Journal of Hydrology 539 (2016) 662–673

Contents lists available at ScienceDirect

Journal of Hydrology

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

Creating a non-linear total sediment load formula using polynomial best subset regression model

Davut Okcu [a,*], Ali Osman Pektas [b], Ali Uyumaz [c]

[a] Bahcesehir University, 34353 Besiktas, Istanbul, Turkey
[b] Bahcesehir University, Civil Engineering Department, 34353 Besiktas, Istanbul, Turkey
[c] Istanbul Technical University, Department of Hydraulics and Water Resources, Maslak, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 4 February 2016
Received in revised form 28 April 2016
Accepted 30 April 2016
Available online 2 June 2016
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor

Keywords:
Sediment transport
River hydrology
Polynomial best subset regression (PBSR)
Bed material load
Suspended sediment
Bedload sediment

SUMMARY

The aim of this study is to derive a new total sediment load formula which is more accurate and which has less application constraints than the well-known formulae of the literature. 5 most known stream power concept sediment formulae which are approved by ASCE are used for benchmarking on a wide range of datasets that includes both field and flume (lab) observations. The dimensionless parameters of these widely used formulae are used as inputs in a new regression approach. The new approach is called Polynomial Best subset regression (PBSR) analysis. The aim of the PBRS analysis is fitting and testing all possible combinations of the input variables and selecting the best subset. Whole the input variables with their second and third powers are included in the regression to test the possible relation between the explanatory variables and the dependent variable. While selecting the best subset a multi-step approach is used that depends on significance values and also the multicollinearity degrees of inputs. The new formula is compared to others in a holdout dataset and detailed performance investigations are conducted for field and lab datasets within this holdout data. Different goodness of fit statistics are used as they represent different perspectives of the model accuracy. After the detailed comparisons are carried out we figured out the most accurate equation that is also applicable on both flume and river data. Especially, on field dataset the prediction performance of the proposed formula outperformed the benchmark formulations.

1. Introduction

Sediment transport is one of the most powerful agents on river environment. It transfigures the river morphology by Long-term degradation and aggradation of channel beds via erosion and deposition. Such issues can have a direct effect on the level of the river during flooding. A change in the morphology of a river can threaten channel stability which can create local scour around hydraulic installations. In addition, an accurate prediction of the total sediment load is a key factor in managing sedimentation in reservoirs. Therefore, for almost a hundred years sediment transport prediction is one of the most studied issues in different disciplines.

The total sediment load includes the wash load and the bed-material load. The bed-material load consists of bed-load and suspended load. Generally, two approaches are available for predicting the bed-material load in a river. One is to estimate the bed-load and suspended load in separate calculations. This is based on the fact that the hydrodynamics of each mode of sediment transport is different. The methods developed by Einstein (1950), van Rijn (1984), and Toffaleti (1969) fall into this approach. The other approach is to estimate the bed-material load directly without dividing the transport mode into two parts (e.g. Engelund and Hansen, 1967; Brownlie, 1981; Ackers and White, 1973; Karim and Kennedy, 1990; Choi and Lee, 2015). This approach is simple and sometimes preferred in the sense that the two modes of sediment transport cannot easily be distinguished from one another, in reality.

In fact, choosing the approach is constrained by the available data, practical engineering purpose and the precision level of the study. It can be asserted that if the data availability is not a constraint the appropriate approach can be selected by using the Shields-Parker diagram. The Shields-Parker diagram (for more details see García, 2008, p: 60–65) shows that in gravel bed rivers, bed material is transported mainly as bed load. In this diagram the critical condition for suspension is plotted with an additional curve, which is derived from the ratio of shear velocity and the

* Corresponding author.
E-mail addresses: davut.okcu@eng.bau.edu.tr, okcud@itu.edu.tr (D. Okcu), aliosman.pektas@eng.bau.edu.tr (A.O. Pektas), uyumaz@itu.edu.tr (A. Uyumaz).

http://dx.doi.org/10.1016/j.jhydrol.2016.04.069
0022-1694/

sediment fall velocity (also see, Niño and García, 1998; Lopez and Garcia, 2001). On the other hand, in sand bed rivers, suspension and bed load transport of bed material coexist, particularly at high flows. So decision on the approach of the study if the bed load and the suspended load are investigated separately or together can be depend on the criteria that if the river bed type is gravel or sand. However, there are some suggested empirical equations on the decision of the river bed type (e.g. García et al., 2000), this decision is a bit complicated and including the bed type in the formulation (model) could add extra uncertainty to the applied numerical model. In this study, the second approach is used to predict the total sediment load. The flume and the field conditions are very different from each other and it is a big deal to find a model which gives good performance for both situations. Therefore some studies focused on only flume data. Smart (1984), Damgaard et al. (1997), Ackers and White (1973). Dogan et al. (2009) investigated if the sediment transport in natural alluvial channels can be predicted from observations at the laboratory scale.

