IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 May 2024, accepted 8 June 2024, date of publication 17 June 2024, date of current version 25 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3414996

RESEARCH ARTICLE

Suspended Sediment Concentrate Estimation From Landsat Imagery and Hydrological Station in Poyang Lake Using Machine Learning

KAITAO LIAO[1,2,3], YUEJUN SONG[2], XIAOFEI NIE[2], LINGJIA LIU[1,3], AND SHUHUA QI[1,3]

[1]School of Geography and Environment, Jiangxi Normal University, Nanchang 330022, China
[2]Key Laboratory of Soil Erosion and Prevention, Jiangxi Academy of Water Science and Engineering, Nanchang 330029, China
[3]Key Laboratory of Poyang Lake Wetland and Watershed Research (Ministry of Education), Jiangxi Normal University, Nanchang 330022, China

Corresponding author: Shuhua Qi (qishuhua11@jxnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 42330108, Grant 42361062, and Grant 42267058; and in part by the Water Conservancy Science and Technology Project of Jiangxi Province, China, under Grant 202425TGKT07.

ABSTRACT Poyang Lake is not only a globally important stopover for migratory birds and a habitat for fish, but also one of the main sand mining areas in China. However, sand mining activities significantly affect the suspended sediment concentration in Poyang Lake, thereby impacting water quality, altering lake bed topography, and potentially disturbing the living environment of plants and animals. Therefore, understanding the suspended sediment concentration and the spatiotemporal patterns of sand mining activities in Poyang Lake holds vital significance for effective lake management and biodiversity conservation. Utilizing surface reflectance (SR) data derived from Landsat satellite imagery sourced via Google Earth Engine spanning from 1989 to 2018, coupled with SSC data recorded by daily observation at Hukou hydrological stations, we conducted a comparative analysis of five distinct methods: Linear Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Classification and Regression Trees (CART), and Back Propagation Neural Network (BPNN). Moreover, the applicability of machine learning methods was analyzed and compared across four scenarios (#01 Landsat 5, #02 Landsat 7, #03 Landsat 8, and #04, Landsat 5, 7, and 8). Using the optimal SSC estimation model, we investigated the spatial and temporal trends of sand mining activities in Poyang Lake from 1989 to 2021, along with their potential impact on suspended sediment levels in the water body. The results showed that the prediction accuracy of machine learning is better than that of linear regression models, and the RF model has the best performance. The RF model generated R^2 >0.9 for four scenarios (#01 Landsat 5, #02 Landsat 7, #03 Landsat 8, and #04 Landsat 5, 7, and 8) and showed little to no overfitting. The generated SSC map can clearly show the distribution of SSC in Poyang Lake, unveiling the impact of sand mining activities on suspended sediment, especially around sand dredgers where SSC can exceed 0.15 g/L. Sand mining activities in Poyang Lake emerged after 2000 and gradually shifted southward and expanded, reaching a peak in 2016. Fortunately, under government regulation, illegal sand mining has been effectively controlled and is currently concentrated near Songmenshan Island. Despite the fact that sand mining has been brought under control, there remains a necessity for heightened oversight to prevent any impact on national nature reserves. The Random Forest (RF) model demonstrates significant potential in utilizing Landsat satellite data to predict SSC in Poyang Lake, as well as to monitor sand mining activities in the area.

INDEX TERMS Machine learning, suspended sediment, Landsat, sand mining, Poyang Lake.

The associate editor coordinating the review of this manuscript and approving it for publication was Stefania Bonafoni.

I. INTRODUCTION

Suspended sediment concentration (SSC) is an important indicator of the water quality conditions in lakes and can be

used to determine the degree of sedimentation, erosion, and pollution in the water body [1], [2], [3]. Suspended sediment makes the lake water turbid, affecting sunlight entering the water column and affecting the growth and reproduction of aquatic plants [4], [5]. Additionally, excessive suspended sediment may also lead to the accumulation or blockage of flow space in rivers and lakes, affecting navigation and water conservancy facilities [6], [7]. Therefore, suspended sediment monitoring can be used to better understand the lake ecosystem, provide information for management strategies and decision-making, and help protect the overall health and sustainability of the lake [8].

Field water quality sampling and laboratory analysis methods to detect the concentration and type of suspended sediment are accurate [9]. However, collecting field samples is time consuming and costly method and cannot explain the spatial distribution of the SSC in large lakes [10].Using mathematical models and computer simulations, combined with other hydrological environmental parameters, to estimate the suspended sediment content is more efficient and requires many parameters [11]. Nevertheless, those parameters are not easy to obtain and make the method difficult to apply [12], [13]. Currently, researchers use remote sensing images to estimate SSC at the regional scale because satellite remote sensing has the ability to capture long time series images of large areas [14], [15]. Satellite sensors like Worldview, Landsat and MODIS have been used for SSC inversion and monitoring [16], [17]. Landsat images are suitable image sensors for monitoring suspended sediment in medium-sized inland waters [18].

Poyang Lake in South China is one of the largest wetlands in Asia. Serving as a crucial hub for shipping and agriculture, it is a major source of sediment for the Yangtse River [19]. Additionally, it is one of the most important habitats for migratory birds [20] and Yangtze finless porpoise, also known as the ''panda of the water'' [21]. Fed by fiver rivers that flow into the Yangtze River at Hukou County, Poyang Lake covers an area of approximately 3675 km^2 in the wet season, with a water storage capacity is about 1.24 10^8 m^3. Driven by the substantial demand for sand and gravel from the real estate and construction industries in the middle and lower reaches of the Yangtze River in China, and influenced by the comprehensive ban on sand mining in the main channel of the Yangtze River since 2000, numerous sand mining vessels have shifted their operations to Poyang Lake [22], [23]. The Chinese government has proposed the construction of the Poyang Lake Hydraulic Project (PLHP) in recently year [24]. On January 1, 2021, the Chinese government issued a ten-year fishing ban for Poyang Lake. The construction of hydraulic facilities and sand mining activities have the potential to alter the transportation and dispersal of sediment within the lake, thereby impacting the sustenance of fish and the nesting grounds of migratory birds. However, research on the impact of sand mining activities on the concentration and transport of suspended sediment in Poyang Lake is limited, and the spatio-temporal distribution characteristics of long-term sand mining activities are unclear. Hence, there is an imminent requirement to devise a suspended sediment estimation model that determines the dispersal of suspended sediment in Poyang Lake, offering a benchmark for the lake's management and preservation [25].

Estimating the SSC at Poyang Lake from remote sensing images is usually done by establishing the correlations between the SSC and surface reflectance of the image. With relatively high correlations resulting from previous studies in Poyang Lake, a high correlation between SR and SSC can be inferred. Most models for estimating SSC are simple linear regressions [26], and the other regression models include logarithmic [27], exponential [28], and power functions [23]. However, due to the influence of complex environmental conditions, various water environmental parameters often exhibit complex, high-dimensional nonlinear relationships and even have the possibility of collinearity.

Machine learning is a data-based algorithm that analyzes and learns data to predict and classify unknown data [29]. The most common machine learning algorithms include support vector machine (SVM), neural network (NN), classification and regression tree (CART), and random forest (RF). Machine learning has been widely used in hydrological research, such as water quality [30], streamflow forecasting [31], rainfall-runoff simulation [32], and sediment transport in rivers [33]. AI methods were found to deliver the best performance in these studies. Machine learning is based on an adequate amount of data but collecting a large amount of measurement data in large lakes is not easy. Therefore, few studies have applied machine learning techniques to estimate the SSC at Poyang Lake by satellite imagery. The hydrological station can provide daily observed suspended sediment data with a long time series. The Landsat satellite was launched in 1972 and is still in operation with a return period of 16 days. One Landsat image is approximately 185 km 180 km in size and has an area of approximately 33000 km^2 and can be obtained free of charge. These conditions provide the possibility for machine learning to invert suspended sediment in Poyang Lake.

