2012, Scienceline Publication

Journal of Civil Engineering and Urbanism

Volume 2, Issue 4: 164-169 (2012) (Received: June 6, 2012; Accepted: June 19, 2012; Published: July 25, 2012) ISSN-2252-0430

Gene Expression Programming for Prediction of Flow Discharge in Compound Channels

Abdol Reza Zahiri[1] and Parvin Eghbali[2*]

[1] *Assistant Professor, Department of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran*
[2] *MSc Student, Water Structures Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran*

*Corresponding author's Email: p.eghbali66@yahoo.com

ORIGINAL ARTICLE

ABSTRACT: There are many studies on the hydraulic calculation of steady uniform flows in compound open channels. Based on these studies, various methods have been developed with different assumptions. In general, these methods either have long computations or need numerical solution of differential equations. Furthermore, their accuracy for all compound channels with different geometric and hydraulic conditions may not be guaranteed. In this paper, to overcome with these limitations, a new simple and dimensionless equation has been proposed based on gene-expression programming (GEP), which is an extension to genetic programming (GP). In this equation only three parameters (e.g. depth ratio, coherence and ratio of computed total flow discharge to bankful discharge) have been used to simplify its applications for hydraulic and river engineers. By compiling 394 stage-discharge data from laboratories and fields of 30 compound channels, this new equation has been applied to estimate the flow conveyance capacity. The proposed GEP approach gives satisfactory results (R2 = 0.98 and RMSE = 0.32) compared to traditional method (VDCM) with R2 = 0.69 and RMSE = 1.72 for flow discharge computations.

Keywords*:* Compound Channels, Gene Expression Programming, Stage-Discharge Curve.

INTRODUCTION

During recent decades, there has been considerable interest in compound channel research beside to the considerable damages caused by flood events and the increased awareness of environmental and social issues. It has thus become more important for hydraulic engineers worked in river engineering projects to find hydraulically efficient solutions to river training and flood protection works and to ensure that these solutions are environmentally sustainable.

In rivers, hydrological measurements such as the discharge and velocity are crucial for the design of river engineering projects, calculating pollutant fluxes, interpretation of water quality data and for water resource management. A major issue of uncertainty in river channel analysis is that of accurately predicting the conveyance capability of river channels with floodplains, which are well known as compound channels. Cross-sections of these channels are generally characterized by a deep main channel, flanked by two relatively shallow floodplains, which is rougher than the main channel, due to their vegetation cover.

At base flow discharges and rather higher flows, when the flow is lower than the bankful depth, conventional resistance formula such as Manning equation may be used to assess discharge capacity with acceptable accuracy. However, in flood events, when overbank flow occurs, the classical formulae do not yield reliable solutions and usually lead to overestimation or underestimation of discharge capacity, which is dangerous for lives and properties located on the floodplains. This problem has led to a considerable investigation of the flow mechanism in compound channels, leading to development of new computational methods for an accurate prediction of discharge capacity. Off the many studies carried out by various authors, works of Shiono and Knight (1988, 1991), Ackers (1992) and Bousmar and Zech (1999) have more reliable results. However, all of these methods are not straightforward to be generally applicable by hydraulic engineers, and may suffer from long computational time or requiring numerical methods and solution of differential equations. To simplify the computations of overbank flows, in recent years some simple and new approaches have been proposed which lead to rather accurate results. MacLeod (1997) and Liu and James (2000) used the neural networks for flow discharge prediction in meandering compound channels. Lambert and Myers (1998) observed that the VDCM with vertical divisions tends to over-estimate the main channel mean velocity while the VDCM with a horizontal division under-estimates the main channel velocity, and vice versa for the floodplain velocity. Using this fact, they applied a weighting factor (ξ) to both the main channel and floodplain areas to give improved mean velocity estimates for these areas. Haiydera and Valentine (2002) using combination of Single Channel Method (SCM) and Coherence method, developed a simple equation for total flow discharge of straight rigid and mobile compound channels. Harris et al. (2003) based on genetic programming, solved the lateral profile of depth-averaged velocity distribution in vegetated floodplains. Sharifi (2009) applied genetic algorithm for conveyance estimation in compound channels. Zahiri and Dehghani (2009), Unal et al. (2010) and Zahiri et al. (2012) used

To cite this paper: Zahiri AR, Eghbali P. 2012. Gene Expression Programming for Prediction of Flow Discharge in Compound Channels. *J. Civil Eng. Urban.* 2(4): 164-169.
Journal homepage: http://www.ojceu.ir/main/

neural networks for discharge prediction in straight compound channels with high accuracy in the range of SKM, COH and EDM.

