S & M 4539

Remote Sensing Inversion Method and Spatiotemporal Distribution of Chlorophyll-a Concentration in Reservoirs of Hebei Province Using Multispectral and Environmental Parameters

Minghua Zhang,[1] Tianyu Qin,[1*] Yu Li,[2] Jiayuan Zhang,[3]
Zheng Zhao,[1] Mengyu Zhang,[1] Shiyuan Zhu,[2] and Ying Jiao[2]

[1]Hebei Provincial Ecological Environmental Monitoring Center, Shijiazhuang 050051, China
[2]School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China
[3]College of Environmental Sciences and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China

(Received March 10, 2026; accepted June 29, 2026)

Keywords: chlorophyll-a concentration inversion, environmental parameters, multisource data fusion, machine learning, reservoir water quality

To address the limitations of traditional remote sensing methods, which often inadequately consider environmental parameters and result in limited accuracy for Chlorophyll-a (Chl-a) concentration inversion in inland waters, in this study, we developed a high-precision multisource inversion method tailored for reservoirs in Hebei Province. Sentinel-2 multispectral sensor data were integrated with three key environmental parameters—water temperature (Temp), pH, and electrical conductivity (EC)—to construct enhanced multisource features. Various machine learning models were systematically evaluated and optimized to improve both model robustness and predictive accuracy. Results demonstrate that incorporating environmental parameters significantly strengthens feature representational capacity, with correlation coefficients between multisource combinations and Chl-a reaching 0.830. The optimal linear regression model based on the Temp × pH × Mean_Rbands feature achieved a testing-set coefficient of determination (R^2) of 0.847. Across all multisource features, the mean testing-set R^2 was 0.700 ± 0.087, compared with 0.339 ± 0.164 for spectral-only methods. The spatiotemporal analysis of the 2025 growing season revealed a distinct seasonal rhythm—increasing in spring and summer, and differentiating in autumn. Spatially, Lincheng Reservoir recorded the highest annual average concentration (0.0103 mg/L), with near-shore areas identified as sensitive high-concentration zones. This multisource fusion strategy provides robust methodological support for high-precision water quality monitoring and dynamic eutrophication assessment in inland reservoirs.

*Corresponding author: e-mail: qintianyu7@163.com
https://doi.org/10.18494/SAM6330

ISSN 0914-4935 © MYU K.K.
https://myukk.org/

1. Introduction

Water quality monitoring serves as a cornerstone for aquatic environmental protection, ecological balance maintenance, and the sustainable utilization of water resources. As a primary indicator of water eutrophication and algal growth, the concentration of Chlorophyll-a (Chl-a) is a critical metric for assessing water quality,[1] and monitoring Chl-a is important for water-quality evaluation and the early warning of algal blooms.[2] Under the escalating pressures of anthropogenic activities and climate change, eutrophication in inland water bodies, such as reservoirs, has emerged as a prominent environmental challenge.[3] Therefore, the effective monitoring of Chl-a is a prerequisite for scientific early warning and remediation efforts.[4]

Traditional in-situ monitoring methods are often characterized by high costs and limited spatial coverage, making it difficult to satisfy the demands for dynamic spatiotemporal monitoring.[5,6] Although remote sensing offers the advantages of wide-scale and high-efficiency observation, traditional inversion methods based on water color remote sensing are susceptible to optical interference in complex inland waters, which compromises their accuracy and robustness.[7,8] Consequently, developing high-precision and efficient remote sensing inversion methods has become a focal point of current research.[9,10] The significant spectral response of Chl-a in red and NIR bands provides the physical basis for remote sensing inversion.[11] However, Chl-a concentration is governed not only by optical properties but also directly by the physiological status of algal growth.[12]

The results of existing studies have demonstrated that environmental parameters, such as water temperature (Temp), pH, and electrical conductivity (EC), significantly affect the growth and development of phytoplankton, thereby indirectly modulating Chl-a concentrations.[13,14] Variations in Temp and pH not only affect algal proliferation but also regulate the availability of nutrients in the water column.[15] Specifically, elevated temperatures and fluctuating pH levels can alter the photosynthesis and nutrient uptake of aquatic vegetation, ultimately shifting Chl-a levels.

Despite preliminary attempts to integrate environmental parameters with spectral data for water quality monitoring, most research remains confined to the application of single bands or spectral indices. However, a systematic modeling framework that jointly uses multisource spectral data and environmental parameters to estimate Chl-a concentration is still lacking.[4,7,16] For instance, although Wang *et al.* improved inversion accuracy by combining Sentinel-2 data with environmental parameters,[1] and Pan *et al.* optimized the inversion process using modified equivalent spectra (MES) and machine learning,[17] critical limitations remain. In particular, without a comprehensive framework to synthesize environmental data and remote sensing spectra, the complex relationship between phytoplankton growth and Chl-a concentration cannot be adequately characterized.[4,16] Thus, leveraging multisource data and advanced algorithms to enhance the precision of water quality monitoring remains a critical research priority.

To address the above gaps, this study is focused on six typical drinking water reservoirs in Hebei Province. The primary objective is to develop a high-precision remote sensing inversion method for Chl-a by integrating Sentinel-2 multispectral data with

Temp, pH, and EC. By systematically constructing multisource feature combinations and comparing various regression and machine learning models, we aim to identify the optimal inversion scheme and validate the efficacy of multisource data fusion in enhancing accuracy for regional water-quality monitoring applications. Furthermore, the optimal model will be applied to the study area to reveal the spatiotemporal distribution and seasonal evolution characteristics of Chl-a during the 2025 growing season. This research is intended to provide a methodological reference for the operational remote sensing monitoring of reservoir water quality in Hebei Province and insights for Chl-a inversion in similar inland water bodies.

