International Journal of Engineering and Geosciences

https://dergipark.org.tr/en/pub/ijeg

e-ISSN 2548-0960

Modeling of annual maximum flows with geographic data components and artificial neural networks

Esra Aslı Çubukçu*1, Vahdettin Demir 1, Mehmet Faik Sevimli 1

[1]KTO Karatay University, Faculty of Engineering and Natural Sciences, Civil Engineering Department, Türkiye

Keywords
Modeling
Flood
Artificial Neural Networks
Annual Maximum Flow
Geographical Information Systems

Research Article
DOI: 10.26833/ijeg.1125412

Received: 04.06.2022
Accepted: 07.10.2022
Published: 19.10.2022

Abstract

The flow rate at which the instantaneous maximum flow is recorded throughout the year is called the Annual Maximum Flow (AMF). These flow rates often cause disasters such as floods. Snow melts and extreme precipitation associated with temperature fluctuations are the two most important factors that occurred flooding. The deluge that follows kills people and destroys property in communities and agricultural lands. As a result, it's critical to predict the flow that causes flooding and take appropriate precautions to limit the damage. The prediction of the probability of a flood event in advance is very important for the safety of life and property of large masses and agricultural lands. Early warning systems, disaster management plans and minimizing these losses are among the important goals of the country's administration. This study was used in five Current Observation Stations (COS) located in Yeşilırmak Basin in Turkey. By using 8 input data including geographical location, altitude and area information of these stations, AMF data were tried to be estimated for each COS. A total of 240 input data was used in the study. The data period covers the years 1964-2012. Unfortunately, AMF values cannot be monitored for all 5 stations used after 2012. Therefore, the data period was stopped in 2012. In this study, Multilayer Artificial Neural Networks (MANN), Generalized Artificial Neural Networks (GANN), Radial Based Artificial Neural Networks (RBANN) and Multiple Linear Regulation (MLR) methods were used. Input data sets were made into 4 packets and these packages were used respectively in both training and testing stages. In these packages, the AMF data measured for the 5 stations mentioned above between 1965 and 2012 were divided into 4 and used by creating 25% (test) and 75% (training) packages. Root Means Square Error (RMSE), Mean Absolute Error (MAE) and correlation coefficient (R) were used as the comparison criteria. The results are as follow; MANN (RMSE = 119.118, MAE = 93.213, R = 0.808), and RBANN (RMSE = 111.559, MAE = 81.114, R = 0.900). These results show that AMF can be predicted with artificial intelligence techniques and can be used as an alternative method.

1. Introduction

A flood is defined as a rapid increase in the amount of water in a river or stream bed due to more than-normal rainfall or sudden snowmelt. Floods have the power to greatly damage living creatures and land around the beds according to their flow rates. The most dangerous floods are the sudden rise of the water in the side branches as a result of the sudden temperature change in the heavily sloping and impermeable soil areas and the basins with heavy snowfall after heavy rainfall. While changes occur in the water cycle of the atmosphere as a result of global warming and climate change, an increase is observed in the occurrence of heavy precipitation events and meteorological natural disasters. Warming climatic conditions increase the risks of drought and flood at different times and places [1]. Another reason is the precipitation regime of the basin. While the rate of floods seen in rivers between 1967-1987 among all hydrometeorological disasters was 33%; Between 1998 and 2008, this rate decreased to 14%. The increase in the number of dams established in the region in recent years, stream improvement works, and migrations from villages to cities have led to a decrease in floods originating from rivers in this way. However, despite this, according to Directorate General For State

* **Corresponding Author**

*(cubukcuasli@gmail.com) ORCID ID 0000-0003-4159-205X
(vahdettin.demir@karatay.edu.tr) ORCID ID 0000-0002-6590-5658
(mehmet.faik.sevimli@karatay.edu.tr) ORCID ID 0000-0002-4676-8782

Cite this article

Çubukçu, E. A., Demir, V., & Sevimli, M. F. (2023). Modeling of annual maximum flows with geographic data components and artificial neural networks. International Journal of Engineering and Geosciences, 8(2), 200-211

Hydraulic Works data, 1209 flood events occurred between 1975 and 2015, 720 deaths occurred as a result of these floods 893.933 hectares of agricultural land were flooded [2].

According to the data of the Turkish State Meteorological Service (TSMS), 109 of the 2372 floods that occurred between 1940 and 2010 occurred in the provinces where the stations selected in the study are located. It has been observed that this rate covers approximately 9.317% of all floods that occur [3]. It is very important in terms of the measures to be taken and the infrastructure systems to be established in order to remove the resulting flood from the settlements and agricultural lands with the least damage. At this point, predicting the flood helps us to minimize the unexpected loss of life and property. Predicting a possible future flood has always been a difficult issue as it depends on precipitation and snowmelt parameters due to sudden weather changes. While hydrological forecasting and flood response is a very important issue, current atmospheric forecasts often do not provide a sufficient level of accuracy for climate forecasts or hydrological forecasts [1]. This study aims to estimate this difficulty with Artificial Neural Networks (ANN) or regression methods based on the closed box model, apart from known functions, and to complete the missing data.

ANNs are computer systems that perform the learning function, which is the most basic feature of the human brain. They perform the learning process with the help of examples [4]. The learning and generalization ability of ANN presents this method as an ideal tool for solving complex problems [5].

