Journal of Oceanology and Limnology

https://doi.org/10.1007/s00343-026-5442-8

Research Paper

Classification-based inversion of suspended sediment concentration using in-situ hyperspectral data in Qingshui River Basin of Huanghe River*

Zhiru JIN[1], Xinyue LIU[1], Qianguo XING[2, 3, 4], Weichen ZHANG[1], Xiao LIU[1], Xing MING[1, **]

[1] *College of Physical Science and Technology, Hebei University, Baoding 071002, China*

[2] *CAS Key Laboratory of Coastal Environmental Processes and Ecological Remediation, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai 264003, China*

[3] *Shandong Key Laboratory of Coastal Environmental Processes, Yantai 264003, China*

[4] *University of Chinese Academy of Sciences, Beijing 100049, China*

Received Nov. 11, 2025; accepted in principle Dec. 30, 2025; accepted for publication Feb. 26, 2026

Abstract The Qingshui River Basin is the largest tributary entering the Huanghe (Yellow) River within Ningxia of China, contributing about 49% of the sediment load from this section. In this basin, the suspended sediment concentration (SSC) exhibits a short-term variability with values rising by orders of magnitude within hours. This phenomenon is primarily driven by the precipitation-induced runoff, which induces the surface soil erosion, sediment transport and the local bed erosion. Such rapid fluctuations of SSC are difficult to be captured using manual sampling or satellite remote sensing methods. To address this issue, a classification-based inversion model was developed in this paper using the ground-based in-situ hyperspectral data from the Wangtuan station beside Qingshui River between March and September 2024, covering both normal and flood periods. At the beginning, we utilized the K-means clustering algorithm to classify these water spectral characteristics into three types. These water types exhibited different spectral characteristics in the position and number of their reflectance peaks. Subsequently we built empirical and machine learning models for each water type to retrieve SSC. By comparison to Global models, the classification-based inversion models substantially improved inversion performance: a linear empirical model performed the best for low SSC water type (R^2=0.92), while support vector regression and random forest achieved an R^2 of 0.92 under medium and high SSC conditions, respectively. Therefore, this study we conducted in this paper achieved a high inversion accuracy of SSC especially for medium and high SSC water types. Our results revealed that the ‘classify first, then model’ method is reliable for estimating SSC of high-sediment water bodies in the Qingshui River Basin, and enhances the feasibility of ground-based in-situ hyperspectral monitoring technology for refined Huanghe River watershed management.

Keyword: Qingshui River Basin; Huanghe (Yellow) River; suspended sediment concentration (SSC); in-situ hyperspectral; K-means clustering; inversion model

1 INTRODUCTION

The Huanghe (Yellow) River, with a total length of approximately 5 464 km, is recognized as the world’s fifth-longest river and serves as a critical water resource and ecological corridor in northern China (Wu et al., 2004; Chen et al., 2020). Its basin

* Supported by the Guiding Program for Technology Innovation of Shandong Province, China (No. YDZX2024021), the Key R&D Program of Shandong Province, China (No. 2022CXPT019), the Key R&D Program of Ningxia Hui Autonomous Region, China (No. 2023BEG02052), the Instrument Developing Project of the Chinese Academy of Sciences (No. YJKYYQ20170048), and the National Natural Science Foundation of China (Nos. 42076188, 42476175)

** Corresponding author: mingxing@hbu.edu.cn

is characterized by complex and dynamic water-sediment interactions, which profoundly influence the regional ecological security and sustainable development. As a key transitional zone within the upper reaches of the Huanghe River, the Ningxia section contains numerous tributaries. Among these tributaries, the Qingshui River is the largest one which basin covers an area of 14 481 km^2. It contributes approximately 49% of the total sediment load delivered from Ningxia section to the Huanghe River, significantly affecting the water and sediment dynamics of the main stream (Jiao et al., 2020).

The suspended sediment, primarily composed of inorganic mineral particles, is widely used as a key indicator of water erosion and sediment transport capacity (Bowers and Binding, 2006; Gao, 2008). High suspended sediment concentration (SSC) not only reduce water transparency and inhibit phytoplankton photosynthesis but also directly damage aquatic communities and change fish population structures. Furthermore, the suspended sediment can carry pollutants, thereby increasing drinking water treatment costs and posing threats to aquatic ecosystems (Rabeni, 1997). Additionally, the suspended sediment plays a crucial role in riverbed formation and channel morphology evolution (Wen et al., 2025). Consequently, the implementation of scientific water-sediment regulation strategies is considered essential for optimizing sediment transport, stabilizing riverbed geometry and maintaining water security and ecological balance in the Huanghe River Basin (Hou et al., 2021).

In recent years, large-scale ecological restoration projects have significantly altered sediment transport mechanisms throughout the Huanghe River Basin (Yang et al., 2022a). Researches indicate that although the sediment transport under normal flow conditions has decreased, it may intensify during extreme flow events, particularly in small basins, thereby increasing sediment transport risks during heavy rainfall (Sun et al., 2020). In the Ningxia section of Huanghe River, sediment transport processes are highly dynamic. Hydrological station observations from the Qingshui River Basin demonstrate that the SSC can surge from tens to hundreds of kg/m^3 within hours. Such rapid variations make it difficult for traditional manual sampling and laboratory analysis methods to accurately capture transient sediment concentration changes (Rai and Kumar, 2015; Matos et al., 2024).

In view of the unavailability of conventional methods in monitoring such highly dynamic sediment processes, the remote sensing technology has been widely applied in SSC monitoring due to its advantages of large spatial coverage and short periodic observations. For example, Yu et al. (2022) used Sentinel-3A/OLCI data to improve the monitoring accuracy of SSC in Hangzhou Bay. Jally et al. (2021) demonstrated that multi-band regression models outperformed single-band models using Landsat-8 OLI data for Chilika Lake in India. Cao et al. (2022) combined multi-source satellite imagery to reveal the spatiotemporal evolution of SSC in the Zhujiang (Pearl) River estuary from 1986 to 2020. Duan et al. (2024) used Landsat data to analyze the spatiotemporal characteristics of sediment transport in the lower Huanghe River and its response to the flow-sediment regulation of the Xiaolangdi Reservoir. The satellite remote sensing technology mentioned above offers an irreplaceable advantage in providing large-scale spatial distribution data, which is particularly crucial for studying dynamic processes such as the seasonal and interannual variations of the large-scale sediment distribution. However, facing short-lived sediment transport events in the Qingshui River Basin, where concentration changes dramatically within a few hours and the spatial extent is relatively small, satellite remote sensing, due to its inherent limitations such as long revisit cycles, susceptibility to cloud and rain interference, and limited spatial resolution, may not adequately capture these short-term changes (Zhu et al., 2018). In contrast, the proximal remote sensing technology using ground-based in-situ above-water hyperspectral sensors, enables non-contact optical measurements, which can capture localized changes with high temporal resolution. For example, Yang et al. (2022b) developed a hybrid model combining spectral preprocessing and deep learning to improve the estimation accuracy of soil organic matter content in desert areas. Sun et al. (2022) used hyperspectral technology and a BP neural network to achieve high-frequency, real-time monitoring of water quality parameters such as total nitrogen, total phosphorus and chemical oxygen demand in inland waters. Hou et al. (2024) established an integrated space-ground monitoring network that combines satellites, drones and ground-based hyperspectral stations. This integrated system not only enabled the water and sediment to be monitored but also supported their sustainable management in the Huanghe River Basin.

As we know, there are three commonly used inversion algorithms as empirical, semi-analytical, and analytical algorithms based on hyperspectral data. Semi-analytical and analytical algorithms have applicability limitations due to parameter acquisition difficulties and complex modeling processes (Mao et al., 2012; Bernardo et al., 2019), which are further complicated by the variable optical properties of natural water bodies like those in the Huanghe River. While empirical algorithms can directly establish quantitative relationships between optimal parameters and SSC values through the statistical regression. Such empirical algorithms are most widely used in current SSC inversion studies owing to their operational simplicity and high efficiency (Feng et al., 2012; Zhang et al., 2014; Hou et al., 2023). In recent years machine learning methods have demonstrated significant advantages in SSC inversion in complex water bodies, as they can more effectively capture nonlinear relationships between spectral features and SSC (Dehkordi et al., 2021; Hu et al., 2023; Zaghian et al., 2023).

In the Qingshui River Basin, heavy rainfall during the flood period is prone to causing fluctuations in SSC and the occurrence of flood peaks. For example, according to the measured data from the Wangtuan Hydrological Station of the Qingshui River, the SSC was only 0.331 kg/m^3 at 8:00 on June 27, 2024, which sharply increased to 214 kg/m^3 by 8:00 on June 28, and then decreased to 66.1 kg/m^3 at 17:00 on June 29. Such drastic changes in concentration alter the absorption and scattering processes of incident light in water bodies, leading to differences in the spectral characteristics of water with varying SSC (Wang et al., 2009). Using a single inversion model often fails to adequately represent the relationship between spectral features and SSC, resulting in limited inversion accuracy (Hou et al., 2025). To improve the accuracy of water quality inversion, researchers suggest classifying water bodies into different types and developing suitable inversion models for each type. For example, Shen et al. (2015) classified optically complex waters in China into four optical types using an improved fuzzy c-means clustering method based on in-situ remote sensing reflectance; Montanher et al. (2014) developed multiple regression models using Landsat 5/TM data to estimate SSC in Amazonian white-water rivers, showing that regional models could improve the estimation accuracy; Zhang et al. (2022) used Sentinel-2 MSI data and the Forel-Ule Index to classify water types, developed an SSC estimation model for the Changjiang (Yangtze) River main stream.

While these studies confirmed that classification techniques improve SSC inversion accuracy, their models were primarily developed using satellite data or applied in large river systems. To enhance the model adaptability in high SSC, optically complex water bodies such as the tributaries of the Huanghe River, this study used the K-means clustering algorithm to classify ground-based hyperspectral data collected from the Wangtuan Hydrological Station into three types. For each water type, both empirical and machine learning models were developed. This approach improved the inversion accuracy of SSC in the regions particularly for medium and high SSC water types. It also enhanced the feasibility of applying ground-based hyperspectral monitoring technology for refined management in the Huanghe River.