2. Data, Materials, and Methods

2.1 Study area

In this study, six typical reservoirs in Hebei Province were selected as the primary research subjects: Huangbizhuang, Shihe, Douhe, Longmen, Lincheng, and Zhuzhuang. These reservoirs serve as centralized surface water sources for drinking water and play a pivotal role in regional water supply and ecological security. Historical monitoring data indicate that while the overall water quality remains satisfactory—primarily characterized by a mesotrophic state—there is still a latent risk of eutrophication. These reservoirs are distributed across various river basins, range in size from medium to large, and fulfill diverse functions including water supply, flood control, and irrigation. The dynamic monitoring of water quality in these reservoirs provides critical insights into the regional aquatic environmental health. Therefore, investigating Chl-a concentration through remote sensing inversion and analyzing its spatiotemporal distribution in these specific sites are essential for providing a scientific basis for water quality supervision and early warning systems in drinking water source areas. The geographical locations of the six reservoirs are shown in Fig. 1.

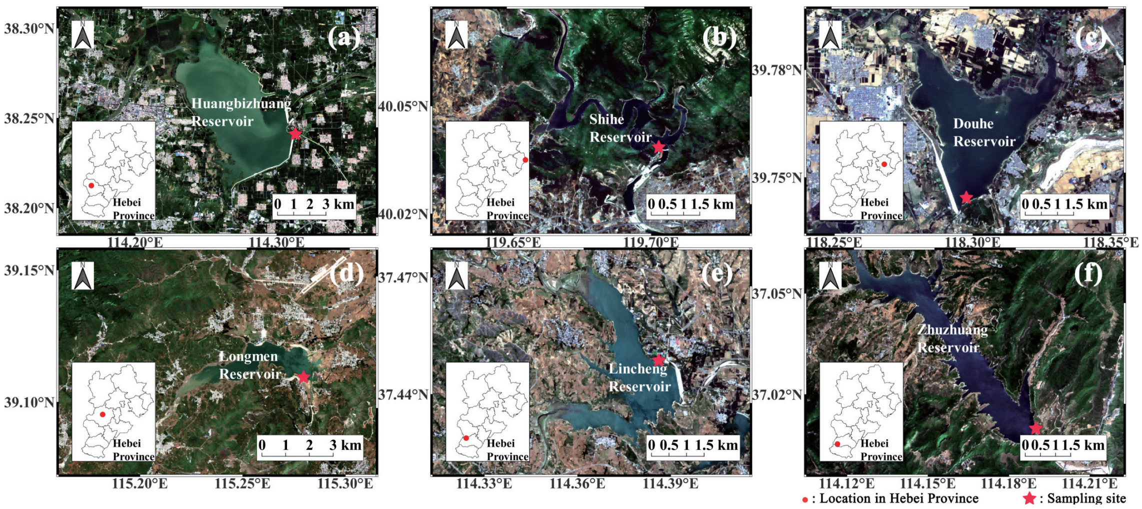

Fig. 1. (Color online) Geographical locations of study areas.

2.2 Data acquisition and preprocessing

The dataset used in this study comprises field-measured water quality parameters and satellite remote sensing imagery.

Field-measured data: A total of 175 valid monthly water samples were collected from six typical reservoirs in Hebei Province—Huangbizhuang, Shihe, Douhe, Longmen, Lincheng, and Zhuzhuang—between January 2020 and September 2025. The sampling covered multiple years and seasons during the ice-free period, thereby capturing temporal variability across the major monitoring months of the reservoirs. No samples were collected during the ice-covered period, because surface ice not only hinders field access and *in situ* sampling, but also exhibits spectral reflectance characteristics markedly different from those of open water, making such observations unsuitable for Chl-a inversion modeling. Water samples were collected within 1 m of the surface and analyzed in the laboratory following standard protocols. Chl-a concentration was determined using the spectrophotometric method specified in HJ 897–2017 (Water Quality—Determination of Chlorophyll a—Spectrophotometric Method), and water temperature was measured using a thermometer in accordance with GB/T 13195–1991 (Water Quality—Determination of Water Temperature—Thermometer or Reversing Thermometer Method). pH and EC were measured using electrode methods in accordance with national environmental standards.

Satellite data: Accurate radiometric and atmospheric correction is essential for the reliable remote sensing inversion of water environmental parameters in inland waters.[18] Sentinel-2 L2A imagery (10 and 20 m spatial resolutions) was selected for its high revisit frequency and spectral suitability. The Level-2A bottom-of-atmosphere (BOA) reflectance products were used directly, and their atmospheric correction was implemented within the Copernicus operational processing chain using the Sentinel-2 atmospheric correction processor (Sen2Cor). Fifty-two cloud-free images were obtained from the European Space Agency Copernicus Data Space Ecosystem, with an acquisition window within ±6 days of field sampling. This matching window was used because same-day cloud-free Sentinel-2 observations were often unavailable, and it represented a practical balance between temporal consistency and sample availability for the studied reservoirs. Image preprocessing included resampling all bands to 10 m and extracting surface reflectance values from a 3 × 3 pixel window centered on each sampling site to mitigate positioning errors. Additionally, a water mask was generated using the modified normalized difference water index (MNDWI) to precisely delineate the reservoir boundaries. This procedure helped reduce potential adjacency effects.

2.3 Multisource feature construction and correlation analysis

To systematically establish a feature framework for Chl-a inversion and ensure a comprehensive model comparison, a hierarchical and progressive strategy was adopted for feature construction. First, multisource feature combinations encompassing univariate, binary, and ternary variables were constructed to fully explore the synergistic effects between multispectral data and environmental parameters. In these feature expressions, the notation reflects direct arithmetic combinations between variables, whereas underscores are used only within variable names.

In the univariate modeling stage, in addition to using single-band reflectance directly, a series of commonly used water quality remote sensing indices were introduced as benchmarks to evaluate the relative performance of the proposed multisource features. These indices and their corresponding inversion algorithms are summarized in Table 1.

Following the construction of multisource feature combinations, the Pearson correlation coefficient (r) was calculated to quantify the linear relationship between variables. The coefficient r is a standard statistical metric used to measure the strength of the association between two variables, calculated as follows:

Table 1
Commonly used remote sensing indices and inversion algorithms for Chl-a inversion.

