

Predictive modeling of discharge in compound open channel by support vector machine technique

Abbas Parsaie¹ · Hojatt Allah Yonesi¹ · Shadi Najafian¹

Received: 17 February 2015 / Accepted: 20 April 2015 / Published online: 9 May 2015
© Springer International Publishing Switzerland 2015

Abstract Modeling of river phonemes is one of the main parts in Hydraulic engineering studies. Predicting the discharge in the compound open channel is the main subject of the river engineer researches. In this study the support vector machine (SVM) technique is used to predict the discharge in the compound open channel. The result of the error indexes shows that the best accuracy of the SVM by correlation coefficient 0.99 is related to the SVM model which includes the radial basis function (RBF) as kernel function. During the SVM model process obtained 0.00005 for the gamma coefficient, epsilon is equal to 0.1 and $C = 1$.

Keywords Compound open channel · Support vector machine · Discharge prediction · River hydraulic

Introduction

Predicting the discharge in the natural streams is one of the main parameter in the flood engineering problems (Greco et al. 2004). Calculating the discharge in the streams especially in the flooding leads to better management of hydraulic structure operation such as gated spillways operation. Discharge estimation in rivers helps to calculate the

downstream water surface elevation that is an important parameter in river engineering projects such as river intake design, defining the height of dams and also determining the area of floodplains (Te Chow 2009). River hydraulic modeling by commercial software such as Mike and Hec Ras and so on needs to have an accurate value of flow discharge or head-discharge curve (Knight and Shamseldin 2005). The concept of the compound open channel is the accurate approach for modeling the natural streams which is developed recently. As shown in the Fig. 1 the compound open channels cross section includes a main channel and floodplains. The normal value of discharges flows in the main channel and when the flood occurs additional discharge flows in the floodplains. Because of covering the floodplains by flow in flooding conditions, the floodplains usually are covered by vegetation so usually it is rougher than the main channel (Myers 1987; Yang et al. 2014; Khatua et al. 2012; Ackers 1993; Wormleaton and Hadjipanous 1985). As shown in the Fig. 1 the main channel and floodplains roughness shows by n_{mc} and n_{fp} sequentially. This difference between the roughness leads to create a momentum transferring and appearance an interaction area between main channel and floodplains (Myers 1987; Ackers 1993; Al-Khatib et al. 2012). Flow structure in compound open channel is fully turbulence and 3D dimensionality especially in flooding condition, so calculating the discharge by classical formulas such as Manning and Chezy makes to appearance the incredible error in river engineering studies (Wormleaton and Hadjipanous 1985; Stephenson and Kolovopoulos 1990).

Several ways as analytical approaches and artificial intelligent (AI) techniques have been proposed to achieve more accuracy in rivers discharge prediction. In this regards the Single-Channel Method, Divided-Channel Method, and Coherence Method can be mentioned as