This study aims to address the challenge of limited remote sensing inversion samples for lake suspended sediment by innovatively integrating hydrological data with Landsat imagery reflectance from the Google Earth Engine (GEE) platform. To enhance accuracy, we compare the applicability of various machine learning methods across four satellite sensor scenarios, with the goal of establishing a highly predictive SR–SSC modeling paradigm specifically tailored for Poyang Lake. Leveraging this approach, we have successfully generated a suspended sediment map of Poyang Lake spanning from 1989 to 2021. These comprehensive maps not only shed light on the distribution of sand mining vessels within the lake and the flow patterns of disturbed suspended sediment, but also provide valuable insights and technical support for the effective

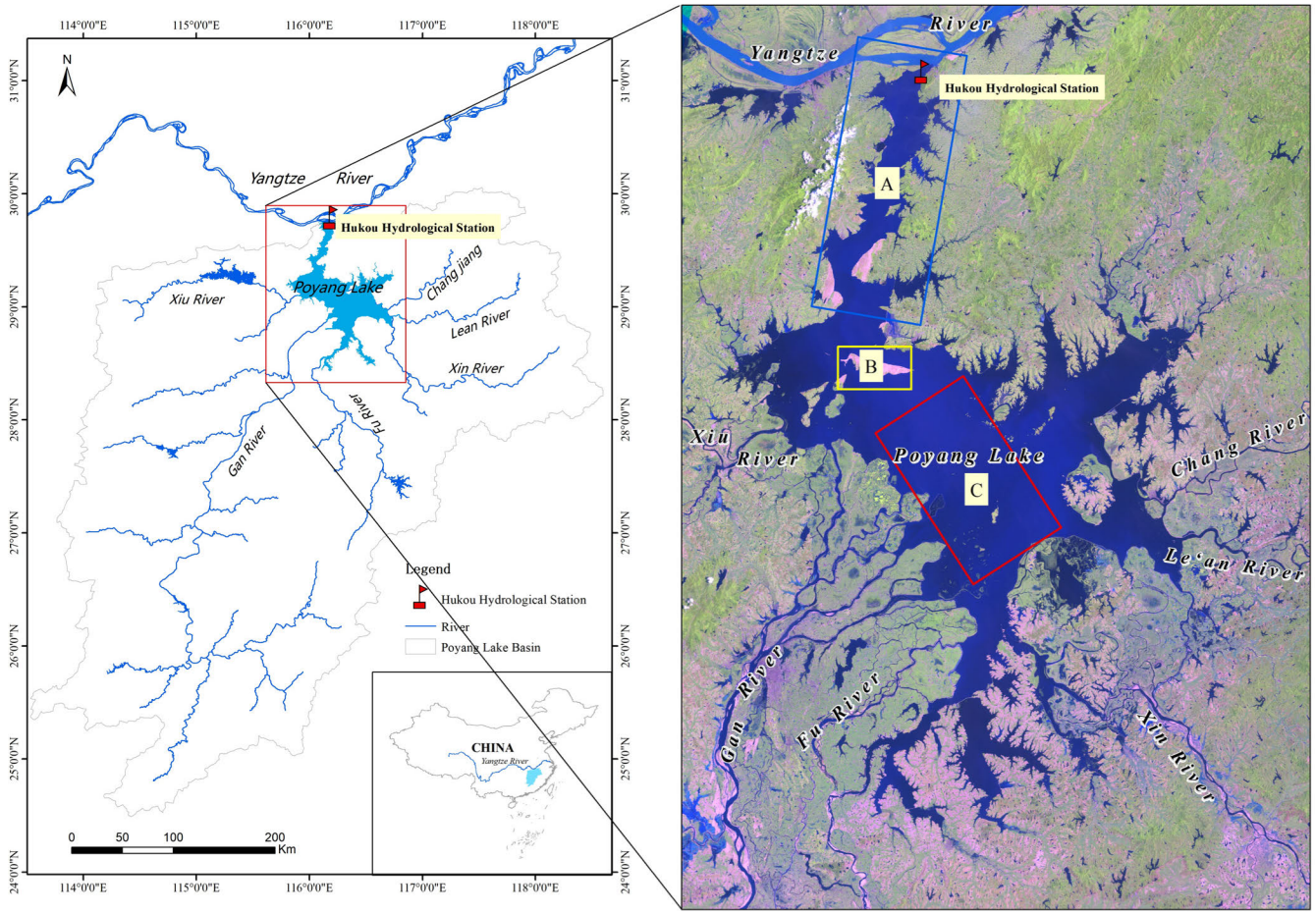

FIGURE 1. The location of the study area. The true-color Landsat 5 image from July 25, 1989. Area A means water exchange channel of Poyang Lake and Yangtze River, Area B means Songmenshan Island, Area C means main lake area of Poyang Lake.

management, conservation, and sustainable development of Poyang Lake.

II. MATERIALS AND METHODS

A. STUDY AREA AND SUSPENDED SEDIMENT DATA

Hukou hydrological station was established in 1922 located on Poyang Lake in China at 29°45′N and 116°13′E (Figure 1). A full section suspended sediment concentration test was used since the suspended load sediment discharge rate test was carried out at this station in 1952. Samples were taken at 0.2, 0.6 and 0.8 relative water depths along the vertical line and mixed vertically at 2:1:1 to calculate the average vertical suspended sediment concentration. The observed SSC data on the day of Landsat satellite transit will be screened out for the estimating of suspended sediment concentration.

As the largest freshwater lake in China, Poyang Lake has a water area of about 3,500 km^2 and a watershed of approximately 167,000 km^2. The water and sediment of Poyang Lake mainly come from the five tributary rivers Ganjiang, Fuhe, Xinjiang, Rao and Xiuhe. The annual average sediment inflow of the five rivers is 1.689×10^7 tons, accounting for more than 80% of the sediment inflow into the lake [34]. The terrain of Poyang Lake is high in the south and low in the north, and its waters flow into the Yangtze River from Hukou station. The large inflows of sand dredgers are a major driver of suspended sediment concentration in the waters of Poyang Lake [35], and sediment retention by reservoirs and improvement of water and soil conservation measures in the waters are the main reasons for the reduced sediment inflow into the lake [36]. Poyang Lake is a highly dynamic and sediment-laden lake dominated by throughput, so it is necessary to monitor the concentration of suspended sediment in the lake.

TABLE 1. Characteristics of landsat sensors.

Spectral region	Wavelength range (μm)			Resolution (m)
	Landsat 5 TM	Landsat 7 ETM	Landsat 8 OLI	
Blue	B1 0.45-0.52	B1 0.45-0.52	B2 0.45-0.52	30
Green	B2 0.52-0.60	B2 0.52-0.60	B3 0.52-0.60	30
Red	B3 0.63-0.69	B3 0.63-0.69	B4 0.63-0.68	30
NIR	B4 0.76-0.90	B4 0.76-0.90	B5 0.84-0.89	30
SWIR 1	B5 1.55-1.75	B5 1.55-1.75	B6 1.56-1.66	30

B. REMOTE SENSING DATA

Landsat imagery allows for reliable data acquisition from lakes [37]. To match the long-term time series data of

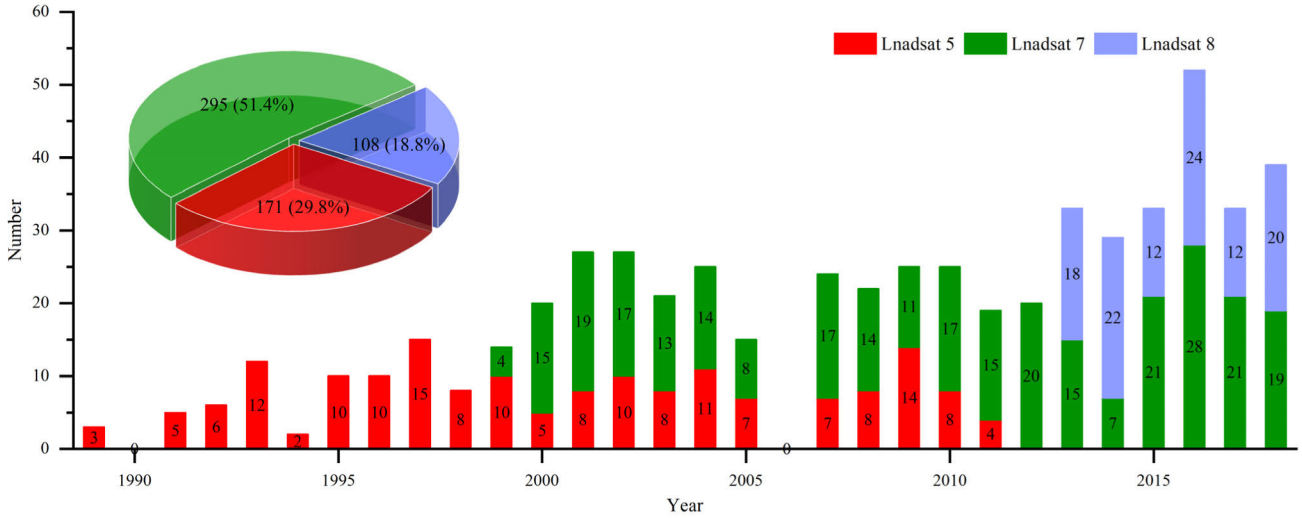

FIGURE 2. Temporal distribution of data in the Poyang Lake study area, 1989–2018.

hydrological stations with the spatial range of lakes, we chose Landsat 5, 7 and 8 satellite records. These Landsat missions have the same temporal resolution of a 16-day repeat cycle and the same spatial resolution of 30 m. The Google Earth Engine (GEE) platform was used to analyze the Landsat images. The GEE platform was developed by Carnegie Mellon University and the U.S. Geological Survey and allows users to process remote sensing images on a global scale [38]. The GEE platform hosts Landsat 5 (1989-2011), Landsat 7 (1999-2018) and Landsat 8 (2013-2018) imagery. We used the Landsat Tier 2 SR product to generate the calibration datasets. Table 1 shows the spectral bands of those satellites. The Landsat SR product also has quality assurance codes for each pixel in the ''pixel-qa'' band, which helps us eliminate cloud and rain pixels [14].

