Mesopotamia Environmental Journal
Mesop. environ. j. 2016, Vol.2, No.3:57-66.

ISSN 2410-2598

Assessment and comparing of support vector machines model and regression equations for predicting alluvial channel geometry

Azadeh Gholami [1,2*] Hossein Bonakdari [1,2] Salma Fenjan[1]
Isa Ebtehaj [1,2]

1 Department of Civil Engineering, Razi University, Kermanshah, Iran
2Water and Wastewater Research Center, Razi University, Kermanshah, Iran

***Corresponding author**: gholamiazadeh1@gmail.com

To cite this article:
Gholami, A.; Bonakdasi, H.;.Fenjan, S and Ebtehal, I. Assessment and comparing of support vector machines model and regression equations for predicting alluvial channel geometry. *Mesop. environ. j.*, 2016, Vol. 2, No.3, pp. 57-66.

Received Date: 20/12/2016, **Accepted Date**: 3/2/2016, **Publishing Date**: 15/5/2016

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Abstract

Determine the stable channel geometry of the river is one of the most important topics in river engineering. Various relationships (based on statistical and theoretical methods) to predict the stable channels dimensions are expressed by many scientists. In this study, three Support Vector Machines (SVM) models are designed to predict width (w), depth (h) and slope (s) of stable channel. 85 cross-section river field data is used in training and testing models. The models input parameters are the flow discharge (Q), median sediment diameter (d_{50}) and affecting Shields parameter (τ^*). Furthermore, the width, depth and slope values are calculated by Afzalimehr regression relationship. Several statistical indexes are used to check the accuracy of the models in comparison with field data. Results show that SVM models with correlation coefficient (R) 0.86, 0.66 and 0.646 in width, depth and slope prediction respectively have a good agreement with observational data. Also, the models comparison show a considerably better performance of the SVM models over the available regressions equations with a mean absolute relative error (*MARE*) decreasing of 72%, 20% and 11% in width, depth and slope prediction, respectively. The presented methodology in this paper is a good approach in predicting cross section geometry of alluvial rivers also it can be used to design stable irrigation and water conveyance channels.

Keywords; SVM model, stable channel geometry, regression equation, observed data.

Introduction

Hydraulic geometry determination of a stable channel is studied by many scientists in the last hundred years. The most important geometric parameters of stable channel are apex width, average depth and longitudinal slope. Various relationships to predict the stable channels dimensions are expressed by many scientists. Hey et al. (1986) through multiple regressions method obtained width, depth and channel slope and concluded that in addition to the effective role of flow discharge, plant coverage of channel walls affect the channel width equation and the bed grain size affects the depth and channel slope equation as well [8]. Afzalimehr et al. (2010) using non-linear regression analysis obtained the width, depth and slope of the stable channel and concluded that the average grain size (d_{50}) and Shields parameter (τ^*) are not considered as effective variables in the channel width and depth prediction and only depends on the flow discharge [1]. Lee and Julien (2006) [12] achieved the hydraulic dimension of alluvial channels by non-linear regression and their results were compared with the Julien and Wargadalam (2006) results which was based on empirical methods and concluded that the regression equations results are in more compatible with experimental data [10]. In recent decades, soft computing methods to study the flow pattern of hydraulic structures, estimate and predict the complex phenomena in hydrology and hydraulics are used (e.g. in: curved channel (Gholami et al. 2014 [5]; Gholami et al. 2015a [6]; Gholami et al. 2015b [7]; Karimi et al. 2015 [11]); Sediment transportation (Ebtehaj & Bonakdari, 2013 and 2014a [3], [4]); Discharge capacity (Karimi et al. 2015)). Khadangi et al. (2009) used Multi-Layer Perceptron Neural Network (MLPNN) for modeling the alluvial channels hydraulic geometry and their results were compared with semi-empirical equations [15]. Tahershamsi et al. (2012) using MLPNN model predicted the stable channel width and compared their results with regression equations. Among all computational intelligence methods, the SVM model is a novel algorithm from the machine learning community [13]. Javadi et al. (2015) estimated river bed form dimension using Artificial Neural Network (ANN) and SVM models, their result show that SVM model had a higher capability for estimating and simulating height of the bed form than ANN model [9]. Beechie and Imaki (2015) Predict natural channel patterns in stable state based on landscape and geomorphic controls by SVM models [2]. Their models show good accuracy when consider all affected parameter such as discharge, valley confinement, sediment supply, and sediment caliber. In the present research, SVM model has been used to predict the width, depth and slope of stable channel. Afzalimehr's field data in 85 cross section rivers are used for training and testing models [1]. The Afzalimehr's regression equations in width, depth and slope prediction are used to evaluate efficiency of models [1]. Different statistical indexes are applied for evaluation of present models.

