



# Estimation of Suspended Sediment Concentration by M5 Model Tree Based on Hydrological and Moderate Resolution Imaging Spectroradiometer (MODIS) Data

Javad Zahiri<sup>1</sup>  · Zeynab Mollaei<sup>1</sup> · Mohammad Reza Ansari<sup>2</sup>

Received: 30 November 2019 / Accepted: 25 May 2020 /

Published online: 12 August 2020

© Springer Nature B.V. 2020

## Abstract

The Estimation of suspended sediment concentration (SSC) is an important factor in river engineering, which is used as an indicator of land-use change, water quality studies, and all projects related to constructions in rivers. In this research, the M5 model tree and the Moderate Resolution Imaging Spectroradiometer (MODIS) data were utilized to estimate the SSC at Ahvaz station on the Karun River. In this study, 135 cloud-free images of the MODIS sensor on the Terra satellite were taken for days corresponding to field SSC data, during the years 2000 to 2015. Input parameters of the model tree in this study were flow discharge, derived from hydrological data, and red (R), near-infrared (NIR) bands, and NIR/R ratio extracted from MODIS imagery. The results of statistical analysis illustrate that the M5 model outperforms the sediment rating curve (SRC) method, which is the most common method of estimating suspended sediment load. The Nash-Sutcliffe efficiency index for the M5 model tree of 0.58 was achieved, which was much better than that of the SRC method (0.26). At high fluxes, the efficiency of the SRC method significantly reduced, while the model tree provides acceptable results. The global sensitivity analysis on the M5 model pointed out that 93% of output variance was established by the main effects of input parameters, and less than 7% belong to the interaction effects. 73% and 12% of output variance specified by the main effects of flow discharge and NIR/R ratio, respectively.

**Keywords** Suspended sediment concentration · M5 model tree · MODIS · Global sensitivity

---

✉ Javad Zahiri  
j.zahiri@asnrukh.ac.ir

<sup>1</sup> Department of Water Engineering, Agricultural Sciences and Natural Resources University of Khuzestan, Khuzestan, Iran

<sup>2</sup> Department of Soil Science, Agricultural Sciences and Natural Resources University of Khuzestan, Khuzestan, Iran

## 1 Introduction

Suspended sediment concentration (SSC) is an essential factor in the quality of the river and estuarine water. It has been considered by river engineering, ecology, and environmental researchers (Li and Li 2016), due to its significant role in the evolution of the geomorphology of canals, flood plains, and biogeochemical cycles (Park and Latrubesse 2014). The high concentration of suspended sediment affects the transmission of aquatic organisms and the fertility of the phytoplankton, and the whole aquatic system (Min et al. 2012). To study the environmental changes of sediment such as river morphological and water quality changes, and negative impacts on aquatic ecosystems, it is necessary to carefully monitor the sediment transport in rivers (Wang and Lu 2010). Most of the sediment transport formulas assume that the sediment transport rate can be determined by individual flow parameters. In most cases, the limited data collected in laboratory conditions confirmed the accuracy of these formulas. Due to the lack of generalization of assumptions, the compatibility of such equations for other situations is often uncertain. This limitation has led to a substantial difference in the results of various sediment transport equations, often with each other and with the measured values (Yang 1996).

Remote sensing technology consists of analyzing and describing the measurements obtained from the amount of electromagnetic radiation emitted from a target or reflected by a viewer or a device without contacting the target from a suitable viewing point or recording (Mather 2009). The use of satellite imagery for evaluating water quality shows the ability of synoptic and inexpensive estimations by satellites. As a result, satellite remote sensing can be a quick alternative and an economic approach to assess the SSC in the oceans, seas, and rivers (Moridnejad et al. 2015). To estimate the sediment concentration by the output emission from the surface of the water and reflected the sensor, the relationship of the radiation transmission between the water optical properties and the radiation estimated by the sensor needs to consider theoretically. The transmission relation is modeled by the statistical relationships between radiometric data and field measurements after eliminating atmospheric effects (Kazemzadeh et al. 2013). Regression analysis is a common empirical approach that attempts to develop the best correlation between field measurements and remote sensing data (Mollaei 2018). Various studies have shown that there is a significant relationship between the radiances from satellite imagery and SSC. Jiang et al. (2009) used the correlation between the infrared wavelengths of the MODIS images and the sediment concentration of Taihu Lake in China, developing a logarithmic regression to estimate the SSC. Moridnejad et al. (2015) predicted the SSC using artificial neural networks (ANN), MODIS images, and field hydrologic data on the southern shorelines of the Caspian Sea. The results showed that using ANN and images of the MODIS sensor to monitor sediment concentration in the Caspian Sea coast is acceptable. Cai et al. (2015), Using TM Landsat OLI sensor data and on-site measurement, investigated spatial variations in the SSC due to Hangzhou Bay Bridge in the eastern coastal China Sea. They found that the sides of the bridge have a significant difference in the SSC rate. Moreover, the results of this study indicate the suitability of satellite images in estimating suspended sediment load. Robert et al. (2016) estimated the turbidity and SSC in the Bagre Dam reservoir in Burkina-Faso by applying MODIS, MOD09Q1, and MYD09Q1 products. They found that the near-infrared to red band ratio was the most appropriate combination for assessing the SSC and turbidity for both spectroscopy and the MODIS sensor radiances.

