



Random forest, support vector machine, and neural networks to modelling suspended sediment in Tigris River-Baghdad

Mustafa Al-Mukhtar 

Received: 18 March 2019 / Accepted: 11 September 2019 / Published online: 25 October 2019
© Springer Nature Switzerland AG 2019

Abstract Suspended sediment is one of the most influential parameters on the water bodies' pollution. It can carry different pollutants with different concentration through the suspension movement in the flow. Therefore, it is of utmost importance to monitoring or modelling these loads so that an accurate sediment reduction strategy can be adopted. However, the monitoring process is laborious and time-consuming task. Thus, modelling is suggested as an alternative method. In this study, three different methods of artificial intelligence (i.e., random forest, support vector machine (Radial Basis Function), and artificial neural network) were employed to model and predict the suspended load at Sarai Station in Baghdad. To this end, observed flow rate (m^3/s) and the corresponding suspended sediment concentration (mg/l) measured over the periods 1962–1981 and 2000–2010 were collected. Auto and partial correlation was used to identify the best combinations of input model data. The data was randomly partitioned into 75% for training and 25% for validation. The confidence interval was hypothesized to assess the uncertainty in the observed and predicted data. Whereas, the k -fold cross validation was used to quantify the uncertainty in the modelling results. The predictive modelling results for the three evaluated methods were assessed based on R^2 , RMSE, and NSE coefficient. Results show that random forest has the superior performance among

the others. The total suspended sediment transported was estimated to be 72,734,852 ton during the period 2000–2010.

Keywords Artificial intelligence · Uncertainty analysis · k -fold partitioning · Sediment load · Tigris

Introduction

Sediment transport by rivers is the main cause of critical issues related to water quality, reservoir siltation, design of dams, channel navigability, fish habitat, hydropower plant malfunctioning, and soil loss (Kaveh et al. 2017; Francke et al. 2010). Therefore, it represents an important constraint for the hydraulic projects management, environmental issues, and watershed management (Kisi and Zounemat-Kermani 2016). Typically, sediment load is classified into suspended sediment load (SSL), bed load, or wash load (Efthimiou 2019). SSL constitutes a serious problem in water resources field as it acts as a physical pollutant leading to turbidity increase or as chemical pollutant through chemical adsorption (Doğan et al. 2007).

An estimation of SSL in the field is laborious and time-money consuming. Thus, in most countries, suspended sediment data with short sampling intervals are rarely available (Al-Mukhtar and Al-Yaseen 2019). In contrast, streamflow data are often available at daily, hourly, or even shorter. The streamflow discharge is the main driving force for the initiation of sediment motion where the turbulent currents pendulant sediments by the

M. Al-Mukhtar (✉)
Civil Engineering Department, University of Technology,
Baghdad, Iraq
e-mail: mmalmukhtar@gmail.com

upward flow components (Vafakhah 2013). Hence, it is of great importance to determine the most suited modeling method of the suspended sediment-streamflow discharge relationship, which can be applied for present estimation and future projection.

In the literature, several techniques have been applied to estimate the SSL. These techniques are classified into (i) physically process based-distributed (e.g., Ascough et al. 1997; Arnold and Srinivasan 1998), (ii) lumped conceptual (Wichmeier and Smith 1978; Williams 1975), (iii) empirical/regression (Renard et al. 1996), and (iv) data driven based.

Physical-based and lumped conceptual models are required a large amount of input data and thus can be difficult to apply. For better application of these models, the data availability in the area should exceed the model requirements (Kalbus et al. 2012).

Because of the inherent non-stationarity, dynamism, and noise in the sediment mechanism, the conventional methods such sediment rating curve or linear regression models have shown mostly incompetency to achieve an accurate prediction of suspended sediment. Given their simplicity, the above methods are not able to understand the behavior of sediment transport in rivers (Afan et al. 2016). In their study, Shiau and Chen (2015) pointed out that “the sediment rating curve is insufficient to describe the inevitable scatter between sediment and discharge.”

Artificial intelligence (AI) as a type of data-driven models has been widely used to modeling the sediment transport as it has demonstrated capacity to address the complexity and noise data problems (Nourani et al. 2014). However, the physical processes are not considered in these models; alternatively, AI models are only tools used to capture the relationships between the relevant input and output variables (Olyaie et al. 2015). Hence, they are likely be more accurate than process models because they are dependent on data (Solomatine et al. 2008). AI models encompass, to name a few, Artificial Neural Networks (ANNs), Fuzzy Rule-Based Systems (FRBS), Random Forest (RF), and Support Vector Machines (SVM).

ANNs have been widely used to solve hydrology-related problems (Tayfur 2002). For example, in rainfall-runoff modelling (Minns and Hall 1996; Mason et al. 1996; Rajurkar et al. 2004; Harun et al. 2002), streamflow forecasting (Dolling and Varas 2002; Kişi 2007), groundwater modelling (Coppola Jr et al. 2003; Daliakopoulos et al. 2005; Maiti and Tiwari

2014), water quality (Palani et al. 2008; Wen and Lee 1998; Singh et al. 2009; Maier and Dandy 1996), sediment transport (Licznar and Nearing 2003; Ouellet-Proulx et al. 2016; Kisi 2004; Zhu et al. 2007; Alp and Cigizoglu 2007; Nagy et al. 2002; Rai and Mathur 2008; Rajaei et al. 2009), and many other hydrological aspects.

The applications of RF and SVM have been successfully reported in the water-related studies. For example, in the USA, Çimen (2008) applied SVM for predicting SSL in two rivers using the discharges as inputs. The study showed that SVM outperformed the fuzzy differential evolution and fuzzy logic. Francke et al. (2010) used traditional sediment rating curves (SRC), generalized linear models (GLM), and RF and Quantile Regression Forests (QRF) techniques to relate SSL to discharges. They concluded that the AI methods provided the best performance. Moreover, they were superior in calculating of confidence levels for the predictions, which in turn useful in the computation of sediment yields and the associated uncertainties. Kakaei Lafdani et al. (2013) used the four different kernels of nu-SVM, i.e., linear, polynomial, sigmoid, and Radial Basis Function (RBF), to predict SSL. They used streamflow and rainfall data as input variables. They concluded that RBF kernel function for SVM model has more capability for prediction and thus represents the reasonable and promising method than the other kernel functions. Li et al. (2015) applied RF for forecasting lake water level variations in Poyang Lake, China. They demonstrated that for daily forecasting, the RF model could attain more reliable and accurate forecasting results than ANN, SVM, and linear modeling in terms of RMSE and R^2 .

