Modeling velocity distributions in small streams using different neuro-fuzzy and neural computing techniques

Onur Genc, Ozgur Kisi and Mehmet Ardiclioglu

ABSTRACT

Accurate estimation of velocity distribution in open channels or streams (especially in turbulent flow conditions) is very important and its measurement is very difficult because of spatio-temporal variation in velocity vectors. In the present study, velocity distribution in streams was estimated by two different artificial neural networks (ANN), ANN with conjugate gradient (ANN-CG) and ANN with Levenberg–Marquardt (ANN-LM), and two different adaptive neuro-fuzzy inference systems (ANFIS), ANFIS with grid partition (ANFIS-GP) and ANFIS with subtractive clustering (ANFIS-SC). The performance of the proposed models was compared with the multiple-linear regression (MLR) model. The comparison results revealed that the ANN-CG, ANN-LM, ANFIS-GP, and ANFIS-SC models performed better than the MLR model in estimating velocity distribution. Among the soft computing methods, the ANFIS-GP was observed to be better than the ANN-CG, ANN-LM, and ANFIS-SC models. The root mean square errors (RMSE) and mean absolute errors (MAE) of the MLR model were reduced by 69% and 72%, respectively, using the ANFIS-GP model to estimate velocity distribution in the test period.

Key words | ANFIS, ANN, modeling, velocity distribution

Onur Genc
Department of Civil Engineering,
Meliksah University,
Kayseri,
Turkey

Ozgur Kisi (corresponding author)
Faculty of Natural Sciences and Engineering,
Ilia State University,
0162, Tbilisi,
Georgia
E-mail: *ozgur.kisi@iliauni.edu.ge*

Mehmet Ardiclioglu
Department of Civil Engineering,
Erciyes University,
Kayseri,
Turkey

INTRODUCTION

Stream flow that can be described as quite difficult is an important study area for water resources and hydraulic engineers. Flows in streams are usually expressed by 1-D hydraulic equations. Many studies have been performed to determine the detailed properties of the hydrodynamics of complex flows using conventional methods, empirical formulas, and velocity samples (Hsu *et al.* 1998; Thomas & Williams 1999; Huang *et al.* 2002; Kar *et al.* 2015).

Three approaches, namely, experimental measurement, the theoretical method, and computer simulation are used to investigate flow properties in hydraulic engineering. Hydraulic systems usually show very complicated nonlinear behavior so it is not easy to get an analytical solution to describe the characteristics of these systems. Theoretical methods may be used to determine some simple flow cases (Kerh *et al.* 1994, 1997). Computer simulation using numerical methods such as the Computational Fluid Dynamics package is another approach to solve the fluid mechanics problems. It can be used to detect the properties of fluid motion in hydraulics engineering when the boundary conditions are properly defined. Flow measurement data are always extremely valuable for researchers who study in the field of hydraulic engineering. Usable measurement data are needed to corroborate its accuracy and to check its reliability in computer simulation (Kerh 2000, 2002). Prediction of velocity distribution is one of the basic properties of an open channel flow to analyze flow characteristics, particularly such as flow discharge, in the estimation of erosion and sediment transport in alluvial channels, shear stress, and watershed runoff which is used by hydraulic engineers. Also, recent researches have expressed that the profile of velocity in streams is the driver of habitat quality for aquatic species (Booker 2003). In this case, distribution of velocity must be investigated

doi: 10.2166/wcc.2019.103

and determined as a priority for solving hydraulics problems in open channels.

Numerous analytical and experimental studies have been conducted to obtain velocity distributions in stream flows (Kirkgoz 1989; Smart 1999; Ferro 2003). The power law and the Prandtl–von Karman universal velocity distribution law are well-known velocity distribution equations for open channel flows (Prandtl 1925; von Karman 1930). Unfortunately, the existing formulas cannot fully reveal the velocity profile, particularly near the channel bed and water surface. Most recently, an entropy concept based on the probabilistic approach was used to investigate velocity distributions in open channels (Chiu 1988; Xia 1997). According to the entropy method, there is a linear relationship between the mean and maximum velocity and it is described as an entropy parameter. Xia (1997) demonstrated that the relationship between the mean and maximum velocities was linear for all the river sections considered. The entropy concept, which is an alternative to the traditional method, is used to forecast flow properties. In the last decade, an artificial neural network (ANN) is another method that has been used to determine the velocity profile. The ANN and adaptive neuro-fuzzy inference system (ANFIS) techniques have been satisfactorily used to solve problems in water resources and hydraulic engineering. Yang & Chang (2005) simulated velocity profiles and velocity contours and estimated the discharges by ANN. Kocabas & Ulker utilized the ANFIS approach for predicting the critical submergence for an intake in a stratified fluid media (Kocabas & Ulker 2006). Dogan *et al.* (2007) utilized the ANN approach to forecast concentration of the sediment acquired by an experimental study. Mamak *et al.* (2009) successfully analyzed bridge afflux through arched bridge constrictions by ANFIS and ANN techniques. Kocabas *et al.* (2009) used the ANN method for estimating the critical submergence for an intake in a stratified fluid medium. Bilhan *et al.* (2010) estimated the lateral outflow over rectangular side weirs by using two different ANN techniques. Emiroglu *et al.* (2011) utilized the ANN approach for predicting the discharge capacity of a triangular labyrinth side weir situated on a straight channel. Emiroglu & Kisi (2013) used a neuro-fuzzy method to predict the discharge coefficient of trapezoidal labyrinth side weirs located on a straight channel. The flow discharge of weirs has been successfully predicted by Kisi *et al.* (2013) by ANFIS. Genc *et al.* (2014) analyzed the accuracy of ANN and ANFIS in determination of mean velocity and discharge of natural streams. They demonstrated that the ANFIS model, which has a determination coefficient (R^2) of 0.996, can be successfully predicted to mean velocity and discharge. In this paper, the applicability of two different ANN and ANFIS approaches for estimating velocity distribution of streams is investigated and the results are compared with the multiple-linear regression (MLR) model. For this purpose, field studies were carried out at different cross-sections in Kayseri by an acoustic Doppler velocimeter (ADV). These techniques have not been used for this purpose before.

