Journal of Soft Computing in Civil Engineering 6-2 (2022) 51-67

Contents lists available at **SCCE**

Journal of Soft Computing in Civil Engineering

Journal homepage: www.jsoftcivil.com

Assessment of Statistical Models for Rainfall Forecasting Using Machine Learning Technique

L. Gowri[1], K.R. Manjula[1], K. Sasireka[2], Durairaj Deepa[2*]

1. School of Computing, SASTRA Deemed to be University, Thanjavur, India
2. School of Civil Engineering, SASTRA Deemed to be University, Thanjavur, India

Corresponding author: *deepa@src.sastra.edu*

https://doi.org/10.22115/SCCE.2022.304260.1363

ARTICLE INFO

Article history:
Received: 11 September 2021
Revised: 07 April 2022
Accepted: 08 April 2022

Keywords:
ARIMA;
ETS;
Holt-winters model;
Time series.

ABSTRACT

As heavy rainfall can lead to several catastrophes; the prediction of rainfall is vital. The forecast encourages individuals to take appropriate steps and should be reasonable in the forecast. Agriculture is the most important factor in ensuring a person's survival. The most crucial aspect of agriculture is rainfall. Predicting rain has been a big issue in recent years. Rainfall forecasting raises people's awareness and allows them to plan ahead of time to preserve their crops from the elements. To predict rainfall, many methods have been developed. Instant comparisons between past weather forecasts and observations can be processed using machine learning. Weather models can better account for prediction flaws, such as overestimated rainfall, with the help of machine learning, and create more accurate predictions. Thanjavur Station rainfall data for the period of 17 years from 2000 to 2016 is used to study the accuracy of rainfall forecasting. To get the most accurate prediction model, three prediction models ARIMA (Auto-Regression Integrated with Moving Average Model), ETS (Error Trend Seasonality Model) and Holt-Winters (HW) were compared using R package. The findings show that the model of HW and ETS performs well compared to models of ARIMA. Performance criteria such as Akaike Information Criteria (AIC) and Root Mean Square Error (RMSE) have been used to identify the best forecasting model for Thanjavur station.

How to cite this article: Gowri L, Manjula KR, Sasireka K, Deepa D. Assessment of statistical models for rainfall forecasting using machine learning technique. J Soft Comput Civ Eng 2022;6(2):51–67. https://doi.org/10.22115/scce.2022.304260.1363

1. Introduction

Agriculture is considered to be a back bone of countries such as India. One of the leading states for agriculture is Tamil Nadu. Thanjavur is depicted as a rice bowl of TamilNadu from its historical era. Surface water and ground water are the main sources for the development of agriculture. The Cauvery River surface water supply is used for the cultivation of major crops such as paddy, pulses, gingelly, groundnut, and sugarcane. The increase in surface water is mainly based on the distribution of rainfall across the region. Due to inadequate water from the Cauvery River, most of the farming area in Thanjavur district depends on the seasonal rainfall. Taking these into account, rainfall forecasting over a prolonged duration will help to plan the management of irrigation water and associated preparation.

To unravel hydrological problems, including forecasting rainfall, the Machine Learning (ML) approach is widely used. The value of this modelling is that the ability of the software to plot input-output patterns without the aforementioned knowledge of the factors affecting the forecast parameters is important [1–3].

This forecast primarily benefits farmers and it is possible to use water supplies effectively as well. Rainfall forecasting is a difficult job and the findings should be correct. By using weather conditions including temperature, humidity, pressure, there are several hardware devices for predicting rainfall. These conventional approaches do not work efficiently, so we can achieve precise results by using machine learning techniques. By using historical data analysis of rainfall in machine learning, it can forecast rainfall for future seasons. Many techniques can be applied, such as classification, regression according to requirements, and we can also quantify the error between the actual and forecast, as well as the precision. Different methods produce different accuracies, so choosing the right algorithm and modelling it according to the requirements is crucial.

Researchers [4–7], Developed Autoregressive Integrated Moving Average (ARIMA) for prediction of monthly rainfall data forecast in the Indonesian region of Wagis and Pujion. Hoa [4] developed a technique to predict weather forecasting with the help of image fuzzy clustering and spatiotemporal using satellite appearance. By using the fuzzy clustering method, the satellite image pixels were divided into clusters. The Fourier transformation method was used to filter out random images, using the regression method to forecast the expected sequence of appearance. The combine prediction model for monthly mean prediction used to increase the accuracy of precipitation prediction along with error correction [8]. Using cross validation with models to try to predict the optimal prediction for rainfall data with difference time horizons [9,10].