2 MATERIAL AND METHOD

2.1 Study area

The Ningxia section of the upper Huanghe River, located in the arid inland region of northwest China, is one of the most water-scarce and ecologically vulnerable areas in the entire Huanghe River Basin. Bordered by deserts on three sides and the Loess Plateau to the south, this section functions as a critical ecological transition zone. It is characterized by a gradual shift from fluvial landforms in the south to wind-eroded landscapes in the north, along with a climatic transition from semi-arid to arid. In recent decades, water and sediment fluxes in this section have been significantly affected because of global climate change and intensive human activities, including reservoir construction, soil conservation practices and large-scale water diversion projects. Both water discharge and sediment load have experienced sharp declines, resulting in a series of ecological and environmental issues (Miao et al., 2022).

The Qingshui River Basin spans the southern mountainous and central arid zones of Ningxia. Vegetation in this area is sparse, and soil erosion is severe. The multi-year average precipitation is 349 mm, with frequent localized heavy rainfall events (Ma et al., 2020). The precipitation is the primary driving force for sediment production in the watershed. It is directly linked to both the runoff volume and sediment transport (Duan and Ma, 2021). Data from the basin's outlet hydrological station show the multi-year average SSC is 218 kg/m^3 (Yang et al., 2025). As a result, the Qingshui River Basin is identified as one of the most severely eroded and sediment-concentrated areas in Ningxia section of Huanghe River (Zhu et al., 2022).

Wangtuan region locates in the middle reaches of the Qingshui River. A hydrological station (105°59′34″E, 36°51′37″N) was built by Ningxia Department of Water Resources in 2013. This station collects the SSC values periodically by manual samplings. According to the measured SSC data from the station, it records a long-term average annual sediment discharge of 11.5 million t, and the maximum single-sample sediment concentration of 968 kg/m^3 (on 8 July 2015), indicating a typical high-sediment characteristics (Hou et al., 2025). The in-situ hyperspectral sensor was also built at Wangtuan hydrological station and could acquire spectral data for SSC inversion at meantime with manual samplings. The location of the study area is shown in Fig.1.

2.2 Data

2.2.1 In-situ hyperspectral data

The hyperspectral data for SSC inversion were collected using the in-situ hyperspectral instrument (Model ZK-UVIR-I; Zhike Yuanda, Beijing) built at Wangtuan hydrological station mentioned above (Fig.2a). Those data were acquired from March to September 2024, covering both normal and flood periods, so they could represent the spectral characteristics of the water bodies under varied hydrological conditions and different periods. The in-situ hyperspectral data have a spectral range of 350 to 950 nm, with a spectral resolution of 0.5 nm. A single data collection takes less than 10 s, and the collection interval is adjustable to meet continuous acquisition requirements. The hyperspectral instrument consists of an upward-looking module that collects downwelling solar radiation and a downward-looking module that synchronously

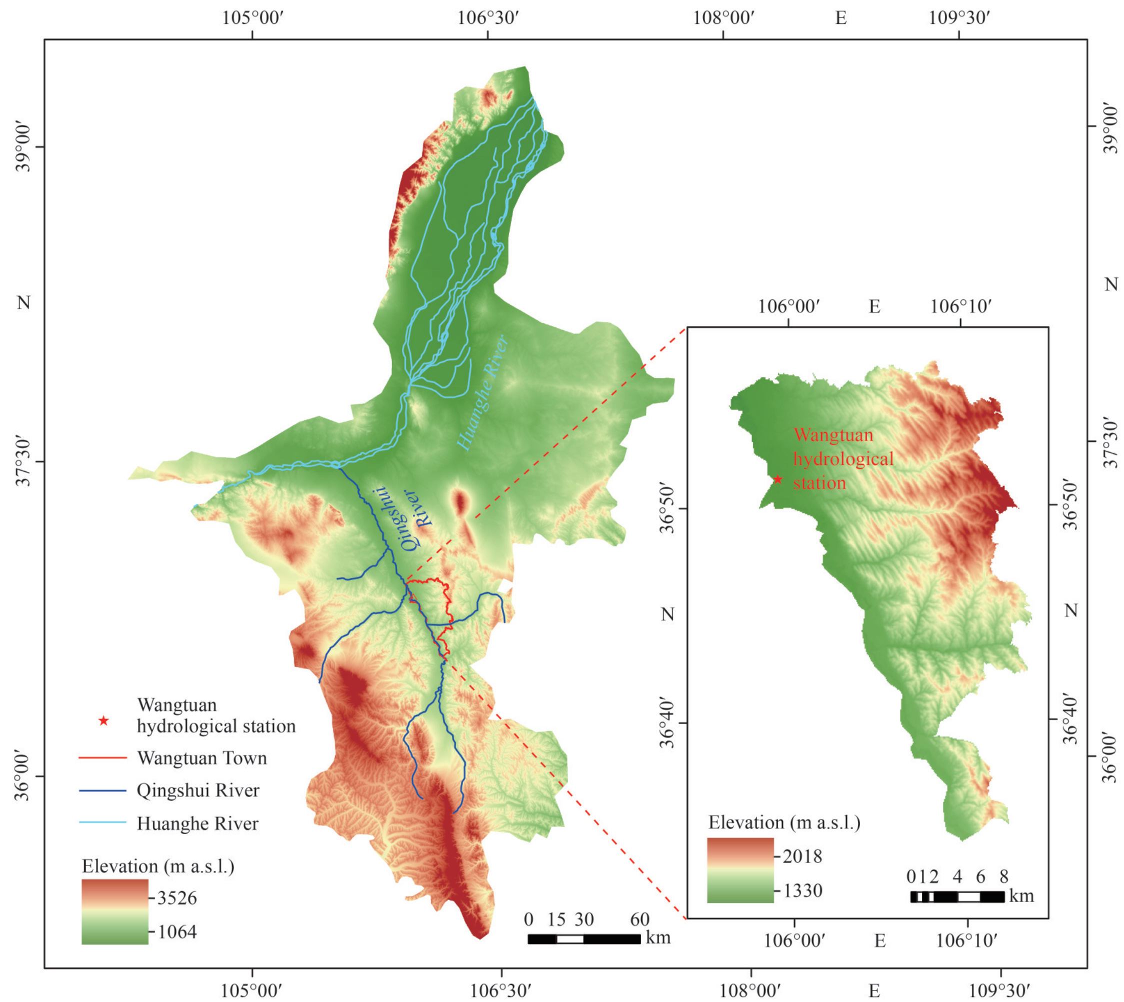

Fig.1 Location of the study area

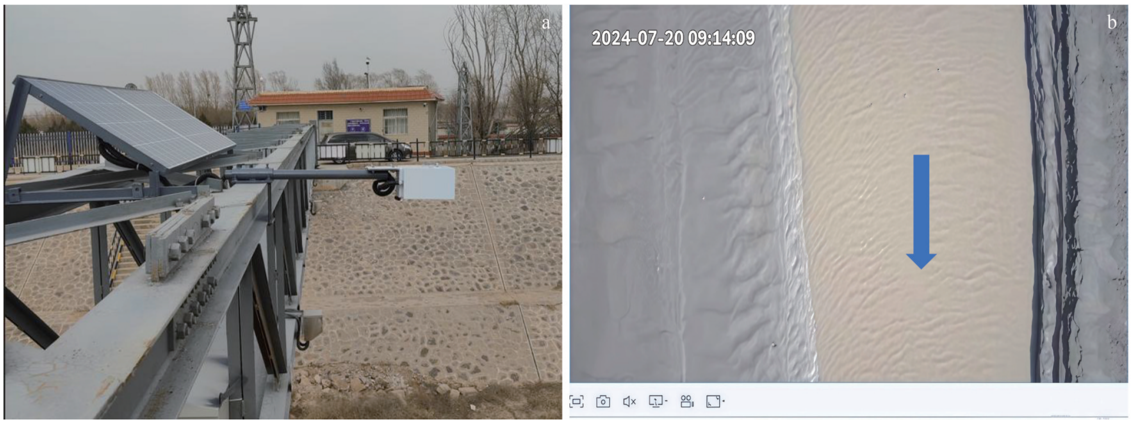

Fig.2 In-situ hyperspectral data acquisition

a. installation of the ZK-UVIR-I in-situ hyperspectral instrument; b. real-time water view at the observation site. In (b), the blue arrow indicates the flow direction of the river.

records upwelling water-leaving radiance. The water-leaving reflectance spectrum is calculated as the ratio of the water-leaving radiance to the synchronous solar radiation (Wang et al., 2012). All spectral data are transmitted in real-time to a cloud platform through a wireless network for further processing.

2.2.2 Measured suspended sediment concentration

The manual measured SSC data, covering the same period with the in-situ hyperspectral data were obtained through standardized field sampling and laboratory procedures. Field sampling followed the one-line sampling method along the riverbank. During normal flow periods, one sample was collected daily at 8:00 AM, while sampling frequency was increased to 8 to 10 times per day during flood periods to capture dynamic sediment transport processes. Laboratory analysis was conducted using the filtration-drying-weighing method. Each water sample was poured into a graduated cylinder to record its volume. After sedimentation, the supernatant was carefully decanted, and the settled sediment was transferred to filter paper. Following natural drainage, the sediment was oven-dried at 100 to 110 °C for 2 h before its dry mass was measured. The SSC was calculated as the ratio of sediment dry mass to the original water sample volume. All procedures strictly adhered to the ‘River Suspended Sediment Testing Specification’ (GB/T 50159-2015) and ‘Water Quality-Determination of Suspended Substance-Gravimetric method’ (GB/T 11901-1989) to ensure data accuracy and reliability.