Hundreds of studies have been carried out since the introduction of ANNs, the studies started in the first half of the 20th century and continued at a rapid pace until today. The first artificial neural network model was developed in 1943 by Doctor Warren McCulloch and mathematician Walter Pitts. McCulloch and Pitts modeled a simple neural network with electrical circuits, inspired by the human brain's ability to calculate and learn through experience. The 1970s became a turning point for ANNs, and many problems that were not previously considered possible to be solved began to be solved in this period [6]. Dibike et al. in their work in 2001, started to apply rainfall-runoff modeling, which can be used to obtain data on river systems design, operation and river flow, with closed box methods. When the results were compared, it was observed that ANN methods gave successful results [7]. Dawson and Wilby made precipitation-flow modeling and flood forecasting using ANN architecture in 2001. They demonstrated that the model and results could be improved [8]. In their 2002 study, Lim and Lye examined the flood records of the river basin in Sarawak, Malaysia, using an index-flood forecasting model based on the L-moment method. The results presented here are useful for engineering applications in Sarawak. It can also be used in other areas, provided that generally flood records are available. It can also be applied to unaffected basins in or near a defined homogeneous region in Sarawak [9]. Dahamsheh in 2008, In his study modeled the future monthly precipitation with Markov chain-added conditional ANNs by using the precipitation data of the previous days and months of Amman, Baqura and Safawi stations selected from different regions of Jordan. As a result, it has been seen that ANN and MLR models and models strengthened with synthetic series are insufficient, and by combining them with Markov chain, ANN and MLR models have achieved success in drought-precipitation forecasts. He concluded that the conditional ANN and conditional MLR models can be used in future studies of precipitation and drought or to complete missing data [10]. Hu et al., in their research in 2009, conducted studies to investigate new measures to improve the generalization performance of the precipitation flow model based on ANN and to evaluate the applicability of new measures. They chose the basin areas in China as the study area and applied them in 7 basins. As a result of this study, they recommended combining previous hydrological knowledge with a neural network learning algorithm instead of using neural networks as a closed-box model for rain conversion, providing consistent improvements in all seven watersheds. Second, changing the training target function offers variable results, despite improved model performance for high and low flows. Finally, work should continue to impose additional sensitivity constraints to reduce network sensitivity to input errors, improve neural network model performance, and especially generalization features. It was observed that the improved models performed better than the BP, and the 3 proposed models improved performance [11]. Ren et al. in their study in 2010, combined fuzzy logic and ANN methods with a hydrological model and modeled real-time flood forecasting for a basin in Liaoning province of China. By looking at the results, they showed that the classified models were more accurate than the others [12]. Ahmad et al., in their study in 2015, performed frequency analysis using the hydrological characteristics of 18 different regions of 5 different rivers in Pakistan. When looking at the results, Generalized Pareto, Generalized Logistics, and GEV were accepted as the best three distributions out of 10 probability distributions, L Moments method was found to be the most appropriate estimation method. For flood frequency analysis, they are in close agreement with observed flows, while also estimating different return times associated with given flood magnitudes [13]. Haktanır et al. in 2016 investigated regional frequency analyses with annual maximum rainfall data and L-moments method [14]. In the study conducted by Akkaya and Doğan in 2016, 2-dimensional flood modeling of Meriç and Tunca Rivers passing through Edirne city center was made and flood maps were created. With the analysis of the results obtained from the flood maps, a drainage channel that can discharge the flood flows exceeding the maximum flow rate that the Meriç River main bed can carry was designed and the downstream conditions of the channel were evaluated [15]. Çıtakoğlu et al. in 2017, performed regional frequency analysis using the L-Moments method using the AMF data of 52 current observation stations in the Black Sea. The application covers the five major basins. It has been revealed that the study area is not a homogeneous region as a whole. Considering the physiographic and hydrological characteristics of the basins and according to the H_1 criterion the study area is divided into 11 different homogeneous regions. According to Z_{DIST} conformity test results, there were 11 sub-regions. In five of them, the Generalized Normal distribution was determined as the most appropriate probability distribution. Generalized The Extreme Values, Pearson Type 3, and Generalized Logistic distributions are the most common in three, two, and one subregion, respectively. As a result of the index-flood frequency analyzes made with these distributions, the study Regional AMF with grid method in recurrence periods between 2 years and 1000 years for the area created maps [16]. Aghayev took a sample of the flood damage assessment, using 2010 flood data downstream of the Kura River. Using ArcGIS 10.2.1 software, maps of flooded areas and possible infrastructures in flooded areas were created. The study showed that combining multiple spatial data can significantly aid in flood damage estimates [17]. Oyebode and Stretch gave information about the advantages and disadvantages of ANN in hydrological modeling. They revealed that ANN is an important structure for obtaining good predictions in hydrological models [18]. Ovcharuk et al., used materials from 54 water measuring stations (WGS) to characterize the maximum flow during rain and melt-rain floods in rivers in the Crimean Mountains. A modified reduction structure of a calculation formula was used to evaluate the maximum flow of flash floods of different origin for rivers in mountainous Crimea. The main parameters of the proposed model are summarized as dependencies on the mean height of the basins and generalized in the form of a map. It is also possible to use the second variant of the proposed method, taking into account that the underlying surface factor is introduced. A comparison of the calculated maximum flow values shows good convergence with both the initial information and the largest values in the observation period [19]. Taylan et al. [20] developed hybrid wavelet-artificial intelligence models used in meteorological drought forecasting. Wavelet transform (W), adaptive neural-based fuzzy inference system (ANFIS), support vector machine (SVM) and ANN were used to develop drought forecasting models of Çanakkale province. Hezarani and colleagues created drought prediction models using ANN in 2021. ANN and SPI combination meteorological predicted drought with high accuracy. However, the combination of ANN and SDI isn't good for predicting hydrological drought [21]. Demir and Keskin, who did not have enough flow measurement in their studies, modeled the flood repetition rates of the Mert River in Samsun by using unit hydrograph methods and the flood propagation created by the flood hydrographs with certain recurrences in the study area. While applying the methods, precipitation data from 3 meteorological stations covering the basin, representing the Mert River basin, were used. They compared using various statistical distributions. The optimal distribution was determined by the Kolmogorov-Smirnov fitness test. Obtained flood flow rates were modeled with the FLO-2D program. Flood spreading areas and water heights were examined in different recurrences, and suggestions were made for the bridges on the river and the areas where there is construction [22].

Based on the surveyed literature in the Scopus database, it is worth mentioning that over 48 research articles were observed on the domain of maximum flow forecasting using ANN models. Figure 1 reports the VOSviewer algorithm on the interconnection between the keywords. It can be observed from this figure, this topic has major importance on water resources engineering domain addition, the connection of the "ANN and maximum flow forecast" keywords revealed the implication of those two words for the majority of the displayed keywords.

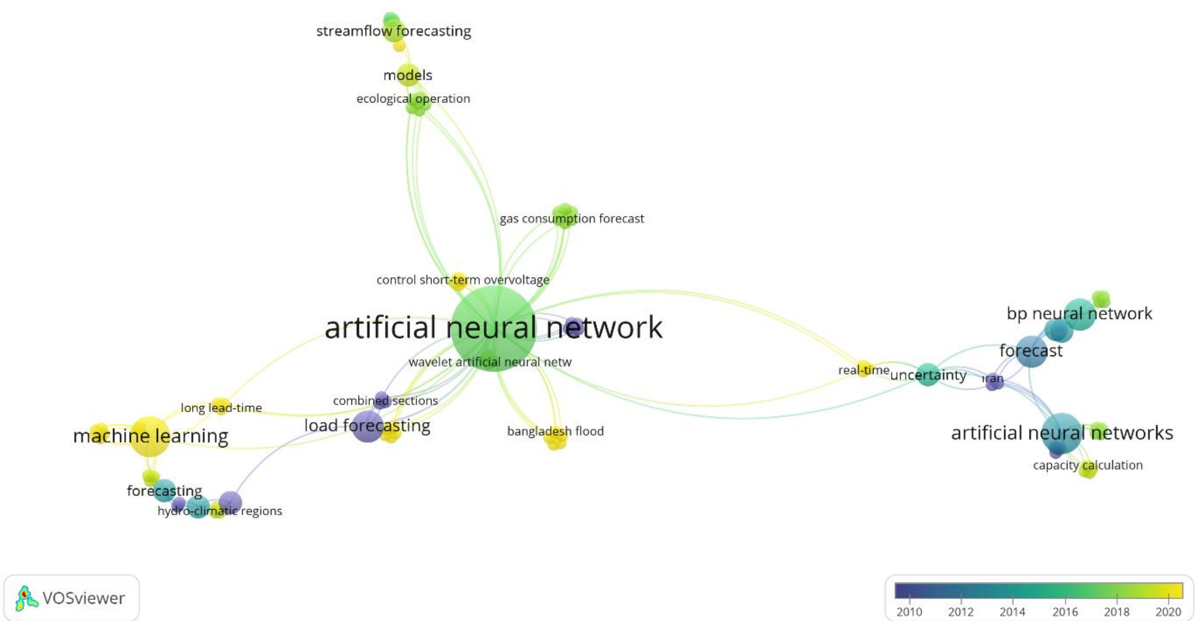

Figure 1. Relationship of ANN and maximum flow forecast keywords on Scopus database