✉ Abbas Parsaie
abbas_parsaie@yahoo.com

Hojatt Allah Yonesi
yonesi.h@lu.ac.ir

Shadi Najafian
sh_najafyan@yahoo.com

¹ Department of Water Engineering, Lorestan University, Khorramabad, Iran

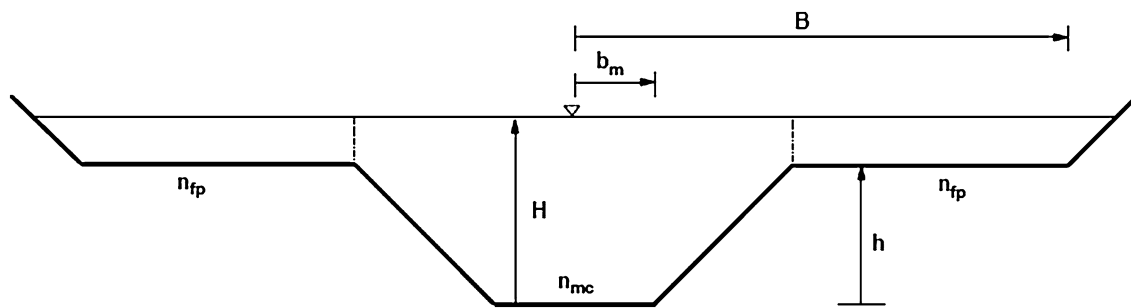


Fig. 1 Compound open channel cross section

analytical approaches (Yang et al. 2014; Bousmar et al. 2004; Huthoff et al. 2008; Yonesi et al. 2013; Al-Khatib et al. 2012; Mohanty and Khatua 2014; Hosseini 2004; Naot et al. 1993; Al-Khatib et al. 2013; Liao and Knight 2007; Atabay and Knight 2006). Recently by advancing the data mining approaches, some artificial intelligence models such as artificial neural network; Genetic Programming, and M5 tree models was applied to predict the flow discharge in compound open channel. Based on the scientific reports the performances of the Artificial intelligence (AI) models are better than the analytical approaches (Sahu et al. 2011; Seckin 2004; Unal et al. 2010; Zahiri and Azamathulla 2014). Using the AI model AI models together numerical methods in a computer simulation of the hydraulic phenomena leads to increase the accuracy of the numerical modeling (Parsaie and Haghiabi 2015). Sahu et al. (2011) predicted the discharge in compound open channel by using the neural network technique and state that the ratio of the depth of the flow in the main channel to the depth of the flow in the floodplains (D_r), ratio of the area of the main channel to the floodplains (A_r) and ratio of the hydraulic radius of the main channel to the floodplains (R_r) are the most influence parameters for predicting the flow discharge in the compound open channel. Zahiri and Azamathulla (2014) by applying the genetic programming technique on the single-channel method results proposed an equation to predict the discharge in compound open channel. In this paper another types of AI models which is named support vector machine (SVM) is used for predicting the flow discharge in the compound open channel.

Support vector regression

SVMs are set of related supervised learning methods used for classification and regression. In many applications a non-linear classifier provides better accuracy. In SVR, the input x is first mapped onto an m -dimensional feature space using some fixed (nonlinear) mapping, and then a linear model is constructed in this feature space. The naive way

of making a non-linear classifier out of a linear classifier is to map our data from the input space X to a feature space F using a non-linear function $\varphi : x \rightarrow f$. In the space F the discriminate function is:

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

Using mathematical notation, the linear model (in the feature space) $f(x, w)$ is given by

$$w = \sum_{i=1}^n \alpha_i x_i \tag{2}$$

$$f(x, w) = \sum_{j=1}^n \alpha_j \varphi_j(x) + b \tag{3}$$

$$f(x) = \sum_{i=1}^n \alpha_i x_i^T x + b \tag{4}$$

In the feature space, F this expression takes the form:

$$f(x) = \sum_{i=1}^n \alpha_i \varphi(x_i)^T \varphi(x) + b \quad 0 \leq \alpha_i \leq C \tag{5}$$

$$k(x, x') = \varphi(x)^T \varphi(x') \tag{6}$$

$$f(x) = \sum_{i=1}^n \alpha_i k(x, x_i) + b \tag{7}$$

There are many kernel functions in SVM, so how to select a good kernel function is also a research issue. However, for general purposes, there are some popular kernel functions.

- (i) Linear kernel: $k(x_i, x_j) = x_i^T x_j$
- (ii) Polynomial kernel: $k(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- (iii) RBF kernel: $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$
- (iv) Sigmoid kernel: $k(x_i, x_j) = \tanh(\gamma x_i^T x_j + r), \gamma > 0$

Here C, γ and r and d are kernel parameters. It is well known that SVM generalization performance (estimation

accuracy) depends on a good setting of meta-parameters parameters C , γ and r and the kernel parameters. The choices of C , γ and r control the prediction (regression) model complexity. The problem of optimal parameter selection is further complicated because the SVM model complexity (and hence its generalization performance) depends on all three parameters. Kernel functions are used to change the dimensionality of the input space to perform the classification.

