DOI: 10.2478/eces-2026-0003 ECOL CHEM ENG S. 2026;33(1):47-62

Dehong SUN[1], Senwei ZHENG[1*] and Yinhui YU[1]

MACHINE LEARNING ANALYSIS OF COASTAL WATER POLLUTION IN CHINA: DRIVERS AND COMPLEX RELATIONSHIPS

Abstract: This study investigates the nonlinear effects of socioeconomic, pollution control, and technological innovation factors on seawater pollution using a Generalised Additive Model (GAM) and panel data from 11 Chinese coastal provinces (2018-2023). Key findings indicate: (1) Chemical Oxygen Demand Emissions (CODE) from direct marine discharge sources showed a temporal pattern of "initial increase, subsequent decrease, and gradual stabilisation", with significant spatial heterogeneity - Zhejiang recorded the highest emissions, Tianjin the lowest; (2) GAM revealed significant nonlinear relationships between most variables and CODE; (3) Factor impacts exhibited distinct range-dependent characteristics. These findings provide a scientific basis for identifying key pollution sources and formulating differentiated control strategies, leading to targeted policy recommendations.

Keywords: seawater pollution, sustainable development, machine learning, generalised additive model (GAM), nonlinear effects

Introduction

Marine pollution presents a severe threat to global marine ecosystems and poses a critical challenge to achieving the United Nations (UN) Development Goals (SDGs). As a key indicator of coastal environmental health, seawater pollution directly reflects ecosystem status and is intrinsically linked to Sustainable Development Goal 14 (SDG 14), which aims to "conserve and sustainably use the oceans, seas and marine resources". This goal explicitly commits to preventing and significantly reducing all forms of marine pollution by 2025, with a specific focus on land-based activities, marine debris, and nutrient pollution. It further encompasses systematic governance tasks such as mitigating eutrophication and addressing ocean acidification. Ultimately, these objectives seek to implement integrated land-sea coordination in comprehensive governance. This approach reduces pollutant damage to marine biodiversity and ecosystem functions, safeguards marine ecosystem integrity and services, and ensures the sustainable use of marine resources for human well-being and global ecological security.

As a complex and systemic environmental issue, seawater pollution arises from the interplay of multidimensional and multifactorial drivers. Traditional studies, often approaching the problem from a single perspective such as industrial emissions or

[1] Minnan University of Science and Technology, No. 1, Longjing Road, 362000, Quanzhou, China, phone +86-187-5051-6366, fax +86-595-2324-5687, email: 155065349@163.com, ORCID: DS 0009-0006-5337-3101, SZ 0009-0007-7920-8638, YY 0000-0001-6021-472X
[*] Corresponding author: 84246354@qq.com

agricultural non-point source pollution, have been limited in their ability to fully elucidate its comprehensive driving mechanisms. In reality, socioeconomic factors - including coastal economic growth patterns, industrial pollution control efforts, agricultural fertiliser application, technological innovation levels, and urbanisation processes-collectively form a complex network influencing coastal environmental quality. Therefore, unravelling the complex causal pathways between these multidimensional drivers and seawater pollution-particularly their potential nonlinear dynamics-holds critical significance for developing nuanced coastal environmental management policies and represents the fundamental rationale for this investigation.

Compared to existing research, this paper demonstrates innovation in the following three aspects. First, it breaks new ground in research content. Previous studies have often focused on single-dimensional variables, lacking an integrated analytical framework. This paper systematically investigates the combined effects of economic, pollution control, pollution source, technological, and urbanisation factors on seawater pollution, aiming to reveal the comprehensive mechanisms of multi-factor interactions. Second, it introduces methodological innovation. Existing studies on the influencing factors of seawater pollution predominantly rely on traditional econometric models, which typically assume simple linear relationships among variables, despite the complex interconnections in reality. To address this, the paper employs a Generalised Additive Model（GAM) from the field of machine learning to identify and analyse potential nonlinear effects of variables on seawater pollution. Compared to conventional regression methods, machine learning offers superior fitting accuracy, providing a more reliable and efficient analytical tool for studying influencing factors in the marine environment. The application of machine learning in this domain also represents a valuable extension of the existing methodological toolkit. Finally, it offers advantages in the practical value of research conclusions. By using the GAM model, the dynamic evolution paths of different variables affecting seawater pollution can be delineated in detail. This analysis not only deepens the understanding of the complex mechanisms between various factors and seawater pollution but also provides a scientific basis and precise decision-making support for the design and implementation of relevant policies. Based on the nonlinear patterns and dynamic characteristics revealed by the GAM, critical insights can be offered for developing zoned and categorised coastal environmental management strategies, thereby enhancing the targeting and effectiveness of marine pollution governance.

This paper comprises five distinct sections. The subsequent section offers a review of existing scholarship concerning marine environmental pollution and the roles played by different variables in the marine ecosystem. Following that, the following section outlines the research design, detailing the methodological approach, criteria for indicator selection, and data sources employed. The penultimate section analyses the spatiotemporal distribution characteristics of seawater pollution, with a particular focus on the specific impacts exerted by various influencing factors. The final section synthesises the principal findings and advances targeted policy recommendations. These suggestions focus on critical areas including economic development, urbanisation, and the management of industrial and agricultural pollution, offering a scientific basis for reconciling coastal region development with the conservation of the marine environment.

Literature review

Marine environmental pollution

Marine environmental pollution, as a global challenge with diverse sources and far-reaching impacts, has become one of the major impediments to achieving SDG 14 [1]. Existing studies have extensively documented the diverse categories of marine pollutants and their associated ecological risks. Among these, land-based pollution is widely recognised as the primary source of marine pollution [2]. Among various pollutants, eutrophication caused by nutrient salts and organic pollutants is particularly prominent. For instance, Zhang et al. [3] found in their study of Tieshan Bay that indicators such as chemical oxygen demand were significantly influenced by land-based inputs [3]. Further research by Nan et al. [4] in the Pearl River Estuary revealed that the synergistic effect between nutrient inputs and reduced suspended sediments is a key factor driving the expansion of hypoxic zones in the region, highlighting the complex environmental effects of pollutants after they enter the sea [4]. Meanwhile, emerging contaminants pose an increasingly prominent threat. Numerous studies have reported the widespread presence of microplastics in coastal waters [5, 6]. Through Mike 21 modelling of microplastic transport in the Zhoushan Archipelago, Li et al. [6] not only confirmed their ecological risks but also demonstrated the complex influence of hydrodynamic forces and seasonal variations on pollutant dispersion, suggesting potential nonlinear dynamic processes between driving factors and pollution outcomes. Furthermore, issues such as heavy metals and antibiotic resistance genes mediated through microplastic transmission collectively constitute multiple environmental pressures on coastal ecosystems [7, 8].