Estimating the flood flow rate is the first stage of flood modeling studies. With this modeling, we can present the flood visually. We need data continuity for predictive modeling with ANN. The floods experienced may damage the Current Observation Stations (COS) and disrupt the data continuity. In the study, we tried to estimate the AMF data of 5 COS (1402, 1412, 1413, 1414, 1418) located in the Yeşilırmak Basin of the Black Sea Region in Turkey with their geographical components. The biggest reason for choosing this area is the flood events experienced in the past years and the continuity we can provide in the data. In this study, using the data sets created with AMF data and geographic information between 1964 and 2012, all data sets were subjected to training and testing stages, respectively. There is a problem of data deficiency in stations caused by natural disasters in the past. This study aims to complete missing data while testing past packages.

2. Method

2.1. Material

AMFs are flows recorded by COSs that show the largest instantaneously recorded flow rate in a water year, often resulting in flooding. Descriptive statistical information on the parameters examined in the study is given in Table 1.

Looking at the statistical information of the AMF data, it is seen that the maximum value is 1252 m^3/sec and the minimum value is 5.010 m^3/sec. This shows that the differences between the data can be very large and the AMF is too variable. While the average of the data is 259.571 m^3/sec, the standard deviation of the data is 246.695. The skewness coefficient is 1.542. Because this number is more than 1, the data is skewed to the right.

Right skewness makes modeling difficult, while the margin of error in modeling the result is expected to be high. Since the North (Degrees) parameter is at the same latitude for each station, its maximum value, minimum value and average are equal.

The workstations and flow networks in Figure 2 were created in the Geographical Information Systems (GIS) environment [23]. River networks and basin boundaries can be determined via ArcGIS and Basin toolbox, which are GIS software. GIS is used to solve many water engineering problems and to prepare base data such as digital elevation model data [17, 22, 24-31].

The study area is in the Yeşilırmak Basin no. 14 and the flow observation stations on the Yeşilırmak River and Çorum Çat River are shown on the map with their numbers. When the elevation map of the basin is examined, it is understood that the elevation increases from west to east [23]. Maximum current data from 1964 to 2012 were used for COS coded 1402, 1412, 1413, 1414, and 1418 as study data, but since there was no data for all stations in 1968, estimates were made by excluding the data for 1968. Since the data after 2012 were the stations that were closed or damaged by the flood, there was data discontinuity, analysis could not be made after this year because it could not be provided by all stations.

According to the information in Table 1 and Table 2, since the parameter expressed as North (Degrees) is located at 40° latitude for all COS, no correlation was found with this variable and this parameter was not used by subtracting it for all stations.

Table 1. Descriptive statistical information of the studied parameters

Variable	Unit	Number of data	Max Value	Min Value	Average	Standard deviation	Skewness coefficient
Area	(Km^2)	240	33904.00	1608.00	13251.44	12566.14	0.66
East (Degree)	°	240	37.00	35.00	36.00	0.63	0.00
East (Minutes)	'	240	30.00	6.00	15.00	10.35	0.47
East (Seconds)	''	240	45.00	3.00	27.80	19.49	-0.40
North (Degree)	°	240	40.00	40.00	40.00	0.00	0.00
North (Minute)	'	240	46.00	18.00	32.20	10.94	0.15
North (Seconds)	''	240	42.00	3.00	21.40	16.61	0.22
Height	m	240	820.00	190.00	470.20	217.09	0.32
Year	1964-2012	240	2012.00	1964.00	1988.41	14.01	-0.02
AMF	m^3/s	240	1252.00	5.01	259.57	246.69	1.54

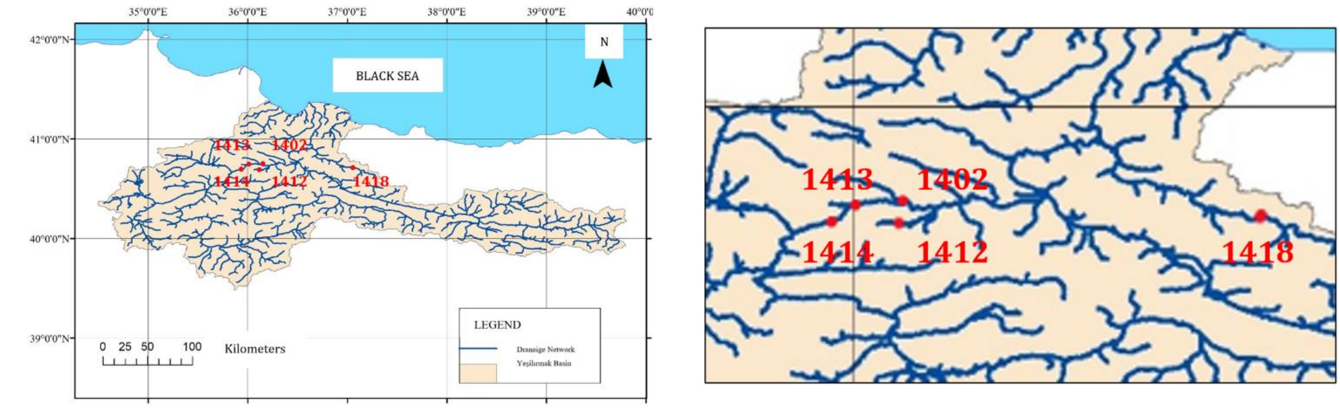

Figure 2. Yeşilırmak Basin drainage network [23]

Table 2. Summary information about the stations used in the study (*The year 1968 was not used)

Code	Province	Station Name	Latitude	Longitude	Height	Data Period
1402	Tokat	Kale	40° 46' N	36° 30' E	190 m	1964-2012*
1412	Çorum	Şeyhoğlu	40° 27' N	35° 25' E	530 m	1964-2012*
1413	Amasya	Durucasu	40° 44' N	36° 30' E	301 m	1964-2012*
1414	Tokat	Sütlüce	40° 26' N	36° 7' E	510 m	1964-2012*
1418	Tokat	Gömeleönü	40° 18' N	37° 7' E	820 m	1964-2012*

2.2. Method

ANN can be defined as machine systems that perform the learning function, which is the most basic feature of the human brain. ANNs perform learning processes with the help of examples. ANNs can be defined as the transfer of the learning mechanism in the human brain to the machine by experience. This learning mechanism, unlike what is known, brings the computational feature to the computer by using the ability to adapt to the environment, to adapt, to work according to past experiences or incomplete information in times of uncertainty.

