

Article

Revealing Spatiotemporal Evolution and Driving Mechanisms of Grey Water Footprint in Land Consolidation Areas Using Explainable Machine Learning Models: Evidence from Yan'an Region, Shaanxi Province

Qiaoyang Yang^{1,2,3}, Hui Qian^{1,2,3,*} , Qi Long^{1,2,3}, Yicheng Duan^{1,2,3} and Zhiming Cao^{1,2,3}

¹ School of Water and Environment, Chang'an University, Xi'an 710054, China; 2020029004@chd.edu.cn (Q.Y.); 2023029004@chd.edu.cn (Q.L.); 2023029005@chd.edu.cn (Y.D.); 2024029003@chd.edu.cn (Z.C.)

² Key Laboratory of Subsurface Hydrology and Ecological Effect in Arid Region of the Ministry of Education, Chang'an University, Xi'an 710054, China

³ Key Laboratory of Eco-Hydrology and Water Security in Arid and Semi-Arid Regions of Ministry of Water Resources, Chang'an University, Xi'an 710054, China

* Correspondence: qianhui@chd.edu.cn

Abstract

The grey water footprint (GWF) is a critical indicator for assessing the impact of socio-economic activities on the water resources environment. To address the dual challenges of economic growth and water pollution associated with Land Consolidation Projects (LCPs) in the Loess Plateau, this study systematically analyzes the spatiotemporal distribution of GWF in the Yan'an region from 2000 to 2023 and employs the eXtreme Gradient Boosting (XGBoost) model to comprehensively explore its driving mechanisms. The SHapley Additive Explanations (SHAP) method was employed to quantify the dynamic contributions of the driving factors of GWF, while the threshold effects of these factors were assessed using partial dependence plot analysis. Additionally, spatial matching patterns between agricultural GWF (GWF_{agr}) and economic factors were examined using the Gini coefficient and imbalance index. These findings indicate that the total GWF (TGWF) peaked at 1.347 billion m³ in 2004 and declined due to improvements in water management efficiency. Spatially, TGWF is higher in the central and eastern regions, where GWF_{agr} is predominant. The permanent population and per capita GDP are the key driving factors, accounting for 21.1% and 15% of the total change in TGWF, respectively. In the spatial coupling relationship between agricultural GDP and GWF_{agr}, the overall imbalance index has significantly decreased. The synergistic effect between the Grain for Green Project and LCPs is becoming increasingly evident. These insights provide scientific support and policy guidance for the ecological protection and high-quality development of the Yellow River Basin.

Keywords: GWF; spatial matching pattern; machine learning; SHapley Additive Explanations; land consolidation area



Received: 30 December 2025

Revised: 8 February 2026

Accepted: 9 February 2026

Published: 11 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Water resources are fundamental and strategic assets essential to human development and the stability of ecosystems [1–4]. However, water scarcity and worsening pollution have emerged as key constraints to ecological protection and high-quality development in the Yellow River Basin (YRB) [5,6]. The YRB accounts for approximately 6% of the nation's wastewater discharge and 7% of its COD emissions [7]. Wastewater discharge represents

about 8% of the basin's current runoff and has shown a continuous upward trend over the past 30 years [8]. The sharp contradiction between the pollutant-carrying capacity of water resources and the water demands of socio-economic development further intensifies the pressure on the basin's water resources [9,10].

Accurate quantification of water quality is essential for sustainable water resource use, effective environmental management, and evidence-based decision-making. In the field of water quality assessment and pollution quantification, researchers have approached the topic from multiple perspectives [11,12]. To provide a more precise assessment of water pollution, Hoekstra introduced the concept of the "Grey Water Footprint" (GWF) [13], which quantifies the volume of freshwater needed to dilute pollutants to natural background concentrations or to meet specified water quality standards. This approach effectively reflects the real impact of pollutants on the environment and has become a vital tool for water quality management [14,15], particularly in densely populated and water-stressed regions such as the YRB. Chen et al. (2021) applied the GWF and the Logarithmic mean division index (LMDI) method to assess the water pollution status in the YRB irrigation area [14]. Kong et al. (2022) employed the factor decomposition standard deviation ellipse and the generalized divisor index method to analyze the decoupling effects of China's agricultural GWF [15]. Zhang et al. (2019) systematically analyzed the spatiotemporal characteristics and driving mechanisms of China's provincial-scale GWF and proposed targeted policy recommendations [16]. These findings provide valuable insights for addressing the global water resource challenges that China is likely to face in the future [16]. According to 2023 statistics from the Yellow River Conservancy Commission of the Ministry of Water Resources, agricultural water consumption in the basin reached 25.347 billion m³, representing 64.6% of the total water use, followed by industrial consumption of 4.428 billion m³ (11.3%) (<http://fgw.shandong.gov.cn>, accessed on 14 March 2023). This clearly indicates an agricultural-dominated water use structure. Hence, this study analyzes water pollution from agricultural, industrial, and domestic sources, emphasizing its spatial-temporal variation and the mechanisms driving its evolution. These insights offer an evidence-based foundation for decision-making aimed at mitigating water scarcity, implementing targeted pollution control, and promoting long-term sustainable development in the YRB.

Identifying the key drivers of regional water pollution is essential for developing targeted and effective water resource management strategies [7]. Although conventional approaches such as regression analysis and factor decomposition have advanced our understanding of GWF [14,17,18], their inherent linearity assumptions and insufficient consideration of variable interactions constrain their effectiveness in elucidating nonlinear driving mechanisms [19]. By contrast, machine learning (ML) models demonstrate superior capability in handling nonlinear relationships and complex interdependencies [20], offering a more effective approach for identifying and quantifying the key drivers of GWF. Nonetheless, increasing model complexity can compromise the interpretability of feature variable–outcome relationships, resulting in "black box" issues [21,22]. To overcome this limitation, this study integrates ML models with the SHapley Additive Explanations (SHAP) method, grounded in cooperative game theory. Leveraging Shapley values, the approach facilitates an accurate decomposition of the GWF into contributions from individual driving factors, enhancing the interpretability and transparency of ML predictions [22,23]. For instance, Liu et al. (2025) focused on the Guangdong–Hong Kong–Macao Greater Bay Area, systematically analyzing the spatial distribution of the grey water footprint in the region [17]. By integrating random forest and interpretable ML methods, he revealed its driving mechanisms and proposed targeted practical implications [17]. Huang et al. conducted an identification and attribution analysis of the comprehensive ecological zone in Chengdu, China, using the XGBoost–SHAP model [21]. Liu et al. (2025) employed the

PMF–ML–SHAP framework combined with a transmission model to analyze the pollution characteristics and causes of PM_{2.5} at different pollution levels in Linyi City, China [20]. Accordingly, the ML–SHAP integrated framework facilitates a rigorous elucidation of the driving mechanisms underlying water pollution in the YRB, delivering both scientifically sound and operationally applicable insights for governance.

The global shortage of water resources and the increasing severity of water pollution are critical scientific challenges that hinder regional sustainable development [1,2]. The eutrophication of water bodies and the degradation of aquatic ecosystems due to agricultural non-point source pollution have become key research areas [4]. The Loess Plateau is a typical area in China where water scarcity and pollution overlap, which makes the management of its aquatic ecosystems a crucial component of the national ecological security strategy. Yan'an, located in the heart of the hilly and gully region of the Loess Plateau along the middle reaches of the Yellow River, combines both energy and agricultural production attributes, which positions it as a key area for ecological protection within the YRB. Its water resource and pollution characteristics serve as representative examples for the region.

To mitigate soil erosion and advance sustainable development, the region pioneered the implementation of the Grain for Green Project (GFGP) in the late 1990s. Nonetheless, the dual impact of vegetation restoration and the expansion of economically utilized land has resulted in decreased cropland and a decreased watershed area, further intensifying the imbalance between water and soil resource availability and demand in the Loess Plateau [24]. Since 2000, Yan'an has implemented a comprehensive ecological restoration initiative to confront these challenges and foster sustainable development within the YRB [25]. Specific interventions include slope management, integrated gully and slope management, small watershed management, GFGP [26], the Campaign to Bulldoze Mountains to Build Cities Project [27], and the Gully Land Consolidation Project (GLCP) [17,28]. These interventions have produced notable outcomes: the ecological environment of the Loess Plateau has steadily improved, vegetation coverage has increased, sediment concentration and discharge in the YRB have declined, soil erosion has been effectively controlled, and the expansion of cropland has helped alleviate food scarcity [27]. However, ecological restoration and land consolidation have also generated new environmental pressures [29,30]. On the one hand, although the expansion of cropland has helped alleviate the population–land–food conflict, the intensification of agricultural activities has led to a marked increase in ammonia nitrogen emissions [17,31,32]. On the other hand, intensive engineering projects have further aggravated water pollution, generating a renewed imbalance between ecological governance and environmental pressures [32]. Although the GWF method has been widely applied both domestically and internationally [33–36], studies specifically addressing land consolidation engineering areas remain limited. Consequently, this has led to a gap in understanding the dynamics of water pollution in the region before and after the governance project. Addressing this gap can provide both practical guidance for efficient water resource management in future LCPs and a scientific basis for reconciling economic development with ecological sustainability. Additionally, as an agriculture-dominated region, agricultural non-point source pollution has intensified with the expansion of cropland [34]. Currently, research on spatial matching patterns primarily focuses on blue and green water footprints in relation to cropland resources, economic factors, and land elements. However, less attention has been given to the spatial mismatch between agricultural GWF (GWF_{agr}) and economic resource factors. Therefore, it is crucial to study the spatial matching pattern between GWF_{agr} and economic resource factors from the perspective of economy–ecology–resources. This will not only advance theoretical and practical progress in this field but also offer targeted recommendations for promoting green agricultural development in LCP areas.

Against this backdrop, this study conducted a comprehensive investigation of the GWF in the Yan'an region from 2000 to 2023. The objectives are as follows: (1) assessing the GWF in Yan'an using GWF theory and Hoekstra's calculation methods; (2) examining the temporal evolution of the GWF and its spatial distribution characteristics; (3) exploring the primary drivers of the GWF using the ML-SHAP models; (4) applying the Gini coefficient and imbalance index to assess the spatial matching pattern between GWF_{agr} and economic resource factors from the three-dimensional perspective of economy–ecology–resources; and (5) proposing targeted policy recommendations. The identified key driving factors can serve as a basis for policymakers to design targeted governance measures. This study also supports the development of a resilient water security system in the YRB, providing guidance for sustainable water management and balanced regional development.

2. Case Study

2.1. Physiography and Climate

Yan'an is situated on the west bank of the middle reaches of the Yellow River, within the hilly and gully region of the central and southern Loess Plateau ($35^{\circ}21'–37^{\circ}31' N$, $107^{\circ}41'–110^{\circ}31' E$) (Figure 1b). The predominant land use types in Yan'an City are grassland and forest, encompassing an area of approximately 27,200 km² (Figure 1b). The terrain of the region is characterized by a high elevation in the northwest and a low elevation in the southeast, with an average altitude of approximately 1200 m and a total area of 37,036 km² (Figure 1c). Influenced by the warm temperate semi-arid continental monsoon climate, the region experiences distinct seasonal changes. The annual average temperature, precipitation, evaporation, and sunshine duration in Yan'an are 9.9 °C, 524.8 mm, 1400 mm, and 2408.9 h, respectively.

2.2. Process of Ecological Restoration

The problem of soil erosion has intensified in the Yan'an region, thereby constraining the sustainable development of the regional socio-economic system. Driven by accelerated urbanization and population expansion, the tension among population–land use–food security has continued to intensify. The excessive application of fertilizers and pesticides, together with unsustainable agricultural practices such as cultivation on steep slopes, has further exacerbated soil degradation and water pollution. In the 1950s and 1960s, the Ministry of Water Resources and the Chinese Academy of Sciences jointly conducted a systematic investigation of soil erosion on the Loess Plateau, proposing scientific strategies for slope erosion control and agricultural productivity enhancement [27,28]. The large silt dams constructed during the 1960s and 1970s played a crucial role in sediment retention, cropland restoration, flood mitigation, and ecological conservation. Between the 1970s and 1990s, the Yan'an region implemented comprehensive small watershed management by integrating engineering, agricultural, and biological measures, thereby replacing the previous fragmented and single-faceted governance approach. In the 21st century, the GFGP has promoted natural vegetation recovery and provided both poverty alleviation and ecological compensation benefits. Since 2013, Yan'an has promoted the GLCP to compensate for the cropland loss from the GFGP, facilitating the integrated development of ecological restoration and agriculture [37]. In summary, Yan'an's ecological governance has evolved from passive management to a high-quality and collaborative model (Figure 2).