2.3 Method

2.3.1 Hyperspectral data preprocessing

Initially, the hyperspectral data were preprocessed to extract spectral features related to suspended sediment through three steps: (1) calculating the spectral reflectance $R_{rs}(\lambda)$ by Eq.1, where $L_w(\lambda)$ representing the water-leaving radiance, $E_d(\lambda)$ representing the downwelling irradiance; (2) restricting the spectral range to 400–900 nm to exclude the poor quality signals at both ends of the spectral curve; (3) resampling the raw data at 5 nm intervals to reduce computational load; (4) smoothing resampled data using the Savitzky-Golay filter to reduce noise (Rinnan et al., 2009; Wang et al., 2022; Hou et al., 2025).

$$R_{rs}(\lambda)=\frac{L_w(\lambda)}{E_d(\lambda)}. \tag{1}$$

2.3.2 Water spectral classification

Before SSC inversion, we used the K-means clustering algorithm to classify the preprocessed hyperspectral reflectance data in this study (Haut et al., 2017; Ranjan et al., 2017; Christilda and Manoharan, 2023). The optimal number of clusters K was determined using both the Silhouette Coefficient (S_i, Eq.2) and the Elbow Method based on the Sum of Squared Errors (SSE, Eq.3), to achieve the intra-cluster consistency and inter-cluster separation in spectral characteristics (Dinh et al., 2019; Lai et al., 2025). In the algorithm, K number of cluster centers were first initialized randomly. Each sample of the hyperspectral dataset was then assigned to the nearest cluster based on the Euclidean distance to the centers, after which the cluster means were updated as new centers. This procedure was repeated iteratively until reaching the convergence.

$$S_i = \frac{b_i - a_i}{\max\left(a_i, b_i\right)}, \tag{2}$$

where a_i is the average distance between sample i and all other samples in the same cluster, and b_i is the minimum average distance from sample i to all samples in any other cluster.

The Elbow Method determines the optimal number of clusters K by analyzing the trend of SSE as K varies (Humaira and Rasyidah, 2020; Umargono et al., 2020). The calculation formula is shown in Eq.3:

$$\mathrm{SSE} = \sum_{i=1}^{K} \sum_{x \in C_i} \left\| x - \mu_i \right\|^2, \tag{3}$$

where K is the number of clusters, C_i represents the i^{th} cluster, x denotes a sample point, and μ_i is the centroid of cluster C_i.

2.3.3 Data augmentation

Following spectral classification, the data augmentation based on linear interpolation was applied separately to each cluster to increase sample size for improving the model training performance (Zhang et al., 2017). Specifically, for any two original samples (x_i, y_i) and (x_j, y_j), where x_i, x_j represents the spectral feature vector and y_i, y_j denotes the corresponding SSC value, new augmented samples (X_{new}, Y_{new}) were generated using the following Eqs.4 & 5:

$$X_{\text{new}} = \alpha \times x_i + (1-\alpha) \times x_j, \tag{4}$$

$$Y_{\text{new}} = \alpha \times y_i + (1-\alpha) \times y_j, \tag{5}$$

where α is a random interpolation coefficient uniformly distributed within the interval (0, 1). In this study, we took 0.5 as center. Then α was chosen from four symmetric intervals ([0.1, 0.9], [0.2, 0.8], [0.3, 0.7], and [0.4, 0.6]). The intervals were centered at 0.5 to ensure that the interpolated results were uniformly distributed between the two original samples. For each interval, three statistical indicators (skewness, median, and mean) were used to evaluate how closely the augmented SSC distribution matched the original SSC. Results indicated that α values within [0.3, 0.7] produced the augmented data exhibits statistical indicators closest to those of the original data. Meanwhile, the augmented data preserved statistical consistency avoiding the introduction of extreme samples that could result from an overly wide interpolation range.

2.3.4 Modeling methods and evaluation

Before model training, this study first checked whether the SSC of each cluster followed a normal distribution. If the distribution was obviously skewed, a logarithmic transformation was performed on the SSC values to improve the distribution shape and stabilize the variance. Next, to identify the spectral features for model training, all possible band difference (BD) and band ratio (BR) combinations were tested. The correlation between each combination and SSC was quantified by calculating the R^2 using linear regression, and the three combinations with the highest R^2 were selected.

All models were trained using both spectral features and SSC: for the empirical models, the selected features were served as independent variables. Relationships including linear, power, quadratic and exponential forms were then fitted using the least squares method (Hou et al., 2022). In contrast, the machine learning models incorporated both the selected band combinations and full-band original reflectance as input features (Ortiz-Lopez et al., 2022). These models were implemented using three algorithms: random forest (RF) configured with 150 trees, XGBoost with a learning rate of 0.1 and 100 trees, and support vector regression (SVR) employing the radial basis function (RBF) kernel. Hyperparameter optimization was performed using 3-fold cross-validation combined with random search (Yang et al., 2023).

The model accuracy was evaluated using the coefficient of determination (R^2) (Fisher, 1930), mean absolute error (MAE) (Willmott, 1981), and root mean square error (RMSE) (Montgomery et al., 2012), which were calculated using Eqs.6–8:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \tag{6}$$

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right|, \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \tag{8}$$

where y_i represents the measured SSC of the i^{th} sample, $\hat{y}_i$ denotes the model predicted value, $\bar{y}$ is the mean of measured SSC, and n is the total number of samples.

3 RESULT

3.1 Water spectral classification

In this study, we obtained 120 SSC samples from the Wangtuan station and collected the same amount of hyperspectral data by matching the water sampling time. Excluding of spectral anomalies due to the intense water surface reflection, 76 valid samples (approximately 63.3% of the total) were retained for subsequent preprocessing. The preprocessed spectral data curves are shown in Fig.3a. Subsequently, the spectral classification was carried out using the K-means cluster algorithm. The optimal number of clusters K was determined as 3 based on the silhouette coefficient and the elbow method (Fig.3b). Thereby the water spectral samples were divided into three distinct types characterized by differing spectral features.

As shown in Fig.3c–f, the spectral curves reflect obvious relationships between spectral shapes and SSC values across the three water types. The spectral curve of Type I shows a significant dual-peak shape, with the primary reflection peak located in the yellow-green band (580–620 nm) and a secondary peak in the red band (690–710 nm), while reflectance tends to stabilize in the near-infrared region. Differently, although the spectral curve of Type II retains a dual-peak shape, the primary reflectance peak shifts to the red band (680–720 nm), and a distinct secondary peak emerges in the near-infrared band (780–820 nm). Compared with Type I, the overall spectral reflection intensity of Type II is significantly higher. However, the spectral curve of Type III significantly changes into a single-peak shape centered at the range of 780 to 820 nm. Notably, the spectral reflectance with high values tends to have lower SSC values.

For each of the above water types, the corresponding SSC range and distribution characteristics also differ (Fig.4): the SSC range of water Type I is from 0.098 to 0.542 kg/m^3 and its skewness is -0.12, showing a slight left-skewed trend; the SSC of water Type II ranges from 0.135 to 3.26 kg/m^3 and its skewness is 0.46, exhibiting a slight right-skewed feature that reflects an increase in SSC. The SSC range of water Type III spans nearly three orders of magnitude, ranging from 2.53 to 740 kg/m^3. It has a skewness of 1.84 and shows a strongly right-skewed distribution, which indicates the presence of a few extremely high SSC values in

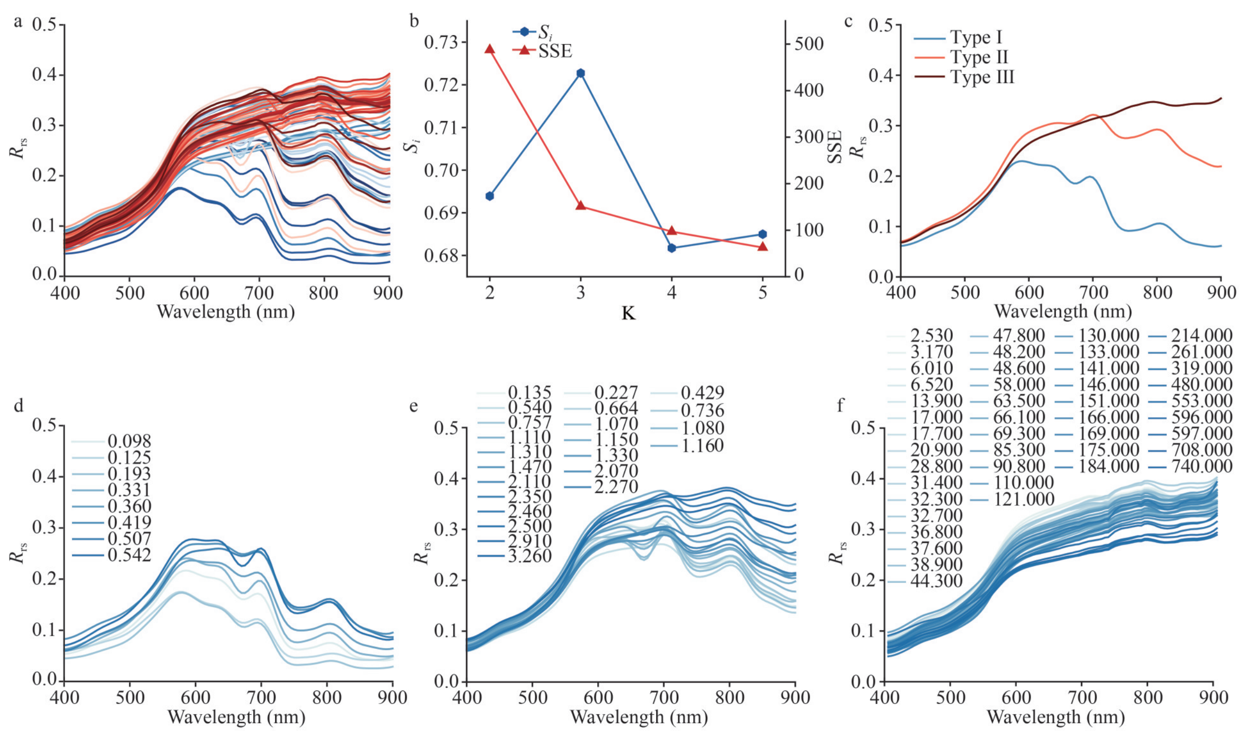

Fig.3 Clustering analysis of water spectral curves

a. preprocessed hyperspectral curves of water bodies; b. determination of the optimal cluster number K; c. mean spectral curves of the three water types; d. spectral curves of Type I; e. spectral curves of Type II; f. spectral curves of Type III. In (d–f), the color depth of the spectral curves corresponds to SSC (kg/m^3) levels, with darker colors indicating higher concentrations.