Thanjavur, often known as Tamil Nadu's rice bowl, has been noted for paddy production since the Chola dynasty. It is situated in the Cauvery Delta region, which has both the necessary criteria for paddy cultivation, namely abundant water and alluvial soil. The North-East monsoon brings roughly 37cm of rain to this region, and the rivers are also a source of water. Due to insufficient water from the Cauvery River, most of the farming area in Thanjavur district depends on the seasonal rainfall. Taking these into account, rainfall forecasting over an extended period can help to plan the management of irrigation water and associated planning. Instant comparisons between past weather forecasts and observations can be processed using machine learning. Weather models can better account for prediction flaws, such as overestimated rainfall, with the help of machine learning, and create more accurate predictions. The proposed research of time series analysis and rainfall forecasting at Thanjavur station is being performed in an open-source data mining environment called R. In order to find the best model for the research field, a comparative study of the three models was carried out: ARIMA, ETS and Holt-winters [11,12]. The performance assessment revealed that the HW model outperformed the ARIMA and ETS model.

2. Study area

Thanjavur is a city with the population close to 225,000 people, located in the state of Tamil Nadu, South India. The latitude of Thanjavur, Tamil Nadu, India is 10.7816° N, and the longitude is 79.1390° E. The Cauvery Delta Zone's daily rainfall is 956 mm, and the Cauvery River is the main source of irrigation for cultivation in this district. With its fertile soil, the Thanjavur District is not only one of the largest paddy cultivation areas in Tamilnadu but also in South India. For the present analysis, 17 years of historical rainfall data from 2000 to 2016 were collected and seasonal trend of the rainfall in this study area is represented in the time series plot is shown in Fig.1.

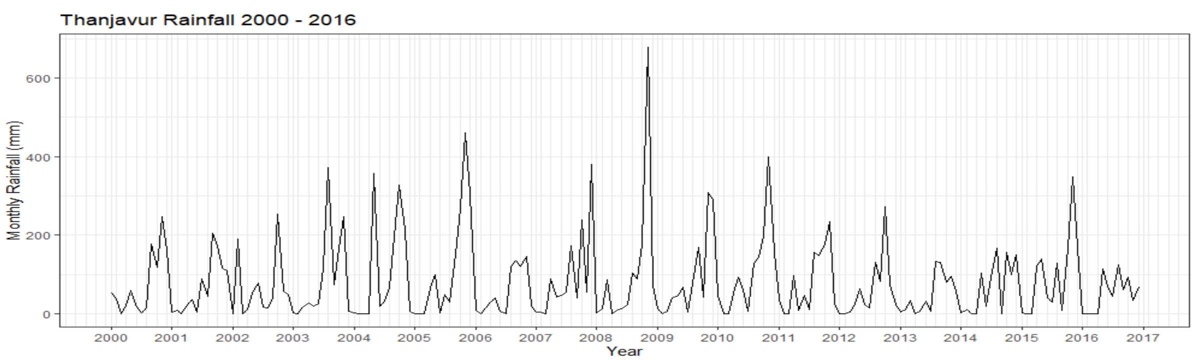

Fig. 1. Thanjavur station Rainfall Time Series plot.

The plot of the time series reveals that rainfall has a seasonality pattern without any trends. Fig.1 illustrates that two peaks are observed per year in the time series map. In the North-East monsoon (October-December), rainfall always hits its higher value and this pattern is always repeated from year to year during the periods 2000-2016. The study area taken for rainfall prediction is depicted in Fig 2 and the flow diagram, the details of the methods adopted for current research work are explained in Fig.3.