VELOCITY DISTRIBUTIONS

Velocity profile should be determined for stream flows to better understand the structures of turbidity, sediment discharge, energy loss, and shear stress distributions (Ardiclioglu *et al.* 2012). Velocity distribution is influenced by vegetation, channel geometry, channel slope, roughness, and the presence of bends in rivers. In river flow, the velocity profile is not consistent at diverse depths. It increases from zero at the bottom of the channel to highest velocity near to the free water surface.

One of the most well-known velocity distribution models is the log-law (Sarma & Lakshminaraynan 1998). This logarithmic model is widely used to determine the two-dimensional velocity profile, particularly for hydraulic smooth and rough flow conditions. Two-dimensional open channel flows are divided into two zones, the inner and outer region, because of the existence of turbulence and the impact of a rigid boundary as shown in Figure 1. u_{sw} indicates the velocity at water surface, u_{max} shows maximum velocity in boundary layer height δ, water depth is shown as H, z_0 is depth where u equals zero. k represents the roughness coefficient.

Velocity distribution in the inner region is characterized as good by logarithmic distribution. The velocity profile in the inner region, presumed to be limited to $z/H < 0.20$, for uniform and steady nonuniform open channel flows, is

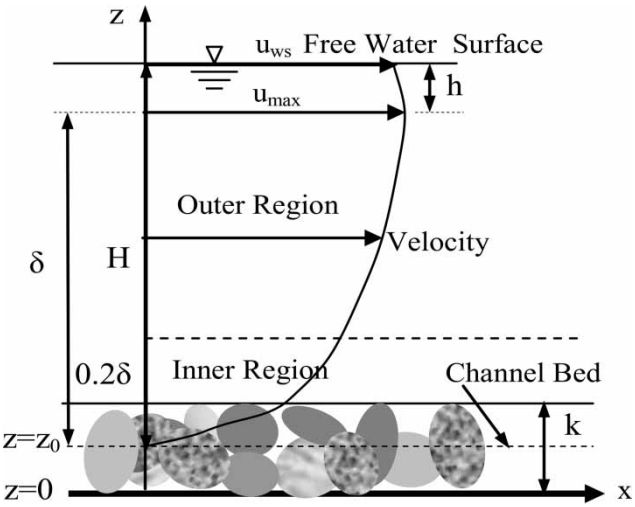

Figure 1 | The turbulent boundary layer and the velocity profile on the rough surface.

presented using the log law in Equation (1):

$$\frac{u}{u_*} = \frac{1}{c}\ln\left(\frac{z}{k_s}\right) + B_p \tag{1}$$

where u shows the streamwise, time-mean flow velocity, $u_* = \sqrt{\tau_0/\rho}$ indicates the shear velocity, (ρ = density) and τ_0 is the boundary shear stress, $\chi = 0.40$ ($1/\chi = 2.5$) is the von Karman's constant (the value of χ has a range of variation between 0.40 and 0.41), z is the distance from the bed, k_s is the equivalent sand roughness, and B_p is a constant of integration, being $B_p = 8.5 \pm 15\%$ (Song & Graf 1996). Although the value of B_p depends on the nature of the wall surface, the value of χ does not. When using this constant, Equation (1) is obtained as shown in Equation (2):

$$\frac{u}{u_*} = \frac{1}{c}\ln\left(\frac{z}{k_s/30}\right) \tag{2}$$

The logarithmic law is valid throughout the inner region except on the channel bed. In practical applications, it is still widely presumed that the logarithmic law explains the velocity profile along the whole depth of uniform, steady open channel flows (Kundu & Ghoshal 2012).

The power law is an alternative model to represent the vertical distribution of the stream-wise velocity in open channel flows. Many applications in water resources and hydraulic engineering have shown that velocity distribution measured in open channels can be expressed well by the power law (Montes 1998; Chanson 2004). The power law can be explained as a simple data-based equation in current theoretical studies. The power law exponent and constant can be obtained as to be quite empirical (Cheng 2007). Chen (1991) modified this law as it appeared as presented by the following equation:

$$\frac{u}{u_*} = a\left(\frac{z}{z_0}\right)^{1/m} \tag{3}$$

where the power law exponent is presented as $1/m$ and a means constant. Different m values have been presented in literature studies for different flows to determine velocity profile measurements with the power law. González *et al.* (1996) reported that $m = 1/6$ in their studies in open channels. Ardiclioglu *et al.* (2005) investigated power law equation constant *(a)* and exponent ($1/m$) in a stream and found them to be 4.0 and 1/5, respectively.

FIELD MEASUREMENTS

The four data sets of fixed ADV measurements presented in this paper were collected in the center of Turkey by a team comprising the first and third authors. Twenty-two field measurements were taken at four diverse cross-sections in the Kızılırmak and Seyhan basins. The first data set was collected between 2009 and 2010 at the Sosun station, which is in the Seyhan basin, and the stream is a tributary of the Zamantı River. The other data sets were obtained at the Barsama, Şahsenem, and Bünyan stations, which are in the Kızılırmak basin, a tributary of Kızılırmak River, between 2005 and 2010. Kızılırmak basin, which is the second biggest basin in Turkey, is located in the center of Turkey and the Black Sea region (Figure 2). Six site visits were carried out to the Barsama, Bünyan, and Şahsenem stations and three visits to Sosun station. In Figure 3, a sample of measurements at the Şahsenem station is shown.

The ADV was utilized to gather three-dimensional velocity data at the four stations. The ADV measures three-dimensional flow velocities (u, v, w) for x, y, z dimensions in a sampling volume utilizing the Doppler shift principle.

