Environmental Science

AIMS Environmental Science, 12(4): 721–743.
DOI: 10.3934/environsci.2025032
Received: 17 April 2025
Revised: 30 July 2025
Accepted: 31 July 2025
Published: 05 August 2025

https://www.aimspress.com/journal/environmental

Research article

The grey water footprint of the Yangtze River Economic Belt, China: Spatial patterns, driving mechanism, and implications

Guangming Yang[1,2,*], Darong Li[1,2], Yizhi Qin[1,2] and Hongxia Sheng[1,2]

[1] School of Management, Chongqing University of Technology, Chongqing 400054, China
[2] Rural Revitalization and Regional High-Quality Development Research Center, Chongqing University of Technology, Chongqing 400054, China

* **Correspondence:** Email: yangguangming@cqut.edu.cn; Tel.: +8617783473357.

Abstract: Deteriorating water ecosystems pose a critical constraint to high-quality development in the Yangtze River Economic Belt (YREB). In this study, we analyzed the grey water footprint (GWF) in the YREB from 2009 to 2022 using panel data from 11 provinces/municipalities. A multi-sector pollutant load model was employed to calculate the GWF. Driving mechanisms were investigated using Random Forest (RF) and SHapley Additive exPlanations (SHAP) analysis, while spatial dynamics were assessed via standard deviational ellipse. The results indicated: (1) The regional GWF peaked in 2015 (489.9 billion m^3), then decreased by 35.5% after 2016 due to the Yangtze River Protection initiative. Agriculture became the dominant contributor (53.1% share), especially in midstream provinces such as Hubei and Jiangxi. (2) Spatially, the GWF exhibited a northeast–southwest axial pattern. The ellipse area initially contracted to 80.22 km^2 (2018) before rebounding, while the distribution centroid shifted 105 km northwestward. (3) The RF–SHAP model revealed key non-linear drivers; the GWF significantly decreased when the secondary/tertiary industry share exceeded 70% or solid waste utilization surpassed 85%. Environmental investment showed diminishing marginal returns beyond 35% of GDP. Urban green space per capita had an optimal GWF-reduction range of 12–16 m^2; beyond 16 m^2, it increased the GWF due to ecological encroachment effects. We recommend targeted agricultural non-point source control, industrial circular economy upgrades, and region-specific ecological management to achieve sustainable development.

Keywords: grey water footprint; random forest model; Yangtze River Economic Belt; threshold; influencing factors

1. Introduction

The Yangtze River Economic Belt (YREB), a core engine for China's development, generates over 45% of the national GDP while occupying only 21% of the land area [1]. Its ecological conservation and green development hold national strategic significance. However, rapid industrialization and urbanization have pushed the water ecosystem to the brink of collapse. A significant proportion of rivers suffer from inferior Class V pollution, directly threatening regional sustainable development [2]. Sustainable water resource utilization, acting as a vital link coordinating regional economic, social, and ecological development [3], fundamentally relies on a deep understanding of water quality dynamics and their driving mechanisms. Therefore, systematically identifying the spatiotemporal patterns and underlying drivers of water pollution across sectors in the YREB is not only scientifically valuable but also a focal point for policymakers and society [4].

Accurate water quality quantification is essential for sustainable water management and scientific decision-making. While various assessment methods exist (e.g., fuzzy mathematics [5], grey system theory [6], artificial neural networks [7], and machine learning [8,9]), the grey water footprint (GWF) is widely adopted for its direct reflection of pollution loads from human activities on water bodies. Originally proposed by Hoekstra et al. (2011) [10], GWF is defined as the theoretical freshwater volume required to dilute specific pollutants to meet environmental quality standards [11,12]. It is recognized as an effective tool for measuring water pollution intensity from regional production and consumption [13,14], providing critical information for addressing water scarcity and formulating sustainable management strategies [15–17].

GWF research focuses on three pollution sources: Industrial [18], agricultural [19], and domestic [20]. In agriculture, point source pollution arises from irrigation return flows and livestock wastewater, while non-point source pollution is driven by runoff carrying fertilizers and pesticides [21]. Livestock farming significantly contributes through nutrient and waste discharges [22]. Industrial pollution mainly stems from point-source emissions of inadequately treated wastewater from manufacturing and chemical processing [23]. The service sector's GWF is primarily influenced by concentrated domestic wastewater discharges and diffuse urban runoff, which introduce nutrients, organic compounds, and other pollutants into water bodies [24]. However, research on driving mechanisms has limitations: First is the driving mechanism "Black Box". Furthermore, traditional methods such as regression analysis [25], factor decomposition [19], and Logarithmic Mean Division Index models [4] struggle to fully capture complex non-linear interactions among drivers [26], particularly lacking a systematic investigation into key driver threshold effects [25–27]. Second, there are blind spots in ecological cognition. Existing frameworks often concentrate on socio-economic-technical dimensions, neglecting the decisive role of natural ecosystem characteristics (e.g., water self-purification capacity and environmental carrying capacity) on GWF environmental impacts, leading to incomplete assessments. Third, a limitation of static space is that researchers ignore the upstream pollution migration caused by industrial transfer.

To address these limitations, we innovatively integrate the Random Forest (RF) algorithm with SHapley Additive exPlanations (SHAP) values. RF excels at handling high-dimensional, non-linear data, resisting overfitting, demonstrating robustness, and simultaneously providing variable importance and marginal effects [28–30], making it ideal for revealing the complex GWF driver network. SHAP quantitatively explains the contribution and dynamic changes of each driver across samples or contexts. Their combination not only identifies key drivers but also deeply analyzes their non-linear effects, interactions, and critical thresholds [29–31], enabling precise targeting of policy interventions. Furthermore, we construct a systematic "Technology-Social-Economic-Ecological" multidimensional analysis framework, incorporating natural ecosystem carrying capacity as a driver,

aiming for conclusions more closely aligned with YREB realities and enhanced reliability.

Despite the broad application of GWF globally, in-depth analysis of complex water pollution dynamics using advanced machine learning, particularly concerning threshold effects and multidimensional integration, remains insufficient. Thus, we aim to address water quality challenges and support regional sustainable development strategies. Against this backdrop, using YREB panel data (2009–2022), our core objectives are: (1) Quantification: Systematically calculate GWF for the YREB and its sub-units applying GWF theory and Hoekstra's methodology. (2) Spatiotemporal analysis: Reveal the spatiotemporal distribution patterns and evolution trends of GWF in the YREB. (3) Mechanism exploration: Employ the RF–SHAP integrated model to identify, interpret, and elucidate key drivers of GWF, their non-linear mechanisms (including threshold effects), and interactions, thereby proposing targeted water pollution control policy recommendations.

2. Materials and methods

2.1. Study area and data sources

2.1.1. Study area

The YREB spans nine provinces and two municipalities in China (97°21′E to 122°25′E, 21°08′N to 35°20′N) (see Figure 1). As a coordinated development belt, it effectively promotes interaction and cooperation between eastern, central, and western regions. It is an inland economic belt with global influence and a leading demonstration zone for ecological civilization. The YREB possesses abundant water resources, accounting for 35% of China's total [28], yet its per-capita water availability is only a quarter of the world's average [32]. Furthermore, it serves as the main arena for China's large-scale industrial development, contributing over 50% to national GDP while covering more than 21.4% of the country's land area. In 2023, industrial water consumption in the YREB reached 585.56 billion m³, exceeding 60% of the national total. Persistent high wastewater discharge has led to increasingly prominent ecological and environmental pollution issues requiring urgent resolution [33].

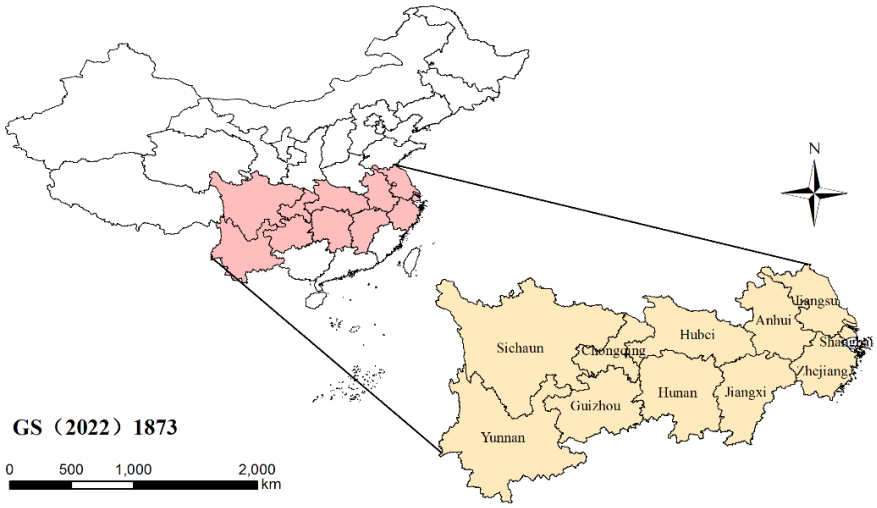

Figure 1. Location map of the study area (note: Based on the standard map of the Ministry of Natural Resources GS (2022) No. 1873, the boundaries of the base map have not been modified.