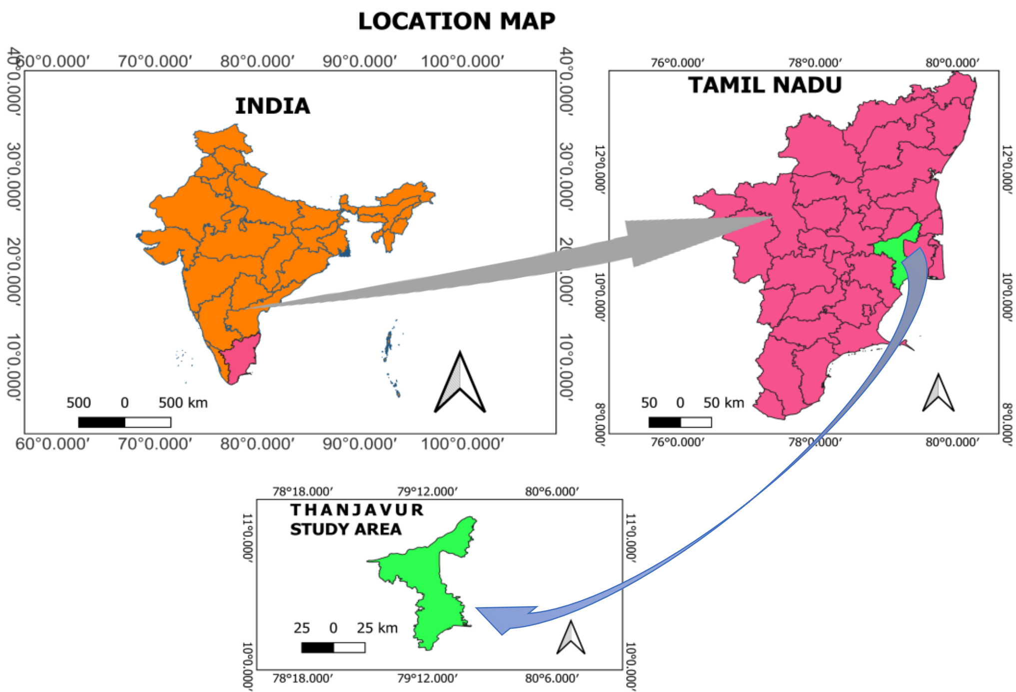

Fig. 2. Study area map.

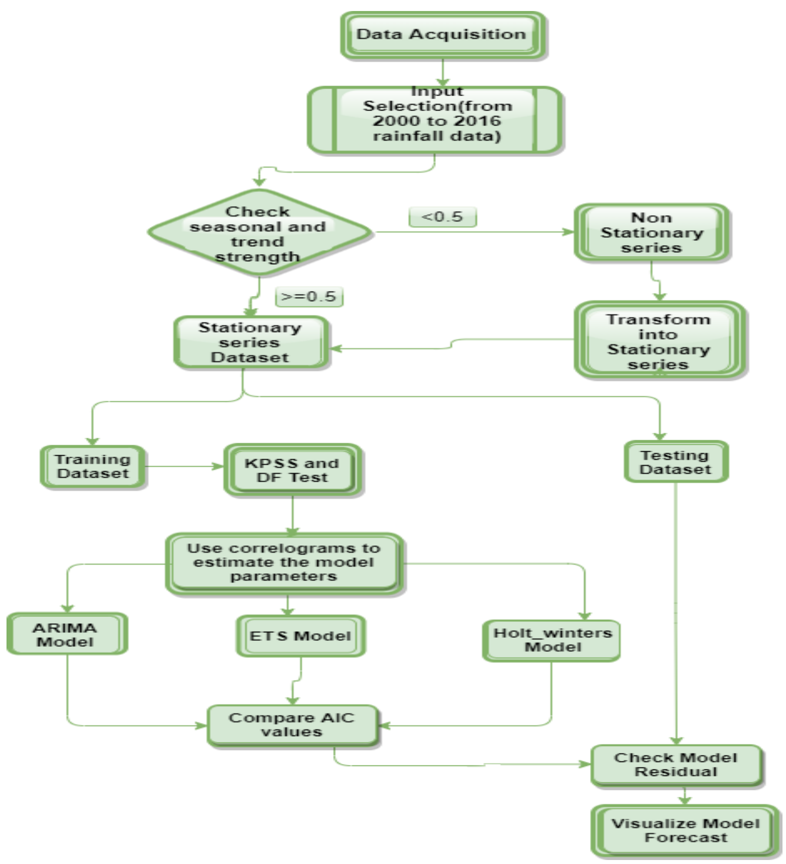

Fig. 3. Flow diagram for rainfall forecasting.

