



OPEN ACCESS

EDITED BY

Zohar Barnett-Itzhaki,
Ruppin Academic Center, Israel

REVIEWED BY

Jiaxing Xin,
Liaoning Normal University, China
Yanqu Cai,
Guangdong Pharmaceutical University, China

*CORRESPONDENCE

Rong-Jie Wei
✉ www08www@163.com
Dian-Ju Kang
✉ kangdianjuSCCDC@163.com

RECEIVED 24 March 2025

ACCEPTED 01 July 2025

PUBLISHED 16 July 2025

CITATION

Zhang Y, Wang Q-L, Peng W, Zhang M-Y,
Qin Y, Zhang L, Wei R-J and Kang D-J (2025)
Interpretable machine learning analysis of
environmental characteristics on bacillary
dysentery in Sichuan Province.
Front. Public Health 13:1598247.
doi: 10.3389/fpubh.2025.1598247

COPYRIGHT

© 2025 Zhang, Wang, Peng, Zhang, Qin,
Zhang, Wei and Kang. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Interpretable machine learning analysis of environmental characteristics on bacillary dysentery in Sichuan Province

Yao Zhang¹, Qiao-Lin Wang², Wei Peng³, Meng-Yuan Zhang¹,
Yao Qin¹, Lun Zhang¹, Rong-Jie Wei^{1*} and Dian-Ju Kang^{1*}

¹Department of Emergency Management, Sichuan Center for Diseases Control and Prevention, Chengdu, China, ²West China School of Public Health/West China Fourth Hospital, Sichuan University, Chengdu, China, ³Department of Health Education Institute, Sichuan Center for Diseases Control and Prevention, Chengdu, China

Background: Bacterial dysentery (BD) is a leading cause of diarrhea-related mortality globally, with its incidence heavily influenced by environmental factors. However, a climate zone-specific predictive model for BD was currently lacking in Sichuan Province.

Objective: This study aims to employ interpretable machine learning to explore the influence of environmental factors on BD incidence across different climate zones and to elucidate their interaction mechanisms.

Methods: Monthly data on meteorological and ecological factors, along with BD case reports, were collected from 183 counties in Sichuan Province (2005–2023). The eXtreme Gradient Boosting (XGBoost) algorithm was employed to assess the influence of key environmental features, including precipitation, temperature, PM10, potential evaporation, vegetation cover, and NDVI, on BD incidence. To enhance interpretability, the model's outputs were visualized and explained using SHapley Additive Explanations (SHAP).

Results: A machine learning model was developed to assess the impact of environmental factors on BD incidence across different climate zones. The findings revealed significant spatial heterogeneity in key drivers of BD. In the Central Subtropical Humid Climate Zone, BD incidence was predominantly influenced by average temperature, PM10, and minimum temperature. In the Subtropical Semi-Humid Climate Zone, potential evaporation, PM10, and precipitation emerged as the primary determinants. In the Plateau Cold Climate Zone, PM10, minimum temperature, and precipitation were the most significant factors. Notably, PM10 consistently showed a positive correlation with BD across all climate zones. Furthermore, average temperature showed a positive association with BD in the Central Subtropical Humid Climate Zone, while potential evaporation and minimum temperature demonstrated similar positive relationships in the Subtropical Semi-Humid and Plateau Cold Climate Zones, respectively. Additionally, precipitation displayed a U-shaped relationship with BD risk in both the Subtropical Semi-Humid and Plateau Cold Climate Zones.

Conclusion: This study developed a climate zone-specific predictive model for BD, systematically evaluating the interactions between environmental factors and BD dynamics. The findings provide a scientific basis for refining targeted public health intervention strategies.

KEYWORDS

bacterial dysentery, climate zones, environmental characteristics, XGBoost, SHAP

1 Introduction

Bacterial dysentery (BD), caused by *Shigella*, is an intestinal infectious disease transmitted through contaminated food, water, and person-to-person contact (1). It poses a significant public health challenge globally, particularly in developing countries (2). Although the incidence of BD has been effectively reduced in many regions worldwide over the past few decades through improved sanitation and public health interventions, it remains one of the leading causes of diarrheal mortality globally (3). According to statistics, BD caused 210,000 deaths in 2016, with over 90% of cases occurring in developing countries, particularly among children under 5 years old and adults over 70 years old (4, 5).

BD exhibits distinct seasonal and geographical patterns. The peak periods of BD vary significantly across regions. For instance, in Bangladesh, the peak typically occurs between September and November (6); in Vietnam, it is between May and October (7); and in Sweden, between July and October (8). Studies have shown that temperature and precipitation directly influence pathogen survival and transmission (9, 10), highlighting the role of meteorological factors. In China, due to regional differences in meteorological conditions, the transmission patterns and peak periods of BD also vary (11, 12). Northern regions typically experience peaks in early summer, while southern regions experience peaks in summer and autumn (13, 14).

Recent extreme weather events, such as floods and El Niño, have exacerbated BD outbreak risks (15, 16). These events often lead to abnormal temperature increases, which are associated with BD epidemics (17, 18). Notably, in addition to meteorological conditions, ecological factors also play a significant role in BD incidence. For example, increased forest cover can help prevent BD by improving water quality (19).

Although many studies have examined the relationship between meteorological factors and BD, most relied on single-factor analyses and short-term, large-scale data (20–23), limiting understanding of local transmission patterns and characteristics. The impact of local meteorological and ecological factors on disease dynamics may differ significantly from findings based on large-scale studies (24, 25). Furthermore, short-term data cannot reflect long-term trends. Therefore, developing models that integrate long-term, fine-scale meteorological and ecological data is crucial for predicting and analyzing BD activity.

Traditional statistical methods are limited in capturing the complex nonlinear relationships between meteorological, ecological, and disease incidence data (26). Therefore, adopting more advanced machine learning methods can more effectively capture these complex relationships, providing more accurate predictions and assessments for BD prevention and control. As an efficient machine learning method, eXtreme Gradient Boosting (XGBoost) has demonstrated significant potential in various fields in recent years, particularly in handling large-scale data and modeling nonlinear relationships among multidimensional variables, offering greater flexibility and accuracy compared to traditional regression models (9, 27). It can efficiently process high-dimensional, nonlinear, and

heterogeneous complex data and, through the integration of multiple decision trees, exhibits strong predictive capabilities, accurately capturing the relationships between diseases and multiple influencing factors (28, 29). This provides a scientific basis for the formulation of health intervention strategies (30, 31). Therefore, this study employs the XGBoost model to assess the impact of meteorological and ecological factors on BD activity across different climate zones at the county and monthly scales, providing data for targeted public health interventions.