Remote sensing index	Expression	Inversion model	Description
Two-band algorithm (2BDA)	$B5/B4$	Linear regression (LR)	$y = ax + b$
Three-band algorithm (3BDA)	$\left(\frac{1}{B4} - \frac{1}{B5}\right) \times B7$	Multiple linear regression (MLR)	$y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n$
Maximum chlorophyll index (MCI)	$B5 - B4 - 0.2 \times (B6 - B4)$	Quadratic polynomial (Quadratic)	$y = ax^2 + bx + c$
Normalized difference chlorophyll index (NDCI)	$(B5 - B4)/(B5 + B4)$	Logarithmic function (Log)	$y = a\ln(x) + b$
Red-edge normalized difference vegetation index (Red-edge NDVI)	$(B7 - B4)/(B7 + B4)$	Exponential function (Exp)	$y = ae^{bx}$
Floating algae index (FAI)	$B8 - B4 - (B8A - B4) \times 0.885$	Power function (Power)	$y = ax^b$
Turbidity index (TI)	$B4/B3$	Decision tree (DT)	The feature space is partitioned into several sub-regions, where each leaf node represents a predicted value.
Enhanced vegetation index (EVI)	$2.5 \times \frac{B8 - B4}{B8 + 6 \times B4 - 7.5 \times B2 + 1}$	Random forest (RF)	$y = \frac{1}{n}\sum_{i=1}^{n} h_i(x)$, where h_i represents an individual decision tree.
Normalized difference vegetation index (NDVI)	$(B8 - B4)/(B8 + B4)$	Support vector machine (SVM)	$f(x) = \omega^T\phi(x) + b$, where the kernel function ϕ performs high-dimensional mapping.
Surface algal bloom index (SABI)	$(B8 - B4)/(B2 + B3)$	Gaussian process regression (GPR)	$f(x) \sim GP(m(x), k(x, x'))$ (based on a mean function and a covariance kernel function)
Mean spectral reflectance (Mean_Rbands)	$\frac{1}{8}\sum_{i=1}^{8} B_i$		

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}, \tag{1}$$

where n represents the sample size; X_i and Y_i denote the i observed values of the two variables; and $\overline{X}$ and $\overline{Y}$ represent their respective sample means. On the basis of the absolute value of r, the strength of the relationship was classified into five categories: extremely strong ($|r| > 0.8$), strong ($0.6 < |r| < 0.8$), moderate ($0.4 < |r| \leq 0.6$), weak ($0.2 < |r| \leq 0.4$), and negligible correlation ($|r| < 0.2$).

By calculating the correlation between Chl-a concentration and various parameters—including spectral band values, Temp, pH, EC, and remote sensing indices—the underlying inter-relationships were analyzed to provide a rigorous basis for model development and subsequent water quality monitoring.

2.4 Model development and validation

In this study, two types of methods—traditional regression and machine learning models—were employed to perform Chl-a inversion modeling and validation based on the selected feature combinations with high correlation coefficients, as shown in Table 1. The model performance was evaluated using the coefficient of determination (R^2).

During the model training and validation stage, data preprocessing was first performed on each feature combination and Chl-a concentration by removing missing values to ensure data validity. To mitigate the impact of spatiotemporal autocorrelation on model validation, a stratified sampling strategy integrating reservoir locations and sampling periods was adopted to partition the training and testing sets. Specifically, samples were stratified by reservoir, sampling year, and season, and a 70/30 random split was then applied within each stratum to maintain the distributional representativeness of both subsets across space and time.

Model training was conducted using the training set and traditional regression and machine learning methods. Model performance was then validated on the testing set, primarily using R^2 to evaluate the goodness of fit. The calculation formula is

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}, \tag{2}$$

where n is the number of samples, y_i is the field-measured Chl-a concentration, $\hat{y}_i$ is the value predicted by the inversion model, and $\overline{y}$ is the mean of the measured values. Finally, for each feature combination, the model with the highest R^2 on the testing set was selected as the optimal model for that specific combination.

3. Results and Analysis

3.1 Correlation analysis between multisource data and Chl-a concentration

In this study, the inter-relationships among Sentinel-2 spectral reflectance, Mean_Rbands, and three environmental parameters (Temp, pH, and EC) were evaluated by calculating r using the Chl-a concentration. This approach systematically assessed the sensitivity of individual variables to Chl-a and characterized the univariate linear associations between parameters. The resulting correlation matrix is presented in Fig. 2.

The results indicate that the correlation coefficients between individual spectral bands (B2–B8A) and Chl-a concentration ranging from 0.396 to 0.586 (specifically, B2 = 0.425, B3 = 0.533, B4 = 0.548, B5 = 0.586, B6 = 0.444, B7 = 0.446, B8 = 0.409, and B8A = 0.396). Although these bands show statistically significant correlations with Chl-a, the moderate correlation strength suggests that the predictive capacity of any single band for Chl-a inversion is inherently limited. Mean_Rbands yield an r value of 0.484, slightly exceeding most individual bands. This reflects the potential of integrated multiband information to indicate water quality, yet it remains within the threshold of moderate correlation.

