Arabian Journal of Geosciences (2021) 14: 86
https://doi.org/10.1007/s12517-020-06363-x

ICCESEN 2017

Precipitation forecasting in Marmara region of Turkey

Veysel Coban[1] · Ezgi Guler[1] · Taner Kilic[2] · Suheyla Yerel Kandemir[1]

Received: 26 October 2020 / Accepted: 7 December 2020 / Published online: 15 January 2021

Abstract

Precipitation regimes that change with global warming and climate changes affect the countries in environmental, economic, and social dimensions. The Marmara region is an important region located in the northwest of Turkey. The impact of economic, environmental, and social dimensions in the region is high. For this reason, the Marmara region is in a situation that can be affected more by climate change and drought. Precipitation forecasting is the first step for the management of agricultural planning, flood controls, and use of drinking water resources. Time series analysis is an important statistics tool that allows forecasting the amount of future precipitation based on the historical data analysis. Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), and Seasonal Autoregressive Integrated Moving Average (SARIMA) models are the most common statistical methods used to estimate precipitation based on time series. The ARMA, ARIMA, and SARIMA models are based on the assumption that past conditions will remain the same in the future. In this study, precipitation for the 9 cities in Turkey's Marmara region is examined based on the 51-year (1969–2019) historical data and the ARMA, ARIMA, and SARIMA models are used to predict the precipitation in the next 60 months (up to 2024). While determining the model, the lowest AIC (Akaike information criterion) and AIC_c (corrected Akaike information criterion) are preferred and, generally, the AIC_c value is used to select the prediction model. After, the forecast measure errors of the models are checked with mean absolute error (MAE), root mean squared error (RMSE), and mean absolute scaled error (MASE) indicators. Finally, the ARIMA model is chosen as the most suitable model with the lowest estimation error.

Keywords Forecasting · Time series · Precipitation · Marmara region · Turkey

Introduction

Water sources are one of the important natural resources (Chan 2012; Tan et al. 2017). In the last centuries, significant changes and developments have arisen in the world (Kandemir and Ozbay 2013). Rainfall is an important parameter to adjust the condition of the climate system and has a high potential that affects the environment. Precipitation is a discontinuous climate parameter associated with different forms (Wang et al. 2017).

Climate change has been an important concern for people as it influences their livelihood and living environments to a significant extent. Due to its risks and effects, many studies have been conducted to climate change and the ways to mitigate its impacts. According to the results, it is predicted that the temperature will rise and precipitation will be more intense and less (Strauch et al. 2015; Kristo et al. 2017).

Coastal areas are among the most significant natural formations and more than a one of third of the people in the world live here (Solomon et al. 2007). These areas are affected at bigger risk from present and probable climate change and weather extreme events. The mean temperature has exposed a 0.85 °C rise over the last century (Toros et al. 2019). One of the coastal regions of Turkey is the Marmara region. The Marmara region covers 40% of the Turkey's population that manufacturing, commerce, and cultivation are the main economic resources for the region. The regions' air is intensely affected by unplanned urbanization and building of high constructions (Kahya et al. 2017).

Forecasting methods are often used in positive science. Some models have been developed (Uzun and Yaylı 2020). In the literature, besides the fields such as sales and production, forecasting methods are used even in basic sciences and

This article is part of the Topical Collection on *Geo-Resources-Earth-Environmental Sciences*

✉ Suheyla Yerel Kandemir
syerel@gmail.com

[1] Department of Industrial Engineering, Bilecik Şeyh Edebali University, Bilecik, Turkey

[2] Department of Geography, Bilecik Şeyh Edebali University, Bilecik, Turkey

materials science (Hong et al. 2020; Arslankaya 2020a; Arslankaya 2020b). Some studies in which forecasting methods have been applied for meteorological data in recent years are given below.

Dimri et al. (2020) investigated the seasonal analysis of monthly average minimum and maximum temperatures and precipitation for the Bhagirathi river basin in Uttarakhand state of India. For this purpose, they used 100 years of precipitation data. In addition to time series analysis, they tested the compatibility between methods using the Seasonal Autoregressive Integrated Moving Average (SARIMA) model (Dimri et al. 2020). Zeydalinejad et al. (2020) predicted the climate variables of Lali region in southwestern Iran. They used Artificial Neural Networks (ANN) as a forecasting method (Zeydalinejad et al. 2020). Praveen et al. (2020) analyzed and predicted the changes in precipitation using space data from 1901 to 2015 in India. For this purpose, they used Mann-Kendall (MK) test and Sen's innovative trend analysis. Artificial Neural Network-Multilayer Sensor (ANN-MLP) was used to predict precipitation for the next 15 years (Praveen et al. 2020). Achour et al. (2020) conducted a drought analysis using the Standardized Precipitation Index (SPI) for the northwestern region of Algeria. Drought prediction has been made with Artificial Neural Networks (ANN). The success of the method was tested using root mean squared error (RMSE) and mean absolute error (MAE) criteria (Achour et al. 2020). Khan et al. (2020) proposed a new hybrid model to predict future drought in their studies. The proposed model includes wavelet transform, Autoregressive Integrated Moving Average (ARIMA), and ANN. In the study, rainfall data of 30 years from 1986 to 2016 were analyzed for the Langat River Basin of Malaysia (Khan et al. 2020).

There are many studies in the literature comparing models created for fields such as physics, basic mathematical sciences, and nanotechnology according to different criteria (Yaylı et al. 2016). The importance of this study is to be able to predict precipitation for preventive plans that can be developed for global warming and climate change problems and to reveal the most suitable forecasting methods that can be used.

Time series analysis is a significant tool that allows forecasting the amount of future precipitation. In this paper, precipitation for the Marmara region (Turkey) is examined based on the 51-year historical data and to predict the precipitation in the next 60 months. For this purpose, the methodology of forecasting methods used in the study is given in the "Materials and methods" section of the study. Equations and parameters used in the methods are defined in this section. The performance evaluation criteria used to evaluate the forecasting methods are summarized. The results obtained by applying the forecasting methods to the 51-year average precipitation values in the Marmara region are included in the "Results and discussions" section of the study. In the study, it is aimed to propose the most advantageous method for estimation according to the calculated values of performance evaluation criteria. The ARIMA method, one of the forecasting methods, has been obtained as an advantageous method in terms of estimation success. Discussion takes place in the "Conclusion" section of the study.