In order to maximize the availability of data, we retrieved 574 valid Landsat satellite remote sensing images from 1989 to 2009 using the coordinates of the Hukou Hydrological Station on the Google Earth Engine (GEE). These filtered images were then cross-referenced with the Suspended Sediment Concentration (SSC) database to identify matching pairs for regression modeling analysis. The 574 valid data points encompassed three distinct satellite sensors: Landsat 5 with 171 samples, Landsat 7 with 295 samples, and Landsat 8 with 107 samples. To assess the inversion effectiveness of the different satellite sensors, we categorized the data into four distinct scenarios based on sensor type and their combinations: Scenario #01 (Landsat 5TM), Scenario #02 (Landsat 7ETM), Scenario #03 (Landsat 8OLI), and Scenario #04 (Landsat 5TM, 7ETM, 8OLI). This approach enabled us to systematically evaluate the performance of each sensor and their combinations in SSC estimation.

C. LINEAR REGRESSION AND MACHINE LEARNING MODELS

To select a straightforward yet efficient approach, we opted to use only the reflectance from 5 bands of Landsat images as input parameters. Linear regression and four common machine learning models, support vector machine (SVM), back propagation neural network (BPNN), classification and regression tree (CART) and random forest (RF), were selected to predict SSC. To ensure that our regression parameter estimators exhibit robust statistical properties and enhance model accuracy, we calculated the natural logarithm of suspended sediment concentration (denoted in Ln (SSC)) and subsequently fitted it with the image surface reflectance. All models were calculated in SPSSPRO (Scientific Platform Serving for Statistics Professional) software, an online data processing and analysis platform that integrates professional statistical methods and data algorithms. Although further parameter optimization is feasible, the primary focus of this study is to offer a general comparison of various machine learning models in estimating SSC based on Landsat image surface reflectance. The observed 574 SSC data were randomly divided into two groups, with 70% used for training the model and the remaining 30% for validation.

D. QUANTITATIVE EVALUATION

We used the Pearson correlation coefficient (PCC) to describe the correlation between Landsat satellite surface reflectance and the SSC data. A larger PCC value means a higher correlation. The equation is as follows:

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

where x_i and $\bar{x}_i$ are the observed SSC value and average of the observed SSC value, y_i and $\bar{y}_i$ are the estimated SSC value and average of the estimated SSC value, and n is the total number of testing samples.

We used the coefficient of determination (R^2), the root mean square error (RMSE error index type), and the median absolute percentage error (MAPE) to evaluate the performance of the proposed methods. The equations are as

follows:

$$R^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 - \sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$RMS = \frac{\sqrt{\sum_{i=1}^{n}(y_i - x_i)^2}}{n}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{y_i - x_i}{x_i}$$

where x_i and $\bar{x}_i$ are the observed SSC value and the estimated values, respectively, y_i is the estimated SSC value, and n is the total number of testing samples. Larger R^2 and smaller RMSE and MAPE values indicate a higher overall accuracy and lower estimated error of the model.

III. RESULTS

A. THE PCC BETWEEN SSC DATA AND SR OF LANDSAT SATELLITE IMAGES

Figure 3 shows the PCC between the SR of Landsat satellite images and the SSC data in the four different scenarios. Obviously, as seen from the figure, the PCC is above 0, and there were strong significant positive correlations among the SSC data and SR data at a significance level (α) of 0.001 in all four scenarios. The PCC between the NIR band and SSC is the highest among the five bands; scenario #01 and scenario #03 are greater than 0.7, scenario #04 is 0.623 and scenario #02 is 0.509. The other bands sorted from large to small are red, blue, green and SWIR in terms of the PCC.

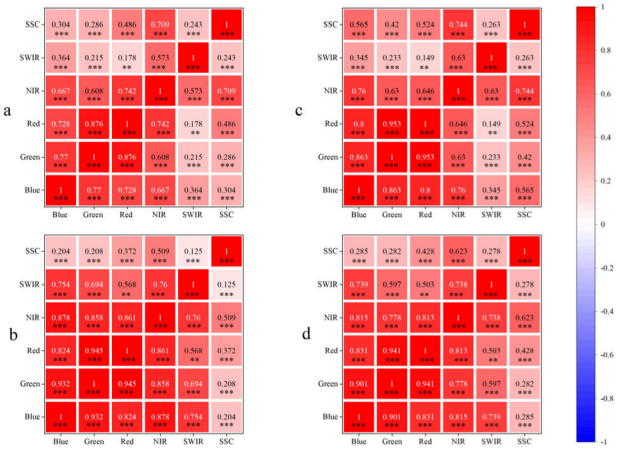

FIGURE 3. The PCC between SSC data and SR of Landsat images in four different scenarios. a represents scenario #01, b represents scenario #02, c represents scenario #03, and d represents scenario #04.

B. ACCURACY OF REGRESSION MODELS

Table 2 lists the performance statistics obtained from the training data for each scenario and the five methods. The machine learning models (RF, SVM, BPNN, CART) performed better than linear regression. Among the individual models, the CART and RF models had the highest overall fit accuracy.

LR represents linear regression, SVM represents support vector machine, BPNN represents back propagation neural network, CART represents classification and regression tree, and RF represents random forest.

TABLE 2. Statistical performances of the evaluated models.

Model name	Satellite	RMSE (g/L)	MAPE (%)	R^2
LR#01	Landsat 5	0.564	14.194	0.747
LR#02	Landsat 7	0.448	8.377	0.751
LR#03	Landsat 8	0.355	7.327	0.772
LR#04	Landsat 5, 7, 8	0.558	11.336	0.672
SVM#01	Landsat 5	0.427	9.047	0.856
SVM#02	Landsat 7	0.371	6.871	0.829
SVM#03	Landsat 8	0.197	4.140	0.930
SVM#04	Landsat 5, 7, 8	0.479	10.544	0.758
BPNN#01	Landsat 5	0.550	12.687	0.735
BPNN#02	Landsat 7	0.439	8.292	0.770
BPNN#03	Landsat 8	0.338	7.743	0.816
BPNN#04	Landsat 5, 7, 8	0.461	8.705	0.719
CART#01	Landsat 5	0.143	2.588	0.984
CART#02	Landsat 7	0.187	3.638	0.956
CART#03	Landsat 8	0.008	0.083	0.988
CART#04	Landsat 5, 7, 8	0.329	7.271	0.886
RF#01	Landsat 5	0.247	5.842	0.952
RF#02	Landsat 7	0.221	4.148	0.940
RF#03	Landsat 8	0.173	3.487	0.946
RF#04	Landsat 5, 7, 8	0.295	6.402	0.908

During scenario #01 training, the observations closely matched those modeled by CART, RF and SVM. RMSE, MAPE and R^2 from CART#01-RF#01-SVM#01 were 0.143 g/L, 2.588 and 0.984; 0.247 g/L, 5.842%, and 0.952; and 0.427 g/L, 9.047% and 0.856, respectively. The lowest R^2 (0.735) was obtained from BPNN#01, and the highest RMSE (0.564) with MAPE (14.194%) was obtained from RF#01.

According to scenario #02, the lowest RMSE (0.187 g/L) and MAPE (3.638%) with the highest R^2 (0.956) were obtained from CART. The performances of LR#02 and RF#02 were not significantly different from those of LR#01 and RF#01 during the training period. The performances positively changed for BNPP#02, while the performances of CART#02 and SVM#02 negatively changed compared to scenario #01 during the training period.

With respect to scenario #03, all evaluated methods generally produced better results than the other scenarios, implying a stronger relationship between SSC measurements and Landsat 8 data in Poyang Lake. The results of SVM#03 show higher values than those of SVM#02, with RMSE, MAPE and R^2 values equal to 0.197 g/L, 4.14%, and 0.93, respectively. CART#03 demonstrates the best performance with the highest R^2 (0.98) and lowest RMSE (0.008 g/L) and MAPE (0.083%), and it is closely followed by RF#03 (R^2 = 0.946, RMSE = 0.173 g/L, MAPE = 3.487%).

Finally, the results of scenario 4 show a performance compared to the above scenario. Unlike before, the lowest PAME and RMSE with the highest R^2 were obtained from RF instead of CART in scenario #04, and the values were 6.402%, 0.295 g/L and 0.908, respectively.

Different scenarios have inconsistent results. According to the statistical metrics, the results of the individual Landsat

satellite scenarios of CART outperformed those of RF, SVM, BPNN and LR; nevertheless, RF performed best in all scenarios during the training period.

By sensor, the lowest overall results were obtained from all the Landsat 5, 7, 8 data with an average R^2 of 0.779, RMSE of 0.444 g/L and MAPE of 9.38% between modeling techniques. Landsat 8 produced the best results with an average R^2 of 0.88, RMSE of 0.218 g/L and MAPE of 4.474%. Although Landsat 5 and Landsat 7 showed similar correlation coefficients (0.857 and 0.845, respectively), the Landsat 7 models produced much lower RMSE and MAPE values, averaging 0.33 g/L and 6.283%, compared to the Landsat 5 average RMSE and MAPE values of 0.389 g/L and 9.175%, respectively.