3. Methodology and data analysis

The Holt-Winters model, the ETS model and the ARIMA model are the models used in this analysis ([7,8]. Monthly rainfall data for Thanjavur station for 17 years (2000 to 2016) was used to verify the best method for rainfall forecasting in the study area. The data to be processed is imported into the R environment using the Time series ts() function and then translated to a time series object.

The rainfall data collected in the study area should be tested for seasonal and trend strength. The seasonal or trend strength is greater than 0.5 and is then taken into account as a seasonal or trend analysis. This verification is used to find that either a stationary or a non-stationary dataset belongs to the given dataset. If the dataset is not stationary, the differentiating approach should be modified to stationary. Then the data set is split into training and testing dataset. The training dataset is used test the Kwiatkowski-Phillios-Schmidt_Shin (KPSS) and Augmented Dickey–Fuller test (DF) test (R. J. Hyndman, 2019).

A. ARIMA model

For rainfall model estimation and univariate forecasting, the ARIMA model is used. It has three elements (p,d,q). p stands for the number of lags of autoregressive (AR); d stands for the degree of differencing (I) that helps as a stationary sequence and can be determined between previous values and data values; q stands for the number of lags of moving average (MA). The 'MA' terms are called error terms, which help to predict observations of current and future data. This eliminates the random movements of time series values [13,14].

The ARIMA model components:

$$RF_t = a + \sum_{i=1}^{p} b_i\, RF_{t-i} + d_0 e_t + \sum_{j=1}^{q} d_j\, e_{t-j} \quad (1)$$

Where RF_t is monthly rainfall in time t.$The\ e_t$ and e_{t-1} is the value of error term and immediate past error known at time t. The p and q are number of lags of dependent variable and error term respectively.

B. ETS Model

Trends and seasonal components are the focus of the ETS model. The components of the trend are expressed as N(none), A (Additive), Ad (Additive Damped), M(Multiplicative), Md (Multiplicative Damped) [15,16]. The season is seen in the series as repeating the short-term pattern of the cycle. The seasonal components are expressed as N(none), A (Additive),M(Multiplicative). The forecast distributions are usual for models with only additive components, so the medians and means are equal. In ETS, the default is AICc. The model that minimizes the standard is chosen as acceptable for the information criteria.

AIC (Alkies' Information Criteria) is: $AIC = -2(L) + 2k$ (2)

$$AICc = AIC + 2(k+1)(k+2)n - k \quad (3)$$

$$AICc = AIC + 2(k+1)(k+2)n - k \quad (4)$$

Forecasting Technique is used to do forecasting with the help of the ETS function, which can be used with R. The following steps are taken to obtain a generally applicable and robust ETS Model for autonomous forecasting: 1. For each series, apply all methods that are appropriate, optimising the model (both the Smoothing Parameter and, as a result, the starting state variable) in each case. 2. Choose the best model based on the AICc value. 3. Create a point forecast after selecting the model with improved parameters. 4. To acquire the prediction intervals for the most effective model.

C. Holt-Winters Model

The Holt Winters model uses an exponential smoothing of the performance and forecasting distribution of time series. Three aspects of the time series were used in this model: level, trend and seasonal values. The future value is predicted using several parameters, such as alpha (a), gamma (γ) and beta (β) . It also utilizes frequency seasonality to be denoted as M. Two variations that help to differ in the nature of the seasonal components were used by this method. When seasonal variations are constant, the additive method is chosen. When seasonal variations change in proportion to the average of the time series, the multiplicative method is chosen.

Holt-Winters additive method components:

Level formula:

$$L_t = \alpha\left(\frac{y_t}{St^{-M}}\right) + (1-\alpha)(\alpha_{t-1} + T_{t-1}) \quad (5)$$

Trend formula:

$$T_t = \beta\left({}^{L_t}/_{L_{t-1}}\right) + (1-\beta)(T_{t-1}) \quad (6)$$

Seasonal formula:

$$S_t = \gamma\left(\frac{y_t}{\alpha_t}\right) + (1-\gamma)s_t - M \quad (7)$$

The level formula shows a weighted average between the seasonal observation and the non-seasonal forecast forT_t. The trend formula is matching to Holt's linear method. The seasonal formula shows an average between current seasonal index and the seasonal index of the same seasonal year (M).

