and vapor injection performance. In order to compare with the theoretical value, the isentropic efficiency of compressor-1 and compressor-2 is based on the theoretical value, which is .757, .831, .813 and .782, respectively. In the simulation model, the splitter is used to replace the shunt valve group, and the shunt coefficient x is set to .96. The initial refrigerant mass flow rate at node 6 in Figure 7 was set to .2236 kg/s. After setting the initial conditions, the parameters of the variable pressure

element (compressor, expansion valve) in the simulation model will be debugted until steady state. The state of each node is shown in Figure 15 below.

It can be seen from Figure 15 that under the cold utilities and the hot utilities and restrictions, the simulation system has no errors and warnings, indicating that the theoretical model is indeed feasible. According to the numerical relationship between the heating energy and compressor power consumption, the COP of the simulation system is 3.716, which is only 2.16% different from the theoretical calculation value. These errors are due to the system performance deviation caused by the inability to control the precise circulating refrigerant in the simulation system, and do not affect the feasibility and performance analysis of the system.

Conclusion

The dual-flash compound circulation system was used to realize the efficient heating process of hot water to steam. The system reduced the heat loss of key components such as compressor and expansion valve during heating process, and had better thermodynamic performance under the same working condition.

(1) The thermodynamic model of double flash compound cycle system is established, and the limiting conditions of the thermodynamic model are also established. When the temperature lift from 50°C to 120°C, the shunt coefficient in the system should not be less than .958.
(2) Combined with simulated annealing algorithm, a multi-objective parameter optimization model with COP as the optimization objective was established. Through optimization calculation, it is found that the pressure of FT-1 and FT-2 increases almost linearly with the increase of T_{cond1}. However, the shunt coefficient x and the system undercooling degree largely depend on the technological conditions of the system, especially the system undercooling degree.
(3) According to theoretical calculation, the COP of the dual-flash compound circulation system was in the range of 2.81–4.49 when the condensation temperature was in the range of 110–140°C, which was 5.71%–12.13% higher than that of the traditional two-stage compression system. The exergy efficiency of the system was 5.11%–10.71% higher than that of the dual-flash compound circulation system. The comprehensive analysis showed that the dual-flash compound circulation system had more advantages when facing the heating condition with high condensation temperature.
(4) In addition, Aspen was used to simulate the dual-flash compound circulation system. The results showed that the COP calculated by the simulation results was only 2.16% different from the theoretical value after the system was in steady state, which proved that the system had certain feasibility.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

ZH conceived and wrote the paper; ZY was responsible for scheme establishment and detail optimization; WL conducted model inspection and logical sorting; XJ provided funds and established the direction; HZ for data programming; ZG carried out image optimization; ZY does grammar check.

Funding

This study was financially supported by the National Natural Science Foundation of China (Grant Nos. 51876055, 51806060, and U1504524) and the Natural Science Foundation of Henan Province (Grant Nos. 182300410233) and the Shanghai Heimdallr Energy Saving Technology Co., Ltd. (05N14030820).

Conflict of interest

Author XJ was employed by the company Shanghai Heimdallr Energy Saving Technology Co., 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.

Publisher's note

References

Bai, X., and Liu, Y. (2016). Reliability analysis on civil engineering project based on integrated adaptive simulation annealing and gray correlation method. *Front. Struct. Civ. Eng.* 10 (4), 462–471. doi:10.1007/s11709-016-0361-y

Bamigbetan, O., Eikevik, T. M., Neks, P., Bantle, M., Schlemminger, C., Dallai, M., et al. (2018). "Experimental investigation of a hydrocarbon piston compressor for high temperature heat pumps," in 24th International Compressor Engineering Conference.

Bamigbetan, O., Eikevik, T. M., Neksa, P., Bantle, M., and Schlemminger, C. (2019). Experimental investigation of a prototype R-600 compressor for high temperature heat pump. *Energy (Oxford, U. K.)* 169 (15), 730–738. doi:10.1016/j.energy.2018.12.020

Cao, X.-Q., Yang, W.-W., Zhou, F., and He, Y.-L. (2014). Performance analysis of different high-temperature heat pump systems for low-grade waste heat recovery. *Appl. Therm. Eng.*

Chamoun, M., Rulliere, R., Haberschill, P., and Peureux, J. L. (2014). Experimental and numerical investigations of a new high temperature heat pump for industrial heat recovery using water as refrigerant. *Int. J. Refrig.* 44, 177–188.

China National Bureau of Statistics (2019). *China statistical Yearbook*.

Dai, Baomin, Liu, Xiao, Liu, Shengchun, Wang, D., Meng, C., Wang, Q., et al. (2020). Life cycle performance evaluation of cascade-heating high temperature heat pump system for waste heat utilization: Energy consumption, emissions and financial analyses. *Energy* 261, 125314. doi:10.1016/j.energy.2022.125314

Gao, J. T., Xu, Z. Y., and Wang, R. Z. (2021). An air-source hybrid absorption-compression heat pump with large temperature lift. *Appl. Energy*, 291.

Hao, Z., Yanting, Z., Jingyu, X., Lin, W., and Zheng, H. (2022). Performance analysis of IHX-based quasi-two-stage vapor compression heat pump system for high-temperature steam production. *ENERGY Technol.* 8, 200063.

Hoon, K. D., Sun-Ik, N., Woo, Y. J., Lee, J. H., and Kim, M. S. (2019). Experimental study on the performance of a steam generation heat pump with the internal heat exchanging effect. *Int. J. Refrig.* 108, 154–162. doi:10.1016/j.ijrefrig.2019.09.003

Kaida, T., Sakuraba, I., Hashimoto, K., and Hasegawa, H. (2015). Experimental performance evaluation of heat pump-based steam supply system. *IOP Conf. Ser. Mat. Sci. Eng.* 90 (1), 012076. doi:10.1088/1757-899x/90/1/012076

Kang, D., Sun-Ik, N., Kim, M. S., et al. (2017). Recent researches on steam generation heat pump system. *Int. J. Air-Cond. Ref.* 25 (4), 1730005–1730008. doi:10.1142/s2010132517300051

KOBELCO, TEPCO (2011). *Steam generator*.

Kryukova, A. E., Konarev, P. V., Volkov, V. V., and Asadchikov, V. (2019). Restoring silicasol structural parameters using gradient and simulation annealing optimization schemes from small-angle X-ray scattering data. *J. Mol. Liq.* 283, 221–224. doi:10.1016/j.molliq.2019.03.070

Liu, B., Gong, Y., Lu, Z, and Yao, Y (2015). Several working fluids for high temperature heat pump steam system. *Kezaisheng Nengyuan* 33 (12), 1755–1761.

Liu, B., Gong, Y., Lu, Z., Qu, Z., and Gao, Y. (2017). Performance analysis and optimization of a coupling sub-cooler heat pump steam system in a quasi two-stage compression cycle. *Huagong Jinzhan* 36 (07), 2360–2367.

Liu, C. C., Han, Wei, and Xue, X. D. (2022). Experimental investigation of a high-temperature heat pump for industrial steam production. *Appl. Energy*, 312.

Meggers, F., and Leibundgut, H. (2011). The potential of wastewater heat and exergy: Decentralized high-temperature recovery with a heat pump. *Energy Build.* 43 (4), 879–886. doi:10.1016/j.enbuild.2010.12.008

Wu, Xiaokun, Xing, Ziwen, He, Zhilong, Wang, X., and Chen, W. (2016). Performance evaluation of a capacity-regulated high temperature heat pump for waste heat recovery in dyeing industry. *Appl. Therm. Eng.* 93 (C), 1193–1201. doi:10.1016/j.applthermaleng.2015.10.075

Xu, Z. Y., Gao, J. T., Hu, Bin, and Wang, R. (2021). Multi-criterion comparison of compression and absorption heat pumps for ultra-low grade waste heat recovery. *Energy* 238, 121804. doi:10.1016/j.energy.2021.121804

Zhang, Y. T., Zhang, Hao, Wang, Lin, Jingyu, X., Lumeng, H., Jingkai, C., et al. (2022). Application and analysis of multi-stage flash vaporization process in steam production in high-temperature heat pump system with large temperature difference. *Int. J. Refrig.* 133, 123–132. doi:10.1016/j.ijrefrig.2021.09.035

Zheng, Nan (2017). *Research on novel two-stage heat pump with vapor expander and its key thermal process using zeotropic mixtures* Tianjin University

Nomenclature

COP Coefficient of performance (-)
h Enthalpy (kJ/kg)
s Entropy (kJ/(K·kg))
m Mass flow rate (kg/s)
T Temperature (K)
p pressure(MPa)
E Exergy (kW)
Q Heat transfer (kW)
W Power consumption (kW)
η Efficiency (-)
a, b Vapor ratio (-)
T_{Amb} Ambient temperature (K)
x The split coefficient (-)

Subscripts

Cond Condensation
Eva Evaporation
Ev Expansion valve
Comp Compressor
FT Flash tank
sub-co Subcooler
sink Thermal product
c Heat source
in Intake
out Exhaust
mid Middle-pressure position
is Isentropic state

Abbreviations

HTHP High-temperature heat pump
IHX Internal heat exchanger
SA Simulated annealing algorithm

Other symbols

m_1 Compressor-1 refrigerant displacement
m_2 Compressor-2 refrigerant displacement
$R_{condition1}$ Cyclic requirement condition one
$R_{condition2}$ Cyclic requirement condition two
Q_1 Heat transfer of the Condenser-1
Q_2 Heat transfer of the Subcooler
Q_3 Heat transfer of the Condenser-2
cond1 Condenser-1
cond2 Condenser-2
a_{Kim}, b_{Kim}, m_n, V_{Ins} SA algorithm parameters

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 09 January 2023
DOI 10.3389/fenrg.2022.1073271

OPEN ACCESS

EDITED BY
Yongming Han,
Beijing University of Chemical Technology, China

REVIEWED BY
Zhiying Wu,
Hong Kong Institute of Science and Innovation (CAS), Hong Kong, Hong Kong, SAR China
Huadong Mo,
University of New South Wales, Australia
Yuqiu Chen,
Technical University of Denmark, Denmark

*CORRESPONDENCE
Yiqi Liu,
aulyq@scut.edu.cn

SPECIALTY SECTION
This article was submitted to Process and Energy Systems Engineering, a section of the journal Frontiers in Energy Research

RECEIVED 18 October 2022
ACCEPTED 11 November 2022
PUBLISHED 09 January 2023

CITATION
Lu Y, Fang G, Huang D, Cai B, Chen H and Liu Y (2023), Shaping energy cost management in process industries through clustering and soft sensors. *Front. Energy Res.* 10:1073271. doi: 10.3389/fenrg.2022.1073271

Shaping energy cost management in process industries through clustering and soft sensors

Yu Lu[1], Gang Fang[1,2], Daoping Huang[1,2], Baoping Cai[3], Hongtian Chen[4] and Yiqi Liu[1,2]*

[1]Key Laboratory of Autonomous Systems and Networked Control, Ministry of Education, The School of Automation Science and Engineering, South China University of Technology, Guangzhou, China, [2]Unmanned Aerial Vehicle Systems Engineering Technology Research Center of Guangdong, The School of Automation Science and Engineering, South China University of Technology, Guangzhou, China, [3]College of Mechanical and Electronic Engineering, China University of Petroleum, Qingdao, China, [4]Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Canada

With the ever-increasing growth of energy demand and costs, process monitoring of operational costs is of great importance for process industries. In this light, both financial budget management and local operational optimization supposed to be guaranteed properly. To achieve this goal, a support vector machine recursive feature elimination (SVM-RFE) method together with clustering algorithm was developed to extract features while serving as importance measurements of each input variable for the sequential prediction model construction. Then, the four variants of autoregressive and moving average (ARMA), i.e., ARMA with exogenous input (ARMAX) based on recursive least squares algorithm (RLS), ARMAX based on recursive extended least squares algorithm (RELS), nonlinear auto-regressive neural network (NARNN) and nonlinear auto-regressive neural network with exogenous input (NARXNN), were applied, respectively, to predict the costs incurred in the daily production for process industries. The methods were validated in the Benchmark Simulation Model No.2-P (BSM2-P) and a practical data set about steel industry energy consumption from an open access database (University of California, Irvine (UCI)), respectively. The nonlinear model, NARXNN, was validated to achieve better performance in terms of mean square error (MSE) and correlation coefficient (R), when used for multi-step prediction of the aforementioned datasets with strong nonlinear and coupled characteristics.

KEYWORDS

industry, operational costs prediction, ARMAX, NARNN, NARXNN, SVM-RFE combined with clustering algorithm

1 Introduction

In recent decades, smart industrial concept or industry 4.0 has gained popularity as an initiative to upgrade traditional manufacture to an intelligent facility with the help of artificial intelligence and machine learning. However, smart concept is always focusing on quality control through instrumentations and controllers, without sufficiently focusing on energy consumption management or operational costs reduction prediction (Ansari et al., 2011). Operational costs reduction, such as minimizing dosage costs, optimizing energy consumption, subsequent optimizing control or operational strategies, can intuitively promote green production of process industries, thereby helping enterprises or sectors achieve sustainable manufacture. With the globalization, the continuous growth trend in energy consumption received significant attentions. As the largest energy end-use sector, industrial currently accounts for nearly 40% of total global final energy consumption (International Energy Agency, 2021). Moreover, energy consumption accounts for a large proportion of total costs in most industrial processes (Han et al., 2018). Excessive energy consumption usually implies more environmental pollutions and more production costs due to the environmental regulations. Therefore, given the potentials to improve industrial energy efficiency, substantial research on energy-efficiency indicators has been proposed to support energy-intensive enterprises and governments to assess energy consumptions and optimize management (Chan et al., 2014; Li and Tao, 2017).

Specifically, constructing energy consumption or operational costs prediction models can help and support decision-making about costs management properly. The motivation behind establishing a predictive model is essentially to make a model able to reflect and mimic the true system characteristics as closer as possible. In general, two types of approaches are typically used for modeling. One is mechanistic model, also known as the white-box approach, in which the mechanism of the system is completely clear and the model construction generally depends on the specific physical, chemical, biological and other behaviors of a process. Such a model is intuitively explainable but difficult to be generalized to other fields. Jia et al. (2018) established an energy consumption model based on motion-study for activities related to equipment and operators, showing the effectiveness of the approach in a case study. Also, Altıntas et al. (2016) combined mechanistic and empirical models to optimize the machine operations in a milling process. This model was used to estimate the theoretical energy consumption in the milling process of prismatic parts with satisfactory prediction accuracy. In fact, to construct mechanism-based simulation models, a certain number of input data associated with the predicted targets are required, and then assumptions about the distribution of corresponding parameters or features related to these inputs are usually made relying on the prior knowledge (Hsu, 2015). However, most industrial processes are difficult to derive a specific mechanistic model, because of extremely nonlinear, coupled, multivariate characteristics and even combination of physical, chemical and biological reactions.

The data-driven modeling approach, called the black-box approach, is another way to address the above issues. The data using for prediction validation usually have similar patterns to those exhibited in the historical data. Data-driven methods has gained popularity since the past decades. This is mainly because data-driven methods can achieve better performance without process mechanisms compared to mechanistic models if the sufficient historical data sets are collected (Wei et al., 2018). With respect to the different types of data, data-driven models can be generally classified into linear and nonlinear forecasting models (Xiao et al., 2018). The ARMA model and its variants, as typical linear models, are one of the most popular methods in time series forecasting, especially for linear and stationary time series scenarios. Even though non-stationary data can be solved by resorting to de-seasoning and de-trending strategies, ARMA could still fail for most of cases (Juberias et al., 1999). Non-linearity in data can be approached by resorting to the nonlinear ARMA properly (Kun and Weibing, 2021). An autoregressive-based time varying model was developed to predict electricity short-term demand, while the performance of the original model depends a lot on the updated coefficients (Vu et al., 2017). The aforementioned variants mainly focused on autoregression and took other correlated variables unusefulness for granted. Fang and Lahdelma (2016) applied the ARMA model to predict heating demand by combining weather variations, social components and other exogenous factors, and the results showed that the proposed method outperformed the model only considering weather components. In the actual industrial process, the predicted targets are influenced by other exogenous variables besides themselves. Therefore, ARX and ARMAX were proposed to improve ARMA model by incorporating the impacts of exogenous variables into the time series model, and have been studied by academic communities such as meteorology, finance, etc (Huang and Jane, 2009; Silva et al., 2022). Recently, with the rapid development of artificial intelligence, artificial neural network techniques were broadly used to tackle with nonlinear problems (Liu et al., 2020; Deng et al., 2021). They perform much better than linear time series model especially when the input data is kept current or the model functions at more than one-step-ahead prediction (De Gooijer and Hyndman, 2006). Therefore, the neural network model is dominant data-driven model that has been widely applied in modeling and predicting (Han et al., 2018). In order to cater for working in various circumstances and conditions, diverse neural network structures and algorithms were continuously developed. The network type and the optimization algorithm of undetermined parameters also need to be selected appropriately for different purposes (Car-Pusic et al., 2020). Shi et al. (2021) designed a model based on convolutional

neural networks to predict coal and electricity consumption simultaneously, and this model also eliminated the negative effects of the coupling between variables. Kahraman et al. (2021) proposed a data-driven method based on the deep neural network, which provided a highly accurate prediction performance for energy consumption of industry machines. The NARXNN adopted in this study is a neural network that combines autoregression and exogenous input series, and this model has the additional advantage of handling nonlinear time series compared to the ARMAX model.

In general, models with exogenous inputs outperform those using autoregressive methods directly, especially for real industrial processes. However, inappropriate input selection may lead to many problems such as overfitting or collinearity (Wu et al., 2020; Liu et al., 2021). Therefore, the selection of features is a critical step before modeling. Principal component analysis (PCA) is one of the most extensively used methods for feature reconstruction, which is able to refine new features by mapping the original high-dimensional vector space onto a new low-dimensional space. However, the use of this method requires to ensure that the collected data must follow Gaussian distributions and also the new features generated by PCA are difficult to interpret. Feature ranking methodologies, as another type of feature selection means, are mainly composed of filter-based, wrapper-based, and embedded methods. These methods rank the importance of each individual feature according to the scores of diverse feature subsets and are effective in interpretability problems (Guyon and Elisseeff, 2003). SVM-RFE, as an embedded method based on backward elimination, was firstly proposed by Guyon et al. (2002) for feature ranking of binary classification. In this study, based on this approach together with clustering algorithm, the feature importance of continuous labels can be derived, and then exogenous inputs can be chosen.

The main objective of this research is to develop energy consumption and operational costs prediction models by using variants of ARMA models to optimize management of process industries. The accuracy of the methodologies was validated in two case studies. Different from the traditional ways for energy prediction, the proposed methods are able to make multiple steps ahead prediction, thus supporting energy consumption and operational costs analysis over a short-term period. This will, in turn, facilitate the controller manipulations and management behaviors in advance if the demand from markets changes. Also, due to the collaboration with SVM-RFE in the proposed method, useful features can be well refined and interpreted by the importance measurement.

The rest of the paper is organized as follows. The methods of predictive modeling and input feature selection are briefly introduced to provide the basic knowledge in Section 2. The dataset performance, prediction performance analysis and discussion of the two cases are presented in Section 3. The conclusions are finally drawn in Section 4.

2 Methods and materials

2.1 The autoregressive and moving average with exogenous input model

The ARMA model is usually suitable for short-term forecasts of time series data, and is widely applied in business, economics, engineering and other areas (Box et al., 2008). The ARMA is usually formulated as the following equations:

$$\rho(B)y_t = \tau(B)a_t, \tag{1}$$

Where,

$$\rho(B) = 1 - \rho_1 B - \rho_2 B^2 - \cdots - \rho_p B^p,$$
$$\tau(B) = \tau_0 + \tau_1 B + \tau_2 B^2 + \cdots + \tau_q B^q,$$

y_t is the output of the model at time t, and a_t is random shocks, such as white noise of Gaussian distribution. B is defined as the backward shift operator, i.e., $y_{t-1} = By_t$. When $\rho(B) = 1$, the ARMA model can be degenerated into the MA model. Similarly, when $\tau(B) = 1$, the model will become AR model. It is worth noting that AR and MA models are both special cases of an ARMA model. $[\rho_1 \cdots \rho_p, \tau_0, \tau_1 \cdots \tau_q]$ represents the weight value of the items corresponding to $[y_{t-1} \cdots y_{t-p}, a_t, a_{t-1} \cdots a_{t-q}]$. Parameters, $[\rho_1 \cdots \rho_p, \tau_0, \tau_1 \cdots \tau_q]$, are unknown and need to be estimated by using the collected data. In this paper, RLS and RELS algorithms are used for parameter identifications. The RLS is used to minimize the cost function as follows:

$$J(\theta) = \sum_{i=1}^{t} \left[y_i - {\varphi_i}^T \theta \right]^2, \tag{2}$$

Where,

$$\varphi_i = \left[y_{i-1}, \cdots, y_{i-p}, a_i, a_{i-1}, \cdots, a_{i-q} \right]^T,$$
$$\theta = \left[\rho_1, \cdots, \rho_p, \tau_0, \tau_1, \cdots, \tau_q \right]^T,$$

The RLS algorithm for estimating parameter θ can be expressed as (Ding, 2010):

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \varphi_t \left[y_t - {\varphi_t}^T \hat{\theta}_{t-1} \right], \tag{3}$$

$$P_t = P_{t-1} + \frac{P_{t-1} \varphi_t {\varphi_t}^T P_{t-1}}{1 + {\varphi_t}^T P_{t-1} \varphi_t}, \tag{4}$$

Where $P_0 = p_0 I_{p+q+1}$, $P_i (i = 1, \cdots, t) \in \mathbb{R}^{(p+q+1)\times(p+q+1)}$ are the covariance matrix and I_{p+q+1} is an identity matrix of the order $p + q + 1$. p_0 is assumed as a large positive number, e.g., $p_0 = 10^6$. $\hat{\theta}_t$ is the estimated value of θ at time t. On the basis of RLS, RELS additionally take into account innovation $[e_{t-1}, \cdots, e_{t-r}]$. The innovation in this study indicates the difference between the real value and the predictive, where $e = y - \varphi^T \theta$. The general form of ARMA-RELS can be extended as shown in Eq. 5.

$$\rho(B)y_t = \tau(B)a_t + c(B)e_t, \tag{5}$$

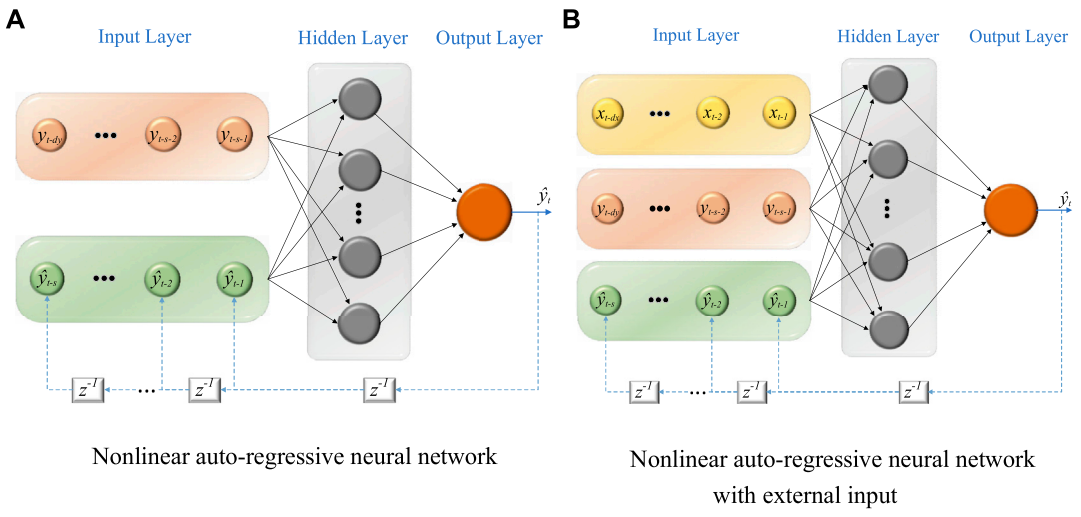


FIGURE 1
Two kinds of neural network structure, where z^{-1} is the unit time delay. **(A)** Nonlinear auto-regressive neural network. **(B)** Nonlinear auto-regressive neural network with external input.

Where,

$$c(B) = c_1 B + c_2 B^2 + \cdots + c_r B^r,$$

The sets of other parameters also need to be updated,

$$\varphi_i = \left[y_{i-1}, \cdots, y_{i-p}, a_i, a_{i-1}, \cdots, a_{i-q}, e_{i-1}, \cdots, e_{i-r}\right]^T,$$
$$\theta = \left[\rho_1, \cdots, \rho_p, \tau_0, \tau_1, \cdots, \tau_q, c_1, \cdots, c_r\right]^T,$$

$P_0 = p_0 I_{p+q+r+1}$, $P_i\,(i = 1, \cdots, t) \in \mathbb{R}^{(p+q+r+1)\times(p+q+r+1)}$, while the recursive form that is used to estimate the parameter θ remains unchanged.

However, aforementioned ARMA models omit the interference of exogenous variables on the prediction results. The ARMAX method was proposed to tackle this nuisance. The ARMAX model takes into account not only the effects from the historical series of output itself, but additionally the effects of exogenous inputs. The general expressions for ARMAX-RLS and ARMAX-RELS can be presented as Eqs 6, 7, respectively.

$$\rho(B) y_t = \tau(B) a_t + \sigma(B) u_t, \tag{6}$$
$$\rho(B) y_t = \tau(B) a_t + \sigma(B) u_t + c(B) e_t, \tag{7}$$

Where,

$$\sigma(B) = \sigma_0 + \sigma_1 B + \sigma_2 B^2 + \cdots + \sigma_j B^j,$$

u_t is the exogenous inputs set. This set of exogenous inputs is possibly multidimensional, whose dimensionality depends on the number of input variables selected. The method used in this research to determine the input variables is described in Section 2.3. On the other hand, the formulations of the two parameter estimation algorithms, RLS and RELS, for the ARMAX model are basically consistent with the ARMA model, respectively. However, mainly due to the effect of the additional u_t, some parameters are updated accordingly on the original basis as follows:

$$\varphi_i = \left[y_{i-1}, \cdots, y_{i-p}, a_i, a_{i-1}, \cdots, a_{i-q}, u_i, u_{i-1}, \cdots, u_{i-j}, \right]^T,$$
$$\theta = \left[\rho_1, \cdots, \rho_p, \tau_0, \tau_1, \cdots, \tau_q, \sigma_0, \sigma_1, \cdots, \sigma_j\right]^T,$$

$P_0 = p_0 I_{p+q+j+2}$, $P_i\,(i = 1, \cdots, t) \in \mathbb{R}^{(p+q+j+2)\times(p+q+j+2)}$ for RLS and,

$$\varphi_i = \left[y_{i-1}, \cdots, y_{i-p}, a_i, a_{i-1}, \cdots, a_{i-q}, e_{i-1}, \cdots, e_{i-r}, u_i, u_{i-1}, \cdots, u_{i-j}, \right]^T,$$

$$\theta = \left[\rho_1, \cdots, \rho_p, \tau_0, \tau_1, \cdots, \tau_q, c_1, \cdots, c_r, \sigma_0, \sigma_1, \cdots, \sigma_j\right]^T,$$

$P_0 = p_0 I_{p+q+r+j+2}$, $P_i\,(i = 1, \cdots, t) \in \mathbb{R}^{(p+q+r+j+2)\times(p+q+r+j+2)}$ for RELS. Then, p, q, r and j are calculated by the Akaike Information Criterion method as shown in Eq. 8.

$$AIC(p, q, r, j) = -2\ln(L) + 2Q, \tag{8}$$

Where Q is the number of parameters, L is the likelihood function. The optimal order of model is the $\{p, q, r, j\}$ value satisfying the minimum AIC. Accordingly, due to the introduction of lag r, iterative k-step ahead prediction can be formulated as Eqs 9, 10.

$$\hat{\rho}(B)\hat{y}_{t+k|t} = \hat{\tau}(B) a_{t+k|t} + \hat{\sigma}(B) u_{t+k|t} + \sum_{i=k}^{r} \hat{c}_i e_{t+k-i}, 1 \le k \le r, \tag{9}$$
$$\hat{\rho}(B)\hat{y}_{t+k|t} = \hat{\tau}(B) a_{t+k|t} + \hat{\sigma}(B) u_{t+k|t}, k \ge r, \tag{10}$$

Where $[\hat{\rho}, \hat{\tau}, \hat{\sigma}, \hat{c}_i, \hat{y}_{t+k|t}]$ are the corresponding estimated values of $[\rho, \tau, \sigma, c_i, y_{t+k}]$. If $k \le 0$, $\hat{y}_{t+k|t} = y_{t+k}$, meaning that the

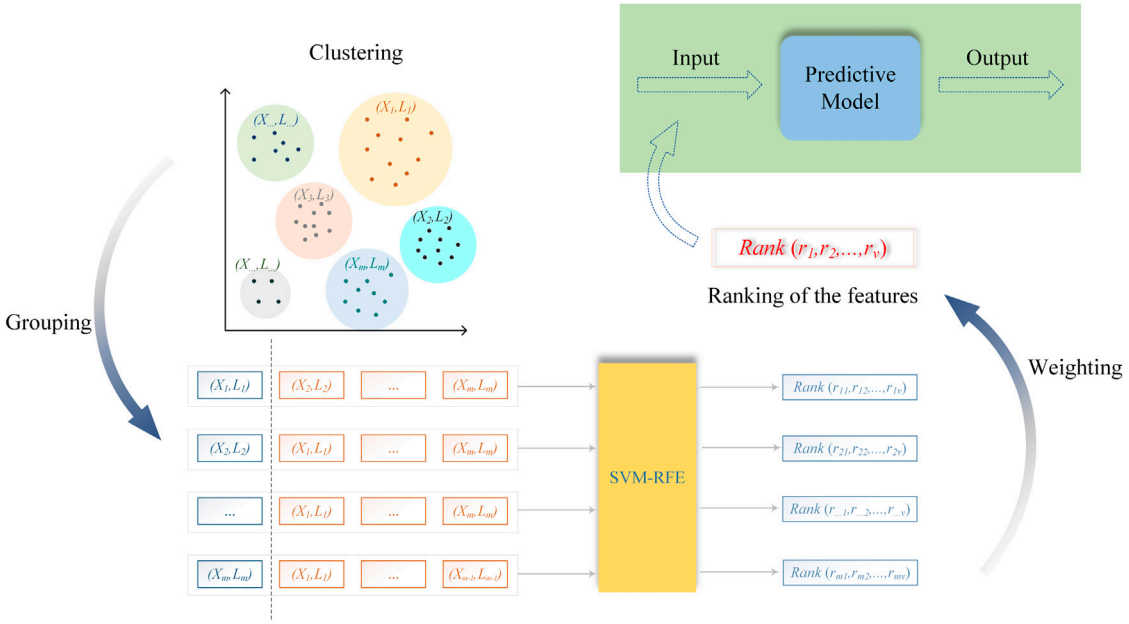


FIGURE 2
SVM-RFE combined with clustering algorithm for sequential feature ranking.

estimated output of the prediction model is equal to the real value.

Both RLS and RELS algorithms are simple but powerful to estimate unknown parameters without needing to calculate matrix inversion during iterative learning. These make them suitable for online model identification. RELS is actually a direct extension of RLS, aiming to reduce the influence of colored noise by adding residuals to the information vector φ and the parameter vector θ. Compared to the standard least squares method, RLS and RELS algorithm improve the identification performance of time series model at the expense of higher computational complexity.

2.2 The nonlinear auto-regressive neural network model

Time series data derived from real industrial processes usually exhibit strong nonlinearity and high dynamics, which renders the monitoring of such data unsuitable if using linear models. Therefore, nonlinear methods based on neural networks are highly recommended for modeling such dataset. The standard NARNN is formulated as follows:

$$\hat{y}_t = \mathcal{N}\left(y_{t-1}, y_{t-2}, \cdots, y_{t-d_y}\right), \tag{11}$$

Where $\hat{y}_t$ is the estimation of the output by a specific neural network at the t moment, $[y_{t-1}, y_{t-2}, \cdots, y_{t-d_y}]$ is the time series dataset, d_y is the maximum output-memory order and $\mathcal{N}(\cdot)$ means a specific neural network. The distinction between NARNN and ordinary neural networks for multi-step prediction is that several observed data have to be replaced by the estimate of the network, so that Eq. 11 can also be reformulated as follows:

$$\hat{y}_t = \mathcal{N}\left(\hat{y}_{t-1}, \hat{y}_{t-2}, \cdots, \hat{y}_{t-s}, y_{t-s-1}, \cdots, y_{t-d_y}\right), \tag{12}$$

Where $[\hat{y}_{t-1}, \hat{y}_{t-2}, \cdots, \hat{y}_{t-s}]$ are estimates of the output over the time period from $t-s$ to $t-1$, respectively, s is the number of delay steps for autoregression, and $[y_{t-s-1}, \cdots, y_{t-d_y}]$ are the observations from time $t-d_y$ to time $t-s-1$. The structure of NARNN is presented as Figure 1A.

The NARNN model is primarily concerned with historical series of the target variables as shown in Eq. 12. The information carried by exogenous inflow data is ignored in this modeling process, and then NARXNN model is proposed to make use of this information. The NARXNN can be formulated as follows:

$$\hat{y}_t = \mathcal{N}\left(\begin{matrix}\hat{y}_{t-1}, \hat{y}_{t-2}, \cdots, \hat{y}_{t-s}, y_{t-s-1}, \\ \cdots, y_{t-d_y}, x_{t-1}, \cdots, x_{t-d_x}\end{matrix}\right), \tag{13}$$

Where $[x_{t-1}, \cdots, x_{t-d_x}]$ is a matrix consisting of exogenous input variables, the dimension of x depends on the quantity of exogenous

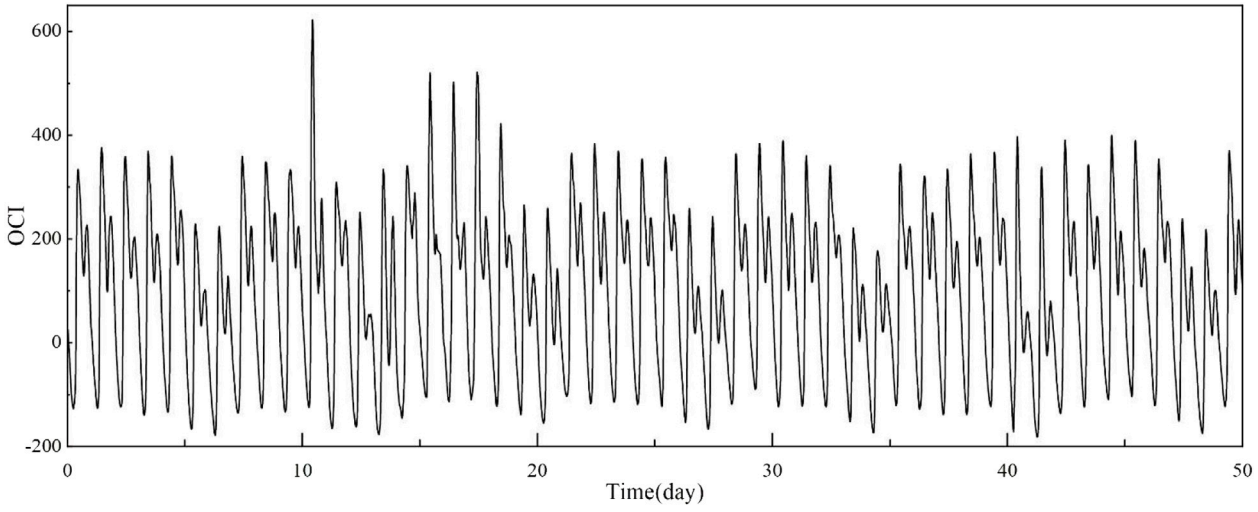


FIGURE 3
Operational cost index (OCI) over 50 days.

TABLE 1 Exogenous input variables related to OCI and descriptions.

Variables	Description	Units
t	Time of simulation	day
S_F	Fermentable substrate	g COD.m^{-3}
S_I	Soluble inert organic matter	g COD.m^{-3}
S_{PO_4}	Phosphate	g.m^{-3}
X_I	Particulate inert organic matter	g COD.m^{-3}
Q_{in}	Influent flow rate	m^3.day^{-1}
S_{N_a}	Sodium	g.m^{-3}

*COD, chemical oxygen demand.

input variables, and d_x is the maximum delay index of exogenous input variables. The structure of NARXNN is different from the NARNN slightly, mainly with the addition of several extra inputs. The structure of NARXNN is presented as Figure 1B.

The two neural networks, NARNN and NARXNN, update weights in each layer by using the Bayesian regularization backpropagation algorithm (MacKay, 1992). Training samples are shown as following set, $D = \{(\bar{x}_1, \bar{y}_1), (\bar{x}_2, \bar{y}_2), \cdots, (\bar{x}_N, \bar{y}_N)\}$, where $\bar{x}_i (i = 1, 2, \cdots, N)$ and $\bar{y}_i (i = 1, 2, \cdots, N)$ represent the input and output of the neural network, respectively. Given a neural network, called $\mathcal{M}$, let $g(x; w, \mathcal{M})$ be the response of network $\mathcal{M}$ with respect to the input x, and w denotes the weight of network. The optimal parameters can be achieved by minimizing the quadratic cost function:

$$E_D = \sum_{i=1}^{N} [g(\bar{x}_i; w, \mathcal{M}) - \bar{y}_i]^2. \tag{14}$$

The objective function is extended from $F = E_D$ to $F = \alpha E_D + \beta E_w$ to prevent the overfitting. The regularization term E_w is denoted as:

$$E_w = \sum_n \omega_n^2. \tag{15}$$

Note that α and β are unknown parameters of the objective function F. When α is larger, the accuracy of the model to the training samples is enhanced, and similarly when β is larger, the generalization ability of the model is enhanced. The Bayesian regularization generally treats the network weights as random variables and the detailed methods for estimating the values of weights w, α and β can be found in (Dan Foresee and Hagan, 1997).

NARNN and NARXNN are variants of ARMA together with neural network, combing both the dynamic and recurrent properties. Both methods do not require strict stationarity of the target time series. On the other hand, it is should be noted that NARXNN needs more reasonable computational cost (Cadenas et al., 2016).

2.3 The selection of features

Feature selection is of vital importance to improve the performance of the model, especially whose predictions depend on a number of extrinsic inputs to some extent. Excellent choices of inputs not only help to provide accurate

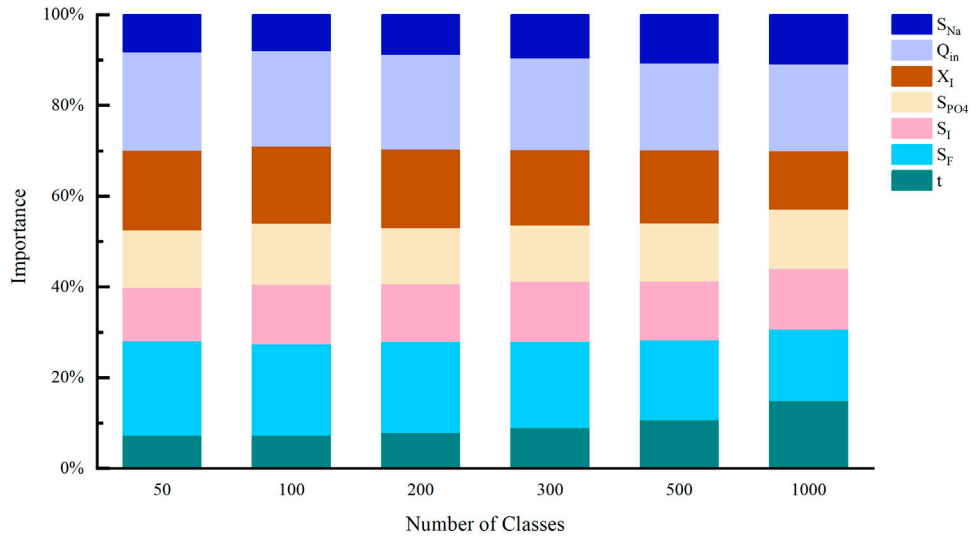


FIGURE 4
Importance of each variable to OCI with different number of clusters.

results, but also speed up calculations and reduce the number of sensor installations, all of which lead to operational cost savings. The SVM-RFE combined with clustering method proposed in this study ranks the features of the continuous processes based on backward elimination.

The first thing worth noting is that the support vector machine classifier achieves the distinction between two classes by searching for the optimal hyperplane in a high-dimensional space (Rakotomamonjy, 2003). For the binary classification problem with training data set $\{\boldsymbol{X}, \boldsymbol{Y}\}$, where $\boldsymbol{X} \in \mathbb{R}^n$ are the features and $\boldsymbol{Y} \in \{-1, 1\}$, there exists a hyperplane or decision function of the following form.

$$f(\boldsymbol{X}) = \langle \boldsymbol{w}, \Phi(\boldsymbol{X}) \rangle + b, \tag{16}$$

Where $\Phi(\boldsymbol{X})$ refers to the mapping relationship from features $\boldsymbol{X}$ to the high dimensional space. The parameters $(\boldsymbol{w}, b)$ are determined by minimizing the weights and the distance of each misclassified data to the hyperplane, before which the features $\boldsymbol{X}$ need to be normalized. The optimization problem can be written as:

$$\begin{gathered} \min_{w,\xi} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{k=1}^{m} \xi_k^2, \\ y_k f(\boldsymbol{x}_k) \geq 1 - \xi_k \\ \text{s.t } y_k \in \boldsymbol{Y},\ \boldsymbol{x}_k \in \boldsymbol{X} \text{ for } \forall k \end{gathered} \tag{17}$$

Where C is used as a penalty factor to weigh the importance of misclassification. SVM-RFE compares the impact of different remaining subsets on the classification by backward elimination of features, with the aim of preserving the subset of features that are most beneficial to the classification. The ranking of features is achieved through multiple iterations of elimination in the above work until the remaining feature subset is empty, and the criterion of elimination at each step for a given feature i can be expressed as:

$$R_c(i) = \left|\nabla \left\|\boldsymbol{w}^{(i)}\right\|^2\right|, \tag{18}$$

Where ∇ denotes the weight difference between the previous subset and the one whose i^{th} feature is eliminated. The feature that minimizes R_c will be removed after one round of loops, which means that the remaining feature subset has the least difference in classification performance from the feature subset containing the removed feature. It is worth mentioning that SVM-RFE was widely used in feature selection for binary classification problems, while it was rarely used in continuous classification problems.