Influencing factors of seawater pollution

Seawater pollution results from the combined effects of natural processes and human activities, with intricately intertwined drivers. Existing literature has primarily explored the following dimensions: Firstly, socioeconomic factors serve as fundamental drivers. Sun et al. [9] demonstrated that tourism concentration and fintech development were found to exacerbate coastal carbon footprints and water pollution, indirectly reflecting the environmental impacts of economic development patterns. Fan et al. [10] identified coastal aquaculture expansion as a key driver of mangrove vulnerability, further illustrating the direct environmental effects of economic activities. These findings collectively suggest that the relationships between economic factors and environmental pressure may not follow simple linear patterns. Secondly, pollution control investment serves as a crucial regulatory mechanism. Industrial pollution control investment represents an important indicator of societal commitment to addressing environmental issues. However, the effectiveness of such investments may not be immediately realised. International experience demonstrates that effective pollution control requires not only financial input but also a sound legal regulatory framework and public awareness, suggesting that the efficiency of governance investment may exhibit diminishing marginal returns or an optimal range [11, 12]. Thirdly, agricultural production constitutes a major non-point source. The application of agricultural fertilisers is closely associated with non-point source pollution. A meta-analysis by Adebanjo-Aina and Oludoye [13] confirmed the significant impact of nitrogen fertiliser use on water quality. Further research by Hu et al. in a karst basin identified cultivated land, particularly sloping farmland, as "critical source areas" for nitrogen and phosphorus loss [14].

Methodologically, conventional studies on the driving factors of seawater pollution have predominantly employed correlation analyses or linear regression models which fail to capture the sophisticated nonlinear dynamics among variables [15]. Although emerging research has begun adopting numerical models or spatial econometric approaches to elucidate complex mechanisms, the systematic application of the GAM from machine learning to analyse pollution drivers in China's coastal waters remains scarce in the literature. GAM, as a powerful nonlinear modelling tool, has demonstrated its applicability and effectiveness across numerous environmental and ecological studies. For instance, in the public health domain, Li et al. [16] successfully utilised GAM to reveal nonlinear relationships between hospital environmental hygiene and infection rates; in species distribution prediction, the ensemble model by Melese et al. [17] demonstrated GAM as one of the reliable algorithms for predicting climate-induced nonlinear impacts on species habitats; in ecological restoration research, Xie et al. [18] applied GAM to disentangle complex nonlinear relationships between post-fire vegetation recovery and multiple environmental drivers; moreover, in fisheries resource management and studies on the impact of land use on disease transmission [19-21], GAM has consistently proven effective in capturing nonlinear effects among variables while maintaining excellent explanatory power. These collective findings demonstrate GAM's unique advantages in handling multifactorial nonlinear interaction mechanisms within complex "social-economic-environmental" systems. With its demonstrated exceptional nonlinear fitting capability and interpretability in environmental fields [22], these cross-disciplinary successful applications provide crucial methodological references and innovative opportunities for this study's adoption of GAM to investigate pollution drivers in China's coastal waters.

Recent advancements in 2024-2025 underscore a methodological shift towards data-driven approaches in environmental research. Machine learning (ML) demonstrates broad utility, from quantifying urbanisation impacts on forest carbon [23] and identifying key drivers of emissions [24], to disentangling complex atmospheric processes like ozone transport [25] and constructing high-resolution emission inventories [26]. These applications highlight ML's strength in handling nonlinear, high-dimensional, and spatiotemporal data. This trend is exemplified by significant concurrent advances in GAMs, a powerful and interpretable class of ML models. Recent GAM research showcases their refined capabilities across diverse domains. In ecological studies, Li et al. [27] employed GAM to reveal nonlinear and seasonal relationships between species resources and environmental factors in the Yangtze River Estuary. For air quality analysis under complex anthropogenic influences, Ji et al. [28] demonstrated the enhanced predictive capability of GAM combined with transfer learning. To address spatial heterogeneity and improve interpretability, Cheng et al. [29] developed the Geographical Gaussian Process GAM (GGP-GAM) for high-precision soil mapping, while Comber et al. [30] systematically elaborated on using GAMs to construct spatially varying coefficient models. Furthermore, addressing model interpretation needs, Lai et al. [31] introduced the gam.hp R package for evaluating predictor importance within GAMs. These focused developments indicate that GAMs continue to be refined in handling nonlinearity, spatial complexity, and interpretability challenges. The demonstrated strengths of ML, coupled with the specific and ongoing advancements in GAM frameworks, strongly reinforce the rationale for adopting GAM in this study to uncover the nuanced, non-linear relationships between anthropogenic drivers and environmental pressures in China's coastal waters.

Research design

Research methodologies

GAM is an important class of interpretable models in machine learning, effectively capturing nonlinear relationships between variables and demonstrating strong applicability in handling complex nonlinear relationships within environmental systems. Compared to 'black-box' machine learning models such as Random Forest or XGBoost, GAM offers a superior balance between fitting flexibility and result interpretability. This characteristic makes it particularly suitable for the primary objective of this study: to not only achieve accurate predictions but, more importantly, to understand and visually interpret the marginal effects and complex nonlinear dynamics of each driver on seawater pollution through its smooth function terms. This model has been successfully applied across multiple environmental research domains: in water supply system studies, GAM has been used to analyse the nonlinear effects of pipe materials and environmental factors on pipe failure, visually representing variable relationships through partial dependence plots [32]; in air quality research, GAM has effectively quantified the contributions of meteorological factors and precursors to PM2.5 and ozone concentrations, revealing their nonlinear response mechanisms [33, 34].