Materials and methods

Study area

Marmara region, in northwest Turkey, is situated around the Sea of Marmara. Economic and social care from Turkey, which is the most developed region in Marmara region, is hosting a population of 25 million (https://www.icisleri.gov.tr/). There are 11 provinces in the Marmara region. Istanbul is one of the most important industrial and commercial centers of Turkey. The Thrace basin of the Marmara region has suitable lands for agriculture as well as for industry. Agriculture and animal husbandry are the livelihoods of this region. The Marmara region has also developed in the tourism sector with its natural and historical beauties. While Çatalca Kocaeli Department, including Istanbul, has been developed in terms of industry-service sectors, South Marmara and Ergene departments are the departments where agricultural activities for the market have developed as well as these sectors. Therefore, the rainfall conditions that may occur due to changes in both global warming and climate change may adversely affect Turkey in terms of both economic and social. In this study, located in Marmara region, long-term precipitation data (1969–2019) received from Kırklareli, Edirne, Tekirdağ, Istanbul, Kocaeli, Sakarya, Bilecik, Bursa, Yalova, Balıkesir, and Çanakkale meteorology stations were evaluated, and future rainfall amount estimations were made. Precipitation forecasting is important for water planning to meet the agricultural irrigation and consequently the food demand in the Marmara region, where the urban population density is high (Fig. 1).

Time series models

Generally, prediction models can have different forms for time series models. Many time series modeling is done on a linear basis (Mirzavand and Ghazavi 2015). The Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), and Seasonal Autoregressive Integrated Moving Average (SARIMA) estimation models are used in this study. After, the forecasting models are tested with MAE, RMSE, and mean absolute scaled error (MASE) model evaluation criteria.

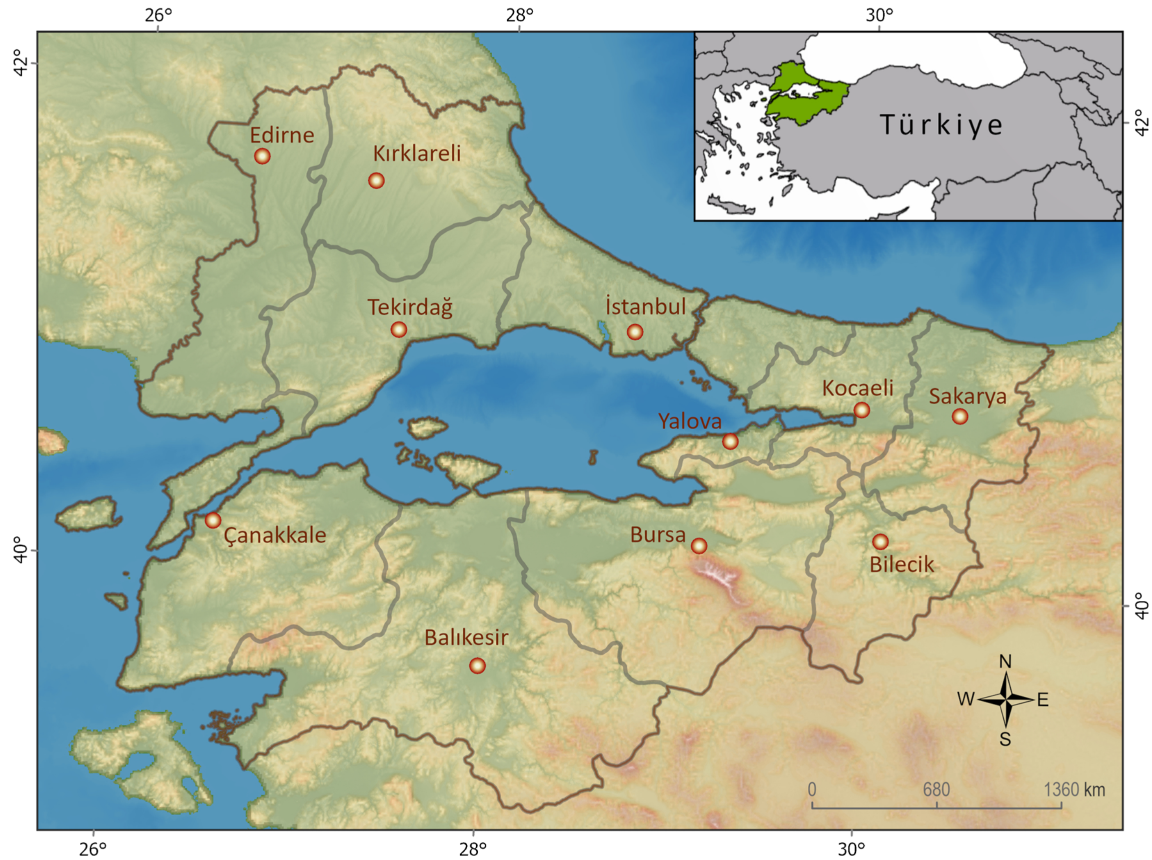

Fig. 1 Map of Marmara region

Autoregressive Moving Average models

ARMA model stands for Autoregressive Moving Average model. The method is also called the Box-Jenkins method as developed by George Box and Gwilym Jenkins in 1976 (Nugroho and Simanjuntak 2014). It basically consists of a combination of autoregressive (AR) and moving average (MA) models (Bang et al. 2019).

The ARMA models consider the error term to predict future conditions based on previous data (Babu and Reddy 2014; Wagena et al. 2020). It is one of the statistical models used in precipitation estimation (Collischonn et al. 2007; Liu and Han 2012; Duangdai and Likasiri 2017).

An ARMA (p, q) model can also be represented by its total notation as in Eq. (1) (Dastorani et al. 2016).

$$x(t) = \delta + \sum_{i=1}^{p} \theta_i x_{(t-1)} + \sum_{j=1}^{q} \varphi_j e_{(t-1)} + e_{(t)} \quad (1)$$

where δ is the stationary part of the ARMA model; θ_i points out the autoregressive coefficient; $e(t)$ is the error term; φ_j is the jth moving average coefficient; it shows the error part at time period t; and $x(t)$ refers the value of precipitation observed or predicted at time period t (Behnia and Rezaeian 2015).

Autoregressive Integrated Moving Averages models

ARIMA is an Autoregressive Integrated Moving Averages model put forward by Box et al. (1970) (Dawood et al. 2020). The ARIMA model, one of the Box and Jenkins methods, includes the approaches to identify, fit, and control single variable models (Lu and AbouRizk 2009; Balasmeh et al. 2019). The ARIMA model is an important statistical approach in forecasting and analyzing time series data related to precipitation (Wang et al. 2014; Balibey and Turkyilmaz 2015; Klaus et al. 2015; Feng et al. 2016).

In practice, the ARMA is considered as ARIMA without the need for an ARMA differentiation process for fixed serialization. That is, the ARMA model can be written as ARIMA (p, d, q) which is more common where p is the order of the autoregressive process, q is the order of the moving average process, and d is the differentiation process in the case of ARMA being 0, so ARMA models are often written as ARIMA $(p, 0, q)$ (Nugroho and Simanjuntak 2014).