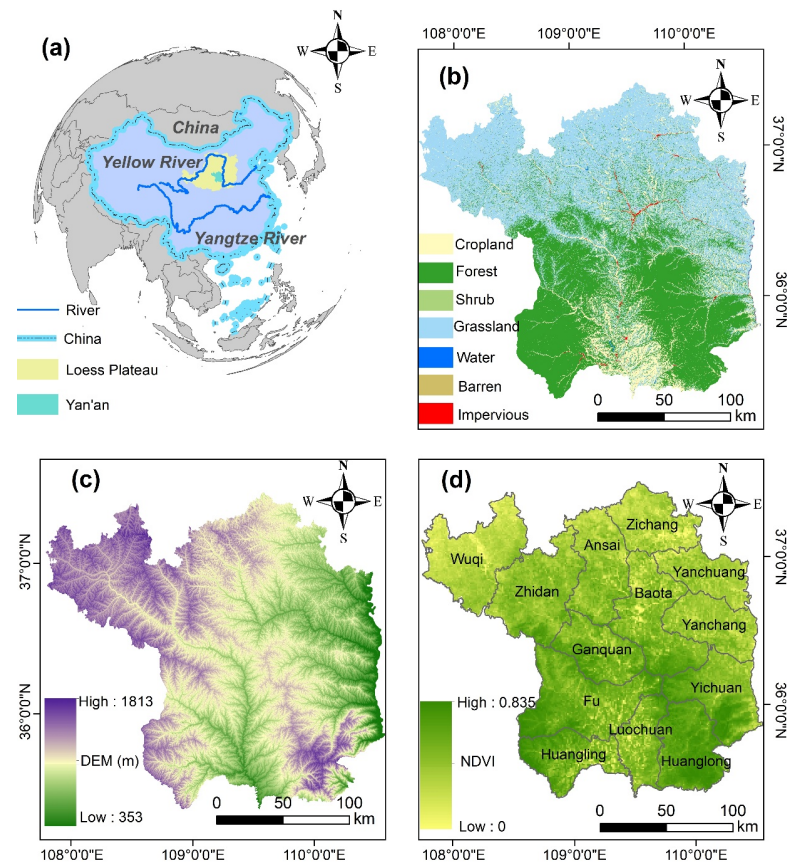


Figure 1. (a) Location of the Loess Plateau and Yan'an; (b) land use in Yan'an region in 2023; (c) digital elevation model of Yan'an; and (d) normalized difference vegetation index (NDVI) value of Yan'an region in 2023.

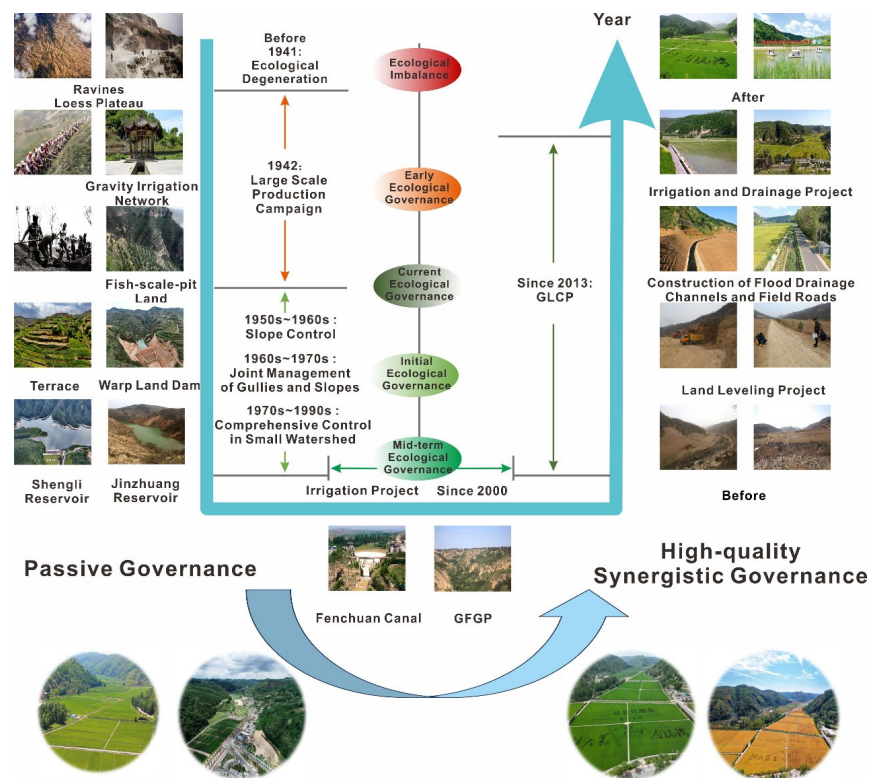


Figure 2. Evolution of ecological restoration and LCPs in Yan'an since 1942.

2.3. Data Sources

This study utilized raw data from the social, economic, and environmental sectors of the Yan'an region, covering the period from 2000 to 2023. Specifically, data on nitrogen fertilizer use, chemical oxygen demand (COD) emissions from industrial and residential sectors, and Gross Domestic Product (GDP) for the study area were sourced from the Shaanxi Statistical Yearbook (Shaanxi Provincial Bureau of Statistics, 2000–2023). Water consumption data were obtained from the Shaanxi Provincial Water Resources Bulletin (Shaanxi Provincial Department of Water Resources, 2000–2023). The data on water yield were obtained from the research findings of [25]. In this study, the mask extraction and spatial analysis tools in ArcGIS 10.8 were employed to conduct spatial processing and refinement of the dataset. To address the gaps in the dataset, particularly the missing values for industrial and domestic sewage discharge in certain years, this study employs interpolation techniques, estimating missing data based on the averages of adjacent years. When adjacent years' data is insufficient or unavailable, this study applies multi-year averages or adjusts the data proportionally to align with the regional context of the Yan'an region, ensuring the consistency and reliability of temporal analyses.

3. Methods

3.1. Measurement of GWF

The industrial structure of Yan'an region is primarily driven by the secondary industry, followed by the tertiary and primary industries as minor contributors, respectively (Shaanxi Provincial Bureau of Statistics, 2005–2023). The GWF calculation for the study area adheres to the guidelines outlined in The Water Footprint Assessment Manual [38]. The detailed calculation method is provided in the Supplementary Information Text S1. This structured analytical framework provides rigorous methodological support for regional water footprint assessment by systematically integrating influencing factors from agriculture, industry, and domestic contributions. Of these, the GWF_{agr} includes the planting sector and the livestock sector.

According to Hoekstra [13], the GWF calculation method involves dividing the pollutant load by the difference between the pollutant's water quality standard and its natural concentration in the receiving water. The calculation formula is

$$GWF_i = \max \frac{L_i}{C_{max} - C_{nat}} \quad (1)$$

where GWF_i and L_i represent the water required to assimilate the pollutant load (m^3/yr) and the pollutant load (kg/yr) of a pollutant i , respectively. C_{max} and C_{nat} are the highest acceptable concentration and the background concentration of the pollutant (mg/L), respectively.

Based on the latest version of China's National Environmental Quality Standards for Surface Water (GB 3838-2002) [39], the maximum allowable concentration (C_{max}) for Grade III surface water is adopted as the corresponding value. This level represents the minimum water quality requirements, with standard limits of 20 mg/L for COD and 1 mg/L for nitrogen and NH_3-N . Given the widespread adoption of the 10 mg/L threshold in GWF_{agr} assessments, it is used as the maximum permissible concentration of nitrogen in the calculation. As defined in The Water Footprint Evaluation Manual [38], C_{nat} refers to the natural background concentration in a water body prior to human disturbances. Therefore, we assume 0 mg/L as the natural background concentration for COD, N, and NH_3-N .

3.2. Integrated ML-SHAP Framework

We employed an integrated analytical framework that combines ML and SHAP algorithms to quantify the impacts of various local anthropogenic factors on GWF. The model construction process involves the following three main steps:

Step 1: Identification of influencing factors

The selection of potential driving factors for spatiotemporal variations in GWF was informed by an extensive literature review, expert insights, and the historical context of ecological protection and high-quality development in the YRB, with careful consideration of data availability. This study focuses on six key dimensions: demographic changes [40], economic growth [15], industrial allocation [41], government intervention, agricultural modernization, and water resource endowment. Ten key environmental and socio-economic variables are selected, with their specific meanings detailed in Table 1.

Table 1. The driving factors and decomposition variables affecting GWF.

Category	Indicator	Specific Meaning	Hypothesis
Demographic changes	P—resident population (ten thousand persons)	It represents the population of permanent residents.	H1a: Population growth negatively impacts the level of regional coordinated development.
	Ur—urbanization level (%)	It indicates the ratio of the urban population to the total population.	
	PD—population density (person/km ²)	It is a key geographic indicator that reflects the degree of population distribution and density within a region.	
Economic growth	PGDP—per capita GDP (10,000 CNY)	It is a key indicator of the economic condition and development level of a region.	H1b: Economic growth positively impacts the level of regional coordinated development.
Industrial allocation	GDP _{agr} —agricultural GDP (hundred million CNY)	It measures the overall outcome of generalized agricultural production activities in a region.	H1c: Industrial allocation positively impacts the level of regional coordinated development.
	Ind 3/Ind 2—The ratio of the value added by the tertiary industry to that of the secondary industry.	It is a measure to assess the degree of regional industrial structure upgrading and economic transformation.	
Government intervention	LFE/GDP—the proportion of local fiscal expenditure relative to GDP.	It is an important indicator for assessing the level of regional economic intervention and public service investment by local governments	H1d: Government intervention positively impacts the level of regional coordinated development.
	AA—afforestation area (km ²)	It is a key indicator for evaluating regional ecological development and the growth of forest resources.	
Agricultural modernization	ADC/AGS—the ratio of agricultural diesel consumption to the area of grain sowing (t/km ²).	It reflects both the extent and scope of the transformation of regional agriculture from a traditional, extensive model to a modern, efficient system.	H1e: Agricultural modernization positively impacts the level of regional coordinated development.
Water resource endowment	WY—water yield (mm)	It refers to the overall conditions of quantity, quality, spatiotemporal distribution, and the natural factors (such as terrain, climate, and hydrological systems) that influence water resources in a region. It serves as the foundation for the sustainable utilization of water resources and directly impacts ecological health, economic development, and social stability.	H1f: Water resource endowment positively impacts the level of regional coordinated development.

Step 2: ML model selection

ML models have shown strong performance and potential in tackling complex non-linear problems [15,20]. In this study, input data from various sources in Step 1 were fed into four popular ML models: eXtreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), random forest (RF), and Support Vector Machine (SVM) in Python 3.14. The dataset was randomly split into a training set and a test set in an 8:2 ratio. The optimal model was selected through iterative optimization of the training strategy and multiple rounds of random experiments. This study employs the RandomizedSearchCV (randomized search cross-validation) technique to minimize uncertainty in the model parameter selection process [42]. Additionally, the K-Fold technique effectively reduces potential bias in data partitioning, optimizes ML model parameters accurately, improves overall performance, and minimizes the risk of overfitting [43,44]. Model performance was assessed using the coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE). The specific calculation formulas for these metrics are provided in Supplementary Information Text S2.

Step 3: Interpretation analysis

The SHAP value is used to quantify the contribution of individual features to the prediction model generated by the ML algorithm [33]. One key advantage of SHAP is its ability to provide both global and local interpretability simultaneously. It not only captures the overall importance of each feature across all predictions but also accurately quantifies the positive and negative contributions of each feature [43]. Compared to traditional feature importance metrics, which typically focus on the independent effects of each feature while ignoring interactions, SHAP evaluates both aspects simultaneously using a cooperative game theory framework. From this, it is evident that SHAP interpretation not only quantifies the contribution of individual features but also clearly reveals the correlation mechanisms and synergistic effects between features, which makes it a more comprehensive and accurate tool for model interpretation [44–46]. The formula for calculating the Shapley value is given by Equations (2) and (3):

$$f(x) \approx g(z') = \varphi_0 + \sum_{i=1}^N \varphi_i z'_i \quad (2)$$

$$\varphi_j = \sum_{S \subseteq \{x_1, \dots, x_p\} / \{x_j\}} \frac{|S|!(p - |S| - 1)!}{p!} (f_x(S \cup \{x_j\}) - f_x(S)) \quad (3)$$

where $f(x)$ represents the model output for a given input x and $g(z')$ is the explanation function that calculates the contribution of individual features to the prediction. $z' \in \{0, 1\}^N$ is an N -dimensional vector of 0/1 values indicating which features in the set of N features are included in the sample; φ_0 is the mean prediction across all samples, often referred to as the base value; φ_i represents the SHAP value for feature i , which indicates the contribution of feature i to the model prediction; and $\{x_1, \dots, x_p\}$ is the set of all input features. p is the number of input features; $\{x_1, \dots, x_p\} / \{x_j\}$ represents all possible subsets of input features excluding feature $\{x_j\}$; and $f_x(S)$ denotes the total contribution of all features in a subset S .

3.3. Gini Coefficient

The Gini coefficient, derived from the Lorenz curve, is a fundamental statistical measure widely applied globally to evaluate disparities in population income distribution [47]. In recent years, its application has gradually expanded beyond the economic domain to encompass allocation problems and equilibrium analysis [48], including research contexts such as environmental assessment and the spatial matching of water resources [49]. Given the agricultural-dominated development characteristics of Yan'an, this study employs the

Gini coefficient to examine the matching relationship between the GWF_{agr} and economic resource factors. The specific calculation formula is presented in Equation (4). Meanwhile, drawing on existing research and accounting for regional differences in water resource endowments, GDP_{agr} and cropland area [50] are selected as key economic-resource indicators to quantitatively evaluate the matching and coordination between GWF_{agr} and economic resource factors:

$$G_j = 1 - \sum_{i=1}^n (X_i - X_{i-1})(Y_i + Y_{i-1}) \quad (4)$$

where G_j denotes the Gini coefficient and i represents the study area. After arranging the data in ascending order, X_i and X_{i-1} denote the cumulative proportion of GWF_{agr} in the i -th and $(i - 1)$ -th study areas relative to the total GWF_{agr} of the region, respectively. Similarly, Y_i and Y_{i-1} represent the cumulative proportion of economic resource factors in the i -th and $(i - 1)$ -th study areas relative to the regional total, respectively. When $i = 1$, both X_{i-1} and Y_{i-1} are defined as zero.

