World Academy of Science, Engineering and Technology
International Journal of Civil and Environmental Engineering
Vol:17, No:7, 2023

Determining of Stage-Discharge Relationship for Meandering Compound Channels Using M5 Decision Tree Model

Mehdi Kheradmand, Mehdi Azhdary Moghaddam, Abdolreza Zahiri, Khalil Ghorbani

Abstract**—**In modeling phenomena, the presence of local conditions may cause the use of a general relation not to produce good results and thus fail to demonstrate local changes. If possible, identifying homogenous limits and providing simple linear relations for each of these limits will increase the accuracy of models. Accordingly, the models are divided into simpler and smaller problems to solve complicated problems, and the obtained answers will be combined. This simple idea can be applied to decision tree models. For this aim, the input data values are divided into several sub-intervals or sub-regions, and an appropriate model is extracted for an appropriate model or equation. This research proposes the M5 decision tree method as a solution to accurately compute the flow discharge in meandering compound channels.

Keywords**—**Stage-discharge relationship, decision tree, M5 decision tree model, meandering compound channels.

I. Introduction

THE decision trees serve as methods to demonstrate a series of laws that lead to a rank or value. Decision trees are made using subsequent separation of data into a separate group series, as attempts are made to increase the distance between the groups within the separation process. The structure of a tree model includes the root, internal nodes, and leaves. Decision tree models are used to solve many classification and regression problems.

II. Materials and Methods

A. The Data Used

To evaluate the accuracy of the M5 decision tree method four sets of data related to the four laboratory meandering compound channels have been used. The first and second sets of data are collected from two channels with a meandering compound section belonging to the University of Rourkela in India [4]. The third category is the data related to a channel with a meandering compound section used at the Loughborough University in England [5], [6]. The fourth category is the channel data with a meandering compound section related to the University of Glasgow [7].

B. Parameters Used

A list of the parameters used in this research to construct the M5 decision tree model has been placed in Table I.

TABLE I
LIST OF PARAMETERS USED IN THE RESEARCH

Parameter	Description
S_0	Longitudinal slope of the channel
L_w(m)	Length of a meandering wave
B_w(m)	Width of the meandering belt
F_w(m)	Total width
S_r	Degree of curvature of the channel (sinuous)
b(m)	The upper width of the main channel
h(m)	Depth of the full section of the main channel
S_{mc}	Slope of the walls
n_c	Manning's roughness coefficient of the main channel
n_f	Manning's roughness coefficient of the floodplain
H(m)	Total depth of the flow
Dr	Relative depth
Qm	Discharge measurement

C. Statistical Methods

This research used the root-mean-square-error (RMSE), average error (AE), coefficient of determination (R^2), and the mean absolute deviation (δ) as evaluation criteria for the results. These parameters are expressed as:

$$R^2 = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum (X-Y)^2}{n}} \tag{2}$$

$$AE = \frac{\sum \left(\frac{X-Y}{X}\right) * 10}{n} \tag{3}$$

$$\delta = \frac{\sum |X-Y|}{\sum X} * 100 \tag{4}$$

where $x = X - \bar{X}$, $y = Y - \bar{Y}$, X is the observed value, Y is the calculated value, $\bar{X}$ is the average observed value, $\bar{Y}$ is the average calculated value, and n is the number of data.

Mehdi Kheradmand is with Department of Civil Engineering, International University of Chabahar, Iran (corresponding author, phone: +98- 937-937-2010, e-mail: from@yahoo.com).
Mehdi Azhdary Moghaddam is with Department of Civil Engineering, University of Sistan and Baluchestan, Iran (e-mail: mazhdary@usb.ac.ir).
Abdolreza Zahiri and Khalil Ghorbani are with Department of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Iran (e-mail: zahiri@gau.ac.ir, kghorbani@ut.ac.ir).
Mohadeseh Kheradmand is with Department of Accounting, Islamic Azad University of Yazd, Iran (e-mail: mohadesehkheradmand@yahoo.com).