In general, the ARIMA model for time series data can be modeled as ARIMA (p, d, q) and is expressed as Eq. (2).

$$\mathrm{x(t)} = \delta + \varphi(1)\mathrm{x(t-1)} + \varphi(2)\mathrm{x(t-2)} + \varphi(3)\mathrm{x(t-3)} \dots + \varphi(\mathrm{p})\mathrm{x(t-p)} + \mathrm{e(t)} - \theta(1)\mathrm{e(t-1)} - \theta(2)\mathrm{e(t-2)} - \theta(3)\mathrm{e(t-3)} \dots - \theta(\mathrm{q})\mathrm{e(t-q)} \quad (2)$$

where δ is the stationary part of the ARIMA model, $x(t)$ and $e(t)$ are the actual time series and white noise (error), and φ and θ are the model parameters, where it is distributed with mean zero and constant variance (Balasmeh et al. 2019).

Seasonal Autoregressive Integrated Moving Average models

The SARIMA model reflecting the feature of seasonal variation in time series can be divided into a simple model and multiple models (Wang et al. 2013). Since it is seasonally based, the SARIMA model has been used in most studies for precipitation estimation (Wang et al. 2013; Han et al. 2013; Park et al. 2018; Parviz and Rasouli 2019).

The SARIMA model notation is written as SARIMA (p, d, q) x $(P, D, Q)s$**.** The general form of multiplicative seasonal model SARIMA (p, d, q) x $(P, D, Q)s$ is given by Eq. (3) (Murthy et al. 2018).

$$\Phi\mathrm{p}\ (\mathrm{Bs})\ \Phi\mathrm{p}\ (\mathrm{B})\ \nabla_s^D \nabla\mathrm{d}\ \mathrm{xt} = \mu + \Theta\mathrm{Q}(\mathrm{Bs})\ \theta\mathrm{q}\ (\mathrm{B})\ \mathrm{at} \quad (3)$$

where at is the nonstationary time series and it is the usual Gaussian white noise process; s is the period of the time series; B is the backshift operator; Φp (Bs) is the seasonal autoregressive operator of order P; Φp (B) is the regular autoregressive operator of order p; $Q(Bs)$ is the seasonal moving average operator of order Q; and $\theta q(B)$ is the regular moving average operator of order q (Chang et al. 2012; Murthy et al. 2018).

Model performance criteria

Performance efficiency of prediction models can be evaluated using model insides such as MAE, RMSE, and MASE (Mohanasundaram et al. 2017). In this study, evaluation indicators MAE, RMSE, and MASE are determined to measure the performance of the prediction models. In addition, Akaike information criterion (AIC) was used as estimator of out-of-sample prediction error and thereby relative quality of statistical models for a given set of data.

Root mean squared error RMSE is a measure of the deviation between the model's estimated values and actual values (Todorovski and Džeroski 2003). In general, high RMSE value indicates low performance. However, comparisons on different plot values have no validity (Wang et al. 2019). Thus, RMSE can be expressed as Eq. (4).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_t' - y_t\right)^2} \quad (4)$$

where y_t' is the predicted value, y_t is the actual observed value, and n is the quantity of observations.

Mean absolute scaled error MASE scales the absolute average error regardless of the scale of the data. If a proportional error is derived from a prediction that is better than that for the average one-step baseline prediction, then the proportional error is less than 1; otherwise, the value is greater than 1. The MASE is expressed as Eq. (5) (Wang et al. 2019).

$$\mathrm{MASE} = \mathrm{mean}\left(\left|q_j\right|\right) \quad (5)$$

The q_j formulation in the equation is given in Eq. (6).

$$q_j = \frac{e_j}{\frac{1}{T-m}\sum_{t=m+1}^{T}\left|y_t - y_{t-m}\right|} \quad (6)$$

where e_j is the difference between the predicted value and the actual value at time j, y_t is the actual value at time t, m is the seasonal period, and $y_{t-m}$is the actual value at time $t-m$ (Wang et al. 2019).

Mean absolute error MAE is obtained by dividing the total error absolute value by the quantity of observations**.** In general, high MAE value indicates low performance. MAE can be expressed as Eq. (7) (Bruster-Flores et al. 2019).

$$\mathrm{MAE} = \frac{1}{n}\sum_{\mathrm{I}=1}^{n}\left|y_t' - y_t\right| \quad (7)$$

where y_t' is the predicted value, y_t is the actual observed value, and n is the quantity of observations.

Akaike's information criterion Akaike information criterion (AIC) ensures the best distribution in a particular set of distributions by relative comparisons with one another. AIC's formulation is given Eq. (8).

$$AIC = 2k - 2.ln(L) \quad (8)$$

where k is the number of parameters of a probability distribution and L is the maximized value of the likelihood function for the distribution (Monish and Rehana 2020).

AIC can be expressed in another notation as in Eq. (9).

$$AIC = Tlog\left(\frac{SSE}{T}\right) + 2(k+2) \quad (9)$$

Table 1 AIC_c values for an alternative ARMA (*p, q*) model

AIC_c	MA0	MA1	MA2	MA3	MA4	MA5
AR0	6064.98	6032.83	6026.4	6028.38	6029.83	6018.61
AR1	6027.56	6029.22	6028.4	6016.45	6015.12	6010.32
AR2	6028.92	6029.43	5897.99	5900.43	5902.48	5899.94
AR3	6025.57	6002.43	5899.67	5902.12	5903.91	5904.97

T is the number of observations used for prediction and *k* is the number of predictors in the model. The $k+2$ part of the equation represents the $k+2$ parameters in the model.

The purpose of the calculation of the mentioned criterion is to punish the fit of the model with the number of parameters to be estimated. The model with the smallest AIC value usually points to the best model for prediction. Minimizing the AIC for large T values is equivalent to minimizing the cross-validation (CV) value that defines the predictive ability of the model.

According to these definitions, AIC can be expressed in Eq. (10).

$$AIC = -2\log(L) + 2(p+q+k+1) \tag{10}$$

L is the likelihood of data. If $c \neq 0$, then $k=1$, and if $c=0$, then $k=0$.

Corrected Akaike's information criterion For small T values, AIC tends to choose too many predictors, so a bias-corrected version of AIC has been developed. Formulation of AIC_c can be expressed in Eq. (11).

$$AIC_c = AIC + \frac{2(k+2)(k+3)}{T-k-3} \tag{11}$$

As with AIC, the lowest AIC_c value indicates the best model. The last term in parentheses refers to the number of parameters in the model. Equation notation can be generated as Eq. (12).

$$AIC_c = AIC + \frac{2(p+q+k+1)(p+q+k+2)}{T-p-q-k-2} \tag{12}$$

Information criteria tend to be a good guideline for choosing only the *p* and *q* values, not for choosing the appropriate order of difference for a model. This is because these information criteria change the data that calculates the probability of difference and cannot compare AIC values between different ordered models. Therefore, a different approach should be used to determine *d*, and then AIC_c should be used to select *p and q*.

Results and discussions

In this section, the approaches used in modeling time series are evaluated. Precipitation for the 9 cities in Marmara region is examined and the precipitation in the next 60 months is predicted. After, the created models are tested with RMSE, MASE, and MAE forecasting accuracy tools.

