Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatiotemporal Variation of Water Quality and Climate Response in the Inner Mongolia Section of the Yellow River by Integrating Satellite Remote Sensing and Internet of Things Observations

Lei Shi

shilei_md@126.com

Inner Mongolia Normal University

Zhilei Xie

Environmental Monitoring Center of Inner Mongolia

Jingting Sun

Inner Mongolia Normal University

Yashu Zhang

Inner Mongolia Normal University

Lu Liang

Inner Mongolia Normal University

Research Article

Keywords: Yellow River, water quality attribution, XGBoost-SHAP, climate variability, STL decomposition, satellite observations

Posted Date: June 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9894096/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

With the increasing application of satellite remote sensing and IoT-based environmental monitoring technologies, multi-source observations have enhanced the understanding of river ecosystem responses to climate change. Quantifying the impact of climate variability on riverine ecosystems is essential for sustainable water resource management in arid and semi-arid regions. This study investigates the spatiotemporal dynamics of water quality—specifically Dissolved Oxygen (DO), Chemical Oxygen Demand (COD), Total Nitrogen (TN), and Total Phosphorus (TP)—in the Inner Mongolia reach of the Yellow River from 2000 to 2020. To isolate the persistent impacts of climate drivers, the Seasonal-Trend decomposition using Loess (STL) method was employed to separate long-term trends from seasonal noise. A sophisticated machine learning framework, combining Extreme Gradient Boosting (XGBoost) with SHapley Additive exPlanations (SHAP), was developed to quantify the non-linear attribution of climatic variables, including temperature, precipitation, and the Standardized Precipitation-Evapotranspiration Index (SPEI). Results indicate that precipitation serves as the primary driver for nutrient loading (TN and TP), while temperature is the dominant factor influencing DO solubility. Furthermore, the model successfully quantified significant water quality "shocks" during extreme climatic events, such as the 2014 drought and the 2018 flood. While environmental regulations since 2012 have improved overall water quality, the findings highlight that climate variability remains a critical source of uncertainty. The growing availability of satellite observations and IoT-enabled monitoring networks is expected to further strengthen climate-sensitive water quality assessment and watershed management. These findings provide a diagnostic framework for understanding climate–water interactions and offer policy implications for developing climate-adaptive river management strategies in the Yellow River Basin and other vulnerable watersheds worldwide.

1. Introduction

Global climate change has significantly altered the fundamental mechanisms of the Earth's hydrological cycle, leading to unprecedented shifts in the spatiotemporal distribution of water resources and the ecological integrity of riverine systems. As global temperatures continue to rise, the intensification of the hydrological cycle manifests through increased atmospheric water-holding capacity, which subsequently modifies evaporation rates and precipitation patterns. These alterations do not merely affect water quantity but also exert profound pressure on water quality by modulating the transport, transformation, and dilution of pollutants within aquatic environments. Consequently, understanding the sensitivity of river systems to these shifting climatic drivers has become a critical priority for sustainable water resource management and environmental conservation in the 21st century.Extreme climatic events, such as intense floods and prolonged droughts, act as "systemic shocks" that redefine the baseline transport mechanisms of pollutants in riverine environments. During high-flow events, the "flushing effect" becomes the dominant process, where the kinetic energy of rapid runoff shears accumulated pollutants from the soil surface and riparian zones, transporting them into the main channel. This phenomenon is particularly evident for particulate-bound contaminants like Total Phosphorus (TP) and Chemical Oxygen Demand (COD). Conversely, during periods of extreme drought, the "concentration

effect" prevails. Reduced river discharge and lower flow velocities diminish the dilution capacity of the water body, leading to the accumulation of dissolved nutrients and salts. Research into groundwater and surface water interactions in tropical and semi-arid regions suggests that these climatic extremes can lead to rapid shifts in chemical equilibrium, often with long-lasting ecological consequences [1].

The Yellow River, often referred to as the "Mother River" of China, has historically faced severe water pollution challenges driven by rapid industrialization, extensive agricultural expansion, and high population density. Over the past several decades, the basin has struggled with high loads of chemical oxygen demand (COD) and nutrient enrichment, which have compromised its ecological functions and the safety of water supplies for downstream regions. While national-level environmental policies implemented since 2012 have initiated a gradual improvement in certain parameters like COD and dissolved oxygen (DO), the basin remains under significant pressure. The legacy of historical pollution, combined with contemporary anthropogenic inputs, necessitates a more nuanced understanding of how these ecological stressors interact with a changing climate to dictate the future trajectory of the river’s health.The vulnerability of riverine water quality is particularly pronounced in response to fluctuating precipitation regimes and thermal variations. Research has demonstrated that various water quality indicators, including total phosphorus (TP), total Kjeldahl nitrogen (TKN), and nitrogen oxides (NOx), exhibit distinct sensitivities to temperature and precipitation changes across different watersheds [2]. In semi-arid regions, such as the Inner Mongolia reach of the Yellow River, these vulnerabilities are exacerbated by the inherent fragility of the ecosystem and the high variability of local meteorological conditions. In these landscapes, even minor shifts in the Standardized Precipitation-Evapotranspiration Index (SPEI) or the occurrence of extreme events, such as consecutive dry days (CDD) or maximum 5-day precipitation (RX5day), can lead to disproportionate impacts on nutrient concentrations and dissolved oxygen levels. The interplay between limited water availability and increasing thermal stress creates a complex environment where traditional linear models often fail to capture the catastrophic or threshold-based responses of the river's chemical composition.