The objective of this study was to determine the most suited predictive model of SSL that can be satisfactory applied in Tigris River-Sarai station. To attain the objective of this study, an evaluation and comparison of three different AI methods were made. The evaluated methods were random forest, support vector machine, and ANN. An intermittent suspended sediment concentration (SSC) data with their counterparts streamflow discharges for the period 1962–1981 were employed for the purpose of this study. As these period represents the only available measured SSC data in the study area. In addition, data set of daily flow discharges collected over 11 years (2000–2010) was used to predict the SSL from the best evaluated method. The auto and partial correlation was utilized for determining the most

effective predicting variables considering as far as possible a parsimonious predictive model. The uncertainty in modelling results and population parameters was quantified through the k -fold cross validation method and confidence interval, respectively. All the computational modellings were done under R packages version 3.4.1.

Materials and methods

Study area and data

The Sarai station is a gaging station established in 1931 located on Tigris River in Baghdad at latitude $33^{\circ} 18' N$ and $44^{\circ} 23' E$ longitude (Fig. 1). Tigris River flows through Turkey, Syria, and Iraq with catchment areas of 57614, 834, and 253000 km², respectively (Bozkurt and Sen 2013). The total drainage area of Tigris River up to Sarai station is 134,000 Km². The river enters Baghdad city from the north where the river characterized as alluvial plain and multimeanders (Al-Ansari et al. 1979). The bed of the river is composed of sand and clay with slope about 7 cm/km. Due to the impacts of climate change, anthropogenic activities, and water policy of Turkey, the mean daily discharge in Tigris River dropped from 1140 to 546 m³/s in 2000s (Ali et al. 2017). Hence, recently, many islands and point bars emerged in the river which ultimately affected on the hydraulic performance and ecological behavior in the river. Thus, it is of utmost importance to adopt some strategies to mitigate the sediment transport amount. Such strategies necessitate firstly an accurate quantification of that amount.

However, for the purpose of this study, the data collected for the Sarai Station were daily discharge (m³/s) and their corresponding suspended sediment concentrations (mg/l), measured intermittently, at best four readings per month. The only period where both data sets were measured simultaneously is 1962–1981. Despite the station is currently in use to measure the water discharges, unfortunately, no data available about the SSC after that period. However, in total, only 111 measurements were exist from the above period. Additionally, the daily discharges from 2000 to 2010 were collected.

The observed SSC and water discharges data were randomly portioned into 75% for training the AI models and 25% for validation. Table 1 below

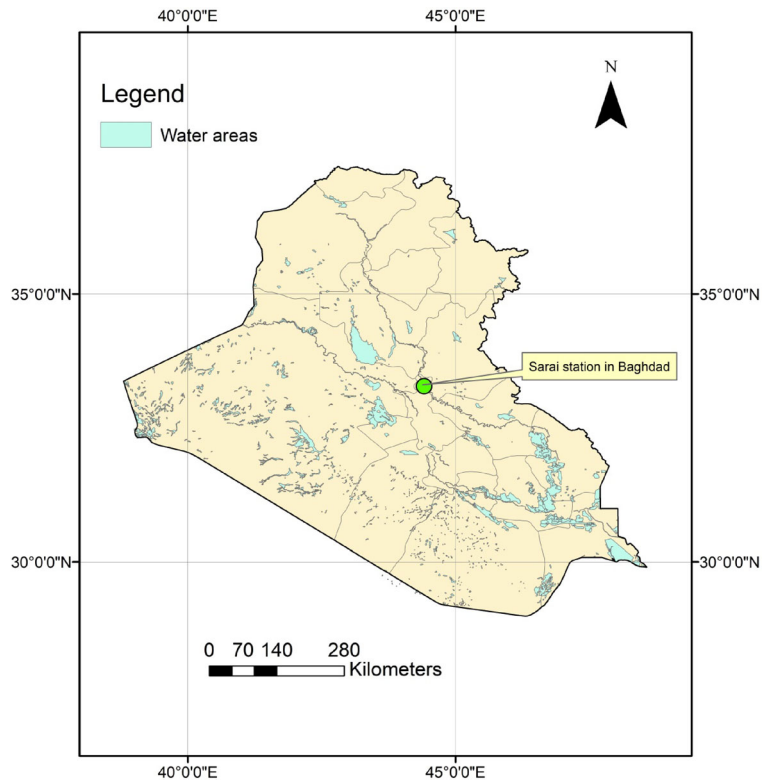
describes the statistical summary of SSC and flow rates used in this study. The minimum and maximum values of training/validation SSC data were 27/181 mg/l and 3071/1262 mg/l, respectively. The first quartile and third quartile values of training/validation data were 206/475 and 808/818 mg/l, respectively, and 50% of the data during training and validation were greater than 502 and 659 mg/l, respectively. On the other side, the minimum and maximum values of training/validation flow rates data were 294.19/331.54 m³/s and 2647.59/1487.05 m³/s, respectively. The first quartile and third quartile values of training/validation data were 471.225/472.577 and 1179.81/962.9 m³/s, respectively. Additionally, the mean values are greater than the median in both data sets, which implies that the right tail is longer than the left and hence proving that the data are right skewed.

Random forest (RF)

RF is one of the most powerful ensemble-learning algorithms. Breiman (2001) proposed the RF algorithm by adding additional randomness layer to bagging method. It functions by constructing multiple decision trees and final predictions are obtained from the averaged results. Each tree is constructed using different bootstrap sample of the data by adopting changes in how the classification or regression trees are constructed. These changes are represented by randomly sample of the candidate predictors and choosing the best split among the variables (Breiman 2001). Hence, two parameters are paramount in RF, which are n_{tree} and m_{try} . n_{tree} is the number of trees in the forest, while m_{try} is the number of variables in the random subset at each node. In this study, the default n_{tree} and m_{try} values were considered, which is $p/3$ and 500, respectively, where p is the number of predictors.

