

## Prediction of bed load transport in Kurau River based on genetic programming

ZAHRA ZANGENEH SIRDARI, *PhD Student, River Engineering and Urban Drainage Research Centre (REDAC), Universiti Sains Malaysia, Engineering Campus, Seri Ampangan, 14300 Nibong Tebal, Penang, Malaysia. Email: zzs10\_red06@student.usm.my*

AMINUDDIN AB. GHANI, *Professor & Deputy Director, REDAC, Universiti Sains Malaysia, Engineering Campus, Seri Ampangan, 14300 Nibong Tebal, Penang, Malaysia. Email: redac02@eng.usm.my*

ZORKEFLEE ABU HASAN, *Senior Lecturer, REDAC, Universiti Sains Malaysia, Engineering Campus, Seri Ampangan, 14300 Nibong Tebal, Penang, Malaysia.. Email: redac04@eng.usm.my*

### ABSTRACT

This paper investigates the ability of genetic programming (GP) as a new approach for estimation of bed load transport in Kurau River in Malaysia. Data collections including bed load and flow discharge have been carried out at six study sites which are situated at Kuru River. The values of  $R^2$ , RMSE, and MAE are equal to 0.989, 0.055, and 0.045, respectively, for training sets and 0.915, 0.090, and 0.069, respectively, for testing sets. For all data sets,  $R^2$  is equal to 0.961 while RMSE and MAE are equal to 0.082 and 0.060 respectively. The values of  $R^2$ , RMSE, and MAE for validation data set were obtained equal to 0.933, 0.096, and 0.067, respectively. In fact, the evolved model has obtained high accuracy for both testing and validation sets in order to confirm that enough generalization was obtained. Therefore it was found that the GP model could be employed successfully in predicting bed load transport.

*Keywords:* Bed load transport; genetic programming; RMSE; MAE.

### 1 Introduction

The difficulties associated with bedload measurement field causes a long history of interest in developing equations for the prediction (e.g., duBoys, 1879; Schoklitsch, 1934; Bagnold, 1980).

Formula is usually based on hydraulic principles of ideal and attempted to relate the level of bedload transport to some correlate of flow (such as water discharge, shear stress or stream power) (Yvonne Martin, 2003). Because of relationship between the reliability and representativeness of the data utilized in defining reference values, constants, and other relevant coefficients and the performance of a particular equation, most sediment transport equations do not represent a fundamental or complete correlation and it is really difficult, if not possible, to recommend a single formula for engineers and geological to use in the field under all conditions. (Wu et al., 2000; Camenen and Larson, 2005).

Genetic programming, which is a powerful tool for pattern recognition and data interpretation, was employed in this study to present an explicit predictive equation for bed load transport. Genetic Programming (GP) the extension of genetic algorithms (Poli, Langdon, & McPhee, 2008) is one of the well-known methods in artificial intelligence (AI) that plays an important role in modeling and simulation of many non-convex and complex phenomena to explain nonlinear relationships between some parameters (Liu

et al., 2002; Tabesh & Dini, 2009; Naseri, M et al., 2011).

Prior to its inherent optimized behavior and touchable resulted formula, GP has been applied to a wide range of problems in artificial intelligence, engineering and science applications, industrial, and mechanical models such as water resources, hydraulic processes and electricity demand, etc. (Azamathullah et al., 2010; Ashour et al., 2003; Babovic and Bojkov, 2001; Babovic and Keijzer, 2003; Drecourt, 1999; Harris et al., 2003; Keijzer and Babovic, 2002; Khu et al., 2001; Lee et al., 1997; Muttill and Lee, 2005; Sivapragasam et al., 2007; Whigham and Crapper, 2001 and Zhang et al., 2005).

