International Journal of Sediment Research 40 (2025) 91–107

Contents lists available at ScienceDirect

International Journal of Sediment Research

journal homepage: www.elsevier.com/locate/ijsrc

Original Research

Interpreting machine learning models based on SHAP values in predicting suspended sediment concentration

Houda Lamane [a, b, c, *], Latifa Mouhir [a], Rachid Moussadek [b, c], Bouamar Baghdad [d], Ozgur Kisi [e, f], Ali El Bilali [g]

[a] Department of Process Engineering and Environment, Faculty of Sciences and Techniques of Mohammedia, Hassan II University of Casablanca, Mohammedia 28806, Morocco
[b] Department of Environment and Natural Resources, National Institute for Agricultural Research (INRA), Rabat 10000, Morocco
[c] International Centre for Agriculture Research in the Dry Areas (ICARDA), Rabat 10000, Morocco
[d] School of Architecture and Landscape, Casablanca 20100, Morocco
[e] Department of Civil Engineering, Luebeck University of Applied Sciences, Lübeck 23562, Germany
[f] Department of Civil Engineering, Ilia State University, Tbilisi 0162, Georgia
[g] River Basin Agency of Bouregreg and Chaouia, Benslimane 13000, Morocco

ARTICLE INFO

Article history:
Received 11 June 2023
Received in revised form
4 September 2024
Accepted 9 October 2024
Available online 12 October 2024

Keywords:
Interpretability
Machine learning (ML)
Shapley values
Suspended sediment concentration (SSC)
Soil erosion
Bouregreg watershed (BW)

ABSTRACT

Machine learning (ML) has become a powerful tool for predicting suspended sediment concentration (SSC). Nonetheless, the ability to interpret the physical process is considered the main issue in applying most of ML approaches. In this regard, the current study presents a novel framework involving four standalone ML models (extra trees (ET), random forest (RF), categorical boosting (CatBoost), and extreme gradient boosting (XGBoost)) and their combination with genetic programming (GP). Three metrics (coefficient of correlation (r), root mean square error (RMSE), and Nash–Sutcliffe model-fit efficiency (NSE)) and a more advanced interpretation system SHapley Additive exPlanations (SHAP) are used to assess the performance of these models applied to hydro-climatic datasets for prediction of SSC. The calibration process was based on data from 2016 to 2020, and the validation was done for 2021 data. Further description and application of the framework are provided based on a case study of the Bouregreg watershed. The results revealed that all implemented models are efficient in SSC prediction with NSE, RMSE, and r varying from 0.53 to 0.86, 1.20–2.55 g/L, and 0.83–0.91 g/L respectively. Box plot diagrams confirm the enhanced performance of these combined models, and the best-performing ones for the four hydrological stations being the combined RF + GP model at the Aguibat Ziar station, the combined XGBoost + GP model at the Ain Loudah station, the CatBoost model at the Ras Fathia station, and the RF model at the Sidi M^{ed} Cherif station. The interpretability results showed that flow (Q) and seasonality (S) are the features most impacting SSC. These outcomes indicate that the applied models can extract accurate and detailed information from the interactions between the hydroclimatic factors and the generation of sediment by erosion (output). ML approaches illustrated the good reliability and transparency of the models developed for predicting SSC in a semi-arid setting, offered new perspectives for reducing ML models' "black box" character, and provided a useful source of information for assessing the consequences of SSC on water quality. The SHAP system and exploring other interpretable techniques are recommended to provide further information in future research. In addition, incorporating additional input data could enhance SSC predictions and deepen understanding of sediment transport dynamics.

1. Introduction

Estimation of the suspended sediment concentration (SSC) in rivers is essential against the backdrop of erosion, sedimentation, flood control, and long-term morphological assessment, and to

* Corresponding author.
E-mail address: lamanehoudageo@gmail.com (H. Lamane)
Peer review under the responsibility of International Research and Training Centre on Erosion and Sedimentation.

https://doi.org/10.1016/j.ijsrc.2024.10.002
1001-6279/

understand the effects of global climate change (Syvitski et al., 2003). Sediment transport by rivers causes critical issues in reservoir siltation, water quality, channel connectivity, hydropower station failure, soil loss, and fish habitat (Kaveh et al., 2017). Sediment yield is the total sediment outflow from a watershed per unit of time (Liu et al., 2015). It is a complex and non-linear process influenced by many interrelated input variables (Khankhoje & Choudhury, 2023). Thus, finding a model for estimating it is challenging in hydrological sciences (Zounemat-Kermani et al., 2020).

A plethora of models exist to study soil erosion and sediment yield processes (Borrelli et al., 2021; Bussi et al., 2014; de Vente et al., 2013; Labbaci et al., 2021; Lamane et al., 2022, 2023; Merritt et al., 2003). Several studies have compared and examined the benefits, strengths, limitations, and shortcomings of different sediment yield or SSC models (Jimeno-Sáez et al., 2022; Merritt et al., 2003; Pandey et al., 2016; Tan et al., 2018). It has been noted that empirical models are simple to use and conceptually easy to understand, still, they can be ineffective when applied beyond a range of developed and calibrated environmental conditions (de Vente et al., 2013). On the other hand, several studies have noted the difficulty of applying physically-based models, as data are sparse and in large drainage basins where the models have not been carefully assessed (de Vente et al., 2013; Merritt et al., 2003). Indeed, a drawback of physically based models is their requirement for extensive information to be developed, calibrated, and validated (Jimeno-Sáez et al., 2022). In addition, the estimation of SSC is subject to the influences introduced by any errors in streamflow and sediment concentration measurements, and direct measurements of these variables are expensive (Kisi et al., 2012). Furthermore, the calibration of the many parameters of these models is very complex, due to the non-linearity of sediment transport processes, and requires specialized knowledge and high computational time (Khosravi et al., 2020). Such restrictions require the development of more efficient SSC prediction models.

Data-driven methods have been extensively used to model sediment transport with a proven ability to address data complexity and noise (Nourani et al., 2014). These methods, such as Machine Learning, have emerged as a powerful alternative to physically-based models due to their ability to model complex nonlinear systems (Jimeno-Sáez et al., 2022). In contrast to hydrological models, data-driven methods have higher prediction accuracy with simple input data and low computational costs (Ji et al., 2021) and are data-dependent (Solomatine et al., 2008).

A review of all ML applications in SSC prediction is beyond the scope of the current paper, and only a few pertinent papers are discussed here. For instance (Kisi, 2005), used a neuro-fuzzy (NF) and artificial neural network (NN) approaches, along with the sediment rating curve and multi-linear regression, for estimating daily suspended sediment load (SSL). It was found that the NF model provided better estimates and was more flexible than ANN, regression, and rating curve models. Aytek & Kişi, 2008 applied genetic programming (GP) to suspended sediment transport fluxes and concluded that it performed better than conventional rating curves and multi-linear regression techniques. Jimeno-Sáez et al. (2022) predicted sediment load in a forest basin of northern Spain using SWAT and random forest (RF) models with the (M5 Pruning tree) M5P model tree algorithm. The results revealed that ML techniques were more accurate than the SWAT model. Al-Mukhtar (2019) modeled SSL in the Tigris–Baghdad River using RF, support vector machine (SVM) and ANN techniques. The results indicated that RF had higher performance than the other methods. Choubin et al. (2018) evaluated the use of a classification and regression tree (MUSLECART) model to estimate SSL based on hydro-meteorological data. The results showed that the CART model could be useful in basins with readily available hydrometeorological data. Hanoon et al. (2022) developed four ML techniques, namely gradient boosting regression (GBR), RF, SVM, and ANN to predict SSL at the Rantau Panjang station in the Johor River basin, Malaysia. The ANN model showed more reliable results than the other models. ANN, support vector regression (SVR), adaptive neuro-fuzzy inference system (ANFIS), and radial basis function (RBF) were used by Asadi et al. (2022) to forecast SSC in the Boostan Dam watershed in Iran, and the results showed that the SVR model had high accuracy in estimating SSC.

Boosting and bagging integration algorithms with regression capabilities have also been successfully applied to hydrology in recent years, due to their high predictive accuracy, computational efficiency, and optimization with various loss functions (Georganos et al., 2018). These algorithms include extra tree (ET), RF, categorical boosting (CatBoost), and extreme gradient boosting (XGBoost). For instance, GBR, AdaBoost regression (ABR), and random forest regression (RFR) were used by (Sharafati et al., 2020) to forecast SSL in the Mississippi River with the RFR model performing best in terms of predictive efficiency.

The different ML methods used in all the studies do not lead to a scientific consensus on the most appropriate method for simulating and predicting SSC. Consequently, it is necessary to identify the ML methods most commonly used to simulate SSC processes. Furthermore, even though in all studies, the predictive performance of ML methods has proved consistently satisfactory in different regions and fields, there is still a lack of transparency and interpretability (Sun & Scanlon, 2019), affecting its wide application to determine the influence of driving factors (Wang, Peng, & Liang, 2022). This causes many studies to decline using ML for attribution analysis (Nearing et al., 2021). Indeed, a number of them are regarded as uninterpretable “black boxes”, this is due to a large number of layers, or nodes, which makes it difficult to reveal the relations between inputs and outputs (El Bilali et al., 2023; Pereira et al., 2018). In addition, they cannot explain the process modeled or evaluate the influence of inputs and their interactions with the model target. Therefore, it is necessary to investigate the generalizability of several popular ML models, stand-alone and combined, for predicting SSC, to identify the relative contribution of different inputs to the model simulation, and to identify and interpret the drivers of SSC.

The field of interpretable ML has begun to blossom in the recent past and has gained a great deal of research attention due to the ability to improve ML applications in water resources research (Ding et al., 2020; El Bilali et al., 2023; Schmidt et al., 2020). However, the ability to interpret a model decision remains a precious property of a predictive system, though it has not received much attention so far (Pereira et al., 2018). Indeed, the evolution of ML models has emphasized the necessity and importance of their interpretability (Kolyshkina & Simoff, 2021; Krishnan, 2020). Nonetheless, compared to the focus on developing ML techniques and models themselves, as well as the focus on obtaining better performance metrics, research on interpretable ML remains a relatively small subsection of all ML research (Carvalho et al., 2019), especially in the soil erosion and SSC field. To address this issue (Shapley, 1953), introduced a SHapley additive exPlanation (SHAP) to make these ML models more interpretable. SHAP does interpretations of the activity of the predictions of complex ML models (Rodríguez-Pérez & Bajorath, 2020) and reveals the positive or negative impact of each variable on the model output (Lundberg & Lee, 2017). This approach demonstrates that evaluating the conditional expectation of a model and its inputs using SHAP values is the only consistent approach to evaluate feature attribution. The SHAP method is utilized for interpreting ML methods in different domains (Rodríguez-Pérez & Bajorath, 2020; Viana et al., 2021), including hydrological studies.

Zhang et al. (2022) developed an interpretable ML method that integrates the RF model and the SHAP method to investigate the water quality/land use relations in the Potomac River basin. In addition, interpretable ML models based on XGBoost and SHAP to predict estuarine water quality were developed by Wang, Peng, and Liang (2022). Recently, SHAP, local interpretable model-agnostic explanations (LIME), and Sobol sensitivity methods were embedded to interpret extra tree, XGBoost, deep neural network (DNN), and support vector regression (SVR) for modeling daily pan evaporation (El Bilali et al., 2022). However, to the author's knowledge, the existing studies on SSC prediction using ML algorithms addressing interpretability are limited. In this regard, the SHAP method was applied to identify the contribution of driving factors to the generation of SSC, the following subsection discusses the limits, objectives, and contributions of the current study.

In the Bouregreg watershed (BW), the area under study, the newest studies of SSC including (El Bilali et al., 2020), compared ANN and MUSLE (modified universal soil loss equation) models to predict reservoir sedimentation and found that the ANN outperformed MUSLE. Also (Ezzaouini et al., 2022), applied ML models involving adaptive boosting (AdaBoost), RF, K-nearest neighbor (K-NN), SVM, and ANN to predict SSC in the same watershed. The results showed that most ML models yielded satisfactory performance in SSC prediction and were good to excellent in generalization ability. However, to the authors knowledge, interpretability testing by combining models in SSC forecasting has not been previously examined.

In highly dynamic semi-arid environments, sediment transport is subject to several uncertainties related to the hydroclimatic input variables (Yang, 1977), seasonality (Ellison et al., 2016), and the delay response to the hydrological events (lag-time). Consequently, there is a need to develop an interpretable ML method by assessing the importance of the input variables for the best models. The current study marks the first attempt to introduce an interpretable model for predicting SSC in a semi-arid watershed to bridge this gap. The innovative aspect of this research is providing the first report to indicate the successful application of an interpretable model driven by standalone and combined ML methods, as gradient decision tree algorithms. The results and framework can be applied to other typical watersheds worldwide, making it possible to scientifically forecast trends in SSC assessment and future drivers and provide a decision-making basis for water resources management, ecological environment protection, and policy formulation in the BW and other semi-arid regions. Specifically, the current study has three objectives: (i) Evaluating and comparing the effectiveness of four standalone ML models with their combination with GP model. (ii) Using the SHAP method to interpret ML models and understand which input variables have a greater impact on SSC and how they affect SSC. Therefore, additional analysis was done to evaluate the bending inputs of hydroclimatic parameters to assess the physical agreement of the ML models with the erosion process. (iii) Emphasizing the power of interpretable ML models to accelerate assumptions and insights, and describe their crucial challenges and opportunities in predicting SSC.

