Science of the Total Environment 639 (2018) 826–840

Contents lists available at ScienceDirect

Science of the Total Environment

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

Estimating suspended sediment load with multivariate adaptive regression spline, teaching-learning based optimization, and artificial bee colony models

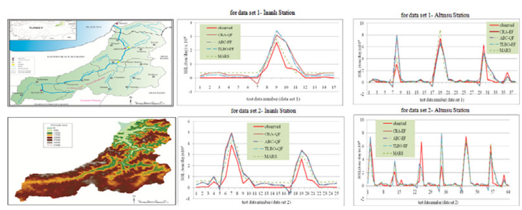
Banu Yilmaz [a], Egemen Aras [a,*], Sinan Nacar [b], Murat Kankal [c]

[a] Karadeniz Technical University, Faculty of Technology, Department of Civil Engineering, Trabzon, Turkey
[b] Karadeniz Technical University, Faculty of Engineering, Department of Civil Engineering, Trabzon, Turkey
[c] Uludağ University, Faculty of Engineering, Department of Civil Engineering, Bursa, Turkey

HIGHLIGHTS

- Implementation of different regression models to estimate SSL in Çoruh River Basin.
- MARS, TLBO, ABC and CRA techniques were developed for two different stations.
- Two different data sets were used to reinforce the validity of model successes.
- Performance of MARS's method is more successful than the others in SSL prediction.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 23 January 2018
Received in revised form 20 April 2018
Accepted 13 May 2018
Available online xxxx

Editor: D. Barcelo

Keywords:
Çoruh River Basin
Heuristic regression
Optimization algorithm
Reservoir life

ABSTRACT

The functional life of a dam is often determined by the rate of sediment delivery to its reservoir. Therefore, an accurate estimate of the sediment load in rivers with dams is essential for designing and predicting a dam's useful lifespan. The most credible method is direct measurements of sediment input, but this can be very costly and it cannot always be implemented at all gauging stations. In this study, we tested various regression models to estimate suspended sediment load (SSL) at two gauging stations on the Çoruh River in Turkey, including artificial bee colony (ABC), teaching-learning-based optimization algorithm (TLBO), and multivariate adaptive regression splines (MARS). These models were also compared with one another and with classical regression analyses (CRA). Streamflow values and previously collected data of SSL were used as model inputs with predicted SSL data as output. Two different training and testing dataset configurations were used to reinforce the model accuracy. For the MARS method, the root mean square error value was found to range between 35% and 39% for the test two gauging stations, which was lower than errors for other models. Error values were even lower (7% to 15%) using another dataset. Our results indicate that simultaneous measurements of streamflow with SSL provide the most effective parameter for obtaining accurate predictive models and that MARS is the most accurate model for predicting SSL.

1. Introduction

Although sediment transported in rivers ultimately originates in their drainage basins (via runoff), some the sediment transported in streams is a re-suspension of sediments from their stream beds and erosion of channel banks. Water and wind energy initiate runoff by

* Corresponding author.
E-mail address: egemen@ktu.edu.tr (E. Aras).

https://doi.org/10.1016/j.scitotenv.2018.05.153
0048-9697/

dislodging particles on land, but once in the channel, water energy moves sediments downstream. Due to erosion in both the drainage basin and channels, materials suspended in water constantly change the shape of the channel, causing sediment to scour (erode) from some sections and accumulate (deposit) in other sections. Although channel erosion is a natural feature of a river, increased runoff (e.g., due to increases in impervious surfaces or loss of vegetation in its watershed) often leads to a variety of undesirable effects, such as more frequent and destructive flooding, increased suspended loads, deeper channel incision, and more rapid infilling of reservoirs. These changes can negatively impact infrastructure (e.g., bridges, roads, and dam lifespan), public safety, economics (e.g., power production), and aesthetics. In addition, water quality can also be negatively impacted. An estimate of the amount of suspended sediment being carried by a river (at various stages) can help predict the potential severity of such problems (Müftüoğlu 1980; Yılmaz et al. 2016).

Reservoir sedimentation is a serious problem worldwide, leading to a gradual loss of reservoir capacity, which reduces the effective life of dams, thus diminishing their capacity to supply irrigation water, generate hydropower, control flooding, maintain a dependable water supply, and provide recreational benefits (Bashar et al. 2010; Ghernaout and Remini 2014).

The Çoruh River (Turkey) with 427 km length supports many dams, most built to generate power. There are fifteen dams either already in operation or under construction, ten of which are located on the mainstream of the Çoruh, with a total head height (from the uppermost Laleli dam to the lowermost Muratli dam) of 1420 m. The Deriner dam (252 m high) and the Yusufeli dam (270 m high) are relatively major dams. These dams are important to Turkey's energy production because they supply 3169.7 MW of capacity and generate 10,724.16 kWh annually (Kankal et al. 2014). Therefore, knowing the suspended sediment load of various reaches of the Çoruh drainage is essential for designing and operating dams in the basin.

Estimating suspended sediment loads (SSL) is a difficult and complex undertaking because sediment transport is highly nonlinear and governed by many factors, including intensity of flow, sediment supply, river bed morphology, and locations of reservoirs (McBean and Al-Nassri 1988; Demirci and Baltaci 2013; Chen and Chau 2016). However, the development of soft computing methods in recent decades has enabled dam designers to use artificial intelligence and heuristic regression techniques to better predict SSL. These new computing methods have recently taken the place of traditional time series techniques, including conventional sediment rating curves (SRC), multi-linear regressions, and autoregressive models. In fact, such artificial intelligence techniques have proven to be a powerful substitute for traditional hydrological forecast models.

