ASSESSING PRECIPITATION PREDICTION MODELS: ARTIFICIAL NEURAL NETWORKS (ANN)

Halil Ibrahim Yildiz[*,1]

[*]hibrahimyldz@outlook.com
[1] 7 boulevard Copernic, Residence Montesquieu 77420 Champs-sur-Marne, Paris, France

Abstract

Precipitation data are important for solving engineering problems. Missing data can be predicted, and historical data can be used to construct precipitation models. This study used monthly precipitation data, temperature, relative humidity, wind, and evaporation data from the Afyon Meteorological Observatory station between 1929 and 2018. The ANN (Artificial Neural Network) method was used to predict precipitation data, and the results were compared with MLR (Multilinear Regression) models.

Keywords

Artificial Neural Networks, Multilinear Regression, Precipitation, Hydrology

1 INTRODUCTION

Water, one of the most basic resources for sustaining life, is a critical element that must be carefully managed to ensure the continuity of ecosystems. Its unpredictability poses a significant challenge to the stability of life on Earth. As a result, decisions aimed at transforming water use into more efficient use, particularly concerning the distribution and availability of existing resources, can be guided by hydrological models.

Today, water infrastructure costs are high, and their development often requires long-term planning. Before such structures can be built for various purposes, it is essential to model factors such as precipitation, which is a key input for flow calculations. To create accurate hydrological models, it is essential to have complete and reliable data.

In areas where direct measurements are lacking, a temporary data collection system can be established to collect data over several years. The data can then be integrated with measurements from other stations in the same catchment to predict historical data. Methods for estimating precipitation are therefore of great importance. With proper analysis, historical datasets can be predicted with a minimum Mean Square Error (MSE) using only a few years of measured data [1].

This study predicted total monthly precipitation using ANN and MLR methods. The study was based on monthly data from a meteorological station in Turkey, including total precipitation, wind, evaporation, temperature, and relative humidity from 1929 to 2018.

Literature Review

Precipitation prediction models can be divided into two main categories: experimental and dynamic. Experimental approaches are generally based on historical data with significant correlations between them. The most important ones are ANN, stochastic models, fuzzy logic, and data-based group models [2]. Dynamic approaches, another method used in precipitation estimation, are based on sets of equations formed by combining climatic conditions and atmospheric changes. A powerful regression technique was developed for modeling long-duration and summer monsoon rains [3].

The model used snowfall in Asia, temperature values in northwestern Europe, pressure in Europe and Asia, surface temperatures in the Arabian Sea, and the previous year's temperature values in the Indian Ocean [3]. Consequently, models were constructed to forecast rainfall at the 95% significance level. In another model, prepared by creating linear regression models, sea level, sea surface temperature, and the Southern Oscillation Index of El Niño were used. As a result of the model, a correlation of 0.60 was obtained [4].

Received 10/01/2025/Accepted 12/05/2025/Published 30/09/2025 DOI 10.13164/juniorstav.2025.25047

Two precipitation observation stations were developed using an FFBP (Feed-Forward Back-Propagation) neural network. The study used 38 years of precipitation data, and the FFBP neural network successfully predicted the following year's precipitation values [5].

An ANN was employed to forecast daily runoff using daily precipitation, temperature, and snowmelt data for the watershed [6], and data from 154 meteorological stations in the United States were used to predict 24-hour precipitation. Five different methods were used, with the most accurate results obtained using ANN-trained models [7].

ANN models for precipitation prediction were developed using data from three precipitation observation stations in Turkey. In the study of the FFBP neural network and radial basis function ANNs, a closer approximation to the actual values was obtained with FFBP neural network models [8].

An ANN model containing meteorological parameters predicted monthly average soil temperature for subsequent years. In comparing the model's precipitation results with the regression analysis results, the ANN model was the closest fit to the actual measured values [9].

2 METHODOLOGY

The data used in the study were analyzed for skewness and kurtosis, as shown in Fig. 3, and normalized so that the data sets would be scaled between 0 and 1.

The prepared data were compared with each data set in Tab. 1 for the MLR method, and with the data in Tab. 2 from the previous month. For each month, the data sets were shifted back four months, and nine models were created in Tab. 3 that could be meaningful. The equations for the nine models in Tab. 4 were created by MLR.