In terms of model structure, the GAM extends the additive framework of traditional linear models by replacing linear terms with smooth nonlinear functions, thereby more accurately describing the actual relationships between variables. The standard multiple linear regression form is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \quad (1)$$

where Y is the dependent variable, β_0 is the intercept term, β_i $(i = 1,2,\dots,p)$ are the regression coefficients, X_i is the i-th independent variable, and ε is the error term. The GAM replaces each term $\beta_i X_i$ in the linear predictor with a smooth nonlinear function $f_i(X_i)$. The model expression is:

$$Y = \beta_0 + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p) + \varepsilon \quad (2)$$

This study fits the model using the gam function from the mgcv (Mixed GAM Computation Vehicle) package in the R environment. This method employs an iterative process to gradually optimise the smooth functions of each variable, effectively capturing the independent influence of variables while also maintaining the overall interpretability of the model. By constructing nonlinear smooth terms for each predictor variable, GAM can uncover complex data patterns overlooked by traditional linear specifications. Consequently, while improving prediction accuracy, the model clearly delineates the marginal effects of various drivers on seawater Chemical Oxygen Demand Emissions (CODE), thereby yielding more insightful analytical tools for formulating precise coastal environmental management strategies.

Definition of variables

The study area of this research comprises China's coastal provinces (autonomous regions and municipalities directly under the central government). Data were obtained from the China Statistical Yearbook and the China Marine Ecological and Environmental Status Bulletin published by the National Bureau of Statistics of China, covering the period from 2018 to 2023 across 11 provinces (autonomous regions and municipalities) in China, with the data type being panel data.

Chemical oxygen demand serves as a key comprehensive indicator for measuring the total amount of organic pollutants in water bodies. Therefore, to directly assess the marine environmental pressure from land-based activities, this study adopts CODE from direct marine discharge sources as the dependent variable, aiming to directly capture the terminal pressure exerted by land-based socioeconomic activities on the coastal marine environment. Focusing on the linkage between coastal development and the environment, the independent variables are specified as follows: Gross Domestic Product (GDP), Industrial Pollution Control Completed Investment (IPCCI), Output of Agricultural Nitrogen, Phosphorus and Potassium Chemical Fertilisers (OANPKCF), Number of Domestic Patent Applications Accepted (NDPAA), and Urbanisation Rate (UR). Definitions of all variables are presented in Table 1, and their descriptive statistics are provided in Table 2. The data show that the value of CODE ranges widely from 552 tonnes (tonne = 10^6 g = Mg; [t]) to 61,204 t, with a mean value of 13,327 t.

Table 1

Variables and definitions

Variables	Unit	Definition
CODE	[t]	Chemical oxygen demand emissions
GDP	[100 million CNY]	Gross domestic product
IPCCI	[10,000 CNY]	Industrial Pollution Control Completed Investment
OANPKCF	[10,000 t]	Output of agricultural nitrogen, phosphorus and potassium chemical fertilisers
NDPAA	[item]	Number of domestic patent applications accepted
UR	[%]	Urbanisation rate

Table 2

Variables description statistics

Variables	Min	Q1	Median	Mean	Q3	Max
CODE	552	2,047	7,453	13,327	18,758	61,204
GDP	5,004	25,400	42,839	53,562	79,769	137,905
IPCCI	476	72,682	128,491	212,067	280,529	987,539
OANPKCF	0.84	24.31	49.29	99.49	166.48	472.69
NDPAA	6,451	87,021	161,868	286,305	450,663	993,480
UR	51.83	61.71	71.21	70.33	74.37	89.47

Results and discussion

Spatiotemporal distribution of seawater pollution

Temporal evolution of seawater pollution

Based on data from the 2018-2023 China Marine Ecological Environment Status Bulletin, Figure 1 illustrates the temporal changes in CODE for direct marine pollution sources during the same period. Each year is represented as an independent sector uniformly distributed around the circumference, with its radial scale corresponding to the emission volume value. Specifically, emissions reached 147,625 t in 2018, rose significantly to a peak of 161,493 t in 2019, then declined to 148,902 t in 2020 and further dropped to 141,843 t in 2021. Subsequently, emissions declined further to a new low of 136,291 t in 2022, before rebounding slightly to 143,404 t in 2023.

Analysis of the CODE trajectory reveals a clear three-phase evolutionary pattern: initial surge, consecutive decline, and eventual stabilisation. The 2019 peak was primarily driven by expansion of high-pollution industries, regulatory delays, and insufficient treatment facilities. The 2020-2021 decline can be attributed to strengthened environmental policies coupled with industrial production reductions during pandemic restrictions. The 2022-2023 stabilisation phase reflects the cumulative effects of pollution control technology advancements and industrial restructuring toward greener development.

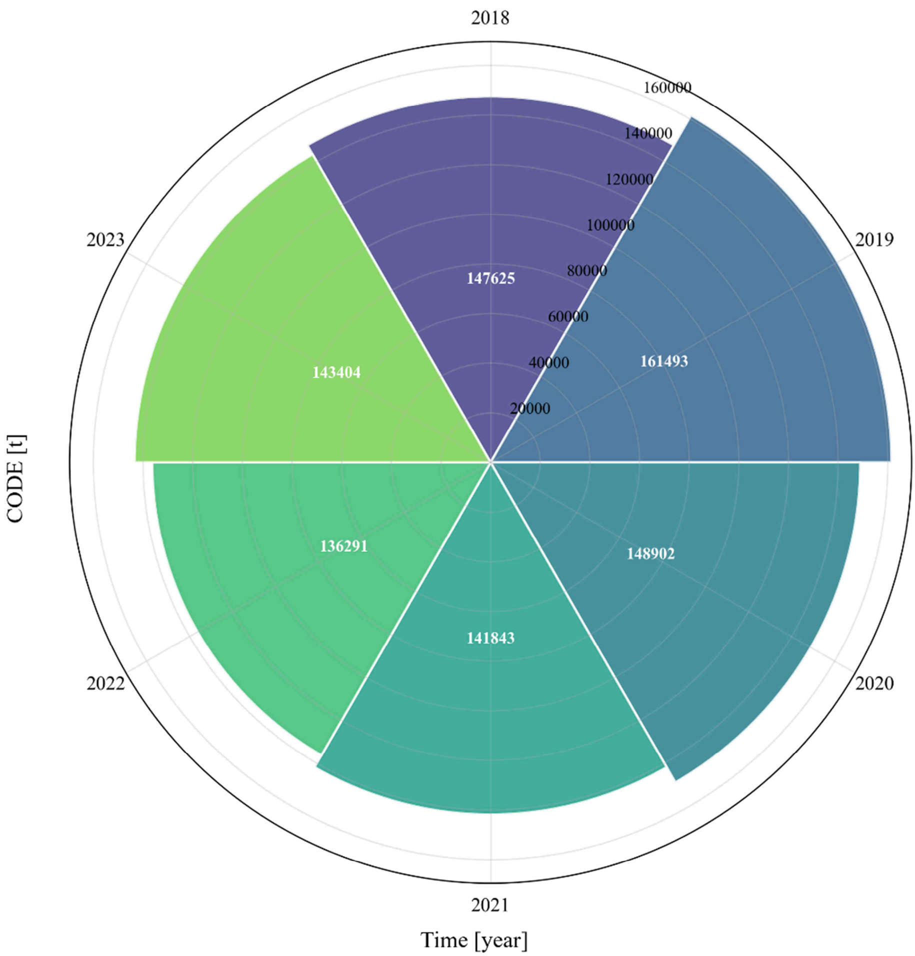

Fig. 1. Temporal evolution of CODE (2018-2023). Source: authors' calculations based on data from the 2018-2023 China Marine Ecological Environment Status Bulletin