Materials and methods

Overview of SVM model

Support Vector Machine (SVM) was first introduced based on Vapnik theory and used for regression by Vapnik et al. (1997) [14]. SVM is based on the construction of the Lagrange multipliers equation. For a certain data set:

$$G = \{(x_i, d_i)\}_i^n$$

SVM estimated the following function:

$$f(x) = w\varphi(x) + b \tag{1}$$

Where d_i is the desired value, x_i is the input vector, n is the data patterns number $\varphi(x)$ is the high dimensional feature space mapping nonlinearly from input space (x) and the w and b are calculated using regularizing risk function defied as follows:

$$R_{SVMs}(C) = C\frac{1}{n}\sum_{i=1}^{n} L(d_i, y_i) + \frac{1}{2}\|w\|^2 \tag{2}$$

The first term of above equation $\left(C(1/n)\sum_{i=1}^{n} L(d_i, y_i)\right)$ is known as empirical risk measuring by $L\varepsilon$:

$$L\varepsilon(d, y) = \begin{cases} |d-y| - \varepsilon|d-y| \geq \varepsilon \\ 0 \quad otherwise \end{cases} \tag{3}$$

The second term of Equation (2) $\left(0.5\times\|w\|^2\right)$ known as regularization term, C is the reqularized constant determining the trade-off between regularization term and empirical risk. Increasing the C, leads to decreasing the relative importance of regularization term in comparison to empirical risk. The ε variable is equal to estimation precision on training dataset. By introducing the positive slack variables $\zeta\imath$ and $\zeta\imath^*$ and transformed the Equation (2) in the primal function as Equation (3), w and b is estimated.

$$\text{Minimize} \quad R_{SVMs}(w, \xi^{(*)}) = \frac{1}{2}\|w\|^2 + C\frac{1}{n}\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{4}$$

$$\text{Subject to} \begin{cases} d_i - w\varphi(x_i) + b_i \leq \varepsilon + \xi_i \\ w\varphi(x_i) + b_i - d_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1.............l \end{cases} \tag{5}$$

By exploiting the optimally constraints and introducing Lagrange multipliers, the decision function (Equation (1)) in the explicit form is presented as follows:

$$f(x, a_i a_i^*) = \sum_{i=1}^{n}(a_i - a_i^*)K(x, x_i) + b \tag{6}$$

where $K(x,xi)=\varphi(xi)\varphi(xj)$ and the term $K(x,xi)$ is called the kernel function, which is product of the two inner vector x_i and x_j in the feature space $\varphi(x_i)$ and $\varphi(x_j)$, respectively. All calculating related $\varphi(x)$ is performed using kernel function. The used kernel function in this study is Radial basis function (RBF) kernel function which is calculated as follows:

$$K(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right) \tag{7}$$

Where, γ is the kernel parameter which should be carefully determined to reach the acceptance results.

Data presentation for stable channel design and experimental design

The channel hydraulic parameters at stable channel include water level width (w), average flow depth (h) and channel slope (s). The study of previous researches show that two flow discharge (Q) and average particle size of sediment (d_{50}) parameters and later Shields parameter (τ^*) (sediment particle moving) are important parameter to predict w, h and s at stable channel in most studies.