The FFBP tool in MATLAB was used for the ANN models [10]. Calculation matrices were created for the input values prepared in Tab. 3. The output was created to predict the current month. While building the model, it was decided to use 40% for the test and validation sets and 60% for the training set. Supervised learning was performed on these models using the Bayes, Levenberg, and Conjugate training algorithms with 4, 6, and 10 hidden layers.

Data

The data used in this study were obtained from the observation station No. 17190, located in Afyon under the Meteorological Regional Directorate V. The station is situated at a latitude of 38.738°, longitude of 30.5604°, and an altitude of 1034 m. The data collected include precipitation and evaporation (mm), temperature (°C), and wind speed (m/s).

The Afyon catchment is a sub-catchment of the Akarcay closed basin, with a total drainage area of 7,337 km² and a sub-catchment drainage area of 818.5 km² (Fig. 1) [11]. The average annual precipitation in the catchment is 436.8 mm.

DOI 10.13164/juniorstav.2025.25047

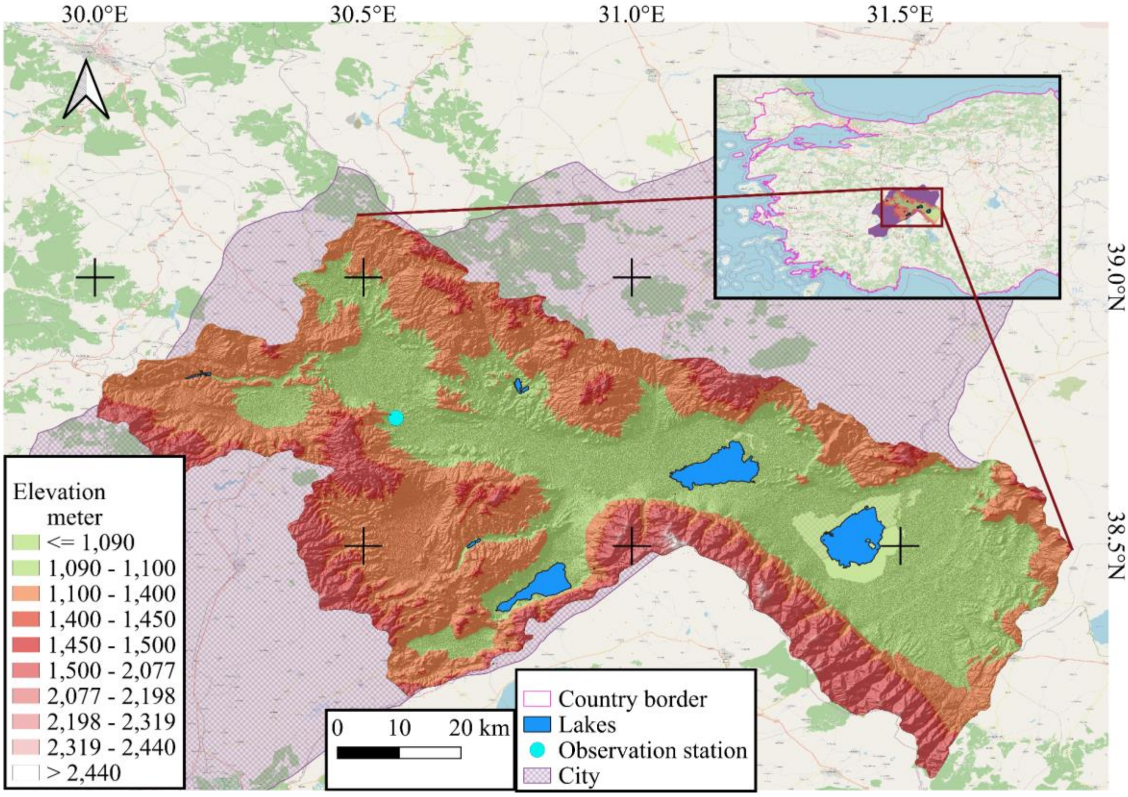

Fig. 1 Akarcay closed basin in Turkey.