The algorithm of RF starting by drawing n_{tree} bootstrap sample from the data. Subsequently, an unpruned classification or regression tree is grown for each bootstrap sample (Ouedraogo et al. 2019). Then, at each node, a random sample of the predictors is to be taken and the best split from among those variables (predictors) is chosen. Lastly, a new data is predicted by aggregating the prediction of n_{tree} trees (Liaw and Wiener 2003). For more detail description and mathematical equations on RF, the reader is referred to Breiman (2001), Breiman (1996) and Liaw and Wiener 2003.

Fig. 1 The study area location



Support vector machine (SVM)

SVM is a soft computing AI method developed by Vapnik (1995). The method has been successfully used

Table 1 Descriptive statistics of suspended sediment concentration (mg/l) and river discharge (m³/s) at Sarai station during 1962–1981 and 2000–2010

		1962–1981		2000–2010
		Training	Validation	All
SSC mg/l	Min.	27.88	181.1	
	1st Qu.	206.03	475.1	
	Median	502.79	659.9	
	Mean	597.24	667.9	
	3rd Qu.	808.08	818.4	
	Max.	3,071.47	1262.90	
Flow rate cumecs	Min.	294.19	331.54	280
	1st Qu.	471.225	472.577	520
	Median	690.69	750.58	520
	Mean	898.06	771.12	546.23
	3rd Qu.	1179.81	962.9	520
	Max.	2647.59	1487.05	1315

in classification and recently in regression (Kecman 2001). There are different types of kernel function in support vector regression (SVR), i.e., linear, polynomial, and radial basis function (RBF) (Lan 2014). RBF is known to better handle the case when the relationship between inputs and outputs is non-linear and it encompasses fewer numerical difficulties than others (Lin et al. 2006). Hence, the commonly used RBF kernel is adopted in this study. Below is a brief description on the RBF-SVM, also called ϵ -SVR. The regression form for the SVR is

$$f(x) = w_i \varphi_i(x) + b \tag{1}$$

where w_i and b are the weight vector and bias, respectively, and φ_i is the nonlinear converter function used to map the input vectors into high dimensional space. Minimizing Eq. 2 is done through a convex optimization function given in Eq. 3 with an ϵ -insensitivity loss function to ultimately produce the nonlinear kernel RBF in Eq. 5.

$$\frac{1}{2} \|W\|^2 + C(\sum_i^N (\xi_i + \xi_i^*)) \tag{2}$$

$$\text{Subject to } \begin{cases} w_i\varphi_i(x) + b_i - y_i \leq \varepsilon + \xi_i^*, & i = 1, 2, \dots, N \\ y_i - w_i\varphi_i(x) - b_i \leq \varepsilon + \xi_i^*, & i = 1, 2, \dots, N \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (3)$$

where C is the cost factor that defines the empirical error in the optimization problem. $\|W\|^2$ is Euclidean smooth vector. ξ_i and ξ_i^* are covariate variables which causes training error for points out the tolerance error ε by loss function.

$$|\xi_\varepsilon| = |y - f(x)|_\varepsilon = \begin{cases} 0 & \text{if } |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & \text{if otherwise} \end{cases} \quad (4)$$

$$K(x, y) = \exp(-\alpha |x - x_i|^2) \quad (5)$$

Artificial neural networks (ANNs)

ANN is a biological computational model inspired from the human brain functions (Solomatine et al. 2008). Typically, it consists of three layers, i.e., input, hidden (neurons), and output. The relationships between input-output and the state of network are extracted from the data itself during the training of the network (Dumedah et al. 2014). An input-output mapping is performed using a set of interconnected simple processing through the hidden layer (Ghumman et al. 2018). Each neuron in the hidden layer receives signals externally or from other neurons and processes it through an activation transferable function. The common used activation function is logistic, linear, or sigmoid curve. The data are processed from the input to output through the hidden layer successively in what is called feedforward. The backpropagation algorithm (BPA), which was principally developed by Werbos in 1974, is the most commonly used learning algorithm in feedforward neural networks (Kasabov 1996). This algorithm minimizes the error between the modelled and actual output values through a gradient descent optimization algorithm. An adjustment of the weighted connections between layers is set after each training episode until the error in the validation data set begins to increase (Dawson and Wilby 2001).

In this study, a multilayer perception network with one hidden layer using the BPA learning algorithm was trained to establish the ANN model. The activation

function used in the hidden layer is a log-sigmoid function and a linear function in the output layer.

Selection of the predictors

In this study, the daily discharges of the Tigris River at Sarai station with different time lags were considered as inputs (predictors), and the current suspended sediment concentration (mg/l) as outputs. Commonly, using antecedent values of water discharges or SSL might improve the model performance. Predictor variable selection was performed to find the optimal model and building concise powerful models by preventing overfitting and eliminating collinearity in the predictors (Harrell 2001). Subsequently, the auto (ACF) and partial auto correlation (PACF) were employed to determine the optimum lag of antecedent predictors using the default time lag (up to 20 values). ACF is defined as a statistical analysis used to determine the degree of correlation between adjacent values correlation (McCuen 2002). While, the PACF is the partial correlation of a time series with its own lagged values without considering the influences of intervening lagged autocorrelation (Al-Mukhtar 2016).

Uncertainty analysis

Confidence interval

Confidence interval (CI) is defined as a plausible range of the population's parameters values. It is commonly used to assess the uncertainty in the sampling distribution based on the central limit theorem (CLT) or bootstrapping method. The difference between the above methods is arisen from the way of the sampling. In CLT, we sample from the original population while in bootstrapping, the sampling is carried out from the sample itself (DiCiccio and Efron 1996). Typically, it is almost impossible to collect the entire population data (as in our case). Therefore, and in order to quantify the uncertainty in the population mean of the observed SSL (1962–1981) and in the predicted SSL (2000–2010) from the best-evaluated method, 95% CI was constructed using above condition based on bootstrap samples using the standard error method. This process was to guarantee that the true population parameter (mean SSL) is within the range of confidence levels in the

respective intervals (1962–1981 and 2000–2010) with 95% confidence.