2. Data

2.1. Study area

Located in the northwestern part of Morocco (5.4°–6.8° W and 32.8°–34° N) (Fig. 1), BW is one of the main watersheds in the country, featuring topographic diversity, supporting the main

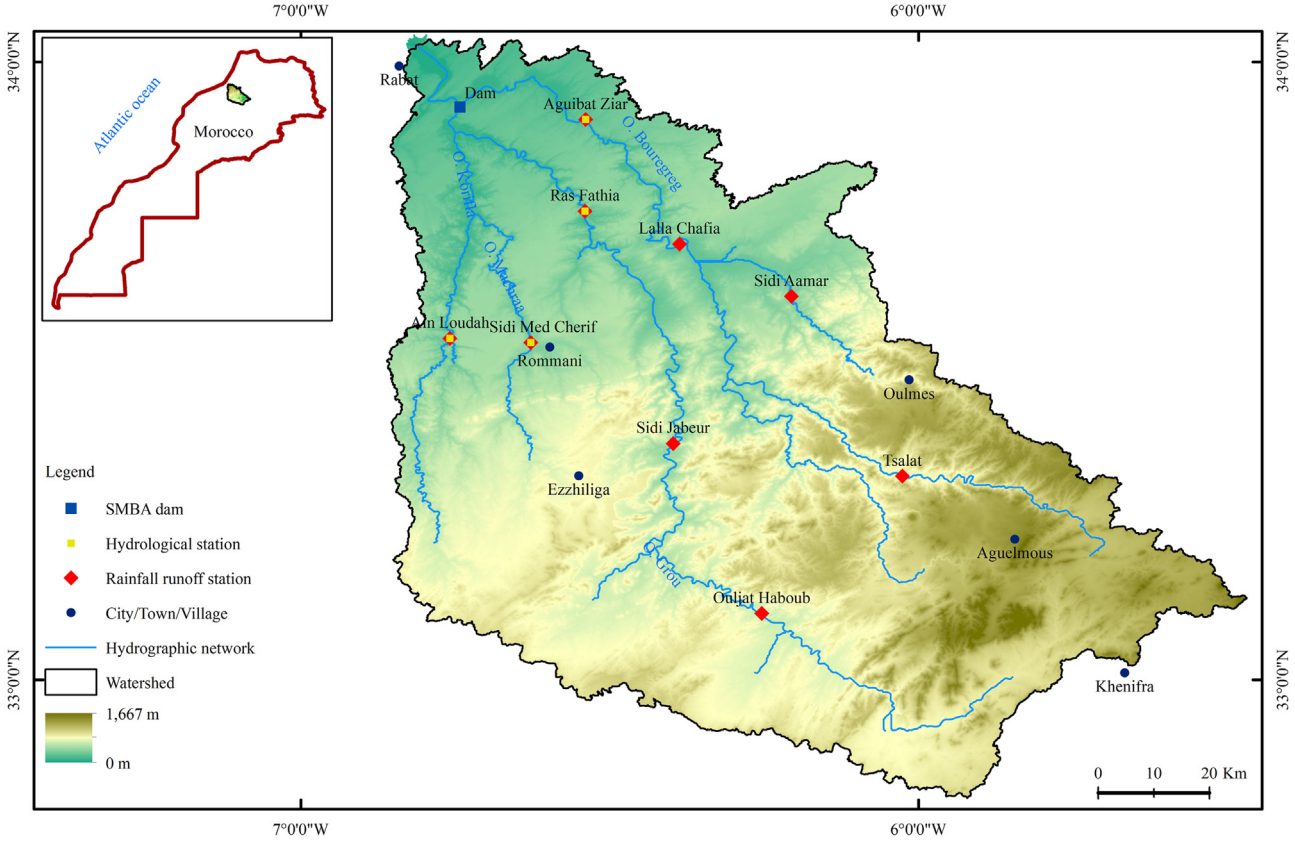

Fig. 1. Regional setting and topography of studied watershed and its different hydraulic and rainfall-runoff stations.

climatic zones in the country, and is home to the major vegetation and crops. It covers an area of over 9,500 km^2 and extends over 240 km. This zone is characterized by a highly varying geographic and geomorphologic framework. It includes three sub-basins: Bouregreg, Grou, and Korifla, and consists of three main rivers that form the hydrographic network, that is Bouregreg, Grou, and Korifla and its tributary, Machraa. There are 9 monitoring rainfall-runoff stations, as shown in Fig. 1, four of which also are hydrological stations with SSC data: Aguibat Ziar, Ras Fatia, Sidi M^{ed} Cherif, and Ain Loudah. Pasture and agriculture are the main socio-economic activities linked to water availability at the scale of BW.

Water resources in BW are very much dependent on the rainfall regime, which is very inconsistent due to the variance between the flood season (from December to April) and the dry season (summer) (Brouziyne et al., 2021). Climatically, the study area is semi-arid with an oceanic influence. The westernmost part of the basin has a sub-humid climate with 500 mm of cumulative rainfall and an annual mean temperature of 18 °C. The basin center has a semi-arid climate with less than 400 mm of rainfall and an annual mean temperature of 18 °C. The southeastern region of the basin has a continental influence (arid region) with 400 mm of cumulative rainfall and an annual mean temperature of 15 °C. The extreme northeastern part of the basin, the Oulmes Highlands, is a subhumid mountainous region with precipitation higher than 600 mm and an annual mean temperature of 15 °C (El Bilali & Taleb, 2020).

According to the land conditions and the topography of the study area, a wide variety of forms of erosion exist (El Aoula et al., 2021). In particular, gullies affect most of the basin, especially in clay and schist formations, and steep slopes. These are some of the formations most responsible for the siltation of the Sidi Mohammed Ben Abdellah (SMBA) Dam (Rhouichi, 1996). Indeed, this dam is suffering from serious siltation problems (Lahlou, 1986), reported an average of 2.5 million m^3/y siltation rate. As for sediment transport studies in this area, Ezzaouini et al. (2020) indicate a 9.49 million m^3/y siltation rate and the main source of sediment for the dam was the erosion of the dam's banks. Thus, the increase in the siltation rate is attributed to the sediment trapping that occurred as a result of the dam's reconstruction in 2006 (El Aoula et al., 2021).

2.2. SSC measurement

SSC was measured at four hydrologic stations: Aguibat Ziar, Ras Fatia, Sidi M^{ed} Cherif, and Ain Loudah. These monitoring stations are located immediately upstream of the SMBA Dam and monitor 87% of BW. The sediment samples were collected daily during low water flow and hourly during flood events; then they were analyzed in the laboratory, filtered under vacuum using membranes (0.45 μm), and weighed. Table 1 lists the characteristics of the SSC measurements.

2.3. Hydrological datasets

To model SSC, daily rainfall and runoff data were used, based on their importance as the main factors influencing the erosion process (Williams & Berndt, 1977; Wischmeier & Smith, 1965), measured at each of the nine rainfall stations previously mentioned. Fig. 2 shows and Table 2 lists the hydrological datasets from September 1, 2016 to August 31, 2021. The rainfall and runoff variability in BW are high due to flash flood events; the duration of the rainfall event is significant.

3. Methods

3.1. Modeling framework description

ET, RF, CATBoost, and XGBoost models were trained and validated in the current research for predicting daily SSC using the previous data: SSC (g/L), climatological and hydrological data (rainfall and runoff), seasonality, and lag-time. Fig. 3 is the methodological flowchart and an overview of the approach developed in the current study to predict daily SSC at four measuring stations in BW. Initially, python (Jupyter notebook) was used for implementing the four ML models. These were trained using a daily SSC series, from 2016 to 2020 and then combined with GP to compare their prediction accuracy with the standalone models. The validation process was carried out done 2021. Therefore, the SHAP method is embedded and applied to interpret the trained models. The following subsections are 3.2 to 3.6 describe the various components of the framework.

3.2. Machine learning methods

RF, ET, XGBoost, CatBoost, and GP were applied to predict the daily SSC. The ML models were trained and tested using the Python language within the Anaconda platform. Dataset loading and visualization processes were accomplished using the Pandas and Matplotlib libraries, respectively. Meanwhile, the implementation of regressor ML models was done using the Scikit-learn library. Based on their fundamental principles, the proposed ML models have been briefly described in the following subsections 3.2.1 to 3.2.5.

3.2.1. Random forest

RF is a supervised ML technique and a bagging ensemble of randomized linear decision tree model families (Breiman, 2001). It is one of the most powerful ensemble learning algorithms (Al-Mukhtar, 2019) and a trendy method that can often produce good predictions even in high-dimensional settings, without the need to fine-tune its internal parameters (Scornet, 2017), and operates reliably on large databases. Fig. 4 shows the RF flowchart.

3.2.2. Extra tree

ET is a relatively new ML technique, developed as an extension of the RF algorithm, and is less likely to overfit a dataset (Geurts et al., 2006). It is an ensemble learning method that averages decision trees (DT) to achieve significant improvements in accuracy and reduce computational complexities (Geurts et al., 2006). It

Table 1
SSC sample datasets at studied stations (source: RBABC).

Station	Aguibat Ziar	Ras Fathia	Sidi M^{ed} Cherif	Ain Loudah
River	Bouregreg	Grou	Machraa	Korifla
Area (km^2)	3,681	3,485	656	699
Number of SSC sample	1,253	1,634	1,662	470
Period of observation considered	01/09/2016 to 31/08/2021			
Average SSC (g/L)	0.83	1.20	0.63	1.05
Maximum SSC (g/L)	32.72	86.9	16.89	24.7
Standard deviation (STD) (g/L)	1.08	1.70	0.76	1.47

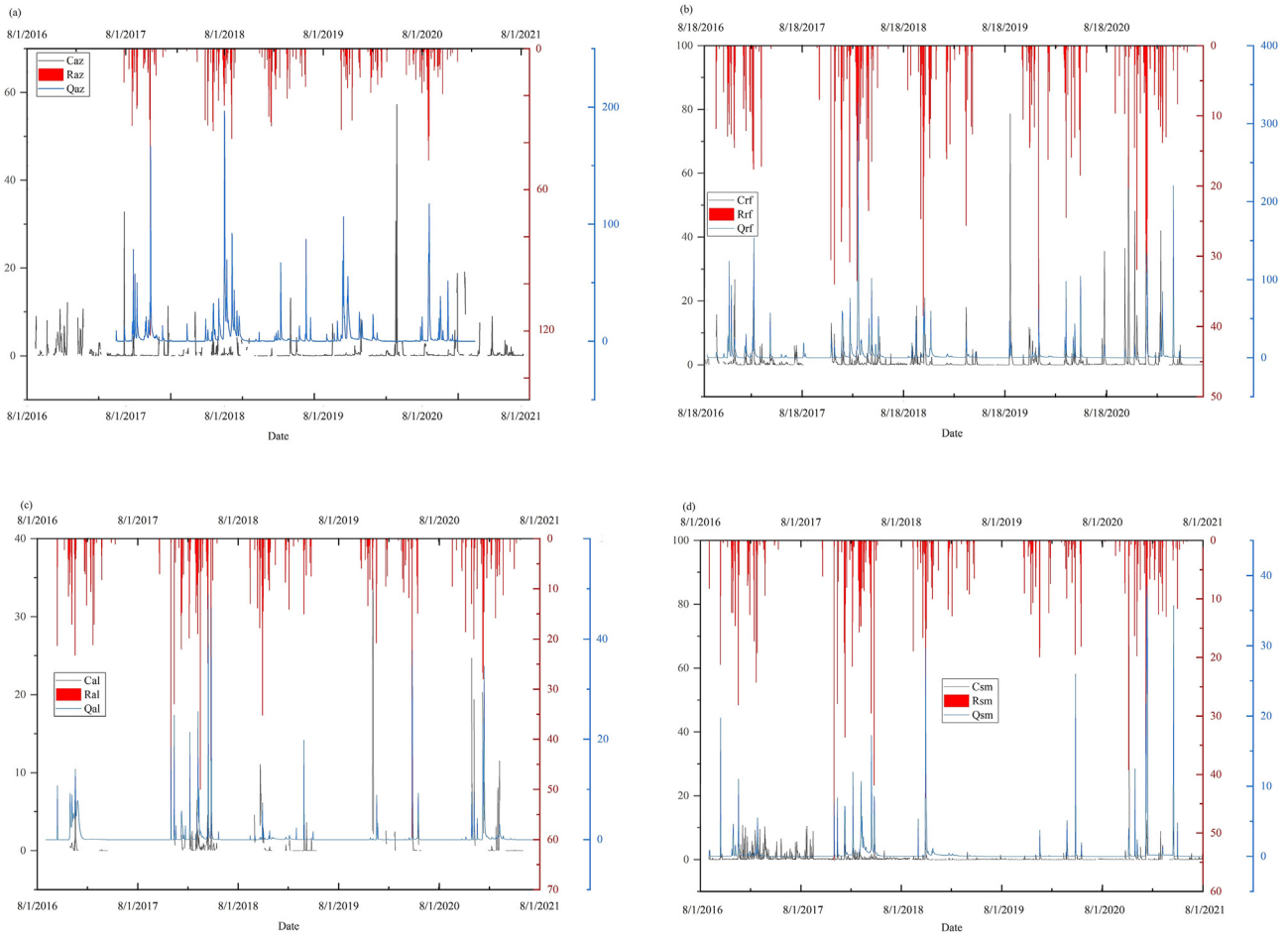

Fig. 2. Lines showing rainfall (*R*) on the top (and having the right-hand side *y*-axis in red), runoff (*Q*) at the bottom (and the right-hand site *y*-axis in blue), and suspended sediment concentration (C) at the bottom (and the left-hand site *y*-axis in black) at the four stations; (a) Aguibat Ziar (az), (b) Ras Fathia (rf), (c) Ain Loudah (al), and (d) Sidi M^{ed} Cherif (sm).

provides decision support by splitting the node attributes into a random format (John et al., 2016).

