



## Regression model for sediment transport problems using multi-gene symbolic genetic programming



Bimlesh Kumar<sup>a,\*</sup>, Anjaneya Jha<sup>a</sup>, Vishal Deshpande<sup>a</sup>, Gopu Sreenivasulu<sup>b</sup>

<sup>a</sup> Indian Institute of Technology Guwahati, Guwahati, India

<sup>b</sup> Department of Civil Engineering, Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal 518501, India

### ARTICLE INFO

#### Article history:

Received 14 August 2013

Received in revised form 12 January 2014

Accepted 17 February 2014

#### Keywords:

Genetic programming

Incipient motion

Sediment transport

Total bed load

Vegetated flow

### ABSTRACT

Sediment transport modeling problems are complex due to the multi-dimensionality of the problems, along with their nonlinear interdependence. Also, in river hydraulics, phenomena are stochastic and variables are measured with uncertainties which are unavoidable. Dimensional and regression analyses have been employed in the past but have associated limitations. As a robust modeling tool, genetic programming was used to develop predictor models for three different but related problems of sediment transport-vegetated flow, incipient motion and total bed load prediction. A relatively new development over the conventional genetic programming-multi-gene symbolic regression was used to model functional relationships that were able to generalize highly nonlinear variations in data as well as predict system behavior from independent input data in all the three cases. The algorithmic parameters for genetic programming technique were resolved iteratively, varying based on problems in context. For all the three models developed, model efficiency criteria were found out and presented and the performance of the present model was compared with several past models for the same data points. The models developed herein were able to generalize the underlying relationships in the presented data as well as were able to predict values for unknown data with high accuracy.

© 2014 Elsevier B.V. All rights reserved.

### 1. Introduction

Sediment transport problems form an essential part of civil engineering practice with regard to river hydraulics related challenges faced in field. Solving sediment transport problems is indispensable in planning and managing water resources. However, system parameters in these problems are multiple in numbers with complex and exhibit nonlinear interdependence. This complexity combined with huge spatiotemporal variations and an inherent nonlinearity makes it difficult to analyze the system analytically. Besides, the variables seem to assume values specific to geographies and climates. This compels one to take assumptions in an analysis that are rendered false when the model may be used for disparate regions. This has also prevented development of universal models which offer satisfactory prediction capabilities irrespective of environments of application. As an example of interdependence, sediment and vegetation make the flow complicated, affecting the velocity profile. This then affects the bed and wall shear stresses and vegetation shapes, causing further changes in sediment loads and velocity profiles. Some or all of the model

variables are subjected to the sources of uncertainty, like errors of measurement, absence of information and poor or partial understanding of the driving forces and mechanisms. This imposes a limit on our confidence in the response of the model. Also, models may have to cope with the natural intrinsic variability of the system, such as the occurrence of stochastic events.

Almost all of the existing equations for sediment transport problems are empirical in nature due to such limitations. However, regression and dimensional analyses have been used extensively in the past. These approaches have certain limitations that keep them from being used widely for field applications. Regression has inadequacies pertaining to a first-hand functional form determination and clustering effect of influential points and groups of points. Dimensional analysis is also inadequate due to high number of variables and problems of multiple forms of the same equation. In problems of river hydrology, the system often reflects a stochastic nature and the variables cannot be measured without uncertainty. It has therefore been realized that there is a need for developing new and robust models that can overcome the restrictions posed by the conventional techniques.

Soft computing is an emerging paradigm based on the backbone of artificial intelligence, evolutionary/bio-inspired computing and probabilistic computing. These allow developing of statistical

\* Corresponding author. Tel.: +91 361 2582420.

E-mail address: [bimk@iitg.ernet.in](mailto:bimk@iitg.ernet.in) (B. Kumar).

**Notation**

$y$	channel height	$\tau_c$	critical bed shear stress
$S_f$	friction slope	$h$	flow depth
$\rho$	sediment density	$D$	diameter of cylindrical vegetation
$G$	specific gravity	$k$	height of vegetation
$\gamma$	unit weight of water	$i$	channel slope
$\gamma_s$	unit weight of sediment	$m$	number of vegetation cylinders per unit horizontal area
$d$	sediment particle diameter	$C_d$	drag coefficient
$\nu$	kinematic viscosity	$g$	acceleration due to gravity
$u$	mean flow velocity	$Gr$	gradation
$C$	sediment concentration	$R^2$	coefficient of correlation
$Q$	water discharge	$I_d$	index of agreement
$B$	channel width	$E$	Nash–Sutcliffe efficiency
$\tau_b$	bed shear stress		

