

Performance comparison of filtering methods on modelling and forecasting the total precipitation amount: a case study for Muğla in Turkey

Serdar Neslihanoglu, Ecem Ünal and Ceylan Yozgatligil

ABSTRACT

Condensed water vapor in the atmosphere is observed as precipitation whenever moist air rises sufficiently enough to produce saturation, condensation, and the growth of precipitation particles. It is hard to measure the amount and concentration of total precipitation over time due to the changes in the amount of precipitation and the variability of climate. As a result of these, the modelling and forecasting of precipitation amount is challenging. For this reason, this study compares forecasting performances of different methods on monthly precipitation series with covariates including the temperature, relative humidity, and cloudiness of Muğla region, Turkey. To accomplish this, the performance of multiple linear regression, the state space model (SSM) via Kalman Filter, a hybrid model integrating the logistic regression and SSM models, the seasonal autoregressive integrated moving average (SARIMA), exponential smoothing with state space model (ETS), exponential smoothing state space model with Box-Cox transformation-ARMA errors-trend and seasonal components (TBATS), feed-forward neural network (NNETAR) and Prophet models are all compared. This comparison has yet to be undertaken in the literature. The empirical findings overwhelmingly support the SSM when modelling and forecasting the monthly total precipitation amount of the Muğla region, encouraging the time-varying coefficients extensions of the precipitation model.

Key words | ETS, Kalman filter, NNETAR, precipitation, prophet, TBATS

Serdar Neslihanoglu (corresponding author)
Department of Statistics,
Eskisehir Osmangazi University,
TR-26040 Eskisehir,
Turkey
E-mail: sneslihanoglu@ogu.edu.tr

Ecem Ünal
Ceylan Yozgatligil
Department of Statistics,
Middle East Technical University,
TR-06800 Ankara,
Turkey

HIGHLIGHTS

- The modelling and forecasting of precipitation amount are difficult because of its highly parametrized and varied nature.
- The performances of filtering methods, namely the multiple linear regression, the state space model (SSM), hybrid, SARIMA, ETS, TBATS, NNETAR and Prophet models on monthly total precipitation amount are investigated.
- The results support SSM when modelling and forecasting the total precipitation amount.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

doi: 10.2166/wcc.2021.332

INTRODUCTION

One of the most common problems in the world is the remarkable changes observed in climate. The effects of these changes on Earth and human beings cannot be ignored if we want to create a more habitable future. There are many parameters listed which change everyday climate. One of the most known is global warming, which has a direct effect on climate. This has caused an imbalance in the world's climate. The seasons are shorter or longer than in previous years, relatively speaking. Accordingly, it does not make sense to expect normal seasonal weather anymore. There may be no solutions to this change in temperature, but if the amount of precipitation can be predicted, that will help make lives easier and the world more liveable. Planning for these types of future events is crucial during this tumultuous time. Making predictions about such unseasonable and changeable factors affecting the Earth requires some scrutiny and investigation.

Especially, when global warming has combined with the variability of nature itself, predicting the amount of precipitation in its various forms will be a tough process. Making such a predictive device will have untold benefits for the people and animals living on Earth. To be able to predict the amount of precipitation with all of its effective parameters, it is necessary to make some changes to agricultural activities, to plan engineering processes, or to be prepared for the conditions caused by a severe amount of precipitation. For example, having a good prediction model would enable farmers to predict the amount of precipitation for their area, thereby enabling them to make innovations on, for example, their existing irrigation systems. According to the updated information of precipitation, the updated irrigation systems plans would be able to improve seeding and cropping mechanisms, thereby allowing farmers to make a profit in terms of time and manpower costs when compared to previous seeding systems. [Keefer \(2003\)](#) mentioned that knowledge of the amount of precipitation also enables the selection of the correct agricultural equipment for the purpose of better handling severe precipitation types. Indeed, the importance of predicting and forecasting the amount of precipitation for agricultural activities is related to the power of production.

As a result, having a perfect model of prediction translates to a strong agricultural system which improves the economies of countries in every respect. For instance, the significance of predicting the amount of precipitation also shows itself in the energy area. The most popular field of energy associated with precipitation is hydroelectric power, which is a source of electricity. The generation of this type of electricity is processed by large dams ([Harting 2010](#)). Thus, precipitation zones are taken into consideration whilst selecting the correct fertile areas for the construction of plants for the purpose of producing energy by hydroelectrical means. Large dams should be constructed in regions which receive sufficient precipitation so that maximal energy is reached. These plants are not only constructed in precipitation zones but also in drainage basins. There also exists a relationship between the amount of precipitation and the basins in such a way that river basins are prioritized in those regions known to receive more precipitation than others. Constructing these basins in these regions also prevents freshet cases, which result from severe amounts of precipitation. As a result, by predicting the correct amount of precipitation, proper drainage basins are then set up and, with the source of water which is gained from the river, compounded by the amount of precipitation which occurs naturally, hydroelectrical energy is then produced. At this point, it should not be ignored that a good model for predicting precipitation potentially lends itself to preventing huge natural disasters. People can take proper precautions against such disasters if they have prior knowledge of possible amounts of extreme precipitation. Predicting precipitation amounts can also aid tourism. Good predictions regarding precipitation are beneficial for travelers wishing to better manage their holidays. It also directly affects countries' economies. As a result, a good mechanism for predicting precipitation has comprehensive effects on daily life and the long-run planning of countries' economies.

Range, intermittency, concentration, and temporal and spatial distribution type problems create massive variety and complexity in precipitation variables which do not allow easy descriptions, modelling and prediction. This difficulty can be explained by the association between the

changing amount of precipitation and the variability in the climate, with both its causes and consequences (Ezenwaji *et al.* 2017). In addition to the effects of climatological variability on the total amount of precipitation, some natural causes can be listed as factors which have effects on the total precipitation. Those factors can be considered as different parameters of nature which, in turn, can both be results of climate change and effects on the total precipitation amount over varying periods of time. According to this nature, the observed amount of precipitation changes. Furthermore, different mechanisms, such as the rate of humidity, observed temperature, and cloudiness may affect the time, duration, or intensity of the precipitation. As a result, an accurate and precise modelling of total precipitation series is difficult to achieve because of its being highly parametrized, not to mention the highly varied nature of the data. While modelling total precipitation, it is important to take the maximum and minimum values of those parameters into account in order to obtain the most efficient structures for the model. To exert dominance on the different factors which affect the total precipitation makes it easy for one to understand that the structure of the data is important in order to generate a model with forecasts which are good enough.