3.2.3. CatBoost

CatBoost is an algorithm proposed by (Prokhorenkova et al., 2018), which applies ordered boosting and an additional algorithm for categorical feature processing and deals with the problems of gradient bias and prediction shift. It represents an innovative algorithm that automatically processes categorical features as numerical ones. Moreover, it utilizes a perfectly symmetric tree model to mitigate overfitting, improving accuracy and generalization (Hancock & Khoshgoftaar, 2020; Huang et al., 2019).

3.2.4. XGBoost

XGBoost is an evolving supervised ML algorithm developed by (Chen & Guestrin, 2016). It allows the user to avoid potential overfitting problems (Molnar et al., 2020), and adapts to sparse data using a split search approach that considers sparseness (Chemura et al., 2020). It is renowned for its distinctive objective function and flexible choice of loss functions and offers high-performance, fast, and reliable processing of large data sets via its sophisticated block technology. Fig. 5 is the XGBoost flowchart. Further details of the XGBoost algorithm can be found in (Chen & Guestrin, 2016).

3.2.5. Genetic programming (GP)

The GP technique was originally introduced by (Koza, 1994). It is a stochastic technique for finding and optimizing the solution of a problem based on the principles of genetics (Wang & Yin, 2020). One notable advantage of applying GP compared to other data-driven methods is its ability to generate explicit formulations describing the governing relations in physical phenomena. Improving the interpretability of GP models mitigates the risk of over-fitting the training data, thus, improving model generalization. This GP process serves as an advancement, particularly in specific case studies (Kisi et al., 2012).

3.3. Input variables

SSC depends on several hydroclimatic variables, such as rainfall, runoff, and baseline conditions, in addition to the physical characteristics of BW. Further, the response of the sediment load to hydrological events is not instantaneous and the lag effect should be considered as an input variable. Similarly, the seasonality (*S*) is an important factor in the soil erosion phenomenon in semi-arid watersheds, as it impacts the land cover by the existence of vegetation combined with pastoral activities (Ferreira & Panagopoulos, 2014). In this paper, the seasonality associated with day of year (*j*) is calculated using Eq. (1):

$$S = \mathrm{Sin}\left(2\frac{2\pi}{365.25}j\right) \tag{1}$$

In the current study, data-driven methods are an effective way of analyzing and predicting the SSC using daily hydrological datasets such as rainfall (mm), stream flow (m^3/s), and SSC (g/L), that were recorded from September 1, 2016 to August 31, 2021. Rainfall and

Table 2
Daily rainfall and runoff data by sub-basin recorded from September 1, 2016 to August 31, 2021 (source: RBABC).

Sub-basin	Rainfall station	Rainfall (mm)			Runoff (m^3/s)		
		Average	Maximum	STD	Average	Maximum	STD
Bouregreg	Aguibat Ziar	1.3	122	2.23	4.16	199	5.64
	Lalla Chafia	0.9	43	1.64	3.06	197	4.85
	Sidi Aamar	1.1	50	1.94	0.48	24	0.61
	Tsalat	1.3	60	2.11	0.81	23	1.04
Grou	Ras Fathia	1.0	41	1.81	4.58	349	6.59
	Sidi Jabeur	0.8	44	1.50	3.74	327	5.29
	Ouljat Haboub	0.8	40	1.38	3.52	262	5.43
Korifla	Ain Loudah	0.9	60	1.57	0.52	49	0.88
Machraa	Sidi M^{ed} Cherif	0.9	55	1.56	0.42	39	0.62

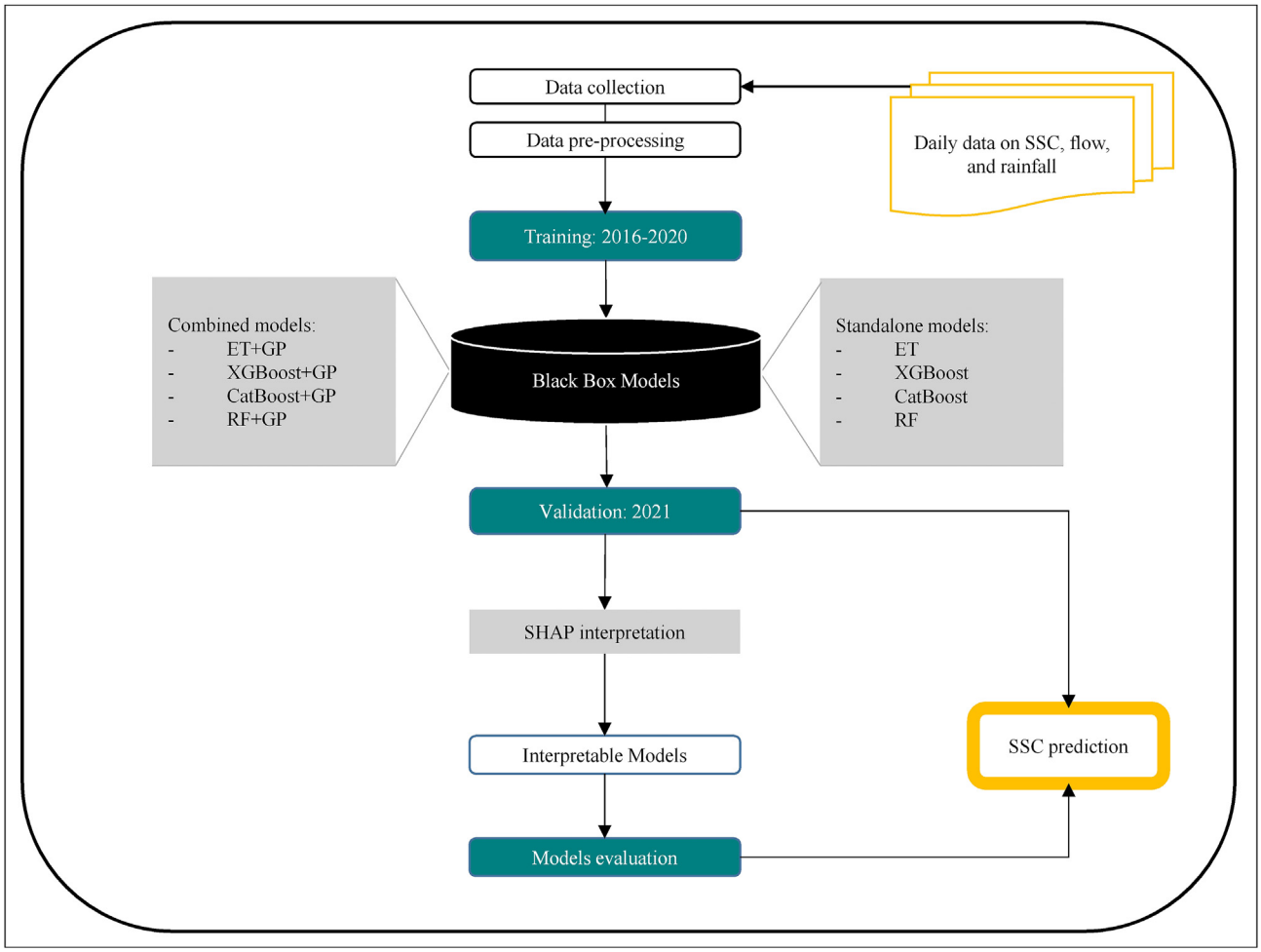

Fig. 3. Flowchart of proposed approach.

streamflow data are recorded at the nine BW stations (Aguibat Ziar, Lalla Chafia, Sidi Aamar Tsalat, Ras Fathia, Sidi Jabeur, Ouljat Haboub, Ain Loudah, and Sidi M^{ed} Cherif), and SSC is monitored at four stations (Aguibat Ziar, Ras Fathia, Ain Loudah, and Sidi M^{ed} Cherif). This dataset was analyzed to ensure no gaps in the time series. The calibration periods (recorded from September 1, 2016 to August 31, 2020) and the validation data sets (recorded from September 1, 2020 to August 31, 2021) were selected for each site. Statistical parameters for the full, calibration, and validation of SSC and *Q* are listed in Table 3.

3.4. Data preprocessing and tunning process

Data pre-processing steps include file merging and preparation, filling-in missing data, data partitioning, and feature scaling. These steps aim to prepare the datasets fed into the ML algorithms for model training and testing. Since most ML algorithms produce errors when they find missing values in a data set, the impute TS package, developed by (Moritz & Bartz-Beielstein, 2017) and available in the RStudio environment, was used for imputation filling-in using interpolation to solve the problem of missing or unavailable data

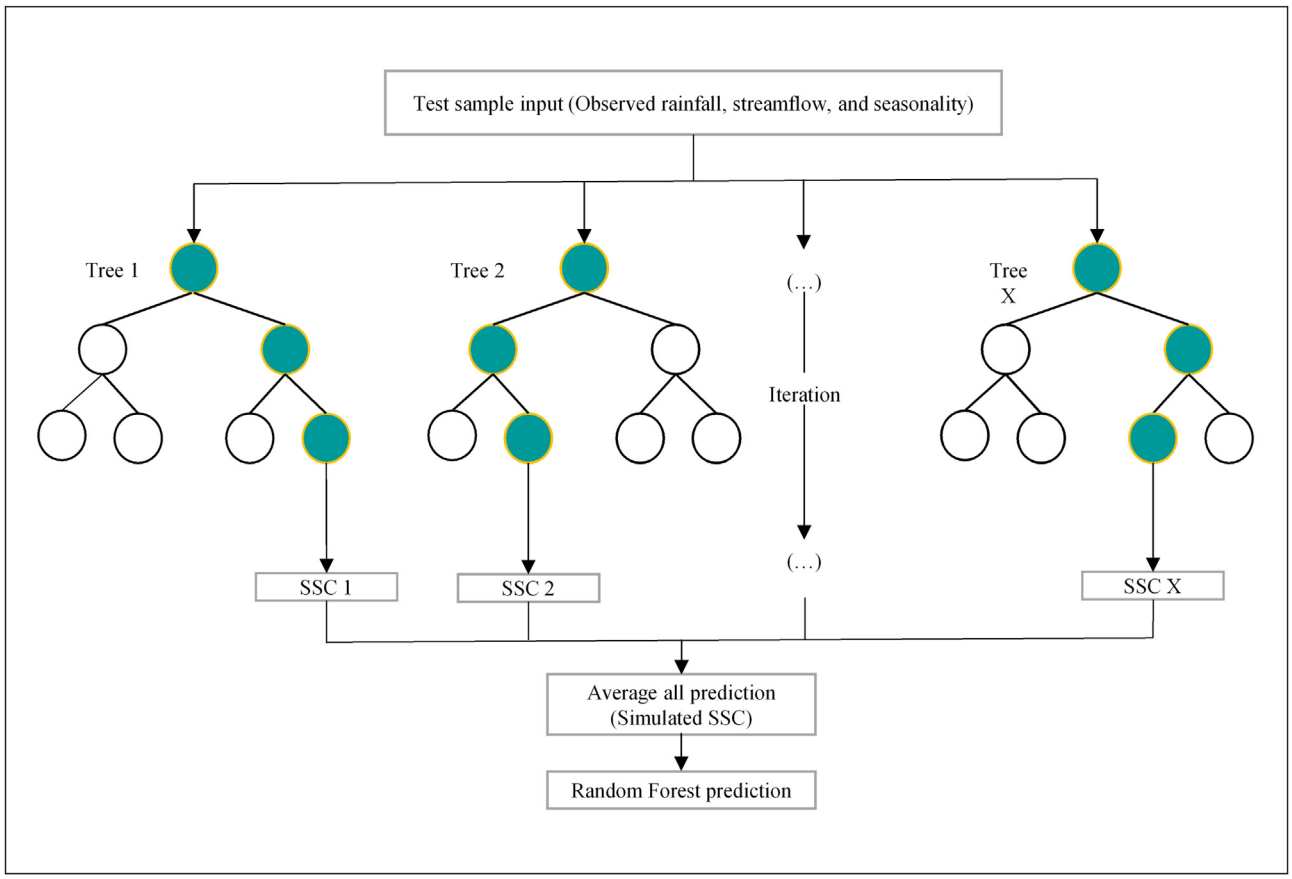

Fig. 4. Random forest flowchart.