The regional Gini coefficient is illustrated in Figure S1 as twice the area of the shaded region under the Lorenz curve within the range of (0,1). Following international conventions, a value of 0.4 is generally regarded as the threshold indicating an alert level of inequality. The classification criteria for the Gini coefficient are presented in Table 2.

Table 2. Standard of Gini coefficient.

Gini Value	$0 < G < 0.20$	$0.20 \leq G < 0.30$	$0.30 \leq G < 0.40$	$0.40 \leq G < 0.50$	$G \geq 0.50$
Degree	High match	Better match	Relative match	Relative mismatch	High mismatch

3.4. Imbalance Index

The Gini coefficient captures only the overall spatial alignment between GWF_{agr} and economic resource factors [34]. To address this limitation, this study introduces the imbalance index to quantify the spatial mismatches between GWF_{agr} and economic resource factors at the regional level. The calculation formula is as follows:

$$I_k = \frac{Y_k}{X_k} \quad (5)$$

where I_k represents the imbalance index of region k , with a value of 1 indicating absolute equilibrium. Values of $I_k > 1$ signify that the consumption of GWF_{agr} per unit of economic resource factors (including agricultural GDP and cropland area) exceeds the regional average, whereas values of $I_k < 1$ indicate that it is below the regional average. The closer I_k is to 1, the higher the spatial alignment between GWF_{agr} consumption and economic resource factors.

In summary, this study developed the following research framework (Figure 3):

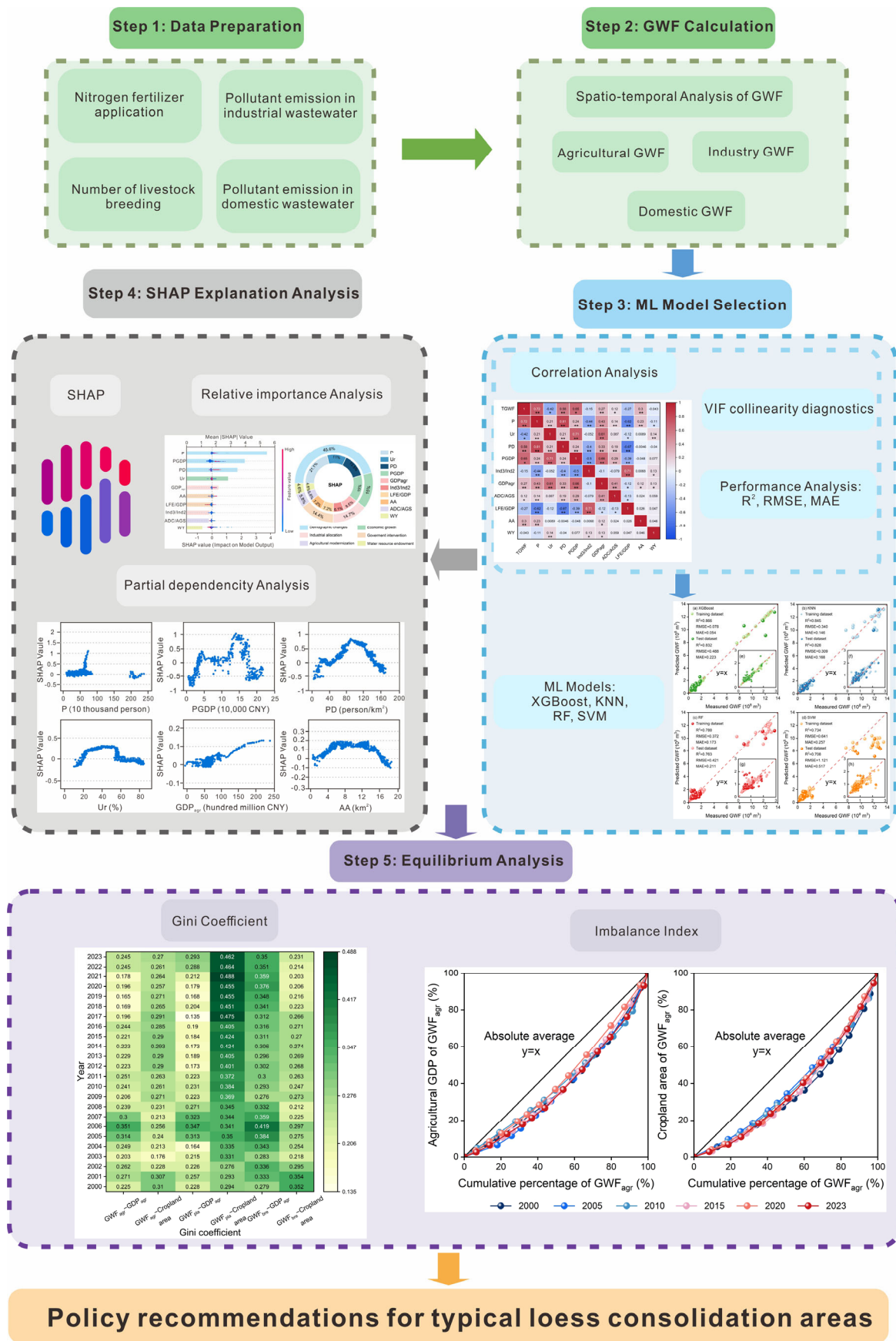


Figure 3. Research framework of this study. (* represents significance at the 95% confidence level and ** at the 99% confidence level).

4. Results

4.1. GWF Measurements

The total GWF (TGWF) peaked at 1.347 billion m³ in 2004, before decreasing to 902 million m³ in 2023. The Mann–Kendall trend test revealed a highly significant downward trend in TGWF over the study period ($Z = -2.853$, $p < 0.01$), as shown in Table 3. The peak in 2004 was primarily driven by a sharp rise in ammonia nitrogen emissions from the agricultural sector, particularly concentrated in the more developed central and northern agricultural regions. In terms of TGWF composition, GWF_{agr} accounted for the largest share, averaging 93.6% between 2000 and 2023, followed by GWF_{dom} at 5.4%, while industrial GWF contributed the least, averaging only 0.9%. The trend of TGWF closely mirrors that of GWF_{agr}, with both exhibiting a significant decline, indicating that the TGWF of Yan'an is largely dominated by the agricultural sector. In the composition of GWF_{agr} (Figure 4e–h), the average GWF_{bre} of Yan'an and its three subregions accounted for 62.92%, 75.32%, 60.39%, and 48.09% of the total GWF_{agr} emissions, respectively. This proportion declined slightly in 2008, followed by a slow but fluctuating upward trend. The Mann–Kendall trend test indicated that, during the study period, the GWF_{agr} from the livestock and poultry breeding industry in Yan'an generally declined, whereas the GWF from the planting sector increased. The GWF_{ind} exhibited a non-significant overall decreasing trend, but with distinct phases: a significant increase occurred from 2008 to 2015, likely associated with the large-scale implementation of small watershed management projects in the Yan'an region, followed by a generally fluctuating decline thereafter.

Table 3. Mann–Kendall trend test for GWF change in the study area from 2000 to 2023.

Area	GWF	Significance Level	Z-Statistics	Trend
Yan'an	TGWF	0.01	−2.853	Extremely significant decreasing trend
	GWF _{agr}	0.01	−3.101	Extremely significant decreasing trend
	GWF _{pla}	0.05	2.009	Significantly increasing trend
	GWF _{bre}	0.01	−3.349	Extremely significant decreasing trend
	GWF _{ind}	n.s.	−0.074	Non-significant decreasing trend
	GWF _{dom}	n.s.	0.769	Non-significant increasing trend
Northern Yan'an	TGWF	0.01	−2.630	Extremely significant decreasing trend
	GWF _{agr}	0.05	−2.339	Significantly decreasing trend
	GWF _{pla}	0.1	−1.910	Significantly decreasing trend
	GWF _{bre}	n.s.	−1.513	Non-significant decreasing trend
	GWF _{ind}	0.01	2.782	Extremely significant increasing trend
	GWF _{dom}	0.01	−3.536	Extremely significant decreasing trend
Middle Yan'an	TGWF	n.s.	0.016	Non-significant increasing trend
	GWF _{agr}	n.s.	0.537	Non-significant increasing trend
	GWF _{pla}	0.01	5.432	Extremely significant increasing trend
	GWF _{bre}	0.01	−3.051	Extremely significant decreasing trend
	GWF _{ind}	0.01	−6.859	Extremely significant decreasing trend
	GWF _{dom}	0.01	−3.722	Extremely significant decreasing trend
Southern Yan'an	TGWF	0.05	1.901	Significantly increasing trend
	GWF _{agr}	n.s.	1.570	Non-significant increasing trend
	GWF _{pla}	n.s.	1.265	Non-significant increasing trend
	GWF _{bre}	0.01	−4.291	Extremely significant decreasing trend
	GWF _{ind}	0.01	7.686	Extremely significant increasing trend
	GWF _{dom}	0.01	4.924	Extremely significant increasing trend

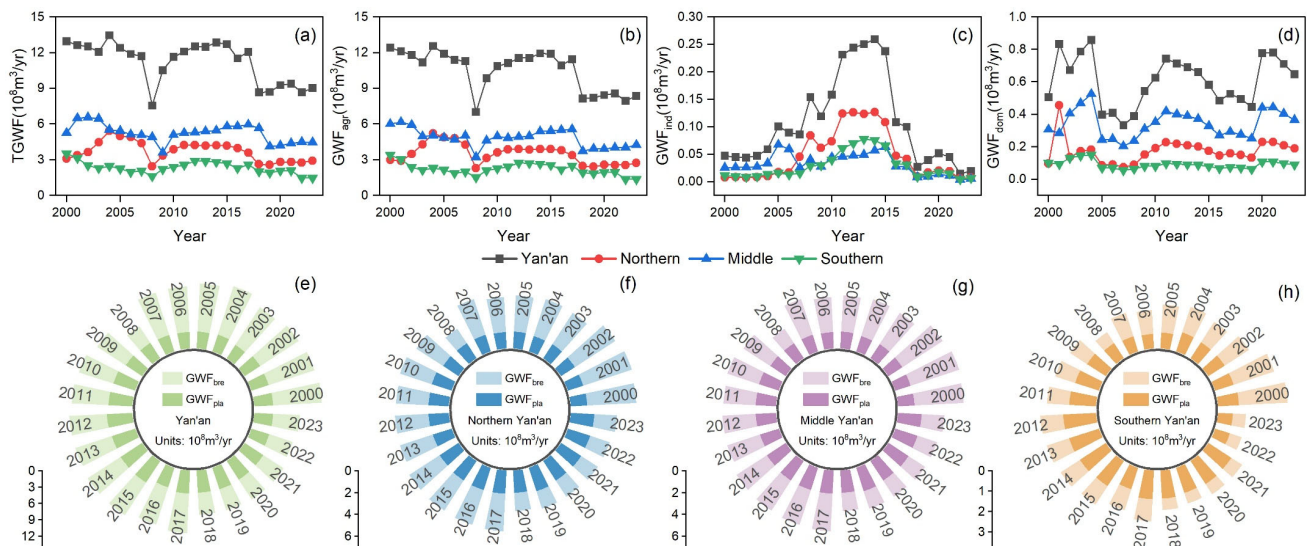


Figure 4. Dynamics and composition characteristics of the GWF in Yan'an, 2000–2023: (a) total GWF; (b) agricultural GWF; (c) industrial GWF; (d) domestic GWF; and (e–h) agricultural GWF in Yan'an, northern Yan'an, middle Yan'an, and southern Yan'an, respectively.

4.2. Spatial–Temporal Variation Analysis of GWF

Between 2000 and 2023, Yan'an's districts and counties exhibited pronounced spatial heterogeneity in GWF distribution and composition. From a spatial perspective, Baota District was identified as the core high-value area, with TGWF, GWF_{agr} , GWF_{ind} , and GWF_{dom} all reaching the region's highest levels (Figure 5). Overall, the spatial distribution of GWF in Yan'an is characterized by a distinct north–central high and southeast low pattern, reflecting significant regional heterogeneity. These regional variations are primarily driven by GWF_{agr} . This may be attributed to the small watershed management projects, such as the construction of check dams, which serve as key measures to enhance agricultural production conditions in the Yan'an area. Following the implementation of these projects, the expansion of cropland has directly facilitated an increase in the scale of agricultural planting. Secondly, to secure crop yields, the intensity of fertilizer application has increased markedly, contributing to higher GWF_{agr} . Together, these factors explain the observed spatial and temporal variations in agricultural water pollution across the region.

Compared with GWF_{agr} , GWF_{ind} is characterized by higher levels in the central and western regions and lower levels in the southern region. This distribution pattern is closely linked to the industrial layout of the Yan'an region. The central and eastern areas, such as Baota District, and the northwest, including Wuqi and Zhidan Counties, serve as the primary industrial hubs, particularly for the energy, chemical, and equipment manufacturing sectors. The wastewater produced during industrial operations in these areas has contributed to elevated levels of industrial GWF. By contrast, the southern districts and counties exhibit relatively high vegetation coverage and an industrial structure dominated by ecotourism, with limited industrial activity. Consequently, industrial GWF in these areas has consistently remained at low levels.