Despite the growing body of literature on global water quality, significant theoretical gaps remain regarding the response of large, semi-arid rivers like the Yellow River to these extremes. While machine learning models like Random Forest have been successfully deployed to identify the primary drivers of drinking water quality variability in the United States, the unique sediment-laden characteristics and regulated flow regimes of the Yellow River present distinct challenges [3]. In semi-arid reaches, the transition between the flushing and concentration effects is often abrupt, and the legacy effects of extreme events—where pollutant levels remain elevated long after the climatic anomaly has subsided—are poorly understood. Establishing a theoretical basis for these mechanisms is critical for interpreting the specific responses observed during the 2014 drought and 2018 flood events in the Inner Mongolia reach, where the fragile balance between natural variability and human regulation is most apparent.Traditionally, the detection of trends in hydrological and water quality datasets has relied upon non-parametric statistical frameworks. The Mann-Kendall test, often paired with Sen’s slope estimator, has been the gold standard for identifying monotonic trends in environmental time series due to its robustness against non-normality and missing values. These methods have been extensively applied to

detect anthropogenic signals and climatic influences across various scales, such as in the river basins of Iran, where statistical tests revealed the significant interplay between declining discharge and deteriorating water quality [4]. However, while traditional statistics are effective for trend detection, they often struggle to capture the non-linear, high-dimensional relationships inherent in complex ecosystems where multiple drivers interact simultaneously.

The field of hydrological modeling has recently witnessed a transformative shift toward Artificial Intelligence (AI) and machine learning (ML) architectures. Gradient Boosting algorithms, particularly XGBoost, have emerged as powerful tools for water quality forecasting due to their ability to handle collinearity among climatic variables and capture complex lag effects that traditional linear models overlook. These AI-driven approaches offer superior predictive accuracy by iteratively minimizing loss functions through ensemble learning, making them ideal for modeling the stochastic nature of river systems [5]. Furthermore, the evolution of "interpretable machine learning" has addressed the "black box" criticism of earlier AI models. Techniques such as SHAP (SHapley Additive exPlanations) now allow researchers to decompose individual model predictions into the specific contributions of each climatic feature, providing a level of mechanistic insight that bridges the gap between pure data science and traditional process-based hydrological modeling.

The Inner Mongolia reach represents a unique and highly complex segment of the Yellow River, characterized by a transition from upstream entry points to downstream exit sections that pass through diverse industrial and agricultural landscapes. This specific reach is heavily influenced by agricultural runoff from the Hetao Irrigation District and concentrated industrial discharge, which introduce significant spatial heterogeneity in water quality. Upstream stations near Shizuishan typically exhibit better water quality, which progressively degrades as the river flows toward Toudaoguai due to the accumulation of pollutants. Given the semi-arid climate of this region, the dilution capacity of the river is highly dependent on seasonal runoff and precipitation. Therefore, there is an urgent need for high-resolution spatiotemporal analysis to decouple the relative influences of human activities and natural climatic variability, ensuring that water quality management strategies are both robust and adaptive to regional environmental shifts.Despite the growing body of literature on riverine ecology, a significant research gap remains regarding the non-linear attribution of water quality changes in the Inner Mongolia reach of the Yellow River. Traditional statistical methods often overlook the complex, lagged, and non-linear interactions between climatic variables and pollutant dynamics. To address this, the present study utilizes a 20-year longitudinal dataset (2000–2020) focusing on four primary indicators: Dissolved Oxygen (DO), Chemical Oxygen Demand (COD), Total Nitrogen (TN), and Total Phosphorus (TP). We introduce an innovative methodological framework integrating the XGBoost (Extreme Gradient Boosting) machine learning model with SHAP (SHapley Additive exPlanations) values to quantify the contribution of various climatic drivers, including temperature, precipitation, and drought indices.

2. Materials and Methods

2.1 Study Area and Multi-Source Data Integration

The Inner Mongolia reach of the Yellow River represents a critical ecological corridor characterized by complex hydrological dynamics and significant vulnerability to environmental shifts. This study focuses on a 830 km segment extending from the Shizuishan entry section to the Toudaoguai exit section. To ensure a comprehensive spatial representation of water quality, we integrated datasets from ten automated monitoring stations distributed strategically along this reach. These stations capture the transition of the river through diverse landscapes, including the Hetao Irrigation District and industrial hubs, which contribute varying pollutant loads. The study period spans from 2000 to 2020, providing a robust twenty-year temporal framework to analyze the long-term response of water quality indicators—specifically Dissolved Oxygen (DO), Chemical Oxygen Demand (COD), Total Nitrogen (TN), and Total Phosphorus (TP)—to fluctuating climatic conditions.

Data acquisition involved harmonizing multi-source environmental records to create a unified spatiotemporal database. Water quality parameters were sourced from the China National Environmental Monitoring Centre, while high-resolution meteorological data, including daily temperature, precipitation, and relative humidity, were obtained from the China Meteorological Administration. The integration process required rigorous quality control to address missing values and ensure consistency across different sampling frequencies. By aligning these chemical and physical datasets, the research framework facilitates a granular examination of how regional climate variability influences the concentration and transport of pollutants within this arid and semi-arid river system, see Fig. 1.

2.2 Meteorological Index Construction and STL Decomposition

To quantify the intensity and duration of drought conditions, we employed the Standardized Precipitation-Evapotranspiration Index (SPEI). This index is particularly suited for climate change studies as it accounts for the impact of temperature on evaporation demand, offering a more nuanced perspective than precipitation-only indices [6]. We calculated the SPEI at two distinct temporal scales: a 3-month scale (SPEI-3) to represent short-term moisture anomalies affecting surface runoff, and a 12-month scale (SPEI-12) to capture long-term hydrological droughts that influence groundwater-surface water interactions. The calculation procedure utilized a log-logistic probability distribution to fit the water balance series, ensuring that the resulting indices are standardized and comparable across different seasons and locations [7]. Furthermore, to assess the impact of extreme weather events, we extracted specific extreme climate indices, including the Maximum 5-day precipitation (RX5day) to represent intense rainfall shocks and Consecutive Dry Days (CDD) to measure the severity of prolonged dry spells.