2.1.2. Data sources

Data from nine provinces and two municipalities within the YREB (2009–2022) were selected to calculate GWF for agricultural, industrial, and domestic sectors. The national average nitrogen leaching rate was used to represent the YREB agricultural nitrogen leaching rate. Nitrogen fertilizer application and wastewater discharge data came from the China Environmental Statistical Yearbook. Annual livestock and poultry breeding numbers were sourced from the China Rural Statistical Yearbook. Annual manure/urine output per animal, pollutant content per unit of manure/urine, and loss rates of manure pollutants into water bodies were obtained from the National Technical Report on Pollution Investigation of Large-Scale Livestock and Poultry Breeding. COD and NH_3-N emissions from industrial and domestic sectors were sourced from the China Environmental Statistical Yearbook. Data on GDP, total water resources, and total population were obtained from the China Statistical Yearbook. To address missing values in the dataset, particularly for industrial and domestic pollution emissions in certain years, interpolation using averages from adjacent years was applied to ensure consistency and reliability in the temporal analysis [34].

2.2. Methodology

2.2.1. Methodology framework

We employed an integrated GWF calculation model to assess the GWF within the YREB from 2009 to 2022. The GWF was calculated across three sectors: Agriculture, industry, and domestic use. This integrated model enabled the evaluation of the spatiotemporal evolution of the GWF, providing detailed insights into its trends and variations over time. Beyond GWF quantification, the RF model was employed to identify and analyze the driving factors behind GWF fluctuations. These drivers encompassed socioeconomic and environmental variables. Through annual contribution analysis, relative importance analysis, and partial dependence analysis, we investigated how these factors influenced GWF changes. Driver impacts were evaluated based on model performance metrics. The SHAP method was further applied to enhance the interpretability of RF predictions.

With the insights gained from this research, we aimed to provide a solid, data-driven foundation for developing targeted policies to optimize water resource management and strengthen environmental governance within the YREB.

2.2.2. Measurement of GWF

The calculation of the GWF for the eleven provinces and municipalities in the YREB strictly follows the guidelines outlined in the Water Footprint Assessment Manual [10]. This calculation covers three major categories of GWF: Agricultural, industrial, and domestic. Among these, the agricultural GWF includes crop farming and animal husbandry.

The GWF for each sector is calculated based on the pollutant load quotient and the concentration difference between the environmental standards and the actual measured values in the receiving water bodies.

$$GWF_k = max\frac{L_k}{C_{max} - C_{nat}} \tag{1}$$

where L_k represents the emission load of pollutants, C_{max} denotes the maximum allowable

concentration of pollutants specified by environmental quality standards, and C_{nat} refers to the natural background concentration of pollutants, which is assumed to be zero.

The reason is that, in accordance with the latest version of China's National Environmental Quality Standards for Surface Water (GB 3838–2002) [35], the Class III surface water discharge standard was selected as the corresponding C_{max} value. This level corresponds to the minimum water quality requirement, with a standard limit of 20 mg/L for COD and 1 mg/L for nitrogen and NH_3-N. Since 10 mg/L is a widely adopted threshold in agricultural GWF assessment, it was set as the maximum allowable nitrogen concentration for calculating agricultural GWF [36]. Moreover, in line with the content of the Water Footprint Assessment Manual, the natural background concentration C_{nat} was assumed to be zero [10].

2.2.3. Random forest model

The RF algorithm excels at handling non-linear problems and is used to evaluate the impacts of various variables. It is particularly notable for its effectiveness in determining variable importance, outperforming other machine learning techniques [28–30]. In this study, the sklearn module on the Python platform was employed for RF regression. After multiple rounds of debugging, the parameters were set, with n_estimators = 800, criterion = MSE, and the remaining parameters kept at their default values. This parameter configuration ensures optimal model performance, providing a sufficiently large training dataset for the RF model to capture complex relationships while maintaining reliable model validation.

In this study, the coefficient of determination (R^2) and mean squared error (MSE) were used as evaluation metrics for the accuracy of the regression model. R^2 measures the degree of fit between the predicted values and the observed values after regression, with a value range of [0, 1]. An R^2 value closer to 1 indicates a better fit of the obtained regression model. MSE is the average of the sum of squared residuals and is one of the most commonly used loss functions in regression; a smaller MSE indicates a smaller error of the obtained model. Additionally, variable importance is determined by the increase in MSE when the influence of a variable is randomized within the model [37]. This metric highlights the relative significance of each factor; a larger change in MSE indicates that the variable plays a key role in the model results, thereby providing substantial insights into the potential patterns of the dependent variable. The importance score of a given feature variable was calculated using the following formula:

$$VI = \frac{\sum_{i=1}^{N}(err2 - err1)_i}{N} \tag{2}$$

where err1 refers to the out-of-bag error calculated using the original feature values, err2 denotes the out-of-bag error of feature j after random permutation, and N represents the number of trees in the forest.

2.2.4. Standard deviation ellipse method

As a statistical method in geography, the standard deviation ellipse (SDE) can assess the directional attributes of the spatial distribution of GWF within a region. Moreover, it quantitatively illustrates changes in centrality, directionality, and geographical distribution patterns from a global perspective. The SDE enables intuitive observation of the centroid location, distribution direction, and variability of GWF and its temporal patterns [38]. This method serves as an effective tool for clarifying the geographical dynamics of GWF, contributing to the formulation of evidence-based policies for

improving water resources and the achievement of sustainable and environmentally friendly development goals. In this study, the SDE model was used to investigate the evolutionary trends of GWF.

The calculation process is as follows:

$$Gravity(\bar{X},\ \bar{Y}):\ \bar{X} = \frac{\sum_{i=1}^{m} q_i x_i}{\sum_{i=1}^{m} q_i}, \bar{Y} = \frac{\sum_{i=1}^{m} q_i y_i}{\sum_{i=1}^{m} q_i} \tag{3}$$

Azimuthθ:

$$tan\theta = \frac{(\sum_{i=1}^{m} q_i^2 \tilde{x}_i^2 - \sum_{i=1}^{m} q_i^2 \tilde{y}_i^2) + \sqrt{(\sum_{i=1}^{m} q_i^2 \tilde{x}_i^2 - \sum_{i=1}^{m} q_i^2 \tilde{y}_i^2)^2 + 4\sum_{i=1}^{m} q_i^2 \tilde{x}_i \tilde{y}_i}}{\sum_{i=1}^{m} 2q_i^2 \tilde{x}_i \tilde{y}_i} \tag{4}$$

Herein, $\tilde{x}_i$ and $\tilde{y}_i$ respectively represent the coordinate deviations of each province or municipality from the centroid.

$$\tilde{x}_i = x_i - \bar{X}, \tilde{y}_i = y_i - \bar{Y} \tag{5}$$

The standard deviations of the X-axis and Y-axis are σ_x and σ_y, respectively:

$$\sigma_x = \sqrt{(2\sum_{i=1}^{m} (q_i \tilde{x}_i cos\theta - q_i \tilde{y}_i sin\theta)^2)/(\sum_{i=1}^{m} q_i^2)} \tag{6}$$

$$\sigma_y = \sqrt{(2\sum_{i=1}^{m} (q_i \tilde{x}_i sin\theta + q_i \tilde{y}_i cos\theta)^2)/(\sum_{i=1}^{m} q_i^2)} \tag{7}$$

Ellipse Area S:

$$S = \pi \sigma_x \sigma_y \tag{8}$$

The calculation formula of spatial intensity I is as follows:

$$I = \frac{W}{S} \tag{9}$$

$(\bar{X},\ \bar{Y})$ represent the longitude and latitude coordinates of each city, q_i denotes the index value of the research object corresponding to each city, and W is the total index value of the research object within the study area.

2.2.5. SHapley Additive exPlanations

SHAP values elucidate the contribution of individual features to predictions generated by machine learning models [39]. Originating from cooperative game theory, SHAP quantifies each feature's impact on model output by accounting for both individual effects and interactions among all features [40]. This approach decomposes a model's prediction into the sum of contributions from each feature, delivering detailed, interpretable insights into how each feature influences the final prediction [41].

A key advantage of SHAP is its capacity to provide both global and local interpretability. It reveals not only a feature's overall importance across all predictions but also how it affects individual predictions. Unlike conventional feature importance metrics, which typically assess features in

isolation and ignore interactions, SHAP evaluates both major effects and interaction effects. This makes SHAP a more comprehensive and precise model interpretation tool, offering a clearer understanding of feature interdependencies.

The SHAP value for a specific feature j was computed using the Shapley formula from cooperative game theory:

$$\emptyset_i(\mathrm{V}) = \sum_{S \epsilon N} \frac{(|S|-1)!\,(n-|S|)}{n!}[V(S) - V(S\{i\}] \tag{10}$$

where N denotes the complete set of features ("players"), S represents a subset of features, $|S|$ is the number of features in subset S, n is the total number of features, and V(S) is the model's output using only the feature subset S. $S\{i\}$ represents the set obtained by removing the i-th participant from S. The result calculated using this equation is a vector V with n components, where each component represents the contribution of the corresponding participant.

The resulting SHAP value $\emptyset_i(\mathrm{V})$ is a vector where each component quantifies the direction (positive/negative) and magnitude of a feature's contribution to the final prediction [42]. This provides explicit visibility into how each feature drives the model's decision process. Consequently, SHAP is particularly valuable for interpreting complex models where feature interactions critically influence predictions, enhancing transparency and model trustworthiness.

2.2.6. Selection of driving factors

The spatiotemporal variation of the GWF is influenced by multiple factors. Guided by data availability and extensive literature support [27,43], we identified key drivers, including population dynamics [44], economic growth [19], technological progress [45], industrial configuration [46], and urban–rural structural transformation [47]. Other researchers have predominantly focused on socioeconomic drivers, largely overlooking the natural ecosystem's self-purification capacity and carrying capacity for water pollution, which are critical determinants of the GWF's environmental impact. Consequently, we introduced a new indicator system representing the regional ecological baseline and ecological construction level. Vegetation coverage (especially forests) effectively conserves water sources, filters pollutants, and improves water quality, while green spaces and protected areas safeguard critical catchment zones and ecologically sensitive regions.