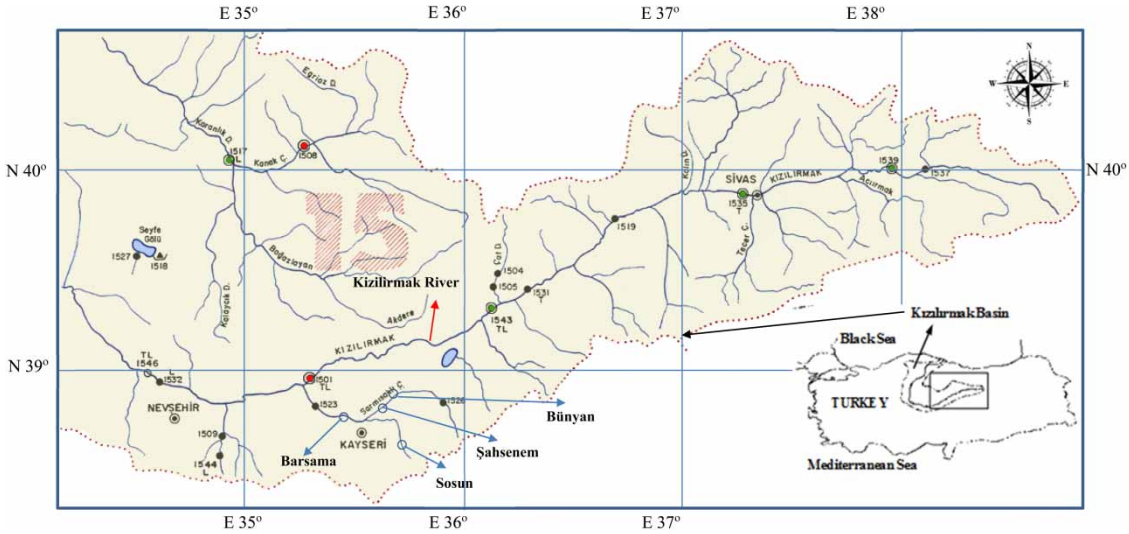

Figure 2 | Location of the study area and measurement stations at Bünyan, Şahsenem, Barsama, and Sosun.

Figure 3 | Field measurements at Şahsenem station [30].

At each measurement cross-section, the ADV records velocity data, location information, and water depth. The ADV sampling volume is found 10 cm before the probe head. Accordingly, the probe head itself has least effect on the flow field surrounding the measurement volume. Velocity reach is ±0.001 m/s to 4.5 m/s, resolution 0.0001 m/s, exactness ±1% of measured velocity (SonTek 2002). To decide the distribution of velocity in river flow, the experimental devices must be properly arranged.

During flow measurements, cross sections were divided into a number of slices for each flow condition according to the water surface width. Point velocity measurements were taken at different positions in the vertical direction starting 4 cm from the streambed for each vertical. The velocities of free water surface in all verticals were estimated by extrapolating the last two measurements of the verticals. Also, mean water surface velocities u_{ws} were measured at each of the visited stations. Water surface velocity can be effortlessly computed with an object that is movable on the water surface and is not too heavy, such as leaves, twigs, and so on.

The flow characteristics at every site are given in Table 1. In this table, the first and second columns show the station visit numbers and dates visited, U_m $(=Q/A)$ is the mean velocity. A is the area of the cross section, u_{ws} is the measured water surface velocity, H_{max} is the maximum flow depth, T is the surface water width, S_{ws} is water surface slope, Re $(=4U_mR/\upsilon)$ is the Reynolds number, and υ is the kinematic viscosity, and Fr $(=U_m/(gH_{max})^{1/2}$ represents the Froude number. When Froude and Reynolds numbers are calculated for all flow measurements, subcritical and turbulent flow conditions have been encountered.

ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM

ANFIS was initially presented by Jang (1993). It is a universal approximator and is equipped for approximating any real continuous function. The ANFIS structure is made out of

Table 1 | Flow characteristics for Barsama, Bünyan, Şahsenem, and Sosun stations

Stations	Dates (d/m/y)	U_m (m/s)	u_{ws} (m/s)	H_{max} (m)	T (m)	S_{ws}	Re ($\times 10^6$)	Fr
Barsama_1	28/05/2005	0.890	1.60	39.0	8.3	0.0091	0.76	0.481
Barsama_2	19/05/2006	1.051	1.85	40.0	9.0	0.0036	0.94	0.531
Barsama_3	19/05/2009	1.214	2.08	45.0	9.0	0.0094	1.47	0.578
Barsama_4	31/05/2009	0.590	1.14	26.0	8.4	0.0092	0.40	0.333
Barsama_5	24/03/2010	0.806	1.55	38.0	8.6	0.0097	0.61	0.417
Barsama_6	18/04/2010	0.865	1.63	38.2	8.8	0.0120	0.85	0.421
Bünyan_1	24/06/2009	0.354	0.65	72.0	4.0	0.0020	0.71	0.133
Bünyan_2	08/02/2010	0.214	0.40	66.0	4.0	0.0030	0.40	0.084
Bünyan_3	27/09/2009	0.301	0.54	72.0	3.9	0.0022	0.50	0.113
Bünyan_4	04/04/2010	0.405	0.74	85.0	4.0	0.0018	0.78	0.140
Bünyan_5	16/05/2010	0.426	0.54	86.0	4.0	0.0024	0.85	0.147
Bünyan_6	20/06/2010	0.286	0.53	79.0	3.9	0.0010	0.53	0.103
Şahsenem_1	29/03/2006	0.600	1.04	28.0	6.0	0.0059	0.47	0.350
Şahsenem_2	20/10/2007	0.529	0.93	32.0	5.4	0.0061	0.46	0.298
Şahsenem_3	22/03/2008	0.565	0.80	33.0	6.0	0.0037	0.49	0.314
Şahsenem_4	03/05/2008	0.518	1.00	32.0	5.4	0.0045	0.39	0.307
Şahsenem_5	11/10/2008	0.536	1.01	32.0	5.5	0.0046	0.44	0.303
Şahsenem_6	08/11/2008	0.516	1.00	34.0	5.6	0.0064	0.51	0.282
Sosun_1	19/05/2009	0.561	0.96	62.0	3.2	0.0032	0.84	0.227
Sosun_2	31/05/2009	0.285	0.63	43.0	3.0	0.0016	0.32	0.144
Sosun_3	24/03/2010	0.327	0.63	45.0	2.9	0.0026	0.37	0.156
Sosun_4	18/04/2010	0.541	0.93	54.0	2.3	0.0034	0.67	0.235

various nodes associated through directional connections and each node has a function comprising fixed or flexible parameters (Jang *et al.* 1997).

