AINFALL FORECASTING USING NEURAL NETWORKS IN THRACE (TURKEY)

Cercis İKİEL and Ömer ÖZYILDIRIM

Faculty of Arts and Sciences, Sakarya University, Turkey.
cikiel@sakarya.edu.tr, omero@sakarya.edu.tr

Abstract

Among the climatic elements rainfall data show the most temporal and spatial variability. Rainfall prediction is the most intensely studied phenomenon, nevertheless due to its nonlinear nature it yields low predictability ratios. Artificial neural networks are increasing in importance in rainfall forecasting in recent years. In this study rainfall data are analyzed as a time series using artificial neural networks. The data set used in this study is the daily rainfall data of Edirne, Çorlu, Tekirdağ, Florya (İstanbul) meteorological stations during the period of 1970 - 2000. The data is analyzed using an artificial neural network (ANN), trained using feed-forward back-propagation (FFBP) technique and the optimum network topology is determined. During the analysis, 4 years of monthly rainfall data are used for training, 4 years for testing and 3 years for running processes. Results of daily total values (sum of 10 days) were obtained better rather than the daily values results.

Keywords: *Thrace, Rainfall, Artificial Neural Networks (ANN)*

Introduction

In this study daily rainfall data (1970 – 2000) were used from Edirne, Çorlu, Tekirdağ and Florya (İstanbul) meteorological stations which located on Thrace Peninsula in northwest of Turkey. The research area has Black Sea Rainfall Regime and Mediterranean Rainfall Regime (İkiel, 2005). Semi humid (C_2) and Arid – Semi humid (C_1) climatic conditions are observed in the study area according to Thornthwaite climatic classification (Türkeş, 2010, Çiçek, 1995). Rainfall data show temporal and spatial variability among the stations (Ustaoglu, 2011).

The ANN approach has been successfully employed in the climatological researches. In particular, rainfall forecasting has already been analyzed. In some papers were analyzed in daily scale (Kim and Pachepsky, 2010, Kumarasiri and Sonnadara, 2008, Rajurkar et al., 2004) and other ones were analyzed in monthly and annual scale (Abbot and Marohasy, 2012, Ilunga, 2010, Wu et al., 2010).

The purpose of the present paper is the employment of ANN method, feed-forward back propagation (FFBP) to forecast daily rainfall re and to provide a best-fit prediction with the calculated data.

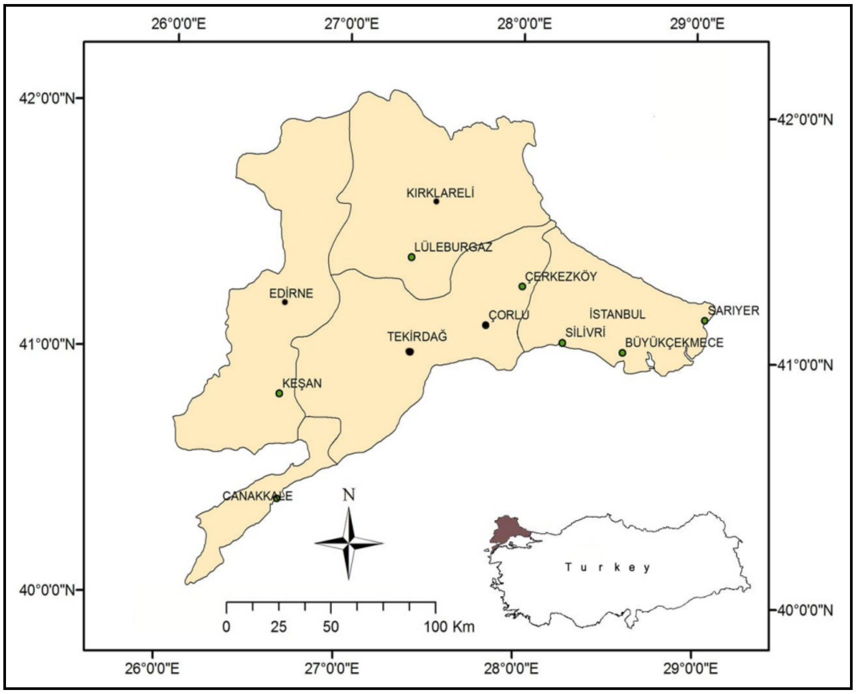


Figure 1: Location map of Thrace.