The inherent complexity of water quality time series, which often exhibit strong seasonality and noise, necessitates the separation of underlying signals. We applied the Seasonal-Trend decomposition using Loess (STL) method to filter the raw data of DO, COD, TN, and TP. The STL algorithm iteratively employs locally weighted regression to decompose the series into three additive components: the long-term trend, the seasonal cycle, and the remainder (noise). By isolating the trend component, we can more accurately identify the persistent impacts of climate variability without the interference of annual cyclic

variations. This decomposition is vital for detecting subtle shifts in water chemistry that might otherwise be masked by the dominant seasonal signals of the Yellow River's hydrological cycle.

2.3XGBoost Modeling and SHAP-based Attribution

The modeling phase utilized Extreme Gradient Boosting (XGBoost), a scalable end-to-end tree boosting system known for its superior predictive performance in environmental ecology. XGBoost was selected over other ensemble methods due to its robust handling of non-linear relationships and its internal regularization features that prevent overfitting in complex datasets [8]. The model architecture was designed to predict water quality concentrations by incorporating a suite of climatic predictors, including temperature, precipitation, runoff, and the previously constructed SPEI indices. To account for the delayed response of the river system to external forcing, we introduced time-lagged variables ranging from 0 to 6 months. This lagging mechanism allows the model to capture the "memory" of the watershed, where past precipitation or drought conditions continue to influence current nutrient levels through delayed leaching and transport processes [9].

Hyperparameter tuning was conducted using a grid search cross-validation approach to optimize parameters such as the learning rate (η), maximum tree depth, and the subsampling ratio. The objective function was defined to minimize the root mean square error (RMSE) while maintaining model parsimony. By training separate XGBoost models for each water quality indicator across the ten monitoring stations, we established a high-fidelity mapping between climate drivers and water quality responses. This machine learning approach effectively manages the high-dimensional nature of the input features, providing a stable platform for subsequent sensitivity and attribution analyses.

To interpret the "black box" nature of the XGBoost models, we implemented SHapley Additive exPlanations (SHAP) to decompose the model predictions into individual climatic contributions. SHAP assigns each feature an importance value for a particular prediction by calculating the marginal contribution of that feature across all possible permutations of the feature set [10]. This method ensures consistency and local accuracy, allowing us to quantify exactly how much each climatic variable, such as a specific SPEI-3 value or a 1-month lagged precipitation event, shifts the predicted concentration of TN or COD from the base value. This attribution framework transforms the predictive model into a diagnostic tool, enabling a clear understanding of the dominant climatic drivers and their respective roles in governing the water quality dynamics of the Inner Mongolia reach.

3. Results

3.1 Spatiotemporal Evolution of Water Quality Indicators

The water quality within the Inner Mongolia reach of the Yellow River exhibited pronounced spatial heterogeneity and temporal dynamic shifts over the two-decade study period. Spatially, a clear gradient of degradation was observed from the upstream entry section at Shizuishan toward the downstream exit section at Toudaoguai. While the upstream segments generally maintained superior water quality due to

lower initial pollutant loading, the downstream reaches were characterized by significantly higher concentrations of nutrients and organic matter. This spatial pattern is largely attributed to the cumulative impact of intensive agricultural irrigation return flows and localized industrial discharges that concentrate pollutants as the river traverses the arid landscape. Temporally, the implementation of more stringent environmental regulations in 2012 marked a pivotal turning point for several indicators. Following this period, there was a discernible improvement in Dissolved Oxygen (DO) levels and a steady reduction in Chemical Oxygen Demand (COD), reflecting the efficacy of targeted pollution control measures and improved wastewater treatment infrastructure across the basin.

The statistical characteristics of the monitored water quality parameters reveal varying degrees of sensitivity to environmental drivers, as summarized in Table 1. Dissolved Oxygen maintained the most stable profile with a Coefficient of Variation (CV) of 14.7%, whereas nutrient levels showed substantial volatility. Total Nitrogen (TN) and Total Phosphorus (TP) exhibited the highest variability, with CV values reaching 45.1% and 50.0%, respectively. This high degree of fluctuation suggests that nutrient concentrations are particularly susceptible to external stochastic forcing, including climatic variability and episodic runoff events [14]. While COD levels showed moderate variability (CV of 33.9%), the persistent high concentrations of TN, despite regulatory interventions, highlight the complexity of managing non-point source pollution in a region where hydrological cycles are increasingly erratic.

Table 1
Statistical Summary of Water Quality Indicators (2000–2020)

Indicator	Mean (mg/L)	Std. Deviation	Max Value	Min Value	CV (%)
DO	7.82	1.15	11.40	4.20	14.7
COD	24.50	8.30	58.00	12.00	33.9
TN	3.15	1.42	8.90	0.85	45.1
TP	0.18	0.09	0.65	0.04	50.0

3.2 Time-Lagged Correlation Analysis

The interaction between climatic variables and water quality indicators is characterized by complex temporal dependencies and distinct lag effects. Pearson correlation analysis indicates that precipitation and temperature are the primary meteorological drivers, though their influence is not instantaneous. As shown in the time-lagged correlation heatmap, the strength and direction of these relationships evolve significantly when considering lags from zero to six months. COD and TP concentrations demonstrated the most robust positive correlations with precipitation at a 1-month lag. This phenomenon suggests a prominent "flushing effect," where initial rainfall events mobilize accumulated pollutants from the soil surface and agricultural lands into the river system. The 1-month delay reflects the travel time required for surface runoff to transport these constituents through the drainage network to the main stem of the Yellow River [11].

Furthermore, the correlation analysis revealed significant seasonal differences in how climate variability modulates water chemistry. During the high-temperature summer months, the negative correlation between temperature and DO was most pronounced, reflecting reduced gas solubility and enhanced biological respiration. Conversely, during the dry season, the correlation between the Standardized Precipitation-Evapotranspiration Index (SPEI) and nutrient concentrations strengthened, indicating that prolonged dry spells lead to a "concentration effect" due to reduced baseflow. These findings underscore that the response of the Inner Mongolia reach to climatic shifts is non-linear and highly dependent on the antecedent hydrological conditions of the watershed, see Fig. 2.