Although numerous studies on modelling and predicting precipitation amounts exist (e.g. Sigrist *et al.* 2011; Abdul-Aziz *et al.* 2013; Kotowski & Kazmierczak 2013; Yozgatligil & Türkeş 2018; Esteves *et al.* 2019), research conducted regarding modelling and predicting precipitation methods with the help of temperature, relative humidity and cloudiness variables is very recent in the literature. When different studies conducted in the Mediterranean region are considered, it is seen that there are several studies on modelling precipitation data. For instance, Toth *et al.* (2000) and Brath *et al.* (2002) used autoregressive integrated moving average (ARIMA), artificial neural network (ANN) and k-nearest neighbour approaches to forecast rainfall and flooding in Italy and they showed that time series analysis offers an improvement in flood forecasting accuracy. Guldal & Tongal (2010) predicted a lake level by ARIMA and neural network models. Bahadır (2012) used ARIMA models to forecast mean temperatures and total precipitation amounts in Afyonkarahisar using annual data until 2025, and their best MSE value for precipitation train

series is 3,629.23. They did not use cross-validation to observe the forecasting performance of models. Soltani *et al.* (2007) suggested the use of a seasonal ARIMA(0,0,1)(0,1,1)₁₂ model for Isfahan station and used different seasonal ARIMA models for different areas to capture the periodicity and temporal characteristics of rainfall generating mechanisms in Iran. They only used R^2 for the accuracy measure which cannot capture the actual difference between the observed and predicted values. Sun *et al.* (2019) used only a univariate precipitation series and applied neural networks. The obtained best mean absolute error (MAE) was 31.25. The model performance of Liu & Shi (2019) is the Nash–Sutcliffe coefficient of efficiency (NSCE). Ji *et al.* (2019) used Bayesian model averaging using 1–15-day and 24-h accumulated precipitation series over East Asia based on the ensemble prediction system (EPS) outputs of ECMWF, NCEP, and UKMO from the TIGGE datasets. Their minimum MAE value is around 2.5 mm. The research of Pan *et al.* (2019) uses convolutional neural networks and they compared their study by linear regression, nearest neighbour and random forest. Their root mean square error (RMSE) values, which were better than the performance of other methods they compared, were around 6 mm for the 3-hour precipitation series. Kumar *et al.* (2019) used recurrent neural networks to predict and forecast monthly precipitation series and their RMSE and MAE values were around 400–500 and 200–300 mm, respectively. Around the world there are many studies using deep learning techniques to predict and forecast precipitation amount series. Sadeghi *et al.* (2019) developed the PERSIANN-CNN algorithm to forecast satellite data using convolutional neural networks. The smallest MAE and RMSE scores they obtained were 0.12 and 0.88, respectively. The study of Parviz (2020) used the hybrid of ARIMA and support vector machines, and hybrid of seasonal ARIMA (SARIMA) and ANN on the monthly precipitation series of two stations in northern Iran. The best MAE value that Parviz obtained was 18.02. Nourani *et al.* (2019) used an ANN-type model to forecast precipitation in the Northern Cyprus region, and the best normalized RMSE value they observed was 0.111. Wei *et al.* (2020) examined the Zhengzhou, China region. They used a wavelet and hybrid model of the complementary ensemble empirical mode decomposition, recurrent neural

networks and ARIMA models on an annual precipitation series. They found the average relative error of the forecasting in 2013–2017 was 14.1%.

The most-well known filtering method is the Kalman Filter algorithm (Kalman 1960), which is a commonly used estimator for linear systems in many different areas, including engineering (Hun *et al.* 2016; Deep *et al.* 2018) and economics (Grewal 2011; Neslihanoglu & Date 2019). Provided that the conditional densities are Gaussian and that the Kalman filter is used as a closed-form solution, the Kalman filter is seen as being the most satisfactory model for these purposes. It is rare to find instances in the literature of using this filtering method for the purpose of predicting precipitation amounts. For example, Asemota *et al.* (2016) conducted a study modelling the seasonal behaviour of rainfall in North Nigeria. They used monthly rainfall data collected from 1981 to 2013 in order to pave the way for new agricultural planning in North Nigeria. They accomplished this by utilizing the state space model (SSM) via the Kalman filter. Zulfı *et al.* (2018) also published a study on the development of a rainfall forecasting device using the Kalman filter. In their study, the ARIMA and Kalman filter methods were compared in terms of performance with relation to forecasting rainfall. The in-sample data collected from 2005 to 2015 was divided into clusters using a k-means algorithm, with the Kalman filter algorithm being applied for modelling and forecasting in each cluster. At the end, the study concluded that the performance attained by the Kalman filter was better than that attained by the ARIMA model for forecasting rainfall. Maşazade *et al.* (2019), on the other hand, focused on the amount of rainfall estimated by the Kalman filter with radar reflectivity measurements. The amount of rainfall obtained from the automatic weather observation stations was assumed to be the unknown state vector, with the radar reflectivity values also being used in the measurement model. The aim for applying the Kalman filter in that study was to model for true rainfall amounts. To sum up, although the number of studies examining the effect of different filtering methods on the ability to predict rainfall totals is relatively limited in the literature thus far, based on the studies which have examined it, the Kalman filter is thought to be one of the most preferred methods for predicting precipitations due to the accuracy of the results which it yields.

The main motivation for this study was to develop a high-performance precipitation prediction model for Turkey. Due to high variation within the precipitation amount series, many time series models cannot capture the true nature of the series. To the best of the author's knowledge, the Kalman filter approach has not been used to predict precipitation series. Hence, in this study, the modelling and forecasting performances of different types of filtering methods on monthly total precipitation series of the Muğla region, with covariates including temperature, relative humidity, and cloudiness, are compared. For this purpose, the performance of multiple linear regression, SSM via the Kalman filter algorithm, a hybrid model which integrates the logistic regression and SSM models, the seasonal ARIMA, exponential smoothing with state space model (ETS), TBATS (exponential smoothing state space model with Box–Cox transformation, ARMA errors, trend and seasonal components), feed-forward neural network with one hidden layer (NNETAR) and Prophet developed by Facebook are compared while conducting a modelling and 12- and 24-month rolling window forecasting procedure. This comparison has yet to be undertaken in the literature. The performance of the aforementioned models is evaluated using MAE and mean square error (MSE). The methods were applied using forecast, TSA and prophet R packages. For SSM via Kalman filter, the codes were developed in R software (R Core Team 2018).

The remainder of this paper is organized as follows. Data description and methodology sections outline the several characteristics of the data and the compared models' algorithms; and the results and discussions section presents the model evaluation results from the compared models in modelling and forecasting procedures and provides details about the best model. The final section presents a conclusion based on the results obtained.

DATA DESCRIPTION AND METHODOLOGY

Data description

The monthly series of total precipitation amounts, including three covariates (average temperature, relative humidity, and cloudiness), recorded from the weather stations at

Muğla and which were obtained from the Turkish Meteorology Services between 1950 and 2010, are analyzed here. There were several missing values in both precipitation and other covariates series. Although for the Kalman filter algorithm imputation is not necessary, to be able to use other forecasting tools we need to impute the missing values. Missing monthly values were imputed by the expectation maximization (EM)–Markov Chain Monte Carlo (MCMC) multiple imputation method as suggested by [Yozgatligil *et al.* \(2013\)](#). The EM algorithm, developed by [Little & Rubin \(1987\)](#), is a method of finding the maximum likelihood estimates of parameters from a certain probability density function when the data are missing. These estimates are used as a starting value for beginning the MCMC process. The parameters of the joint posterior distribution of

the incomplete values are estimated by simulation in two steps: the imputation step (I-step) and the posterior step (P-step). In step I, missing values with the estimated mean vector and covariance matrix are simulated, and in step P, the posterior population mean vector and covariance matrix from the complete sample estimates are simulated. These two steps are iterated enough times to obtain reliable results ([Schafer 1997](#)). [Yozgatligil *et al.* \(2013\)](#) suggested inputting missing values using this method when at most 50% of the data is not observed.