Firstly, the ARMA model is used in the forecasting step. AIC_c values for the alternative ARMA (*p*, *q*) model are included in Table 1. The ARMA (2, 2) model with the lowest AIC_c value (5897.99) is chosen as the most suitable model. Autocorrelation Function (ACF) plot is merely a bar chart of

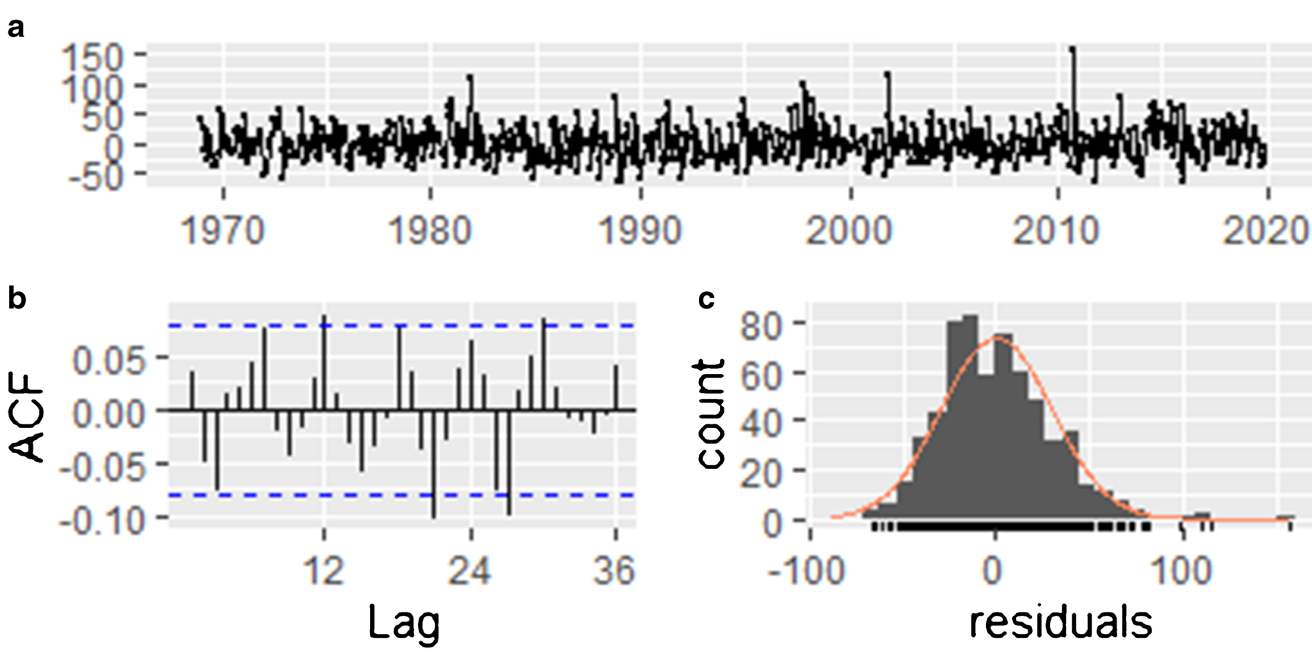

Fig. 2 **a** The time plot of the ARMA model series. **b** The ACF plot of the residuals from the ARMA (2,2) model. (c) The distribution of residuals from the ARMA (2,2) model

Table 2 AIC_c values for an alternative ARMA ($p,1,q$) model

AIC_c	MA0	MA1	MA2	MA3	MA4	MA5
AR0	6300.63	6062.49	6030	6023.37	6025.37	6026.97
AR1	6198.82	6024.56	6026.18	6025.38	6014.29	6012.91
AR2	6178.53	6025.85	6026.37	6027.45	6029.46	6018.14
AR3	6163.29	6022.69	6029.13	6026.26	5900.57	5903.15
AR4	6159.59	6019.52	5997.11	6004.19	5903.32	5904.35
AR5	6152.42	6005.69	5989.42	**5897.16**	5902.69	5905.09

the coefficients of correlation between a time series and lags of itself. The results of the ARMA (2, 2) graphs are given in Fig. 2. Figure 2a is the plots of series. This plot shows some sudden changes. It can be seen from Fig. 2a that there were some particularly changes in precipitation amounts in the early 2000s and 2010s. Figure 2b shows that all autocorrelations are within the threshold limits, indicating that the residuals are behaving like white noise. Figure 2c is obtained from portmanteau test, and the portmanteau test returns a large p value, also suggesting that the residuals are white noise.

The (p, d, q) combination that gives the lowest AIC_c value is searched to obtain the most suitable ARIMA model. Simulations show that high d values cause to increase the AIC_c values. The AIC_c values of the combinations based on $(p, 1, q)$ are shown in Table 2. In this study, the ARIMA (5, 1, 3) model with the lowest AIC_c value (5897.16) is preferred as the most suitable model.

The results of the ARIMA (5, 1, 3) graphs are given in Fig. 3. Figure 3a is the plots of model series. This plot shows some sudden changes. It can be seen from the figure that there were some particularly changes in precipitation amounts in the early 1980s, 2000s, and 2010s. Figure 3b shows that all autocorrelations are within the threshold limits, indicating that the residuals are behaving like white noise. Figure 3c is obtained from portmanteau test, and the portmanteau test returns a large p value, also suggesting that the residuals are white noise.

The ARIMA (5, 1, 3) model has a large p value (0.1331) than the AR (1), MA (1) , and ARMA (2, 2) models. If all the autocorrelation values are within the limits (blue dashed lines) in ACF plot, it indicates that the residuals behave like white noise. ACF plot of residuals also shows that the residuals are partially in the blue dashed lines.

The seasonal ARIMA model is created by adding seasonal terms to the ARIMA models. The $(p, d, q)(P, D, Q)[s]$ combination that gives the lowest AIC_c value is searched to obtain the most suitable SARIMA model. The evaluation is carried out by observing the change of other parameters according to constant d and D values. The SARIMA models are sorted by AIC_c values and the model with the lowest AIC value is chosen as the most suitable model. The SARIMA (0,1,2) (5,2,1) [12] model that has the lowest AIC_c value (5639.488) is preferred as the most suitable model. The AIC_c values of the combinations based on (p, d, q) (P, D, Q) are shown in Table 3.