The literature review further revealed that prior studies primarily emphasized the "pressure" from economic activities (e.g., GDP, industry) and partial resource consumption (water use volume) but inadequately characterized environmental governance intensity and effectiveness. Given the immense environmental pressure facing the YREB, where "prioritizing ecological protection and restraining excessive development" is the core strategy, analyzing its GWF necessitates incorporating environmental regulation intensity and the actual effectiveness of pollution control as core variables. These indicators are directly linked to the efficacy of policy interventions and the actual reduction of the GWF, serving as a crucial window for evaluating the success of the Yangtze River Protection initiative. Therefore, we added a new dimension: Environmental regulation and governance efficacy.

The specific indicators selected were as follows: Technology market transaction value as % of GDP (x1); R&D expenditure of large-scale industrial enterprises (10,000 CNY) (x2); urbanization rate (%) (x3); proportion of secondary and tertiary industries (%) (x4); GDP growth rate (%) (x5); sulfur dioxide (SO_2) emissions (tons) (x6); total industrial wastewater discharge (10,000 tons) (x7); comprehensive utilization rate of industrial solid waste (%) (x8); environmental protection investment

as % of GDP (%) (x9); built-up area green coverage rate (%) (x10); per-capita public green space area (m^2/person) (x11); forest coverage rate (%) (x12); protected area as % of total land area (%) (x13); total import and export volume (10,000 USD) (x14); and fiscal expenditure as % of GDP (%) (x15). The specific meanings of these variables are detailed in Table 1.

Table 1. The index meanings of GWF and driving variables.

Indicator	Specific meaning
GWF—grey water footprint (10^8 m^3)	Measures the degree of water pollution; the larger this value is, the more serious the water pollution is
Technology market transaction value as % of GDP（%）	Reflects the marketization level of technological achievements.
R&D expenditure of large-scale industrial enterprises (10^3 CNY)	Total R&D investment by large/medium industrial enterprises
Urbanization rate (%)	Urban permanent population as % of total population
Proportion of secondary and tertiary industries (%)	Share of non-primary industries in economic output
GDP growth rate (%)	Comparable price growth rate of the national economy
Sulfur dioxide (SO_2) emissions (tons)	Total emissions of key air pollutants
Total industrial wastewater discharge (10^3 tons)	Volume of wastewater discharged from industrial activities
Comprehensive utilization rate of industrial solid waste (%)	Resource recovery efficiency of industrial solid waste
Environmental protection investment as % of GDP (%)	Environmental governance input relative to economic size
Built-up area green coverage rate (%)	Vegetation coverage % in urban developed areas
Per-capita public green space area (m^2/person)	Average public green space per urban resident
Forest coverage rate (%)	Forest area as % of total land area
Protected area as % of total land area (%)	Proportion of protected natural areas
Total import and export volume (10,000 USD)	Gross value of cross-border trade (customs basis)
Fiscal expenditure as % of GDP (%)	Government expenditure relative to economic size

3. Results

3.1. Spatiotemporal patterns of the GWF

The GWF of the YREB reached its historical peak of 489.9 billion m^3 in 2015 (see Figure 2). Hubei, Hunan, and Sichuan provinces were the primary drivers, contributing 39.2% collectively, with agriculture dominating (>60% share in each), highlighting the ecological cost of food security strategies. A significant structural shift occurred in 2016 following the implementation of the Yangtze River Conservation Strategy. The total regional GWF decreased sharply by 35.5% to 316.0 billion m^3. The dramatic declines in Hubei (-45.9%) and Hunan (-21.8%) confirmed the policy's effectiveness. Notably, Yunnan exhibited an opposite trend (decline then surge). After a short-term decrease of 37.0% in 2016, its GWF rebounded to 42.97 billion m^3 by 2022, representing a 224% increase compared to 2019. This reveals a deep-seated conflict between upstream plateau agricultural expansion and basin management goals.

Integrated analysis reveals three distinct phases in YREB's GWF evolution. The extensive expansion

phase (2009–2013) was driven by rapid agricultural growth in the central and western regions. Jiangxi's agricultural GWF increased by 10.6%, Hubei maintained high levels (27.6–28.9 billion m^3), while industrial peaks in Anhui (6.44 billion m^3) and Chongqing (1.63 billion m^3) created a dual-pressure "agriculture-industry" system. The policy shock phase (2014–2016) exposed governance gaps. Hunan's agricultural GWF peaked in 2014 (29.7 billion m^3), and Sichuan's peaked in 2015 (35.2 billion m^3), indicating lagging non-point source pollution control. The total industrial GWF plummeted by 56.2% in 2016. The differing reduction magnitudes between Hubei (-84.5%) and Yunnan (-59.5%) revealed a gradient in policy implementation intensity. The divergent adjustment phase (2017–2022) highlighted regional imbalances. Jiangsu achieved deep industrial emission reductions (66.3%) through industrial upgrading. Conversely, the central and western regions faced transformation difficulties; Hunan's agricultural GWF accounted for 64.7% in 2022, reflecting slow progress in water-saving retrofits for rice-growing areas. Yunnan's unbalanced pattern (an agricultural surge of 274% and an industrial slow decline of 52%) intensified ecological unsustainability.

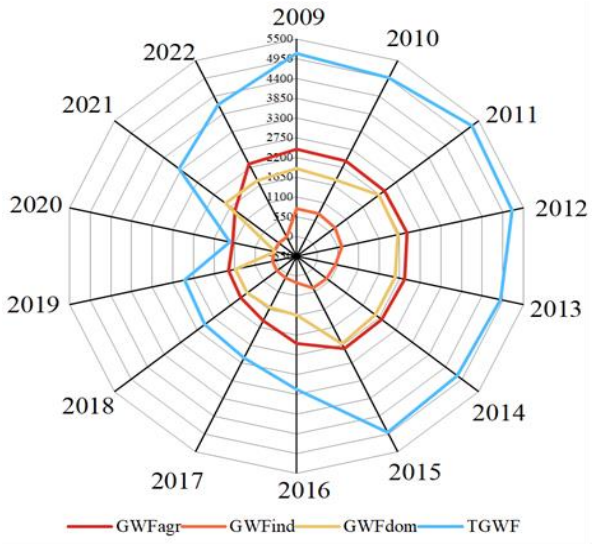

Figure 2. Dynamics and compositional characteristics of the total GWF (TGWF) in the YREB, 2009–2022 (unit: 10^8 m^3).

Sectoral analysis identifies three key conflicts (see in Figure 3). First is strengthening agricultural dominance. Agriculture's contribution to TGWF rose to 53.1% in 2022 (up 4.6% from 2009). High shares in Hunan (64.7%) and Jiangxi (50.6%) reflect persistent non-point source pollution control challenges in the middle reaches. Second are divergent industrial emission reduction paths. Eastern regions saw significant success via technology-driven approaches. Shanghai reduced industrial GWF by 86.3% over 14 years, while Jiangsu achieved a reduction of 89.2%. Central and western regions relied more on stringent policy enforcement; Hubei's drastic 84.5% drop in 2016 was linked to the "Ten Small Polluting Enterprises Elimination" campaign. Large fluctuations, such as Guizhou's 218% industrial GWF swing, indicate difficulties in transitioning resource-dependent economies. Third is the dual effect of domestic sources. Population growth increased domestic GWF in Anhui (+54.4%) and Sichuan (+9.3%), while Shanghai achieved an 85.7% reduction through advanced wastewater treatment upgrades. The sharp decline in domestic GWF in multiple provinces during 2020 (e.g., Hubei's -98.2%) may relate to exogenous shocks such as the pandemic, warranting vigilance for post-pandemic pollution rebound risks.

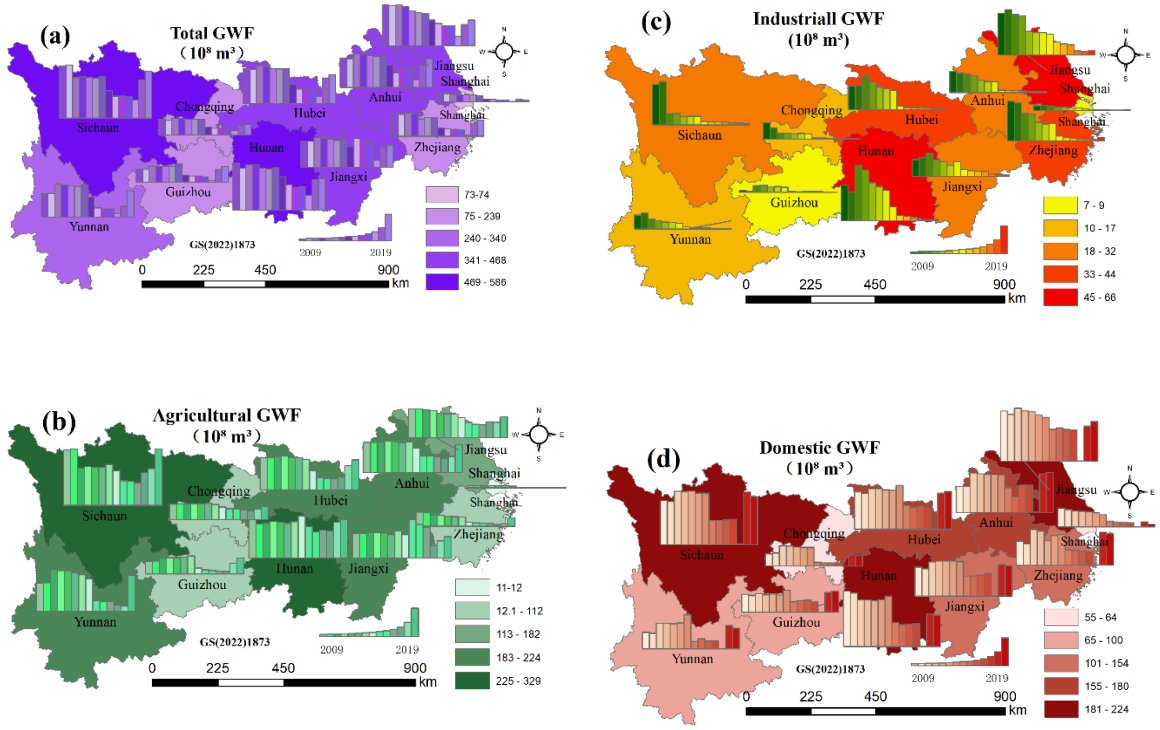

Figure 3. Spatiotemporal evolution and spatial distribution of GWF and its compositional structure in the YREB, 2009–2022 (note: Based on the standard map of the Ministry of Natural Resources GS (2022) No. 1873, the boundaries of the base map have not been modified).