Assume that a fuzzy inference system with three inputs, x, y, and z and one output, f and the rule base composed of two fuzzy IF-THEN rules as:

$$\text{Rule 1: IF x is } A_1, \text{ y is } B_1 \text{ and z is } C_1 \text{ THEN } f_1 = p_1x + q_1y + r_1z + t_1 \quad (4)$$

$$\text{Rule 2: IF x is } A_2, \text{ y is } B_2 \text{ and z is } C_2 \text{ THEN } f_2 = p_2x + q_2y + r_2z + t_2 \quad (5)$$

Here, f_1 and f_2 allude to the output function of rule 1 and rule 2, separately. The ANFIS structure is shown in Figure 4. The node functions of each layer will be explained next.

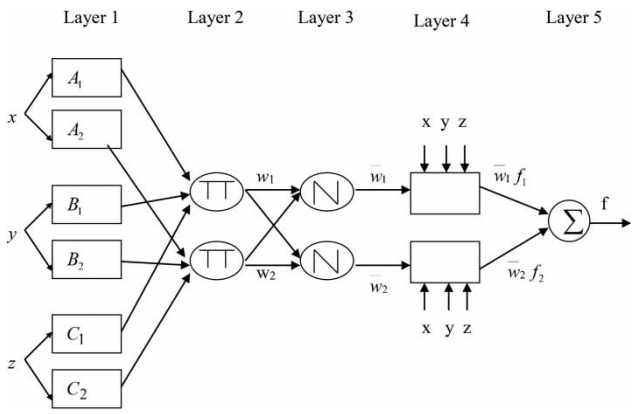

Figure 4 | ANFIS architecture.

Every node i in layer 1 is composed of an adaptive node function:

$$O_{l,i} = \varphi A_i(x), \quad \text{for i} = 1, 2 \quad (6)$$

where x is the i_{th} node's input and A_i is a linguistic label, for example, 'low' or 'high' connected with this node function. $O_{1,i}$ is the membership function of a fuzzy set A (=A_1, A_2, B_1, B_2, C_1, or C_2). It indicates the degree to which the given input x satisfies the quantifier A_i. $\varphi A_i(x)$ is typically chosen to be a Gaussian function with a minimum equivalent to 0 and maximum equivalent to 1:

$$\varphi A_i(x) = \exp\left(-\left(\frac{x - a_i}{b_i}\right)^2\right) \tag{7}$$

In Equation (7), a_i and b_i indicate the parameters. When the values of these parameters change, the Gaussian function also varies accordingly, thus exhibiting various forms of membership functions on linguistic label A_i (Jang 1993). This layer's parameters are called the premise parameters (Emiroglu & Kisi 2013).

Each node in the second layer multiplies the incoming signals and sends the product out. For instance:

$$w_i = \varphi A_i(x)\, \varphi B_i(y)\, \varphi C_i(z), \quad i = 1, 2. \tag{8}$$

The output of each node shows the firing strength of a rule.

The ratio of the i_{th} rule's firing strength to the sum of all rules' firing strengths in layer 3 is figured by i_{th} node as:

$$\bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2. \tag{9}$$

In the fourth layer, each node has a function as:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i z + t_i) \tag{10}$$

In Equation (10), $\bar{w}_i$ refers to the output of the third layer, and p_i, q_i, r_i, and t_i are the parameters. This layer's parameters are called the consequent parameters.

The single node in the fifth layer processes the final output as the summation of all incoming signals:

$$O_{5,i} = \sum_{i=1} \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{11}$$

In this manner, an ANFIS has been constructed which is practically equal to a first-order Sugeno FIS. More details on ANFIS can be obtained from the related references (Jang 1993).

ARTIFICIAL NEURAL NETWORKS

Artificial neural networks (ANN) are motivated by the natural nervous system but by neglecting a great part of the biological details. ANN is composed of numerous processing elements. The ANN is composed of layers comprising parallel processing components, called neurons. Each layer in the network is fully associated with the proceeding layer by interconnection. Arbitrarily assigned initial weight values are logically corrected during a training process. This process compares computed outputs to actual outputs and backpropagates any errors. Thus, the last weights are adjusted by minimizing the errors (Kisi 2005; Emiroglu & Kisi 2013). Each neuron in the second layer j gets the x input which is the weighted sum of outputs from the past layer. As a case, y for layer j is given by:

$$y_{pj} = \sum_{i=1}^{I} W_{ij} O_{pi} + \theta_j \tag{12}$$

where θ_j = a bias for neuron j, O_{pi} = i_{th} output of the past layer, W_{ij} = weight between the first layer i and j. An output $f(y)$ is calculated from each neuron in the second layer j and third layer k by passing its value of y through a non-linear activation function. An ordinarily utilized activation function is the logistic function:

$$f(y) = \frac{1}{1 + e^{-y}} \tag{13}$$

More details about ANN can be obtained from the related references (Haykin 2009).