A number of applications of GP have been reported, which include sediment transport modelling, (Babovic and Abbott, 1997); effect of flexible vegetation on flow in wetlands (Harris et al., 2003; Babovic and Keijzer, 2000); sedimentary particle settling velocity equations (Babovic et al. 2001); emulating the rainfall runoff process (Whigham and Crapper, 2001; Liong et al., 2002) evolutionary computation approach to sediment transport modeling (Kizhisseri et al., 2005). Multigene genetic programming is a recently developed approach for enhancing accuracy of GP that was developed by Hinchliffe et al. (1996) and Hiden (1998).

In multigene GP, several trees may describe the model but in traditional GP a single tree expresses the model and amount of tree that can be employed is the main difference between the multigene GP and GP.

All of genes have specific optimal weights and summation of weighted genes plus a bias term would form the final formula as the best obtained numerical model. Multigene GP can be written as:

$$Y = d_0 + d_1 * gene_1 + d_2 * gene_2 + \dots + d_n * gene_n$$

Where  $d_0$  is the bias term and  $d_i$  is weight of the  $i$ th gene. Actually, multigene GP is a linear combination of nonlinear terms, and this feature makes it possible to recognize the pattern of engineering problems in a highly precise manner (Hinchliffe et al., 1996).

GPTIPS was utilized in the current study to perform a multigene genetic programming for precise estimation of bed load transport. This is a new “Genetic Programming & Symbolic Regression” code written based on multigene GP for use with MATLAB (Searson, 2009a).

## 2 Site description

Kurau River sub-basin (Figure 1) lies between latitude 570,000 (N) and 530,000 (N), longitude 290,000 (E) and 330,000 (E). The catchment area is approximately 40,000 hectares, consisting of two main river tributaries namely Kurau River and Ara River. Kurau River represents the main drainage artery of the basin, draining an area of approximately 682 km<sup>2</sup> that is generally low lying. The river originates partly in the Bintang Range and partly in the Main Range where the terrain in the upper reaches is steep and mountainous. Mid valleys of the river are characterized by low to undulating terrain, which give way to broad and flat floodplains. Ground elevations at the river headwaters are moderately high, being 1,200 m and 900 m. The slopes in the upper 6.5 km of the river averaged 12.5% whilst those lower down the valleys are much lower, of the order of 0.25% to 5%.



Figure 1 Kurau River catchment

## 3 Field method

Six study sites were chosen in the kurau River for data collection. Hydraulic and bed load data were collected for detailed analysis on river (Ab. Ghani et al., 2003). All cross-sections in the kurau River are single thread channel width. The top with ranging between 7 and 19 m and the slops are between 0.00076 and 0.001 indicating small-medium river and the cross-sections are still natural.

During the field season, each section was sampled 6 times. At the beginning of each sampling event, water surface height was surveyed. The channel cross- section

was then divided into 8 equally spaced increments on the basis of flow width at time sampling. At each increment, flow depth and velocity was measured. Bed load was sampled immediately after velocity at each increment. Bedload was collected with a Helley-Smith bedload sampler which had a square 7.6 cm orifice, 0.25 mm mesh bag with frame. Table 1 shows the range of data collected from kurau River. According to U.S. Geology Survey protocol, in order to account for immediate transport variation while avoiding the risk of clogging the sampler bag, sampling time with a Helley-Smith is typically between 30 and 60 s (Glysson, 1993). Emmett (1980) has found that sampler efficiencies drop

off the sampler is 40% full. Otherwise, the sampling time up to 10 min has been used under low-flow conditions (Ashworth and Ferguson, 1980).

These data used for estimation of bed load rate in GP. The measured bed load rating curves for these six sites are illustrated in Figure 2.