In this paper, a new and more accurate method based on evolutionary algorithm known as Gene Expression Programming (GEP) has been presented to solve the main problem of flow hydraulics in straight compound channels. The promising point of this study which leads to general applicability of the results is selecting only 3 input variables in GEP model structure.

MATERIAL AND METHODS

Vertical Divided Channel Methods (VDCM)

In Fig. 1, a compound section with three subdivided zones by classical methods of channel dividing is shown. Total flow discharge is the sum of discharges calculated separately in each subsection using an appropriate conventional friction formula, e.g. Manning equation, by assumption that each zone is homogeneous (Chow, 1959):

$$Q = \sum_{i=1}^{N} Q_i = \sum_{i=1}^{N} \frac{A_i R_i^{2/3} S_0^{1/2}}{n_i} \qquad (1)$$

Where Q is total flow discharge, A is area, R is hydraulic radius, S_0 is bed slope and *n* is Manning roughness coefficient. In this equation, *i* refer to each subsection (e.g. main channel or floodplains) and N is total number of subsections.

Among the three methods of vertical, horizontal and inclined divided channel methods, the most applicable method in many engineering tools and softwares (e.g. HEC-RAS, MIKE11, ISIS, SOBEC, etc.), is Vertical Divided Channel Method (VDCM). This method has great over-prediction error for discharge estimation in field and laboratory compound sections. This formula, as mentioned by many researchers (Martin and Myers, 1991; Ackers, 1992) is very erroneous. This error may be reaches up to 40% and 60% in field and laboratory compound channels, respectively (Martin and Myers, 1991; Lai and Bessaih, 2004).

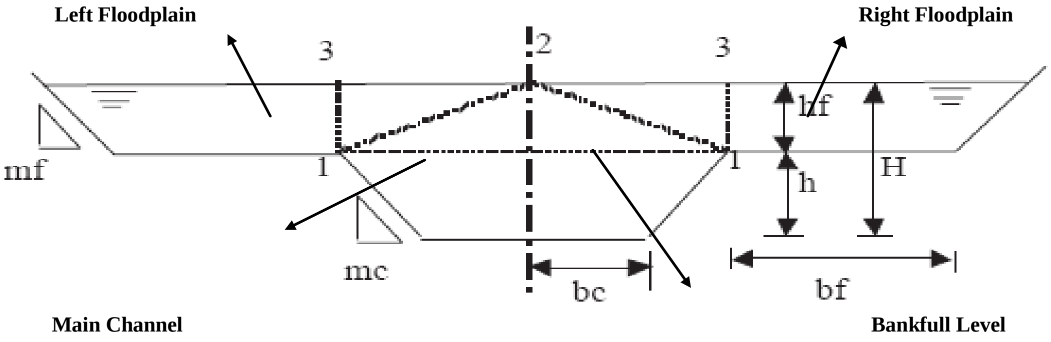

Figure 1. Compound channel cross-section with horizontal (H), diagonal (D) and vertical (V) planes shown as 1-1, 1-2 and 1-3, respectively (Wormleaton and Merrett, 1990).

Coherence parameter

The concept of coherence was introduced by Ackers (1992). Coherence is defined by Ackers as the ratio of the basic conveyance (treating the compound channel as a single unit, SCM) to that of computed by VDCM. By this definition, the following equation Manning roughness coefficient can be used to calculate coherence parameter:

$$COH = \frac{A^{5/3} \Big/ \left(\sum_{i=1}^{N} n_i^{1.5} P_i \right)^{2/3}}{\sum_{i=1}^{N} \frac{1}{n_i} \frac{A_i^{5/3}}{P_i^{2/3}}} \qquad (2)$$

where P_i and A_i are the wetted perimeter and the area of each subsections, respectively. P and A are total wetted perimeter and area of the channel, and N is the number of separate zones.