D. M5 Decision Tree Model

Quinlan (1992) was the first to provide the M5 Decision Tree for the prediction of continuous data. Unlike conventional decision tree models that provide discontinuous class or ranks, this model creates a multivariate linear model for the data at each node of the tree model. The formation of decision tree model structures includes stages of the creation of a tree and its pruning [1].

Zahiri and Ghorbani (2013) presented detailed relationships for calculating the total flow discharge in straight compound sections using the decision tree model [2].

In Fig. 1, the performance of M5 decision tree model for a hypothetical problem of each model indicates a linear regression equation [3]. For example, if $X_1>2.5$, and $X_2>2.5$, then the third model is used as below:

$$Y = a_0 + a_1X_1 + a_2X_2 \tag{5}$$

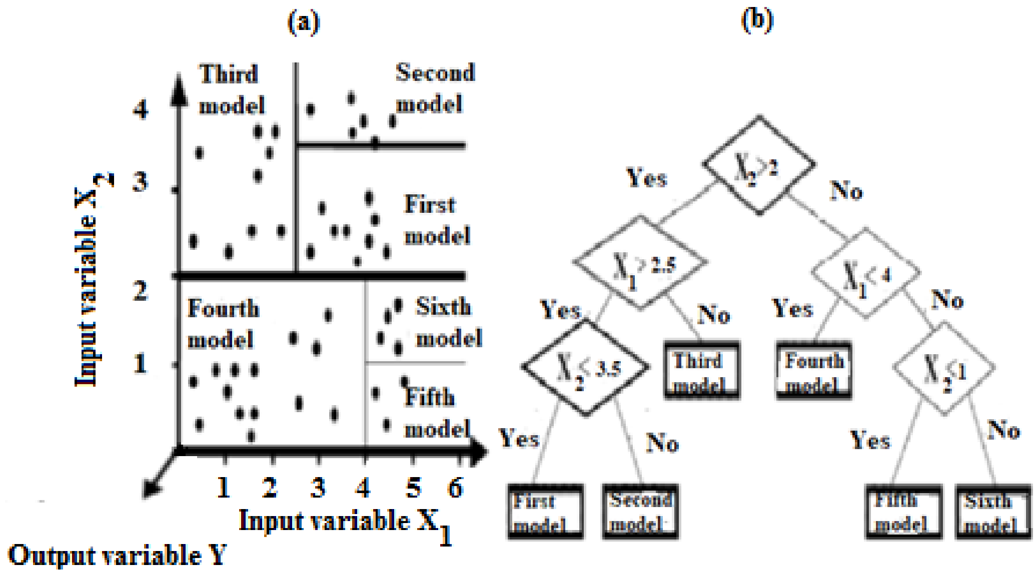

Fig. 1 M5 decision tree model performance (a) division of input parameter spaces (x1*x2) into 6 regions, (b) Describing the Division Criterion of Input Parameters Space in the Form of Tree

III. Results

Figs. 2-11 illustrate the results of the M5 decision tree model for the calibration data by considering various parameters of D_r, H_m, n_f, S_r, B_m, B_w, and L_w.

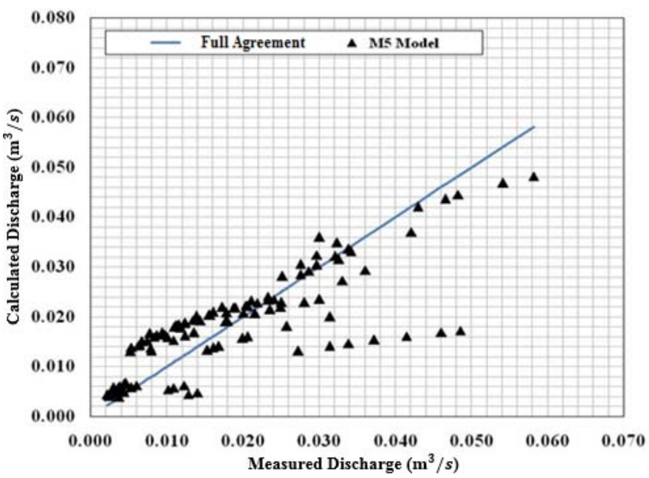

Fig. 2 M5 model results for calibration data to calculate the flow discharge (Dr parameter)