In practical industrial processes, the variation in feature values often leads to continuous variation in the output. This begs a question that needs to be addressed. When each output point is treated directly as a separate label, the volume of the feature set corresponding to each label is so small that each feature set does not have the ability to characterize a specific label. Therefore, such continuous processes cannot be directly classified as a multi-label feature classification problem. In order to rank the features for this type of data, this study proposed SVM-RFE combined with clustering algorithm, as shown in Figure 2.

The successive outputs are first clustered to form new classes, and thus similar outputs can be grouped into homogeneous classes to enhance the differences between the new classes, e.g., $\{y_1, y_2, \cdots, y_n\} \rightarrow \{L_1, L_2, \cdots, L_m\}$, $n \gg m$. The features corresponding to the same type of label are also merged into the same group, e.g., $\{x_1, x_2, \cdots, x_n\} \rightarrow \{X_1, X_2, \cdots, X_m\}$, $n \gg m$. In this way, the original problem is successfully transformed into a

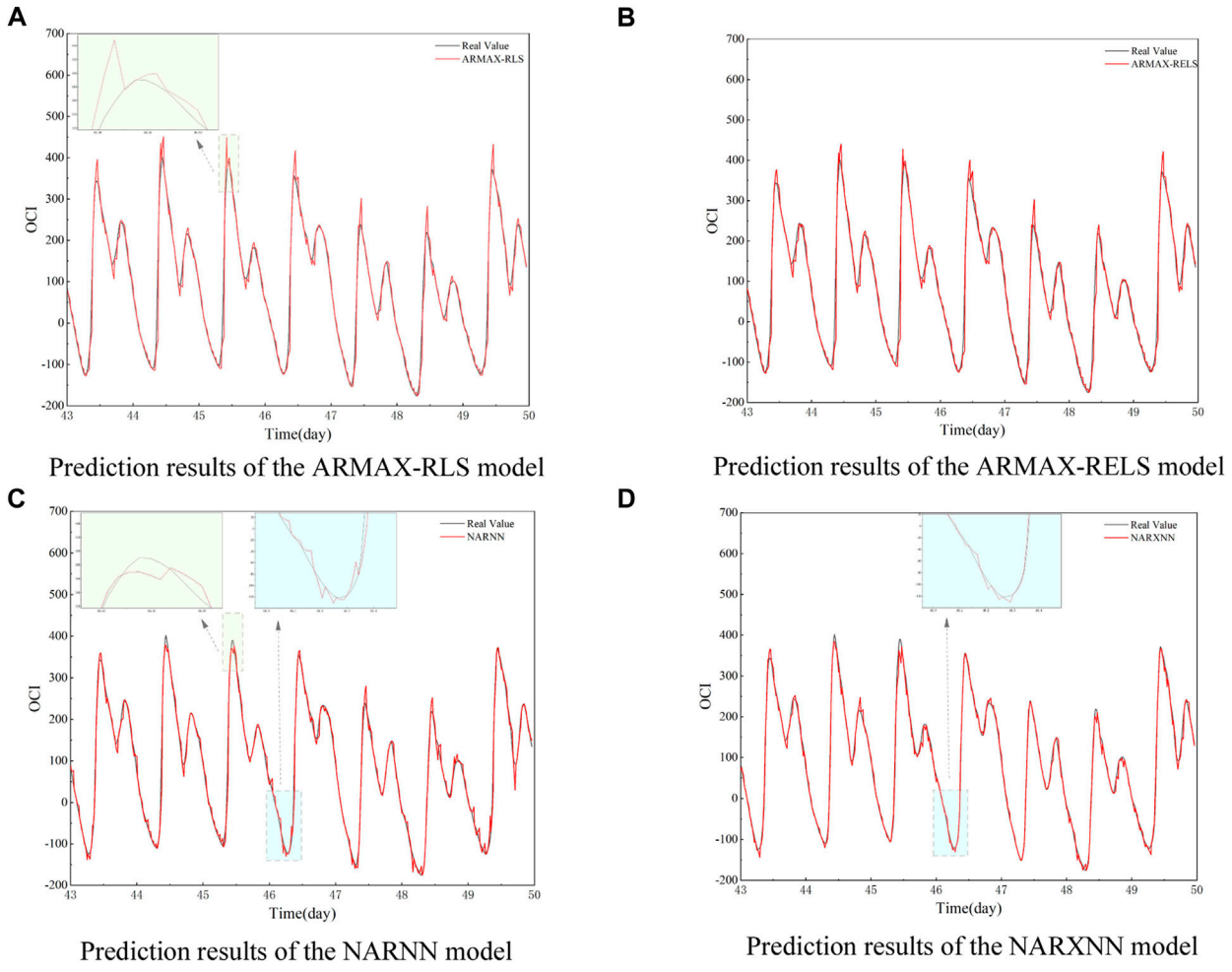


FIGURE 5
Prediction results of OCI for the last 7 days over four steps ahead. **(A)** Prediction results of the ARMAX-RLS model. **(B)** Prediction results of the ARMAX-RELS model. **(C)** Prediction results of the NARNN model. **(D)** Prediction results of the NARXNN model

multi-label feature classification problem. When SVM is used for multi-classification problems, it is usually transformed into a series of binary classification problems that are handled separately and then summarized for the final result. The new data set $\{(X_1, L_1), (X_2, L_2), \cdots, (X_m, L_m)\}$ is partitioned using the so-called the one-versus-all method, and subsequently the features are individually ranked using SVM-RFE. The final summary of the ranking results for each group is $\mathcal{R}_{m\times v}$ as shown in Eq. 19.

$$\mathcal{R} = \begin{pmatrix} \mathcal{R}_1 \\ \vdots \\ \mathcal{R}_m \end{pmatrix} = \begin{pmatrix} r_{11} & \cdots & r_{1v} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mv} \end{pmatrix}, \tag{19}$$

Where $\mathcal{R}_i\,(i = 1, \cdots, m)$ represents the ranking of the features that distinguish $\{X_i, L_i\}$ from the rest of dataset using SVM-RFE and the order, r_{i1} to r_{iv}, is decreasing according to the importance of the v optional features. For the ranking matrix $\mathcal{R}$, this research considers that the importance index of the same column is consistent, which unifies the weights assigned to the same column. Then the final ranking of all features, $\{r_1, r_2, \cdots, r_v\}$, can be obtained by counting the frequency of each feature in each column.

In the final step, the model is trained by sequentially increasing the number of exogenous inputs to the model, depending on the importance of the features. Then, the appropriate training set is obtained by comparing the model test results under the AIC criterion. SVM-RFE combined with clustering algorithm migrates the feature ranking method for binary classification problems to a new application scenario and solves the problem of feature ranking for continuous processes. This approach implements feature selection while keeping the original data of the features intact and visually explaining the input variables selection.

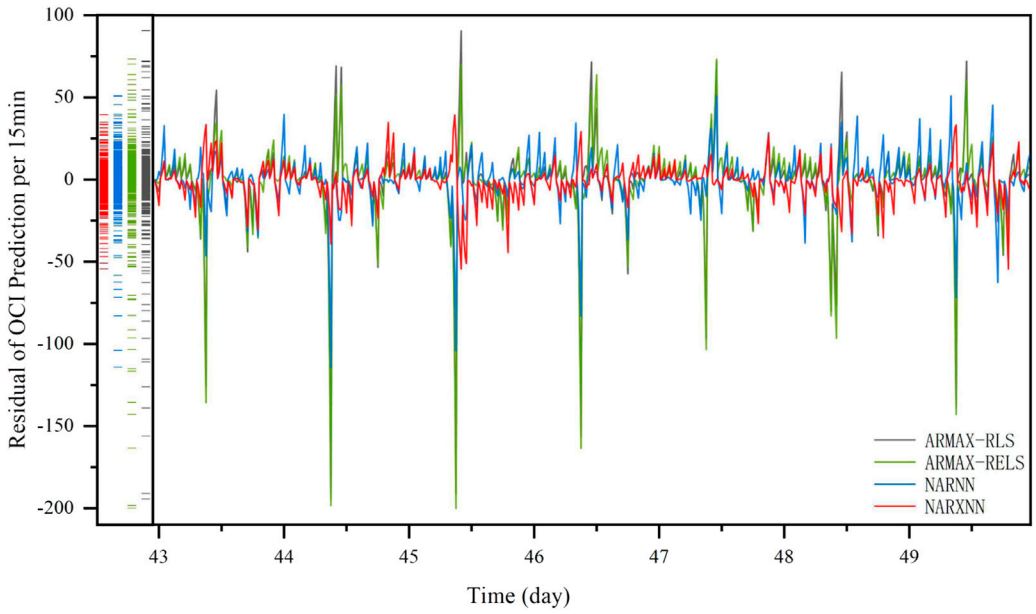


FIGURE 6
Prediction residual of OCI for the last 7 days over four steps ahead.

TABLE 2 Comparison of the Prediction Performance on OCI over four steps ahead.

Models	MSE	R
ARMAX-RLS	502.6552	0.9886
ARMAX-RELS	527.8635	0.9879
NARNN	219.4406	0.9946
NARXNN	108.0847	0.9974

3 Case studies

3.1 Performance evaluation index

In this study, MSE and R are used as the performance evaluation metrics of the model, which can be calculated as follows:

$$MSE = \frac{\sum_{t=1}^{n} (y(t) - y_{et}(t))^2}{n}, \tag{20}$$

$$R = \frac{\sum_{t=1}^{n} (y(t) - \bar{y})(y_{et}(t) - \bar{y}_{et})}{\sqrt{\sum_{t=1}^{n} (y(t) - \bar{y})^2} \sqrt{\sum_{t=1}^{n} (y_{et}(t) - \bar{y}_{et})^2}}, \tag{21}$$

Where y and y_{et} refer to the true values of the system output and the estimated values of the prediction model, respectively; $\bar{y}$ and $\bar{y}_{et}$ are the mean values of y and y_{et}. n is the total number of evaluation samples. The smaller the MSE, the smaller the error of the model. R is in the range of [0, 1], the closer it is to 1, the better the performance of model.

TABLE 3 Exogenous input variables related to the steel plant and descriptions.

Variables	Abbreviations	Units
Lagging current reactive power	Lag.CRP	kVarh
Leading current reactive power	Lead.CRP	kVarh
$tCO_2(CO_2)$	tCO2	ppm
Lagging current power factor	Lag.CPF	%
Leading current power factor	Lead.CPF	%

3.2 Operational cost index from wastewater treatment processes

3.2.1 Data processing

The data for the case study in this section mainly came from the wastewater treatment platform, BSM2-P Simulink simulation model, which adds the phosphorus treatment process based on BSM2. The actual collected inflow parameters (e.g., Qin: influent flow, SF: readily biodegradable substrate) were input into the simulation platform, and OCI was calculated every 15 min according to Eq. 22 based on the data collected from the simulation model.

$$OCI = 3SP + AE + ME + PE + EC + \max(0, HE - 7MP) + 24MT - 6MP, \tag{22}$$

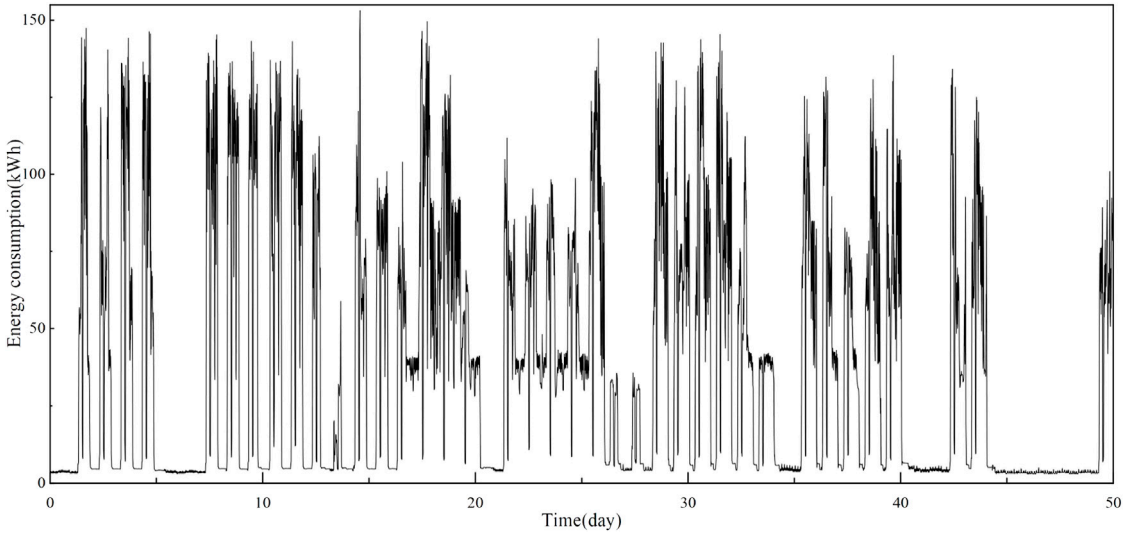


FIGURE 7
Operational cost index data for 50 days.

Where SP is the sludge production for disposal, AE is the aeration energy, ME is the mixing energy, PE is the pump energy, EC is external carbon addition, HE is the heating energy for increasing the temperature of the anaerobic digester, MP is the methane production, and MT is the metal salt to be added. A total of 50 days of data are illustrated in Figure 3, which has significant non-linearity. It should be noted that the reason for the negative OCI is that the methane produced in the water treatment process has a certain compensation for the operating cost.

After the data of OCI and influent data were obtained, additional preparation for feature selection was required in addition to outlier removal and normalization. The operational cost of WWTP is closely related to the parameters of influent, but excessive parameters are not conducive to the further selection of features, so more preprocessing of the input data is required. Specifically, the selection of features practically implies exploring the correlation between exogenous inputs and target variables. The dataset with no or little change has little influence but will increase the subsequent computation, so that variables related to this kind of data need to be eliminated. Such problem can be solved by excluding feature arrays with variance less than a certain threshold. Furthermore, variables with strong linear correlation would make the SVM-RFE's judgment of importance seem unreasonable. Therefore, some correlation analysis methods, such as Pearson correlation analysis, need to be used to isolate the variables with strong linear correlation before using SVM-RFE. The variables $\{t, S_F, S_I, S_{PO_4}, X_I, Q_{in}, S_{N_a}\}$ were finally selected in this study as the alternative input features for predicting OCI, and further ranked by using SVM-RFE combined with clustering algorithm. Table 1 shows the full overview of these variables.

To test the reliability of the method under different number of clusters, the diverse number of OCI clusters was set and the importance of each variable based on different number of clusters is presented in Figure 4. The results indicate that changes in the number of clusters over a wide range have some but little effect on the importance of exogenous variables, such that there is no change in the importance ranking of the variables. The decreasing ranking of variables importance is $\{S_F, Q_{in}, X_I, S_I, S_{PO_4}, S_{N_a}, t\}$, where the importance indicators of S_F and Q_{in}, S_I and S_{PO_4} are similar, respectively. It is worth noting that the importance of t behaves unusually at a clustering number of 1000, which is caused by the fact that the number of clusters is close to the raw data labels. As mentioned earlier, in order to accurately identify feature importance, classes with significant differences need to be generated by clustering.

Generally, the model prediction accuracy will improve somewhat as the number of exogenous variables increases, but the rate of improvement is limited when the number of selected variables reaches a specific value. Fewer variables can be selected from alternative variables set to save computational power and avoid overfitting. Finally, according to the AIC criterion, S_F and Q_{in} were chosen as exogenous input variables in this study case.

3.2.2 Results and discussion

The OCI values for each time period were calculated using the data collected in the BSM2-P simulation model. A total of 4,799 samples from 50 days were retained. The sample set was split, with the data of the first week being the training set and the remaining as the testing set. Each model was applied to predict OCI over four steps ahead.

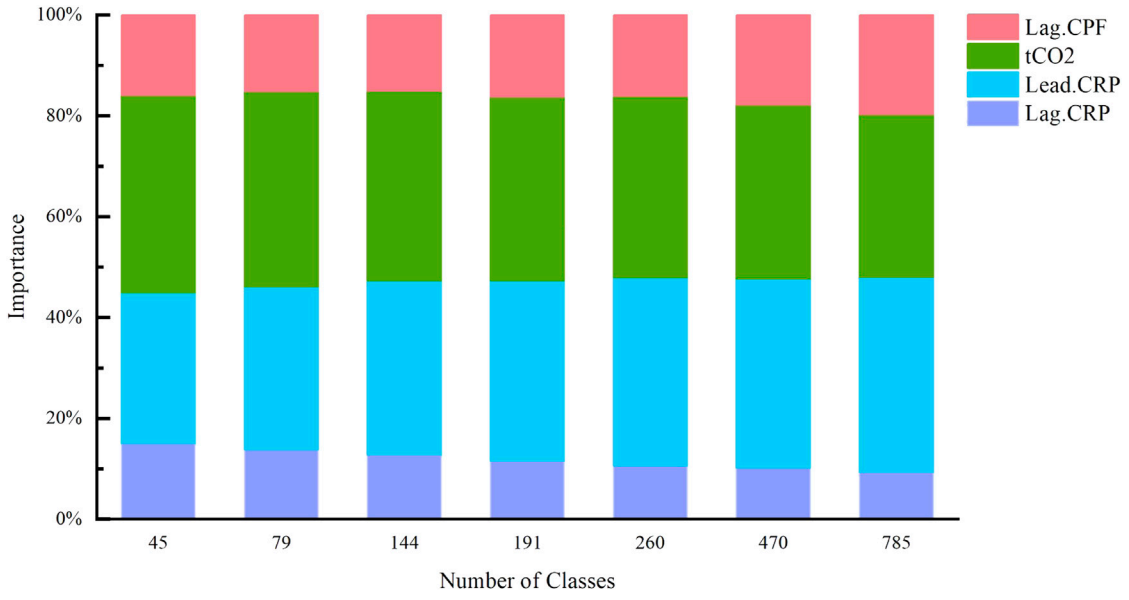


FIGURE 8
Importance of each variable to the energy consumption with different number of clusters.

Figure 5 shows the prediction performance of four ARMA variant models on the last week of OCI values and compares them with the test set, respectively. As shown in Figures 5A,B, the ARMAX models based on two different algorithms are similar in overall prediction performance. Under relatively stable conditions, the predicted values of ARMAX model are in good agreement with some original data with linear characteristics. However, comparing the peak locations marked in Figures 5A,C, the ARMAX model has slightly worse performance. This mainly results from the fact that the linear model is not competitive for data predictions with significant nonlinearity. On the other hand, although the RELS algorithm takes into account the effect of residual information, the performance does not improve significantly compared to RLS. This is due to the limited effect provided by the residuals of the previous moment during the nonlinear change phase of the data.

Compared to the ARMAX model, the prediction values of the NARNN model displayed in Figure 5C are more in line with the real values. However, there are still deviations, as shown in the half-day period after the 46th day. The NARXNN model fits the real data better at the locations of the peaks, troughs, as well as under the other linear conditions as shown in Figure 5D. The residual distribution of the prediction results for the four variants is plotted in Figure 6. The NARXNN model produces a smaller span of error, which indicates the better prediction performance of the model.

It is worth noting that the training of the neural network model has strong randomness. This can be solved by trials and errors. To well illustrate the performance of the proposed method, the results of NARNN and NARXNN are listed in Table 2, average values of which were calculated from the prediction more than 10 times. As can be seen from Table 2, the results are as follows: Compared with ARMAX-RLS, ARMAX-RELS and NARNN model, MSE of the NARX model is reduced by 78.50%, 79.52%, 50.75%, respectively, and R is improved by 0.89%, 0.96%, 0.28%. Based on the above results, it can be observed that NARXNN model has a better performance in predicting the wastewater treatment cost index.

3.3 Energy consumption from a full-scale steel plant

3.3.1 Data processing

The real data in this subsection, concerning energy consumption in the steel industry, were taken from University of California, Irvine (UCI). The data were collected from DAEWOO Steel Co., Ltd. in Gwangyang, South Korea. Energy consumption information for the industry is stored on the Korea Electric Power Corporation's website (pccs.kepco.go.kr), and daily, monthly and annual data are computed and displayed. Figure 7 presents the data of energy consumption for every 15 min over a total of 50 days.

Obviously, the energy consumption data in this case are more non-linear due to the fact that they are collected directly from the actual steel plant. This data set even contains some coarse data or outliers, and exhibits significant dramatics over

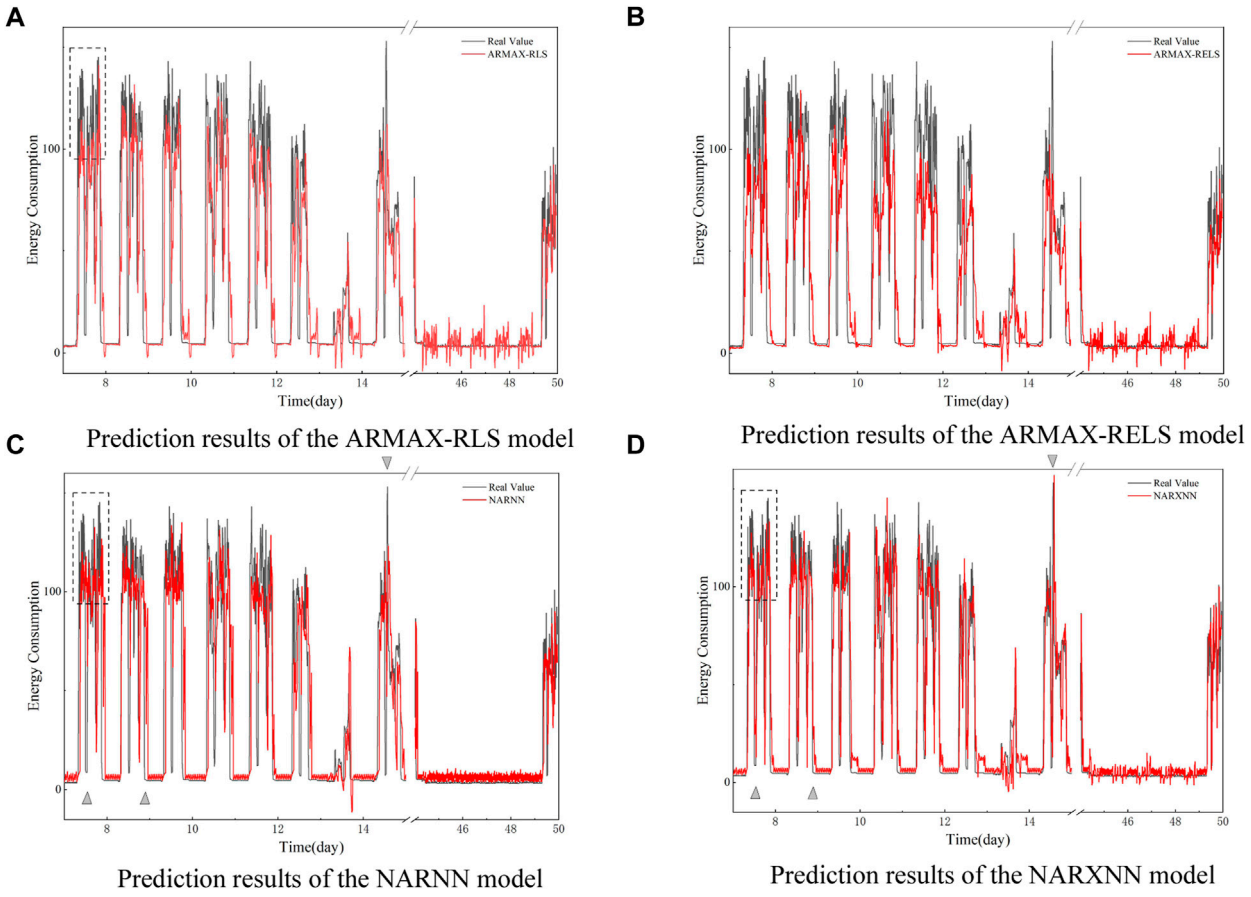


FIGURE 9
Prediction results of energy consumption over four steps ahead. **(A)** Prediction results of the ARMAX-RLS model. **(B)** Prediction results of the ARMAX-RELS model. **(C)** Prediction results of the NARNN model. **(D)** Prediction results of the NARXNN model

time. The data performance varied obviously at different time intervals, particularly from the 5th to the 7th day, the processes were stable relatively, while the stable processes changed completely from the 20th to the 21st day and from the 45th to the 50th day. The data from the 10th to the 15th day, the 15th to the 20th day and the 21st to the 27th day also changed completely with different fluctuating trends, all of which added the difficulty in the sequential modeling and predicting.

Table 3 lists the alternative exogenous variables provided in the data source file that present continuous numeric variation. Similarly, as mentioned earlier, the selection of model inputs is essential before building models with exogenous inputs. This work is based entirely on data relationships without considering the mechanism. The linear correlations between the variables in Table 3 were first examined by the Pearson correlation analysis, and there was a strong linear relationship between Lead. CRP and Lead. CPF. After removing Lead. CPF, the remaining four variables were analyzed for importance in this study case and the results are shown in Figure 8. It is noticeable that Lead. CRP and tCO2 have significantly higher importance indicators than the other two variables, and therefore Lead. CRP and tCO2 were identified as exogenous input variables for the prediction model of energy consumption in steel plant.

3.3.2 Results and discussion

The dataset was split, with the first 7 days of data used as the training set and the remaining data used as the test set to evaluate the performance of each model for four steps ahead prediction. The prediction performance of the four variants of the ARMA model is shown in Figure 9, and in general, the prediction performance on real data all deteriorates compared to the prediction on the simulated data in the last study case. The main reason for this occurrence is still that the steel plant data set contains the rough data, as well as its own strong nonlinearity and sharp changes over time.

TABLE 4 Comparison of the Prediction Performance on Energy Consumption over four steps ahead.

Models	MSE	R
ARMAX-RLS	522.7977	0.8215
ARMAX-RELS	460.9122	0.8588
NARNN	684.5822	0.7666
NARXNN	374.0288	0.8741

Nevertheless, in this study case, the neural network models perform much better than the general time series models. Specifically, the difference between the prediction results of two ARMAX models is small, and both have the tendency that the predicted data obviously cannot track the true data with high oscillation, as shown in the dashed rectangular box in Figure 9A. The predicted data from ARMAX-RELS are less overshooting compared to ARMAX-RLS when the energy consumption data change rapidly from a declining state to a flat state, as can be seen in the comparison of Figures 9A,B. On the contrary, two autoregressive neural network models perform much better in these issues.

As shown in Figures 9C,D, the results predicted by the two autoregressive neural network models oscillate less between the 44th and the 49th day, and the data perform more smoothly. However, the NARXNN model has a higher prediction accuracy than other models for parts with steep variations, such as the rising or falling edges indicated by the triangle symbol in the figure. Different from the NARXNN model, NARNN model performs even worse than ARMAX in these regions indicated by the triangle symbol, showing the positive impact of introducing exogenous inputs on the prediction results. Although the prediction for smooth data in the period from the 44th to the 49th day is slightly inferior to that of the NARNN model, the NARXNN still performs better overall.

The results are tabulated in Table 4 as follows: Compared with ARMAX-RLS, ARMAX-RELS and NARNN model, MSE of the NARX model is reduced by 28.46%, 18.85%, 45.36%, respectively, and R is improved by 6.40%, 1.79%, 14.02%. Based on the above results, it can be observed that NARXNN model has a better performance in predicting the energy consumption from a full-scale steel plant. It is worth noting that the performance of the NARNN model is quite different from that in the previous study case, which is mainly due to the lack of intervention from exogenous inputs. It is difficult for the general neural network model to predict the practical data with complex characteristics such as strong nonlinearity, strong volatility, and outliers in this case.

4 Conclusion

Process monitoring of operational costs can benefit for the operational costs reduction and other financial budget management in industries. This paper provides a comparative analysis on the performance of four ARMA model variants (i.e., ARMAX-RLS, ARMAX-RELS, NARNN and NARXNN), using operational costs and energy consumption predictions as a baseline for real applications. In addition, a method based on SVM-RFE combined with clustering algorithm was developed to extract useful features that are important for the construction of the above-mentioned models and to provide a way to measure and explain how important the corresponding features are.

The analysis of the data and the evaluation of the prediction results lead to the following conclusions: The two time series models, ARMAX-RLS and ARMAX-RELS, have acceptable prediction performance under conditions where the data exhibit stable patterns. But if the predicted data have strong nonlinearities as well as irregular changes, two ARMAX models can only meet the minimum prediction needs. Compared to the other three variants, the NARXNN model achieves the most accurate prediction results in both study cases, due to the help of the neural network for nonlinear data prediction on the one hand and the choice of exogenous inputs on the other.

In future research, the method of feature selection will be further explored and the interpretability of the method will be enhanced. Another aspect is that predictive models will be further incorporated into control strategies for costs reduction in industrial processes.

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

YL was responsible for the specific work of this manuscript. GF and YL collaborated to the analysis of the data and to the writing of the manuscript. DH guided the work of this manuscript. BC and HC reviewed the content of the paper.

Funding

This research was funded by the National Natural Science Foundation of China (62273151, 61873096, and 62073145), Guangdong Basic and Applied Basic Research Foundation (2020A1515011057, 2021B1515420003), Guangdong Technology International Cooperation Project Application (2020A0505100024, 2021A0505060001). Fundamental Research Funds for the central Universities, SCUT (2020ZYGXZR034). Yiqi Liu also thanks for the support of Horizon 2020 Framework Programme-Marie Skłodowska-Curie Individual Fellowships (891627).

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/fenrg.2022.1073271/full#supplementary-material

References

Altıntaş, R. S., Kahya, M., and Ünver, H. Ö. (2016). Modelling and optimization of energy consumption for feature based milling. *Int. J. Adv. Manuf. Technol.* 86, 3345–3363. doi:10.1007/s00170-016-8441-7

A. A. Ansari, S. Singh Gill, G. R. Lanza, and W. Rast (Editors) (2011). *Eutrophication: Causes, consequences and control* (Dordrecht: Springer Netherlands). doi:10.1007/978-90-481-9625-8

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008). *Time series analysis: Forecasting and control*. 4th Edition1st ed. Hoboken, New Jersey: John Wiley & Sons. doi:10.1002/9781118619193

Cadenas, E., Rivera, W., Campos-Amezcua, R., and Cadenas, R. (2016). Wind speed forecasting using the NARX model, case: La mata, oaxaca, méxico. *Neural comput. Appl.* 27, 2417–2428. doi:10.1007/s00521-015-2012-y

Car-Pusic, D., Petruseva, S., Zileska Pancovska, V., and Zafirovski, Z. (2020). Neural network-based model for predicting preliminary construction cost as part of cost predicting system. *Adv. Civ. Eng.* 2020, 1–888617013. doi:10.1155/2020/8886170

Chan, D. Y.-L., Huang, C.-F., Lin, W.-C., and Hong, G.-B. (2014). Energy efficiency benchmarking of energy-intensive industries in Taiwan. *Energy Convers. Manag.* 77, 216–220. doi:10.1016/j.enconman.2013.09.027

Dan Foresee, F., and Hagan, M. T. (1997). "Gauss-Newton approximation to Bayesian learning," in Proceedings of International Conference on Neural Networks (ICNN'97), Houston, TX, USA, April 1997 (IEEE). doi:10.1109/ICNN.1997.614194

De Gooijer, J. G., and Hyndman, R. J. (2006). 25 years of time series forecasting. *Int. J. Forecast.* 22, 443–473. doi:10.1016/j.ijforecast.2006.01.001

Deng, H., Yang, K., Liu, Y., Zhang, S., and Yao, Y. (2021). Actively exploring informative data for smart modeling of industrial multiphase flow processes. *IEEE Trans. Ind. Inf.* 17, 8357–8366. doi:10.1109/TII.2020.3046013

Ding, F. (2010). Several multi-innovation identification methods. *Digit. Signal Process.* 20, 1027–1039. doi:10.1016/j.dsp.2009.10.030

Fang, T., and Lahdelma, R. (2016). Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Appl. Energy* 179, 544–552. doi:10.1016/j.apenergy.2016.06.133

Guyon, I., and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182. doi:10.1162/153244303322753616

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422. doi:10.1023/A:1012487302797

Han, Y., Zeng, Q., Geng, Z., and Zhu, Q. (2018). Energy management and optimization modeling based on a novel fuzzy extreme learning machine: Case study of complex petrochemical industries. *Energy Convers. Manag.* 165, 163–171. doi:10.1016/j.enconman.2018.03.049

Hsu, D. (2015). Identifying key variables and interactions in statistical models of building energy consumption using regularization. *Energy* 83, 144–155. doi:10.1016/j.energy.2015.02.008

Huang, K. Y., and Jane, C.-J. (2009). A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories. *Expert Syst. Appl.* 36, 5387–5392. doi:10.1016/j.eswa.2008.06.103

International Energy Agency (2021). *World energy outlook 2021*, Paris: International Energy Agency. Available at; https://www.iea.org/reports/world-energy-outlook-2021.

Jia, S., Yuan, Q., Cai, W., Li, M., and Li, Z. (2018). Energy modeling method of machine-operator system for sustainable machining. *Energy Convers. Manag.* 172, 265–276. doi:10.1016/j.enconman.2018.07.030

Juberias, G., Yunta, R., Garcia Moreno, J., and Mendivil, C. (1999). "A new ARIMA model for hourly load forecasting," in Proceedings of the 1999 IEEE Transmission and Distribution Conference (Cat. No. 99CH36333), New Orleans, LA, USA, June 1999 (IEEE), 314–319. doi:10.1109/TDC.1999.755371

Kahraman, A., Kantardzic, M., Kahraman, M. M., and Kotan, M. (2021). A data-driven multi-regime approach for predicting energy consumption. *Energies* 14, 6763. doi:10.3390/en14206763

Kun, Z., and Weibing, F. (2021). Prediction of China's total energy consumption based on bayesian ARIMA-nonlinear regression model. *IOP Conf. Ser. Earth Environ. Sci.* 657, 012056. doi:10.1088/1755-1315/657/1/012056

Li, M.-J., and Tao, W.-Q. (2017). Review of methodologies and polices for evaluation of energy efficiency in high energy-consuming industry. *Appl. Energy* 187, 203–215. doi:10.1016/j.apenergy.2016.11.039

Liu, Y., Huang, D., Liu, B., Feng, Q., and Cai, B. (2021). Adaptive ranking based ensemble learning of Gaussian process regression models for quality-related variable prediction in process industries. *Appl. Soft Comput.* 101, 107060. doi:10.1016/j.asoc.2020.107060

Liu, Y., Yang, C., Zhang, M., Dai, Y., and Yao, Y. (2020). Development of adversarial transfer learning soft sensor for multigrade processes. *Ind. Eng. Chem. Res.* 59, 16330–16345. doi:10.1021/acs.iecr.0c02398

MacKay, D. J. C. (1992). A practical bayesian Framework for backpropagation networks. *Neural Comput.* 4, 448–472. doi:10.1162/neco.1992.4.3.448

Rakotomamonjy, A. (2003). Variable selection using SVM-based criteria. *J. Mach. Learn. Res.* 3, 1357–1370. doi:10.1162/153244303322753706

Shi, X., Huang, G., Hao, X., Yang, Y., and Li, Z. (2021). A synchronous prediction model based on multi-channel CNN with moving window for coal and electricity consumption in cement calcination process. *Sensors* 21, 4284. doi:10.3390/s21134284

Silva, V. L. G. da, Oliveira Filho, D., Carlo, J. C., and Vaz, P. N. (2022). An approach to solar radiation prediction using ARX and ARMAX models. *Front. Energy Res.* 10–822555. doi:10.3389/fenrg.2022.822555

Vu, D. H., Muttaqi, K. M., Agalgaonkar, A. P., and Bouzerdoum, A. (2017). Short-term electricity demand forecasting using autoregressive based time varying model incorporating representative data adjustment. *Appl. Energy* 205, 790–801. doi:10.1016/j.apenergy.2017.08.135

Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., et al. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renew. Sustain. Energy Rev.* 82, 1027–1047. doi:10.1016/j.rser.2017.09.108

Wu, J., Cheng, H., Liu, Y., Huang, D., Yuan, L., and Yao, L. (2020). Learning soft sensors using time difference–based multi-kernel relevance vector machine with applications for quality-relevant monitoring in wastewater treatment. *Environ. Sci. Pollut. Res.* 27, 28986–28999. doi:10.1007/s11356-020-09192-3

Xiao, J., Li, Y., Xie, L., Liu, D., and Huang, J. (2018). A hybrid model based on selective ensemble for energy consumption forecasting in China. *Energy* 159, 534–546. doi:10.1016/j.energy.2018.06.161

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 12 January 2023
DOI 10.3389/fenrg.2022.1091350

OPEN ACCESS

EDITED BY
Xingxing Zhang,
Dalarna University, Sweden

REVIEWED BY
Yihe Wang,
University of North Carolina at Charlotte, United States
Jichao Bi,
Zhejiang University, China

*CORRESPONDENCE
Junyu Wang,
junyuwang@fudan.edu.cn

SPECIALTY SECTION
This article was submitted to Process and Energy Systems Engineering, a section of the journal Frontiers in Energy Research

RECEIVED 06 November 2022
ACCEPTED 29 December 2022
PUBLISHED 12 January 2023

CITATION
Wang B, Xu J, Ke J, Chen CLP, Wang J, Wang N, Li X, Zhang F and Li L (2023), CE-SDT: A new blockchain-based distributed community energy trading mechanism. *Front. Energy Res.* 10:1091350. doi: 10.3389/fenrg.2022.1091350

CE-SDT: A new blockchain-based distributed community energy trading mechanism

Bozhi Wang[1,2,3], Jinfei Xu[1,2], Jin Ke[4], C. L. Philip Chen[1,5], Junyu Wang[1,2]*, Nanxin Wang[6], Xi Li[6], Fangfei Zhang[7] and Lulu Li[1,6]

[1]Zhuhai Fudan Innovation Institute, Zhuhai, China, [2]State Key Laboratory of ASIC and System, Fudan University, Shanghai, China, [3]School of Computer Science and Engineering, University of New South Wales, Sydney, NSW, Australia, [4]School of Economics and Management, South China Normal University, Guangzhou, China, [5]School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong, China, [6]School of Computer Science, Fudan University, Shanghai, China, [7]School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW, Australia

With the spread of distributed renewable energy, residents are shifting from being mere consumers to being energy producers and consumers. This role shift poses challenges to the electricity trading mechanism that connects distributed renewable energy sources to the grid. In this paper, a new efficient and secure blockchain-based distributed community energy trading mechanism is proposed, called CE-SDT. Our system is proved to be stable and scalable. It can also help shift loads and power peaks and reduce customer costs by 60%. As a result, our proposed blockchain-based trading mechanism, as compared to the centralized trading mechanism, is applied to microgrids formed by distributed renewable energy sources, not only obtaining greater economic benefits but also reducing the carbon footprint of residents, and furthermore, it promotes low or zero-carbon configurations of the power system, thereby achieving certain environmental benefits.

KEYWORDS

energy trading, blockchain-based, smart contract, microgrids, solar power, zero-carbon communities

1 Introduction

Most of the current residential energy needs are met by non-renewable fossil fuel generation, including coal, oil, and natural gas (Shindell and Smith, 2019). However, the combustion of fossil fuels and the huge energy losses associated with long-distance transmissions produce large carbon emissions that greatly affect the climate and environment (Zeyringer et al., 2018). In order to fulfill the goal of carbon neutrality, countries around the world are strongly encouraging the development of low-carbon communities (LCCs) and zero-carbon communities (ZCCs) and promoting the integration of renewable energy systems into LCCs and ZCCs (Surana and Jordaan, 2019; Zhu et al., 2022). Modern buildings have a large flexibility to implement self-supply in daily life with the prevalence of distributed renewable energy resources (e.g., rooftop solar panels and wind turbines). In this context, the integration of distributed renewable energy systems and advanced networks, such as advanced metering infrastructure, solar panels, and a residential battery energy storage system, allows individuals to generate, store, and transfer energy among demand-side sources (Zhang et al., 2011; Gai et al., 2019). Traditional residents are transformed from electricity consumers to "electricity producers and consumers," who can generate electricity from renewable energy sources and participate in active grid distribution.

Neighboring energy trading enables energy transfer and transactions in the neighborhood to achieve greater economic benefits (Logenthiran et al., 2012). This microgrid, which integrates the distributed load and power generation, can minimize the loss of electrical energy and achieve cost reduction. Likewise, it can reduce greenhouse gas emissions, achieve a low or zero-carbon configuration of the power system, and reduce the carbon footprint in the production and management of energy (Hanna et al., 2017; Du et al., 2019; Su et al., 2021). To fully realize the economic and environmental benefits of microgrids, it is critical to establish an efficient power trading mechanism to integrate environmentally friendly, reliable, and flexible distributed renewable energy into the power grid.