Data and Methods

Daily rainfall data of 3 stations, namely Edirne, Tekirdağ and Florya, are gathered for a span of 3 decades from 1970 to 2000. Data from another station, namely Çorlu are gathered for 2 decades from 1970 to 1990 making a total of 4 stations.

A feed forward back propagation type artificial neural network (ANN) has been used with a time series approach. “All ANN modules are based on the FFBP model. Given a training set of input–output data, the most common learning rule for multilayer perceptions is the back propagation algorithm (BPA). Back propagation involves two phases: a feed-forward phase in which the external input information at the input nodes is propagated forward to compute the output information signal at the output unit, and a backward phase in which modifications to the connection strengths are made based on the differences between the computed and observed information signals at the output units (Alp and Cigizoglu, 2007).

The neural network structure in this study possessed a three-layer learning network consisting of an input layer, a hidden layer and an output layer (Figure 2).

The FFBP configuration consists of an input layer, one or more hidden layers and an output layer. In a feed-forward network, the input quantities are fed to the input nodes, which in turn pass them on to the hidden layer nodes after multiplying by a weight. A hidden layer node, the function of which is to intervene between the external input and the network output, adds up the weighted input received from each input node, associates it with a bias, and then passes the result on to the nodes of the next hidden layer or the output, through a non-linear transfer function. The learning process works in small iterative steps. The output is compared to the known-good output, and a mean square error signal is calculated. The error value is then propagated backwards through the network, and small changes are made to the weights in each layer. The weight changes are calculated to reduce the error signal for the case in question. The cycle is repeated until the overall error value drops below some predetermined threshold (Ustaoglu et al. 2008)"

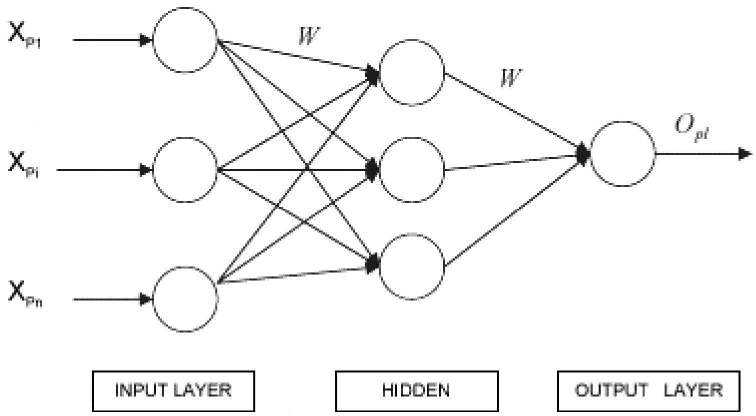


Figure 2: The structure of the feed-forward back-propagation neural network (FFBP).

The input consists of a time series of consecutive normalized rainfall data, daily or sum of ten days data, which ends with the day before the calculated plus a single normalized value that indicates the day count inside the year starting with September 1st. Two different normalization schemes have been utilized. First, the day count value has been divided by 366 to yield a value within the range [0, 1]. The reason 366 is chosen instead of 365 is to avoid the leap years to overload the value. Secondly a sinusoidal value with a period of 366 days and amplitude of 1 has been utilized. The reason for utilizing two different normalization schemes is to investigate whether the linear scheme would have a possible unwanted effect. Following the day indicator is a series of rainfall data from the previous days to fill the remaining input layer neurons.

The number of input and hidden layer neurons has been increased progressively and independently in 5 neuron steps starting with 5 ending with 30 neurons each. Along with the linear and sinusoidal normalization scheme for the day of the year value, this gives 2x6x6=72 ANN's per decade per station for daily calculations. In addition to the daily calculations, the ANNs have been trained for the sum of 10 days rainfall data. Within each decade, ANN's have been trained using the first 4 years, tested using the second 4 years and ran for the remaining 2 years for the daily and sum of 10 days calculations. The effects of the following parameters have been examined:

- The length of the time series (the number of input layer neurons)
- The number of the hidden layer neurons
- The linear vs sinusoidal normalized day indicator
- Summation (single day, 10 days)
- The rainfall characteristics of the station

The implementation of the method is realized by writing a custom C++ code utilizing the FANN library.