2 Materials and methods

2.1 Study area

Sichuan Province, located in southwestern China, encompasses 183 counties with diverse climatic conditions. It is divided into three distinct climate zones based on variations in temperature, precipitation, and sunlight (32): (a) Central Subtropical Humid Climate Zone (Zone 1): Covering 128 counties, this zone is characterized by warm, humid conditions year-round. Average annual temperatures range from 16°C to 18°C, with mild winters and hot summers. Rainfall is abundant, averaging 1,000–1,200 mm annually. Over half of the precipitation occurs during the summer months. (b) Subtropical Semi-Humid Climate Zone (Zone 2): This zone includes 23 counties and features relatively high temperatures throughout the year, averaging 12°C to 20°C. The region experiences a pronounced dry season lasting 7 months, with annual precipitation of 900–1,200 mm, 90% of which falls between May and October. (c) Plateau Cold Climate Zone (Zone 3): Comprising 32 counties, this zone is marked by significant elevation changes and a cold temperate climate. Average annual temperatures range from 4°C to 12°C, with cool summers and cold winters. Annual precipitation is lower, ranging from 500 to 900 mm, but the region benefits from ample sunshine (Figure 1).

2.2 Data collection

In China, BD is classified as a Category B notifiable infectious disease, requiring reporting to the local Center for Disease Control and Prevention within 24 h of diagnosis. The BD case data in this study were obtained from the National Notifiable Diseases Reporting System. Case definitions adhered to the standardized criteria established by the National Health and Family Planning Commission of the People's Republic of China.¹ Both clinically diagnosed and laboratory-confirmed cases were included in the analysis. Meteorological and ecological data were sourced from the National Earth System Science Data Center, National Science and Technology Infrastructure of China² and the National Tibetan Plateau/Third Pole Environment Data

1 <http://www.nhc.gov.cn/zwgkzt/s9491/200802/39040.shtml>

2 <http://www.geodata.cn>

3 Results

3.1 Descriptive analysis

Between 2005 and 2023, Sichuan Province exhibited considerable variability in BD case counts, ultimately demonstrating a general decrease. The yearly incidence rate decreased from 34.72 to 3.25 per 100,000 individuals, exhibiting significant seasonal variation: 67.25% of incidences transpired between May and October, with a peak in June. The peak incidence periods varied across different climate zones: the Central Subtropical Humid Climate Zone peaked from August to September, the Subtropical Semi-Humid Climate Zone peaked from May to June, and the Plateau Cold Climate Zone peaked from July to August (Figure 2). Cases were documented in all 183 counties, albeit the distribution exhibited geographical variation. The Central Subtropical Humid Climate Zone included 55.86% of cases (mean: 938 cases per county), succeeded by the Subtropical Semi-Humid Climate Zone (33.11%; 3,094 cases per county) and the Plateau Cold Climate Zone (11.03%; 741 cases per county) (Supplementary Figure 1). We performed statistical calculations on meteorological and ecological data from 183 counties. Key metrics, including the mean, median, and standard deviation, were obtained for eight factors: precipitation, average temperature, minimum temperature, maximum temperature, PM10, potential evaporation, NDVI, and vegetation cover (250 m). Detailed results are provided in Supplementary Table 2.

3.2 Performance of machine learning models

Using monthly data from 2005 to 2023, we conducted XGBoost model analyses for the entire province as well as for each individual climate zone. The performance metrics of the models on the test set were presented in Table 1. The MAE values for the overall

(province-wide), Zone 1, Zone 2, and Zone 3 were 4.40, 3.21, 10.87, and 2.77, respectively, while the RMSE values were 9.96, 6.05, 20.21, and 5.30, reflecting the average and overall error levels of the models. The R^2 values, representing the models' explanatory power, were 0.76, 0.89, 0.77, and 0.81, respectively. In summary, the models demonstrated strong predictive performance.

3.3 Feature analysis

Through SHAP analysis, this study systematically revealed the importance ranking of the influencing factors included in the model (see Supplementary Table 3 for details) and effectively distinguished the positive and negative correlations between these factors and BD incidence (Figure 3). Province-wide, the factor importance ranking was as follows: potential evaporation, maximum temperature, PM10, vegetation cover (250 m), minimum temperature, NDVI, average temperature, and precipitation. Among these, potential evaporation, maximum temperature, and PM10 showed positive associations with BD incidence, while vegetation cover (250 m) and NDVI exhibited negative associations (Figure 4).

Based on the SHAP analysis of the climate zone-specific models, the results indicated the following: In the Central Subtropical Humid Climate Zone, the factor importance ranking was: average temperature, PM10, minimum temperature, maximum temperature, potential evaporation, vegetation cover (250 m), NDVI, and precipitation. Among these, average temperature, PM10, and maximum temperature were positively correlated with BD incidence, while vegetation cover (250 m) and NDVI were negatively correlated (Figure 5). In the Subtropical Semi-Humid Climate Zone, the factor importance ranking was: potential evaporation, PM10, precipitation, minimum temperature, average temperature, vegetation cover (250 m), NDVI, and maximum temperature. PM10 and potential evaporation showed positive associations, while vegetation cover (250 m) and NDVI remained negatively correlated (Figure 6). In the Plateau Cold Climate Zone, the factor importance ranking was: PM10,