Due to the fact that Muğla is receiving precipitation on average the whole year round, compared to the rest of the regions in Turkey, the dataset was used as one of the application stations. The descriptive statistics belonging to the research data are shown in [Table 1](#). Furthermore, the time series plot of monthly total precipitation amounts is displayed in [Figure 1](#).

The various descriptive statistics are given in [Table 1](#). The minimum observation for precipitation in Muğla is 0 mm as, on days without any precipitation, the amount of precipitation remains at 0. The mean amount of precipitation is 96.72 mm, which is characterized as a moderate value taking into consideration the dataset's minimums and maximums. The temperature never drops below 0 °C and never exceeds 29 °C. The average temperature during the time period under study is 14.97 °C. This indicates a moderate

Table 1 | Descriptive statistics for Muğla

	Precipitation (mm)	Temperature (°C)	Relative humidity (%)	Cloudiness
Minimum	0.00	2.50	29.10	0.10
1st quartile	11.10	7.90	51.38	1.50
Median	55.85	14.10	64.80	3.50
Mean	96.72	14.97	62.47	3.39
3rd quartile	143.40	22.20	73.83	5.00
Maximum	645.30	29.00	89.40	7.90

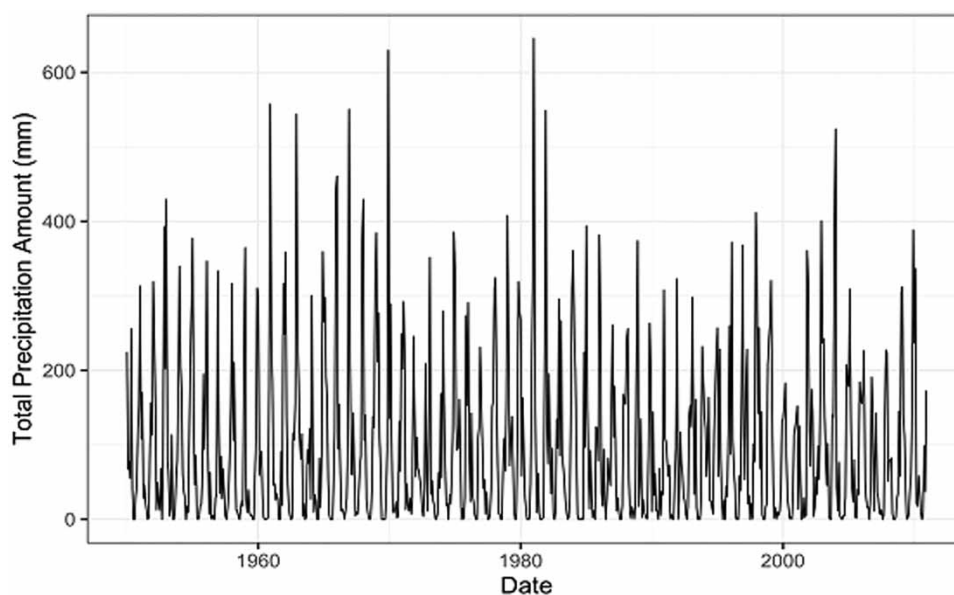


Figure 1 | The time series plot of monthly total precipitation.

zone in terms of weather conditions. This moderateness may be due to Muğla's location as its capital city is by the sea. Furthermore, it could be said that the region has a maritime climate given its humidity. The minimum and maximum relative humidity values are observed as 29.10 and 89.40%. These figures reflect the fact that the region is mostly bordered by the Aegean Sea. Although the average value of relative humidity is approximately 65%, the cloudiness observed in the city of Muğla remains mostly at low levels.

As seen from Figure 1, there are no conclusive patterns or trends in precipitation. After deciding about trend, periodicity and variation are taken into consideration. Variation derives from the peaks and troughs of the data presented. It can clearly be seen that the times series precipitation data rises and falls across the span of the studied 62 years. The regular peaks and troughs which occur every 12 months are due to seasonality. To be able to observe the forecast performance of models, the series is divided into a train (in-sample) and test set (out-of sample). First, only the last year of the series is taken as a test set because ARIMA type stochastic models converge to its process mean and cannot be used for long-run forecasting. Performance measures are calculated both for train and test sets, then, to see the effect of the forecast horizon the last two years are used as the test set and accuracy calculations are just given for the test set.

Methodology

The precipitation models in the form of a multiple linear regression (MLR) model, SSM, a hybrid model, the ARIMA (Ünal 2019), ETS, TBATS, NNETAR and Prophet models are discussed below.

Multiple linear regression

The precipitation model in the form of the multiple linear regression model is defined as follows:

$$\begin{aligned} \text{Precipitation}_t = & \kappa + \beta_1(\text{Temperature})_t \\ & + \beta_2(\text{Relative Humidity})_t + \beta_3(\text{Cloudiness})_t + \varepsilon_t, \varepsilon_t \sim N(0, H) \end{aligned} \quad (1)$$

where κ is the regression intercept and β_1, β_2 and β_3 are unknown coefficients. ε_t ($t = 1, \dots, T$), are normally distributed

errors with mean 0 and constant variance H . The unknown β values given in Equation (1) are estimated via ordinary least squares (OLS), where $\hat{\beta}$ is written as follows:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \| \text{Precipitation}_t - (\kappa + \beta_1(\text{Temperature})_t + \beta_2(\text{Relative Humidity})_t + \beta_3(\text{Cloudiness})_t) \|_2^2 \quad (2)$$

where

$$\begin{aligned} & \| \text{Precipitation}_t - (\kappa + \beta_1(\text{Temperature})_t \\ & + \beta_2(\text{Relative Humidity})_t + \beta_3(\text{Cloudiness})_t) \|_2^2 \\ & = \sum_{t=1}^n (\text{Precipitation}_t - (\kappa + \beta_1(\text{Temperature})_t \\ & + \beta_2(\text{Relative Humidity})_t + \beta_3(\text{Cloudiness})_t))^2 \end{aligned}$$

State space model and Kalman filter

The extension of the regression model given in Equation (1) allows time-varying coefficients, β_i ($i = 1, 2, 3$), into the mean reverting specification of the SSM using the Kalman filter algorithm present in Equation (3):

$$\begin{aligned} \text{Precipitation}_t = & \kappa + \beta_{1t}(\text{Temperature})_t \\ & + \beta_{2t}(\text{Relative Humidity})_t + \beta_{3t}(\text{Cloudiness})_t \\ & + \varepsilon_t, \varepsilon_t \sim N(0, H). \end{aligned} \quad (3)$$