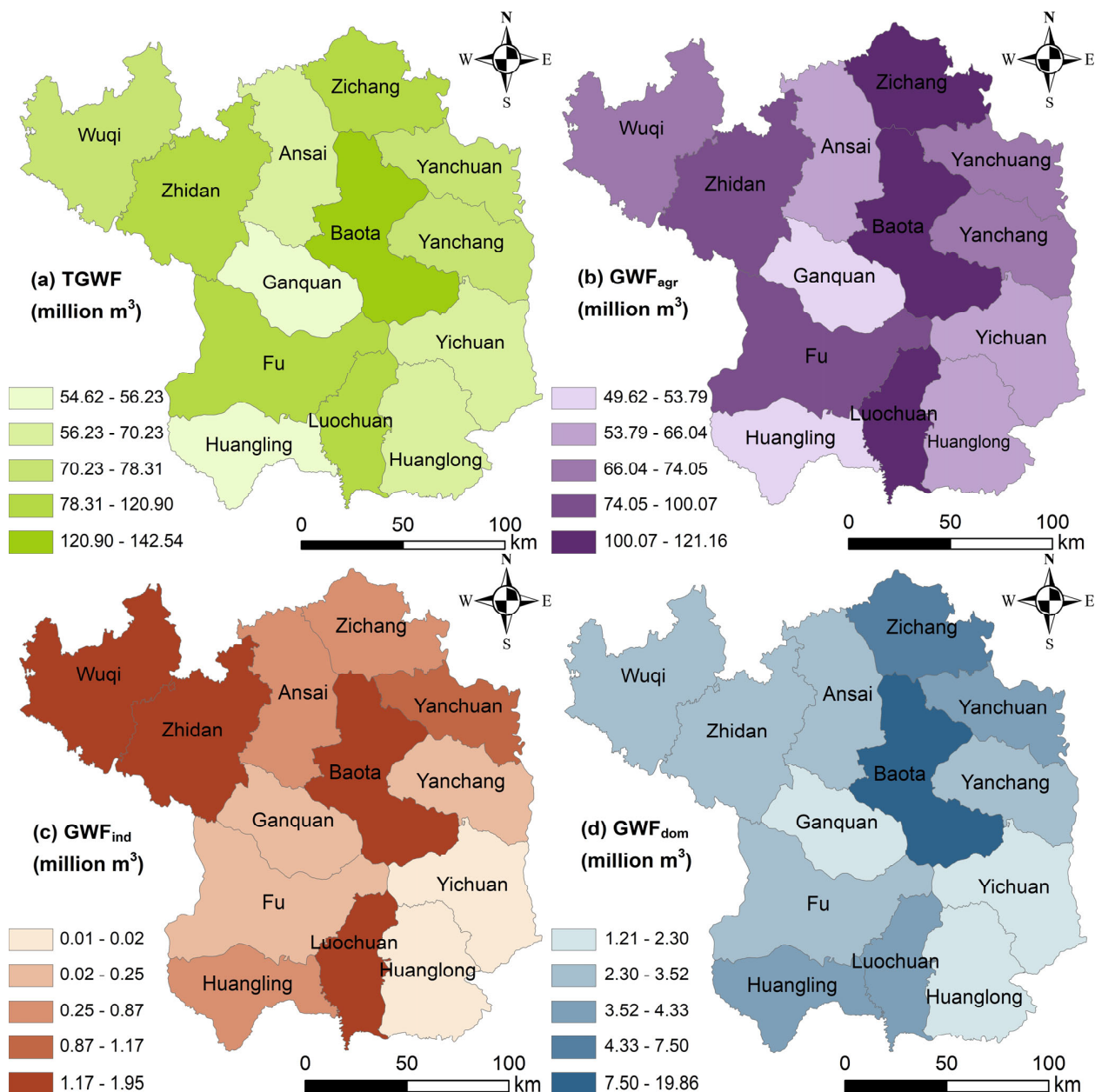


Figure 5. Spatial distribution of GWF and its compositional structure in the Yan'an during 2000–2023. Note: Bottom map colors represent multi-year averages.

4.3. Correlation Between TGWF and Influencing Factors

To investigate the statistical correlation between TGWF and the 10 driving factors, this study initially performed a preliminary Spearman correlation analysis (Figure 6). The results revealed that TGWF had a significant positive correlation with P ($r = 0.72$, $p < 0.01$), PGDP ($r = 0.65$, $p < 0.05$), PD ($r = 0.58$, $p < 0.01$), AA ($r = 0.30$, $p < 0.01$), and GDP_{agr} ($r = 0.27$, $p < 0.01$). In particular, the correlation between P , PGDP, and TGWF was very strong ($r > 0.6$), while AA and GDP_{agr} exhibited a moderate positive correlation. However, these factors did not show a significant correlation in this linear correlation analysis. It is important to note that the correlation coefficient captures only the linear relationship between variables and serves as a preliminary reference for the potential connections between TGWF and its driving factors. Therefore, future research will delve into the more complex nonlinear interactions between the two, with the aim of providing a more comprehensive understanding of the underlying driving mechanisms.

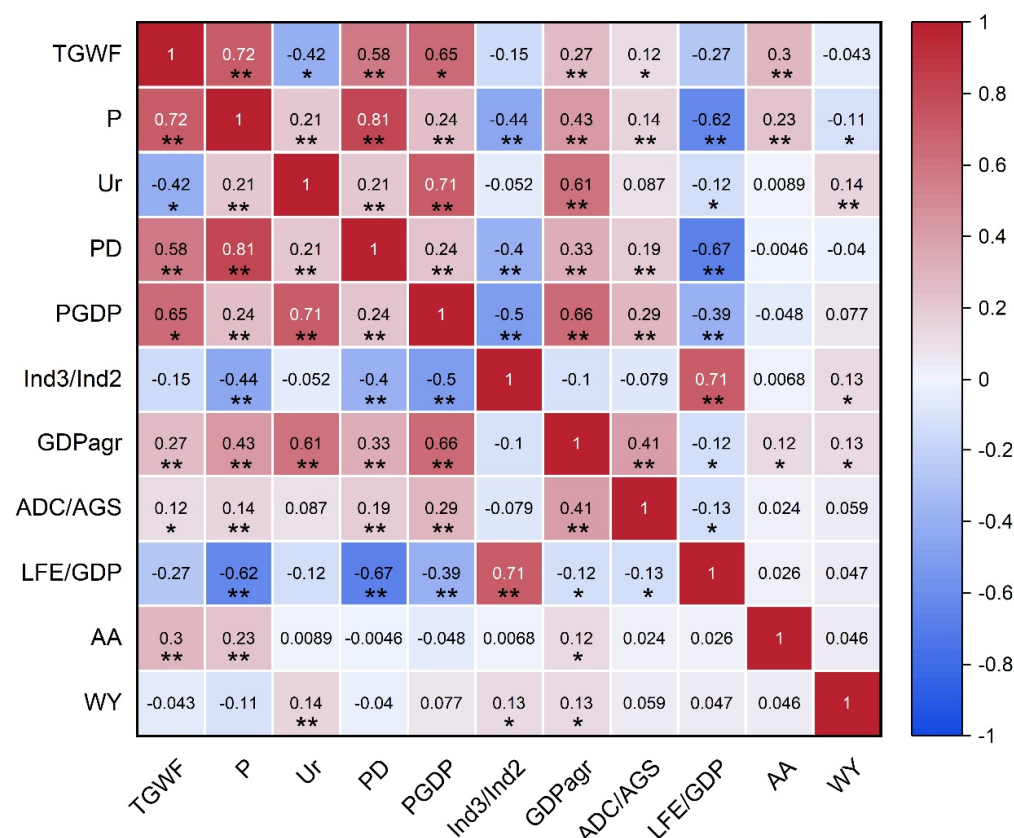


Figure 6. Heatmap of eigenvector Spearman correlation analysis. (The significance of correlation coefficients is denoted by asterisks, where * represents significance at the 95% confidence level and ** at the 99% confidence level).

4.4. Equilibrium Analysis

4.4.1. Imbalance Among Different GWFs

Figure 7 depicts the spatial correspondence between GWF_{agr} (GWF_{pla} , GWF_{bre}) and economic resource factors over time. The findings reveal that the mismatch between GWF_{pla} and cropland area reached its highest level in 2021, with a Gini coefficient of 0.488, surpassing the warning threshold of 0.4. This suggests that cropland resources are highly concentrated, displaying significant agglomeration in a limited number of regions, while the coordination between water and land resources has progressively shifted from relative matching to relative mismatching. The imbalance between GWF_{pla} and GDP_{agr} was minimal, with the Gini coefficient at only 0.173 in 2012 and 2014, further declining to 0.135 in 2017. This reflects a relatively balanced economic spatial distribution of the planting industry from the perspective of agricultural economic output. The Gini coefficients of GWF_{agr} , GDP_{agr} , and cropland area are mostly within the range of 0.160–0.300, suggesting a relatively consistent and stable overall distribution pattern. The Gini coefficient of GWF_{bre} ranges from 0.276 to 0.419, while those of GDP_{agr} and cropland area range from 0.203 to 0.354, all exhibiting a gradual upward trend.

4.4.2. Trend in Temporal Fluctuation

The Gini coefficients of various agricultural GWFs exhibit significant inter-annual fluctuations. In particular, the Gini coefficient of GWF_{pla} –cropland area displayed an overall upward trend from 2000 to 2021, reaching a maximum in 2021 and subsequently declining slightly. This trend indicates that the linkage between GWF_{pla} and cropland resources has strengthened over time, accompanied by an intensification of resource centralization. By contrast, the Gini coefficient of GWF_{agr} relative to GDP_{agr} decreased significantly from

0.351 in 2006 to 0.178 in 2021, indicating a substantial reduction in inequality. The Gini coefficient of the GWF_{bre} relative to cropland area reached its peak at 0.354 in 2001 and subsequently showed a gradual decline. It stabilized within the range of 0.2 to 0.3 for an extended period thereafter. This indicates a decreasing spatial imbalance in the distribution of GWF_{agr} relative to cropland area over time.

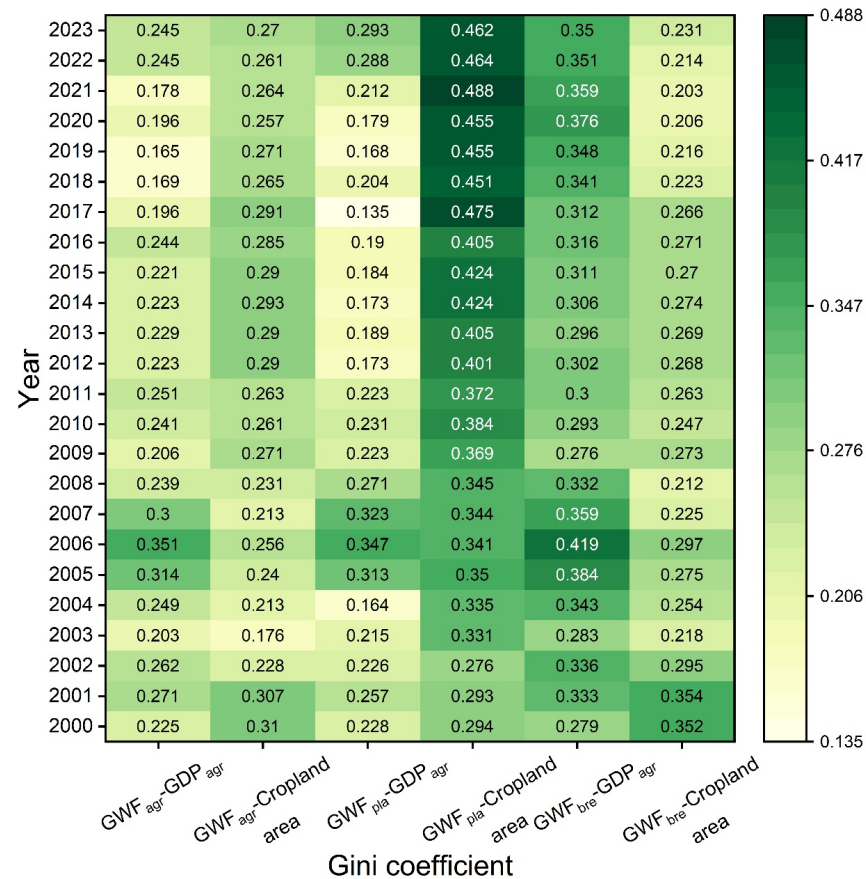


Figure 7. Gini coefficient of GWF_{agr} (GWF_{pla} , GWF_{bre}) in Yan'an region.

4.4.3. Trends in Spatial Distribution

As presented in Figure 8, the $GDP_{agr}-GWF_{agr}$ imbalance index across the Yan'an region consistently deviates from the "absolute equilibrium line," and this deviation has expanded significantly in recent years. Significant spatial differentiation is observed in the imbalance index of cropland area- GWF_{agr} . Northern counties (Wuqi, Zhidan, and Zichang Counties) display a trend of initial increase followed by stabilization. Central counties, including Fu and Yichuan Counties, exhibit significantly higher index levels than other regions, both exceeding the absolute equilibrium line. Southern counties have maintained relatively low levels for an extended period, with minimal spatial variation.