Results

Daily rainfall results

The minimum RMSE values for the training and the testing phases are given in Table 1. It is seen that the linear and the sinusoidal schemes showed no apparent differences whether during the training or the testing phase. Examining Table 2, in which the ANN topologies with the minimum RMSE values are given, it is apparent that the training and the testing phases showed minimum RMSE values on quite the opposite sides of the ANN topology range. That is the higher the number of input and hidden layer neurons the training RMSE value is lowest while the lowest testing RMSE values are obtained with lower number of input and hidden layer neurons.

Table 1. Minimum RMSE values for daily calculations.

Day count		**Sinusoidal normalization**			**Linear normalization**		
Year span		**1970-1079**	**1980-1989**	**1990-2000**	**1970-1079**	**1980-1989**	**1990-2000**
Training	**Edirne**	0.668	0.667	0.913	0.667	0.668	0.914
	Tekirdağ	0.720	0.721	1.398	0.721	0.720	1.400
	Florya	0.765	0.765	1.109	0.764	0.765	1.109

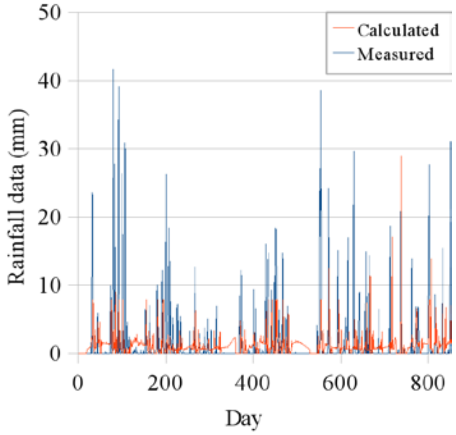
	Çorlu	0.817	0.816	-	0.816	0.816	-
Testing	**Edirne**	5.219	4.741	5.596	5.076	4.968	5.892
	Tekirdağ	5.428	4.772	6.171	5.487	4.470	6.754
	Florya	5.244	5.500	5.235	5.276	5.404	4.986
	Çorlu	5.902	3.925	-	5.746	4.157	-

Table 2. ANN topologies with the minimum RMSE values for daily calculations.

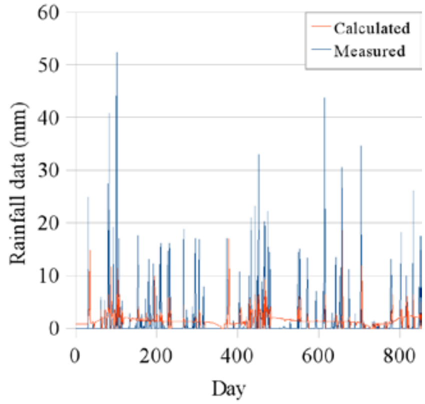
Day count		**Sinusoidal normalization**			**Linear normalization**		
Year span		**1970-1079**	**1980-1989**	**1990-2000**	**1970-1079**	**1980-1989**	**1990-2000**
Training	**Edirne**	30-30-1	30-25-1	25-30-1	30-25-1	30-30-1	30-30-1
	Tekirdağ	25-25-1	30-20-1	30-25-1	25-30-1	25-20-1	30-30-1
	Florya	30-25-1	25-25-1	25-30-1	30-30-1	25-25-1	30-30-1
	Çorlu	25-25-1	30-25-1	-	25-30-1	25-25-1	-
Testing	**Edirne**	5-5-1	10-5-1	5-5-1	10-5-1	5-5-1	5-5-1
	Tekirdağ	5-5-1	5-10-1	5-5-1	10-5-1	5-5-1	5-20-1
	Florya	5-10-1	5-5-1	5-5-1	5-5-1	5-5-1	5-10-1
	Çorlu	5-5-1	10-5-1	-	5-5-1	15-5-1	-

The RMSE value in training phase decrease with both the increasing number of input neurons and the number of neurons in the inner layer and minimum RMSE values are obtained with high number of input and inner layer neurons (25 or 30).