In this study, MODIS and hydrological field data along with the M5 model tree used to estimate SSC of Karun River. The M5 model is one of the reliable data-driven models, which

present comprehensible formulas that describe the structure of the phenomenon more clearly (Etemad-Shahidi and Bonakdar 2009). This model was previously used successfully in different approaches, including flood modeling (Solomatine and Xue 2004), longitudinal dispersion coefficient prediction (Etemad-Shahidi and Taghipour 2012), modeling daily dissolved oxygen concentration (Heddami and Kisi 2018), and transverse mixing coefficient prediction (Zahiri and Nezaratian 2020). Based on the behavior of sediment concentration and sediment transport, the M5 model was applied in this research to divide the problem space and provide a regression equation for each subdomain. The novelty of this study lies in building a tree model based on the MODIS and the hydrological data and presenting regression equations for SSC prediction. Moreover, a global sensitivity analysis was applied to determine the main and interaction effects of input variables on sediment concentration. Finally, the prediction maps were used to show the behavior of sediment concentration under the influence of input variables for measured and estimated concentrations.

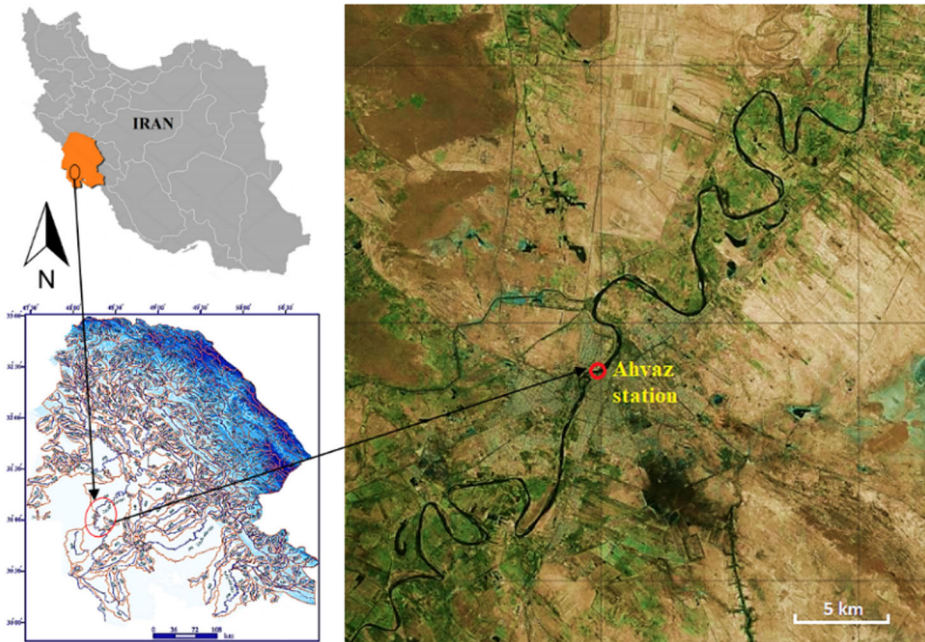
## 2 Data and Methodology

### 2.1 Study Area

Karun River, with an annual yield of about 22 billion cubic meters and an average instantaneous discharge of 736 m<sup>3</sup>/s, was considered a study area. The area of this river basin is 66352 km<sup>2</sup>, its average height is 1537 m, and its average slope is 0.3%. The width of this river in the mountainous parts is between 25 and 40 m, and in the plain of Ahwaz upstream is 250–400 m. The Karun River was formed after the confluence of the Dez, Shoteyt, and Gargar Rivers in the area called Band-e Qir above Molasani city. The hydrometric station of Ahwaz is located in the coordinates (48° 41' 41") longitude and (31° 20' 16") northern latitudes on the Karun River, depicted in Fig. 1.

### 2.2 Data Processing

High sensitivity radiometric data are available at nominal spatial resolutions of 250 m (bands 1–2), 500 m (bands 3–7), and 1000 m (bands 8–36). The spectral range of MODIS band 1 (red) is 620–670 nm and band 2 (near-infrared) has a range of 841–876 nm (Miller and McKee 2004). In this study, 135 cloud-free images of the MODIS sensor on the Terra satellite were taken for days corresponding to field SSC data, during the years 2000 to 2015. The georeferenced and radiometrically corrected MODIS/Terra images for all 36 spectral bands were obtained from the <http://earthexplorer.usgs.gov/> website. The top-of-atmosphere (TOA) radiance is derived in several narrow spectral bands spanning the visible and near-infrared sections of the spectrum by a sensor installed on a satellite. The radiance departing from the water, derived from the TOA, includes information on the water constituents. (Pinkerton et al. 2003). Based on the river width and the spatial precision of the MODIS, the pixel with the lowest TOA radiance was selected because water radiance is considerably lower than land radiance due to the powerful absorption of water. The TOA radiance was obtained straight from the MODIS images (Wang and Lu 2010). The example of the estimated TOA radiance of Ahwaz hydrometric-station showed in Fig. 2.