APPLICATION AND RESULTS

Two different ANN methods, ANN with conjugate gradient (ANN-CG) and ANN with Levenberg–Marquardt

(ANN-LM), and two different ANFIS methods, ANFIS with grid partition (ANFIS-GP) and ANFIS with subtractive clustering (ANFIS-SC) were employed by using MATLAB program codes. The input parameters used for estimating the velocity distributions of the streams are the water surface velocity $u_{ws,}$ water surface slope S_{ws}, z/H, and y/T. The 2,184 measured data were used for the ANN-CG, ANN-LM, ANFIS-GP, ANFIS-SC, and regression analyses (MLR). After being randomly permutated, the data were split into two parts, training and testing. The first part (1,747 values, 80% of the whole data) was utilized for training and the second part (437 values, 20% of the whole data) was utilized for testing. Before application of the ANN models, the training input and output values were standardized using Equation (14):

$$a_1 \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} + a_2 \tag{14}$$

in which x_{max} and x_{min} are the maximum and minimum of the training and test data. In the present study, the a_1 and a_2 values were individually assigned as 0.6 and 0.2 and the input and output data were standardized somewhere around 0.2 and 0.8. Different ANN structures were tried to obtain the optimal models. The assessing criteria utilized in the applications are the root mean square errors (RMSE), mean absolute errors (MAE), and determination coefficient (R^2). The expressions of the RMSE and MAE are provided in the following equations:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{i,\,measured} - y_{i,\,estimate})^2} \tag{15}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_{i,\,measured} - y_{i,\,estimate}| \tag{16}$$

where N and y_i refer to the number of data sets and velocity, respectively.

For estimating the velocity distribution of the streams, four different input combinations were utilized. The correlations between the inputs u_{ws}, S_{ws}, z/H, and y/T and output are 0.714, 0.547, 0.362, and 0.010, respectively. According to the correlation values, u_{ws} seems to be the most effective variable on velocity distribution while y/T is the least effective one. The optimal hidden node numbers were obtained for the ANN models by the trial and error method. Two different ANN models were obtained by using the CG and LM algorithms which are more powerful and faster than the conventional gradient descent technique (Kisi 2007). The sigmoid activation functions were used for the hidden and output nodes. The training of the ANN networks was stopped after 1,000 iterations. Table 2 reports the training and test results of the optimal ANN models in estimating velocity distribution. The numbers given in the second column of the table indicate the optimal number of hidden nodes for each ANN model. It is clear from the table that the ANN-CG (4,7,1) model comprising four inputs corresponding to u_{ws}, S_{ws}, z/H, and y/T, seven hidden and one output nodes, has the lowest RMSE, MAE, and the highest R^2 than the other model both in training and test periods. Out of four ANN-LM models, the ANN-LM (4,9,1) model comprising four inputs performs better than the other models. The relative RMSE and MAE differences between the optimal ANFIS-LM and ANFIS-CG models are 12% in the test period.

The training and test results of the ANFIS-GP and ANFIS-SC models are given in Table 2 for each input combination. The optimal number of membership functions and parameter values are also provided in the second column of the table. From Table 2, it is clear that the ANFIS-GP model comprising four Gaussian membership functions for the inputs, u_{ws}, S_{ws}, z/H, and y/T performs better than the other ANFIS-GP models for both periods. There is a harmony between training and test results and this proves the proper calibration of the applied models. The comparison of ANFIS approaches reveals that the optimal ANFIS-GP model has a better accuracy than the optimal ANFIS-SC model with four inputs. According to the ANN-CG, ANN-LM, ANFIS-GP, and ANFIS-SC results given in Table 2, the y/T seems to be the most effective variable on velocity distribution in streams. However, y/T was found to be the least effective variable with respect to correlation (correlation = 0.010). This implies the strong nonlinear relationship between y/T and velocity distribution. This is also valid for z/H which also has considerable nonlinear effect on velocity distribution even though it has low correlation (correlation = 0.010). The main reason for this is the fact that these variables

Table 2 | Training and test results of the ANN, ANFIS, and MLR models in estimating velocity distribution

		Training			Test		
Input	**Parameters**	**RMSE**	**MAE**	**R^2**	**RMSE**	**MAE**	**R^2**
ANN-CG							
u_{ws}	9	0.244	0.187	0.512	0.243	0.190	0.604
u_{ws} and S_{ws}	8	0.238	0.180	0.536	0.237	0.186	0.626
u_{ws}, S_{ws} and z/H	7	0.171	0.134	0.758	0.167	0.128	0.810
u_{ws}, S_{ws}, z/H and y/T	7	0.106	0.081	0.908	0.098	0.074	0.934
ANN-LM							
u_{ws}	10	0.241	0.183	0.533	0.243	0.189	0.606
u_{ws} and S_{ws}	5	0.238	0.181	0.532	0.238	0.186	0.623
u_{ws}, S_{ws} and z/H	5	0.169	0.132	0.764	0.167	0.129	0.812
u_{ws}, S_{ws}, z/H and y/T	9	0.095	0.073	0.925	0.086	0.065	0.950
ANFIS-GP							
u_{ws}	(gaussmf, 3)	0.247	0.190	0.499	0.245	0.196	0.599
u_{ws} and S_{ws}	(gaussmf, 4)	0.237	0.180	0.536	0.239	0.187	0.618
u_{ws}, S_{ws} and z/H	(gaussmf, 2)	0.177	0.138	0.743	0.167	0.127	0.810
u_{ws}, S_{ws}, z/H and y/T	(gaussmf, 4)	0.037	0.027	0.989	0.066	0.046	0.971
ANFIS-SC							
u_{ws}	(gaussmf, 0.2)	0.240	0.183	0.523	0.242	0.189	0.610
u_{ws} and S_{ws}	(gaussmf, 0.2)	0.237	0.179	0.539	0.239	0.187	0.617
u_{ws}, S_{ws} and z/H	(gaussmf, 0.4)	0.174	0.137	0.752	0.171	0.132	0.802
u_{ws}, S_{ws}, z/H and y/T	(gaussmf, 0.4)	0.107	0.082	0.906	0.103	0.081	0.928
MLR							
u_{ws}	(0.58)	0.250	0.193	0.488	0.247	0.196	0.591
u_{ws} and S_{ws}	(0.56;4.72)	0.250	0.192	0.489	0.246	0.195	0.593
u_{ws}, S_{ws} and z/H	(0.42;4.51;0.31)	0.223	0.172	0.621	0.217	0.168	0.726
u_{ws}, S_{ws}, z/H and y/T	(0.48;4.68;0.37; − 0.21)	0.216	0.164	0.624	0.211	0.163	0.712