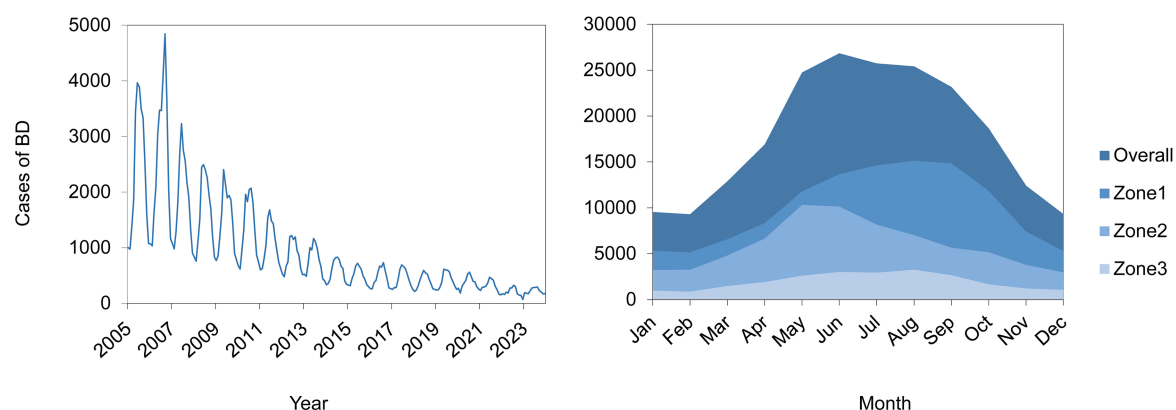
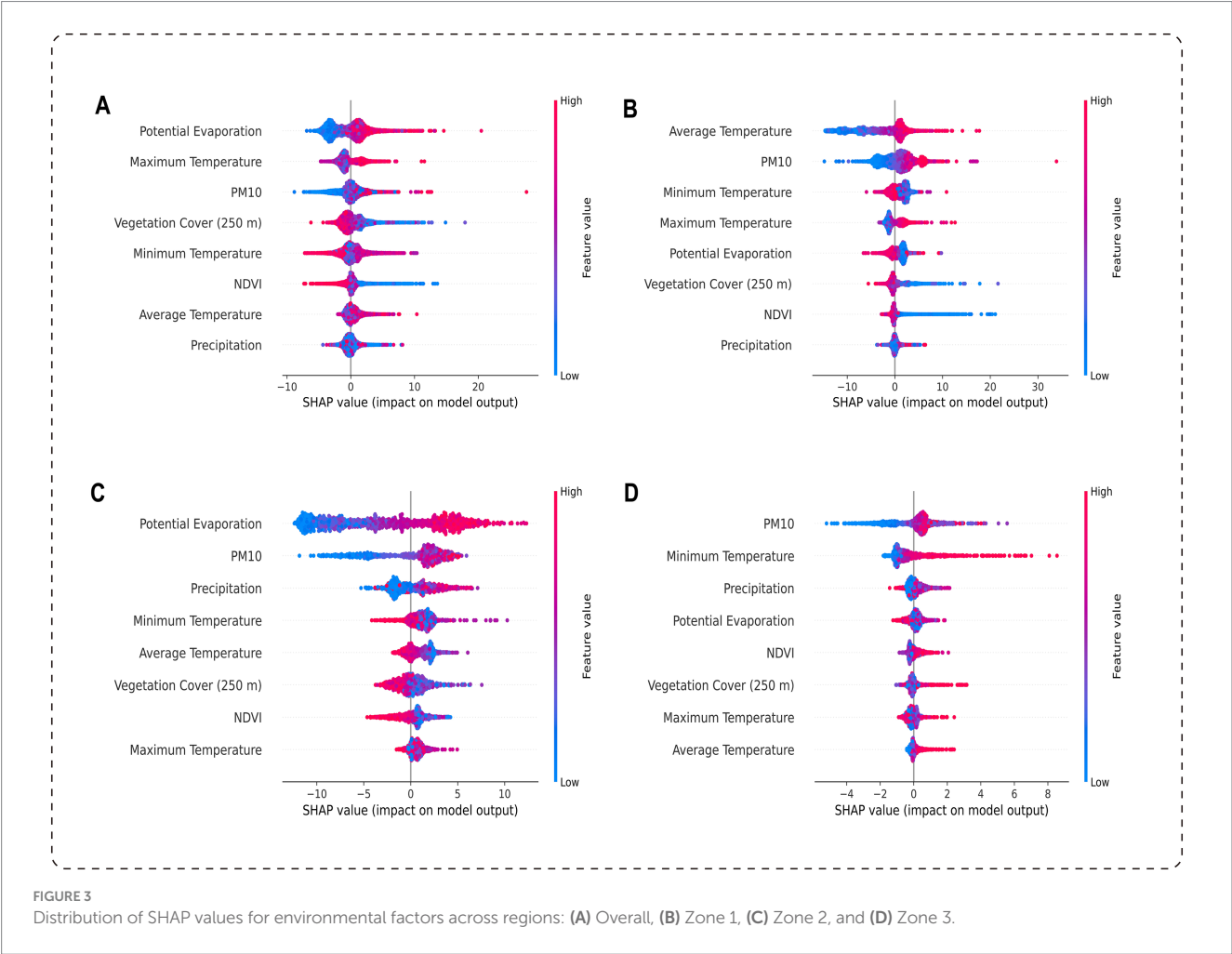


FIGURE 2

Time series distribution of bacillary dysentery (BD) cases and monthly case distribution by region: Overall (province-wide), Zone 1 (Central Subtropical Humid Climate Zone), Zone 2 (Subtropical Semi-Humid Climate Zone), and Zone 3 (Plateau Cold Climate Zone).

TABLE 1 Performance metrics of the XGBoost model across climate zones.

Metrics	Overall	Zone 1	Zone 2	Zone 3
MAE	4.40	3.21	10.87	2.77
RMSE	9.96	6.05	20.21	5.30
R2	0.76	0.89	0.77	0.81



minimum temperature, precipitation, potential evaporation, NDVI, vegetation cover (250 m), maximum temperature, and average temperature. Minimum temperature, average temperature, and PM10 were positively correlated with BD incidence, but no significant negative associations were detected (Figure 7). The lagged effect analysis (1–3 months) for the top two most important variables revealed minimal impacts of different lag periods on the model's MAE values (see Supplementary Table 4 for details). Furthermore, to further quantify the relationships between influencing factors and BD incidence, this study analyzed the threshold effects of key environmental parameters using SHAP dependence curves. The results showed that the interaction patterns of these factors across climate zones were highly consistent with the findings above. Detailed threshold values will be thoroughly discussed in the subsequent analysis section.

4 Discussion

Based on long-term, fine-scale research data, this study systematically revealed the spatiotemporal heterogeneity of BD incidence and its environmental driving mechanisms across different climate zones in Sichuan Province. Analysis using the XGBoost machine learning model demonstrated that environmental factors significantly influenced BD transmission in a climate zone-specific manner, highlighting the importance of fine-resolution climate zoning in assessing BD incidence risk.

To elucidate the contribution of each environmental factor, this study employed the game theory-based SHAP analysis method. The core principle of this method involves calculating Shapley values to quantify the contribution of each feature to the prediction outcome, thereby enhancing model transparency and interpretability. This