Cross validation

Data partition constitutes the main component in developing an accurate and reliable suspended sediment model. This fact emerges from the complexity of sediment transport phenomenon, and hence, it is mostly unreliable to build a single predictive model that is able to capture the entire system behavior based on one group. Therefore, the k -fold cross validation was used in this study to reduce the uncertainty in the modelling results. In k -fold cross validation, the training data were randomly partitioned into five equal-sized subsets where the predictive model is trained on all, except one for testing. The procedure is repeated k times where k is the number of subsets and the evaluation criteria were averaged to obtain the final performance (Casanueva et al. 2014). Subsequently, the best performance model was applied to simulate SSC using the training and validation data sets.

Evaluation criteria

The quantitative statistics in hydrological modelling are divided into three major categories (Moriassi et al. 2007), i.e., standard regression, dimensionless, and error index. For better assessment of the predictive model accuracy, it is highly recommended that the statistical metrics must consider these various types. Therefore, in this study, the following statistics were used.

- 1- Determination coefficient R^2 (standard regression type)

$$R^2 = \left[\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right]^2 \quad (6)$$

where O_i is the actual value, \bar{O} is the average actual value, P_i is the predicted value, and \bar{P} is the average predicted value. Values of R^2 range from 0 to 1. The closer the value to 1, the better the model is.

- 2- Nash and Sutcliffe coefficient efficiency NSE (dimensionless type)

$$NSE = 1 - \frac{\sum_1^n (O_i - S_i)^2}{\sum_1^n (O_i - \bar{S})^2} \quad (7)$$

The values of NSE range from $-\infty$ to 1. A perfect fit between the modelled and measured data is represented by a value of 1. In general, a value of > 0.6 is considered satisfactory (Zounemat-Kermani et al. 2016).

- 3- Root mean square error RMSE (error index type)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P)^2}{n}} \quad (8)$$

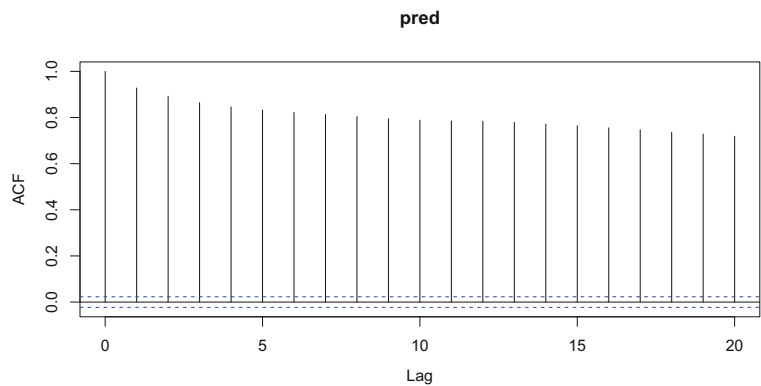
The RMSE values range from 0 to ∞ . The less this index, the better is the performance of the model.

Results and discussion

Model building and the parsimonious case

The auto correlation correlogram and the PACF plot of water discharges over the period 1962–1981 were plotted with their 95% confidence intervals as shown in Figs. 2 and 3, respectively. Obviously, as it can be seen from Fig. 2 that there were strong correlations among the various antecedent values up to 20 time lags at a significance level (α) of 0.05. The ACF starts from a value of 0.9 at time lag-0 and gradually decreased up to ~ 0.8 at time lag-20. On the other side, PACF plot (Fig. 3) show that there were a significant correlation at $\alpha = 0.05$ up to 6-time lag. Thereafter, the values fall within the confidence intervals and thus indicated an evidence to accept the null hypothesis. Therefore, initially, the common time lag between the two metrics was adopted in model building. In other words, the time lag-6 antecedent discharge values were used to model the current SSC. But, according to the statistical metrics used, adding the fifth and sixth values (time lags-5 and 6) to the models has not improved the performance of predictive models. Hence, and in order to maintain as far as

Fig. 2 Auto correlation plot of daily discharged in Tigris River at Sarai Station (1962–1981)



possible a parsimonious model, all the predictive models were built based on time lag-1 to 4 (Table 2).

The evaluated methods

Table 3 shows the statistical performances of the evaluated methods according to their different scenarios during training and validation. All the models with four and three input combinations show a superior performance in comparison to the other structures. Thus, time lag of discharge has a substantial influence on sediment load variations. The observations were closely matched with those modelled from RF, SVM, and ANN during training according to scenario 1. R^2 , NSE, and RMSE during the training period from RF#01-SVM#01-ANN#01 were 0.9, 0.75, and 189.65; 0.73, 0.57, and 265.22; and 0.85, 0.78, and 197.3, respectively. During the validation period, R^2 , NSE, and RMSE from RF#01-SVM#01-ANN#01 were 0.76, 0.71, and 144.98; 0.67, 0.60, and 194.02; and 0.68, 0.59, and 178.3, respectively, which implies a satisfactory performance to capture the observed variations. The highest R^2 and NSE with the lowest RMSE were obtained from RF. According to

scenario 2, R^2 , NSE, and RMSE during the training period from RF#02-SVM#02-ANN#02 were 0.9, 0.75, and 186.96; 0.62, 0.31, and 311.75; and 0.85, 0.78, and 197.3, respectively. During the validation period, R^2 , NSE, and RMSE from RF#02-SVM#02-ANN#02 were 0.76, 0.71, and 144.98; 0.67, 0.6, and 194.02; and 0.78, 0.67, and 240.84, respectively. Obviously, the performance of RF#02 was not considerably different on that from RF#01 during the training. However, the performance positively changed of RF#02 and ANN#02 on that from scenario 1 during the validation. The values of R^2 , NSE, and RMSE obtained from RF#02-ANN#02 models were 0.8, 0.75, and 130.71, and 0.68, 0.63, and 170.75, respectively. While results from SVM#02 show lower values of that from SVM#01 with R^2 , NSE, and RMSE equal to 0.53, 0.45, and 227.87, respectively. With respect to scenario 3, generally all the evaluated methods introduced poorer results than those from the above scenarios, which implies that the sediment transport process in the study area is highly dynamics and thus associated to higher range of antecedent discharges. R^2 , NSE, and RMSE obtained from RF#03-SVM#03-ANN#03 models during the training period were 0.9,