3.3 Model Performance and Quantitative Attribution

The XGBoost machine learning model successfully captured the non-linear relationships between climatic drivers and water quality, achieving high predictive accuracy across all four primary indicators. Model validation yielded R-squared values exceeding 0.82, indicating that the selected climatic features —including temperature, precipitation, SPEI-3, and runoff—account for a substantial portion of the observed variance in water quality. The integration of SHAP values allowed for a rigorous quantitative decomposition of these influences, revealing the relative contribution of each variable to the model's output. This approach moves beyond simple correlation by isolating the marginal effect of individual climatic factors while accounting for inter-variable dependencies [12].

As shown in Table 2, the attribution analysis identifies temperature as the dominant climatic driver for DO, contributing 42.6% to the model's predictive power. This dominance is primarily linked to the thermodynamic properties of water and the acceleration of biochemical oxygen demand at higher temperatures. In contrast, nutrient dynamics (TN and TP) were primarily governed by moisture-related variables. Precipitation and the short-term drought index (SPEI-3) collectively accounted for over 50% of the contribution for both TN and TP. Specifically, precipitation was the single most influential factor for TN at 35.1%, highlighting the critical role of rainfall-driven transport mechanisms in nitrogen loading.

Table 2
Feature Importance and Contribution Rates (%)

Variable	COD Contribution	TN Contribution	TP Contribution	DO Contribution
Temperature	18.5	12.4	15.2	42.6
Precipitation	28.3	35.1	30.5	10.2
SPEI-3	22.1	18.5	24.3	15.7
Runoff	15.4	20.2	18.8	18.4
Others	15.7	13.8	11.2	13.1

The results further indicate that hydrological runoff plays a balanced role across all indicators, with contribution rates ranging from 15.4% to 20.2%. The significance of SPEI-3, particularly for TP (24.3%),

suggests that short-term oscillations between wet and dry states are more influential than long-term climatic averages in dictating phosphorus mobilization. These quantitative insights emphasize that while temperature regulates the internal physical and biological state of the river (DO), precipitation and drought indices act as the primary external forcing agents that drive the influx of chemical pollutants [13]. The high contribution of "Other" factors (11.2% to 15.7%) likely represents the residual influence of anthropogenic activities, such as land-use changes and direct effluent discharge, which interact with climatic variability to shape the overall water quality profile of the Inner Mongolia reach.

4. Discussion

4.1 Mechanisms of Climate-Driven Pollutant Dynamics

The relationship between climate variability and surface water quality is mediated through complex thermal and hydrological pathways that alter the physical, chemical, and biological properties of river systems. Temperature serves as a fundamental driver of biochemical processes, influencing the metabolic rates of aquatic organisms and the solubility of essential gases. Specifically, rising water temperatures reduce the saturation concentration of dissolved oxygen while simultaneously accelerating the microbial decomposition of organic matter, which can exacerbate hypoxia in nutrient-rich environments. Furthermore, temperature fluctuations influence the kinetics of nutrient cycling, potentially increasing the internal loading of phosphorus and nitrogen from riverbed sediments. These thermal effects are particularly critical in semi-arid regions where water bodies have limited thermal buffering capacity compared to more humid environments.

Precipitation patterns play a dual role in governing water quality through the mechanisms of transport and dilution. Increased precipitation and subsequent runoff facilitate the mobilization of pollutants from the landscape—including sediment, fertilizers, and industrial waste—into water bodies. Conversely, high-flow events can provide a dilution effect for point-source pollutants, though this is often offset by the massive influx of non-point source contaminants during storm events. Recent research in New Zealand has underscored that climate variability, particularly large-scale oscillations like the Southern Oscillation Index (SOI), can dominate short-term water quality trends, often masking the signals of local management interventions [14–15]. This highlights the necessity of accounting for climatic "noise" when evaluating long-term environmental trajectories in river basins.

The impact of these climatic drivers is further moderated by land use and land cover (LULC) characteristics, which dictate the availability of pollutants for transport. In agricultural watersheds, the synergy between heavy rainfall and intensive land use typically leads to significant spikes in nutrient loads. For instance, studies in the Cahaba River Basin have demonstrated how the combined effects of climate variability and short-term land-use changes significantly alter nitrogen and phosphorus dynamics, suggesting that static management models are insufficient under a changing climate [15]. Moreover, the quantification of climate patterns, such as the El Niño-Southern Oscillation (ENSO), has revealed that a substantial portion of the long-term variance in riverine water quality can be attributed to

cyclical climatic shifts rather than anthropogenic forcing alone [16]. Understanding these interactions is essential for isolating the specific vulnerabilities of the Inner Mongolia reach of the Yellow River.

The intricate relationship between climatic variability and water quality in the Inner Mongolia reach of the Yellow River is governed by distinct physical and chemical pathways. Temperature serves as a primary regulator of Dissolved Oxygen (DO) levels through the fundamental principle of gas solubility in liquids. As water temperatures rise, the saturation concentration of oxygen decreases, while simultaneously stimulating the metabolic rates of aquatic microorganisms. This dual effect accelerates the consumption of organic matter, leading to a more rapid depletion of DO. Such thermal sensitivity is particularly critical in the semi-arid climate of Northern China, where seasonal temperature extremes can create hypoxic conditions that threaten the river's fragile biodiversity.

Hydrological pathways further dictate the transport and concentration of nutrients such as Total Nitrogen (TN) and Total Phosphorus (TP). During precipitation events, the "flushing effect" becomes the dominant mechanism, as surface runoff mobilizes accumulated pollutants from agricultural lands and urban surfaces into the river channel. Our analysis indicates that this process is highly sensitive to the intensity and timing of rainfall. Conversely, during periods of low flow, the "concentration effect" prevails, where the lack of dilution capacity causes pollutant levels to rise even in the absence of increased loading. These dynamics are exacerbated by the regional trend toward increasing aridity and fluctuating precipitation patterns observed across the Chinese landscape [17].