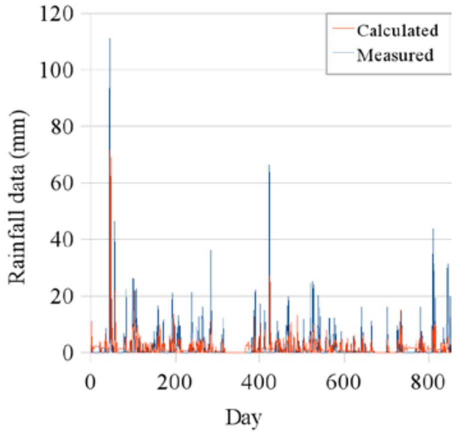
On the other hand the testing phase RMSE values do not show the same characteristics and tend to reach the minimum values with the lowest number of input and inner layer neurons (5 and 10). This is the result of the seeming irregularity of the daily rainfall data.



a

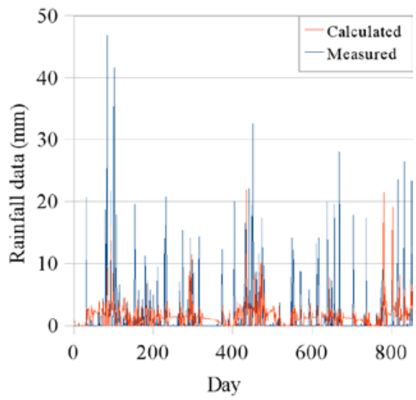


b



z

c



d

Figure 3. Results for station Edirne (a), Tekirdağ (b), Florya (c) and Çorlu (d) for daily rainfall data

The results of the daily runs for each station with the lowest RMSE values are given in Figure 3. In Figure 3a, the result for Edirne station for the data of 1980 – 1989 period with the topology 10-5-1 and sinusoidal day count normalization scheme is shown. In Figure 3b the result for Tekirdağ station with the topology 5-5-1 for the data of 1980 – 1989 period with linear day count normalization scheme is shown. In Figure 3c, the result for Florya station with the topology 5-10-1 with a linear date count normalization scheme for the period 1990 – 1999 is shown. Figure 3d shows the result for Çorlu station for period 1980 – 1989 with sinusoidal day count normalization scheme and a topology of 10-5-1.

Examining Figure 3, it is seen that the plain ANN method with time series approach does not yield day to day accurate predictions of the rainfall data. Although the rainfall calculated results seem to show a qualitative match with the observed data, they show a poor performance in the quantitative sense. This is partly due to the discontinuous and non-periodic nature of the daily rainfall data. There exists rainless days along with the individual days in which the data shows a sharp peak following a sudden fall the next day, rather than a smooth, periodic, continuous curve.

10 days rainfall results

The minimum RMSE values for the training and the testing phases and the topologies resulting in these values are given in Table 3 and 4 respectively. The results of the 10 daily summed runs for each station with the lowest RMSE values are given in Figures 5 through 8. In Figure 5a, the result for Edirne station for the data of 1980 – 1989 period with the topology 25-15-1 and linear day count normalization scheme is shown. In Figure 5b the result for Tekirdağ station with the topology 15-25-1 for the data of 1980 – 1989 period with sinusoidal day count normalization scheme is shown. In Figure 6a, the result for Florya station with the topology 10-20-1 with a sinusoidal date count normalization scheme for the period 1990 – 1999 is shown. And Figure 6b shows the result for Çorlu station for period 1980 – 1989 with sinusoidal day count normalization scheme and a topology of 30-30-1.

Table 3. Minimum RMSE values for 10 days summed calculations.

Day count		**Sinusoidal normalization**			**Linear normalization**		
Year span		**1970-1079**	**1980-1989**	**1990-2000**	**1970-1079**	**1980-1989**	**1990-2000**
Training	**Edirne**	0.659	0.658	0.884	0.662	0.655	0.908
	Tekirdağ	0.713	0.709	1.382	0.714	0.707	1.365
	Florya	0.748	0.756	1.093	0.748	0.759	1.102
	Çorlu	0.805	0.809	-	0.807	0.812	-
Testing	**Edirne**	2.189	2.086	2.198	2.161	1.967	2.384
	Tekirdağ	2.321	2.067	2.598	2.447	2.194	2.614
	Florya	2.048	2.462	1.907	1.963	2.762	2.008
	Çorlu	2.398	1.747	-	2.355	1.901	-

Table 4. ANN topologies with the minimum RMSE values for 10 days summed calculations.