Where D_r is the relative depth, H_m is the total depth of the flow, n_f is manning's roughness coefficient of the floodplain, S_r is the degree of curvature of the channel, B_m is the upper width of the main section B_w is the width of the meandering belt, and L_w is the length of a meandering wave of the meandering channel [8].

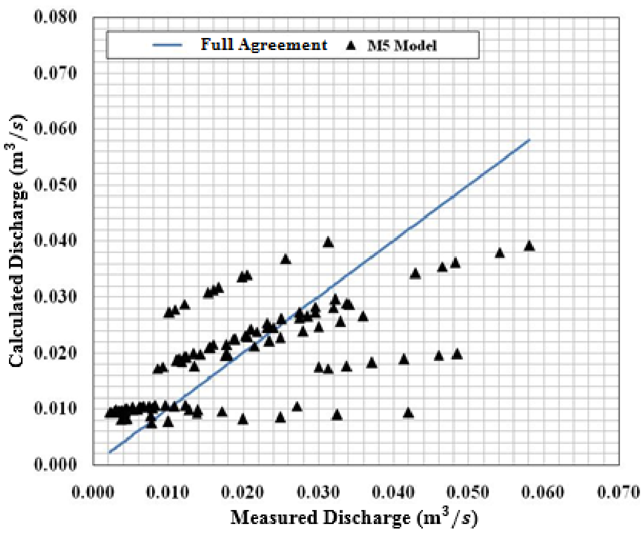

Fig. 3 M5 model results for calibration data to calculate the flow discharge (Hm parameter)

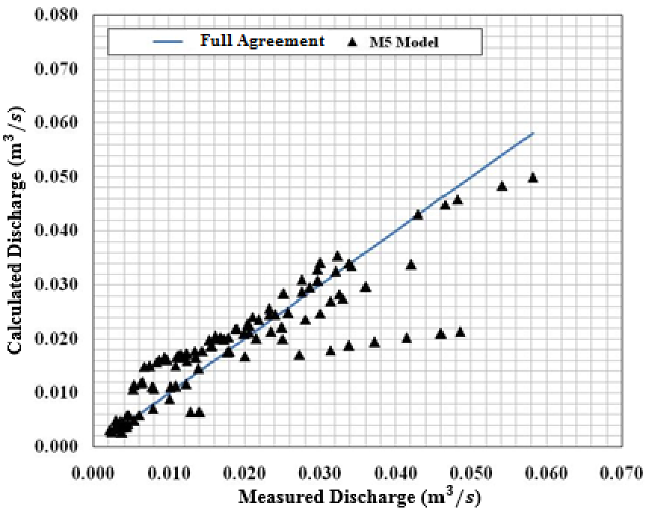

Fig. 4 M5 model results for calibration data to calculate the discharge (Dr, and Hm parameters)

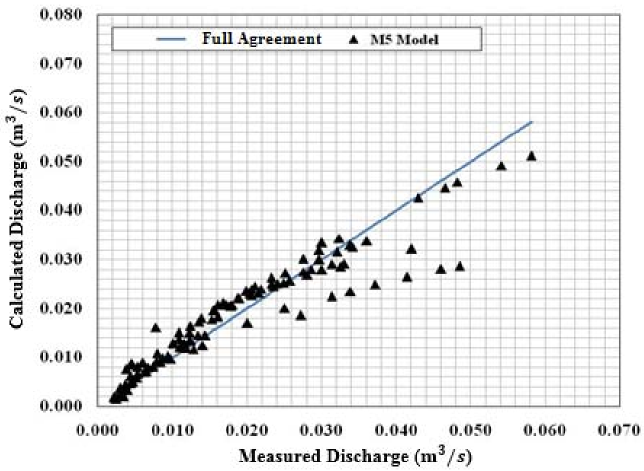

Fig. 5 M5 model results for calibration data to calculate the flow discharge (Dr, Hm, and nf parameters)