The results of the SARIMA (0, 1, 2) (5, 2, 1) [12] graphs are given in Fig. 4. Figure 4a is the plots of series. This plot shows some sudden changes. It can be seen from the figure that there were some particular changes in precipitation amounts in the 1975s, 1995s, and early 2010s. Figure 4b shows that all autocorrelations are within the threshold limits,

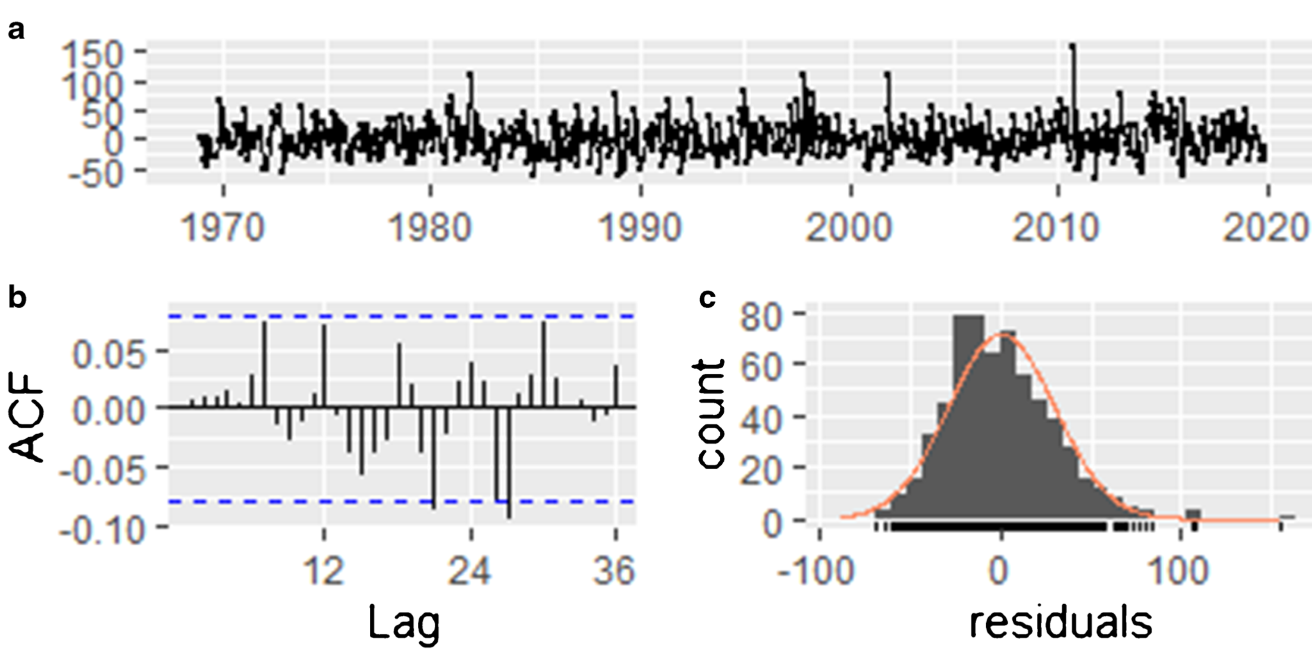

Fig. 3 **a** The time plot of the ARIMA model series. **b** The ACF plot of the residuals from the ARIMA (5,1,3) model. **c** The distribution of residuals from the ARIMA (5,1,3) model

Table 3 AIC$_c$ values for an alternative SARIMA model

Models	AIC$_c$	Models	AIC$_c$
ARIMA (0,1,2) (5,2,1) [12]	5639.488	ARIMA (0,0,0) (0,1,1) [12]	5727.524
ARIMA (0,1,1) (5,2,1) [12]	5640.895	ARIMA (0,0,0) (0,1,2) [12]	5728.914
ARIMA (0,1,3) (5,2,1) [12]	5641.312	ARIMA (0,1,1) (1,1,2) [12]	5730.433
ARIMA (0,1,2) (4,2,1) [12]	5685.766	ARIMA (0,1,1) (1,1,1) [12]	5731.642
ARIMA (0,1,1) (4,2,1) [12]	5688.6	ARIMA (0,1,1) (1,1,3) [12]	5732.359
ARIMA (0,0,0) (1,1,1) [12]	5719.816	ARIMA (0,1,2) (1,1,2) [12]	5732.37
ARIMA (0,0,0) (2,1,1) [12]	5720.602	ARIMA (0,1,1) (2,1,2) [12]	5742.982
ARIMA (0,0,1) (1,1,1) [12]	5721.567	ARIMA (1,1,1) (1,1,2) [12]	5743.668
ARIMA (1,0,0) (1,1,1) [12]	5722.019	ARIMA (0,1,1) (0,1,1) [12]	5744.766
ARIMA (4,2,2) (2,1,2) [12]	5723.765	ARIMA (0,1,1) (0,1,2) [12]	5746.577

indicating that the residuals are behaving like white noise. Figure 4c is obtained from portmanteau test, and the portmanteau test returns a large p value, also suggesting that the residuals are white noise.

The SARIMA (0,1,2) (5,2,1) [12] model has the largest p value (0.6807) among all models. All autocorrelation values are within the limits in ACF plot that means the residuals behave like white noise. The estimated values obtained from the ARMA (2,2), ARIMA (5,1,3), and SARIMA (0,1,2) (5,2,1) [12] models with a prediction interval of 80% and 95% for 2020–2024 are shown in Fig. 5a, b, and c which show graph of estimated precipitation values, respectively.

Future forecasts have a generally similar view as in Fig. 5. Seasonal effects are observed in the SARIMA model. Models that are like each other are suitable as a prediction algorithm, but the models that can give the best results can be reached with model evaluation criteria. MAE, RMSE, and MASE are calculated from the success evaluation criteria according to the estimation results obtained with the forecasting models. These values are given in Table 4.

The MAE, RMSE, and MASE values obtained with the ARMA (2,2) model are calculated as 23.3117, 29.4888, and 0.7542, respectively. These values of criteria obtained with the ARIMA (5,1,3) model are calculated as 23.0641, 29.4330, and 0.7417, respectively. These values of criteria obtained with the SARIMA (0,1,2) (5,2,1) [12] model are calculated as 23.1118, 30.6224, and 0.7417, respectively. As can be seen, close results are obtained from each performance criteria according to the forecasting methods. The methods give extremely sensitive results. The similarity in the forecasting algorithms is an important reason for this situation. For the