Today, the most-widely used approaches for SSL modeling are artificial neural networks (ANN), adaptive neuro-fuzzy inference systems (ANFIS), support vector machine (SVM) models, and hybrid models. Often more than one model is used in SSL studies and usually outputs of various models are compared with one another for accuracy. For example, Jain (2001), Cigizoglu (2004) and Rajaee et al. (2009) found that the ANN models gave better results than SRC, whereas Alp and Cigizoglu (2007), Rajaee et al. (2009), Melesse et al. (2011) and Taşar et al. (2017) found that ANN gave more realistic results than multi-linear regression (MLR) approaches. Similarly, Cobaner et al. (2009) and Kisi et al. (2009) found that the ANFIS models provided lower error values than ANN and SRC models. However, Kisi and Shiri (2012) and Emamgholizadeh and Demneh (2018) showed that gene expression programs can predict SSL better than either ANN or ANFIS models.

When Cimen (2008) tested SVM, fuzzy differential evolution, and fuzzy-logic methods to estimate SSL, he found that the SVM method gave the best results. Likewise, Buyukyildiz and Kumcu (2017) showed that the SVM method yielded better results than the ANN and ANFIS methods. In contrast, Lafdani et al. (2013) determined that the ANN method gave lower error values than the SVM method.

In recent years, hybrid methods have been successfully used to predict SSL. For example, Rajaee (2011) and Olyaie et al. (2015) tested a combination of wavelet and ANN methods (WANN). In both studies, WANN achieved better results than other methods (SRC, MLR, ANN, and ANFIS) used alone. Chen and Chau (2016) predicted SSL using hybrid and classical feedforward neural network (FNN) models, namely hybrid double FNN (HDFNN), multilayer FNN, double parallel FNN, and Hybrid FNN models. In fact, HDFNN has proven to be more accurate than the other models in predicting SSL. When Nourani and Andalib (2015) compared a hybrid wavelet-based least square SVM (WLSSVM) and WANN models with least-square SVM (LSSVM) and ANN models, they found that WLSSVM behaved better than WANN, whereas both WLSSVM and WANN showed a more robust performance than did either LSSVM or ANN models alone.

In the past decade, heuristic regression methods, namely model trees, support vector regression (SVR) models, and multiple adaptive regression splines (MARS) models have been used most often for estimating SSL. Reddy and Ghimire (2009) determined that an M5 model tree proved more successful than SRC, MLR, or GEP models. Likewise, Kumar et al. (2012) found that the M5 model tree algorithm provided lower error values than ANN, fuzzy logic, or reduced-error pruning-tree (REPTree) methods. Similarly, Goyal (2014) found that the M5 model tree and wavelet regression methods provided better predictions of sediment yields and gave more consistent results than ANN. Kumar et al. (2016) showed that least-square SVR and ANN models provided lower error values than other models (i.e., radial basis functions ANN, MLR, classification and regression trees, and M5 model trees). When Bharti et al. (2017) evaluated five machine-learning models (ANN Levenberg-Marquad, ANN scaled conjugate gradient, least square SVR, REPTree, and M5 model tree) to predict runoff and sediment load using hydro-meteorological variables as inputs, they found that the ANN Levenberg-Marquad model marginally outperformed the other models. However, when Talebi et al. (2017) applied the regression tree and model tree approaches to estimate SSL, they found that both approaches outperformed the ANN and SRC methods. The MARS method is very rarely used to predict SSL relative to other heuristic models. Vali et al. (2010) investigated the capabilities of using a geomorphologically-based ANN, power relation regression (PRR), traditional ANN, and MARS models for predicting SSL. They found that the performance of PRR and MARS models was inferior to ANN models. However, Chachi et al. (2016) found that a hybrid MARS fuzzy regression model was more accurate than fuzzy least squares regression and fuzzy least absolute regression in predicting SSL.

The above discussion demonstrates that artificial intelligence techniques, especially ANN-based ones, have been used widely to estimate SSL and that heuristic regression models have been more frequently used recently. In our study, a MARS heuristic regression model and meta-heuristic methods (such as artificial bee colony (ABC) and teaching-learning based optimization (TLBO) techniques) were selected to estimate the amount of SSL in the Çoruh River. The final output of the MARS approach is a set of simple equations (functions) that are very convenient to use and are comparable in structure to traditional sediment transport models. In contrast, ANN models are more complex; they are also embedded in computer code and so are not transparent to users.

ABC and TLBO algorithms are usually used to solve combinatorial optimization problems; therefore, we used them to optimize coefficients of regression equations used to estimate SSL. Similar studies have been conducted by Altunkaynak (2009) and Pour et al. (2011), but they used genetic algorithms to optimize solutions. The ABC algorithm was used by Kisi et al. (2012) to train the ANN method, but this algorithm has never been applied to regression equations used to model SSL. The TLBO method has also never been is used before now to estimate SSL.

This paper is organized as follows. The next section provides detailed information on the methods and available data used in the study. Then,