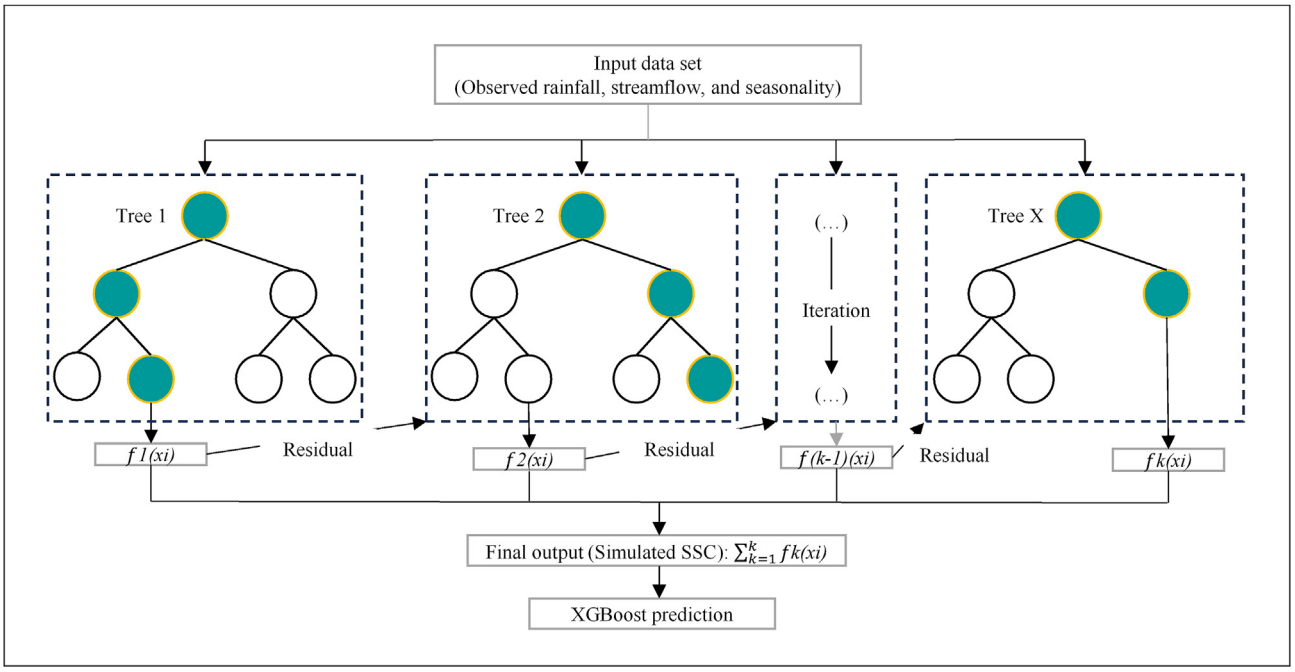

Fig. 5. XGBoost flowchart.

Table 3
Hyperparameters and structures of developed machine learning models.

Model	Parameters, functions, and typical range	Optimal parameter
Extra tree	Estimator number: 2,000–3,000	2,600
	n_jobs = 3–8	4
	minimum_sample_split = 10–40	25
	minimum_sample_leafs = 20–60	35
XGBoost	Estimator number: 1,000–4,000	3,000
	Maximum depth: 100–1,500	1,000
	Learning rate: 0.01–0.15	0.1
Random forest	Estimator number: 500–3,000	1,500
	n_jobs = 2–8	5
	minimum_sample_split = 10–40	27
	minimum_sample_leafs = 20–60	32
	Estimator number: 250–1,000	700
CatBoost	Estimator number: 250–1,000	650
	Maximum depth: 400–100	300
	Learning rate: 0.01–0.15	0.1

values with likely and reasonable values that were efficient in the field of modeling SSC (Essam et al., 2022; Mesfin et al., 2021; Schulz & Gerkema, 2017). Partitioning of the data is applied to the datasets to separate the daily data for each station into a training and testing set. The training set is used to develop and equip the ML models with the ability to make SSC predictions. In contrast, the testing set is used to evaluate the ability and prediction accuracy of the ML models using selected performance measures.

3.5. Metrics evaluation

The prediction performance of the implemented ML models is evaluated using three common evaluation metrics: root mean square error (RMSE), Nash–Sutcliffe model-fit efficiency (NSE), and correlation coefficient (r). RMSE (Eq. (2)) is a function of the model residuals which represents the average distance between the residuals and 0 or the average distance between the observed values and the model predictions has been also used. NSE (Eq. (3)) measures the degree of correspondence between the observed data and the simulated output. A perfect match between observed data and simulated values is indicated by an NSE value of 1. NSE values < 0 indicate that the average of the observed data is a more accurate predictor than the simulated output. r in Eq. (4) reflects the strength of the linear relationship between the predicted values and the observed data (Krause et al., 2005). It is a variable between 0 and 1, where 0 indicates no correlation, and 1 indicates perfect correlation. These performance metrics measure the deviation of the model from the observations and evaluate the predictive power of the model (Hyndman & Koehler, 2006; Nash & Sutcliffe, 1970). They are calculated using Eqs. (2)–(4):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (O_i - S_i)^2}{n}} \tag{2}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n} (O_i - S_i)^2}{\sum_{i=1}^{n} (O_i - \widehat{S})^2} \tag{3}$$

$$r = \frac{\sum_{i}^{n} (S_i - \widehat{S})(O_i - \widehat{O})}{\sqrt{\sum_{i}^{n} (S_i - \widehat{S})^2} \sqrt{\sum_{i}^{n} (O_i - \widehat{O})^2}} \tag{4}$$

where O_i and S_i are the observed and simulated SSC, respectively; n is the sample size of the dataset, $\widehat{O}$ is the average value of the observed SSC, and $\widehat{S}$ is the average value of the simulated SSC. As a model never is perfect, the perfect values are $r = 1$; $\in [-1, 1]$, RMSE = 0, (RMSE $\in$ [0, inf)), and NSE = 1, (NSE $\in$ (−inf, 1])).

3.6. SHAP implementation

The importance of features in conventional ML only indicates the degree of influence of input variables on model output but does not reveal how input variables influence model output (Lund et al., 2022; Wang, Peng, & Liang, 2022). Therefore, it is important to understand which input variables have the greatest impact on the results and how they affect them. In this sense, research in data science has advanced techniques to compute feature importance by incorporating “game theory” into SHAP (Štrumbelj & Kononenko, 2014), to interpret the output of any black-box model, such as most ML models by considering its inputs. This additive, consistent, and locally accurate approach is part of the explainable branch of artificial intelligence (AI) and describes the performance of any ML model (Lundberg & Lee, 2017). The SHAP method was used to interpret the applied ML methods and construct a simpler SSC model. The SHAP method treats all characteristics as "contributors" (Wang et al., 2022), and measures the importance of variables based on the Shapley values (Shapley, 1953). In effect, using the SHAP analysis, each predictive sample represents a predictive value, and the SHAP value is the value derived for each characteristic in the predictive sample (Wang et al., 2022). The primary advantage of the SHAP values lies in their ability to reflect the impact of features on each sample, illustrating both positive and negative effects (Molnar et al., 2020). For example, supposing an ML model is trained with the following input data: $X_i = \{x_1, x_2, \ldots x_n\}^t$ To evaluate the contribution of each input variable to the ML model, the SHAP method uses an explanatory model (EM). These are described in Eqs. (5) and (6):

$$\mathrm{EM} = \varphi_0 + \sum_{i=1}^{n} \varphi_i \, t_i \tag{5}$$

$$\varphi_i(\mathrm{ML}, x) = \sum_{t \subseteq x}^{n} \frac{|t|!(n - |t| - 1)!}{n!} [\mathrm{ML}(t) - \mathrm{ML}(t \backslash i)] \tag{6}$$

where n is the number of input variables, t_i is the abbreviation for the input variable (i). φ_0 represents the model output with all simplified inputs toggled off, $\varphi_i \in R$ is the contribution of the variable i to the ML model, t is the subset of the features used in the model that does not include the feature value i, ML() is the original prediction model, and \ is the difference set notation for set operations.

4. Results

4.1. Tuning process

In the current study, CatBoost, extra tree, XGBoost, and random forest were tuned to predict SSC using streamflow and rainfall datasets. First, the tuning process was done using trial-error procedures to determine the hyperparameters of the models. These models were validated for new datasets unseen during the tuning process. Second, the GP algorithm was applied to optimize these parameters. Then, the optimized ML models were retrained for the best-selected hyperparameters and evaluated through the simulation of SSC during the validation process. Table 3 lists the best

parameters and architectures of the developed ML models for predicting SSC at the studied monitoring stations.

4.2. Training performances

In this stage, daily SSC data from 2016 to 2020 were utilized for training. Table 4 lists the model performance involving NSE, RMSE, and *r*. It can be observed that all models yielded satisfactory accuracy at the four monitoring stations. Interestingly, NSE ranges from 0.55, as recorded by the XG model for predicting SSC at the Aguibat Ziar station, to 0.9, achieved by the CAT model for the Ain Loudah station. RMSE varies between 0.3 g/L for the ET model at Ras Fathia station and 1.31 g/L for the CAT model at the Ain Loudah station. The *r* ranges from 0.85 for XG at Aguibat Ziar to 0.96 for the CAT model at Sidi M^{ed} Cherif station. Overall, these results clearly show the superiority of ET over the other models at the four stations.

4.3. Validation performance

The comparison between the forgoing models and their combination with the GP model was done. Briefly, an example of a comparison of observed (blue lines) and predicted (green lines) SSC at the four stations is shown in Fig. 6 for the RF + GP model at Aguibat Ziar station, the XGBoost + GP model at Ain Loudah station, the CatBoost + GP model at Ras Fathia station, and the RF model at Sidi M^{ed} Cherif station. Table 5 lists the performance validation of the stand-alone and combined models. NSE, RMSE, and *r* range, respectively, from 0.54 to 0.75, 1.20–1.66 g/L, and 0.83 to 0.88 at Aguibat Ziar, with CAT + GP being the best model and XGBoost the worst. However, the XGBoost model combined with GP delivered a highly similar performance to CAT + GP. At Ain Loudah, NSE, RMSE, and *r* ranged from 0.82 to 0.86, 2.28–2.55 g/L, and 0.88 to 0.91, respectively. The optimal model was XG + GP and the poorest one was RF. The results ranged from 0.67 to 0.78, 1.79–2.20 g/L, and 0.84 to 0.90, respectively, at Ras Fathia. Finally, they varied from 0.75 to 0.80, 1.42–1.62 g/L, and 0.88 to 0.90, respectively, NSE, RMSE, and *r*. The best model for Sidi M^{ed} Cherif was RF and the poorest was CAT + GP, for the most part, the models combined with GP show better results than those for the stand-alone models. For the latter, the best model was ET at Aguibat Ziar, and the most effective model was XG at Ain Loudah. At Ras Fathia, the RF and CAT models yielded the best results. At Sidi M^{ed} Cherif, the RF model yielded the best results. Overall, the models were accurate and sufficiently generalizable to reproduce the daily SSC over the entire validation period.

Globally, except for the Ras Fathia station, it was observed that the ensemble models are more accurate than the standalone ones. Graphically, the bar plots shown presented in Fig. 7 illustrate the comparison of the models' performances at the studied stations. Fig. 7 demonstrates the superiority of the combined models for the stations Aguibat Ziar, Ain Loudah, and Sidi M^{ed} Cherif, excluding the station Ras Fathia. Thus, the combined models yielded high efficiencies for SSC prediction. Additionally, Fig. 8 shows the model error distributions and shows that the errors were evenly distributed over 0 for most models indicating their stability for predicting daily SSC. Excluding the Ain Loudah station, where all models did not exceed 0. On the other hand, the median value is equal between the stand-alone and combined models. However, the upper and lower values of the stand-alone models are larger than those of the combined models, demonstrating the enhanced performance of the combined models.

Table 4
Training performance for predicting daily SSC at 4 hydrological stations.

Station	Index	ET	RF	CAT	XG
Aguibat Ziar	NSE	0.78	0.65	0.67	0.55
	RMSE	0.50	0.80	0.80	0.98
	r	0.88	0.86	0.89	0.85
Ain Loudah	NSE	0.88	0.85	0.90	0.87
	RMSE	0.40	1.20	1.31	1.22
	r	0.95	0.91	0.93	0.90
Ras Fathia	NSE	0.82	0.81	0.79	0.78
	RMSE	0.30	0.70	0.95	1.05
	r	0.95	0.94	0.96	0.91
Sidi M^{ed} Cherif	NSE	0.84	0.87	0.81	0.79
	RMSE	0.45	0.60	1.10	1.01
	r	0.93	0.93	0.96	0.91

Note: NSE is the Nash–Sutcliff efficiency, RMSE (g/L) is the root mean square error, *r* is the coefficient of correlation, ET is the extra tree, RF is the random forest, CAT is the CATBoost, and XG is the XGBoost.

4.4. Model interpretability

For brevity, Fig. 9 shows the interpretability of the accurate standalone models such as ET for Aguibat Ziar, RF for Ras Fathia, XGBoost for Ain Loudah, and RF for Sidi M^{ed} Cherif. The ranking of the variables, as presented on the *y*-axis, indicates the importance of the input variables in predicting SSC. A feature is classified as important when the error increases after being omitted from the model, reflecting that the model has relied on this feature for prediction (Breiman, 2001). The SHAP values, as presented on the *x*-axis, are a unified index to address the influence of a given variable in the trained model. The colored bars in each row show more details on how each input variable affects the changes in SSC. Meanwhile, the red dot (blue) indicates that the value of the input variable is positive (negative). Hence, the SHAP summary plot shows the importance of the features and their influence on SSC. Each figure (Figs. 9(a)–9(d)) is a set of scatter plots arranged in order of importance. The *y*-axis refers to the names of the variables in descending order of importance, and the *x*-axis shows the SHAP values for each characteristic, ranked from lowest to highest. Each point symbolizes a sample in the data set, and its gradient color indicates the original value for that feature. For example, Fig. 9(a) shows 20 input parameters that impact the SSC using the ET model for Aguibat Ziar. From Fig. 9, it is observed that the flow (Q_{az}), flow with lag = 1 (Q_{az1}), flow with lag = 2 (Q_{az2}), followed by seasonality (S) are the input parameters having the most influence on SSC. Similarly, for the Ras Fathia station, 20 important input parameters impact SSC using the RF model.

