



EPiC Series in Engineering

Volume 3, 2018, Pages 155–161

HIC 2018. 13th International
Conference on Hydroinformatics



Weekly Flow Prediction of Ergene River using an Artificial Neural Network Based Solution Approach

M. Tamer Ayvaz^{1*}, Ulas Tezel², Elcin Kentel³ and Recep Kaya Goktas⁴

¹ Department of Civil Engineering, Pamukkale University, Denizli, Turkey

² Institute of Environmental Sciences, Bogazici University, Istanbul, Turkey

³ Department of Civil Engineering, Middle East Technical University, Ankara, Turkey

⁴ Department of Environmental Engineering, Kocaeli University, Kocaeli, Turkey

tayvaz@pau.edu.tr, ulas.tezel@boun.edu.tr

ekentel@metu.edu.tr, rkaya.goktas@kocaeli.edu.tr

Abstract

The objective of this study is to develop an artificial neural network (ANN) based solution approach to predict the weekly flows of Ergene River which is the largest river in Thrace region of Turkey. In the developed approach, precipitation – flow data relationships have been investigated in order to establish the best model structure to predict streamflow at the selected basin. The developed relationships are then evaluated using a feed forward neural network where back propagation algorithm is used to determine the associated network weights. The performance of the developed ANN based solution approach is evaluated by using the weekly precipitation and flow data collected from different monitoring sites in Ergene River basin. The model results are also compared with HEC-HMS model outputs which is calibrated using the same precipitation and flow data. Results indicate that the proposed ANN based solution approach can be effectively used to predict the weekly flows of Ergene River.

1 Introduction

Prediction of river flows is an important part of water resources planning and management. Since the amount of appropriate quality water is decreasing each day, it is required to develop sustainable strategies for managing the limited amount of water resources. One of the most important elements of water resources management studies is the prediction of the available water, especially the flow of the rivers (Rezaei et al., 2015). Although hydrological models can be used in predicting streamflow,

* Corresponding author

development of a well-calibrated model generally requires large amount of data to simulate rainfall-runoff process in the basin. Rainfall-runoff process is a complex process and is hard to mathematically model all of its components (i.e. surface runoff, infiltration, evapotranspiration, interflow, groundwater flow, etc.). Moreover, hydrologic models often require various basin-specific data including soil type, curve number, infiltration rate, percolation rate, soil moisture, evaporation, etc. In the absence of these different types of basin-specific data, calibration of the hydrological model may turn out to be very difficult or impossible. In such situations, data driven approaches become useful tools in predicting the river flow (Jayawardena and Fernando, 2001). Data driven approaches are based on establishing nonlinear relationships between input and output variables. One of the well-known data driven approaches is ANNs since they can effectively establish the relationship between input and output data.

This study is a part of a TUBITAK project entitled “Development of a Geographical Information System based Decision-Making Tool for Water Quality Management of Ergene Watershed using Pollutant Fingerprints”. Within the scope of this project, the first task is to study the water resources in Ergene River Basin. For this purpose, stream and rainfall gages in the basin are identified and the associated flow and precipitation data are obtained. An ANN model is developed to predict the streamflow at stream gage, E01A012 that is located very close to the outlet of the basin (see Figure 1). Different model architectures using weekly precipitation and streamflow data at different gaging stations within the basin are developed in this study. Results of these ANN models are compared with those obtained from the hydrological model developed in HEC-HMS for Ergene River Basin. It is observed that the results of the ANN model are better than those of the HEC-HMS model for the selected stream gage.

2 Study Area

Ergene River Basin is located in the Thrace region of Turkey (Figure 1). The basin has a drainage area of 12,438 km² which corresponds to 1.6% of the Turkey’s total surface area. At the outlet of the basin, Ergene River merges with Meriç River and flows into the Aegean Sea. Figure 1 shows the location map of Ergene River basin together with the drainage network under consideration.