C. VALIDATION RESULTS OF THE MODELS

Figure 4 shows the performance of the predicted natural logarithm of SSC (estimated Ln(SSC)) and the actual natural logarithm (in situ Ln(SSC)) of 5 models of scenario#01. The results showed that there was a high correlation between the predicted values estimated by the 5 models and the measured values, with $R^2 > 0.55$, RMSE< 0.78 g/L, and MAPE $<$ 16%. Unlike the training period, the comprehensive accuracy of the RF model is the best, with the highest R^2 (0.72) and lowest RMSE (0.58 g/L) and MAPE (11.92%) during the validation period. The results for the SVM showed significant R^2 correlations but produced much higher RMSE and MAPE values than those for the RF, BPNN, CART and LR.

The validation results of the regression models for scenario #02 are displayed in Figure 5. Generally, the model validation results of Scenario #02 perform better than Scenario #01. SVM generated the highest R^2 (0.82) with the lowest RMSE (0.47 g/L) and MAPE (10.85%) values in all cases when applied to the validation data. The RF values were 0.76, 0.53 g/L, and 11.20%, respectively. CART performs poorest among the five models with the lowest R^2 (0.61) and highest RMSE (0.47 g/L) and MAPE (10.85%) values, while it performs best during the training period.

The validation results of the regression model for scenario #03 are displayed in Figure 6. The BPNN showed significant R^2 (0.92) correlations but produced a much higher RMSE (0.33 g/L) and MAPE (7.47%) values than LR ($R^2 = 0.88$, RMSE $= 0.29$ g/L and MAPE $= 6.46$%, respectively). The R^2 generated from RF was lower than that of SVM, while MAPE was higher than that of SVM and CART. Similar to scenario #02, CART still performs worst during the validation period.

The validation results of the regression model for scenario 4 are displayed in Figure 7. The RF presented the highest R^2 value of 0.67 (0.49-0.66 for other methods) and smallest MAPE value of 12.83% (12.94-15.38% for other methods) and RMSE value of 0.6 g/L in Ln SSC (0.61-0.76 g/L for other methods). The second-best overall validation result is the SVM model ($R^2 = 0.55$, RMSE $= 0.69$ g/L and MAPE $=$ 15.83%), while the worst overall accuracy is the CART model ($R^2 = 0.49$, RMSE $= 0.76$ g/L and MAPE $= 15.13$%).

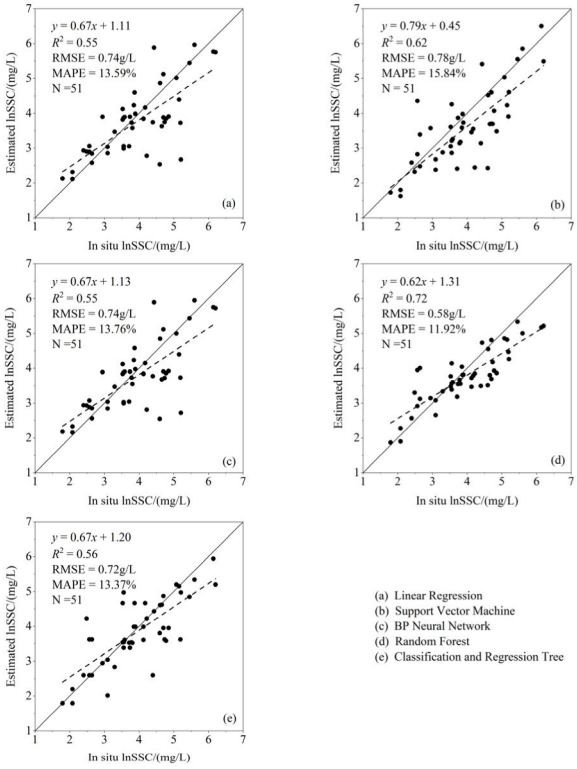

FIGURE 4. Plots of the in situ versus estimated SSC values when applied to the test dataset in scenario #01. The straight line represents the 1:1 relationship. R^2, RMSE and MAPE values represent the overall fit of the model.

Overall, CART had the best modeling accuracy, but during validation, the overall accuracy was the worst. The accuracy of the RF model is the second-best, and the highest comprehensive accuracy of model verification is observed in the first three models in the case of different scenarios. Other models sometimes have good performance in verifying accuracy, but they are not stable enough, and the modeling accuracy is lower than that of the CART and RF models. In general, the RF model accuracies between the SR of Landsat images and SSC were the best among those machine learning models.

D. SUSPENDED SEDIMENT MAPPING OF POYANG LAKE

Utilizing our developed optimal models, we mapped the suspended sediment concentration (SSC) in Poyang Lake across seven years: 1989, 2000, 2001, 2005, 2011, 2016, and 2021 (Figure 8). Data for 1989, 2001, 2005, and 2011 were acquired using the Landsat 5 satellite sensor, while 2000 data originated from the Landsat 7 satellite sensor. The 2016 and 2021 datasets were sourced from the Landsat 8 satellite sensor.

As illustrated in Figure 8, the predicted SSC ranges between 0.019 and 0.205 g/L. During 1989 and 2000, the SSC predominantly remained below 0.03 g/L, exhibiting minimal spatial variation within the lake. From 2001 onwards, a substantial SSC increase was observed in Poyang Lake compared to pre-2000 levels. Specifically, SSC levels around the mouths of the Five Rivers surpassed 0.03 g/L, peaking

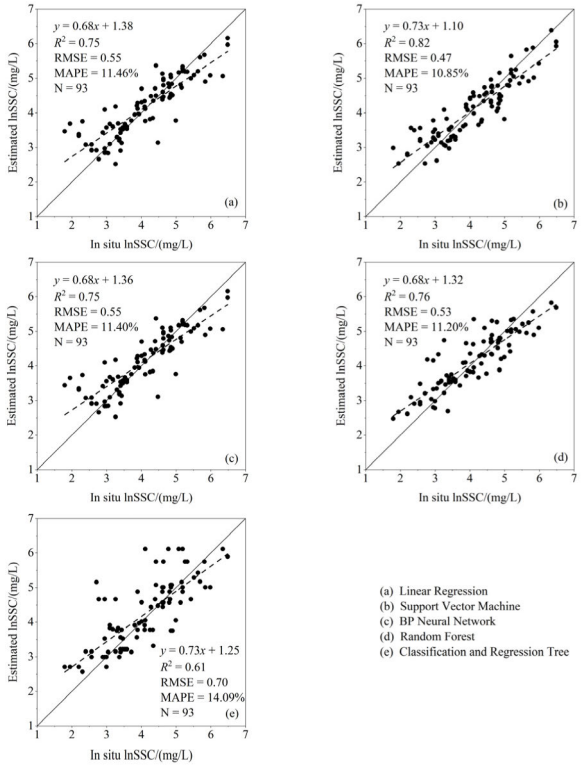

FIGURE 5. Plots of the in situ versus estimated SSC values when applied to the test dataset in scenario #02. The straight line represents the 1:1 relationship. R^2, RMSE and MAPE values represent the overall fit of the model.

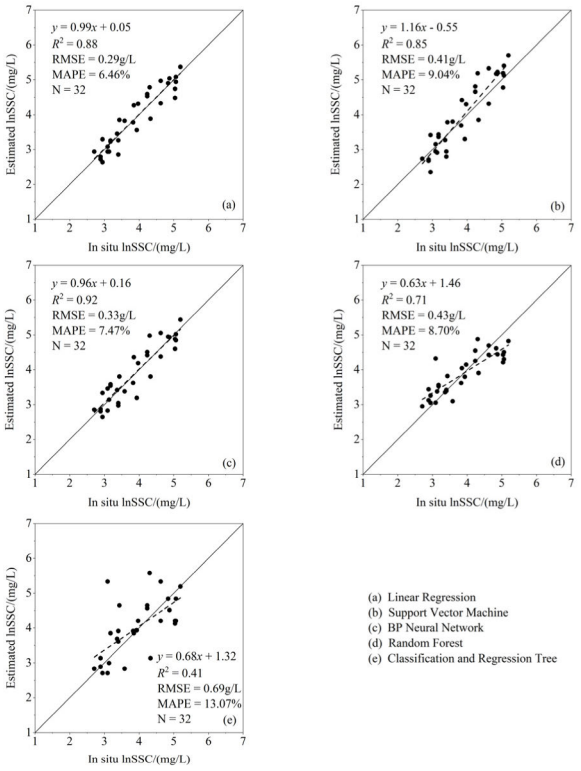

FIGURE 6. Plots of the in situ versus estimated SSC values when applied to the test dataset in scenario #03. The straight line represents the 1:1 relationship. R^2, RMSE and MAPE values represent the overall fit of the model.

at 0.2 g/L in certain areas. Result from 2001 indicates the emergence of sand mining vessel activity, mainly concentrated in the water exchange channel of Poyang Lake and Yangtze River, with suspended sediment concentrations below 0.1 g/L. By 2005, sand mining operations had shifted southward, concentrating near Songmen Mountain Island, with SSC exceeding 0.15 g/L in some regions. By 2011, sand mining vessels had expanded their operations significantly southward, extending past Songmen Mountain Island towards the center of Poyang Lake. In 2016, sand mining activities continued to spread further south, reaching the southernmost tip of Poyang Lake, with main lake area of the Poyang Lake(Area C) becoming the primary mining foci. Numerous sand mining vessels are visible in these regions through Landsat imagery, highlighting the main sand mining zones, and nearly one-fifth of Poyang Lake's water bodies exhibited an SSC exceeding 0.05 g/L. However, by 2021, the SSC in Poyang Lake had decreased significantly. Except for the northern channel connecting to the Yangtze River, where SSC ranged from 0.03 to 0.05 g/L, SSC levels in other lake areas dropped below 0.03 g/L.