Flow and seasonality are the features most impacting SSC. However, only 7 major characteristics affected SSC at Ain Loudah and Sidi M^{ed} Cherif (Figs. 9(c) and (9d)). With flow (Q_{al}), flow with lag = 1 (Q_{al1}), and seasonality (S) being the most influential input parameters on SSC at Ain Loudah, while the flow (Q_{sm}) and seasonality (S) were most influential on SSC at Sidi M^{ed} Cherif station. Overall, the flow (Q) was found to be the dominant variable affecting SSC, followed by seasonality at all stations. Furthermore, increased parameters are associated with increased SHAP values for almost all models (Fig. 9). Exceptionally, an increase in seasonality corresponds to a decline in SHAP values. This leads to a comprehensive representation of the physical process of SSC. These outcomes indicate that ET, RF, CatBoost, and XGBoost models can extract accurate and detailed information from the interactions between the hydrological factors and the generation of sediment by erosion (output).

Further details are provided in Fig. 10, which illustrates the range of values for each parameter with the corresponding SHAP value. Each point corresponds to a prediction and each dot represents a row of the data. The horizontal graph represents the measured value of the data set and the vertical graph shows the effect of this measured value on the prediction. Fig. 10(a) shows the dependence plot between the flow at Aguibat Ziar (Q_{az}) and Tsalat

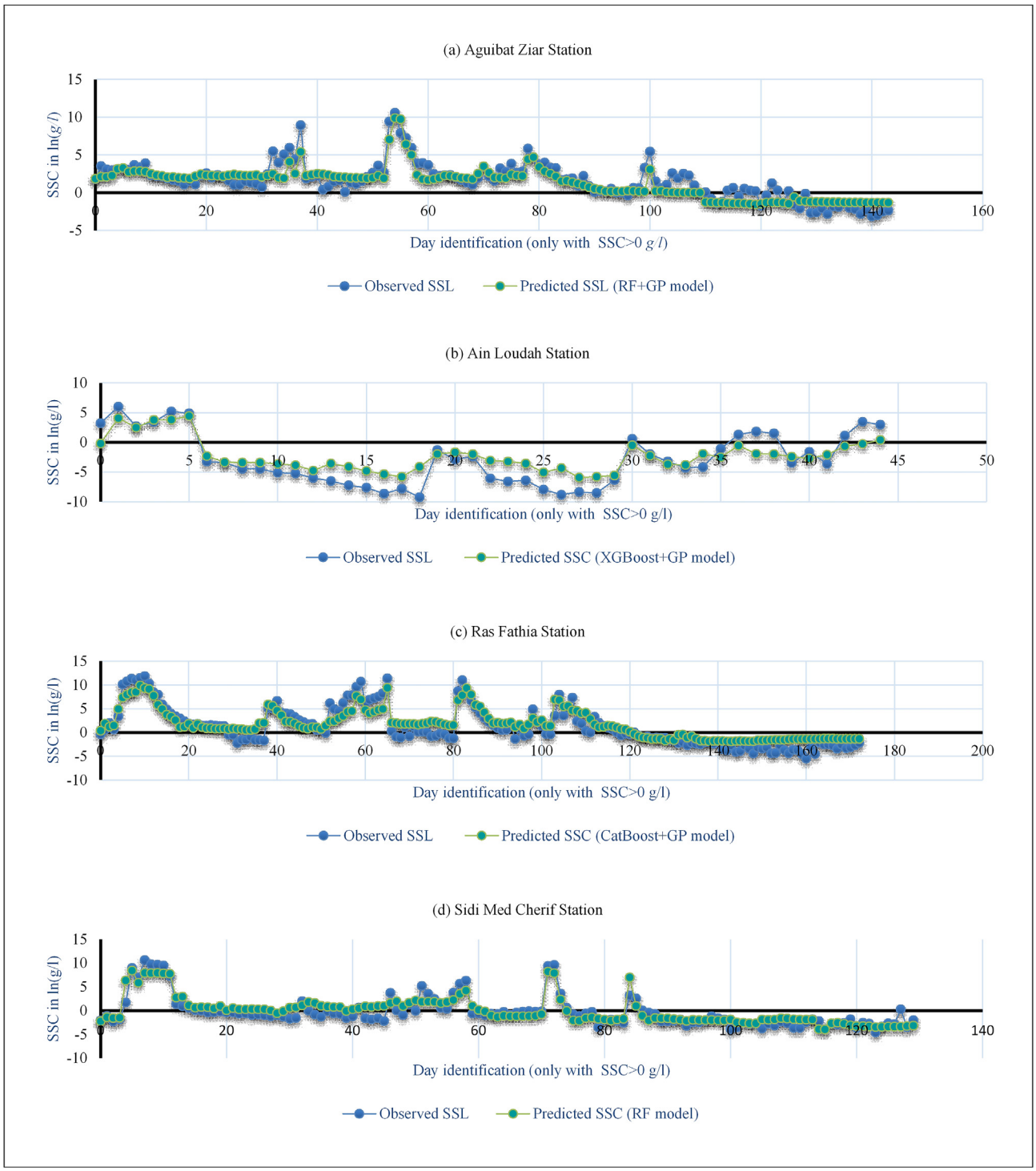

Fig. 6. Comparison of observed (blue lines) and predicted (green lines) SSC at four monitoring stations, by: RF + GP model at Aguibat Ziar, XGBoost + GP model at Ain Loudah, CatBoost + GP model at Ras Fathia, and Sidi M^{ed} Cherif.

(Q_{ts3}). The *x*-axis shows Q_{az}, and the *y*-axis shows the predicted SHAP value (how much does knowing the value of this feature change the model output for the prediction of this sample), and the color represents the value of Q_{ts3}. Therefore, it can be seen if having a specific Q_{az} and having Q_{ts3} positively or negatively impact the output. From Fig. 10(a), it can deduced that at lower values of Q_{az}, there is a wide range of SHAP values, with some being negative, indicating a decrease in the model's output, and others being positive, indicating an increase. As Q_{az} increases, the SHAP values stabilize and generally become positive, suggesting that higher Q_{az} values consistently increase the model's output. The color variation indicates that the SHAP value's impact is also influenced by Q_{ts}. For instance, at lower Q_{az} values, data points with higher Q_{ts} values (in red) tend to have higher SHAP values. Overall, the plot suggests that Q_{az} significantly impacts the model's output, particularly at higher values, and this impact is further influenced by the Q_{ts} feature.

Table 5
Validation results at 4 hydrological stations.

Station	Index	ET	ET + GP	RF	RF + GP	CAT	CAT + GP	XG	XG + GP
Aguibat Ziar	NSE	0.69	0.74	0.54	0.75	0.62	0.75	0.53	0.74
	RMSE	1.36	1.24	1.63	1.21	1.49	1.20	1.66	1.22
	r	0.86	0.86	0.84	0.87	0.85	0.88	0.83	0.87
Ain Loudah	NSE	0.83	0.85	0.82	0.84	0.84	0.86	0.85	0.86
	RMSE	2.52	2.38	2.55	2.39	2.39	2.31	2.38	2.28
	r	0.89	0.90	0.88	0.90	0.90	0.91	0.88	0.91
Ras Fathia	NSE	0.77	0.67	0.78	0.69	0.78	0.69	0.77	0.68
	RMSE	1.84	2.20	1.79	2.16	1.79	2.14	1.86	2.16
	r	0.90	0.84	0.90	0.85	0.90	0.86	0.89	0.85
Sidi M^{ed} Cherif	NSE	0.79	0.77	0.80	0.78	0.77	0.75	0.77	0.77
	RMSE	1.47	1.53	1.42	1.50	1.54	1.62	1.54	1.52
	r	0.89	0.89	0.90	0.90	0.90	0.89	0.88	0.89

Note: NSE is the Nash–Sutcliff efficiency, RMSE (g/L) is the root mean square error, *r* is the coefficient of correlation, ET is the extra tree, RF is the random forest, CAT is the CATBoost, XG is the XGBoost, and GP is the genetic programming.

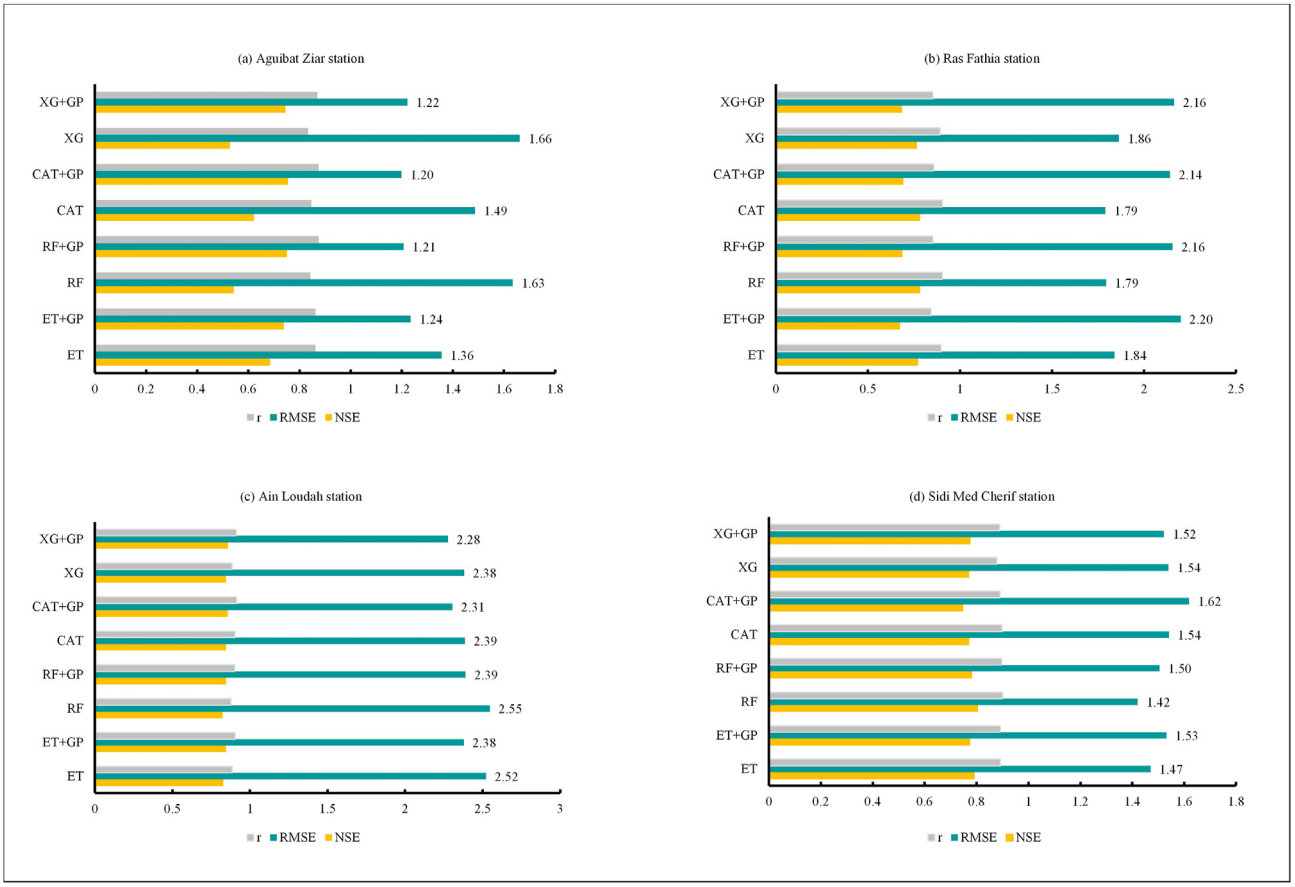

Fig. 7. Model performance comparison at studied stations considering RMSE, NSE, and *r*.

Similarly, Fig. 10(b) shows the SHAP plot showing the influence of Q_{rf} on the predicted SSC for the RF model at Ras Fathia. The color gradient represents Q_{rf1}, highlighting the effect of lag time on the model's predictions. Higher Q_{rf} values generally increase the SHAP values, indicating a positive contribution to SSC, particularly when Q_{rf1} values are also high. The same is true for Fig. 10(c) illustrating the interaction between Q_{al} and Q_{al1}, where there is little positive interaction between the two when $Q_{al} > 10$. Fig. 10(d) shows the plot of the effect of Q_{sm} at Sidi M^{ed} Cherif on the predicted SSC as a function of seasonality. However, in this figure, there are always positive values of seasonality, regardless of the value of Q_{sm}. For the large basins Bouregreg and Grou, the antecedent flow upstream was more important than the current flows. This is consistent with the physical erosion process where the sediment load's response to the flood from the upstream of the watershed is not instantaneous. Meanwhile, for the Korifla basin, the antecedent rainfall was more important than the current day's rainfall.

