

EVALUATION OF EMPIRICAL MODELLING TECHNIQUES FOR THE ESTIMATION OF SEDIMENT AMOUNT IN RIVERS

*Başak GÜVEN **
*Zeynep AKDOĞAN **

Received: 22.07.2016; revised: 16.11.2016; accepted: 17.11.2016

Abstract: The sediment transport processes of streams have been the subject of research for many years. Sediment amount carried by a river is strongly correlated with the river's flow rate and sediment concentration. This study aims to represent this correlation and to estimate the sediment amount using four different modelling techniques: MLR, PLS, SVM, and ANN. Records of river flow, sediment concentration and sediment amount obtained from the Göksu River, located in the Eastern Mediterranean region of Turkey, are used as input data in the models. The aim of this study is to evaluate the effectiveness of ANN modelling in the estimation of sediment amount carried by river flow. Fifty percent of the data are used as training set to develop the models. The other half of the data is used for verification set. The performance of the four models is evaluated by determination coefficient of prediction set (r^2_{pred}). The results indicate that ANN is the most effective method ($r^2_{pred} = 0.94$), followed by SVM ($r^2_{pred} = 0.72$). MLR and PLS methods are the least effective techniques ($r^2_{pred} = 0.67$) for estimating sediment amount in the Göksu River. Therefore, ANN approach is further studied to propose the best configuration for the prediction of river sediment amount.

Keywords: Sediment amount, River, Modelling, ANN

Nehirlerde Sediment Miktarının Belirlenmesinde Ampirik Modelleme Tekniklerinin Değerlendirilmesi

Öz: Nehirlerdeki sediment taşınım süreçleri uzun yıllardır önemli bir araştırma konusu olmuştur. Nehirlerde taşınan sediment miktarı, nehrin akımı ve sediment konsantrasyonu ile güçlü bir ilişki içerisindedir. Bu çalışma, bu ilişkiyi göstermeyi ve dört farklı modelleme tekniği olan MLR, PLS, SVM ve ANN metotlarını kullanarak sediment miktarını hesaplamayı amaçlamaktadır. Türkiye'nin Doğu Akdeniz bölgesinde yer alan Göksu Nehri'ne ait akım, sediment konsantrasyonu ve sediment miktarı modellerde girdi verisi olarak kullanılmıştır. Bu çalışmanın amacı, nehir akımıyla taşınan sediment miktarının tahmin edilmesinde ANN modelleme tekniğinin etkisini değerlendirmektir. Verilerin yüzde ellisi modelin geliştirilmesi için öğrenme seti olarak, kalan veriler ise modelin validasyonu için test seti olarak kullanılmıştır. Test setinin belirleme katsayısı (r^2_{pred}) dikkate alınarak dört modelin performansı değerlendirilmiştir. Sonuçlar ANN'nin en etkili yöntem olduğunu ($r^2_{pred} = 0.94$) ve onu SVM'nin takip ettiğini ($r^2_{pred}=0.72$) göstermektedir. MLR ve PLS ise Göksu Nehri'ndeki sediment miktarının belirlenmesinde en az etkili yöntemlerdir ($r^2_{pred} = 0.67$). Bu nedenle, nehirdeki sediment miktarını tahmin etmek için en etkili yöntem, ANN'nin farklı konfigürasyonları çalışılarak araştırılmıştır.

Anahtar Kelimeler: Sediment miktarı, Nehir, Modelleme, ANN

* Institute of Environmental Sciences, Boğaziçi University, 34342 Istanbul, Turkey
Correspondence Author: Zeynep Akdoğan (zeynepakd@hotmail.com)

1. INTRODUCTION

The sediment transport of streams is a complex phenomenon, which have been the subject of research for many years due to its importance in planning the management of water resources. Process-based numerical models based on the relation between sediment concentration values and streamflow data have been widely used for prediction of sediment amount (Engelund and Fredsoe, 1976; Dietrich et al., 1999; Nelson et al., 2006; Jarritt and Lawrence, 2007; Kettner and Syvitksi, 2008). However, a river system is a complex network including various physical and morphologic dynamics, thereby modelling such systems requires a detailed spatial and temporal data. For this reason, a simpler, user-friendly approach is required and preferable for modelling sediment transport in rivers.