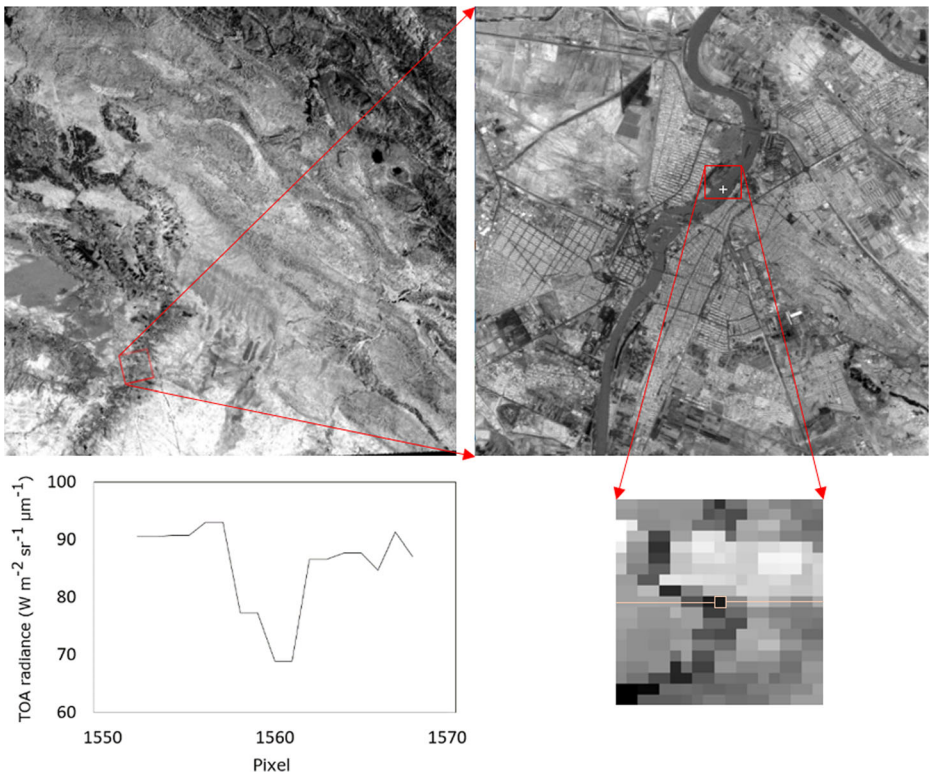


**Fig. 1** Ahvaz hydrometric station on Karun River (Google Earth)

The spectral bands of red (R), near-infrared (NIR), and NIR/R band ratio were calculated to estimate the SSC. Band ratios are usually used to illustrate differences that are not indicated by singular spectral bands (Rangzan and Fattahi-Moghadam 2013). Satellite images were first reviewed using ENVI 5.1 image analysis software, and afterward, all pixels within the study area with cloud conditions and sun glint effects were eliminated from the dataset. Surface reflectance products, using the absolute atmospheric correction model, predict the surface spectral reflectance for each band and correct the effects of atmospheric gases, aerosols, and thin cirrus clouds. These images also have radiometric and geometric corrections, and therefore do not require preprocessing. Images were subset to a geographic region bounded by the latitude and longitude to limit the area of interest. Then cloud-free pixel values corresponding to the location of the sampling station were extracted to be evaluated with the field data. River width, spatial resolution, and station geographic coordinates were considered to choose suitable pixels. Moreover, the suspended sediment concentration data reviewed using SPSS software and outlier data were excluded from the samples. The Shapiro–Wilk statistical test was employed to consider the normality of the data. After removing cloud images and other radiometric errors and outlier data from SSC and flow discharge, 110 images were left for Ahvaz station.

### 2.3 Hydrological Data

Flow discharge and SSC of Ahvaz hydrometric station, provided by Khuzestan Water and Power Authority, were used as hydrological data in this study. These data include 110 samples of flow discharge and sediment concentration within the wet and dry seasons on Ahvaz station during the years 2000 to 2015, corresponding to MODIS images. Among all the SSC data, 106



**Fig. 2** Pixel selection and radiance extraction at Ahvaz hydrometric station using MODIS Terra 250 m imagery

samples varied between 20 mg/l to 500 mg/l, and just four samples have the SSC values more than 500 mg/l. The maximum values of the SSC and flow discharges were 933 mg/l and 2167 m<sup>3</sup>/s, respectively, which generally occur during flood events in Karun River. The statistics of the parameters applied in this study are shown in Table 1.

There are some uncertainties in the amount of sediment concentration, flow discharge, and bands of the MODIS sensor. Poor measuring methods and hardware errors are some sources of uncertainty in sediment concentration and flow discharge measurements.

**Table 1** Statistics of parameters used in this study

Parameter	Range (Min Max)	Average	Standard deviation
$Q_w$ (m <sup>3</sup> /s)	98.1–2167	402.03	341.76
$R$	0.08–0.27	0.17	0.03
$NIR$	0.03–0.32	0.17	0.06
$NIR/R$	0.38–1.71	0.97	0.24
$SSC$ (mg/l)	20–933.33	122.86	150.01

### 2.4 M5 Model Tree

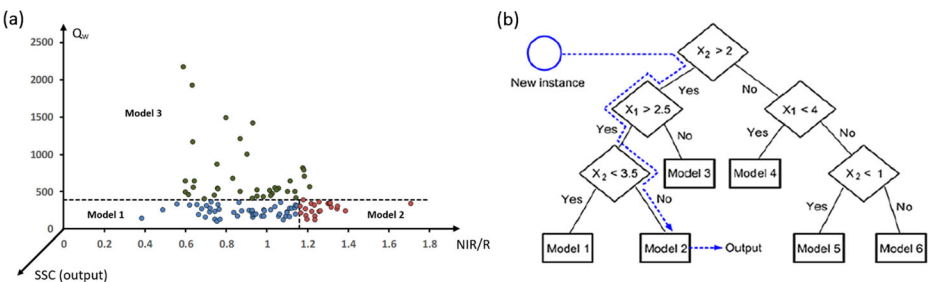
The M5 algorithm divided the problem space into subdomains and proposed a multivariate linear model for each subdomain (Quinlan 1992). The decision tree structure is like a tree, composed of roots, branches, nodes, and leaves. The roots are as the first nodes, set above, and a series of branches and nodes close in leaves. In the M5 model, each parent node splits into two branches. The branches consist of a number range that branches out of the parent node and reaches the child node (Emamifar et al. 2014). Three phases applied by the M5 model are the building, pruning, and smoothing. Standard deviation reduction (SDR) index applied by the M5 to split the problem space.