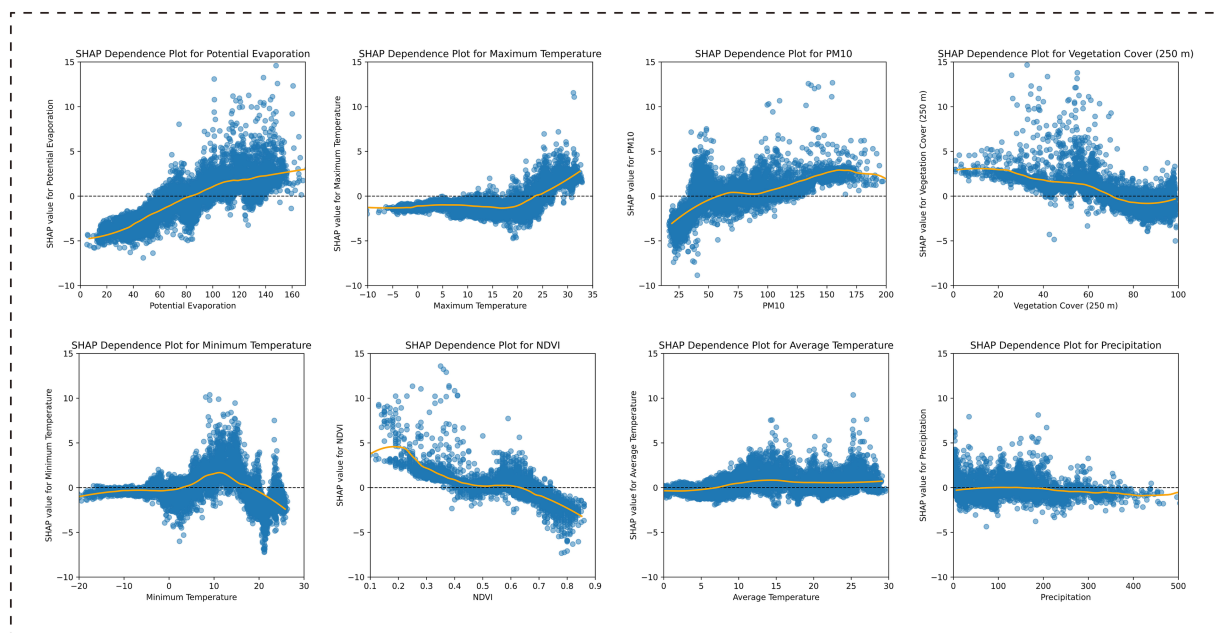


FIGURE 4
SHAP dependence plots for environmental features in the XGBoost model: Province-wide analysis.

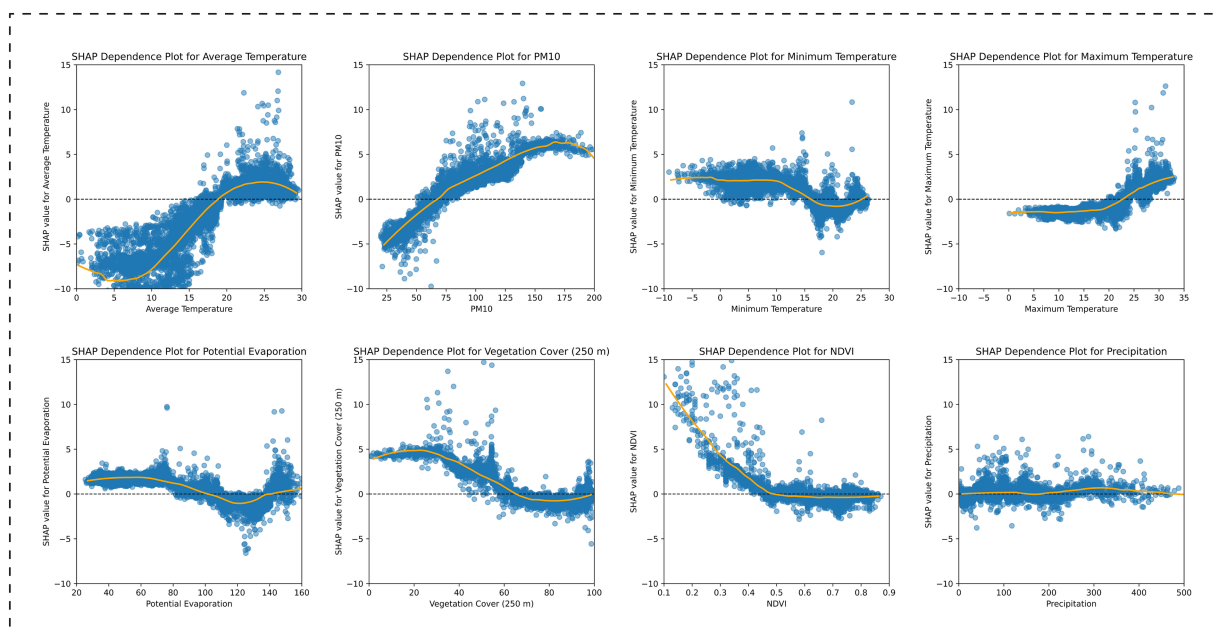


FIGURE 5
SHAP dependence plots for environmental features in the XGBoost model: Central Subtropical Humid Climate Zone analysis.

approach assists researchers and policymakers in gaining a deeper understanding of how various factors influence prediction results, providing valuable insights for BD monitoring and prevention.

The association between ambient temperature and BD incidence exhibited significant spatial heterogeneity. In the Central Subtropical Humid Climate Zone, BD incidence risk significantly increased when the average temperature exceeded a threshold of 18°C, which aligns with the growth temperature of *Shigella* (6–8°C to 45–47°C, optimal

around 37°C) (36), suggesting that warm environments facilitate bacterial proliferation and transmission both within and outside hosts. In the Plateau Cold Climate Zone, risk increased when the minimum temperature exceeded 2°C, a phenomenon that may not directly stem from pathogen biological characteristics but rather be associated with increased outdoor activities among local residents at this temperature threshold (37). Rising temperatures prompted local residents to engage in more outdoor activities, thereby increasing the risk of