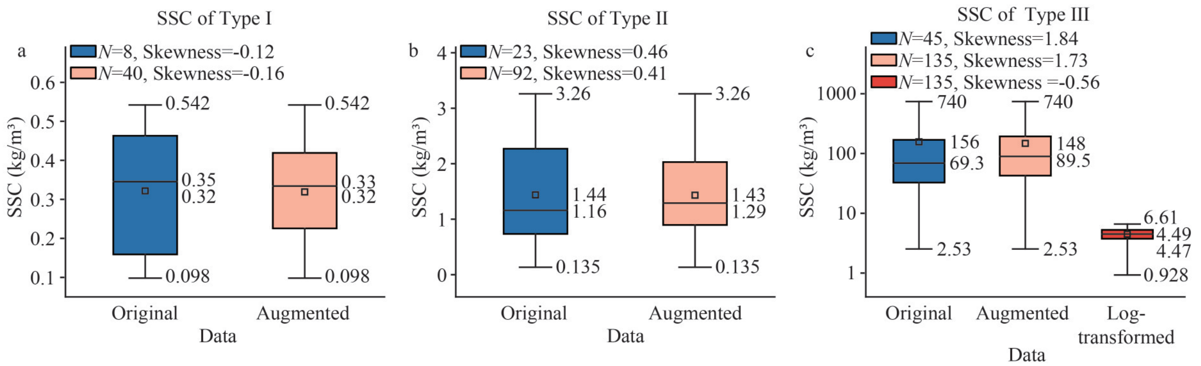

Fig.4 Box plots of SSC for the three water types

The blue represents original data, the pink denotes augmented data, and the dark red indicates log-transformed SSC. The box represents the interquartile range encompassing the middle 50% of the data. The solid line inside the box indicates the median, the square symbol represents the mean, and the upper and lower whiskers represent the maximum and minimum values, respectively.

this type. Therefore, these results confirm that classifying water samples before SSC inversion is indeed necessary.

Additionally, data augmentation was separately applied to each water type. Taking the condition of the augmented SSC as an example: the SSC ranges remain unchanged and the distribution shapes show only minimal variation. For instance, after augmentation, Type I contains 40 samples with a skewness of -0.16; Type II contains 92 samples with a skewness of 0.41; Type III contains 135 samples with a skewness of 1.73, which also closely match the original distribution. This indicates that the augmented data remain similar with the original data in terms of both statistical metrics and distribution patterns.

3.2 SSC inversion model and evaluation

For water Type I, the SSC model was developed based on 40 samples. These samples are characterized by generally low SSC with an approximately normal distribution (Fig.5a). The heatmaps (Fig.5b–c) of correlation results between all band combinations and SSC values show that BD combinations within the 700–830-nm and 450–550-nm ranges appear dark red, indicating higher R^2. Similarly, BR combinations in the 650–700-nm and 700–800-nm ranges also exhibit deep red colors, reflecting similarly high R^2. Guided by these high-correlation regions observed in the heatmap, we ranked all the correlation results. The three band combinations with the highest correlations were identified as $R_{rs}(825)$-$R_{rs}(770)$, $R_{rs}(795)$-$R_{rs}(450)$, and $R_{rs}(725)/R_{rs}(670)$, with R^2 of 0.87, 0.85, and 0.85, respectively. These three band combinations were selected to be the independent variables for developing both empirical and machine learning inversion models. For training and validation, the dataset was randomly divided into training and test sets at a 7:3 ratio to ensure statistical reliability of model evaluation (Lu and Su, 2020; Song et al., 2025). Table 1 lists the established models and accuracy evaluations. The scatter plots in Fig.5d–i describe the performance of different inversion models in terms of estimating SSC.

The linear empirical model using the band combination of $R_{rs}(795)$-$R_{rs}(450)$ achieves the best accuracy (R^2=0.92, RMSE=0.03 kg/m^3) on the test set. The XGBoost model approaches a similar accuracy (R^2=0.91). Compared to XGBoost model, linear empirical model has a simpler algorithm structure and less mathematical calculations. Therefore, the linear empirical model was ultimately employed to invert the SSC for Type I.

For Type II, the SSC inversion model was developed using 92 samples. The SSC distribution in these samples is characterized by dispersed and slightly right-skewed properties (Fig.6a). Following the same procedure as described for Type I, the correlation heatmaps were generated (Fig.6b–c). The heatmaps for both BD and BR combinations show deep red colors within the 720–800-nm range, indicating strong correlations with SSC values. Guided by this conclusion, the optimal band combinations were identified as $R_{rs}(750)$-$R_{rs}(745)$, $R_{rs}(775)$-$R_{rs}(725)$, and $R_{rs}(775)/R_{rs}(725)$, with the R^2 values of 0.89, 0.88, and 0.87, respectively. Table 2 lists the established models and accuracy evaluations. The scatter plots in Fig.6d–i describe the performance of different inversion models in terms of estimating SSC.

The comparative evaluation shows that all the

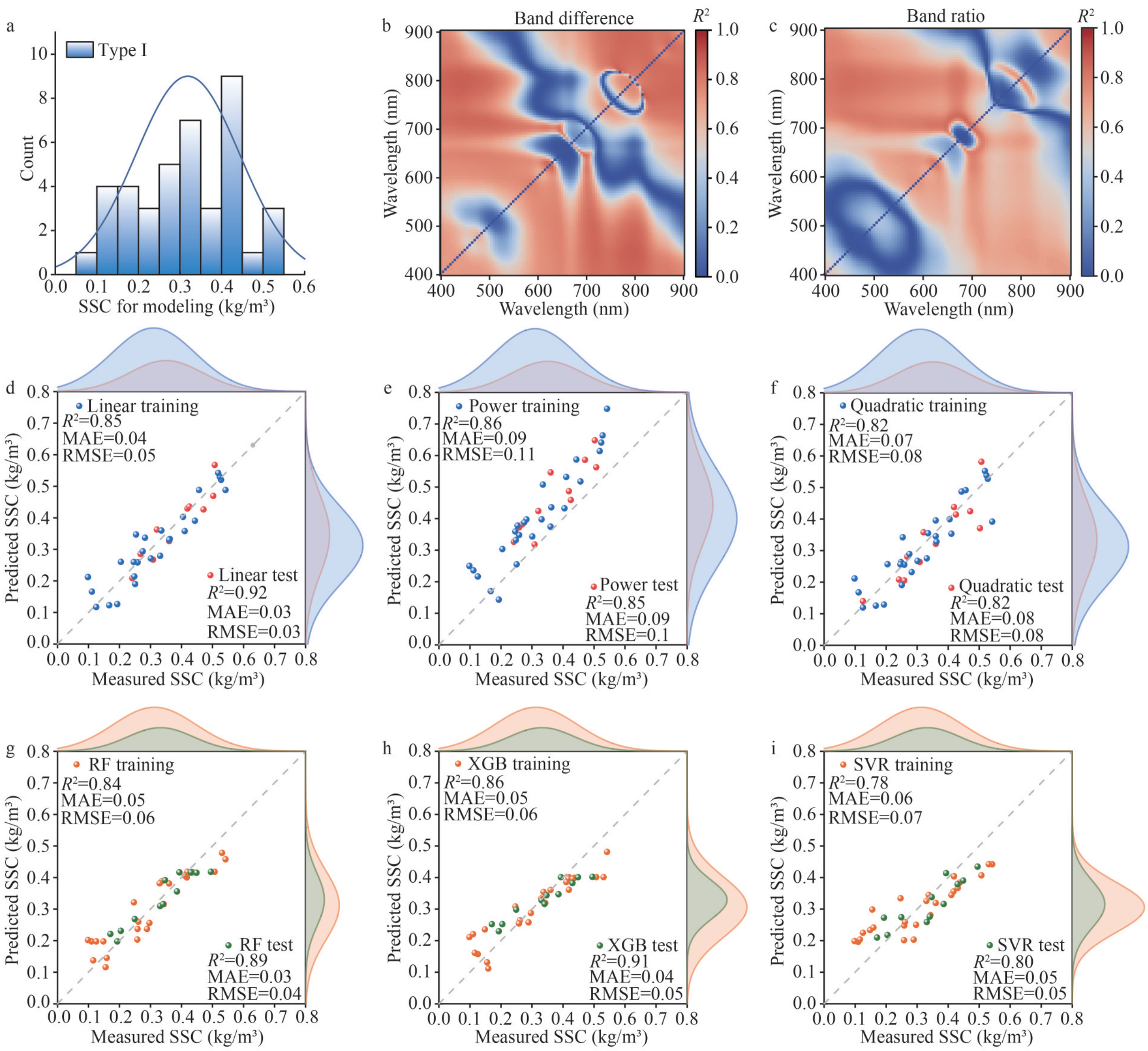

Fig.5 SSC inversion modeling process for Type I

a. SSC distribution used for modeling; b. heatmap between BD and SSC; c. heatmap between BR and SSC; d. measured SSC versus predicted SSC by the linear model; e. measured SSC versus predicted SSC by the power model; f. measured SSC versus predicted SSC by the quadratic model; g. measured SSC versus predicted SSC by the RF model; h. measured SSC versus predicted SSC by the XGBoost model; i. measured SSC versus predicted SSC by the SVR model. In (d–i), the units of RMSE and MAE are kg/m³.