$$SDR = sd(T) \sum \left| \frac{T_i}{T} \right| sd(T_i) \tag{1}$$

in which  $T$  is a series of data points, before dividing,  $T_i$  is a vector of data that yield from dividing the space and falls within subspace based on the selected dividing factor, and  $sd$  is a standard deviation. The branching process in each node is repeated to reach the end node (leaf), where the standard deviation reaches zero (Etemad-Shahidi and Taghipour 2012). In Fig. 3, splitting the input space by the M5 algorithm and extracting the knowledge from the construction are demonstrated.

In this study, the test-and-train procedure was performed to develop the tree algorithm based on the hydrological and MODIS data. According to the test-and-train technique, 88 data sets were applied for training, and the rest of the data records were used for verifying the M5 algorithm. In this research, the sediment rating curve (Campbell and Bauder 1940) was used to determine the efficiency of the model tree in estimating sediment concentration. The coefficients of determination ( $R^2$ ), discrepancy ratio ( $DR$ ), root mean squared error ( $RMSE$ ), and Nash-Sutcliffe Efficiency ( $NSE$ ), were used to determine the efficiency of the M5 algorithm, and compare it with the results of sediment rating curve (SRC) method. These parameters are defined as:

$$R^2 = \frac{[\sum (SSC_{me} - \overline{SSC}_{me})(SSC_{es} - \overline{SSC}_{est})]^2}{\sum (SSC_{me} - \overline{SSC}_{me})^2 \sum (SSC_{es} - \overline{SSC}_{est})^2} \tag{2}$$



**Fig. 3** Splitting of the input domain by the model tree in this study, and prediction for new instance by M5 algorithm (Jung et al. 2010)

$$DR = \log \frac{SSC_{es}}{SSC_{me}} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (DR_i)^2} \tag{4}$$

$$NSE = 1 - \frac{\sum (SSC_{me} - SSC_{es})^2}{\sum (SSC_{me} - \overline{SSC_{me}})^2} \tag{5}$$

in which  $N$  is the number of observations,  $SSC_{me}$  is the measured SSC, and  $SSC_{es}$  is the estimated SSC. For  $DR$  equal to zero, the results from the field measurements and model estimation match perfectly. An overestimation occurred for  $DR$  higher than zero; otherwise, underestimation happened. Percentage of  $DR$  amounts with a range of -0.3 and 0.3 considered as the accuracy index in this study (Seo and Cheong 1998).  $NSE$  indicates the relative magnitude of the residual variance compared to the variance of the observed data. The performance of the model is good if  $NSE > 0.7$ , and the efficiency is an acceptable if  $0.4 < NSE \leq 0.7$ ; otherwise, the efficiency of the model is an unacceptable (Wu et al. 2017).

### 2.5 Global Sensitivity Analysis

In this study, the Monte-Carlo simulation method was applied to estimate the first order and total-effect indices, according to Saltelli et al. (2008). In the Monte-Carlo method, the main effect of factor  $X_i$  on output is indicated by the first-order sensitivity index, which ranged between zero and one. Entire participation of variable  $X_i$  to the output variation considered as the total effect indices that includes the first order and all higher-order effects through interactions (Saltelli et al. 2008). In this study, sensitivity indices were estimated for four variables, containing:  $Q_w$ ,  $R$ ,  $NIR$  and  $NIR/R$ . Hence, two ( $N$ , 4) matrices (A, B) of random components were produced, which  $N$  is defined as a base sample and determined to be 10000. Another matrix was established by the matrix B with the  $i$ th column of the matrix A for each input variable. The first order and total-effect indices were predicted by Eqs. (6) and (7), respectively.

$$S_i = \frac{V[E(Y|X_i)]}{V(Y)} = \frac{y_A \bullet y_{C_i} - f_0^2}{y_A \bullet y_A - f_0^2} = \frac{(1/N) \sum_{j=1}^N y_A^{(j)} \bullet y_{C_i}^{(j)} - f_0^2}{(1/N) \sum_{j=1}^N (y_A^{(j)})^2 - f_0^2} \tag{6}$$

$$S_{T_i} = 1 - \frac{V[E(Y|X_{\sim i})]}{V(Y)} = 1 - \frac{y_B \bullet y_{C_i} - f_0^2}{y_A \bullet y_A - f_0^2} = 1 - \frac{(1/N) \sum_{j=1}^N y_B^{(j)} \bullet y_{C_i}^{(j)} - f_0^2}{(1/N) \sum_{j=1}^N (y_A^{(j)})^2 - f_0^2} \tag{7}$$

where  $V[E(Y|X_i)]$  is a conditional variance, defined as the first-order effect of  $X_i$  on output ( $Y$ ).  $V[E(Y|X_{\sim i})]$  is a conditional variance regarded to all the variables except one,  $X_{\sim i}$ .  $V(Y)$  is an

unconditional variance and  $f_0^2 = \left( \frac{1}{N} \sum_{j=1}^N y_A^{(j)} \right)^2$ .  $S_i$  is the first-order sensitivity index of  $X_i$  on  $Y$  and  $S_{T_i}$  is the total-effect indices of  $X_i$ . The  $y_A, y_B,$  and  $y_{C_i}$  are vectors of model output.