3.2. Dynamic evolution trends of the GWF

The spatial distribution pattern of the GWF exhibited an overall contraction followed by a rebound. The area of the SDE decreased from 84.69 km^2 in 2009 to 80.22 km^2 in 2018, representing a 5.3% reduction (see Figure 4). This contraction signified increased spatial concentration of the GWF, indicating that pollution became more clustered. However, by 2022, the SDE area rebounded to 84.69 km^2, returning to the 2009 level. This rebound suggests a subsequent increase in pollution dispersion, potentially linked to regional industrial relocation or temporary relaxation of environmental policies. Concurrently, the orientation of the ellipse remained remarkably stable, with its rotation angle consistently maintained between 78° and 81° (northeast–southwest orientation). This stable axial alignment corresponds closely with the direction of the Yangtze River's main channel and the spatial layout of major urban clusters, such as the Chengdu–Chongqing and Middle Yangtze River agglomerations. Furthermore, the ellipse shape showed fluctuating contraction in 2018. The major axis shortened significantly (from 9.38 km to 8.72 km), while the minor axis experienced a slight increase (from 2.87 km to 2.93 km). This change reflects reduced pollution spread along the Yangtze River direction during this period, coupled with a marginal increase in spread perpendicular to it, likely reflecting enhanced industrial pollution control along the riverbanks under the Yangtze River Conservation Strategy.

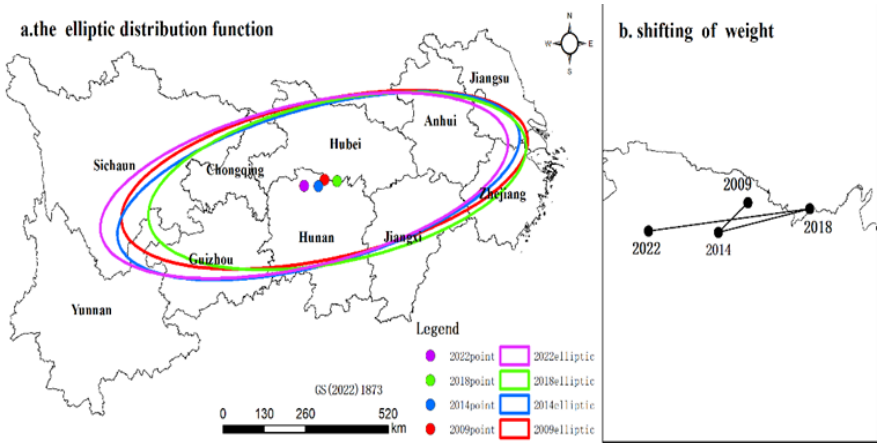

Figure 4. Spatial evolution trend of the GWF in the YREB.

The migration trajectory of the distribution center of gravity (CoG) reveals dynamic regional shifts, following a distinct zigzag path southwest, then northeast, and finally northwest (see Table 2). Between 2009 and 2014, the CoG shifted southwestward (from 111.86°E to 111.59°E longitude and from 29.68°N to 29.45°N latitude), covering an offset distance of approximately 27 km. This movement suggests an increased pollution load within the middle river reaches, particularly in Hunan and Jiangxi provinces. The period from 2014 to 2018 saw a significant northeastward jump of the CoG (to 112.44°E, 29.63°N), displacing it by about 88 km. This shift points directly toward the Wuhan metropolitan area in Hubei, likely associated with accelerated industrialization within the middle-reach urban agglomerations. Subsequently, between 2018 and 2022, the CoG migrated northwestward (to 110.94°E, 29.47°N), a substantial offset of roughly 140 km. The final position lies near the border area between Zhangjiajie in Hunan and Enshi in Hubei, reflecting a rising contribution to pollution from the transitional zone connecting the upper reaches (Chongqing, Sichuan) and the middle reaches.

Considering the period from 2009 to 2022, the CoG displayed a net displacement northwestward (see Table 3). The final longitude decreased by 0.93° and the final latitude decreased by 0.21°, resulting in a total offset distance of approximately 105 km. This net northwestward movement confirms the increasing weight and contribution of the upper Yangtze River region within the overall regional pollution pattern.

Table 2. Key parameters of the SDE for the TGWF in the YREB, 2009–2022.

Year	Shape_Length	Shape_Area	CenterX	CenterY	XStdDist	YStdDist	Rotation
2009	41.262672	84.687861	111.864228	29.676111	9.381412	2.874067	80.417332
2014	41.148223	84.045191	111.588374	29.454362	9.359225	2.859017	78.057259
2018	38.876945	80.221844	112.441705	29.629601	8.715302	2.930484	80.75369
2022	41.262672	84.687861	111.864228	29.676111	9.381412	2.859017	79.609225

Table 3. Key parameters of the distribution centroid for the TGWF in the YREB, 2009–2022.

Year	XCoord	YCoord	latitude	longitude
2009	111.864298	29.676163	29.676163N	111.864298E
2014	111.588435	29.454414	29.454414N	111.588435E
2018	112.441758	29.629646	29.629646N	112.441758E
2022	110.935137	29.466634	29.466634N	110.935137E

3.3. Decomposition of GWF driving factors

3.3.1. Applicability assessment of the RF model

We employed an RF model to analyze the drivers influencing the GWF, utilizing the previously mentioned 15 potential driving factors as predictors and GWF as the response variable. The dataset was split into a training set (70%) and a test set (30%). The model demonstrated excellent fitting performance on the training data, with an MSE of 1002.744, root mean squared error (RMSE) of 31.666, mean absolute error of 23.585, and an R^2 value as high as 0.975. This indicates the model successfully captured the vast majority of the variation in GWF within the training data, confirming the strong theoretical explanatory power of the selected feature combination. For the test set, the R^2 value was 0.843. Although this represents a decrease compared to the training set, an R^2 value of 0.843 signifies moderately strong explanatory power. These results confirm the fundamental effectiveness of the RF model in identifying and capturing key socio-economic and environmental drivers of GWF, establishing a reliable foundation for the subsequent use of SHAP models to delve into the specific contribution patterns and directions of each factor.

3.3.2. Relative importance of driving factors

Based on the feature importance calculated by the RF model (illustrated in Figure 5), the relative importance of each driving factor of GWF was quantified, revealing key drivers and their priorities. The dominant drivers were x6 (SO_2 emissions) and x7 (total industrial wastewater discharge). SO_2 emissions not only reflect air pollution levels but also indicate high-pollution industrial activity intensity through dry and wet deposition into water bodies, making it a core indicator directly linked to aquatic environmental pollution load. Total industrial wastewater discharge directly represents the primary volume of pollutant discharge entering the water environment. The high importance of these two factors strongly suggests that the current intensity of industrial pollution emissions, particularly SO_2 (linked to energy consumption) and industrial wastewater, is the most critical factor driving regional GWF increases.

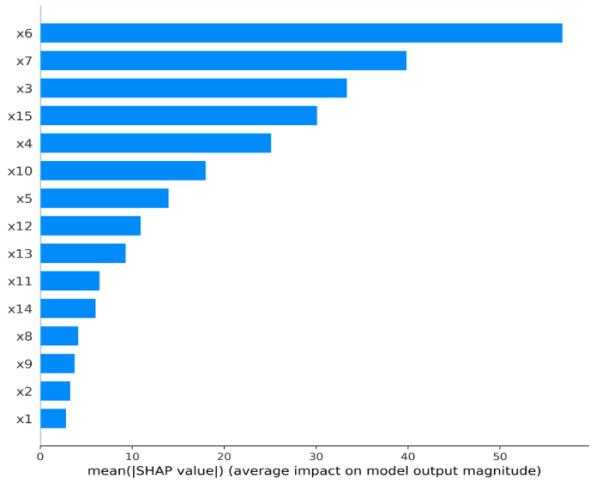

Figure 5. Ranking of the importance of each single factor feature.

Important influencing factors included x3 (urbanization rate), x15 (general public budget

revenue), x4 (proportion of secondary and tertiary industries), x10 (green coverage rate in built-up areas), and x5 (GDP growth rate). This combination highlights the multi-dimensional nature of GWF influences. Factors x3 and x4 jointly point to the stage and structure of economic development, where high urbanization rates and a large secondary/tertiary industry share typically accompany increased resource consumption and pollution pressure. The importance of x15 may reflect government regulatory capacity, suggesting the potential influence of public finance on environmental governance investment or industrial restructuring. Factor x10, representing urban ecological infrastructure, showed positive importance. This potentially stems from short-term negative effects associated with land-use changes during initial greening construction, or it may reflect the link between green coverage and urban expansion/population density. Factor x5 is directly linked to the expansion of overall economic activity, and its importance confirms that the rate of economic growth itself significantly drives GWF increases.