Empirical modelling is an alternate method to estimate the sediment amount in rivers using the regression techniques to fit the measured data. Such methods facilitate to control the data inputs and identify the irrelevant variables and provide a flexible approach to produce reasonable solution from small data sets (Abrahart and White, 2001). Different regression models have been studied in literature for modelling sediment transport in rivers. For example, Sinnakaudan et al. (2006) developed a model to estimate the total bed material for rivers in Malaysia using Multiple Linear Regression (MLR) analyses. Shi et al. (2013) used Partial Least Squares (PLS) regression to explore the relationship between the landscape characteristics and sediment amount. A study carried out by Kisi (2012) investigated the ability of Least Square Support Vector Machine (LSSVM) for modelling discharge-suspended sediment relationship.

Artificial Neural Network (ANN) is an alternative data-driven modelling, which has been widely applied in a variety of areas, especially for the last decades. Recent studies reveal that ANN has become an effective methodological approach for modelling sediment transport (Abrahart and White, 2001; Tayfur, 2002; Yitian and Gu, 2003; Bhattacharya et al., 2005; Yang et al., 2009; Yenigün et al., 2010; Van Maanen, 2010; Arı Güner et al., 2013). Abrahart and White (2001) carried out a study on the comparison of ANN and MLR techniques using small data sets, and proposed ANN was able to exceed the limitations of MLR method. Tayfur (2002) modelled the sheet sediment transport using ANN and tested the performance against that of the most commonly used physically-based models, whose transport capacities were based on flow velocity, shear stress, stream power, and unit stream power. The results revealed that ANN performed as well as the physically-based models for simulating nonsteady-state sediment loads from different slopes. Yitian and Gu (2003) applied ANN for modelling daily and annual sediment discharges in the Yangtze River and Dongting Lake, China. The comparison of the predicted and observed data demonstrated that ANN technique was a powerful tool for real-time prediction of flow and sediment transport in complex network of rivers. Arı Güner et al. (2013) applied ANN method for modelling longshore sediment transport (LST) in Karaburun, Turkey and evaluated the accuracy of the ANN predictions against the measured values. They also compared ANN with two well-known empirical formulas (CERC, Kamphuis), and a numerical model (LITPACK). According to the results, ANN followed the most successful method “Kamphuis” for estimation of LST rates and provided a practical and accurate determination of the LST rate for most regions.

This paper aims to develop four different regression models; MLR, PLS, SVM, ANN, and test the performance of these models for the estimation of sediment amount in the Göksu River. In addition, the effect of different network topologies of ANN are studied and the best configuration for the prediction of river sediment amount is assessed. Here, we aimed at proposing an effective and simple regression model, which could provide a reliable alternative to more complicated process-based models for the estimation of sediment amount in the study area.

2. METHODOLOGY

2.1. Data Requirements

The classical and commonly used method in the estimation of sediment amount is based on the relation between measured suspended sediment concentration values and measured water discharge, which can be represented by the below formula:

$$Q_s = Q_w C_s k \quad (1)$$

where Q_s is the sediment amount (ton/day), Q_w is the flow-rate (m^3/s), C_s is the sediment concentration (ppm) and k is a coefficient.

The data for the Göksu River including river flow, sediment concentration, and sediment amount is obtained from Turkish General Directorate of Electrical Power Resources Survey and Development Administration (EIE). A total number of 493 data including daily flow and monthly sediment concentrations between years 1999 and 2010 are entered to regression models as independent variables, while monthly sediment amount are used as dependent variable.

2.2. Regression Models

Molegro Data Modeller (MDM) software is used to estimate the sediment amount by the application of four different regression models: MLR, PLS, SVM, and ANN. Finally, three different network topologies of ANN methods are further assessed to determine the best configuration for the prediction of river's sediment amount.

MLR model assumes that the dependent variable y is a linear function of the independent variables, x_i , which can be written as:

$$y = c_0 + c_1 x_1 + c_2 x_2 + \dots + c_N x_N \quad (2)$$

where the c_i 's are the regression coefficients in the linear model (MDM User Manual, 2013).