Analytical approach

The Divided Channel method (DCM) is the accurate analytical approach which divides the cross section into the main channel and floodplains. In this the method the discharge is calculated for each subsection and total discharge of the compound channel section is calculated by adding the subsections discharge. The discharge in subsections calculates by Eq. (8). As shown in this equation the classical equation such as manning is used for calculating the discharge and i subscription is related to the discharge in each sub sections. The separation line between the main channel and floodplains as shown in the Figs. 1 may be considered as vertical, diagonal or horizontal in this paper the divided channel method with horizontal division lines which are included within the calculation of wetted perimeter (DCM (h-i)) is considered (Huthoff et al. 2008).

$$Q_i = \left(\sum_{i=1}^N \frac{A_i R_i^{\frac{2}{3}}}{n_i} \right) S_0^{\frac{1}{2}} \tag{8}$$

where N is the number of the subsections, S_0 is the longitudinal slope of the compound open channel, A_i is the area, R_i is the Hydraulic radius, and n_i is the roughness coefficient (Manning coefficient) of the subsections.

Model development

Developing the SVM model as similar to other neural network models is based on data set. To this purpose 202 data set related to the flow discharge in compound open channel which published in the creditable journals was collected. Some of the resources that data derived from the are given as follow (Naot et al. 1993; Sahu et al. 2011; Seckin 2004; Unal et al. 2010; Yonesi et al. 2013; Tang et al. 1999; Knight et al. 1984). For developing the SVM model, the collection data divides in two groups, training and testing data. The range of the data collection gives as gives in the Table 1. The main parameters which are influences on the flow discharge in the compound open channel gives in the Eq. (9).

Table 1 Range of data collection related to the discharge

	n_r	A_r	R_r	S	D_r	Q_m
Min	1.000	0.047	0.020	0.000	0.041	0.005
Max	6.408	2.909	0.195	0.002	0.773	1.114
Avg	1.306	0.651	0.066	0.001	0.303	0.144
STDEV	0.999	0.424	0.038	0.001	0.167	0.218

$$Q_{cmp} = f(f_{fp}, f_{mc}, A_{fp}, A_{mc}, R_{fp}, R_{mc}, S, h, (H - h)) \tag{9}$$

where f_{fp} and f_{mc} is the roughness of main channel and floodplains, A_{fp} and A_{mc} is the area of main channel and floodplains, R_{fp} and R_{mc} is the hydraulic radius of main channel and floodplains, S is the longitudinal slope of compound open channel, H is the depth of flow in the main channel and h is the depth of the flow in the floodplains. Based on the researches which implement the neural network techniques to predict the flow discharge in the compound open channel, developing the AI models based on the dimensionless parameter leads to prepare an optimal structure of the model. Based on these researches the most popular dimensionless parameters are given in the Eq. (10).

$$Q = f(f_r, A_r, R_r, S, H_r) \tag{10}$$

where S is the longitudinal slope of the compound open channel, n_r is defined as $\frac{f_{mc}}{f_{fp}}$, A_r is defined as $\frac{A_{mc}}{A_{fp}}$, R_r is defined as $\frac{R_{mc}}{R_{fp}}$, D_r is defined $\frac{H-h}{H}$, and Q_m is the measured discharge. For developing the SVM the n_r , A_r , R_r , S and D_r is considered as inputs parameters and Q_m is considered as outputs parameter.

Results and discussion of SVM

The Divided Channel method (DCM) as common analytical approach is used to calculate the discharge of flow in compound open channel. The performance of the DCM model shows in the Fig. 2 and to assess the performance of the DCM by error indices such as correlation coefficient (R^2) and root mean square (RSME) of error was applied. It

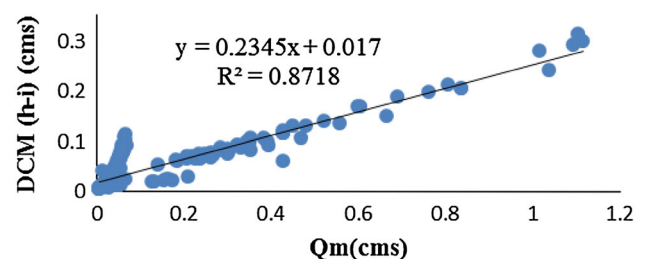


Fig. 2 Results of the divided channel method

should be noted that all the data collection (202 data set) is used for assessing the performance of the DCM method. As shown in the Fig. 2, the results of the DCM method in low values of the discharge especially values which are less than the 0.1 (cms) is more than the measured data and it is related to increase the area of the flow cross section in compare to hydraulic radius. The accuracy of the DCM method in the values between the 0.2 and 1 is suitable. By increasing the value of the discharge [values more than the 1(cms)] due to increase the intensity of the momentum transferring between the main channel and floodplains, usually the accuracy of the other analytical approaches are decreases but as shown in the Fig. 2 the performance of the DCM method is suitable and it is related to the considering the effect of the momentum transferring in the method development. As shown in the Fig. 2, the correlation