Spatial distribution of seawater pollution

Seawater pollution demonstrates significant spatial heterogeneity. Figure 2 presents CODE from direct marine discharge sources across different coastal regions of China from 2018 to 2023. As shown in Figure 2, substantial disparities in CODE emissions exist among regions. Zhejiang Province recorded the highest emissions, approaching 350,000 t and significantly exceeding all other regions. This was followed by Shandong Province (approximately 150,000 t) and Guangdong Province (around 120,000 t). Subsequent regions including Fujian and Hainan provinces, showed sequentially decreasing emissions, with Tianjin Municipality registering the lowest recorded levels. The substantial disparities in CODE from direct marine discharge sources across China's coastal regions primarily stem from spatially heterogeneous distributions of industrial structure, pollution control capacity, and demographic-geographic conditions.

Zhejiang Province leads in CODE due to concentrated high-pollution industries (chemical and textile printing/dyeing), significant urbanisation pressure, and partial direct wastewater discharge. As major industrial and agricultural bases, Shandong and Guangdong provinces follow in emission volumes-the former characterised by substantial industrial wastewater volumes and numerous riverine discharges, the latter by dense manufacturing in the Pearl River Delta region and inadequate pollution control in small-to-medium enterprises. Regions like Fujian and Hainan demonstrate progressively declining emissions, attributable to their economic focus on light industry and tourism, coupled with substantial pollution control investments and stringent regulatory enforcement. Tianjin maintains the lowest emission levels, achieved through coastal buffer zones for high-pollution industries, high wastewater collection-treatment rates, and rigorous environmental supervision. These spatial differentiations necessitate tailored approaches: high-emission regions require industrial upgrading, enhanced treatment standards, and strengthened oversight, while low-emission regions offer transferable governance models for pollution mitigation.

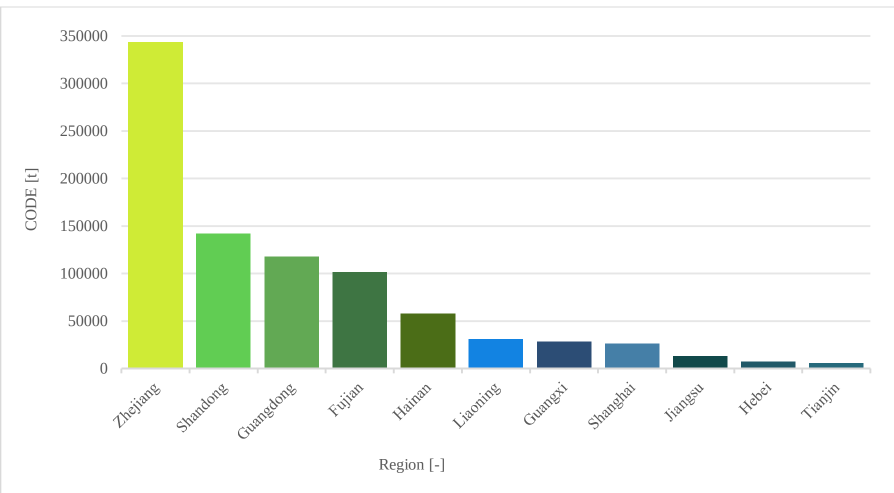

Fig. 2. Comparison of CODE in coastal provinces (2018-2023). Source: Authors' calculations based on data from the 2018-2023 China Marine Ecological Environment Status Bulletin

GAM evaluation

Correlation analysis

To examine the statistical relationships between the variables and CODE from direct marine discharge sources, this study first conducted a correlation analysis. Figure 3 presents a heatmap of the correlation coefficient matrix among the variables, where colour intensity intuitively reflects the strength of correlations. The results indicate the following correlation strengths with CODE, in descending order: NDPAA (0.38), GDP (0.34), IPCCI (0.25), OANPKCF (0.04), and UR (–0.03).

Although correlation coefficients can measure the strength of linear associations between variables, they are inherently descriptive tools offering only preliminary insights. A fundamental limitation of this metric lies in its inability to establish causal links between variables or to capture the complex nonlinear dynamics that may underlie their relationships. Therefore, to gain deeper insights into the influence mechanisms of various variables on CODE and the intrinsic nature of their relationships, this study further employs regression analysis and specifically selects the GAM as the primary analytical tool. This model not only enables a more reliable quantification of variable relationships but also overcomes the constraints of linear models by effectively uncovering potential nonlinear associations among variables. Consequently, it can offer more scientifically grounded and effective decision support for formulating optimised seawater pollution control strategies.