These figures show the distribution of all instances' SHAP values for each feature, with the SHAP value dots' color dependent on the feature values. The vertical axis is the list of input parameters whereas the horizontal axis shows the effects of the input parameters on SSC. The blue color refers to low characteristic values while the red color shows positive values. Q_{al}: flow at Ain Loudah, Q_{al1}: flow at Ain Loudah with lag = 1, Q_{al2}: flow at Ain Loudah with

Fig. 8. Box plots of error distributions for predicted SSC produced by different ML models (ET, RF, CAT, and XG) at four stations. The blue boxes represent predictions without GP, while the green boxes represent predictions with GP. Interquartile range (IQR) is shown by boxes, the median by the horizontal line within boxes, and the whiskers extend to the smallest and largest values within 1.5 times IQR. Outliers are shown as points beyond whiskers. This visual comparison highlights the impact of GP on variability and central tendency of predictions.

lag = 2, Q_{az}; flow at Aquibat Ziar, Q_{az1}: flow at Aquibat Ziar with lag = 1, Q_{az2}: flow at Aquibat Ziar with lag = 2, Q_{lc}: flow at Lalla Chafia, Q_{lc1}: flow at Lalla Chafia with lag = 1, Q_{lc2}: flow at Lalla Chafia with lag = 2, Q_{lc3}: flow at Lalla Chafia with lag = 3, Q_{oh}; flow at Ouljat Haboub, Q_{oh1}: flow at Ouljat Haboub with lag = 1, Q_{oh2}: flow at Ouljat Haboub with lag = 2, Q_{oh3}: flow at Ouljat Haboub with lag = 3, Q_{rf}: flow at Ras Fathia, Q_{rf1}: flow at Ras Fathia with lag = 1, Q_{rf2}: flow at Ras Fathia with lag = 2, Q_{sa}: flow at Sidi Aamar, Q_{sa1}: flow at Sidi Aamer with lag = 1, Q_{sa2}: flow at Sidi Aamer with lag = 2, Q_{sa3}: flow at Sidi Aamer with lag = 3, Q_{sj}: flow at Sidi Jabeur, Q_{sj1}: flow at Sidi Jabeur with lag = 1, Q_{sj2}: flow at Sidi Jabeur with lag = 2, Q_{sj3}: flow at Sidi Jabeur with lag = 3, Q_{sm}: flow at Sidi M^{ed} Cherif, Q_{sm1}: flow at Sidi M^{ed} Cherif with lag = 1, Q_{sm2}: flow at Sidi M^{ed} Cherif with lag = 2, Q_{ts}: flow at Tsalat, Q_{ts1}: flow at Tsalat with lag = 1, Q_{ts2}: flow at Tsalat with lag = 2, Q_{ts3}: flow at Tsalat with lag = 3, R_{al}: Rainfall at Ain Loudah, R_{al1}: Rainfall at Ain Loudah with lag = 1, R_{al2}: Rainfall at Ain Loudah with lag = 2, R_{az2}: rainfall at Aquibat Ziar with lag = 2, R_{oh}: Rainfall at Ouljat Haboub, R_{oh2}: Rainfall at Ouljat Haboub with lag = 2, R_{oh3}: Rainfall at Ouljat Haboub with lag = 3, R_{rf}: Rainfall at Ras Fathia, R_{rf1}: Rainfall at Ras Fathia with lag = 1, R_{rf2}: Rainfall at Ras Fathia with lag = 2, R_{sj1}: Rainfall at Sidi Jabeur with lag = 1, R_{sj3}: Rainfall at Sidi Jabeur with lag = 3, R_{sm}: Rainfall at Sidi M^{ed} Cherif, R_{sm1}: Rainfall at Sidi M^{ed} Cherif with lag = 1, R_{sm2}: Rainfall at Sidi M^{ed} Cherif with lag = 2, R_{ts1}: rainfall at Tsalat with lag = 1, R_{ts2}: rainfall at Tsalat with lag = 2, R_{ts3}: rainfall at Tsalat with lag = 3, S: seasonality.

5. Discussion

In this technological age, advances in data science have improved SSC modeling, predictive accuracy, and interpretation techniques (Lund et al., 2022). Using an appropriate model to estimate the SSL is crucial to determine the severity of soil erosion (Alewell et al., 2019; Benavidez et al., 2018) and to prioritize mitigation measures (Ezzaouini et al., 2022). ML algorithms have become powerful tools in modeling sediment yield (Al-Mukhtar, 2019; Jimeno-Sáez et al., 2022). However, they are less interpretable, which can lead to a lack of confidence (Pereira et al., 2018). The complexities of interpreting ML models and their predictions limit the practical applicability and confidence in ML model for the prediction of SSL (Rodríguez-Pérez & Bajorath, 2020). This is especially important as these systems are introduced ubiquitously in critical areas, such as SSC prediction. As ML methods have strengths and weaknesses (Wolpert, 1996), no single ML model is the best method for solving all engineering problems. Therefore, using multiple ML models to predict water quality may be better

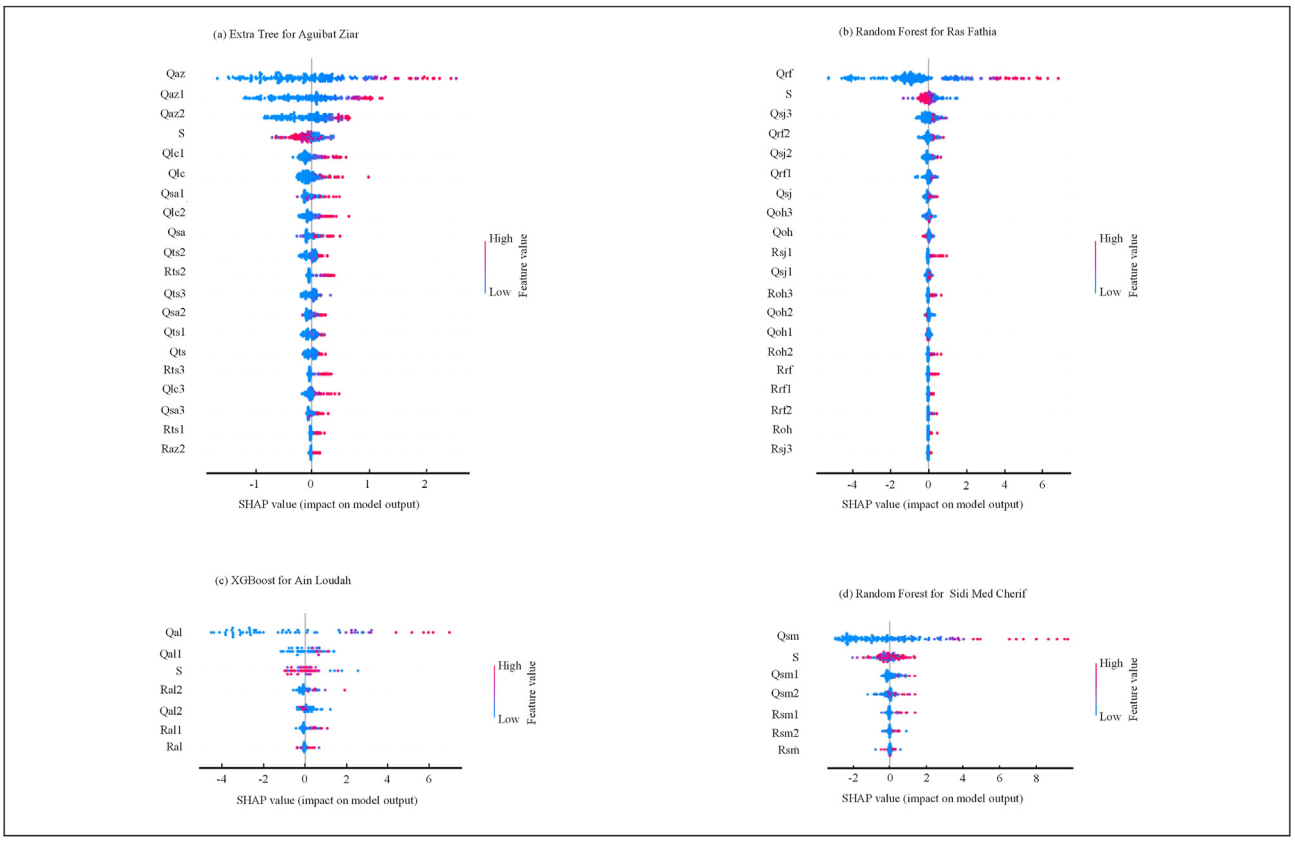

Fig. 9. Global interpretability plots of (a) extra tree, (b, d) random forest, and (c) XG boost models ranking input parameters.

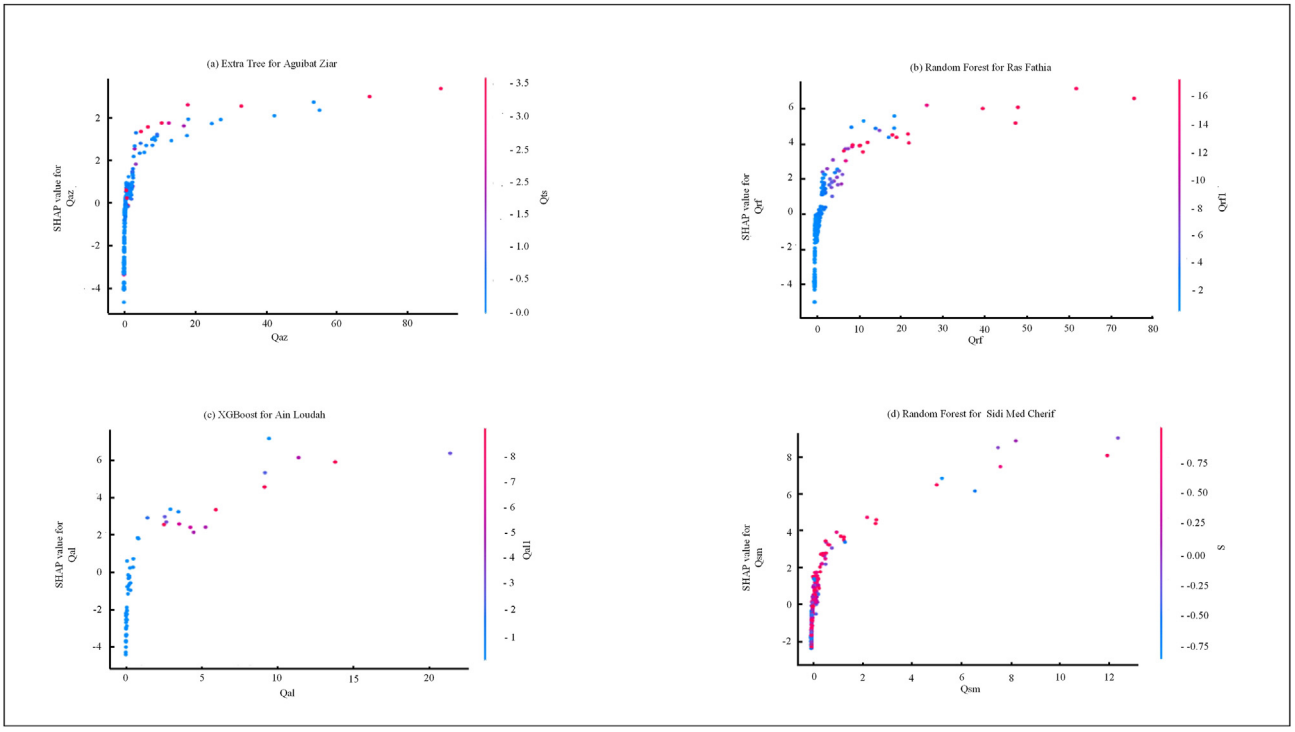

Fig. 10. SHAP dependence plot illustrates distribution of values for each parameter with the respective SHAP value. The horizontal graph depicts the value of actual data set, and the vertical graph represents the effect of this value on prediction.

than using a single model. For this reason, ET, RF, CatBoost, and XGBoost models are applied and combined with the GP model to determine the best model. The forecasting results indicate that all the stand-alone and combined ML models have an NSE greater than 0.5, which indicates that they are suitable for simulation purposes, given that an NSE >0.5 is required to accept the simulation of sediment load (Moriasi et al., 2007). Compared with the stand-alone models, the combined models showed a higher potential for improving forecast accuracy.

The explanatory results showed that the runoff and seasonality factors most affected SSC at all stations. Further, runoff has been confirmed as the most sensitive factor affecting the SSC (Li et al., 2017). Thus, by controlling runoff and sequestering sediment reservoirs and dams could significantly alter water and sediment conditions in rivers, leading to changes in sediment fluxes on land and to the ocean (Chen et al., 2012; Li et al., 2020; Miao et al., 2024). Furthermore, the current explanatory findings confirm the conclusions of other researchers that SSC fluctuates with seasonality, with the highest concentrations occurring in summer and the lowest in winter (Murphy & Voulgaris, 2006). Climate and land use changes can explain the strong contribution of runoff and seasonality on erosion and sediment transport processes (Miao et al., 2024). Interestingly, these factors were already depicted in studying the seasonality of soil erosion under Mediterranean conditions using the empirical model the revised universal soil loss equation (RUSLE) (Ferreira & Panagopoulos, 2014) in Alentejo, Portugal. Physically, the importance of the seasonality factor is justified by the fact that land cover changes over the seasons by the dynamic process of the vegetation growth, especially for Mediterranean watersheds (Panagos et al., 2014). Accordingly, during the autumn, the Mediterranean watershed is characterized by low vegetation and flash flood events, leading to severe erosion (Chinnasamy et al., 2020; Diodato & Bellocchi, 2019). All these issues the need to implement appropriate strategies for preserving environmental streamflow, considering each characteristic of rivers, especially in BW which will be affected by climate change by the end of this century (Brouziyne et al., 2021).