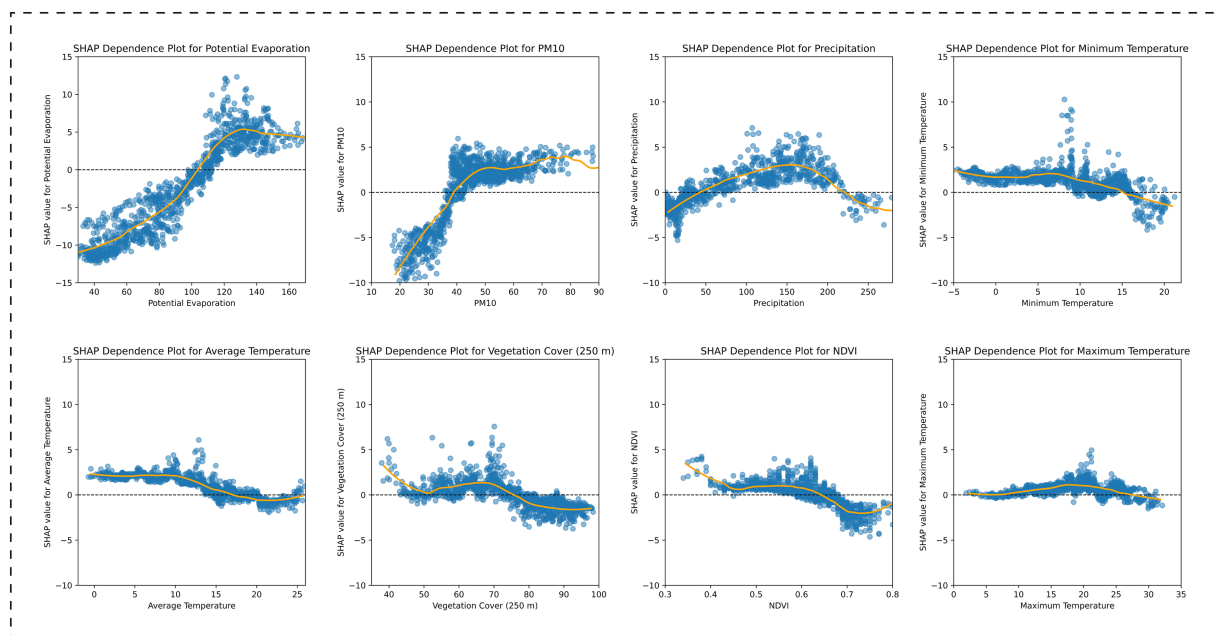


FIGURE 6
SHAP dependence plots for environmental features in the XGBoost model: Subtropical Semi-Humid Climate Zone analysis.

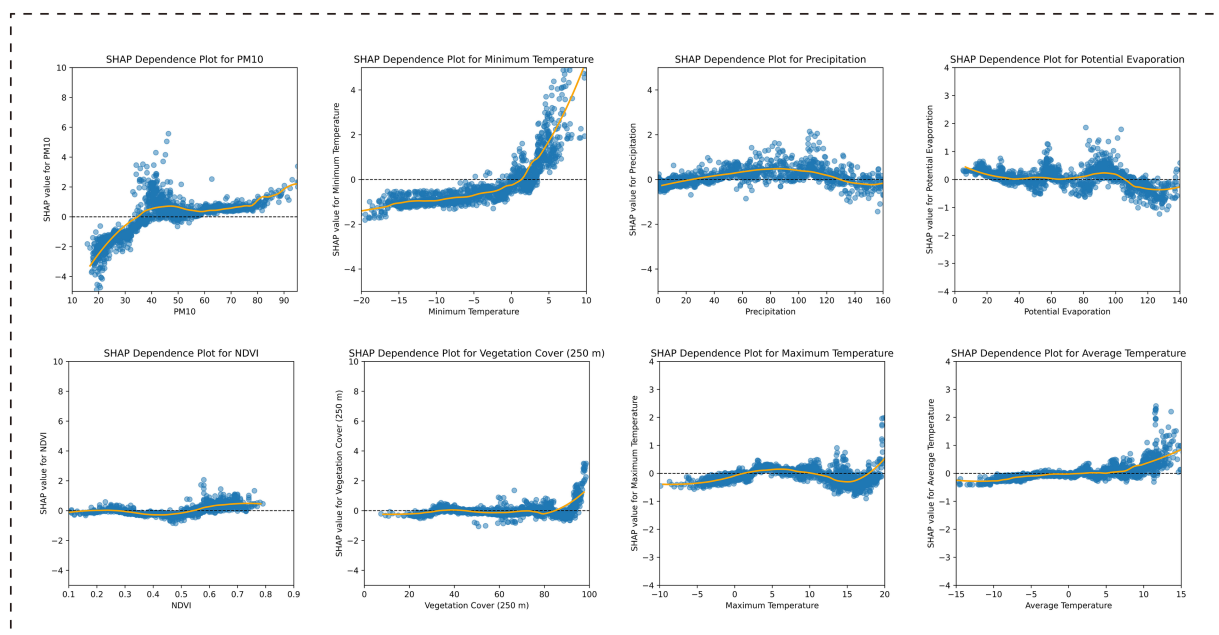


FIGURE 7
SHAP dependence plots for environmental features in the XGBoost model: Plateau Cold Climate Zone analysis.

pathogen exposure. Conversely, in the Subtropical Semi-Humid Climate Zone, the predictive importance of temperature factors significantly decreased due to consistently high annual temperatures, suggesting that temperature is no longer a primary limiting factor for BD transmission in this region, and disease spread may be more regulated by other environmental factors, such as water availability or precipitation patterns.

The impact of precipitation on BD incidence also exhibited significant climate zone dependence. In the Central Subtropical Humid Climate Zone, well-developed sanitation infrastructure effectively blocked precipitation-related waterborne transmission routes, resulting in no clear correlation between precipitation and BD incidence. This finding highlights that robust water, sanitation, and hygiene (WASH) infrastructure can effectively mitigate the

influence of climatic factors (e.g., heavy rainfall) on waterborne diseases, decoupling environmental triggers from disease outcomes. In contrast, in the Subtropical Semi-Humid Climate Zone, a sharp increase in precipitation after the dry season (monthly precipitation > 50 mm in May) coincided significantly with the peak BD incidence period. This pattern is consistent with the outbreaks of diarrheal diseases triggered by heavy rainfall after droughts observed in other global regions [e.g., Ecuador (38), Eswatini (39), the United Kingdom (40), Japan (41), and Vietnam (7)]. This may be related to the burst of microbial activity caused by rewetting dry soil (42), which, during heavy rainfall, facilitates the flushing of pathogens into water bodies, subsequently triggering disease outbreaks. This phenomenon suggests that in regions prone to alternating dry and heavy rainfall periods, public health strategies should not only focus on immediate flood response but also manage the environmental consequences of drought. In the Plateau Cold Climate Zone, BD incidence risk significantly increased when monthly precipitation exceeded 30 mm. This threshold may reflect a critical point of the region's sanitation infrastructure capacity. Heavy rainfall can overload wastewater treatment plants, leading to the overflow of untreated or partially treated sewage into the environment, or overwhelm septic tank systems, increasing the risk of fecal contaminant entry into water sources. This suggests that even moderate precipitation can pose a public health threat in regions with relatively weak infrastructure.

Potential evapotranspiration exhibited the highest predictive importance in the Subtropical Semi-Humid Climate Zone. As a comprehensive indicator reflecting meteorological factors such as temperature, humidity, wind speed, and solar radiation, an increase in potential evaporation was associated with the synergistic effects of high temperature, low humidity, and strong solar radiation (43). High potential evaporation values indicate increased atmospheric demand for moisture, which, when precipitation is insufficient, exacerbates regional water stress. Under these circumstances, residents may be forced to rely on suboptimal or unsafe water sources, thereby increasing their risk of exposure to pathogenic microorganisms. As a comprehensive meteorological indicator, potential evaporation can serve as a valuable early warning indicator for hydrological stress and impending water scarcity, supporting early warning systems for waterborne disease risks.