In ANNs, various pre-processes are applied to the inputs and outputs of the network cells, and the training process of the data that is included in the ANN cycle and trained can become more efficient. This pre-process is called "Normalization". The training process of ANN data can take a very long time when applied without normalization. There are many data normalization methods in the literature. Min-Max normalization is used in this study. For the application of the ANN model, all data were first normalized between 0.2 - 0.8 using Eq. 1 [32].

$$X' = 0.6 * \frac{X_1 - X_{min}}{X_{max} - X_{min}} + 0.2 \qquad (1)$$

X' = Normalized data
X_1 = Input value
X_{min} = Minimum data
X_{max} = Maximum data

2.2.1. *Multi-Layered Artificial Neural Networks*

Multi-layer Artificial Neural Networks (MANN) consist of an input layer, one or more hidden (intermediate) layers, and an output layer where information is input. MANN has transitions between layers called forward and backward propagation. In the forward propagation phase, the output and error values of the network are calculated. In the back-propagation phase, the inter-layer link weight values are updated to minimize the calculated error value [33]. The MANN model uses the backpropagation learning algorithm, which is the generalization of the least squares algorithm in linear perception. The input layer contains neurons that receive these inputs. Therefore, the number of neurons in the input layer must be the same as the number of input values in the data set. The neurons in the input layer pass the input values directly to the hidden layer. Each neuron in the hidden layer calculates the total value by adding the threshold value to the weighted input values and processes them with an activation function and passes them to the next layer or directly to the output layer. Weights between layers are usually randomly selected at the beginning. The error value is calculated by comparing the output values of the network with the expected output values. Figure 3 shows the structure of the MANN. The multi-layer sensor model consists of an input (X_1, X_2, X_3,....., X_D), a hidden, and an output layer (Y). Each layer may also have one or more processing elements [34].

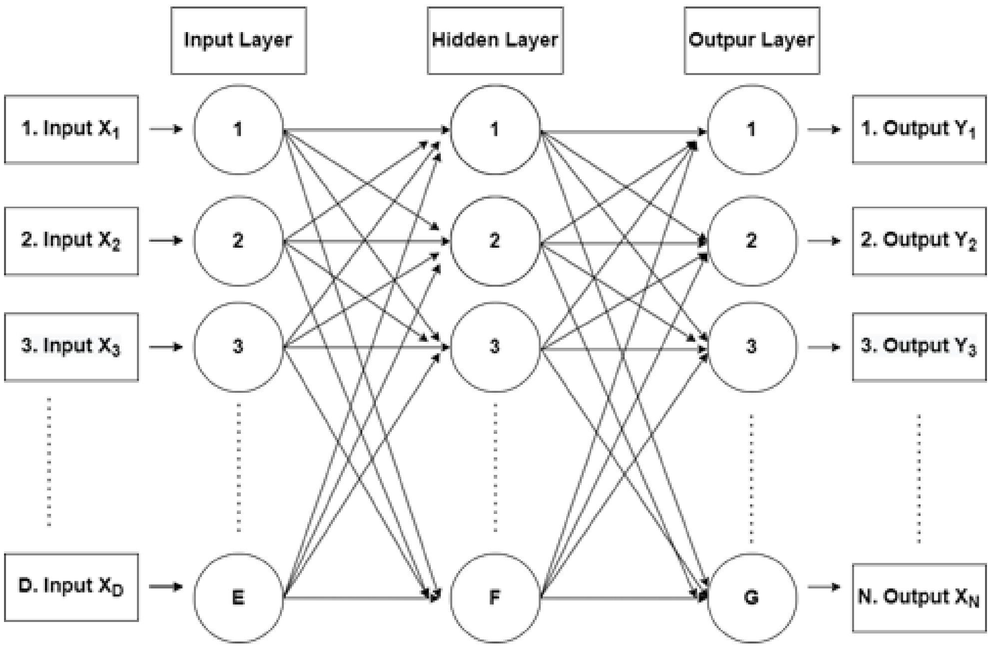

Figure 3. Structure of MANN

In Figure 2, each cell in the hidden and output layers takes the NET-weighted total outputs from the previous layer as input. NET values are found in Eq. 2.

$$NET_{xb} = \sum_{a=1}^{D} A_{ab} Ç_{xa} + \theta_b \tag{2}$$

θ_b is the bias constant (bias), A_{ab} is the set of weights between the input and hidden layers, D is the size of the input vector, $Ç_{xa}$ is the output set of the input layer for the x sample. Each cell from the second layer, the hidden layer, and the third layer, output layer, passes the NET value through a non-linear sorting function. As a result, the output f(NET) is produced in Eq. 3.

$$f(NET) = \frac{1}{1+e^{-NET}} \tag{3}$$

The total error "H_t" for sample "t" is calculated as in Eq. 4 during the training phase, depending on the difference of squares between the estimated and actual outputs.

$$H_t = \sum_{c=1}^{N} (G_{tc} + T_{tc})^2 \tag{4}$$

"G_{tc}" is the actual output value for the "c" sample, "T_{tc}" is the estimated output value for the c sample, "N" is the number of iterations. Depending on the total error, each connection weight, "A_{ab}", is renewed with the help of the equation in Eq. 5.

$$A_{ab}^{new} = A_{ab}^{old} - \left[J^T J + \mu I \right]^{-1} J^T H_t \tag{5}$$

"J" is a parameter that affects the Jacobian matrix, "J^T" is the transpose of the Jacobian matrix, "I" is the unit matrix, and "μ" is a parameter that affects the convergence rate.

In Matlab
net=train
Y1(output) = sim;

Net value obtained by running the code Y1(output) = sim; The training and testing process using the code is modeled in MANN.

2.2.2. Radial-Based Artificial Neural Networks

Radial-Based ANN (RBANN) concept was introduced into the literature in 1988 by Broomhead and Lowe [35]. ANN model and Radial-based functions have been developed by considering the effect-response states of neuron cells in human nervous system. It is possible to see the education of RBANN models as a curve fitting approach in multidimensional space [36, 37]. Thus, the educational performance of the RBANN sample turns into a problem of finding the closest result to the data in the output vector space and thus an interpolation problem [38]. RBANN structure generally consists of input layer, hidden layer and output layer similar to ANN structure. However, unlike other ANNs, the data is subjected to radial based activation functions and a nonlinear cluster analysis when passing from the input layer to the hidden layer. The structure between the hidden layer and the output layer functions as in other ANN types and the actual training takes place in this layer. In the RBANN model we used, the problem was solved with purelin function [39].

In Matlab
net=newrb
Y1(output) = sim;

Net value obtained by running the code Y1(output) = sim; The training and testing process using the code is modeled in RBANN.