### 3 Results and Discussion

#### 3.1 M5 and SRC Results

After preprocessing satellite images, two types of spectral indexes, including single-bands reflectance and the band ratio, were extracted from red and infrared wavelengths. The reflectance of red and near-infrared bands and bands ratio of MODIS images and the flow discharge introduced as input variables and the SSC of the river was considered as an output parameter to the M5 model tree. According to previous studies in the literature, the SSC formulas have exponential forms, and since the M5 model only provides linear regression equations and due to the abnormality of the data distribution based on the Shapiro–Wilk statistical test, the analysis was performed on the natural log of the data. The regression equations, provided by the M5 model, were illustrated in Eqs. (8) to (10) after being transformed from the logarithmic scale. According to the structure of the M5 tree, flow discharge and  $NIR/R$ ratio were used to divide the problem space of the M5 algorithm, having the main effects in estimating the SSC. For lower discharges ( $Q_w \leq 392 \text{ m}^3/\text{s}$ ),  $Q_w, R,$  and  $NIR/R$ ratio were used in regression equations (Eqs. (8) and (9)). The power of  $Q_w$  for the two mentioned equations is 0.08, which is much less than Eq. (10), and was obtained for higher discharges ( $Q_w > 392 \text{ m}^3/\text{s}$ ). The effect of the near-infrared band on the SSC was increased for high discharges, in agreement with the results of Gordon and Morel (1983), who state that spectral sensitivity range moves toward the longer wavelength with increasing the SSC. The M5 equations showed that the effect of MODIS spectral bands has more impact on the SSC values related to lower discharges compared to higher ones.

$$\text{If } Q_w \leq 392 \frac{\text{m}^3}{\text{s}} \text{ and } \left( \frac{NIR}{R} \right) \leq 1.16, \text{ then } SSC = 84.82 R^{0.37} \left( \frac{NIR}{R} \right)^{-0.45} Q_w^{0.08} \quad (8)$$

$$\text{If } Q_w \leq 392 \frac{\text{m}^3}{\text{s}} \text{ and } \left( \frac{NIR}{R} \right) > 1.16, \text{ then } SSC = 78.41 R^{0.43} \left( \frac{NIR}{R} \right)^{-0.53} Q_w^{0.08} \quad (9)$$

$$\text{If } Q_w > 392 \frac{\text{m}^3}{\text{s}}, \text{ then } SSC = 8.87 R^{1.84} NIR^{-1.33} \left( \frac{NIR}{R} \right)^{-0.52} Q_w^{0.54} \quad (10)$$

In the above equations, the  $SSC$  is in  $\text{mg/l}$ ,  $Q_w$  is the flow discharge in  $\text{m}^3/\text{s}$ ,  $R$  is the red band, and  $NIR$  is the near-infrared band of the MODIS. The obtained equations were then used on the test data to determine the performance of the M5 model on the  $SSC$  estimation.

The SRC method is one of the most common procedures for estimating the suspended sediment load, which is the establishment of a relationship between the sediment and flow discharge data. The general form of the SRC equation is as follows:

$$Q_s = aQ_w^b \quad (11)$$



where  $Q_s$  is the suspended sediment discharge in ton/day,  $a$  and  $b$  are the constant factors of the equation. The values of  $a$  and  $b$  are determined by the linear regression between the sediment and the flow discharge logarithms. The training and testing of the SRC method were performed with the same data used for the M5 model. The sediment load was first estimated by Eq. (12) based on the SRC method, and then by using the Eq. (13), converted to the SSC to compare the results of two the M5 and the SRC methods.

$$Q_s = 0.19Q_w^{1.67} \tag{12}$$

$$SSC = c \frac{Q_s}{Q_w} \tag{13}$$

where  $c = 11.574$ , represents the unit conversion factor.

The M5 model divides the input space into three sub-domains due to flow discharge and bands ratio, and provides a regression equation for each sub-domain, while the SRC model provides only one equation for the input space. Figure 4 demonstrates variations in the flow discharge and observed and predicted SSC with the M5 and the SRC methods. Compared to the SRC method, the SSC estimation by the M5 model tree has higher efficiency, especially for the high values of the SSC. In addition, the SRC method lower estimates the SSC in more cases compared to the M5 model, especially for high concentration, which could be a weakness in river engineering and water resources management. Despite the acceptable efficiency of the M5 model in estimating the SSC, based on Fig. 4, this model underestimated the high SSC values with relatively low flow rates. The suspended sediment usually has a straight correlation with the flow discharge. According to the training procedure with limited available data, a straight correlation between the flow discharge and the sediment concentration was established by the tree model. This correlation caused the tree model to have an acceptable efficiency in most cases with a direct relationship between the flow rate and the sediment concentration. However, for high concentrations that have occurred in relatively low flow rates, the M5 model tree underestimated the SSC. With a