The traditional power trading model uses a centralized approach by establishing a centralized trusted third party (TTP) to control the collection and exchange of the entire transmission data (Yang and Wang, 2016). However, great amounts of work have indicated the weakness behind it. For instance, the unified and centralized management of trading networks is similar to other centralized computing solutions, which leads to a series of privacy and security challenges (Aitzhan and Svetinovic, 2018). The electricity consumption behaviors of users and trading price ranges are easily eavesdropped on by attacking the control center (Niyato et al., 2011). Also, it is difficult to adapt to the large-scale utilization of renewable energy in the unified management (Lu et al., 2019). Furthermore, the renewable energy supply differs from the traditional long-term generation trading with fixed characteristics. There is the uncertainty of generation and time-sharing trading (Jiawei et al., 2018). Decentralized energy management solutions based on the blockchain technology are gaining attention to address these issues and better realize the economic and environmental benefits of microgrids.

The decentralized, asymmetric encryption and smart contract applications that characterize blockchains have great potential in many areas, such as finance, computer applications, and resource monetization (Ali et al., 2019; Li et al., 2021). Similarly, the blockchain technology holds great promise for applications in the energy sector. Yinan et al. (2019) proposed a new type of network innovation architecture of software-defined networking based on blockchain computing. By separating all control commands and operations of the energy internet, the system becomes more flexible. The blockchain helps peer-to-peer (P2P) interactions in the electricity market and reduces the threshold of the local retailer participation (S, 1963). Yorozu et al. (1987) studied the conversion problems between the virtual currency and the real currency, when traded in a blockchain. A credit-based payment scheme can be established to enable users to manage and transfer energy coins. Blockchains are also widely used in the energy industry in the area of carbon quotas (Hua and Sun, 2019; Hua et al., 2020; Yuan et al., 2021).

The use of blockchains in decentralized energy management is also being demonstrated. Based on the advantages of the blockchain technology, Li and Li (2019) proposed a microgrid management scheme that incorporates the demand for electric vehicles. Fan et al. (2018) used the blockchain technology to explore the issues of trading and valuation of the energy internet and provided useful references for producers, consumers, and regulators as participants. Pipattanasomporn et al. (2018) implemented a blockchain network-based solar energy exchange project in their laboratory. Hou et al. (2019) created a system that allows nodes to use locally stored power to meet their own power load before participating in the trading market as sellers while still having a large amount of power. There are also studies that have established P2P energy trading platforms based on a blockchain. Alcarria et al. (2018) proposed an empowerment platform based on the blockchain technology to facilitate the exchange of information and the transaction of resources within the community. Park et al. (2018) designed and simulated a blockchain-based P2P energy trading platform and compared the economic benefits of the platform with those of traditional power generation companies based on the simulation results. In the long run, the blockchain-based P2P energy trading platform can lead to a secure, energy-efficient, trustworthy, and sustainable energy trading environment.

However, for community-oriented distributed energy trading mechanisms, there are still gaps between theory and practical applications.

- The sheer volume of data and limited computing resources make it difficult to meet the demand and work properly.
- Although the concept of a decentralized system has been developed, the bid trade mechanism among local producers and consumers is still being ignored. In the neighborhood trade event, individuals would bid simultaneously; so, detailed regulations need to be provided in bid operations.

In this paper, we propose a blockchain-based secure distributed trading (SDT) mechanism that enables individuals in a community to trade energy in a distributed manner, called community energy (CE)-SDT. The smart contract is adopted in CE-SDT to ensure confidentiality, authenticity, incontestability, and integrity of the shared information. Through simulation, we analyze different scales of residential users from the generation side to determine whether the installation of solar panels, batteries, and smart contract-based trading systems increase the smartness and self-sufficiency of these homes and businesses. This will help reduce the carbon footprint of households and bring some economic benefits. This paper offers the following contributions:

- A new blockchain-based secure distributed trading framework is designed to connect households and the grid using smart contracts to automate bidding and decision-making to enable the energy exchange.
- An energy management model is developed to determine the appropriate sell/buy strategy, optimizing the amount of individual value and shifting the system peak demand.
- Our CE-SDT is shown to be stable and scalable. It can also help shift loads and power peaks and reduce customer costs.

The remaining parts of this paper are summarized as follows: Section 2 describes the system structure and workflow of CE-SDT; Section 3 describes the energy management model. Section 4 shows the deployment of CE-SDT in simulation and discusses the simulation results. Section 5 draws conclusions and looks forward to future research.

2 System structure

The CE-SDT trading prototype considered in this paper comprises multiple autonomous residential buildings, which actively respond to

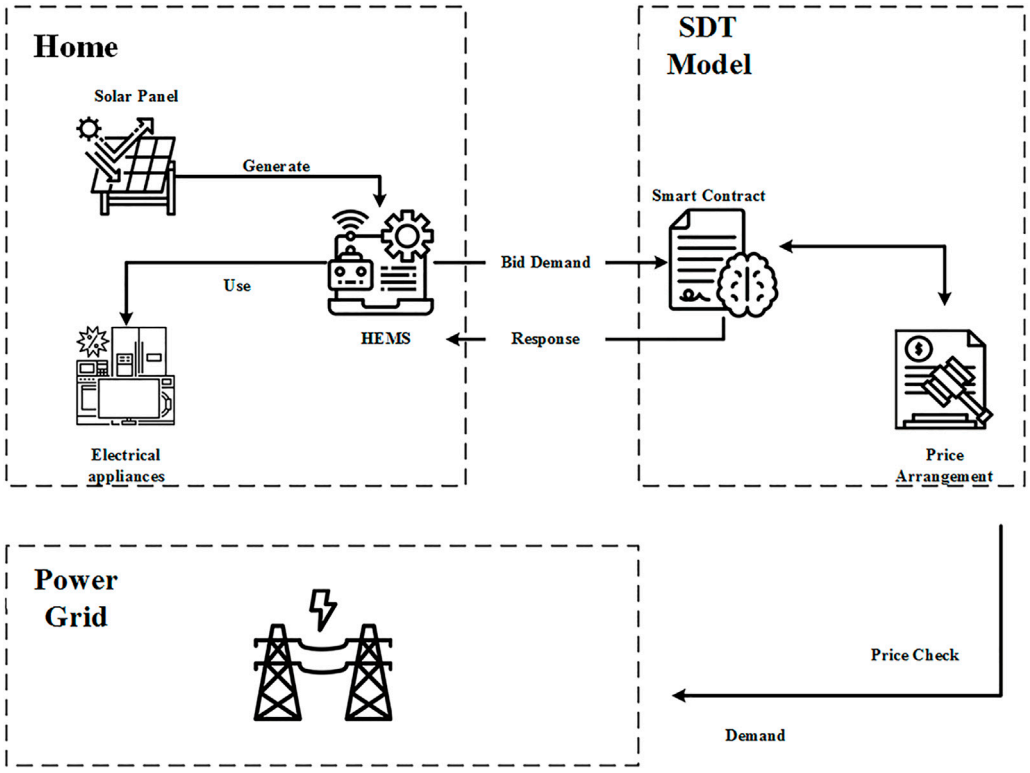


FIGURE 1
System architecture.

demand-side management signals and utilize the user-side energy. There is a regional smart contract that performs the entire energy management in trading and transformation. The control architecture of the proposed system is shown in Figure 1. In the residential end, some smart buildings are equipped with a rooftop PV solar source and a battery energy storage system (BESS). The BESS is used to accommodate solar power and provide power for domestic use. It may also store the extra solar panel or get cheap power from the grid for future usage. The smart meter is installed in every home, taking the role of a communication agent between the smart contract and the home.

A home energy management system (HEMS) is deployed in a smart home, managing home energy resources, including solar penetration data, some electrical appliances, and BESS. It communicates with the smart meter, home energy resources, and smart contract, and it also interacts with the user. When trading happens, the HEMS collects information about the forecasted personal power usage, real-time pricing data from the grid, and real-time personal solar generation data.

When residential trading happens, the HEMS obtains bidding information from the regional smart contract. Simultaneously, it forecasts the load consumption profile in the next stage duration and submits the forecasting excrescent/missing load profiles to the smart contract for selling/buying. Based on this concept, the smart contract sets up an optimal energy management model to determine the proper selling/buying strategy, optimizing the amount of individual value and the shifting system peak demand.

At the simulation level, our model works in the flow, as shown in Figure 2.

(1) The user makes a request to the smart contract, indicating the required electricity consumption and bidding method.
(2) The smart contract confirms the grid power price and gives a suggested price based on the recently traded price (the model is shown in Section 3.3).
(3) It determines whether users request bidding.
(3.1) No bidding, supply power directly from the grid and shift to 9.
(3.2) Requires bidding, shift to 4.
(4) The smart contract broadcasts user requests to other user nodes with power generation capabilities.
(5) After receiving a request, the supply user node determines the amount of power that can be provided. It then autonomously provides an offer to the smart contract based on the suggested price.
(6) The smart contract keeps accepting offers from supply nodes until timeout.
(7) It sorts the prices in the offers.
(8) It accepts offers in order and gives feedback to supply nodes.
(6.1) The power in the accepted offers has met the customer's demand.
(6.2) If all offers have been accepted and the power still does not meet the customer's demand, then the remaining demand will be provided by the grid.

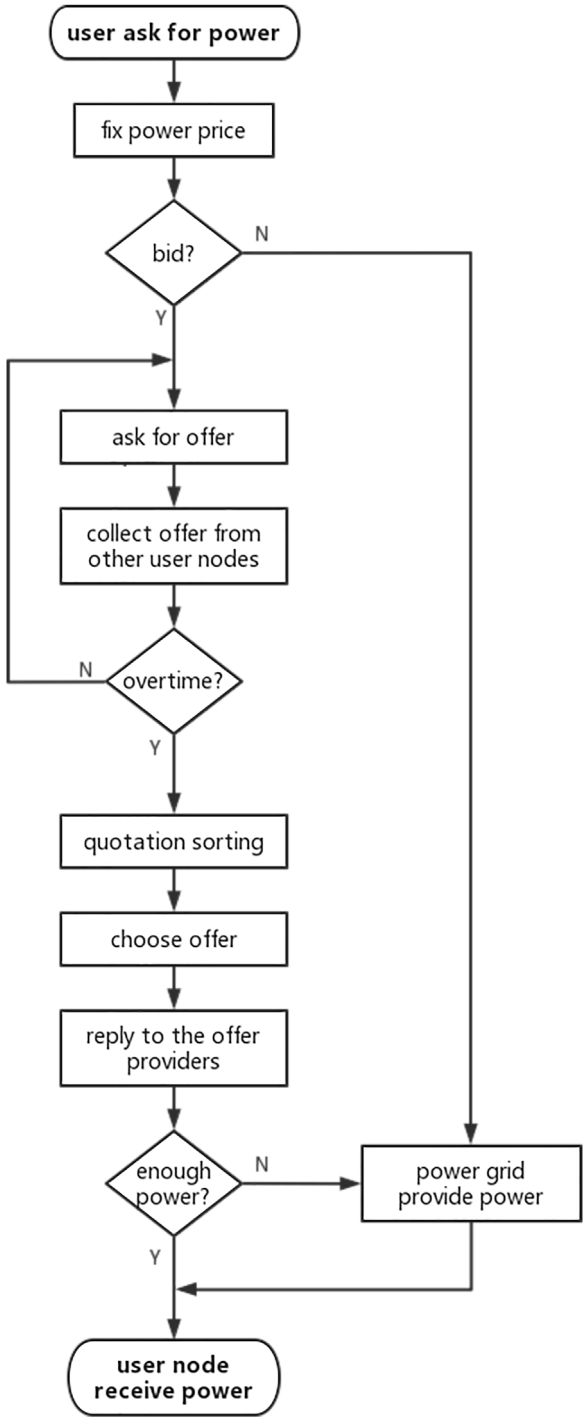


FIGURE 2
Energy trading workflow.

(9) The smart contract calculates the bill and supplies the power to the requested customer.

Following the system schematic representation presented upon, in this section, we first illustrate the generic CE-SDT models; then, we formulate the proposed HEMS with different kinds of operational dependencies due to living habits, which are considered as an additional constraint of the CE-SDT-based real-time trading model.

3 Model

In this section, we introduce the main models in the CE-SDT system, including the user classification, power generation model, and bidding model. These models define and constrain behavior patterns of users in the system.

3.1 Generic classification of resident users

Consider residential units with N controllable users and denote the set of users as Ω, i.e., $|\Omega| = N$. The controllable users are further categorized into the following sets based on their operational characteristics:

Ω^1: Set of users operating with the power in the range of $[P_a{}^{min}, P_a{}^{max}]$, $a \in \Omega^1$ but without the energy generation ability. In addition to that, a disutility function is applied to measure the dissatisfaction of the user on deviating from a nominal operating point. Typical users in this set consume but do not generate electricity.

Ω^2: Set of users generating power in the range of $[P_a{}^{min}, P_a{}^{max}]$, $a \in \Omega^2$ and having a prescribed energy consumption that must be completed in a specific time range. Typical users in this class equip a battery energy storage system (BESS) and generate electricity that they can supply to themselves and other residents.

3.2 Residential photovoltaic solar power models

The power output from the PV solar panel is related to solar radiation, the panel's surface area, and the energy conversion efficiency of the panel, expressed as follows:

$$\boldsymbol{P_t^{pv}} = \boldsymbol{A \cdot \sigma \cdot r_t}, \tag{1}$$

where P_t^{pv} denotes the solar power output at time t (kW); A is the surface area of the PV solar panel (m^2), and σ defines the energy conversion efficiency of the PV solar panel (%); and r_t is the solar radiation at time t (kW/m^2).

3.3 Energy bidding models

Denote the energy consumption schedule for each controllable user as follows:

$$\boldsymbol{P_k} = [\boldsymbol{P_{k,1}}, \boldsymbol{P_{k,2}}, \ldots, \boldsymbol{P_{k,T}}] \forall \boldsymbol{k} \in \boldsymbol{\Omega}. \tag{2}$$

The power consumption schedule of N controllable appliances can then be represented as a matrix with $N \times T$ dimensions, where the entry $P_{a,t}$ represents the power consumption of an appliance a at the time interval t. The net-power consumption of the home can be correspondingly represented as follows:

$$\tilde{\boldsymbol{P}}^{\boldsymbol{h}} = \left[\tilde{\boldsymbol{P}}_{\boldsymbol{1}}^{\boldsymbol{h}}, \ldots, \tilde{\boldsymbol{P}}_{\boldsymbol{t}}^{\boldsymbol{h}}, \ldots, \tilde{\boldsymbol{P}}_{\boldsymbol{T}}^{\boldsymbol{h}}\right], \tag{3}$$

$$\tilde{\boldsymbol{P}}_{\boldsymbol{t}}^{\boldsymbol{h}} = \boldsymbol{P}_{\boldsymbol{t}}^{\boldsymbol{mr}} + \sum_{\boldsymbol{i} \in \boldsymbol{\Omega}} \boldsymbol{P_{k,t}} - \boldsymbol{P}_{\boldsymbol{t}}^{\boldsymbol{pv}} \; \boldsymbol{t} = \boldsymbol{1}: \boldsymbol{T}, \tag{4}$$

where P_t^{mr} is the must-run home load at time t (W).

3.3.1 Objective

The objective of the proposed energy bidding model in this study is to realize the energy self-arrangement of the demand side. All bid strategies will be decided and recorded in the smart contract, and the selling/buying prices will be arranged from high to low.

$$p^S = \lambda \sum_{t=1}^{T} \left(\tilde{P}_t^h\right) + \mu \cdot \sqrt{\sum_{t=1}^{T} \left(p^{S1} + p^{S2} + p^{S3} + \cdots p^{Sn-1}\right)}, \quad (5)$$

where p^S denotes the selling price determined by the smart contract.

There are two components in the model. The first is the redundant energy generated by the users, Ω^2, at time t; the second is the average bidding price. λ and μ are different weighting coefficients.

$$p^B = \beta \sum_{t=1}^{T} \left(P_t^k\right) - \nu \cdot \sqrt{\sum_{t=1}^{T} \left(p^{B1} + p^{B2} + p^{B3} + \cdots p^{Bn-1}\right)}, \quad (6)$$

where p^B denotes the buying price determined by the smart contract.

The first component is the user requirement energy at time t; the second is the average buying price. β and ν are different weighting coefficients.

3.3.2 Mandatory constraints

The energy bidding model is subjected to the following constraints:

(1) Operational constraints of users in Ω^2, i.e., a BESS in this study. The operation of the BESS must satisfy the following constraints:

$$E_{t+1}^{ess} = \begin{cases} E_t^{ess} + \Delta t \eta^c P_t^{ess} - E_t^{ess} \eta^l \Delta t \; P_t^{ess} > 0 \\ E_t^{ess} - \left|P_t^{ess}\right| \eta^d \Delta t - E_t^{ess} \eta^l \Delta t \; P_t^{ess} \le 0 \end{cases}, \; t = 1 : T - 1, \quad (7)$$

$$SOC_t^{ess} = \frac{E_t^{ess}}{E^{ess,rate}} \; t = 1 : T, \quad (8)$$

$$\left|P_t^{ess}\right| \le P^{ess,rate} \; t = 1 : T, \quad (9)$$

$$SOC^{min} \le SOC_t^{ess} < SOC^{max} \; t = 1 : T, \quad (10)$$

$$SOC_t^{ess} \ge SOC^{dsr} \; t = T. \quad (11)$$

Equations 7, 8 model the variation of energy stored in the BESS; constraint (9) specifies the BESS's maximum charging/discharging power; constraint (10) ensures that the BESS's SOC is maintained within an allowable range; constraint (11) ensures that the SOC level of the BESS is larger or equal to a pre-specified threshold at the end of the day.

(2) The bidding ranking constraints of users in Ω^2 in this study. The smart contract arranges the selling/buying price as follows:

$$p_t^{S1} > p_t^{S2} > p_t^{S3} > \ldots > p_t^{Sn}, t = 1, 2 \ldots T, \quad (12)$$

$$p_t^{B1} < p_t^{B2} < p_t^{B3} < \ldots < p_t^{Bn}, t = 1, 2 \ldots T. \quad (13)$$

The smart contract will arrange the selling price from low to high and the buying price from high to low, pairing each of them following this sequence. For example, the smart contract will pair p_t^{S1} and p_t^{B1} first and then pair others and so forth.

The user node sends a request to the smart contract to ask for power. When receiving the request, the smart contract starts bidding and asks for an offer. Other user nodes who have the capacity to provide the power will reply with an offer to the contract. The smart contract will also ask the power grid for its power price. Then, the smart contract will collect these offers and make decisions.

4 Deployment and results

4.1 CE-SDT deployment

We used our own blockchain simulation tool to run early tests on the CE-SDT system. As shown in Table 1, there are several parameters we can change in the simulation.

We obtained some data on power usage from the government publicity. Figure 3 shows the average hourly power usage of 15 different users on 11/2012 in NSW, Australia. We set these 15 users' data as 15 types of user power usage inputs.

After an investigation, the most popular home-use power generator was found to be that with 5 kW. So, we set a default power output as 5 kW. Figure 3 shows the average hourly power generation of three different users with a 5-kW power generator on 11/2012 in NSW, Australia. We set these three users' data as three types of user power generation inputs. To fit the generation capacity of users and meet the need of the settlement, we give each user a 5,000-Wh power storage ability to store the power they have generated. The users are disabled to store the power they get from the power grid, but it can be activated in future research.

Since the number of users will be set to more than 15, each user will randomly select one of these 15 types as its input. It is similar to the input selection of the user power generation. Figure 4 shows the average hourly power usage of 15 different users. To differentiate the same type of users, we set a random settlement interval (ST) instead of an hourly interval. The settled power is calculated as follows:

$$P_{k,T}^s = P_{k,T} \star \frac{ST}{60}, \; P_{k,T}^{s,pv} = P_{k,T}^{pv} \star \frac{ST}{60}. \quad (14)$$

This settlement interval will also help us reduce network congestion.

The price of power in November from Energy AU is shown in Table 2.

The solar feed-in tariff is about 10 cents/kWh.

4.2 Results

With the definition shown previously, we get the following results: Figure 5 shows the power usage of the system under default settings. The user usage line means the quantity of power that users need from the system, which has already been subtracted from the power generated and used by them. The grid usage line means the quantity of power the grid has provided to users. The gap between the two lines is the power transacted through the bidding system.

As shown in Figure 6, when the number of users increases, the average power that each user needs from the system remains stable. On the other hand, in Figure 7, the average power that the grid provides to each user decreases in the middle of the day. It means that the more the users, the more likely are the transactions to be completed between users and the more the system will be activated. Our system illustrates great scalability.

Figure 8 shows that when the percentage of users who have the power generation capacity increases, the power each user needs from the system will decrease from 9 a.m.

As shown in Figure 9, as the percentage rises, the power grid provides a decrease at the high level, which means that we can utilize

TABLE 1 Key parameters in power grid simulation.

Parameter	Description	Default
SIM_TIME	Amount of time that the simulation runs	24 h
Ω	Number of users	60
Ω^2/Ω	Percentage of users who have the ability to generate power	20%
P^{pv}	Power output from the PV solar panel	5,000 W
P_STORAGE	Power storage ability of each user	5,000 Wh
ST	Settlement interval between two transactions	[45,75]

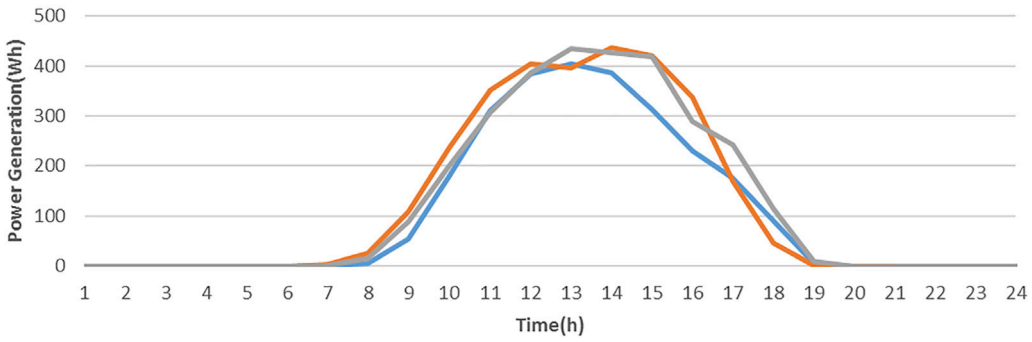


FIGURE 3
Power generation types.

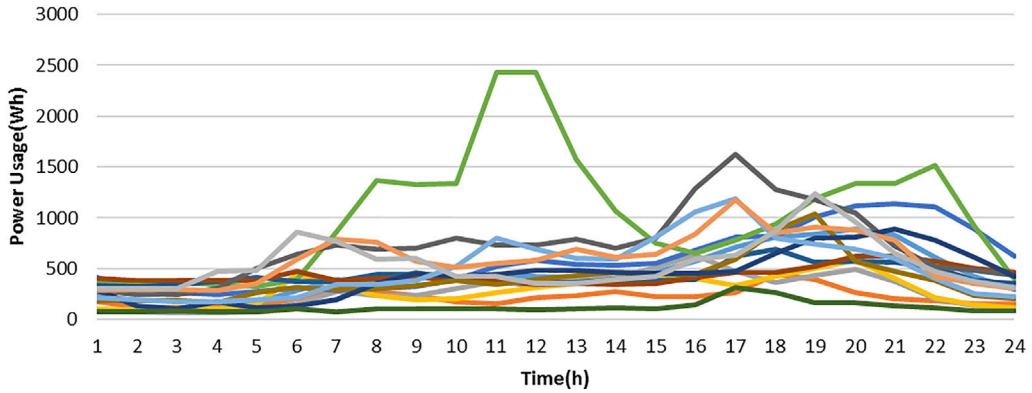


FIGURE 4
Power usage types.

TABLE 2 Australian power price hourly.

Time (h)	0–7	7–14	14–20	20–22	22–24
Price (AU cent/kwh)	17.7	27.71	54.11	27.71	17.7

this to help the grid to conduct peak load shifting at noon. We can also find out that when the percentage is more than 50%, the generation capacity exceeds electricity consumption, which means that some power is wasted. This may be solved by giving users a higher power storage capacity.

In Figure 10, we can find out that, as the percentage rises, the daily power cost has been decreasing, no matter whether the user has the power generation capacity or not. Ω^2 can even earn money when the percentage is up to 70%; at the same time, Ω^1 can reduce the cost to 44.7%. When the percentage is over 30%, Ω^1 in our system can spend less money than Ω^2 in a traditional power system, only 68.3% of the original cost. When half the users have the power generation capacity, the cost for Ω^1 and Ω^2 can be reduced to 60.2% and 19.6%,

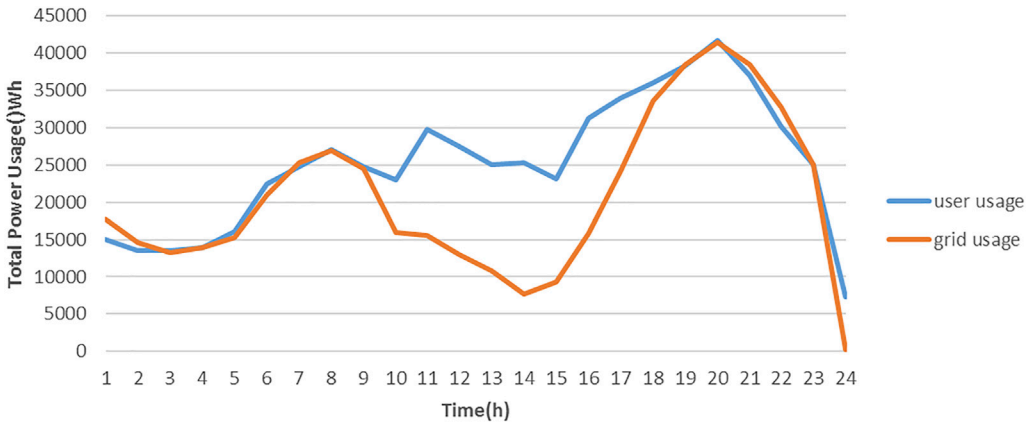


FIGURE 5
Total power usage with default settings.

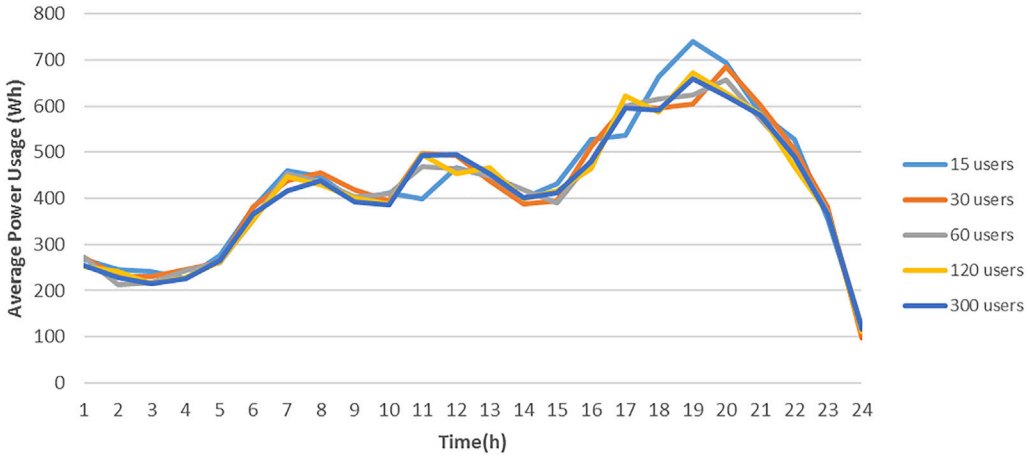


FIGURE 6
User power usage with different user numbers.

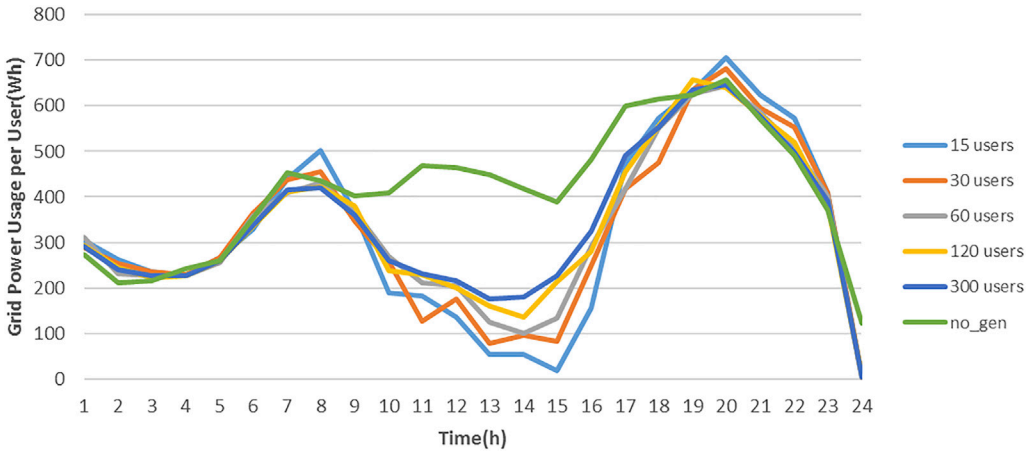


FIGURE 7
Grid power usage with different user numbers.

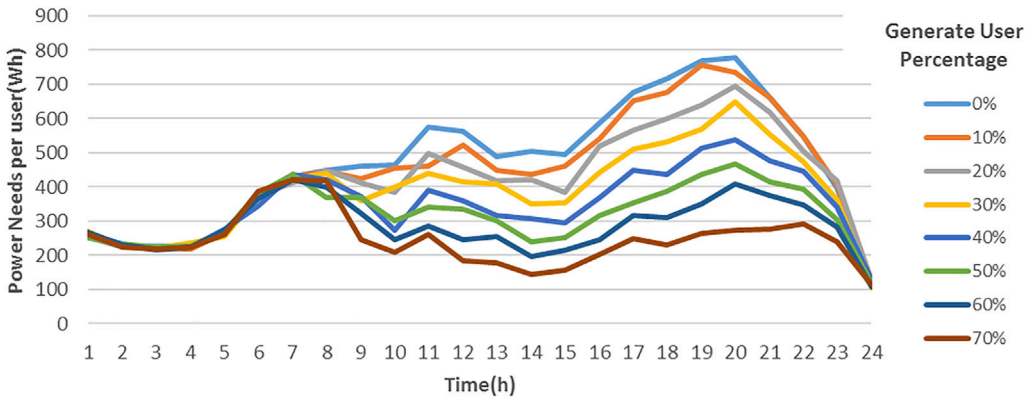


FIGURE 8
User power needs with different generated user percentages.

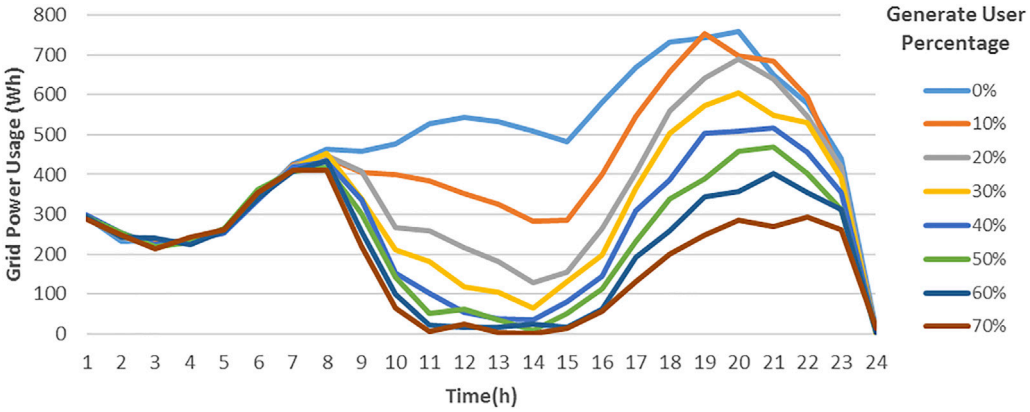


FIGURE 9
Grid power usage with different generated user percentages.

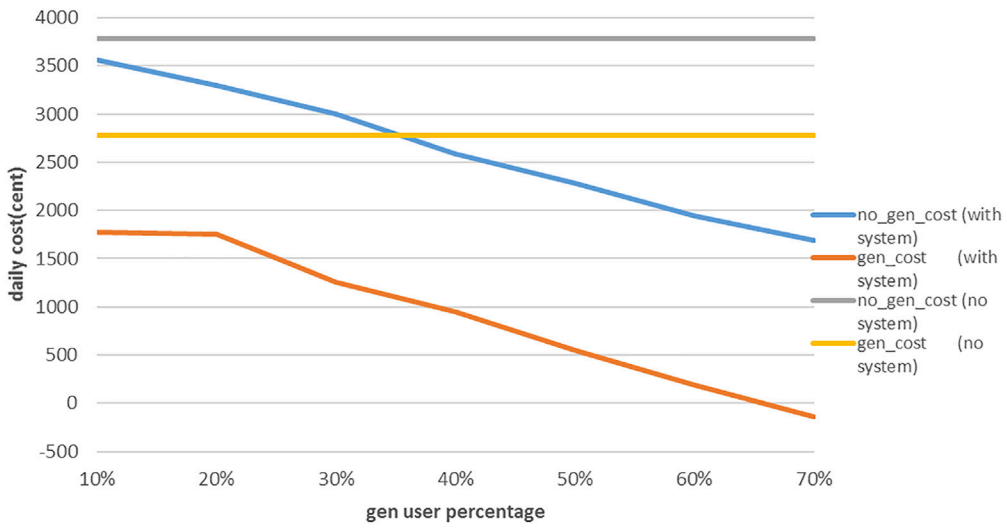


FIGURE 10
Average cost with different settings.

respectively, and 43.1% for all users. It shows that our system can help users reduce their spending on power usage.

5 Conclusion

Intelligent smart metering facilities and two-way communication infrastructure enable residential buildings to actively generate and trade energy on the demand side. However, the traditional centralized power management mode causes unreliability and lack of privacy in information exchange. Energy losses from inefficient management models produce large amounts of carbon emissions and lead to irreversible changes to the climate. In this study, we propose a privacy-preserving, efficient, and robust distributed energy exchange scheme supported by the smart contract, called CE-SDT. An efficient bidding model is first introduced for the demand side to achieve peer-to-peer trading of electricity in the real-time market. The entire exchanges of data are implemented in the smart contract to guarantee information safety and traceability with a decentralized scheme. The proposed solution achieves reliable and stable trading objects under the premise of information security. We deploy it in the simulation framework to test its feasibility. The result shows that our CE-SDT is stable and extendable. Also, it can help shift the load and peak power usage, thereby reducing costs for users. Even users without the power generation capacity can only pay 44% of the original electricity bill. We conducted a systematic and comprehensive applicability analysis of the proposed mechanism, and it further confirms that the system can be applied in a home energy management system. However, the process of blockchain mining requires a large number of hash operations, and in order to obtain more computational power, traditional blockchain computing requires a large amount of power. In future research, 1) we want to find out a way to conduct the verification work using a power grid bidding system instead of a smart contract. 2) We hope to change the way the blockchain reaches consensus. There can be a local division in the power grid itself, and the blockchain will reach a consensus in the region first and then send the results of the local consensus to the whole network for verification. In this way, we can make use of the surplus power in the power grid for mining and realize peak load shifting by transferring the consensus area. 3) We hope to continue to incorporate carbon quotas into our proposed blockchain-based distributed trading mechanisms to enable residents to directly obtain active incentives to reduce carbon emissions, thereby realizing greater environmental benefits.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

Conceptualization, BW and FZ; methodology, BW, JX, and FZ; software, BW; validation, BW and NW; formal analysis, BW and XL; investigation, JK, JX, and BW; resources, JK and JX; data curation, BW; writing—original draft preparation, BW and JK; writing—review and editing, PC, JW, NW, XL, and LL; visualization, BW and JK; supervision, PC, JW, and JX. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Research and Development in the key areas of Guangdong Province (2020B0101090001).

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/fenrg.2022.1091350/full#supplementary-material

References

Aitzhan, N. Z., and Svetinovic, D. (2018). Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams. *IEEE Trans. Dependable Secure Comput.* 15, 840–852. doi:10.1109/TDSC.2016.2616861

Alcarria, R., Bordel, B., Robles, T., Martín, D., and Manso-Callejo, M.-Á. (2018). A blockchain-based authorization system for trustworthy resource monitoring and trading in smart communities. *Sensors* 18, 3561. doi:10.3390/s18103561

Ali, M. S., Vecchio, M., Pincheira, M., Dolui, K., Antonelli, F., and Rehmani, M. H. (2019). Applications of blockchains in the internet of things: A comprehensive survey. *IEEE Commun. Surv. Tutor.* 21, 1676–1717. doi:10.1109/COMST.2018.2886932

Du, Y., Wu, J., Li, S., Long, C., and Paschalidis, I. C. (2019). Distributed MPC for coordinated energy efficiency utilization in microgrid systems. *IEEE Trans. Smart Grid* 10, 1781–1790. doi:10.1109/TSG.2017.2777975

Fan, T., He, Q., Nie, E., and Chen, S. (January 2018)., Vols 1-4. Bristol: Iop Publishing Ltd. doi:10.1088/1755-1315/108/5/052083A study of pricing and trading model of Blockchain & Big data-based Energy-Internet electricity, Proceedings of the 2017 3rd International Conference on Environmental Science and Material Application (esma2017),Chongqing, China 052083,

Gai, K., Wu, Y., Zhu, L., Qiu, M., and Shen, M. (2019). Privacy-preserving energy trading using consortium blockchain in smart grid. *IEEE Trans. Industrial Inf.* 15, 3548–3558. doi:10.1109/TII.2019.2893433

Hanna, R., Ghonima, M., Kleissl, J., Tynan, G., and Victor, D. G. (2017). Evaluating business models for microgrids: Interactions of technology and policy. *Energy Policy* 103, 47–61. doi:10.1016/j.enpol.2017.01.010

Hou, W., Guo, L., and Ning, Z. (2019). Local electricity storage for blockchain-based energy trading in industrial internet of things. *IEEE Trans. Industrial Inf.* 15, 3610–3619. doi:10.1109/TII.2019.2900401

Hua, W., Jiang, J., Sun, H., and Wu, J. (2020). A blockchain based peer-to-peer trading framework integrating energy and carbon markets. *Appl. Energy* 279, 115539. doi:10.1016/j.apenergy.2020.115539

Hua, W., and Sun, H. (September 2019). A blockchain-based peer-to-peer trading scheme coupling energy and carbon markets. Available at: https://www.webofscience.com/wos/alldb/full-record/WOS:000594184400122 (Accessed October 27, 2022).Proceedings of the 2019 2nd International Conference on Smart Energy Systems and Technologies (sest 2019) (New York: Ieee),Porto, Portugal

Jiawei, W., Jingxi, D., Huiqing, L., Yao, W., Shuangmei, J., Yanjuan, Z., et al. (October 2018). Green time-sharing electricity price mechanism and optimization based on demand response. (Available at: https://www.webofscience.com/wos/alldb/full-record/WOS:000468028300041 (Accessed October 27, 2022).Proceedings of the 2018 2nd Ieee Conference on Energy Internet and Energy System Integration (ei2) New York: Ieee), 210–214. Beijing, China

Li, H., Xiao, F., Yin, L., and Wu, F. (2021). Application of blockchain technology in energy trading: A review. *Front. Energy Res.* 9, 671133. doi:10.3389/fenrg.2021.671133

Li, Y., and Li, K. (2019). Incorporating demand response of electric vehicles in scheduling of isolated microgrids with renewables using a Bi-level programming approach. *IEEE Access* 7, 116256–116266. doi:10.1109/ACCESS.2019.2936487

Logenthiran, T., Srinivasan, D., and Shun, T. Z. (2012). Demand side management in smart grid using heuristic optimization. *IEEE Trans. Smart Grid* 3, 1244–1252. doi:10.1109/TSG.2012.2195686

Lu, X., Shi, L., Chen, Z., Fan, X., Guan, Z., Du, X., et al. (2019). Blockchain-based distributed energy trading in energy internet: An SDN approach. *IEEE Access* 7, 173817–173826. doi:10.1109/ACCESS.2019.2957211

Niyato, D., Xiao, L., and Wang, P. (2011). Machine-to-machine communications for home energy management system in smart grid. *IEEE Commun. Mag.* 49, 53–59. doi:10.1109/MCOM.2011.5741146

Park, L. W., Lee, S., and Chang, H. (2018). A sustainable home energy prosumer-chain methodology with energy tags over the blockchain. *Sustainability* 10, 658. doi:10.3390/su10030658

Pipattanasomporn, M., Kuzlu, M., and Rahman, S. (October 2018). A blockchain-based platform for exchange of solar energy: Laboratory-scale implementation. in Proceedings of the 2018 International Conference and Utility Exhibition on Green Energy for Sustainable Development (ICUE), 1–9.Phuket, Thailand doi:10.23919/ICUE-GESD.2018.8635679

S, J. I. (1963). Fine particles, thin films and exchange anisotropy. *Magnetism* 0, 271–350.