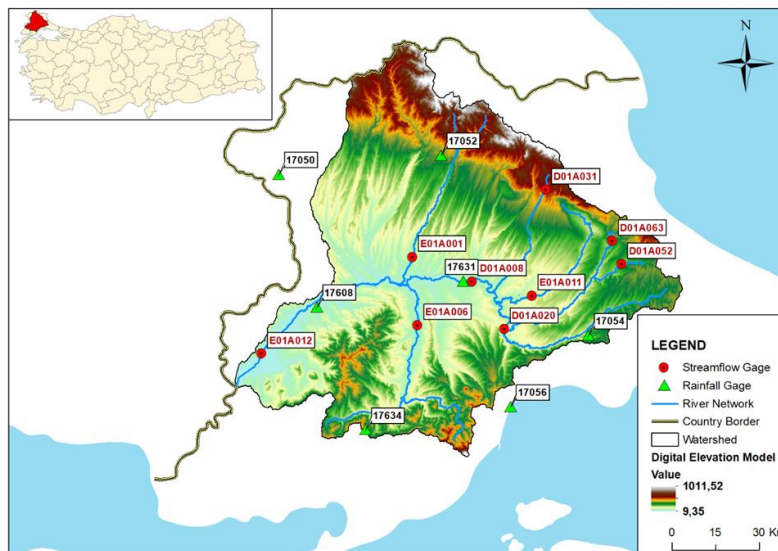


Figure 1: The location map of the Ergene River Basin

Although there are many rainfall and streamflow gages in the basin, the ones given in Figure 1 are used since daily precipitation and streamflow observations are simultaneously available between 1996 and 2005 at these sites. This data is used to calculate weekly precipitation and flow volumes that are used in the training of ANN models. Thiessen polygon approach is used to convert point precipitation data to areal values. Figure 2 shows observed weekly average precipitation volumes over the basin and observed weekly flow volumes at stream gage E01A012.

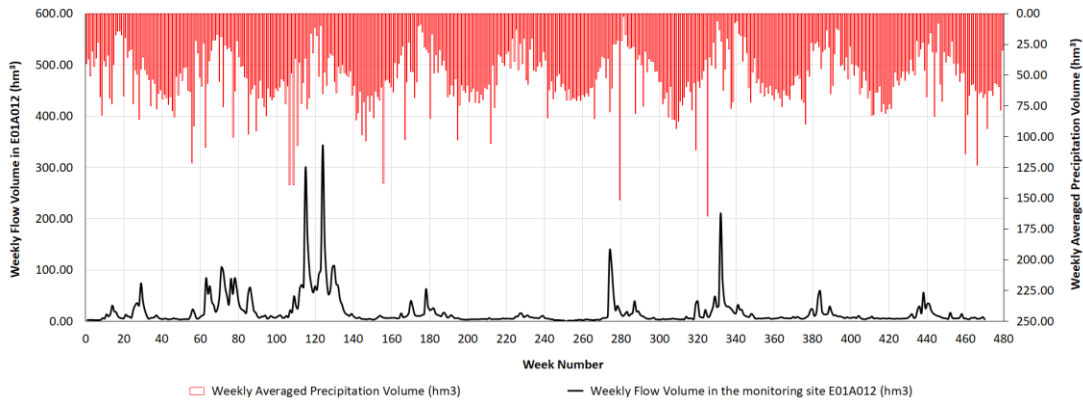


Figure 2: Observed weekly average precipitation and flow volumes in stream gage E01A012

3 Model Development

ANN is a computational tool which is inspired from the behaviour of biological neural network system. By processing the provided historical data, the ANN can adaptively learn the behaviour of the system by adjusting the network weights between input and output data (Coppola et al., 2003). Figure 3 shows a typical multi-layer feed forward ANN structure. By using the terms given in Figure 3, the mathematical structure of the ANN model can be stated as follows:

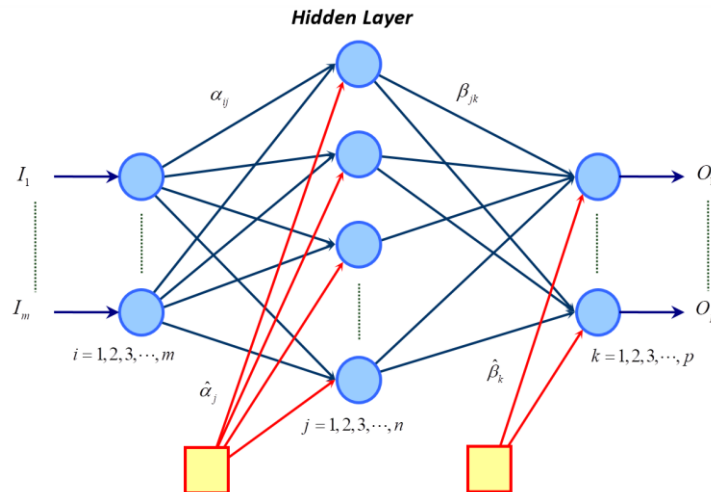


Figure 3: A multi-layer feed forward ANN structure (Ayvaz and Popovici, 2013)

Let m be the number of input variables, n be the number of hidden layer neurons, p be the number of output variables, $\mathbf{I} = \{I_i\}_{i=1}^m$ and $\mathbf{O} = \{O_k\}_{k=1}^p$ be the vectors containing the input and output variables, $\boldsymbol{\alpha} = [\alpha_{ij}]_{i=1|j=1}^{m|n}$ and $\boldsymbol{\beta} = [\beta_{jk}]_{j=1|k=1}^{n|p}$ be the matrices including the network weights between input to hidden and hidden to output neurons, respectively, $\hat{\boldsymbol{\alpha}} = [\hat{\alpha}_j]_{j=1}^n$ and $\hat{\boldsymbol{\beta}} = [\hat{\beta}_k]_{k=1}^p$ be the vectors including bias weights for hidden and output layers, respectively, and $f(\circ)$ be the sigmoid type transfer function. Using these definitions, the elements of the output vector can be calculated as follows:

$$\mathbf{O} = f(\hat{\boldsymbol{\beta}} + \boldsymbol{\beta} \cdot f(\hat{\boldsymbol{\alpha}} + \boldsymbol{\alpha} \cdot \mathbf{I})) \tag{1}$$

It should be noted that the primary objective of the ANN model given in Figure 3 is to establish a relationship between the elements of \mathbf{I} and \mathbf{O} . This task is conducted by determining $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\hat{\boldsymbol{\alpha}}$, and $\hat{\boldsymbol{\beta}}$ by means of the back propagation algorithm. Note that various input – output relationships have been investigated in this study. Among them, the input – output structures given in Table 1 are used.

Input Variables				Number of Hidden Layer Neurons	Output Variable
M_t	P_t^{ave}	$Q_t^{D01A008}$	$Q_t^{D01A020}$	5, 10, 15, 20, 25, 30	$Q_t^{E01A012}$
$Q_t^{D01A031}$	$Q_t^{D01A052}$	$Q_t^{D01A063}$	$Q_t^{E01A001}$		
$Q_t^{E01A006}$	$Q_t^{E01A011}$				

Table 1: Comparison of the estimated \mathbf{T} and \mathbf{S} values for each piezometer

In Table 1, t is the week number, M_t is the corresponding month of week t , P_t^{ave} is the average precipitation volume over the basin at week t , Q_t^x is the weekly flow volume at monitoring site x and week t . As can be seen, the objective of the ANN model is to predict the streamflow volume at E01A012 using input parameters given in Table 1. The network training process is performed by considering 6 different architectures in terms of different number of hidden layer neurons as given in Table 1.