Medium and minor influencing factors included x12 (forest coverage rate), x13 (proportion of protected area), x11 (agricultural water consumption per unit area), x14 (total import and export volume), x8 (industrial solid waste comprehensive utilization rate), x9 (proportion of environmental protection investment), x2 (R&D expenditure of industrial enterprises), and x1 (proportion of technology market transaction value). Ecological protection factors such as forest coverage (x12) and protected area proportion (x13) showed medium to low importance. This indicates that while these ecological elements theoretically possess water purification functions, at the current regional scale, their direct contribution to explaining overall GWF variation is less core than pollution emission and economic structure factors. Their role likely leans more toward long-term ecological barrier functions or localized improvements. Environmental management and technology factors also showed low importance, including industrial solid waste utilization rate (x8), environmental protection investment proportion (x9), industrial enterprise R&D expenditure (x2), and technology market transaction proportion (x1). Notably, the relatively low importance of x8, representing circular economy principles, suggests that under the current analytical framework, solid waste recycling has a relatively limited direct impact on reducing water pollution loads; its emission reduction benefits may be more evident in solid waste disposal itself or associated media. The low influence of x9 may reflect issues with investment efficiency or direction (e.g., not being primarily targeted at water pollution control). Factors x2 and x1, ranking as the lowest, strongly reflect a failure of regional technological innovation activities, whether R&D investment or technology transactions, to effectively translate into significant pollution control technology applications that reduce GWF, revealing a distinct "technology transfer bottleneck". The low importance of economic openness in factor x14 (total import and export volume) suggests that, in this regional context, the environmental pressures (e.g., pollution transfer) or potentially higher standards (e.g., pollution halo effect) associated with foreign trade activities do not prominently influence the overall explanation of GWF.

In summary, industrial pollution emission intensity, particularly SO_2 and industrial wastewater, is the most direct and primary driver of the rising GWF. Key structural drivers shaping the GWF landscape include urbanization, industrial structure, economic growth rate, and government fiscal capacity, representing crucial socio-economic background factors. Environmental management measures (e.g., environmental investment and solid waste utilization) and technological innovation factors (e.g., R&D input and technology market activity) exhibit a relatively weak direct influence on GWF, highlighting deficiencies in end-of-pipe treatment efficiency and technology application, revealing governance and technological bottlenecks. Ecological factors such as forest coverage, protected areas, and urban green space play a secondary role in the model's explanatory power, suggesting their water purification functions may be constrained by spatial scale effects, time lags, or interactions with dominant factors.

Figure 6 visualizes the non-linear impact mechanisms of various indicators on the GWF prediction model using SHAP summary plots. This visualization's core value lies in its synergistic mapping of feature values (red for high, blue for low) and SHAP values (positive/negative direction), intuitively revealing the dynamic changes in impact intensity at different indicator levels and their inherent patterns. This provides profound insights into environmental effects that extend beyond traditional linear associations.

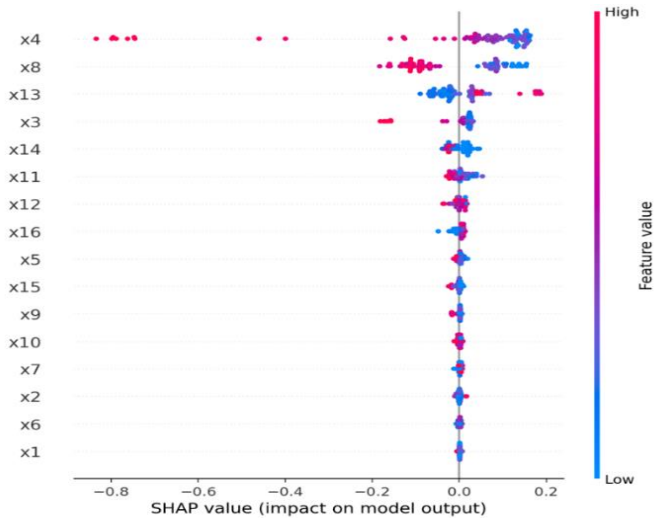

Figure 6. SHAP summary plot for feature analysis.

The analysis reveals distinct patterns. Pollution emission indicators (x6, SO_2; x7, industrial wastewater discharge) exhibit a clear and strong positive driving effect. Their high feature values are densely concentrated in the positive SHAP region, confirming that increased industrial pollution emissions are the most direct and critical pressure source elevating regional GWF in the current development stage. This not only validates the strong causal link between end-of-pipe pollutant inputs and water pollution load but also underscores the extreme urgency of strengthening pollution source reduction and whole-process control. In contrast, ecological construction indicators (x10, built-up area green coverage; x12, forest coverage rate) demonstrate significant environmental regulatory functions. Their high feature values consistently correspond to negative SHAP values, providing strong evidence for the positive benefits of green space and forest ecosystems in pollutant adsorption, purification, and water conservation. This finding offers crucial quantitative support for the "nature-based solutions" path of reducing GWF and enhancing environmental self-purification capacity by expanding and improving ecological spaces, such as increasing urban greening and protecting/restoring forests. Economic structure and development indicators reveal inherent tensions during the transition. A high value of x4 (proportion of secondary and tertiary industries) is associated with significant negative SHAP values, deeply illustrating that industrial upgrading toward high-value-added, low-pollution services and technology-intensive manufacturing industries (industrial structure advancement) can effectively reduce GWF intensity per-unit output by optimizing economic form and resource use efficiency. However, high values of x5 (GDP growth rate) are tightly linked to positive SHAP values, sharply reflecting that rapid economic growth amplifies environmental pressure if the development model fails to break free from dependence on traditional, extensive pathways.

Resource utilization efficiency and policy intervention indicators exhibit complex non-linear effects. The variation pattern in the SHAP values for x8 (solid waste recycling rate) suggests a potential

"scale threshold effect". When the resource utilization rate is low, its effectiveness in reducing GWF is limited. However, once it surpasses a critical level, its negative environmental impact (reducing GWF) strengthens significantly. This indicates that solid waste recycling must reach a substantial scale and operate as an integrated system to fully unlock its synergistic potential to alleviate water pollution loads. Notably, the behavior of x9 (environmental protection investment) is more revealing. At low investment levels, it exhibits a significant positive influence (increasing GWF), reflecting the risks of inadequate environmental governance. As investment increases, however, the expected negative inhibitory effect (reducing GWF) is relatively weak or even diminishes. This strongly suggests that environmental protection investment may face efficiency bottlenecks or structural inefficiencies. Simply increasing total investment is not an effective pathway; there is an urgent need to shift toward systematic optimization of investment targeting precision, technology suitability, and management efficiency. Furthermore, the influence of urbanization (x3) demonstrates a characteristic non-linear trajectory. During moderate development stages, infrastructure improvements and scale efficiencies may alleviate environmental pressure (corresponding to negative SHAP values). Conversely, during excessive expansion phases, resource and environmental overload coupled with concentrated pollution discharge can lead to a sharp increase in GWF pressure (corresponding to positive SHAP values). This pattern, to some extent, aligns with the theoretical expectations of the environmental Kuznets curve.

In summary, industrial pollution discharge is a clear core positive pressure source (increasing GWF), while ecological construction provides effective negative regulation (reducing GWF). Economic development presents a dual dynamic: The load-reducing effect of industrial structure upgrading coexists with the pressure-increasing inertia of extensive growth. Resource recycling faces a scale threshold, and environmental investment encounters efficiency bottlenecks. The urbanization process follows non-linear development laws.

3.3.3. Partial dependence analysis of driving factors

While feature summary plots reveal the overall correlation direction, partial dependence plots (PDPs) provide further insights by depicting the functional relationship between a single variable and the SHAP value (see in Figure 7), offering perspectives on the marginal impact of specific drivers on GWF while controlling for other variables [28–30]. This exposes critical policy sensitivity points and spatial heterogeneity, enriching our understanding of the complexity of water resource management in the YREB.

In technological innovation, the PDP reveals a complex "double-edged sword" effect. The dependence relationship for x2 (industrial enterprise R&D expenditure) shows significant "diminishing marginal returns of green technology innovation". Specifically, when R&D expenditure increases from 0.5×10^7 to 1×10^7 CNY, its contribution to reducing GWF (negative SHAP value) weakens from -10 to -7. However, once investment crosses the 2×10^7CNY threshold, the SHAP value rises to -2, indicating an 80% decay in marginal emission reduction efficiency. Quantitative analysis confirms that at low investment levels (below 2×10^7 CNY), each additional 10,000 CNY in R&D reduces the SHAP value by approximately 0.002 units, while at high levels (above 2×10^7 CNY), the marginal benefit plummets to 0.0004 units, confirming a saturation effect. In stark contrast, the PDP for x8 (Industrial Solid Waste Comprehensive Utilization Rate) exhibits a dramatic scale threshold transition effect at an 85% utilization rate. Below this threshold, SHAP values fluctuate erratically between [-5, 5], with minimal suppression efficiency per unit increase (≤ 0.5). Upon exceeding 85%, the SHAP value decreases linearly to -10, with the unit efficiency jumping to approximately 2.0. This abrupt change indicates that an 85% solid waste utilization rate is the critical threshold for forming an industrial metabolic loop and triggering a qualitative change in synergistic pollutant reduction.