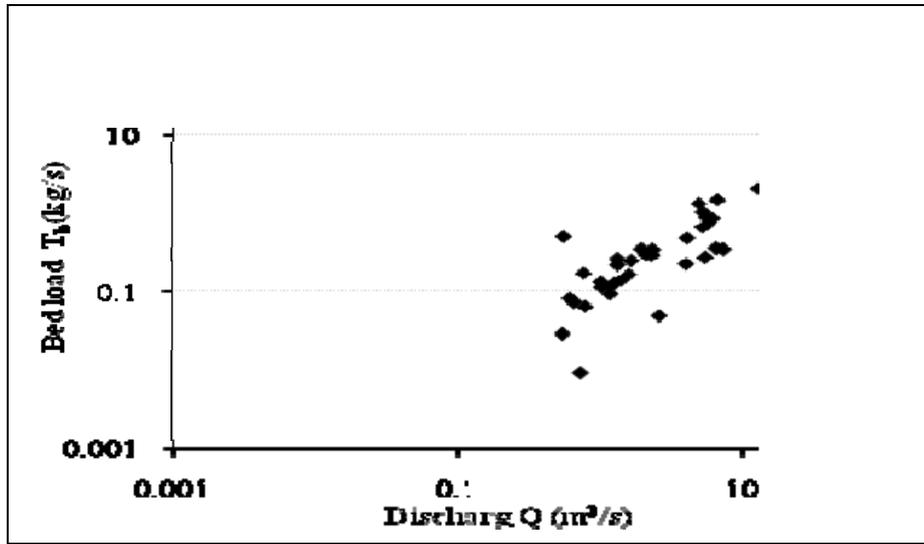


Figure 2 Bed load rating curve

Table 1 Range of Field Data for Kurau River

Study Sites	Discharge, Q (m <sup>3</sup> /s)	V (m/s)	So	B (m)	Y <sub>0</sub> (m)	A (m <sup>2</sup> )	R	d <sub>50</sub> (mm)	Bed Load Transport Q <sub>b</sub> (kg/s)
ku2	3.99-12.79	0.54-0.82	0.0005-0.007	17-19	3.99-12.79	7.2-15.51	0.412-0.885	0.65-1.044	0.23-2.098
ku6	1.6-6.1	0.5-0.73	0.0007-0.0185	9-10.3	1.6-6.1	2.87-8.37	0.313-0.76	0.699-1.084	0.168-
ku8	0.55-1.32	0.31-0.48	0.0006-0.0096	7-9.2	0.55-1.32	1.39-2.84	0.166-0.303	0.99-1.404	0.028-
ku11	0.56-2.59	0.15-1.22	0.001-0.0062	12-13	0.56-2.59	1.99-3.75	0.161-0.286	1.02-1.83	0.009-
ku12	2.32-6.6	0.53-1.56	0.0003-0.0051	12-13	1.27-6.6	2.57-9.78	0.224-0.699	0.74-1.51	0.128-
Ara5	3.99-12.79	0.4-0.69	0.0003-0.0312	11.3-13	0.77-5.25	1.94-7.57	0.167-0.567	1.29-1.84	0.116-1.04

#### 4 Derivation of bed load models based on GP

A GPTIPS run with the following settings was performed: Population size = 500, Number of generations = 25, Tournament size = 7 (with lexicographic selection pressure), D<sub>max</sub> = 3, G<sub>max</sub> = 4, Elitism = 0.01 % of population, function node set = {plus, minus, times, protected/}. The default GPTIPS multigene symbolic regression function was used in order to minimize the root mean squared prediction error on the training data. (Searson, 2009b)

The following (default) recombination operator event probabilities were used: Crossover events = 0.85, mutation events = 0.1, direct reproduction = 0.05. The following sub-event probabilities were used: high level crossover = 0.2, low level crossover = 0.8, subtree mutation = 0.9, replace input terminal with another random terminal = 0.05, Gaussian perturbation of randomly selected constant = 0.05 (with standard deviation of Gaussian = 0.1) (Table 2).

These settings are not considered 'optimal' in any sense but were based on experience with modelling other data sets of similar size.

In genetic programming, as in any data-driven prediction model, the selection of appropriate model input variables is extremely important. The choice of input variables is generally based on a priori knowledge of causal variables and physical/ecological insight into the problem. Moreover, the use of lagged input variables also leads to better predictions in a dynamical system (Muttill and Lee, 2005).