Table 1 Comparison of SSC inversion models for Type I

No.	Band combination	Band combination R^2	Model	Training R^2	Test R^2
1	$R_{rs}(795)$-$R_{rs}(450)$	0.85	SSC=4.9925x+0.2213	0.85	0.92
2	$R_{rs}(725)/R_{rs}(670)$	0.85	SSC=1.4840$x^{4.0982}$	0.86	0.85
3	$R_{rs}(795)$-$R_{rs}(450)$	0.85	SSC=8.1777x^2+4.6426x+0.2201	0.82	0.82
4	$R_{rs}(825)$-$R_{rs}(770)$	0.87	RF	0.84	0.89
5	$R_{rs}(825)$-$R_{rs}(770)$	0.87	XGB	0.86	0.91
6	$R_{rs}(825)$-$R_{rs}(770)$	0.87	SVR	0.78	0.80

machine learning models achieve higher R^2 on the test set than the empirical models. Among them, the SVR model achieves the highest accuracy (R^2=0.92), while the best-performing empirical model only reaches an R^2 of 0.85. This indicates that in Type II water bodies with more complex optical

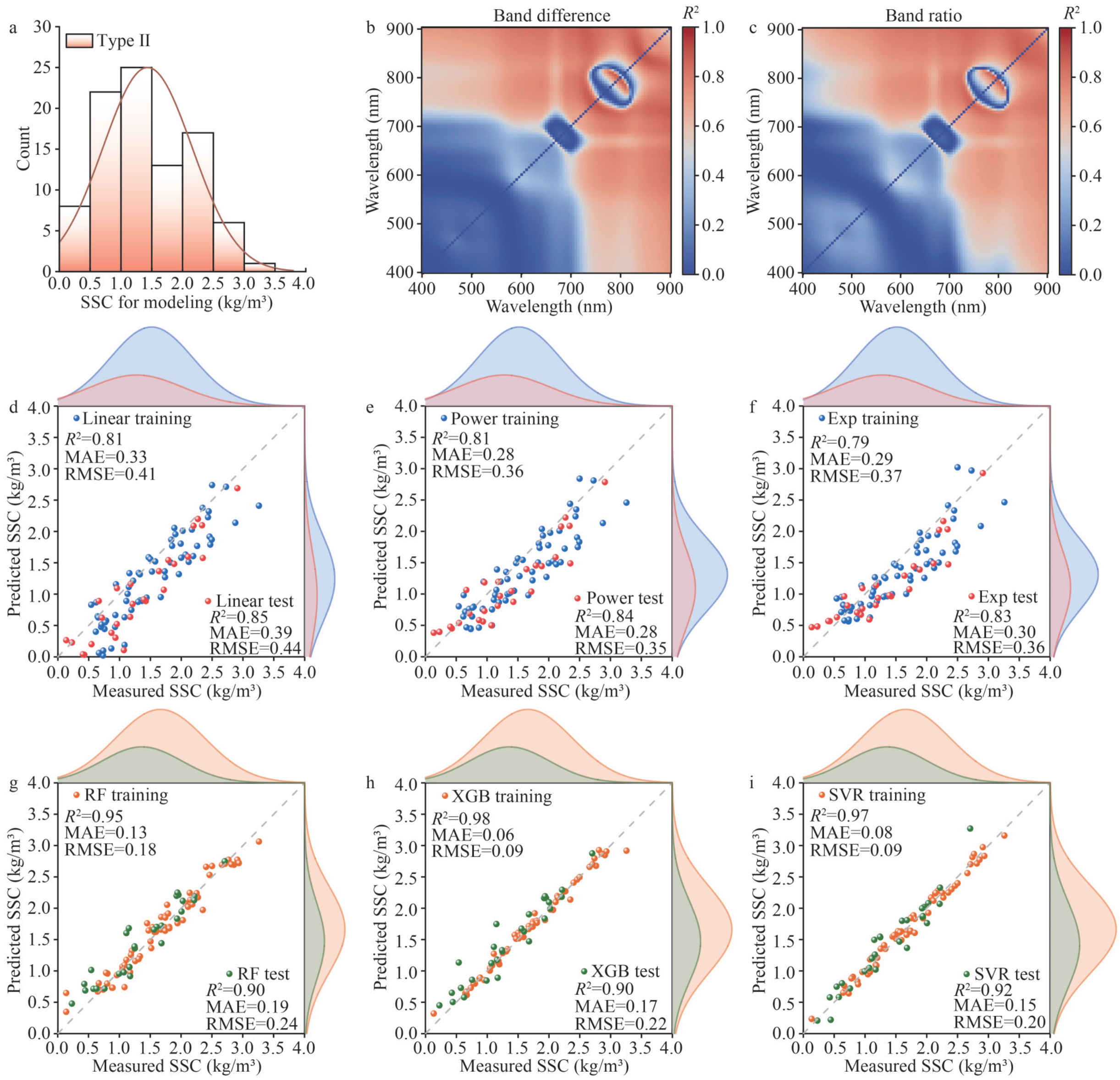

Fig.6 SSC inversion modeling process for Type II

a. SSC distribution used for modeling; b. heatmap between BD and SSC; c. heatmap between BR and SSC; d. measured SSC versus predicted SSC by the linear model; e. measured SSC versus predicted SSC by the power model; f. measured SSC versus predicted SSC by the exponential model; g. measured SSC versus predicted SSC by the RF model; h. measured SSC versus predicted SSC by the XGBoost model; i. measured SSC versus predicted SSC by the SVR model. In (d–i), the units of RMSE and MAE are kg/m³.

properties, the relationship between SSC and spectral features cannot be fully captured by predefined linear or exponential functions. In contrast, the SVR model was selected as the best one to invert the SSC for Type II due to its strong ability to handle complex nonlinear relationships.

For Type III, the SSC model was developed based on 135 samples. The SSC distribution of these samples exhibits a strongly right-skewed pattern. To improve the distribution pattern, a logarithmic transformation was applied. Following this transformation, the SSC distribution is close to normal distribution, with a skewness of -0.56 (Fig.7a). The heatmap for BD combinations within the 600–650 nm range exhibits relatively high correlations with SSC values (Fig.7b). In contrast, the heatmap for BR combinations appears mostly blue, indicating that the R^2 are below 0.5 (Fig.7c). Therefore, $R_{rs}(645)$-$R_{rs}(595)$, $R_{rs}(645)$-$R_{rs}(605)$, and $R_{rs}(645)$-$R_{rs}(585)$ were selected as the spectral features for the models. Table 3 lists the established models and their accuracy metrics. The scatter plots

Table 2 Comparisons of SSC inversion models for Type II

No.	Band combination	Band combination R^2	Model	Training R^2	Test R^2
1	$R_{rs}(775)$-$R_{rs}(725)$	0.88	SSC=57.8974x+2.0803	0.81	0.85
2	$R_{rs}(775)/R_{rs}(725)$	0.87	SSC=2.0678$x^{10.3414}$	0.81	0.84
3	$R_{rs}(775)$-$R_{rs}(725)$	0.88	SSC=2.0118e$^{35.7228x}$	0.79	0.83
4	$R_{rs}(750)$-$R_{rs}(745)$	0.89	RF	0.95	0.90
5	$R_{rs}(750)$-$R_{rs}(745)$	0.89	XGB	0.98	0.90
6	$R_{rs}(750)$-$R_{rs}(745)$	0.89	SVR	0.97	0.92

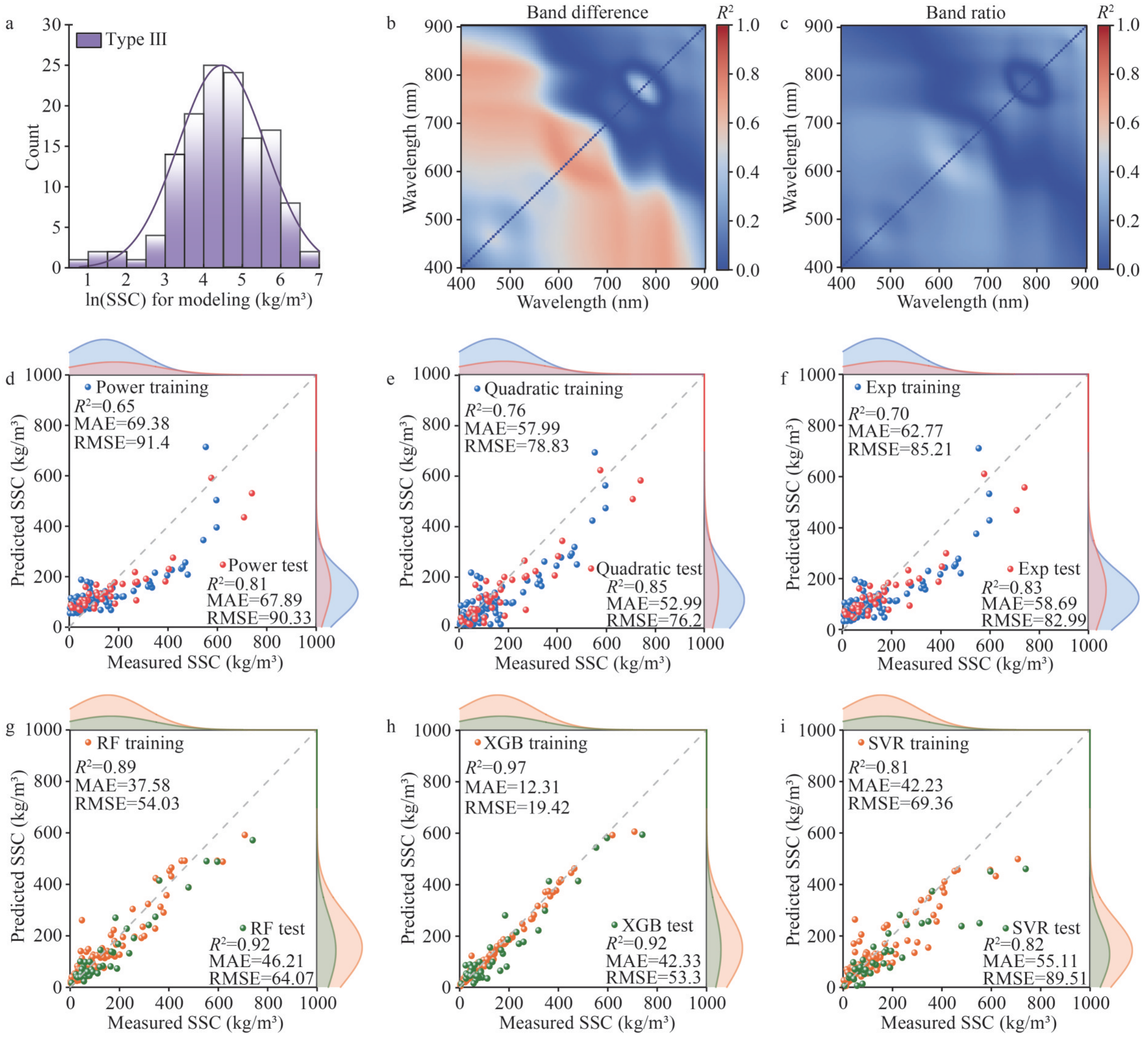

Fig.7 SSC inversion modeling process for Type III

a. log-transformed SSC distribution used for modeling; b. heatmap between BD and SSC; c. heatmap between BR and SSC; d. measured SSC versus predicted SSC by the power model; e. measured SSC versus predicted SSC by the quadratic model; f. measured SSC versus predicted SSC by the exponential model; g. measured SSC versus predicted SSC by the RF model; h. measured SSC versus predicted SSC by the XGBoost model; i. measured SSC versus predicted SSC by the SVR model. In (d–i), the units of RMSE and MAE are kg/m^3.

in Fig.7d–i describe the performance of different inversion models in terms of estimating SSC.