Review of gene expression programming

GEP, which is an extension of GP (Koza, 1992), is a search technique that evolves computer programs of different sizes and shapes encoded in linear chromosomes of fixed lengths. The chromosomes are composed of multiple genes, each gene encoding a smaller subprogram. Furthermore, the structural and functional organization of the linear chromosomes allows the unconstrained operation of important genetic operators such as mutation, transposition and recombination. The advantages of a system like GEP are clear from nature, but the most important are (Ferreira, 2001): (i) the chromosomes are simple entities: linear, compact, relatively small, easy to manipulate genetically (replicate, mutate, recombine, etc.); (ii) the expression trees are exclusively the expression of their respective chromosomes; they are entities upon which selection acts, and according to fitness, they are selected to reproduce with modification. In the present work the GeneXpro program was used for modeling flow discharge in compound channels (Ferreira, 2001). The procedure to predict flow discharge is as follows. The first step is the fitness function. For this problem, the fitness function, fi, of an individual program, i, is expressed as (Ferreira, 2001): ;

$$f_i = \sum_{j=1}^{n} \left(M - \left| C_{i,j} - T_j \right| \right) \qquad (3)$$

In which M is the range of selection, Ci,j is the value returned by the individual chromosome i for fitness case j, and Tj is the target value for fitness case j. For a perfect fit, Ci,j=Tj. The second step consists of choosing the set of terminals T and the set of functions F, to create the chromosomes. In the current problem, the terminal set includes (Depth parameter, coherence parameter, relative discharge). The study examined the

To cite this paper: Zahiri AR, Eghbali P. 2012. Gene Expression Programming for Prediction of Flow Discharge in Compound Channels. *J. Civil Eng. Urban.* 2(4): 164-169.
Journal homepage: http://www.ojceu.ir/main/

various combinations of these parameters as inputs to the GEP models to evaluate the flow discharge. The choice of the appropriate function is not so obvious and depends on the viewpoint and guess of user. In this study, different mathematical functions were utilized, including basic arithmetic operators ({-, /, +, *}) as well as some of the other basic mathematical functions {power, √, ^{3}√, ln, e^{x}, Log, Exp}. The preliminary investigation of parse tree (and choosing the appropriate function set) shows that this function set has more accuracy. However, the full study about the effect of function set and parse tree on the models' performance is beyond the scope of this paper. The third step is to choose the chromosomal architecture. Length of head, h=8, and three genes per chromosomes are employed. The fourth step is to choose the linking function. The linking function must be chosen as "addition" or "multiplication" or "subtraction" for algebraic sub trees (Ferreira, 2001). Here, the sub trees are linked by subtraction. The fifth and final step is to choose the genetic operators. It is noted that these parameters are default values of GeneXpro program.

Data Set

394 flume and field data sets of flow rating curves from 31 different straight compound sections were selected for this study. Most of these data are collected form experimental works carried out by HR Wallingford (FCF) in compound channel flumes with large-scale facility (Knight and Sellin, 1987). Also, some experimental data from Blalock and Sturm (1981), Knight and Demetriou (1983), Lambert and Sellin (1996), Myers and Lyness (1997), Lambert and Myers (1998), Bousmar and Zech (1999), Haidera and Valentine (2002), Lai and Bessiah (2004), Bousmar et al. (2004) and Hu et al. (2010) have been used. Field data were collected from natural rivers including River Severn at Montford bridge (Ackers, 1992; Knight et al. 1989), River Main (Martin and Myers, 1991) and Rio Colorado (Tarrab and Weber, 2004). The cross section of a typical river compound channel is shown in Fig. 2. The ranges of geometric and hydraulic characteristics of compound channels as well as the mean values used in this paper are listed in Table 1. And Table 2 summarizes the ranges of dimensionless parameters used in this study.