Four input parameters as variable data including discharge(Q), water surface slop(s<sub>0</sub>), mean grain size(d<sub>50</sub>) and shields' parameter for initiation of motion (θ) as variable data and T<sub>b</sub> (bed load rate) as invariable data were used in the current study where :

$$T_b = f(Q, S_0, d_{50}, \theta)$$

$$\theta = RS_0 / (G_s - 1)d_{50} \quad (1)$$

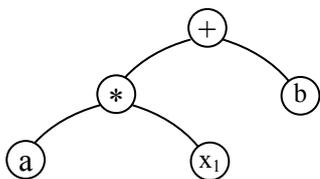
Table 2 Multigene GP Range of Initially Defined Parameters

Parameter	Range
Population size	500
Function set	+, -, *, /
Number of generations	25
Maximum number of genes	4
Maximum number of nodes per tree	13
Maximum depth of trees	3
Probability of GP tree mutation	0.1
Probability of GP tree cross over	0.85
Probability of GP tree direct copy	0.05

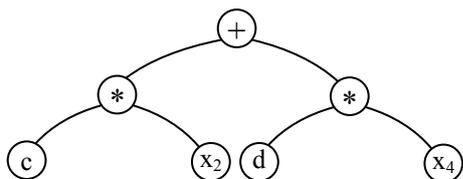
### 5 Results and discussions

Numerous runs were performed with various initial settings and the performance of the developed equation was analyzed for each run. Consequently, the best models were selected according to statistical criteria such as R<sup>2</sup>, RMSE, and MAE. In addition, a comprehensive parametric study was performed to monitor the behavior of each model versus variations of input variables. Figure 3 shows the expression of genes for GP formulation.

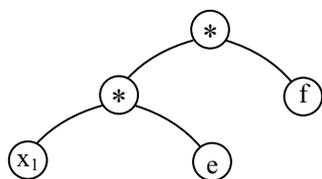
Gene 1 and base term:



Gene 2:



Gene 3:



Gene 4:

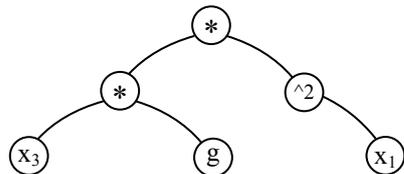


Figure 3 Expression genes for GP formulation

Accordingly, the following relationship was selected as the best model for bed load transport:

$$T_b = 0.03903 x_1 + 0.2587(x_2 + x_4) - 26.29 x_2 x_1 + 0.01781 x_1^2 x_3 - 0.009905 \quad (2)$$

Where  $T_b$  is bed load transport rate (kg/s) and  $x_i$ s are defined as cited in Table 3.

Table 3 Definition of Parameters Used in Eq. (2)

symbol	Definition
$x_1$	Q (m <sup>3</sup> /s)
$x_2$	S <sub>0</sub>
$x_3$	d <sub>50</sub> (mm)
$x_4$	θ

Precision of the developed equation is examined by plotting the measured versus predicted values of t bed load rate for training, testing, and all data as shown in (Figure 4), (Figure 5) and (Figure 6), respectively. The values of R<sup>2</sup>, RMSE, and MAE are equal to 0.989, 0.055, and 0.045, respectively, for training sets (Figure 4) and 0.915, 0.090, and 0.069, respectively, for testing sets (Figure 5).