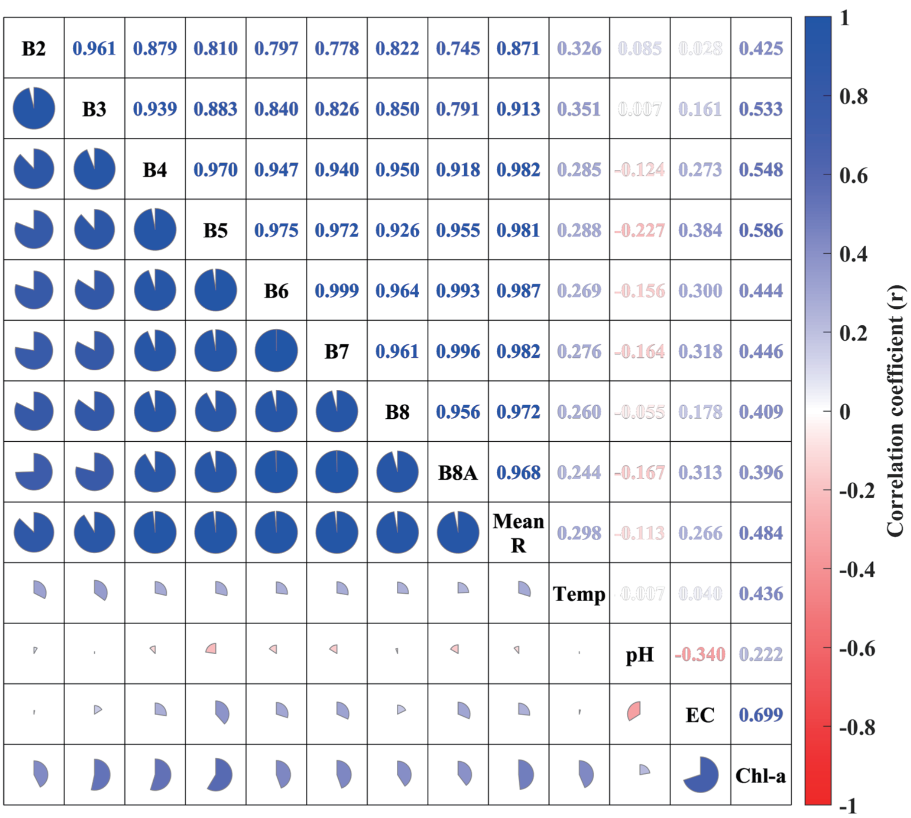

Fig. 2. (Color online) Correlation coefficient matrix between variables and Chl-a concentration.

Regarding environmental parameters, Temp exhibited a moderate positive correlation (r = 0.436) with Chl-a, while pH showed a weak positive correlation (r = 0.222). Notably, EC demonstrated a strong correlation (r = 0.699). This relatively strong association may reflect the shared effect of terrestrial nutrient input and water chemical conditions. In the studied reservoirs, catchment runoff may simultaneously increase EC through dissolved ions and promote algal growth through associated nutrient transport. In addition, variations in ionic strength may be related to changes in algal physiological activity and nutrient uptake efficiency. To reduce the possibility of spurious correlation, a partial correlation analysis was conducted by controlling for the red-edge band (B5), yielding a coefficient of 0.589. This result suggests that the EC–Chl-a relationship was not solely attributable to the selected spectral factor, although the underlying mechanism still requires further investigation. The weak univariate correlation for pH is attributed to the fact that the process of pH alteration via photosynthetic CO_2 consumption is mediated by the initial alkalinity and dissolved CO_2 concentration. The heterogeneity of these factors across different reservoirs and periods in the study area weakens the linear association, suggesting a more complex, nonlinear physical relationship. All correlations discussed above were statistically significant, with p values below 0.05.

In summary, the limited correlation between single spectral bands and Chl-a highlights the challenge of achieving high-precision inversion using traditional methods. Conversely, the stronger correlations exhibited by environmental parameters such as EC suggest that fusing multiband spectral features with environmental auxiliary variables can significantly enhance the model's explanatory power and predictive stability. From the perspective of intervariable associations, no extremely strong linear correlations ($|r| > 0.8$) were observed among the parameters, indicating the absence of significant multicollinearity and providing a solid foundation for multisource feature combinations. Although pH displayed a weak univariate correlation, it remains a critical indicator of the physicochemical coupling between algal photosynthesis and the water environment. When integrated with Temp and spectral reflectance, pH provides synergistic information that captures the full dimensions of algal growth dynamics, photosynthetic processes, and biomass characterization, offering a rigorous mechanistic basis for subsequent multisource inversion modeling.

3.2 Construction of multisource features and comparative correlation analysis

To further enhance the inversion capability for Chl-a concentration, we systematically constructed and screened multivariate combinations merging Sentinel-2 spectral features with environmental parameters (Temp, pH, and EC). These were compared against commonly used remote sensing water quality indices, as illustrated in Fig. 3. Incorporating environmental parameters generally increased the correlation coefficients, with most combinations exceeding 0.800. The highest r reached 0.830 (B5 × Temp × pH), exceeding the correlations obtained for individual bands (r = 0.396–0.586) and conventional remote sensing indices (the highest being 0.742 for MCI). This indicates that integrating environmental parameters with spectral information captures the biochemical processes affecting Chl-a dynamics more comprehensively, thereby strengthening both the physical significance and statistical explanatory power of the features.

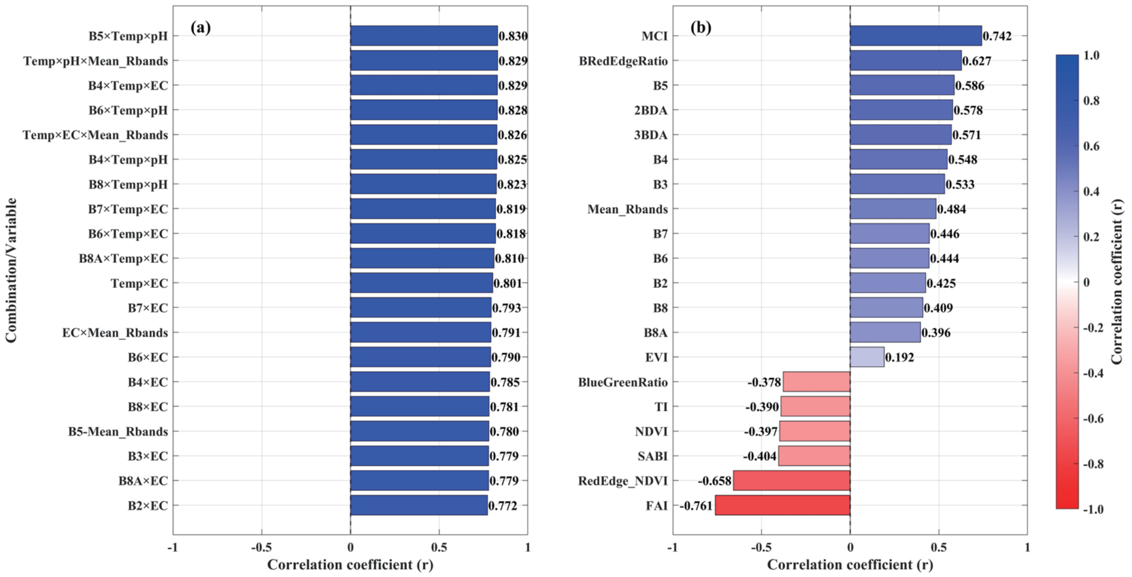

Fig. 3. (Color online) Correlation analysis of multisource feature combinations, single bands, and remote sensing indices with Chl-a concentration.