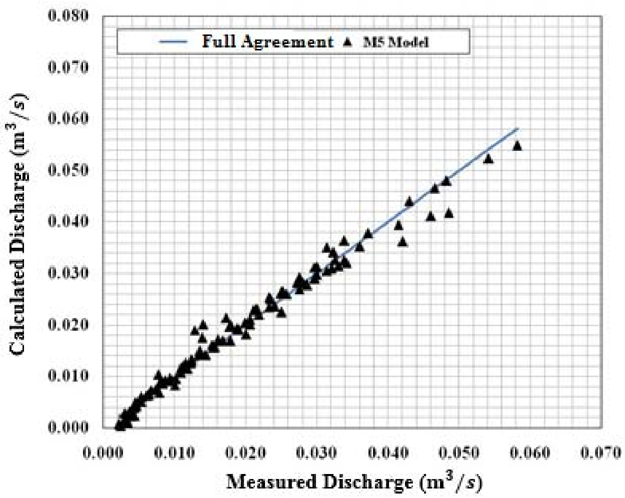

Fig. 6 M5 model results for calibration data to calculate the flow discharge (Dr, Hm, nf, and Sr parameters)

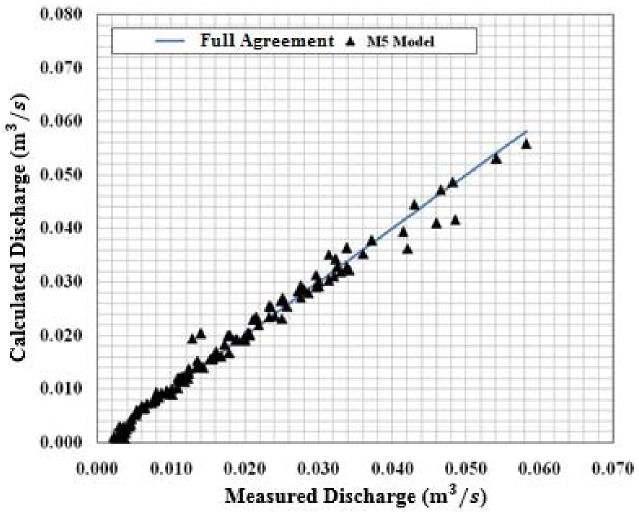

Fig. 7 M5 model results for calibration data to calculate the discharge (Dr, Hm, nf, Sr, and Bm parameters)

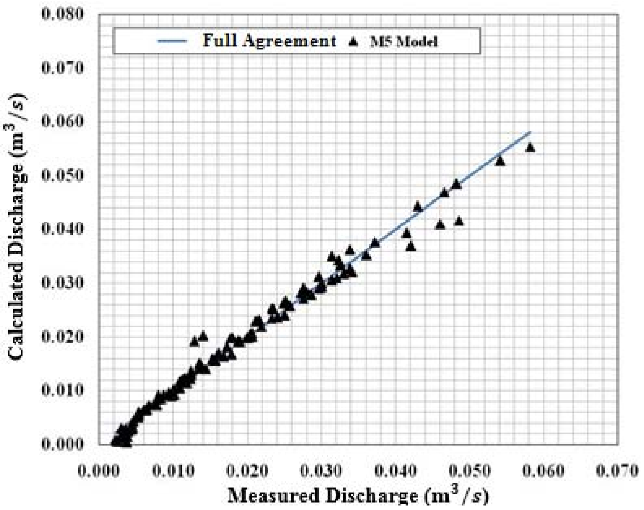

Fig. 8 M5 model results for calibration data to calculate the flow discharge (Dr, Hm, nf, Sr, and Bw parameters)