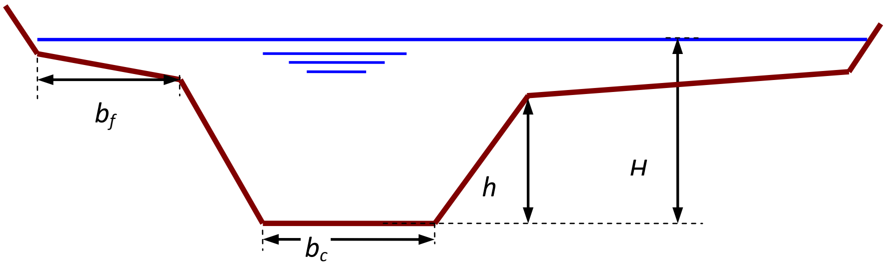

Figure 2. Typical natural river compound channel cross section.

Table1. Range of geometric and hydraulic variables of compound channels

Symbol	Variable Definition	Variable Range	Mean Value
h(m)	Bankfull height	0.031 – 6	0.811
b_c(m)	Main channel width	0.152 – 21.4	3.2
b_f(m)	Floodplain width	0 – 63	6.5
n_c	Manning’s roughness coefficient of main channel	0.01 – 0.036	0.0133
n_f	Manning’s roughness coefficient of floodplains	0.01 – 0.05	0.0166
S_0	Bed slope	0.000185 – 0.005	0.0011
H(m)	Flow depth	0.036 – 7.81	0.985
Q_b(m^3/s)	Bankfull discharge	0.00268 – 172.048	20.99
Q_t(m^3/s)	Total flow discharge	0.003 – 560	30.486

Table 2. Range of dimensionless parameters which are used in this study.

	Dr	COH	Q_{VDCM}/Q_b	Q_t/Q_b
Min	0.01549	0.6692	1.0462	1.0598
Max	0.73418	0.9119	17.924	11.786

RESULTS AND DISCUSSIONS

GEP Modeling

In this paper, using Gene Expression Programming, a dimensionless equation has been developed to obtain total flow discharge in compound channels with high accuracy. It is assumed; somewhat similar to Ackers (1992) approach, that discharge ratio in compound open channels is dependent on some dimensionless parameters through following equation:

$$\frac{Q_t}{Q_b} = f\left(Dr, COH, \frac{Q_{VDCM}}{Q_b}\right) \qquad (4)$$

where Q_t is total flow discharge, Q_b is bank-full discharge, Dr is depth ratio (ratio of water depth in floodplain to that of main channel), COH is coherence parameter and Q_{VDCM} is flow discharge calculated from Manning equation assuming vertical divided planes between main channel and floodplains. Of the total data set, approximately 70% (274 sets) were selected randomly and used for training. The remaining 30% (120 sets) were employed for testing. These three dimensionless parameters are most important factors explaining the complex features of compound channels hydraulics and hence, have been used by many authors such as Ackers (1992), Lambert and Myers (1998), Haiydera and Valentine (2002), Atabay and Knight (2006) and Huthoff et al (2008).

In this research, a two-point string crossover was utilised in the GEP. A segment of random length and random position is used in both parents and exchanged between parents. The crossover is abandoned and restarted by exchanging equalised segments when one of the resulting children exceeds the maximum length. The instruction operator is modified by mutation into another symbol over the same set (Brameier and Banzhaf 2001). In GEP, to avoid overgrowing programs, the maximum size of the program is generally restricted (Brameier and Banzhaf, 2001). This configuration was tested for the proposed GEP model and was found to be sufficient. The best individual (program) of a trained GEP can be converted into a functional representation by successive replacements of variables starting with the last effective instruction (Oltean and Groşan 2003). In this paper, only two basic mathematical functions (multiplication and power) and a large number of generations were used for testing. Table 3 shows the operational parameters and functional set used per run the GEP modelling.