$((w,h,s)=f(Q,d_{50},\tau^*)$ This parameter is defined as follows:

$$\tau_* = \frac{\tau}{(\rho_s-\rho)\,.g.\,d_s} \tag{8}$$

Where ρ , the mass density of water (ρ =1000 kg/m^3), ρs mass density of sediment (ρs=2650 kg/m^3), g the gravitational force (g= 9.81 m^2/s), d_s the particle diameter or that relative density and v kinematics viscosity (v= 1×10-6 m^2/s), and τ the shear stress that shear stress model with zero-pressure gradient is estimated as follows:

$$\tau = \rho ghs \tag{9}$$

In the present paper, three different SVM models are presented to predict width, depth and slope of stable channel cross sections. The input variables are Q, d_{50} and τ^* parameters. The observational data related to 85 cross section river with gravel-bed in IRAN (Afzalimehr et al. 2010 [1]) that are located in stability regime condition, are used to SVM models train and test. Then Afzalimehr' s regression relationships [1] are used to predict geometry of stable channel and also for evaluating of SVM model' s efficiency, These equations are extended based on statistical calculations on a Afzalimehr's dataset which these equations are as follow [1]:

$$w = 5.876\,Q^{0.743} \tag{10}$$

$$h = 0.226\,Q^{0.345} \tag{11}$$

$$s = 1.565\,d_{50}^{0.821}\,\tau_*^{0.851} \tag{12}$$

Statistical Analysis

The performance of SVM models is evaluated with the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Relative Error (MARE), Correlation coefficient (R) and BIAS statistical parameters. Equations 13, 14, 15, 16 and 17 are used to compute RMSE, MAE, MARE, R, and BIAS respectively:

$$RMSE = \left[\frac{\sum_{i=1}^{N} (Y_{i(model)} - Y_{i(Observed)})^2}{N} \right]^{1/2} \tag{13}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| Y_{i(model)} - Y_{i(Observed)} \right| \tag{14}$$

$$MARE = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\left| Y_{i(model)} - Y_{i(Observed)} \right|}{Y_{i(Observed)}} \right) \tag{15}$$

$$R = \frac{\sum_{i=1}^{n} (Y_{i(Observed)} - \overline{Y_{i(Observed)}}).(Y_{i(model)} - \overline{Y_{i(model)}})}{\sqrt{\sum_{i=1}^{n} (Y_{i(Observed)} - \overline{Y_{i(Observed)}})^2 \sum_{i=1}^{n} (Y_{i(model)} - \overline{Y_{i(model)}})^2}} \tag{16}$$

$$BIAS = \frac{\sum_{i=1}^{N} (Y_{i(model)} - Y_{i(Observed)})}{N} \tag{17}$$

Where $Y_{i(Observed)}$ is the output observational parameter, $Y_{i(Model)}$ is the parameter predicted by the SVM models, $\overline{Y_i}_{(Model)}$ is the mean SVM model parameter and N is the number of parameters. RMSE and MAE show the difference between the modeled and observed data in the same unit of them. Higher model accuracy leads to RMSE and MAE values closer to zero. R provides a measure of how well the observed outcomes are replicated by the model.