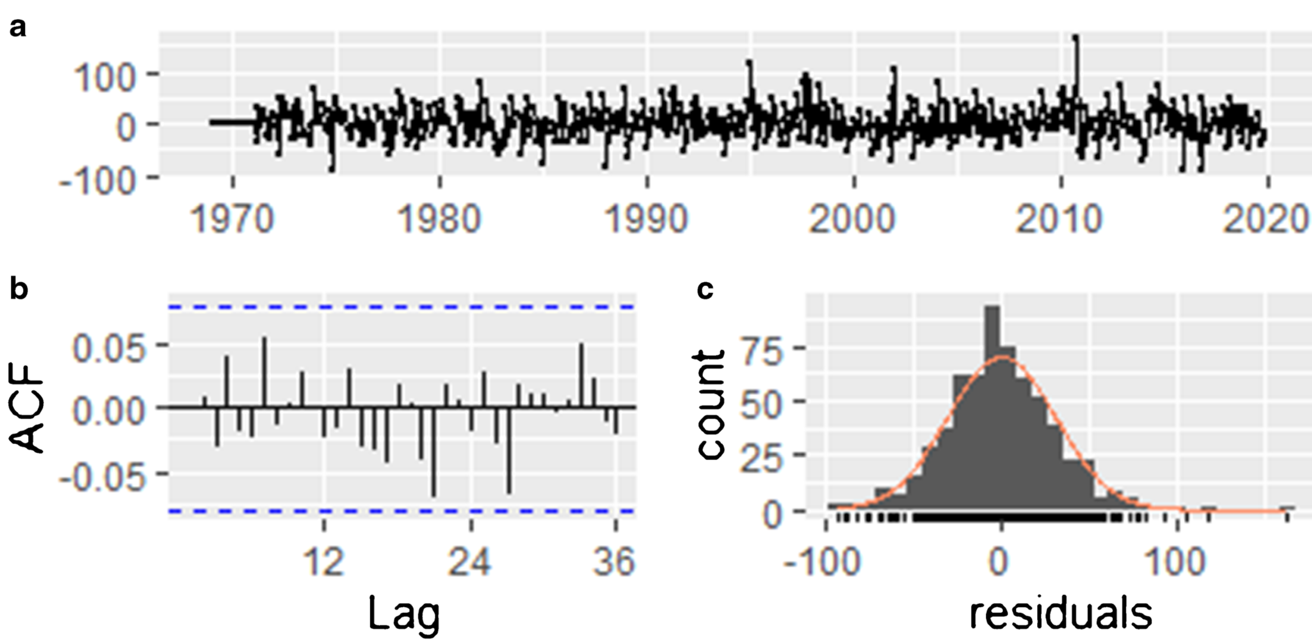

Fig. 4 **a** The time plot of the SARIMA model series. **b** The ACF plot of the residuals from the SARIMA (0,1,2) (5,2,1) [12] model. **c** The distribution of residuals from the SARIMA (0,1,2) (5,2,1) [12] model

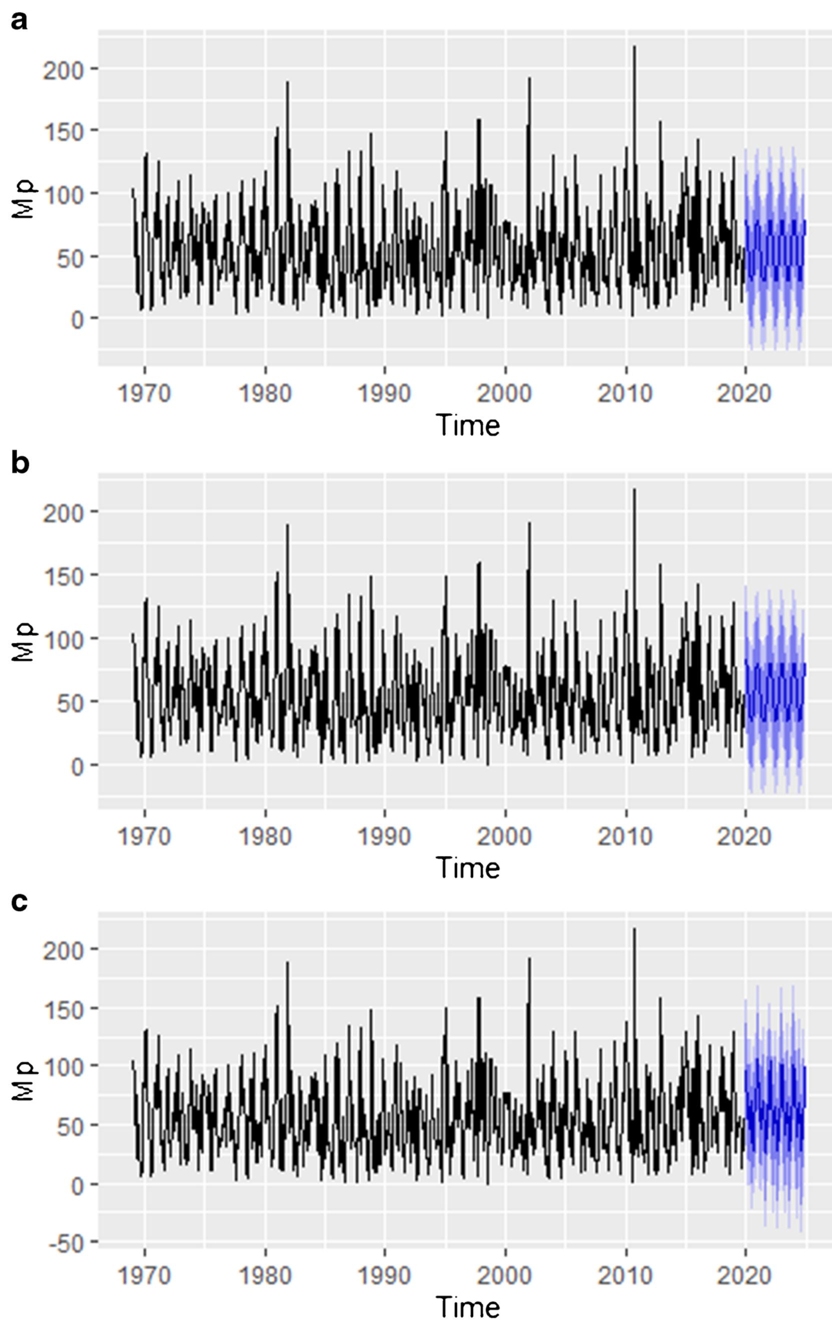

Fig. 5 **a** Graph of estimation values obtained from the ARMA (2,2) model. **b** Graph of estimation values obtained from ARIMA (5,1,3) model. **c** Graph of estimation values obtained from SARIMA (0,1,2) (5,2,1) [12] model

decision-makers, the forecasting method that can minimize the MAE, RMSE, and MASE values is considered successful. According to the obtained values in Table 4, the ARIMA (5,1,3) model can be evaluated as the best model in terms of MAE, RMSE, and MASE criteria. The SARIMA (0,1,2) (5,2,1) [12] model has a similar performance to the ARIMA (5,1,3) model in terms of the MASE criterion. In this context, the SARIMA (0,1,2) (5,2,1) [12] model is the closest alternative to the ARIMA (5,1,3) model in terms of the convenience of the forecasting methods.

Table 4 shows that forecasting models give very close results in terms of model performance criteria. According to the obtained values, the ARIMA (5, 1, 3) model can be evaluated as the best model in terms of MAE, RMSE, and MASE criteria. The SARIMA model has a similar performance to the ARIMA model in terms of the MASE criterion.