2.2.3. Generalized Regression Neural Network

The generalized regression neural network proposed by Specht does not require an iterative training procedure like the back propagation method [40]. Unlike the feedback ANN models, it can perform all calculations in a single pass. It can produce results faster than others. However, in standard GRNN, where no preprocessor algorithm is used, the number of neurons in the pattern layer is equal to the amount of data in the training set. Therefore, in problems with large training data, the network structure becomes larger at the same rate as the training data set, and the number of operations increases.

GRNN is a 4-layer, feed-forward ANN model with input layer, pattern layer, collection layer and output layer. Unlike it does not require iterative. Each layer in its structure consists of different numbers of neurons [41].

net = newgrnn
Y1 (output) = sim;

Net value obtained by running the code Y1(output) = sim; The training and testing process using the code is modeled in GRNN.

2.2.4. Multi-Layered Regression

According to the equation "y" dependent, "x" is an independent variable, "ε" can be expressed as an error [42]. Equations where variable number increases. Linear regression between two variables can express with Eq. 6.

$$y = a + bx + \varepsilon \tag{6}$$

Linear regression biger then two variables can express with equation.

$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots\ldots b_n x_n + \varepsilon \tag{7}$$

These equations are called multiple linear regression equations.

3. Results

In the correlation relationship obtained by using the data of the stations selected within the Yeşilırmak basin no 14 selected as the study area; The correlation coefficient between AMF and Area was found to be 0.803, making it Minute (D) 0.677, Second (K) -0.616, Second (D) 0.603, Z (m) 0.384, Minute (K) 0.151, Degree (D) It follows with 0.147 and finally the year parameter with a correlation coefficient of 0.113 (Table 3).

Between 1964 and 2012, it was divided into packages equally and the packages it represented are given in Table 4.

Table 3. Correlation of input parameters

	Area (km^2)	Minutes'(N)	Z(m)	Second''(E)	Degree°(N)	Minute'(E)	Degree°(E)	Second''(N)	Year	AMF
Area (km^2)	1.00									
Minutes'(N)	-0.05	1.00								
Z(m)	0.45	-0.55	1.00							
Second''(E)	0.59	0.65	-0.07	1.00						
Degree°(N)	***	***	***	***	1.00					
Minute'(E)	0.94	-0.26	0.37	0.44	***	1.00				
Degree°(E)	0.08	0.69	-0.50	0.80	***	0.04	1.00			
Second''(N)	-0.89	0.42	-0.46	-0.21	***	-0.96	0.22	1.00		
Year	0.01	0.00	0.01	0.02	***	0.01	0.02	-0.00	1.00	
AMF	**0.80**	**0.15**	**0.38**	**0.60**	***	**0.67**	**0.14**	**-0.61**	**-0.11**	**1.00**

***5. Column (Degree (N)) values will not be used, since all of our stations are located at 40° latitude, no relationship could be established.

Table 4. Data packets used

Data Package	Years
A_1	2012-2001
A_2	2000-1989
A_3	1988-1977
A_4	1976-1964

Correlation coefficients were analyzed according to both sign and order of magnitude in the first place and the results were compared. Since there were very small differences in the results, the study was continued by ordering them from the largest to the smallest regardless of the sign according to the coefficient relationship, and the input order of the parameters was determined. While the correlation coefficients were found, since the Degree (N) parameter was located at 40° latitude for all stations and their data, no relationship could be established with this parameter and Degree (N) data was not used in the analysis.

The distribution of the test packages and the analysis names named according to each case are given in Table 4. The A_1 package represents the years 2012- 2001, the A_2 package represents the years 2000- 1989, the A_3 package represents the years 1988 - 1977, the A_4 package represents the years 1976 - 1964.

Table 5. Data packages used in analysis

Analysis	Training Package	Test Package
M_1	A_2+ A_3+ A_4	A_1
M_2	A_1+ A_3+ A_4	A_2
M_3	A_1+ A_2+ A_4	A_3
M_4	A_1+ A_2+ A_3	A_4

The grouping of data packages as training and test packages is given in Table 5. As seen in the Table 5, in M_1 analysis, A_1 package is used in testing, A_2+ A_3+ A_4 package is used in education. In M_2 analysis, A_2 package is used in testing, A_1+ A_3+ A_4 package is used in education. In M_3 analysis, A_3 package is used in testing, A_1+ A_2+ A_4 package is used in education. In M_4 analysis, A_4 package is used in testing, A_1+ A_2+ A_3 packages are used in training.

RMSE, MAE and R were used as comparison criteria. According to these criteria, the methods giving the best results were determined and the results were tabulated.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Z_e - Z_o)^2} \tag{8}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|Z_e - Z_o\right| \tag{9}$$

$$R = \left(\frac{N*\left(\sum Z_o * Z_e\right)-\left(\sum Z_o\right)*\left(\sum Z_e\right)}{\sqrt{\left(N*\sum Z_o^2\right)-\left(\sum Z_e\right)^2*\left(N*\sum Z_e^2\right)-\left(\sum Z_e\right)^2}}\right) \tag{10}$$

In the Equations 8-10, "Z_e" and "Z_o" show the estimated and observed AMF values and "N" represents the number of data [43]. M_1 analysis test results are shown in Table 6.

When the test results are examined, in Table 6 the MANN gives the lowest error and the highest accuracy rate in the averages of all inputs. M_2 analysis test results are shown in Table 7.

In Table 7, the averages give close results according to the methods. Individually, RBANN (8 Entries) gave the best results. M_3 analysis test results are shown in Table 8.

In Table 8, the averages give close results according to the methods. RBANN (1 Input) gave the best results locally. M_4 analysis test results are shown in Table 9.

In Table 9, the averages give close results according to the methods. RBANN (2 Inputs) gave the best results individually.

Since AMF data could not be obtained at all stations for the year 1968, entries were made in the data sets used in the analysis, excluding that year. In Table 10, the study of completing the missing data of 1968 was carried out. The estimated flow rates for each station according to the methods are as in the table. However, since we do not have data for that year, there is no situation that we can control [44].

In Table 10, MLR method makes higher estimations for all stations compared to other methods. It is observed that the MANN and GRNN methods give very close results as in the results of other packages.