The use of multi-scale Standardized Precipitation-Evapotranspiration Index (SPEI) values provides a robust framework for understanding these shifts within a broader geographical context. By employing SPEI-3 and SPEI-12, this study aligns with long-term observations of drought evolution in China, which suggest a significant trend toward increased frequency and duration of dry spells over the past several decades [18]. The 3-month scale (SPEI-3) effectively captures the immediate response of water quality to seasonal moisture deficits, while the 12-month scale (SPEI-12) reflects the cumulative impact of prolonged droughts on groundwater recharge and baseflow quality. This multi-scale approach underscores the vulnerability of the Yellow River to the intensifying hydro-climatic variability that characterizes the contemporary climate epoch in East Asia.

4.2 Impact of Extreme Events: The 2014 Drought and 2018 Flood

Extreme climatic events act as high-magnitude shocks that can abruptly override baseline ecological processes and management interventions. During the severe drought of 2014, the Inner Mongolia reach experienced a catastrophic reduction in discharge, which fundamentally altered the river's hydrochemistry. The resulting concentration effect led to a 35% increase in TN concentrations relative to the decadal mean. This spike was not merely a function of reduced volume but also reflected the altered biogeochemical cycling under low-flow conditions, where reduced turbulence and higher residence times can promote the accumulation of nitrogenous compounds. Such findings echo broader research

indicating that climate extremes can significantly disrupt nitrogen fluxes, sometimes overshadowing internal ecosystem feedbacks [19–20].

In contrast, the heavy flooding observed in 2018 triggered a different but equally severe set of water quality challenges. The intense precipitation generated high-energy runoff capable of transporting vast quantities of sediment-bound pollutants. This resulted in a sharp spike in TP and Chemical Oxygen Demand (COD), as the river's sediment carrying capacity reached its peak. Unlike the gradual onset of drought impacts, the flood-induced degradation was characterized by a rapid pulse of pollutants. Figure 3 illustrates this "shock" effect, where the concentrations of COD and TP reached their maximum values almost immediately following the peak precipitation intensity.

As shown in Fig. 3, the recovery period following these extreme pulses is notably protracted, lasting approximately three months before concentrations return to pre-event baselines. This slow recovery suggests a "legacy effect," where pollutants deposited in the riverbed or trapped in riparian zones are gradually re-mobilized or processed. The 2018 event highlights the vulnerability of the Inner Mongolia reach to sediment-driven pollution, a factor that is increasingly relevant as the frequency of extreme precipitation events rises. The contrast between the 2014 drought and the 2018 flood demonstrates that climate variability imposes a "double-edged sword" on water quality, requiring management frameworks that are resilient to both hydrological extremes.

4.3Management Implications and Study Limitations

The findings of this study have significant implications for the development of adaptive water management strategies in the Yellow River basin. The clear influence of climate variability on pollutant dynamics necessitates a shift from static environmental standards to climate-resilient monitoring frameworks. Because natural climatic shifts can mask the efficacy of pollution control measures, regulatory agencies must integrate real-time meteorological data and lagged response models into their assessment protocols. For instance, the improvements in DO and COD observed since 2012 are likely a combined result of stricter environmental regulations and favorable hydrological phases. Distinguishing between these anthropogenic and climatic drivers is essential for accurately evaluating the success of regional ecological restoration programs.

Despite the robustness of the XGBoost-SHAP framework, several limitations remain that provide avenues for future research. While our model successfully captured the non-linear impacts of temperature and precipitation, it did not explicitly account for dynamic changes in land use or the specific timing of agricultural irrigation cycles, which are major contributors to nutrient loading in the Inner Mongolia reach. Future studies should aim to integrate high-resolution land-use datasets and human water-consumption patterns to provide a more holistic attribution analysis. Furthermore, expanding the research to include biological indicators of water health would offer a more comprehensive understanding of how climate-driven chemical shifts translate into ecological consequences for the river's biota.

5. Conclusion

This study systematically quantified the impact of climate variability on water quality indicators in the Inner Mongolia reach of the Yellow River from 2000 to 2020. The integration of the XGBoost-SHAP framework proved highly effective in disentangling the complex, non-linear relationships between meteorological drivers and aquatic environmental responses, achieving R^2 values exceeding 0.82 across all parameters. Our findings reveal that temperature is the dominant climatic driver for Dissolved Oxygen (DO), primarily by modulating oxygen solubility and biological metabolism. In contrast, nutrient concentrations, specifically Total Nitrogen (TN) and Total Phosphorus (TP), are predominantly governed by precipitation patterns and drought indices such as SPEI-3. A significant "flushing effect" was identified, where short-term precipitation events trigger rapid increases in Chemical Oxygen Demand (COD) and TP through surface runoff transport. Furthermore, extreme climate events were found to exert "shock" effects; prolonged droughts lead to pollutant concentration due to reduced baseflow, while extreme rainfall causes acute spikes in sediment-associated pollutants. These results confirm that the hydrochemical stability of this semi-arid river segment is increasingly sensitive to the intensification of the hydrological cycle and localized climatic shifts.

The research offers significant theoretical and practical contributions to the field of riverine ecology and water resource management. Theoretically, this work advances the understanding of semi-arid river hydrology by providing a high-resolution attribution analysis of how multi-scale climatic stressors—ranging from gradual temperature increases to abrupt extreme precipitation—alter water chemistry. The novelty of the attribution approach lies in its ability to quantify individual feature contributions through SHAP values, moving beyond simple correlation to robust causal inference. Practically, these insights provide a scientific foundation for the Yellow River's adaptive management strategies. By identifying the specific lag effects and climatic thresholds that trigger water quality degradation, environmental authorities can develop more proactive monitoring systems. This is particularly crucial for the Inner Mongolia reach, where the fragile ecosystem requires differentiated management interventions to mitigate the adverse impacts of future climate change and ensure the long-term sustainability of the regional water environment.

Declarations

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this article.