It should be noted that one of the most important problems on ANN model training is the over-fitting problem which causes to lose the generalization ability of the ANN models. The best way to control this problem is to divide the available data into three parts; training, validation, and testing. Training data has the largest portion among the others and is directly used for training the model. Validation data is used to control if over-fitting occurs. This control is conducted by checking the trend of the validation error during the training stage. Over-fitting occurs when the value of validation error starts continuously increasing while training error decreases (Bhattacharjya and Datta, 2005). At this point, the training process should be stopped and the current network weights should be used (Ayvaz and Popovici, 2013). Finally, the testing data which is different than the training and validation data is used to test the prediction performance of the ANN model.

4 Results

In this section, results of the ANN based solution approach is presented. Note that model training, validation, and testing is performed by using the weekly input – output data between 1st Oct 1996 and 30th Sep 2005. 70% of the data is randomly selected for model training and the remaining 15% + 15% are selected for validation and testing, respectively. This means that 30% of the data is not used during calculation of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\hat{\boldsymbol{\alpha}}$, and $\hat{\boldsymbol{\beta}}$. Using these data partitions, the developed ANN model is trained by using the back propagation algorithm for 10,000 iterations. For each case in Table 1, Figure 4 compares the

mean absolute training errors (MATE) and the locations of the points where the problem of over-fitting starts. As can be seen, the problem of over-fitting is first observed in $10 \times 30 \times 1$ and the last observed for $10 \times 25 \times 1$. When the error values of this points are compared, the maximum training error is observed in $10 \times 5 \times 1$ whereas the minimum is observed for $10 \times 25 \times 1$. For the calculated network weights, the Nash-Sutcliffe model efficiency (NS) values are compared in Figure 5 for training, validation, and testing data.

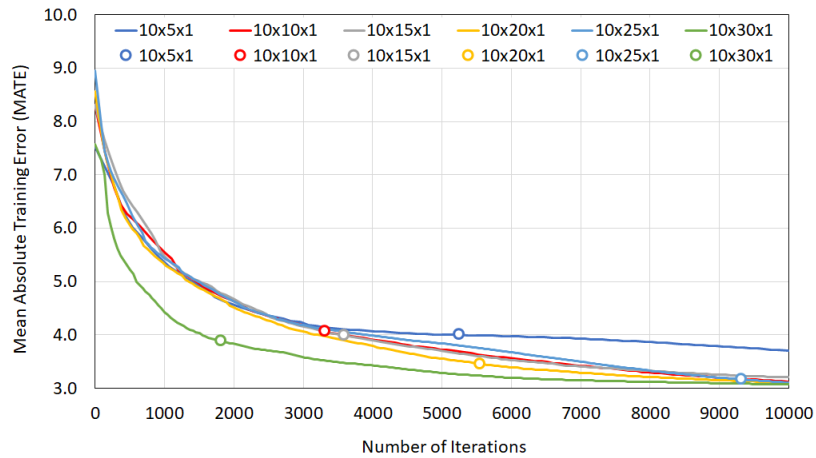


Figure 4: Comparison of the MATE values for each case (the points on the plots represent the starting locations of the over-fitting problem)

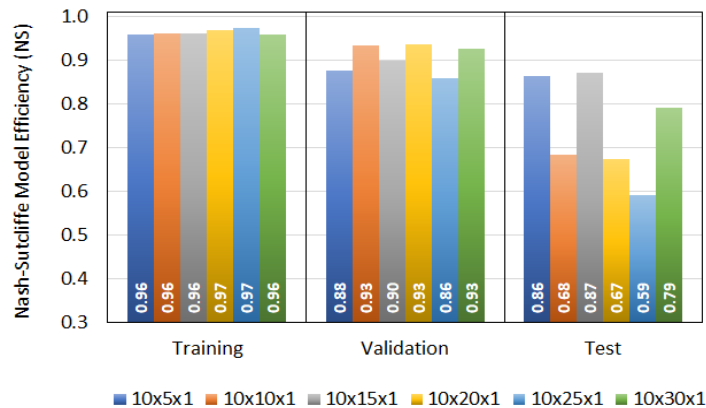


Figure 5: Comparison of the calculated NS values for each architecture

As can be seen from Figure 5, the calculated NS values in the training stage are over 0.96 for each case. In the validation and testing parts, the calculated NS values change from 0.86 ($10 \times 25 \times 1$) to 0.93 ($10 \times 20 \times 1$) and from 0.59 ($10 \times 25 \times 1$) to 0.87 ($10 \times 15 \times 1$), respectively. Evaluating these results through an engineering point of view, the network structure of $10 \times 15 \times 1$ is identified as the best network architecture for the given data since it has the best generalization performance in terms of the calculated NS values in training, validation, and testing parts. For the selected architecture of $10 \times 15 \times 1$, the observed and predicted flow hydrographs are compared in Figure 6 by considering entire data and also the data of the training, validation, and testing stages, separately. These comparisons are conducted by plotting all the flow hydrographs by using the same network weights determined in

the training stage. Note that horizontal axis of each chart includes 470 weeks since the flow data is randomly partitioned into three parts. Figure 6 also provides the results of the HEC-HMS hydrologic model which is calibrated by using the same data collected at Ergene River basin.