Among the ternary combinations of environmental parameters and spectral bands, B5 × Temp × pH (r = 0.830), Temp × pH × Mean_Rbands (r = 0.829), and B4 × Temp × EC (r = 0.829) exhibited the most prominent performance. These results demonstrate that the synergy between environmental parameters and specific bands (particularly red-edge and red bands) exhibits strong explanatory power for Chl-a concentration. Binary combinations, such as Temp × EC (r = 0.801) and B7 × EC (r = 0.793), also showed high correlations, further validating the critical role of EC in Chl-a inversion.

By contrast, among the conventional remote sensing indices, only MCI (r = 0.742) and BRedEdgeRatio (B2/B5, blue-to-red-edge band ratio, r = 0.627) maintained moderate positive correlations with Chl-a. Other indices, such as NDVI, SABI, and FAI, exhibited either negative or weak correlations, with explanatory capacities significantly lower than those of the multisource feature combinations. This outcome highlights the limitations of existing remote sensing indices for Chl-a inversion in complex inland waters and underscores that fusing multisource environmental information is a key pathway to overcoming the precision bottlenecks in current water quality remote sensing.

Notably, multiple multisource feature combinations achieved r values above 0.80 with minimal variance among them. This suggests that the inclusion of environmental parameters did not merely enhance specific combinations but rather universally improved the representational capacity of the features. Consequently, subsequent modeling is not restricted to a single high-correlation feature; instead, a multimodel adaptability analysis was conducted using this suite of high-correlation multisource features to identify the most accurate and robust inversion schemes. By systematically constructing and screening these combinations, we clarify the significant advantages of synergistic modeling between environmental parameters and spectral bands, providing essential input features and a theoretical foundation for developing high-precision inversion models.

3.3 Model selection and inversion performance analysis

To determine the optimal modeling approach for Chl-a concentration across different variables, we conducted the training and testing of various regression and machine learning models. These evaluations were applied to the high-correlation multisource feature combinations selected in Sect. 3.2, as well as to individual bands and remote sensing indices. Following the criterion of maximizing the testing set R^2, the best-performing model for each variable was identified, as summarized in Table 2. For comparative purposes, the feature combinations in Table 2 are ranked in descending order in accordance with their testing set R^2.

The analysis of data in Table 2 reveals a diversity in model selection for the multisource feature combinations that integrate environmental parameters with spectral information. LR, RF, and GPR were most frequently selected as the optimal models for the multisource feature combinations. Specifically, for ternary interactive combinations such as Temp × pH × Mean_Rbands and B4 × Temp × pH, the optimal models were predominantly LR with simplified parameters, suggesting a strong linear association between these combinations and Chl-a concentration. Conversely, numerous binary or ternary combinations, such as B4 × Temp × EC and B2 × EC, were better suited to the RF model, which is capable of capturing nonlinear relationships. Furthermore, GPR achieved optimal results for combinations such as Temp × EC × Mean_Rbands and B5 × Temp × pH, demonstrating its advantage in handling specific complex data distributions. In contrast, the optimal models for individual bands and remote sensing indices (e.g., MCI, B7, and EVI) were relatively concentrated, primarily consisting of SVM, Quadratic, and DT. SVM showed robust performance across multiple indices, while Quadratic

Table 2
Optimal models for Chl-a concentration inversion with different feature combinations.

No.	Multisource feature combination	Optimal model	Individual band/Index	Optimal model
1	Temp × pH × Mean_Rbands	LR	MCI	SVM
2	Temp × EC × Mean_Rbands	GPR	FAI	SVM
3	B4 × Temp × pH	LR	B7	RF
4	B6 × Temp × pH	LR	B4	RF
5	B4 × Temp × EC	RF	EVI	SVM
6	B4 × EC	RF	2BDA	SVM
7	Temp × EC	Exp	Mean_Rbands	Quadratic
8	B3 × EC	SVM	B6	DT
9	B6 × Temp × EC	RF	BRedEdgeRatio	DT
10	B8A × Temp × EC	RF	B2	Quadratic
11	B8 × EC	SVM	B8	Quadratic
12	B6 × EC	GPR	3BDA	Quadratic
13	B2 × EC	RF	NDVI	RF
14	B7 × EC	RF	SABI	Quadratic
15	B8A × EC	Exp	Red-edge_NDVI	GPR
16	B8 × Temp × pH	RF	B3	GPR
17	EC × Mean_Rbands	Quadratic	BlueGreenRatio	Quadratic
18	B5 × Temp × pH	GPR	B5	SVM
19	B7 × Temp × EC	RF	B8A	Log
20	B5-Mean_Rbands	LR	TI	Power

regression became the optimal choice for Mean_Rbands, B2, and SABI, reflecting a quadratic curvilinear relationship between certain spectral features and Chl-a concentration.

The disparity in model selection between the two categories of features reveals an intrinsic link between feature complexity and the adapted mathematical models. Because multisource feature combinations integrate multidimensional information, their relationships with the target variable may encompass both linear and nonlinear components, requiring flexible models (such as ensemble learning and nonlinear regression). In contrast, relatively simple remote sensing indices may follow more fixed relationship patterns that can be adequately fitted by specific models such as SVM and Quadratic. To further quantify differences in predictive accuracy, the following section compares the testing-set R^2 values of the optimized models.

For the optimal models determined in Table 2, the corresponding testing set performance characteristics are shown in Fig. 4. Figure 4(a) visually presents the R^2 ranking for the top 20 models of both feature categories. Overall, the blue performance curve of the multisource feature combinations remains consistently and significantly higher than the red curve of the individual bands and remote sensing indices, confirming the decisive role of integrated environmental parameters in enhancing model precision. The Temp × pH × Mean_Rbands combination (LR) achieved the highest R^2 of 0.847, which is 0.114 higher than the control MCI index (SVM, R^2 = 0.733). The corresponding training-set R^2 values of the optimal models were generally comparable to those on the testing set, suggesting acceptable generalization ability and no evidence of severe overfitting. As the ranking descends, the R^2 values for both categories decrease; however, the performance gap (the vertical distance between the curves) gradually widens. By the 20th rank, the difference between the B5-Mean_Rbands combination (LR, R^2 = 0.571) and TI (Power, R^2 = 0.069) reaches 0.502, indicating that for weakly correlated features, the introduction of environmental information is especially critical in preventing model failure.