Multilinear Regression

The linear relationship between variables is shown in Equation (1):

$$\Upsilon i = \varphi 1\mathrm{X}1 + \varphi 2\mathrm{X}2 + \varphi 3\mathrm{X}3 + \cdots + \varphi i\mathrm{Xi} + \varepsilon\mathrm{t} \tag{1}$$

In Equation (1), ϒi denotes the estimated value, and X1, X2, etc., Xi represent the types of variables related to the output parameter. The correlation coefficients of these inputs are φi, and the sum of the error values is denoted by εt. In regression analysis, the data have to be normally distributed. Logarithmic transformation can be applied to convert non-normally distributed data into a normal distribution as shown in Equations (2) and (3).

$$\mathrm{Zi} = \frac{\mathrm{Xi} - \mathrm{X}}{\mathrm{S}} \tag{2}$$

Standardization is performed using this transformation, and the skewed data are then transformed as follows;

$$\log\left[\mathrm{Zi} + \left(1 - \min(\mathrm{Zi})\right)\right] \tag{3}$$

The conversion in Equation (3) is feasible, where Zi denotes the value of the scores, and min (Zi) is the minimum of the current data set. In Equation (2), Xi is the observation value, X is the mean of the observation set, and S indicates the standard deviation.

The methods used to evaluate model success include the Correlation Coefficient (R), Coefficient of Determination (R^2), and Mean Squared Error (MSE). The MSE is defined in Equation (4):

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Hgi - Hti)^2 \tag{4}$$

DOI 10.13164/juniorstav.2025.25047

In Equation (4), MSE is the mean squared error value, n is the number of observations, Hgi is the observation value, and Hti is the estimation.

Artificial Neural Networks

An FFBP neural network has one input layer, one output layer, and at least one hidden layer (Fig. 2). During the learning process, forward scanning is performed across the network, and the output of a successive sequence of nodes is calculated and transferred to the output layer. The desired values are compared with those assigned to the output layer. Then, to reduce the error rate between these two values, the network returns to the beginning and repeats the process by adjusting the weight values between the layer elements. In short, the main purpose of the network structure is to minimize the error between the output layer values and the desired result values [4], [5].

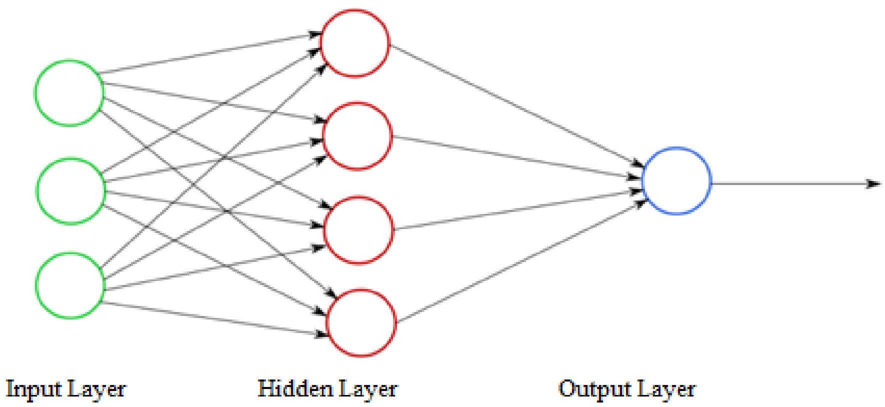

Fig. 2 Structure of the ANN.

3 RESULTS

Multilinear Regression (MLR) model parameters were analyzed by considering the correlation values obtained in SPSS software [12].

Tab. 1 Correlation values for precipitation.

	Precipitation	Temperature	R-Humidity	Wind	Evaporation
Precipitation	1				
Temperature	-0.32	1			
R-Humidity	0.62	-0.77	1		
Wind	-0.13	0.05	-0.18	1	
Evaporation	0.21	0.85	-0.60	0.17	1

Due to the low correlation and lack of data, evaporation and wind effects are not included in the models.