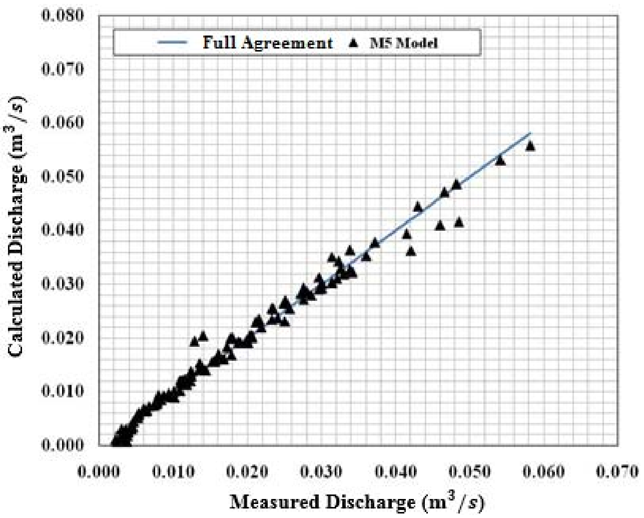

Fig. 9 M5 model results for calibration data to calculate the flow discharge (Dr, Hm, nf, Sr, and Lw parameters)

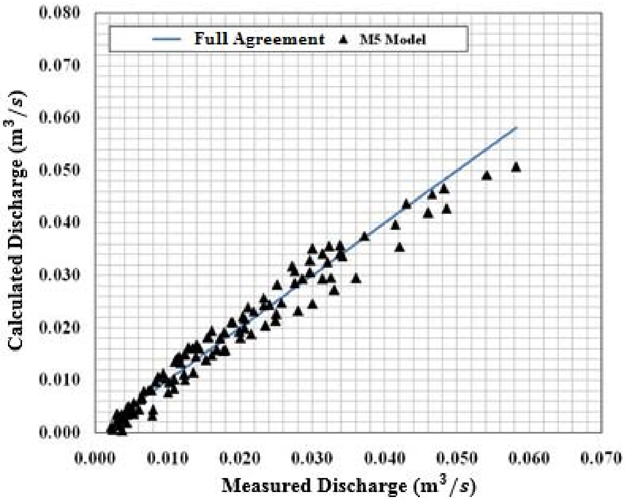

Fig. 10 M5 model results for calibration data to calculate the discharge (Dr, Hm, and Sr, parameters)

IV. Conclusion

To accurately investigate the subject, Table II shows the calculates statistical parameters for the M5 model by considering the use of various parameters. These results are given for the whole calibration data.

A review of results in Table II indicates that the M5 model with parameters such as Dr, H_m, n_f, S_r, and L_w has significantly increased the accuracy of the calculations, with the RMSE, coefficient of determination, AE and mean absolute deviation to 20%, 0.981, 14.35% and 6.36 respectively. To ensure the accuracy of the proposed model, parameters of the longitudinal slope and relative roughness of the compound sections (ratio of roughness coefficient of the floodplain to the main section) were also added to the previous parameters, and the M5 model was implemented again. The results suggested that a significant increase in the accuracy of statistical parameters was not achieved; for this, consistent with statistical parameter results of M5 models for the validation and calibration data, it was decided to introduce the M5 model with various parameters of Dr, H_m, n_f, S_r, and L_w as the optimal model.