The statistical results in Fig. 4(b) further quantify this advantage. The mean R^2 for multisource feature combinations was 0.700 ± 0.087, which is significantly higher than the 0.339 ± 0.164 observed for individual bands and indices. The higher mean and smaller standard deviation of the former indicate not only superior overall inversion accuracy but also more concentrated and stable performance, whereas the latter exhibits limitations characterized by low precision and high volatility.

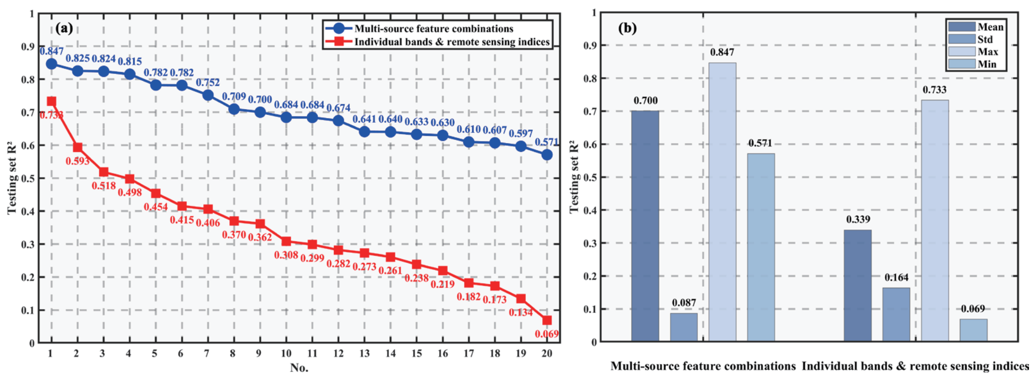

Fig. 4. (Color online) Performance and statistical analysis results of optimal models for different feature combinations on the testing set.

In conclusion, Fig. 4 demonstrates at both the individual and statistical levels that multisource feature combinations integrating environmental parameters outperform traditional spectral methods in both accuracy and stability for Chl-a inversion, fully validating the effectiveness of the multisource data fusion strategy proposed in this study.

3.4 Spatiotemporal distribution characteristics of Chl-a concentration

To visually reveal the spatiotemporal variation patterns of Chl-a concentration in typical reservoirs of Hebei Province, we performed remote sensing inversion for six reservoirs during the 2025 growing season (May, July, and September) using the previously determined global optimal model. For full-image inversion in 2025, Temp and pH were treated as reservoir-level auxiliary variables. For each reservoir and acquisition period, the corresponding representative Temp and pH values were uniformly assigned to all valid water pixels within the reservoir mask, while Mean_Rbands was calculated pixel-wise from Sentinel-2 reflectance. The results are illustrated in Fig. 5.

The inversion results clearly demonstrate significant spatiotemporal heterogeneity and dynamic shifts in Chl-a concentration across the study area. Overall, the concentration levels exhibited a seasonal rhythm—initial growth in spring, a significant increase in summer, and differentiated evolution in autumn—which aligns with the water temperature fluctuations and algal growth cycles in northern China. Furthermore, distinct spatial variations were observed both among different reservoirs and within individual water bodies, reflecting the integrated effect of local environmental and hydrological conditions.

From a time-series perspective, in May (late spring), Chl-a concentrations remained at annually low or moderate levels with relatively uniform spatial distribution, indicating the early stages of algal growth. By July (with peak summer heat), most reservoirs showed a widespread and significant increase in concentration. Notably, localized areas in the Huangbizhuang, Douhe, and Zhuzhuang Reservoirs formed distinct high-concentration cores, indicating peak

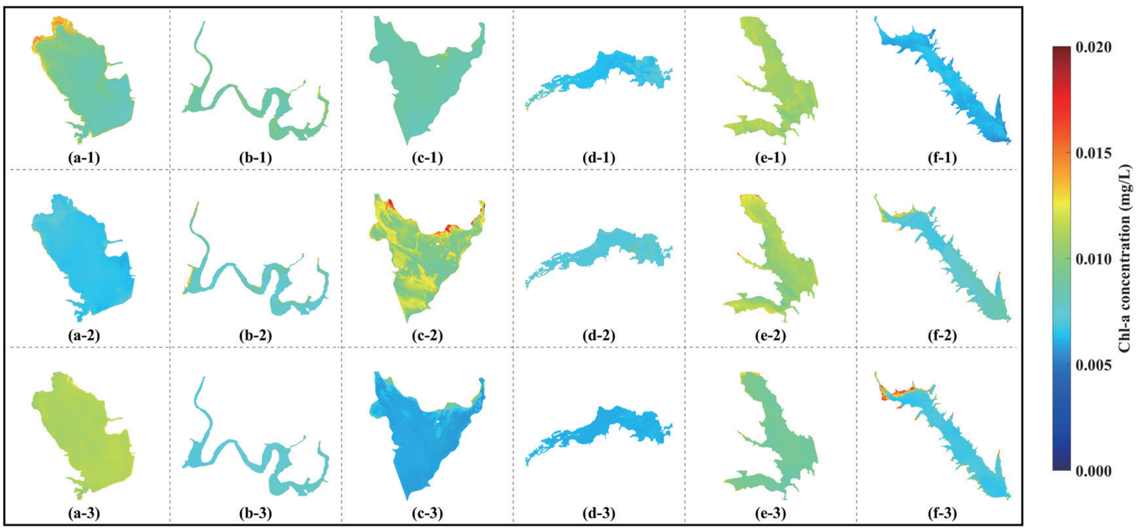

Fig. 5. (Color online) Spatiotemporal distribution of inverted Chl-a concentration in typical reservoirs of Hebei Province.

algal biomass and heightened eutrophication risk. In early September (autumn), the patterns diverged: the high-value areas in some reservoirs (e.g., Shihe and Longmen) contracted or decreased in concentration, while portions of the Zhuzhuang and Lincheng Reservoirs maintained high levels. This was driven by the synergy of optimal autumn water temperatures and accumulated summer nutrient supplies, leading to secondary peaks or the delayed senescence of algae.