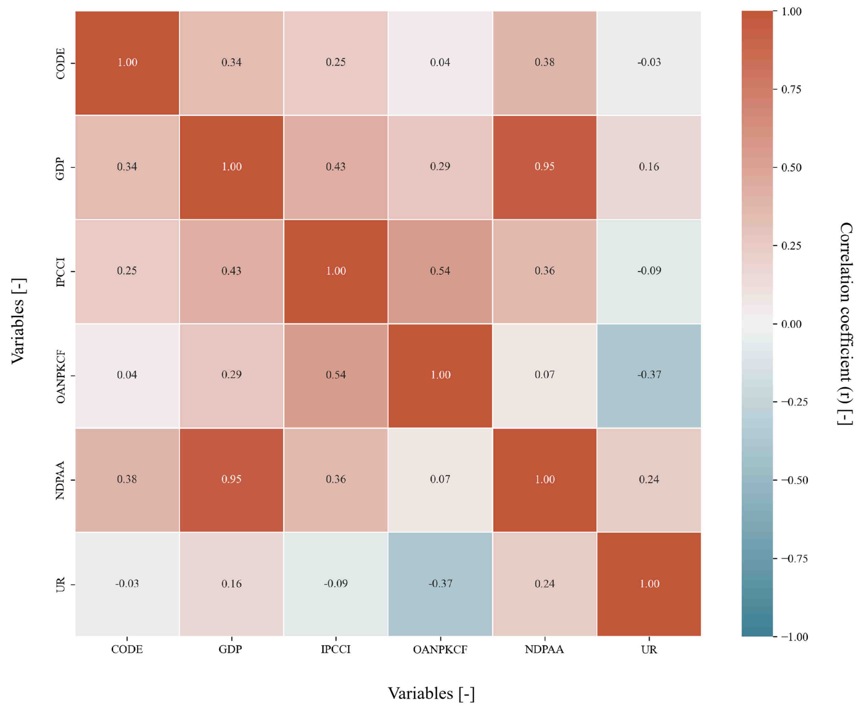

Fig. 3. Heat map of variable correlation

GAM results

The results of the GAM are presented in Table 3. Regarding *p*-values, with the exception of OANPKCF, which has a *p*-value of 0.056629 (approaching 0.05), all other variables exhibit *p*-values below 0.05. This indicates that the smooth functions for all variables except OANPKCF passed the test for statistical significance, confirming that the selected variables significantly influence the CODE from direct marine discharge sources. The Effective Degrees of Freedom (EDF) provides a key metric for evaluating the contribution of individual variables within the model. Statistically significant smooth terms with EDF values exceeding 1 generally suggest the presence of nonlinear relationships between predictors and the response variable. As evidenced in Table 3, all fitted smooth terms demonstrate EDF values substantially above 1, confirming pervasive nonlinear effects across the examined variables. Based on the specific numerical values, the ranking of variable contributions to the model is as follows: UR (8.837), GDP (8.200), IPCCI (7.590), OANPKCF (4.180), and NDPAA (3.979).

The adjusted *R*-squared statistic for the generalised additive model was 0.96, with deviance explained of 98 %. These metrics demonstrate the model's exceptional explanatory power for the variation in CODE from direct marine discharge sources, further validating the superior performance of the constructed nonlinear regression model in reliably capturing the relationships between the variables and CODE.

Table 3

GAM results

Variables	**EDF**	**F**	***p*-value**
S(GDP)	8.200	6.921	1.71e-05
S(IPCCI)	7.590	2.956	0.013022
S(OANPKCF)	4.180	2.440	0.056629
S(NDPAA)	3.979	6.287	0.000409
S(UR)	8.837	4.332	0.000951

Impacts of various variables on CODE

The impact of GDP on CODE

As shown in Figure 4a, the impact of GDP on CODE exhibits complex nonlinear characteristics. Overall, it demonstrates a trend of initial fluctuation, followed by an increase, and eventually a significant decline. When GDP values are below 60,000 (100 million CNY), their impact on CODE emissions remains relatively weak and unstable. However, as GDP values fall within the range of 60,000-80,000 (100 million CNY), their positive driving effect on CODE emissions becomes significantly enhanced. Notably, when GDP exceeds 80,000 (100 million CNY), the rate of change in CODE rapidly decreases and shifts to a negative direction, indicating that GDP growth beyond this threshold transitions from promoting to inhibiting CODE emissions, thereby revealing a distinct nonlinear threshold effect.

As a core indicator of regional economic development levels, the scale and structure of GDP play a decisive role in pollutant emissions. The pattern displayed in Figure 4a suggests that economic growth exerts the most substantial pressure on CODE when GDP ranges between 60,000 and 80,000 (100 million CNY), likely corresponding to the accelerated industrialisation phase. When GDP surpasses the critical threshold of 80,000

(100 million CNY), its impact shifts from positive to negative. This finding provides empirical support for the Environmental Kuznets Curve (EKC) hypothesis, indicating that advanced economic development can achieve decoupling from environmental pollution. The conclusion highlights a clear phased threshold effect between economic growth and marine pollution. Therefore, in future environmental governance practices, differentiated emission reduction strategies should be implemented based on the specific stage of regional economic development to achieve synergistic development of economic growth and marine environmental protection in coastal areas.

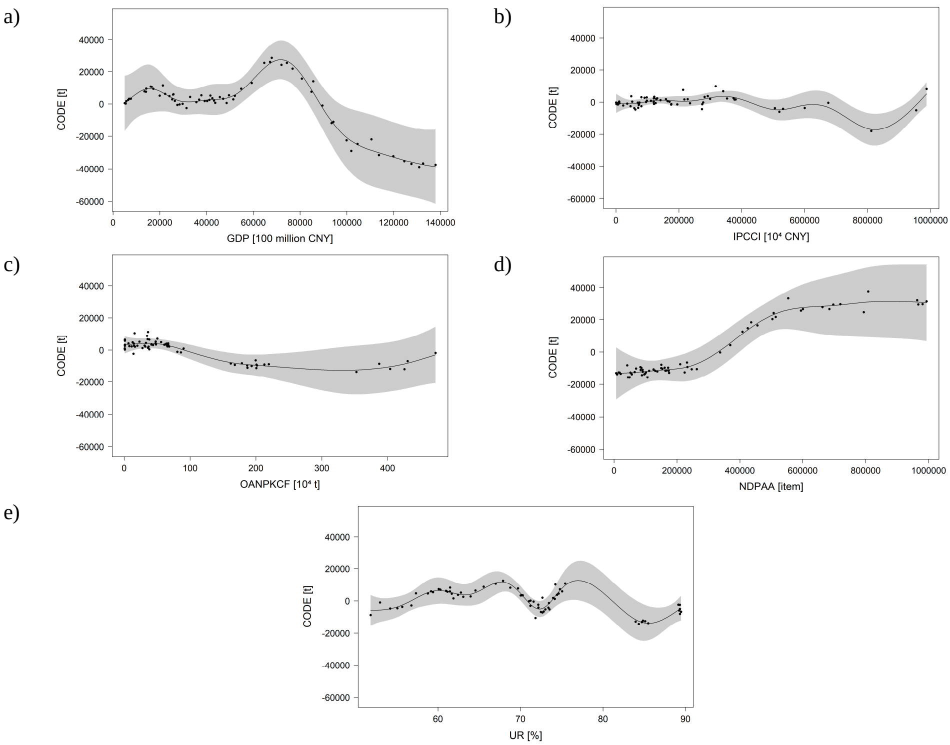

Fig. 4. Impact of variables on CODE: a) GDP, b) IPCCI, c) OANPKCF, d) NDPAA, e) UR. The estimated degrees of freedom for the smooth functions are: s (GDP, 8.2), s (IPCCI, 7.6), s (OANPKCF, 4.2), s (NDPAA, 4.0), and s (UR, 8.9)

The impact of IPCCI on CODE

As shown in Figure 4b, IPCCI demonstrates a significant nonlinear impact on CODE. Specifically, when IPCCI is below $4\cdot 10^5$ (10,000 CNY), its suppressive effect on CODE emissions remains relatively limited. As IPCCI increases to the range of $4\cdot 10^5$ - $8\cdot 10^5$ (10,000 CNY), the emission reduction effect becomes substantially enhanced, reaching its optimum near $8\cdot 10^5$ (10,000 CNY). However, when the investment scale exceeds this critical value, diminishing marginal abatement benefits are observed.