Results and Discussion

Geometric and hydraulic parameters of stable channels are predicted using SVM models in the present study. 85 cross section rivers field data of Afzalimehr et al. (2010) are utilized for training and testing models which 60 data (70% of whole data) and 25 data (30% of whole data) are used for training and testing models, respectively [1]. Figure 1 shows the regression plot to predict the width, depth and slope by SVM models in comparison with observed data. Table 2 shows the R, MARE, RMSE, MAE, BIAS error indexes between SVM models and observed data. Also, the width, depth and slope values are estimated by Afzalimehr et al. (2009)'s equations (based on statistical regression methods [1]) and are compared with observed values and this equation's error indexes are shown in Table 1. In Figure 1, data compression has been more around exact line and also R value (in Table 1) in three models shows high accuracy of SVM models in prediction (R= 0.86, 0.66 and 0.646 in width, depth and slope prediction models, respectively). In all figures, data distribution is around the ±10% error line that shows the range of data errors. In Table 1, in width prediction model, SVM model with high R value is more accurately than the Afzalimehr' s equation (R= 0.860). Also in this model, relative and absolute error of MARE and RMSE have the less value in SVM model (MARE =0.062 and RMSE =0.20). In depth prediction model, SVM model with high correlation coefficient and low error values is more accurately than Afzalimehr' s equation (R= 0.66, MARE= 0.042, RMSE= 0.202). In slope prediction model like width and depth prediction models, relative and absolute error values in SVM model are more less than Afzalimehr's equation (MARE= 0.078, RMSE= 0.597). in general, MARE value in SVM model are more less than Afzalimehr' s regression equations by 72%, 20% and 11% in prediction width, depth and slope parameters, respectively. Therefore, regression models performance improve using SVM models. Performance improvement of SVM model in width prediction is more than another models because in the Afzalimehr' s width equation are existed only discharge parameter but input variables in SVM model are all three Q, d_{50}, τ_* parameter. The lowest efficiency improvement is slope prediction model because the presence of all three affected parameters in both models. Specifically, presence of τ^* parameter in slope prediction according on shear stress relation with zero-pressure gradient is more affected. BIAS indices show the underestimation and overestimation of models. This index value in SVM model Afzalimehr et al. (2009) equations [1] to predict the width, depth and slope is close to zero. It can be said that, this index with the same average value predict the estimation value as underestimation and overestimation. Negative and positive value indices show the underestimation and overestimation model, respectively. In present SVM and regression models, in both width and slope models, BIAS values are positive, indicating that the models are as overestimate. And negative values indicate an underestimation in predicting the depth models.

Table 1. SVM models and Afzalimehr et al. (2009) equations [1] evaluation to predict the width, depth and slope of stable channel section in comparison with observed values.

Variables	Models	R	MARE	RMSE	MAE	BIAS
Width	SVM	0.860	0.062	0.20	0.152	**0.035**
	Afzalimehr	0.802	4.520	12.622	11.978	**11.978**
Depth	SVM	0.660	0.042	0.202	0.147	**-0.001**
	Afzalimehr	0.562	0.902	3.162	3.153	**-3.153**
Slope	SVM	0.646	0.078	0.597	0.404	**0.075**
	Afzalimehr	0.383	0.952	5.532	5.482	**0.504**

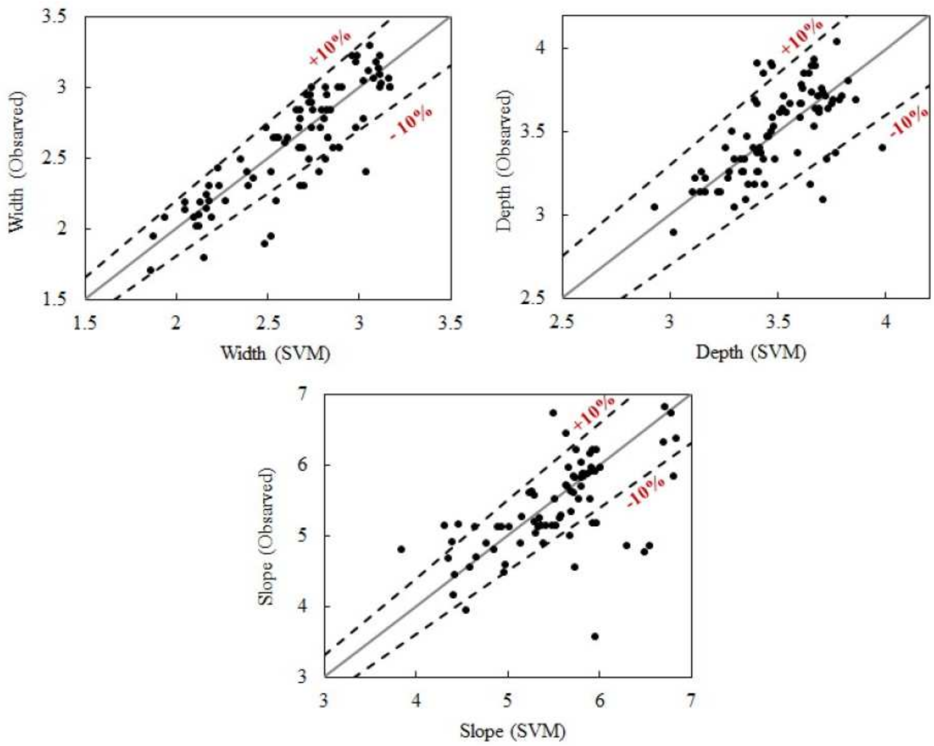

Fig.1: The scatter plot diagrams to predict the width, depth and slope by SVM model in comparison with observational values.