The empirical models show a large gap in R^2 between the training and test sets, indicating poor generalization. In contrast, all the three machine learning models achieve better performance on the

Table 3 Comparison of SSC inversion models for Type III

No.	Band combination	Band combination R^2	Model	Training R^2	Test R^2
1	$R_{rs}(645)$-$R_{rs}(595)$	0.73	SSC=0.0009$x^{-3.3558}$	0.65	0.81
2	$R_{rs}(645)$-$R_{rs}(595)$	0.73	SSC=188146.4108x^2–137249.5366x+2515.1516	0.76	0.85
3	$R_{rs}(645)$-$R_{rs}(595)$	0.73	SSC=9785.8357e$^{-150.2443x}$	0.70	0.83
4	$R_{rs}(645)$-$R_{rs}(605)$	0.74	RF	0.89	0.92
5	$R_{rs}(645)$-$R_{rs}(605)$	0.74	XGB	0.97	0.92
6	$R_{rs}(645)$-$R_{rs}(605)$	0.74	SVR	0.81	0.82

test set. Both RF and XGBoost achieve an R^2 of 0.92, reflecting excellent fitting stability. However, there is a difference in their training performance: XGBoost reaches a high R^2 of 0.97, while RF shows a lower but more stable R^2 of 0.89. This suggests that RF may offer better generalization stability with a lower risk of overfitting. Therefore, considering both accuracy and reliability, the RF model was identified as a reliable choice to predict SSC for Type III.

3.3 Comparison with global modeling

To validate the effectiveness of the classification-based modeling approach, a comparison was conducted between the classification-based and global modeling strategies for SSC inversion. Under the global modeling strategy, samples were not classified. All samples were used to develop both empirical and machine learning models. The R^2, RMSE, and MAE of each model were calculated and compared with the results from the best classification-based models in Section 3.2 (Table 4).

The results show that inversion accuracy is significantly improved by the classification-based approach. In the global modeling approach, all models perform poorly. The highest R^2 is only 0.57 on the test set, obtains by the RF model. The linear empirical model achieves an R^2 of just 0.45, while its MAE is as high as 107.06 kg/m^3. In contrast, using classification-based modeling, the optimal models for each water type show notable improvements: the linear model for Type I achieves an R^2 of 0.92 with an MAE of only 0.03 kg/m^3 on the test set; the SVR model for Type II reaches an R^2 of 0.92 with an MAE of 0.20 kg/m^3; and the RF model for Type III also attains an R^2 of 0.92. In summary, the classification-based modeling strategy improved both prediction accuracy and practical value by building separate models based on the optical properties of each water type.

Table 4 Comparison among different modeling methods

Method	Model	Test R^2	RMSE (kg/m^3)	MAE (kg/m^3)
Global modeling	Linear	0.45	137.49	107.06
	Power	0.43	130.47	99.97
	Exponential	0.49	127.29	99.43
	Quadratic	0.53	132.97	101.35
	RF	0.57	92.58	48.88
	XGBoost	0.53	96.91	49.52
	SVR	0.49	99.32	48.44
Classification-based modeling	Type I: linear	0.92	0.03	0.03
	Type II: SVR	0.92	0.20	0.15
	Type III: RF	0.92	64.07	46.21

4 DISCUSSION

The Qingshui River Basin in the upper Huanghe River has sparse vegetation and severe soil erosion. Sudden increases in SSC frequently occur during flood periods (Ma et al., 2020), making SSC monitoring critically important for this area (Sun et al., 2020). However, existing research primarily focused on SSC inversion for low-sediment water (Cao et al., 2022; Yu et al., 2022), with only a few studies dedicated to SSC inversion for medium or high sediment concentrations. These studies attempted to establish a single SSC inversion model for high-sediment water, but the results were not ideal. To further improve the inversion accuracy for high-sediment water, this study used the K-means clustering algorithm to classify the water samples into three types based on their spectral characteristics. There exist distinct relationships between spectral shapes and SSC values across these three water types. Water Type I has a low SSC range (0.098–0.542 kg/m^3). Its spectral curve shows a double-peak shape, and its spectral reflectance increases with rising SSC. This spectral signature is primarily determined by the combined effects of absorption by dissolved substances and scattering by fine sediment particles. Due to the low SSC and small particle size in this water type, particle scattering is more pronounced in the red to near-infrared bands (He et al., 2013).

The SSC of water Type II ranges from 0.135 to 3.26 kg/m^3. Its spectral reflectance peak shifts to the red band, and the overall reflectance is higher. This phenomenon is mainly attributed to increased particle backscattering resulting from higher SSC (Jiang et al., 2013). Furthermore, the weak absorption of Huanghe River fine sediment in the near-infrared band further accentuates the reflectance peak in this spectral region (Doxaran et al., 2009; Lu et al., 2025).

Water Type III has a high SSC range (2.53–740 kg/m^3). Its spectral curve changes into a single-peak shape, and its spectral reflectance with high values tends to have lower SSC values. This behavior reflects the substantially enhanced absorption in the visible range caused by high-sediment concentrations, which suppresses effective backscattering. Additionally, light-absorbing constituents within the particles themselves may further dampen reflectance, leading to saturation of the spectral signal (Chen et al., 1991; Nechad et al., 2010).

Following the classification, separate SSC inversion models were established for each water type. For water Type I, a linear empirical model was used to reflect the relationship between spectral features and low SSC values. For water Type II and III, machine learning models were employed to establish the complex nonlinear relationship between the saturated spectral features and medium to high SSC values. This approach aligns with the demonstrated advantages of machine learning in inverting complex water (Dehkordi et al., 2021; Hu et al., 2023; Zaghian et al., 2023).

Furthermore, the classification-based modeling methodology took many factors into account, such as season changes, time periods and hydrological conditions of the study area. This study used the ground-based in-situ hyperspectral data and measured SSC data from the Wangtuan station between March and September 2024, which covered both normal and flood periods. The measured SSC values cover the sediment concentration range of the study area and represent the normal sediment conditions. During normal flow periods (March–May), the spectral curves of water bodies appear to have the features of water Type I. This makes it possible to use linear empirical models for monitoring subtle SSC variations. In the flood periods (July–August), the spectral curves of water bodies are likely to have the features of water Type II or Type III. During this period, machine learning models could capture the rapid fluctuations of SSC. Figure 8 presents the measured and predicted SSC values in chronological order. The data points show a clear transition from low SSC during normal flow periods to high SSC during flood events. These results illustrate clear seasonal relevance of our proposed classification-based modeling methodology.

This study proposed and validated a practical methodology for SSC inversion in Wangtuan region of the Qingshui River. To improve its stability for long-term monitoring and applicability to other regions, more in-situ hyperspectral instruments

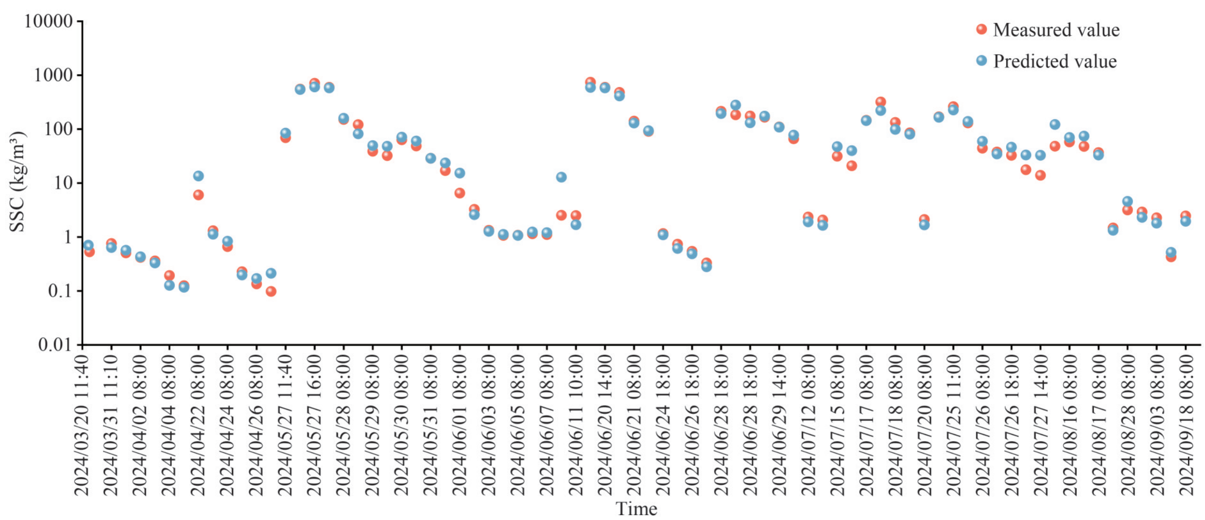

Fig.8 Chronological comparison of measured and predicted SSC

should be deployed in key areas of the Qingshui River Basin, such as the main stem and typical tributaries. These hyperspectral instruments can obtain more representative spatiotemporal samples, and thereby expand the dataset with hyperspectral data and corresponding measured SSC data under medium and high SSC conditions. Subsequently, the expanded dataset will be used for optimizing the parameters of the SSC inversion models proposed in this study. Eventually, the optimized inversion models are expected to be stable across other tributaries or river sections of the Huanghe River. Moreover, if the optical characteristics of the added hyperspectral data into the expanded dataset are consistent with the regularities summarized in this section mentioned above, the proposed water spectral classification algorithm in this study can be directly applied to the added hyperspectral data with no more re-clustering. However, if the optical characteristics of the added hyperspectral data are affected by local factors such as topography or extreme weather and differ from the regularities summarized in this section mentioned above, it is necessary to follow the steps of ‘local data collection-spectral classification-model building’ to develop the corresponding SSC inversion model suitable for that specific region. The performance of the inversion models should also be evaluated again.