Increased PM10 concentration were positively correlated with BD incidence. Based on existing research, this study proposes that PM10 may influence BD incidence through the following mechanisms: First, particulate matter can serve as a carrier for pathogenic microorganisms (44), facilitating their direct transmission through aerosol deposition or water contamination. Second, PM10 can alter the local microenvironment, affecting pathogen survival and transmission efficiency. Chemical pollutants within particulate matter may inhibit microbial growth at high concentrations but could provide a suitable microenvironment and nutrients at moderate concentrations (45). Furthermore, long-term exposure to PM10 may impair the barrier function of the intestinal mucosa, thereby increasing host susceptibility to pathogens. PM10 exposure can lead to alterations in the gut microbiota, reducing the abundance of beneficial microorganisms and promoting the overgrowth of pro-inflammatory species, which in turn contributes to intestinal barrier dysfunction, oxidative stress, and inflammatory responses, all associated with the development and progression of gastrointestinal inflammatory diseases (46, 47).

Ecological factors, including vegetation cover (250 m) and NDVI, were negatively correlated with BD incidence in both the Central Subtropical and Subtropical Climate Zones. These results have been corroborated in previous studies (48, 49). This study hypothesizes that the mechanisms by which green spaces reduce BD incidence may include: Vegetation, particularly riparian vegetation, reduces runoff of sediments, fertilizers, and pesticides from agricultural fields through physical buffering, and helps capture and cycle nutrients, preventing their excessive entry into water bodies that could lead to eutrophication (50). Furthermore, vegetation can lower ambient temperatures by providing shade and enhancing evapotranspiration, thereby mitigating the risk of BD transmission associated with rising temperatures (51, 52).

Furthermore, this study also found that BD case numbers in Sichuan Province exhibited significant seasonal variations, with differing incidence peaks across climate zones, suggesting region-specific seasonal driving mechanisms. In the Central Subtropical Humid Climate Zone, the peak BD incidence occurred from August to September. This correlated with the average temperature in this region exceeding the 18°C threshold. Despite the well-developed sanitation infrastructure in this region, which effectively blocked direct waterborne transmission routes related to precipitation, high summer temperatures may promote pathogen transmission by accelerating bacterial growth and increasing the frequency of outdoor activities and water body contact. In the Subtropical Semi-Humid Climate Zone, the peak BD incidence occurred from May to June. This peak period significantly coincided with a sharp increase in precipitation after the dry season (monthly precipitation > 50 mm in May), indicating that this 'dry-wet transition' effect was a key driving factor for seasonal BD outbreaks in this region (42). Additionally, this region may experience water stress, compelling residents to rely on suboptimal water sources during the dry season, while runoff pollution from initial rainfall in the wet season exacerbates the incidence risk. In the Plateau Cold Climate Zone, the peak BD incidence occurred from July to August. This correlated with the minimum temperature in this region exceeding the 2°C threshold, where rising temperatures may encourage local residents to increase outdoor activities. Furthermore, increased monthly precipitation may also contribute to regulating seasonal BD incidence in this region. These climate zone-specific seasonal patterns further emphasize that BD epidemiology is not determined by a single factor, but rather results from complex interactions among environmental conditions, infrastructure, and human behavior. Understanding these refined seasonal driving mechanisms is crucial for developing more targeted and timely public health interventions.

Despite utilizing the XGBoost model to elucidate the influence of meteorological and ecological factors on BD incidence, this study has several limitations, specifically: First, no statistically significant lagged effects were identified in this study; however, this does not imply the absence of lagged impacts of environmental factors on BD incidence. The inconspicuous lagged effects may be attributed to inherent limitations of the current dataset and modeling methodology. Second, the model relies on reported data, which may be subject to underreporting issues. Third, BD transmission is also influenced by socioeconomic factors, sanitation infrastructure, and public health awareness (53, 54), which were not fully considered in this study. Finally, while SHAP analysis effectively identified key

predictive factors, the method itself only captures associations between features and predicted outputs, without establishing causal relationships. Therefore, the results of this study should be interpreted with caution.

In conclusion, through an in-depth analysis of the spatiotemporal heterogeneity of BD and its environmental driving mechanisms in Sichuan Province, this study revealed the influence of environmental factors on BD transmission. These findings provide a scientific basis for developing climate zone-specific BD monitoring, prevention, and intervention strategies. In the future, a more comprehensive BD surveillance and reporting system should be established, integrating socioeconomic factors to more comprehensively assess the potential influencing factors of BD incidence, thereby achieving more precise and effective disease control and prevention.

Data availability statement

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

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

YZ: Data curation, Validation, Investigation, Supervision, Methodology, Project administration, Visualization, Conceptualization, Funding acquisition, Software, Writing – review & editing, Resources, Formal analysis, Writing – original draft. Q-LW: Software, Methodology, Writing – original draft, Investigation, Formal analysis, Data curation. WP: Writing – original draft, Investigation, Software. M-YZ: Writing – original draft, Data curation, Conceptualization. YQ: Conceptualization, Writing – original draft, Investigation. LZ: Methodology, Writing – original draft, Investigation. R-JW: Supervision, Software, Writing – review & editing. D-JK: Conceptualization, Supervision, Writing – review & editing.

References

- Kotloff KL, Riddle MS, Platts-Mills JA, Pavlinac P, Zaidi A. Shigellosis. *Lancet*. (2018) 391:801–12. doi: 10.1016/S0140-6736(17)33296-8
- Kotloff KL. The burden and etiology of diarrheal illness in developing countries. *Pediatr Clin N Am*. (2017) 64:799–814. doi: 10.1016/j.pcl.2017.03.006
- Hosangadi D, Smith PG, Kaslow DC, Giersing BK. Who consultation on etec and shigella burden of disease, Geneva, 6–7th april 2017: meeting report. *Vaccine*. (2019) 37:7381–90. doi: 10.1016/j.vaccine.2017.10.011
- Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, et al. Morbidity and mortality due to shigella and enterotoxigenic escherichia coli diarrhoea: the global burden of disease study 1990–2016. *Lancet Infect Dis*. (2018) 18:1229–40. doi: 10.1016/S1473-3099(18)30475-4
- Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A, Reiner RC, et al. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the global burden of disease study 2015. *Lancet Infect Dis*. (2017) 17:909–48. doi: 10.1016/S1473-3099(17)30276-1
- Cash BA, Rodó X, Emch M, Yunus M, Faruque AS, Pascual M. Cholera and shigellosis: different epidemiology but similar responses to climate variability. *PLoS One*. (2014) 9:e107223. doi: 10.1371/journal.pone.0107223
- Lee HS, Ha HT, Pham-Duc P, Lee M, Grace D, Phung DC, et al. Seasonal and geographical distribution of bacillary dysentery (shigellosis) and associated climate risk factors in Kon tam province in Vietnam from 1999 to 2013. *Infect Dis Poverty*. (2017) 6:113. doi: 10.1186/s40249-017-0325-z

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the National Institutes of Health (award no. 5RO1AI125842-02), Sichuan Science and Technology Program (2022YFS0052), and Chongqing Science and Technology Program (cstc2020jscxylhX0003).