In this context, it is recommended to apply large-scale water conservation projects and soil and water conservation measures, such as dams, enclosures, and grassing, to all parts of watersheds to control SSL, which is essential for sustainable ecological protection and high-quality development over the watershed. To effectively control sediment in the stream course, it is key to apply soil and water conservation management measures in channels, such as buckle and check dams. In addition, the current study confirmed the ability of the modeling approach developed to monitor suspended sediment factors for the first time in Morocco and further highlighted the necessity of proactive strategies to sustain the long-term stream ecosystems in North African watersheds.

Even though the current results make an interesting contribution to the SSC modeling, it is essential to recognize the inherent limitations of the current research to provide insight and guidance for future research. Thus, the current study is limited by the omission of SSC from the intermediate basin and the erosion of the banks of the SMBA reservoir in the current analysis (Ezzaouini et al., 2020). In addition, the ML model is based on input-output statistical equations rather than explaining the mechanisms underlying the process. Furthermore, although this is an accurate approach for estimating the SSC upstream of the SMBA reservoir, the ML models cannot simulate best management practices (BMPs) at the basin scale. This limitation arises due to the absence of essential physical parameters, such as land use, soil type, and anthropogenic activities specific to the studied basin, in the input characteristics. Furthermore, the limited data set may not reflect all the complex factors that significantly impact predictive accuracy.

The data used in this study also were collected in restricted locations. The geographic constraint may impede the relevance of the findings in regions with diverse environmental and infrastructural attributes, potentially neglecting an entire set condition. Future research could overcome these limitations by delving into a broader dataset, integrating more variables, incorporating data from various geographical areas, and exploring alternative modeling techniques.

The absence of addressing uncertainty in predictive models because of insufficient data presents another limitation. Integrating uncertainty considerations into predictive models improves their relevance for risk assessment and decision-making processes (Stødle et al., 2023). Suggesting Bayesian methods or Monte Carlo simulations will offer a more nuanced grasp of uncertainty, enhancing model reliability and providing decision-makers with a more robust foundation for risk assessment (Taiwo et al., 2024).

Future studies also can be extended to explore the effects of spatial variability of soil erosion and suspended sediment over the watershed and to apply more relevant input variables to SSC forecasting, such as soil data to mitigate uncertainties in the simulation process. The current applicability of various ML models in predicting SSC and doing attribution analysis in a semi-arid watershed was constrained due to limited inputs. Thus, the spatial feature variation of the basin cannot be demonstrated well. Even though ML models have structures and controllable input data dimensions accelerating the modeling process, their ability to accurately reflect the real process is relatively weak. Therefore, it is recommended to investigate the impact of spatio-temporal variability on SSC assessment in the future. Incorporating additional input data, such as daily climate data (precipitation, temperature (minimum & maximum), relative humidity, wind, solar radiation), hydrological data (flow, suspended solids (sediment), quality data (nutrients: nitrates & phosphorus), high spatial resolution digital elevation model (DEM), land use map, soil map with physicochemical characteristics of each soil type, could provide more accurate predictions. The foregoing constraints could be addressed through the integration of ML and process-based models. Therefore, it may be suggested to study this topic in the future. Interestingly, spatial ML models integrated with GIS tools could be a fruitful approach to simulate BMPs to reduce soil erosion in basins. In addition, it is recommended that the proposed ML models be twinned with a river-based model to assess the sediment input to the SMBA reservoir to improve accuracy, considering that the stations studied are upstream of the reservoir (Ezzaouini et al., 2022; Tadesse & Dai, 2019).

The scope and value of the current study can be extended by applying improved ML models and their attribution frameworks to different spatial settings with consistently reliable, high-precision datasets. In the future, increasingly improved ML techniques will provide powerful tools for exploring and predicting changes in the suspended sediment for different watersheds worldwide. The foregoing recommendations aim to enhance suspended sediment modeling and increase its real-world relevance. The explanatory results of proposed framework provide an important reference for future watershed management, and future work could develop new interpretation methods adapted to spatial data and models, to improve understanding of complex spatial relations.

6. Conclusions

In this manuscript, four ML models, extra trees, random forest, CatBoost, and XGBoost were trained and validated using structured tabular hydrological datasets to predict daily suspended sediment concentrations. Then, the models were combined with genetic programming and coupled with the SHAP method to improve the

prediction accuracy and assess their interpretability. From the findings of the current study, the main conclusions are:

i. The developed models presented acceptable performance for predicting daily SSC during the whole year of 2021;
ii. Compared with the standalone-based models, the average ensemble ML based on genetic programming proved to be a potential approach to improve the prediction accuracy of the models;
iii. The interpretability of the models using the SHAP method indicated a consistency of the ML to predict SSC. The streamflow is the most important factor for all studied basins. This improves the transparency of ML models by determining important factors used as input features;
iv. Seasonality also was among the important input variables for predicting daily suspended sediment concentration.

To sum up, the outcomes of the current research, it can be concluded that the proposed method is an efficacious tool for improving the transparency of the applied ML models, such as the XGBoost, CatBoost, RF, and ET models used in predicting daily SSC in the Bouregreg basin. Globally, the proposed approach is valuable for interpreting ML models in other hydrological and environmental sciences. Despite the consistency of the developed ML models with the physical process of soil erosion, these models cannot simulate the processes involved. For instance, they cannot simulate the physical actions to reduce soil erosion from watersheds such as best practices management because they are not trained with a physical dataset related to characteristics of the studied watersheds. Consequently, using these factors for training ML models in predicting SSC is suggested for future research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Houda Lamane: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Latifa Mouhir:** Supervision. **Rachid Moussadek:** Supervision. **Bouamar Baghdad:** Supervision. **Ozgur Kisi:** Writing – review & editing. **Ali El Bilali:** Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Methodology.

Acknowledgements

This research was supported by the River Basin Agency of Bouregrge and Chaouia (ABHBC) in Morocco by providing the study dataset. So, the authors thank the ABHBC teams for their help and support. The authors would like to thank the Bouregreg and Chaouia basin agency for its help in collecting the data, as there was no funding or subsidy provided by this agency. Also, the authors thank the editor and three anonymous reviewers for their constructive comments during the review phase.

Nomenclature

ABR	AdaBoost regression
ANFIS	Adaptive neuro fuzzy inference system
ANN	Artificial neural network
BMPs	Best management practices
BW	Bouregreg watershed
CatBoost	Categorical boosting
CART	Classification and regression tree
DEM	Digital elevation model
DNN	Deep neural network
DT	Decision tree
EM	Explanatory model
ET	Extra tree
GBR	Gradient boosting regression
GP	Genetic programming
K-NN	K-nearest neighbor method
LIME	Local interpretable model-agnostic explanations
M5P	M5 pruning tree
ML	Machine learning
MUSLE	Modified universal soil loss equation
NF	Neuro-fuzzy model
NSE	Nash–Sutcliffe model-fit efficiency
r	Coefficient of correlation
RBABC	River Basin Agency of Bouregreg and Chaouia
RBF	Radial basis function
RF	Random forest
RFR	Random forest regression
RMSE	Root mean square error
RUSLE	Revised universal soil loss equation
S	Seasonality
SHAP	SHapely additive exPlanations
SMBA	Dam Sidi Mohammed Ben Abdellah Dam
SSC	Suspended sediment concentration
SSL	Suspended sediment load
STD	Standard deviation
SVM	Support vector machine
SVR	Support vector regression
XGBoost	Extreme gradient boosting

References

Al-Mukhtar, M. (2019). Random forest, support vector machine, and neural networks to modelling suspended sediment in Tigris River-Baghdad. *Environmental Monitoring and Assessment, 191*(11), 673.
Alewell, C., Borrelli, P., Meusburger, K., & Panagos, P. (2019). Using the USLE: Chances, challenges and limitations of soil erosion modelling. *International Soil and Water Conservation Research, 7*(3), 203–225.
Asadi, H., Dastorani, M. T., Khosravi, K., & Sidle, R. C. (2022). Applying the C-Factor of the RUSLE model to improve the prediction of suspended sediment concentration using smart data-driven models. *Water, 14*(19), 19.
Aytek, A., & Kişi, Ö. (2008). A genetic programming approach to suspended sediment modelling. *Journal of Hydrology, 351*(3), 288–298.
Benavidez, R., Jackson, B., Maxwell, D., & Norton, K. (2018). A review of the (Revised) Universal Soil Loss Equation ((R)USLE): With a view to increasing its global applicability and improving soil loss estimates. *Hydrology and Earth System Sciences, 22*(11), 6059–6086.
Borrelli, P., Alewell, C., Alvarez, P., Anache, J. A. A., Baartman, J., Ballabio, C., Bezak, N., Biddoccu, M., Cerdà, A., Chalise, D., Chen, S., Chen, W., De Girolamo, A. M., Gessesse, G. D., Deumlich, D., Diodato, N., Efthimiou, N., Erpul, G., & Fiener, P. (2021). Soil erosion modelling: A global review and statistical analysis. *Science of the Total Environment, 78*, 146494.
Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
Brouziyne, Y., De Girolamo, A. M., Aboubdillah, A., Benaabidate, L., Bouchaou, L., & Chehbouni, A. (2021). Modeling alterations in flow regimes under changing climate in a Mediterranean watershed: An analysis of ecologically-relevant hydrological indicators. *Ecological Informatics, 61*, 101219.
Bussi, G., Francés, F., Montoya, J. J., & Julien, P. Y. (2014). Distributed sediment yield modelling: Importance of initial sediment conditions. *Environmental Modelling & Software, 58*, 58–70.
Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics, 8*(8), 8.
Chemura, A., Rwasoka, D., Mutanga, O., Dube, T., & Mushore, T. (2020). The impact of land-use/land cover changes on water balance of the heterogeneous Buzi sub-catchment, Zimbabwe. *Remote Sensing Applications: Society and Environment, 18*, 100292.
Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings, (pp. 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 785–794)*. New York: Association for Computing Machinery.

Chen, J., Zhou, W., & Chen, Q. (2012). Reservoir sedimentation and transformation of morpho-logy in the Lower Yellow River during 10 year's initial operation of the Xiaolangdi Reservoir. *Journal of Hydrodynamics, 24*(6), 914–924.

Chinnasamy, P., Honap, V. U., & Maske, A. B. (2020). Impact of 2018 Kerala Floods on soil erosion: Need for post-disaster soil management. *Journal of the Indian Society of Remote Sensing, 48*(10), 1373–1388.

Choubin, B., Darabi, H., Rahmati, O., Sajedi-Hosseini, F., & Kløve, B. (2018). River suspended sediment modelling using the CART model: A comparative study of machine learning techniques. *Science of the Total Environment, 615*, 272–281.

Ding, Y., Zhu, Y., Feng, J., Zhang, P., & Cheng, Z. (2020). Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing, 403*, 348–359.

Diodato, N., & Bellocchi, G. (2019). Reconstruction of seasonal net erosion in a Mediterranean landscape (Alento River basin, Southern Italy) over the past five decades. *Water, 11*(11), 2306.

El Aoula, R., Mhammdi, N., Dezileau, L., Mahe, G., & Kolker, A. S. (2021). Fluvial sediment transport degradation after dam construction in North Africa. *Journal of African Earth Sciences, 182*, 104255.

El Bilali, A., Abdeslam, T., Ayoub, N., Lamane, H., Ezzaouini, M. A., & Elbeltagi, A. (2023). An interpretable machine learning approach based on DNN, SVR, Extra Tree, and XGBoost models for predicting daily pan evaporation. *Journal of Environmental Management, 327*, 116890.

El Bilali, A., Moukhliss, M., Taleb, A., Nafii, A., Alabjah, B., Brouziyne, Y., Mazigh, N., Teznine, K., & Madark, M. (2022). Predicting daily pore water pressure in embankment dam: Empowering Machine Learning-based modeling. *Environmental Science and Pollution Research, 29*, 1–17.

El Bilali, A., & Taleb, A. (2020). Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment. *Journal of the Saudi Society of Agricultural Sciences, 19*(7).

El Bilali, A., Taleb, A., El Idrissi, B., Brouziyne, Y., & Mazigh, N. (2020). Comparison of a data-based model and a soil erosion model coupled with multiple linear regression for the prediction of reservoir sedimentation in a semi-arid environment. *Euro-Mediterranean Journal for Environmental Integration, 5*(3), 1–13.

Ellison, C. A., Groten, J. T., Lorenz, D. L., & Koller, K. S. (2016). *Application of dimensionless sediment rating curves to predict suspended-sediment concentrations, bedload, and annual sediment loads for rivers in Minnesota*. U.S. Geological Survey Scientific Investigations Report 2016–5146.

Essam, Y., Huang, Y., Birima, A., Najah, A.-M., & El-Shafie, A. (2022). Predicting suspended sediment load in Peninsular Malaysia using support vector machine and deep learning algorithms. *Scientific Reports, 12*, 302.

Ezzaouini, M. A., Mahé, G., Kacimi, I., El Bilali, A., Zerouali, A., & Nafii, A. (2022). Predicting daily suspended sediment load using machine learning and NARX hydro-climatic inputs in semi-arid environment. *Water, 14*(6), 862.

Ezzaouini, M., Mahé, G., Kacimi, I., & Zerouali, A. (2020). Comparison of the MUSLE model and two years of solid transport measurement, in the Bouregreg Basin, and impact on the sedimentation in the Sidi Mohamed Ben Abdellah Reservoir, Morocco. *Water, 12*(7), 1882.