The state equations given in Equation (3) can be written as follows:

$$\begin{aligned} \beta_{1t} = & \bar{\beta}_1 + \varphi_1(\beta_{1t-1} - \bar{\beta}_1) + w_{1t}, w_{1t} \sim N(0, Q_1), \\ \beta_{2t} = & \bar{\beta}_2 + \varphi_2(\beta_{2t-1} - \bar{\beta}_2) + w_{2t}, w_{2t} \sim N(0, Q_2), \\ \beta_{3t} = & \bar{\beta}_3 + \varphi_3(\beta_{3t-1} - \bar{\beta}_3) + w_{3t}, w_{3t} \sim N(0, Q_3), \end{aligned} \quad (4)$$

with prior distributions for the parameters:

$$\beta_{10} \sim N(\mu_{\beta_1}, \Sigma_{\beta_1}), \beta_{20} \sim N(\mu_{\beta_2}, \Sigma_{\beta_2}), \beta_{30} \sim N(\mu_{\beta_3}, \Sigma_{\beta_3}).$$

Here, the initial estimates of $\mu_{\beta_{10}}, \mu_{\beta_{20}}$ and $\mu_{\beta_{30}}$ and $\Sigma_{\beta_{10}}, \Sigma_{\beta_{20}}$ and $\Sigma_{\beta_{30}}$ are obtained from the data as part of the estimation and $\bar{\beta}_i = 1/T \sum_{t=1}^T \beta_{it}$ ($i = 1, \dots, 3$). The error terms for observation (ε_t) and state equations (w_{it}) are assumed to be mutually independent of each other and independent in time t ($t = 1, \dots, T$) and normally distributed with

0 mean and variances H and Q_i , respectively. Also, φ_i quantifies the temporal autocorrelation in β_{it} in precipitation.

A Kalman filter can be used anywhere uncertain information exists about some dynamic system. The aim of the Kalman filter is to obtain as much information from the uncertain measurements as possible. A Kalman filter is an optimal estimator that infers parameters of interest from indirect, inaccurate and uncertain observations. It deals with the uncertainty associated with the system by adding some new uncertainty after every prediction step. It combines a prediction of the true data with the new measurement, using a weighted average that is an estimate lying between the prediction and the measurement. Hence, it has a better estimated uncertainty. This process is repeated at every time step, with the new estimate informing the prediction used in the preceding iteration. The relative certainty of the measurements and current estimate yields the Kalman filter gain which is the relative weight given to the measurements and current state estimate, and can be 'tuned' to achieve particular performance. Because the uncertainty is too much in the precipitation series, this method handles this well. For example, Weeks (1999) proposed the Kalman filter, which is also particularly suitable for missing data in time series due to its recursive algorithm.

During the Kalman filtering and smoothing process, three types of problems are defined as follows: if $t > n$, this is a prediction problem, if $t = n$, this is a filtering problem, and if $t < n$, this is a smoothing problem. The flow chart of general state space model via Kalman filtering (forward step ($t = 1, \dots, n$)) and smoothing (backward step ($t = n, n-1, \dots, 1$)) algorithm are given in Figure 2. In this process, the parameters of Equations (3) and (4) are modified as follows:

$$Y_t = \text{Precipitation}_t, \alpha_0^0 = (\mu_{\kappa_0}, \mu_{\beta_{10}}, \mu_{\beta_{20}}, \mu_{\beta_{30}})'$$

$$A_t = (1, \text{Temperature}_t, \text{Relative Humidity}_t, \text{Cloudiness}_t)$$

$$P_0^0 = \begin{pmatrix} \Sigma_{\kappa_0} & 0 & 0 & 0 \\ 0 & \Sigma_{\beta_{10}} & 0 & 0 \\ 0 & 0 & \Sigma_{\beta_{20}} & 0 \\ 0 & 0 & 0 & \Sigma_{\beta_{30}} \end{pmatrix}, \Phi = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \varphi_1 & 0 & 0 \\ 0 & 0 & \varphi_2 & 0 \\ 0 & 0 & 0 & \varphi_3 \end{pmatrix}$$

$$Q = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & Q_1 & 0 & 0 \\ 0 & 0 & Q_2 & 0 \\ 0 & 0 & 0 & Q_3 \end{pmatrix}, H = H$$

Note that the whole mechanism for the application of the Kalman filter is based on this procedure. This model is used for the application to obtain the smoothed values for precipitation in Equations (3) and (4). As a result of the estimation, some of the values are predicted as negative values that are very close to 0. To overcome this, logarithmic transformation was considered, but this did not solve the problem. At that point, it was decided to count those negative values as zero because they were already very close to zero. This means that there is no expectation for precipitation on that day. This model is called an SSM throughout this research.

Hybrid model

The hybrid data is created from the actual data to handle negative predictions observed in the smoothing part by using the logistic regression method. The steps listed below are followed for generating the hybrid model:

Step 1: The precipitation amount series are arranged as 0 if the amount is 0; otherwise, it is recorded as 1;

Step 2: A logistic regression model is fitted by using explanatory variables. The logistic regression model is defined as follows:

$$\frac{P(x)}{1 - P(x)} = e^{g'x} \quad (5)$$

Step 3: After classifying observations as rain and no rain cases, the estimated 0's (no-rain cases) fixed as 0 and the estimated 1's (rain cases) are put in the Kalman filter algorithm as precipitation amounts;

Step 4: The accuracy measures and forecasts for future observations are then obtained by using the state space model via the Kalman filter algorithm.

This approach is called a hybrid model throughout this research.

ARIMA model

The SARIMA model is defined as follows:

$$\begin{aligned} \Phi(B_s)\varphi(B)(1-B)^d(1-B^s)^D \text{Precipitation}_t \\ = \Theta(B^s)\theta(B)\varepsilon_t \end{aligned} \quad (6)$$