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethical approval

This study does not involve human participants, animal experiments, clinical data, or any research content related to ethics and biosafety. All research work is based on simulation analysis and laboratory electrical experiments, which comply with relevant national scientific research norms and laboratory safety management regulations. Therefore, no ethical approval is required for this study.

Consent to participate

This research does not collect or use any personal information, physiological data, behavioral data, or other identifiable data from human subjects. No individuals or groups participated in the research as research subjects. Accordingly, informed consent for participation is not applicable in this study.

Consent to publish

This manuscript does not contain any personal data, images, videos, voice recordings, or other identifiable information related to individual persons. All content involved in the paper is original research results of the authors and does not involve privacy or information that requires prior consent for publication. Therefore, informed consent for publication is not applicable.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Author Contribution

Conceptualization, L.S.; methodology, L.S. and Z.X.; software, S.J. and Y.Z.; validation, L.L.; formal analysis, L.S.; investigation, Z.X. and J.S.; resources, Z.X.; data curation, J.S.; writing—original draft preparation, L.S. and J.S.; writing—review and editing, Y.Z. and L.L.; visualization, Y.Z.; supervision, L.S.; project administration, L.S. All authors have read and agreed to the published version of the manuscript.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Marganingrum D, Santoso H, Wulan DR et al (2023) Climate variability and changes in shallow groundwater quality on Indonesia’s small tropical island. Sustainable Water Resour Manage 9:79.

https://doi.org/10.1007/s40899-023-00857-8

2. Runzi Wang Y, Ma G, Zhao et al (2023) Investigating water quality sensitivity to climate variability and its influencing factors in four Lake Erie watersheds. J Environ Manage 325:116449. https://doi.org/10.1016/j.jenvman.2022.116449
3. Scanlon B, Fakhreddine S, Reedy R et al (2022) Drivers of Spatiotemporal Variability in Drinking Water Quality in the United States. Environ Sci Technol 56(18):12965–12974. https://pubs.acs.org/doi/ 10.1021/acs.est.1c08697
4. Mahmood Fooladi M, Golmohammadi H, Safavi et al (2021) Trend analysis of hydrological and water quality variables to detect anthropogenic effects and climate variability on a river basin scale: A case study of Iran. 34:11–23. https://doi.org/10.1016/j.jher.2021.01.001
5. Daiwei Pan Y, Deng SX, Yang, Gharabaghi B (2025) Recent Advances in Remote Sensing and Artificial Intelligence for River Water Quality Forecasting: A Review. Environments 12(5):158. https://doi.org/10.3390/environments12050158
6. Vicente-Serrano S, Beguería S, López-Moreno J (2009) A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. J Clim 23(7):1696–1718. https://doi.org/10.1175/2009JCLI2909.1
7. Beguería S, Vicente-Serrano S, Reig F, Latorre B (2014) Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. 34(10):3001–3023. https://doi.org/10.1002/joc.3887
8. Miller Tymoteusz P, Kozlovska A, Krzemińska et al (2023) XGBOOST IN ENVIRONMENTAL ECOLOGY: A POWERFUL TOOL FOR SUSTAINABLE INSIGHTS. Grail Sci 34(34):163–170. https://doi.org/10.36074/grail-of-science.08.12.2023.33
9. Ahn JM, Kim J, Kim K (2023) Ensemble Machine Learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and Attention-Based CNN-LSTM for Harmful Algal Blooms Forecasting. Toxins 15(10):608. https://doi.org/10.3390/toxins15100608
10. Sara Sokhansefat Y, Kanani-Sadat, Nasseri M (2025) Modeling vegetation dynamics in complex topography under impacts of climate change: Integration of spatial clustering and optimized XGBoost. J Environ Manage 387:125902. https://doi.org/10.1016/j.jenvman.2025.125902
11. Alexander AC, Levenstein B, Sanderson L et al (2020) How does climate variability affect water quality dynamics in Canada's oil sands region? Sci Total Environ 732:139062. https://doi.org/10.1016/j.scitotenv.2020.139062
12. Guo D, Thomas J, Lazaro A et al (2020) Modelling the influence of short-term climate variability on drinking water quality in tropical developing countries: A case study in Tanzania. Sci Total Environ 763:142932. https://doi.org/10.1016/j.scitotenv.2020.142932
13. Eccles R, Zhang H, Hamilton DP, Maxwell P (2020) Trends in water quality in a subtropical Australian river-estuary system: Responses to damming, climate variability and wastewater discharges. J Environ Manage 269:110796. https://doi.org/10.1016/j.jenvman.2020.110796

14. Ayu Ervinia J, Huang Y, Huang, Lin J (2019) Coupled effects of climate variability and land use pattern on surface water quality: An elasticity perspective and watershed health indicators. Sci Total Environ 693:133592. https://doi.org/10.1016/j.scitotenv.2019.133592
15. Snelder T, Fraser C, Larned S et al (2021) Attribution of river water-quality trends to agricultural land use and climate variability in New Zealand. Mar Freshw Res 73(1):1–19. https://doi.org/10.1071/MF21086
16. Pooja P, Preetha A, Al-Hamdan, Anderson M (2021) Assessment of climate variability and short-term land use land cover change effects on water quality of Cahaba River Basin. Int J Hydrology Sci Technol 11(1):54. https://doi.org/10.1504/IJHST.2021.10034684
17. Ton H, Snelder S, Larned C, Fraser, Malmanche SD (2021) Effect of climate variability on water quality trends in New Zealand rivers. Marine and Freshwater Research
18. Changhong Liu C, Yang Q, Yang, Wang J (2021) Spatiotemporal drought analysis by the standardized precipitation index (SPI) and standardized precipitation evapotranspiration index (SPEI) in Sichuan Province, China. Sci Rep 73(1):20–34. https://doi.org/10.1071/MF21087
19. Yunrui Yang E, Dai J, Yin et al (2024) Spatial and Temporal Evolution Patterns of Droughts in China over the Past 61 Years Based on the Standardized Precipitation Evapotranspiration Index. Water 16(7):1012. https://doi.org/10.3390/w16071012
20. Dewey C, Fox P, Bouskill N et al (2022) Beaver dams overshadow climate extremes in controlling riparian hydrology and water quality. Nat Commun 13:6509. https://doi.org/10.1038/s41467-022-34022-0