Fig. 3 Partial correlation plot of daily discharged in Tigris River at Sarai Station (1962–1981)

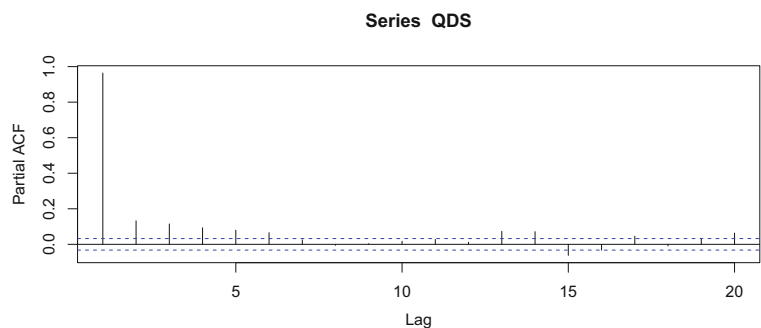


Table 2 Model descriptions

Scenarios	Model structure
Scenario 1	SSC-Q _t ,Q _{t-1} ,Q _{t-2} ,Q _{t-3}
Scenario 2	SSC-Q _t ,Q _{t-1} ,Q _{t-2}
Scenario 3	SSC-Q _t ,Q _{t-1}
Scenario 4	SSC-Q _t

0.77, and 184.51; 0.62, 0.14, and 322.38; and 0.78, 0.63, and 240.79, respectively. During the validation period, the performances were, 0.76, 0.69, and 145; 0.41, 0, and 266.01; and 0.58, 0.45, and 211.74, respectively. Finally, results from scenario 4 also show poor performance in comparison to scenarios 1 and 2. R^2 , NSE, and RMSE during the training period from RF#04-SVM#04-ANN#04 were 0.87, 0.77, and 192.76; 0.58, 0.26, and 330.24; and 0.68, 0.44, and 287.98, respectively. During the validation period, R^2 , NSE, and RMSE from RF#04-SVM#04-ANN#04 were 0.68, 0.66, and 175.11; 0.42, 0.11, and 257.91; and 0.5, 0.34, and 233.1, respectively. Results from various scenarios of RF outperformed those from SVM and ANN according to statistical metrics, indicating a superior performance in modelling SSC in the study area. Among the aforementioned model structures, the RF model with a time lag of 3 days (RF#02) for the case

Table 3 Statistical performances of the evaluated models

Model name	Training			Validation		
	R^2	NSE	RMSE	R^2	NSE	RMSE
rf#01	0.9	0.75	189.65	0.76	0.71	144.98
rf#02	0.9	0.75	186.96	0.8	0.75	130.71
rf#03	0.9	0.77	184.51	0.76	0.69	145
rf#04	0.87	0.77	192.76	0.68	0.66	175.11
svm#01	0.73	0.57	265.22	0.67	0.6	194.02
svm#02	0.62	0.31	311.75	0.53	0.45	227.87
svm#03	0.62	0.14	322.38	0.41	0	266.01
svm#04	0.58	0.26	330.24	0.42	0.11	257.91
ann#01	0.85	0.78	197.3	0.68	0.59	178.3
ann#02	0.78	0.67	240.84	0.68	0.63	170.75
ann#03	0.78	0.63	240.79	0.58	0.45	211.74
ann#04	0.68	0.44	287.98	0.5	0.34	233.1

#01: scenario 1; #02: scenario 2; #03: scenario 3; #04: scenario 4

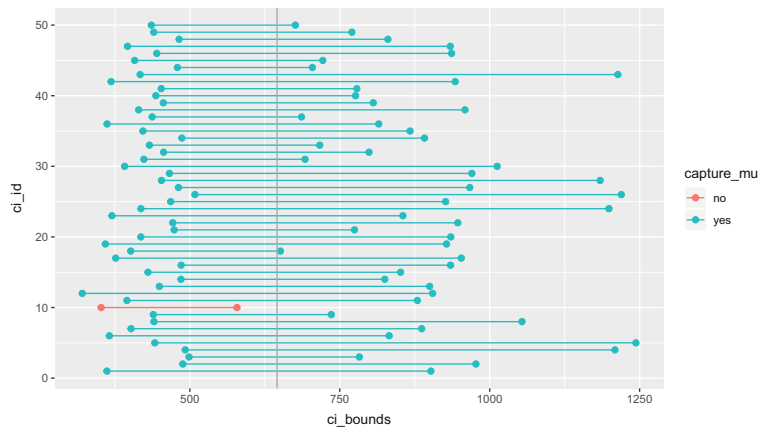
study exhibited the best performance with the lowest RMSE values of 130.71 mg/l during validation. Therefore, this scenario was adopted to predict the monthly SSL using daily observed discharges along the period 2000–2010.

The superior performance of random forest might be attributed to the fact that they do not consider easy interpretation of the effects of single predictors (De'ath 2007). Instead, a random subset of the predictors is used for each tree and at each node and hence is free of overfitting problems as the number of trees increases. Ultimately, and as pointed by Francke et al. (2010), predictions are made from a weighted average of the training data making the model predictions are always within the range of the observations. This precludes implausible values but inhibits extrapolation". On the other side, SVR shows poorer performance than that from RF and ANN. In their studies, Haji et al. (2014), Kakaei Lafdani et al. (2013), and Nourani and Andalib (2015) also reported that "the performance of SVM is worse than other AI methods such as ANN in modeling sediment load". Even the ANN model performed better than SVM but worse than RF. It is well known that the ANN models are unable to extrapolate beyond the range of the data used for training (Flood and Kartam 1994; Minns and Hall 1996). Additionally, when the validation data contain values outside the range of those used for training; poor predictions can be expected. In other words, it is necessary that the training and validation sets are representative of the same population. Moreover, it is noteworthy that the ANNs are highly dependent on the amount of trained data to generate prediction. These drawbacks might represent the limitation of using ANNs in prediction procedure and could explain the unsatisfactory performance of the ANNs in this study.