DOI 10.13164/juniorstav.2025.25047

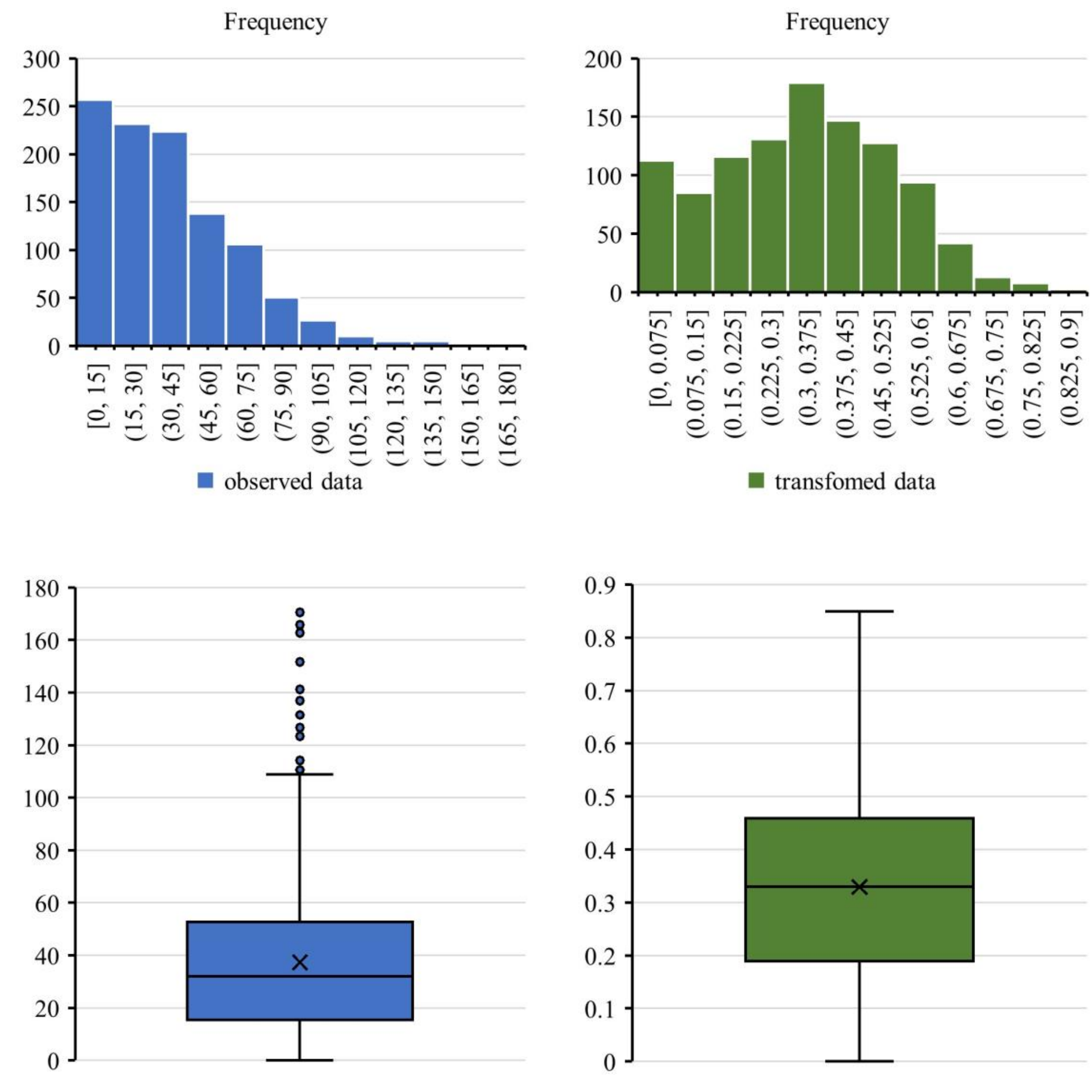

Fig. 3 Frequency of precipitation after observation and transformation. The boxes show the interquartile range (25^{th}–75^{th} percentiles) with the median line inside. Whiskers represent the 10^{th} and 90^{th} percentiles. Dots indicate outliers. Crosses indicate the average.

Since the data were not normally distributed and contained extreme outliers, the data were standardized. A logarithmic function correction was applied to the data to reduce the effect of skewness. It was observed that there is a dependency between precipitation, temperature, and relative humidity. The correlations of the variables were analyzed according to the time-lagged values: $P_{(t-1)}$, $R\text{-}H_{(t)}$, $R\text{-}H_{(t-1)}$, $R\text{-}H_{(t-2)}$, $T_{(t)}$, $T_{(t-1)}$, $T_{(t-2)}$, and $T_{(t-3)}$, and the best correlation was found for the monthly precipitation variable. Based on the results of the variance analysis for these variables, models were created by reducing the significance coefficient. The variables T_t represent temperature, $R\text{-}H_{(t)}$ relative humidity, and P_t precipitation, which were used in Tab. 2 for the current month. At the same time, the sub-indices 1, 2, 3, and 4 contain data for the retrospective lagged month. The highest correlation is between precipitation and humidity.