Shindell, D., and Smith, C. J. (2019). Climate and air-quality benefits of a realistic phase-out of fossil fuels. *Nature* 573, 408–411. doi:10.1038/s41586-019-1554-z

Su, J., Li, Z., and Jin, A. J. (2021). Practical model for optimal carbon control with distributed energy resources. *IEEE Access* 9, 161603–161612. doi:10.1109/ACCESS.2021.3130550

Surana, K., and Jordaan, S. M. (2019). The climate mitigation opportunity behind global power transmission and distribution. *Nat. Clim. Chang.* 9, 660–665. doi:10.1038/s41558-019-0544-3

Yang, Z., and Wang, L. (July2016). Demand Response Management for multiple utility companies and multi-type users in smart grid. in Proceedings of the 2016 35th Chinese Control Conference (CCC), 10051–10055.Chengdu, China doi:10.1109/ChiCC.2016.7554946

Yinan, L., Wentao, Y., Ping, H., Chang, C., and Xiaonan, W. (2019). Design and management of a distributed hybrid energy system through smart contract and blockchain. *Applied Energy*, 248, 390-405. doi:10.1016/j.apenergy.2019.04.132https://scholarbank.nus.edu.sg/handle/10635/191169 (Accessed October 28, 2022).

Yorozu, T., Hirano, M., Oka, K., and Tagawa, Y. (1987). Electron spectroscopy studies on magneto-optical media and plastic substrate interface. *IEEE Transl. J. Magnetics Jpn.* 2, 740–741. doi:10.1109/TJMJ.1987.4549593

Yuan, L., Dong, J., Wang, N., Chen, Z., and Gong, G. (May 2021). Nanjing, China: IEEE, 245. –248. doi:10.1109/iSPEC53008.2021.9735453Blockchain-based carbon allowance trading market construction, Proceedings of the 2021 IEEE Sustainable Power and Energy Conference (iSPEC)

Zeyringer, M., Price, J., Fais, B., Li, P.-H., and Sharp, E. (2018). Designing low-carbon power systems for Great Britain in 2050 that are robust to the spatiotemporal and inter-annual variability of weather. *Nat. Energy* 3, 395–403. doi:10.1038/s41560-018-0128-x

Zhang, Y., Yu, R., Xie, S., Yao, W., Xiao, Y., and Guizani, M. (2011). Home M2M networks: Architectures, standards, and QoS improvement. *IEEE Commun. Mag.* 49, 44–52. doi:10.1109/MCOM.2011.5741145

Zhu, Y., Koutra, S., and Zhang, J. (2022). Zero-carbon communities: Research hotspots, evolution, and prospects. *BUILDINGS-BASEL* 12, 674. doi:10.3390/buildings12050674

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 19 January 2023
DOI 10.3389/fenrg.2022.1115769

OPEN ACCESS

EDITED BY
Zhiqiang Geng,
Beijing University of Chemical Technology, China

REVIEWED BY
Xianning Meng,
Daegu University, South Korea
Yan Guo,
Laboratory of Nuclear Energy Chemistry (CAS), China
Yingying Jiang,
University for the Creative Arts, United Kingdom

*CORRESPONDENCE
Ying Wang,
DNHBLWTG@163.com
Jiajun Liu,
xianligong_88@163.com

SPECIALTY SECTION
This article was submitted to Process and Energy Systems Engineering, a section of the journal Frontiers in Energy Research

RECEIVED 04 December 2022
ACCEPTED 23 December 2022
PUBLISHED 19 January 2023

CITATION
Wang Y and Liu J (2023), Optimal capacity allocation of wind-light-water multi-energy complementary capacity based on improved multi-objective optimization algorithm.
Front. Energy Res. 10:1115769.
doi: 10.3389/fenrg.2022.1115769

Optimal capacity allocation of wind-light-water multi-energy complementary capacity based on improved multi-objective optimization algorithm

Ying Wang[1,2]* and Jiajun Liu[1]*

[1]School of Electrical Engineering, Xi'an University of Technology, Xi'an, China, [2]School of Electric Power, Yinchuan University of Energy, Yinchuan, China

As a basic industry for national economic development, the power industry is closely related to the overall economic and environmental development of China. At present, China is still dominated by thermal power generation. In order to reduce carbon emissions, promote the realization of the "double carbon" goal, and improve the level of clean energy utilization and the operating efficiency of the power system, a wind-light-water storage complementary power generation system is built, and a mathematical model of multi energy complementation is established. The minimum economic cost and the minimum battery capacity are proposed as the objective functions of system capacity configuration. Then a multi-objective evolutionary algorithm based on Pareto optimal space of the NDWA-GA and the PCA is proposed for optimal capacity allocation of multi energy complementary systems in this paper. Compared with the traditional multi-objective optimization algorithm, the correctness and effectiveness of the proposed method are verified. In addition, according to the actual research object, the optimal capacity configuration of the multi energy complementary system is given, which can guide the production and has an important promotion significance for energy saving and emission reduction.

KEYWORDS

multi energy complementary model, multi-objective evolutionary algorithm, capacity configuration, conserve energy, reduce emissions, production guidance

1 Introduction

At present, because of the energy shortage, serious environmental pollution increasingly and climate change in the world, the issue of energy development has become a hot topic of interest. Therefore, improving the rational utilization of energy can provide support for the sustainable development of social economy (Li et al., 2017). The electric power industry is one of the basic industries supporting the national production, and its power generation process has also brought serious problems such as energy consumption and excessive pollutant emissions. Therefore, improving the energy utilization efficiency and power generation efficiency on the power generation side, while minimizing pollutant emissions, is the key work on the power generation side.

At present, thermal power generation accounts for the largest proportion of installed capacity in Chinese power industry and its power mainly comes from burning coal, which results in the emission of pollutants and low efficiency. Therefore, compared with energy conservation, the thermal power industry should first consider the carbon emission reduction goal (Wang et al., 2013). In the past few years, the proportion of coal consumption has

decreased by 8.4%, and the proportion of non-fossil energy consumption has increased to 14.3%, of which the proportion of electricity in terminal energy consumption has increased to 25.5%. In order to achieve the "dual carbon" goal, a clean, low-carbon, safe and efficient energy system should be built to improve the utilization level of clean energy and the operating efficiency of the power system, and better play the role of multi energy complementation in ensuring energy security. In the 13th Five-Year Plan for Energy Development (National Development and Reform Commission and National Energy Board, 2016), it was pointed out clearly that centering on the "low-carbon" energy system, industrial reform was promoted and comprehensive energy utilization was promoted through continuous improvement of energy utilization efficiency. The state will accelerate the development of non-fossil energy, vigorously increase the scale of wind power and photovoltaic power generation, build a number of clean energy bases with multiple and complementary energy sources, and raise the share of non-fossil energy in total energy consumption to about 20 percent (State Council of the People's Republic of China, 2021). The National Development and Reform Commission and the National Energy Administration have also issued relevant policy document (National Development and Reform Commission and National Energy Board, 2021) clearly proposing "promoting multi energy complementation and improving the consumption level of renewable energy".

In order to reduce carbon emissions and improve the energy pattern dominated by thermal power plants in China, and also improve the application rate of clean energy represented by wind, light and hydropower, a wind-light-water storage complementary power generation system by clean energy is constructed, to establish a mathematical model of multi energy complementation, which can propose the minimum economic cost and the minimum battery capacity as the objective function of system capacity configuration. A multi-objective evolutionary algorithm based on Pareto optimal space of the NDWA-GA and the PCA is proposed for optimal capacity allocation of multi energy complementary systems in this paper. Compared with the traditional multi-objective algorithm, the results show that the proposed algorithm has better convergence and distribution. In addition, compared with the light-water storage model and the wind-water storage model, it is proved that the wind-light-water storage model proposed in this paper has the lowest economic cost under the premise of accurate prediction of hydropower station output. The capacity configuration results show that in February to March each year, the wind power generation can meet the load demand, light and water will be stored in the form of energy storage battery. When the system output cannot meet the load, the energy storage unit is used to supplement.

The rest of the paper is organized as follows: The second section introduces the research status of multi energy complementary system, capacity allocation of multi energy complementary generation system and multi-objective evolutionary algorithm. The third section introduces the topology of the wind-light-water storage multi-energy complementary system, and analyzes the wind-light-water complementary situation. The mathematical model of the multi energy complementary system is given in this section. In the fourth section, a multi-objective evolutionary algorithm based on the Pareto optimal space of the NDWA-GA and the PCA is proposed to carry out the optimal capacity allocation for the multi energy complementary system. In order to verify the feasibility of the wind-light-water storage model in practical application, this section constructs the light-water storage model and the wind-water storage model, and compares the economy of their capacity configuration and the capacity of the battery. According to the actual case, the optimal capacity allocation for the multi energy complementary system in this case is given. The fifth section is the conclusion.

2 Literature review

Multi-energy complementarity is generally based on the characteristics of a variety of energy sources, it is a way to integrate them and use them effectively. The random and intermittent nature of wind, light and water resources as well as their complementary characteristics in space and time make it possible to establish a wind-light-water multi-energy complementary power generation system. Heide et al. believed that renewable energy generation across Europe, it shows that strong seasonality and will dominate in the future (Heide et al., 2010). Combining with the existing characteristics of various energy sources in northwest China, Zhang analyzed the necessity, feasibility and rationality of multi-energy complementary development (Zhang, 2012). At present, the more mature development are the photovoltaic and wind power complementary system. The reference (Deetjen et al., 2018) developed a model for calculating optimal transmission, wind and photovoltaic capacity for a region in Germany. However, photovoltaic and wind power are complementary, and the output fluctuation is still large, so it is necessary to configure a certain energy storage system on this basis. The research object of the reference (Parastegari et al., 2015) was the optimal scheduling of the complementary optical storage energy. The optimal economic scheduling mathematical model was established by taking the optimal economic efficiency of the system as the objective function and integrating various power conditions and balance states, and the particle swarm optimization algorithm was applied to solve the problem, the reference (Yahyaoui et al., 2017) designed an independent photovoltaic energy storage system and successfully provided electricity for an irrigation system. A multi-objective optimization model of wind energy storage was proposed in the reference (Xu et al., 2018), which can be applied in wind farms to adjust output fluctuations and reduce wind abandonment phenomenon.

The cost of energy storage system is relatively high, and it is difficult to apply in large-scale production. Hydropower, as the leader of clean energy, has large storage capacity and strong regulation capacity. It is the current development trend to use it in the multi-energy complementary system. Because of the economic underdevelopment in Ethiopia, the total power supply is insufficient, the reference (Bekele and Tadesse, 2012) studied a small wind-light-water system to make up for the shortage of power supply in rural areas. Spain's Canary island climate is conducive to the development and utilization of wind and water energy, so a new wind-hydro system was proposed in the reference (Portero et al., 2015), and the results showed that its performance was significantly better than that of independent wind power systems. The research object of the reference (Liu et al., 2017) was the optimal scheduling of the three complementary energy sources of water, light and wind, on this basis, the system optimization model was established and the solution method was determined. The reference (Mahmoudimehr and Shabani, 2018) designed a PV/hydro hybrid power station for the energy system in the northern and southern

coastal areas of Iran and configured its capacity. Small hydropower stations in Guizhou province is taken as the research object in the reference (Zhang et al., 2018), and wind and photovoltaic power generation systems are combined to build a wind-light-water storage complementary power generation system. The configuration scheme of the system is discussed in remote mountainous rural areas with electricity (off-grid mode) and with grid-connected conditions (grid-connected mode) based on HOMER.

The economy of the power generation system is the basis for allocating the capacity of each power source, which is of great practical significance for ensuring the economy and reliability of the complementary power generation system. Therefore, domestic and foreign scholars have established capacity allocation models with the minimum total investment cost or the maximum investment income as the objective function (Chang, 2020). Based on the minimum system cost and considered the seasonal load variation, a capacity optimization configuration model of the wind-solar storage system was established (Sanajaoba and Fernandez, 2016). The results showed that wind-solar complementarity can reduce the energy storage requirements of the system. In reference (Dou et al., 2016), the evaluation indexes of distributed capacitor configuration scheme and energy storage system configuration scheme is discussed, that is, the former is evaluated from economy, reliability and environmental protection, and the latter is evaluated from functionality and economy. Taking Longyangxia Hydropower Station in China as an example, a capacity configuration model of water-light complementary system was proposed to maximize the net income in the whole life cycle, considering the short and long term decisions of hydropower operation (Ming et al., 2017). In reference (Abdin and Merida, 2019), a wind-light multi-energy complementary model was constructed based on HOMER Pro solving for the lowest economic cost of energy and capacity configuration for each in five different locations around the world. The results showed that hydrogen energy is more economically advantageous than battery storage in off-grid energy systems. A wind/light/storage grid connected microgrid was built (Li et al., 2020). Based on considering the life-cycle cost economy, it compared four energy storage modes: single battery energy storage, hydrogen storage hybrid battery energy storage, single lithium battery energy storage, and single liquid flow battery energy storage. The results showed that the hybrid energy storage system composed of hydrogen storage and battery had more significant advantages in energy utilization efficiency and the economy of the microgrid. In reference (Alghussain et al., 2020), scholars set the lowest energy cost and maximized the proportion of renewable energy use as the objective function, comparing the capacity configuration of the power grid system under four different energy storage scenarios, which were without energy storage modules, pumped storage, hydrogen fuel cells, and pumped storage hybrid hydrogen fuel cells. The results showed that the proportion of supply and demand increased from 46.5% to 89.4%, and the proportion of renewable energy increased from 62.6% to 91.8% when the hybrid energy storage system with pumped storage and hydrogen fuel cell was used.

The hybrid energy storage of batteries and super capacitors were applied in a grid-connected wind-light complementary power generation system with a two-step capacity configuration (Hu et al., 2016), and its capacity configuration is divided into two steps. First, the HOMER software is used to obtain preliminary results with the minimum system cost as the optimization objective function, and then genetic algorithm is used to further optimize the hybrid energy storage system. In reference (Sheng and Zhang, 2017), an improved binary bat algorithm with the continuous optimization algorithm was directly applied in the local optimization problem generated by the wind photovoltaic diesel battery microgrid capacity configuration optimization model. The comparison with genetic algorithm and particle swarm optimization showed that the proposed algorithm achieved better results in solving effect and convergence speed. In the reference (Yu et al., 2019), a grid-connected wind-light complementary system was constructed using pumped storage power plants as energy storage devices, and a variable-step cyclic discrete algorithm was proposed in which the number of wind turbines and photovoltaic arrays were discrete integers and the pumped storage capacity was considered as a discrete of a certain step size. In reference (Liu et al., 2020), photovoltaic energy storage diesel micro grid was taken as the research object in southwest of China, the state of charge of the energy storage system and the reliability of the self-power supply of the system were taken as constraints, and the particle swarm optimization (PSO) algorithm was used to solve capacity allocation. The relationship between the theoretical output of wind farms and retired energy storage batteries was comprehensively considered (Wang et al., 2020), and the charging and discharging threshold range of ideal retired batteries with a piecewise probability distribution function were set, which reduced the recovery cost of retired batteries. At the same time, the cycle life consumed by the hierarchical control strategy was reduced by 27.7% compared with the overall energy storage control strategy. The actual village coal to electricity project in Qinhuangdao area was taken as the research object (Zhao et al., 2021), considering the complementary system composed of thermal storage heating and photovoltaic, proposed a joint configuration method based on PSO algorithm. The results showed that the investment and operation costs of photovoltaic and thermal storage heating equipment were the least. In reference (Chen et al., 2021), researchers introduced the incentive demand side response in the optimal configuration of wind light hydrogen multi energy complementary microgrid, optimized the load curve using mixed integer programming, and solved the optimal configuration based on the PSO algorithm. The configuration results showed that with the reduction of photovoltaic and wind turbine construction costs, the equivalent annual cost of the system could be significantly reduced. In addition, the planning method, enumeration method and hybrid method can also be used to solve the optimal capacity of each source in the complementary generation system.

There are multiple objective functions in the mathematical model of capacity allocation of multi energy complementary systems, which is a typical multi-objective optimization problem that can be solved based on multi-objective evolutionary algorithms. In recent years, multi-objective evolutionary algorithms have been developed rapidly. The multi-objective evolutionary algorithm based on Pareto dominance uses the non-dominated sorting and diversity maintenance mechanism to guide the population to the Pareto front. Typical algorithms include non-dominated sorting genetic algorithm II, Pareto envelop-based selection algorithm-II (PESA-II) (Corne et al., 2001) and Strength Pareto evolutionary algorithm II (SPEAII). Although the above algorithms can solve many multi-objective optimization problems, when the objective function increases or Pareto front characteristics are complex, the solving efficiency of multi-objective evolutionary algorithm based on

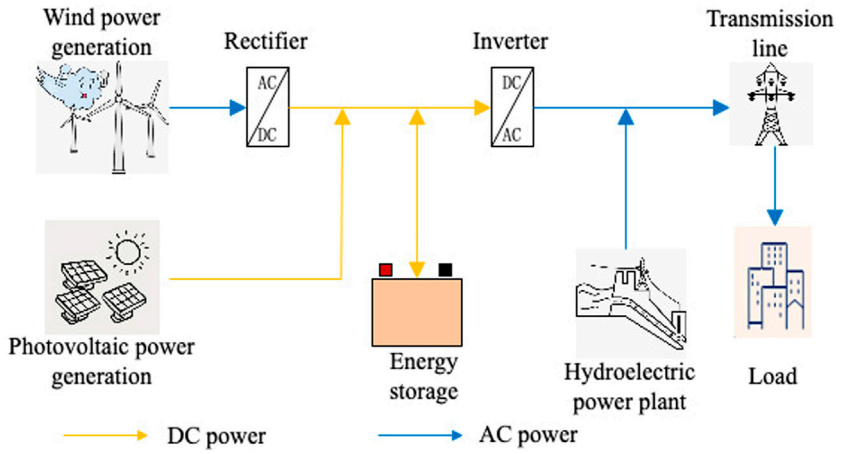


FIGURE 1
Structure diagram of wind/PV/hydropower/storage coupling system.

dominance relationship becomes low. With the increase of objective function, the selection pressure of individuals on Pareto front drops sharply. When the objective function is more than a certain threshold, almost all individuals of the same generation are non-dominated, which makes it difficult for evolutionary algorithms based on Pareto dominance to distinguish these individuals. In order to solve the multi-objective optimization problem of Pareto frontier with complex characteristics, many improved algorithms have been proposed including fuzzy Pareto dominant algorithm (He et al., 2014), Vector angle-based evolutionary algorithm (Xiang et al., 2017) and Many-objective evolutionary algorithm based on a rotation of grid (Zou et al., 2018). The common feature of these above methods is that the preference for a specific area of the Pareto front is introduced, so that the population can converge to a certain part of the Pareto front. However, it is difficult for these above methods to find the entire Pareto front.

Through the decomposition multi-target evolution algorithm transforming multi-target optimization problems into a series of single target optimization sub-problems, the evolution algorithm is used to optimize these sub-problems based on a certain number of adjacent problems. Representative algorithms include non-dominated sorting genetic algorithm III (NSGA-III) (Deb and Jain, 2014), and modified particle swarm optimization based on decomposition with different ideal points (MPSO/DD) (Qin et al., 2020).

In order to make full use of the complementary characteristics of each power source, weaken the impact of the instability of new energy output on the power grid, improve the application rate of clean energy represented by wind, light and hydropower, and improve the energy pattern dominated by thermal power plants in China, a complementary wind-light-water storage power generation system is constructed and a multi-objective evolutionary algorithm based on the NDWA-GA and the PCA with Pareto optimal space is proposed for the optimal capacity allocation of the multi-energy complementary system in this paper. Through the above method the dependence on thermal power generation and carbon emissions can be reduced, which is of great significance for the implementation of the national energy conservation and emission reduction and the "double carbon" policy.

3 Wind, light, water and storage multi-energy complementary model

3.1 Multi-energy complementary system topology design

Because wind power and photovoltaic power generation are characterized by randomness, intermittency, and volatility, especially the daily output characteristics of wind power are just opposite to the load, it is difficult to directly connect them to the large power grid for users. The units in the hydropower station have the ability to start and stop quickly, which can be well used to supplement the photovoltaic and wind output characteristics and can effectively solve the problem of new energy grid connection consumption. The structure of the wind-light-water storage coupling system is shown in Figure 1, which mainly includes power generation module, energy storage module, energy conversion module and user module. The power generation module consists of wind turbine, photovoltaic turbine and hydroelectric turbine, which is a power generation system composed entirely of clean energy without carbon emission. The energy conversion module includes rectifier and inverter. The rectifier converts the alternating current generated by the wind turbine into direct current, and the direct current generated by the photovoltaic power is transmitted through the direct current transmission line. The energy is transferred in both directions between the battery in the storage module and the DC transmission line, and the DC power is inverted and sent to the customer through the AC transmission line. The AC power from the hydropower station is transmitted through the AC transmission line.

3.2 Water-wind-light complementarity analysis

According to the mechanism of wind and photovoltaic power generation, its output is influenced by the natural environment (such as wind power level, light intensity, and temperature.) and is highly

volatile. Wind power generation relies on wind generating units, which converts wind kinetic energy into electric energy by rotating windmill blades driven by wind. Its power generation is mainly affected by uncertain environmental factors such as wind speed, sunshine, terrain and air pressure. Photovoltaic power generation uses photovoltaic effect to convert solar energy into electricity. Its power output is non-linear and mainly depends on the external environment such as light intensity and temperature, which makes photovoltaic power generation has significant weather sensitivity. In clear weather, photovoltaic output is generally consistent with the changing trend of light intensity and temperature, showing a parabolic shape. However, under the condition of cloudy or rainy weather, photovoltaic output shows an obvious decline.

The Hanzhong area in southern Shaanxi Province is taken as the research object in this paper. This region is the subtropical climate with abundant rainfall and sufficient power output from hydropower plants. Due to its location at the southern foot of the Qinling Mountains, the wind power rises along the mountains, resulting in increased wind power and sufficient wind energy. According to the distribution of wind speed and air density during the year, the local wind turbine output is maximum in winter and spring, and minimum in summer. The climate is humid and mild, with four distinct seasons, long frost-free period, warm and dry spring, cool and wet autumn, and abundant light energy. The average annual flat source temperature is about 15°C, the sunshine hours are between 1495.6 h and 1836.2 h, and the frost-free period is more than 8 months on average, so there are more superior conditions for photovoltaic wind turbine power generation.

For the wind-light-water storage complementary power generation system, the battery capacity configuration is relatively small, mainly for smoothing the fluctuation of the short-time wind-light power, which can avoid the phenomenon of frequent start and stop of the hydropower unit due to the fluctuation of the total wind-light power. However, the hydropower unit is mainly used to maintain the balance of the system power supply, the power balance between the power supply system and the load can be expressed by the following formula.

$$P_L(t) = P_{PV}(t) + P_W(t) + P_{HP}(t) - P_B(t) \quad (1)$$

where, $P_L(t)$–load power; $P_{PV}(t)$–photovoltaic power output; $P_W(t)$–wind turbine power output; $P_{HP}(t)$–hydroelectric power output; $P_B(t)$–battery discharge power.

In terms of system power complementarity, according to the analysis of the natural complementary characteristics of water-wind-light in the region, the local hydropower resources are very unevenly distributed in time, with sufficient water resources in the rainy season and scarce in the dry season, which makes the power grid in the dry period having a huge risk of energy supply shortage. The wind power and photovoltaic energy supply and demand are in line with hydropower's dry season and flood season, which can provide considerable power support for the dry season when hydropower is insufficient, effectively alleviating the contradiction of seasonal supply and demand balance, which is a useful complement to hydropower on a long-term scale. In terms of capacity complementary characteristics, hydropower with certain regulation capacity can use the reservoir storage capacity to smooth out the natural water changes to achieve rapid and accurate control of power output, which has good capacity characteristics. Wind power photovoltaic is affected by natural factors with strong uncertainty, wind power daytime low night high power characteristics make it having a significantly anti-peak characteristics, large-scale wind power access often increase the pressure of grid peaking at the same time also bring greater fluctuations in system power. However, the complementary operation of wind power and photovoltaic can not only effectively alleviate the intensification of peak load regulation during the day, but also effectively make hydropower transfer its low power to peak load during the low load period, converting the power benefit into capacity benefit. So that hydropower in the peak load period has sufficient capacity space for full peaking operation, to improve the hydropower peaking capacity at the same time also effectively suppress the short-term volatility of wind power photovoltaic. When the wind-light output can meet the load demand, hydropower units do not carry out power generation operations, the use of reservoirs to store natural water. When the wind-light output does not meet the load demand, the use of hydropower devices to supplement the missing power, and when the full generation of hydropower still cannot meet the load demand, load cutting operations should be carried out.

In summary, the wind-light-water resources in the region have strong complementary characteristics throughout the year, so it is suitable to add wind-light storage power generation devices on the basis of existing hydropower plants. Not only the extraction of fossil energy can be reduced, but also the utilization rate of clean renewable energy can be effectively increased. The combined operation of wind, light and water mainly uses the energy storage characteristics of the hydropower plant itself and the flexible start-stop feature to calm the sharp fluctuations of wind power and photovoltaic, hydropower and wind-light combined operation can improve the impact of wind power and photovoltaic volatility on the power system, reduce wind power abandonment. On the other hand, it can also make up for the lack of seasonal characteristics of hydropower to a certain extent by using the complementarity between the two, which provides an important support for future research on multi-energy complementary coordination.

3.3 Mathematical model of multi-energy complementary system

In this paper, wind-light-water storage is used as the research object to configure the optimal wind-light storage power supply capacity for existing hydropower plants to meet the regional power supply demand. According to the characteristics of each power source within the power supply system, a capacity configuration model is established, which takes power system investment cost and battery capacity minimization as the target and power balance, SOC, and hydropower unit as the constraints. The calculation results are also evaluated and analyzed in terms of load shortage rate, battery charging and discharging times and the number of starts and stops of hydropower units.

3.3.1 Systematic investment model

The prerequisite for capacity configuration of the system is to calculate the investment cost of the system. The system investment mainly includes wind turbines, photovoltaic installations, and battery storage systems. Wind turbines and PV installations have a long service life (generally considered to be up to 20–25 years for wind turbines and 25 years for PV installations), so the time value of the fund needs to be considered. In addition, when PV installations exceed

their useful life, the power generated is only about 80% of their rated power, which affects the power generation revenue. The investment costs of PV systems, wind power systems and energy storage systems are described in turn in the following section.

1) Photovoltaic systems

In the construction of a PV system, its initial construction cost needs to be considered and equated to years (i.e., the average annual initial investment cost). In addition, PV power is DC power, and in order to convert it into AC power that can be used directly by the system, a converter matching the power generation capacity needs to be purchased, which likewise needs to be equated to years (i.e., the annual equivalent cost of the PV converter). During the working of the system, its implementation of strict maintenance measures is needed, so the operation and maintenance costs of the system also need to be considered. In summary, the investment cost of PV system mainly considers the average annual initial investment cost C_{pv} of PV system, the operation and maintenance cost C_{PVOM} of PV system and the annual equivalent cost C_{PVCON} of PV converter.

The average annual initial investment cost of a PV system can be expressed as follows.

$$C_{pv} = \frac{P_{pv}U_{pv}\gamma(1+\gamma)^{lpv}}{(1+\gamma)^{lpv}-1} \tag{2}$$

where, P_{PV}–Total PV capacity; U_{pv}–Installed cost of PV unit capacity; γ–Interest rate; l_{pv}–Life of PV installation.

The annual operation and maintenance (O&M) cost of a PV system can be expressed as the product of the PV capacity and the O&M cost per unit of capacity.

$$C_{PVOM} = P_{pv}M_{pv} \tag{3}$$

where, M_{pv}–unit capacity PV O&M costs.

The annual equivalent cost of a PV converter is related to the equipped PV capacity and can be expressed as the following.

$$C_{PVCON} = \frac{P_{pv}U_{PVCON}\gamma(1+\gamma)^{L_{PVCON}}}{(1+\gamma)^{L_{PVCOM}}-1} \tag{4}$$

where, U_{PVCON}–PV converter unit power cost; L_{PVCON}–PV converter lifetime.

2) Blower system

Consistent with the PV system, the investment cost of the wind turbine power generation system is mainly the average annual initial investment cost C_w of the wind turbine system, the operation and maintenance cost C_{WOM} of the wind turbine system and the annual equivalent cost C_{WCON} of the wind turbine converter.

The average annual initial investment cost of the wind turbine system can be expressed as the following.

$$C_w = \frac{P_W U_W \gamma(1+\gamma)^{lw}}{(1+\gamma)^{lw}-1} \tag{5}$$

where, P_W–total capacity of the wind turbine system; U_W–installed cost per unit capacity of the wind turbine system; l_w–life of the wind turbine.

The annual operation and maintenance cost of the wind turbine system can be expressed as the product of the capacity of the wind turbine system and the operation and maintenance cost per unit capacity.

$$C_{WVOM} = P_W M_W \tag{6}$$

where, M_W–unit capacity wind turbine operation and maintenance costs.

The wind turbine should be installed with the same power converter device, so the equivalent annual cost of the converter should also be included in the primary investment in the wind turbine system.

$$C_{PVCON} = \frac{P_{pv}U_{PVCON}\gamma(1+\gamma)^{L_{PVCON}}}{(1+\gamma)^{L_{PVCOM}}-1} \tag{7}$$

where, U_{PVCON}–PV converter unit power cost; L_{PVCON}–PV converter lifetime.

3) Energy Storage Systems

Consistent with PV and wind turbine systems, the investment cost of the energy storage systems are mainly the average annual initial investment cost C_B for energy storage systems, the operation and maintenance cost C_{BOM} for energy storage systems, and the annual equivalent cost C_{BCON} for converters.

Remaining useful life (RUL) of the battery in energy storage system is closely related to its own operating conditions and external ambient temperature. Overcharging and discharging of the battery, overvoltage, and high temperature of the external environment will reduce the RUL of the battery, and even require frequent battery replacement, which will inevitably increase the expenses and impact on the investment. Battery performance degradation is a complex multidisciplinary coupled problem of electrochemistry and electro-physics, and it is difficult to establish a unified RUL prediction model with universal applicability. In order to simplify the calculation, multiple overcharging and discharging will shorten the life of the battery in this paper, so the RUL of the battery is simplified to the number of times with the battery being charged and discharged to equivalently describe the life of the battery. At this time, the average annual investment cost of the battery can be calculated based on the average number of cycles per year.

$$C_B = \frac{N_E}{N_B}(U_{BP}P_B + U_{BE}E_B) \tag{8}$$

where, N_E–battery annual average charge and discharge times, U_{BP} -battery unit power cost, P_B -battery rated power, U_{BE}–battery unit energy cost, E_B–battery rated capacity and N_B–battery cycle life.

The annual O&M cost of an energy storage system can be expressed as the product of the capacity of the energy storage system and the O&M cost per unit of capacity.

$$C_{BOM} = P_B M_B \tag{9}$$

where, M_B–unit power battery O&M cost.

The installation of the energy storage device should be configured with a converter device with the same total power, so the equivalent annual cost of the converter should also be included in the primary investment of the energy storage system.

$$C_{BCON} = \frac{P_B U_{BCON} \gamma (1+\gamma)^{L_{BCON}}}{(1+\gamma)^{L_{BCON}} - 1} \tag{10}$$

where, U_{BCON} - battery converter unit power cost; L_{BCON} - battery converter life.

3.3.2 Capacity optimal allocation model

In this section, the capacity optimization allocation model of the wind-light-water storage system will be studied, mainly including the objective function, constraint function and evaluation index of the system.

1) Objective function

In this section, the objective function is defined as the minimum total system investment cost and the minimum battery capacity. The total system investment cost C minimization can be expressed as:

$$\min C = \min (C_A + C_{OM} + C_{CON}) \tag{11}$$

where, C_A - total annual initial investment cost of wind-light storage power, C_{OM} - annual operation and maintenance cost of wind-light storage power, and C_{CON} - annual equivalent cost of wind-light storage converter.

$$C_A = C_w + C_{pv} + C_B \tag{12}$$

$$C_{OM} = C_{WOM} + C_{PVOM} + C_{BOM} \tag{13}$$

$$C_{CON} = C_{WCON} + C_{PVCON} + C_{BCON} \tag{14}$$

When configuring the system capacity, wind power, photoelectricity and hydropower should be made to meet the load power demand as much as possible to avoid repeated charging and discharging of the battery which affects the battery life and system investment cost. In the constraint of minimum battery capacity, it is first necessary to define the power mismatch ratio.

$$\Delta (t) = \gamma \left[P_{pv}(t) + P_W(t) + P_{hp}(t) - P_{bat}(t) \right] - P_L(t) \tag{15}$$

where, γ - new energy relative load ratio, $P_{pv}(t)$, $P_W(t)$, $P_{hp}(t)$, - photovoltaic, wind power, hydropower output power at each moment, P_{bat} - battery charging and discharging power, and $P_L(t)$ - load power.

The energy stored in the battery is not constant, the size of the energy changes from time to time with the process of storing or releasing energy, so the size of its energy is a function of time. The maximum energy that can be stored in the storage system is related to the size of its own capacity, which can be expressed in the ideal case as the following.

$$H\_\text{store}(t) = H\_\text{store}(t-1) + \begin{cases} \eta_{mi} \Delta (t) \\ \eta^{-1} \text{oun} \Delta (t) \end{cases} \tag{16}$$

where, $H\_store(t)$–relation of energy with time in the ideal case.

When wind, light and water power generation meet the consumption of energy storage devices and load demand, there is still a surplus will be stored in the battery, and the capacity of the battery increases with time.

$$E_H = \max_T \left(H\_store(t) - \min_{t' \geq t} H\_store(t') \right) \tag{17}$$

Then the objective function to ensure the minimum capacity of the battery is the following.

$$E_H = \min\{\max (H\_store(t) - \min H\_store(t'))\} \tag{18}$$

2) Constraints

In this section, three constraints including instantaneous power constraint, state of charge (SOC) constraint and supply reliability constraint are set.

The instantaneous power constraint defines the daily energy inequality measure W_i as:

$$W_i = E_B(i-1)\eta_{\text{in}} + E_{W(i)} + E_{PV(i)} + E_{HP(i)} - Q_{L(i)} \\ W_i > 0 \tag{19}$$

where, $E_B(i\text{-}1)$ - the previous day's battery surplus power, η_{in} - battery charging efficiency, $E_{W(i)}$, $E_{PV(i)}$, $E_{HP(i)}$ - the total power generation of wind, light and water power sources on that day, and $Q_{L(i)}$ - the total load on that day.

The energy storage system charge state constraint SOC can be expressed as the following.

$$\text{SOC} = \frac{H_C(t)}{E_H} \tag{20}$$

where, $Hc(t)$ - the actual capacity of the battery storage device can store.

$$C_{low} \leq SOC \leq C_{up} \tag{21}$$

The reliability of power supply can be expressed as the following.

$$LPSP = \frac{\sum_{t=0}^{T} Time(P_{out}(t) < P_L(t))}{T} \tag{22}$$

where, $P_{out}(t)$ - the total output of the wind-light-water system at the moment t, and $P_L(t)$ - the load demand at the moment.

3) Assessment indicators

In this paper, two objective functions are established as the minimum economic cost and the minimum battery capacity, so the Pareto optimal solution set is approximated as a curve on a two-dimensional plane or uniformly distributed around a curve. In order to select the optimal solution from the optimal solution set as the solution result of this problem, the evaluation scheme is developed as following. In the two-dimensional right-angle coordinate system, the horizontal axis is the economic cost and the vertical axis is the battery capacity, at this time each solution in the optimal solution set can correspond to a point on the two-dimensional plane. Then all the candidate solutions are put on the two-dimensional plane and the Euclidean distance to the origin is calculated. Finally, the solution with the smallest distance is the optimal solution. The distance is calculated by Eq. 23.

$$d_i = \sqrt{f_{1,i}^2 + f_{2,i}^2} \tag{23}$$

where, d_i - the distance of the ith solution from the origin, $f_{1,i}$ - The ith solution corresponds to the value of objective function one, and $f_{2,i}$ - The ith solution corresponds to the value of objective function two.

4 Analysis of algorithms based on multi-objective evolutionary algorithms

In this section, the capacity of wind-light-water storage units are configured to make full use of the complementary characteristics of each power source and weaken the impact of the instability of new energy output on the grid. Based on the wind-light-water storage mathematical model established in the previous paper, the most economical storage units of wind-light are configured for existing hydropower plants based on an improved multi-objective evolutionary algorithm and compared with the traditional multi-objective optimization algorithm.

4.1 Improved multi-objective evolutionary algorithm

In this paper, an improved multi-objective optimization algorithm is proposed, which mainly consists of non-dominated dynamic weight aggregation by using a genetic algorithm (NDWA-GA) and PCA-based Pareto optimal space. The proposed NDWA-GA algorithm is used to search for a given set of Pareto optimal solutions and learn the Pareto optimal space from the found Pareto optimal solutions.

1) Non-dominated dynamic weight aggregation by using a genetic algorithm

Individuals with good suitability in single objective optimization have difficulty in convergence in the search space of the corresponding multi-objective optimization problem. The NDWA-GA proposed in this paper adopts the non-dominated sorting method, which is the most commonly used method in multi-objective optimization to solve this problem.

The NDWA-GA uses non-dominated sorting for solution selection, that is, before the selection operation is executed, the populations are sorted according to the dominant and non-dominated relationships, which can be outlined as the following steps. First, the individuals in the parent and replacement populations are combined, and all individuals are sorted according to the Pareto dominance relationship. The non-dominated solutions are selected with the marker order of 1. Then, based on the individuals in the first non-dominated front, the non-dominated solutions are selected among the remaining individuals with the marker order of 2 (that is the second non-dominated front), and so on until every individual is marked to form multiple dominance front arrays. The individuals in each non-dominated front after sorting are considered to be non-dominated by each other. The individuals on the first non-dominated front are considered to be the best, and the former non-dominated front dominate the latter non-dominated frontier.

The NDWA-GA algorithm search process can be summarized as follows. Given the objective function $f(x)$ to be optimized and initialize the maximum number of iterations T and the population size L. Subsequently, the weight vector is randomly generated and the population is initialized. After the iteration, the non-dominated solution is selected from the file as the result of the algorithm search. Each iteration of the NDWA-GA algorithm consists of four parts.

Step 1, periodically transform $f(x)$ into a single-objective optimization problem using the weight vector *ind* and encode it.

Step 2, calculate individual fitness based on the encoding results and classify individuals according to their dominance relationships. Then generate offspring using the genetic operator.

Step 3, count the non-dominated solutions into the file.

Step 4, perform iterations and select individuals with good convergence performance from the population as parents to participate in the next iteration.

2) Pareto optimal space based on the PCA

Unlike traditional PCA which uses the mean to populate the non-dominant components, the median value is used to populate the non-dominant components in this paper. The reason is that non-all non-dominated solutions happen to lie in the Pareto optimal front surface. In dimensionality reduction, these non-dominated solutions that do not lie in the Pareto optimal front plane are called outliers. Removing outliers from the entire dataset is not an easy task, and outliers usually affect the results. In summary, using the median will generate a more accurate Pareto optimal space than using the mean.

The basic process of the PCA-based Pareto optimal space algorithm can be outlined as follows.

First, the matrix representation of the Pareto optimal solution takes the form of $M \in R^{|P|\times n}$, where each row represents a solution and each column represents the same component of all solutions. The mean of M is calculated and made to save M minus its own mean as the new M.

Second, the eigenvalues of the covariance matrix of M are calculated, and the principal components are found according to the given threshold.

Third, the median of each column of M is calculated, and then the lower and upper limits of x are set to the corresponding medians. Finally the Pareto optimal subspace is generated by sampling all points in the updated M.

The overall flow chart of the algorithm is shown in Figure 2.

The Pareto optimal solution set can be found based on the NDWA-GA. The convergence of the solution set can be promoted by dimensionality reduction of the optimal space. Then, the diversity problem of the solution set is dealt with by solving a set of single-objective optimization problems in the optimal Pareto subspace. Finally, the stability of the optimal solution is guaranteed not to be eliminated in the iterative process.

4.2 Capacity configurations

The load data of the region in recent years show that the monthly average load curve in a year has certain fluctuations, with the peak value of monthly average load being about 1084021 MW and the lowest value of daily average load being 518037.8 MW. The electricity load mainly includes industrial electricity, urban transportation, agriculture/forestry/pastoralism/fishery/water conservancy, lighting, and household appliances. The load change curves of two consecutive typical years is shown in Figure 3.

In this paper, the parameters related to each power source of wind-light-water storage in the complementary power generation system are set as summarized in Table 1.