Uncertainty analysis

Given that the SSL data are highly complex and uncertain, it was essential to make inference on what the SSL population looks like. Figure 4 shows 50 confidence levels bootstrapped from 1000 random samples of size n ($n = 111$) per each one of them. The vertical line represents the true population mean of SSL across the period 1962–1981, which is equal to 645 mg/l. Each horizontal line depicts a confidence calculated based on different random sampling, which ranges from 385 to 1009 mg/l. There are 50 interval plots and 49 of them contain the true

Fig. 4 Fifty 95% confidence interval for the mean observed SSC (mg/l) from 1000 samples during 1962–1981 (μ is the mean)



population mean. Therefore, the proportion of confidence intervals is 98%, which implies that 98% of random samples of SSL during the period 1962–1981 will yield confidence intervals that contain the true population mean. Thus, it can be judged that the sample data are representative of the entire population. In addition, Fig. 5 depicts 50 confidence level bootstrapped from 1000 random samples of size n ($n = 3649$, daily modelled SSC from 2000 to 2010) from RF#02 model structure during 2000–2010. It can be clarified that the mean SSC during 2000–2010 was determined as 389 mg/l with range of values from 366 to 447 mg/l. The whole 50 confidence intervals captured the true population mean, indicating that the uncertainty in SSC during the period 2000–2010 was fairly identified by RF#02 model. Narrower range of confidence intervals obtained from the modelling results implies that the model precisely predict SSC during 2000–2010.

Based on the above, the total sediment amount from October 2001 to 2010 was predicted using the optimal model, i.e., RF#02. The suspended sediment

concentrations (mg/l) were converted into suspended sediment load (ton) using their corresponding discharges and a conversion factor. Figure 6 presents the temporal variations of modelled monthly SSL (ton) using box and whisker plot. Some outlier values were detected mostly in March where higher values of flow are released from Turkey due to the snow melting and subsequently increasing the water release from the upstream dams and regulators. The monthly sediment load in March ranges between 5073.75 and 50218.13 ton. The interquartile range ranges between 12064.88 and 30146.05 ton, which implies that the most likely sediment load of 50% probability might occur within this range. Also, it can be illustrated from Fig. 6 that the summer months (July and August) recorded the highest amounts of sediment load. The monthly sediment load during July and August ranges between 5021.56 to 60960.28 and 4686.80 to 60874.21 ton, respectively. The interquartile fluctuated between 9040.73 to 31008.70 and 9025.23 to 30693.95, respectively. The total summed sediment load over the period 2001–2010 was estimated to be 72,734,852 ton, indicating the

Fig. 5 Fifty 95% confidence interval for the mean modelled SSC (mg/l) from 1000 samples of RF#02 during 2000–2010

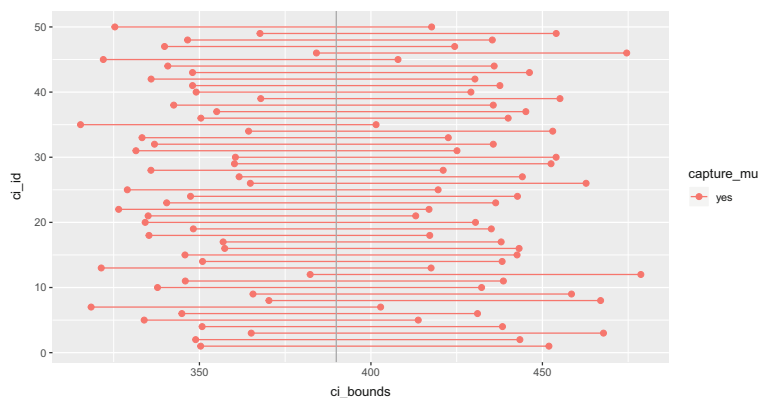
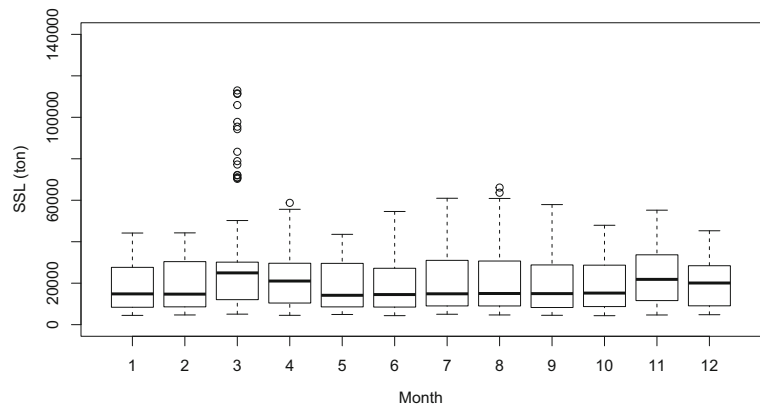


Fig. 6 Temporal variations of monthly sediment load over period 2001–2010



urgent need in adopting some strategies to reduce and mitigate the massive amount of SSL.

Conclusions

This study investigated three kinds of artificial intelligence methods in modelling suspended sediment concentration in Tigris River-Baghdad. The evaluation of the AI methods was performed based on two datasets: training and validation. During the training, the datasets was split into 5-subsets using k -fold cross validation. The best performance was determined based on the optimal average statistical metrics from the k -fold cross validation and subsequently used with an independent data for validation. Model structure was identified based on auto and partial correlation with a significant level of 0.05.

Results demonstrated that for SSC, the RF model attained more reliable and accurate forecasting results than SVR and ANN in terms of R^2 , NSE, and RMSE. Best prediction performance was obtained by incorporating input data with 3-day time lag of discharge from the river, showing the vital role of discharge time lag on the current value of SSC. The outcomes from this work would be providing a better insight on the amounts of sediment carrying by the river, and ultimately better water quality management.

Acknowledgments The author is grateful to the ministry of higher education and scientific research in Iraq for their support during this study. In addition, I would like to extend my gratitude to Dr. Rasul Khalaf from Al-Mustansiriya University-Baghdad for providing the sediment data.