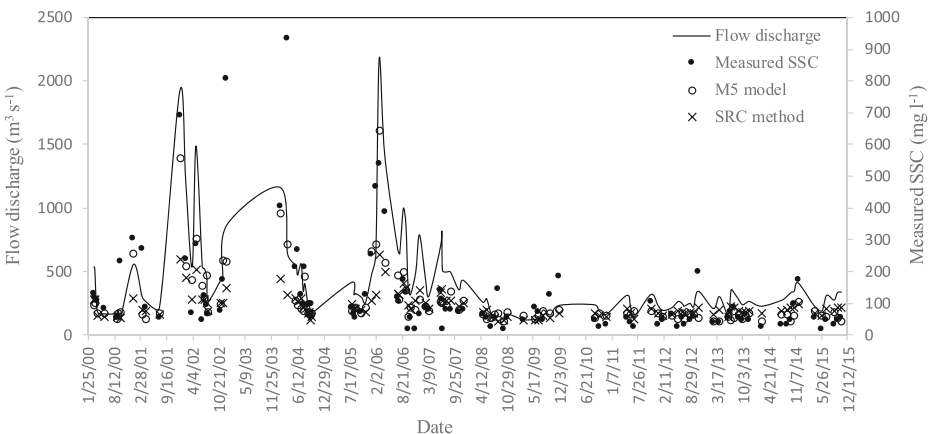


Fig. 4 Comparison between measured and predicted SSC by the M5 and the SRC methods

large amount of training data, the efficiency of the tree model can be improved to some extent in these cases, because the model will be able to provide a separate formula for such data.

### 3.2 Statistical Analysis

The results of the statistical analysis of the SSC estimation using the M5 model tree and the SRC method are presented in Table 2. According to the statistical analysis, the  $R^2$  coefficient for the M5 model is closer to one compared to the SRC method. Moreover, the  $RMSE$  in the M5 model is 37% and 20% less than the SRC method for the training and test stages, respectively. The variance of the measured SSC, the M5, and the SRC method is estimated to be  $2.27E + 04$ ,  $1.18E + 04$ , and  $4.19E + 03$ , respectively. Compared to the SRC method, the variance of the M5 model is closer to the variance of the measured SSC. The variance index, along with the  $RMSE$ , shows the superiority of the M5 model over the SRC method in SSC estimation. The  $NSE$  of the M5 model was estimated about 0.58, indicates the appropriate accuracy of the M5 model. However, the  $NSE$  of the SRC method has a value of about 0.26 that illustrates the low efficiency of this method.

Further statistical analysis, such as the discrepancy ratio ( $DR$ ), accuracy index, standard deviation ( $\sigma_{DR}$ ), maximum  $DR$  ( $\overline{DR_{max}}$ ) and  $DR$  skewness ( $SK_{DR}$ ) were also used to determine the performance of the two methods (Table 3).

According to Table 3, the M5 model is more accurate (78%) compared to the SRC method with an accuracy index of 69%. As mentioned above,  $DR$  values greater than zero represent a percentage of the computational values that estimate the SSC more than the measured values. The SRC method underestimated the SSC in about 39% of samples, and 13% of all samples was estimated with  $DR$  (the estimated SSC is less than half the measured values). In the M5 model, only 7% of all samples have a  $DR$ , which is due to the importance of the SSC, and considering the safety factor for the river engineering projects, the M5 model tree offers more acceptable results than the SRC method.  $DR$  distribution of the M5 model varies between  $-0.70$  and  $0.75$ , which indicates that the M5 model is almost non-skewed ( $SK_{DR} = -0.03$ ) towards positive and negative values.  $DR$  of the SRC method varies between  $-0.6$  and  $0.9$ , which indicates that the SRC method is skewed to negative numbers ( $SK_{DR} = -0.37$ ).  $DR$  standard deviation indicates the scattering of  $DR$  values around the mean value of  $DR$ . The closer this value to zero, the less data scattering. The M5 model has a lower  $\sigma_{DR}$  and  $\overline{DR_{max}}$  compared to the SRC method based on Table 3.

**Table 2** Statistical indices of the M5 and the SRC methods for estimating the SSC

Model	Train			Test			All data		
	NSE	$R^2$	RMSE	NSE	$R^2$	RMSE	NSE	$R^2$	RMSE
SRC	0.23	0.45	0.40	0.35	0.42	0.20	0.26	0.44	0.30
M5	0.57	0.58	0.25	0.58	0.79	0.02	0.58	0.64	0.14

**Table 3** Comparing the performance of the M5 model and the SRC method

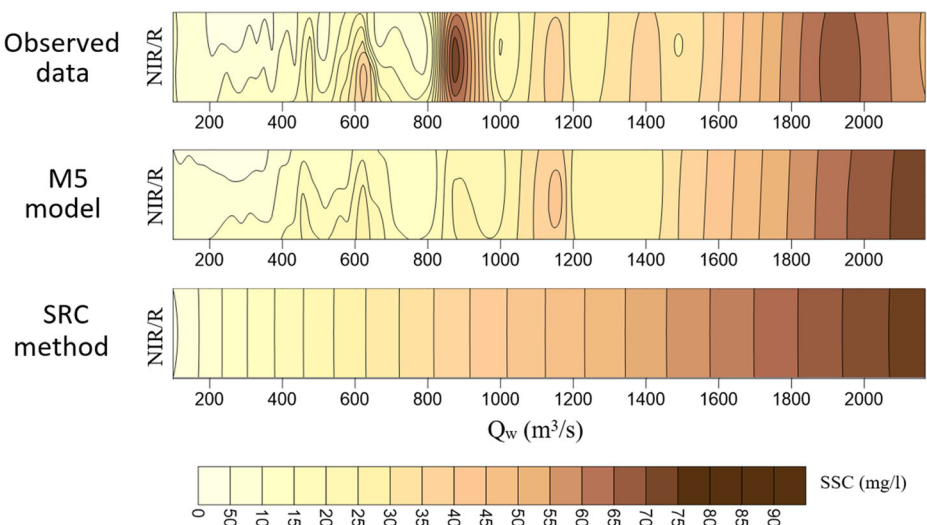
Model	<i>DR</i>	- 0.3	0	<i>DR</i> > 0.3	Accuracy	$\sigma_{DR}$	$\overline{DR_{max}}$	<i>SK</i> <sub><i>DR</i></sub>
SRC	13	26	43	18	69	0.30	0.716	- 0.37
M5	7	28	50	15	78	0.24	0.656	- 0.03