Analysis of data

- The time series has been decomposed to get more detail about Trend, Seasonality, and Remainder component and flow diagram is explained in the fig.4.

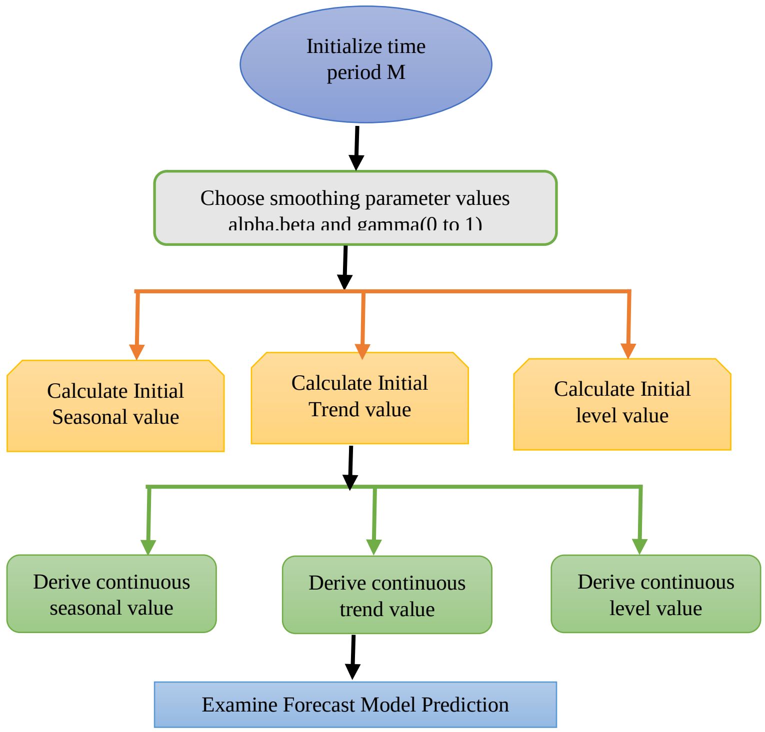

Fig. 4. Proposed model flow diagram.

The Akaike information criterion (AIC) is a precise method for estimating how well a model fits using the rainfall forecast data. It is used to compare different conceivable model samples and govern which one is the best fit for the rainfall forecast data. This is named entropy maximization principle and minimizing AIC values is equivalent to maximizing entropy and helps to measure the relative loss of information. Generally, AIC is calculated from the number of independent variables used to form the model and the maximum likelihood approximation of the model.

$$AIC = 2k - 2\ln(\hat{L})$$

K is the number of estimated parameter variables used and **L** is the log-likelihood estimate parameter which is used for the model measure.

Mean Absolute Error (MAE) are metrics used to evaluate the average of absolute value of the errors. The metrics helps to know how the model prediction rainfall forecast values are accurate and calculate the amount of deviation from the actual rainfall forecast values. This helps to predict the rainfall forecast based on the numbers of rainfall samples consider for the measurement.

$$MAE = \sum_{i=1}^{n} y_i - x_i$$

Where, n is the total number of rainfall samples, y_i is the model rainfall forecasts values and x_i is the true rainfall samples.

Root Mean Squared Error (RMSE) is the square root of mean squared error, used as a standard statistical parameter to measure the model performance of rainfall forecast data. The model parameter indicated the standard deviation of residuals of rainfall forecast data.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(f_i - o_i)^2}{n}}$$

Where, n is the number of rainfall samples, f is the model rainfall forecasts values and o is the observed rainfall samples. The RMSE is a good indicator to evaluate the performance of the interpolation values. Decomposition is performed using the stl() function and divides the time series automatically into three components (Trend, Seasonality, Remainder) shown in Fig. 5