References

- Afan, H. A., El-shafie, A., Mohtar, W. H. M. W., & Yaseen, Z. M. (2016). Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction. *Journal of Hydrology*, *541*, 902–913.
- Al-Ansari, N., Ali, S., & Taqa, A. (1979). Sediment discharge of the River Tigris at Baghdad (Iraq). *Canberra Symposium: The Hydrology of Areas of Low Precipitation*, (July).
- Ali, A. A., Al-Ansari, N. A., Al-suhail, Q., & Knutsson, S. (2017). Spatial measurement of bed load transport in Tigris River. *Journal of Earth Sciences and Geotechnical Engineering*, *7*(4), 55–75.
- Al-Mukhtar, M. (2016). Modelling the root zone soil moisture using artificial neural networks, a case study. *Environmental Earth Sciences*, *75*(15), 1124.
- Al-Mukhtar, M., & Al-Yaseen, F. (2019). Modeling water quality parameters using data-driven models, a case study Abu-Ziriq Marsh in South of Iraq. *Hydrology*, *6*(1), 24.
- Alp, M., & Cigizoglu, H. (2007). Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data. *Environmental Modelling & Software*, *22*(1), 2–13.
- Arnold, J., & Srinivasan, R. (1998). Large area hydrologic modeling and assessment part I: model development. *JAWRA Journal of the American Water Resources Association*, *34*(1), 73–89.
- Ascough, J., Baffaut, C., Nearing, M., & Liu, B. (1997). The WEPP watershed model. I. Hydrology and erosion. *Transactions of the ASAE*, *40*(4), 921–933.
- Bozkurt, D., & Sen, O. L. (2013). Climate change impacts in the Euphrates-Tigris Basin based on different model and scenario simulations. *Journal of Hydrology*, *480*, 149–161.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(421), 123–140.
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
- Casanueva, A., Frías, M. D., Herrera, S., San-Martín, D., Zaninovic, K., & Gutiérrez, J. M. (2014). Statistical down-scaling of climate impact indices: testing the direct approach. *Climatic Change*, *127*(3–4), 547–560.

- Çimen, M. (2008). Estimation of daily suspended sediments using support vector machines. *Hydrological Sciences Journal*, 53(3), 656–666.
- Coppola, E., Jr., Poulton, M., Charles, E., Dustman, J., & Szidarovszky, F. (2003). Application of artificial neural networks to complex groundwater management problems. *Natural Resources Research*, 12(4), 303–320.
- Daliakopoulos, I. N., Coulibaly, P., & Tsanis, I. K. (2005). Groundwater level forecasting using artificial neural networks. *Journal of Hydrology*, 309(1–4), 229–240.
- Dawson, C. W., & Wilby, R. L. (2001). Hydrological modelling using artificial neural networks. *Progress in Physical Geography*, 25(1), 80–108.
- De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology*, 88(1), 243–251.
- DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 189–212.
- Doğan, E., Yüksel, İ., & Kişi, Ö. (2007). Estimation of total sediment load concentration obtained by experimental study using artificial neural networks. *Environmental Fluid Mechanics*, 7(4), 271–288.
- Dolling, O. R., & Varas, E. A. (2002). Artificial neural networks for streamflow prediction. *Journal of Hydraulic Research*, 40(5), 547–554.
- Dumedah, G., Walker, J. P., & Chik, L. (2014). Assessing artificial neural networks and statistical methods for infilling missing soil moisture records. *Journal of Hydrology*, 515, 330–344.
- Efthimiou, N. (2019). The role of sediment rating curve development methodology on river load modeling. *Environmental Monitoring and Assessment*, 191(2), 108.
- Flood, I., & Kartam, N. (1994). Neural networks in civil engineering. II: Systems and application. *Journal of Computing in Civil Engineering*, 8(2), 149–162.
- Francke, T., Opez-Taraz', J. A. L., & Oder, B. S. (2010). Estimation of suspended sediment concentration and yield using linear models, random forests and quantile regression forests. *Hydrological Processes*, 2274(2008), 2267–2274.
- Ghumman, A. R., Ahmad, S., & Hashmi, H. N. (2018). Performance assessment of artificial neural networks and support vector regression models for stream flow predictions. *Environmental Monitoring and Assessment*, 190(12), 704.
- Haji, S., Mirbagheri, S. A., Javid, A. H., & Najafpur, G. D. (2014). A wavelet support vector machine combination model for daily suspended. *International Journal of Engineering*, 27(6), 855–864.
- Harrell, F. E. (2001). *Regression modeling strategies, with applications to linear models, survival analysis and logistic regression*. Berlin: Springer.
- Harun, S., Nor, N. I. A., & Kassim, A. H. M. (2002). Artificial neural network model for rainfall-runoff relationship. *Jurnal Teknologi*, 37(1), 1–12.
- Kakaei Lafdani, E., Moghaddam Nia, A., & Ahmadi, A. (2013). Daily suspended sediment load prediction using artificial neural networks and support vector machines. *Journal of Hydrology*, 478, 50–62.
- Kalbus, E., Kalbacher, T., Kolditz, O., Krüger, E., Seegert, J., Röstel, G., Teutsch, G., Borchardt, D., & Krebs, P. (2012). Integrated water resources management under different hydrological, climatic and socio-economic conditions. *Environmental Earth Sciences*, 65(5), 1363–1366.
- Kasabov, N. K. (1996). Foundations of neural networks, fuzzy systems, and knowledge engineering. Marcel Alencar.
- Kaveh, K., Bui, M. D., & Rutschmann, P. (2017). A comparative study of three different learning algorithms applied to ANFIS for predicting daily suspended sediment concentration. *International Journal of Sediment Research*, 32(3), 340–350.
- Kecman, V. (2001). *Learning and soft computing: support vector machines, neural networks, and fuzzy logic models*. Cambridge: MIT Press.
- Kisi, Ö. (2004). Multi-layer perceptrons with Levenberg-Marquardt training algorithm for suspended sediment concentration prediction and estimation. *Hydrological Sciences Journal*, 49(6), 37–41.
- Kişi, Ö. (2007). Streamflow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering*, 12(5), 532–539.
- Kisi, O., & Zounemat-Kermani, M. (2016). Suspended sediment modeling using neuro-fuzzy embedded fuzzy c-means clustering technique. *Water Resources Management*, 30(11), 3979–3994.
- Lan, Y. (2014). Forecasting performance of support vector machine for the Poyang Lake's water level. *Water Science & Technology*, 70(9), 1488–1495.
- Li, B., Yang, G., Wan, R., Dai, X., & Zhang, Y. (2015). Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China Bing. *Hydrology Research*, 47(S1), 69–83.
- Liaw, A., & Wiener, M. (2003). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Licznar, P., & Nearing, M. A. (2003). Artificial neural networks of soil erosion and runoff prediction at the plot scale. *Catena*, 51(2), 89–114.
- Lin, J., Cheng, C., & Chau, K. (2006). Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, 51(4), 599–612.
- Maier, H. R., & Dandy, G. C. (1996). The use of artificial neural networks for the prediction of water quality parameters. *Water Resources Research*, 32(4), 1013–1022.
- Maiti, S., & Tiwari, R. K. (2014). A comparative study of artificial neural networks, Bayesian neural networks and adaptive neuro-fuzzy inference system in groundwater level prediction. *Environmental Earth Sciences*, 71(7), 3147–3160.
- Mason, J. C., Price, R. K., & Tem'Me, A. (1996). A neural network model of rainfall-runoff using radial basis functions. *Journal of Hydraulic Research*, 34(4), 537–548.
- McCuen, R. H. (2002). *Modelling hydrological change: statistical methods*. Washington, D.C.: Lewis Publishers A.
- Minns, A. W., & Hall, M. J. (1996). Artificial neural networks as rainfall-runoff models. *Hydrological Sciences Journal*, 41(3), 399–417.
- Moriassi, D. N., Arnold, J. G., Van Liew, M. V., Binger, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *American Society of Agricultural and Biological Engineers*, 50(3), 885–900.
- Nagy, H. M., Watanabe, K., & Hirano, M. (2002). Prediction of sediment load concentration in rivers using artificial neural network model. *Journal of Hydraulic Engineering*, 128(6), 588–595.