Industrial pollution control investment, serving as a direct reflection of environmental regulatory stringency, represents a crucial approach for controlling the discharge of industrial pollutants into the sea. Figure 4b reveals that when IPCCI falls within the range of $4 \cdot 10^5$ - $8 \cdot 10^5$ (10,000 CNY), the environmental benefits generated by the investment reach their maximum. However, beyond $8 \cdot 10^5$ (10,000 CNY), emission reduction efficiency not only fails to improve but actually declines, potentially indicating the technological efficiency ceiling of existing treatment methods or suggesting that further substantial emission reductions require systemic transformations such as industrial restructuring and source control. This pattern clearly demonstrates a diminishing marginal returns relationship between pollution control investment and environmental benefits. Consequently, in environmental management practice, it is essential to scientifically determine differentiated optimal investment levels according to regional pollution loads and techno-economic conditions, avoiding indiscriminate excessive expenditure to achieve synergistic optimisation of environmental and economic benefits.

The impact of OANPKCF on CODE

As illustrated in Figure 4c, OANPKCF exhibits a composite nonlinear relationship with CODE. Specifically, when OANPKCF exceeds 200 (10,000 t), its promoting effect on CODE demonstrates an increasingly enhanced pattern with continued dosage increases. Notably, within the range of 300 - 400 (10,000 t), OANPKCF manifests the most substantial positive driving effect on CODE emissions.

Agricultural non-point source pollution represents a major contributor to marine pollutants, with OANPKCF directly influencing the flux of nitrogen and phosphorus elements into marine systems. Figure 4c reveals a critical environmental safety threshold in agricultural production (approximately 200 (10,000 t)): below this threshold, the impact on the marine environment remains manageable; however, once this critical point is exceeded, the driving effect on coastal CODE becomes significantly amplified. This underscores the importance of balancing food security and environmental protection in agricultural management practices. We recommend implementing precision fertiliser technologies to maintain fertiliser application rates within ecological safety thresholds, while simultaneously optimising cropping structures and adopting complementary measures to achieve coordinated development of agricultural production and coastal environmental protection.

The impact of NDPAA on CODE

As shown in Figure 4d, NDPAA exhibits a positive nonlinear relationship with CODE. Specifically, when NDPAA remains below 200,000 items, the curve remains relatively flat; as NDPAA enters the range of 200,000 - 800,000 items, the curve shows a marked upward trajectory; when NDPAA exceeds 800,000 items, the curve levels off and stabilises.

This finding suggests that during the developmental stage, the growth in patent numbers has not led to a reduction in CODE but instead shows a significant positive correlation. This phenomenon may reflect several critical aspects: First, a structural imbalance exists wherein both technological innovation and patent growth are predominantly concentrated in non-environmental sectors; the industrial expansion in these sectors consequently leads to increased pollutant emissions. Second, the translation and application of green technologies remain inadequate, as environmentally-focused patents likely constitute a relatively small share of the total, coupled with an extended transition period from patent application to practical implementation. Furthermore, at this specific phase of economic development, technological innovation primarily serves economic growth objectives, while its potential environmental benefits have not yet been fully realised.

This finding carries significant policy implications: when promoting the innovation-driven development strategy, it is imperative to optimise the patent structure, enhance support for green technology innovation, establish rapid transformation mechanisms for green patents, and ensure that technological innovation genuinely contributes to green development goals.

The impact of UR on CODE

As shown in Figure 4e, the impact of UR on CODE demonstrates a complex multistage nonlinear dynamic. Specifically, when the urbanisation rate remains below 65 %, its impact exhibits an overall fluctuating upward trend, indicating the progressive accumulation of environmental pressure during the initial stages of urbanisation, driven by infrastructure expansion and industrialisation. As the urbanisation rate enters the 65 % - 75 % range, its influence displays a V-shaped pattern - characterised by an initial decline followed by a rebound - reflecting the dynamic interplay between short-term achievements in environmental governance and persistent pressures stemming from long-term economic growth. When UR falls within the 75 % - 85 % range, the relationship transitions into a consistent downward trajectory, likely driven by the emission reduction effects of technological advancements beginning to systematically surpass the pollution effects of scale expansion. Once UR exceeds 85 %, the relationship reverses again to an upward trend.

This non-monotonic trajectory indicates that the environmental effects of urbanisation result from the dynamic interplay of scale, structural, and technological effects. Policy formulation must be precisely tailored to specific urbanisation stages: in highly urbanised regions (UR > 85 %), efforts should prioritise developing circular economies and enhancing resource efficiency to counter rising urban environmental pressures; whereas in the upper-middle development phase (UR 75 % - 85 %), measures should consolidate and reinforce the emission reduction achievements driven by technological progress.