In PLS, a smaller set of factors called latent components is extracted from the set of available descriptors (independent variables x_i), which models the dependent variable y . PLS regression creates latent components from the independent variables, x_i , while taking the dependent variable y into account (MDM User Manual, 2013).

SVM is used for linear classification. MDM considers that different types of objects are positioned on a 2D plane and is interested in a classifier capable of predicting the type of an object given its position in the plane. In this case the data are linearly separable with several possible choices of lines dividing the plane into regions according to class of objects. Support vector machines try to find the maximum separating hyperplane, which in 2D corresponds to the line with the widest borders (MDM User Manual, 2013).

ANN consists of input, hidden and output neurons arranged in layers. The neural network is constructed by assigning each independent variable to a neuron in the input layer. Each input is connected to a number of neurons, which constitute the hidden layer (Van Maanen et al., 2010). The network is first trained, whereby the target output neuron in each output neuron is minimized by adjusting the weights and biases through some training algorithm. During training, each connection multiplies the neuron output by a weight before the output enters the connected neuron. The combination of the weighted inputs can be expressed as (Tayfur, 2002):

$$net_j = \sum x_i w_{ij} - b_j \quad (3)$$

where net_j is the summation of the weighted input for the j th neuron, x_i is the input from the i^{th} neuron to the j^{th} neuron, w_{ij} is the weight from the i^{th} neuron in the previous layer to the j^{th} neuron in the current layer, and b_j is the threshold value, also called the bias, associated with node j . The sigmoid function is applied as an activation function in the training of network to understand if the activation of a neuron is strong enough and produces a successive output that is sent to other neurons as an input. The sigmoid function is represented below (Tayfur, 2002):

$$f(net_j) = \frac{1}{1+e^{-net_j}} \quad (4)$$

In this study, flow rate and sediment concentration are entered to ANN model as input layer and the connections from the hidden layer are connected to the output layer, which is trained to estimate the dependent variable: sediment amount. The number of layers and neurons in hidden layers are adjusted by considering different network configurations, which are given in Figure 1.