Day count		Sinusoidal normalization			Linear normalization		
Year span		1970-1079	1980-1989	1990-2000	1970-1079	1980-1989	1990-2000
Training	Edirne	15-25-1	20-10-1	20-20-1	20-20-1	25-30-1	30-25-1
	Tekirdağ	25-30-1	30-10-1	25-10-1	30-25-1	25-25-1	15-10-1
	Florya	30-15-1	25-25-1	25-30-1	20-30-1	25-10-1	10-20-1
	Çorlu	30-20-1	30-25-1	-	30-5-1	25-20-1	-
Testing	Edirne	10-25-1	15-20-1	10-30-1	15-15-1	25-15-1	15-30-1
	Tekirdağ	15-30-1	15-25-1	15-30-1	5-20-1	10-10-1	15-25-1
	Florya	15-20-1	20-15-1	10-20-1	15-25-1	10-15-1	20-10-1
	Çorlu	15-30-1	30-30-1	-	10-20-1	25-30-1	-

As with the daily calculations, the linear and the sinusoidal schemes showed no apparent differences. Comparing the values given in Table 1 and Table 3 it is seen that the summation over a 10 days period yielded RMSE values that are approximately half of the daily calculations. On a predictive point of view on the other hand, Figures 5 and 6 show that for Tekirdağ and Çorlu stations although the peaks on the observed rainfall data seemed to be matched by the calculated results on the basis of magnitude, there exists a clear time lag between the observed and the calculated results. For Edirne and Florya stations however, the results could not match even in magnitude. These results show that a summation over a ten day period is not sufficient to smooth the irregularity and the discontinuity of the daily rainfall data to be used with plain ANN calculations. Different rainfall characteristics of different stations also caused a distinction in the predictive capacity of the method used.

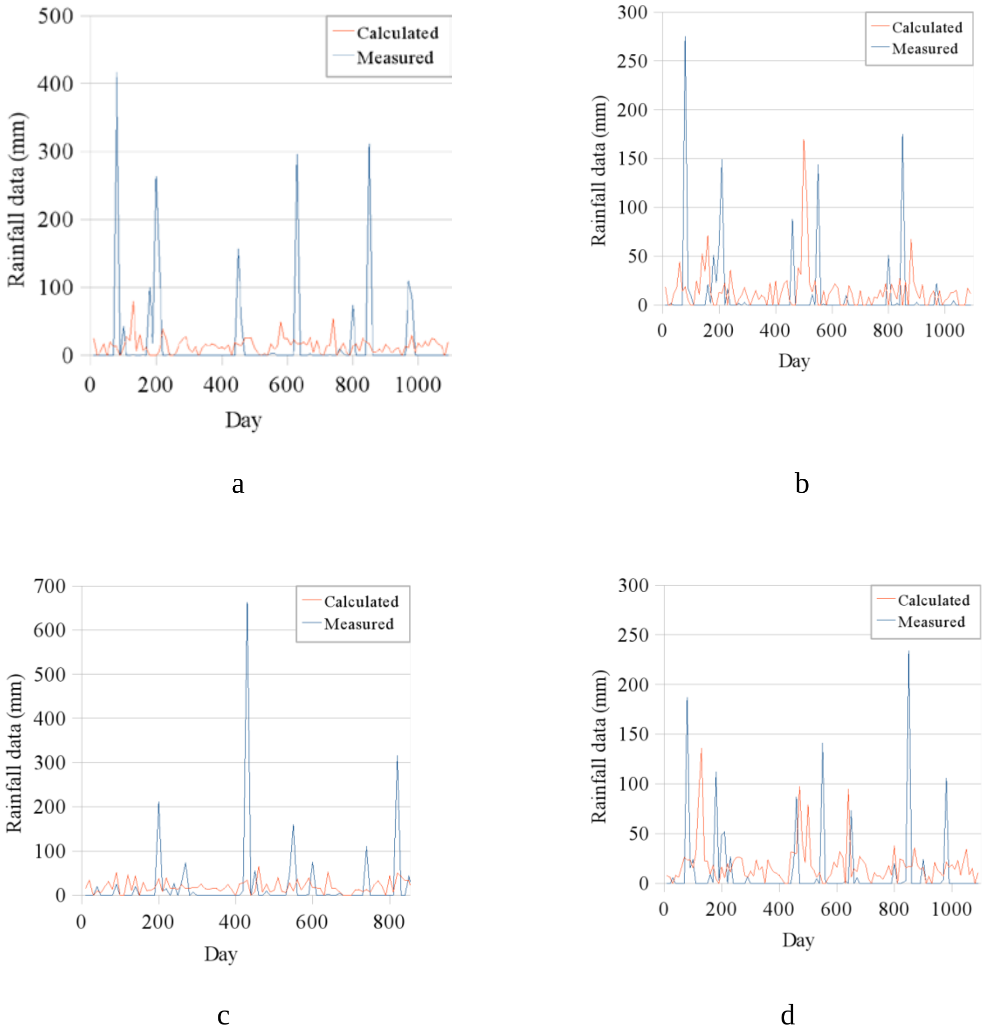


Figure 4. Results for station Edirne (a), Tekirdağ (b), Florya (c) and Çorlu (d) for 10 days rainfall data