The impact of ecological protection measures on GWF is highly non-linear and requires precise adaptation, with some results challenging conventional understanding. The dependence relationship for x12 (forest coverage rate) shows an unexpected positive correlation: As coverage increases from 10% to 60%, the SHAP value continuously rises from 10 to 60. The rate of increase notably accelerates from 0.5 units per percentage point to 1.5 units per percentage point once coverage surpasses 40%. This exposes the hidden cost of "ecological space competition", where excessive forest expansion may encroach on riparian buffer zones, compressing hydrological connectivity space and diminishing natural water purification functions. The PDP for x11 (per-capita park green space area) displays a typical bell-shaped distribution response, peaking in environmental benefit (SHAP ≈ -25) at 15 m^2, forming a strict optimal interval. In the insufficient green space zone (<12 m^2), SHAP values rise sharply from -20 to 0, reflecting the failure of ecological regulation functions. Within the optimal zone (14–16 m^2), SHAP values remain stably below -20, indicating peak water purification efficiency. However, in the overload zone (> 18 m^2), SHAP values surge to 35, signifying that negative effects from maintenance demands such as irrigation water consumption begin to outweigh the environmental benefits. The steep curvature of this U-shaped curve reveals that environmental benefit losses increase exponentially as the deviation from the optimal interval grows. The PDP for x10 (built–up area green coverage rate) exhibits a significant threshold response, with a clear inflection point at 40%. Below this threshold, SHAP values fluctuate violently between -40 and 20, indicating highly unstable ecosystem functions. Above 40%, SHAP values decrease steadily to -40, strongly validating the existence of a rigid scale threshold for ecosystems to exert stable purification functions.

Socio-economic structural transformation reveals critical breakpoint responses. The dependence plot for x4 (proportion of secondary and tertiary industries) shows a "cliff–like mutation" at 70%. Below this threshold, SHAP values fluctuate within [-20, 20], indicating that quantitative changes in industrial structure had not yet triggered a qualitative change in environmental performance. Once exceeding 70%, the SHAP value plummets sharply to -60 (a change magnitude of Δ= 80), conclusively proving that the structural shift toward a service-dominated economy triggers systemic pollution reduction effects. The evolution of x3 (urbanization rate) follows an "ecological window period" law. During the rapid expansion phase (30–50%), SHAP values drop from -40 to -80, reflecting the heavy resource and environmental costs of disordered development. Entering the optimization window period (50%–65%), SHAP values reach their trough (≈ -100), signifying the peak efficiency effect of green infrastructure. Upon entering the overload stage (> 70%), SHAP values rebound to 20, highlighting a resurgence of environmental pressure dominated by intensified urban heat island effects and water resource overexploitation. This distinct U-shaped curve identifies the 55%–65% urbanization range as the final opportunity window for implementing irreversible green transformation strategies; deviation from this window triggers significant ecosystem backlash.

Finally, pollution emissions and related economic policies demonstrate rigid constraints and structural traps. The PDP for x7 (total industrial wastewater discharge) exhibits a strong positive linear driving relationship, with SHAP values consistently high in the 50 to 100 range, confirming its role as a rigid core factor driving GWF growth. The U-shaped dependence of x14 (total import and export volume) exposes the "low-end foreign trade trap." Within the optimization zone (below 2×10^7 USD), SHAP values decrease from 0 to -20, reflecting the positive effect of international environmental standards forcing industrial upgrading. However, upon entering the risk zone (exceeding 5×10^7 USD), SHAP values surge to 30, revealing the environmental cost of trade models dominated by high-pollution transportation and low-end processing. The PDP for x9 (environmental protection investment as % of GDP) further reveals an "investment saturation effect". When the investment proportion exceeds 35%, SHAP values rebound significantly from -100 to -40, reflecting a clear diminishing marginal

environmental return on investment after existing technological carrying capacity reaches its limit.

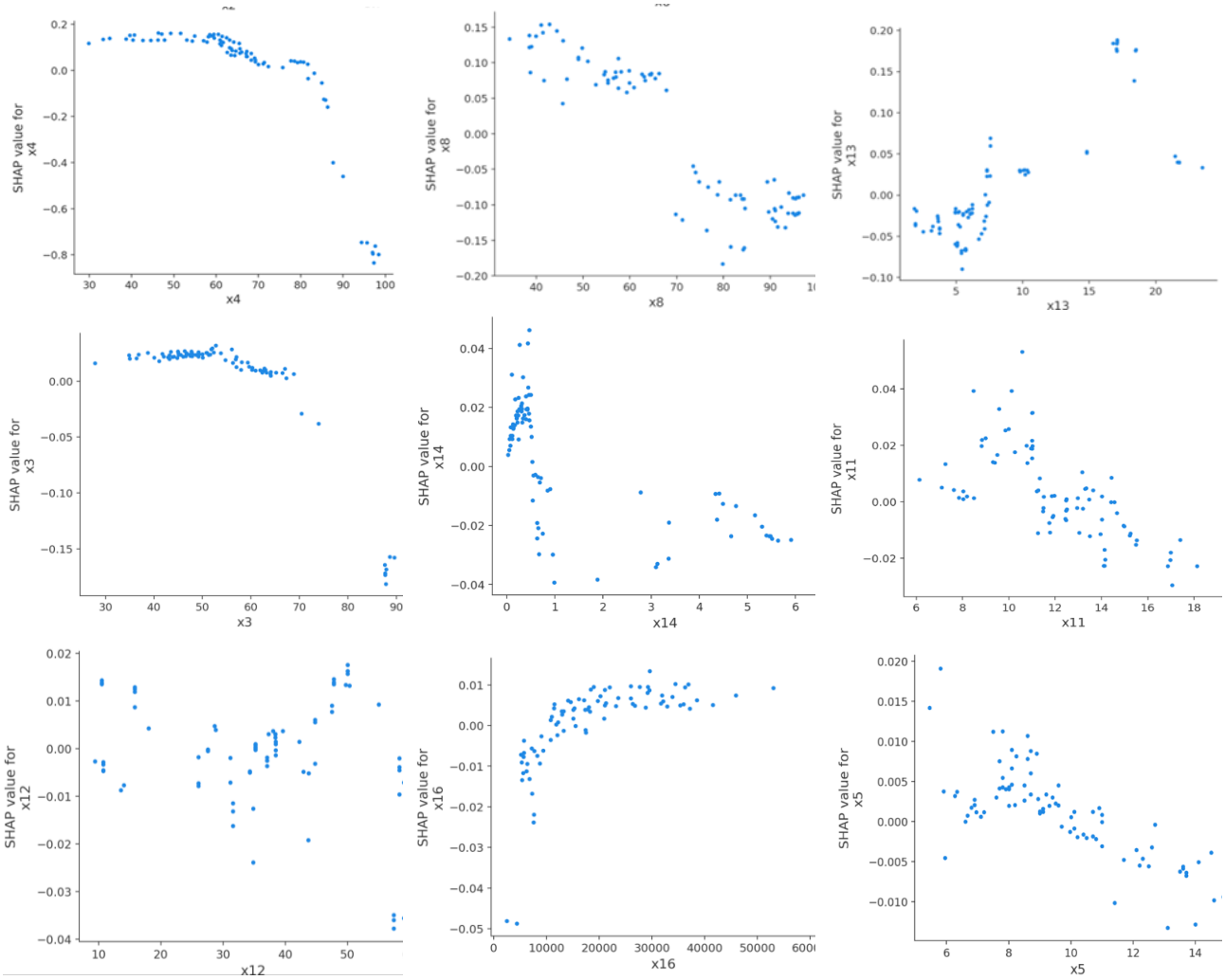

Figure 7. Partial dependency graph of the constructed RF model.

4. Discussion

4.1. Comparison with research

Accurately identifying pollution hotspots and their driving mechanisms is crucial for balancing economic development and ecological protection. Regarding the spatial pattern of GWF, studies by Li et al. (2021) and Zhang et al. (2019) using panel data from 31 Chinese provinces also found a lower GWF in the downstream YREB compared to the middle and upper reaches, aligning with our findings [43,44]. However, discrepancies exist concerning the dominant pollution source. While Zhang et al. (2019) identified domestic sewage as the primary contributor, our study, along with studies by Chai et al. (2022) and Li et al. (2021), confirms the absolute dominance of the agricultural sector. This difference mainly stems from varying accounting scopes [45]. Zhang et al. (2019) covered all Chinese provinces, including highly urbanized areas such as Beijing and Shanghai (where domestic sewage contribution is high). Their GWF calculation primarily considered pollutants such as COD and NH_3-N and excluded

livestock farming. Given the increasing prominence of livestock pollution, its inclusion in GWF accounting is essential. Regarding spatiotemporal evolution, Cui et al. (2020) calculated the highest TGWF in the southwest region. Both Cui et al. (2020) and Xu et al. (2022) identified the agricultural sector, specifically livestock and crop farming, as the core GWF contributor, consistent with our results [4,48]. Notably, most researchers focus on static spatial comparisons, such as Xu et al. (2022), who use Moran's I index to show high-efficiency "high-high" clustering in the Yangtze River Delta and "low-low" clustering upstream [4]. Our study complements this by revealing the dynamic spatial evolution of GWF patterns across the YREB using SDE and centroid migration models, addressing a gap in the literature.