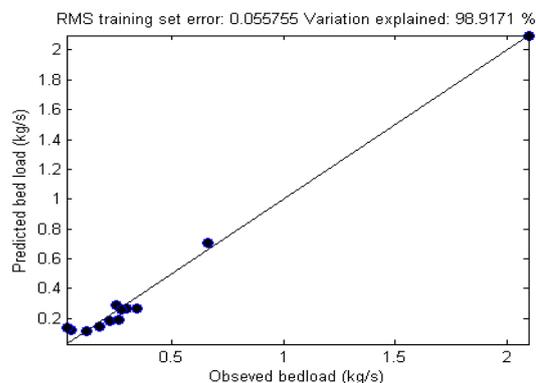


Figure 4 Measured versus predicted values of  $T_b$  for training data set

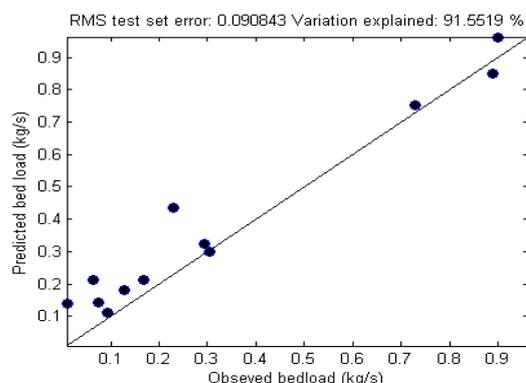


Figure 5 Measured versus predicted values of  $T_b$  for testing data set

For all data sets (Figure 6),  $R^2$  is equal to 0.961 while RMSE and MAE are equal to 0.082 and 0.060 respectively. Figure 7 illustrates the measured versus predicted values of bed load rate for validation set. The values of  $R^2$ , RMSE, and MAE for this data set were obtained equal to 0.933, 0.096, and 0.067, respectively. In fact, the evolved model has obtained high accuracy for both testing and validation sets in order to confirm that enough generalization was obtained. Figure 8 shows the frequency of input data, the selected parameter have the same weight of frequency, therefore all parameters affected mostly the same on predicted function of genetic programming.

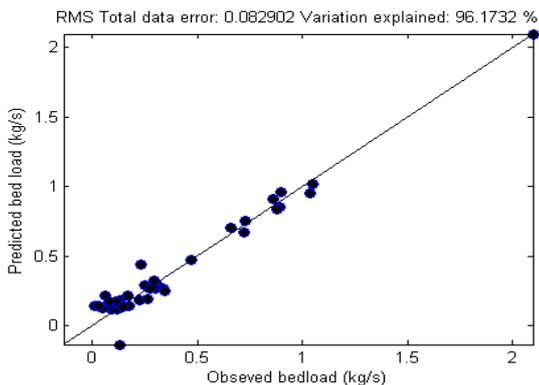


Figure 6 Measured versus predicted values of  $T_b$  for all data set

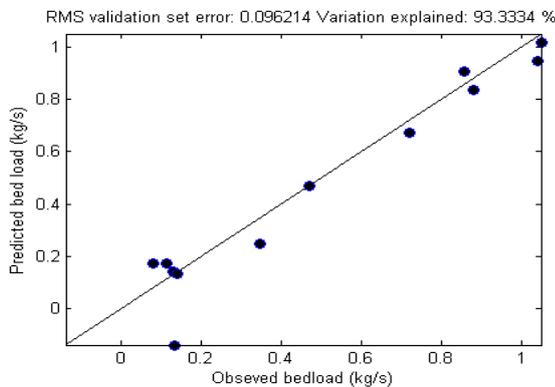


Figure 7 Measured versus predicted values of  $T_b$  for validation data set

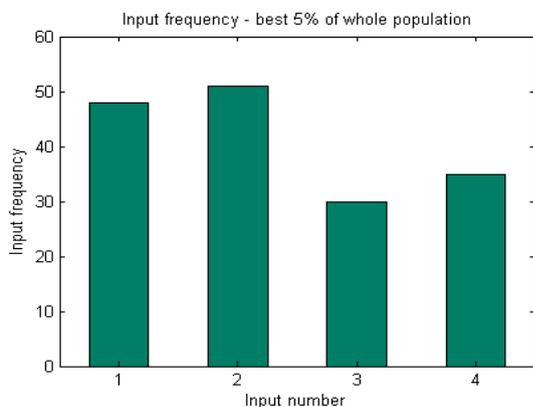


Figure 8 Frequency of input data –best 5% of whole population

## 5 Conclusions

Correct estimation of sediment volume carried by a river is very important for many water resources projects. The prediction of river sediment load also constitutes an important issue in hydraulic engineering. This study presents the prediction of bed load transport in Kurau River in Malaysia by using genetic programming.