In Figure 2, width, depth and slope values predicted by SVM model are compared with observed data. SVM model predicts well velocity data trend like observed data and have a good agreement with observed data. In all three figures, SVM model predicts higher and lower values than those observed data in maximum and minimum points, respectively (except of 15-25 data range).

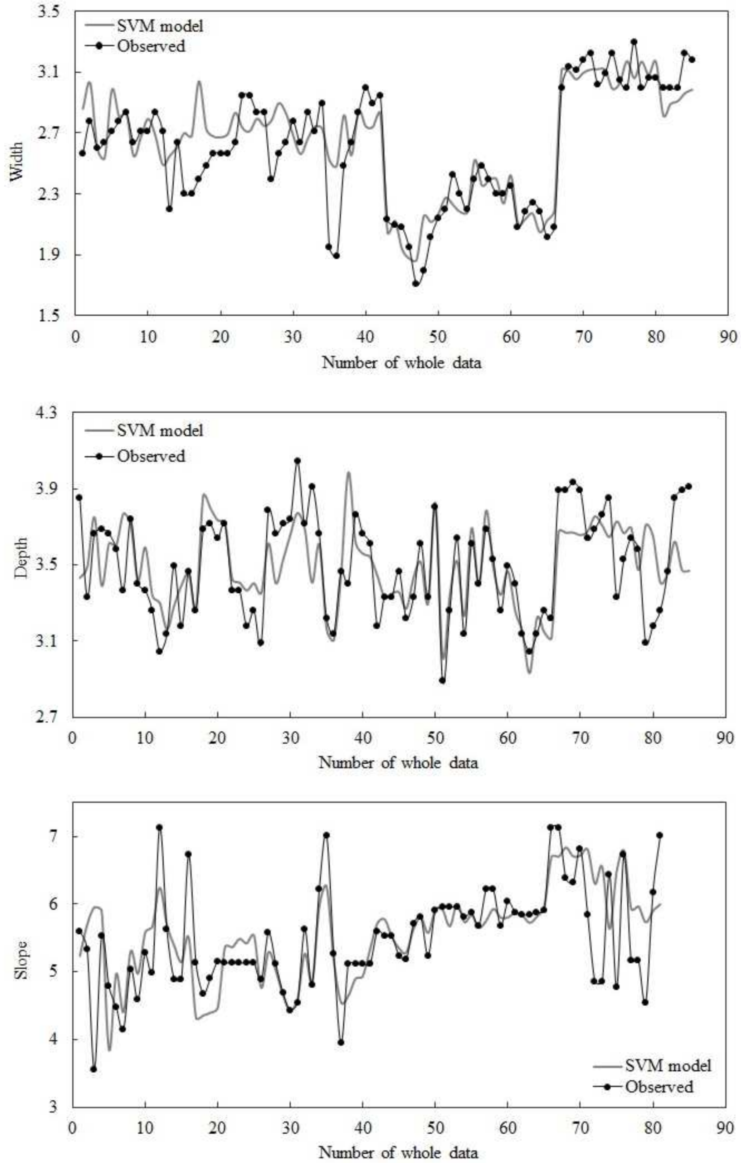

Fig .2: Comparison of width, depth and slope predicted by SVM model with observed values.

Conclusions

In the present paper, soft computing methods are used in prediction of hydraulic and geometric parameters of stable channels. Three SVM models are designed to predict width, depth and slope of stable channel cross section. Available observed data are used for training and testing SVM models. To models evaluation, the SVM results are compared with the regression equations results which fitted on the same observed data that using in present study. The comparison of SVM models with observed data shows the high accuracy of models. Also the SVM models with RMSE values of 0.152, 0.147 and 0.404 are act better than regression equation with RMSE value of 12.622, 3.162 and 5.532 in prediction width, depth and slope, respectively. The noticeable point in this study is that despite of the regression relation are fitted on the same field data that uses for training and testing SVM model, but the proposed SVM models still better than the regression equations. It is suggested that the other soft computing techniques such as, Adaptive Neuro Fuzzy Computing Technique (ANFIS), Gene Expression Programming (GEP) models and etc will be used in prediction of stable channel parameters.