Using optimization procedure, following relationship has been obtained for training data:

$$\frac{Q_t}{Q_b} = A + B - C \qquad (5)$$

$$A = 3.9538Dr \times 0.457^{Dr + (1-COH)^{Dr}}$$

$$B = \frac{Q_{VDCM}}{Q_b}^{(1-COH)^{(1-COH)^{2^{Dr}}}}$$

$$C = Dr^{(1-COH)^{\frac{Q_{VDCM}}{Q_b}}} (1-COH)^{Dr^{2.462}}$$

Training and testing results of GEP modeling

To validate the results for the training and testing data sets, several common statistical measures are used, such as R^2 (coefficient of determination), RMSE (root mean square error), the average error (AE) and mean absolute deviation.

$$R^2 = \left(\frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \right)^2 \qquad (6)$$

$$RMSE = \sqrt{\frac{\sum (X - Y)^2}{n}} \qquad (7)$$

$$AE = \frac{\sum \frac{X - Y}{X} \times 100}{n} \qquad (8)$$

$$\delta = \frac{\sum |X - Y|}{\sum X} \times 100 \qquad (9)$$

Where x= $(X - \overline{X})$, y= $(Y - \overline{Y})$, X are the observed values, $\overline{X}$ is the mean of X, Y is the predicted value, $\overline{Y}$ is the mean of Y, and n is the number of samples. These parameters have been calculated for training, testing and all data for VDCM and proposed GEP model and are depicted in Table 4.

Table 3. The parameters used per run inGEP Model

Parameter	Description of parameter	Parameter amount
P1	Chromosomes	30
P2	Genes	3
P3	Mutation Rate	0.044
P4	Inversion Rate	0.1
P5	Function set	× , power
P6	One-Point Recombination Rate	0.3
P7	Two-Point Recombination Rate	0.3
P8	Gene Recombination Rate	0.1
P9	Gene Transposition Rate	0.1
P10	Linking Function	Subtraction
P11	Program size	33

Table 4. Evaluation of models by statistical parameters

Method	Training				Testing			
	R^2	RMSE	%AE	Dev	R^2	RMSE	%AE	Dev
VDCM	0.69	1.723	-60.9	35.8	0.82	1.215	-38.4	47.2
GEP	0.98	0.243	1.07	7.13	0.97	0.323	-1.23	7.88

CONCLUSIONS

The computed discharge ratios (Q_t/Q_b) resulted from new GEP model for both training and testing data are presented in Figs. 3. It is clearly seen that this model in all variable ranges of selected data (laboratory and field sections), has very promised accuracy.

In Fig. 4, the calculated discharge ratios for VDCM are presented. As can be seen, this approach, in all over the data ranges, over-estimates the discharge ratios with large errors. This is due to ignoring the momentum exchange phenomenon developed at the main channel-floodplains interface lines. Based on these predictions, maximum relative error of discharge ratios for VDCM and GEP model have been calculated as 355 and 80 percent, respectively. The tabular results of statistical parameters are shown in Table 4. Based on error statistics and scatter plots, it is revealed that the GEP model with $R^2=0.98$, RMSE=0.243, AE=+1.07 and Dev=7.13 for training data sets and $R^2=0.97$, RMSE=0.323, AE=-1.23 and Dev=7.88 for testing data sets has the best accuracy, while the DVCM with $R^2=0.69$, RMSE=1.723, AE=-60.9 and Dev=35.8 for training data sets and $R^2=0.82$, RMSE=1.215, AE=-38.4 and Dev=47.2 for testing data sets has the worst accuracy. This comparison clearly shows that VDCM has failed to produce accurate results compared to the proposed GEP model.