DOI 10.13164/juniorstav.2025.25047

	$P_{(t)}$	$P_{(t-1)}$	$P_{(t-2)}$	$P_{(t-3)}$	$P_{(t-4)}$	$T_{(t)}$	$T_{(t-1)}$	$T_{(t-2)}$	$T_{(t-3)}$	$T_{(t-4)}$	$R\text{-}H_{(t)}$	$R\text{-}H_{(t-1)}$	$R\text{-}H_{(t-2)}$	$R\text{-}H_{(t-3)}$	$R\text{-}H_{(t-4)}$
$P_{(t)}$	1	-0.03	-0.02	-0.22*	-0.09	-0.49**	0.13	0.07	0.04	-0.02	0.75**	-0.05	-0.10	0.04	-0.04
$P_{(t-1)}$		1	0.15	0.21*	0.146	0.12	-0.47**	-0.11	0.18	0.12	0.02	0.70**	0.12	-0.01	-0.03
$P_{(t-2)}$			1	0.27**	-0.05	0.20	-0.11	-0.21	0.11	0.05	-0.03	0.25*	0.52**	0.13	-0.08
$P_{(t-3)}$				1	0.16	0.18	-0.16	-0.29**	-0.04	0.34**	-0.11	0.22*	0.30**	0.41**	0.02
$P_{(t-4)}$					1	0.22*	-0.12	-0.14	0.08	-0.05	-0.10	0.13	0.19	0.15	0.38**
$T_{(t)}$						1	-0.06	-0.04	0.07	0.15	-0.68**	0.05	0.09	-0.09	0.05
$T_{(t-1)}$							1	0.35**	0.08	-0.02	0.12	-0.65**	-0.39**	-0.19	0.01
$T_{(t-2)}$								1	0.35**	-0.03	-0.05	-0.28**	-0.57**	-0.26*	-0.02
$T_{(t-3)}$									1	0.41**	-0.02	0.02	-0.23*	-0.47**	-0.05
$T_{(t-4)}$										1	-0.07	0.07	-0.05	-0.25*	-0.20
$R\text{-}H_{(t)}$											1	0.12	0.07	0.20	0.03
$R\text{-}H_{(t-1)}$												1	0.49**	0.27**	0.13
$R\text{-}H_{(t-2)}$													1	0.53**	0.28**
$R\text{-}H_{(t-3)}$														1	0.30**
$R\text{-}H_{(t-4)}$															1

*. Correlation is significant at the 0.05 level (2-tailed) and **. Correlation is significant at the 0.01 level (2-tailed)

Tab. 2 Correlation coefficients of the inputs.

The variables with the highest variance among the selected variables were chosen while building the model. Thus, the models in Tab. 3 were created based on of Tab. 2.

Tab. 3 Models for ANN and MLR.

MODELS	INPUTS	OUTPUT
1	R-H_t, P_{t1}, R-H_{t4}, T_{t2}, T_{t4}, P_{t2}, T_{t1}, T_{t3}, P_{t3}, T_t, R-H_{t2}, R-H_{t1}	P_t
2	R-H_t, P_{t1}, R-H_{t4}, T_{t4}, P_{t2}, T_{t1}, T_{t3}, P_{t3}, T_t, R-H_{t2}, R-H_{t1}	P_t
3	R-H_t, P_{t1}, R-H_{t4}, T_{t4}, P_{t2}, T_{t1}, P_{t3}, T_t, R-H_{t2}, R-H_{t1}	P_t
4	R-H_t, P_{t2}, R-H_{t4}, T_{t4}, T_{t1}, P_{t3}, T_t, R-H_{t2}, R-H_{t1}	P_t
5	R-H_t, P_{t2}, T_{t4}, T_{t1}, P_{t3}, T_t, R-H_{t2}, R-H_{t1}	P_t
6	R-H_t, P_{t2}, T_{t1}, P_{t3}, T_t, R-H_{t2}, R-H_{t1}	P_t
7	R-H_t, P_{t2}, T_{t1}, P_{t3}, R-H_{t2}, R-H_{t1}	P_t
8	R-H_t, P_{t2}, T_{t1}, R-H_{t2}, R-H_{t1}	P_t
9	R-H_t, P_{t2}, R-H_{t2}, R-H_{t1}	P_t

Tab. 4 shows the equation sets of the MLR (Multilinear Regression Analysis).

Tab. 4 Model equations obtained from MLR analysis.