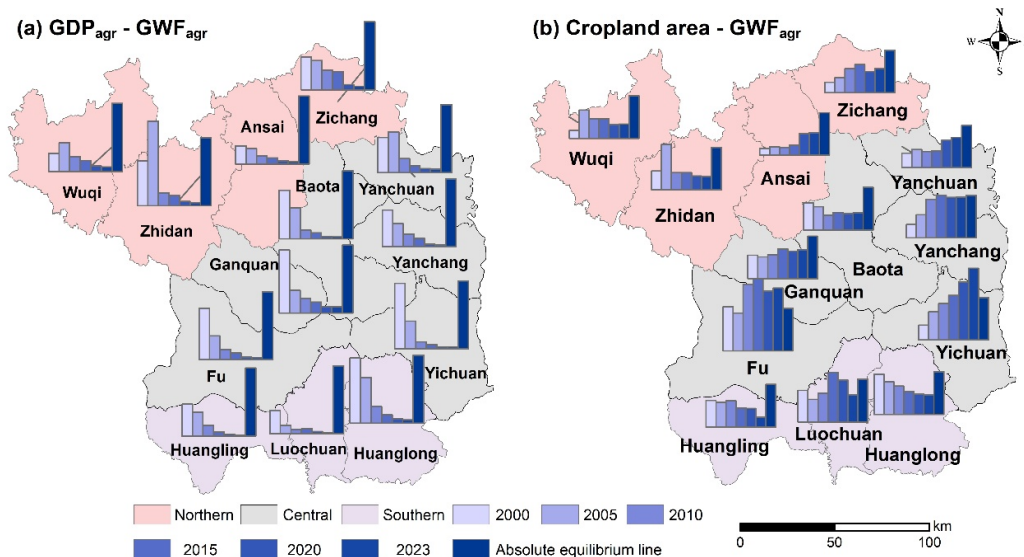


Figure 8. Region's imbalance index of GWF_{agr} with economy resource factors in different years.

5. Discussion

5.1. Driving Factor Decomposition Analysis of TGWF

5.1.1. ML Model Selection

To ensure the reliability of the analysis, this study performed a collinearity diagnosis based on the principle of indicator independence, examining 10 explanatory variables across six dimensions: demographic change, economic growth, industrial allocation, government intervention, agricultural modernization, and water resource endowment. To quantify collinearity among explanatory variables, the Variance Inflation Factor (VIF) is used to test for collinearity and prevent interference from redundancy in the analysis [44]. The results of the collinearity test (Table 4) indicate that the average VIF for the 10 explanatory variables is 2.60, with a maximum value of 5.70, both significantly lower than the collinearity threshold of 10. This indicates the absence of multicollinearity among the explanatory variables, thereby confirming that all variables can be incorporated into the subsequent model analysis [51].

Table 4. Results of the multicollinearity diagnostic test.

Explanatory Variables	VIF	1/VIF
P	5.70	0.18
Ur	1.81	0.55
PD	2.20	0.45
PGDP	2.26	0.44
Ind3/Ind2	1.33	0.75
GDP_{agr}	3.62	0.28
LFE/GDP	1.22	0.82
AA	2.62	0.38
ADC/AGS	4.18	0.24
WY	1.04	0.96

In this study, we employed four machine learning models—XGBoost, KNN, RF, and SVM—to analyze the determinations of the GWF. The ten potential driving factors listed in Table 1 were used as predicted variables, with the TGWF serving as the response variable. Figure 9 illustrates the correlation between the measured and predicted GWF values for both the training and testing datasets. The XGBoost model outperformed the

other models, exhibiting an excellent fit on the training set with $R^2 = 0.866$, $RMSE = 0.078$, and $MAE = 0.054$ (Figure 9a). After verification through the test set, the model achieved an R^2 of 0.832, with an $RMSE$ of 0.488 and an MAE of 0.223. These indicators underscore the accuracy of the model's predictions and confirm the effectiveness of the XGBoost model in identifying the key drivers of the GWF.

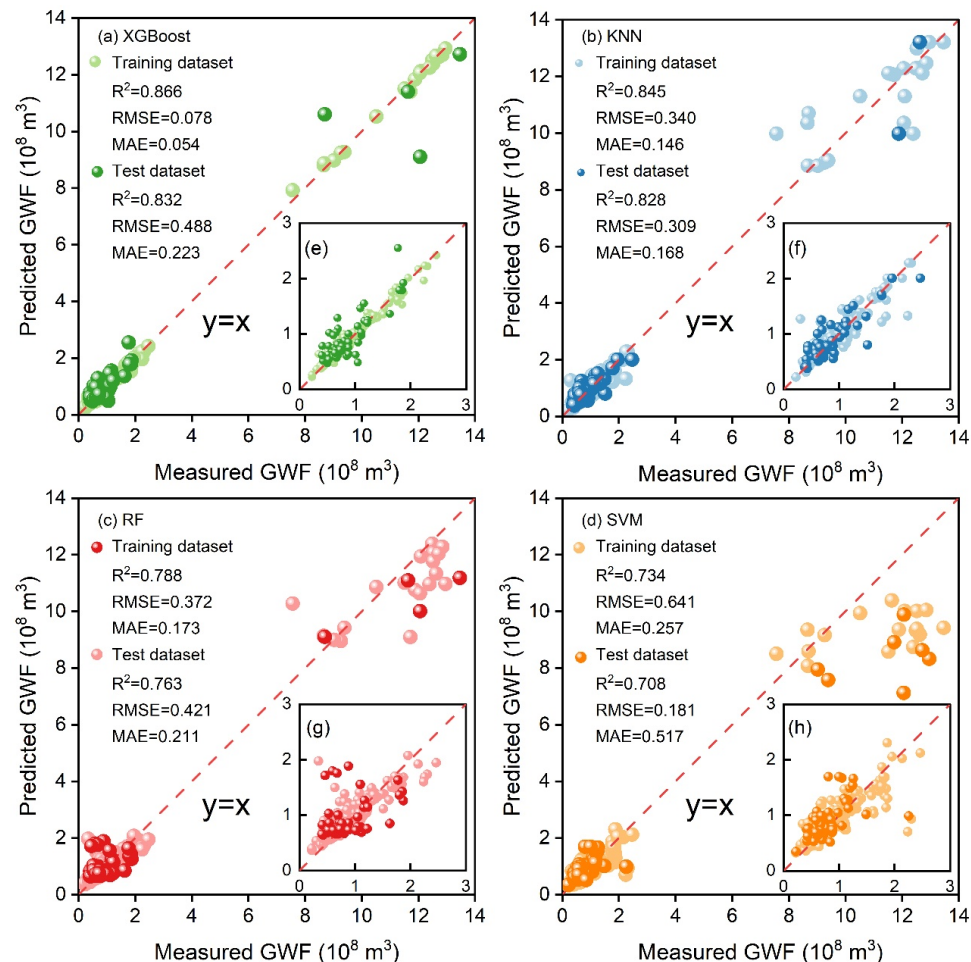


Figure 9. (a–d) Evaluation of the predictive performance of XGBoost, KNN, RF and SVM on GWF of the Yan'an region; and (e–h) Evaluation of the predictive performance of different ML models XGBoost, KNN, RF and SVM on GWF of the Yan'an districts/counties. R: coefficients of determination. RMSE: root mean square error. MAE: mean absolute error normalized mean bias. The units of RMSE and MAE are 10^8 m^3 .

5.1.2. Model Interpretation Based on SHAP

To examine the contribution mechanism of each driving factor to the TGWF, this study calculates the SHAP values using the trained XGBoost model and integrates the variable response patterns (Figures 6 and 10a) for further analysis. Specifically, this study focuses on the following two dimensions. From the perspective of the action mechanisms of positive driving factors, an increase in P, PD, PGDP, AA, and GDP_{agr} and the advancement of agricultural modernization all contribute to the promotion of TGWF. This pattern suggests that a larger regional population size and density, higher levels of economic development, broader afforestation efforts, and more pronounced agricultural intensification and modernization are associated with a greater TGWF. This reflects the positive driving influence of population agglomeration, expanded economic activity, and enhanced agricultural efficiency on the TGWF [7,16,52]. From the perspective of negative driving factors and their regulatory impact, the indicators of industrial structure optimization

(Ind3/Ind2), government intervention (LFE/GDP), and water resource endowment (WY) exhibit a negative correlation with TGWF. This suggests that optimal industrial resource allocation (e.g., promoting high value-added, low water-consuming industries), effective government regulations (e.g., water conservation and pollution control), and adequate water resources can effectively reduce TGWF, providing a clear direction for regional intervention [15]. From the perspective of the phased response characteristics of key factors, it is noteworthy that the influence of P and PGDP on TGWF is nonlinear, displaying significant phased patterns.

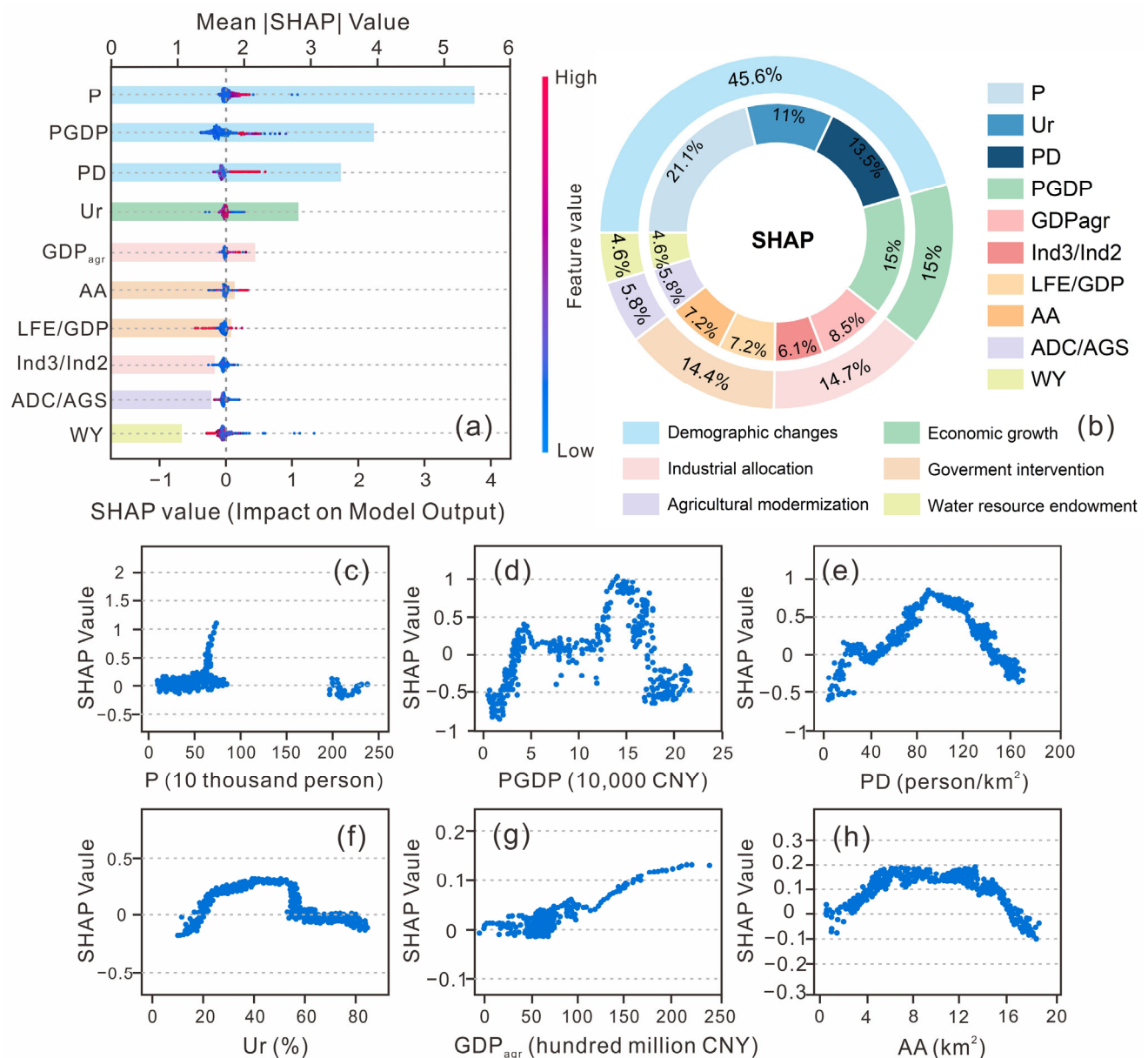


Figure 10. (a,b) Factor importance and influence mechanism based on the XGBoost-SHAP model; (c–h) temporal dynamics of TGWF driving factors analyzed via SHAP values in 2000–2023.

To clearly illustrate the impact and relative relationships of the 10 explanatory variables on TGWF, this study quantifies each variable's contribution using the mean SHAP value and its absolute value. The results indicate that in the TGWF model (Figure 10b), demographic change has the largest contribution, accounting for 45.6%, which makes it the most significant variable affecting TGWF. It is followed by economic growth (15%), industrial allocation (14.7%), and government intervention (14.4%). The combined contribution

of these four factors exceeds 89%, which highlights the key role of population dynamics, economic output, industrial structure, and government regulations in driving the evolution of the regional GWF. This conclusion aligns with existing research, which identifies economic development and population factors as key variables in water resource management. Additionally, agricultural modernization (5.8%) and water resource endowment (4.6%) are important factors, though their individual contributions are relatively modest. Together, they reflect the complex interplay between agricultural production, urban development and expansion, and the efficiency of water resource utilization.