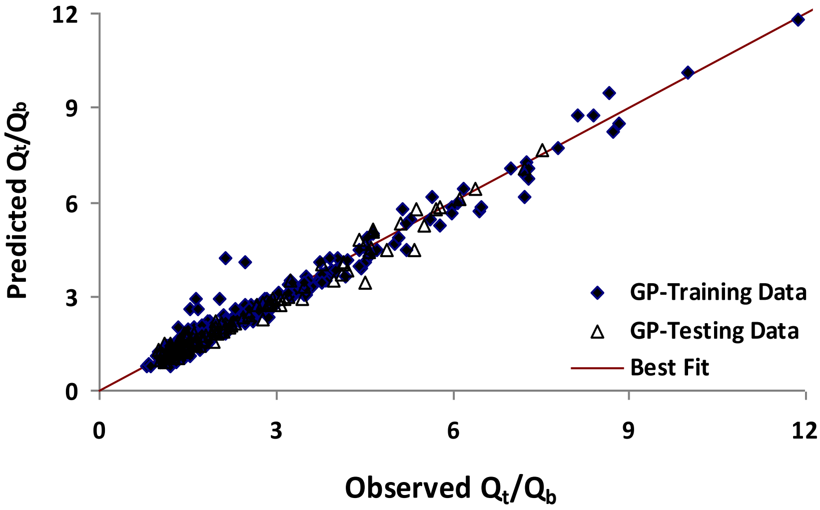

Figure 3. Comparison of observed and predicted discharge ratios from GEP model for training and testing data

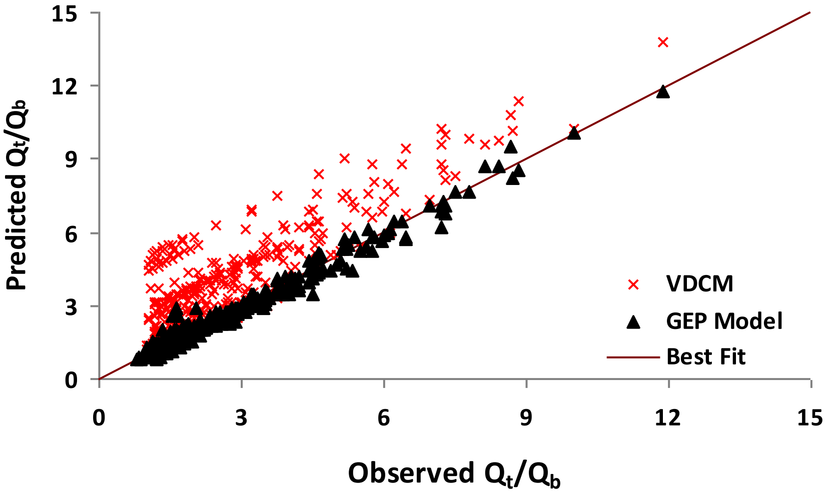

Figure 4. Comparison of observed and predicted discharge ratios from GEP model and VDCM

REFERENCES

Ackers P. (1992). Hydraulic design of two-stage channels. Journal of Water and Maritime Engineering, Vol. 96: 247-257.

Atabay S., Knight D.W. (2006). 1-D modeling of conveyance, boundary shear and sediment transport in overbank flow. J. Hydraul. Res., 44:6: 739-754.

Blalock M. E. and Sturm T. W. (1981). Minimum specific energy in compound channel. Journal of the Hydraulics Division, ASCE, 107: 699–717.

Bousmar D. and Zech Y. (1999). Momentum transfer for practical flow computation in compound channels. Journal of Hydraulic Engineering, ASCE, 125(7): 696-70.

Bousmar D., Wilkin N., Jacquemart, H. and Zech, Y. (2004). Overbank flow in symmetrically narrowing floodplains. Journal of Hydraulic Engineering, ASCE, 130(4): 305-312.

Brameier M., Banzhaf W. (2001). A comparison of linear genetic programming and neural networks in medical data mining. IEEE Trans. Evol. Comput., 5: 17–26.

Chow V. T. (1959). Open channel hydraulics, McGraw-Hill, London.

Ferreira C. (2001). Gene expression programming: a new adaptive algorithm for solving problems. Complex Systems, 13(2): 87-129.

Haidera M. A. and Valentine E. M. (2002). A practical method for predicting the total discharge in mobile and rigid boundary compound channels. International Conference on Fluvial Hydraulics, Belgium, 153-160.

Hu C., Ji Z., and Guo Q. (2010). Flow movement and sediment transport in compound channels. J. Hydraul. Res., 48:1: 23-32.