coefficient of the DCM method is equal to (0.87) and RMSE is (0.05) and in overall the performance of the DCM is suitable.

Developing the SVM model based on the data set requires evaluating the quality of the data collection. Because of AI models development such as almost types of neural network models are based on the data set. The data distribution has influence impact on the performance of the AI models. To prepare a suitable ANN model by high reliability it is necessary to collect data which is regular distribution on the range of the data collection. Another way addition to the collecting the regular data is selecting the suitable data from the data collected for training process which becomes as suitable proxy for all data collection. To this purpose randomly selection may be considered. Although to evaluate the distribution of the

Fig. 3 Histogram of discharge values of collected data

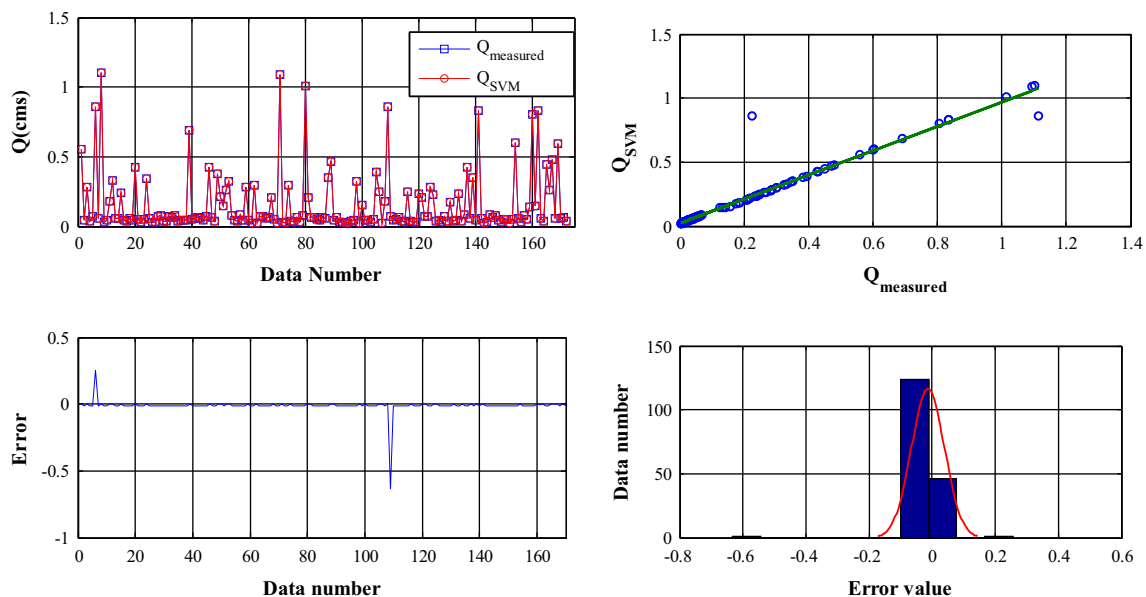
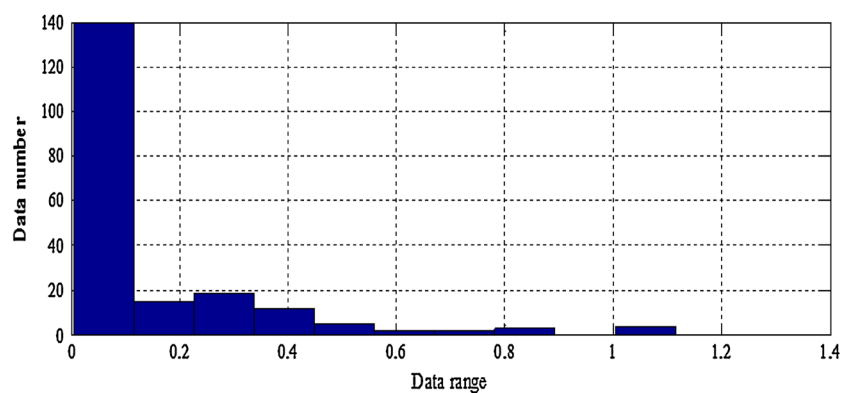