Model	Cons.	$T_{(t-4)}$	$T_{(t-3)}$	$T_{(t-2)}$	$T_{(t-1)}$	$T_{(t)}$	$P_{(t-3)}$	$P_{(t-2)}$	$P_{(t-1)}$	R-$H_{(t-4)}$	R-$H_{(t-2)}$	R-$H_{(t-1)}$	R-$H_{(t)}$
1	-0.15	0.19	-0.10	0.07	-0.55	0.40	-0.19	0.19	0.11	0.03	-0.07	-0.10	0.29
2	-0.06	0.187	-0.09		-0.53	0.39	-0.19	0.19	0.10	0.03	-0.07	-0.09	0.28
3	-0.13	0.15			-0.53	0.39	-0.17	0.17	0.09	0.03	-0.07	-0.09	0.28
4	-0.13	0.15			-0.54	0.41	-0.17	0.18		0.03	-0.07	-0.08	0.28
5	-0.10	0.13			-0.50	0.43	-0.17	0.16			-0.06	-0.07	0.28
6	-0.02				-0.49	0.46	-0.14	0.16			-0.07	-0.07	0.28
7	0.73				-0.41		-0.13	0.18			-0.06	-0.06	0.25
8	0.73				-0.42			0.16			-0.07	-0.07	0.26
9	-0.06							0.13			-0.06	-0.03	0.24

DOI 10.13164/juniorstav.2025.25047

Tab. 5 R^2 and MSE values obtained from ANN–MLR analyses.

		MLR		ANN					
				Bayes		Levenberg		Conjugate	
SCENARIOS		R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE
	10			0.94	0.42	0.01	0.34	0.92	0.43
1	**6**	0.80	0.64	0.80	0.23	0.82	0.41	0.82	0.35
	4			0.83	0.38	0.76	0.32	0.73	0.43
	10			0.80	0.43	0.79	0.32	0.78	0.43
2	**6**	0.80	0.64	0.80	0.36	0.80	0.38	0.86	0.35
	4			0.81	0.36	0.98	0.40	0.77	0.32
	10			0.79	0.37	0.83	0.30	0.72	0.45
3	**6**	0.80	0.60	0.84	0.36	0.85	0.33	0.83	0.45
	4			0.81	0.34	0.86	0.39	0.80	0.44
	10			0.82	0.35	0.93	0.23	0.81	0.35
4	**6**	0.80	0.57	0.78	0.37	0.88	0.16	0.85	0.47
	4			0.79	0.32	0.81	0.34	0.79	0.46
	10			0.77	0.34	0.99	0.23	0.81	0.33
5	**6**	0.79	0.58	0.79	0.26	0.89	0.36	0.82	0.39
	4			0.81	0.33	0.86	0.26	0.80	0.44
	10			0.78	0.32	0.84	0.40	0.80	0.47
6	**6**	0.79	0.59	0.81	0.34	0.86	0.29	0.79	0.49
	4			0.79	0.37	0.81	0.32	0.83	0.50
	10			0.80	0.39	0.83	0.35	0.74	0.43
7	**6**	0.79	0.61	0.79	0.33	0.85	0.41	0.79	0.39
	4			0.79	0.33	0.86	0.26	0.82	0.44
	10			0.81	0.36	0.81	0.46	0.70	0.43
8	**6**	0.78	0.49	0.78	0.28	0.82	0.31	0.74	0.39
	4			0.78	0.44	0.81	0.36	0.80	0.36
	10			0.76	0.36	0.84	0.25	0.75	0.46
9	**6**	0.77	0.46	0.80	0.32	0.87	0.37	0.83	0.41
	4			0.79	0.35	0.81	0.31	0.72	0.45