References

[1] **Afzalimehr, H.; Abdolhosseini, M. and Singh, V.P**. Hydraulic geometry relations for stable channel design. Journal of Hydrologic Engineering, Vol.15,No.10,pp.859-864.2009.

[2] **Beechie, T. and Imaki, H.** Predicting natural channel patterns based on landscape and geomorphic controls in the Columbia River basin, USA. Water Resource Research, Vol.50,No.1,pp.39-57.2015.

[3] **Ebtehaj, I. and Bonakdari, H**. Evaluation of sediment transport in sewer using artificial neural network. Engineering Applications of Computational Fluid Mechanics, Vol.7,No.3,pp.382-392.2013.

[4] **Ebtehaj, I. and Bonakdari, H**. Performance Evaluation of Adaptive Neural Fuzzy Inference System for Sediment Transport in Sewers. Water Resource Management, Vol.28,No.13,pp. 4765–4779.2014.

[5] **Gholami, A.; Akhtari, A.A.; Minatour, Y.; Bonakdari, H. and Javadi, A.A.** Experimental and numerical study on velocity fields and water surface profile in a strongly-curved 90° open channel bend. Engineering Applications of Computational Fluid Mechanics, Vol.8,No.3,pp. 447-461.2014.

[6] **Gholami, A.; Bonakdari, H.; Zaji, A.H. and Akhtari, A.A.** Simulation of open channel bend characteristics using computational fluid dynamics and artificial neural networks. Engineering Applications of Computational Fluid Mechanics, Vol.9,No.1,pp. 355-361.2015.

[7] **Gholami, A.; Bonakdari, H.; Zaji, A.H.; Akhtari, A.A. and Khodashenas, S.R.** Predicting the Velocity Field in a 90º Open Channel Bend Using a Gene Expression Programming Model. Flow measurement and instrumentation, doi:10.1016/j.flowmeasinst.2015.10.006, 2015.

[8] **Hey, R.D. and Colin, R.T.** Stable channels with mobile gravel beds. Journal of Hydraulic Engineering, Vol.112,No.8,pp. 671-689.1986.

[9] **Javadi, F.; Ahmadi, M.M. and Qaderi, K.** Estimation of River Bedform Dimension Using Artificial Neural Network (ANN) and Support Vector Machine (SVM). Journal of Agricultural science and technology, Vol.17,No.4,pp. 859-868.2015.

[10] **Julien, P. Y. and Wargadalam, J**. Alluvial channel geometry: Theory and applications. Journal of Hydraulic Engineering, Vol.1214,pp. 312-325.1995.

[11] **Karimi, S.; Bonakdari, H. and Gholami, A.** Determination Discharge Capacity of Triangular Labyrinth Side Weir using Multi-Layer Neural Network (ANN-MLP). Current World Environment, Special volume, 2015.

[12] **Lee, J. S. Julien, P.Y.** Downstream Hydraulic Geometry of Alluvial Channels. Journal of Hydraulic Engineering, Vol.132,No.12, 2006.

[13] **Tahershamsi, A.; Tabatabai, M.R.M, and Shirkhani, R.** An evaluation model of artificial neural network to predict stable width in gravel bed rivers. International Journal of Environmental Science and Technology, Vol.9,No.2,pp. 333-342.2012.

[14] **Vapnik, V.; Golowich, S.E. and Smola, A**. Support vector method for function approximation, regression estimation, and signal processing. MIT Press, Cambridge, MA, neural information processing systems edition, 1997.

[15] **Khadangi, E.; Madvar, H.R. and Kiani, H.** Application of artificial neural networks in establishing regime channel relationships. 2nd International Conference on Computer, Control and Communication, IC4 2009.