Acknowledgments

The authors want to thank the family for their support.

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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/fpubh.2025.1598247/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Temporal distribution of BD cases across counties in Sichuan Province.

8. Ekdahl K, Andersson Y. The epidemiology of travel-associated shigellosis--regional risks, seasonality and serogroups. *J Infect.* (2005) 51:222–9. doi: 10.1016/j.jinf.2005.02.002
9. Liang D, Wang L, Liu S, Li S, Zhou X, Xiao Y, et al. Global incidence of diarrheal diseases--an update using an interpretable predictive model based on xgboost and shap: a systematic analysis. *Nutrients.* (2024) 16:217. doi: 10.3390/nu16183217
10. Wu X, Liu J, Li C, Yin J. Impact of climate change on dysentery: scientific evidences, uncertainty, modeling and projections. *Sci Total Environ.* (2020) 714:136702. doi: 10.1016/j.scitotenv.2020.136702
11. Zhao Z, Yang M, Lv J, Hu Q, Chen Q, Lei Z, et al. Shigellosis seasonality and transmission characteristics in different areas of China: a modelling study. *Infect Dis Model.* (2022) 7:161–78. doi: 10.1016/j.idm.2022.05.003
12. Liu Z, Tong MX, Xiang J, Dear K, Wang C, Ma W, et al. Daily temperature and bacillary dysentery: estimated effects, attributable risks, and future disease burden in 316 chinese cities. *Environ Health Perspect.* (2020) 128:57008. doi: 10.1289/EHP5779
13. Chang Q, Wang K, Zhang H, Li C, Wang Y, Jing H, et al. Effects of daily mean temperature and other meteorological variables on bacillary dysentery in Beijing-Tianjin-Hebei region, China. *Environ Health Prev Med.* (2022) 27:13. doi: 10.1265/ehpm.21-00005
14. Li C, Wang X, Liu Z, Cheng L, Huang C, Wang J. El niño southern oscillation, weather patterns, and bacillary dysentery in the yangtze river basin, China. *Glob Health Res Policy.* (2024) 9:45. doi: 10.1186/s41256-024-00389-4
15. Li X, Zhang K, Gu P, Feng H, Yin Y, Chen W, et al. Changes in precipitation extremes in the yangtze river basin during 1960–2019 and the association with global warming, enso, and local effects. *Sci Total Environ.* (2021) 760:144244. doi: 10.1016/j.scitotenv.2020.144244
16. Chen NT, Chen YC, Wu CD, Chen MJ, Guo YL. The impact of heavy precipitation and its impact modifiers on shigellosis occurrence during typhoon season in Taiwan: a case-crossover design. *Sci Total Environ.* (2022) 848:157520. doi: 10.1016/j.scitotenv.2022.157520
17. Liu Z, Ding G, Zhang Y, Lao J, Liu Y, Zhang J, et al. Identifying different types of flood-sensitive diarrheal diseases from 2006 to 2010 in Guangxi, China. *Environ Res.* (2019) 170:359–65. doi: 10.1016/j.envres.2018.12.067
18. Zhang X, Wang Y, Zhang W, Wang B, Zhao Z, Ma N, et al. The effect of temperature on infectious diarrhea disease: a systematic review. *Heliyon.* (2024) 10:e31250. doi: 10.1016/j.heliyon.2024.e31250
19. Rasolofson RA, Ricketts TH, Johnson KB, Jacob A, Fisher B. Forests moderate the effectiveness of water treatment at reducing childhood diarrhea. *Environ Res Lett.* (2021) 16:64035. doi: 10.1088/1748-9326/abff88
20. Li T, Yang Z, Wang M. Temperature and atmospheric pressure may be considered as predictors for the occurrence of bacillary dysentery in Guangzhou, southern China. *Rev Soc Bras Med Trop.* (2014) 47:382–4. doi: 10.1590/0037-8682-0144-2013
21. Zhang H, Si Y, Wang X, Gong P. Patterns of bacillary dysentery in China, 2005–2010. *Int J Environ Res Public Health.* (2016) 13:164. doi: 10.3390/ijerph13020164
22. Liu J, Wu X, Li C, Xu B, Hu L, Chen J, et al. Identification of weather variables sensitive to dysentery in disease-affected county of China. *Sci Total Environ.* (2017) 575:956–62. doi: 10.1016/j.scitotenv.2016.09.153
23. Li R, Liu D, Wang T, Li D, Shi T, Zhao X, et al. Lagged effects of climate factors on bacillary dysentery in western China. *Trans R Soc Trop Med Hyg.* (2025) 119:33–41. doi: 10.1093/trstmh/trae064
24. Balasubramani K, Prasad KA, Kodali NK, Abdul RN, Chellappan S, Sarma DK, et al. Spatial epidemiology of acute respiratory infections in children under 5 years and associated risk factors in India: district-level analysis of health, household, and environmental datasets. *Front Public Health.* (2022) 10:906248. doi: 10.3389/fpubh.2022.906248
25. Jagai JS, Sarkar R, Castronovo D, Kattula D, McEntee J, Ward H, et al. Seasonality of rotavirus in south asia: a meta-analysis approach assessing associations with temperature, precipitation, and vegetation index. *PLoS One.* (2012) 7:e38168. doi: 10.1371/journal.pone.0038168
26. Keshavamurthy R, Dixon S, Pazdernik KT, Charles LE. Predicting infectious disease for biopreparedness and response: a systematic review of machine learning and deep learning approaches. *One Health.* (2022) 15:100439. doi: 10.1016/j.onehlt.2022.100439
27. Kassaw AK, Bekele G, Kassaw AK, Yimer A. Prediction of acute respiratory infections using machine learning techniques in Amhara region, Ethiopia. *Sci Rep.* (2024) 14:27968. doi: 10.1038/s41598-024-76847-3
28. Fang ZG, Yang SQ, Lv CX, An SY, Wu W. Application of a data-driven xgboost model for the prediction of covid-19 in the USA: a time-series study. *BMJ Open.* (2022) 12:e56685. doi: 10.1136/bmjopen-2021-056685
29. Chen Z, Liu H, Zhang Y, Xing F, Jiang J, Xiang Z, et al. Identifying major depressive disorder among us adults living alone using stacked ensemble machine learning algorithms. *Front Public Health.* (2025) 13:1472050. doi: 10.3389/fpubh.2025.1472050