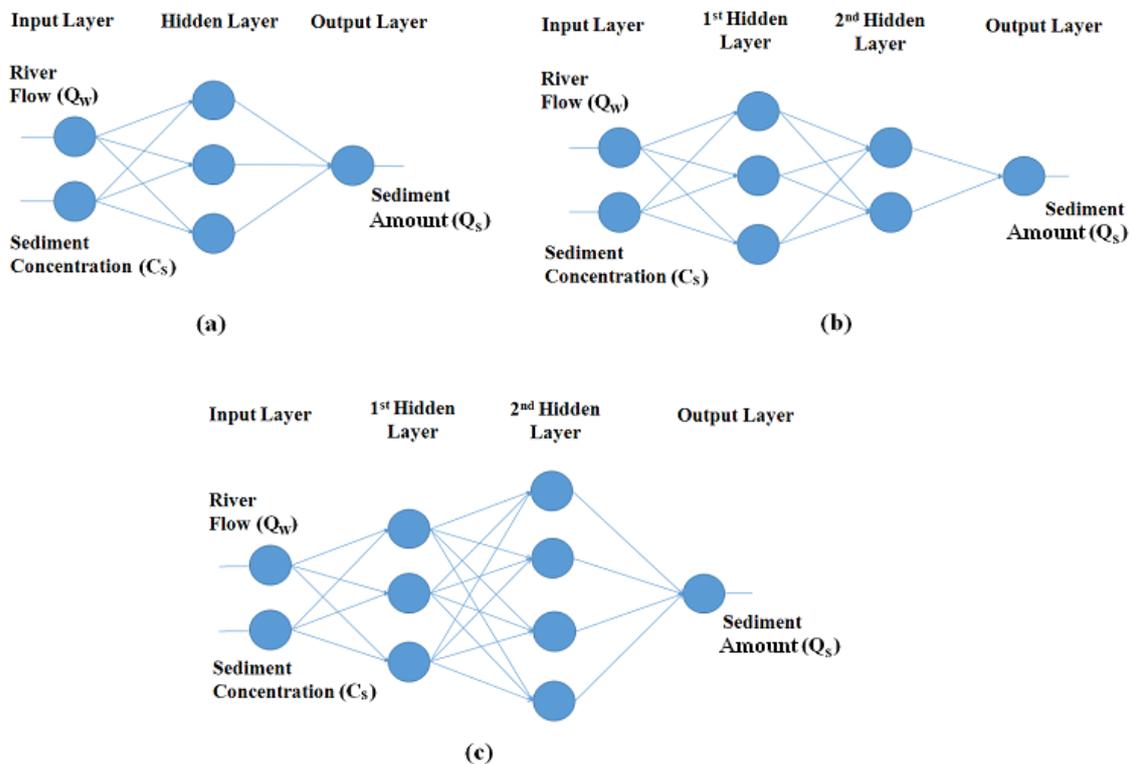


Figure 1: Backpropagation configuration of (a) ANN (3-0), (b) ANN (3-2), and (c) ANN (3-4) models

3. RESULTS AND DISCUSSION

MLR, PLS, SVM, and ANN analyses are applied to investigate the relationship between dependent variable and independent variables (descriptors) and to predict sediment amount in the Göksu River. Depending on the availability of field data, model validation is undertaken based on the predicted and observed sediment amounts. MDM divides the existing database into two groups for all regression models. One is used for training, and the other for validation purposes. Hence, the existing data sets are splitted into two subsets where 50% of them are used for training and the other 50% are used for prediction and validation. The same training/

prediction sets are used for generation of all models. The regression results for MLR, PLS and SVM are illustrated in Figure 2a, 2b and 2c, respectively. The model outcomes for SVM fit the observed values better, whereas more outliers are observed for MLR and PLS model results. Outliers are observations that have large residual values and may be originated from errors or from initially accepting marginal or unacceptable data (Sinnakaudan et al., 2006). Parameter settings for SVM are given in Table 1.

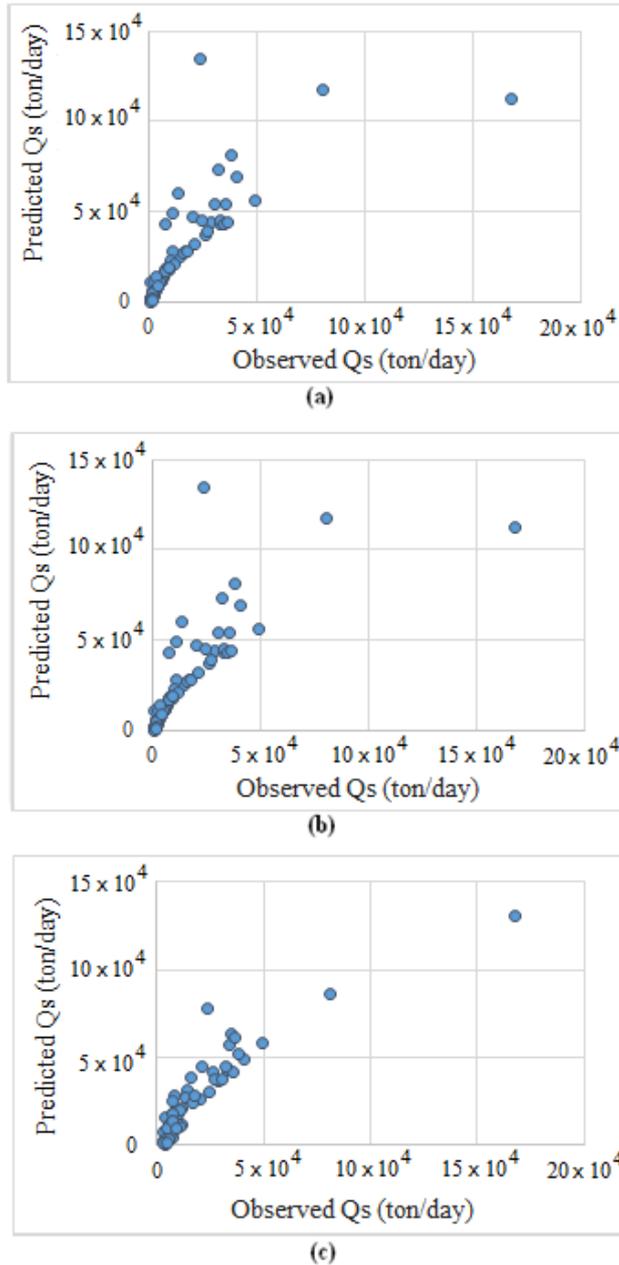


Figure 2:
Predicted vs. observed sediment amounts for (a) MLR, (b) PLS, and (c) SVM models