Huthoff F., Roose P. C., Augustijn, D. C. M., Hulscher, S.J.M.H. (2008). Interacting divided channel method for compound channel flow. Journal of Hydraulic Engineering, ASCE, 134(8): 1158-1165.

Johari A., Habibagahi G., Ghahramani, A. (2006). Prediction of soil-water characteristic curve using genetic programming. *Journal. Geotech. Geoenviron. Eng*., 132(5): 661–665.

Knight D. W. and Demetriou J. D. (1983). Flood plain and main channel flow interaction. Journal of Hydraulic Division, ASCE, 109(8): 1073-1092.

Knight D. W. and Sellin R. H. J. (1987). The SERC flood channel facility. Journal of Institution of Water and Environment Management, 1(2): 198-204.

Knight D. W., Shiono K., Pirt, J. (1989). Prediction of depth mean velocity and discharge in natural rivers with overbank flow. International Conference on Hydraulics and Environmental Modeling of Coastal, Estuarine and River Waters. England, 419-428.

Koza J.R. (1992). Genetic Programming: On the programming of computers by means of Natural Selection. The MIT Press, Cambridge, MA 840 pp.

Lai S. H. and Bessaih N. (2004). Flow in compound channels. 1st International Conference on Managing Rivers in the 21st Century, Malaysia, 275-280.

Lambert M. F. and Myers R. C. (1998). Estimating the discharge capacity in straight compound channels. Water, Maritime and Energy, 130: 84-94.

Lambert M. F. and Sellin R. H. J. (1996). Discharge prediction in straight compound channels using the mixing length concept. Journal of Hydraulic Research, IAHR, 34: 381-394.

Liu W and James C. S. (2000). Estimating of discharge capacity in meandering compound channels using artificial neural networks. Can. J. Civil. Eng., 27:2: 297-308.

MacLeod A. B. (1997). Development of methods to predict the discharge capacity in model and prototype meandering compound channels, PhD Thesis in Civil Engineering, University of Glasgow, 513 p.

Martin L. A. and Myers R. C. (1991). Measurement of overbank flow in a compound river channel. Journal of Institution of Water and Environment Management, 645-657.

Myers R. C. and Lyness J. F. (1997). Discharge ratios in smooth and rough compound channels. Journal of Hydraulics Engineering, ASCE, 123(3): 182-188.

Oltean M., Groşan C. (2003). A comparison of several linear genetic programming techniques. Complex Systems, 14(1): 1-29.

Sharifi S. (2009). Application of evolutionary computation to open channel flow modeling. PhD Thesis in Civil Engineering, University of Birmingham, 330 p.

Shiono K. and Knight D. W. (1991). Turbulent open-channel flows with variable depth across the channel. Journal of Fluid Mechanics, 222: 617-646.

Shiono K., Knight D.W. (1988). Two-dimensional analytical solution for a compound channel. 3rd International Symposium on Refined Flow Modeling and Turbulence Measurements, Japan, 503-510.

Tarrab L., Weber J. F. (2004). Predicción del coeficiente de mezcla transversal en cauces aturales. Mecánica Computacional, XXIII, Asociación Argentina de Mecanica Computacional, San Carlos de Bariloche, 1343-1355.

Unal B., Mamak M., Seckin, G., and Cobaner, M. (2010). Comparison of an ANN approach with 1-D and 2-D methods for estimating discharge capacity of straight compound channels, Advances in Engineering Software, 41: 120-129.

Wormleaton P.R. and Merrett D. J. (1990). An improved method of calculation for steady uniform flow in prismatic main channel/floodplain sections. Journal of Hydraulic Research, IAHR, 28: 157-174.

Zahiri A., and Dehghani A.A. (2009). Flow discharge determination in straight compound channels using ANN. World Academy of Science, Engineering and Technology, 58: 1-8.

Zahiri A., Ghabai M. and Mosaedi, A. (2012). Determination of Influence Parameters and its Optimal Combination in Order to Model Flood Rivers Discharge. Journal of Water and Soil, Vol. 25, No. 6: 1480-1493.