In terms of spatial distribution, the effect of reservoir morphology and hydrodynamic conditions was particularly prominent. River-type reservoirs, such as Shihe, exhibited lower overall concentrations and uniform distributions owing to stronger water exchange, which inhibits algal aggregation. Conversely, near-shore areas and bays in several reservoirs (e.g., Douhe Reservoir bays and Zhuzhuang Reservoir shores) consistently became intermittent high-value zones. These areas are typically characterized by significant terrestrial inputs and weak hydrodynamics, creating micro-environments conducive to nutrient enrichment and dominant algal growth. Larger reservoirs, such as Huangbizhuang and Lincheng, showed more complex internal spatial fluctuations, reflecting the heterogeneity of their hydrodynamic processes, exogenous inputs, and ecological responses.

In summary, the inversion results in Fig. 5 reveal the spatiotemporal coupling of Chl-a concentration, fundamentally driven by seasonal climatic rhythms (light and temperature) and the intrinsic attributes of the reservoirs. This pattern is of clear significance for precision reservoir management: spring, the summer–autumn transition, and stagnant areas such as shores and bays are critical spatiotemporal units for eutrophication monitoring and control.

To quantitatively reveal the overall levels and variability of the inversion results obtained with the optimal model (Temp × pH × Mean_Rbands–LR), a statistical analysis was conducted on Chl-a concentrations across different reservoirs and months, as shown in Fig. 6. Figure 6(a) displays the mean concentration changes for the six reservoirs as a line chart, while the overall pixel-level concentration distribution for each month is shown in Fig. 6(b) using box plots.

The seasonal variation curves for each reservoir show distinct fluctuation modes, reflecting their unique environmental responses. Huangbizhuang Reservoir exhibited sharp fluctuations in

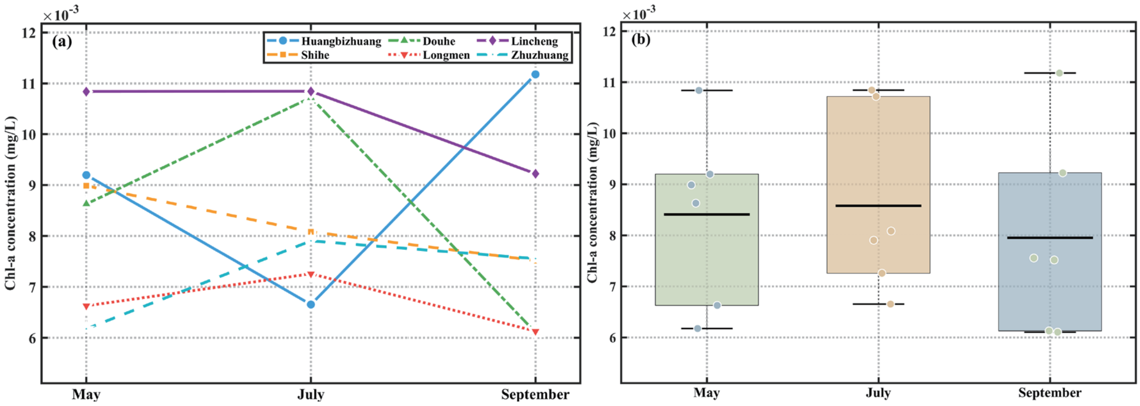

Fig. 6. (Color online) Statistical characteristics analysis of Chl-a concentration based on inversion results.

a "decrease then increase" pattern, with concentrations dropping from 0.0092 mg/L in May to 0.0067 mg/L in July, before surging to 0.0112 mg/L in September; its annual mean (0.0090 mg/L) remained high. Lincheng Reservoir maintained high levels throughout all three periods (approaching 0.0110 mg/L in May and July, with a slight drop in September), with the highest annual mean (0.0103 mg/L) among the six reservoirs. Douhe Reservoir followed an "increase then decrease" trend, peaking in July (0.0107 mg/L). In contrast, Longmen Reservoir had the lowest overall concentration and stable fluctuations (annual mean: 0.0067 mg/L), while Shihe and Zhuzhuang Reservoirs showed gradual declines or minor fluctuations.

Monthly statistics across all reservoirs showed mean concentrations of 0.0084, 0.0086, and 0.0080 mg/L for May, July, and September, respectively, indicating slightly higher overall levels in July. The box plots reveal further distribution details: in May, the median was 0.0085 mg/L with relatively concentrated data. In July, the boxes widened significantly, the interquartile range (IQR) expanded, and high-value outliers appeared, indicating that July had not only the highest mean concentration but also the most significant spatial heterogeneity within and among reservoirs. By September, the median declined, but the upper whiskers remained long—primarily because of the extremely high values in Huangbizhuang—while the lower whiskers corresponded to low values in the Douhe and Longmen Reservoirs, reflecting the divergence of the autumn concentration pattern.

These statistical results corroborate the spatiotemporal distribution maps (Fig. 5), providing quantification from both mean values and distribution dimensions. Lincheng Reservoir faces high eutrophication pressure across both space and time. July is the critical period when overall risk peaks and spatial variance is greatest, and the anomalous high values in Huangbizhuang during September highlight the need to monitor specific post-autumn algal proliferation phenomena.

4. Conclusions

This study was focused on six typical reservoirs in Hebei Province. By systematically integrating Sentinel-2 multispectral remote sensing data with three environmental parameters (Temp, pH, and EC), a suite of multisource feature combinations was constructed and screened. Various regression and machine learning models were utilized for Chl-a concentration inversion modeling and validation. Finally, the spatiotemporal distribution characteristics were revealed by using the optimal model. The main conclusions are as follows.