5.1.3. Dependencies Between Influencing Factors

To further explore these relationships, the impact of the top six factors (Figure 10c–h) on TGWF is presented in order of importance to offer insights into their complex interactions. The partial dependency plot from the XGBoost model highlights the marginal impact of specific drivers on TGWF. These findings indicate that *P* plays a key role in TGWF and exhibits phased characteristics. Between 40,000 and 600,000, population growth positively correlates with TGWF; From 600,000 to 800,000, this relationship weakens and stabilizes, and when *P* exceeds 800,000, the relationship becomes negative, with population growth inhibiting TGWF. From a mechanistic perspective, *P* not only drives pollutant generation but also increases water resource consumption. Firstly, the growth of *P* directly increases domestic water demand, thereby leading to higher sewage discharge. Secondly, population growth has heightened the demand for agricultural production. As the conflict among population–land–food intensifies in Yan'an, large-scale LCPs have expanded cropland areas. However, the extensive use of fertilizers and pesticides has exacerbated agricultural non-point source pollution [24]. Additionally, while population growth and urbanization have driven regional economic growth, they have also exacerbated the spatial imbalance of TGWF (Figure 8). It is noteworthy that the TGWF of the entire Yan'an region did not show an upward trend along with population growth. This can be attributed to the refined water resource management policies implemented in recent years. Measures such as water quota control, tiered pricing, and total water use regulation have constrained water consumption, reduced inefficiencies and pollutant emissions, and attenuated the positive effect of population growth on TGWF [15].

As a key indicator of regional socio-economic development, the relationship between PGDP and TGWF generally exhibits a positive 'M-shaped' correlation, which mirrors the typical pattern of the Environmental Kuznets Curve (EKC) [46]. Specifically, in the initial phase of economic development, TGWF increases in tandem with PGDP. However, once PGDP surpasses a certain threshold, TGWF exhibits a declining trend. In the early stages of PGDP growth, economic development was primarily driven by the expansion of industry and agriculture, which resulted in a substantial increase in the pollution-related consumption of water resources. Since 2010, changes in industrial structure, technological progress, and environmental policies have reduced water pollution emissions, promoting the shift to sustainable water resource development [15]. The relationship between *PD* and TGWF mirrors that of PGDP, reflecting the intricate interactions among population agglomeration–industrial optimization–policy management.

Ur primarily reflects the evolution of regional population and consumption patterns, indirectly influencing the TGWF by shaping the trajectory of economic development [52]. Overall, it exhibits a significant negative correlation with TGWF (Figure 6). As *Ur* increases, the SHAP value exhibits an initial increase, followed by stabilization and a subsequent decline. During the rapid urbanization phase ($10\% < Ur < 40\%$), the rapid increase in urban population density, intensified industrial activities, and rising domestic demands, combined with the delayed development of environmental protection infrastructure, become the pri-

mary drivers of TGWF growth. In the mature stage of urbanization ($40\% < Ur < 80\%$), as the region enters a phase of high-quality development, the growth rate of TGWF slows significantly, even showing signs of inhibition. This transformation results from the tripartite collaborative pollution control mechanism. First, industrial structure optimization has reduced high-pollution, high-water-consumption sectors, while green and low-carbon industries have taken the lead [24]. Second, improvements in environmental infrastructure have boosted sewage treatment and reclaimed water use [53–55]. Third, stronger policy controls and heightened environmental awareness have driven water conservation and emission reduction measures, effectively managing pollution sources. Together, these factors mitigate the positive impact of Ur on TGWF.

GDP_{agr} shows a significant positive correlation with TGWF (Figure 6), with its SHAP value dependency analysis following an inverted Z-shaped trend (Figure 10g). GDP_{agr} primarily affects TGWF by reflecting the scale, structure, and technological level of agricultural production. In terms of positive drivers, the growth of GDP_{agr} is typically associated with an expansion in TGWF, closely linked to the implementation of regional LCPs. The expansion of cropland often leads to excessive use of fertilizers and pesticides, with surplus nutrients entering water bodies through runoff, thus driving the increase in agricultural non-point source pollution and elevating TGWF [56]. However, industrial optimization and efficient production models can significantly weaken the positive correlation, reducing the driving effect of GDP_{agr} growth on TGWF at the source [57].

AA is positively correlated with TGWF (Figure 6), with its SHAP value dependency analysis exhibiting an inverted trapezoidal relationship (Figure 10h). The impact of AA on TGWF varies significantly across different ranges, influenced by the interplay between GFGP and LCP. In the early stage (AA between 2 and 6 km²), the lack of synergy between the two projects leads to increased pollution pressure and a rise in TGWF. As AA increases to 6–14 km², both projects enter a cooperative phase, reducing short-term negative effects and stabilizing the SHAP value, weakening AA's positive driving effect on TGWF. When AA expands further to 18 km², afforestation and land consolidation are deeply coupled, forming a systematic ecological restoration model. The SHAP value decreases and even becomes negative, with AA suppressing TGWF growth. Consequently, large-scale afforestation through land consolidation has enhanced water resource management and mitigated agricultural non-point source pollution, ultimately shifting AA from a positive driver to a negative regulator of TGWF.

5.2. Spatial Matching and Regional Disparities of GWF_{agr}

The Yan'an Small Watershed Management Project (2011–2018) has had a profound impact on the spatial distribution and regional disparities of the GWF_{agr}. Prior to the governance project in 2011, the Gini coefficients of different factors in the study area were generally at a better match level, reflecting a balanced spatial distribution between the GWF_{agr} and economic factors. As the contradictions among population–land–food pressures became increasingly pronounced, the Yan'an region in 2011 designated Zichang, Yanchuan, and Baota as pilot sites for the GLCP. This initiative pioneered an ecological governance approach that combined “returning cropland to forests on the mountains to protect the ecology” with “ditch and land reclamation at the foothills to enhance livelihoods.” In 2013, Yan'an launched a comprehensive LCP for gully control and reclamation, integrating systematic land consolidation, supporting water conservancy infrastructure, and small watershed ecological restoration. However, the Gini coefficient between the GWF_{pla} and cropland area has declined from relative matching to relative mismatching since 2013 (Figure 7), indicating a growing spatial imbalance between agricultural development and environmental pressure. While the intensive LCP has driven the intensification of cropland

resources, it has simultaneously aggravated spatial imbalances in land resource allocation (Figure 11). Secondly, the expansion of cropland has been accompanied by increased overuse of fertilizers and pesticides, intensifying the issue of agricultural non-point source pollution [1,58]. In addition, deficiencies may exist in the application of agricultural non-point source pollution control technologies and in the implementation of environmental management policies in the region [59]. Collectively, these factors contribute to a spatial imbalance between environmental and economic factors.

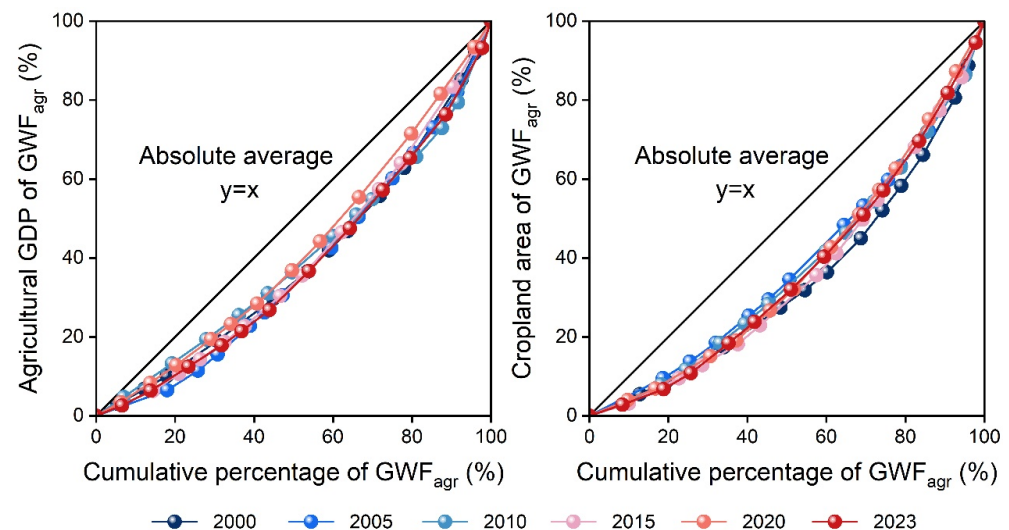


Figure 11. Lorenz curve between GWF_{agr} and economy resource factors in Yan'an region.

As of 2018, the GLCP in Yan'an encompassed 369.86 km², adding 78.49 km² of cropland area and 282.33 km² of high-standard cropland. The implementation scale of these projects across Yan'an is detailed in Table 5. Between 2019 and 2023, the Gini indices of GWF_{agr} , GDP_{agr} , and cropland area remained stable and became increasingly optimized, reflecting a relative matching state. This suggests that the long-term ecological and economic impacts of small watershed management projects are gradually emerging. Simultaneously, the projects have promoted a more sustainable and balanced allocation of agricultural water, economic, and cropland resources.

Table 5. Implementation scale of GLCP in Yan'an.

Region	District/County	Number of Projects	Construction Scale	Newly Added Cropland Area	High-Standard Cropland Area	Investment Amount
	Units	/	km ²	km ²	km ²	Ten Thousand CNY
Northern	Wuqi	10	21.60	8.38	12.45	19,880.75
	Zhidan	3	3.31	0.96	1.34	3375.71
	Ansai	20	24.06	6.32	19.10	21,229.54
	Zichang	29	49.20	8.17	32.14	50,370.23
Central	Ganquan	19	32.79	5.29	26.26	30,824.60
	Baota	34	77.35	11.91	61.83	80,892.24
	Yanchuan	14	35.62	8.02	27.95	38,754.82
	Fu	16	30.71	4.28	24.60	28,438.50
	Yanchang	16	34.03	8.31	26.99	34,277.77
	Yichuan	8	18.73	6.12	14.24	16,289.26
Southern	Huangling	8	12.36	2.98	10.54	11,637.61
	Luochuan	13	21.91	6.17	17.83	22,917.90
	Huanglong	7	8.21	1.57	7.07	75.19
Sum		197	369.86	78.48	282.33	366,407.68

In the spatial coupling between GDP_{agr} and GWF_{agr} , the regional absolute equilibrium line is pronounced, with the overall imbalance index showing a significant decline. These findings imply that regional agricultural production is becoming increasingly environmentally efficient, achieving a stronger decoupling between economic growth and water pollution [60–62]. This improvement is closely associated with the small watershed management project, which implemented measures such as check dam construction, slope-to-terrace conversion, and water-efficient irrigation [63,64]. These measures not only enhance water use efficiency and pollutant control in agriculture but also promote the adjustment and optimization of the agricultural industrial structure, fostering more coordinated development between the agricultural economy and the environment [65,66].

5.3. Policy Recommendations

To achieve a regional decoupling of economic development and water pollution, targeted measures should be implemented based on a water endowment, pollution drivers, and governance framework. The water resources in Yan'an are limited and unevenly distributed, with a significant dependence on groundwater, which exacerbates the instability of the water supply. As an agriculture-dominated region, each district and county must balance the three primary goals of economy–ecology–resources while fostering precise agricultural development and implementing effective water pollution control measures (Figure 12). Specific suggestions are as follows:

- (1) Strong decoupling between industrial development and GWF (oil, coal, and natural gas-rich regions in the north)

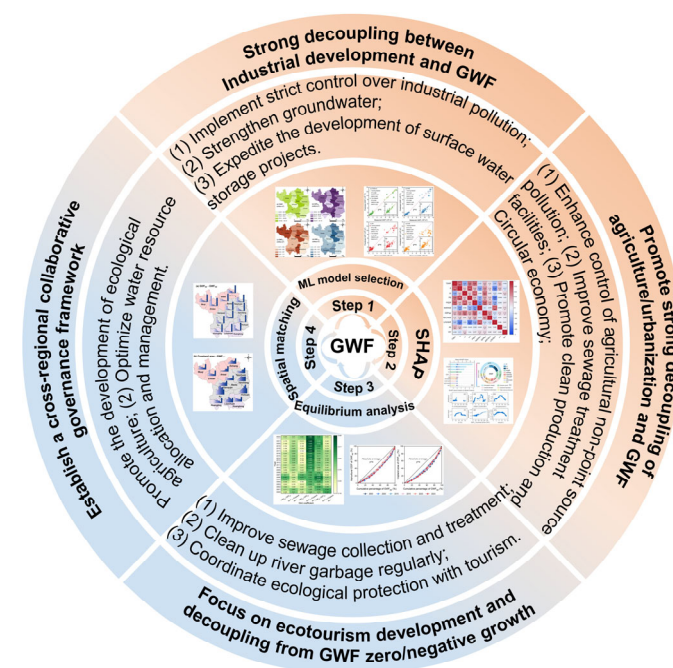


Figure 12. Policy suggestions for Yan'an region.

This region's economy is largely dependent on mineral extraction and the chemical industry, with industrial wastewater being a major source of groundwater pollution. Furthermore, the area experiences both engineering-related and water quality issues contributing to water scarcity, alongside the overexploitation of groundwater resources. It is recommended to implement strict control over industrial pollution and enhance the enforcement of environmental regulations. Second, strengthen groundwater protection by designating no-extraction and restricted-extraction zones. Third, expedite the development

of surface water storage and other backup water supply projects to mitigate the risks associated with relying on groundwater as the primary drinking water source.