Conclusion

This study analysed the effects of various variables on CODE using data from 11 coastal provinces (autonomous regions and municipalities) in China from 2018 to 2023. The main conclusions are as follows: (1) CODE emissions demonstrate distinct spatiotemporal heterogeneity. Temporally, they exhibit a three-phase evolutionary trend of "rapid surge - consecutive decline - gradual stabilisation", with the peak observed in 2019 and a transition to a lower plateau phase after 2021. Spatially, a north-south gradient distribution is evident, where Zhejiang Province shows significantly higher emission intensity than other regions, while Tianjin records the lowest emissions. This pattern reflects systematic differences in industrial structure, environmental governance, and geographical conditions across coastal areas. (2) The GAM effectively captured complex nonlinear mechanisms among variables. The model achieved an adjusted R^2 of 0.96, with smooth terms for most variables being statistically significant ($p < 0.05$). All EDF values exceeded 1, confirming nonlinear influences of all variables on CODE. Specifically, UR (EDF = 8.837) and GDP (EDF = 8.200) demonstrated the most complex nonlinear patterns. (3) The impact patterns and threshold effects of the variables exhibit distinct interval characteristics: GDP shows minimal influence below 60,000 (100 million CNY), demonstrates the strongest promoting effect within the range of 60,000-80,000 (100 million CNY), and transitions to inhibitory impact when exceeding 80,000 (100 million CNY), displaying typical Environmental Kuznets Curve (EKC) characteristics. IPCCI achieves optimal emission reduction benefits within the range of $4 \cdot 10^5$ - $8 \cdot 10^5$ (10,000 CNY), beyond which diminishing marginal benefits occur. OANPKCF exhibits a continuously strengthening promoting effect on CODE after exceeding 200 (10,000 t), with the most substantial effect observed in the range of 300-400 (10,000 t). NDPAA maintains a positive driving effect throughout the entire observed range. UR demonstrates a "complex multi-stage nonlinear fluctuation", with pressure escalation phases at 65 % - 75 % and above 85 %, while showing mitigation within the 75 % - 85 % interval.

Based on the findings, the following actionable recommendations are proposed. (1) Implement differentiated management strategies based on developmental stages. Regions with higher economic levels (e.g., GDP > 80,000 (100 million CNY)) should further strengthen environmental standards and promote green technologies, while areas still in the accelerated industrialisation phase should strictly control the addition of high-pollution production capacity and facilitate the transition to a lighter and greener industrial structure. (2) Optimise the structure and efficiency of industrial pollution control investment. When allocating funds, priority should be given to projects within the empirically identified "cost-effective investment interval" (IPCCI: $4 \cdot 10^5$ - $8 \cdot 10^5$ (10,000 CNY)). The investment focus within this interval should be strategically shifted from end-of-pipe treatment towards clean production technological transformation and circulating water system upgrades to achieve pollution reduction at source. Furthermore, a robust post-evaluation mechanism for the operational efficiency of treatment facilities must be established and institutionalised. This mechanism should link continued funding support to performance outcomes, ensuring the effective utilisation of funds and preventing diminishing marginal returns from excessive or misplaced investment. (3) Establish and enforce an agricultural ecological safety threshold management system. In coastal agricultural areas, variable-rate fertilisation technology based on soil testing and crop demand should be actively promoted, and demonstration zones for the "ecological safety fertilisation threshold" should be established. A monitoring system integrating satellite remote sensing, drone patrolling, and sensor networks should be developed to achieve grid-based precision management of fertiliser application. Furthermore, it is recommended to incorporate fertiliser application intensity into local environmental protection performance evaluation indicators, ensuring that the output of agricultural non-point source pollution is controlled within the 200 (10,000 t) environmental safety threshold. (4) Guide technological innovation toward green transformation. While encouraging patent growth, implement mechanisms such as tax incentives and green patent fast-track channels to increase the proportion of green technologies and shorten the transition period from research and development to practical application of environmental technologies. (5) Develop urbanisation stage-adapted environmental governance pathways. For highly urbanised areas (UR > 85 %), focus on developing circular economy and venous industries to enhance resource metabolic efficiency; for regions with UR between 75 % - 85 %, consolidate and promote the emission reduction effects achieved through technological progress to prevent the resurgence of environmental pressure.

In summary, by establishing a targeted governance system characterised by zoning, categorisation, staging, a shift can be achieved from single end-of-pipe treatment to systematic governance that integrates the "structure-technology-institution" triad. This approach provides a scientific pathway for implementing UN Sustainable Development Goal 14 and advancing land-sea coordination.

References

[1] Vierros M, Nelson G, Caldeira M, Polejack A, Veiga JS, Sekinairai A. Building global momentum towards managing marine plastic pollution through SDG 14. Ocean Soc. 2024;1:8388. DOI: 10.17645/oas.8388.

[2] Pan X. A strategy for risk control of marine pollution from land-based sources: Establishment of multilateral guarantee mechanism. Front Mar Sci. 2025;12:1681518. DOI: 10.3389/fmars.2025.1681518.

[3] Zhang P, Wu S, Xu M, Luo X, Peng X, Ren C, et al. Spatiotemporal nutrient patterns, stoichiometry, and eutrophication assessment in the Tieshan Bay coastal water, China. J Mar Sci Eng. 2023;11:081602. DOI: 10.3390/jmse11081602.

[4] Nan Y, Chen Z, Wang B, Liang B, Hu J. Human activities caused hypoxia expansion in a large eutrophic estuary: Non-negligible role of riverine suspended sediments. Biogeosciences. 2025;22:5573-89. DOI: 10.5194/bg-22-5573-2025.

[5] Buhhalko N, Kuddithamby G, Vianello A, Rotander A, Vidal-Liñán L, Beiras R, et al. Microplastics (10 µm-5 mm) in European Atlantic Coastal Waters. Environ Adv. 2025;21:100644. DOI: 10.1016/j.envadv.2025.100644.

[6] Li Y, Cao L, Bian J, Chen W, Leng Z, Zhao X, et al. Investigation of the migration, distribution, and influence of microplastic leakage on Tuanjishan Island in Zhoushan using Mike. Prog Fish Sci. 2024;45(2):70-81. DOI: 10.19663/j.issn2095-9869.20231023002.

[7] Chaolu, Miao B, Qian N. Geospatial analysis of heavy metal concentrations in the coastal marine environment of Beihai, Guangxi during April 2021. Coasts. 2025;5(3):27. DOI: 10.3390/coasts5030027.

[8] Bhikhoo R, Bezuidenhout K, Molale-Tom L, Mienie C, Bezuidenhout C. Marine outfall discharges contribute to coastal microplastic pollution and the spread of antimicrobial resistance. PLoS ONE. 2025;20(8):e0329599. DOI: 10.1371/journal.pone.0329599.

[9] Sun Q, Wang X, Cheng L, Yang M. Over-tourism and green investments: Spatial MMQR insights on China's coastal pollution and carbon emissions. Front Mar Sci. 2025;12:1605039. DOI: 10.3389/fmars.2025.1605039.

[10] Fan B, Li Y, Zhang Z, Yang Y, Li Y. Exploring cumulative vulnerability of mangrove forests to intensive coastal anthropogenic threats. Ecosyst Health Sustain. 2024;10:0153. DOI: 10.34133/ehs.0153.