30. Adnan M, Alarood A, Uddin MI, Ur RI. Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models. *PeerJ Comput Sci.* (2022) 8:e803. doi: 10.7717/peerj-cs.803
31. Nduka IC, Huang T, Li Z, Yang Y, Yim S. Long-term trends of atmospheric hot-and-polluted episodes (hpe) and the public health implications in the Pearl River Delta region of China. *Environ Pollut.* (2022) 311:119782. doi: 10.1016/j.envpol.2022.119782
32. Sichuan statistical yearbook. *China Statistics Press* (2023). Available at: <https://tjj.sc.gov.cn/scstj/tjnjnew/2023/zk/indexeh.htm>
33. Peng S. 1-km monthly precipitation dataset for China (1901–2023) National Tibetan Plateau Data Center (2024). doi: 10.5281/zenodo.3114194
34. Wei JLZ. ChinaHighPM2.5: High-resolution and High-quality Ground-level PM2.5 Dataset for China (2000–2023). *National Tibetan Plateau Data Center* (2024). doi: 10.5281/zenodo.3539349
35. Gao S, Shi Y, Zhang H, Chen X, Zhang W, Shen W, et al. China regional 250m fractional vegetation cover data set (2000–2023). *National Tibetan Plateau Data Center* (2024). doi: 10.11888/Terre.tpd.300330
36. Payne SM. Laboratory cultivation and storage of shigella. *Curr Protoc Microbiol.* (2019) 55:e93. doi: 10.1002/cpmc.93
37. Kovats RS, Edwards SJ, Hajat S, Armstrong BG, Ebi KL, Menne B. The effect of temperature on food poisoning: a time-series analysis of salmonellosis in ten european countries. *Epidemiol Infect.* (2004) 132:443–53. doi: 10.1017/S0950268804001992
38. Carlton EJ, Eisenberg JN, Goldstick J, Cevallos W, Trostle J, Levy K. Heavy rainfall events and diarrhea incidence: the role of social and environmental factors. *Am J Epidemiol.* (2014) 179:344–52. doi: 10.1093/aje/kwt279
39. Effler E, Isaacscon M, Arntzen L, Heenan R, Canter P, Barrett T, et al. Factors contributing to the emergence of *Escherichia coli* O157 in Africa. *Emerg Infect Dis.* (2001) 7:812–9. doi: 10.3201/eid0705.017507
40. Nichols G, Lane C, Asgari N, Verlander NQ, Charlett A. Rainfall and outbreaks of drinking water related disease and in England and wales. *J Water Health.* (2009) 7:1–8. doi: 10.2166/wh.2009.143
41. Yamamoto N, Urabe K, Takaoka M, Nakazawa K, Gotoh A, Haga M, et al. Outbreak of cryptosporidiosis after contamination of the public water supply in Saitama prefecture, Japan, in 1996. *Kansenshogaku Zasshi.* (2000) 74:518–26. doi: 10.11150/kansenshogakuzasshi1970.74.518
42. Meisner A, Rousk J, Bååth E. Prolonged drought changes the bacterial growth response to rewetting. *Soil Biol Biochem.* (2015) 88:314–22. doi: 10.1016/j.soilbio.2015.06.002
43. Zhang Y, Peña-Arancibia JL, McVicar TR, Chiew FH, Vaze J, Liu C, et al. Multi-decadal trends in global terrestrial evapotranspiration and its components. *Sci Rep.* (2016) 6:19124. doi: 10.1038/srep19124
44. Romano S, Fragola M, Alifano P, Perrone MR, Talà A. Potential human and plant pathogenic species in airborne pm10 samples and relationships with chemical components and meteorological parameters. *Atmosphere.* (2021) 12:654. doi: 10.3390/atmos12050654
45. Chen X, Kumari D, Achal V. A review on airborne microbes: the characteristics of sources, pathogenicity and geography. *Atmosphere.* (2020) 11:919. doi: 10.3390/atmos11090919
46. Salim SY, Jovel J, Wine E, Kaplan GG, Vincent R, Thiesen A, et al. Exposure to ingested airborne pollutant particulate matter increases mucosal exposure to bacteria and induces early onset of inflammation in neonatal il-10-deficient mice. *Inflamm Bowel Dis.* (2014) 20:1129–38. doi: 10.1097/MIB.0000000000000066
47. Farah A, Paul P, Khan AS, Sarkar A, Laws S, Chaari A. Targeting gut microbiota dysbiosis in inflammatory bowel disease: a systematic review of current evidence. *Front Med.* (2025) 12:1435030. doi: 10.3389/fmed.2025.1435030
48. Zhang Y, Odo DB, Li J, Hu L, Qiu H, Xie Y, et al. Greenspace and burden of infectious illnesses among children in 49 low- and middle-income countries. *Cell Rep Sustain.* (2024) 1:100150. doi: 10.1016/j.crsus.2024.100150
49. Pienkowski T, Dickens BL, Sun H, Carrasco LR. Empirical evidence of the public health benefits of tropical forest conservation in Cambodia: a generalised linear mixed-effects model analysis. *Lancet Planet Health.* (2017) 1:e180–7. doi: 10.1016/S2542-5196(17)30081-5
50. Dosskey MG, Vidon P, Gurwick NP, Allan CJ, Duval TP, Lowrance R. The role of riparian vegetation in protecting and improving chemical water quality in streams. *JAWRA J Am Water Resour Assoc.* (2010) 46:261–77. doi: 10.1111/j.1752-1688.2010.00419.x
51. Lee H, Mayer H, Chen L. Contribution of trees and grasslands to the mitigation of human heat stress in a residential district of Freiburg, Southwest Germany. *Landsc Urban Plann.* (2016) 148:37–50. doi: 10.1016/j.landurbplan.2015.12.004
52. Wu J, Yunus M, Ali M, Escamilla V, Emch M. Influences of heatwave, rainfall, and tree cover on cholera in Bangladesh. *Environ Int.* (2018) 120:304–11. doi: 10.1016/j.envint.2018.08.012
53. Wang L, Xu C, Xiao G, Qiao J, Zhang C. Spatial heterogeneity of bacillary dysentery and the impact of temperature in the Beijing-Tianjin-Hebei region of China. *Int J Biometeorol.* (2021) 65:1919–27. doi: 10.1007/s00484-021-02148-3
54. Huang D, Guan P, Guo J, Wang P, Zhou B. Investigating the effects of climate variations on bacillary dysentery incidence in Northeast China using ridge regression and hierarchical cluster analysis. *BMC Infect Dis.* (2008) 8:130. doi: 10.1186/1471-2334-8-130