DOI 10.13164/juniorstav.2025.25047

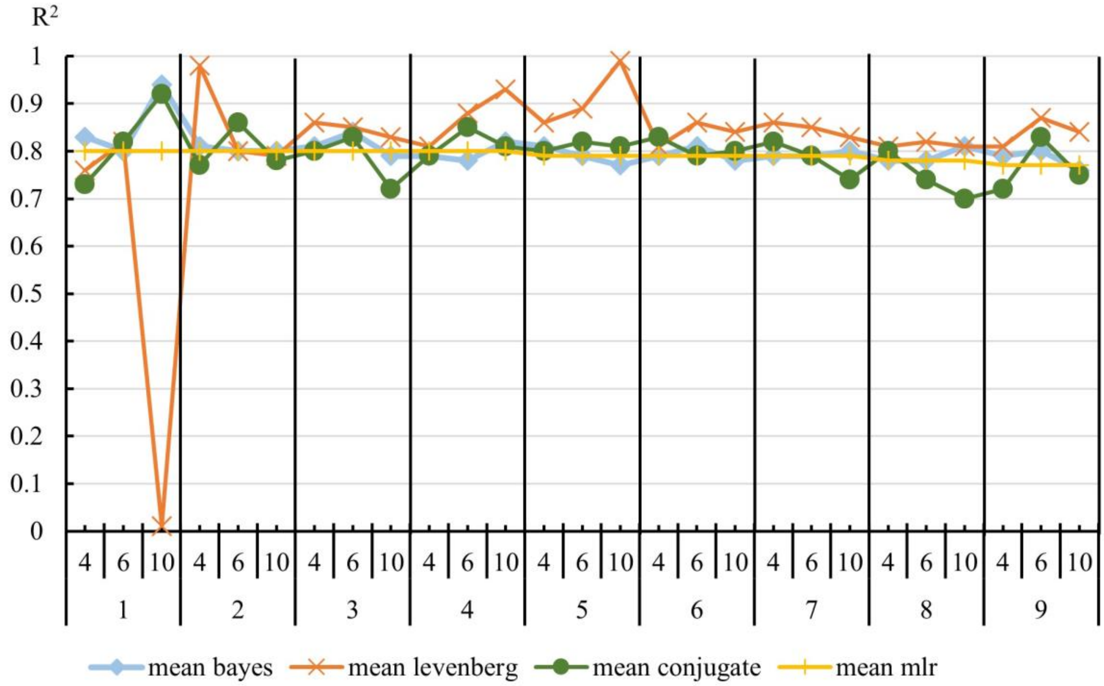

Fig. 4 Comparison of R^2 values for the mean of Bayes, Levenberg, Conjugate, and MLR.

4 DISCUSSION

In this study, monthly precipitation values were predicted in one step using temperature, relative humidity, and precipitation data, considering the correlation values in Tab. 1. The values of the month to be predicted and the significant relationships were logarithmically transformed according to equations (2) and (3), and nine different models in Tab. 3 were created. Three different algorithms – Bayes, Levenberg, and Conjugate – were tested in the FFBP structure, with each algorithm tested on models with 4, 6, and 10 hidden layers. Comparing the results in Fig. 4, the Levenberg algorithm showed slower convergence than the other algorithms and MLR at the 10^{th} hidden layer for the high input values in the first model.

Comparing the types of inputs used in different studies shows that it is necessary to test the effect of input types when creating models [1], [5], [8], [9].

The first and fourth models with six hidden layers provided better approximation results, and the Levenberg algorithm predicted monthly precipitation more effectively (Tab. 5). While starting the process with classical stochastic models, as shown in Tab. 4, certain conditions must be met to reduce the effects of extreme values and make the data meaningful. These are normal distribution, stationarity, and constant variance. Since natural events exhibit statistically non-normal behavior, pre-processing is required to meet these conditions before modeling. Applications were conducted using logarithmic transformation, and among the models prepared with the ANN, the four-layer Levenberg structure gave more accurate precipitation results than the best results from multilinear regression models. However, in future studies, it would be beneficial to test the effect of different transformation methods. As a result, it was observed that the ANN model converged better. In this case, hydrological researchers are advised to use models built with ANN.

5 CONCLUSION

While examining the research results, it was observed that the ANN method prediction scenarios produced better results than the MLR model. It has been found that precipitation data are less affected than wind and evaporation data. In future studies, the long-term effects of variables such as wind and evaporation should be monitored, and parameters correlated with linear bias may play a crucial role in predicting extreme values.

DOI 10.13164/juniorstav.2025.25047

It is expected that recent data will show closeness to the prediction curve. While creating the prediction dataset, data from the previous week should be sufficient to predict the target day. However, the size of the training set allows us to test the prediction power of the model.

It may be beneficial to fit the data to a standard distribution curve or test noise reduction methods to reduce the noise effect in the prediction curve.

The prediction model created with the ANN method shows values closer to those of the MLR model while predicting the three distinct variables. The impact of relative humidity and sudden changes in temperature on precipitation prediction is highlighted.