IV. DISCUSSION

A. THE ADVANTAGES OF THE RANDOM FOREST MODEL

Linear regression models can simulate the relationship between response variables and explanatory variables,

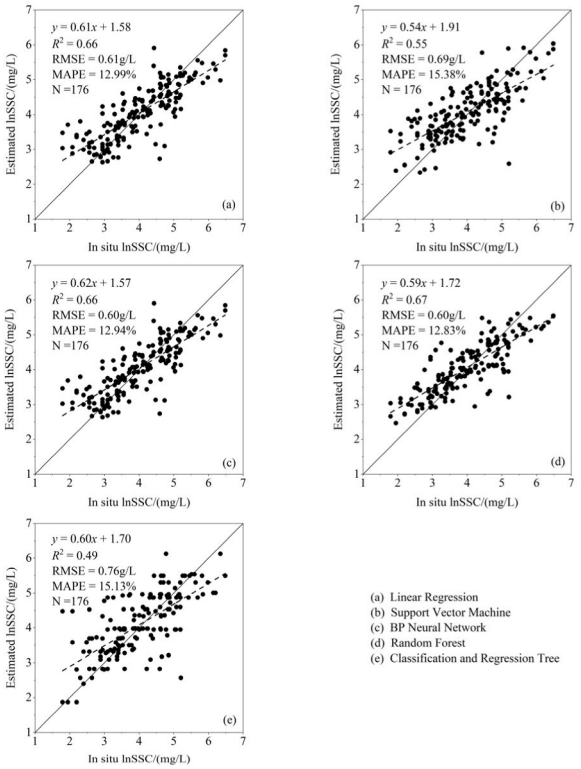

FIGURE 7. Plots of the in situ versus estimated SSC values when applied to the test dataset in scenario #04. The straight line represents the 1:1 relationship. R^2, RMSE and MAPE values represent the overall fit of the model.

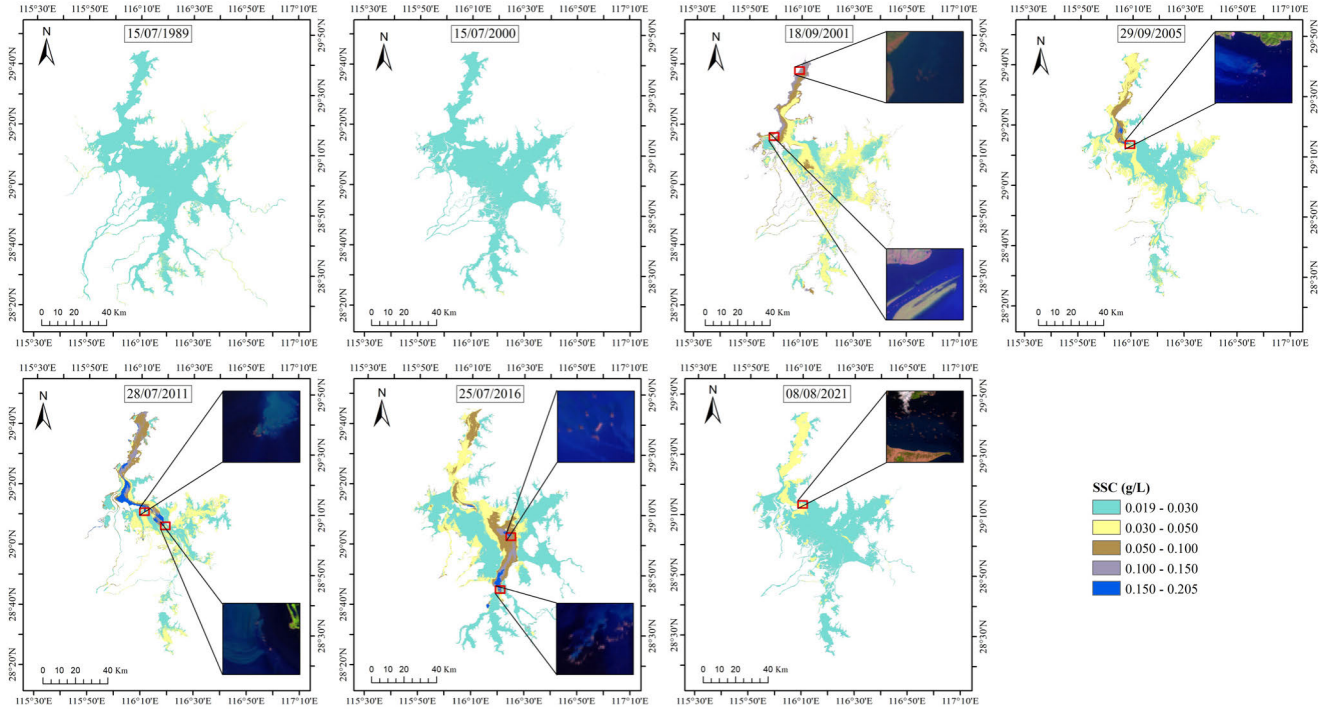

FIGURE 8. Maps of predicted SSC based on the RF model in Poyang Lake from year 1989-2021 the two cases. Blue represents lower levels of SSC, while white represents higher levels of SSC.

TABLE 3. The accuracy results of RF models after adding band ratio parameters.

Scenario		#01		#02		#03		#04	
		parameters		parameters		parameters		parameters	
		5	10	5	10	5	10	5	10
estimation	R^2	0.952	0.97	0.94	0.962	0.946	0.978	0.908	0.934
	RMSE (g/L)	0.247	0.196	0.221	0.176	0.173	0.111	0.295	0.25
	MAPE (%)	5.842	4.586	4.148	3.291	3.487	2.347	6.402	5.366
validation	R^2	0.72	0.709	0.76	0.856	0.71	0.836	0.67	0.776
	RMSE (g/L)	0.58	0.563	0.53	0.407	0.43	0.318	0.6	0.484
	MAPE (%)	11.92	11.34	11.2	8.048	8.7	5.779	12.83	9.68

achieving the prediction of variables in spatial or temporal dimensions [39]. However, due to the inherent non-stationary, dynamic and noise nature of the sediment mechanism, conventional methods such as sediment rating curves or linear regression models have largely been unsuccessful in accurately predicting suspended sediment [40], [41]. Machine learning models can address these issues, so in SSC estimation, the overall accuracy of machine learning models is higher than that of linear regression. Among the four machine learning models, the RF model has superior performance.

SVM and BPNN models are black box models, and researchers are unable to understand the actual operating mechanism within the model, nor can they intuitively determine the overall importance of the predicted variables [42]. Although RF regression is also a black box model, it provides other effective and helpful explanations, such as the importance of variables to model prediction. Compared with the other machine learning models, the RF model is easier to explain. Therefore, some researchers tend to regard the RF model as a gray box model ()()() [40], [43]. Moreover, the CART is prone to ''overfitting'' [44], [45] and ignores the correlation of attributes in the dataset. Two random parameters (k, m) are introduced into the RF model, which has good anti-noise ability and does not easily fall into overfitting. Therefore, in the remote sensing estimation of suspended sediment concentration in Poyang Lake, the RF model based on satellite-sensor division can be considered. However, the RF algorithm cannot be perfect because the model it constructs is nonparametric. There are still certain limitations in its application, and further research is needed in later attempts and explorations.