Concerning driving factors, conclusions from Cui et al. (2020) and Xu et al. (2022) also identify industrial upgrading (particularly industrial efficiency gains) as the strongest driver for GWF reduction. They similarly emphasize circular economy technologies (e.g., solid waste/wastewater recycling) as vital pathways and note the limited effectiveness of pure end-of-pipe emission controls, consistent with our findings. However, differences exist regarding the role of technology. Xu et al. (2022) found technological upgrades effective in reducing GWF, whereas our study showed that the ratio of technology market transaction value to GDP (x1) had a weaker impact. This divergence likely arises from differing indicator definitions; our metric captures technology market activity, while Xu et al. focused more on the actual effects of technology application. Furthermore, existing research seldom quantifies the direct impact of ecological space on GWF. In contrast, we introduce ecological resilience indicators (x13, protected area ratio; x11, per-capita green space), providing empirical evidence for quantifying the potential inhibitory effects of nature-based solutions on GWF. Crucially, our use of the RF–SHAP model and PDPs revealed the non-linear threshold effects of drivers, moving beyond linear assumptions. For instance, we identify the critical point for qualitative industrial change; SHAP values plunge by 80 units when the secondary/tertiary industry share exceeds 70%, triggering systemic reductions, while below this threshold, fluctuations are high (±20 units). We also identify the circular economy inflection point at which GWF suppression efficiency surges 300% after solid waste utilization reaches 85%, confirming scale effects as a prerequisite for circular technologies. The ecological scale trap, evidenced by SHAP values turning positive when per-capita green space exceeds 16 m^2, provides a mechanistic explanation for Cui et al.'s (2020) observation of "decoupling between green space expansion and water quality improvement", attributing it to riparian zone encroachment and increased irrigation-related water consumption/pollution.

4.2. Limitations

Despite extensive exploration of the spatiotemporal variations in the YREB's GWF, some limitations remain. First, aggregation of city-level data may mask intra-city heterogeneity (e.g., differences between industrial clusters and rural areas), while the interpolation of missing values, though unavoidable, introduces additional uncertainty into temporal trends. Second, by focusing only on COD and nitrogen, the current framework overlooks emerging pollutants (e.g., microplastics, pharmaceuticals, and persistent organic pollutants), which are increasingly recognized as critical stressors on water quality. While we focus on the GWF as a key indicator of water pollution pressure, future research should integrate green and blue water footprints to fully assess the YREB's water resource sustainability.

To address these gaps, in future studies, researchers should utilize higher-resolution spatial data (district or watershed scale) combined with sub-annual temporal observations to more accurately

target local pollution hotspots. Incorporating new water quality indicators into GWF calculations will enable a more comprehensive assessment of evolving environmental pressures. These methodological improvements will enhance assessment precision and strengthen the empirical foundation for sustainable water resource management in rapidly urbanizing large regions.

5. Conclusions

We systematically analyzed the spatiotemporal evolution patterns and non-linear driving mechanisms of the GWF in the YREB from 2009 to 2022. By examining the dynamic contributions of individual driving factors to GWF, we identified key patterns and influences shaping the GWF landscape. The relationships between GWF and its drivers at varying thresholds were further explored, enabling tailored policy recommendations for different urban contexts. The major findings are summarized below.

(1) The evolution of GWF exhibited three distinct phases: Extensive expansion (2009–2013), policy-induced fluctuation (2014–2016), and differentiated adjustment (2017–2022). The 2015 peak (489.9 billion m^3) was primarily driven by agricultural non-point source pollution in central and western provinces (Hubei, Hunan, and Sichuan collectively accounting for 39.2%). Post-2016 policy interventions reduced the TGWF by 35.5%, but a concerning post-decline surge occurred in Yunnan (a 224% increase from 2019 to 2022), highlighting conflicts between upstream agricultural expansion and basin-wide governance. The dominance of the agricultural sector continued to strengthen, reaching 53.1% by 2022. Industrial emission reductions showed an east–west divergence, with eastern regions achieving deep cuts through technological upgrades (e.g., Shanghai reduced GWF by 86.3%), while central and western regions relied heavily on mandatory policy constraints.

(2) Spatially, GWF distribution displayed periodic fluctuations while maintaining a stable extension along the northeast–southwest axis (rotation angle: 78°–81°). The dispersion underwent "contraction-rebound" cycles. The SDE area decreased to 80.22 km^2 in 2018, indicating pollution concentration under the Yangtze River Conservation initiative. However, it rebounded to 84.69 km^2 by 2022 (equivalent to the 2009 level). This rebound, coupled with a 105 km northwest shift of the centroid (to 29.47°N, 110.94°E), signals pollution dispersion risks in upstream industrial relocation zones.

(3) The RF–SHAP model confirmed significant threshold effects for core drivers. A structural transformation threshold was identified where the secondary/tertiary industry share exceeding 70% caused SHAP values to plummet by 80 units, triggering systemic emission reductions. A circular economy inflection point occurred when solid waste utilization reached 85%; beyond this, the GWF suppression efficiency per unit increase jumped by 300% (SHAP slope increased from ≤0.5 to 2.0). An ecological scale trap emerged when per-capita green space exceeded 16 m^2, causing SHAP values to turn positive, revealing that excessive greening can harm water environments (irrigation water consumption/pollution outweighing purification benefits). Policy saturation effects were observed where marginal emission reduction benefits from R&D investment diminished by 80% after exceeding 200 million CNY. Efficiency significantly decreased when environmental investment exceeded 35% of total investment, reflecting technological conversion gaps and imbalanced investment structures.

To reduce the GWF and enhance water resource sustainability in the YREB, the following countermeasures and suggestions are proposed:

(1) Precision pollution control is needed for agricultural non-point sources. In high-GWF agricultural provinces, such as Hubei and Jiangxi, drip irrigation subsidies and nitrogen fertilizer

quotas need to be implemented. Moreover, cross-provincial compensation mechanisms for ecologically fragile upstream areas such as Yunnan need to be established to restrict the unregulated expansion of plateau-specific agriculture.

(2) Industrial circular economy upgrades require threshold-based management. Moreover, solid waste utilization standards for industries, such as chemicals and textiles, supported by VAT reductions, need to be mandated. Shanghai's "Clean Production Park" model in central and western regions need to be replicated by establishing solid waste recycling technology promotion centers to activate circular economy scale effects.

(3) Spatially differentiated governance strategies are essential. Thus, promoting AI-driven industrial water recycling technologies and increasing the service sector share in the east are needed. "Green Investment Access Zones" in central and western regions with green space scale limits (12–16 m^2 per capita) that require foreign-funded projects to maintain GWF per-unit output below the regional average must be established. Thereafter, implementing strict "Pollution Increment Replacement" and "Protected Area Expansion" plans upstream are needed to restore the ecological filtration function of riparian zones.

(4) To optimize the scale of ecological infrastructure. Aiming for 12–16 m^2 of per-capita green space in urban clusters and avoiding excessive greening that encroaches on natural water systems are needed. Moreover, it is crucial to conduct farmland-to-forest programs in provinces with forest coverage below 40%.

(5) To foster cross-regional coordination mechanisms. Establishing a YREB GWF trading market with cross-provincial compensation standards based on centroid migration trajectories is needed. Furthermore, it is crucial to create a basin-wide technology transfer alliance to address the R&D investment saturation effect.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors are grateful for the financial support provided by the National Social Science Foundation Post-Funded Project (24FGLB092); the Key Science and Technology Project of Chongqing Municipal Education Commission (2024CJZ015); the Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN202504022); the Higher Education Teaching Reform Research Project of Chongqing (233337); and the Planning Project of Chongqing Academy of Educational Science (K23YG2110387); Chongqing Graduate Research Innovation Project (Cys25800).

Guangming Yang: Supervision, writing-review and editing. Darong Li: Data curation, supervision, writing-original draft, writing-review and editing. Yizhi Qin: Supervision, visualization, writing-review and editing, methodology. Hongxia Sheng: Methodology, software. All authors have read and approved the final version of the manuscript for publication.

Conflict of interest

The authors declare no conflict of interest.

References

1. **S**hi C, Li L, Chiu YH, et al. (2022) Spatial differentiation of agricultural water resource utilization efficiency in the Yangtze River Economic Belt under changing environment. *J Clean Prod* 346: 131200. https://doi.org/10.1016/j.jclepro.2022.131200
2. Chen Y, Zhang S, Huang D, et al. (2017) The development of China's Yangtze River Economic Belt: how to make it in a green way? *Sci Bull* 62: 648–651. http://dx.doi.org/10.1016/j.scib.2017.04.009
3. Graymore MLM, Sipe NG, Rickson RE (2010) Sustaining human carrying capacity: A tool for regional sustainability assessment. *Ecol Eco*n 69: 459–468. https://doi.org/10.1016/j.ecolecon.2009.08.016
4. Xu C, Liu Y, Fu T (2022) Spatial-temporal evolution and driving factors of grey water footprint efficiency in the Yangtze River Economic Belt. *Sci Total Environ* 844: 156930. https://doi.org/10.1016/j.scitotenv.2022.156930
5. Wang L, Wang X (2022) A holistic assessment of spatio-temporal pattern and water quality in the typical basin of northeast China using multivariate statistical methods. *Process Saf Environ* 168: 1009–1018. https://doi.org/10.1016/j.psep.2022.10.079
6. Long X, Wu S, Wang J, et al. (2022) Urban water environment carrying capacity based on VPOSR-coefficient of variation-grey correlation model: a case of Beijing, China. *Ecol Indic* 138: 108863. https://doi.org/10.1016/j.ecolind.2022.108863
7. Zhao X, Liu X, Xing Y, et al. (2022) Evaluation of water quality using a Takagi-Sugeno fuzzy neural network and determination of heavy metal pollution index in a typical site upstream of the Yellow River. *Environ Res* 211: 113058. https://doi.org/10.1016/j.envres.2022.113058
8. Mohammadpour A, Gharehchahi E, Golaki M, et al. (2025) Advanced water quality assessment using machine learning: Source identification and probabilistic health risk analysis. *Results Eng* 27: 105421. https://doi.org/10.1016/j.rineng.2025.105421
9. Jin S, Zhang K, Zhang T, et al. (2025) Establishment of an innovative machine learning-driven drinking water quality assessment model with health considerations. *J Environ Sci* in press. https://doi.org/10.1016/j.jes.2025.03.019
10. Hoekstra AY, Chapagain AK, Mekonnen MM, et al. (2011) *The water footprint assessment manual: Setting the global standard*, London: Earthscan.
11. Franke NA, Boyacioglu H, Hoekstra AY (2013) *Grey water footprint accounting: Tier 1 supporting guidelines*, Delft: Unesco-IHE Institute for Water Education.
12. Tuninetti M, Tamea S, D'Odorico P, et al. (2015) Global sensitivity of high-resolution estimates of crop water footprint. *Water Resour Res* 51: 8257–8272. https://doi.org/10.1002/2015wr017148
13. Corredor JAG, González GLV, Granados MV, et al. (2021) Use of the gray water footprint as an indicator of contamination caused by artisanal mining in Colombia. *Resour Policy* 73: 102197. https://doi.org/10.1016/j.resourpol.2021.102197
14. Ansorge L, Stejskalová L, Dlabal J (2020) Grey water footprint as a tool for implementing the water framework directive-Temelín nuclear power station. *J Clean Prod* 263: 121541. https://doi.org/10.1016/j.jclepro.2020.121541
15. Jamshidi S, Imani S, Delavar M (2022) An approach to quantifying the grey water footprint of agricultural productions in basins with impaired environment. *J Hydrol* 606: 127458. https://doi.org/10.1016/j.jhydrol.2022.127458