As shown in Tab. 3, even minor fluctuations in temperature or relative humidity dramatically impact the accuracy of the precipitation forecast, showing the delicacy and sensitivity of the weather. Such drastic changes, generally difficult to predict with linear models, were predicted better by the ANN model and provided a better precipitation forecast. By addressing these areas, it is expected that prediction models will become even more accurate, benefiting fields such as agriculture, disaster management, and environmental monitoring.

Acknowledgements

The research data were provided by the Afyon Meteorology Department.

References

[1] SAPLIOĞLU, K., and ÇİMEN, M. Yapay Sinir Ağlarını Kullanarak Günlük Yağış Miktarının Tahmini [Predicting of Daily Precipitation Using Artificial Neural Network]. Journal of Engineering Science and Design [Online]. 2010, vol: 1, no: 1, pp. 14-21. [Accessed 10/03/2025]. Available at:https://dergipark.org.tr/en/download/article-file/195364.

[2] ZAW, W.T.T., and NAING, T.T. Empirical statistical modeling of rainfall prediction over Myanmar. International Journal of Computer and Information Engineering. 2008, vol: 2, no: 10, pp. 3418-3421. Available at: https://doi.org/10.5281/zenodo.1084254.

[3] RAJEEVAN, M., PAI, D.S., KUMAR, R.A., and LAL, B. New statistical models for long-range forecasting of southwest monsoon rainfall over India. Climate Dynamics. 2007, vol: 28, no: 7-8, pp. 813-828. Available at: https://doi.org/10.1007/s00382-006-0197-6.

[4] SINGHRATTNA, N., RAJAGOPALAN, B., CLARK, M., and KUMAR, K. K. Seasonal forecasting of Thailand summer monsoon rainfall. International Journal of Climatology. 2005, vol: 25, no: 5, pp. 649–664. Available at: https://doi.org/10.1002/joc.1144.

[5] BODRI, L., and ČERMÁK, V. Prediction of extreme precipitation using a neural network: application to summer flood occurrence in Moravia. Advances in Engineering Software 2000, vol: 31, no: 4, pp. 311-321. Available at: https://doi.org/10.1016/S0965-9978(99)00063-0.

[6] TOKAR, A. S., and JOHNSON, P. A. Rainfall-runoff modeling using artificial neural networks. Journal of Hydrologic Engineering [Accessed 10/03/2025]. 1999, vol: 4 no: 3, pp. 232–241. Available at: https://doi.org/10.1061/(ASCE)1084-0699(1999)4:3(232).

[7] APPLEQUIST, S., GAHRS, G. E., PFEFFER, R. L., and NIU, X. Comparison of Methodologies for Probabilistic Quantitative Precipitation Forecasting. Weather and Forecasting. 2002, vol. 17, pp. 783–799. Available at: https://doi.org/10.1175/1520-0434(2002)017<0783:COMFPQ>2.0.CO;2.

[8] PARTAL, T., KAHYA, E., and CIĞIZOĞLU, K. Estimation of precipitation data using artificial neural networks and wavelet transform. İTÜ Journal/D Engineering [Online] [Accessed 10/03/2025]. 2008, vol: 7, no: 3, pp. 73–85. Available at: http://www.itudergi.itu.edu.tr/index.php/itudergisi_d/article/viewFile/321/274.

[9] ASLAY, F., and ÖZEN, Ü. Meteorolojik parametreler kullanılarak yapay sinir ağları ile toprak sıcaklığının tahmini [Estimating soil temperature with artificial neural networks using meteorological parameters]. Journal of Polytechnic. 2013, vol: 16, no: 4, pp. 139-145. Available at: https://dergipark.org.tr/tr/pub/politeknik/issue/33068/367992.

[10] THE MATHWORKS INC. MATLAB. Version 9.3 (R2017b). Natick, Massachusetts: The MathWorks Inc., 2017. Available at: https://www.mathworks.com.

[11] TEZCAN, L., et. al. Akarçay Havzası hidrojeolojisi ve yeraltısuyu akım modeli [Hydrogeology and Groundwater Flow Model of the Akarçay Catchment]. Ankara: Hacettepe Üniversitesi Uluslararası Karst Su Kaynaklan Uygulama ve Araştırma Merkezi-Devlet Su İşleri Genel Müdürlüğü, 2002, p. 339.

[12] IBM CORP. IBM SPSS STATISTICS FOR WINDOWS. Version 24.0. Armonk, NY: IBM Corp., 2016. Available at: https://www.ibm.com.