(2) Promote strong decoupling of agriculture/urbanization and GWF (central urban area)

Following land consolidation, the expansion of cropland, and rapid urbanization, agricultural non-point source pollution, industrial emissions, and domestic sewage have collectively contributed to an increase in GWF. It is crucial to shift from a “weak decoupling” between agricultural intensification, urbanization, and GWF growth to a “strong decoupling”. To achieve this, the following recommendations are proposed: First, enhance control of agricultural non-point source pollution, promote precision fertilization and agricultural waste recycling, and strengthen irrigation water quality monitoring. Second, improve sewage treatment facilities, particularly by increasing treatment capacity in rural areas. Third, strengthen emission monitoring in the chemical and metallurgical industries, and promote clean production and a circular economy. Additionally, plan and construct green infrastructure, such as rainwater collection systems and ecological wetlands, to enhance the self-purification capacity of water bodies.

(3) Focus on ecotourism development and decoupling from GWF zero/negative growth (southern mountainous areas)

This region benefits from abundant precipitation and favorable water resources, but some areas face issues such as inadequate sewage pipe networks and river pollution from domestic waste. It is recommended to improve sewage collection and treatment, regularly clean up river garbage, and coordinate ecological protection with tourism development.

In general, each region should optimize its development strategy based on the differences in water resource distribution. In the northern and central regions, where water resources are scarce, priority should be given to resource and economic goals, with a focus on fully utilizing water conservation potential to achieve a strong decoupling between economic development and water pollution. In the southern regions, where water resources are relatively abundant, the cultivation of high-yield crops should be appropriately expanded, and the aquaculture industry should be adjusted while ensuring water resources and ecological balance, aiming for a decoupling between agricultural income growth and zero/negative GWF growth.

6. Conclusions and Limitations

6.1. Conclusions

Based on the theory of GWF, this study combines ML models and interpretability analysis to investigate the spatiotemporal evolution and driving mechanisms of GWF in the Yan'an region. From the perspectives of “Economy—Ecology—Resources”, this study evaluated the spatial matching characteristics between GWF_{agr} and economic resource factors (GDP_{agr} and cropland area) using the Gini coefficient, imbalance index, and decoupling theory. The main conclusions are summarized as follows:

(1) Spatial and temporal heterogeneity of GWF

From 2000 to 2023, the TGWF in Yan'an exhibited a significant downward trend, primarily driven by the GWF_{agr} , which accounted for an average of 93.6%. The GWF_{ind} showed no significant decrease and exhibited a phased pattern, while the GWF_{dom} remained the smallest. Spatially, TGWF and GWF_{agr} were the most pronounced in the northern and central regions, with Baota District as the high-value area. By contrast, GWF_{ind} was the most prominent in the central and eastern regions, while being relatively lower in the southern areas.

(2) Identification of driving factors and ML model selection

This study investigates the driving factors of TGWF from six dimensions: demographic changes, economic growth, industrial allocation, government intervention, agricultural modernization, and water resource endowment. Following the screening of the 10 explanatory variables through VIF collinearity analysis, four ML models, including XGBoost, KNN, RF, and SVM, were employed to analyze the driving factors of TGWF. Among these, XGBoost exhibited the best performance, with an R^2 of 0.866 on the training set and 0.832 on the test set, thereby effectively identifying the key driving factors.

(3) Causal interpretation of the driving mechanism based on SHAP

To address the ‘black box’ nature of complex models, the SHAP method is employed to decompose and visualize the contributions of various variables. Among these, demographic change emerges as the most significant factor (45.6%), followed by economic growth, industrial allocation, and government intervention, which together account for over 89% of the total effect.

(4) Imbalance analysis in GWF_{agr}

During the initial stages of the LCPs, the expansion of cropland and the intensification of agricultural non-point source pollution led to a prominent spatial imbalance between the GWF_{agr} and economic development. In the subsequent stages of governance, a gradual alignment emerged between the GWF_{agr} , agricultural GDP, and cropland area. This suggests that the long-term ecological and economic impacts of small watershed management projects are beginning to materialize, fostering a stronger decoupling between economic growth and water pollution.

(5) The GWF analysis framework with the LCP

We incorporated relevant variables such as changes in agricultural land use, water management practices, and the implementation of the Grain for Green Project (GFGP). These variables were integrated into the XGBoost model to examine their direct and indirect impacts on GWF. The findings reveal that the total GWF (TGWF) peaked at 1.347 billion m^3 in 2004, subsequently declining due to improvements in water management efficiency and land use changes driven by LCP and GFGP. Spatially, TGWF is higher in the central and eastern regions, where agricultural GWF (GWF_{agr}) predominates. The permanent population and per capita GDP emerged as key driving factors, contributing 21.1% and 15% to the total change in TGWF, respectively. The imbalance index in the spatial coupling between agricultural GDP and GWF_{agr} has significantly decreased, indicating improvements in land use efficiency and resource management. The synergistic effects of the Grain for Green Project and LCPs are increasingly evident, contributing to reductions in water consumption and pollution.

6.2. Limitations and Further Study

Although this study systematically analyzed the spatiotemporal changes in GWF in the Yan'an region, several research gaps persist. Firstly, the accounting methods employed lack standardization, focusing predominantly on conventional indicators such as COD, TN, and TP while neglecting emerging water quality stressors, including heavy metals, microplastics, antibiotics, and persistent organic pollutants. Additionally, discrepancies exist in the environmental standard concentrations of pollutants and the natural background concentrations of water bodies. And the self-purification capacity of natural water bodies (e.g., biodegradation and physical adsorption) has not been considered, leading to deviations in GWF calculations. Secondly, the paucity of data on agricultural non-point source pollution emissions, coupled with the absence of long-term monitoring networks,

constitutes a significant gap. Thirdly, the use of “annual average” calculations overlooks seasonal variations, such as increased runoff pollution during the agricultural rainy seasons and higher concentrations of industrial drainage in the dry season. Moreover, the spatial heterogeneity of hydrological characteristics (e.g., river flow velocity and water area) within the basin has not been taken into account. Finally, the synergistic effects of grey water–blue water–green water have not been correlated, and cross-scale transfer effects have been neglected, with no coupled analysis of ecosystem responses.

To address these deficiencies, future research should focus on five key areas—method standardization, data system construction, spatiotemporal model optimization, multidimensional coupling, and policy tool transformation—to ultimately form a closed-loop process of “accounting—evaluation—application.” Specifically, (1) standardize accounting methods by developing a unified framework and systematically selecting pollutants for accurate measurement; (2) enhance data support through the establishment of an integrated multi-source data system; (3) refine spatiotemporal models by incorporating seasonal or monthly dynamics and integrating basin hydrological characteristics; (4) encourage multidimensional coupling of water environment, ecosystem, and economic development, focusing on the interactions between grey–blue–green water; and (5) link policy to practice by supporting ecological protection and sustainable development in the YRB through “total quantity control,” industry standards, and incentive mechanisms.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/su18041854/s1>, Text S1: Definition of greywater footprint; Text S2: Methodology for GWF calculation; Text S3: R2, mean absolute error (MAE), and root mean square error (RMSE); Text S4: Multiple Linear Regression; Figure S1: Schematic diagram of regional Gini coefficient; Table S1: Values of parameters for livestock and poultry breeding grey water footprint; Table S2: Fitting results of the multiple linear regression model.

Author Contributions: Q.Y.: conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft. H.Q.: resources project administration, supervision, review and editing. Q.L.: investigation. Y.D.: investigation. Z.C.: investigation. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 41931285; Forestry Science and Technology Innovation Project of Shaanxi Province, grant number SXLK2023-02-1; Fundamental Research Funds for the Central Universities, CHD, grant number 300102294905.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: We are grateful for the constructive comments and suggestions provided by the anonymous reviewers and the support from MDPI Author Services. We thank the editors at MDPI for their assistance throughout the submission process. We also appreciate that the completion of this article is inseparable from the contributions of all authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, Y.X.; Qian, H.; Gao, Y.Y.; Li, S.Q.; Shi, X.X.; Cao, Z.M.; Long, Q.; He, Y.; Ma, Y.D.; Wu, P.X.; et al. Identifying nitrate pollution of groundwater based on aquifer vulnerability and surface nitrogen loads in the Northern Foot of the Qinling Mountains, China. *J. Hazard. Mater.* **2025**, *495*, 138939. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Najah Ahmed, A.; Binti Othman, F.; Abdulmohsin Afan, H.; Khaleel Ibrahim, R.; Ming Fai, C.; Shabbir Hossain, M.; Ehteram, M.; Elshafie, A. Machine learning methods for better water quality prediction. *J. Hydrol.* **2019**, *578*, 124084. [\[CrossRef\]](#)
3. Vörösmarty, C.J.; McIntyre, P.B.; Gessner, M.O.; Dudgeon, D.; Prusevich, A.; Green, P.; Glidden, S.; Bunn, S.E.; Sullivan, C.A.; Liermann, C.R.; et al. Global threats to human water security and river biodiversity. *Nature* **2010**, *467*, 555–561. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Wang, B.S.; Yu, N.N.; Sher, A.; Cui, D.; Yang, S.L.; Si, J.S.; Ren, B.Z.; Zhang, J.W. Optimized integrated soil-crop system management enhances crop yield while reducing water resource consumption. *Agric. Water Manag.* **2025**, *319*, 109792. [\[CrossRef\]](#)
5. Chen, L.; Yu, W.; Zhang, X.B. Spatio-temporal patterns of High-Quality urbanization development under water resource constraints and their key Drivers: A case study in the Yellow River Basin, China. *Ecol. Indic.* **2024**, *166*, 112441. [\[CrossRef\]](#)
6. Han, X.X.Q.; Hua, E.; Guan, J.J.; Engel, B.A.; Liu, R.; Bai, Y.W.; Sun, S.K.; Wang, Y.B. Development of a method to assess synergy and competition for water use among water-energy-food nexus in the Yellow River basin: Water quantity-quality dimensions. *J. Hydrol.* **2024**, *639*, 131607. [\[CrossRef\]](#)
7. Feng, H.Y.; Yang, M.J.; Zeng, P.; Liu, Y.Y.; Sun, F.Y.; Wang, K.; Yang, K.; Che, Y. Transprovincial water quality impacts and the environmental inequity of grey water footprint transfer in China. *Resour. Conserv. Recycl.* **2022**, *186*, 106537. [\[CrossRef\]](#)
8. Xie, L.M.; Gao, R.Z.; Wang, X.X.; Duan, L.M.; Fang, L.J.; Tong, H.; Yue, C.; Liu, T.X. Spatiotemporal evolution of surface water quality and driving factors across varying levels of human interference in a major subbasin of the Yellow River Basin, China. *J. Hydrol. Reg. Stud.* **2025**, *59*, 102327. [\[CrossRef\]](#)
9. Xiao, F.N.; Zhang, R.Q.; Jian, Z.Q.; Liu, W.; Sun, T.B.; Pang, W.C.; Han, L.; Qin, H.P. Using ensemble machine learning to predict and understand spatiotemporal water quality variations across diverse watersheds in coastal urbanized areas. *Ecol. Indic.* **2025**, *178*, 113976. [\[CrossRef\]](#)
10. Zeng, L.W.; Sun, D.Z. Accurate calculation method for urban domestic sewage centralized collection rate based on water quality and quantity variations. *J. Clean. Prod.* **2025**, *516*, 145809. [\[CrossRef\]](#)
11. Long, X.; Wu, S.G.; Wang, J.Y.; Wu, P.Z.; Wang, Z. Urban water environment carrying capacity based on VPOSR-coefficient of variation-grey correlation model: A case of Beijing, China. *Ecol. Indic.* **2022**, *138*, 108863. [\[CrossRef\]](#)
12. Ren, J.; Han, G.L.; Gao, X.; Zhang, S.T. Water quality assessment in a megacity river: Water chemical analysis integrated with AHP-Entropy weighted fuzzy comprehensive evaluation model. *J. Environ. Sci.* **2025**, *162*, 121–130. [\[CrossRef\]](#)
13. Hoekstra, A.Y. Water scarcity challenges to business. *Nat. Clim. Chang.* **2014**, *4*, 318–320. [\[CrossRef\]](#)
14. Chen, J.; Gao, Y.Y.; Qian, H.; Jia, H.; Zhang, Q.Y. Insights into water sustainability from a grey water footprint perspective in an irrigated region of the Yellow River Basin. *J. Clean. Prod.* **2021**, *316*, 128329. [\[CrossRef\]](#)
15. Kong, Y.; He, W.J.; Zhang, Z.F.; Shen, J.Q.; Yuan, L.; Gao, X.; An, M.; Ramsey, T.S. Spatial-temporal variation and driving factors decomposition of agricultural grey water footprint in China. *J. Environ. Manag.* **2022**, *318*, 115601. [\[CrossRef\]](#)
16. Zhang, L.; Dong, H.J.; Geng, Y.; Francisco, M.J. China's provincial grey water footprint characteristic and driving forces. *Sci. Total Environ.* **2019**, *677*, 427–435. [\[CrossRef\]](#)
17. Liu, B.F.; He, Y.H.; Tan, Q.; Zhang, Y. The grey water footprint of the Guangdong-Hong Kong-Macao Greater Bay Area, China: Spatial patterns, driving mechanism and implications. *J. Environ. Manag.* **2025**, *389*, 126063. [\[CrossRef\]](#)
18. Fan, Y.P.; Fang, C.L. A comprehensive insight into water pollution and driving forces in Western China—Case study of Qinghai. *J. Clean. Prod.* **2020**, *274*, 123950. [\[CrossRef\]](#)
19. He, W.J.; Zhang, K.; Kong, Y.; Yuan, L.; Peng, Q.L.; Mulugeta Degefu, D.; Stephen Ramsey, T.; Meng, X. Reduction pathways identification of agricultural water pollution in Hubei Province, China. *Ecol. Indic.* **2023**, *153*, 110464. [\[CrossRef\]](#)
20. Liu, S.; Wang, G.; Kong, F.H.; Zhao, N.; Gao, W.K.; Zhang, H.Y. PM_{2.5} pollution characteristics, drivers, and regional transport during different pollution levels in Linyi, China: An integrated PMF-ML-SHAP framework and transport models. *J. Hazard. Mater.* **2025**, *494*, 138534. [\[CrossRef\]](#)
21. Huang, X.B.; Liu, X.S.; Jin, Y.H.; Gao, X.; Chen, Y.L. Identification and attribution analysis of integrated ecological zones based on the XGBoost-SHAP model: A case study of Chengdu, China. *Ecol. Indic.* **2025**, *177*, 113787. [\[CrossRef\]](#)
22. Liu, X.; Lu, D.W.; Zhang, A.Q.; Liu, Q.; Jiang, G.B. Data-Driven Machine Learning in Environmental Pollution: Gains and Problems. *Environ. Sci. Technol.* **2022**, *56*, 2124–2133. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Niu, X.Y.; Lu, C.P.; Zhang, Y.; Zhang, Y.; Wu, C.C.; Saidy, E.; Liu, B.; Shu, L.C. Hysteresis response of groundwater depth on the influencing factors using an explainable learning model framework with Shapley values. *Sci. Total Environ.* **2023**, *904*, 166662. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Yan, Y.B.; Yang, Y. Revealing the synergistic spatial effects in soil heavy metal pollution with explainable machine learning models. *J. Hazard. Mater.* **2025**, *482*, 136578. [\[CrossRef\]](#)