Tayfur et al. (2013) and Pektas (2015) used explanatory analysis like principal component analysis or cluster analysis to identify the significant non-dimensional parameters of sediment transport. In recent studies most frequently Machine learning based models like Neural networks, Fuzzy logic, Support vector machines, are used in sediment modeling (e.g. Pektas and Dogan, 2013; Cigizoglu, 2002). Kisi and Cigizoglu, 2007 studied to improve the neural network performance in suspended sediment estimation. But machine learning models have a black box nature, so very small amount of information could be gained inside the model and most of these models are not suitable to generate a formula. Therefore the regression models are still popular.

Sinnakaudan et al. (2006) developed a total bed material formula by using multiple linear regression model. The authors focused on high gradient river sediment transport (Sinnakaudan et al., 2010) by using Regression models. Neter et al. (1989) discussed the use of all-possible-subset regression (Best Subset regression) in conjunction with stepwise regression. Howard et al. (2010) used best subset regression in their rainfall-runoff response models. Loomis et al. (2012) developed a new calibration that used best subsets regression model on lake sediment. Recently Lacombe et al. (2014) used best subset regression model for stream flow prediction. At the same time there is no sediment transport study using best subset regression model.

In the present study Best subset regression technique is modified and used to find the optimum input combination that keeps the nonlinear relationships. The aim to obtain the most parsimonious and the most accurate model to predict the total bed material concentration by considering the nonlinear relationships. Then we compare the new generated formula with the well-known and widely used traditional formulas of the literature. All the benchmark formulations are referenced by ASCE (2008). These are The Yang (1979), The Karim (1998), The Engelund and Hansen (1967), The Ackers and White (1973), The Molinas and Wu (2001) formulas. The Yang (1979) and The Karim (1998) regression based formulas.

2. Total bed material load sediment transport formulas

In the literature, there are many sediment transport formulas which have different specifications. Therefore many studies have attempted to find the best performing formulas for determining the total sediment load in rivers. Alonso (1980) compared eight formulas using both flume and field data and concluded that Yang's (1973), Ackers and White's (1973), Engelund and Hansen's (1967), and Laursen's (1958) formulas are all reliable. Brownlie (1981) compared 14 formulas using a compendium of sediment transport data from the laboratory and field records. He concluded that Brownlie's (1981), Ackers and White's (1973), and Engelund and Hansen's (1967) formulas are acceptable. Woo and Yoo (1991) carried out extensive performance tests with 10 selected sediment transport formulas and found that Engelund and Hansen's (1967), Ackers and White's (1973), and van Rijn's (1984) formulas are more reliable than the others. Nakato (1990) tested 11 total sediment load formulas using field data. Wu and Wang (2003) tested Engelund and Hansen's (1967), Ackers and White's (1973), Yang's (1979), Wu et al.'s (2000) formulas and found that the performance of all of these formulas are comparable, when uniform sediment is being considered. García (2008) recommended six total sediment load formulas, namely the Engelund and Hansen's (1967), Brownlie's (1981), Karim and Kennedy (1983), Ackers and White's (1973), Yang's (1973), and Molinas and Wu's (2001) formulas. Recently, Yang et al. (2009) compared the predictive performance of neural networks and the selected sediment formulas.

In this study, five total sediment load formulas, including Engelund and Hansen's (1967), Ackers and White's (1973), Yang's (1979), Brownlie's (1981), and Karim's (1998) formulas, are used. Hereafter, they are referred to as EH, AW, YANG, PBSR, and KARIM formulas, respectively. Although data used for development of some of these formulas include gravel, the formulas are designed for use in sand-bed rivers. Herein, the total sediment load formulas estimate either total sediment load per unit width (qt) or total bed-material concentration in parts per million by weight (flux based mass concentration) (C) and are related by

$$q_t = \frac{1}{(G_s)}\frac{C}{(1-C)}q_w \tag{1}$$

where q_w is water discharge per unit width, G_s is specific gravity of sediment.

Each formulation has special constraints and in comparison part these constraints are applied.

2.1. The Engelund and Hansen Formula (EH-1967)

The Engelund and Hansen (1967) Relation is a semi-empirical equation based on energy concepts. It is derived for sandy streams. This relation was developed from a small set of laboratory data (ASCE, 2008).

$$\frac{q_t}{\sqrt{(G_s-1)d_{50}^3}} = \frac{1}{C}0.05(\tau^*)^{2.5} \tag{2}$$

where q_t is total sediment load per unit width, $G_s - 1$ is the submerged specific gravity, and τ^* is dimensionless Shields stress.

The equation can be determined in another form:

$$C = 0.05\left(\frac{G_s}{G_s-1}\right)\frac{US}{\sqrt{(G_s-1)gd_{50}}}\frac{rS}{(G_s-1)d_{50}} \tag{3}$$

where C is flux-based mass Concentration, d_{50} is the median size of particle diameter, G_s is specific weight of sediment, U is velocity of water, S is slope, r is hydraulic radius.

The Engelund and Hansen equation (EH) is applicable to :

$$\begin{cases} d_{50} \geqslant 0.15\text{mm} \\ \text{Re}^* \geqslant 12 \\ Gradation(\sigma_s) \leqslant 2 \end{cases}$$

Implicitly, The EH formula can be written as a function of dimensionless parameters: