Union of Mathematicians of Macedonia - ARMAGANKA

IX SEMINAR OF DIFFERENTIAL EQUATIONS AND ANALYSIS

and

1st CONGRESS OF DIFFERENTIAL EQUATIONS, MATHEMATICAL ANALYSIS AND APPLICATIONS

CODEMA 2020

Proceedings of the CODEMA 2020
Зборник на трудови од CODEMA 2020

Skopje, 2021

FLOOD FORECASTING USING ARTIFICIAL NEURAL NETWORKS

ISBN 978-608-4904-09-0 UDC:004.89.032.26:519.248]:556.166.06(497.7)

Drenushe Fidani, Violeta Gjeshovska and Silvana Petrusheva

Abstract. Floods, as natural disasters, cause huge material damages and often result in loss of human lives. Their early prediction is necessary in order to take appropriate actions to reduce economic losses and risks for people. Albeit there is no universal method for flood modeling, significant advances in the technology of flood modeling techniques provide opportunities for flood prediction. Their modeling usually requires a large amount of data. In cases where only a specific part of the river basin is explored for more accurate modeling, the time and the effort to implement such complicated models is not justified. Therefore, the use of intelligent techniques such as Artificial Neural Networks (ANN) can be a practical alternative. The purpose of the investigations presented in this paper has been to make flood forecast for part of the Polog region using ANN. The forecast has been based on a model developed for this purpose. Modeling has been performed by use of four artificial neural networks in time series: Support Vector Machine (SVM), Radial Basis Function Neuron Network (RBFNN), General Regression Neural Network (GRNN) and Multilayer Perception (MP). Data on maximum annual flows of Vardar river, recorded at the Radusha measuring station throughout a period of 58 years, have been used as an input for the models and the output of the ANN is the maximum annual flow forecast for a 5-year (1951-2008) period. The results presented show that the ANN method, in this case the GRNN, can be useful and can provide sufficient accuracy in solving problems related to hydrological extremes.

1. INTRODUCTION

Water is vital for life, health and safety on Earth. There is no life without water. Since water resources are unevenly distributed in time and space, they need to be managed in order to avoid occurrence of floods and droughts. Floods are hydrological extremes known as natural hazards. Often, they can cause huge material damages including loss of human lives. Taking the dramatic climate changes into account, almost every country must be concerned with floods. As it is the case with many countries, floods are one of the most common natural phenomena in the Republic of North Macedonia, as well. The country is exposed to two types of floods: river - regional and flash- local sudden floods [1].

2010 *Mathematics Subject Classification.* Primary: 92B20, Secondary: 68T05, 62P30.

Key words and phrases. Artificial neural networks, maximum annual flow, time series, general regression neural network, floods.

Technically and financially, it is impossible to prevent all causes of flooding. However, early forecasting can provide good preparedness from the aspect of minimizing damages and reducing the consequences to nil.

Albeit there is no universal method for flood modeling, significant advances in the technology of flood modeling techniques provide opportunities for flood prediction. Their modeling usually requires a large amount of data. In cases where only a specific part of the river basin is explored for more accurate modeling, the time and the effort to implement such complicated models is not justified. Therefore, the use of intelligent techniques such as ANN can be a practical alternative.

Due to the above, the purpose of the investigations presented in this paper has been to predict the occurrence of hydrological extremes - floods based on the Maximum Annual Flow (MAF) of part of the Vardar river basin, the biggest river in North Macedonia. Actually, this has been done as part of doctoral thesis analyses performed for the Vardar river. Using four different ANNs, namely Support Vector Machine (SVM), Radial Basis Function Neuron Network (RBFNN), General Regression Neural Network (GRNN) and Multilayer Perception (MP) forecasting of the Vardar river flow has been done for four hydrological stations (Radusha, Gevgelija, Skopje, Jegunovce) for a 5 years' period of time. The forecasting has been done using the DTREG software for time series in combination with ANN. Presented in this paper is only the model of one station established by applying the General Regression Neural Network. For this model, data on flows of Vardar river measured at hydrological stations in the village of Radusha, Polog region, over a period of 58 years ($1951-2008$), have been used. MAF has been used as an alternative to detect changes in order to identify possible floods early enough for the purpose of responding appropriately.

The study has been based on recent scientific studies of the implementation of ANN in hydrology. The use of ANN models is becoming increasingly common in hydrological analyses and solving problems with water resources [2]. This is mainly because of the ability of ANN to model both linear and nonlinear systems without the need to make assumptions like those in implicit traditional statistical approaches [3].

The ANN technology is an alternative software approach inspired by studies of the brain and nervous system. Like the human brain, ANN behavior demonstrates the ability to learn and generalize from training data [2] and is a flexible structure capable of making nonlinear mapping between input and output layers [4]. ANN has already been successfully applied in some hydrological problems such as: rainfall forecasting, flood disaster prediction, modeling of engineering variables for water resources, river sediments and flow, prediction of river water quality, prediction of river flow, etc. Numerous studies show that ANN can offer a promising alternative to hydrological river flow forecasting and flood forecasting: Elsafi has used ANN to predict flood levels along the Nile river. According to him, this method is advantageous because only one variable can be used as a predictor, while other models require several variables to produce accurate predictions [5]. To predict floods in Indonesia, an ANN has been used by Sanubari et al. In this case, the Radial Basis Function neural network has been used. It is a network whose architecture consists of three layers, namely, an input layer, a hidden layer and an output. The analysis has been based on the water level from the 2015 rainfall data recorded at Dayeuhkolot, with results for the mean absolute percentage error-MAPE in the process of training and testing amounting to $4,97\%$ and $29,1\%$ for the rainfall and $0,047\%$ and $1,05\%$ for the water level [6]. In India, by selection of two different networks, namely, the feed forward network and the recurrent neural network, ANNs have been used to predict monthly river flows. The recurrent neural network has given better results and has therefore been recommended as a tool for predicting river flows [7].

Although the application of ANN is well developed, predicting time series events still remains the most challenging task for many engineers and scientists, "forcing engineers to constantly try to optimize existing solutions in order to obtain more accurate results" [8]. Hrnjica and Bonaci, in their paper on Vrana lake, located on the island of Cres in Croatia, present results from models for predicting from extending time series. In the paper, based on monthly measurements of the lake level during the last 38 years, ANN has been used to predict the levels for 6 and 12 months. Two types of ANN have been used: the Long-Short Term Memory (LSTM) recurrent neural network (RNN) and the Feed Forward Neural Network (FFNN). The investigations presented in this paper have confirmed the possibility and efficiency of ANN in forecasting hydrological phenomena [8], etc.

Annual maximum flow modeling is a key tool for early warning of flood hazards [5]. This has been proved by the following studies: In a study performed by Singo et al., MAF data from 8 stations involving hydrological data recorded in the course of 50 years were used to analyze the flood frequencies in the river basin. To rule out the likelihood of flooding, frequency distributions have been tested and have best described the past characteristics and magnitude of such floods [9]. Similar research has been done by Seyam and Othman. They have conducted a long-term analysis of variations in the annual river flow regime over a period of 50 years. The purpose of their analysis has been to identify long-term variations in the annual flow regime of the Selangor river, which is one of the major rivers in Malaysia, over a 50-year period [10]. From the given examples of scientific research, it can be concluded that ANNs that use annual maximum flow can be a very useful tool that can be used with satisfactory accuracy for certain forecasts in solving water resources problems.

2. METHODOLOGY

Artificial Neural Network. An artificial neural network (ANN) is a flexible mathematical tool, inspired by the biological neural networks of human brain.

As in the human brain, in form of signals neurons receive external information, in the same way artificial neural networks receive external data or input. Consistently the neurons are arranged in a layer, with the output of one layer serving as the input to the next layer and possibly other layers. Different layers may perform different transformations on their inputs. From here, neural networks consist of input and output layers, and in most case a hidden layer [11].

A single layer neural network is called a Perceptron. It gives a single output as shown in Figure 1.

In the Figure 1, $x_0, x_1, x_2, x_3, ..., x_n$ represents various inputs, independent variables, to the network. Each of these inputs is multiplied by a connection weight or synapse. The weights are represented as $w_0, w_1, w_2, w_3, ..., w_n$. Weight shows the strength of a particular node. In the simplest case, these products are summed, fed to a transfer activation function to generate a result, and this result is sent as output [12].

Mathematically it can be written as

$$x_1 w_1 + x_2\, w_2 + x_3\, w_3 + x_4 w_4 + ... + x_n w_n = \sum_{i=1}^{n} x_i w_i \ .$$

Activation function which is applied $f\left(\sum_{i=1}^{n} x_i w_i\right)$.

Activation function decides whether a neuron should be activated or not by calculating the weighted sum. The motive is to introduce non-linearity into the output of a neuron. Neural Network is considered Universal Function Approximators which means they can learn and compute any function at all [12]. Due to this feature, they are used to identify complex nonlinear relationships between input and output data sets.

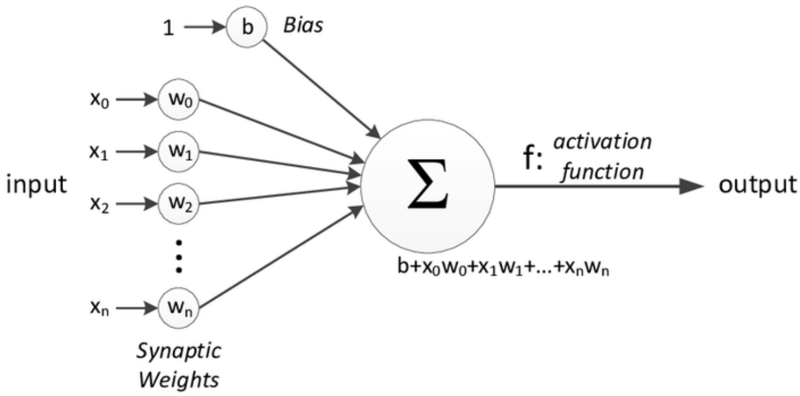

Figure 1. A single layer neural network: Perceptron
(Source [12])

There are many types of neural network, each with their own specific architecture and levels of complexity. In Figure 2 is presented a typical multilayer artificial neural network showing the input layer for ten different inputs, the hidden layer, and the output layer having three outputs. Generally, the neurons in the input layer receive an input from the external environment and without any transformations upon the inputs they send their weighted values to the neurons in the hidden layer. The neurons of the hidden layer receive the transferred weighted inputs from the input, perform the needed transformations on it, and pass the output to the next hidden layer or the output layer. The output layer consists of neurons that receive the hidden layer output and send it to the user [11].

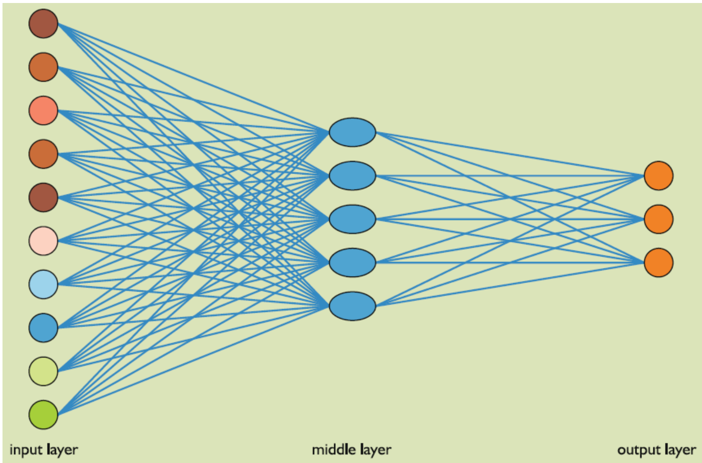

Figure 2. A typical multilayer artificial neural network with input layer, hidden layer and output layer (Source [11])

GRNN. GRNN networks have four layers input layer, pattern layer, summation layer and output layer. They do not require an iterative training procedure. Categorized by a layer that feeds back upon itself using adaptable weights, even with a constant input, they do not necessarily settle to a constant output. They can exhibit limit cycles and even chaotic behavior [13]. As schematically given in Figure 3, the input layer is connected to the pattern layer where ach neuron in the pattern layer represents a training pattern. The pattern layer performs a nonlinear transformation on the input data [14]. There are only two neurons in the summation layer. One neuron is the denominator summation unit, the other is the numerator summation unit. The denominator summation unit (in pink) figures out the weight values coming from each of the hidden neurons while the numerator summation unit (in green) figures out the weight values multiplied by the actual target value for each hidden neuron [11]. The neuron in the output layer divides the value accumulated in the numerator summation unit by the value in the denominator summation unit to yield the predicted result.

GRNN-Performance Measures: Some important performance measures for the NN model are: Mean Square Error (MSE), Normalized Mean Square Error (NMS) and correlation coefficient (r)

Mean Square Error can be determined by the following equation

$$MSE = \frac{\sum_{j=0}^{P} \sum_{i=0}^{N} (d_{ij} - y_{ij}{}^2)}{N \cdot P},$$

where (P) is number of outputs, (N) is number of exemplars in the data set, (y_{ij}) is network output for exemplar (i) at processing elements (PE_j), (d_{ij}) is desired output for exemplar (i) at (PE_j).

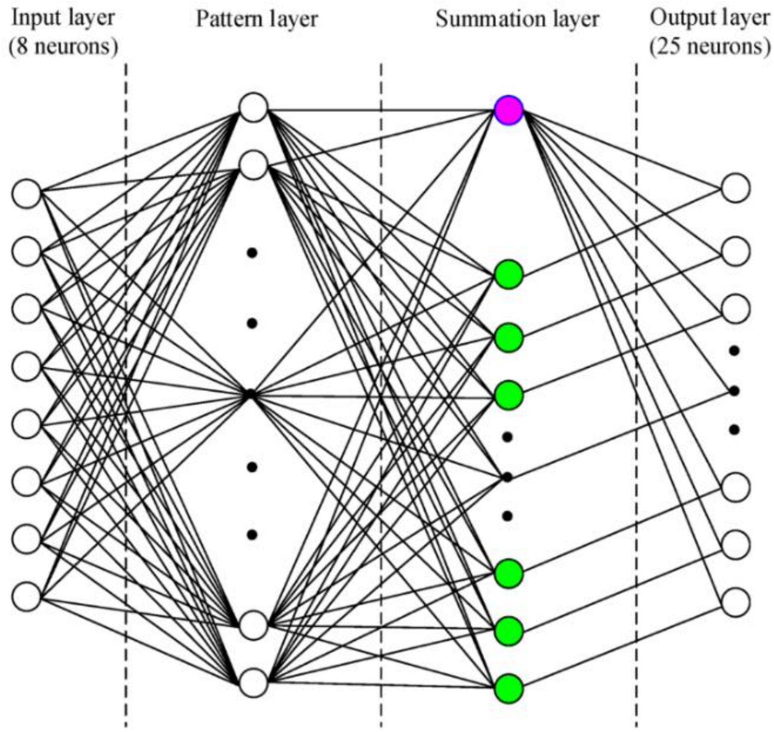

Figure 3. General regression neural network-GRNN architecture
(Source [14])

Normalized Mean Square Error can be determined by the following equation

$$NMSE = \frac{P \cdot N \cdot MSE}{\sum_{j=0}^{P} \frac{N \sum_{i=0}^{N} d_{ij}^2 - \left(\sum_{i=0}^{N} d_{ij}\right)^2}{N}},$$

where (P) is number of output processing elements, (N) is number of exemplars in the data set, (MSE) is mean square error, (d_{ij}) = desired output for exemplar (i) at processing element (j).

The equation for determining the correlation coefficient is

$$r = \frac{\frac{\sum_i (x_i - \bar{x})(d_i - \bar{d})}{N}}{\sqrt{\frac{\sum_i (d_i - \bar{d})^2}{N}} \cdot \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N}}},$$

where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ and $\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i$.

Time Series. Forecasting presents the transformation of information across time. The time series is a chronological sequence of excitations for a particular variable, usually, observed in regular intervals days, periods, months, years. In the time series with regular pattern, a value of the series is a function of the previous values. If (Y) is the value we are trying to model and predict, and (Y_t) is the value of (Y) at time t, then the goal is to create a model of the form

$$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, ..., Y_{t-n}) + et$$

where (Y_{t-1}) is the value of (Y) for the previous observation, (Y_{t-2}) is the value two observations ago, etc., and *et* represents noise that does not follow a predictable pattern [11]. Time series *forecasting* is the use of a model to predict future values based on previously observed values. The purpose for constructing a time series model is to create a model such that the error between the predicted value of the target variable and the real value is as small as possible.

The ANN approach does provide a viable and effective time series approach for developing input- output simulation and forecasting models. A proper design of the architecture of Artificial Neural Network (ANN) models can provide a robust tool in water resources modeling and forecasting**.**

3. STUDY AREA AND DATA SET

The Basin of River Vardar. The Vardar River is the largest river in the Republic of North Macedonia which provides 75% of the total water resources in the country [15]. Its spring is found in the Polog valley in the village of Vrutok, municipality of Gostivar. It passes through the cities of Gostivar and Tetovo, and then passes through Skopje. It flows through the central part of the country, enters Greece and finally reaches the Aegean Sea. The total annual average flow is estimated as 4289×106 m³ / year [16].

In this paper are presented the analyzes of the flow of the river Vardar measured at the measuring hydro station Radusa. For that purpose, the watershed upstream of this station is first analyzed, Figure 4. The section on the river Vardar is located in the watershed of Goren Vardar or in the lower part of the Polog valley. Dominant participation in the outflow and formation of flows in the river Vardar have surface waters coming from the northwestern massif, i.e. from the left tributaries of the river Vardar. Larger tributaries of the river Vardar in this part of the watershed are the rivers Pena, Vratnicka and Bistrica. The total area of the watershed upstream of the station in Raduse is about 1489 km^2. There are several categories of land use in the watershed, the so-called forests and semi-natural areas, artificial areas (urban areas, industrial, commercial and transport facilities, then mines, landfills and construction sites), agricultural areas, wetlands and swamps, water bodies, etc. In general, the studied region belongs to a modified type of Mediterranean climate which is a result of the influence of the continental climate from the central and eastern regions of Europe [17].

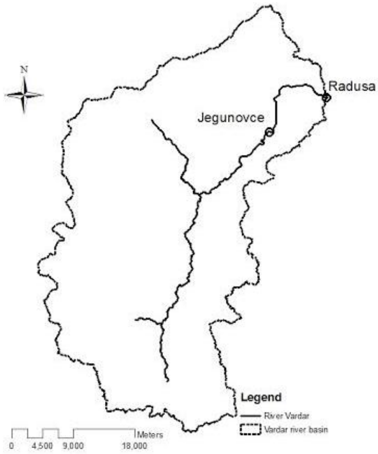

Figure 4. The watershed of the river Vardar upstream from Radusa

Data. The data used in the paper, are measured at the hydrological station Radusa, measuring station on the river Vardar for which there is a sufficiently long and quality series of data on the flow. The data were measured by the authorized institution for monitoring of hydrological stations – National Hydro meteorological service of the Republic of North Macedonia (УХМР). Data for maximum, minimum and average flow for the river Vardar in the village of Radusa for the period from 1951 to 2008 [18]. In this paper the analyzes are made with the maximum annual flows.

Using ANN, with the help of forecasting modeling software DTREG [19] a forecast of maximum annual flow for 5 years was obtained. The data were analyzed with several types of neural networks, namely Support Vector Machine (SVM), Radial Basis Function Neuron Network (RBFNN), General Regression Neural Network (GRNN) and Multilayer Perception (MP). To improve the accuracy of the model, the data can be normalized (the values of all data may be mapped at some intervals, usually [0,1]). The most accurate results were obtained with the General Regression Neural Network (GRNN). In this case, the data have been logarithmized, Figure 5.

AA1 fx Ln(max VQ)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	year	1	Ln(1)	2	Ln(2)	3	Ln(3)	4	Ln(4)	5	Ln(5)	6	Ln(6)
2	1951	40.5	3.701302	31	3.433987	49	3.89182	47.4	3.858622	94	4.543295	68.3	4.22391
3	1952	45	3.806662	32.3	3.475067	78.4	4.361824	78.4	4.361824	28	3.332205	19.1	2.949688
4	1953	38.2	3.642836	62	4.127134	23	3.135494	74.7	4.31348	58.4	4.067316	72.9	4.289089
5	1954	9.6	2.261763	15.3	2.727853	102	4.624973	92	4.521789	100	4.60517	51.4	3.939638
6	1955	51.4	3.939638	70.1	4.249923	66.5	4.197202	51.4	3.939638	49.8	3.908015	33	3.496508
7	1956	31	3.433987	30.4	3.414443	59.3	4.082609	134	4.89784	72.9	4.289089	56.6	4.036009

Figure 5. Logarithmized data set

4. RESULTS

The data have been analyzed by use of several types of neural networks: Support Vector Machine (SVM), Radial Basis Function Neuron Network (RBFNN), General Regression Neural Network (GRNN) and Multilayer Perception (MP), using DTREG software [19]. The most accurate results have been obtained with the General Regression Neural Network (GRNN). An optimal solution has been obtained. The accuracy of the model, represented by the standard accuracy assessors, is: the Mean Absolute Percentage Error (MAPE) is $3,25\%$, while the Coefficient of Determination (R^2), which expresses the general convenience of the model, is $85,153\%$. The correlation coefficient is $0,962$, Table 1. The forecasting has been done for a 5-year period, as given in Figure 6.

Table 1. Validation Data

Validation Data	
CV -Coefficient of variation	0.032869
NMSE -Normalized mean square error	0.148471
Correlation between actual and predicted	0.962000
Maximum error	0.1682587
RMSE-Root Mean Squared Error	0.1396308
MSE- Mean Squared Error	0.0194968
R2- Coefficient of determination	0.85153 (85.153%)
MAE - Mean Absolute Error	0.1370405
MAPE- Mean Absolute Percentage Error	3.2503804

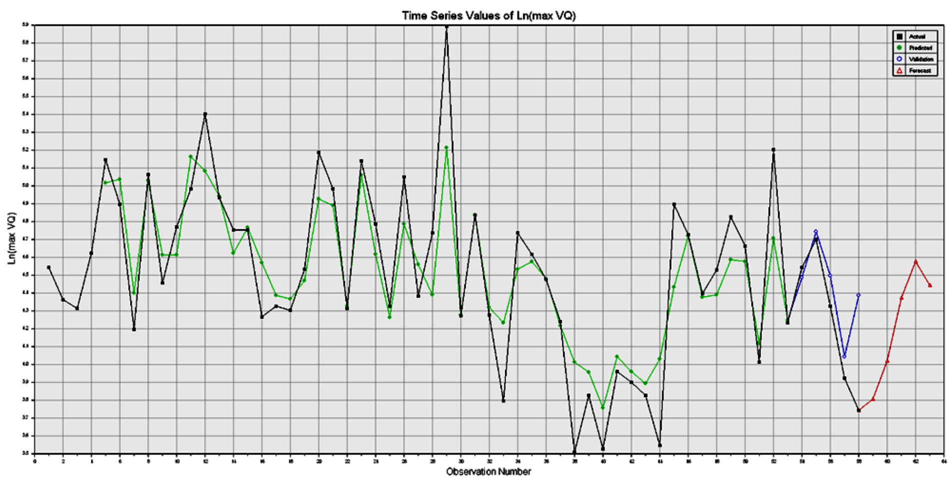

Figure 6. Time series of the maximum annual flow with a forecast for a 5-year period.

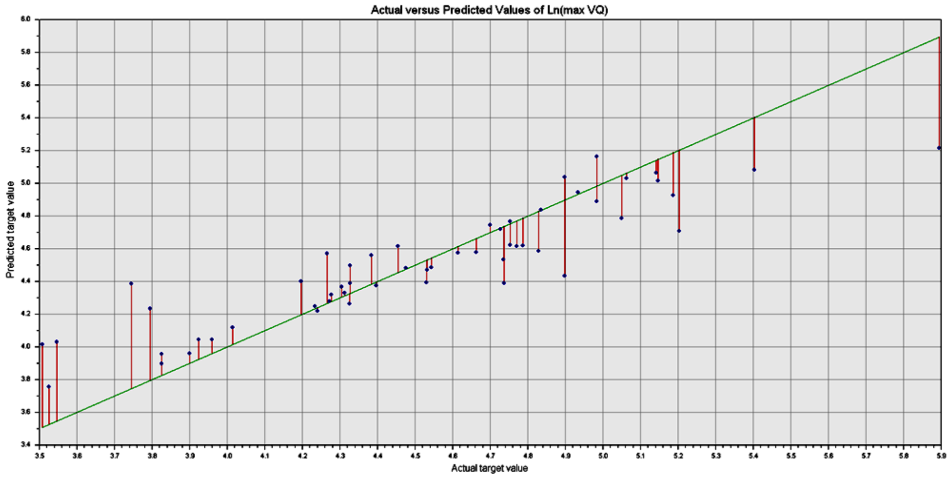

Figure 7. Actual target values vs Predicted target values

5. CONCLUSION

Artificial neural networks have shown a good ability to model a hydrological process. With artificial neural networks, using DTREG software for time series, MAF has been forecasted for 5 years. The results that have been obtained in this study, namely, the mean absolute percentage error is 3,25%, the coefficient of determination R^2, which expresses the general suitability of the model is 85,153% and the correlation coefficient is 0,962 prove that the General Regression Neural Network can be used to obtain a model with a satisfactory accuracy in forecasting maximum river flows that can be used as a basis for prediction of floods. For further research, it is recommended to develop a model of a greater accuracy using data on maximum monthly or daily flows.

References

[1] N. Dragovic, R. Ristic, H. Pulsel, B. Wolfschner, *Natural resources management in SEE: forests, soils and waters*, Deutsche Gesellschaft für Internationale, 2018.

[2] S. Riad, J. Mania, L. Bouchaou, Y. Najjar, *Rainfall-runoff model usingan artificial neural network approach*, Mathematical and Computer Modelling, 40(7-8), (2004), 839-846.

[3] I. Aichouri, A. Hani, N. Bougherira, L. Djabri, H. Chaffai, S. Lallahem, *River Flow Model Using Artificial Neural Networks*, Energy Procedia, 74, (2015), 1007-1014.

[4] M. Zakermoshfegh, M. Ghodsian, S.A.A. Salehi Neishabouri, M. Shakiba, *River Flow Forecasting Using Neural Networks and Auto-Calibrated NAM Model with Shuffled Complex Evolution*, Journal of Applied Sciences, 8: (2008), 1487-1494.

[5] S. Elsafi, *Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan*, Alexandria Engineering Journal, 53(3), (2014), 655-662.

[6] A. R. Sanubari, P. D. Kusuma, C. Setianingsih, *Flood Modelling and Prediction Using Artificial Neural Network,* IEEE International Conference on Internet of Things and Intelligence System (IOTAIS), (2018), 227-233.

[7] K. D. Nagesh, R. K. Srinivasa, T. Sathish, *River Flow Forecasting using Recurrent Neural Networks,* Water Resources Management, 18(2), (2004), 143-161.

[8] B. Hrnjica, O. Bonacci, *Lake Level Prediction using Feed Forward and Recurrent Neural Networks,* Water Resources Management, 33(7), (2019), 2471-2484

[9] L. Singo, P. Kundu, J. Odiyo, F. Mathivha, T. R. Nkuna, *Flood Frequency Analysis of Annual Maximum Stream Flows for Luvuvhu River Catchment,* Limpopo Province, South Africa, 2012.

[10] M. Seyam, F. Othman, *Long-term variation analysis of a tropical river's annual streamflow regime over a 50-year period,* Theoretical and Applied Climatology, 121(1-2), (2014), 71-85.

[11] D. Loucks, E. van Beek, *Water Resource Systems Planning And Management*, Gewerbestrasse , Springer International Publishing AG, Switzerland, 2017.

[12] N. Singh Chauhan, *Introduction to Artificial Neural Networks (ANN),* Medium, 2019.

[13] R. Deshmukh, A. Ghatol, *Short term flood forecasting using General Recurrent neural network modeling a comparative study,* International Journal of Computer Applications, 8(12), (2010), 5-9.

[14] L. Li, T. Xu, Y. Chen, *Improved Urban Flooding Mapping from Remote Sensing Images Using Generalized Regression Neural Network-Based Super-Resolution Algorithm,* Remote Sensing, 8(8), (2016), 625-636.

[15] Wfd-twinning.info. (2019). *Draft Vardar River Basin Management Plan* (*VRBMP),*2019. [online]

[16] S. Popov, T. Stafilov, R. Šajn, C. Tănăselia, K. Bačeva, Applying of Factor Analyses for Determination of Trace Elements Distribution in Water from

River Vardar and Its Tributaries, Macedonia/Greece, The Scientific World Journal, (2014), 1-11

[17] *Technical documentation on Vardar River Regulation in Polog region, Book 1-Hydrology,* Faculty of Civil Engineering, Skopje, 2016.

[18] *The Study on Integrated Water Resources Development and Management Master Plan in the Republic of Macedonia- Final Report,* NIPPON KOEI CO., LTD. KRI INTERNATIONAL CORPORATION, Japan International Cooperation Agency, 1999.

[19] P. Sherrod, *DTREG Predictive Modeling Software and Tutorial*, 2014, (www.dtreg.com)

Department of Architecture, International Vision University, Gostivar, Macedonia
E-mail address: drenushefidani@gmail.com

Faculty of Civil Engineering, Ss. Cyril and Methodius University, Skopje, Macedonia
E-mail address: violetag@gf.ukim.edu.mk

Faculty of Civil Engineering, Ss. Cyril and Methodius University, Skopje, Macedonia
E-mail address: silvana@gf.ukim.edu.mk

Publisher
Union od Mathematicians of Macedonia – ARMAGANKA

Editor in chief
Prof. d-r Aleksa Malcheski

CIP - Каталогизација во публикација
Национална и универзитетска библиотека "Св. Климент Охридски", Скопје

51(082)

PROCEEDINGS of CODEMA 2020 = Зборник на трудови од CODEMA 2020. -
Skopje : Armaganka, 2021. - 375 стр. : илустр. ; 25 см

Текст на мак. и англ. јазик. - Фусноти кон текстот. - Библиографија кон трудовите

ISBN 978-608-4904-09-0

а) Математика -- Зборници

COBISS.MK-ID 53570309