The developed model has shown reasonable performance under field conditions according to the verifications demonstrated in Figure 7. From the results of prediction of bed load with  $R^2= 0.96$  and root mean square error (RMSE= 0.082) and mean absolute error (MAE=0.06) respectively, it can be concluded that genetic programming can be used to predict bed load transport rate for Kurau River with only 4 parameters. The simplified equation is recommended as alternative equation for Kurau River in Malaysia for predicts the bed load transport.

## References

1. Ab. Ghani, A., Zakaria, N.A., Abdullah, R., Chang, C.K., Sinnakaudan, S. K. & Mohd Sidek, L. (2003). *River Sediment Data Collection and Analysis Study*, Contract Research No. JPS (PP)/SG/2/2000, Department of Irrigation and Drainage, Malaysia, Kuala Lumpur.
2. Ashour, A. F., Alvarez, L. F., & Toropov, V.V. (2003). Empirical modelling of shear strength of RC deep beams by genetic programming. *Computers and Structures*, 81, 331–338.
3. Ashworth and Ferguson (1989). Size-selective entrainment of bed load in gravel bed streams. *Water Resour. Res.* 25, 627–634.
4. Azamathulla HMD, Ab Ghani A, Zakaria NA, Chang C.K, Abu Hassan Z. Genetic programming approach to predict sediment concentration for Malaysian Rivers. *Int J Ecol Econ Stat* 2010b;16(10):53–64.
5. Bagnold, R.A. (1980). *An empirical correlation of bedload transport rates in flumes and natural rivers*. Proceedings of the Royal Society of London A372, 453–473.
6. Babovic, V., & Abbott, M. B. (1997). The evaluation of equations from hydraulic data, Part II: Applications. *Journal of Hydraulics Research*, 35(3), 411–430.
7. Babovic, V., & Bojkov, V. H. (2001). *Runoff modeling with genetic programming and artificial neural networks*, DHI. Technical Report D2k TR 0401-1.
8. Babovic, V., & Keijzer, M. (2000). Genetic programming as a model induction engine. *Journal of Hydroinformatics*, 2(1), 35–60.
9. Babovic, V., & Keijzer, M. (2003). Rainfall runoff modeling based on genetic programming. *Nordic Hydrology*, 34(1), 331–346.