(y/T and z/H) are related to wetted area and velocity is proportional to the wetted area (continuity equation, $V = Q/A$, where V, Q, and A are mean velocity, discharge, and wetted area, respectively). It is apparent from Table 2 that the first two input combinations provide low accuracy in the applied models. Although the u_{ws} and S_{ws} have high correlations with velocity, they are not sufficient for accurate estimation of velocity distribution. Table 2 also reports the RMSE, MAE, and R^2 statistics of the MLR models. The second column of this table indicates the regression coefficients of the MLR models. The MLR model with four inputs performs better than the other MLR models. Comparison of ANN-CG, ANN-LM, ANFIS-GP, ANFIS-SC, and MLR models reveals that the data-driven ANN and ANFIS based models perform better than the MLR model in estimating velocity distribution. The optimal ANFIS-GP model comprising four input combinations has the lowest RMSE (0.066) and MAE (0.046) and the highest R^2 (0.971) values. The main advantage of the ANFIS-GP over ANFIS-SC method is that it uses all possible rule combinations in its structure while ANFIS-SC decreases the possible rules to some limited numbers by using a clustering algorithm. ANFIS-SC has simpler structure compared to ANFIS-GP but it has less accuracy than the latter. Table 3 reports the cross-correlations among the estimates of the optimal models with

Table 3 | The cross-correlations among the estimates of the optimal models with four input variables in the test period

	ANN-CG	ANN-LM	ANFIS-GP	ANFIS-SC	MLR
ANN-CG	1				
ANN-LM	0.991	1			
ANFIS-GP	0.963	0.970	1		
ANFIS-SC	0.975	0.974	0.961	1	
MLR	0.855	0.856	0.838	0.857	1

four input variables in the test period. It is apparent that the MLR has the lowest correlations with soft computing models. The ANN-LM has the highest correlation with ANFIS-GP and this indicates that the ANN-LM has the second rank in estimating velocity distribution.

The measured and estimated velocity distributions by the ANN-CG, ANN-LM, ANFIS-GP, ANFIS-SC, and MLR models are shown in Figure 5 for the test period. It is clear that the estimates of the ANN and ANFIS models are closer to the corresponding measured velocities than the MLR model. As seen from the figure, ANN-CG, ANN-LM, and ANFIS-SC models underestimate some peaks while the ANFIS-GP model generally well estimates. The significantly under- and over-estimations of the MLR model are clearly seen. The scatterplots of the estimates in the test period are illustrated in Figure 6. It is clear from the fit line equations (assume that the equation is $y = ax + b$) given in the scatterplots that the a and b coefficients of the ANFIS-GP model are individually closer to 1 and 0 with a higher R^2 than those of the ANN-CG, ANN-LM, ANFIS-SC, and MLR models. It is clear from Figure 6 that the MLR model is insufficient in estimating the velocity distribution of the natural streams.

In order to verify the robustness (the significance of differences between the model estimates and measured velocity values) of the models, the results are likewise tested by utilizing one-way ANOVA. The test is set at a 95% significant level. The test statistics are provided in Table 4. The ANN and ANFIS models yield small testing values with high significance levels in respect to the MLR. According to the ANOVA results, the ANFIS and ANN models are more robust (the closeness between the measured velocity values and model estimates are significantly high) in estimating velocity distribution than the MLR model. Among the data-driven models, the ANN-CG and ANFIS-SC models seem to be better than the ANN-LM and ANFIS-GP regarding robustness.

CONCLUSION

Estimating velocity distribution in streams by ANN-CG, ANN-LM, ANFIS-GP, ANFIS-SC, and MLR approaches was examined in this study. The 2,184 field data gauged from four diverse cross sections at four destinations on the Sarımsaklı and Sosun streams in central Turkey were used

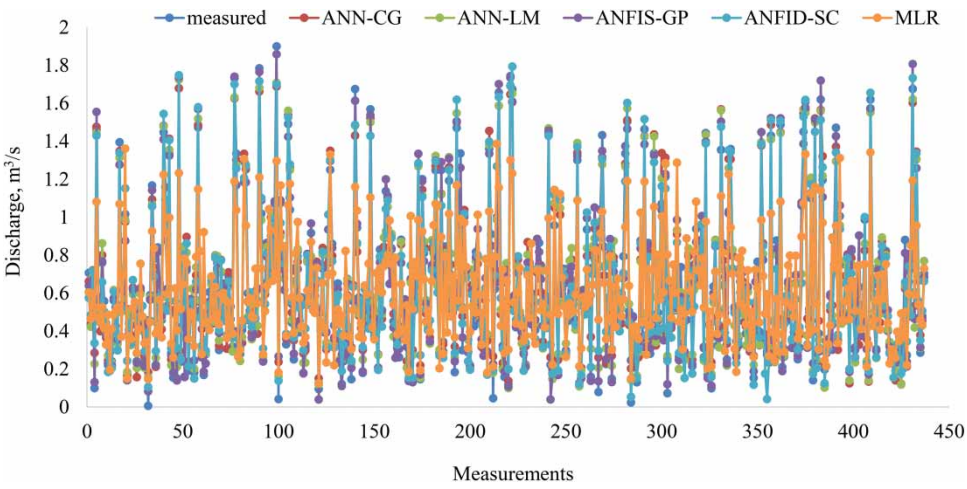

Figure 5 | Time variation of the measured and estimated velocity distribution by ANN-CG, ANN-LM, ANFIS-GP, ANFIS-SC and MLR models.