Ferreira, V., & Panagopoulos, T. (2014). Seasonality of soil erosion under Mediterranean conditions at the Alqueva Dam watershed. *Environmental Management, 54*(1), 67–83.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., & Wolff, E. (2018). Very high-resolution object-based land use–land cover urban classification using extreme gradient boosting. *IEEE Geoscience and Remote Sensing Letters, 15*(4), 607–611. (IEEE Geoscience and Remote Sensing Letters)

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*(1), 3–42.

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: An interdisciplinary review. *Journal of Big Data, 7*(1), 94.

Hanoon, M. S., Abdullatif, B. A. A., Ahmed, A. N., Razzaq, A., Birima, A. H., & El-Shafie, A. (2022). A comparison of various machine learning approaches performance for prediction suspended sediment load of river systems: A case study in Malaysia. *Earth Science Informatics, 15*(1), 91–104.

Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., & Zhou, H. (2019). Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology, 574*, 1029–1041.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688.

Ji, H., Chen, Y., Fang, G., Li, Z., Duan, W., & Zhang, Q. (2021). Adaptability of machine learning methods and hydrological models to discharge simulations in data-sparse glaciated watersheds. *Journal of Arid Land, 13*(6), 549–567.

Jimeno-Sáez, P., Martínez-España, R., Casalí, J., Pérez-Sánchez, J., & Senent-Aparicio, J. (2022). A comparison of performance of SWAT and machine learning models for predicting sediment load in a forested Basin, Northern Spain. *Catena, 212*, 105953.

John, V., Liu, Z., Guo, C., Mita, S., & Kidono, K. (2016). Real-time lane estimation using deep features and extra trees regression. In T. Bräunl, B. McCane, M. Rivera, & X. Yu (Eds.), *Image and video technology* (pp. 721–733). Cham: Springer International Publishing.

Kaveh, K., Duc Bui, M., & 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.

Khankhoje, T., & Choudhury, P. (2023). River system sediment flow modeling using artificial neural networks. *International Journal of Sediment Research, 39*(2), 222–229.

Khosravi, K., Cooper, J. R., Daggupati, P., Thai Pham, B., & Tien Bui, D. (2020). Bedload transport rate prediction: Application of novel hybrid data mining techniques. *Journal of Hydrology, 585*, 124774.

Kisi, O. (2005). Suspended sediment estimation using neuro-fuzzy and neural network approaches/Estimation des matières en suspension par des approches neurofloues et à base de réseau de neurones. *Hydrological Sciences Journal, 50*(4), 683–696.

Kisi, O., Dailr, A. H., Cimen, M., & Shiri, J. (2012). Suspended sediment modeling using genetic programming and soft computing techniques. *Journal of Hydrology, 450–451*, 48–58.

Kolyshkina, I., & Simoff, S. (2021). Interpretability of machine learning solutions in public healthcare: The CRISP-ML Approach. *Frontiers in Big Data, 4*, 660206.

Koza, J. R. (1994). Genetic programming as a means for programming computers by natural selection. *Statistics and Computing, 4*(2), 87–112.

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences, 5*, 89–97.

Krishnan, M. (2020). Against interpretability: A critical examination of the interpretability problem in machine learning. *Philosophy & Technology, 33*(3), 487–502.

Labbaci, A., Moukrim, S., Lahssini, S., Laaribya, S., Alaoui, H. M., & Hallam, J. (2021). Estimation of land degradation loss by water erosion: Case of the site of biological and ecological interest of Ain Asmama (Western High Atlas, Morocco). *Advances in Science, Technology and Engineering Systems Journal, 6*(3), 241–247.

Lahlou, A. (1986). *Etude actualisée de l'envasement des barrages au Maroc*. http://www.abhatoo.net.ma/maalama-archives/archives-textuelles1/developpement-economique-et-social/developpement-economique/equipements-et-infrastructures/barrages/etude-actualisee-de-l-envasement-des-barrages-au-maroc. *(in french)*

Lamane, H., Mouhir, L., Moussadek, R., Baghdad, B., Briak, H., Zouahri, A., & Bilali, A. E. (2023). Statistical analysis of a systematic review on soil water erosion assessment in Morocco. In *Proceedings, EGU general assembly 2023, (PP-EGU23-9800). 23-28, 2023, Vienna, Austria*.

Lamane, H., Moussadek, R., Baghdad, B., Mouhir, L., Briak, H., Laghlimi, M., & Zouahri, A. (2022). Soil water erosion assessment in Morocco through modeling and fingerprinting applications: A review. *Heliyon, 8*(8), e10209.

Li, T., Wang, S., Liu, Y., Fu, B., & Zhao, W. (2020). A retrospective analysis on changes in sediment flux in the Mississippi River system: Trends, driving forces, and implications. *Journal of Soils and Sediments, 20*(3), 1719–1729.

Li, G., Zheng, T., Fu, Y., Li, B., & Zhang, T. (2017). Soil detachment and transport under the combined action of rainfall and runoff energy on shallow overland flow. *Journal of Mountain Science, 14*(7), 1373–1383.

Liu, X., Qi, S., Huang, Y., Chen, Y., & Du, P. (2015). Predictive modeling in sediment transportation across multiple spatial scales in the Jialing River basin of China. *International Journal of Sediment Research, 30*(3), 250–255.

Lund, J. W., Groten, J. T., Karwan, D. L., & Babcock, C. (2022). Using machine learning to improve predictions and provide insight into fluvial sediment transport. *Hydrological Processes, 36*(8), e14648.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*.

Merritt, W. S., Letcher, R. A., & Jakeman, A. J. (2003). A review of erosion and sediment transport models. *Environmental Modelling & Software, 18*(8–9), 761–799.

Mesfin, S., Mullu, A., & Kassie, K. (2021). Micro-watershed hydrological monitoring and evaluation. A case study at Lake Tana sub-basin, Ethiopia. In A. M. Melesse, W. Abtew, & S. A. Moges (Eds.), *Nile and grand Ethiopian renaissance dam: Past, present and future* (pp. 493–517). Cham: Springer International Publishing.

Miao, J., Zhang, X., Zhang, G., Wei, T., Zhao, Y., Ma, W., Chen, Y., Li, Y., & Wang, Y. (2024). Applications and interpretations of different machine learning models in runoff and sediment discharge simulations. *Catena, 238*, 107848.

Molnar, C., Casalicchio, G., & Bischl, B. (2020). Interpretable machine learning – a brief history, State-of-the-Art and Challenges. In I. Koprinska, M. Kamp, A. Appice, C. Loglisci, L. Antonie, A. Zimmermann, R. Guidotti, Ö. Özgöbek, R. P. Ribeiro, R. Gavaldà, J. Gama, L. Adilova, Y. Krishnamurthy, P. M. Ferreira, D. Malerba, I. Medeiros, M. Ceci, G. Manco, & E. Masciari (Eds.), *ECML PKDD 2020 workshops* (pp. 417–431). Ghent: Springer International Publishing.

Moriasi, D. N., Arnold, J. G., van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE, 50*(3), 885–900.

Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time series missing value imputation in R. *The R Journal, 9*(1), 207.

Murphy, S., & Voulgaris, G. (2006). Identifying the role of tides, rainfall and seasonality in marsh sedimentation using long-term suspended sediment concentration data. *Marine Geology, 227*(1), 31–50.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — a discussion of principles. *Journal of Hydrology, 10*(3), 282–290.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research, 57*(3), e2020WR028091.

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.

Panagos, P., Meusburger, K., Ballabio, C., Borrelli, P., & Alewell, C. (2014). Soil erodibility in europe: A high-resolution dataset based on LUCAS. *Science of the Total Environment, 479–480*, 189–200.

Pandey, A., Himanshu, S. K., Mishra, S. K., & Singh, V. P. (2016). Physically based soil erosion and sediment yield models revisited. *Catena, 147*, 595–620.

Pereira, S., Meier, R., McKinley, R., Wiest, R., Alves, V., Silva, C. A., & Reyes, M. (2018). Enhancing interpretability of automatically extracted machine learning features: Application to a RBM-Random Forest system on brain lesion segmentation. *Medical Image Analysis, 44*, 228–244.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems, 31*.

Rhouichi, H. (1996). *Processus et facteurs d'érosion dans le bassin versant du Boureggreg (s. S.) : Approche sédimentologique pour la quantification de l'érosion. (Ph.D. dissertation)*. Rabat (Morocco): University Mohamed V.

Rodríguez-Pérez, R., & Bajorath, J. (2020). Interpretation of machine learning models using Shapley values: Application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design, 34*(10), 1013–1026.

Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems, 41*(3), 647–665.

Schmidt, L., Heße, F., Attinger, S., & Kumar, R. (2020). Challenges in applying machine learning models for hydrological inference: A case study for flooding events across Germany. *Water Resources Research, 56*(5), e2019WR025924.

Schulz, K., & Gerkema, T. (2017). An inversion of the estuarine circulation by sluice water discharge and its impact on suspended sediment transport. *Estuarine, Coastal and Shelf Science, 200*, 31–40.

Scornet, E. (2017). Tuning parameters in random forests. *ESAIM: Proceedings and Surveys, 60*, 144–162.

Shapley, L. S. (1953). 17. A value for n-person games. In *In contribution to the theory of games volume II* (pp. 307–318). Princeton, NJ, U.S.: Princeton University Press.

Sharafati, A., Haji Seyed Asadollah, S. B., Motta, D., & Yaseen, Z. M. (2020). Application of newly developed ensemble machine learning models for daily suspended sediment load prediction and related uncertainty analysis. *Hydrological Sciences Journal, 65*(12), 2022–2042.

Solomatine, D., See, L. M., & Abrahart, R. J. (2008). Data-driven modelling: Concepts, approaches and experiences. In R. J. Abrahart, L. M. See, & D. P. Solomatine (Eds.), *Practical hydroinformatics: Computational intelligence and technological developments in water applications* (pp. 17–30). Berlin: Springer.

Stødle, K., Flage, R., Guikema, S. D., & Aven, T. (2023). Data-driven predictive modeling in risk assessment: Challenges and directions for proper uncertainty representation. *Risk Analysis, 43*(12), 2644–2658.

Sun, A. Y., & Scanlon, B. R. (2019). How can big data and machine learning benefit environment and water management: A survey of methods, applications, and future directions. *Environmental Research Letters, 14*(7), 073001.

Syvitski, J. P. M., Peckham, S. D., Hilberman, R., & Mulder, T. (2003). Predicting the terrestrial flux of sediment to the global ocean: A planetary perspective. *Sedimentary Geology, 162*(1), 5–24.

Tadesse, A., & Dai, W. (2019). Prediction of sedimentation in reservoirs by combining catchment based model and stream based model with limited data. *International Journal of Sediment Research, 34*(1), 27–37.

Taiwo, R., Yussif, A.-M., Ben Seghier, M. E. A., & Zayed, T. (2024). Explainable ensemble models for predicting wall thickness loss of water pipes. *Ain Shams Engineering Journal, 15*(4), 102630.

Tan, Z., Leung, L. R., Li, H.-Y., & Tesfa, T. (2018). Modeling sediment yield in land surface and earth system models: Model comparison, development, and evaluation. *Journal of Advances in Modeling Earth Systems, 10*(9), 2192–2213.

de Vente, J., Poesen, J., Verstraeten, G., Govers, G., Vanmaercke, M., Van Rompaey, A., … Boix-Fayos, C. (2013). Predicting soil erosion and sediment yield at regional scales: Where do we stand? *Earth-Science Reviews, 127*, 16–29.

Viana, C. M., Santos, M., Freire, D., Abrantes, P., & Rocha, J. (2021). Evaluation of the factors explaining the use of agricultural land: A machine learning and model-agnostic approach. *Ecological Indicators, 131*, 108200.

Wang, S., Peng, H., Hu, Q., & Jiang, M. (2022). Analysis of runoff generation driving factors based on hydrological model and interpretable machine learning method. *Journal of Hydrology: Regional Studies, 42*, 101139.

Wang, S., Peng, H., & Liang, S. (2022). Prediction of estuarine water quality using interpretable machine learning approach. *Journal of Hydrology, 605*, 127320.

Wang, H.-L., & Yin, Z.-Y. (2020). High performance prediction of soil compaction parameters using multi expression programming. *Engineering Geology, 276*, 105758.

Williams, J. R., & Berndt, H. D. (1977). Sediment yield prediction based on watershed hydrology. *Transactions of the ASAE, 20*(6), 1100–1104.

Wischmeier, W. H., & Smith, D. D. (1965). Predicting rainfall-erosion losses from cropland east of the rocky mountains: Guide for selection of practices for soil and water conservation. In *Agricultural handbook No. 282*. Washington, DC: U.S. Department of Agriculture.

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation, 8*(7), 1341–1390.

Yang, C. T. (1977). The movement of sediment in rivers. *Geophysical Surveys, 3*(1), 39–68.

Zhang, Z., Huang, J., Duan, S., Huang, Y., Cai, J., & Bian, J. (2022). Use of interpretable machine learning to identify the factors influencing the nonlinear linkage between land use and river water quality in the Chesapeake Bay watershed. *Ecological Indicators, 140*, 108977.

Zounemat-Kermani, M., Mahdavi-Meymand, A., Alizamir, M., Adarsh, S., & Yaseen, Z. M. (2020). On the complexities of sediment load modeling using integrative machine learning: Application of the great river of Loíza in Puerto Rico. *Journal of Hydrology, 585*, 124759.