B. THE IMPACT OF THE NUMBER OF THE NUMBER OF MODEL PARAMETERS ON THE RESULTS

In addition to the single-band SR of Landsat image in SSC prediction of Poyang Lake, several band combinations

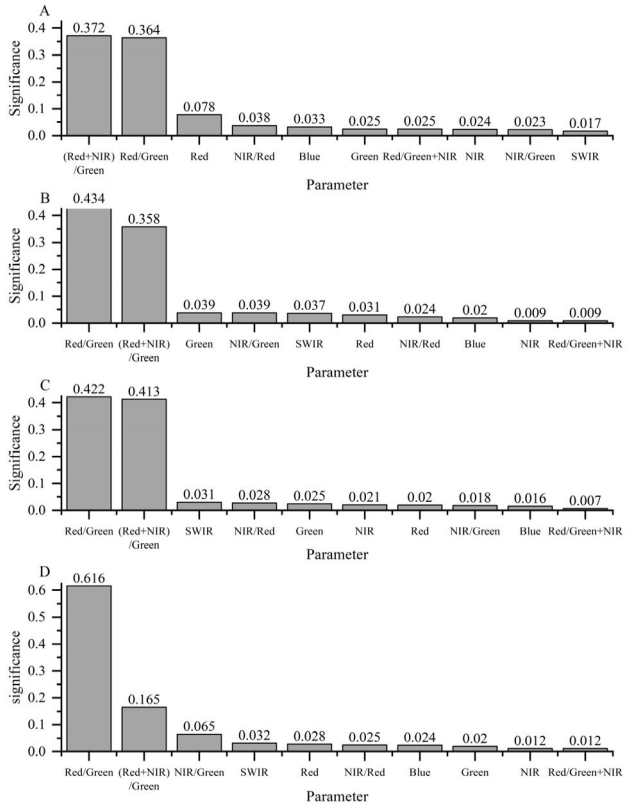

FIGURE 9. Feature ranking was performed on the importance of parameters in the four scenarios.

can also be used for SSC prediction research. For example, in band ratioing models and difference models, commonly used indices include NIR/Green [28], NIR/Red [46], Red/Green [47], (Red+NIR)/Green [48], Red/Green+NIR [47], etc. The RF model can estimate the importance of variables, so we selected several commonly used band ratios and single-band SR of Landsat images together to estimate the suspended sediment and observe the accuracy of the estimation results.

Similarly, we divided the Landsat images into four scenarios and used 70% of the data to establish the model and 30% of the data for model validation. The accuracy results of the RF model are shown in Table 3. After adding the band ratio parameters, the accuracy of the models in all four scenarios improved, with Scenario #03 having the highest improvement in accuracy, exceeding 0.02.

We further examined whether all five newly added parameters contributed to the improvement of model accuracy. A feature ranking was performed on the importance of parameters in four scenarios (Figure 9). The top two contributors of parameters in scenarios #01, #02, and #03 are (Red+NIR)/Green and Red/Green. The contribution rates of the two parameters are not significantly different, but they have a positive advantage compared to other parameters, all of which are below 10%. In scenario #04, Red/Green and (Red+NIR)/Green also rank in the top two in terms of contribution rate, but there is a significant difference in values. Red/Green is 61.6%, (Red+NIR)/Green is only 16.5%, while the remaining parameter values are still below 10%.

A change in wavelength can lead to a change in reflectivity. When sediment particles appear in a water body, the reflectance spectral curve will be affected by sediment scattering and correspondingly change. The blue and green bands absorb less sunlight, while the red band (with a wavelength greater than 600 nm) has stronger absorption ability, and the red band is sensitive to the prediction of SSC in water and can effectively identify sediment [48], [49]. Therefore, the (Red+NIR)/Green and Red/Green parameters are key parameters that can be added to the model when using Landsat satellite surface reflectance to predict SSC.

C. THE SPATIOTEMPORAL CHARACTERISTICS OF SAND DREDGERS AND SUSPENDED SEDIMENT IN POYANG LAKE

Before 2000, the suspended sediment concentration in Poyang Lake was at a very low level, with an overall concentration below 0.03g/L. Since 1980, Jiangxi has undertaken diverse artificial afforestation projects, including the ''Mountain-River-Lake'' initiative, the Sloping Land Conversion Program, and the Sino-German Afforestation Project [36]. These efforts have increased Jiangxi's forest cover to 60%, securing the province's second place ranking nationwide. Concurrently, China has implemented soil and water conservation measures in Jiangxi, specifically the ''National Key Control Project for Soil and Water Loss'' and ''Comprehensive Management of Small Watersheds.'' These conservation and afforestation endeavors have diminished sediment inflow from rivers into Poyang Lake. During this period, sand mining primarily occurred along the Yangtze River, leaving Poyang Lake undisturbed and maintaining low SSC levels.

After the ban on sand mining in China's Yangtze River in 2000, Poyang Lake emerged as a primary site for such activities. Initially, between 2001 and 2005, mining was focused in the Water Exchange Channel connecting Poyang Lake and the Yangtze River (Area A). This mining generated suspended sediment disturbances that flowed into the Yangtze River via the Exchange Channel, but had minimal effect on the main lake area of Poyang Lake (Area C) [22]. By 2011, sand dredging operations had shifted significantly southward, extending beyond Songmenshan Island (Area B) to the northern fringes of Area C. Despite regulatory measures imposed by the Jiangxi Provincial Government, illegal mining persisted due to the lucrative nature of the sand mining industry. Consequently, the junction of Areas A and B suffered the most severe impact from suspended sediment [50]. By 2016, sand dredging activities had further expanded south, reaching the southernmost tip of Poyang Lake, with Area C becoming the focal point of mining, marking the apex of these activities [27], [51]. The northeastern section of Area C was most severely impacted by suspended sediment, with even the

national nature reserves in the northern part of Poyang Lake being affected.

However, a notable decline in both mining activities and suspended sediment concentration was observed in 2021 compared to 2016. This shift occurred after the Chinese government acknowledged the gravity of the situation and introduced stricter sand mining bans in 2019, along with intensified crackdowns on illegal mining. Currently, sporadic mining mainly takes place in Area B, primarily affecting suspended sediment levels in Areas A and B, albeit at concentrations below 0.5g/L. Nevertheless, given that the western part of Area B and the northern part of Area C constitute national nature reserves of Poyang Lake, it is imperative to prohibit mining in Area C and reinforce regulations in Area B to safeguard these ecologically sensitive regions [24], [46].

V. CONCLUSION

Accurate estimation of the SSC in large lakes is considered a critical challenge in water resource planning and management. The response of the SSC in lakes is non-linear and highly dynamic, especially in shifting lakes. Using free Landsat data in conjunction with long term hydrological station data, SR–SSC models were developed for Poyang Lake. SSR-SC modeling results showed that RF model comprehensively outperformed BPNN, CART, SVM and LR when evaluated by R^2, RMSE and MAPE. The RF models for the four scenarios produced R^2 values in excess of 0.9 and showed little or no overfitting. The accuracy of the model does not undergo a substantial improvement with the increase in the parameters of the parameters, the effective parameters being (red + NIR)/green and red/green parameters. The RF model proved to be an effective model for predicting a wide range of SSCs from surface reflectance data.

Using the developed RF model, SSC maps for Poyang Lake from 1989 to 2021 were generated for the Landsat 5, 7, and 8 sensors. Unlike the traditional in situ modelling methods, these SSC maps provide a continuous spatial coverage for monitoring sediment in Poyang Lake, effectively identifying sand mining activities and their impact areas within the lake. Based on the RF model, long-term monitoring of suspended sediments in Poyang Lake is achievable. The results indicate that sand mining activities in Poyang Lake have gradually shifted southward from Area A to Area C since 2000, peaking in 2016. Sand dredgers have a significant impact on suspended sediment in lake, especially in their immediate vicinity where the SSC can exceed 0.15 g/L. By 2021, sand mining activities maintained at a relatively low level due to strengthened regulation by the Chinese government. Due to the influence of water currents, disturbances caused by suspended sediments from sand mining activities, especially in Area C, may affect the national nature reserves in the north of Poyang Lake. In future studies, we plan to combine the SR-SSC model with high-resolution satellite sensors like Sentinel-2 to further enhance the accuracy of the SSC estimation model.

REFERENCES

[1] P. W. Moran, L. H. Nowell, N. E. Kemble, B. J. Mahler, I. R. Waite, and P. C. Van Metre, "Influence of sediment chemistry and sediment toxicity on macroinvertebrate communities across 99 wadable streams of the midwestern USA," *Sci. Total Environ.*, vols. 599–600, pp. 1469–1478, Dec. 2017, doi: 10.1016/j.scitotenv.2017.05.035.

[2] B. Granger, B. Laval, S. Vagle, E. L. Petticrew, P. N. Owens, and S. A. Baldwin, "Initial distribution and interannual decrease of suspended sediment in a two-basin lake following a massive mine tailings spill: Quesnel Lake, BC, Canada," *Water Resour. Res.*, vol. 58, no. 5, pp. 1–17, May 2022, doi: 10.1029/2021wr030574.

[3] F. Jourdin, P. R. Renosh, A. A. Charantonis, N. Guillou, S. Thiria, F. Badran, and T. Garlan, "An observing system simulation experiment (OSSE) in deriving suspended sediment concentrations in the ocean from MTG/FCI satellite sensor," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5423–5433, Jul. 2021, doi: 10.1109/TGRS.2020.3011742.

[4] L. N. Patterson, D. M. Paulson, V. J. Colucciello, and J. A. Covi, "Sediment from lake with missing egg bank is toxic to hatchlings of model zooplankton: A reason to consider obligate dormancy in toxicological assessment," *Aquatic Toxicol.*, vol. 236, Jul. 2021, Art. no. 105862, doi: 10.1016/j.aquatox.2021.105862.