Currently, large-scale satellite projects such as SpaceX’s Starlink and China’s constellation plans are advancing rapidly (Ding et al., 2026). More and more satellites are expected to provide remote sensing data with more spectral dimensions and higher ground spatial resolution. Also, the increase in the number of satellites will inevitably raise the frequency of satellite crossings. Therefore, the satellite remote sensing will play an increasingly important role in SSC monitoring of the Huanghe River. When the satellite crosses the river, the data obtained synchronously from the ground-based in-situ hyperspectral instruments can serve as the reference measured SSC data for further inversion with satellite remote sensing data. The proposed SSC inversion models can be applied to satellite remote sensing data by matching ground-based spectral data with satellite data bands, which allows for the establishment of an integrated space-earth collaborative SSC monitoring network. It should be noted that in practical application, ground-based hyperspectral data need to be resampled according to the specific spectral response function of the satellite sensor (Yang et al., 2023). Additionally, common bands sensitive to SSC should be selected for model calibration to ensure the feasibility and accuracy of extending this monitoring method from ground observation to space-earth monitoring application.

5 CONCLUSION

In this study, we proposed a ‘classify first, then model’ method to develop and validate SSC inversion models in Qingshui River Basin of Ningxia section of the Huanghe River. In the Qingshui River Basin, the SSC varies with the seasons and hydrological conditions in a wide dynamic range, even reaching hundreds of kilograms per cubic meter. Additionally, the SSC exhibits a short-term variability with values rising by orders of magnitude within hours in flood periods. Therefore, conventional methods are inadequate for monitoring such highly dynamic sediment processes and such rapid fluctuations of SSC. While the proximal remote sensing technology using ground-based in-situ above-water hyperspectral sensors enables non-contact optical measurements and can capture SSC changes with high temporal resolution. In this paper we used in-situ hyperspectral data with synchronously measured SSC and proposed the ‘classify first, then model’ method to develop SSC inversion models. This method identified three types of water bodies with distinct spectral characteristics through K-means clustering and established optimal inversion models for each water type. The results demonstrated that inversion accuracy was significantly improved by our proposed classification-based method.

The methodology proposed in this study aims to enhance the capability to capture transient sediment flood events and to improve the inversion accuracy of remote sensing methods in optically complex water bodies. The application of this classification-based method in the Qingshui River Basin demonstrates its feasibility and provides a methodological reference for SSC monitoring in other similarly high SSC, optically complex water bodies within the Huanghe River Basin. In the future, with the continuous advancement of hyperspectral remote sensing and artificial intelligence, it is expected that more robust and adaptable hydrological informatization technologies will be developed for accurate monitoring and intelligent management of watershed environments, and will promote ecological protection and high-quality development of the Huanghe River Basin.

6 DATA AVAILABILITY STATEMENT

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

References

Bernardo N, Do Carmo A, Park E et al. 2019. Retrieval of suspended particulate matter in inland waters with widely differing optical properties using a semi-analytical scheme. *Remote Sensing*, **11**(19): 2283, https://doi.org/10.3390/rs11192283.

Bowers D G, Binding C E. 2006. The optical properties of mineral suspended particles: a review and synthesis. *Estuarine*, *Coastal and Shelf Science*, **67**(1-2): 219-230, https://doi.org/10.1016/j.ecss.2005.11.010.

Cao B W, Qiu J L, Zhang W X et al. 2022. Retrieval of suspended sediment concentrations in the Pearl River estuary using multi-source satellite imagery. *Remote Sensing*, **14**(16): 3896, https://doi.org/10.3390/rs14163896.

Chen Y P, Fu B J, Zhao Y et al. 2020. Sustainable development in the Yellow River Basin: Issues and strategies. *Journal of Cleaner Production*, **263**: 121223, https://doi.org/10.1016/j.jclepro.2020.121223.

Chen Z M, Hanson J D, Curran P J. 1991. The form of the relationship between suspended sediment concentration and spectral reflectance: its implications for the use of Daedalus 1268 data. *International Journal of Remote Sensing*, **12**(1): 215-222, https://doi.org/10.1080/01431169108929647.

Christilda A J, Manoharan R. 2023. Enhanced hyperspectral image segmentation and classification using K-means clustering with connectedness theorem and swarm intelligent-BiLSTM. *Computers and Electrical Engineering*, **110**: 108897, https://doi.org/10.1016/j.compeleceng.2023.108897.

Dehkordi A T, Ghasemi H, Zoej M J V. 2021. Machine learning-based estimation of suspended sediment concentration along Missouri River using remote sensing imageries in Google Earth Engine. *In*: 2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS). IEEE, Tehran, Islamic Republic of Iran. p.1-5, https://doi.org/10.1109/ICSPIS54653.2021.9729382.

Ding X Y, Li X Q, Li S Y et al. 2026. The industrial innovation ecosystem construction of SpaceX's "Starlink" program and its implications for the development of China's commercial aerospace. *Forum on Science and Technology in China*, (1): 172-180, https://doi.org/10.3969/j.issn.1002-6711.2026.01.019. (in Chinese with English abstract)

Dinh D T, Fujinami T, Huynh V N. 2019. Estimating the optimal number of clusters in categorical data clustering by silhouette coefficient. *In*: 20th International Symposium on Knowledge and Systems Sciences. Springer, Da Nang, Vietnam. p.1-17, https://doi.org/10.1007/978-981-15-1209-4_1.

Doxaran D, Froidefond J M, Castaing P et al. 2009. Dynamics of the turbidity maximum zone in a macrotidal estuary (the Gironde, France): observations from field and MODIS satellite data. *Estuarine*, *Coastal and Shelf Science*, **81**(3): 321-332, https://doi.org/10.1016/j.ecss.2008.11.013.

Duan L T, Ma Y G. 2021. Characteristic and change of sediment analysis in Ningxia in the Past 60 Years. *Journal of Northwest Engineering Technology*, **20**(1): 34-38, https://doi.org/10.3969/j.issn.1671-7244.2021.01.007. (in Chinese with English abstract)

Duan M W, Qiu Z Q, Li R R et al. 2024. Monitoring suspended sediment transport in the lower yellow river using landsat observations. *Remote Sensing*, **16**(2): 229, https://doi.org/10.3390/rs16020229.

Feng L, Hu C M, Chen X L et al. 2012. Human induced turbidity changes in Poyang Lake between 2000 and 2010: observations from MODIS. *Journal of Geophysical Research*: *Oceans*, **117**(C7): C07006, https://doi.org/10.1029/2011JC007864.

Fisher R A. 1930. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.

Gao P. 2008. Understanding watershed suspended sediment transport. *Progress in Physical Geography*: *Earth and Environment*, **32**(3): 243-263, https://doi.org/10.1177/0309133308094849.

Haut J M, Paoletti M, Plaza J et al. 2017. Cloud implementation of the K-means algorithm for hyperspectral image analysis. *The Journal of Supercomputing*, **73**(1): 514-529, https://doi.org/10.1007/s11227-016-1896-3.

He X Q, Bai Y, Pan D L et al. 2013. Using geostationary satellite ocean color data to map the diurnal dynamics of suspended particulate matter in coastal waters. *Remote Sensing of Environment*, **133**: 225-239, https://doi.org/10.1016/j.rse.2013.01.023.

Hou C Y, Yi Y J, Song J et al. 2021. Effect of water-sediment regulation operation on sediment grain size and nutrient content in the lower Yellow River. *Journal of Cleaner Production*, **279**: 123533, https://doi.org/10.1016/j.jclepro.2020.123533.

Hou Y K, Zhang A B, Lv R L et al. 2022. A study on water quality parameters estimation for urban rivers based on ground hyperspectral remote sensing technology. *Environmental Science and Pollution Research*, **29**(42): 63640-63654, https://doi.org/10.1007/s11356-022-20293-z.

Hou Y Z, Ma Y G, Hou Z et al. 2024. Space-ground remote sensor network for monitoring suspended sediments in the Yellow River Basin. *Sensors*, **24**(21): 6888, https://doi.org/10.3390/s24216888.

Hou Y Z, Xing Q G, Zheng X Y et al. 2023. Monitoring suspended sediment concentration in the Yellow River Estuary and its vicinity waters on the basis of SDGSAT-1 multispectral imager. *Water*, **15**(9): 3522, https://doi.org/10.3390/w15193522.

Hou Z, Hou Y Z, Xing Q G et al. 2025. Investigation of monitoring models for extremely high suspended sediment concentration based on in-situ hyperspectral data. *Yellow River*, **47**(4): 28-31, 37, https://doi.org/10.3969/j.issn.1000-1379.2025.04.005. (in Chinese with English abstract)

Hu J L, Miao C Y, Zhang X P et al. 2023. Retrieval of suspended sediment concentrations using remote sensing and machine learning methods: a case study of the lower Yellow River. *Journal of Hydrology*, **627**: 130369, https://doi.org/10.1016/j.jhydrol.2023.130369.

Humaira H, Rasyidah R. 2020. Determining the appropiate cluster number using elbow method for K-means algorithm. *In*: Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA). EAI, Padang, Indonesia. p.1-8.

Jally S K, Mishra A K, Balabantaray S. 2021. Retrieval of suspended sediment concentration of the Chilika Lake, India using Landsat-8 OLI satellite data. *Environmental Earth Sciences*, **80**(8): 298, https://doi.org/10.1007/s12665-021-09581-y.