Table 6. M_1 Analysis Comparison of test data

Test	Method	Input Data (M_1 analysis)								
		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	Average
RMSE	MANN	167.98	167.35	167.35	167.35	167.35	167.35	167.35	119.11	161.40
	RBANN	167.35	163.16	163.67	158.92	167.79	167.35	167.35	140.25	161.98
	GRNN	170.56	167.66	167.57	167.36	167.36	167.35	167.35	161.31	167.06
	MLR	182.56	180.60	164.27	167.23	167.35	167.35	167.35	158.53	169.40
Average		172.11	169.69	165.71	165.21	167.46	167.35	167.35	144.80	164.96
MAE	MANN	120.79	118.73	118.73	118.73	118.73	118.73	118.73	93.21	115.79
	RBANN	118.73	120.30	117.93	111.05	119.73	118.73	118.73	108.64	116.73
	GRNN	126.14	120.03	119.77	118.81	118.80	118.73	118.73	120.15	120.14
	MLR	144.62	146.54	114.11	118.40	118.73	118.73	118.73	108.50	123.54
Average		127.57	126.40	117.63	116.75	119.00	118.73	118.73	107.62	119.05
R	MANN	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.80	0.78
	RBANN	0.77	0.78	0.78	0.77	0.77	0.77	0.77	0.63	0.76
	GRNN	0.75	0.77	0.77	0.77	0.77	0.77	0.77	0.79	0.77
	MLR	0.71	0.70	0.78	0.77	0.77	0.77	0.77	0.77	0.76
Average		0.75	0.75	0.78	0.77	0.77	0.77	0.77	0.75	0.77

Table 7. M_2 Analysis Comparison of test data

Test	Method	Input Data (M_2 analysis)								
		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	Average
RMSE	MANN	131.20	164.07	130.48	130.48	130.48	130.48	130.48	170.29	139.74
	RBANN	130.48	130.48	130.48	130.48	133.41	130.28	169.72	123.33	134.83
	GRNN	130.48	130.48	130.48	130.48	130.48	130.48	130.48	174.20	135.94
	MLR	147.73	146.86	129.86	130.47	130.48	130.48	130.48	129.83	134.52
Average		134.97	142.97	130.32	130.48	131.21	130.43	140.29	149.41	136.26
MAE	MANN	85.03	111.66	81.44	81.44	81.44	81.44	81.44	124.25	91.02
	RBANN	81.44	81.44	81.44	81.44	91.15	82.61	134.36	78.38	89.03
	GRNN	81.44	81.44	81.44	81.44	81.44	81.44	81.44	116.36	85.80
	MLR	101.83	103.26	80.57	81.37	81.44	81.44	81.44	80.96	86.54
Average		87.43	94.45	81.22	81.42	83.87	81.73	94.67	99.99	88.10
R	MANN	0.85	0.78	0.85	0.85	0.85	0.85	0.85	0.82	0.84
	RBANN	0.85	0.85	0.85	0.85	0.84	0.85	0.73	0.87	0.83
	GRNN	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.77	0.84
	MLR	0.81	0.80	0.85	0.85	0.85	0.85	0.85	0.85	0.84
Average		0.84	0.8	0.85	0.85	0.84	0.85	0.82	0.83	0.84

Table 8. M_3 Analysis Comparison of test data

Test	Method	Input Data (M_3 analysis)								
		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	Average
RMSE	MANN	157.16	157.16	157.16	157.16	157.16	157.16	157.16	133.83	154.25
	RBANN	156.76	156.10	156.22	157.16	157.09	162.57	194.34	168.39	163.58
	GRNN	157.16	157.16	157.16	157.16	157.16	157.16	157.16	131.45	153.95
	MLR	172.97	171.99	157.54	157.10	157.16	157.16	157.16	150.69	160.22
Average		161.02	160.60	157.02	157.15	157.15	158.52	166.46	146.09	158.00
MAE	MANN	94.76	94.76	94.76	94.76	94.76	94.76	94.76	90.31	94.21
	RBANN	95.49	93.05	93.43	94.76	94.62	115.80	139.45	109.61	104.53
	GRNN	94.76	94.76	94.76	94.76	94.76	94.76	94.76	85.03	93.55
	MLR	117.92	112.79	95.54	94.71	94.76	94.76	94.76	89.35	99.32
Average		100.74	98.84	94.63	94.75	94.73	100.02	105.94	93.58	97.90
R	MANN	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.89	0.90
	RBANN	0.91	0.90	0.90	0.90	0.90	0.89	0.81	0.86	0.87
	GRNN	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.88	0.90
	MLR	0.88	0.87	0.90	0.90	0.90	0.90	0.90	0.90	0.89
Average		0.90	0.89	0.90	0.90	0.90	0.87	0.88	0.89	0.89

Table 9. M_4 Analysis Comparison of test data

Test	Method	Input Data (M_4 analysis)								
		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	Average
RMSE	MANN	113.92	113.92	113.92	113.92	113.92	113.92	113.92	462.57	157.50
	RBANN	112.86	111.55	112.05	113.51	112.53	112.26	113.92	163.24	118.99
	GRNN	113.92	113.92	113.92	113.92	113.92	113.92	113.92	129.81	115.90
	MLR	131.54	131.84	114.17	113.81	113.92	113.92	113.92	158.06	123.90
Average		118.06	117.81	113.52	113.79	113.57	113.50	113.92	228.42	129.07
MAE	MANN	79.70	79.70	79.70	79.70	79.70	79.70	79.70	371.90	116.22
	RBANN	80.16	81.11	78.97	82.84	80.10	80.76	79.70	133.71	87.17
	GRNN	79.70	79.70	79.70	79.70	79.70	79.70	79.70	98.56	82.06
	MLR	97.14	95.16	80.12	79.78	79.70	79.70	79.70	138.36	91.21
Average		84.17	83.92	79.62	80.51	79.80	79.96	79.70	185.63	94.17
R	MANN	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.69	0.86
	RBANN	0.89	0.90	0.89	0.89	0.89	0.89	0.89	0.84	0.89
	GRNN	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
	MLR	0.86	0.85	0.89	0.89	0.89	0.89	0.89	0.89	0.88
Average		0.88	0.88	0.89	0.89	0.89	0.89	0.89	0.83	0.88

Table 10. Estimates of 1968 AMF data

Method	Station No	Station Name	Input number	Estimated Flow Rates (m^3/s)
MANN	1402	Kale	7	619.47
	1412	Şeyhoğlu		51.42
	1413	Durucasu		326.57
	1414	Sütlüce		109.79
	1418	Gömeleönü		182.58
RBANN	1402	Kale	2	636.50
	1412	Şeyhoğlu		105.20
	1413	Durucasu		283.47
	1414	Sütlüce		109.31
	1418	Gömeleönü		155.34
GRNN	1402	Kale	7	619.47
	1412	Şeyhoğlu		51.42
	1413	Durucasu		326.57
	1414	Sütlüce		109.79
	1418	Gömeleönü		182.58
MLR	1402	Kale	8	746.45
	1412	Şeyhoğlu		178.40
	1413	Durucasu		453.55
	1414	Sütlüce		236.77
	1418	Gömeleönü		309.56

4. Discussion

A flood is defined as a rapid and uncontrolled increase in the amount of water in a river or stream bed as a result of more than normal rainfall or sudden snowmelt. Floods have the power to greatly damage the living creatures and lands around the beds according to their flow rates. The sudden rise of water in the side branches in sloping and impermeable soil areas caused by snow melts after heavy rain or sudden warming in snowy areas causes the most dangerous floods. Estimating the flood flow rate is the first stage of flood modeling studies. We need data continuity for predictive modeling with ANN. Experienced floods can damage COS and disrupt data continuity. In our study, the AMF data of 5 COSs (1402, 1412, 1413, 1414, 1418) located in the Yeşilırmak Basin 14 in the Black Sea Region were tried to be estimated with their geographical components.