Table 1. Parameter settings for SVM model

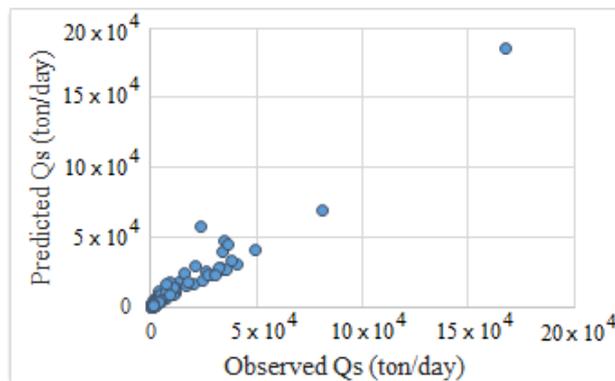
Parameter Settings	
Model type	Epsilon- SVR
Kernel	Radial basis function
Termination criterion tolerance	0.001
Cost	1
Gamma	0
Epsilon	0.1
Data range normalization	-1 – 1

The same method is followed for development of ANN model. ANN configuration given in Figure 1a is set up to predict the sediment amount in the Göksu River. Determination coefficient of prediction set (r^2_{pred}) is used to compare the performance of the four models and select the best method. The model that have maximum r^2_{pred} value is selected for further analysis. The model results reveal that ANN is the most effective method for estimating sediment amount in the Göksu River. Previous studies also revealed that ANN is a powerful tool for prediction of flow and sediment transport in river systems and preferable to exceed the limitations of other regression methods and physically-based models (Abrahart and White, 2001; Tayfur, 2002; Arı Güner et al., 2013).

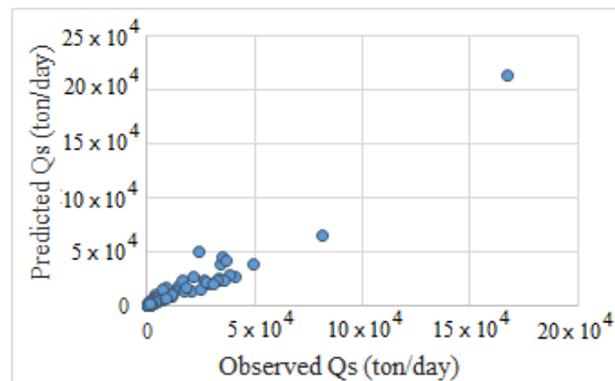
Two different network topologies are also applied to determine the best configuration, one of which includes 2nd hidden layer with two neurons and the other also contains the 2nd layer with four neurons. Initial weight range values between 0.2 and 0.8 are entered to ANN model. The best regression outcomes are obtained for the weight value 0.5 ($r^2_{pred}=0.94$), so this value is maintained for all ANN methods. Parameter settings of the models and outcomes are given in Table 2 and Figure 3, respectively. Overall statistics of four models are also given in Table 3. According to the model results, it is observed that increased number of neurons in the 2nd layer does not have a significant influence on regression outcomes.

Table 2. Parameter settings for ANN models

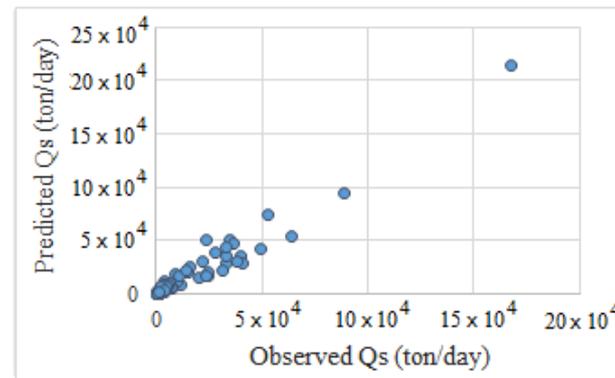
Parameters	ANN (3-0)	ANN (3-2)	ANN (3-4)
Max training epochs	1000	1000	1000
Learning rate	0.3	0.3	0.3
Output layer learning rate	0.3	0.3	0.3
Momentum	0.2	0.2	0.2
Data range normalization	0.1-0.9	0.1-0.9	0.1-0.9
Number of neurons in 1st hidden layer	3	3	3
Number of neurons in 2nd hidden layer	0	2	4
Initial weight range (+/-)	0.5	0.5	0.5