Jiang L L, Zhao D Z, Wang L et al. 2013. Methods and research progress on backscattering coefficient in the water. *Remote Sensing Technology and Application*, **28**(1): 150-156, https://doi.org/10.11873/j.issn.1004-0323.2013.1.150. (in Chinese with English abstract)

Jiao P, Wei W Y, Bao H Z. 2020. Variation and trend analysis of rainfall in Qingshui River Basin of Ningxia in China. *IOP Conference Series*: *Earth and Environmental Science*, **526**(1): 012040, https://doi.org/10.1088/1755-1315/526/1/012040.

Lai H X, Huang T, Lu B L et al. 2025. Silhouette coefficient-based weighting k-means algorithm. *Neural Computing and Applications*, **37**(5): 3061-3075, https://doi.org/10.1007/s00521-024-10706-0.

Lu J, Cheng Y Y, Zhang H X et al. 2025. Erosion and sedimentation characteristics of grouped sediment during flood process in the Ningxia-Inner Mongolia reach of the Yellow River. *Journal of Sediment Research*, **50**(6): 38-45, https://doi.org/10.16239/j.cnki.0468-155x.2025.06.006. (in Chinese with English abstract)

Lu X M, Su H. 2020. Retrieving total suspended matter concentration in Fujian coastal waters using OLCI data. *Acta Scientiae Circumstantiae*, **40**(8): 2819-2827, https://doi.org/10.13671/j.hjkxxb.2020.0176. (in Chinese with English abstract)

Ma J G, Zheng Y S, Zhang X H et al. 2020. Analysis on the characteristics of flow and sediment variation in Qingshuihe River basin of Ningxia. *Hydro-Science and Engineering*, (4): 57-63, https://doi.org/10.12170/20200213002. (in Chinese with English abstract)

Mao Z H, Chen J Y, Pan D L et al. 2012. A regional remote sensing algorithm for total suspended matter in the East China Sea. *Remote Sensing of Environment*, **124**: 819-831, https://doi.org/10.1016/j.rse.2012.06.014.

Matos T, Martins M S, Henriques R et al. 2024. A review of methods and instruments to monitor turbidity and suspended sediment concentration. *Journal of Water Process Engineering*, **64**: 105624, https://doi.org/10.1016/j.jwpe.2024.105624.

Miao J D, Zhang X M, Zhao Y et al. 2022. Evolution patterns and spatial sources of water and sediment discharge over the last 70 years in the Yellow River, China: a case study in the Ningxia Reach. *Science of the Total Environment*, **838**: 155952, https://doi.org/10.1016/j.scitotenv.2022.155952.

Montanher O C, Novo E M L M, Barbosa C C F et al. 2014. Empirical models for estimating the suspended sediment concentration in Amazonian white water rivers using Landsat 5/TM. *International Journal of Applied Earth Observation and Geoinformation*, **29**: 67-77, https://doi.org/10.1016/j.jag.2014.01.001.

Montgomery D C, Peck E A, Vining G G. 2012. Introduction to Linear Regression Analysis. Wiley, Hoboken.

Nechad B, Ruddick K G, Park Y. 2010. Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sensing of Environment*, **114**(4): 854-866, https://doi.org/10.1016/j.rse.2009.11.022.

Ortiz-Lopez C, Bouchard C, Rodriguez M. 2022. Machine learning models with potential application to predict source water quality for treatment purposes: a critical review. *Environmental Technology Reviews*, **11**(1): 118-147, https://doi.org/10.1080/21622515.2022.2118084.

Rabeni C F. 1997. Sediment in streams: sources, biological effects, and control. *Transactions of the American Fisheries Society*, **126**(6): 1048-1051, https://doi.org/10.1577/1548-8659-126.6.1048.

Rai A K, Kumar A. 2015. Continuous measurement of suspended sediment concentration: technological advancement and future outlook. *Measurement*, **76**: 209-227, https://doi.org/10.1016/j.measurement.2015.08.013.

Ranjan S, Nayak D R, Kumar K S et al. 2017. Hyperspectral image classification: a k-means clustering based approach. *In*: 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, Coimbatore, India. p.1-7, https://doi.org/10.1109/ICACCS.2017.8014707.

Rinnan Å, Van Den Berg F, Engelsen S B. 2009. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, **28**(10): 1201-1222, https://doi.org/10.1016/j.trac.2009.07.007.

Shen Q, Li J S, Zhang F F et al. 2015. Classification of several optically complex waters in China using *in situ* remote sensing reflectance. *Remote Sensing*, **7**(11): 14731-14756, https://doi.org/10.3390/rs71114731.

Song G X, Jiang Y J, Lei X Y et al. 2025. Measurement of suspended sediment concentration at the outlet of the Yellow River Canyon: using Sentinel-2 images and machine learning. *Remote Sensing*, **17**(14): 2424, https://doi.org/10.3390/rs17142424.

Sun P C, Wu Y P, Gao J N et al. 2020. Shifts of sediment transport regime caused by ecological restoration in the Middle Yellow River Basin. *Science of the Total Environment*, **698**: 134261, https://doi.org/10.1016/j.

scitotenv.2019.134261.

Sun X, Zhang Y L, Shi K et al. 2022. Monitoring water quality using proximal remote sensing technology. *Science of the Total Environment*, **803**: 149805, https://doi.org/10.1016/j.scitotenv.2021.149805.

Umargono E, Suseno J E, Gunawan V S K. 2020. K-means clustering optimization using the elbow method and early centroid determination based on mean and median formula. *In*: The 2nd International Seminar on Science and Technology (ISSTEC 2019). Atlantis Press. p.121-129, https://doi.org/10.2991/assehr.k.201010.019.

Wang C L, Shi K Y, Ming X et al. 2022. A comparative study of the COD hyperspectral inversion models in water based on the maching learning. *Spectroscopy and Spectral Analysis*, **42**(8): 2353-2358, https://doi.org/10.3964/j.issn.1000-0593(2022)08-2353-06. (in Chinese with English abstract)

Wang F, Zhou B, Xu J M et al. 2009. Surface spectral measurement and characteristics analysis of turbid water in Hangzhou Bay. *Spectroscopy and Spectral Analysis*, **29**(3): 730-734, https://doi.org/10.3964/j.issn.1000-0593(2009)03-0730-05. (in Chinese with English abstract)

Wang X Y, Tang J W, Li T J et al. 2012. Key Technologies of water spectra measurements with above-water method. *Journal of Ocean Technology*, **31**(1): 72-76, https://doi.org/10.3969/j.issn.1003-2029.2012.01.018. (in Chinese with English abstract)

Wen Y F, Li P F, Li M Z et al. 2025. Changes in River cross-section morphology and response to streamflow and sediment processes in middle reaches of Yellow River, China. *Chinese Geographical Science*, **35**(1): 161-174, https://doi.org/10.1007/s11769-025-1488-3.

Willmott C J. 1981. On the validation of models. *Physical Geography*, **2**(2): 184-194, https://doi.org/10.1080/02723646.1981.10642213.

Wu B S, Wang Z Y, Li C Z. 2004. Yellow River Basin management and current issues. *Journal of Geographical Sciences*, **14**(1): 29-37, https://doi.org/10.1007/BF02841104.

Yang C X, Li Y, Yang J et al. 2023. Remote sensing inversion and regularity analysis of suspended sediment in Pearl River Estuary based on machine learning model. *Bulletin of Surveying and Mapping*, (9): 117-123, https://doi.org/10.13474/j.cnki.11-2246.2023.0245. (in Chinese with English abstract)

Yang J, Xu J H, Zeng C J et al. 2025. Water and sediment effects of soil and water conservation in Qingshui River Basin, Ningxia. *Journal of Changjiang River Scientific Research Institute*, **42**(6): 94-101, https://doi.org/10.11988/ckyyb.20240439. (in Chinese with English abstract)

Yang J, Zhang H L, Yang W Q. 2022a. Coupling effects of precipitation and vegetation on sediment yield from the perspective of spatiotemporal heterogeneity across the Qingshui River Basin of the Upper Yellow River, China. *Forests*, **13**(3): 396, https://doi.org/10.3390/f13030396.

Yang P M, Hu J, Hu B F et al. 2022b. Estimating soil organic matter content in desert areas using in situ hyperspectral data and feature variable selection algorithms in Southern Xinjiang, China. *Remote Sensing*, **14**(20): 5221, https://doi.org/10.3390/rs14205221.

Yu Z F, Wang J W, Li Y et al. 2022. Remote sensing of suspended sediment in high turbid estuary from sentinel-3A/OLCI: a case study of Hangzhou Bay. *Frontiers in Marine Science*, **9**: 1008070, https://doi.org/10.3389/fmars.2022.1008070.

Zaghian S, Mohammadzadeh A, Ghorbanian A et al. 2023. Enhancing suspended sediment concentration retrieval by integrating thermal infrared and optical bands of Landsat-8 and machine learning algorithms. *International Journal of Remote Sensing*, **44**(18): 5814-5844, https://doi.org/10.1080/01431161.2023.2255350.

Zhang C L, Liu Y X, Chen X W et al. 2022. Estimation of suspended sediment concentration in the Yangtze main stream based on sentinel-2 MSI data. *Remote Sensing*, **14**(18): 4446, https://doi.org/10.3390/rs14184446.

Zhang H Y, Cisse M, Dauphin Y N et al. 2017. mixup: beyond empirical risk minimization. arXiv preprint arXiv: 1710.09412, https://arxiv.org/abs/1710.09412.

Zhang M W, Dong Q, Cui T W et al. 2014. Suspended sediment monitoring and assessment for Yellow River estuary from Landsat TM and ETM+ imagery. *Remote Sensing of Environment*, **146**: 136-147, https://doi.org/10.1016/j.rse.2013.09.033.

Zhu L L, Suomalainen J, Liu J B et al. 2018. A review: remote sensing sensors. *In*: Rustamov R B, Hasanova S, Zeynalova M H eds. Multi-Purposeful Application of Geospatial Data. IntechOpen, London, https://doi.org/10.5772/intechopen.71049.

Zhu W Y, Li K P, Li L K et al. 2022. Evolution law of water and sediment and spatial distribution characteristics of erosion and deposition in the Ningxia Section of the Yellow River. *Yellow River*, **44**(8): 28-33, https://doi.org/10.3969/j.issn.1000-1379.2022.08.006. (in Chinese with English abstract)