16. Rong Q, Wu H, Otkur A, et al. (2023) A novel uncertainty analysis method to improve the accuracy of agricultural grey water footprint evaluation considering the influence of production conditions. *Ecol Indic* 154: 110641. https://doi.org/10.1016/j.ecolind.2023.110641
17. Liu B, He Y, Tan Q, Zhang Y (2025) The grey water footprint of the Guangdong-Hong Kong-Macao Greater Bay Area, China: Spatial patterns, driving mechanism and implications. *J Environ Manage* 389: 126063. https://doi.org/10.1016/j.jenvman.2025.126063
18. Huang Y, Han R, Qi J, et al. (2022) Health risks of industrial wastewater heavy metals based on improved grey water footprint model. *J Clean Prod* 377: 134472. https://doi.org/10.1016/j.jclepro.2022.134472
19. Kong Y, He W, Zhang Z, et al. (2022) Spatial-temporal variation and driving factors decomposition of agricultural grey water footprint in China. *J Environ Manage* 318: 115601. https://doi.org/10.1016/j.jenvman.2022.115601
20. Dong H, Zhang L, Geng Y, et al. (2021) New insights from grey water footprint assessment: An industrial park level. *J Clean Prod* 285: 124915. https://doi.org/10.1016/j.jclepro.2020.124915
21. Wang M, Jiang T, Mao Y, et al. (2023) Current situation of agricultural non-point source pollution and its control. *Water Air Soil Poll* 234: 471. https://doi.org/10.1007/s11270-023-06462-x
22. Xu Y, Ma T, CNY Z, et al. (2023) Spatial patterns in pollution discharges from livestock and poultry farm and the linkage between manure nutrients load and the carrying capacity of croplands in China. *Sci Total Environ* 901: 166006. https://doi.org/10.1016/j.scitotenv.2023.166006
23. Tariq A, Mushtaq A (2023) Untreated wastewater reasons and causes: a review of most affected areas and cities. *Int J Chem Biochem Sci* 23: 121–143.
24. Feng H, Schyns JF, Krol MS, et al. (2024) Water pollution scenarios and response options for China. *Sci Total Environ* 914: 169807. https://doi.org/10.1016/j.scitotenv.2023.169807
25. Fan Y, Fang C (2020) A comprehensive insight into water pollution and driving forces in Western China—case study of Qinghai. *J Clean Prod* 274: 123950. https://doi.org/10.1016/j.jclepro.2020.123950
26. He W, Zhang K, Kong Y, et al. (2023) Reduction pathways identification of agricultural water pollution in Hubei Province, China. *Ecol Indic* 153: 110464. https://doi.org/10.1016/j.ecolind.2023.110464
27. Feng H, Sun F, Liu Y, et al. (2021) Mapping multiple water pollutants across China using the grey water footprint. *Sci Total Environ* 785: 147255. https://doi.org/10.1016/j.scitotenv.2021.147255
28. Wang C, Wang X, Zhang G, et al. (2023) Identification of critical effect factors for prediction of spatial and intra-annual variability of shallow groundwater nitrate in agricultural areas. *Sci Total Environ* 891: 164342. https://doi.org/10.1016/j.scitotenv.2023.164342
29. Yang R, Meng J (2022) Using advanced machine-learning algorithms to estimate the site index of Masson pine plantations. *Forests* 13: 1976. https://doi.org/10.3390/f13121976
30. Liu M, Hu S, Ge Y, et al. (2021) Using multiple linear regression and random forests to identify spatial poverty determinants in rural China. *Spatial Stat* 42: 100461. https://doi.org/10.1016/j.spasta.2020.100461
31. Zhu X, Li Y, Wang X (2019) Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. *Bioresource Technol* 288: 121527. https://doi.org/10.1016/j.biortech.2019.121527
32. Jing P, Sheng J, Hu T, et al. (2022) Spatiotemporal evolution of sustainable utilization of water resources in the Yangtze River Economic Belt based on an integrated water ecological footprint model. *J Clean Prod* 358: 132035. https://doi.org/10.1016/j.jclepro.2022.132035

33. Yang G, Gui Q, Liu J, et al. (2024) Spatiotemporal evolution characteristics and influencing factors of energy-ecology-economy complex system efficiency: Case study of Yangtze River Economic Belt in China. *Energy* 312: 133526. https://doi.org/10.1016/j.energy.2024.133526
34. Xiang Y, Shao W, Wang S, et al. (2022) Study on regional differences and convergence of green development efficiency of the chemical industry in the Yangtzeriver economic Belt based on grey water footprint. *Int J Environ Res* 19: 1703. https://doi.org/10.3390/ijerph19031703
35. Ministry of Environmental Protection of China (2002) *Environmental quality standard for surface water (GB3838-2002)*, Beijing: China Environmental Science Press.
36. Liu W, Antonelli M, Liu X, et al. (2017) Towards improvement of grey water footprint assessment: With an illustration for global maize cultivation. *J Clean Prod* 147: 1–9. https://doi.org/10.1016/j.jclepro.2017.01.072
37. Breiman L (2001) Random forests. *Mach Learn* 45: 5–32. https://doi.org/10.1023/A:1010933404324
38. Zeng G, Guo Y, Nie S, et al. (2024) Spatial characteristic of carbon emission intensity under "dual carbon" targets: evidence from China. *Global Nest J* 26: 06164. https://doi.org/10.55555/gnj.006164
39. Kanani-Sadat Y, Safari A, Nasseri M, et al. (2024) A novel explainable PSO-XGBoost model for regional flood frequency analysis at a national scale: exploring spatial heterogeneity in flood drivers. *J Hydrol* 638: 131493. https://doi.org/10.1016/j.jhydrol.2024.131493
40. Lundberg SM, Lee SI (2017) *A unified approach to interpreting model predictions*, In: Proc 31st Int Conf Neural Inf Process Syst (NIPS'17), Red Hook, NY: Curran Associates Inc, 4768–4777. https://doi.org/10.48550/arXiv.1705.07874
41. Aldrees A, Khan M, Taha ATB, et al. (2024) Evaluation of water quality indexes with novel machine learning and SHapley Additive Explanation (SHAP) approaches. *J Water Process Eng* 58: 104789. https://doi.org/10.1016/j.jwpe.2024.104789
42. Cui H, Li J, Sun Y, et al. (2024) A novel framework for quantitative attribution of particulate matter pollution mitigation to natural and socioeconomic drivers. *Sci Total Environ* 926: 171910. https://doi.org/10.1016/j.scitotenv.2024.171910
43. Li H, Liang S, Liang Y, et al. (2021) Multipollutant based grey water footprint of chinese regions. Resour Conserv Recycl 164: 105202. https://doi.org/10.1016/j.resconrec.2020.105202
44. Zhang L, Dong H, Geng Y, et al. (2019) China's provincial grey water footprint characteristic and driving forces. *Sci Total Environ* 677: 427–435. https://doi.org/10.1016/j.scitotenv.2019.04.318
45. Chai M, Chen Y (2022) Spatio-temporal variations and driving factors of greywater footprint in the Yangtze River Economic Belt, China. *Pol J Environ Stud* 31: 1577–1586. https://doi.org/10.15244/pjoes/143247
46. Yang G, Cheng S, Huang X, et al. (2024) What were the spatiotemporal evolution characteristics and influencing factors of global land use carbon emission efficiency? A case study of the 136 countries. *Ecol Indic* 166: 112233. https://doi.org/10.1016/j.ecolind.2024.112233
47. Liao X, Chai L, Liang Y (2021) Income impacts on household consumption's grey water footprint in China. *Sci Total Environ* 755: 142584. https://doi.org/10.1016/j.scitotenv.2020.142584
48. Cui S, Dong H, Wilson J (2020) Grey water footprint evaluation and driving force analysis of eight economic regions in China. *Environ Sci Pollut R* 27: 20380–20391. https://doi.org/10.1007/s11356-020-08450-8

AIMS Press