Journal of Hydrology 372 (2009) 68–79

Contents lists available at ScienceDirect

Journal of Hydrology

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

Evolutionary fuzzy models for river suspended sediment concentration estimation

Özgür Kişi *

Erciyes University, Engineering Faculty, Civil Engineering Department, Hydraulics Division, 38039 Kayseri, Turkey

ARTICLE INFO

Article history:
Received 28 April 2008
Received in revised form 8 January 2009
Accepted 30 March 2009

This manuscript was handled by K. Georgakakos, Editor-in-Chief, with the assistance of Taha Ouarda, Associate Editor

Keywords:
Suspended sediment
Fuzzy Modelling
Differential evolution
Neuro-fuzzy
Neural networks
Rating curve

SUMMARY

This paper proposes the application of evolutionary fuzzy models (EFMs) for suspended sediment concentration estimation. The EFMs are improved by the combination of two methods, fuzzy logic and differential evolution. The accuracy of EFMs is compared with those of the adaptive neuro-fuzzy, neural networks and rating curve models. The daily streamflow and suspended sediment data belonging to two stations, Quebrada Blanca Station and Rio Valenciano Station, operated by the US Geological Survey (USGS) are used as case studies. The mean square errors and determination coefficient statistics are used for evaluating the accuracy of the models. Based on the comparison of the results, it is found that the EFMs give better estimates than the other techniques.

Introduction

Correct estimation of sediment volume carried by a river is important with respect to pollution, channel navigability, reservoir filling, hydroelectric-equipment longevity, fish habitat, river aesthetics and scientific interests. It is a well known fact that all reservoirs are designed to a volume known as "the dead storage" to accommodate the sediment income that will accumulate over a specified period called the economic life. The underestimation of sediment yield results in insufficient reservoir capacities while the overestimation will lead to over-capacity reservoirs. Only the appropriate reservoir design and operation is sufficient to justify every effort to determine sediment yield accurately. In environmental engineering the prediction of river sediment load has an additional significance, especially if the particles also transport pollutants.

McBean and Al-Nassri (1988) examined the uncertainty in suspended sediment rating curves and concluded that the practice of using sediment load versus discharge is misleading because the goodness of fit implied by this relation is spurious. They have instead recommended that the regression be established between sediment concentration and discharge.

Neural networks (NN) have been successfully applied in a number of diverse fields including water resources. In the hydrological forecasting context, recent experiments have reported that artificial neural networks (ANNs) may offer a promising alternative for rainfall–runoff modelling (Shamseldin, 1997; Solomatine and Dulal, 2003; Giustolisi and Laucelli, 2005), streamflow prediction (Zealand et al., 1999; Chang and Chen, 2001; Sivakumar et al., 2002; Kisi, 2004a; Hu et al., 2005; Kisi, 2007, 2008), reservoir inflow forecasting (Jain et al., 1999; Bae et al., 2007), and suspended sediment estimation (Jain, 2001; Tayfur, 2002; Cigizoglu, 2004; Kisi, 2004b, 2005; Cigizoglu and Kisi, 2006; Tayfur and Guldal, 2006). Jain (2001) used a single ANN approach to establish sediment-discharge relationship and found that the ANN model could perform better than the rating curve. Tayfur (2002) developed an ANN model for sheet sediment transport and indicated that the ANN could perform as well as, in some cases better than, the physically-based models. Cigizoglu (2004) investigated the accuracy of a single ANN in estimation and forecasting of daily suspended sediment data. Kisi (2004b) used different ANN techniques for daily suspended sediment concentration prediction and estimation and he indicated that multi-layer perceptron could show better performance than the generalized regression neural networks and radial basis function. Kisi (2005) developed an ANN model for modelling suspended sediment and compared the ANN results with those of the rating curve and multi-linear regression. Cigizoglu and Kisi (2006) developed some methods to improve ANN performance in

* Tel.: +90 352 437 0080; fax: +90 352 437 5784.
E-mail address: kisi@erciyes.edu.tr

0022-1694/$ - see front matter
doi:10.1016/j.jhydrol.2009.03.036

suspended sediment estimation. Tayfur and Guldal (2006) predicted total suspended sediment from precipitation.

Also, fuzzy logic has also been used successfully for prediction of suspended sediment during recent years (Kisi, 2003; Tayfur et al., 2003; Kisi, 2004c, 2005; Kisi et al., 2006; Lohani et al., 2007; Kisi et al., 2008). Tayfur et al. (2003) used a fuzzy logic algorithm for runoff-induced sediment transport from bare soil surfaces. Kisi (2004c) developed fuzzy models to estimate daily suspended sediments. He compared the fuzzy estimates with those of the sediment rating curves and found that the fuzzy models performed better than the rating curves. The fuzzy models proposed in his study are site-specific and do not simulate the hysteresis effects. Kisi (2005) used a neuro-fuzzy model for daily suspended sediment estimation. Kisi et al. (2006) showed that fuzzy rule-based models using triangular membership functions perform better than the sediment rating curve models in suspended sediment concentration prediction. They used expert considerations and trial–error method for the derivation of the membership functions. The hysteresis effects were not considered in their study. Lohani et al. (2007) used fuzzy logic for deriving stage-discharge-sediment concentration relationships. They used a fuzzy system based on Tagaki-Sugeno technique and subtractive clustering approach for the derivation of the membership functions. Kisi et al. (2008) investigated the accuracy of a neuro-fuzzy technique in modelling daily suspended sediment of rivers in Turkey and compared with three different ANN techniques.