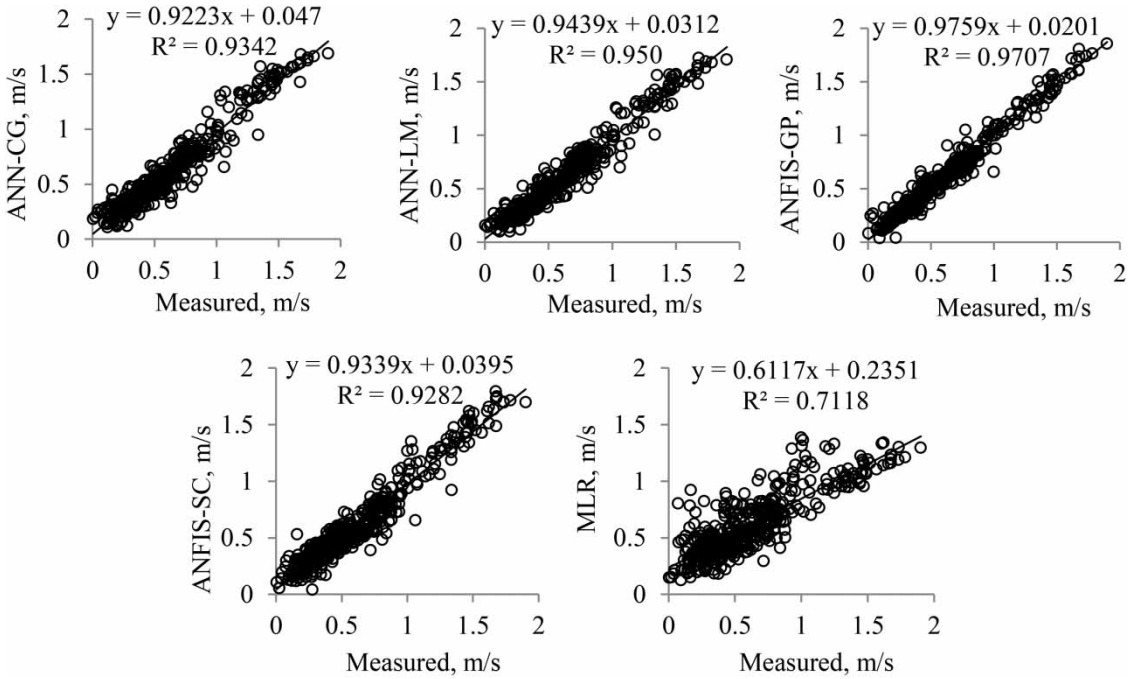

Figure 6 | The scatterplots of the measured and estimated velocity distributions by ANN-CG, ANN-LM, ANFIS-GP, ANFIS-SC and MLR models.

Table 4 | ANOVA of ANN-CG, ANN-LM, ANFIS-GP, ANFIS-SC, and regression models in the test period

Model	F-statistic	Resultant significance level
ANN-CG (4,7,1)	0.003	0.959
ANN-LM (4,9,1)	0.021	0.885
ANFIS-GP (gaussmf, 4)	0.038	0.845
ANFIS-SC (gaussmf, 0.4)	0.004	0.950
MLR	0.076	0.783

in the applications. To predict velocity distribution, the water surface velocity u_{ws}, water surface slope S_{ws}, z/H, and y/T were utilized as inputs to the models. The accuracy of the ANN-CG, ANN-LM, ANFIS-GP, and ANFIS-SC models was compared with MLR models. Comparison results showed that the ANN and ANFIS models provided better accuracy than the regression model in estimating velocity distribution. The ANFIS-GP model was observed to be better than the ANN-CG, ANN-LM, and ANFIS-SC models. The optimal ANFIS-GP model separately reduced the RMSE and MAE by 69% and 72% and increased the determination coefficient by 36% with respect to the optimal MLR model. The study suggests that the ANFIS and ANN techniques can be effectively utilized for estimating the velocity distribution of the streams.

In this study, 2,184 field data measured in four cross-sections at four destinations were used for model development. More data from different places may improve the models' accuracy. In this way, generalization of the applied models may be improved.

REFERENCES

Ardiclioglu, M., Araujo, J. C. & Senturk, A. I. 2005 Applicability of velocity distribution equations in rough-bed open-channel flow. *La. Houille Blanche* **4**, 73–79.

Ardiclioglu, M., Genc, O., Kalin, L. & Agiralioglu, N. 2012 Investigation of flow properties in natural streams using the entropy concept. *Water and Environment Journal* **26** (2), 147–154.

Bilhan, O., Emiroglu, M. E. & Kisi, O. 2010 Application of two different neural network techniques to lateral outflow over rectangular side weirs located on a straight channel. *Advances in Engineering Software* **41** (6), 831–837.

Booker, D. J. 2003 Hydraulic modeling of fish habitat in urban rivers during high flows. *Hydrological Processes* **17**, 577–599.

Chanson, H. 2004 *The Hydraulics of Open Channel Flow: An Introduction*, 2nd edn. Elsevier Butterworth-Heinemann, Oxford, UK and Burlington, MA, USA, p. 585.

Chen, C. L. 1991 Unified theory on power laws for flow resistance. *Journal of Hydraulic Engineering, ASCE* **117** (3), 371–389.

Cheng, N. S. 2007 Power-law index for velocity profiles in open channel flows. *Advances in Water Resources* **30** (8), 1775–1784. doi: 10.1016/j.advwatres.2007.02.001.

Chiu, C. L. 1988 Entropy and 2-D velocity distribution in open channel. *Journal of Hydraulic Engineering* **114** (7), 738–756.

Dogan, E., Yuksel, I. & Kisi, O. 2007 Estimation of sediment concentration obtained by experimental study using artificial neural networks. *Environmental Fluid Mechanics* **7**, 271–288.

Emiroglu, M. E. & Kisi, O. 2013 Prediction of discharge coefficient for trapezoidal labyrinth side weir using a neuro-fuzzy approach. *Water Resources Management* **27** (5), 1473–1488.