TABLE II
Statistical Parameters of M5 Model Results for Calibration Data by Considering Used Parameters

Used parameters	R^2	RMSE	AE%	δ
D_r	0.66	0.87	52.34	27.56
H_m	0.431	1.12	81.51	39.15
D_r - H_m	0.778	0.71	35.85	21.35
D_r - H_m-n_f	0.890	0.52	25.39	15.48
D_r - H_m- Sr	0.960	0.29	21.16	11.05
D_r - H_m-n_f-Sr	0.979	0.21	15.83	6.83
D_r - H_m-n_f-Sr-B_m	0.981	0.20	14.35	6.36
D_r - H_m-n_f-Sr-B_w	0.981	0.20	14.88	6.24
D_r - H_m-n_f-Sr-L_w	0.981	0.20	14.35	6.36

Fig. 11 illustrates results from the optimal decision tree model compared to observed data. As noted, the results of this method have less error compared to the results of previous methods. In this chart, the dispersion of the data around the central points is much less than the previous methods, and this issue indicates a better match between the calculation results and the real data.

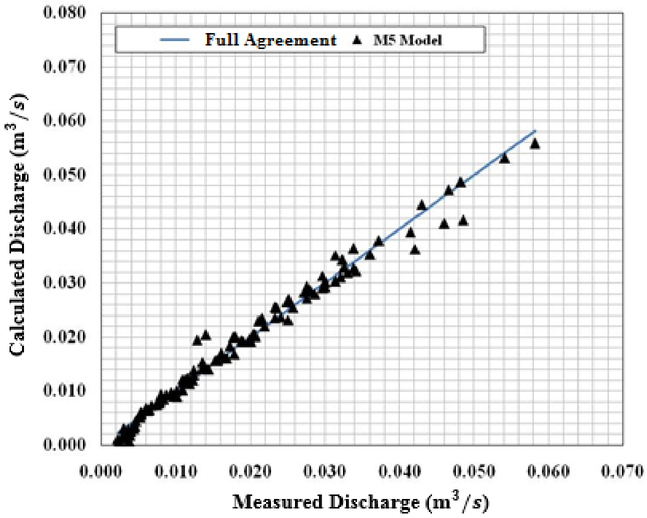

Fig. 12 Optimal tree model results (including Dr, Hm, nf, Sr, and Lw parameters)

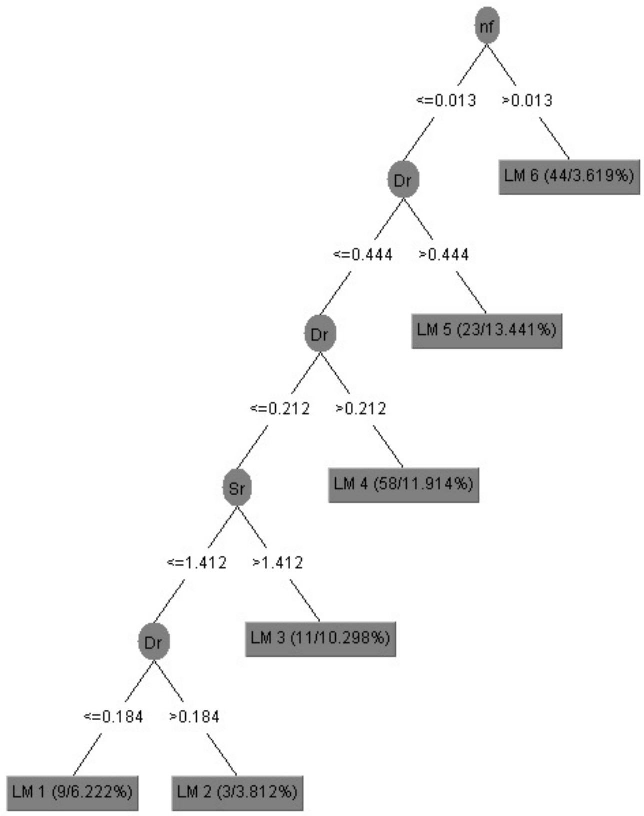

Fig. 12 The tree diagram of decision tree model for estimation of the flow discharge