### 3.3 Prediction Map

Contour maps of the measured SSC values and the results of the M5 model and the SRC method are presented in Fig. 5. Flow discharge and *NIR/R* were used on the two-dimensional coordinate system on the contour maps of the SSC. To decrease the number of variables, *NIR/R*, was applied as a representative of *R* and *NIR* bands. The contour lines indicate the similarities and differences between the results of the M5 and the SRC with measured data. According to Fig. 5, the SSC is affected by the flow discharge and band ratio for the discharges below 1000 m<sup>3</sup>/s. For the flow discharges higher than 1000 m<sup>3</sup>/s, the effect of the band ratio decreased, and the SSC is just affected by the flow discharge. Although the performance of the M5 model being poor for the interval range 700–1000 m<sup>3</sup>/s, the model efficiency for other intervals is acceptable. The SRC method only used flow discharge as input parameter, so the SSC values in this method remain invariant by changing the bands' ratio for constant discharge values.

### 3.4 Global Sensitivity Analysis

The global sensitivity analysis results, as the first-order and the total-effect indices, are stated in Table 4. The sum of the first-order sensitivity indices was calculated at 0.93, indicating the high main effects on the SSC estimation by the M5 model tree, where only 7% of the output variance is influenced by the interaction effects. The flow discharge has the highest first order and total-effect indices, and 73% of output variance is influenced by the main effect of this



**Fig. 5** Contour maps of The SSC of the observed data, the M5 model, and the SRC method

**Table 4** The first-order and the total-effect indices of different variables

Sensitivity indices	Parameter				$\sum S$
	$Q_w$	$R$	$NIR$	$NIR/R$	
$S_i$	0.73	0.05	0.03	0.12	0.93
$S_T$	0.98	0.22	0.21	0.27	1.68

variable.  $NIR/R$  ratio after the flow discharge has the highest values of the first order and total-effect indices. The main effect of the  $NIR/R$  ratio estimated to be 12% of the output variance.

## 4 Conclusions

In this study, the MODIS and the hydrological data, in combination with the M5 model tree in 15 years period, were applied to evaluate the SSC of Karun River in Ahvaz station. The flow discharge, red and near-infrared bands, and the band ratio were applied to construct the M5 model tree. The flow discharge and the band ratio have the highest effect on the tree structure of the M5 model. Three regression equations were developed by the tree model to estimate the SSC. In these equations, in addition to the flow discharge, the MODIS bands were also used as input parameters, consistent with previous studies (Robert et al. 2016; Jiang et al. 2009; Moridnejad et al. 2015), which indicated that there is a strong correlation between the water reflectance and the SSC. For flood conditions, the effect of flow discharge was increased compared to low discharges, based on the M5 equations. Comparing the results of the tree model and the SRC method showed that the traditional method, in most cases, estimates the SSC less than the measured values, which is one of the main weaknesses of this method.  $NSE$  was estimated 0.58 and 0.26 for the M5, and the SRC methods, respectively, showed the superiority of the M5 model. The results of global sensitivity on the M5 model illustrated that 20% of output variance was influenced by the main effects of the MODIS parameters, where the  $NIR/R$  ratio had a noticeable effect on the SSC estimation compared to  $R$  and  $NIR$  bands. Prediction maps were used to investigate the SSC variation, due to the changes in flow discharge, and band ratio. The counter maps showed that, for flow discharge lower than 1000 m<sup>3</sup>/s, the SSC significantly affects the band ratio behavior.

**Acknowledgements** The Deputy of Research and Technology of Agricultural Sciences and Natural Resources University of Khuzestan acknowledged for providing financial support for this study (Project Number 2/411/170).

## Compliance with Ethical Standards

**Conflict of Interest** The authors declare that they have no conflict of interest.

## References

- Cai L, Tang D, Li C (2015) An investigation of spatial variation of suspended sediment concentration induced by a bay bridge based on Landsat TM and OLI data. *Adv Space Res* 56(2):293–303. <https://doi.org/10.1016/j.asr.2015.04.015>
- Campbell FB, Bauder H (1940) A rating-curve method for determining silt-discharge of streams. *Trans Am Geophys Union* 21:603–607