Fig. 4 The performance of the SVM model in training stage

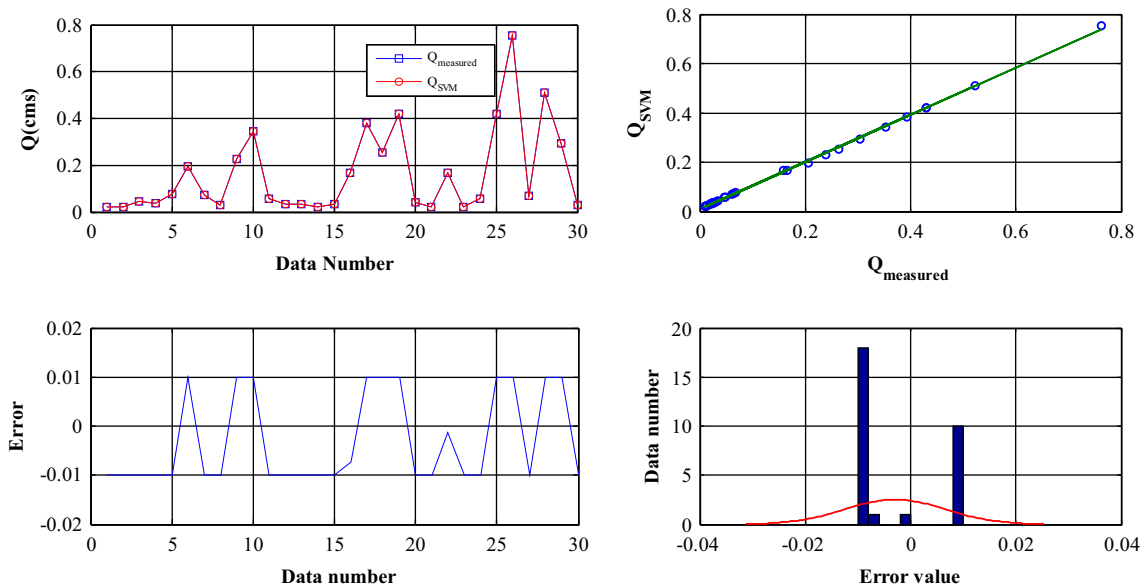


Fig. 5 The performance of the SVM model in testing stage

data some indices such as minimum, maximum, average skewness, and standard division are present but to better understand the distribution of data, the histogram of data was plotted in the Fig. 3. As shown in the Fig. 3, about 95 % of data has a value less than the 0.4 (cms) and less than the 5 % of data has a values more the 0.4 (cms). To prepare the SVM model based on the collected data and by considering the Fig. 3 as distribution of the data, about 85 % (172) of the data are considered for training process and other data about 15 % (30) of data is used in the testing stage. Four types of the kernel function which are mentioned in the support vector regression section such as Linear, secondary Polynomial, Radial Basis Function (RBF), and Sigmoid was used in development process of the SVM. The best performance of the SVM model has been found to be related to the RBF as the kernel function. The results of the SVM model in the training and testing stages shows in the Figs. 4, 5. As shown in the Figs. 4, 5, the accuracy of the SVM model is so suitable by training correlation coefficient 0.94 (i.e. R^2_{train}) and RMSE = 0.035 and testing correlation coefficient 0.99 (i.e. $R^2_{test} = 99$) and RMSE = 0.0001. During the SVM model process obtained 0.00005 for the gamma coefficient, epsilon is equal to 0.1 and C = 1. As shown in the Figs. 4, 5, the performance of the SVM model to predict the low values of the discharges especially discharge values less than the 0.5 (cms) is so suitable. By attention to the distribution of the discharge data (Fig. 3) and performance of the SVM model Figs. 4, 5, it is seems that the performance of SVM model for prediction the discharge values especially in the near to the maximum values 0.8 (cms) is suitable. To assess the performance of SVM model addition to plot the SVM results

vs measured data, the error distribution and histogram of the error is plotted. As shown in the Figs. 4, 5 the accuracy of the SVM model to predict the Maximum values of the discharge.