In the optimal decision tree, for the input data following the running of rapidminer software, six linear regression equations were extracted, the tree diagram of these equations is given in Fig. 12. The conditions to use these linear equations are given in Fig. 13, and the form of the six equations is also given in Fig. 14. As noted in Figs. 12 and 13, the main criterion to divide the input data into several different equations is n_f, which indicates the roughness of the floodplain.

nf <= 0.013 :
| Dr <= 0.444 :
| | Dr <= 0.212 :
| | | Sr <= 1.412 :
| | | | Dr <= 0.184 : LM1 (9/6.222%)
| | | | Dr > 0.184 : LM2 (3/3.812%)
| | | Sr > 1.412 : LM3 (11/10.298%)
| | Dr > 0.212 : LM4 (58/11.914%)
| Dr > 0.444 : LM5 (23/13.441%)
nf > 0.013 : LM6 (44/3.619%)

Fig. 13 The conditions to use equations proposed by the decision tree model for estimation of the flow discharge

LM num: 1

Qm = 0.0025 * Lw(m) + 0.0311 * Sr - 1.2898 * nf + 0.1316 * H(m) + 0.0517 * Dr - 0.0479

LM num: 2

Qm = 0.0025 * Lw(m) + 0.0311 * Sr - 1.2898 * nf + 0.1316 * H(m) + 0.053 * Dr - 0.0479

LM num: 3

Qm = 0.0034 * Lw(m) + 0.0313 * Sr - 1.2898 * nf + 0.1328 * H(m) + 0.0673 * Dr - 0.0498

LM num: 4

Qm = 0.0035 * Lw(m) + 0.0526 * Sr - 2.295 * nf + 0.186 * H(m) + 0.0713 * Dr - 0.0796

LM num: 5

Qm = 0.0066 * Lw(m) + 0.0216 * Sr - 3.0655 * nf + 0.21 * H(m) + 0.0562 * Dr - 0.0303

LM num: 6

Qm = 0.0015 * Lw(m) + 0.0103 * Sr - 8.504 * nf + 0.351 * H(m) + 0.0124 * Dr + 0.0665

Number of Rules: 6

Fig. 14 The six equations proposed by the decision tree model for estimation of the flow discharge

REFERENCES

[1] J.R. Quinlan, "Learning with continuous classes," Proceedings of Australian Joint Conference on Artificial Intelligence, Hobart 16-18 November 1992, pp. 343-348.
[2] A. Zahiri and Kh. Ghorbani, "Flow discharge prediction in compound channels by using decision model tree M5", Journal of Water and Soil Conservation, Gorgan University of Agricultural Sciences and Natural Resources, vol. 20 (3), 2013, pp. 113-132
[3] D.P. Solomatine, and Y. Xue, "M5 model trees and neural networks: Application to flood forecasting in the upper reach of the Huai river in China," J. Hydrol. Eng, vol. 9 (6), 2004, pp. 1-10.
[4] K. K. Khatua, Interaction of Flow and Estimation of Discharge in Two Stage Meandering Compound Channels, Ph.D. Thesis, National Institute of Technology, Rourkela, India, 2008.
[5] K. Shiono, J. S. Al-Romaih and D. W. Knight. "Stage discharge assessment in compound meandering channels," Journal of Hydraulic Engineering, ASCE, vol. 125 (1), 1999. pp. 66-77.
[6] J. Spooner and K. Shiono. "Modelling of meandering channels for overbank flow," Proc. Institution of Civil Engineers, Water and Maritime Engineering, vol. 156 (3), 2003, pp. 225-233.
[7] A. B. Macleod, Development of Methods to Predict the Discharge Capacity in Model and Prototype Meandering Compound Channels, Ph.D. Thesis, Department of Civil Engineering, University of Glasgow, 1997.
[8] M. Kheradmand, Determining of Stage-Discharge Relationship for Meandering Compound Channels Using New Data Mining Methods, Master Thesis, Department of Civil Engineering, International University of Chabahar, Iran, 2013.