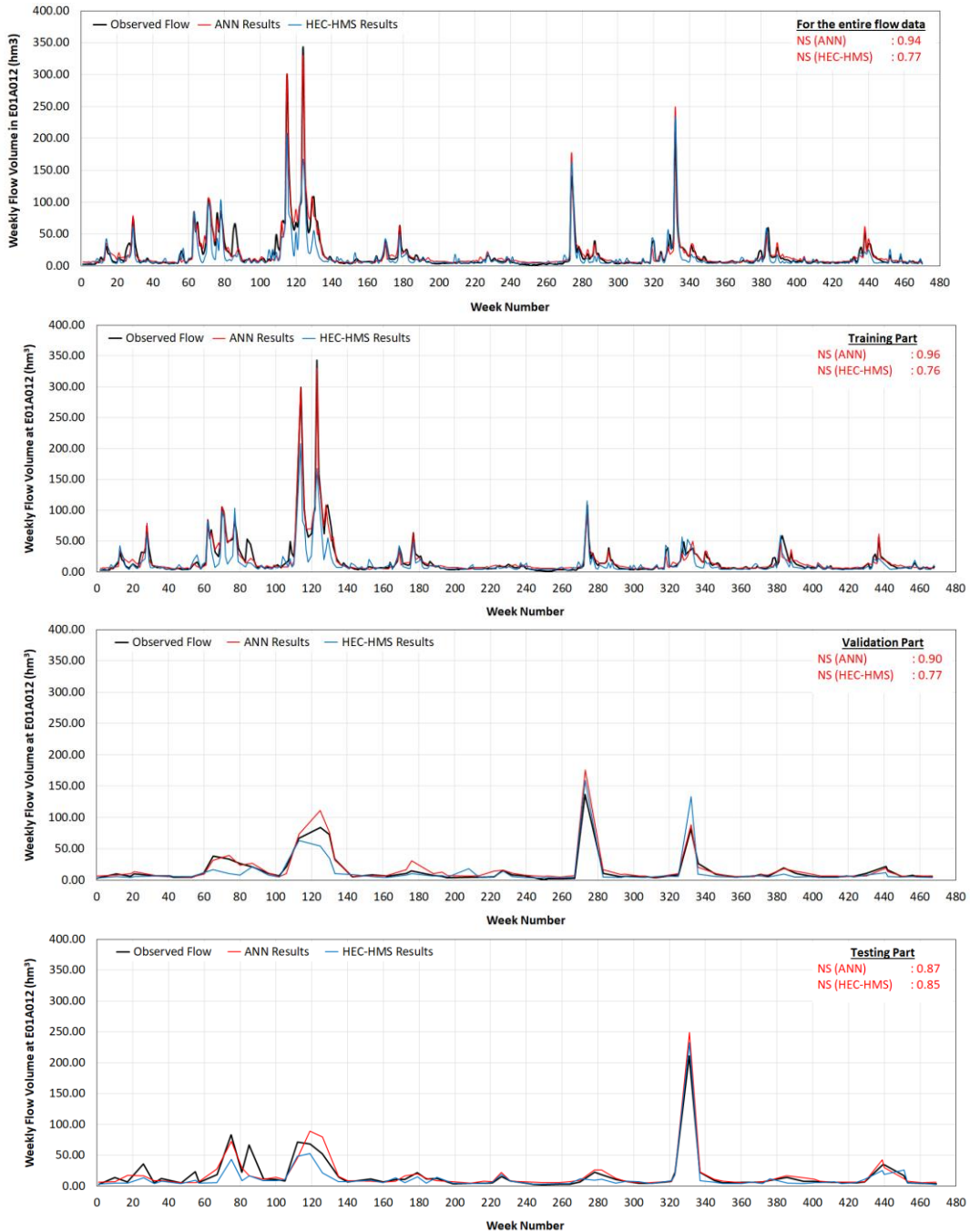


Figure 6: Comparison of the calculated flow values at site of E01A012

As can be seen from Figure 6, for the entire flow data, the calculated NS values are 0.94 and 0.77 for ANN and HEC-HMS, respectively. When data partitions are considered, the NS values for ANN and HEC-HMS models are obtained as 0.96 and 0.76 for training, 0.90 and 0.77 for validation, and 0.87 and 0.85 for testing, respectively. As can be seen both approaches generate acceptable streamflow estimations ($NS > 0.7$), but ANN performs better for this catchment. The reason of this outcome may be intensive data requirement of HEC-HMS and assumptions made in providing these requirements.

5 Conclusions

In this study, an ANN based solution approach is developed for predicting the weekly flows of Ergene River. Various precipitation – flow relationships have been established and tested in the proposed approach in order to establish the best network structure. The developed network is then trained by using the back propagation algorithm to determine the associated network weights between input and output data. During the training process, trend of the training and validation errors are continuously evaluated to determine if over-fitting problem occurs and training is stopped before over-fitting starts. The performance of the developed ANN based solution approach is also evaluated by comparing the results with those of HEC-HMS model. Results indicated that the developed ANN based solution approach resulted in a better prediction performance in training, validation, and testing stages than HEC-HMS model at the selected catchment.

Acknowledgement

This study is funded by The Scientific and Technological Research Council of Turkey (TUBITAK) under Project Number 115Y064.

References

- Ayvaz, M.T., Popovici, A. (2013). Predicting seepage flow through an earthfill dam - foundation system using an artificial neural network based solution approach: Application to Motru dam site (Romania), *6th International Perspective on Water Resour. & Env. Conf. (IPWE 2013)*, January 7-9, Izmir, Turkey.
- Bhattacharjya, R.K., Datta, B. (2005). Optimal management of coastal aquifers using linked simulation optimization approach, *Water Resources Management*, Vol. 19, pp. 295–320.
- Coppola, E., Szidarovszky, F., Poulton, M., Charles, E. (2003). Artificial neural network approach for predicting transient water levels in a multilayered groundwater system under variable state, pumping, and climate conditions. *Journal of Hydrologic Engineering, ASCE*, Vol. 8, pp. 348-360.
- Jayawardena, A.W., Fernando, T.M.K.G. (2001). River flow prediction: an artificial neural network approach, *Regional Management of Water Resources (Proceedings of a symposium held during the Sixth IAHS Scientific Assembly at Maastricht)*, The Netherlands. IAHS Publ. no. 268.
- Rezaei, M., Motlaq, A.A.A., Mahmoudi, A.R., Mousavi, S.H. (2015). River flow forecasting using artificial neural network, *Ciência e Natura, Santa Maria*, Vol. 37, pp. 207–215.