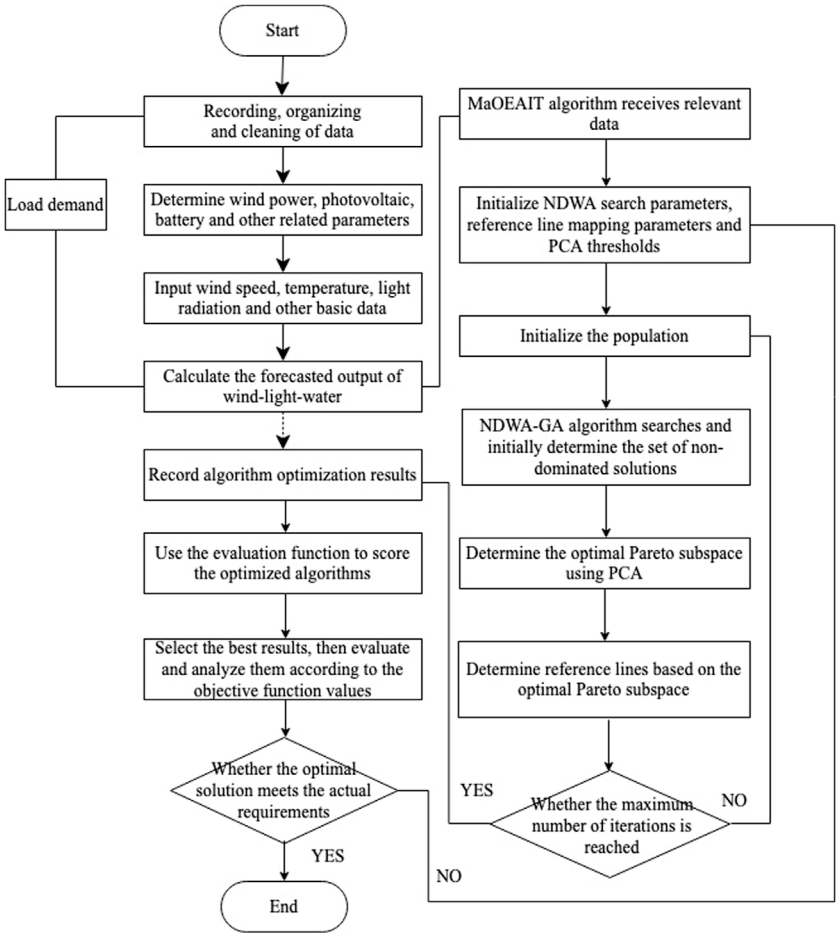


FIGURE 2
Flow chart of independent two-stage multi-objective evolutionary algorithm.

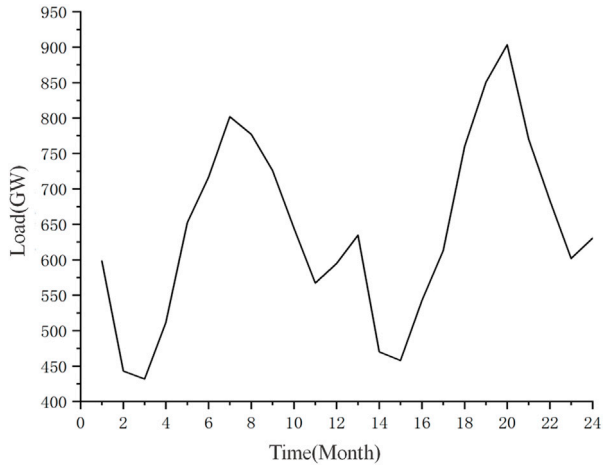


FIGURE 3
Annual load change curve.

According to the capacity optimization configuration model, the output power of wind power and photovoltaic is solved by using monthly climate data, and the capacity configuration of four power sources of wind, light, water and storage is optimized according to the load demand in each time period based on the multi-objective evolutionary algorithm proposed in the previous paper. Considering the energy consumption characteristics of the battery device when charging and discharging, it is assumed that the charging and discharging efficiency of the battery $\eta_{in} = \eta_{out} = .93$. In the algorithm, the initial number of individuals is set to 100, the reference mapping line is set to 6,000, the non-dominated dynamic weight aggregation is set to 20000, the maximum number of iterations is set to 20000, and the crossover rate and variation rate are set to 1.0 and 1/3, respectively. The simulated binary crossover (SBX) and polynomial mutation (PM) are used as the crossover and mutation operators, respectively.

Setting the new energy relative load generation ratio $\gamma = 1.1$ in the region, the optimal configuration of the capacity of the four power sources in the complementary power generation system in the current time scale is shown in Table 2, and the corresponding specific capacity and investment cost values of the wind-light storage are shown in Table 3.

The loss of power supply probability (LPSP) is an important index to evaluate the reliability of distributed complementary power generation system. To improve the reliability of power supply of the system means reducing the rate of load outages.

TABLE 1 The cost of each power supply and related parameters.

Symbol	Name	Unit	Value
U_W	Unit price of wind turbine purchase	yuan/kW	4,800
U_{PV}	Unit price of photovoltaic purchase	yuan/kW	6,000
U_{BP}	Unit price of energy storage power	yuan/kW	1550
M_W	Unit price of wind turbine operation and maintenance	yuan/kW	100
M_{PV}	Unit price of photovoltaic O&M	yuan/kW	40
M_{HP}	Unit price of hydropower O&M	yuan/kW	80
U_{WCON}	Unit price of wind turbine converter	yuan/kW	500
U_{PVCON}	Unit price of photovoltaic converter	yuan/kW	450
U_{BCON}	Unit price of battery converter	yuan/kW	160
γ	Interest rate	—	.12
L_W/L_{PV}	Life	year	25
C_{ci}	Cut-in wind speed	m/s	3
V_r	Rated wind speed	m/s	11
U_{co}	Cut-out wind speed	m/s	25
H	Hub height	m/s	80
$I_{SC(nom)}$	Short circuit current	A	8.33
η_{in}/η_{out}	Charge/discharge efficiency	—	.93

TABLE 2 Optimal allocation of unit quantity.

LPSP	Wind power configuration/unit	Photovoltaic configuration/thousand pieces	Battery configuration/block	Total cost/million yuan
0	349	300	7,490	1,017
.01	365	337	6,581	1,058
.02	332	383	8,132	910
.03	390	225	5,634	1,045
.04	300	214	10365	1,016
.05	348	369	7,358	1,031

TABLE 3 Optimal allocation of unit capacity.

LPSP	Wind capacity/MW	Photovoltaic capacity/MW	Battery capacity/MW
0	698	90	749
.01	730	102	658
.02	664	115	813
.03	780	68	563
.04	600	65	1,036
.05	696	111	736

4.3 Algorithm comparison

Unlike single-objective optimization problems, solving multi-objective optimization problems often yields a set of approximate Pareto optimal solution sets rather than individual solutions. Therefore, it is not possible to determine the performance of the algorithms directly by comparing the magnitude of the function values of the solutions obtained by the algorithms. In order to compare

TABLE 4 Hyper parameters selected by different algorithms.

Algorithm	Parameter settings
NSGA-III	The number of populations n = 100, polynomial variation factor p_m = 1/n, crossover probability η_c = 30, and variation probability η_m = 20.
RvEA-iGNG	The number of populations n = 100, neuron movement learning rate εb = .2, adjacent neuron movement learning rate εn = .006, maximum neuron age amax = 50, crossover distribution index n_c = 20, variation distribution index n_m = 20.
MSOPSII	The number of populations n = 100, differential evolution operator DE = .6, and crossover rate C = .8.

different algorithms, the Inverted Generational Distance (IGD) evaluation metric is introduced. The convergence performance and distribution performance of the algorithm are evaluated by calculating the minimum distance sum between each point on the real Pareto front surface and the set of individuals obtained by the algorithm. The smaller the value, the better the overall performance of the algorithm, including convergence and distribution performance. IGD can be expressed as the following.

$$\mathrm{IGD}(P,Q) = \frac{\sum_{v \in P} d(v,Q)}{|P|} \tag{24}$$

where, P is the set of points uniformly distributed on the real Pareto surface, $|P|$ is the number of individuals in the set of points distributed on the real Pareto surface, Q is the optimal Pareto optimal solution set obtained by the algorithm. d (v,Q) is the minimum Euclidean distance from the individuals v in P to the population Q. Therefore, IGD is used to evaluate the comprehensive performance of the algorithm by calculating the average of the minimum distances from the real Pareto surface set to the obtained population. Meanwhile, it can be seen that if the convergence performance of the algorithm is relatively good, the d (v,Q) is relatively small, so that the convergence performance of the algorithm can be evaluated. However, when the distribution performance of the algorithm is poor and most of the individuals in the population are concentrated in a narrow area, the d (v,Q) of many individuals will be large.

In order to better demonstrate the accuracy and applicability of the methods used in this paper, the reference-point based non-dominated sorting approach (NSGA-III) (Deb and Jain, 2014), adaptive reference vector-guided evolutionary algorithm using growing neural gas (RVEA-iGNG) (Liu et al., 2022), multiple single objective Pareto sampling (MSOPSII) (Hughes, 2007) are compared, and the parameters of each algorithm are selected as shown in Table 4.

The convergence of the new multi-objective algorithm proposed is compared with other algorithms though the mean and standard deviation values of IGD of different multi-objective algorithms for solving wind-light-water storage capacity allocation optimization problems. The IGD indicators of each multi-objective algorithm are shown in Figure 4.

From the characteristics of IGD indicator, it can be concluded that the smaller the IGD value, the better the convergence of the algorithm for the problem. As shown in Figure 4, the IGD value of MSOPSII is higher than other algorithms in the current iteration interval, so it is considered to have the worst performance in solving the wind-light-water storage capacity configuration model proposed in this paper. The IGD value of NSGA-III is higher than that of RvEA-iGNG when the number of iterations are 2,400 and 4,600. However, when the number of iterations is 6,800, it is lower than that of RvEA-iGNG. When the number of iterations is higher than 9,000, the IGD values of

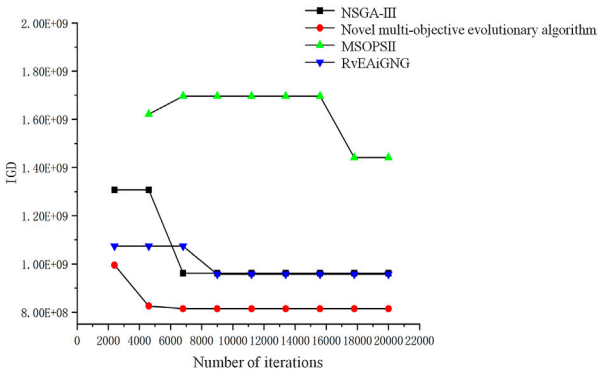


FIGURE 4
Comparison of IGD of different algorithms.

NSGA-III and RvEA-iGNG are consistent. The IGD value of the improved multi-objective evolutionary algorithm proposed in this paper is always the minimum in the current iteration interval, which is the best comprehensive performance compared with other algorithms. Therefore, the proposed algorithm has a greater advantage in dealing with the wind-light storage capacity allocation problem, which provides algorithmic support for the solution of the wind-light-water storage capacity allocation problem. Usage of this algorithm for solving such planning problems can effectively speed up the bi-objective solution speed as well as the accuracy requirements and avoid the phenomenon of premature convergence of the algorithm.

4.4 Multi-energy complementary model comparison

In order to verify the feasibility of the wind-light-water storage model in practical application, this section constructs the light-water storage model and the wind-water storage model, and compares the economy of their capacity configuration and the capacity of the battery. Based on the improved multi-objective optimization algorithm proposed in this paper, the optimal capacity of the light-water storage model and the wind-water storage model are configured.

This paper selects LPSP = .02 to analyze the capacity configuration of each system. The comparison results are shown in Table 5: wind-water storage model is model I, light-water storage model is model II, and wind-light-water storage model is model III.

As shown in Table 5, the part of the hydropower station with insufficient output in the wind-water storage model is provided by wind turbines and batteries, so the number of wind turbines are higher than that of the wind-light-water storage system. In the light-water

TABLE 5 Optimal allocation of different energy system.

Configuration type	Number of wind turbine/ unit	Number of photovoltaic/ thousand pieces	Number of battery/ block	Wind capacity/ MW	Photovoltaic capacity/MW	Battery capacity/ MW	Total cost/ million yuan
I	441	—	4,044	882	—	292	1,299
II	—	4,559	18339	—	1367.7	1,321	1,466
III	332	383	8,132	664	115	813	910

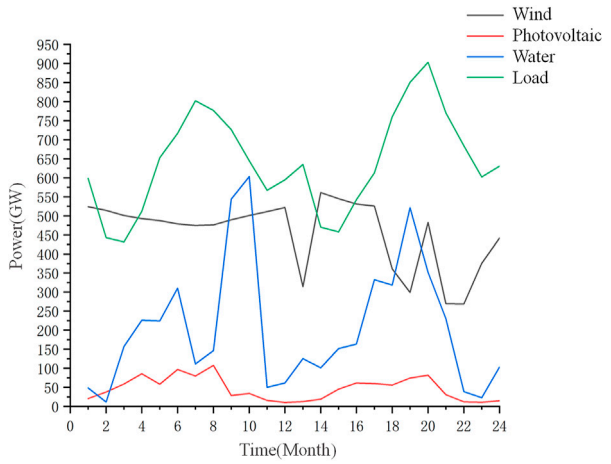


FIGURE 5
Monthly output of wind/PV/hydropower/storage system.

storage model, the insufficient output of the hydropower station is provided by photovoltaic panels and batteries, so the number of photovoltaic systems are higher than that of the wind-light-water storage system.

By comparing the total cost of different energy structures, the wind-light-water storage system is the lowest, the wind-water storage system is the second, and the light-water storage system is the highest. The comparison of storage battery capacity configuration under different energy structures show that wind-water storage is the lowest, wind-light-water storage is the second, and light-water storage is the highest. In conclusion, the wind-light-water storage system established in this paper has the lowest total cost by compared with wind-water storage system and light-water storage system.

4.5 Analysis of monthly outputs

Wind, photoelectric power station output fluctuates frequently, in different seasons and time presents randomness, fluctuation, intermittent and other characteristics, and different areas by wind speed, light and other meteorological factors also have a large difference. The output size of wind and solar power plants is related to numerous factors, including wind speed, wind direction, air pressure, temperature, light intensity, light duration, topography, unit parameters, transmission channels, and grid regulation requirements.

To better illustrate the configuration results and guide production, the output of the wind-light-water load is calculated for a load shortage rate of .02, as shown in Figure 5.

As can be seen from the Figure 5, PV is more out between April and August. According to the local illumination data, the illumination time during the period is about 140 h per month, and the total daily radiation curve is 17000 kJ/m^2/d for the whole year, which is higher than other months year-on-year. At the same time, considering the influence of temperature on photovoltaic power generation devices, the local area is in spring and summer season, with little temperature change, which has less influence on photovoltaic power generation devices and can generate power at a higher level. Wind power generation in the following January and June to December has less power output and fluctuates sharply. From the data, it can be seen that the wind speed is low and the air density is lower than the rest of the months in the same period, so the wind power generation plant produces less power. The hydroelectric power generation fluctuates throughout the year, and the maximum output is from September to October, because the reservoir level increases in the early rainy season, resulting in an increase in water storage and the output capacity fluctuating greatly. Considering the normal operation of reservoirs and downstream safety issues, reservoirs have to be released to reduce risks, and reservoirs will only be released when a certain amount of water is accumulated. Local power generation water consumption increases, and reservoir output increases, but the season brings a high fluctuation of reservoir output due to abundant and dry seasons. The load shows cyclical fluctuations, with low in spring and autumn and high in summer and winter. Considering the increase in electricity consumption of general users by the seasons, industries such as agriculture and industry also face high electricity consumption for irrigation and equipment maintenance. The load is higher than wind and water output in the same period for a long time existing in the form of power shortage.

Considering the demand of wind-light-water and load, in February-March every year, wind power generation meets the load demand and light and water will be stored in the form of energy in the battery. In the rest months, the output of wind-light-water is lower than the load demand, and the accumulated sum is also lower than the load end demand. At this time the battery output is enough to make up for the load demand, to make up for the shortage caused by the lack of local wind-light-water output. Through the battery discharge to meet the shortage, the utilization rate of wind-light-water can be improved, effectively avoiding the waste of resources caused by the volatility of wind-light-water.

5 Conclusion

In this paper, a multi-objective evolutionary algorithm based on the NDWA-GA and the PCA with Pareto optimal space is proposed for the optimal capacity allocation of the multi-energy complementary system. The wind-light-water storage complementary power generation system is constructed, which mainly includes wind turbine, photovoltaic battery array, hydroelectric power station and energy storage part. Then the wind-light-water storage power supply is added to the current hydropower station, and the wind-light-water storage capacity optimization configuration model satisfying various constraints is established with the objective of minimizing the total investment cost and battery capacity of the system. Finally, the proposed novel multi-objective evolutionary algorithm is used to optimize the capacity configuration of each power supply in the complementary power generation system with the load shortage rate as the evaluation index. The configuration results show that the value of LPSP has a direct impact on the number of individual units to be configured. When the load shortage rate is .02, the optimal capacity configurations of wind, light, and storage are 664 MW, 115 MW, and 813 MW, respectively. Moreover, compared with the traditional multi-objective evolutionary algorithm, the results show that the proposed algorithm has better convergence, which verifies the effectiveness and accuracy. At the same time, compared with the light-water storage model and the wind-water storage model, it is proved that the wind-light-water storage model proposed in this paper has the lowest economic cost under the premise of accurate prediction of hydropower station output. The proposed model improves the application rate of clean energy, typically represented by wind-light, on the premise of achieving accurate prediction of hydro power plant output, which helps to improve the energy pattern of Chinese mainly thermal power plants and is of great significance for energy conservation and emission reduction.

The wind-light power output characteristics are closely related to the climatic conditions, with frequent fluctuations in power output, showing random, fluctuating and intermittent characteristics in different seasons and diurnal time sequences. There are also large differences in different regions affected by wind speed, light and other meteorological factors. However, there is no accurate description of the spatial and temporal output law of the wind-light power plant yet, and the capacity allocation of the multi-energy complementary model can only be done by historical data, which has certain limitations, and this will be further studied in our future work.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

YW: Data curation, Methodology, Resources, Software, Writing—original draft, Writing—review and editing. JL: Methodology, Project administration, Resources, Software, Writing—review and editing.

Acknowledgments

Special thanks to colleagues who provided help and guidance on 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

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

Abdin, Z., and Merida, W. (2019). Hybrid energy systems for off-grid power supply and hydrogen production based on renewable energy: A techno-economic analysis. *Energy Convers. Manag.* 196, 1068–1079. doi:10.1016/j.enconman.2019.06.068

Alghussain, L., Samu, R., Taylan, O., and Fahrioglu, M. (2020). Sizing renewable energy systems with energy storage systems in microgrids for maximum cost-efficient utilization of renewable energy resources. *Sustain. Cities Soc.* 55, 102059. doi:10.1016/j.scs.2020.102059

Bekele, G., and Tadesse, G. (2012). Feasibility study of small hydro/PV/wind hybrid system for off-grid rural electrification in Ethiopia. *Appl. Energy* 97, 5–15. doi:10.1016/j.apenergy.2011.11.059

Chang, R. L. (2020). *Research on optimal allocation of capacity for wind-solar-water-storage complementary power generation system*. Xi'an, China: Xi'an University of Technology. [master's thesis]. [Xi'an].

Chen, W. R., Fu, W. Q., Han, Y., Li, Q., Huang, L. J., and Xu, C. P. (2021). Optimal configuration of wind-solar-hydrogen multi-energy complementary microgrid with demand side. *J. Southwest Jiaot. Univ.* 56 (03), 640–649. doi:10.3969/j.issn.0258-2724.20200163

Corne, D. W., Jerram, N. R., Knowles, J. D., and Oates, M. J. (2001). *PESA-II: Region-based selection in evolutionary multi-objective optimization proceedings of the genetic and evolutionary computation conference*. San Francisco: Morgan Kaufmann Publishers, 283–290.

Deb, K., and Jain, H. (2014). An evolutionary many-objective optimization algorithm using reference-point based non-dominated sorting approach, part I: Solving problems with box constraints. *IEEE Trans. Evol. Comput.* 18 (4), 577–601. doi:10.1109/tevc.2013.2281535

Deetjen, T. A., Martin, H., Rhodes, J. D., and Webber, M. E. (2018). Modeling the optimal mix and location of wind and solar with transmission and carbon pricing considerations. *Renew. Energy* 120, 35–50. doi:10.1016/j.renene.2017.12.059

Dou, X. B., Yuan, J., Wu, Z. J., Ni, Y. M., Fan, C., and Xiao, Y. (2016). Improved configuration optimization of PV-wind-storage capacities for grid-connected microgrid. *Electr. Power Autom. Equip.* 36 (3), 26–32. doi:10.16081/j.issn.1006-6047.2016.03.005

He, Z. N., Yen, G. G., and Zhang, J. (2014). Fuzzy-based Pareto optimality for many-objective evolutionary algorithms. *IEEE Trans. Evol. Comput.* 18 (2), 269–285. doi:10.1109/tevc.2013.2258025

Heide, D., Von, B. L., Greiner, M., Hoffmann, C., Speckmann, M., and Bofinger, S. (2010). Seasonal optimal mix of wind and solar power in a future, highly renewable Europe. *Renew. Energy* 35 (11), 2483–2489. doi:10.1016/j.renene.2010.03.012

Hu, L. X., Gu, Y. Y., and Yao, Y. S. (2016). Optimal capacity configuration method for grid-connected wind-solar complementary power system. *Power Syst. Clean Energy* 32 (03), 120–126.

Hughes, E. J. (2007). "MSOPS-II: A general-purpose many-objective optimiser," in Proceedings of the IEEE Congress on Evolutionary Computation, Singapore, 25-28 September 2007.

Li, G. F., Bie, Z. H., Wang, R. H., Jiang, J. F., and Kou, Y. (2017). Research status and prospects on reliability evaluation of integrated energy system. *High. Volt. Eng.* 43 (1), 114–121. doi:10.13336/j.1003-6520.hve.20161227016

Li, Y. Z., Guo, X. J., Dong, H. Y., and Gao, Z. (2020). Optimal capacity configuration of wind/PV/storage hybrid energy storage system in microgrid. *Proc. CSU-EPSA* 32 (06), 123–128. doi:10.19635/j.cnki.csu-epsa.000322

Liu, Q. Q., Jin, Y. C., Heiderich, M., Rodemann, T., and Yu, G. (2022). "An adaptive reference vector-guided evolutionary algorithm using growing neural gas for Many-Objective optimization of irregular problems," in IEEE Transactions on Cybernetics 52 (5), 2698–2711.

Liu, Q., Wei, M. K., Zhou, Q., Cai, S. R., Jiang, L., Zhou, H., et al. (2020). Research on capacity optimization configuration of the Southwestern China microgrid considering electricity cost and system self-power supply reliability. *Power Syst. Prot. Control* 48 (10), 139–145. doi:10.19783/j.cnki.pspc.190782

Liu, Y., Tan, S., and Jiang, C. (2017). Interval optimal scheduling of hydro-PV-wind hybrid system considering firm generation coordination. *Iet Renew. Power Gener.* 11 (1), 63–72. doi:10.1049/iet-rpg.2016.0152

Mahmoudimehr, J., and Shabani, M. (2018). Optimal design of hybrid photovoltaic-hydroelectric standalone energy system for north and south of Iran. *Renew. energy* 115, 238–251. doi:10.1016/j.renene.2017.08.054

Ming, B., Liu, P., Guo, S. L., Zhang, X. Q., Feng, M. Y., and Wang, X. X. (2017). Optimizing utility-scale photovoltaic power generation for integration into a hydropower reservoir by incorporating long-and short-term operational decisions. *Appl. Energy* 204, 432–445. doi:10.1016/j.apenergy.2017.07.046

National Development, Reform Commission, National Energy Board (2021). *Guidelines on promoting the integration of charge and storage of power sources and networks and the complementary development of multiple energies*. Beijing: National Development and Reform Commission of China ,National Energy Board of China.

National Development, Reform Commission, National Energy Board (2016). *The 13th five-year plan for*. Beijing: Hydropower Development.

Parastegari, P., Hooshm, R. A., Khodabakhshian, A., and Zare, A. H. (2015). Joint operation of wind, farm, photovoltaic, pump-storage and energy storage devices in energy and reserve markets. *Electr. Power & Energy Syst.* 64, 275–284. doi:10.1016/j.ijepes.2014.06.074

Portero, U., Velázquez, S., and Carta, J. A. (2015). Sizing of a wind-hydro system using a reversible hydraulic facility with seawater: A case study in the canary islands. *Energy Convers. Manag.* 106, 1251–1263. doi:10.1016/j.enconman.2015.10.054

Qin, S. F., Sun, C. L., Zhang, G. C., He, X. J., and Tan, Y. (2020). A modified particle swarm optimization based on decomposition with different ideal points for many-objective optimization problems. *Complex Intelligent Syst.* 6 (6), 263–274. doi:10.1007/s40747-020-00134-7

Sanajaoba, S., and Fernander, E. (2016). Maiden application of cuckoo search algorithm for optimal sizing of a remote hybrid renewable energy system. *Renew. Energy* 96, 1–10. doi:10.1016/j.renene.2016.04.069

Sheng, S., and Zhang, J. (2017). Capacity configuration optimisation for stand-alone micro-grid based on an improved binary bat algorithm. *J. Eng.* 13, 2083–2087. doi:10.1049/joe.2017.0696

State Council of the People's Republic of China (2021). *The 14th five-year plan for national economic and social development of the People's Republic of China and the outline of the long-range goals to 2035*. Beijing: State Council of the People's Republic of China.

Wang, K. F., Xie, L. R., Qiao, Y., Wang, X. F., and Bao, H. Y. (2020). Curtailed wind consumption mode based on threshold setting and hierarchical control of retired batteries. *Electr. Power Autom. Equip.* 40 (10), 92–98. doi:10.16081/j.epae.202008033

Wang, X. P., Tian, D. D., Meng, M., Liu, J., and Jiang, L. (2013). Total-factor energy efficiency of coal-fired power plants considering environmental constraints. *East China Electr. Power* 41 (6), 1317–1323.

Xiang, Y., Zhou, Y. R., Li, M. Q., and Chen, Z. F. (2017). A vector angle-based evolutionary algorithm for unconstrained many-objective optimization. *IEEE Trans. Evol. Comput.* 21 (1), 131–152. doi:10.1109/tevc.2016.2587808

Xu, F. Q., Liu, J. C., Lin, S. S., Dai, Q. J., and Li, C. B. (2018). A multi-objective optimization model of hybrid energy storage system for non-grid-connected wind power: A case study in China. *Energy* 163, 585–603. doi:10.1016/j.energy.2018.08.152

Yahyaoui, I., Atieh, A., Serna, A., and Tadeo, F. (2017). Sensitivity analysis for photovoltaic water pumping systems: Energetic and economic studies. *Energy Convers. Manag.* 135, 402–415. doi:10.1016/j.enconman.2016.12.096

Yu, D. X., Zhang, J. H., Wang, X. Y., and Gao, Y. (2019). Optimal capacity configuration of grid-connected wind-PV-storage hybrid power generation system. *Proc. CSU-EPSA* 31 (10), 59–65. doi:10.19635/j.cnki.csu-epsa.000127

Zhang, L. P. (2012). Multiple energy complementations-An effective approach to promote renewable energy development. *Northwest Hydropower* (1), 7–12.

Zhang, Z. W., Fan, W., Liu, T., Zhou, J., and Shi, J. K. (2018). Optimal capacity configuration of wind-solar-water-battery complementary power generation system in remote mountainous areas. *J. Power Supply* 16 (05), 138–146. doi:10.13234/j.issn.2095-2805.2018.5.138

Zhao, T. J., Meng, Q., Wang, J., Ren, M. M., and Zhang, M. C. (2021). Optimal allocation strategy for photovoltaic and electric heating thermal storage capacities in interactive mode. *Proc. CSU-EPSA*. 33 (05), 9–15. doi:10.19635/j.cnki.csu-epsa.000514

Zou, J., Fu, L. W., Zheng, J. H., Yang, S. X., Yu, G., and Hu, Y. R. (2018). A many-objective evolutionary algorithm based on rotated grid. *Appl. Soft Comput.* 67 (7), 596–609. doi:10.1016/j.asoc.2018.02.031

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 24 January 2023
DOI 10.3389/fenrg.2022.1104989

OPEN ACCESS

EDITED BY
Yongming Han,
Beijing University of Chemical Technology, China

REVIEWED BY
Yu Ma,
Chang'an University, China
Zhijiang Chen,
Frostburg State University, United States
Yalong Wu,
University of Houston–Clear Lake, United States

*CORRESPONDENCE
Dou An,
douan2017@xjtu.edu.cn

SPECIALTY SECTION
This article was submitted to Process and Energy Systems Engineering, a section of the journal Frontiers in Energy Research

RECEIVED 22 November 2022
ACCEPTED 05 December 2022
PUBLISHED 24 January 2023

CITATION
Lin X, An D, Cui F and Zhang F (2023), False data injection attack in smart grid: Attack model and reinforcement learning-based detection method.
Front. Energy Res. 10:1104989.
doi: 10.3389/fenrg.2022.1104989

False data injection attack in smart grid: Attack model and reinforcement learning-based detection method

Xixiang Lin, Dou An*, Feifei Cui and Feiye Zhang

School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, China

The smart grid, as a cyber-physical system, is vulnerable to attacks due to the diversified and open environment. The false data injection attack (FDIA) can threaten the grid security by constructing and injecting the falsified attack vector to bypass the system detection. Due to the diversity of attacks, it is impractical to detect FDIAs by fixed methods. This paper proposed a false data injection attack model and countering detection methods based on deep reinforcement learning (DRL). First, we studied an attack model under the assumption of unlimited attack resources and information of complete topology. Different types of FDIAs are also enumerated. Then, we formulated the attack detection problem as a Markov decision process (MDP). A deep reinforcement learning-based method is proposed to detect FDIAs with a combined dynamic-static detection mechanism. To address the sparse reward problem, experiences with discrepant rewards are stored in different replay buffers to achieve efficiency. Moreover, the state space is extended by considering the most recent states to improve the perception capability. Simulations were performed on IEEE 9,14,30, and 57-bus systems, proving the validation of attack model and efficiency of detection method. Results proved efficacy of the detection method in different scenarios.

KEYWORDS

state estimation, deep reinforcement learning, attack detection, smart grid, false data injection attack

1 Introduction

Smart grid is a representative cyber-physical system (Pasqualetti et al., 2013), permitting the bidirectional communication of both information and electric power between the utility and users. In the energy management system (EMS), state estimation plays a critical part of information-physical integration. Through state estimation, EMS can recognize the actual state of electricity transmission by filtering out possible noise to improve the reliability of the real-time data (Katiraei and Iravani, 2006).

Due to the decentralized and multi-temporal coupled characteristics of measurement devices (Annaswamy and Amin, 2013), the terminal equipment often lacks effective physical protection, resulting in the susceptibility to attacks. Diversified cyber-physical

system attacks have been proposed and among the attacks, FDIA is a representative one. The FDIA attacker constructs attack vectors through specific algorithms conditioned on the power grid topology information, injects them through weak points of the grid, and avoids being detected to damage data integrity (An et al., 2019). FDIAs can directly affect the state estimation and the subsequent control elements, causing the system to lose stability and even break down. A large number of smart grid security incidents have shown that, compared with traditional attacks, it is more difficult to detect and defend cyber-physical attacks (Liang et al., 2017).

Research efforts against FDIAs can be categorized into two main types: defense and detection. First, to defend against FDIAs before being attacked, researchers have studied the deployment of grid at the cyber-physical level, such as optimizing the distribution of key nodes and devices according to their coupling characteristics (Lei et al., 2020; Wu et al., 2021). Second, to detect FDIAs after attack, substantial efforts has been made, such as dynamic state estimation method (An et al., 2022), tracking the deviation of measurement (Alnowibet et al., 2021; Mohamed et al., 2021; Sinha et al., 2022) and some stochastic game methods (Wei et al., 2018; Oozeer and Haykin, 2019). In summary, above researches had shown that both cyber and physical methods are required in FDIA studies.

However, the grid operation is full of uncertainties, in which the system states and attacks are diverse (An and Liu, 2019). Due to the diversity, enumerating all attacks is not realistic with limited resources. Moreover, empirical or off-line attack detection strategies are not optimal solutions for online network attack detection (Ashok et al., 2018; Tsobdjou et al., 2022). Therefore, reinforcement learning (RL)-based methods are introduced to avoid the complexity of empirical methods and gain the ability of detecting attacks in multiple scenarios (Wang et al., 2018; Kurt et al., 2019; Haque et al., 2021).

Deep reinforcement learning (DRL) learns the optimum strategy of sequential decision problems by exploring and interacting with the environment. The agent gets rewards for guiding the behavior, with the goal of maximizing the long-term return (Sutton and Barto, 1998). DRL combines the feature-extraction capacity of neural networks with the decision-making capability of reinforcement learning in unknown environments to achieve direct control from state to action (Arulkumaran et al., 2017). The Deep Q-Network (DQN), used in conjunction with the replay buffer and a target network, is a representative DRL algorithm that can be adapted to environment with uncertainty (Mnih et al., 2015).

As for detection process of FDIA, after the unknown start of attack, state estimation results are falsified by the attack vector with unknown attack model (Kurt et al., 2019). Moreover, detection process of FDIA has the feature of sequential decision and the transition of state can be described as model-free (An et al., 2019, 2022). Thus, FIDA detection can be described as a MDP and trained utilizing DQN algorithm to achieve detection by neural networks.

The rest part is: In **Section 2**, related studies are reviewed. In **Section 3**, the smart grid state estimation is introduced, including the static and dynamic method. The empirical bad data detection is also introduced. In **Section 4**, an FDIA model is introduced and three types of FDIAs are discussed based on the attack model. In **Section 5**, a DRL-based, combined dynamic-static FDIA detection method is proposed and optimized. In **Section 6**, simulations of attacks and detections are performed on IEEE grid systems in multiple scenarios. In **Section 7**, this paper is concluded.

2 Related work

Weighted Least Squares (WLS) is the basic and widely-used method for power grid state estimation (Schweppe and Rom, 1970; Schweppe and Wildes, 1970). (Debs and Larson, 1970) applied Kalman filter (KF) to the power grid. As the study deepened, the extended Kalman filter (EKF) applied KF to non-linear systems. Moreover, unscented Kalman filter (UKF) and particle filtering (PF) were applied on state estimation, which improved the accuracy and stability of filtering (Wan and Van Der Merwe, 2000; Julier and Uhlmann, 2004).

(Liu et al., 2009) proposed FDIA and proved that the attack vector can bypass the detection element and cause damages on the system (Pang et al., 2016). studied attack method with the minimum cost to avoids anomaly detection (He et al., 2017). constructed a parallel FDIAs detection scheme, utilizing static-dynamic state estimation to detect attacks, which is robust (Li and Wang, 2019). investigated the method to construct a less costly and undetectable attack vector by partial topology information (Li et al., 2019). studied the selection of optimal buses during the attack and proposed a data-driven optimal bus attack method (Jiang et al., 2020). studied two types of FDIAs and proposed a detection-defense method (Chen and Wang, 2020). proposed a new state estimation method that estimates the grid state by sequential Monte Carlo filtering to detect multiple attacks.

Deep reinforcement learning matured later, but is widely used in sequential decision-making problems in recent years (Mnih et al., 2013). proposed DQN algorithm in 2013, and published a paper in 2015, in which DQN reached a high level over human players (Mnih et al., 2015). In the fields of smart grid security, (Wang et al., 2018), proposed an autonomous FDIA method adopting the nearest sequence memory Q-learning (Liu et al., 2020). investigated the vulnerability of power grids with new energy based on DRL (Wang et al., 2021). studied a hybrid cyber-physical topological attack in power grids, and proposed DRL-based method for detecting attacks with minimum cost (Luo and Xiao, 2021). proposed a FDIA method based on reinforcement learning (RL), utilizing measurements,

grid states and other parameters to construct attacks, without dependence on topology information.

As for RL-based FDIA detection, (Kurt et al., 2019), formulated the detection process as a MDP, and proposed a model-free RL-based detection scheme (Zhang and Wu, 2021). proposed a RL-based detection method without the attack model, utilizing a Q-table to detect attacks by Sarsa algorithm. To address the complexity of storing Q-table, (An et al., 2019; Sinha et al., 2022), applied DQN algorithm to detect FDIAs by neural networks. Moreover, (Alnowibet et al., 2021; Mohamed et al., 2021), studied FDIA detection on energy trading and energy management systems by intelligent priority selection-based RL method.

Researches have proved the efficacy of RL-based method in FIDA detection. However, most studies focused on detection against single attack model, and studies on different types of FIDAs are not sufficient. Moreover, few studies combined multiple state estimation methods in the detection scheme, bringing out the focus of this paper.

3 Preliminaries

In this section, static and dynamic state estimation algorithms are introduced, laying the foundation of combined dynamic-static FIDA detection mechanism. Bad data detection method is shown to verify the efficacy of FDIA. Basic information about DQN algorithm is also introduced.

3.1 Measurement equations

Relation between the measured power flows and state of the grid is (Schweppe and Wildes, 1970):

$$\begin{aligned} \boldsymbol{z} &= \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{v} \\ \boldsymbol{x} &= [\varphi_i, V_i]^T \\ \boldsymbol{z} &= [V_i, P_i, Q_i, P_{ij}, Q_{ij}]^T \end{aligned} \tag{1}$$

where, $\boldsymbol{z}$ denotes the system measurement, $\boldsymbol{h}$ denotes the measurement equation, $\boldsymbol{x}$ denotes the state of the grid, v denote the measurement noise, φ_i and V_i denote the voltage phase angle and magnitude of node i, P_i and Q_i denote the power of node i, P_{ij} and Q_{ij} denote the tributary power flow, from i to j, whose detailed equations are:

$$\begin{aligned} P_i &= \sum_{j\in N_i} V_i V_j \left(G_{ij} cos\left(\varphi_i - \varphi_j\right) + B_{ij} sin\left(\varphi_i - \varphi_j\right)\right) \\ Q_i &= \sum_{j\in N_i} V_i V_j \left(G_{ij} sin\left(\varphi_i - \varphi_j\right) - B_{ij} cos\left(\varphi_i - \varphi_j\right)\right) \\ P_{ij} &= V_i^2\left(g_{si} + g_{ij}\right) - V_i V_j cos\left(\varphi_i - \varphi_j\right) - V_i V_j b_{ij} sin\left(\varphi_i - \varphi_j\right) \\ Q_{ij} &= -V_i^2\left(b_{si} + b_{ij}\right) - V_i V_j sin\left(\varphi_i - \varphi_j\right) \\ &\quad - V_i V_j b_{ij} cos\left(\varphi_i - \varphi_j\right) \end{aligned} \tag{2}$$

where, G_{ij}, B_{ij} are the real and imaginary part of the i, j term in the node conduction matrix, g_{ij}, b_{ij} denote the conductance and susceptance between i, j, $g_{si} + jb_{si}$ denotes the conductance of the parallel branch of i, $g_{sj} + jb_{sj}$ denotes the conductance of the parallel branch of j.

3.2 Static state estimation

Due to the pervasive noise, measurements can be unreliable and inconsistent with the actual state. Static state estimation filters the noise based on the current-measured data (Schweppe and Wildes, 1970). Due to this factor, when attacked by FDIAs, results of static state estimation will deviate significantly from the true state. That deviation is utilized in attack detection mechanism of this paper. WLS method is adopted in this paper, whose iterative form is:

$$\begin{cases} \Delta \boldsymbol{z}^{(k)} = \boldsymbol{z} - \boldsymbol{h}\left(\hat{\boldsymbol{x}}^{(k)}\right) \\ \Delta \hat{\boldsymbol{x}}^{(k)} = \left[\boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{H}\right]^{-1} \boldsymbol{H}^T \boldsymbol{R}^{-1} \Delta \boldsymbol{z}^{(k)} \\ \hat{\boldsymbol{x}}^{(k+1)} = \hat{\boldsymbol{x}}^{(k)} + \Delta \hat{\boldsymbol{x}}^{(k)} \end{cases} \tag{3}$$

where, k denotes the step of iteration, $\hat{\boldsymbol{x}}$ denotes the estimated state, $\boldsymbol{H}(\boldsymbol{x})$ denotes the Jacobian matrix of the measurement equation, with the dimension of $m \times n$. The iteration ends when $\Delta \hat{\boldsymbol{x}}$ is sufficiently small.

3.3 Dynamic state estimation

Dynamic state estimation bases on KF to estimate the state and eliminate noise. While static method focuses primarily on real-time states, dynamic method tries to predict the state of the next step and the estimation result is closer to true state under attacks. Due to the non-linearity of power system, EKF and UKF are applied in our works.

3.3.1 Extended kalman filter

EKF is effective for non-linear models, and performs better in systems with weak non-linearities and perturbations (Li et al., 2015).

A second order expansion of $\boldsymbol{h}(\hat{\boldsymbol{x}})$ around $\tilde{\boldsymbol{x}}$ is:

$$\boldsymbol{h}(\hat{\boldsymbol{x}}) = \boldsymbol{h}(\tilde{\boldsymbol{x}}) + \boldsymbol{H}(\tilde{\boldsymbol{x}})\Delta \boldsymbol{x} + \boldsymbol{S} \tag{4}$$

where, $\Delta \boldsymbol{x} = \hat{\boldsymbol{x}} - \tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{x}}$ is the prior estimation of state, $\hat{\boldsymbol{x}}$ is the posterior estimation of state, $\boldsymbol{H}$ denotes the Jacobian matrix of the measurement function, and $\boldsymbol{S}$ is the remainder term of the second and higher order. Omitting $\boldsymbol{S}$, a linearized model of grid

TABLE 1 Equations of EKF.

Step	Formula
Prior estimation	$\tilde{\boldsymbol{x}}_{k+1} = \boldsymbol{F}_k \hat{\boldsymbol{x}}_k$
	$\boldsymbol{M}_{k+1} = \boldsymbol{F}_k \boldsymbol{\Sigma}_k \boldsymbol{F}_k^T + \boldsymbol{Q}_k$
Kalman gain	$\boldsymbol{K}_{k+1} = \boldsymbol{M}_{k+1} \boldsymbol{H}_{k+1}^T \left(\boldsymbol{H}_{k+1} \boldsymbol{M}_{k+1} \boldsymbol{H}_{k+1}^T + \boldsymbol{R}_{k+1}\right)^{-1}$
Post estimation	$\hat{\boldsymbol{x}}_{k+1} = \tilde{\boldsymbol{x}}_{k+1} + \boldsymbol{K}_{k+1}\left(\boldsymbol{z} - \boldsymbol{h}\left(\tilde{\boldsymbol{x}}_{k+1}\right)\right)$
	$\boldsymbol{\Sigma}_{k+1} = \left(\boldsymbol{I} - \boldsymbol{K}_{k+1} \boldsymbol{H}_{k+1}\right) \boldsymbol{M}_{k+1}$

state is obtained.