Table 4 Calculated model performance criteria

Forecasting models	MAE	RMSE	MASE
ARMA (2,2)	23.3117	29.4888	0.7542
ARIMA (5,1,3)	23.0641	29.4330	0.7417
SARIMA (0,1,2) (5,2,1) [12]	23.1118	30.6224	0.7417

Conclusion

The Marmara region is the most densely populated and urbanized region in Turkey. This socio-spatial concentration in the region is influenced by agricultural development in addition to highly developed industrial and service sector. It is important to make low error precipitation forecast in planning the water demanded in the urban population's drinking, usage, and industrial production. More importantly, low error precipitation forecast should be made in planning the water needed by the agricultural sector in food production. Global warming, which occurs as a result of intense economic activities, and the resulting climate change make it difficult to make precipitation forecast for long term as it causes extreme weather events. These critical reasons increase the importance of determining prospective precipitation forecast models.

In this study, the ARMA, ARIMA, and SARIMA methods are used to predict precipitation in Marmara region of Turkey. Precipitation forecasts for the next 60 months have been obtained with forecasting methods and a graphic representation has been presented. The similar results of model evaluation criteria indicate that the methods are sensitive, and the algorithm structures are similar. In this study, calculated values of MAE, RMSE, and MASE model performance criteria are compared for estimation methods. Among the ARMA, ARIMA, and SARIMA methods, the most favorable results were achieved with the ARIMA method. Since model evaluation criteria give better results, the Autoregressive Integrated Moving Average (ARIMA) model is determined as the most suitable precipitation forecast model for the next 60 months (between 2020–2024). The ARIMA model is one of the time series models that consider seasonality as an integral part of the modeling strategy. In this approach, seasonality is modeled by taking into account the statistical properties of the data.

In future studies, besides the precipitation parameter, different parameters such as solar radiation, air pollution, amount of snow, and global warming indicators can be included in the analysis. Relationships between variables can be determined by various statistical methods. Time series analysis can be done with similar estimation methods. Estimation methods can be used as hybrid and different model evaluation criteria can be suggested for the success of the prediction model.

References

Achour K, Meddi M, Zeroual A, Bouabdelli S, Maccioni P, Moramarco T (2020) Spatio-temporal analysis and forecasting of drought in the plains of northwestern Algeria using the standardized precipitation index. J Earth Syst Sci 129(1):42. https://doi.org/10.1007/s12040-019-1306-3

Arslankaya S (2020a) Estimation of hanging and removal times in eloxal with artificial neural networks. Emerg Mater Res 9(2):366–374. https://doi.org/10.1680/jemmr.19.00191

Arslankaya S (2020b) Estimating the effects of heat treatment on aluminum alloy with artificial neural networks. Emerg Mater Res 9(2): 540–549. https://doi.org/10.1680/jemmr.20.00059

Babu CN, Reddy BE (2014) A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. Appl. Soft Comput 23:27–38. https://doi.org/10.1016/j.asoc.2014.05.028

Balasmeh A, Babbar OR, Karmaker T (2019) Trend analysis and ARIMA modeling for forecasting precipitation pattern in Wadi Shueib catchment area in Jordan. Arab J Geosci 12:27. https://doi.org/10.1007/s12517-018-4205-z

Balibey M, Turkyilmaz S (2015) A time series approach for precipitation in Turkey. Gazi Univ J Sci 28:549–559 Corpus ID: 124084732

Bang S, Bishnoi R, Chauhan AS, Dixit AK, Chawla I (2019) Fuzzy logic based crop yield prediction using temperature and rainfall parameters predicted through ARMA, SARIMA, and ARMAX models. 2019 Twelfth International Conference on Contemporary Computing (IC3) Noida India: 1-6. https://doi.org/10.1109/IC3.2019.8844901

Behnia N, Rezaeian F (2015) Coupling wavelet transform with time series models to estimate groundwater level. Arab J Geosci 8: 8441–8447. https://doi.org/10.1007/s12517-015-1829-0

Box GE, Jenkins GM, Reinsel G (1970) Time series analysis: forecasting and control. BoxTime series analysis: forecasting and control. Holden-day, San Francisco

Bruster-Flores JL, Ortiz-Gómez R, Ferriño-Fierro AL, Guerra-Cobián VH, Burgos-Flores D, Lizárraga-Mendiola LG (2019) Evaluation of precipitation estimates CMORPH-CRT on regions of Mexico with different climates. Water 11(8):1722. https://doi.org/10.3390/w11081722

Chan NW (2012) Managing urban rivers and water quality in Malaysia for sustainable water resources. Int J Water Resour D 28(2):343–354. https://doi.org/10.1080/07900627.2012.668643

Chang X, Gao M, Wang Y, Hou X (2012) Seasonal autoregressive integrated moving average model for precipitation time series. J Math Stat 8:500–505

Collischonn W, Tucci CEM, Clarke RT, Chou SC, Guilhon LG, Catald M, Allasia D (2007) Medium-range reserv oirinflow predictions based on quantitative precipitation forecasts. J Hydrol 344:112–122. https://doi.org/10.1016/j.jhydrol.2007.06.025

Dastorani M, Mirzavand M, Dastorani MT, Sadatinejad SJ (2016) Comparative study among different time series models applied to monthly rainfall forecasting in semi-arid climate conditionâ. Nat Hazards 81:1811–1827. https://doi.org/10.1007/s11069-016-2163-x

Dawood M, Rahman A, Ullah S, Mahmood S, Rahman G, Azam K (2020) Spatio-statistical analysis of rainfall fluctuation, anomaly and trend in the Hindu Kushregion using ARIMA approach. Nat Hazards 101:449–464. https://doi.org/10.1007/s11069-020-03881-5

Dimri T, Ahmad S, Sharif M (2020) Time series analysis of climate variables using seasonal ARIMA approach. J Earth Syst Sci 129(1):1–16. https://doi.org/10.1007/s12040-020-01408-x

Duangdai E, Likasiri C (2017) Rainfall model investigation and scenario analyses of the effect of government reforestation policy on seasonal