Conclusions

Prediction of flow discharge in the natural stream flows is the main parameter in the flood management problems. The compound channel concept is the accurate approach for the simulation of the river problems. Predicting the flow discharge in compound open channels by classical formulas such as Manning and Chezy lead to appearance the incredible error in compare to the measured data, so researcher revised these approaches and proposed more advanced analytical approaches such as Divided Channel method. Although the revised approaches have been increased the accuracy of flow discharge calculation but flood management needs to implement more accurate approaches in the river engineering problems. Recently by advancing the AI models in water engineering problems, the artificial neural network models have been used widely for flow discharge prediction. Based on the results of this study, the maximum accuracy (correlation coefficient) of the analytical approaches is about 0.87 but a new model of the ANN such as support vector machine can easily obtain the accuracy about 0.99. The main points of the AI models developing are preparing a suitable set data collection which has suitable distribution along the range of the data collection. In addition to have a suitable data collection, considering the suitable kernels functions and correct

setting the coefficients which uses in the kernel functions is necessary.

References

- Ackers P (1993) Flow formulae for straight two-stage channels. *J Hydraul Res* 31(4):509–531. doi:[10.1080/00221689309498874](https://doi.org/10.1080/00221689309498874)
- Al-Khatib IA, Dweik AA, Gogus M (2012) Evaluation of separate channel methods for discharge computation in asymmetric compound channels. *Flow Meas Instrum* 24:19–25. doi:[10.1016/j.flowmeasinst.2012.02.004](https://doi.org/10.1016/j.flowmeasinst.2012.02.004)
- Al-Khatib I, Hassan H, Abaza K (2013) Application and validation of regression analysis in the prediction of discharge in asymmetric compound channels. *J Irrig Drain Eng* 139(7):542–550. doi:[10.1061/\(ASCE\)IR.1943-4774.0000579](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000579)
- Atabay S, Knight DW (2006) 1-D modelling of conveyance, boundary shear and sediment transport in overbank flow. *J Hydraul Res* 44(6):739–754. doi:[10.1080/00221686.2006.9521725](https://doi.org/10.1080/00221686.2006.9521725)
- Bousmar D, Wilkin N, Jacquemart J, Zech Y (2004) Overbank flow in symmetrically narrowing floodplains. *J Hydraul Eng* 130(4):305–312. doi:[10.1061/\(ASCE\)0733-9429\(2004\)130:4\(305\)](https://doi.org/10.1061/(ASCE)0733-9429(2004)130:4(305))
- Greco M, Carravetta A, Morte RD (2004) River flow 2004. In: Balkema AA
- Hosseini SM (2004) Equations for discharge calculation in compound channels having homogeneous roughness. *Iran J Sci Technol Trans B Eng*
- Huthoff F, Roos P, Augustijn D, Hulscher S (2008) Interacting divided channel method for compound channel flow. *J Hydraul Eng* 134(8):1158–1165. doi:[10.1061/\(ASCE\)0733-9429\(2008\)134:8\(1158\)](https://doi.org/10.1061/(ASCE)0733-9429(2008)134:8(1158))
- Khatua K, Patra K, Mohanty P (2012) Stage-discharge prediction for straight and smooth compound channels with wide floodplains. *J Hydraul Eng* 138(1):93–99. doi:[10.1061/\(ASCE\)HY.1943-7900.0000491](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000491)
- Knight D, Shamseldin A (2005) River basin modelling for flood risk mitigation. CRC Press, UK