25. Yang, Q.Y.; Qian, H.; Gao, Y.Y.; Duan, Y.C.; Cao, Z.M.; Tian, P.P.; Li, K.; Yang, S.H.; Zhao, W.J.; Long, Q. Spatio-temporal evolution and driving mechanism of ecosystem services in typical hilly and gully areas of the loess Plateau: A case study in Yan'an Region, Shaanxi Province. *Ecol. Indic.* **2025**, *177*, 113773. [[CrossRef](#)]
26. Liao, Z.M.; Yue, C.; He, B.B.; Zhao, K.G.; Ciais, P.; Alkama, R.; Grassi, G.; Sitch, S.; Chen, R.; Quan, X.W.; et al. Growing biomass carbon stock in China driven by expansion and conservation of woody areas. *Nat. Geosci.* **2024**, *17*, 1127–1134. [[CrossRef](#)]
27. Cao, S.X. Impact of China's Large-Scale Ecological Restoration Program on the Environment and Society in Arid and Semiarid Areas of China: Achievements, Problems, Synthesis, and Applications. *Crit. Rev. Environ. Sci. Technol.* **2011**, *41*, 317–335. [[CrossRef](#)]
28. Li, P.Y.; Qian, H.; Wu, J.H. Environment: Accelerate research on land creation. *Nature* **2014**, *510*, 29–31. [[CrossRef](#)]
29. Cao, Z.; Li, Y.R.; Liu, Z.J.; Yang, L.F. Quantifying the vertical distribution pattern of land-use conversion in the loess hilly region of northern Shaanxi Province 1995–2015. *J. Geogr. Sci.* **2019**, *29*, 730–748. [[CrossRef](#)]
30. Li, Y.R.; Zhang, X.C.; Cao, Z.; Liu, Z.J.; Lu, Z.; Liu, Y.S. Towards the progress of ecological restoration and economic development in China's Loess Plateau and strategy for more sustainable development. *Sci. Total Environ.* **2021**, *756*, 143676.
31. Wang, S.; Fu, B.J.; Liu, J.B. Soil moisture temporal stability analysis for typical hilly and gully re-vegetated catchment in the Loess Plateau, China. *Environ. Earth Sci.* **2016**, *75*, 789. [[CrossRef](#)]
32. Jia, X.X.; Shao, M.A.; Zhu, Y.J.; Luo, Y. Soil moisture decline due to afforestation across the Loess Plateau, China. *J. Hydrol.* **2017**, *546*, 113–122. [[CrossRef](#)]
33. Lai, Y.W.; Wan, G.H.; Qin, X.X. Decoding China's new-type industrialization: Insights from an XGBoost-SHAP analysis. *J. Clean. Prod.* **2024**, *478*, 143927. [[CrossRef](#)]
34. Naderi, M.; Darzi-Naftchali, A.; Karandish, F.; Razaghian, H.; Šimůnek, J. A comparative assessment of grey water footprint estimation methods in paddy fields. *Agric. Water Manag.* **2025**, *307*, 109269. [[CrossRef](#)]
35. Huang, Y.Y.; Han, R.R.; Qi, J.C.; Duan, H.B.; Chen, C.K.; Lu, X.H.; Li, N. Health risks of industrial wastewater heavy metals based on improved grey water footprint model. *J. Clean. Prod.* **2022**, *377*, 134472. [[CrossRef](#)]
36. Çankaya, S. Evaluation of the impact of water reclamation on blue and grey water footprint in a municipal wastewater treatment plant. *Sci. Total Environ.* **2023**, *903*, 166196. [[CrossRef](#)]
37. Song, M.F.; He, W.J.; An, M.; Fang, X.; Wang, B.; Ramsey, T.S. Toward better agricultural grey water footprint allocation under economy-resource factors constraint. *Ecol. Indic.* **2023**, *154*, 110806. [[CrossRef](#)]
38. Hoekstra, A.Y.; Chapagain, A.K.; Aldaya, M.M.; Mekonnen, M.M. *The Water Footprint Assessment Manual: Setting the Global Standard*; Earthscan: London, UK, 2011.
39. GB3838-2002; Environmental Quality Standard for Surface Water. Ministry of Ecology and Environment. The People's Republic of China: Beijing, China, 2002.
40. Wahba, S.M. A consumption-based approach to trace the effects of income inequality on water pollution responsibility in Egypt: An internal grey water footprint perspective. *Ecol. Econ.* **2025**, *227*, 108404. [[CrossRef](#)]
41. Zhou, Y.B.; Wang, C.; Guo, X.Y.; Zhao, D.L.; Liu, K.; Guo, C.F. Assessment of grey water footprint pressure in the Yangtze River basin and its relationship with water quality. *Ecol. Front.* **2025**, *45*, 154–163. [[CrossRef](#)]
42. Phinzi, K.; Abriha, D.; Szabó, S. Classification efficacy using k-fold cross-validation and bootstrapping resampling techniques on the example of mapping complex gully systems. *Remote Sens.* **2021**, *13*, 2980. [[CrossRef](#)]
43. Han, J.B.; Guzman, J.A.; Chu, M.L. Prediction of gully erosion susceptibility through the lens of the SHapley Additive exPlanations (SHAP) method using a stacking ensemble model. *J. Environ. Manag.* **2025**, *383*, 125478. [[CrossRef](#)] [[PubMed](#)]
44. Hooda, S.; Mondal, P. Predictive modeling of plastic pyrolysis process for the evaluation of activation energy: Explainable artificial intelligence based comprehensive insights. *J. Environ. Manag.* **2024**, *360*, 121189. [[CrossRef](#)] [[PubMed](#)]
45. Feng, H.Y.; Sun, F.Y.; Liu, Y.Y.; Zeng, P.; Deng, L.Z.; Che, Y. Mapping multiple water pollutants across China using the grey water footprint. *Sci. Total Environ.* **2021**, *785*, 147255. [[CrossRef](#)] [[PubMed](#)]
46. Schratz, P.; Muenchow, J.; Iturrutxa, E.; Richter, J.; Breining, A. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecol. Model.* **2019**, *24*, 109–120. [[CrossRef](#)]
47. Crespo, R.; Hernandez, I. On the spatially explicit Gini coefficient: The case study of Chile-high-income developing country. *Lett. Spat. Resour. Sci.* **2020**, *13*, 37–47. [[CrossRef](#)]
48. Khan, H.; Liu, W.L.; Khan, I.; Han, L. The effect of income inequality and energy consumption on environmental degradation: The role of institutions and financial development in 180 countries of the world. *Environ. Sci. Pollut. Res.* **2022**, *29*, 20632–20649. [[CrossRef](#)]
49. Liang, X.D.; Li, J.C.; Guo, G.X.; Li, S.P.; Gong, Q.X. Urban water resource utilization efficiency based on SBM-undesirable-Gini coefficient-kernel density in Gansu Province, China. *Environ. Dev. Sustain.* **2022**, *25*, 13015–13034. [[CrossRef](#)]
50. Liu, G.; Hu, F.; Wang, Y.X.; Wang, H.M. Assessment of Lexicographic Minimax Allocations of Blue and Green Water Footprints in the Yangtze River Economic Belt Based on Land, Population, and Economy. *Int. J. Environ. Res. Public Health* **2019**, *16*, 643. [[CrossRef](#)]

51. Vu, D.H.; Muttaqi, K.M.; Agalgaonkar, A.P. A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *Appl. Energy* **2015**, *140*, 385–394. [[CrossRef](#)]
52. Hubacek, K.; Guan, D.B.; Barrett, J.; Wiedmann, T. Environmental implications of urbanization and lifestyle change in China: Ecological and Water Footprints. *J. Clean. Prod.* **2009**, *17*, 1241–1248. [[CrossRef](#)]
53. Banafa, T.; Eräranta, S.; Yletyinen, J.; Pihlajamäki, M.; Keskinen, M. Collaborative water resources management: A multilayer social network analysis. *J. Hydrol.* **2026**, *664*, 134335. [[CrossRef](#)]
54. Najafi, H.; Huang, J.J.; Naeibi, S.S.; Ramazanilar, M.; Nourani, V. Advancing group decision-making in sustainable water resource management: A Z-number enhanced framework for wastewater reuse. *J. Hydrol.* **2025**, *663*, 134256. [[CrossRef](#)]
55. Wang, Q.; Wu, Y.P.; Yao, T.Y.; Peng, Y.Y.; Chen, X.Y.; Yang, P.; Nie, S.H.; Li, J.H.; Farooq, T.H.; Wang, G.J.; et al. Assessment of water quality and microbial diversity across different land use areas in the Taojia River, China. *J. Clean. Prod.* **2025**, *523*, 146377. [[CrossRef](#)]
56. Cui, S.B.; Dong, H.J.; Wilson, J. Grey water footprint evaluation and driving force analysis of eight economic regions in China. *Environ. Sci. Pollut. Res.* **2020**, *27*, 20380–20391. [[CrossRef](#)]
57. Wang, Y.F.; Fu, B.J.; Chen, L.D.; Lü, Y.H.; Gao, Y. Check Dam in the Loess Plateau of China: Engineering for Environmental Services and Food Security. *Environ. Sci. Technol.* **2011**, *45*, 10298–10299. [[CrossRef](#)]
58. Yan, L.; Zhang, Z.D.; Zhang, J.J.; Gao, Q.; Feng, Z.G.; Abdelrahman, A.M.; Chen, Y. Effects of improving nitrogen management on nitrogen utilization, nitrogen balance, and reactive nitrogen losses in a Mollisol with maize monoculture in Northeast China. *Environ. Sci. Pollut. Res.* **2016**, *23*, 4576–4584. [[CrossRef](#)]
59. Yan, B.; Li, Y.Y.; Gao, Z.W.; Fang, Y.H.; Zhu, Y.Q.; Song, Z.B. Source apportionment of agricultural non-point source pollution in county-level regions with high urbanization rates under water resources management policy orientation. *Ecol. Indic.* **2025**, *178*, 113995. [[CrossRef](#)]
60. Guliyev, H.; Seyfullayev, I. Is the environmental Kuznets curve hypothesis valid for all countries? Fresh insights from bias-corrected dynamic panel data models. *J. Environ. Manag.* **2025**, *393*, 126960. [[CrossRef](#)]
61. Chen, Y.; He, Z.Q.; Niu, X.C.; Huang, D.C. How water resource management policies shape the coupled coordination development of the water-energy-food nexus: Evidence from the dual pathways of taxation and property rights. *J. Environ. Manag.* **2025**, *381*, 125311. [[CrossRef](#)]
62. Ming, J.; Liao, X.W.; Zhao, X. Grey water footprint for global energy demands. *Front. Earth Sci.* **2020**, *14*, 201–208. [[CrossRef](#)]
63. Qin, X.H.; Sun, C.Z.; Han, Q.; Zou, W. Grey Water Footprint Assessment from the Perspective of Water Pollution Sources: A Case Study of China. *Water Resour.* **2019**, *46*, 454–465. [[CrossRef](#)]
64. Shu, R.; Cao, X.C.; Wu, M.Y. Clarifying Regional Water Scarcity in Agriculture based on the Theory of Blue, Green and Grey Water Footprints. *Water Resour. Manag.* **2021**, *35*, 1101–1118. [[CrossRef](#)]
65. Najafi, S.; Kiani-Harchegani, M.; Dragovich, D. Check dams: Distribution, purposes and effectiveness. *Sci. Total Environ.* **2025**, *980*, 179468. [[CrossRef](#)]
66. Fan, R.; Fang, N.F.; Zeng, Y.; Zong, R.J.; Wang, Q.; Zhang, Y.; Shi, Z.H. Machine learning-assisted check dam planning on the Chinese Loess Plateau. *J. Hydrol.* **2025**, *656*, 133015. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.