Emiroglu, M. E., Bilhan, O. & Kisi, O. 2011 Neural networks for estimation of discharge capacity of triangular labyrinth side-weir located on a straight channel. *Expert Systems with Applications* **38** (1), 867–874. DOI: 10.1016/j.eswa.2010.07.058.

Ferro, V. 2003 ADV measurements of velocity distributions in a gravel bed flume. *Earth Surface Processes and Landforms* **28**, 707–722.

Genc, O., Kisi, O. & Ardiclioglu, M. 2014 Determination of mean velocity and discharge in natural streams using neuro-fuzzy and neural network approaches. *Water Resources Management* **28**, 2387–2400. DOI 10.1007/s11269-014-0574-6.

González, J. A., Melching, C. S. & Oberg, K. A. 1996 Analysis of open-channel velocity measurements collected with an acoustic Doppler current profiler. In: *Proceedings of the 1st International Conference on New/Emerging Concepts for Rivers*, Chicago, IL, USA.

Haykin, S. 2009 *Neural Networks and Learning Machines*, 3rd edn. Prentice Hall, Upper Saddle River, NJ, USA.

Hsu, C. C., Wu, F. S. & Lee, W. J. 1998 Flow at 900 equal-width open channel junction. *Journal of Hydraulic Engineering, ASCE* **124** (2), 186–191.

Huang, J., Weber, L. J. & Lai, Y. G. 2002 Three-dimensional numerical study of flows in open-channel junctions. *Journal of Hydraulic Engineering, ASCE* **128** (3), 268–280.

Jang, J. S. R. 1993 ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics* **23** (3), 665–685.

Jang, J. S. R., Sun, C. T. & Mizutani, E. 1997 *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice Hall, Upper Saddle River, NJ, USA.

Kar, K. K., Yang, S. K. & Lee, J. H. 2015 Assessing unit hydrograph parameters and peak runoff responses from storm rainfall events: a case study in Hancheon Basin of Jeju Island. *Journal of Environmental Science International* **24** (4), 437–447.

Kerh, T. 2000 Application of Galerkin Time Scheme to investigate unsteady flow around an inclined plate. *International Journal of Modelling and Simulation* **20** (1), 69–78.

Kerh, T. 2002 Computing of viscous fluid in a backflow preventer with an oscillating boundary. *International Journal of Fluid Mechanics Research* **29** (2), 1–15.

Kerh, T., Chung, W. G. & Ting, C. S. 1994 Studies of varying behaviors for fluid over a checking valve. *Journal of Taiwan Water Conservancy* **42** (4), 67–73.

Kerh, T., Lee, J. J. & Wellford, L. C. 1997 Transient fluid-structure interaction in a control valve. *Journal of Fluids Engineering. ASME* **119**, 354–359.

Kirkgoz, M. S. 1989 Turbulent velocity profiles for smooth and rough open channel flow. *Journal of Hydraulic Engineering* **115** (11), 1543–1561.

Kisi, O. 2005 Suspended sediment estimation using neuro-fuzzy and neural network approaches. *Hydrological Sciences Journal* **50** (4), 683–696.

Kisi, O. 2007 Streamflow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering, ASCE* **12** (5), 532–539.

Kisi, O., Bilhan, O. & Emiroglu, M. E. 2013 ANFIS to estimate discharge capacity of rectangular side weir. *Water Management* **166**, 479–487.

Kocabas, F. & Ulker, S. 2006 Estimation of critical submergence for an intake in a stratified fluid media by neuro-fuzzy approach. *Environmental Fluid Mechanics* **6**, 489–500.

Kocabas, F., Kisi, O. & Ardiclioglu, M. 2009 An artificial neural network model for prediction of critical submergence for an intake in a stratified fluid media. *Civil Engineering and Environmental Systems* **26** (4), 367–375.

Kundu, S. & Ghoshal, K. 2012 Velocity distribution in open channels: combination of log-law and parabolic-law. *World Academy of Science, Engineering and Technology* **6** (8), 1234–1241.

Mamak, M., Seckin, G., Cobaner, M. & Kisi, O. 2009 Bridge afflux analysis through arched bridge constrictions using artificial intelligence methods. *Civil Engineering and Environmental Systems* **26** (3), 279–293.

Montes, S. 1998 *Hydraulics of Open Channel Flow*. ASCE, Reston, VA, USA, p. 697.

Prandtl, L. 1925 Über die Ausgebildate Turbulenz. *Z. Angew Math. Mech. Bd.* **5**, 136–139.

Sarma, K. V. N. & Lakshminaraynan, P. R. 1998 NSL velocity distribution in smooth rectangular open channels. *Journal of Hydraulic Engineering, ASCE* **109** (2), 270–289.

Smart, G. M. 1999 Turbulent velocity profiles and boundary shear in gravel-bed rivers. *Journal of Hydraulic Engineering* **125** (2), 106–116.

Song, T. & Graf, H. 1996 Velocity and turbulence distribution in unsteady open-channel flows. *Journal of Hydraulic Engineering, ASCE* **122** (3), 141–154.

SonTek Flow Tracker Handheld ADV 2002 *Technical Document*, San Diego, CA, USA.

Thomas, T. G. & Williams, J. R. 1999 Large-eddy simulation of flow in a rectangular open-channel. *Journal of Hydraulic Research* **37** (3), 345–361.

von Karman, T. 1930 Mechanische Ahnlichkeitund und Turbulenz. Nachrichten der Gesellschaft der Wissenschaften Göttingen. *Mathematisch-Physikalisch Klasse* 58–60.

Xia, R. 1997 Relation between mean and maximum velocities in a natural river. *Journal of Hydraulic Engineering* **123** (8), 720–723.

Yang, H. C. & Chang, F. J. 2005 Modeling combined open channel flow by artificial neural networks. *Hydrological Processes* **19**, 3747–3762.

First received 12 April 2018; accepted in revised form 3 November 2018. Available online 29 January 2019