- Emamifar S, Rahimi-Khoob A, Noroozi AA (2014) Evaluation of M5 model and artificial neural network for estimating daily mean air temperature based on the data of MODIS surface temperature. *Iran J Water Soil Res* 4:423–433 (In Persian)
- Etemad-Shahidi A, Bonakdar L (2009) Design of rubble-mound breakwaters using M5' machine learning method. *Appl Ocean Res* 31:197–201
- Etemad-Shahidi A, Taghipour M (2012) Predicting Longitudinal Dispersion Coefficient in Natural Streams Using M5 Model Tree. *J Hydraul Eng* 138(6):542–554. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0000550](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000550)
- Gordon HR, Morel AY (1983) Remote assessment of ocean color for interpretation of satellite visible imagery: A review. Lecture notes on coastal and estuarine studies. Springer, New York
- Heddam S, Kisi O (2018) Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *J Hydrol* 559:499–509
- Jiang X, Tang J, Zhang M, Ma R, Ding J (2009) Application of MODIS data in monitoring suspended sediment of Taihu Lake, China. *Chin J Oceanol Limnol* 27(3):614. <https://doi.org/10.1007/s00343-009-9160-9>
- Jung N-C, Popescu N, Kelderman P, Solomatine DP, Price RK (2010) Application of model trees and other machine learning techniques for algal growth prediction in Yongdam reservoir, Republic of Korea. *J Hydroinformatics* 12(3):262–274. <https://doi.org/10.2166/hydro.2009.004>
- Kazemzadeh M, Ayoubzadeh S, Moridnejad A (2013) Estimation of suspended sediment concentration in surface water with high concentrations using remote sensing techniques. 6th National Congress and Exhibition of Environmental Engineering, Tehran, Iran (In Persian)
- Li Y, Li X (2016) Remote sensing observations and numerical studies of a super typhoon-induced suspended sediment concentration variation in the East China Sea. *Ocean Model* 104:187–202. <https://doi.org/10.1016/j.ocemod.2016.06.010>
- Mather PM (2009) Computer processing of remotely-sensed images: An introduction. 3rd edn. Wiley, New York
- Miller RL, McKee BA (2004) Using MODIS Terra 250 m imagery to map concentrations of total suspended matter in coastal waters. *Remote Sensing of Environment* 93(1–2):259–266
- Min JE, Ryu JH, Lee S, Son S (2012) Monitoring of suspended sediment variation using Landsat and MODIS in the Saemangeum coastal area of Korea. *Mar Pollut Bull* 64(2):382–390. <https://doi.org/10.1016/j.marpolbul.2011.10.025>
- Mollae S (2018) Estimation of Phytoplankton Chlorophyll-a Concentration in the Western Basin of Lake Erie Using Sentinel-2 and Sentinel-3 Data. Master's thesis, University of Waterloo. <https://uwspace.uwaterloo.ca/handle/10012/13456>
- Moridnejad A, Abdollahi H, Alavipanah SK, Samani JMV, Moridnejad O, Karimi N (2015) Applying artificial neural networks to estimate suspended sediment concentrations along the southern coast of the Caspian Sea using MODIS images. *Arab J Geosci* 8(2):891–901. <https://doi.org/10.1007/s12517-013-1171-3>
- Park E, Latrubesse EM (2014) Modeling suspended sediment distribution patterns of the Amazon River using MODIS data. *Remote Sens Environ* 147:232–242. <https://doi.org/10.1016/j.rse.2014.03.013>
- Pinkerton MH, Lavender SJ, Aiken J (2003) Validation of Sea WiFS ocean color satellite data using a moored databuoy. *J Geophys Res Oceans* 108:3133
- Quinlan JR (1992) Learning with continuous classes. Proceedings australian joint conference on artificial intelligence. World Scientific, Singapore, pp 343–348
- Rangzan K, Fattahi-Moghadam M (2013) Assessment of water quality the Karun River in Ahvaz area by ground data, spectrometer fieldspec 3 and Hyperion hyperspectral products. *J Adv Appl Geol* 4:91–108 (In Persian)
- Robert E, Grippa M, Kergoat L, Pinet S, Gal L, Cochonneau G, Martinez J-M (2016) Monitoring water turbidity and surface suspended sediment concentration of the Bagre Reservoir (Burkina Faso) using MODIS and field reflectance data. *Int J Appl Earth Obs Geoinf* 52:243–251. <https://doi.org/10.1016/j.jag.2016.06.016>
- Saltelli A et al (2008) Global sensitivity analysis: the primer. Wiley, Hoboken
- Seo IW, Cheong TS (1998) Predicting Longitudinal Dispersion Coefficient in Natural Streams. *J Hydraul Eng* 124(1):25–32. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1998\)124:1\(25\)](https://doi.org/10.1061/(ASCE)0733-9429(1998)124:1(25))
- Solomatine DP, Xue Y (2004) M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China. *J Hydrol Eng* 9:491–501
- Wang JJ, Lu XX (2010) Estimation of suspended sediment concentrations using Terra MODIS: An example from the Lower Yangtze River, China. *Sci Total Environ* 408(5):1131–1138. <https://doi.org/10.1016/j.scitotenv.2009.11.057>
- Wu B, Wang Z, Zhang Q, Shen N, Liu J (2017) Modelling sheet erosion on steep slopes in the loess region of China. *J Hydrol* 553:549–558. <https://doi.org/10.1016/j.jhydrol.2017.07.017>
- Yang CT (1996) Sediment transport: theory and practice. McGraw-Hill, New York
- Zahrieh J, Nezaratian H (2020) Estimation of transverse mixing coefficient in streams using M5, MARS, GA, and PSO approaches. *Environ Sci Pollut Res* 27:14553–14566. <https://doi.org/10.1007/s11356-020-07802-8>