- Nourani, V., & Andalib, G. (2015). Daily and monthly suspended sediment load predictions using wavelet based artificial intelligence approaches. *Journal of Mountain Science*, 12(1), 85–100.
- Nourani, V., Hosseini Baghanam, A., Adamowski, J., & Kisi, O. (2014). Applications of hybrid wavelet-artificial intelligence models in hydrology: a review. *Journal of Hydrology*, 514, 358–377.
- Olyyaie, E., Banejad, H., Chau, K. W., & Melesse, A. M. (2015). A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: a case study in United States. *Environmental Monitoring and Assessment*, 187(4), 189.
- Ouedraogo, I., Defourny, P., & Vanclooster, M. (2019). Application of random forest regression and comparison of its performance to multiple linear regression in modeling groundwater nitrate concentration at the African continent scale. *Hydrogeology Journal*, 27(3), 1081–1098.
- Ouellet-Proulx, S., St-Hilaire, A., Courtenay, S. C., & Haralampides, K. A. (2016). Estimation of suspended sediment concentration in the Saint John River using rating curves and a machine learning approach. *Hydrological Sciences Journal*, 61(10), 1847–1860.
- Palani, S., Liong, S. Y., & Tkalich, P. (2008). An ANN application for water quality forecasting. *Marine Pollution Bulletin*, 56(9), 1586–1597.
- Rai, R. K., & Mathur, B. S. (2008). Event-based sediment yield modeling using artificial neural network. *Water Resources Management*, 22(4), 423–441.
- Rajaei, T., Mirbagheri, S. A., Zounemat-Kermani, M., & Nourani, V. (2009). Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. *Science of the Total Environment*, 407(17), 4916–4927.
- Rajurkar, M. P., Kothiyari, U. C., & Chaube, U. C. (2004). Modeling of the daily rainfall-runoff relationship with artificial neural network. *Journal of Hydrology*, 285, 96–113.
- Renard, K. G., Foster, G. R., Weesies, G. A., McCool, D. K., & Yoder, D. C. (1996). *Predicting soil erosion by water: a guide to conservation planning with the Revised Universal Soil Loss Equation*. (Vol. 703). Washington, DC: United States Department of Agriculture.
- Shiau, J.-T., & Chen, T.-J. (2015). Quantile regression-based probabilistic estimation scheme for daily and annual suspended sediment loads. *Water Resources Management*, 29(8), 2805–2818.
- Singh, K. P., Basant, A., Malik, A., & Jain, G. (2009). Artificial neural network modeling of the river water quality-a case study. *Ecological Modelling*, 220(6), 888–895.
- Solomatine, D., See, L. M., & Abraham, R. J. (2008). Data-driven modelling: concepts, approaches and experiences. In *Practical hydroinformatics* (pp. 17–30). Berlin, Heidelberg: Springer.
- Tayfur, G. (2002). Artificial neural networks for sheet sediment transport. *Hydrological Sciences Journal*, 47(6), 879–892.
- Vafakhah, M. (2013). Comparison of cokriging and adaptive neuro-fuzzy inference system models for suspended sediment load forecasting. *Arabian Journal of Geosciences*, 6(8), 3003–3018.
- Vapnik, V. (1995). *The nature of statistical learning theory*. Berlin: Springer Science & Business Media.
- Wen, C., & Lee, C. (1998). A neural network approach to multiobjective optimization for water quality management in a river basin. *Water Resources Research*, 34(3), 427–436.
- Wichmeier, W. H., & Smith, D. D. (1978). *Predicting rainfall erosion loss: a guide to conservation planning*. Agriculture handbook (USA).
- Williams, J. R. (1975). Sediment-Yield prediction with Universal Equation using runoff energy factor. Present and Prospective Technology for Predicting Sediment Yields and Sources, ARS-S-40, US Department of Agriculture, Agriculture Research Service, pp. 244–252.
- Zhu, Y.-M., Lu, X. X., & Zhou, Y. (2007). Suspended sediment flux modeling with artificial neural network: an example of the Longchuanjiang River in the Upper Yangtze Catchment, China. *Geomorphology*, 84(1–2), 111–125.
- Zounemat-Kermani, M., Kişi, Ö., Adamowski, J., & Ramezani-Charmahineh, A. (2016). Evaluation of data driven models for river suspended sediment concentration modeling. *Journal of Hydrology*, 535, 457–472.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.