[11] Golovko L, Ladychenko V, Yara O. Economic and international legal aspects of the protection of the marine environment from pollution. Balt J Econ Stud. 2025;11(3):339-43. DOI: 10.30525/2256-0742/2025-11-3-339-343.

[12] Verzosa RC, Katipunan FJM, Lumangyao JGB, Antonio ES. Solid waste management awareness and practices in coastal communities. Davao Res J. 2024;15(3):60-77. DOI: 10.59120/drj.v15i3.247.

[13] Adebanjo-Aina O, Oludoye O. Impact of nitrogen fertiliser usage in agriculture on water quality. Pollutants. 2025;5(3):21. DOI: 10.3390/pollutants5030021.

[14] Hu C, Wu Q, Liu G, Ran H, Guo M, Zhu J, et al. Nitrogen and phosphorus non-point source pollution in the upper Wujiang River Karst Basin: Critical source areas identification and influencing factors. Ecol Indic. 2025;170:112989. DOI: 10.1016/j.ecolind.2024.112989.

[15] Shenfang W, Haitao MA, Wei S. Water quality status and pollution prevention measures of rivers flowing into the sea via the Pearl River Estuary. Renmin Zhujiang. 2025;1-11. DOI: 10.3969/j.issn.1001-9235.2025.12.006.

[16] Li R, Wang Z, Huang M, Liao D, Yuan Z, Qian K. Environmental hygiene and healthcare-associated infection: A time-series study based on generalized additive model. Front Public Health. 2025;13:1592700. DOI: 10.3389/fpubh.2025.1592700.

[17] Melese D, Lemessa D, Abebe M, Hailegiorgis T, Nemomissa S. Modelling the distribution of Ekebergia capensis sparrm. (Meliaceae) under the current and future climate change scenarios in Ethiopia. BMC Ecol Evol. 2025;25(1):1-14. DOI: 10.1186/s12862-025-02437-0.

[18] Xie P, Yang Z, Liu F, Wu X. Drivers of short-term recovery in vegetation greenness and canopy height in burned areas of Southwest China. Environ Sustain Indic. 2025;28:100950. DOI: 10.1016/j.indic.2025.100950.

[19] Semedi B, Hardoko H, Dewi SNF, Fatimah A PDN. GAM (Generalized Additive Model) analysis for predicting potential area of Lemuru in Bali Strait. JFMR-J Fish Mar Res. 2022;6(1):33-40. DOI: 10.21776/ub.jfmr.2022.006.01.5.

[20] Nickels CF, Muhling BA, Portner EJ, Snodgrass OE, Dewar H. Ocean conditions drive interannual variability in juvenile albacore tuna (Thunnus alalunga) muscle energy content in the California Current System. PLoS ONE. 2025;20(9):e0331436. DOI: 10.1371/journal.pone.0331436.
[21] Han L, Sun Z, Zhang G, Zhang Y, Ren H, Teng Z, Current and future projection of scrub typhus risk related to land use change in China. GeoHealth. 2025;9(9):e2024GH001203. DOI: 10.1029/2024GH001203.
[22] Ravindra K, Singh T, Mor S. Emissions of air pollutants from primary crop residue burning in India and their mitigation strategies for cleaner emissions. J Clean Prod. 2019;208:261-73. DOI: 10.1016/j.jclepro.2018.10.031.
[23] Chang DY, Jeong S, Shin J. Assessing the impact of urbanization on forest carbon stocks and social costs using a machine learning approach. Sci Total Environ. 2024;954:176521. DOI: 10.1016/j.scitotenv.2024.176521.
[24] Sun J, Wang X, Liang M, Ren X, Liu X. Construction and analysis of China's carbon emission model based on machine learning. Sci Rep. 2025;15(1):13349. DOI: 10.1038/s41598-025-91222-6.
[25] Zhang Y, Zhang S, Gao S, Ning Z, Jiao Z, Hu Q. Data-driven machine learning quantifies ozone transport in the Hangzhou Bay urban cluster. Front Environ Sci Eng. 2025;19(12):169. DOI: 10.1007/s11783-025-2089-1.
[26] Li Z, Zhao B, Li S, Shi Z, Yin D, Wu Q, et al. Tracking county-level cooking emissions and their drivers in China from 1990 to 2021 with ensemble machine learning. Earth Syst Sci Data. 2025;17(10):5113-35. DOI: 10.5194/essd-17-5113-2025.
[27] Li Y, Gao C, Chen J, Wang Q, Zhao J. Spatial-temporal distribution characteristics of Harpadon nehereus in the Yangtze River Estuary and its relationship with environmental factors. Front Mar Sci. 2024;11:1340522. DOI: 10.3389/fmars.2024.1340522.
[28] Ji Y, Zhang X, Cao Y. Urban air quality shifts in china: application of additive model and transfer learning to major cities. Toxics. 2025;13(5):334. DOI: 10.3390/toxics13050334.
[29] Cheng L, Yan M, Zhang W, Guan W, Zhong L, Xu J. Interpretable digital soil organic matter mapping based on geographical gaussian process-generalized additive model (GGP-GAM). Agriculture. 2024;14(9):1578. DOI: 10.3390/agriculture14091578.
[30] Comber A, Harris P, Murakami D, Nakaya T, Tsutsumida N, Yoshida T, et al. Encapsulating spatially varying relationships with a generalized additive model. ISPRS Int J Geo-Inf. 2024;13(12):459. DOI: 10.3390/ijgi13120459.
[31] Lai J, Tang J, Li T, Zhang A, Mao L. Evaluating the relative importance of predictors in generalized additive models using the gam.hp R package. Plant Divers. 2024;46(4):542-6. DOI: 10.1016/j.pld.2024.06.002.
[32] Barton NA, Farewell TS, Hallett SH. Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. NPJ Clean Water. 2020;3(1):31. DOI: 10.1038/s41545-020-0077-3.
[33] Tao G, Chen H, Li W. Beijing PM2.5 influencing factors analysis based on GAM. In: Proceedings of the 2020 IEEE/WIC/ACM Int Joint Conf Web Intelligence Intelligent Agent Technol (WI-IAT). 2020. p. 916-21. DOI: 10.1109/WIIAT50758.2020.00139.
[34] Xu T, Zhang C, Liu C, Hu Q. Variability of PM2.5 and O_3 concentrations and their driving forces over Chinese megacities during 2018-2020. J Environ Sci. 2023;124:1-10. DOI: 10.1016/j.jes.2021.10.014.