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{F}_k \boldsymbol{x}_k + \boldsymbol{Q}_k \\ \boldsymbol{z}_{k+1} &= \boldsymbol{H} \boldsymbol{x}_{k+1} + \boldsymbol{R}_k \end{aligned} \tag{5}$$

where, $\boldsymbol{F}_k$ denotes the state-transition function, $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$ denote the system and measurement noise.

Equations of EKF are shown in **Table 1**, and the explanation is:

1. Prior estimation: Calculate $\tilde{\boldsymbol{x}}_{k+1}$ and covariance matrix of prior estimation $\boldsymbol{M}_{k+1}$ by the post estimation results of step k.
2. Kalman gain: Calculate gain $\boldsymbol{K}_{k+1}$ by $\boldsymbol{M}_{k+1}$ and $\boldsymbol{H}$.
3. Post estimation: Calculate $\hat{\boldsymbol{x}}_{k+1}$ and covariance matrix of post estimation $\boldsymbol{\Sigma}_{k+1}$ for the next step.

3.3.2 Unscented kalman filter

UKF applies KF to non-linear systems utilizing the Unscented Transformation (UT). UKF performs better under systems with strong non-linearity compared with EKF (Julier and Uhlmann, 2004). The non-linear form of grid state is:

$$\begin{cases} \boldsymbol{x}_{k+1} = \boldsymbol{f}\left(\boldsymbol{x}_k\right) + \boldsymbol{\omega}_k \\ \boldsymbol{z}_k = \boldsymbol{h}\left(\boldsymbol{x}_k\right) + \boldsymbol{v}_k \end{cases} \tag{6}$$

where, $\boldsymbol{f}\left(\boldsymbol{x}_k\right)$ is $n \times 1$ dimensional non-linear state-transition function, $\boldsymbol{\omega}_k$ and $\boldsymbol{v}_k$ are $n \times 1$ and $m \times 1$-dimensional Gaussian white noise with the zero mean.

Equations of UKF are shown in **Table 2**, and the explanation is:

1. Generate sigma-points: Generate $2n + 1$ sigma-points (Julier and Uhlmann, 2004).
2. Prior estimation: Utilizing sigma points to calculate the prior estimation $\tilde{\boldsymbol{x}}_{k+1}$, and the prior estimation error covariance $\boldsymbol{M}_{k+1}$.
3. Measurement correction: Calculate the prior estimated measurement $\tilde{\boldsymbol{z}}_{k+1}$. Difference between $\tilde{\boldsymbol{z}}_{k+1}$ and $\boldsymbol{z}_{k+1}$ is used to calculate covariance matrices $\boldsymbol{\Sigma}_{k+1}^{zz}$ and $\boldsymbol{\Sigma}_{k+1}^{xz}$.
4. Kalman gain: Gain $\boldsymbol{K}_{k+1}$ and post estimated state $\hat{\boldsymbol{x}}_{k+1}$ are calculated by $\boldsymbol{\Sigma}_{k+1}^{zz}$ and $\boldsymbol{\Sigma}_{k+1}^{xz}$.

TABLE 2 Equations of UKF.

Step	Formula
Generate sigma points	$\boldsymbol{x}_0, \boldsymbol{x}_i, \omega_i^{(m)}, \omega_i^{(c)}$
Prior estimation	$\boldsymbol{x}_{i,k+1} = \boldsymbol{f}\left(\boldsymbol{x}_{i,k}^a, k\right)$
	$\tilde{\boldsymbol{x}}_{k+1} = \sum_{i=0}^{L} \omega_i^{(m)} \boldsymbol{x}_{i,k+1}$
	$\boldsymbol{M}_{k+1} = \sum_{i=0}^{L} \omega_i^{(c)} \left(\boldsymbol{x}_{i,k+1} - \tilde{\boldsymbol{x}}_{k+1}\right)\left(\boldsymbol{x}_{i,k+1} - \tilde{\boldsymbol{x}}_{k+1}\right)^T$
Measurement correction	$\boldsymbol{z}_{i,k+1} = \boldsymbol{h}\left(\boldsymbol{x}_{i,k+1}\right), \tilde{\boldsymbol{z}}_{k+1} = \sum_{i=0}^{L} \omega_i^{(m)} \boldsymbol{z}_{i,k+1}$
	$\boldsymbol{\Sigma}_{k+1}^{zz} = \sum_{i=0}^{L} \omega_i^{(c)} \left(\boldsymbol{z}_{i,k+1} - \tilde{\boldsymbol{z}}_{k+1}\right)\left(\boldsymbol{z}_{i,k+1} - \tilde{\boldsymbol{z}}_{k+1}\right)^T + \boldsymbol{R}_{k+1}$
	$\boldsymbol{\Sigma}_{k+1}^{xz} = \sum_{i=0}^{L} \omega_i^{(c)} \left(\boldsymbol{x}_{i,k+1} - \tilde{\boldsymbol{x}}_{k+1}\right)\left(\boldsymbol{z}_{i,k+1} - \tilde{\boldsymbol{z}}_{k+1}\right)^T$
Kalman gain	$\boldsymbol{K}_{k+1} = \boldsymbol{\Sigma}_{k+1}^{xz} \left(\boldsymbol{\Sigma}_{k+1}^{zz}\right)^{-1}$
	$\hat{\boldsymbol{x}}_{k+1} = \tilde{\boldsymbol{x}}_{k+1} + \boldsymbol{K}_{k+1}\left(\boldsymbol{z}_{k+1} - \tilde{\boldsymbol{z}}_{k+1}\right)$
	$\boldsymbol{\Sigma}_{k+1} = \boldsymbol{M}_{k+1} - \boldsymbol{K}_{k+1} \boldsymbol{\Sigma}_{k+1}^{zz} \boldsymbol{K}_{k+1}^T$

3.4 Bad data detection

Errors in the initial measurement data can be the source of distortion of estimated states, leading to wrong decisions of EMS. Therefore, detection of bad data in measurements is applied to detect possible errors. The most common method is constructing an empirical threshold and detect by the residual function (Merrill and Schweppe, 1971):

$$\tau < \left\| \boldsymbol{z} - \boldsymbol{h}\left(\hat{\boldsymbol{x}}\right) \right\|_2 \tag{7}$$

where, $\hat{\boldsymbol{x}}$ denotes the state estimation results, $\left\| \boldsymbol{z} - \boldsymbol{h}\left(\hat{\boldsymbol{x}}\right) \right\|_2$ denotes the l^2-norm of residuals and τ denotes the empirical threshold generated form historical data.

Holding of Eq. 7 denotes that residuals of the estimated states exceed the threshold. Then a bad data alarm will be triggered, indicating the existence of bad data.

3.5 DQN algorithm

The DQN algorithm, used with replay buffer and target network, is a representative DRL algorithm. Applying DQN algorithm can overcome the complexity of storing Q-table in Q-learning. Other improvements of DQN over Q-learning are (Mnih et al., 2015):

1. Construct replay buffer: At each step, store the experiences in buffer $\mathbb{D}$. When updating the neural network, a mini batch is extracted to update weights θ. Format of experience $\boldsymbol{e}_t$ is:

$$\boldsymbol{e}_t = \left(s_t, a_t, r_t, s_{t+1}\right) \tag{8}$$

where, s_t, a_t, and r_t denote the state, action and reward of step t during the interacting between the agent and environment.

2. Use target Q network: DQN is a dual-network model. A target network is defined and periodically updated, generating target Q. Thus, the equation of gradient descent is:

$$\nabla_{\theta_i} L(\theta_i) = \mathbb{E}_{s,a,r,s'}\left[\left(r + \gamma max Q\left(s_{t+1}, a_{t+1}, \theta_i^-\right) - Q\left(s_t, a_t, \theta_i\right)\right)\right] \tag{9}$$

where, $Q\left(s_{t+1}, a_{t+1}, \theta_i^-\right)$, $Q\left(s_t, a_t, \theta_i\right)$ are generated by the weights of target and current Q network, respectively.

3. Normalize reward: Restrain the reward r in $(-1, 1)$, which can reduce the gradient during updating.
4. Adopt ε-greedy strategy: Adopt a random strategy at each step with a chance of $1-\varepsilon$, and ε increases with training.

4 Smart grid FDIA

4.1 FDIA model based on complete topology information

In this section, we construct FDIAs under the assumption of complete topology information and unlimited cost. Thus the attacker can extract whole measurement function $\boldsymbol{h}(\boldsymbol{x})$ and construct attack without considering the cost, resulting in the inefficacy of empirical bad data detection mechanism.

Equations for constructing attacks are (Liu et al., 2009):

$$\begin{aligned} \boldsymbol{z}_a &= \boldsymbol{z} + \boldsymbol{a} \\ \hat{\boldsymbol{x}}_a &= \boldsymbol{H}^{-1}\boldsymbol{z}_a = \hat{\boldsymbol{x}} + \boldsymbol{c} \end{aligned} \tag{10}$$

where, $\boldsymbol{z}_a$ denotes the attacked measurement values that the system obtains, $\boldsymbol{z}$ denotes the real measurement values of the grid, $\boldsymbol{a}$ denotes the attack vector, $\hat{\boldsymbol{x}}_a$ denotes the estimated states under attacks, $\hat{\boldsymbol{x}}$ denotes the estimated states without attacks, $\boldsymbol{c}$ denotes the change of state values.

As for a non-linear power system, FIDA can also satisfy the measurement equation $\boldsymbol{z}_a = \boldsymbol{h}(\boldsymbol{x}_a)$ by:

$$\boldsymbol{a} = \boldsymbol{z}_a - \boldsymbol{z} = \boldsymbol{h}(\boldsymbol{x} + \boldsymbol{c}) - \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x}_a) - \boldsymbol{h}(\boldsymbol{x}) \tag{11}$$

$$\begin{aligned} \left\|\boldsymbol{z}_a - \boldsymbol{h}(\hat{\boldsymbol{x}}_a)\right\|_2 &= \left\|(\boldsymbol{z} - \boldsymbol{h}(\hat{\boldsymbol{x}})) + (\boldsymbol{a} + \boldsymbol{h}(\hat{\boldsymbol{x}}) - \boldsymbol{h}(\hat{\boldsymbol{x}}_a))\right\|_2 \\ &= \left\|\boldsymbol{z} - \boldsymbol{h}(\hat{\boldsymbol{x}})\right\|_2 \end{aligned} \tag{12}$$

If the static state estimation is operated as always, there is $\hat{\boldsymbol{x}}_a \approx \boldsymbol{x}_a$ and $\boldsymbol{h}(\hat{\boldsymbol{x}}_a) \approx \boldsymbol{h}(\boldsymbol{x}_a)$. Comparing Eq. 12 with Eq. 7, ignoring the inherent Gaussian noise in Eq. 6, the residuals under valid FDIAs are the same as the residuals without an attack, namely $\left\|\boldsymbol{z}_a - \boldsymbol{h}(\hat{\boldsymbol{x}}_a)\right\|_2 = \left\|\boldsymbol{z} - \boldsymbol{h}(\hat{\boldsymbol{x}})\right\|_2$.

In other words, a FIDA constructed this way doesn't change the residuals in Eq. 7. Therefore, a valid FIDAdoesn't trigger the bad data detection alarm mentioned in Section 3.4.

TABLE 3 Classification of FDIAs.

Duration	Attack intensity	Types of FDIAs
continuous	constant	continuous-constant-intensity attack
continuous	variable	continuous-variable-intensity attack
transient	constant	transient-constant-intensity Attack
transient	variable	transient-variable-intensity attack

Algorithm 1. Strategies of three attacks.

```
Initialize x and x_a = x
for t = 1 to T do
    for t_start < t < t_end do
        if Attack1 is adopted then
            Construct z_a by Equation 13
        else if Attack2 is adopted then
            With a probability ε_attack to construct z_a by Equation 13
            Otherwise z_a = z
        else if Attack3 is adopted then
            Construct z_a by Equation 14
        else
            Break
        end
        Falsify the measurement by z ← z_a
    end
end
```

4.2 Types of attacks

Considering the diversity of attacks and attackers' intentions, types of FDIAs are also diverse. In this paper, we classified FDIAs by duration and variation of intensity.

According to the duration, attacks can be divided into transient attack and continuous attack (Jiang et al., 2020). The transient attack tends to have stronger perturbation in a short period, while the continuous attack can remain undetected for a longer period by applying weak perturbation.

According to the intensity, the attack can be divided into constant-intensity and variable-intensity attack. The constant-intensity attack vectors are similar in magnitude, while the variable-intensity attack vectors can be stochastic or asymptotic.

Detailed classification of FDIA is shown in Table 3. Three types of attacks are selected and studied in this paper, the strategies are shown in Algorithm 1 and the detailed equations are:

1. Attack1. Continuous-constant-intensity attack:

Define the start and end of the attack as t_{start} and t_{end}. While $t_{start} < t < t_{end}$, construct and inject the attack by Eq. 13.

$$\begin{cases} \boldsymbol{x}_a = \boldsymbol{x} + \boldsymbol{c} \cdot \boldsymbol{\omega} \\ \boldsymbol{z}_a = \boldsymbol{h}(\boldsymbol{x}_a) \end{cases} \tag{13}$$

where, $\boldsymbol{c} = [c_{\varphi 1}, c_{\varphi 2}, \cdots c_{\varphi n}, c_{V1}, c_{V2}, \cdots c_{Vn}]$ denotes the intended deviation of phase angle $c_{\varphi i}$ and magnitude c_{Vi} on the node voltage, $\boldsymbol{\omega}$ is a standardized normal variable.

Attack-1 is a typical form of FDIA. When $c_{\varphi i} = c_{Vi} = 0$, no attack will be injected on bus i, and $\boldsymbol{c}$ depends on the intention of attackers. Since we focus on detection, the programming problem of determining $\boldsymbol{c}$ is replaced by a Gaussian variable $\boldsymbol{\omega}$. Multiplying by $\boldsymbol{\omega}$, the attack vector varies in a reasonable range $(\boldsymbol{0}, \boldsymbol{c})$. Thus, diversified attack intention can be included, and value of $\boldsymbol{c}$ can be fixed. For example, if we define $c_{V1} = 0.1$p.u., all attacks with the intensity between 0 and 0.1p.u. on bus one are considered as long as there are enough episodes. Moreover, $\boldsymbol{\omega}$ can also make FDIAs hard to be detected by empirical method.

2. Attack2. Transient-constant-intensity attack:

While $t_{start} < t < t_{end}$, at each step, with a probability of ε_{attack} to construct and inject the attack by Eq. 13. In other cases, no attack is conducted. Attack-2 aims to test the response speed of the detection method.

3. Attack3. Continuous-variable-intensity (incremental) attack:

While $t_{start} < t < t_{end}$, construct the attack by:

$$\begin{cases} \boldsymbol{x}_a = \boldsymbol{x}_a + \dfrac{\boldsymbol{c}}{t_{end} - t_{start}} \\ \boldsymbol{z}_a = \boldsymbol{h}(\boldsymbol{x}_a) \end{cases} \tag{14}$$

Attack3 is valid during steady-state grid operation when state $\boldsymbol{x}$ undergoes little change. One obvious feature of **Attack3** is that $\boldsymbol{x}_a$ is cumulative. Since the cumulation of deviation is slow, Attack3 is hard to be detected at an early stage.

5 DQN-based FDIA detection

In this section, we first introduce a combined dynamic-static detection mechanism. Then, to avoid the complexity and achieve more effectiveness, we proposed a DQN-based FDIA detection method.

5.1 Combined dynamic-static empirical FDIA detection

According to **Section 4.1**, it is hard to detect well-constructed FDIAs by bad data detection. However, when attacked by FDIAs, results of different state estimation methods produce a significant difference: Result of static state estimation will deviate from the true state, since it only depends on the real-time measurements. Result of dynamic method is closer to the true state due to the prediction steps.

So in our works, we combine results of static method (WLS) with dynamic method (EKF and UKF) and detect attack based on their inconsistency:

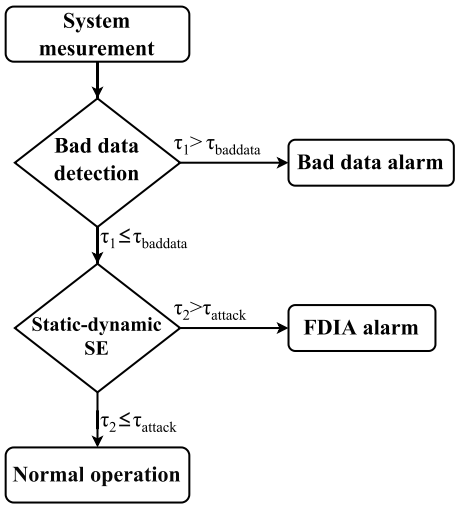


FIGURE 1
FDIA detection mechanism.

$$\tau_1 = \left| \boldsymbol{x}_{KF} - \boldsymbol{x}_{WLS} \right\|_2 > \tau_{attack} \tag{15}$$

where, x_{KF} and x_{WLS} denote the result of KF and WLS, τ_{attack} is the threshold for determining an attack. When Eq. 15 holds, the system is determined to be attacked.

However, grid states change abruptly sometimes due to other factors, which can also leads to the deviation of state estimation. Thus Eq. 15 can be false-positive, so we combine it with bad data detection:

$$\tau_2 = \left\| \boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x}_{WLS}) \right\|_2 > \tau_{baddata} \tag{16}$$

The mechanism is summarized as **Figure 1**, when Eq. 16 holds, the system is determined to have bad data, When Eq. 16 does't hold but Eq. 15 holds, the system is determined to be attacked by FDIAs.

However, since the gird is vulnerable to disturbances, evaluating the performance of the method only by accuracy is incomplete. Considering the time sensitivity, an effective detection of FDIA in this paper is defined in Eq. 17. Utilizing Eq. 17, performance of detection is evaluated by detection rate.

$$t_{start} \leqslant t_{alarm} \leqslant t_{start} + 2 \tag{17}$$

where, t_{start} denotes the start of attack, and t_{alarm} denotes the time that the attack is detected.

If Eq. 17 holds, it shows that the attack is detected within a short period of time. Thus the detection is effective and the safety of grid can be protected.

5.2 DQN-based FDIA detection scheme

Due to the changing load of grid and randomness of attacks, the threshold of Eq. 15 varies greatly in different scenarios. Therefore, it is impractical and costly to apply a certain empirical threshold τ_{attack} in a wide range of grids.

To address the shortcomings of empirical detection, FIDA detection is formulated as a MDP and trained utilizing DQN algorithm. Detection is achieved through neural network, equivalent to a dynamic threshold instead of the empirical threshold τ_{attack} in Eq. 15. The neural network is trained by interacting with the environment during the MDP of FDIA detection (An et al., 2019; Kurt et al., 2019). After an action of detection, agent receives a feedback (reward) from the environment for guiding the actions by updating the neural network (Sutton and Barto, 1998).

5.2.1 MDP-based attack detection model

MDP is the model for sequential decision making (Baxter, 1995). When the state of environment is Markovian, MDP can simulate the strategies and rewards that an agent can achieve. We formulate FIDA detection process as a MDP due to the feature of sequential decision and uncertainty of attack model.

Main components of MDP are state space $\boldsymbol{S}$, action space $\boldsymbol{A}$, state transition $\boldsymbol{P}$ and reward $\boldsymbol{R}$, denoted by $\{\boldsymbol{S},\boldsymbol{A},\boldsymbol{P},\boldsymbol{R}\}$ (Luong et al., 2019). For the FDIA detection, we defined $\boldsymbol{S}$ and $\boldsymbol{A}$ as:

$$\begin{aligned} \boldsymbol{S} &= [s_n, s_a] \\ \boldsymbol{A} &= [a_c, a_s] \end{aligned} \tag{18}$$

where, s_n represents that no attack exists in the grid, s_a represents that the gird is under an attack, a_c denotes that no attack is detected and the system continues to operate, a_s denotes that the attack is detected, and the MDP ends when an attack is detected or time ends.

$\boldsymbol{P}$ represents the state transition function. To address the random and unpredictable characteristics of cyber attacks, a model-free approach is taken to define state transition, i.e., the state transition probability $p(s'|s,a)$ is unknown (An et al., 2019). When the system chooses to continue the operation, the state of the next step is calculated by state estimation and perceived by the agent.

$\boldsymbol{R}$ represents the reward function. For the sparse characteristics of the power grid under attack, we define $\boldsymbol{R}$ by efficacy of detection. When agent detects attacks during normal

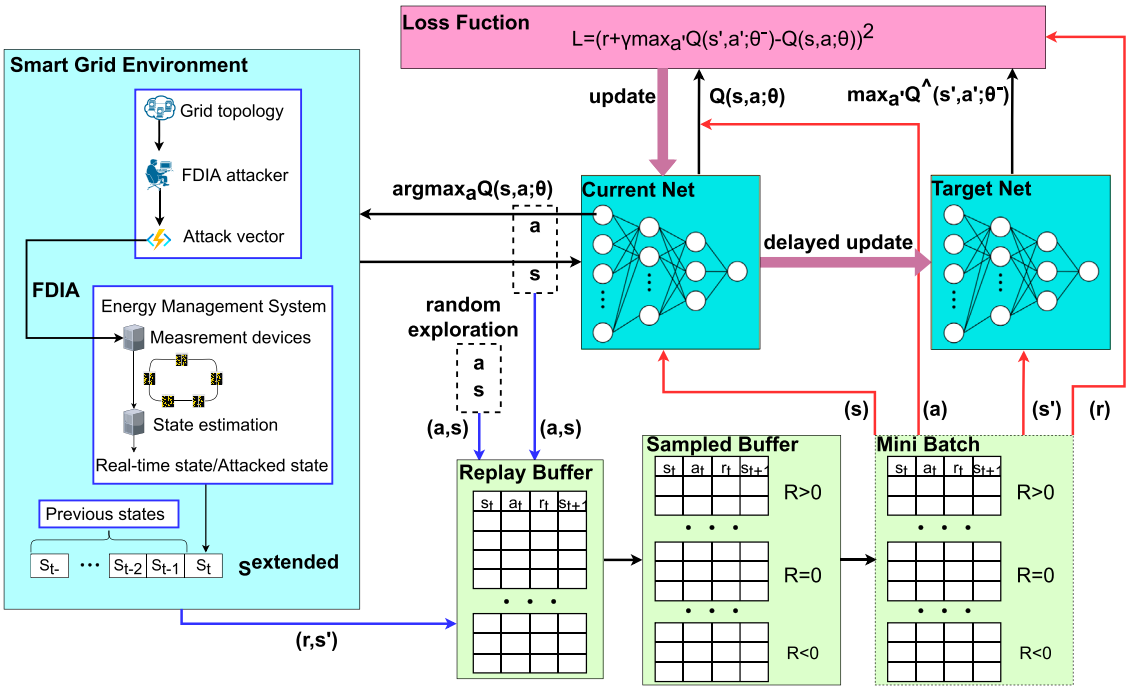


FIGURE 2
Framework of optimized DQN-based detection schem.

Initialize power grid topology
for *episode= 1 to* N_0 **do**
  Initialize system state
  for *time= 1 to* T **do**
    if *in the attack time* **then**
      Construct and inject the attack vector to the measurement
    end
    Get state s_t by state estimation
    Extend state space to get o_t
    Randomly selected actions a_t
    Get reward r_t, preprocess to get s_{t+1} and extend to o_{t+1}
    Store experience (o_t, a_t, r_t, o_{t+1}) in the sampled replay buffer $\mathbb{D}^-$, $\mathbb{D}^+$, and $\mathbb{D}^0$
    if $a_t = a_s$ **then**
      break
    end
  end
end
for *epoch= 1 to* N_1 **do**
  for *episode= 1 to* N_2 **do**
    Initialize system state
    for *time= 1 to* T **do**
      if *in the attack time* **then**
        Construct and add the attack vector to the measurement value
      end
      Get state s_t by state estimation
      Extend state space to get o_t
      Get action a_t by ε-greedy strategy
      Get reward r_t, preprocess to get s_{t+1} and extend to o_{t+1}
      Store experience (o_t, a_t, r_t, o_{t+1}) in the sampled replay buffer $\mathbb{D}^-$, $\mathbb{D}^+$, and $\mathbb{D}^0$
      Updated θ by gradient descent
      if $a_t = a_s$ **then**
        break
      end
    end
    Every certain period, let $\theta^- = \theta$
  end
end

Algorithm 2. Training of the a optimized DQN-based FDIA detection algorithm.

Initialize power grid topology
for *episode= 1 to* N_3 **do**
  Initialize system state
  for *time= 1 to* T **do**
    if *in the attack time* **then**
      Construct and inject the attack vector to the measurement
    end
    Get state s_t by state estimation and extended to get o_t
    Get action a_t by the trained network
    if $a_t = a_s$ **then**
      break
    end
  end
end
Calculate the detection rate

Algorithm 3. Testing of the FIDA detection algorithm.

operation, the reward is negative. When the agent detects attacks under attacks, the reward is positive, and the more timely the detection, the more the rewards. Rewards of the other cases are 0. The detailed function is:

$$r_t(s_t, a_t) = \begin{cases} 0 & s_t = s_n \quad a_t = a_c \\ 0 & s_t = s_a \quad a_t = a_c \\ -\beta^- & s_t = s_n \quad a_t = a_s \\ \beta^+ \dfrac{t - t_{start}}{t_{end} - t_{start}} & s_t = s_a \quad a_t = a_s \end{cases} \tag{19}$$

where, t_{start}, t_{end} denote the start and end of the attack, respectively, β^-, β^+ denote the reward coefficient.

5.2.2 Optimized DQN-based detection scheme

Framework of the detection scheme is shown in **Figure 2**. The detailed training and testing algorithms are given in **Algorithm 2** and **Algorithm 3**.

TABLE 4 Simulation settings.

Environment	Settings
Number of buses on IEEE systems	9, 14, 30, 57
Time steps of each episode T	100
Duration of attack T_a	50
Probability of ε_{attack} in Attack2	0.2
Intended deviation of voltage angle $c_{\varphi i}$	0°
Intended deviation of voltage magnitude c_{Vi}	0.1p.u
Parameters of UKF	$\alpha = 10^{-3}, \beta = 2, \kappa = 0$
Learning rate α_l	0.001
Discount factor γ	0.9
Dimension of extended State space N	4
Reward coefficient β^-, β^+	1:1
ϵ-greddy strategy initial ϵ	0.9
ϵ-greddy strategy increment of ϵ	0.0005
Replay buffer size	1,000
Mini-batch size	32
Sampled mini-batch size $\mathbb{D}^-$, $\mathbb{D}^+$, and $\mathbb{D}^0$	4, 4, 24
Time interval of updating target network	10
Episodes of random exploring $N0$	100
Epoches of training $N1$	10, 20
Episodes of training $N2$	500
Episodes of testing $N3$	100
Random seeds	101,102,103,104,105

6 Simulations

6.1 Simulations setup

Extensive simulations performed to simulate the actual scenarios of FIDA attack-detection process. First, to ensure the practicability, simulations are based on IEEE 9, 14, 30 and 57-bus networks by MATPOWER (Zimmerman et al., 2011). Second, due to the diversity of attacks and attack intentions (An and Liu, 2019), three types of FDIAs are adopted, namely Attack-1, 2, and 3. Cases with single attack and multiple attacks are both considered during simulations. Then, the attacks aim at the magnitude of node voltages, with the intensity to cause voltage violation (Zhu and Liu, 2016; Zheng et al., 2020). Third, WLS is adopted in static estimation while EKF and UKF are adopted for dynamic state estimation in different cases. In addition, random seeds were used to reduce the random error. Details of settings are shown in **Table 4**.

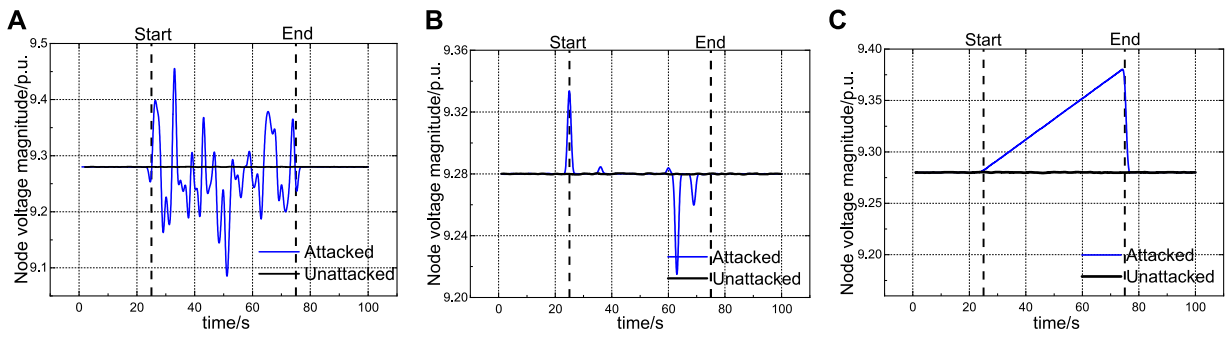


FIGURE 3
Effects of three typical attacks. **(A)** Effect of Attack-1, **(B)** Effect of Attack-1, and **(C)** Effect of Attack-1.

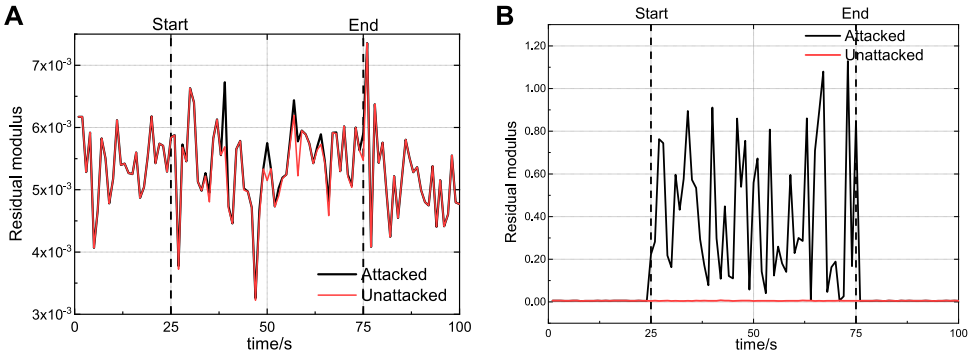


FIGURE 4
Bad data detection results **(A)** Bad data detection results under valid attack and **(B)** Bad data detection results under invalid attack.

6.2 Simulations and effects of attacks

Considering the change in voltage magnitude of static state estimation result, effects of three typical attacks is in Figure 3. The dashed lines indicate the start and end of attacks.

Figure 3A shows effect of Attack-1, namely continuous-constant-intensity attack. Attack-1 injects attack continuously and magnitude of the attack vector follows the same Gaussian distribution. Thus, deviation in voltage magnitude is obvious under Attack-1.

Figure 3B shows effect of Attack-2, namely transient-constant-intensity attack. Attack-2 injects the attack vector intermittently and magnitude of the vector follows the same Gaussian distribution. Deviation generated by Attack-2 is also considerable, but the attack duration is compressed. Thus, for the detector, higher response speed is required.

Figure 3C shows effect of Attack-3, namely continuous-variable-intensity (incremental) attack. Attack-3 injects the attack vector continuously and magnitude of the vector is cumulative. The deviation between two steps generated by Attack-3 is smaller than other attacks, which can avoid being detected. However, the deviation accumulates over a period of time and the amplitude at the end of attack is also considerable, so the consequence of Attack-3 can be severe.

Taking Attack-1 as an example, Figure 4 shows the differences between valid and invalid attacks. In Figure 4A, the residual modulus of the no-attack case is about 0, and almost overlaps with the valid-attack case. But in Figure 4B, the residual modulus fluctuates greatly at a high level under an invalid attack, exposing the attack to detector. In summary, the difference of effectiveness between valid and invalid attacks to bypass the bad data detection is obvious.

Changes in node voltage of IEEE 9-bus system when attacked by the FDIA are shown in Figure 5. Since the intended deviation of angle $c_{\varphi i} = 0°$, the phase angle deviates slightly during the attack in Figure 5A. Moreover, the intended deviation of magnitude $c_{Vi} = 0.1$p.u., so the intensity of attack is within

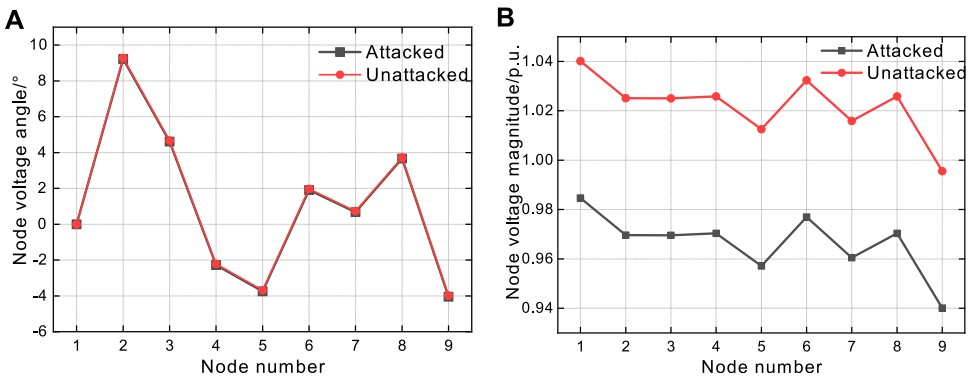


FIGURE 5
Static state estimation results under valid attack. **(A)** Change of node voltage angle after the valid attack and **(B)** Change of node voltage magnitude after the valid attack.

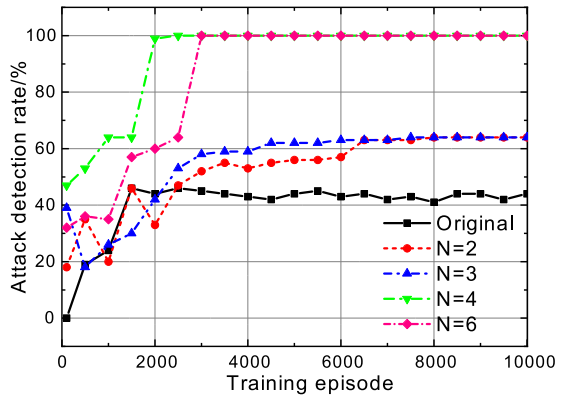


FIGURE 6
Attack detection rate under different state spaces.

$(0, c_j)$. Correspondingly, the bus voltage magnitude undergoes a large deviation in Figure 5B, misleading the following grid operation.

6.3 Effectiveness of optimized DQN-based method

Simulations in this section compare the optimized DQN-based method with the original DQN-based method and empirical threshold method. Moreover, cases with different dimension of state space are also compared Effectiveness of adopting sampled replay buffers and extending state space is proved.

First, to verify that effectiveness of extending state space, we changed the dimension of state space in different cases. Results are shown in Figure 6. The detection of original method is unstable during the training and reaches convergence after 3,000 episodes, the attack detection rate fluctuates at a low level (43%), and the detection rate is unstable with fluctuation.

With the extended state space, the detection rate increases by at least 21%–64%, and the fluctuation decreases significantly. Comparing the above cases, the detection rate reaches near 100% after convergence in cases that $N \geq 4$, so in the rest of this paper $N = 4$.

Second, to prove efficacy of the optimized method, we simulated the detections with optimized DQN-based method, original DQN-based method and empirical method.

The empirical threshold method uses a fixed threshold consturcted from experiences. In this paper the algorithm is: By $\tau = \|x_{KF} - x_{WLS}\|_2$, calculate τ_1 in no-attack cases and τ_2 in attacked cases, $\tau_{attack} = \tau_2 - \tau_1$. So the detection rate is a fixed number and behaves as a horizontal line since the empirical threshold is hardly updated online in practice.

Results are shown in Figure 7. Each case is simulated in five parallel groups utilizing random seeds in Table 4. Detection rates are averaged and the shadows denote the standard deviations between different groups.

Comparing the performances in Figure 7, empirical method doesn't perform well in detection. Detection rate of empirical method is 61% and 80% against Attack-1 and Attack-2, and is only 30% against Attack-3. What's more, the method with original DQN performs well at certain episodes in Figures 7C, F, but the detection rate fluctuates substantially throughout the training in Figures 7A, D, E. The convergence of training with original DQN is difficult, too.

As for the optimized DQN-based method, cases with EKF converges around 8,000 episodes, and the converged detection rate is 98.42% against Attack-1, 99.70% against Attack-2, and 100% against Attack-3, with some fluctuation. Cases with UKF converges around 5,000 episodes, and the converged detection

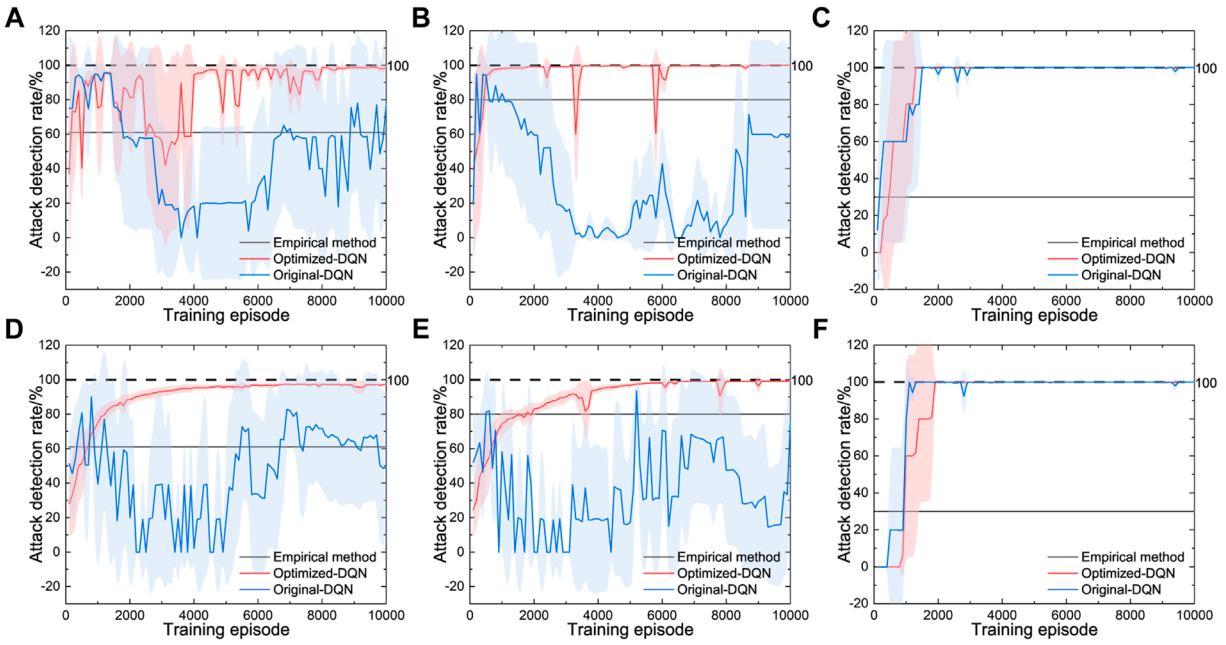


FIGURE 7
Detection with optimized-DQN, original-DQN and empirical method against attacks. **(A)** Detection with EKF against Attack-1, **(B)** Detection with EKF against Attack-2, **(C)** Detection with EKF against Attack-3, **(D)** Detection with UKF against Attack-1, **(E)** Detection with UKF against Attack-2 and **(F)** Detection with UKF against Attack-3.

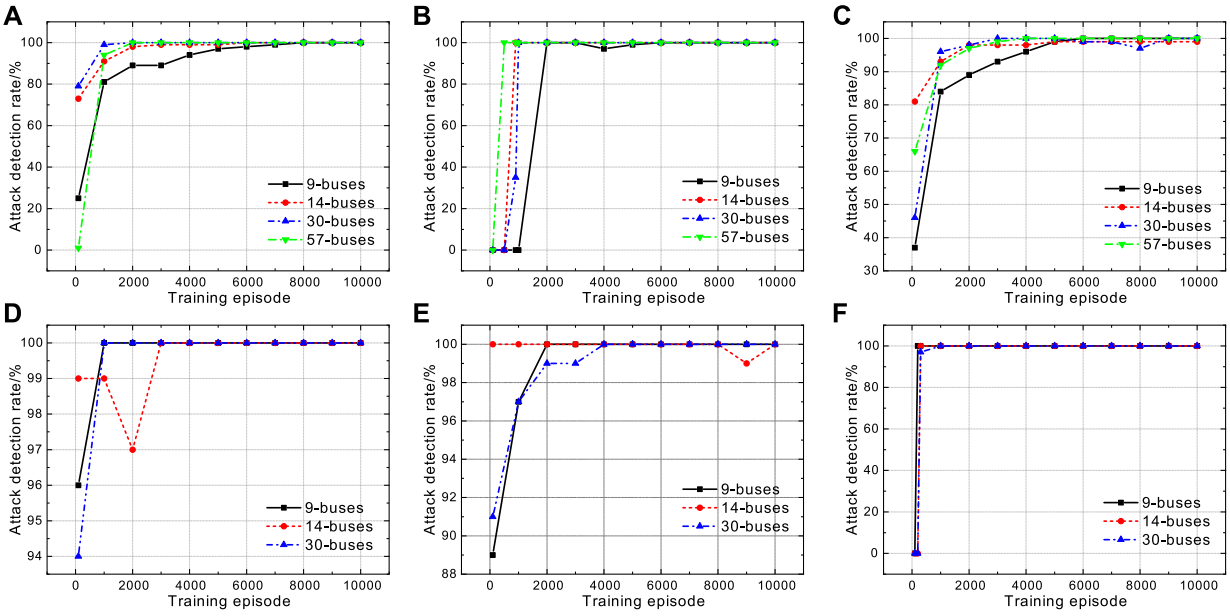


FIGURE 8
Attack detection rate while training against multiple attacks under multiple systems based on EKF or UKF. **(A)** Training against Attack-1 based on EKF, **(B)** Training against Attack-2 based on EKF, **(C)** Training against Attack-3 based on EKF, **(D)** Training against Attack-1 based on UKF, **(E)** Training against Attack-2 based on UKF, and **(F)** Training against Attack-3 based on UKF.