(a)



(b)



(c)

Figure 3:

Predicted vs. observed sediment amounts for (a) ANN (3-0), (b) ANN (3-2), and (c) ANN (3-4) models

In addition to r^2_{pred} , Spearman's rank correlation coefficient (ρ) and Nash-Sutcliffe efficiency coefficient (NS) are also calculated. NS is defined as one minus the sum of the absolute squared differences between the predicted and observed values normalized by the variance of the observed values during the period under investigation (Krause et al., 2005). According to the overall statistics given in Table 3, ANN (3-0) can be suggested as the most reliable model among the four regression techniques and different configurations of ANN.

It is important to define an applicability domain of the proposed models for future applications on different data scales. Applicability domain is a structural space, knowledge, or information on which the training set of the model has been developed, and for which it is

applicable to make predictions for new data points (Roy et al., 2015). The model results reveal that 92% of the predicted values of ANN (3-0) fall within the applicability domain of the proposed model.

Table 3. Overall statistics of regression models

Prediction Statistics	MLR	PLS	SVM	ANN (3-0)	ANN (3-2)	ANN (3-4)
r^2_{pred}	0.67	0.67	0.72	0.94	0.93	0.94
Spearman's rho	0.99	0.99	0.95	0.96	0.95	0.94
NS	0.25	0.25	0.7	0.94	0.89	0.89

4. CONCLUSION

The aim of the present study is modelling the sediment amount in the Göksu River via different black box models by using the water discharge and sediment concentrations as input data. For this purpose, four regression techniques; MLR, PLS, SVM, and ANN are applied to develop the models and the performance of such models are evaluated by determination coefficient of prediction set (r^2_{pred}).

The ANN model gives the most reliable predictions among the regression models tested, with a r^2_{pred} value 0.94, followed by SVM ($r^2_{\text{pred}} = 0.72$). MLR and PLS methods are the least effective techniques ($r^2_{\text{pred}} = 0.67$) for estimating sediment amount in the Göksu River. Further analysis of ANN method is applied for different configurations: ANN (3-0), ANN (3-2), and ANN (3-4). According to r^2_{pred} values given in Table 2, increasing the number of neurons in the 2nd layer does not have a significant influence on model outcomes.

Widely-used process-based models are based on the relationship between water discharge and sediment concentrations, as well as the topographical and geomorphologic properties of the rivers. However, spatial heterogeneity of river systems cause limitations of measured field data and prevent to obtain an accurate and reliable estimation of the sediment amount. For this reason, simpler approaches have been investigated in literature for modelling sediment transport in rivers. This paper focuses the four different empirical models that provide quick simulations with minimum data requirement. ANN (3-0) model may be used as an effective method instead of process-based models for the estimation of sediment amount in rivers.

ACKNOWLEDGMENT

This study was supported by Boğaziçi University Research Fund with Grant Number 916.

REFERENCES

1. Abrahart, R.J. and White, S.M. (2001). Modelling sediment transfer in Malawi: Comparing backpropagation neural network solutions against a multiple linear regression benchmark using small data sets, *Physics and Chemistry of the Earth (B)*, 26(1), 19-24. doi: 10.1016/S1464-1909(01)85008-5
2. Arı Güner, H.A., Yüksel, Y. and Çevik, E.Ö. (2013). Longshore sediment transport-field data and estimations using neural networks, numerical model, and empirical models, *Journal of Coastal Research*, 29(2), 311 – 324. doi: <http://dx.doi.org/10.2112/JCOASTRES-D-11-00074.1>
3. Bhattacharya, B., Price, R.K. and Solomatine, D.P. (2005). Data-driven modelling in the context of sediment transport, *Physics and Chemistry of the Earth*, 30(4), 297–302. doi: 10.1016/j.pce.2004.12.001
4. Dietrich, C.R., Green, T.R. and Jakeman, A.J. (1999). An analytical model for stream sediment transport: application to Murray and Murrumbidgee river reaches, Australia,

- Hydrological Processes*, 13(5), 763-776. doi: 10.1002/(SICI)1099-1085(19990415)13:5<763::AID-HYP779>3.0.CO;2-C
5. Engelund, F. and Fredsoe, J. (1976). A sediment transport model for straight alluvial channels, *Nordic Hydrology*, 7(5), 293-306.
 6. Jarritt, N.P. and Lawrence, D.S.L. (2007). Fine sediment delivery and transfer in lowland catchments: Modelling suspended sediment concentrations in response to hydrological forcing, *Hydrological Processes*, 21(20), 2729-2744. doi: 10.1002/hyp.6402
 7. Kettner, A.J. and Syvitski, J.P.M. (2008). HydroTrend v.3.0: A climate-driven hydrological transport model that simulates discharge and sediment load leaving a river system, *Computers & Geosciences*, 34(10), 1170-1183. doi:10.1016/j.cageo.2008.02.008
 8. Kisi, O. (2012). Modeling discharge-suspended sediment relationship using least square support vector machine, *Journal of Hydrology*, 456-457, 110-120. doi:10.1016/j.jhydrol.2012.06.019
 9. Krause, P., Boyle, D.P. and Base, F. (2005). Comparison of different efficiency criteria for hydrological model assessment, *Advances in Geosciences*, 5, 89-97. doi:10.5194/adgeo-5-89-2005
 10. MDM (Molegro Data Modeller) User Manual, 2013. http://www.clcbio.com/files/usermanuals/MDM_manual.pdf (last accessed in June 2016)
 11. Nelson, P.A., Smith, J.A. and Miller, A.J. (2006). Evolution of channel morphology and hydrologic response in an urbanizing drainage basin, *Earth Surface Processes and Landforms*, 31(9), 1063-1079. doi: 10.1002/esp.1308
 12. Roy, K., Kar, S. and Ambure, P. (2015). On a simple approach for determining applicability domain of QSAR models, *Chemometrics and Intelligent Laboratory Systems*, 145, 22-29. doi: <http://dx.doi.org/10.1016/j.chemolab.2015.04.013>
 13. Shi, Z.H., Ai, L., Li X., Huang, X.D., Wu, G.L. and Liao, W. (2013). Partial least-squares regression for linking land-cover patterns to soil erosion and sediment yield in watersheds, *Journal of Hydrology*, 498, 165-176. doi:10.1016/j.jhydrol.2013.06.031
 14. Sinnakaudan, S. K., Ghani, A. A., Ahmad, M. S. S. and Zakaria N. A. (2006). Multiple linear regression model for total bed material load prediction, *Journal of Hydraulic Engineering*, 132(5), 521-528. doi: 10.1061/(ASCE)0733-9429(2006)132:5(521)
 15. Tayfur, G. (2002). Artificial neural networks for sheet sediment transport, *Hydrological Sciences*, 47(6), 879-892. doi: 10.1080/02626660209492997
 16. Van Maanen, B., Coco, G., Bryan, K. R. and Ruessink, B. G. (2010). The use of artificial neural networks to analyze and predict alongshore sediment transport, *Nonlinear Processes in Geophysics*, 17(5), 395-404. doi:10.5194/npg-17-395-2010
 17. Yang, C.T., Marsooli, R. and Aalami, M.T. (2009). Evaluation of total load sediment transport formulas using ANN, *International Journal of Sediment Research*, 24(3), 274-286. doi: 10.1016/S1001-6279(10)60003-0
 18. Yenigün, K., Bilgehan, M., Gerger, R. and Mutlu, M. (2010). A comparative study on prediction of sediment yield in the Euphrates basin, *International Journal of the Physical Sciences*, 5(5), 518-534. doi: 2B15E0C25933
 19. Yitian, L. and Gu, R.R. (2003). Modeling flow and sediment transport in a river system using an Artificial Neural Network, *Environmental Management*, 31(1), 122-134. doi: 10.1007/s00267-002-2862-9