- Knight DW, Demetriou JD, Hamed ME (1984) Stage discharge relationships for compound channels. In: Smith KVH (ed) Channels and channel control structures. Springer, Berlin pp 445–459. doi:[10.1007/978-3-662-11300-4_35](https://doi.org/10.1007/978-3-662-11300-4_35)
- Liao H, Knight D (2007) Analytic stage-discharge formulas for flow in straight prismatic channels. *J Hydraul Eng* 133(10):1111–1122. doi:[10.1061/\(ASCE\)0733-9429\(2007\)133:10\(1111\)](https://doi.org/10.1061/(ASCE)0733-9429(2007)133:10(1111))
- Mohanty PK, Khatua KK (2014) Estimation of discharge and its distribution in compound channels. *J Hydrodyn Ser B* 26(1):144–154. doi:[10.1016/S1001-6058\(14\)60017-2](https://doi.org/10.1016/S1001-6058(14)60017-2)
- Myers W (1987) Velocity and discharge in compound channels. *J Hydraul Eng* 113(6):753–766. doi:[10.1061/\(ASCE\)0733-9429\(1987\)113:6\(753\)](https://doi.org/10.1061/(ASCE)0733-9429(1987)113:6(753))
- Naot D, Nezu I, Nakagawa H (1993) Calculation of compound-open-channel flow. *J Hydraul Eng* 119(12):1418–1426. doi:[10.1061/\(ASCE\)0733-9429\(1993\)119:12\(1418\)](https://doi.org/10.1061/(ASCE)0733-9429(1993)119:12(1418))
- Parsaie A, Haghiabi A (2015) The effect of predicting discharge coefficient by neural network on increasing the numerical modeling accuracy of flow over side weir. *Water Resour Manage* 29(4):973–985. doi:[10.1007/s11269-014-0827-4](https://doi.org/10.1007/s11269-014-0827-4)
- Sahu M, Khatua KK, Mahapatra SS (2011) A neural network approach for prediction of discharge in straight compound open channel flow. *Flow Meas Instrum* 22(5):438–446. doi:[10.1016/j.flowmeasinst.2011.06.009](https://doi.org/10.1016/j.flowmeasinst.2011.06.009)
- Seckin G (2004) A comparison of one-dimensional methods for estimating discharge capacity of straight compound channels. *Can J Civ Eng* 31(4):619–631. doi:[10.1139/04-053](https://doi.org/10.1139/04-053)
- Stephenson D, Kolovopoulos P (1990) Effects of momentum transfer in compound channels. *J Hydraul Eng* 116(12):1512–1522. doi:[10.1061/\(ASCE\)0733-9429\(1990\)116:12\(1512\)](https://doi.org/10.1061/(ASCE)0733-9429(1990)116:12(1512))
- Tang X, Knight DW, Samuels PG (1999) Variable parameter Muskingum-Cunge method for flood routing in a compound channel. *J Hydraul Res* 37(5):591–614. doi:[10.1080/00221689909498519](https://doi.org/10.1080/00221689909498519)
- Te Chow V (2009) Open-channel hydraulics. Blackburn Press
- Unal B, Mamak M, Seckin G, Cobaner M (2010) Comparison of an ANN approach with 1-D and 2-D methods for estimating discharge capacity of straight compound channels. *Adv Eng Softw* 41(2):120–129. doi:[10.1016/j.advengsoft.2009.10.002](https://doi.org/10.1016/j.advengsoft.2009.10.002)
- Wormleaton P, Hadjipanous P (1985) Flow distribution in compound channels. *J Hydraul Eng* 111(2):357–361. doi:[10.1061/\(ASCE\)0733-9429\(1985\)111:2\(357\)](https://doi.org/10.1061/(ASCE)0733-9429(1985)111:2(357))
- Yang K, Liu X, Cao S, Huang E (2014) Stage-discharge prediction in compound channels. *J Hydraul Eng* 140(4):06014001. doi:[10.1061/\(ASCE\)HY.1943-7900.0000834](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000834)
- Yonesi HA, Omid MH, Ayyoubzadeh SA (2013) The hydraulics of flow in non-prismatic compound channels. *J Civil Eng Urban* 3(6):342–356
- Zahiri A, Azamathulla HM (2014) Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels. *Neural Comput Applic* 24(2):413–420. doi:[10.1007/s00521-012-1247-0](https://doi.org/10.1007/s00521-012-1247-0)