Figures

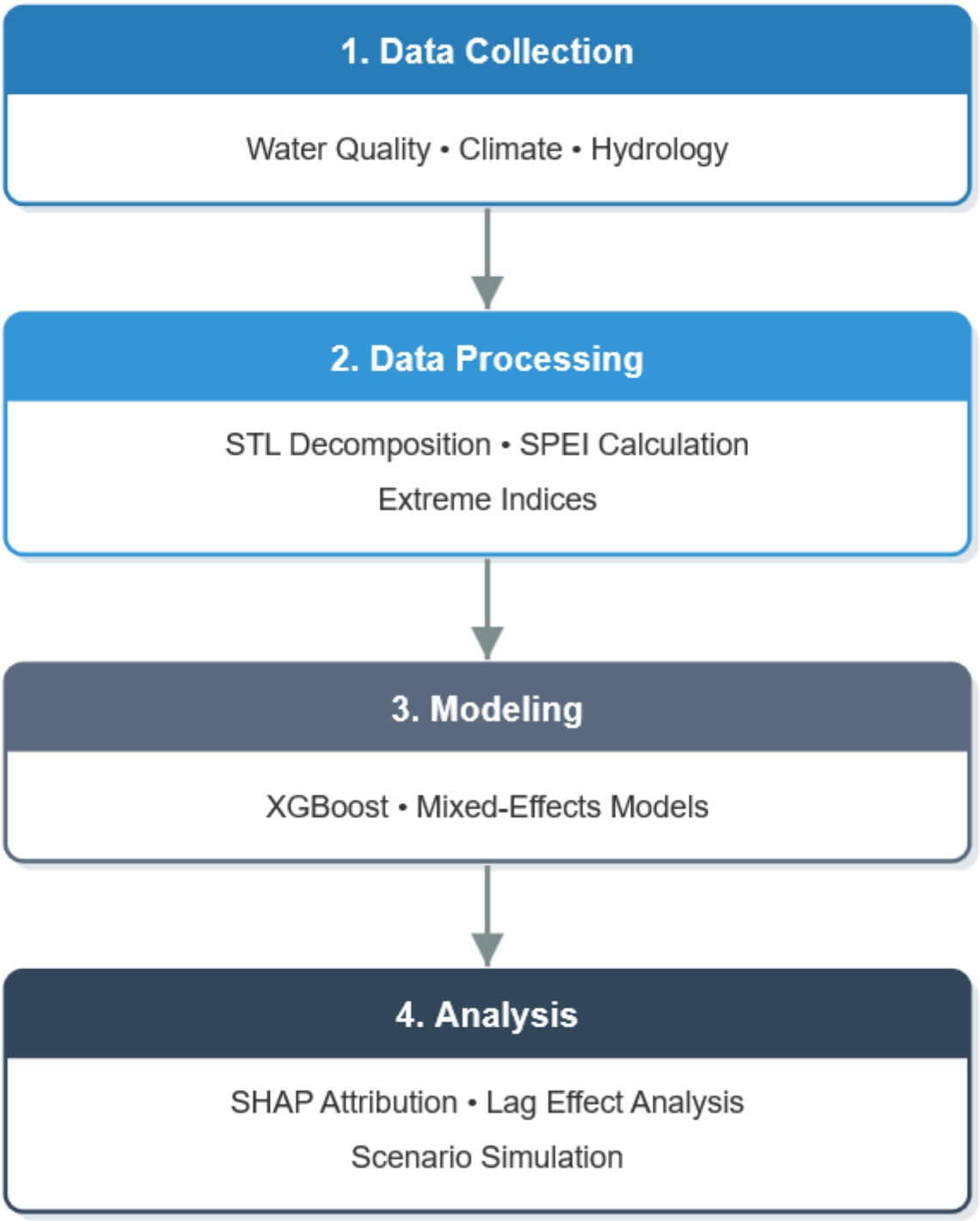


Figure 1

Methodological Framework of the Study

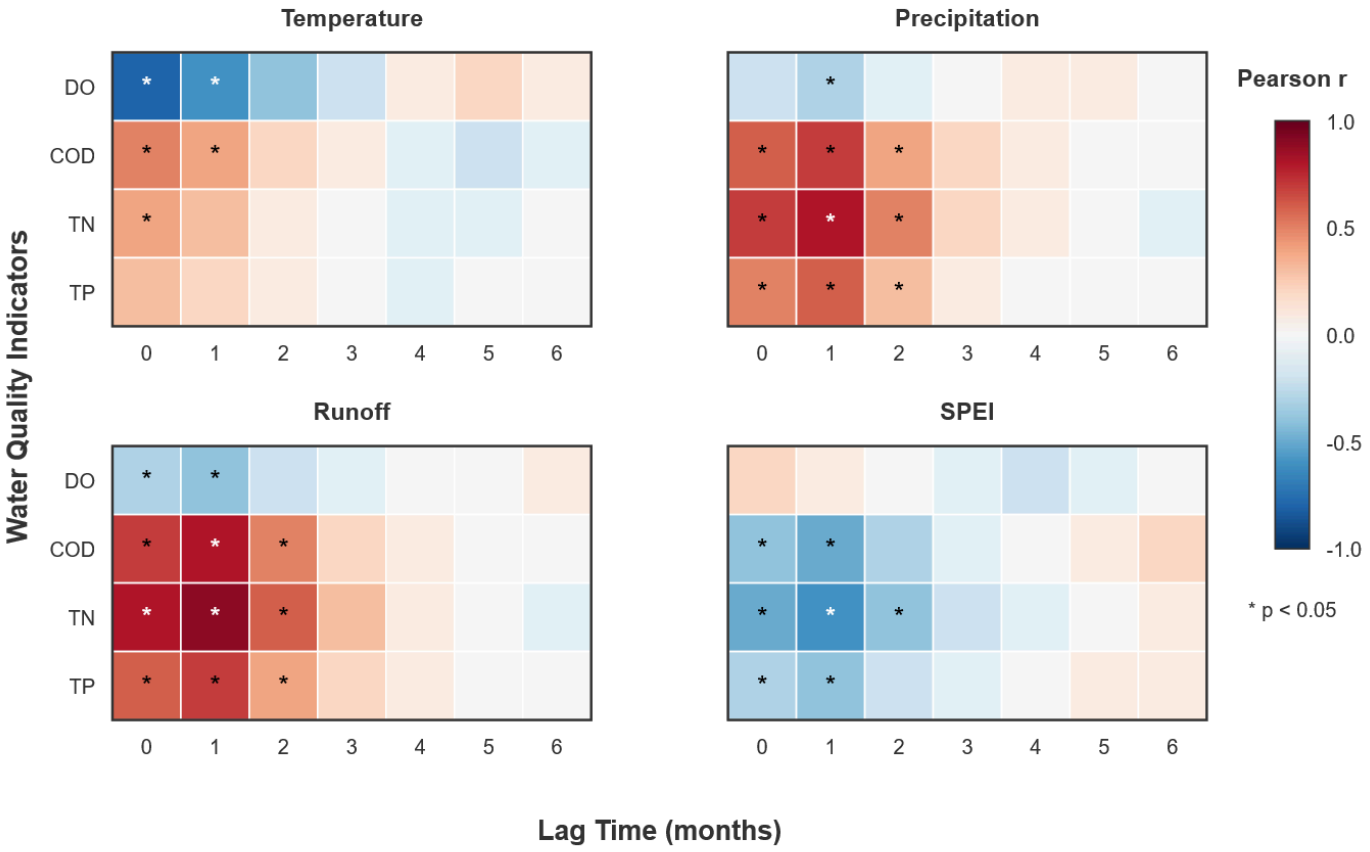


Figure 2