[5] L. Sokoletsky, S. Fang, X. Yang, and X. Wei, "Evaluation of empirical and semianalytical spectral reflectance models for surface suspended sediment concentration in the highly variable estuarine and coastal waters of east China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 11, pp. 5182–5192, Nov. 2016, doi: 10.1109/JSTARS.2016.2582909.

[6] N. P. Majozi, M. S. Salama, S. Bernard, D. M. Harper, and M. G. Habte, "Remote sensing of euphotic depth in shallow tropical inland waters of Lake Naivasha using MERIS data," *Remote Sens. Environ.*, vol. 148, pp. 178–189, May 2014, doi: 10.1016/j.rse.2014.03.025.

[7] J. Liu, J. Liu, X. He, D. Pan, Y. Bai, F. Zhu, T. Chen, and Y. Wang, "Diurnal dynamics and seasonal variations of total suspended particulate matter in highly turbid Hangzhou bay waters based on the geostationary ocean color imager," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2170–2180, Jul. 2018, doi: 10.1109/JSTARS.2018.2830335.

[8] K. N. Markert, C. M. Schmidt, R. E. Griffin, A. I. Flores, A. Poortinga, D. S. Saah, R. E. Muench, N. E. Clinton, F. Chishtie, K. Kityuttachai, P. Someth, E. R. Anderson, A. Aekakkararungroj, and D. J. Ganz, "Historical and operational monitoring of surface sediments in the lower Mekong basin using Landsat and Google Earth Engine cloud computing," *Remote Sens.*, vol. 10, no. 6, p. 909, Jun. 2018, doi: 10.3390/rs10060909.

[9] L. Cai, D. Tang, and C. Li, "An investigation of spatial variation of suspended sediment concentration induced by a bay bridge based on landsat TM and OLI data," *Adv. Space Res.*, vol. 56, no. 2, pp. 293–303, Jul. 2015, doi: 10.1016/j.asr.2015.04.015.

[10] J. Hui, L. Yao, and Z. Wen-bin, "Retrieval and analysis of total suspended solid concentration by MODIS Terra 500 m imagery during flood period in Poyang Lake, China," in *Proc. Int. Conf. Comput. Distrib. Control Intell. Environ. Monitor.*, Feb. 2011, pp. 985–988, doi: 10.1109/CDCIEM.2011.480.

[11] J. Seegert, D. Markova, O. Kolditz, P. Krebs, and D. Borchardt, "IWAS II: Integrated water resources management under different hydrological, climatic and socio-economic conditions," *Environ. Earth Sci.*, vol. 72, no. 12, pp. 4673–4675, Dec. 2014, doi: 10.1007/s12665-014-3879-0.

[12] R. Mao, J. Song, P. Yan, Z. Ouyang, R. Wu, S. Liu, and X. Guo, "Horizontal and vertical distribution of microplastics in the Wuliangsuhai Lake sediment, northern China," *Sci. Total Environ.*, vol. 754, Feb. 2021, Art. no. 142426, doi: 10.1016/j.scitotenv.2020.142426.

[13] M. K. Stroud, G. H. Allen, M. Simard, D. Jensen, B. Gorr, and D. Selva, "Optimizing satellite mission requirements to measure total suspended solids in rivers," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–9, 2024, Art. no. 4200409, doi: 10.1109/TGRS.2023.3337641.

[14] E. N. Dethier, C. E. Renshaw, and F. J. Magilligan, "Toward improved accuracy of remote sensing approaches for quantifying suspended sediment: Implications for suspended-sediment monitoring," *J. Geophys. Res., Earth Surf.*, vol. 125, no. 7, pp. 1–21, Jul. 2020, doi: 10.1029/2019jf005033.

[15] T. Sok, C. Oeurng, V. Kaing, S. Sauvage, G. M. Kondolf, and J. M. Sánchez-Pérez, "Assessment of suspended sediment load variability in the tonle sap and lower Mekong Rivers, Cambodia," *CATENA*, vol. 202, Jul. 2021, Art. no. 105291, doi: 10.1016/j.catena.2021.105291.

[16] J.-E. Min, J.-H. Ryu, S. Lee, and S. Son, "Monitoring of suspended sediment variation using Landsat and MODIS in the Saemangeum coastal area of Korea," *Mar. Pollut. Bull.*, vol. 64, no. 2, pp. 382–390, Feb. 2012, doi: 10.1016/j.marpolbul.2011.10.025.

[17] J. Chen and W. Quan, "Using Landsat/TM imagery to estimate nitrogen and phosphorus concentration in Taihu Lake, China," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 5, no. 1, pp. 273–280, Feb. 2012, doi: 10.1109/JSTARS.2011.2174339.

[18] L. Yang, K. Jia, S. Liang, X. Wei, Y. Yao, and X. Zhang, "A robust algorithm for estimating surface fractional vegetation cover from Landsat data," *Remote Sens.*, vol. 9, no. 8, p. 857, Aug. 2017, doi: 10.3390/rs9080857.

[19] H. Jiang, Y. Liu, and J. Lu, "A new algorithm for monitoring backflow from river to lake (BRL) using satellite images: A case of Poyang Lake, China," *Water*, vol. 13, no. 9, p. 1166, Apr. 2021, doi: 10.3390/w13091166.

[20] L. Zou, B. Hu, S. Qi, Q. Zhang, and P. Ning, "Spatiotemporal variation of Siberian crane habitats and the response to water level in Poyang Lake Wetland, China," *Remote Sens.*, vol. 13, no. 1, p. 140, Jan. 2021, doi: 10.3390/rs13010140.

[21] Q. Li, G. Lai, Y. Liu, A. T. Devlin, S. Zhan, and S. Wang, "Corrigendum to 'assessing the impact of the proposed Poyang Lake hydraulic project on the Yangtze finless porpoise and its calves' [Ecol. Indic. 129 (2021) 107873]," *Ecol. Indicators*, vol. 132, Jan. 2021, Art. no. 108324, doi: 10.1016/j.ecolind.2021.108324.

[22] F. Jiang, S. Qi, F. Liao, X. Zhang, D. Wang, J. Zhu, and M. Xiong, "Hydrological and sediment effects from sand mining in Poyang Lake during 2001–2010," *Acta Geographica Sinica*, vol. 70, no. 5, pp. 837–845, 2001, doi: 10.11821/dlxb201505014.

[23] L. Cui, Y. Qiu, T. Fei, Y. Liu, and G. Wu, "Using remotely sensed suspended sediment concentration variation to improve management of Poyang Lake, China," *Lake Reservoir Manage.*, vol. 29, no. 1, pp. 47–60, Mar. 2013, doi: 10.1080/10402381.2013.768733.

[24] S. Yao, X. Li, C. Liu, D. Yuan, L. Zhu, X. Ma, J. Yu, G. Wang, and W. Kuang, "Quantitative assessment of impact of the proposed Poyang lake hydraulic project (China) on the habitat suitability of migratory birds," *Water*, vol. 11, no. 8, p. 1639, Aug. 2019, doi: 10.3390/w11081639.

[25] S. Qi, M. Xiong, F. Liao, G. Liu, and H. Zheng, "Impacts of human activities on sediments budget in Poyang Lake," *Scientia Geographica Sinica*, vol. 36, no. 6, pp. 888–894, 2016, doi: 10.13249/j.cnki.sgs.2016.06.011.

[26] I.-Y. Yeo, M. Lang, and E. Vermote, "Improved understanding of suspended sediment transport process using multi-temporal Landsat data: A case study from the Old Woman Creek Estuary (Ohio)," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 2, pp. 636–647, Feb. 2014, doi: 10.1109/JSTARS.2013.2265191.

[27] P. Zhang, X. Chen, J. Lu, and W. Zhang, "Assimilation of remote sensing observations into a sediment transport model of China's largest freshwater lake: Spatial and temporal effects," *Environ. Sci. Pollut. Res.*, vol. 22, no. 23, pp. 18779–18792, Dec. 2015, doi: 10.1007/s11356-015-4958-9.

[28] R. Kuang, Y. Zhao, W. Luo, G. Zhang, and Y. Chen, "Study on inversion model of suspended sediment concentration based on optical classification of water body in Poyang lake," *J. China Hydrol.*, vol. 37, no. 6, pp. 23–28, 2017.

[29] D. Kreuzberger, N. Kühl, and S. Hirschl, "Machine learning operations (MLOps): Overview, definition, and architecture," *IEEE Access*, vol. 11, pp. 31866–31879, 2023, doi: 10.1109/ACCESS.2023.3262138.

[30] M. Zeng, J. Peng, L. Jiang, and J. Feng, "Temporal and spatial distribution of suspended sediment concentration in lakes based on satellite remote sensing and Internet of Things," *IEEE Access*, vol. 9, pp. 87849–87856, 2021, doi: 10.1109/ACCESS.2021.3089367.