10. Camenen, B. and M. Larson (2005). A General Formula for Non- Cohesive Bed Load Sediment Transport. *Estuarine Coastal and Shelf Science*, 63,249-260.
11. Drecourt, J. P. (1999). *Application of neural networks and genetic programming to rainfall runoff modeling*, DHI. Technical Report D2k 0699-1.
12. DuBoys, M.P (1879). Etudes du régime et l'action exercé par les eaux sur un lit à fond de graviers indéfiniment affouille. *Annales des Pontset Chaussées*.5, 141–195.
13. Emmett, W.W (1980). *Emmett, A field calibration of the sediment-trapping characteristics of the Helley-Smith bedload sampler*. USGS Prof. Pap.1139, 44.
14. Glysson, G.D (1993). U.S. Geological Survey bedload sampling policy. In: H.W. Shen, S.T. Su and F. Wen, Editors, *Hydraulic Engineering '93*. ASCE 1 701–706.
15. Harris, E. L., Babovic, V., & Falconer, R. A. (2003). Velocity predictions in compound channels with vegetated floodplains using genetic programming. *Journal of River and Basin Management*, 1(2), 117–123.
16. Hiden, H.G. (1998). Data-Based Modeling Using Genetic Programming. *Ph.D. Dissertation*, Department of Chemical and Process Engineering, University of Newcastle-upon-Tyne, UK.
17. Hinchliffe, M.P., Willis, M.J., Hiden, H.G., Tham, M.T., McKay, B., Barton, G.W (1996). *Modelling chemical process systems using a multi-gene genetic programming algorithm*. In: Koza, J. (Ed.) Proceedings of the First Annual Conference on Genetic Programming (late breaking papers), pp. 56–65.
18. Keijzer, M., & Babovic, V. (2002). Declarative and preferential bias in GP-based scientific discovery. *Genetic Programming and Evolutionary Machines*, 3(1), 41–79.
19. Kizhisseri AS, Simmonds D, Rafiq Y, Borthwick M. *An evolutionary computation approach to sediment transport modeling*. Fifth International Conference on Coastal Dynamics; 2005. April 4–8, Barcelona, Spain.
20. Khu, S. T., Liong, S. Y., Babovic, V., Madsen, H., & Muttill, N. (2001). Genetic programming and its application in real-time runoff forecasting. *Journal of American Water Resources Association*, 37(2), 439–451.
21. Liu, J., Savenije, H. G., & Xu, J. (2002). Forecast of water demand in Weinan city in China using WDF-ANN model. *Physics and Chemistry of the Earth*, 28, 219–224.
22. Liong, S.Y., Gautam, T.R., Khu, S.T., Babovic, V., Muttill, N.(2002). Genetic Programming: a new paradigm in rainfall-runoff modelling. *J. Amer. Water Res. Assoc.* 38 (3), 577–584.
23. Martin, Y (2003). Evaluation of bed load transport formulae using field evidence from the Vedder River, British Columbia, *Geomorphology*, 53, 73–95.
24. Muttill, N., & Lee, J. H. W. (2005). Genetic programming for analysis and real-time prediction of coastal algal blooms. *Ecological Modelling*, 189, 363–376.
25. Nasserli, M., Asghari, K., & Abedini, M. J. (2008). Optimized scenario of rainfall forecasting using genetic algorithms and artificial neural networks. *Expert Systems with Applications*, 35(3), 1415–1421.
26. Ploi, R., McPhee, N.F (2008). *Covariant parsimony pressure for genetic programming*. Technical report CES-480, ISSN:1744-8050.
27. Schoklitsch, A (1934). Der geschiebetrieb und die geschiebefracht. *Wasser kraft Wasser wirtschaft*. 5, 1–7.
28. Searson, P., (2009a). GPTIPS: Genetic Programming & Symbolic Regression for MATLAB. /http://gptips.sourceforge.net (accessed 10 February, 2010).
29. Searson, D.P., (2009b). GPTIPS: Genetic Programming & Symbolic Regression for MATLAB, User Guide. (Accessed 10 February, 2010).
30. Sivapragasam, C., Vincent, P., & Vasudevan, G. (2007). Genetic programming model for forecast of short and noisy data. *Hydrological Processes*, 21, 266–272.
31. Tabesh, M., & Dini, M. (2009). Fuzzy and neuro-fuzzy models for short-term water demand forecasting in Tehran. *Iranian Journal of Science and Technology, Transaction B: Technology*, 33, 61–77.
32. Whigham, P. A., & Crapper, P. F. (2001). Modeling rainfall-runoff using genetic programming. *Mathematical and Computer Modelling*, 33, 707–721
33. Wu, W., S.S.Y. Wang, and Y. Jia (2000). Nonuniform Sediment Transport in Alluvial Rivers, *Journal of Hydraulic Research* 38(2), 427–434.
34. Zhang, L., Jack, L. B., & Nandi, A. K. (2005). Fault detection using genetic programming forecasting. *Mechanical Systems and Signal Processing*, 19, 271–289.