5. Conclusion

In this study; using 8 input data including geographical location (north and east; degree, minute, second), altitude and area information of 5 COS (1402, 1412, 1413, 1414, 1418) located in the Yeşilırmak Basin numbered 14, each an attempt was made to estimate the AMF data for a COS. The data period covers the years 1964-2012. 48 years of data are grouped as 75% training and 25% test packages. Analyzes were made 4 times, with 1 package being tested each time. The study aims aims to complete the missing data by modeling the deficiencies of data losses due to natural disasters. MANN, GRNN, RBANN and MLR methods were used to estimate AMF.

When the results are examined;

- For M_1 analysis, the best result was MANN (8 input) (RMSE=119.118, MAE= 93.213, R=0.808),
- For M_2 analysis, the best result is RBANN (8 input) (RMSE= 123.334, MAE= 78.381, R=0.870),
- For M_3 analysis, the smallest error rates were observed as GRNN (8 input) (RMSE= 131.457, MAE= 85.033, R= 0.889), while the largest R coefficient for RBANN (1 input) (RMSE= 156.763, MAE= 95.499, R= 0.910),
- For M_4 analysis, it was observed that the best results were obtained with RBANN (2 input) (RMSE=111.559, MAE= 81.114, R=0.900).

• There are years when COS's cannot save data for various reasons. It is observed that missing data occurs in years when data cannot be recorded in this way. It has been observed that the test results give sufficient accuracy to make predictions instead of missing data.

• Errors according to the input parameters used (Table 1); While the area decreased using minute(E), second(N), second(E) entries, it increased in year, degree(E) and minute(N) parameters. It was observed that there were small differences in the results when the correlation coefficients were ordered from largest to smallest sign. Therefore, in this study, the correlation coefficients were ordered by size, and the data were entered in the order of magnitude of the correlation coefficients while performing the analysis, regardless of the sign.

• Considering the results of the analysis, it was observed that although the MLR method was better than MANN, it could not provide a better estimation than RBANN and GRNN. Therefore, it was concluded that RBANN and GRNN models are an alternative solution to MLR. Therefore, it has been observed that RBANN provides a valid accuracy rate and can be used in anerage flood modeling.

• The reason why the accuracy rate in the A_1 package is low compared to other packages; It can be said as the global warming experienced in recent years and the variation of seasons and precipitation amounts accordingly.

• High estimations from the data observed as a result of the modeling are important in terms of taking precautions. AMF flow often causes flooding at its peaks. Considering this situation, the estimation of a flow rate higher than observed is beneficial in terms of minimizing the damage caused by flooding by taking precautions.

The limitations of this study are the estimation of AMFs and the completion of missing station data using 3 different ANNs. Modeling was carried out in 4 different test packages. In future studies, the authors aim to use more up-to-date data and more stations. In addition, the authors plan to make predictions using spatial and temporal inputs by trying different methods.

Acknowledgement

The author thanks KTO Karatay University for support and thanks the anonymous reviewers for their contributions.

Author contributions

Esra Aslı Çubukçu: Methodology, Application, Writing Original draft preparation, **Vahdettin Demir:** Conceptualization, Methodology, Reviewing and Editing. **Mehmet Faik Sevimli:** Last Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Tonkaz T., Çetin M., Kızıloğlu F., & Fayrap A. (2010). Mixed Eastern Black Sea Water Basin Annual Probabılıty Analysıs of Instant Maxımum Currents. II. National Flood Symposium, 315-321.
2. Republic of Turkey Ministry of Agriculture And Forestry General Directorate of Water Management Flood Management (2017). Ministry of Forestry and Water Affairs, Ankara.
3. URL1: "Natural disasters, floods" access address: https://www.riob.org/fr/file/272232/download?token=PysGjbkb accessed 6 October 2022.
4. Öztemel E. (2012). Artificial Neural Networks, Istanbul.
5. Aydoğan B., Ayay B., & Çevik, E. (2011). Ann Current Profile Forecasting in Straits With an Example:Bosphorus. 7th Coastal Engineering Symposium, 403-409.
6. URL2: "Artificial Neural Networks" access address:https://www.kimnezamanicatetti.com/yapay-sinir-agi/, accessed May 1, 2019.
7. Dibike Y. B., & Solomatine, D. P. (2001). River Flow Forecasting Using Artificial Neural Networks. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere (B), 26*(1), 1-7.
8. Dawson, C. W. & Wilby, R. L. (2001). Hydrological Modelling Using Artificial Neural Networks. *Prog. Phys. Geogr., 25*(1), 80–108. https://doi.org/10.1177/030913330102500104.
9. Lim, Y. H., & Lye, L. M. (2003). Regional flood estimation for ungauged basins in Sarawak , Malaysia Regional flood estimation for ungauged basins in Sarawak. *Malaysia, Hydrological Sciences Journal, 48*(1), 79–94.
10. Dahamsheh, A. (2008). *Forecasting Monthly Precipitation For Arid Regions Using Conditional Artificial Neural Networks Combined With Markov Chain*. (Ph.D. Thesis, ITU Institute of Science and Technology, Turkey).
11. Hu, T. S., Lam, K. C. & Ng, S. T. A. (2009). Modified Neural Network for Improving River Flow Prediction/Un Réseau de Neurones Modifié pour Améliorer la Prévision de L'Écoulement Fluvial. *Hydrological Sciences Journal, 50*(2), 298–318.
12. Ren, M., Wang, B., Liang, Q. & Fu, G. (2010). Classified Real-Time Flood Forecasting By Coupling Fuzzy Clustering and Neural Network. *International Journal of Sediment Research 25*(2), 134-148.
13. Ahmad, I., Fawad, M. & Mahmood, I. (2015). At-Site Flood Frequency Analysis of Annual Maximum Stream Flows in Pakistan Using Robust Estimation Methods. *Polish Journal of Environmental Studies, 24*(6), 2345–2353.
14. Haktanir, T., Citakoglu, H. & Seckin, N. (2016). Regional frequency analyses of successive-duration annual maximum rainfalls by L-moments method. *Hydrological Sciences Journal,* 61(4), 647–668. https://doi.org/10.1080/02626667.2014.966722.
15. Akkaya, U. & Doğan, E. (2016). Generation of 2D Flood İnundation Maps of Meriç and Tunca Rivers