(1) The integration of environmental parameters significantly enhances the correlation between features and Chl-a concentration, serving as a key factor in improving inversion accuracy. While correlation coefficients for individual bands and conventional remote sensing indices were generally below 0.600, the multisource feature combinations (e.g., B5 × Temp × pH) constructed by fusing Temp, pH, and EC achieved r values as high as 0.830. This provides a solid foundation for developing high-precision inversion models.

(2) The multisource feature fusion strategy, combined with optimal model selection, facilitates the development of inversion models that are stable and significantly outperform traditional methods. Among all tested feature combinations, the LR model driven by the Temp × pH ×

Mean_Rbands feature achieved the highest accuracy on the testing set ($R^2 = 0.847$). Statistical analysis demonstrated that the optimal models corresponding to multisource feature combinations reached an average testing set R^2 of 0.700 ± 0.087, significantly higher than those of traditional methods that are based solely on spectral information (average $R^2 = 0.339 \pm 0.164$). This confirms the effectiveness and superiority of the multisource data fusion strategy.

(3) The inversion results obtained with the optimal model (Temp × pH × Mean_Rbands–LR) clearly reveal the spatiotemporal variation patterns of Chl-a concentration in the study area during the 2025 growing season (May, July, and September). Temporally, concentrations exhibited a seasonal rhythm of initial growth in spring, significant elevation in summer, and differentiated evolution in autumn, with July identified as the peak period for both concentration levels and spatial heterogeneity. Spatially, significant differences were observed among and within reservoirs: Lincheng Reservoir exhibited the highest annual mean concentration (0.0103 mg/L), while Longmen Reservoir had the lowest (0.0067 mg/L). River-type reservoirs (e.g., Shihe) showed uniform distributions, whereas the shores and bays of most reservoirs consistently appeared as sensitive areas with high-concentration clusters. These findings provide a direct basis for the precise spatiotemporal identification of eutrophication risks.

In summary, this study substantiates the significant potential of fusing environmental parameters with remote sensing spectral information to improve Chl-a inversion accuracy in inland waters, providing a robust methodological framework for practical satellite-based water-quality monitoring applications. The model framework and spatiotemporal patterns revealed herein provide methodological references and data support for the operational remote sensing monitoring of water quality, dynamic eutrophication assessment, and precise risk identification in Hebei Province and reservoirs under similar climatic and hydrological conditions.

References

1 Wang, B.-H. Tang, Z. Fu, L. Huang, M. Li, G. Chen, and X. Pan: Remote Sens. **14** (2022) 4950. http://doi.org/10.3390/rs14194950

2 Z. Wu, J. Pang, J. Li, Y. Wang, J. Ruan, X. Zhang, L. Yang, Y. Pang, and Y. Gao: Intell. Mar. Technol. Syst. **3** (2025) 24. http://doi.org/10.1007/s44295-025-00075-2

3 G. Wen, X. J. Chen, Y. K. Liu, K. Li, H. Z. Lü, F. Si, and T. L. Huang: Environ. Eng. **43** (2025) 147. https://doi.org/10.13205/j.hjgc.202503013

4 B. W. Asres, M. G. Kebedew, M. D. Nerae, S. Tsegaye, and F. A. Zimale: Front. Water **7** (2025) 1600222. http://doi.org/10.3389/frwa.2025.1600222

5 S. Ngamile, S. Madonsela, and M. Kganyago: Front. Environ. Sci. **13** (2025) 1549301. http://doi.org/10.3389/fenvs.2025.1549301

6 S. I. Schmidt, T. Schröder, R. D. Kutzner, P. Laue, H. Bernert, K. Stelzer, K. Friese, and K. Rinke: Remote Sens. **16** (2024) 3416. http://doi.org/10.3390/rs16183416

7 Y. Sun, D. Wang, L. Li, R. Ning, S. Yu, and N. Gao: Water Res. **267** (2024) 122546. http://doi.org/10.1016/j.watres.2024.122546

8 C. Cheng, Y. Wei, G. Lv, and Z. Yuan: J. Appl. Remote Sens. **7** (2013) 073465. https://doi.org/10.1117/1.JRS.7.073465

9 P. M. Maier and S. Keller: ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. **IV-2/W5** (2019) 609. http://doi.org/10.5194/isprs-annals-IV-2-W5-609-2019

10 J. Llodrà-Llabrés, J. Martínez-López, T. Postma, C. Pérez-Martínez, and D. Alcaraz-Segura: Int. J. Appl. Earth Obs. Geoinf. **125** (2023) 103605. http://doi.org/10.1016/j.jag.2023.103605

11 W. D. Ma, Q. G. Xing, Y. Z. Zhang, P. Shi, and Y. L. Liu: Spectrosc. Spectral Anal. **30** (2010) 313. https://doi.org/10.3964/j.issn.1000-0593(2010)02-0313-05
12 X. Z. Shu, Q. Yin, and D. B. Kuang: J. Remote Sens. **4** (2000) 41. https://doi.org/10.11834/jrs.20000108
13 H. B. Zhang: Optical Properties of Chlorophyll and Its Applications (M.S. thesis, Shandong Normal University, Jinan, 2008). https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&filename=2008133588.nh
14 L. You, L. F. Cui, Z. W. Liu, B. Yang, and Z. F. Huang: Environ. Sci. Technol. **30** (2007) 42. https://doi.org/10.19672/j.cnki.1003-6504.2007.09.015
15 M. Y. Wei, S. W. Xue, Z. Teng, Y. Li, and Y. H. Bi: J. Hydroecology **46** (2025) 1. https://doi.org/10.15928/j.1674-3075.20250801.1356
16 Y. Deng, Y. Zhang, D. Pan, S. X. Yang, and B. Gharabaghi: Remote Sens. **16** (2024) 4196. http://doi.org/10.3390/rs16224196
17 W. Pan, F. Yu, J. Li, C. Li, and M. Ye: Ecol. Inf. **87** (2025) 103061. http://doi.org/10.1016/j.ecoinf.2025.103061
18 C. Niu: Radiometric Correction and Water Environment Parameter Inversion of Inland Water Bodies Based on Hyperspectral Remote Sensing (Ph.D. dissertation, East China Normal University, Shanghai, 2024). https://doi.org/10.27149/d.cnki.ghdsu.2024.000206