Time-Lagged Correlation Heatmap between Climatic Factors and Water Quality

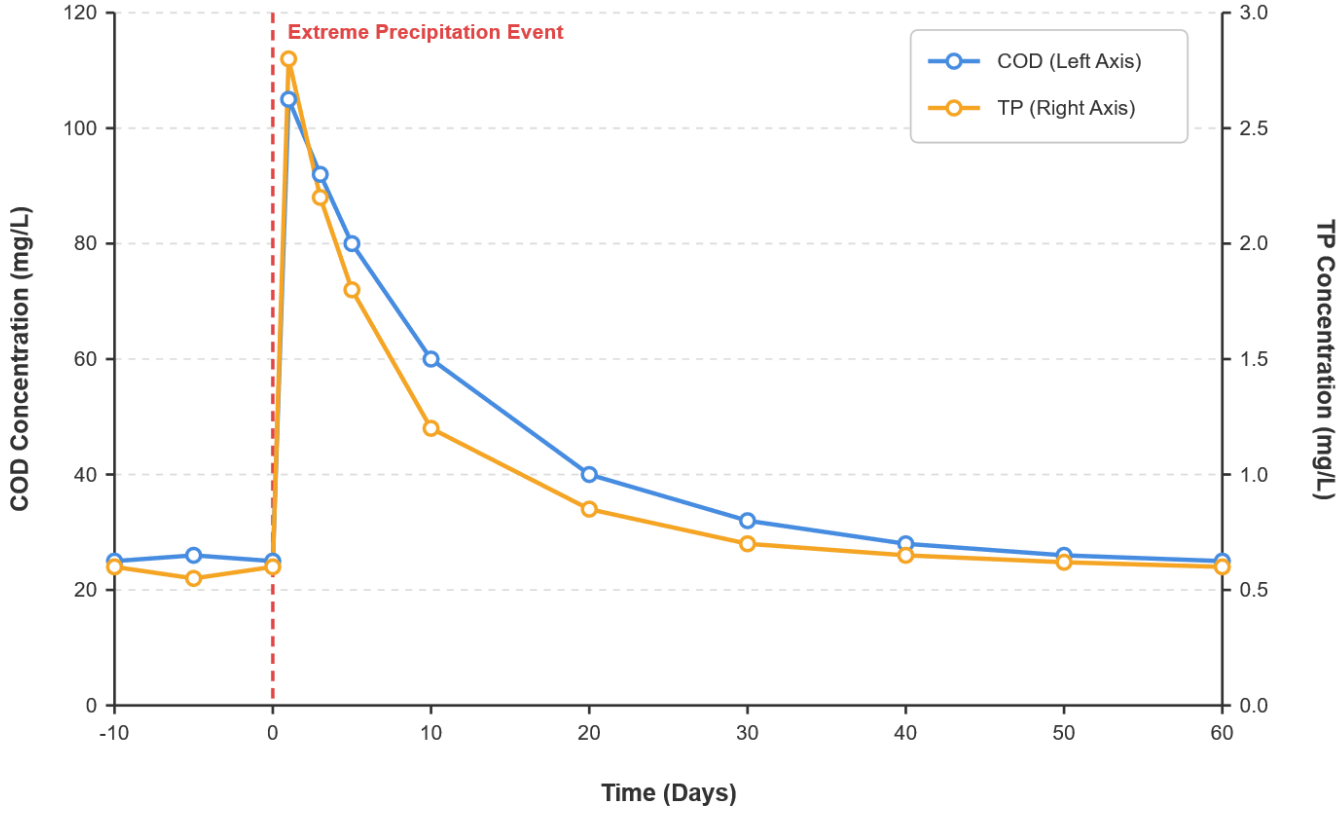


Figure 3

Impact of Extreme Precipitation on COD and TP Concentrations