Passing Through Edirne City Center. *Geofizika 33*(1), 15-34, https://doi.org/10.15233/gfz.2016.33.7
16. Çıtakoğlu, H., Demir, V. & Haktanir, T. (2017). Regional Frequency Analysis Of Annual Flood Peaks Of Natural Streams Discharging To The Black Sea By The L-Moments Method. *Omer Halisdemir Univercity Journal of Engineering Science, 6*(2), 571–580.
17. Aghayev, A. (2018). Determining of Different Inundated Land Use in Salyan Plain During 2010 the Kura River Flood Through Gis and Remote Sensing Tools. *Journal of Engineering and Geosciences, 3*(3) 80–86. https://doi.org/10.26833/ijeg.412348.
18. Oyebode, D. & Stretch, O. (2019). Neural Network Modeling Of Hydrologicalsystems: A Review Of İmplementation Techniques. Natural Resource Modeling. 32(1), 1–14. https://doi.org/10.1002/nrm.12189
19. Ovcharuk, V., Gopchenko, E., Todorova, O., & Myrza, K. (2020). Calculating the Characteristics of Flash FLoods on Small Rivers in the Mountainous Crimea, *Geofizika 37*(1), 27-43. https://doi.org/10.15233/gfz.2020.37.3, July.
20. Taylan, E. D., Terzi, Ö. & Baykal, T. (2021). Hybrid Wavelet–Artificial İntelligence Models İn Meteorological Drought Estimation. *J. Earth Syst. Sci., 130*(38), 2021, https://doi.org/10.1007/s12040-020-01488-9.
21. Boustani Hezarani, A., Zeybekoğlu, U. & Ülke Keskin, A. (2021). Hydrological And Meteorological Drought Forecasting For The Yesilirmak River Basin, Turkey. *Sürdürülebilir Mühendislik Uygulamaları ve Teknol. Gelişmeler Derg., 4*(2), 121–135. 2021, https://doi.org/10.51764/smutgd.993792
22. Demir, V. & Ülke Keskin, A. (2022). Flood flow calculation and flood modeling in rivers that do not have enough flow measurement (Samsun, Mert River sample). *Geomatik, 7*(2), pp. 149–162, https://doi.org/10.29128/geomatik.918502
23. Çay, F. (2018). *Determination of Morphological Characteristics Of Basins in Turkey with The Help of GIS.* (Master Thesis, KTO Karatay University Institute of Science and Technology, Turkey).
24. Akgül, M. A. (2018). Sentetik Açıklıklı Radar verilerinin Taşkın Çalışmalarında Kullanılması: Berdan Ovası Taşkını. Geomatik , 3 (2) , 154-162 . https://doi.org/10.29128/geomatik.378123
25. Oğuz, E., Oğuz, K. & Öztürk, K. (2022). Düzce bölgesi taşkın duyarlılık alanlarının belirlenmesi . *Geomatik*, 7(3) , 220-234 . https://doi.org/10.29128/geomatik.972343
26. Tona, A. U. , Demir, V. , Kuşak, L. & Yakar, M. (2022). Su Kaynakları Mühendisliğinde CBS'nin Kullanımı. T*ürkiye Coğrafi Bilgi Sistemleri Dergisi*, 4 (1), 23-33. https://doi.org/10.56130/tucbis.993807
27. Yılmaz, O. S. (2023). Frekans oranı yöntemiyle coğrafi bilgi sistemi ortamında heyelan duyarlılık haritasının üretilmesi: Manisa, Demirci, Tekeler Köyü örneği. *Geomatik*, 8(1), 42-54. https://doi.org/10.29128/geomatik.1108735
28. Yılmaz, A. & Erdoğan, M. (2018). Designing high resolution countrywide DEM for Turkey . International Journal of Engineering and Geosciences, 3(3), 98-107. https://doi.org/10.26833/ijeg.384822
29. Al Kalbani, K. & Rahman, A. A. (2022). 3D city model for monitoring flash flood risks in Salalah, Oman. International Journal of Engineering and Geosciences, 7(1), 17-23. https://doi.org/10.26833/ijeg.857971
30. Yağmur, N. , Tanık, A. , Tuzcu, A. , Musaoğlu, N. , Erten, E. & Bilgilioglu, B. (2020). Opportunities provided by remote sensing data for watershed management: example of Konya Closed Basin . International Journal of Engineering and Geosciences, 5(3), 120-129. https://doi.org/10.26833/ijeg.638669
31. Yakar, M. (2009). Digital elevation model generation by robotic total station instrument. *Experimental Techniques*, *33*(2), 52-59
32. Yavuz S. & Deveci M. (2012). The Effect of Statistical Normalization Techniques on The Performance of Artificial Neural Network. *Erciyes University Journal of the Faculty of Economics and Administrative Sciences,* 167–187.
33. Arı, A., & Berberler, M. E. (2017). Prediction and Classification with Artificial Neural Networks Interface Design for Solving Problems. *Acta Infologica, 1*(2): 55–73.
34. Çubukçu, E. A., Sancıoğlu S., Demir, V. & Sevimli, M. F. (2019). Sea Water Level Estimation Using Six Different Artificial Neural Networks Training Algorithm, International Civil Engineering Architecture Conference, 716-725.
35. Broomhead, D. D. & Lowe, S. (1988). Multivariable Functional Interpolation and Adaptive Networks. *Complex Syst., 2*, 321–355.
36. Partal, T., Kahya, E. & Cığızoğlu, K. (2008). Prediction of Precipitation Data with Artificial Neural Networks and Wavelet Transformation Methods. *ITU Journal of Engineering, 7*(3), 73–85.
37. Poggio, T. & Girosi, F. (1990). Regularization Algorithms for Learning That Are Equivalent to Multilayer Networks. *Science, 247*(4945), 978–982.
38. Okkan, U. & Dalkılıç, H. Y. (2012). Monthly Runoff Model for Kemer Dam with Radial Based Artificial Neural Networks. *IMO Technical Journal,* 5957–5966.
39. Demir, V., Çubukçu, E. A. & Sevimli, M. F. (2019). Long-Term Month Temperature Forecast With Inverse Distances Weighted, Kriging And Artificial Neural Networks. CISET-2nd Cilicia International Symposium On Engineering And Technology, 10-13.
40. Sürel, A., (2006). *The Use of Generalized Regression Neural Network in Water Resources Engineering.* (Master Thesis, ITU Institute of Science and Technology, Turkey).
41. Oral, M., Kartal, S. & Özyıldırım, M. (2017). A Cluster Based Approach to Reduce Pattern Layer Size for Generalized Regression Neural Network. *Pamukkale University Journal of Engineering Sciences, 24*(5), 857–863.
42. Okkan, U. & Mollamahmutoglu, A. (2010). Daily Runoff Modelling of Yiğitler Stream by Using Artificial Neural Networks and Regression Analysis.

Dumlupınar University Journal of Science Institute, 23: 33–48.

43. Çubukçu, E. A. (2019). *Modelıng Of Annual Maxımum Flows Wıth Geographıc Data Components And Artıfıcıal Neural Networks.* (Master Thesis, KTO Karatay University Institute of Science and Technology, Turkey).

44. Zhang, Q. Yuan, Q. Zeng, C. Li, X. &Wei, Y. (2018). Missing Data Reconstruction In Remote Sensing Image With A Unified Spatial-Temporal-Spectral Deep Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens., 56*(8), 4274–4288.