black-box models based entirely on historical data. Soft computing has been employed extensively in hydrology and hydraulics with varying applications. The suitability of application of soft computing comes from the fact that it allows for uncertainties in measured values. This is critical in river hydraulics due to the inadvertent uncertainties in measuring data from the field and while experimenting. The models developed are not expected to give 100% accurate results but rather to be tolerant to errors in measurement and offer overall better predictability. These “black-box” models are purely statistical models and model parameters are adjusted by providing training data so as to give predictions for independent and new inputs. Primarily, soft computing techniques include artificial neural network (ANN), fuzzy logic, genetic algorithms (GA), particle swarm optimization (PSO), etc. Several soft computing models have been developed in the past. Adib (2008) used ANN for determining water surface elevation in tidal rivers. Sediment load prediction was carried out by Altunkaynak (2009) using genetic algorithms. Goel and Pal (2009) have used support vector machine in scour prediction. GA was also used for parameter identification for modeling river network by Tang et al. (2010). Kumar et al. (2010) has used Radial Basis Function model to design an incipient channel with bed suction. Kumar and Rao (2010) has used metamodel to predict friction factor in alluvial channel. Application of neural networks and fuzzy logic models to long-shore sediment transport was carried out by Samani et al. (2011). Amirabdollahian et al. (2011) used fuzzy genetic algorithm for optimal design of water networks. Kumar (2011) has used ANN model for friction factor prediction in alluvial channel. Krishna et al. (2012) used a wavelet neural network model for river flow time series. Kumar (2012) has applied soft computing technique for bed material load prediction. Ismail et al. (2013) have applied a feed-forward neural network to predict bridge scour. Other recent relevant work done in the field of river hydraulics by employing soft computing techniques include those of Kisi and Hosseinzadeh (2012) for modeling rainfall–runoff process, Kisi and Hosseinzadeh (2012) for suspended sediment modeling, Shiri et al. (2012) for forecasting daily stream flow. Shiri and Kisi (2012) also estimated daily suspended sediment load using wavelet conjunction models. A comparative study was completed by Kisi and Shiri (2012) in river suspended sediment estimation by climatic variables implication where various soft computing techniques were compared.

Genetic programming (GP) proposed in Koza (1992) views the modeling problem as one of program discovery. Genetic programming is a relatively newer domain in soft computing and has gained popularity in a variety of applications, including those in river hydraulics and sediment dynamics in fluvial systems. Singh et al. (2007) applied neural network–genetic programming for

sediment transport. Azamathulla et al. (2008) used genetic programming to predict ski-jump bucket spill-way scour. Aytekin and Kisi (2008) attempted sediment modeling using a genetic programming approach. Kisi and Guven (2010) carried out suspended sediment concentration estimation using a machine code-based genetic programming. Chang et al. (2012) used linear genetic programming for discharge prediction in compound channels. Kisi and Hosseinzadeh (2012) developed suspended sediment models using genetic programming. The paradigm of genetic programming attempts a search for the best program from a search space of programs by evolving generations of genetically bred and mutated populations of programs (mathematical expressions). Indeed, the modeling problem requires one to develop models which may well be an explicit function of the independent variables. In this, the approach of genetic programming differs from that of artificial neural models which do not present an explicit expression and rather utilize a number of network parameters to transform inputs to outputs. However, both ANN and GP help develop black-box models which are not based on the underlying physics or the phenomena of the system but are purely statistical. Genetic programming is different from conventional regression. Rather than finding numeric coefficients of a predetermined functional form as done by regression, symbolic regression attempts to find a symbolic expression containing both, functions as well as independent variables and numeric coefficients. The method is also referred to as symbolic function identification. The major difference lies in the fact that unlike conventional regression, GP does not require predetermined functional forms. Instead, it accepts the library of operators (functions and variables) and evolves generations of expressions to ultimately reach the best expression. The term symbolic regression is used for any technique which fits the measured data using a suitable mathematical formula. GP employs a search heuristic where the algorithm begins with randomized sets of expressions and creating new expressions in each generation (iteration) which perform better than the previous generation. Hence, the expressions are not calculated but generated from parent expressions using the genetic operators (mutation, crossover, etc.). The only calculations that take place are evaluations of expressions to assess their performance. This is done using model performance indicators (correlation coefficient, etc.) on the training data. The indicator helps to assess to what degree the model has been able to generalize the training dataset statistically. A good correlation coefficient, for example, would indicate a good generalization. These river hydraulics models are highly complex, and therefore their underlying relationships may be poorly understood. In such cases, the model can be viewed as a black box, i.e. the output is an opaque function of its inputs.

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات