AN IMPROVED EXPRESSION OF THE SEDIMENT TRANSPORT MODEL OF ENGELUND AND HANSEN

B. BHATTACHARYA, R.K. PRICE and D. P. SOLOMATINE

UNESCO-IHE Institute for Water Education,
P.O. Box 3015, 2601 DA Delft, The Netherlands.
(Tel: +31-15-215-1812, Fax: +31-15-212-2921,
e-mail: bha@ihe.nl, rkp@ihe.nl, seol@ihe.nl)

The paper presents a data-driven approach using methods of machine learning to improve the accuracy of the sediment transport model of Engelund and Hansen.

Engelund and Hansen used a similarity principle to arrive to the following expression:

$$f\phi = \phi((\theta' - 0.06)\sqrt{\theta}) \quad (1)$$

where mobility parameter $\theta = hI/(s-1)D$, mobility parameter related to the grain roughness $\theta' = h'I/(s-1)D$, roughness factor $f = 2g/C^2$, dimensionless transport parameter $\phi = q_t/\sqrt{(s-1)gD^3}$, h = flow depth, I = energy slope, s = specific density parameter, D = mean bed particle size, h' = flow depth related to the grain roughness, g = acceleration due to gravity, C = Chezy's coefficient and q_t = total transport flux per unit width..

Using the data of Fort Collins Engelund and Hansen developed the following expression:

$$\theta' - 0.06 = 0.4\theta^2 \quad (2)$$

Substituting Eq. 2 in Eq. 1 and using the data of Fort Collins the final form of the model is:

$$f\phi = 0.1\theta^{5/2} \quad (3)$$

In order to build a data-driven model to predict total-load transport rate we considered the expression of Engelund and Hansen presented in Eq. 1, and selected θ and θ' as the input variables and $f\phi$ as the output variable.

Brownlie's (1981) compilation of 55 flume and 24 field datasets on sediment transport was considered. Data points with h: 0.01 to 17.3 m, u: 0.12 to 2.91 m/s, D_{50}: 0.011 to 32 mm and I: 0.000003 to 0.0158 were considered. The width/depth ratio was between 1 and 579. Due to a non-uniform distribution of data a log (natural) transformation of the input and output variables was applied. The data was partitioned into training (2814 data points) and testing (1373 data points) by maintaining a statistical homogeneity.

The data-driven modelling was performed using two machine learning methods: artificial neural networks (ANN) and model trees (MT). ANN is, in essence, a non-linear regression model that proved to be accurate in reconstructing functional dependencies of various types. MT is a set of linear regression models that are built using the principles of information theory: the input space is split into areas (subspaces) and in each of them a specialized linear regression model (LM) is built; they are made compatible by "smoothing". *WEKA* software was used to build MT and *NeuroSolutions* – to build multi-layered perceptron ANN trained by the back-propagation algorithm with one hidden layer of 9 nodes with a tangent hyperbolic function, and linear functions in the output layer.

The generated model tree was pruned to make it simpler so that the generalisation capacity improves. The pruned MT can be presented as four rules:

If $\theta > 1$ *then* LM1: $f\phi = 0.17\theta^{2.11}\theta'^{0.6}$

If $\theta' > 0.11$ and $\theta \leq 1$ *then* LM2: $f\phi = 0.24\theta^{1.96}\theta'^{0.908}$

If $0.054 < \theta' \leq 0.11$ *then* LM3: $f\phi = 1.183\theta^{1.15}\theta'^{2.2}$

If $\theta' \leq 0.054$ *then* LM4

It is interesting to note that a separate LM for the *upper regime* (with $\theta > 1$) and another LM for transport conditions with $\theta' \leq 0.054$ were associated. $\theta' = 0.054$ approximately defines the critical value of θ' for the initiation of motion for most particle sizes and therefore, LM4 is associated with flow situations without the initiation of transport. The first three LM can be used for computing the total transport rates.

The values of the exponent of θ and θ' in LM3 are relatively high. MT possibly compensates the higher values of the exponent with a higher value of the multiplying coefficient (1.183).

Table 1. Comparison of the root mean square errors (RMSE) of prediction of the models of Engelund and Hansen, model tree and ANN on the testing dataset. LM1, LM2 and LM3 refer to the linear models found by the model tree algorithm.

	Linear model	Number of examples	Average transport rate (ϕ)	RMSE		
				Engelund-Hansen	Model Tree	ANN
θ	LM1	185	38.7	32.8	23.5	25.7
$\theta \leq$ 1	LM2	445	3.7	5.0	4.1	4.7
	LM3	462	0.04	0.07	0.04	0.04
Total		1092	8.1	13.9	10.0	11.0

Table 2. Comparison of the percentage errors of prediction of the models of Engelund and Hansen, model tree and ANN on the testing dataset.

	Linear model	Number of examples	Percentage errors		
			Engelund-Hansen	Model Tree	ANN
$\theta > 1$	LM1	185	126	118	117
$\theta \leq 1$	LM2	445	109	75	86
	LM3	462	261	104	126
Total		1092	176	95	108

If compared to MT, ANN model has a more complex structure, it is embedded in computer code and is less transparent. The data-driven models predict $f\phi$, which is used to compute ϕ. The predictive accuracy of the data-driven models was compared with the model of Engelund and Hansen on the basis of root mean square errors (RMSE), average absolute percentage errors and the ratio of computed and measured transport rates. Table 1 compares the models based on RMSE. For transport conditions with $\theta' > 0.11$ and $\theta \leq 1$ (LM2) the decrease in RMSE from that of the model due to Engelund and Hansen is 18% for the MT model and 6% for the ANN model. For transport conditions with $0.054 < \theta' \leq$

0.11 (LM3) the decrease in RMSE from that of the model due to Engelund and Hansen is 43% for both the MT and ANN model. The decrease in RMSE for transport in the upper regime is 28% for the MT and 22% for the ANN model. The overall decrease in RMSE is 28% for the MT and 21% for the ANN model. Table 2 shows the comparison of the percentage errors. In general, it is observed that the data-driven models performed better than the model of Engelund and Hansen. The data-driven models were developed and tested using a large dataset with a substantially wide range of the variables. Therefore, it is expected that these models would be applicable for most situations in which a practising engineer needs to assess sediment transport rates.

Among the data-driven models the performance of the ANN and MT models are comparable. The MT model is a collection of equations. The validity of these equations can be reviewed and the equations can be conveniently implemented in a spreadsheet to compute transport rates. ANN, however, is encapsulated in computer code and needs specialised software.

Keywords: sediment transport, artificial neural networks, ANN, model trees, Engelund and Hansen.