[31] J.-W. Kim and Y. A. Pachepsky, "Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation," *J. Hydrol.*, vol. 394, nos. 3–4, pp. 305–314, Nov. 2010, doi: 10.1016/j.jhydrol.2010.09.005.

[32] M. Tahmoures, A. R. Moghadamnia, and M. Naghiloo, "Modeling of streamflow- suspended sediment load relationship by adaptive neuro-fuzzy and artificial neural network approaches (case study: Dalaki River, Iran)," *Desert*, vol. 20, no. 2, pp. 177–195, Jun. 2015, doi: 10.22059/JDESERT.2015.56481.

[33] Q. Xing, M. Lou, C. Chen, and P. Shi, "Using in situ and satellite hyperspectral data to estimate the surface suspended sediments concentrations in the Pearl river estuary," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 731–738, Apr. 2013, doi: 10.1109/JSTARS.2013.2238659.

[34] C. Gu, X. Mu, G. Peng, G. Zhao, and Y. Qiang, "Rainfall erosivity and sediment load over the Poyang Lake basin under variable climate and human activities since the 1960s," *Theor. Appl. Climatol.*, vol. 136, no. 1, pp. 1–16, 2019, doi: 10.1007/s00704-018-2460-2.

[35] L. Haijun, C. Xiaoling, L. Jianzhong, Z. Peng, Q. Hengda, and C. Liqiong, "Numerical simulation of suspended sediment concentration in lake Poyang during flood season considering dredging activities," *J. Lake Sci.*, vol. 28, no. 2, pp. 421–431, 2016.

[36] M. Mo, X. Yang, S. Xiao, and A. Tu, "Analysis on runoff and sediment change characteristics and influence factors of Poyang Lake five rivers basin," *Res. Soil Water Conservation*, vol. 24, no. 5, pp. 197–203, 2017, doi: 10.13869/j.cnki.rswc.2017.05.031.

[37] G. H. Allen and T. M. Pavelsky, "Global extent of rivers and streams," *Science*, vol. 361, no. 6402, pp. 585–588, Aug. 2018, doi: 10.1126/science.aat0636.

[38] H. Tamiminia, B. Salehi, M. Mahdianpari, L. Quackenbush, S. Adeli, and B. Brisco, "Google Earth Engine for geo-big data applications: A meta-analysis and systematic review," *ISPRS J. Photogramm. Remote Sens.*, vol. 164, pp. 152–170, Jun. 2020, doi: 10.1016/j.isprsjprs.2020.04.001.

[39] L. Zhang, L. Wang, X. Zhang, S. Liu, P. Sun, and T. Wang, "The basic principle of random forest and its applications in ecology—A case study of Pinus Yunnanensis," *Acta Ecologicas Sinica*, vol. 34, no. 3, pp. 650–659, 2014.

[40] X. Fang, Z. Wen, J. Chen, S. Wu, Y. Huang, and M. Ma, "Remote sensing estimation of suspended sediment concentration based on random forest regression model," *J. Remote Sens.*, vol. 23, no. 4, pp. 756–772, 2019.

[41] D. Jiang, B. Matsushita, N. Pahlevan, D. Gurlin, M. K. Lehmann, C. G. Fichot, J. Schalles, H. Loisel, C. Binding, Y. Zhang, K. Alikas, K. Kangro, M. Uusõue, M. Ondrusek, S. Greb, W. J. Moses, S. Lohrenz, and D. O'Donnell, "Remotely estimating total suspended solids concentration in clear to extremely turbid waters using a novel semi-analytical method," *Remote Sens. Environ.*, vol. 258, Jun. 2021, Art. no. 112386, doi: 10.1016/j.rse.2021.112386.

[42] Q. Wang, Y. Shen, and Y. Chen, "Rule extraction from support vector machines," *J. Nat. Univ. Defense Technol.*, vol. 10, no. 2, pp. 106–110, 2006.

[43] A. M. Prasad, L. R. Iverson, and A. Liaw, "Newer classification and regression tree techniques: Bagging and random forests for ecological prediction," *Ecosystems*, vol. 9, no. 2, pp. 181–199, Mar. 2006, doi: 10.1007/s10021-005-0054-1.

[44] M. A. Ezzaouini, G. Mahé, I. Kacimi, A. E. Bilali, A. Zerouali, and A. Nafii, "Predicting daily suspended sediment load using machine learning and NARX hydro-climatic inputs in semi-arid environment," *Water*, vol. 14, no. 6, p. 862, Mar. 2022, doi: 10.3390/w14060862.

[45] B. Choubin, H. Darabi, O. Rahmati, F. Sajedi-Hosseini, and B. Kløve, "River suspended sediment modelling using the CART model: A comparative study of machine learning techniques," *Sci. Total Environ.*, vol. 615, pp. 272–281, Feb. 2018, doi: 10.1016/j.scitotenv.2017.09.293.

[46] Z. Huang, L. Lu, G. Jiao, J. Jiang, and Q. Ye, "Analysis of the correlations between environmental factors and rare cranes in the Poyang Lake region of China," *J. Great Lakes Res.*, vol. 44, no. 1, pp. 140–148, Feb. 2018, doi: 10.1016/j.jglr.2017.11.003.

[47] K. Peterson, V. Sagan, P. Sidike, A. Cox, and M. Martinez, "Suspended sediment concentration estimation from Landsat imagery along the lower Missouri and middle Mississippi Rivers using an extreme learning machine," *Remote Sens.*, vol. 10, no. 10, p. 1503, Sep. 2018, doi: 10.3390/rs10101503.

[48] Q. V. Pham, N. T. T. Ha, N. Pahlevan, L. T. Oanh, T. B. Nguyen, and N. T. Nguyen, "Using Landsat-8 images for quantifying suspended sediment concentration in Red River (Northern Vietnam)," *Remote Sens.*, vol. 10, no. 11, p. 1841, Nov. 2018, doi: 10.3390/rs10111841.

[49] L. Mertes, M. Smith, and J. Adams, "Estimating suspended sediment concentrations in surface waters of the Amazon River wetlands from Landsat images," *Remote Sens. Environ.*, vol. 43, no. 3, pp. 281–301, Mar. 1993, doi: 10.1016/0034-4257(93)90071-5.

[50] S. Zheng, H. Cheng, M. Tang, W. Xu, E. Liu, S. Gao, J. Best, Y. Jiang, and Q. Zhou, "Sand mining impact on Poyang Lake: A case study based on high-resolution bathymetry and sub-bottom data," *J. Oceanol. Limnol.*, vol. 4, p. 40, Apr. 2022.

[51] Z. Yongchao, L. Jianzhong, C. Liqiong, and C. Xiaoling, "Spatial–temporal dynamic monitoring of sand dredging activities based on GF-1 WFV in lake Poyang during 2013–2020," *J. Lake Sci.*, vol. 34, no. 6, pp. 2144–2155, 2022, doi: 10.18307/2022.0626.

KAITAO LIAO was born in Linchuan, Jiangxi, China, in 1990. He received the B.S. and M.S. degrees in geographic information system (GIS) from Jiangxi Normal University, Jiangxi, in 2015, where he is currently pursuing the Ph.D. degree in geography.

From 2015 to 2021, he was a Research Assistant with Jiangxi Provincial Key Laboratory of Soil Erosion and Prevention. Since 2021, he has been a Senior Engineer with Jiangxi Academy of Water Science and Engineering. His research interests include remote sensing applications, GIS, artificial intelligence, machine learning, soil, and water conservation.

YUEJUN SONG was born in Songjia, Shandong, China, in 1973. He received the B.S. degree in geographic sciences from Yantai Normal University, in 2004, the M.S. degree in physical geography from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, in 2008, and the Ph.D. degree in soil and water conservation and desertification control from Fujian Agriculture and Forestry University, Fujian, China, in 2017.

From 2008 to 2020, he was a Research Assistant with Jiangxi Provincial Key Laboratory of Soil Erosion and Prevention. Since 2021, he has been a Senior Engineer and a Professor with Jiangxi Academy of Water Science and Engineering. He is the author of three books, more than 30 articles, and more than 20 inventions. His research interests include soil and water conservation, application of remote sensing technology, watershed ecology, benggang erosion monitoring, and big data.

XIAOFEI NIE, photograph and biography not available at the time of publication.

LINGJIA LIU, photograph and biography not available at the time of publication.

SHUHUA QI was born in Wuyuan, Jiangxi, China, in 1973. He received the B.S. degree in mechanical engineering from Northwest Agriculture and Forestry University, Shanxi, China, in 1997, the M.S. degree in soil water science from China Agricultural University, Beijing, China, in 2001, and the Ph.D. degree in remote sensing of the environment from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, in 2004.

From 2006 to 2007, he was a Postdoctoral Fellow with the Department of Natural Resources and Environment, University of Michigan, USA. He is the author of 60 articles. His research interests include the applications of remote sensing technology to ecology, hydrology, resources and the environment, and the ecological environment of Poyang Lake.