TABLE 5 Performance of detection in different systems against multiple attacks.

Type of attack	Number of buses	Detection rate (EKF)/%	Detection rate (UKF)/%
Attack-1	9	99.46	99.88
	14	99.58	99.72
	30	100.00	100.00
	57	100.00	\
Attack-2	9	99.87	100.00
	14	99.00	99.95
	30	99.32	100.00
	57	100.00	\
Attack-3	9	99.46	99.88
	14	100.00	100.00
	30	100.00	100.00
	57	100.00	\

rate is 96.95% against Attack-1, 98.99% against Attack-2, and 100% against Attack-3 with little fluctuation. After convergence, fluctuation of detection rate has been restricted within 4%. In addition, the training process of EKF-based method is less stable than the UKF-based method.

In conclusion, detection rate is improved by at least 15.95% utilizing the optimized DQN-based method. Stability of the training is also improved fundamentally over the original DQN-based method, especially the UKF-based method.

6.4 Simulation in multiple cases

In this section, we compared three types of FDIAs in power systems in different networks to prove that the proposed method is effective for multiple scenarios. In addition, each case was repeated at least three times and results are averaged. Results are shown in Figure 8 and Table 5.

In Figure 8, the training process converges after about 8,000 episodes against Attack-1 and 6,000 episodes against Attack-2. Meanwhile, the speed of convergence is faster in cases based on UKF, especially in the cases against Attack-3. At the final stages of training in above cases, the detection rates fluctuate by 2%. In addition, the convergence speed is slightly faster of a more complex network, since the accumulation of state deviation is faster.

First, in Table 5, the detection performances are similar in different networks, since the detection mechanism only depends on the state estimation performance and is not affected by the network complexity. Second, detection rates in different cases are consistently close to 100% after convergence. Third, UKF-based method performs better in detection than EKF-based method.

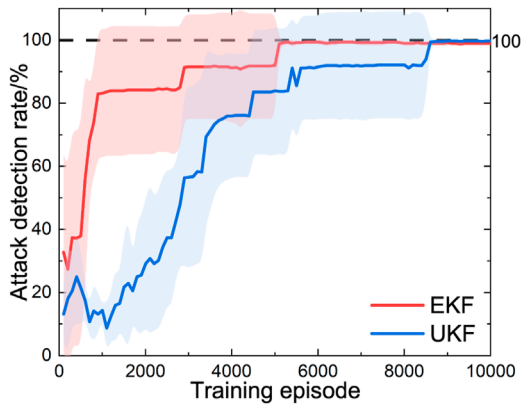


FIGURE 9
Attack detection rate against the hybrid attack.

In summary, the method performs well under different attacks in multiple scenarios.

6.5 Simulation against hybrid attacks

To prove utility of the detection method, a hybrid attack model is constructed. In each episode, the type and start of attack is random and unknown. One of Attack-1, 2, 3 is randomly selected and conducted during the attack based on IEEE 14-bus network.

Rusult of training is shown in Figure 9. Detection rates are also averaged by results of five groups, and standard deviations are given by the shadows.

After training, the detection rate of EKF-based method reaches 99.01%, and the UKF-based method reaches 99.71%. In Figure 9, since the attack is hybrid, the detection rate fluctuates at the early stage of training. Trainings converge more slowly compared to the cases against single attack. EKF-based method converges at about 5,500 episodes and UKF-based method converges at about 8,500 episodes.

7 Conclusion

In this paper, a FDIA model with complete topology information and unlimited cost is introduced first. Attacks constructed under this model is verified to have the ability of bypassing the empirical bad data detection. FDIAs are classified by duration and intensity. Three types of attacks and their effects are performed. Then, a detection mechanism is proposed by combining static and dynamic state estimation. Second, the FIDA detection process was formulated as a MDP, and a DQN-based detection method is constructed. To address the problems while training and detection, optimizations were made to improve the efficacy. The DQN-based method is adaptive and has a non-deterministic threshold. Third, sufficient simulations were conducted, including a variety of cases, laying the foundation for studying multiple types of FDIAs. Simulation results prove that the detection rate against FDIA is improved by at least 15.95% over the empirical threshold method. The fluctuation of detection rate has been restricted within 4% during the final stage of training. Moreover, the highest detection rate reached 99.71% against the proposed hybrid attack.

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

XL developed the methodology, performed the experiment, analyzed the data, and wrote the manuscript; DA contributed to the conception of the study and manuscript preparation; FC helped perform the analysis with constructive discussions. FZ contributed significantly to analysis and manuscript preparation.

Funding

This work was supported in part by the National Natural Science Foundation of China under Grant 62173268, Grant 61803295, Grant 61973247, and Grant 61673315; in part by the Major Research Plan of the National Natural Science Foundation of China under Grant 61833015; in part by the National Postdoctoral Innovative Talents Support Program of China under Grant BX20200272; in part the National Key Research and Development Program of China under Grant 2019YFB1704103; and in part by the China Postdoctoral Science Foundation under Grant 2018M643659.

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

Alnowibet, K., Annuk, A., Dampage, U., and Mohamed, M. A. (2021). Effective energy management via false data detection scheme for the interconnected smart energy hub–microgrid system under stochastic framework. *Sustainability* 13, 11836. doi:10.3390/su132111836

An, D., Yang, Q., Liu, W., and Zhang, Y. (2019). Defending against data integrity attacks in smart grid: A deep reinforcement learning-based approach. *IEEE Access* 7, 110835–110845. doi:10.1109/ACCESS.2019.2933020

An, D., Zhang, F., Yang, Q., and Zhang, C. (2022). Data integrity attack in dynamic state estimation of smart grid: Attack model and countermeasures. *IEEE Trans. Autom. Sci. Eng.* 19, 1631–1644. doi:10.1109/TASE.2022.3149764

An, Y., and Liu, D. (2019). Multivariate Gaussian-based false data detection against cyber-attacks. *IEEE Access* 7, 119804–119812. doi:10.1109/ACCESS.2019.2936816

Annaswamy, A. M., and Amin, M. (2013). Ieee vision for smart grid controls: 2030 and beyond. *IEEE Vis. Smart Grid Controls 2030 Beyond.* doi:10.1109/IEEESTD.2013.6577608

Arulkumaran, K., Deisenroth, M. P., Brundage, M., and Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Process. Mag.* 34, 26–38. doi:10.1109/MSP.2017.2743240

Ashok, A., Govindarasu, M., and Ajjarapu, V. (2018). Online detection of stealthy false data injection attacks in power system state estimation. *IEEE Trans. Smart Grid* 9, 1–1646. doi:10.1109/TSG.2016.2596298

Baxter, L. A. (1995). Markov decision processes: Discrete stochastic dynamic programming. *Technometrics* 37, 353. doi:10.1080/00401706.1995.10484354

Chen, L., and Wang, X. (2020). Quickest attack detection in smart grid based on sequential Monte Carlo filtering. *IET Smart Grid* 3, 686–696. doi:10.1049/iet-stg.2019.0320

Debs, A. S., and Larson, R. E. (1970). A dynamic estimator for tracking the state of a power system. *IEEE Trans. Power Apparatus Syst.* 89, 1670–1678. doi:10.1109/TPAS.1970.292822

Haque, N. I., Shahriar, M. H., Dastgir, M. G., Debnath, A., Parvez, I., Sarwat, A., et al. (2021). "A survey of machine learning-based cyber-physical attack generation, detection, and mitigation in smart-grid," in 2020 52nd North American Power Symposium (NAPS). doi:10.1109/NAPS50074.2021.9449635

He, Y., Mendis, G. J., and Wei, J. (2017). Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism. *IEEE Trans. Smart Grid* 8, 2505–2516. doi:10.1109/TSG.2017.2703842

Jiang, Q., Chen, H., Xie, L., and Wang, K. (2020). Learning-based cooperative false data injection attack and its mitigation techniques in consensus-based distributed estimation. *IEEE Access* 8, 166852–166869. doi:10.1109/ACCESS.2020.3023117

Julier, S., and Uhlmann, J. (2004). Unscented filtering and nonlinear estimation. *Proc. IEEE* 92, 401–422. doi:10.1109/JPROC.2003.823141

Katiraei, F., and Iravani, M. (2006). Power management strategies for a microgrid with multiple distributed generation units. *IEEE Trans. Power Syst.* 21, 1821–1831. doi:10.1109/TPWRS.2006.879260

Kurt, M. N., Ogundijo, O., Li, C., and Wang, X. (2019). Online cyber-attack detection in smart grid: A reinforcement learning approach. *IEEE Trans. Smart Grid* 10, 5174–5185. doi:10.1109/TSG.2018.2878570

Lei, D., Zhao, J., Hu, M., Chang, X., Zhang, X., and Song, X. (2020). "Optimized configuration scheme of harmonic measuring device considering practical situations of grid nodes and monitoring device," in 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2). doi:10.1109/EI250167.2020.9346993

Li, Q., Li, R., Ji, K., and Dai, W. (2015). "Kalman filter and its application," in 2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS), 74–77. doi:10.1109/ICINIS.2015.35

Li, Q., Li, S., Xu, B., and Liu, Y. (2019). Optimal node attack on causality analysis in cyber-physical systems: A data-driven approach. *IEEE Access* 7, 16066–16077. doi:10.1109/ACCESS.2019.2891772

Li, Y., and Wang, Y. (2019). False data injection attacks with incomplete network topology information in smart grid. *IEEE Access* 7, 3656–3664. doi:10.1109/ACCESS.2018.2888582

Liang, G., Zhao, J., Luo, F., Weller, S. R., and Dong, Z. Y. (2017). A review of false data injection attacks against modern power systems. *IEEE Trans. Smart Grid* 8, 1630–1638. doi:10.1109/TSG.2015.2495133

Liu, X., Ospina, J., and Konstantinou, C. (2020). Deep reinforcement learning for cybersecurity assessment of wind integrated power systems. *IEEE Access* 8, 208378–208394. doi:10.1109/ACCESS.2020.3038769

Liu, Y., Reiter, M., and Ning, P. (2009). False data injection attacks against state estimation in electric power grids. *ACM Trans. Inf. Syst. Secur.* 14, 21–33. doi:10.1145/1952982.1952995

Luo, W., and Xiao, L. (2021). "Reinforcement learning based vulnerability analysis of data injection attack for smart grids," in 2021 40th Chinese Control Conference (CCC), 6788–6792. doi:10.23919/CCC52363.2021.9550523

Luong, N. C., Hoang, D. T., Gong, S., Niyato, D., Wang, P., Liang, Y. C., et al. (2019). Applications of deep reinforcement learning in communications and networking a survey. *IEEE Commun. Surv. Tutorials* 21, 3133–3174. doi:10.1109/COMST.2019.2916583

Merrill, H. M., and Schweppe, F. C. (1971). Bad data suppression in power system static state estimation. *IEEE Trans. Power Apparatus Syst.* 90, 2718–2725. doi:10.1109/TPAS.1971.292925

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., et al. (2013). Playing atari with deep reinforcement learning. *CoRR* 1312, 5602.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A., Veness, J., Bellemare, M., et al. (2015). Human-level control through deep reinforcement learning. *Nature* 518, 529–533. doi:10.1038/nature14236

Mohamed, M. A., Hajjiah, A., Alnowibet, K. A., Alrasheedi, A. F., Awwad, E. M., and Muyeen, S. M. (2021). A secured advanced management architecture in peer-to-peer energy trading for multi-microgrid in the stochastic environment. *IEEE Access* 9, 92083–92100. doi:10.1109/ACCESS.2021.3092834

Oozeer, M. I., and Haykin, S. (2019). Cognitive dynamic system for control and cyber-attack detection in smart grid. *IEEE Access* 7, 78320–78335. doi:10.1109/ACCESS.2019.2922410

Pang, Z.-H., Liu, G.-P., Zhou, D., Hou, F., and Sun, D. (2016). Two-channel false data injection attacks against output tracking control of networked systems. *IEEE Trans. Ind. Electron.* 63, 3242–3251. doi:10.1109/TIE.2016.2535119

Pasqualetti, F., Dörfler, F., and Bullo, F. (2013). Attack detection and identification in cyber-physical systems. *IEEE Trans. Autom. Contr.* 58, 2715–2729. doi:10.1109/TAC.2013.2266831

Schweppe, F. C., and Rom, D. B. (1970). Power system static-state estimation, part ii: Approximate model. *IEEE Trans. Power Apparatus Syst.* 89, 125–130. doi:10.1109/TPAS.1970.292679

Schweppe, F. C., and Wildes, J. (1970). Power system static-state estimation, part i: Exact model. *IEEE Trans. Power Apparatus Syst.* 89, 120–125. doi:10.1109/TPAS.1970.292678

Sinha, A., Thukkaraju, A. R., and Vyas, O. P. (2022). "A multi agent framework to detect in progress false data injection attacks for smart grid," in *Advanced network technologies and intelligent computing*. Editors I. Woungang, S. K. Dhurandher, K. K. Pattanaik, A. Verma, and P. Verma (Cham: Springer International Publishing), 123–141.

Sutton, R., and Barto, A. (1998). Reinforcement learning: An introduction. *IEEE Trans. Neural Netw.* 9, 1054. doi:10.1109/TNN.1998.712192

Tsobdjou, L. D., Pierre, S., and Quintero, A. (2022). An online entropy-based ddos flooding attack detection system with dynamic threshold. *IEEE Trans. Netw. Serv. Manage.* 19, 1679–1689. doi:10.1109/TNSM.2022.3142254

Wan, E., and Van Der Merwe, R. (2000). "The unscented kalman filter for nonlinear estimation," in Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, 153–158. doi:10.1109/ASSPCC.2000.882463

Wang, Z., Chen, Y., Liu, F., Xia, Y., and Zhang, X. (2018). Power system security under false data injection attacks with exploitation and exploration based on reinforcement learning. *IEEE Access* 6, 48785–48796. doi:10.1109/ACCESS.2018.2856520

Wang, Z., He, H., Wan, Z., and Sun, Y. (2021). Coordinated topology attacks in smart grid using deep reinforcement learning. *IEEE Trans. Ind. Inf.* 17, 1407–1415. doi:10.1109/TII.2020.2994977

Wei, L., Sarwat, A. I., Saad, W., and Biswas, S. (2018). Stochastic games for power grid protection against coordinated cyber-physical attacks. *IEEE Trans. Smart Grid* 9, 684–694. doi:10.1109/TSG.2016.2561266

Wu, Z., He, L., Li, S., Zhang, H., Hu, S., Zhang, M., et al. (2021). "Reinforcement learning based multistage optimal pmu placement against data integrity attacks in smart grid," in 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS). doi:10.1109/ICPS49255.2021.9468170

Zhang, K., and Wu, Z.-G. (2021). "A reinforcement learning-based detection method for false data injection attack in distributed smart grid," in 2021 8th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS), 38–43. doi:10.1109/ICCSS53909.2021.9722027

Zheng, Y., Hill, D. J., Song, Y., Zhao, J., and Hui, S. Y. R. (2020). Optimal electric spring allocation for risk-limiting voltage regulation in distribution systems. *IEEE Trans. Power Syst.* 35, 273–283. doi:10.1109/TPWRS.2019.2933240

Zhu, H., and Liu, H. J. (2016). Fast local voltage control under limited reactive power: Optimality and stability analysis. *IEEE Trans. Power Syst.* 31, 3794–3803. doi:10.1109/TPWRS.2015.2504419

Zimmerman, R. D., Murillo-Sánchez, C. E., and Thomas, R. J. (2011). Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Trans. Power Syst.* 26, 12–19. doi:10.1109/TPWRS.2010.2051168

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 23 March 2023
DOI 10.3389/fenrg.2023.1077266

OPEN ACCESS

EDITED BY
Yongming Han,
Beijing University of Chemical Technology, China

REVIEWED BY
Yumin Zhang,
Shandong University of Science and Technology, China
Hamid Reza Rahbari,
Aalborg University, Denmark

*CORRESPONDENCE
Chengzhu Gong,
chengzhu.gong@cug.edu.cn

SPECIALTY SECTION
This article was submitted to Process and Energy Systems Engineering, a section of the journal Frontiers in Energy Research

RECEIVED 22 October 2022
ACCEPTED 13 February 2023
PUBLISHED 23 March 2023

CITATION
Liu S, Gong C and Pan K (2023), A combinatorial model for natural gas industrial customer value portrait based on value assessment and clustering algorithm.
Front. Energy Res. 11:1077266.
doi: 10.3389/fenrg.2023.1077266

A combinatorial model for natural gas industrial customer value portrait based on value assessment and clustering algorithm

Sicong Liu[1], Chengzhu Gong[1,2]* and Kai Pan[3]

[1]School of Economics and Management, China University of Geosciences, Wuhan, China, [2]Center for Energy and Environmental Management and Decision-making, China University of Geosciences, Wuhan, China, [3]PetroChina Planning and Engineering Institute, China National Petroleum Corporation, Beijing, China

Frequent geopolitical events have reduced the stability of natural gas supply and caused drastic price fluctuations, which poses a new challenge to the natural gas consumer market. To improve the anti-risk ability of the natural gas industrial market, this study constructs a new customer value portrait framework to discern the industrial customer value based on different types of behavioral features and the emerging trends of the natural gas market. Specifically, we rediscover the value composition of natural gas industrial customers and establish a set of indicators to reflect the customer value in different dimensions with mixed data types. Then, a visualizable customer value classification model has been established by combining Gower's dissimilarity coefficient with the PAM clustering algorithm. To ensure the accuracy of the clustering results, the optimal number of clusters is determined by gap statistics and elbow point, and the average silhouette method is used to detect the clustering effect as well as used in misclassified sample identification. To verify the applicability of the model, we used a certain amount of natural gas industrial customer data from a large state-owned oil and gas enterprise for application analysis and effectively divided customer value into three groups, demand-serving, demand-potential, and demand-incentive, according to their value characteristics and behavioral features. The results indicate that the framework proposed in this study can reasonably reflect and better characterize natural gas industrial customers' value with different types of behavioral feature data, which can provide technical support for big data smart natural gas consumer marketing.

KEYWORDS

natural gas industrial, customer value portrait, value assessment framework, PAM clustering, smart marketing

1 Introduction

Natural gas plays a significant role worldwide as well as in China, which can be used as a transitional energy source to mitigate climate change and reduce pollution, since it produces 50% less CO_2 emissions than coal and 30% less emissions than oil (Howarth, 2014). Natural gas is usually used in different sectors, such as residential, industrial, and commercial, but the industrial customer accounts for the highest proportion (Farzaneh-Gord et al., 2012;

Farzaneh-Gord et al., 2013; Farzaneh-Gord and Rahbari, 2018). In China, the use of natural gas is leading to an era full of opportunities and challenges under the background of globalization, accelerating energy transformation, complexity of the geopolitical environment, trial of energy security, and, especially, the balance between the downside pressure of dual-carbon goals and the need to maintain its economic development (Li, 2022). In early studies that focus on natural gas in China, domestic consumption was considered important by scholars. Later, it was proven that companies deserved more attention because they consumed much more natural gas than families (Nikolaidis et al., 2009). In the last few years, with the ownership unbundling of natural gas pipelines and storage facilities, the upstream gas suppliers have paid more attention to the downstream consumer market, and the competition among gas supply companies in the end-consumer market has become increasingly fierce; the market-oriented transaction mechanism will play an essential role for China to achieve a large-scale and efficient national allocation of new energy (Zhou et al., 2022). How to differentiate gas supply services to attract more customers according to customer characteristics has become a key concern to gas sellers. According to the China Natural Gas Development Report, in 2021, nearly 373 billion cubic meters of natural gas were consumed, that is, up to 12.7 percent year on year. In the natural gas consumption category, industrial fuels make up the bulk, which has become the core driving force of natural gas consumption growth, and industrial customers have become the main consumers. Therefore, accurately identifying the natural gas industrial customer demand characteristics and explicitly giving the industrial customer value portrait, which is of great significance to increase the comprehensive revenue of gas suppliers, enhance market competitiveness and optimize resource allocation.

With the continuous development of new technologies such as Internet and Big Data, customer classification technology, represented by customer profiling technology, has become a focus for companies in various fields to carry out differentiated marketing services. This study presents a data-mining approach to characterize the natural gas consumption behavior of industrial customers, which can visually demonstrate the characteristics of various customers' gas consumption behaviors and provide a basis for gas suppliers to differentiate their marketing services accordingly. Specifically, we summarize the current research status of the customer value-based profiling technology in the natural gas industrial sector and use big data modeling methods to build a value portrait of natural gas industrial customers. Our study could provide differentiated marketing strategy suggestions for gas sales operators based on the customer value profiling results.

Previous studies on customer value assessment are mainly divided into two categories. On the one hand, from a qualitative perspective, exploring the composition of multidimensional attributes of customer value and conducting comprehensive customer value assessment by constructing a customer demand characteristics index system, which is mainly applied to customer segmentation practice, is necessary. On the other hand, from the perspective of quantitative research, the use of mathematical methods to directly account for the currency value of customers, which is mainly used for customer value prediction, is also vital. The results predicted by those methods helped build a regional distributed energy system which could reduce emissions, reduce power consumption, and increase the safety and reliability of power grids (Hou et al., 2021). This study is centered on the construction of the value evaluation system for large industrial customers of natural gas and the classification of customer value based on customer portrait techniques. Due to the variety and diversity of data, it was hard to find out the relationships between daily phenomenon and data, but it was much easier with the help of clustering methods (Doğan, 2018). Clustering could be used to divide customer groups (Vieira et al., 2018), test habits (Garcia et al., 2017), and analyze user needs (Yang et al., 2015). CVIs (cluster validity indices) and PFCM (possibilistic fuzzy C-means algorithm) could be used to predict behaviors of the natural gas distribution network and nodal gas consumptions (Askari et al., 2015). Faced with complicated standards and needs, the PPFCI (projection pursuit fuzzy clustering model) technique was developed by combining the projection pursuit model with a fuzzy clustering iterative model (Wang and Yang, 2021). CFCM (causal fuzzy C-means) could effectively improve the scheduling accuracy and reduce the gas diffusion (Jin et al., 2018). A neuro fuzzy multivariate algorithm was developed to estimate accurate gas consumption with noisy inputs (Azadeh et al., 2013). By analyzing the data and consumer behaviors, the connection between them could be found (Beckel et al., 2014), and an accurate feature-based portrait classification could be made (Stephen et al., 2014). Dividing the samples into different groups helped distinguishing types of different features; thus, the behavioral features of different groups could be found and used to optimize resource allocation (Li, 2022). Prediction of the demand of resources by clustering can get suppliers better prepared (Rahim et al., 2019). Onur (2021) applied the intuitionistic fuzzy C-mean clustering algorithm (IFCM) to the natural gas industry to cluster and analyze consumption data of gas-using companies and develop a visual calendar diagram for corporate natural gas consumption behavior research. In addition to traditional clustering algorithms, data mining and machine learning algorithms in artificial intelligence have also been applied to customer segmentation by most scholars. Data mining could be helpful for gas consumption predictions (Palinski, 2018; Stuivenvolt-Allen and Wang, 2019). Then, the results of clustering and data mining could be used to make accurate customer portraits. Customer portraits distinguished the features of different customers and could be used to better optimize the resource allocation and response to demands (Guan et al., 2021; Chen et al., 2022). In order to effectively handle multidimensional attribute data of electric power customers and improve customer classification performance, an electric power customer classification method based on the symmetric uncertainty of feature subset generation and integrated learning was proposed to generate redundant and significant feature sets and to make customer classification decisions based on these significant feature sets (Piao et al., 2019). Dui et al. (2020) proposed an iterative decision tree (ITDT)-based customer classification method in the study of telecommunication service target customer classification, which extracts customer consumption features to effectively identify student and non-student customer groups by modeling and analyzing customer communication data provided by telecommunication operators. When faced with

resource-consuming customer classifications, data mining worked well in transforming data into behavioral classifications (Cominola et al., 2019; Kang and Reiner, 2022) and could even reduce energy consumption by improving energy efficiency (Shan et al., 2022).

There are some studies on customer value assessment and customer segmentation in industrial fields such as the electric power industry; however, similar research studies on natural gas consumers are not abundant. Due to the exclusivity of electric power–customer relationships, electric power industrial customers have a single consumption choice, while natural gas customers have a variety of consumption choices due to its own commodity, which means the natural gas industry has a variety of accessible suppliers due to the storable nature of the commodity and the development of the LNG market, so there are fundamental differences in the customer value composition and customer segmentation between the two industries. In addition, the aforementioned studies rarely consider the inclusion of the customer industry-type attribute in the construction of a comprehensive customer value assessment system, and the difference in the demand behavior between customers in different industries directly affects the consumption behavior of customers and thus the accuracy of customer classification results. In addition, former studies do not take the influence of mixed data-type indicators of the clustering algorithm in implementing customer value segmentation techniques into consideration and cannot provide a similarity measure involving multiple data-type indicators.

To solve such problems by designing a model that works well in the aforementioned situations, this study developed a new framework of consumer classification and portrait evaluation, thus achieving the goal of accurately classifying industry consumers of natural gas in China by the following innovative aspects.

(1) We rediscover the value composition of natural gas industrial customers and establish a set of indicators to reflect the customer value in different dimensions with mixed data types.
(2) A visualizable customer value classification model has been established by combining Gower's dissimilarity coefficient with the PAM clustering algorithm for better customer classification.
(3) A three-category natural gas industrial customer classification based on user value portrait was developed and was further put forward for the corresponding marketing suggestions.

Our study is closely related to the current natural gas market situation. To verify the applicability of the model, we used a certain amount of natural gas industrial customers' behavioral feature data from a large state-owned oil and gas enterprise. The results indicate that the framework proposed in this study can reasonably reflect and better characterize the natural gas industrial customers' value with different types of behavioral feature data, which can provide technical support for big data natural gas consumer smart marketing.

The remainder of this article is organized as follows. Section 2 elaborates on the construction of the customer value evaluation system and the customer value segmentation model for the natural gas industry. Section 3 presents the application of the model and the empirical analysis of the model using several natural gas industrial customers. Section 4 presents the value portrait and differentiated marketing services of different natural gas industrial customers, and Section 5 concludes the article.

2 Model construction

2.1 Customer value assessment model

2.1.1 Natural gas industrial customer value composition

The current customer value classifications among finance, logistics, retail, and other industrial parameters are basically based on the concept of customer lifetime value, which divides customer value into current value and potential value from the perspective of time and then is subdivided into four indicators of profit contribution, cost occupation, loyalty, and creditworthiness to characterize customer value. In terms of a natural gas industrial customer, the end-consumer markets are mainly characterized by competition among three large state-owned oil and gas companies, i.e., CNPC, SINOPEC, and CNOOC. With the advent of perfectly competitive LNG resources, some large-scale industrial customers have a variety of gas purchase options. Therefore, stability and loyalty have now become important indicators in assessing the value of natural gas industrial customers. Moreover, with the continuous attention of the society to corporate social responsibility, the reputation and social responsibility will also affect customer value. Although customer value is reflected in all aspects, in this article, we refer to other studies on the value assessment of China's power industry customers, combined with the actual situation of the natural gas industry; the value of large industrial customers of natural gas is described in Figure 1.

As shown in Figure 1, the value of a natural gas industrial customer consists of two aspects: currency value and non-currency value. In terms of currency value, it consists of the current economic value and potential economic value, which represent the scale and growth rate, mainly characterizing the direct economic benefits brought to gas supply enterprises by customers' current and future gas consumption behavior during the statistical period, i.e., the currency value of customers. In terms of non-currency value, we call it the orderly gas consumption value and social responsibility value. In the case of orderly gas consumption value, it consists of safety and stability value, and the demand response value, i.e., the safety and stability for an industrial customer of natural gas and the degree of customers' enthusiasm to participate in the regional gas peaking plan of gas supply enterprises and their contribution to demand response. As for the social responsibility value, it contains the contract and credit value, and the sustainable development value. The contract and credit value represents the degree of compliance and customer loyalty value of gas customers in signing transaction contracts with gas supply enterprises, while the sustainable development value is assessed from the perspective of the external environment of the gas-using enterprise and the social responsibility undertaken by the enterprise and is mainly divided into the value of energy-saving contribution and the value of sustainable development potential.

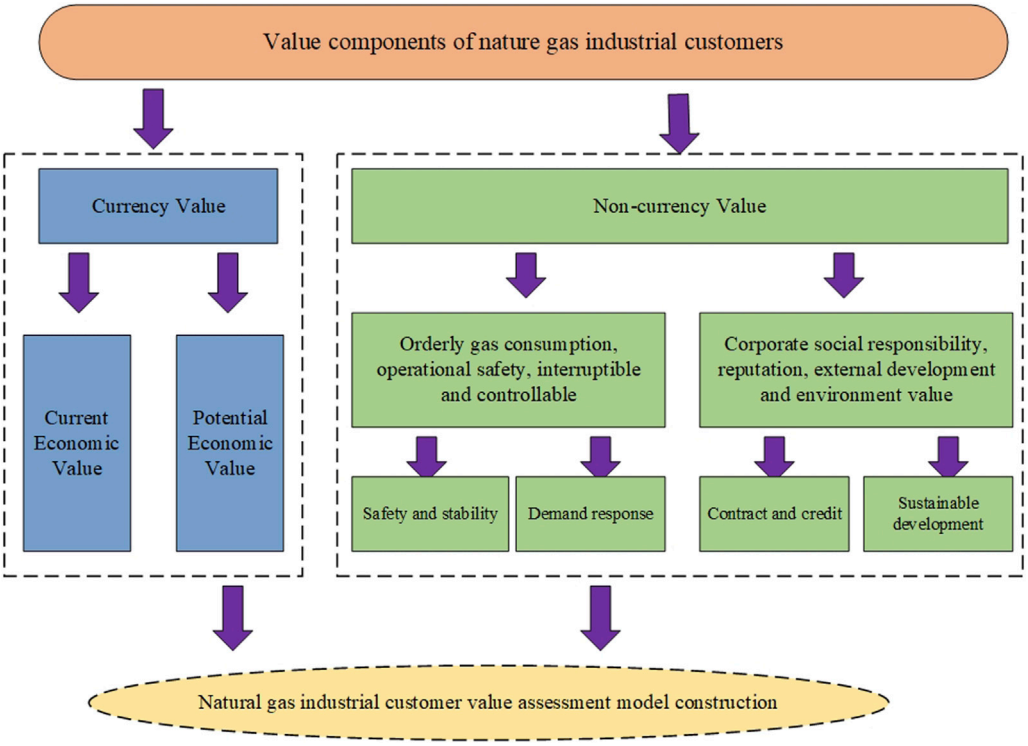


FIGURE 1
Value composition of natural gas industrial customers.

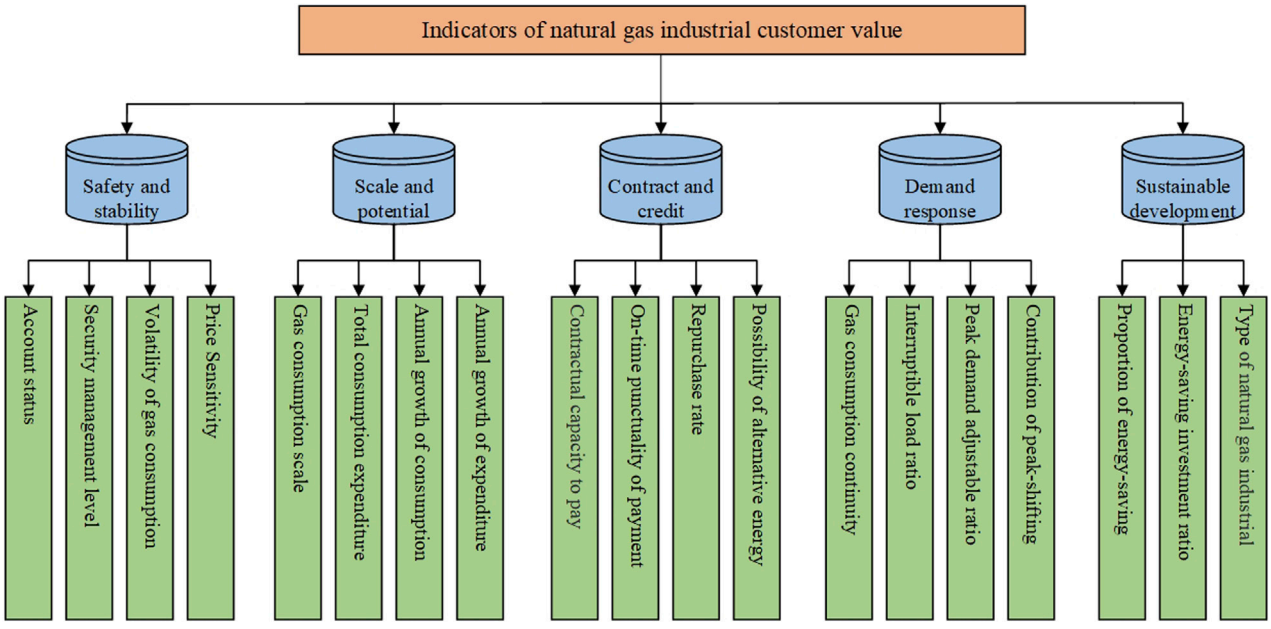


FIGURE 2
Evaluation system of natural gas industrial customer value.

TABLE 1 Meaning of sub-indicators.

Sub-indicator	Symbol	Type	Meaning
Account status	a1	Nominal	Customer account usage
Security management level	a2	Ordinal	Customer safety production operation
Volatility of gas consumption	a3	Numerical	Monthly fluctuations in customer gas consumption
Price sensitivity	a4	Numerical	Maximum price of gas that customers can afford
Gas consumption scale	b1	Numerical	Total amount of gas used by customers
Total consumption expenditure	b2	Numerical	Total consumption of gas by customers
Annual growth of consumption	b3	Proportional	Annual growth rate of gas consumption by customers
Annual growth of expenditure	b4	Proportional	Annual growth rate of gas consumption expenditure
Contractual capacity to pay	c1	Proportional	Turnover performance of customer gas cost
On-time punctuality of payment	c2	Proportional	Punctuality of gas cost turnover for customers
Repurchase rate	c3	Proportional	Customers buy back gas in gas supply enterprises
Possibility of alternative energy	c4	Nominal	Probability of customers using alternative energy
Gas consumption continuity	d1	Nominal	Can customer's production be interrupted
Interruptible load ratio	d2	Numerical	Proportion of customer interruptible load
Ability of peak-shifting	d3	Proportional	Proportion of customers avoiding peak periods
Peaking contribution rate	d4	Proportional	Contribution of customers in the regional peaking plan
Proportion of energy-saving	e1	Proportional	Proportion of energy-saving plant equipment
Energy-saving investment ratio	e2	Proportional	Proportion of investment in energy-saving
Type of natural gas industrial	e3	Nominal	Categories of industrial customers

2.1.2 Indicators of customer value assessment

In order to ensure the scientific accuracy and operability of customer value segmentation modeling, quantifiable segmentation variables need to be selected to measure and evaluate the two aspects of the value composition of large natural gas industrial customers. The selection of segmentation variables is based on the principles of independence, scientificity, measurability, typicality, and comprehensiveness, to ensure the operability of the segmentation process and the accuracy of the results. By using the literature induction method and combining the existing data of relevant gas supply enterprises, this article selects the segmentation variable indicators of each value dimension and constructs a gas customer value assessment index model, as shown in Figure 2, and the meanings and data types of each sub-indicators are shown in Table 1.

2.1.3 Processing of assessment indicators

In order to reduce the impact of autocorrelation between redundant segmentation indicator variables on the final value assessment results of large industrial customers of natural gas, the aforementioned indicator system needs to be further revised and streamlined before conducting data analysis and modeling.

2.1.3.1 Remove the invalid indicators

The missing value ratio method assumes that when a data column contains some missing values, it is less likely to contain useful information. Therefore, in order to ensure the usefulness of the information contained in the segmentation variable indicators in the index system, data columns with missing values greater than a certain threshold can be eliminated. In this study, if there are too many missing values in the sub-indicators, it signifies that the indicators lack the support of actual business data and violate the principle of measurability and operability of indicator selection and should be eliminated.

2.1.3.2 Eliminate indicators with low discrimination

The low variance filtering method assumes that if the values of a column in the dataset are essentially the same, i.e., its variance is very low, the low variance data column carries very little useful information and cannot satisfy the basic condition that there is a clear distinction between different objects in the values of the indicator. In practice, the indicator of sample variance is used to measure each segmentation variable indicator, and the calculation method is shown in Eq. 1.

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}, \tag{1}$$

where S^2 denotes the variance of each indicator variable and x_i and $\bar{x}$ denote the sample value and mean value of each variable, respectively. According to the principle of low variance filtering, the smaller the variance of a subdivided variable indicator, the weaker its ability to contain useful information, and the indicator

variables with smaller variance can be deleted according to the calculation results. Since this method is not applicable to categorical variables, only numerical and binary variable indicators are further reduced in this article, and the indicator data need to be standardized before the operation in order to reduce the influence of the magnitude.

2.1.3.3 Eliminate indicators with high correlation

High correlation filtering considers that if two indicator variables are highly correlated with each other, they have similar trends and may carry similar information. Such variables can lead to a reduction in the performance of the classification model, so the correlation between independent numerical variables can be calculated, and when the correlation coefficient exceeds a certain threshold, the indicator variable with high correlation with the target variable is usually chosen to be retained, and the other indicator is excluded to achieve the indicator simplification goal. In this article, Pearson's correlation coefficient is used to measure the correlation between indicator variables, which is calculated as shown in Eq. 2.

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)\left(\frac{y_i - \bar{y}}{\sigma_y}\right), \tag{2}$$

where r denotes Pearson's correlation coefficient; $\bar{x}$ and σ_x denote the sample mean and sample standard deviation, respectively; and n is the sample size.

2.2 Customer value classification model

The customer value classification model mainly uses the clustering algorithm to rate the value of natural gas industrial customers. Because of the proposed indicators contain multiple data types in the customer value assessment model, a visualizable customer value classification model has been established by combining Gower's dissimilarity coefficient with the PAM clustering algorithm.

2.2.1 Dissimilarity of indicators

Measuring the similarity or dissimilarity among data objects is the basic work of cluster modeling analysis, and the segmentation index variables selected in this study contain a variety of mixed types of data, such as numerical, nominal, and proportional data, so the traditional methods of measuring dissimilarity of data objects cannot be well-applied (Gower, 1971). In this article, we propose to use the *Gower* dissimilarity coefficient to measure the phase dissimilarity of nominal, ordinal, and binary data in the segmentation index variables.

Gower's dissimilarity coefficient assumes that the dataset contains M type variables, and the dissimilarity between data objects a and b, $d(a,b)$, is defined as shown in Eq. 3.

$$d(a,b) = \frac{\sum_{f=1}^{M}\delta_{ab}^{(f)}d_{ab}^{f}}{\sum_{f=1}^{M}\delta_{ab}^{(f)}}, \tag{3}$$

where M is the number of variable types, f is the variable, $d_{ab}^{(f)}$ is the dissimilarity between the data objects a and b under the variable f, $\delta_{ab}^{(f)}$ is the variable indicator, $\delta_{ab}^{(f)}$ takes 0 if x_{af} or x_{bf} data do not exist (no measurement for the object a or the variable f for the object b), or $x_{af} = x_{bf} = 0$, and if the variable f is an asymmetric binary variable, conversely, $\delta_{ab}^{(f)}$ takes 1. The variable f calculates the direct dissimilarity between the objects a and b with respect to its specific data type in the following steps: 1) if the variable f is a binary or symbolic variable, if $x_{af} = x_{bf}$, then $d_{ab}^{(f)} = 0$ and *vice versa* $d_{ab}^{(f)} = 1$; 2) if the variable f is an interval-valued variable, then $d_{ab}^{(f)} = |x_{af} - x_{bf}|/(max_h x_{hf} - min_h x_{hf})$, where h is the range of all possible value changes for the variable f; 3) if the variable f is a sequential variable or a proportional numeric variable, the variable can be converted using a method such as logarithmic conversion, and then the converted variable is treated as an interval numeric variable for calculation.

2.2.2 Optimal number of clustering

Cluster analysis is an unsupervised learning tool, and most clustering algorithms generally require a pre-determined number of clusters; different numbers of clusters are passed to the algorithm and the final output results will vary, so how to determine the optimal number of clusters becomes an urgent problem to be solved before conducting cluster analysis in this study. Tibshirani et al. (2001) proposed the *Gap Statistic* method to solve the problem of determining the optimal number of clusters, which is basically defined as follows.