rainfalls: a case study from Northern Thailand. Atmos Res 185:1–12. https://doi.org/10.1016/j.atmosres.2016.10.019
Feng G, Cobb S, Abdo Z, Fisher DK, Ouyang Y, Adeli A, Johnie NJ (2016) Trend analysis and forecast of precipitation, reference evapotranspiration, and rainfall deficit in the Blackland Prairie of eastern Mississippi. J Appl Meteorol Climatol 55(7):1425–1439. https://doi.org/10.1175/JAMC-D-15-0265.1
Han P, Wang P, Tian M, Zhang S, Liu J, Zhu D (2013) Application of the ARIMA models in drought forecasting using the standardized precipitation ındex. In: Li D, Chen Y (eds) Computer and Computing Technologies in Agriculture VI. CCTA 2012, IFIP Advances in Information and Communication Technology, vol 392. Springer, Berlin, Heidelberg
Hong F, Qiao H, Wang P (2020) Predicting the life of BNC-coated reinforced concrete using the Weibull distribution. Emerg Mater Res:1–11. https://doi.org/10.1680/jemmr.19.00087
Kahya C, Balcik FB, Oztaner YB, Ozcomak D, Seker DZ (2017) Spatio temporal analysis of $PM_{2.5}$ over Marmara region, Turkey. Fresenius Environ Bull 26(1):310–317
Kandemir SY, Ozbay N (2013) Assessment of climate changes: a case study from Bilecik, Turkey. J Selcuk Univ Nat Appl Sci ICOEST Conf. 2013 (Part 2):431–435
Khan MMH, Muhammad NS, El-Shafie A (2020) Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. J Hydrol 590:125380. https://doi.org/10.1016/j.jhydrol.2020.125380
Klaus J, Chun KP, Stumpp C (2015) Temporal trends in δ18O composition of precipitation in Germany: insights from time series modelling and trend analysis. Hydrol Process 29:2668–2680. https://doi.org/10.1002/hyp.10395
Kristo C, Rahardjo H, Satyanaga A (2017) Effect of variations in rainfall intensity on slopest ability in Singapore. Int Soil Water Conserv Res 5(4):258–264. https://doi.org/10.1016/j.iswcr.2017.07.001
Liu J, Han D (2012) On selection of the optimal data time interval forreal-time hydrological forecasting. Hydrol Earth Syst Sci Discuss 9: 10829–10875. https://doi.org/10.5194/hess-17-3639-2013
Lu Y, AbouRizk SM (2009) Automated Box–Jenkins forecasting modelling. Autom Constr 18(5):547–558. https://doi.org/10.1016/j.autcon.2008.11.007
Mirzavand M, Ghazavi R (2015) A stochastic modelling technique for ground water level forecasting in an arid environment using time series methods. Water Resour Manag 81:1811–1827. https://doi.org/10.1007/s11269-014-0875-9
Mohanasundaram A, Narasimhan B, Kumar GS (2017) Transfer function noise modelling of ground water level fluctuation using threshold rainfall-based binary-weighted parameter estimation approach. Hydrol Sci J 62(1):36–49. https://doi.org/10.1080/02626667.2016.1171325
Monish NT, Rehana S (2020) Suitability of distributions for standard precipitation and evapotranspiration index over meteorologically homogeneouszones of India. J Earth Syst Sci 129(25):1–19. https://doi.org/10.1007/s12040-019-1271-x
Murthy KVN, Saravana R, Kumar KV (2018) Modeling and forecasting rainfall patterns of southwest monsoons in North–East India as a SARIMA process. Meteorog Atmos Phys 130:99–106. https://doi.org/10.1007/s00703-017-0504-2
Nugroho A, Simanjuntak BH (2014) ARMA (Autoregressive Moving Average) model for prediction of rainfall in regency of Semarang-Central Java-Republic of Indonesia. Int J Comput Sci Issues (IJCSI) 11(3):27–32 https://search.proquest.com/docview/1543263434?accountid=16382 . Accessed 17 June 2020
Park J, Onof C, Kim D (2018) A hybrid stochastic rainfall model that reproduces rainfall characteristics at hourly through yearly time scale. Hydrol Earth Syst Sci 23:989–1014
Parviz L, Rasouli K (2019) Development of precipitation forecast model based on artificial ıntelligence and subseasonal clustering. J Hydrol Eng 24(12):1–13. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001862
Praveen B, Talukdar S, Mahato S, Mondal J, Sharma P, Islam AR, Rahman A (2020) Analyzing trend and forecasting of rainfall changes in India using non-parametrical and machine learning approaches. Sci Report 10(1):1–21. https://doi.org/10.1038/s41598-020-67228-7
Solomon S, Manning M, Marquis M, Qin D (2007) Climate change 2007-the physical science basis: working group I contribution to the fourth assessment report of the IPCC. Cambridge University Press, Cambridge
Strauch AM, Mackenzie RA, Giardina CP, Bruland GL (2015) Climate driven changes to rainfall and stream flow patterns in a model tropical island hydrological system. J Hydrol 523:160–169. https://doi.org/10.1016/j.jhydrol.2015.01.045
Tan ML, Yusop Z, Chua VP, Chan NW (2017) Climate change impacts under CMIP5 RCP scenarios on water resources of the Kelantan River Basin, Malaysia. Atmos Res 189:1–10. https://doi.org/10.1016/j.atmosres.2017.01.008
Todorovski L, Džeroski S (2003) Combining classifiers with meta decision trees. Mach Learn 50:223–249. https://doi.org/10.1023/A:1021709817809
Toros H, Mokari M, Abbasnia M (2019) Regional variability of temperature extremes in the maritime climate of Turkey: a case study to develop agricultural adaptation strategies under climate change. Model Earth Syst Environ 5(3):857–865. https://doi.org/10.1007/s40808-019-00572-4
Uzun B, Yaylı MÖ (2020) Nonlocal vibration analysis of Ti-6Al-4 V/ZrO 2 functionally graded nanobeam on elastic matrix. Arab J Geosci 13(4):1–10. https://doi.org/10.1007/s12517-020-5168-4
Yaylı MÖ, Kandemir SY, Çerçevik AE (2016) A practical method for calculating eigenfrequencies of a cantilever microbeam with the attached tip mass. J Vibroeng 18(5):3070–3077. https://doi.org/10.21595/jve.2016.16636
Wang XL, Xu H, Qian B, Feng Y, Mekis E (2017) Adjusted daily rainfall and snowfall data for Canada. Atmosphere-Ocean 55(3):155–168
Wagena MB, Goering D, Collick AS, Bock E, Fuka DR, Buda A, Easton ZM (2020) Comparison of short-term stream flow forecasting using stochastic time series, neural networks, process-based, and Bayesian models. Environ Model Softw 126. https://doi.org/10.1016/j.envsoft.2020.104669
Wang H, Wang C, Lin X, Kang J (2014) An improved ARIMA model for precipitation simulations. Nonlinear Process Geophys 21:1159–1168. https://doi.org/10.5194/npg-21-1159-2014
Wang S, Feng J, Liu G (2013) Application of seasonal time series model in the precipitation forecast. Math Comput Model 58(3-4):677–683. https://doi.org/10.1016/j.mcm.2011.10.034
Wang M, Wang H, Wang J, Liu H, Lu R, Duan T, Gong X, Feng S, Liu Y, Cui Z, Li C, Ma J (2019) A novel model for malaria prediction based on ensemble algorithms. PLoS ONE:14. https://doi.org/10.1371/journal.pone.0226910
Zeydalinejad N, Nassery HR, Shakiba A et al (2020) Prediction of the karstic spring flow rates under climate change by climatic variables based on the artificial neural network: a case study of Iran. Environ Monit Assess 192:375. https://doi.org/10.1007/s10661-020-08332-z https://www.icisleri.gov.tr/turkiyenin-nufus-haritasi (20.05.2020)