The process of establishing a rating relationship between flow and suspended sediment is basically a nonlinear mapping problem. The statistical tools that are commonly used in such situations are regression and curve fitting. However, these techniques are not adequate in view of the complexity of the problem and there is room for much improvement. This study is concerned with the application of evolutionary fuzzy models (EFMs) for predicting suspended sediment concentration. The accuracy of EFM is compared with those of the adaptive neuro-fuzzy, neural networks and rating curve models employed in the previous work of Kisi (2005). The fuzzy models proposed in this study based on Mamdani technique and simulate the hysteresis effects. This is the first study that compares the accuracy of the EFMs with those of the neuro-fuzzy and neural network models in the hydrological context.

Methodology

Fuzzy logic approach

Zadeh (1965) first introduced the concepts of the fuzzy logic and pioneered its development. The fuzzy concepts and operational algorithms are given in many textbooks, for example Kosko (1993) and Ross (1995).

Fuzzy logic allows for something to be partly this and partly that, rather than having to be either all this or all that. The degree of "belongingness" to a set or category can be described numerically by a membership number between 0 and 1.0. Fuzzy membership functions can take many forms, but simple straight line functions and they are often preferred. Triangular functions with equal base widths are the simplest possible and these are often selected for practical applications (Russel and Campbell, 1996).

Fuzzy categories can be used to set up rules of the following form for control purposes: "If the value of variable x_1 is "high" and variable x_2 is "medium" then the result, y is "low". It is claimed that such rules more closely resemble the way we think than do more explicit mathematical rules. Fuzzy logic programming can be used in two main ways: as a way of relating a set of outputs to a set of inputs in a "model-free" way-the "fuzzy inference method and as a way of trying to model the behavior of a human expert".

Sets of input data along with the corresponding outputs are provided to the fuzzy system in the fuzzy inference method. And the fuzzy system "learns" how to transform a set of inputs to the corresponding set of outputs through a fuzzy associative map (FAM), sometimes called fuzzy associative memory (Kosko, 1993). ANNs can perform the same function such as regression, but these tend to be "black box" approaches. A fuzzy logic system is more transparent and more flexible.

In the present study fuzzy logic is used for the estimation of sediment concentration from streamflow measurements. As suggested by its name, fuzzy logic does not provide a rigorous approach for developing or combining fuzzy rules which can be achieved through many ways. The approach used in this study is explained as below.

First the input and output variables are divided into a number of subsets with simple triangular fuzzy membership functions. Generally, there are c^n fuzzy rules where c and n denote the numbers of subsets and input variables, respectively. The more subsets the greater the accuracy possible but the larger the rule base which is more difficult to construct (Şen, 1998). In the case, say, of one input, x, with n subsets, the rule base takes the form of an output y_k ($k = 1, 2, \ldots, n^2$). If there is one input variable as x with "low", "medium" and "high" fuzzy subsets then consequently there will be three rules as follows:

R$_1$: IF x has low **THEN** y_1
R$_2$: IF x has medium **THEN** y_2
R$_3$: IF x has high **THEN** y_3

Membership degree, w_k, for x is computed to be assigned to the corresponding output y_k for each rule triggered. Hence the weighted average of the outputs from three rules gives a single weighted output, y, as:

$$y = \frac{\sum_{k=1}^{3} w_k \cdot y_k}{\sum_{k=1}^{3} w_k} \tag{1}$$

Thus, the values of the output y can be computed from Eq. (1) for any combination of input variable fuzzy subsets after setting up the rule base. To use sample data and derive the necessary rule base by the fuzzy inference procedure is a very common method in deciding about the fuzzy rule base (Şen, 1998).

A fuzzy rule base, used also in the present study, can be achieved step-by-step from sets of input and output data as follows:

1. Try to model the problem using minimum number of input variables.
2. Divide the range of each input variable into a number of sub intervals with a membership function (triangular membership function is used generally). Theoretically, different number of fuzzy subsets can be tried and the optimum one can be found by minimizing the total squared error between the predictions and observations. The number of fuzzy subsets has been established empirically between four and eight in practical studies (Şen, 1998).
3. Compute the membership value (w_k) for x in each of the fuzzy subsets for each data point n (one value for x and y).
4. Store the output y_k along with the complete set of rule weights w_k.
5. Repeat for all the other data points.
6. Compute the weighted average with an expression similar to Eq. (1) (Kiszka et al., 1985a, b).

One of the main problems in designing any fuzzy system is construction of the fuzzy subsets because all changes in the subsets