A sample dataset containing n mutually independent observations, with each observation being x_i, $i = 1, 2\cdots, n$. The dataset is clustered into k classes, denoted as C_k, representing the sample points belonging to the $C_r r$ class, and n_r represents the number of observations belonging to the C_r class. The sum of the distances between two observations in the r class is shown in Eq. 4.

$$D_r = \sum_{i,i' \in C_r} d_{ii'}. \tag{4}$$

Eq. 5 defines an in-group offset W_k:

$$W_k = \sum_{r=1}^{k}\frac{1}{2n_r}D_r. \tag{5}$$

The interval statistic $Gap_n(k)$ is defined as in Eq. 6.

$$Gap_n(k) = E_n^*(log(W_k)) - log(W_k), \tag{6}$$

where E_n^* represents the expectation for the reference dataset $log(W_k)$.

The basic idea of the interval statistics method is to compare the expected value of the reference dataset with the observed dataset so that the value of k with the fastest decrease in $log(W_k)$ is used as the optimal number of clusters. The reference dataset is usually the mean distributed data with a sample size of n generated by sampling within the range of values of the study sample using the Monte Carlo algorithm.

In actual practice, determining the optimal number of clusters is also carried out by the *Elbow* method and the average contour method. The *Elbow* method considers the within sum of squares (*WSS*) as a function of the number of clusters, and since *WSS* measures the compactness of the clusters, it seeks to minimize *WSS* as much as possible. Based on this, the *Elbow* method to determine the optimal number of clusters is defined as follows: assuming that the dataset is divided into k classes, for each k, the *WSS* is calculated

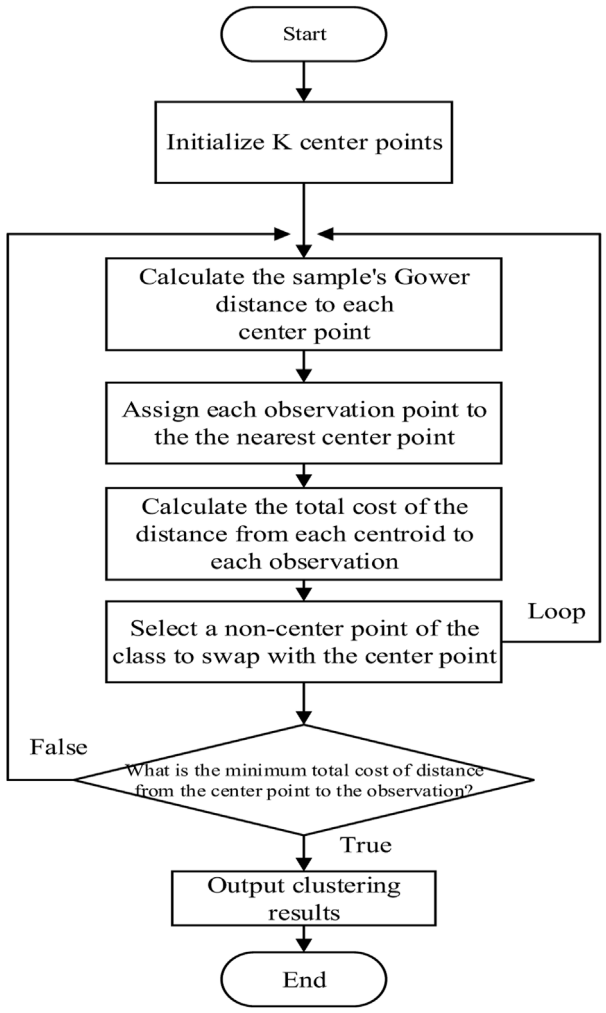


FIGURE 3
Flow chart of the PAM clustering algorithm.

and the corresponding *WSS* curve is plotted, and the inflection point in the curve is usually considered the optimal number of clusters. Similarly, the average contour coefficient is calculated by averaging the contours of observations with different k values, and maximizing the number of average contours within the possible range of k is the optimal number of clusters. In order to avoid a single method being used to determine the optimal number of clusters, which affects the accuracy of clustering results, this study intends to test the determination of the optimal number of clusters by combining the aforementioned three methods for analysis.

2.2.3 PAM clustering algorithm

The main purpose of building a customer classification model based on customer value is to classify and rate the value of large industrial customers of natural gas by choosing a suitable clustering algorithm. The K-means clustering algorithm has the advantage of being simple and fast and is widely used in the field of clustering analysis. However, this algorithm is sensitive to outliers in the sample because it is based on the sample mean, and extreme customer sample points can affect the accuracy of this algorithm in the actual customer segmentation practice. Thus, an improved K-means algorithm, the partitioning around medoids (PAM), has been created and applied in clustering. In the PAM clustering algorithm, we can select the actual observations in the sample as the cluster centers and optimize the new cluster centers by calculating the minimum value of clustering from each point in the class cluster to all other points except the cluster center when modifying the cluster centers. Based on this feature, the PAM makes up for the shortcomings of the K-means which is sensitive to sample noise and isolated points and makes the clustering model more robust. The customer sample information in this study is a mixed data-type attribute, and the PAM clustering algorithm can combine the *Gower* dissimilarity coefficient for clustering analysis with mixed data samples, which is more suitable for dealing with customer segmentation studies of mixed data than K-means clustering. The specific idea of the PAM clustering algorithm is as follows: 1) random selection of k observations as centroids; 2) calculation of the distance of the sample to each centroid at *Gower*; 3) assigning each observation to the nearest centroid; 4) calculation of the total cost of the distance from each centroid to each observation; 5) selection of a non-center point and swapping it with the center point and repeating steps 3) and 4); and 6) choosing the solution with the lowest total cost of distance from the center point to the observation and repeating steps 3), 4), and 5) until the center point no longer changes. The flow chart of the customer value classification algorithm based on the PAM clustering algorithm is shown in Figure 3.

2.2.4 Clustering result examination

In order to test the cluster result of the classification model, this study used Rousseeuw's contour coefficient (*Silhouette Coefficient*) for examination (Rousseeuw, 1987), which combines both cohesion and separation to evaluate the operational effectiveness of the clustering algorithm on the basis of the same data. The basic principle of the contour coefficient is as follows: 1) calculating the average distance of sample i to other samples in the same category (i); $a(i)$ is called the intra-cluster dissimilarity of sample i and the smaller $a(i)$ is, the more the sample i should be classified into that category; 2) calculating the average distance of sample i to all samples of another class $C_j b(i)$; $b(i)$ is called the dissimilarity of sample i to cluster C_j. Let the separation of sample i be $(i) = min\left\{b_{i1}, b_{i2}, \cdots, b_{ij}\right\}$, i.e., the inter-cluster dissimilarity of sample i is the minimum value of the average distance from this sample to all samples of all other clusters. The larger $b(i)$ is, the less the sample i belongs to other clusters; and 3) based on the intra-cluster dissimilarity and inter-cluster dissimilarity of the sample, the contour coefficients are defined as shown in Eq. 7.

$$S(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}. \tag{7}$$

The range of the contour coefficient is [−1, 1] and when the contour coefficient is close to 1,this indicates that the sample i is reasonably clustered, while when the contour coefficient is close to −1, this indicates that the sample i is misclassified and should be divided into clusters that are closer together. If the contour coefficient is close to 0, this indicates that the sample i is at the boundary of two clusters. The mean value of the contour coefficients of all samples is called the

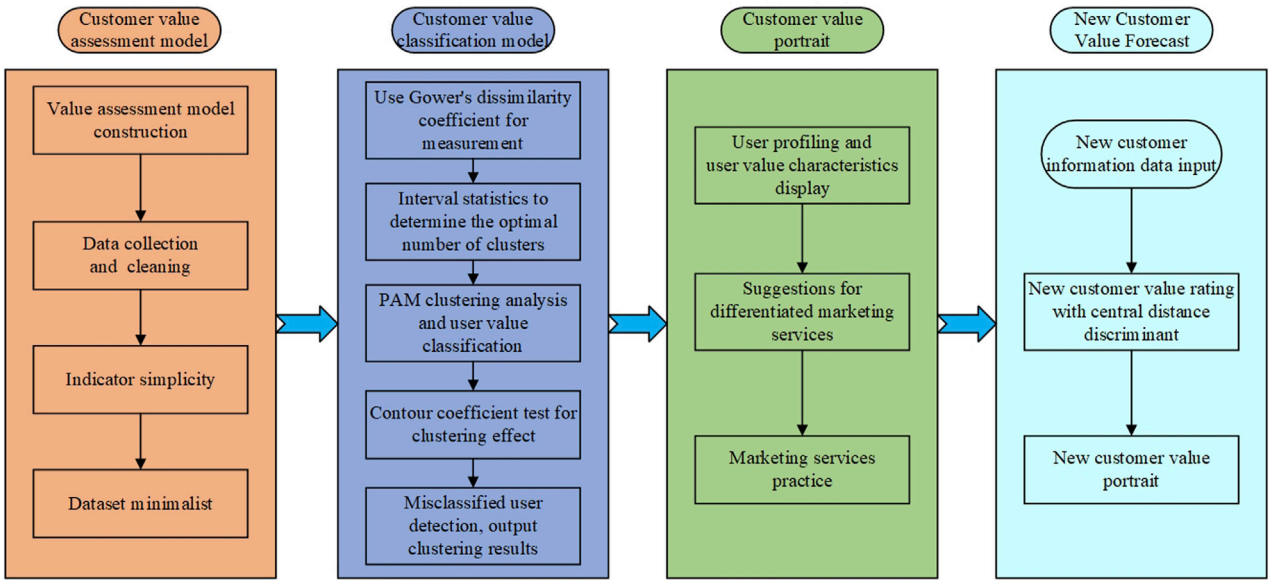


FIGURE 4
Process of the natural gas industrial customer value portrait.

contour coefficient of the clustering result, which is a measure of whether the cluster is reasonable and valid. The range of the contour coefficient of clustering results is [−1, 1], and the larger the value, the closer the similar samples are to each other, and the farther the different samples are from each other, the better the clustering effect.

2.3 Customer value portrait procedure

Based on the natural gas industrial customer value assessment and classification model, we give the specific process of the natural gas industrial customer value portrait, as shown in Figure 4.

Our natural gas industrial customer value portrait includes the value categories of a customer and the value dimension depiction of five aspects. According to the customer value portrait, an optimal marketing plan can be formulated to achieve the purpose of precision marketing. Moreover, when a new industrial customer appears in the market, we can easily get the value portrait of the new customer by inputting the relevant data of the new customer and making comparative analysis with existing industrial customers in our established model framework and then making corresponding marketing plans to establish a better customer relation.

3 Practical application analysis

In order to verify the practicability of the model, this study collected a certain amount of data on natural gas industrial customers from a large state-owned oil and gas enterprise. The model is implemented using R language version 4.1.2 and programmed on the RStudio platform.

3.1 Data collection

Based on the established natural gas industrial customer value assessment model, we collected 19 indicators of each customer. Considering the availability of data, this study has collected 76 natural gas industrial customers' value characteristic data from a large state-owned oil and gas enterprise.

Based on the customer value assessment model, the derived data were sorted, and the indexes were simplified by applying the missing value ratio method, low variance filtering, and high correlation filtering in turn. After data cleaning, 21 samples with more missing values of indicators (missing ratio >80%) and a total of 55 large natural gas industrial customers were selected for analysis finally. Additionally, according to the processing of assessment indicators, the high missing rate of indicators 2 and $e1$ are removed, the indicator $a1$ with zero or low variance is deleted, and the indicator with high correlation $b4$ is removed according to the correlation coefficient between indicators calculated by Eq. 2. Finally, the streamlined dataset is obtained to provide data preparation for the PAM clustering algorithm.

3.2 Clustering analysis

Based on the selected natural gas industrial customers, we further performed the clustering analysis. First, for the approximate dataset, the distance between samples is calculated according to Eq. 3 to complete the measure of mixed-attribute dissimilarity in the sample objects. Then, the interval statistics of the approximately parsimonious dataset were calculated by applying Eqs 4–6 in turn, and the *Elbow* and average contour plots were drawn to determine the optimal number of clusters; the results are shown in Figure 5.

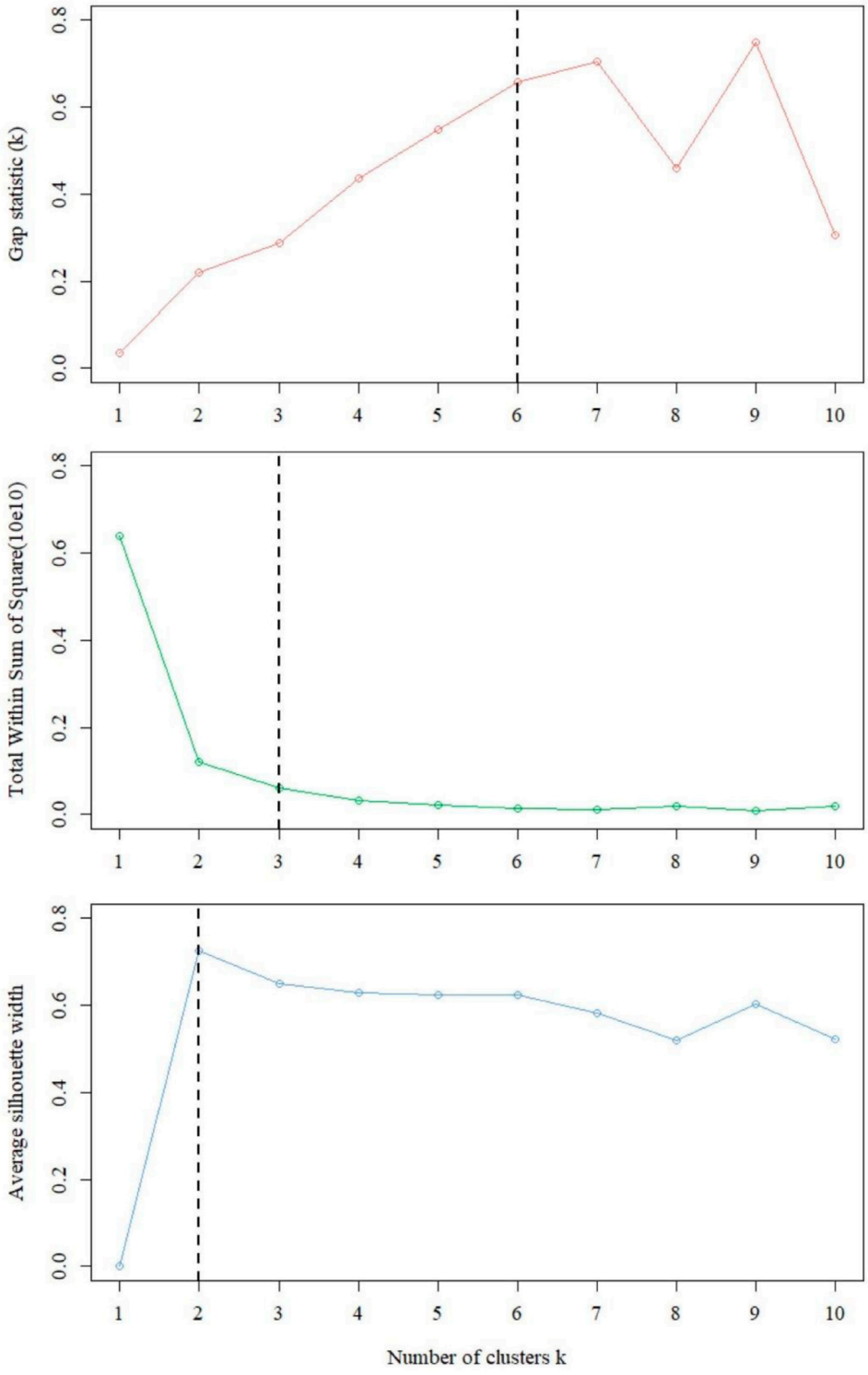


FIGURE 5
Optimal number of clusters in different methods.

As shown in Figure 5, when k is taken as 6, interval statistics takes the maximum value. When k is taken as 3, the decreasing trend of WSS becomes obvious. When k is taken as 2, the average contour coefficient is highest. Combining the aforementioned results and the actual characteristics of the sample data, $k = 3$ is chosen as the optimal number of clusters.

The determined optimal number of clusters $k = 3$ was passed to the PAM clustering algorithm for cluster analysis, and the clustering results are shown in Figure 6.

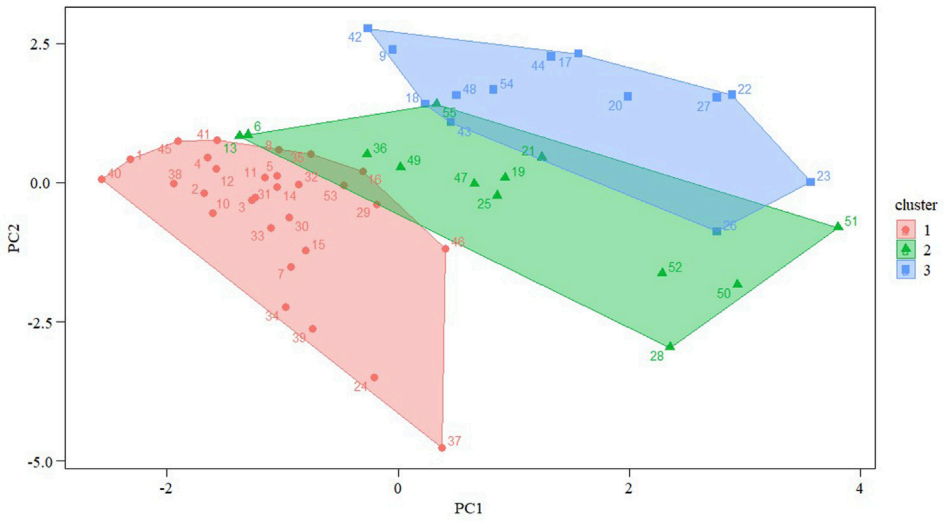


FIGURE 6
PAM clustering effect.

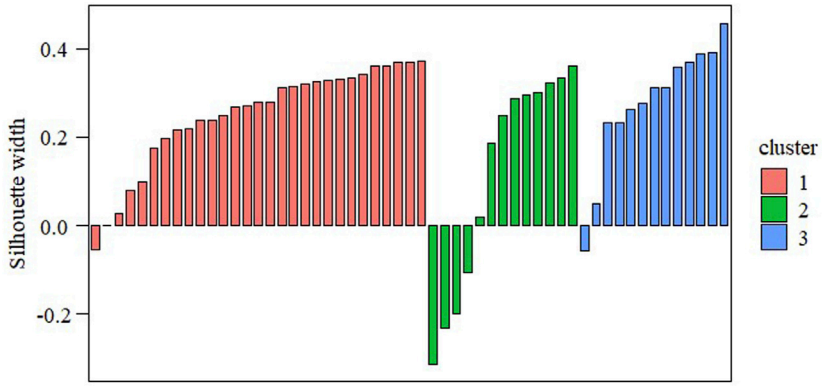


FIGURE 7
PAM clustering contour coefficient.

The PAM clustering effect was tested by applying Eq. 7 to calculate the contour coefficients of each cluster, and the contour coefficients of each cluster are plotted in Figure 7.

As shown in Figure 7, after PAM cluster analysis, the sample points with customer IDs 451, 126, and 79 were used as cluster centroids to classify large industrial customers of natural gas into three categories based on value assessment indicators, with the first category containing 29 customer samples, the second category containing 13 customer samples, and the third category containing 13 customer samples. From an overall perspective, the average contour coefficient of each cluster of clustering results reaches 0.22. Specifically, the average contour coefficient of the first category is 0.25, that of the second category is 0.12, and that of the third category is 0.28. The overall clustering effect is less satisfactory, indicating that some samples have incorrect clustering results. When the misclassified samples are found, we put those whose values are more ambiguous in the closest classification, while those that are close to the boundary point are classified into the glass industry. In terms of individual samples, there are samples with negative contour coefficients in all three clusters, as shown in Figure 7. By accurately identifying the misclassified samples with negative contour coefficients for re-clustering and classifying the customer value ratings, the classification of customer value is achieved.

4 Portrait results and discussion

4.1 Portrait result analysis

According to the customer value portrait modeling analysis, the natural gas industrial customer groups contained

TABLE 2 Typical characteristics of different customers' value.

Value dimension	Indicator	Typical characteristics of customers		
Customer value portrait	Value level	One-star	Two-star	Three-star
Security and stability	Account status	Normal	Normal	Normal
	Volatility of gas consumption	10.86	15.28	7.51
	Volatility of gas consumption	1.97	1.66	1.91
Scale and potential	Gas consumption scale	1,405.06	20,715.49	48,357.88
	Total consumption expenditure	2,404.99	28,380.22	70,773.74
	Annual growth of consumption	15.248%	9.31%	−0.02%
	Annual growth of expenditure	−15.923%	0.61%	−10.68%
Contract and credit	Repurchase rate	100%	100%	99.98%
	On-time punctuality of payment	97%	100%	100%
Demand response	Gas consumption continuity	Non-interruptible	Interruptible	Interruptible
Sustainable development	Industrial type	Steel	Methanol	Synthetic ammonia

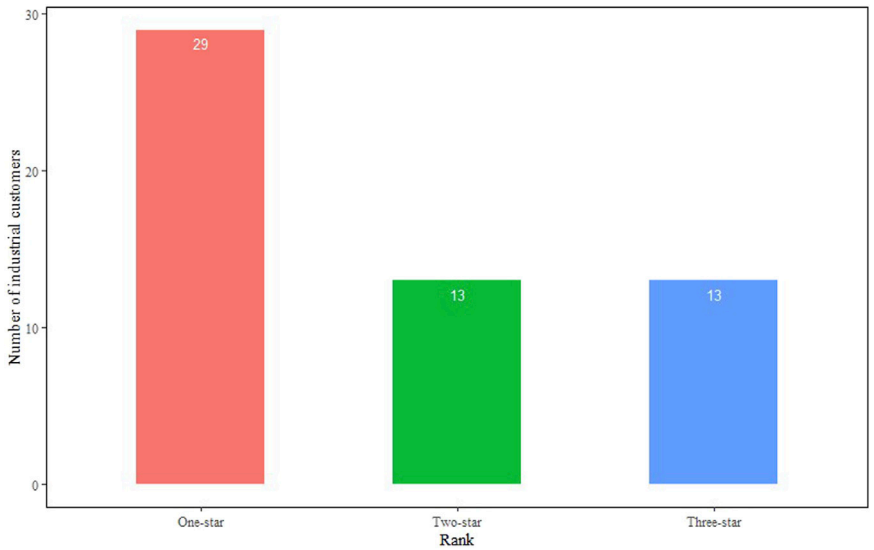


FIGURE 8
Number of natural gas customers by the value level.

information of six types, i.e., glass, steel, ammonia, methanol, aluminum, and ceramics. The value level and typical characteristics of natural gas industrial customers' value portrait are summarized in Table 2.

Based on the modeling results, the customers' value is rated into three levels. First, one-star customers are demand-motivated customers. Second, two-star customers are demand-potential customers. Third, three-star customers are demand-service customers. The number and percentage of different natural gas industrial customer value levels are shown in Figures 8–11.

4.2 Suggestions for marketing

By modeling and profiling the value of large industrial customers of natural gas, combined with specific marketing practices, the following differentiated marketing services are proposed for the three types of demand-based customers.

First, the one-star customers are demand-motivated customers. One-star customers are characterized by a small gas demand and a large number of customer groups, and the marketing service priority of these customers is the lowest. Gas suppliers should actively establish a customer gas incentive system while maintaining basic

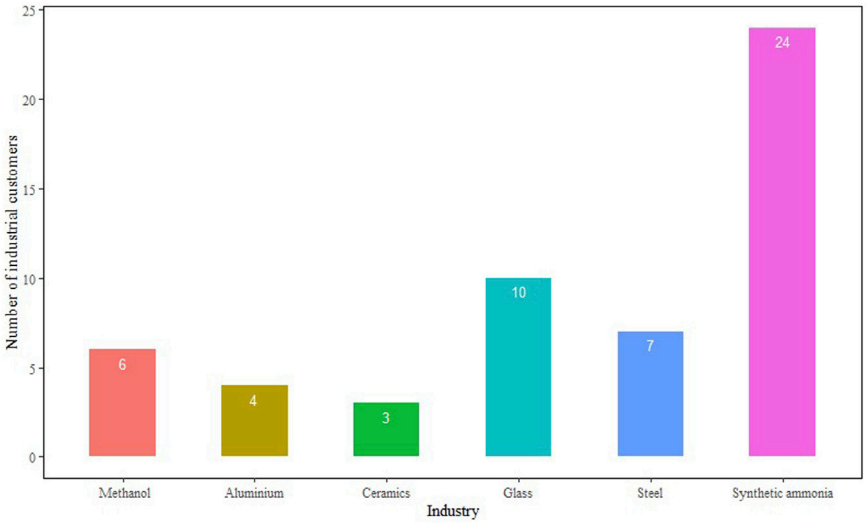


FIGURE 9
Number of natural gas customers by industrial categories.

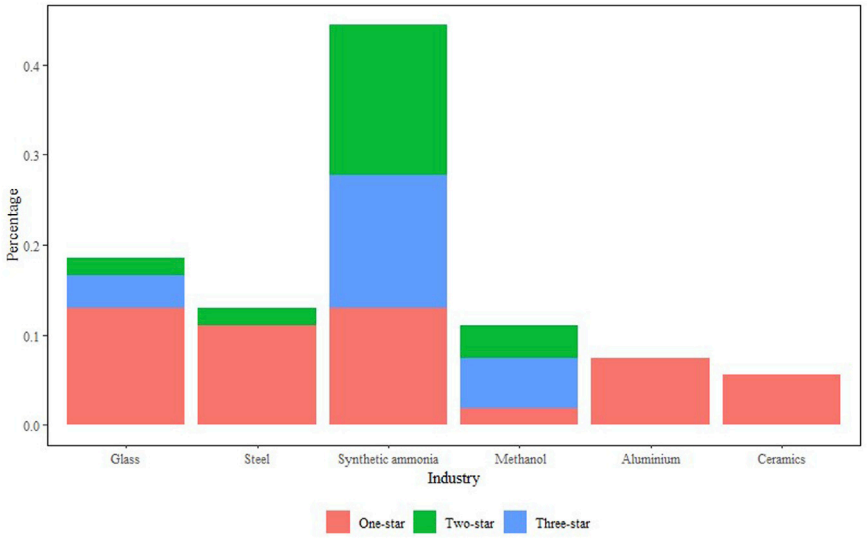


FIGURE 10
Percentage of natural gas customers by the value level and categories.

services in their marketing services and can explore the gas demand of these customers in specific ways such as a gas point system to stimulate the gas demand of these customers.

Second, the two-star customers are demand-potential customers. The main difference between two-star customers and three-star customers is that the volatility of gas consumption varies widely among these customers, followed by high price sensitivity, lower maximum affordable gas prices, and medium priority for marketing services. Therefore, it is recommended that gas suppliers should grasp the seasonal fluctuations of customers' gas consumption in addition to normal marketing services and increase gas supply services during the peak season to explore the gas demand potential of this group. At the same time, since most of the customers in this group are interruptible gas customers, they can be encouraged and guided to actively participate in the demand response programs of gas supply companies so as to bring into play the value of peaking and make this group of customers a potential customer group for future cooperation with gas supply companies.

Third, the three-star customers are demand-service customers. According to the value characteristics of three-star customers, they are the most valuable customers of gas suppliers and have the highest priority in marketing services. Gas suppliers can establish individual customer demand management systems for these

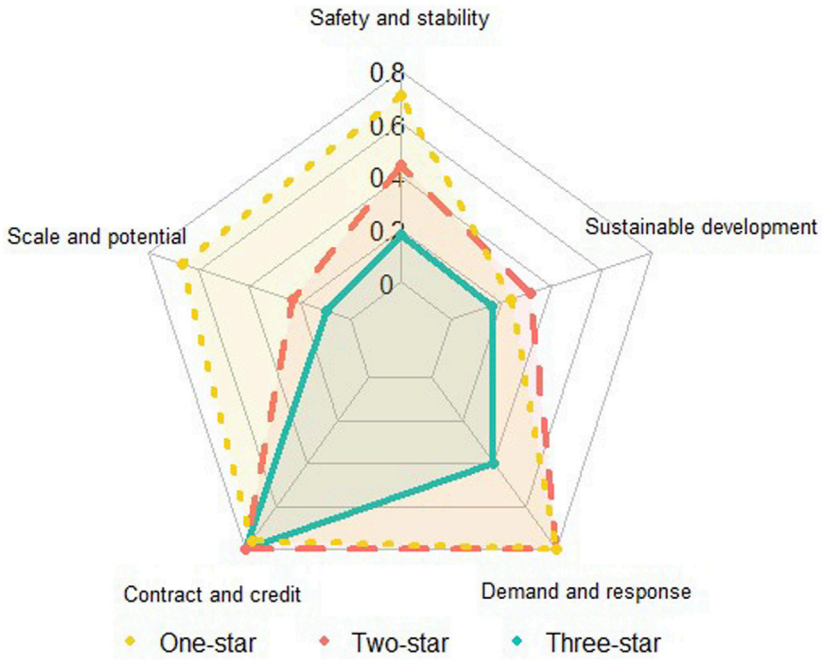


FIGURE 11
Radar chart of natural gas customer value distribution.

customers, set up special marketing staff to provide specialist services, open green service channels, carry out regular gas delivery and maintenance services, consolidate and enhance service satisfaction and loyalty of large gas customers to gas suppliers through VIP personalized services, and increase customer viscosity.

5 Discussion and conclusion

Customers are important assets of an enterprise, to some extent. Using big data modeling to explore the consumption characteristics of industrial customers and to conduct customer value profiling is helpful in identifying customers with value to the enterprise and in providing technical assurance to achieve the dual goals of maximizing enterprise profits and maximizing customer benefits. In this article, we take large industrial customers of natural gas as the research object, summarize the experience of value profiling of industrial customers in the electric power industry and other industries, construct a customer value assessment model and a customer value classification model in two steps, and conduct empirical analysis with actual consumption data of large industrial customers of natural gas. The findings of our research show that natural gas industrial customers can be divided into three groups: one-star, two-star, and three-star customers. The more stars a user has, the higher the performance. Although the number of three-star customers is not significant, they do deserve best service, and long-term trade with them is usually steady and profitable. Paying attention to two-star customers is a good choice as well because they have big potential. We also suggest that efforts on one-star customers should be limited.

The empirical results show that the model constructed in this article can achieve a reasonable rating and classification of customer value, and the customer portrait depicted based on the value characteristics of customer groups can accurately identify valuable industrial customers for gas supply enterprises. Based on the customer results combined with specific marketing practices, this article proposes differentiated marketing services for gas supply enterprises for various types of customers of different value groups, which is important for gas supply enterprises to improve market competitiveness, increase economic benefits, and improve customer relationships. However, there is still room for further development and improvement in the construction of the customer value assessment model and the customer portrait display. Subsequent research can focus on improvement and optimization of customer value segmentation techniques and big data platform-based customer value portrait display systems to better serve the specific marketing practice of gas supply enterprises.

This article realizes the value portrait of large industrial customers of natural gas by constructing a customer value assessment system and a customer value classification model in two steps; however, there are still some aspects of the research that need to be discussed and studied further. First, more corresponding customer business data to support the four levels of value indexes involving orderly gas consumption and sustainable development are needed to stress the importance of these indicators. In follow-up studies, researchers can further improve the customer value evaluation index system and reasonably balance the integrity and accessibility of data indicators to ensure the scientific nature of customer value evaluation. Second, the weights of each indicator can be unequally treated because they have different effects. Also, subsequent research can adopt intelligent algorithms for indicator appointment to improve the model performance. Third, in consideration of the expansion of the customer scale, the

stability and credibility of our framework should be further tested. We hold the opinion that our framework can work well under different situations like a bigger market with more customer information. Hopefully, a more detailed and concrete framework that can better assess customer value can be built with the support of bigger data. Finally, we can also consider trying fuzzy clustering, hierarchical clustering, and other clustering algorithms or data mining class algorithms for technical improvement as clustering methods or robust test methods in order to improve the credibility of the customer value hierarchical classification model.

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.

Funding

This research is funded by the National Natural Science Foundation of China under Grant Nos. 71804167, 72174188.

Conflict of interest

Author KP was employed by China National Petroleum Corporation.

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.

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

Askari, S., Montazerin, N., and Zarandi, M. H. F. (2015). Forecasting semi-dynamic response of natural gas networks to nodal gas consumptions using genetic fuzzy systems. *Energy* 83, 252–266. doi:10.1016/j.energy.2015.02.020

Azadeh, A., Saberi, M., Asadzadeh, S. M., Hussain, O. K., and Saberi, Z. (2013). A neuro-fuzzy-multivariate algorithm for accurate gas consumption estimation in South America with noisy inputs. *Int. J. Electr. Power and Energy Syst.* 46, 315–325. doi:10.1016/j.ijepes.2012.10.013

Beckel, C., Sadamori, L., Staake, T., and Santini, S. (2014). Revealing household characteristics from smart meter data. *Energy* 78, 397–410. doi:10.1016/j.energy.2014.10.025

Chen, S. C., Gan, L., Chen, C. Y., Yu, K., Pi, H. M., Qian, Z. A., et al. (2022). Demand-response oriented multi-dimension refined portrait of adjustable resources based on load and survey data fusion. *Front. Energy Res.* 10, 968368. doi:10.3389/fenrg.2022.968368

Cominola, A., Nguyen, K., Giuliani, M., Stewart, R. A., Maier, H. R., and Castelletti, A. (2019). Data mining to uncover heterogeneous water use behaviors from smart meter data. *Water Resour. Res.* 55, 9315–9333. doi:10.1029/2019wr024897

Doğan, O. (2018). "Heuristic approaches in clustering problems," in *Handbook of research on applied optimization methodologies in manufacturing systems*. Editors Ö. Faruk Yılmaz and S. Tüfekçí (Hershey, PA, USA: IGI Global), 107–124.

Dui, P., Huang, D. S., Liu, Z. X., Zhang, H. Y., Zhang, J. G., and Yan, Q. J. (2020). "Power customer classification based on OCHNN-Kmeans and price setting of TOU," in 5th Asia Conference on Power and Electrical Engineering (ACPEE), Chengdu, P R CHINA, 252–256.

Farzaneh-Gord, M., Deymi-Dashtebayaz, M., Rahbari, H. R., and Niazmand, H. (2012). Effects of storage types and conditions on compressed hydrogen fuelling stations performance. *Int. J. Hydrogen Energy* 37, 3500–3509. doi:10.1016/j.ijhydene.2011.11.017

Farzaneh-Gord, M., Rahbari, H. R., and Nikofard, H. (2013). The effect of important parameters on the natural gas vehicles driving range. *Pol. J. Chem. Technol.* 14, 61–68. doi:10.2478/v10026-012-0104-3

Farzaneh-Gord, M., and Rahbari, H. R. (2018). Response of natural gas distribution pipeline networks to ambient temperature variation (unsteady simulation). *J. Nat. Gas Sci. Eng.* 52, 94–105. doi:10.1016/j.jngse.2018.01.024

Garcia, D., Puig, V., Quevedo, J., and Cugueró, M. A. (2017). "Big data analytics and knowledge discovery applied to automatic meter readers," in *Real-time monitoring and operational control of drinking-water systems*. Editors V. Puig, C. Ocampo-Martínez, R. Pérez, G. Cembrano, J. Quevedo, and T. Escobet (Cham: Springer International Publishing), 401–423.

Gower, J. C. (1971). A genera coefficien of similarity and some of its properties. *Biometrics* 24 (24), 623–637.

Guan, W. L., Zhang, D. L., Yu, H., Peng, B. G., Wu, Y. F., Yu, T., et al. (2021). Customer load forecasting method based on the industry electricity consumption behavior portrait. *Front. Energy Res.* 9, 742993. doi:10.3389/fenrg.2021.742993

Hou, J. M., Wang, J. J., Ji, C. Y., and Zhou, N. (2021). A review of regional distributed energy system planning and design. *Int. J. Embed. Syst.* 14, 89–100. doi:10.1504/ijes.2021.10033849

Howarth, R. W. (2014). A bridge to nowhere: Methane emissions and the greenhouse gas footprint of natural gas. *Energy Sci. Eng.* 2, 47–60. doi:10.1002/ese3.35

Jin, F., Zhao, J., Sheng, C. Y., and Wang, W. (2018). Causality diagram-based scheduling approach for blast furnace gas system. *Ieee-Caa J. Automatica Sinica* 5, 587–594. doi:10.1109/jas.2017.7510715

Kang, J., and Reiner, D. M. (2022). Off seasons, holidays and extreme weather events: Using data-mining techniques on smart meter and energy consumption data from China. *Energy Res. Soc. Sci.* 89, 102637. doi:10.1016/j.erss.2022.102637

Li, L. (2022). Development of natural gas industry in China: Review and prospect. *Nat. Gas. Ind. B* 9, 187–196. doi:10.1016/j.ngib.2022.03.001

Nikolaidis, Y., Pilavachi, P. A., and Chletsis, A. (2009). Economic evaluation of energy saving measures in a common type of Greek building. *Appl. Energy* 86, 2550–2559. doi:10.1016/j.apenergy.2009.04.029

Onur, D. (2021). Natural gas consumption behavior of companies by clustering analysis. *Eng. Appl. Artif. Intell.* 106, 104502. doi:10.1016/j.engappai.2021.104502

Palinski, A. (2018). Hurtownie danych i eksploracja danych w prognozowaniu popytu na gaz i usługi magazynowania gazu. *Nafta-Gaz* 74, 283–289. doi:10.18668/ng.2018.04.04

Piao, M., Piao, Y. J., and Lee, J. Y. (2019). Symmetrical uncertainty-based feature subset generation and ensemble learning for electricity customer classification. *Symmetry-Basel* 11, 498. doi:10.3390/sym11040498

Rahim, M. S., Nguyen, K., Stewart, R., Giurco, D., and Blumenstein, M. (2019). "Predicting household water consumption events: Towards a personalised recommender system to encourage water-conscious behaviour," in 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14-19 July 2019.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. doi:10.1016/0377-0427(87)90125-7

Shan, X., Deng, Q., Tang, Z., Wu, Z., and Wang, W. (2022). An integrated data mining-based approach to identify key building and urban features of different energy usage levels. *Sustain. Cities Soc.* 77, 103576. doi:10.1016/j.scs.2021.103576

Stephen, B., Mutanen, A., Galloway, S., Burt, G. M., and Järventausta, P. (2014). Enhanced load profiling for residential network customers. *IEEE Trans.*29, 88–96. doi:10.1109/tpwrd.2013.2287032

Stuivenvolt-Allen, J., and Wang, S. S. Y. (2019). Data mining climate variability as an indicator of US natural gas. *Front. Big Data* 2, 00020. doi:10.3389/fdata.2019.00020

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B-Statistical Methodol.* 63, 411–423. doi:10.1111/1467-9868.00293

Vieira, P., Jorge, C., and Covas, D. (2018). Efficiency assessment of household water use. *Urban Water J.* 15, 407–417. doi:10.1080/1573062x.2018.1508596

Wang, Q., and Yang, X. (2021). Evaluating the potential for sustainable development of China's shale gas industry by combining multi-level DPSIR framework, PPFCI technique and Raga algorithm. *Sci. Total Environ.* 780, 146525. doi:10.1016/j.scitotenv.2021.146525

Yang, J., Li, Y., Zhang, N., Yang, J., Kuang, K., Hu, Y., et al. (2015). "Analysis of urban residential water consumption based on smart meters and fuzzy clustering," in 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, UK, 26-28 October 2015.

Zhou, Q., Li, J., Lv, Q. Q., Wang, D. M., and Yang, S. Y. (2022). Research on China's new energy cross-provincial marketization mechanism under the background of "double carbon. *Front. Energy Res.* 10, 846870. doi:10.3389/fenrg.2022.846870

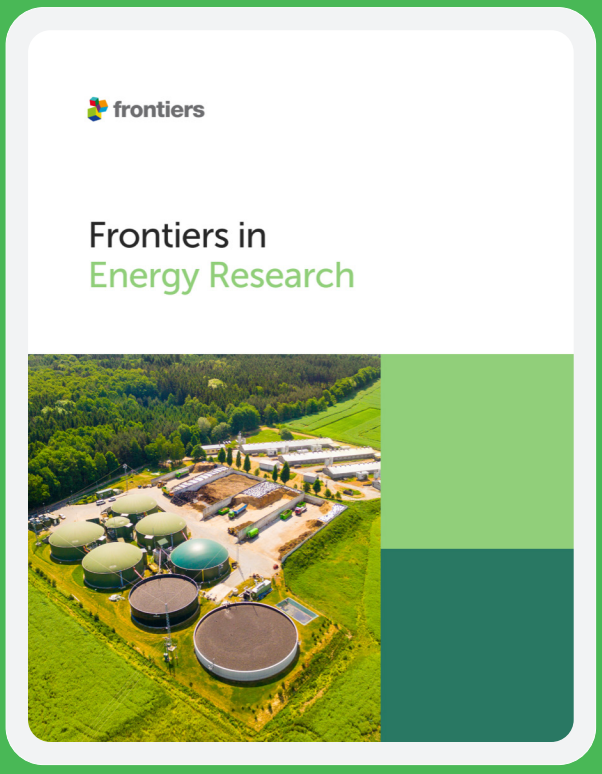
Frontiers in
Energy Research

Advances and innovation in sustainable, reliable and affordable energy

Explores sustainable and environmental developments in energy. It focuses on technological advances supporting Sustainable Development Goal 7: access to affordable, reliable, sustainable and modern energy for all.

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