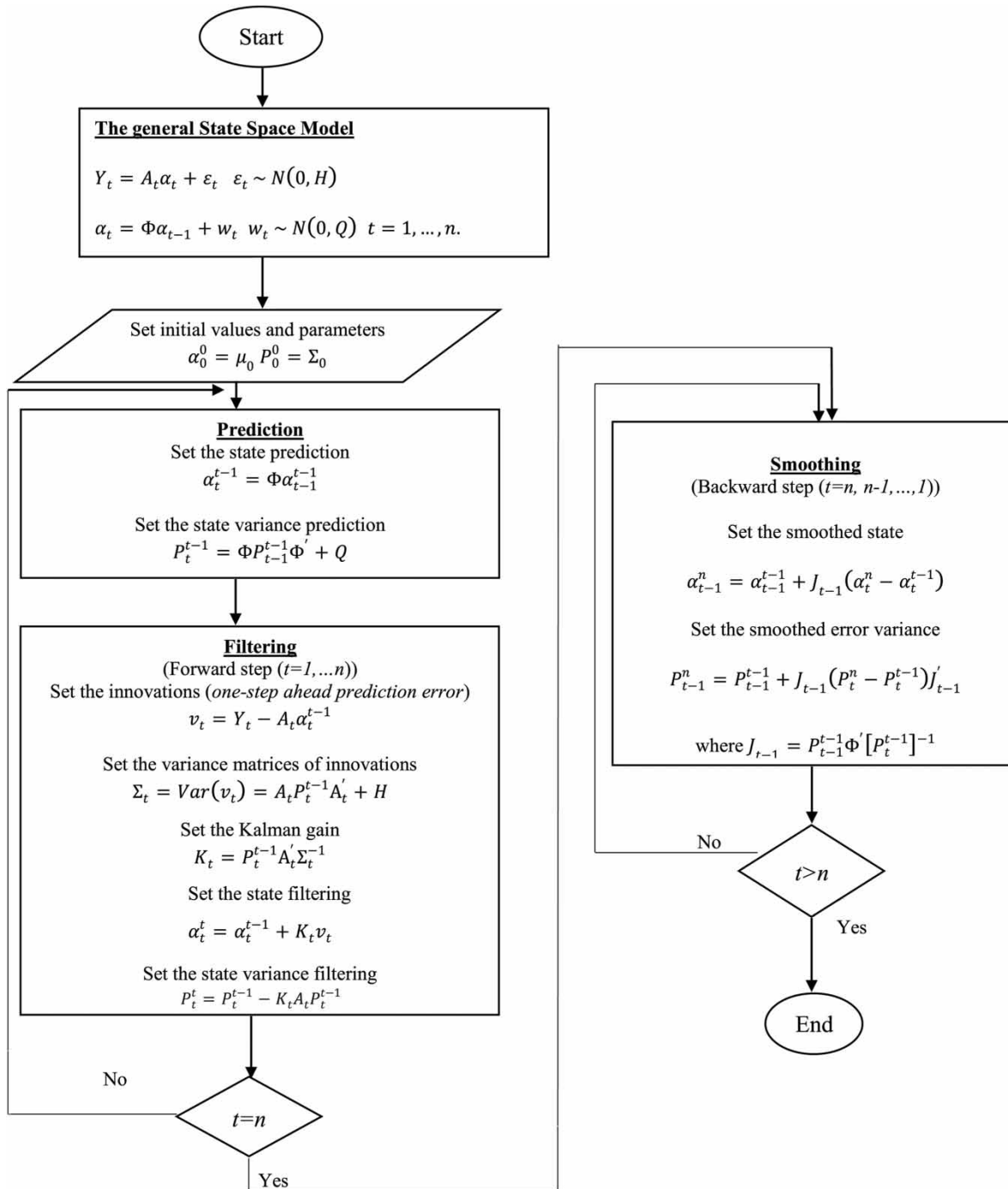


Figure 2 | The flow chart of the algorithm used in Kalman filtering and smoothing.

Here, it is denoted by $ARIMA(p, d, q) \times (P, D, Q)_s$ where the seasonality is represented by s (Box & Jenkins 1970). Φ and Θ are the polynomials with orders P and Q , respectively, each containing no roots inside the unit circle and B

represents the back-shift operator, $B^j y_t = y_{t-j}$. Given the seasonality observed in Figure 1 and autocorrelation and partial autocorrelation plots of the series, the SARIMA model was fitted by using the *arima* function in R software (R Core

Team 2018). The SARIMA model is fitted as SARIMA(1, 0, 1) × (1, 1, 0)₁₂ based on the Bayesian information criteria (BIC). This approach is called the ARIMA model throughout this research.

Exponential smoothing with state space model (ETS)

Exponential smoothing is a deterministic forecasting method which was developed in the late 1950s by the works of Holt (1957), Brown (1959) and Winters (1960). This model is an exponential smoothing model with an underlying state space model consisting of a level component, a trend component (T), a seasonal component (S), and an error term (E). Forecasts are weighted averages of past observations where weights decrease exponentially. This method is widely used in forecasting due to its simplicity, computational efficiency and accurate forecast performance. Types of the exponential smoothing method vary based on the characteristics of the time series. The additive model is given in Equation (7):

$$\text{Precipitation}_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t \quad (7)$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$$

$$b_t = b_{t-1} + \beta \varepsilon_t$$

$$s_t = s_{t-m} + \gamma \varepsilon_t$$

where ℓ_t denotes an estimate of the level of the series at time t and b_t denotes an estimate of the trend of the series at time t , s_t denotes the seasonality with weight coefficients α , β and γ and $\varepsilon_t \sim NID(0, \sigma^2)$. In the multiplicative model the summation sign is replaced by a multiplication sign. Different combinations of this model are given in Hyndman et al. (2008).

TBATS

To be able to handle more than one seasonal pattern, time series, Livera et al. (2011) developed a model which is the combination of exponential smoothing and ARIMA models including the trigonometric representation of the seasonal component on a Box-Cox transformation. The mathematical formulation of the TBATS model is expressed

by the following equations:

$$\text{Precipitation}_t^{(w)} = \begin{cases} \frac{\text{Precipitation}_t^{(w)} - 1}{w}, & w \neq 0 \\ \log \text{Precipitation}, & w = 0 \end{cases} \quad (8)$$

$$\text{Precipitation}_t^{(w)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t$$

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = \phi b_{t-1} + \beta d_t$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)} \text{ where } s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t$$

where m_i denotes the seasonal periods, ℓ_t and b_t represent the level and trend components of the series at time t , respectively, $s_{j,t}^{(i)}$ represents the i th seasonal component at time t with $\lambda_j^{(i)} = 2\pi j/m_i$, d_t denotes an ARMA(p, q) process and ε_t is a Gaussian white noise process with zero mean and constant variance. The smoothing parameters are given by α , β , γ_i for $i = 1, \dots, T$, is the dampening parameter.

Feed-forward neural network (NNETAR)

ANN is a mathematical model imitating human neural biology to solve nonlinear problems. One of its significant properties is containing non-linearity in its structure. Also, the ANN models can be described as universal approximators that can approximate the generating mechanism of the data accurately (Zhang 2003). They do not require any modelling assumptions. However, since they require a high level of modelling complexity, ANN models suffer from being time-consuming (Theodosiou 2011). The algorithm is supervised learning where input and output series need to be provided. At the first step, the given input variables are multiplied by weights which is then learned by the algorithm using the back-propagation. Next, these weighted inputs are summed. Then, a bias term is added

to adjust the threshold. At the final step, the summation of the weighted inputs and bias are transformed into the final output by the activation function to catch the complex structure of the series. One of the common activation functions that is used for time series is the sigmoid function. The mathematical formulation of this process is given as follows:

$$\text{Precipitation}_t = f\left(\sum_{i=0}^m w_{i,t} x_{i,t} + b\right) \quad (9)$$

where $x_{i,t}$ is the input in discrete time t where $i = 1, \dots, m$, $w_{i,t}$ is the weight value at time t , b is bias, f is an activation function, and Precipitation_t is the output value at time t . To update the weights to minimize the loss function like MSE, back-propagation which is based on the chain rule is used. In feed-forward neural network, the procedure is one directional from the input variables to the output variable. For the network to capture the nonlinear structure of the data, hidden layers are added between the input layer and the output layer. In the structure of feed-forward there exists a network with interconnections, but these interconnections do not form any loops (Krenker et al. 2011). In the application of the algorithm, five regular lags and four seasonal lags are given as input variables. The average of 20 networks, each of which is a 9-40-1 network with 441 weights, are used. The algorithm is named NNETAR.

Prophet

The Prophet model was introduced by Facebook in 2017 (Taylor & Letham 2018). It can capture the trend and strong multiple seasonality at day, week, year level etc. with the time series. It uses a decomposable time series model with three main model components: trend, seasonality, and holidays. They are combined in the following equation:

$$\text{Precipitation}(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (10)$$

where $g(t)$ is the piecewise linear or logistic growth curve for modelling non-periodic changes in time series; $s(t)$ is the periodic change (e.g. weekly/yearly seasonality); $h(t)$ is the effect of holidays (user provided) with irregular schedules; ε_t is the error term accounting for any unusual changes not accommodated by the model.

The model can be implemented for user defined change points or irregular holidays. To fit and forecast the effects of seasonality, the model is based on Fourier series. It uses a Bayesian framework to estimate the unknown parameters of the model but the application is very user-friendly. The algorithm obtained the predicted values automatically. In the application of the algorithm, some parameters are tuned to get better results which is known as hyperparameter tuning.

SARIMA, TBATS and NNETAR models are also run including the temperature, humidity and cloudiness covariates. The results of these models are given as ARIMAX, TBATSX and NNETARX.

The performance of fitted models in the in-sample modelling (out-of-sample forecasting) procedure are evaluated in terms of the MAE and MSE criteria and are shown as follows:

$$\text{MAE} = \sum_{t=1}^T \frac{|\widehat{\text{Precipitation}}_t - \text{Precipitation}_t|}{T} \quad (11)$$

$$\text{MSE} = \sum_{t=1}^T \frac{(\widehat{\text{Precipitation}}_t - \text{Precipitation}_t)^2}{T} \quad (12)$$

The lowest values of modelling (forecast) evaluation criteria mean that the in-sample modelling procedure (out-of-sample forecasting) has been estimated properly and that the model constructed at the end is a significant model.

RESULTS AND DISCUSSION

This section presents a comparison of the in-sample (train set) model fit and out-of-sample (test set) forecasting performance for the MLR, SSM, hybrid, SARIMA, ETS, TBATS, NNETAR and Prophet models and the same models applied with the covariates during the given time period. A model performance comparison is conducted in terms of the MAE (Equation (11)) and MSE (Equation (12)), respectively. Those results are presented in Table 2.

As seen in Table 2, the state space model (SSM) has the lowest MAE and MSE values. This means that it outperforms other methods. The performance of the SSM is 57.88% (92.38%) better than that of the hybrid model in terms of MSE (MAE). When the in-sample modelling performance of other models is considered, it is seen that

Table 2 | MAE and MSE of in-sample modelling and out-of-sample forecasting

Period Model	Train: In-sample modelling		Test: Out-of-sample forecasting (12 months forecast)		Test: Out-of-sample forecasting (24 months forecast)	
	MAE	MSE	MAE	MSE	MAE	MSE
MLR	52.4912	5,177.2840	50.3077	4,172.013	45.9838	3,620.9360
SSM	0.1872	0.1744	0.0130	0.00030	0.0088	0.00012
Hybrid	0.4444	2.2887	0.0152	0.00035	0.0094	0.00013
SARIMA	49.8064	5,620.1560	51.2716	5,126.8750	50.2222	5,616.3184
SARIMAX	42.1393	3,813.8159	51.6351	5,291.5440	49.8555	5,425.2063
ETS	49.1069	5,613.7406	49.2087	4,969.4604	44.3153	4,384.5322
TBATS	49.1069	5,605.4720	54.2269	5,545.6617	48.6871	4,924.1938
TBATSX	49.1069	5,605.4720	54.2269	5,545.6617	48.6871	4,924.1938
NNETAR	14.6143	491.9524	51.9587	6,305.0111	73.6834	11,557.8196
NNETARX	7.07677	104.9068	47.8245	5,727.0537	61.0232	7,619.3171
PROPHET	49.9683	5,625.9000	48.9492	5,087.0274	43.8493	4,355.7228

although the values are very close to each other, the performance of the Prophet model is better than the others. Taking the covariates into account improves the results slightly. To sum up, the SSM is the preferable model for the precipitation amount using the train set considering MAE and MSE measures.

In the out-of-sample forecasting procedure, a rolling window technique with one month ahead forward prediction is used for evaluating the predictive performance of these models. The length of the prediction period is chosen as 12 months over the period from January 2010 to December 2010 and 24 months over the period from January 2009 to December 2010. The MAE and MSE values between the predicted and actual returns on the precipitation amounts are calculated over 12 and 24 values during that time period and are given in Table 2.

As seen in Table 2, it is apparent that the performance of the SSM outperforms the others. It also improves on the hybrid model by 14.48% (14.29%) for 12 months precipitation values and by 6.78% (7.69%) for 24 months precipitation values while in the out-of-sample forecasting procedure in terms of MAE (MSE). When the out-of-sample forecasting performance of other models is considered, it is seen that although the values are very close to each other, the performance of the Prophet model is slightly better than the others. The in-sample performance of NNETAR models is good but the out-of-sample results

are very high which is an indication of overfitting. A comparative forecast plot of all methods is given in Figure 3. SSM forecasted values are almost the same as the original observed values. Moreover, SARIMA forecasts are closer to the original values compared to the forecasts of the other methods. The performances of the other methods are similar to each other and do not quite explain the behavior of the original series. Overall, the SSM via the Kalman filter is the preferable model for predicting precipitation amounts while forecasting out-of-sample during the given time period. For the sake of brevity, the SSM via Kalman filter forecasts and actual total precipitation amounts for the 12 months are provided in Table 3.

As seen from Table 3, it is apparent that the difference between the actual precipitation values and the forecast precipitation values from the SSM via Kalman filter is too low. This means that the variation of the precipitation amount has been captured successfully in the model.

Best model fit and forecast

In terms of a comparison of the results between the aforementioned model while modelling and forecasting during these time periods, the SSM via Kalman filter exhibits the best performance. The parameter estimates of SSM in the in-sample model fit procedure are shown in Table 4.

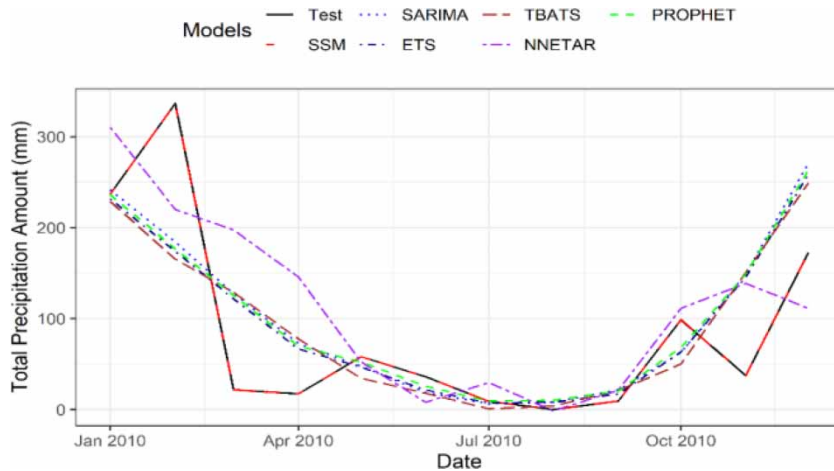


Figure 3 | Forecast plot of all methods compared to the original series.

Table 3 | Forecasts of SSM and actual total precipitation amount for 12 months

Forecast	Actual	Residual
237.0015	237.0	0.0015
336.5844	336.6	-0.0156
21.6244	21.6	0.0244
17.4138	17.4	0.0138
58.2010	58.2	0.0010
36.1162	36.1	0.0161
9.0436	9.0	0.0436
0.0066	0.0	0.0066
9.2087	9.2	0.0087
98.7082	98.7	0.0082
37.4147	37.4	0.0147
172.8015	172.8	0.0015

Overall, the 99% variability in the in-sample model fit of the precipitation values can be explained by the independent variables of temperature, relative humidity, and cloudiness for the hybrid model. This is actually a very extreme value of accuracy and can therefore be counted as proof of the Kalman filter's power of modelling precipitation. The estimated values of \hat{H} are higher than \hat{Q}_1 , \hat{Q}_2 except for \hat{Q}_3 during the given time period. This means that the observation variance captures the volatility of the precipitation values more than the state (including the parameters of temperature, relative humidity). Additionally, the estimated \hat{Q}_3 value is higher than the \hat{Q}_1 and \hat{Q}_2 values, which means that the cloudiness state variance

Table 4 | The precipitation model in the state space form via Kalman filter parameter estimates

Parameter	Value	Parameter	Value
\hat{H}	7.8870 (4.5257)	$\hat{\alpha}$	-19.798 (22.7454)
\hat{Q}_1	1.08×10^{-23} (0.0000)	$\hat{\beta}_1$	0.2636 (0.0878)
\hat{Q}_2	4.26×10^{-43} (0.0000)	$\hat{\beta}_2$	0.0859 (0.0106)
\hat{Q}_3	287.983 (15.5027)	$\hat{\beta}_3$	27.5361 (4.4127)
$\hat{\varphi}_1$	0.0015 (0.0252)	Loglikelihood	-3,821.285
$\hat{\varphi}_2$	0.2754 (0.1132)	AIC	7,664.570
$\hat{\varphi}_3$	0.3016 (0.0183)	BIC	7,715.124
		R^2	0.999986
		Adj R^2	0.999986

captures the volatility of the various covariates. In addition, the temporal autocorrelation (captured by $\hat{\varphi}_1$, $\hat{\varphi}_2$ and $\hat{\varphi}_3$) is close for relative humidity and cloudiness, but that of temperature is far from the relative humidity and cloudiness state parameters, seeing as they are closer to 0 than they are to 1. This suggests that the time-varying state parameters changed rapidly due to low autocorrelation. It is worth noting that the average regression intercept, α , is negative. This result indicates that the actual precipitation amount is lower than the expected precipitation amount during the

given time period. Moreover, the mean value of the time-varying $\hat{\beta}_3$ of cloudiness is greater than that of temperature and relative humidity during the time period. This result indicates that cloudiness is more volatile than the other two parameters.

CONCLUSIONS

The issue of predicting precipitation is a challenging process since there are various natural parameters which are involved in the procedure which directly affect precipitation. The parameters which are taught as deterministic factors for the amount of precipitation chosen include temperature, relative humidity, and cloudiness. The modelling and forecasting performance of the monthly total precipitation amount with these determined factors in Muğla compare the multiple linear regression, state space model via the Kalman filter algorithm, the hybrid model (which integrates the logistic regression and SSM models), and the SARIMA.

It was determined that the performance of the state space model is better than that of the hybrid model and all other models considered in this study. This suggests that the Kalman filter is preferable to that of using the integration of the logistic regression and the Kalman filter. The Kalman filter outperforms the MLR via OLS, SARIMA, ETS, TBATS (having the worst prediction and forecasting performance), NNETAR and Prophet models in modelling and forecasting the amount of precipitation. It is concluded that having too many zeros (no-rain situations) in the data itself, as at the Muğla weather stations, does not affect the performance of the Kalman filter at all. Also, changing the forecast horizon does not have any influence on the forecasting performance of this model.

Our model considers the relationship between the precipitation amount and several other variables via SARIMA, SSM, ETS, TBATS, NNETAR and Prophet forecasting approaches. Hence, it uses more information related to the nature of the series. It is observed that state space modeling via Kalman filtering outperforms the well-known ARIMA and other forecasting models. The MSE value obtained from the Kalman filter method is 0.1744 and the MAE value is 0.1872, which are significantly lower than the performances of the other studies. It is

believed that this model performance can also be obtained in different regions of the world.

To sum up, light was shed on the necessity for time-varying coefficients extensions of the precipitation model and filtering methods and applying the method to high frequency data such as daily and hourly. That type of study would be significant in terms of being a guide for a large number of interdisciplinary applications, such as agricultural, engineering, and preventative (e.g. with regards to natural disasters, such as floods and droughts) applications. Moreover, with the help of Kalman filtering, simulation of the actual precipitation series is also possible with a small error which allows the theoretical studies on the precipitation studies such as the performance of homogeneity test on precipitation series. Imitating the series with the estimated coefficients is not possible with machine learning algorithms and ANN type approaches because a mathematical form of the model cannot be written with them.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

- Abdul-Aziz, A. R., Anokye, M., Kwame, A., Munyakazi, L. & Nsowah-Nuamah, N. N. 2013 Modeling and forecasting rainfall pattern in Ghana as a seasonal ARIMA process. The case of Ashanti region. *International Journal of Humanities and Social Science* 3 (3), 224–233.
- Asemota, O. J., Bamanga, M. A. & Alaribe, O. J. 2016 Modelling seasonal behaviour of rainfall in northeast Nigeria. A state space approach. *International Journal of Statistics and Applications* 6 (4), 203–222.
- Bahadır, M. 2012 The analyse of precipitation and temperature in Afyonkarahisar (Turkey) in respect of Box–Jenkins technique. *Journal of Academic Social Science Studies* 5 (8), 195–212.
- Box, G. & Jenkins, G. 1970 *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- Brath, A., Montanari, A. & Toth, E. 2002 [Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models](#). *Hydrology and Earth System Sciences Discussions* 6 (4), 627–639.

- Brown, R. G. 1959 *Statistical Forecasting for Inventory Control*. McGraw-Hill, New York.
- Deep, A., Mittal, M. & Mittal, V. 2018 Application of Kalman filter in GPS position estimation. In: *IEEE 8th Power India International Conference (PIICON)*. IEEE, Kurukshetra, India.
- Esteves, J. T., de Souza, R. G. & Ferraudo, A. S. 2019 *Rainfall prediction methodology with binary multilayer perceptron neural networks*. *Climate Dynamics* **52**, 2319–2331.
- Ezenwaji, E., Nzoiwu, C. & Chima, G. 2017 Analysis of precipitation concentration index (pci) for awka urban area, Nigeria. *Hydrology: Current Research* **8** (4), 1–6.
- Grewal, M. S. 2011 Kalman filtering. In: *International Encyclopedia of Statistical Science* (M. Lovric, ed.). Springer, Berlin, Heidelberg.
- Guldal, V. & Tongal, H. 2010 *Comparison of recurrent neural network, adaptive neuro-fuzzy inference system and stochastic models in Egirdir Lake level forecasting*. *Water Resources Management* **24**, 105–128.
- Harting, C. 2010 *Rainfall as an Energy Source*. Available from: <http://large.stanford.edu/courses/2010/ph240/harting2/>
- Holt, C. C. 1957 *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages*. ONR Memorandum, Vol. 52. Carnegie Institute of Technology, Pittsburgh. Available from the Engineering Library, University of Texas, Austin.
- Hun, L. C., Yeng, O. L., Sze, L. T. & Chet, K. V. 2016 Kalman filtering and its real-time applications. In: *Real-time Systems* (F. Govaers, ed.). IntechOpen, Rijeka.
- Hyndman, R. J., Koehler, A. B., Ord, J. K. & Snyder, R. D. 2008 *Forecasting with Exponential Smoothing: The State Space Approach*. Springer-Verlag, Berlin.
- Ji, L., Zhi, X., Zhu, S. & Fraedrich, K. 2019 *Probabilistic precipitation forecasting over East Asia using bayesian model averaging*. *Weather and Forecasting* **34**, 377–392.
- Kalman, R. E. 1960 *A new approach to linear filtering and prediction problems*. *Transactions of the ASME–Journal of Basic Engineering* **82**, Series D, 35–45.
- Keefer, T. O. 2003 Precipitation simulation models. In: *Encyclopedia of Water Science* (S. W. Trimble, ed.). Wiley Blackwell, Tuscon, Arizona, United States of America.
- Kotowski, A. & Kazmierczak, B. 2013 *Probabilistic models of maximum precipitation for designing sewerage*. *Journal of Hydrometeorology* **14** (6), 1958–1965.
- Krenker, A., Kos, A. & Janez, B. 2011 *Introduction to the Artificial Neural Networks*. INTECH Open Access Publisher, London.
- Kumar, D., Singh, A., Samui, P. & Jha, R. K. 2019 *Forecasting monthly precipitation using sequential modelling*. *Hydrological Sciences Journal* **64** (6), 690–700.
- Little, R. J. A. & Rubin, D. B. 1987 *Statistical Analysis with Missing Data*. John Wiley & Sons, New York, NY.
- Liu, S. & Shi, H. 2019 *Recursive approach to long-term prediction of monthly precipitation using genetic programming*. *Water Resources Management* **33**, 1103–1121.
- Livera, A. M., Hyndman, R. J. & Snyder, R. D. 2011 *Forecasting time series with complex smoothing exponential using seasonal patterns*. *Journal of the American Statistical Association* **106**, 1513–1527.
- Maşazade, E., Bakır, A. K. & Kırıcı, P. 2019 *A Kalman filter application for rainfall estimation using radar reflectivity*. *Turkish Journal of Electrical Engineering and Computer Sciences* **27**, 1198–1212.
- Neslihanoglu, S. & Date, P. 2019 *A modified sequential Monte Carlo procedure for the efficient recursive estimation of extreme quantiles*. *Journal of Forecasting* **38** (5), 390–399.
- Nourani, V., Molajou, A., Uzelaltinbulat, S. & Sadikoglu, F. 2019 *Emotional artificial neural networks (EANNs) for multi-step ahead prediction of monthly precipitation; case study: Northern Cyprus*. *Theoretical and Applied Climatology* **138**, 1419–1434.
- Pan, B., Hsu, K., AghaKouchak, A. & Sorooshian, S. 2019 *Improving precipitation estimation using convolutional neural network*. *Water Resources Research* **55** (3), 2301–2321.
- Parviz, L. 2020 Comparative evaluation of hybrid SARIMA and machine learning techniques based on time varying and decomposition of precipitation time series. *JAST* **22** (2), 563–578.
- R Core Team. 2018 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Sadeghi, M., Asanjan, A. A., Faridzad, M., Nguyen, P., Hsu, K., Sorooshian, S. & Braithwaite, D. 2019 *PERSIANN-CNN: Precipitation estimation from remotely sensed information using artificial neural networks–convolutional neural networks*. *Journal of Hydrometeorology* **20** (12), 2273–2289.
- Schafer, J. L. 1997 *Analysis of Incomplete Multivariate Data*. Chapman & Hall, New York, NY.
- Sigrist, F., Künsch, H. & Stahel, A. W. 2011 *A dynamic nonstationary spatio-temporal model for short term prediction*. *Annals of Applied Statistics* **6** (4), 1452–1477.
- Soltani, S., Modarres, R. & Eslamian, S. S. 2007 *The use of time series modelling for the determination of rainfall climates of Iran*. *International Journal of Climatology* **27**, 819–829.
- Sun, M., Li, X. & Kim, G. 2019 *Precipitation analysis and forecasting using singular spectrum analysis with artificial neural networks*. *Cluster Computing* **22**, 12633–12640.
- Taylor, S. J. & Letham, B. 2018 *Forecasting at scale*. *The American Statistician* **7** (1), 37–45.
- Theodosiou, M. 2011 *Disaggregation aggregation of time series components: a hybrid forecasting approach using generalized regression neural networks and the theta method*. *Neurocomputing* **74**, 896–905.
- Toth, E., Brath, A. & Montanari, A. 2000 *Comparison of short-term rainfall prediction models for real-time flood forecasting*. *Journal of Hydrology* **239**, 132–147.
- Ünal, E. 2019 *Performance Comparison of Filtering Methods on Modelling and Forecasting Total Precipitation Amount*. M.Sc. Thesis (unpublished), Middle East Technical University, Ankara, Turkey.
- Weeks, M. 1999 *Methods of Imputation for Missing Data*. Cambridge, UK. Available from: <http://www.econ.cam.ac.uk/people-files/faculty/mw217/pdf/mispapnw.pdf>

- Wei, T., Xianqi, Z., Chao, S., Dengkui, H. & Tianyi, L. 2020 [Annual precipitation analysis and forecasting – taking Zhengzhou as an example](#). *Water Supply* **20** (5), 1604–1616.
- Winters, P. R. 1960 [Forecasting sales by exponentially weighted moving averages](#). *Management Science* **6** (3), 324–342.
- Yozgatligil, C. & Türkeş, M. 2018 [Extreme value analysis and forecasting of maximum precipitation amounts in the western black sea subregion of Turkey](#). *International Journal of Climatology* **38** (15), 5447–5458.
- Yozgatligil, C., Aslan, S., Iyigun, C. & Batmaz, I. 2013 [Comparison of missing value imputation methods for Turkish meteorological time series data](#). *Theoretical and Applied Climatology* **112**, 143–167.
- Zhang, I. M. G. P. 2003 [Time series forecasting using a hybrid ARIMA and neural network model](#). *Neurocomputing* **50**, 159–175.
- Zulfi, M., Hasan, M. & Purnomo, K. D. 2018 The development rainfall forecasting using Kalman filter. *Journal of Physics: Conference Series* **1008** (1), 012006.

First received 1 December 2020; accepted in revised form 19 February 2021. Available online 23 March 2021