Discussion

Artificial Neural Networks when used with a time series approach is a powerful tool for the forecasting of climatic phenomena, when the available data is insufficient to allow the usage of more complex and accurate models. The daily calculations show poor performance in the quantitative sense while summation of the consecutive data yields better results. While the number of input and hidden layer neurons showed almost insignificant effect on the results as the normalization scheme of the day count, the results are best improved by the summation process. It should be stressed however that each individual station should be taken into account separately and the ANN model should be improved accordingly.

On the basis or RMSE values, the 10 days summed calculations yielded better results, lower than half the values of daily calculations. For daily results, the order of the stations with increasing minimum RMSE values is Çorlu, Tekirdağ, Edirne and Florya; with the values 3.925, 4.470, 4.741 and 4.986 mm and topologies 10-5-1, 5-5-1, 10-5-1 and 5-10-1 respectively. For the 10 days rainfall data the order is Çorlu, Florya, Edirne and Tekirdağ with the values 1.747, 1.907, 1.967 and 2.067 mm and topologies 30-30-1, 10-20-1, 25-15-1 and 15-25-1 respectively. For both methods the best results are the predictions of Çorlu station, the worst place is shared between Florya and Tekirdağ stations for the daily and 10 days results respectively. For further improvement of the method, a wider range of stations from different climatic properties should be experimented and the optimum summation parameter i.e. the number of days to be summed should be determined.

References

Abbot, J., Marohasy, J., 2012. Application of artificial neural networks to rainfall forecasting in Queensland, Australia. Advances In Atmospheric Sciences Volume: 29 Issue: 4 Pages: 717-730.

Alp M., Cigizoglu, HK., 2007. Suspended sediment load simulation by two artificial neural network methods using hydro meteorological data. Environmental Modelling & Software 22(1): 2–13.

Çiçek, İ., 1995. Türkiye'deki Kurak Dönemin Yayılışı ve Süresi (Thornthwaite Metoduna Göre) Ankara Üniversitesi, Türkiye Coğrafyası Araştırma ve Uygulama Merkezi Dergisi, S:4, s:77-101.

Ilunga, M., 2010. Infilling annual rainfall data using feed forward back-propagation Artificial Neural Networks (ANN): Application of the standard and generalized back-propagation techniques. Journal of The South African Institution of Civil Engineering Volume: 52 Issue: 1.

İkiel, C.,2005. Rainfall Regime Regions in Turkey (A Statistical Climate Study),Forest Impact on Hydrological Process and Soil Erosion, Yundola, Bulgaria.

Kim, JW., Pachepsky, YA., 2010. Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT stream flow simulation. Journal of Hydrology Volume: 394 Issue: 3-4 Pages: 305-314.

Kumarasiri, A D., Sonnadara, U J., 2008. Performance of an artificial neural network on forecasting the daily occurrence and annual depth of rainfall at a tropical site Hydrological Processes Volume: 22 Issue: 17 Pages: 3535-3542.

Rajurkar, M.P., Kothyari, U.C., Chaube, U.C., 2004. Modeling of the daily rainfall-runoff relationship with artificial neural network. Journal of Hydrology Volume 285, Issues 1–4, 15 January 2004, Pages 96–113.

Türkeş, M ., 2010, *Klimatoloji ve Meteoroloji* Kriter Yayınevi - Yayın No. 63, Fiziki Coğrafya Serisi No. 1 ISBN: 978-605-5863-39-6.

Ustaoglu B, Cigizoglu HK, Karaca M., 2008. Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. Meteorol. Appl. 15: 431–445.

Ustaoglu B., 2011. Türkiye'de A2 Emisyon Senaryosuna Göre Ortalama Yağış Tutarlarının Olası Değişimi,", Türk Coğrafya Kurumu Yayınları, Fiziki Coğrafya Araştırmaları Sistematik ve Bölgesel. Prof. Dr. M. Y. Hoşgören'in 40. meslek yılı makaleler kitabı, ISBN : Yayın No: 6.

Wu, C. L., Chau, K. W., Fan, C., 2010. Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques Journal Of Hydrology Volume: 389 Issue: 1-2 Pages: 146-167.