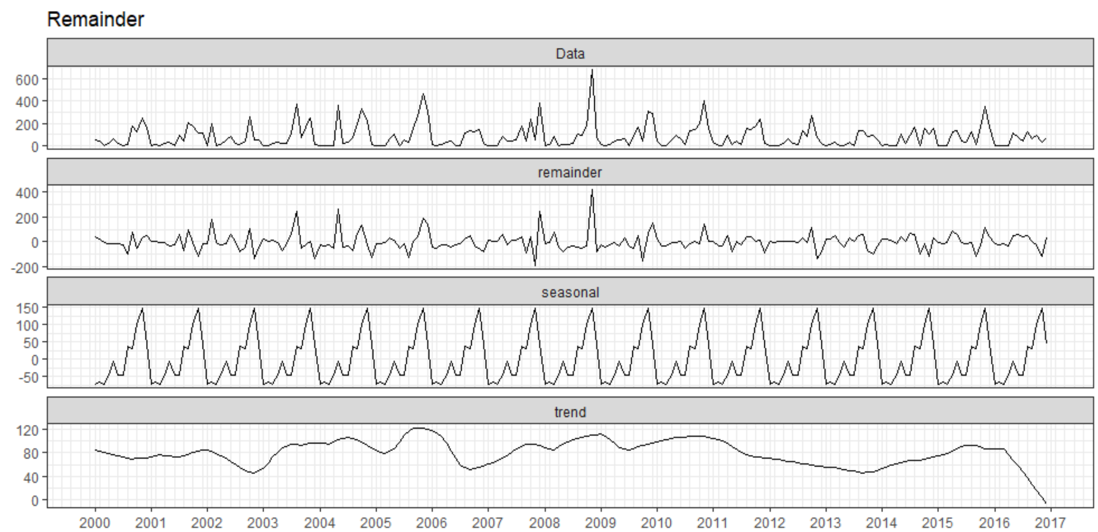

Fig. 5. Time series decomposition.

- Calculation to assess trend and strength of seasonality

Ft: Trend Strength

$$F_T = max(0,1 - \frac{Var(R_t)}{Var(T_t+R_t)}) \quad (7)$$

Fs: Seasonal Strength

$$F_s = max(0,1 - \frac{Var(R_t)}{Var(S_t+R_t)}) \quad (8)$$

The strength of the seasonal and trend ranged between 0 and 1, while ,1, indicates that the trend and seasonal occurred very strongly. In the present study the Trend strength is 0.1 and Seasonal strength is 0.5, it shows that the dataset follows seasonal pattern alone and it doesn't follow the trend pattern. It shows that our data is comes under stationary dataset. In Fig.5 the seasonal subseries plot will provide a much more informative interpretation of our data. Seasonal subseries plots are a tool for detecting seasonality in a time series.

Pseudocode: Best model selection:

Input: rainfall data for Thanjavur region

Output: Best fit for forecast model

1.If seasonal_strength>=0.5 and/or trend_strength>=0.5 then Dataset is stationary series.

Else Tansform as stationary series.

2. Split the dataset into training and testing sets.

3. Calculate statistical values using KPSS and DF method.

4. visualize ACF and PACF lag values for model parameters.

5. Train the dataset using different models:

5.1 ARIMA(p,d,q)(P,D,Q)

5.1.1 (p,q)= (i, i) where i= 0 to 4

If p=1 and d=0 and q=0 then AR model

else if p=0 and d=0 and q=1 then MA model

else if p=1 and d=0 and q=1 then ARMA model

5.2 ETS(A,Ad,A)

5.2.1 compare the seasonality component with remainder values.

5.2.2 if output_components = independent then additive series parameters

Else

multiplicative series parameters

5.3 Holt_Winters (L, T, S)

5.3.1 fix initial seed value of α, γ and, β

5.3.2 calculate initial seasonal (S), Level (L), Trend (T) factors

5.3.3 check the parameters as additive or

multiplicative components

6. Find the residuals and apply diagnostic test. If the residuals are good then fit the model. Otherwise repeat the same process go to 5 and change the parameter values.

7. Custom the fitted model for forecasting.

4. Result and discussion

The prediction of rainfall at Thanjavur for the time series is carried out by the construction of ETS, ARIMA and Holt-winters models. Out of the available 17-year monthly data, 10-year data from 2000 to 2009 is taken as training, 2010 to 2014 is taken as testing, and the prediction for the next two years from 2014 to 2016 is attained. The resulting prediction is correlated with the real rainfall data and plotted against it.

Fig. 6 shows that the rainfall gradually increases from October and reach its maximum value in the month of November due to NE (North-East) monsoon season and decreases gradually and reach its minimum value in the month of March. Rainfall will begin to increase again after March and reach its maximum value in the month of August and September due to SW (South-West) monsoon. It depicts monthly average rainfall data for four time periods (based on industrial development and urbanisation phases). Significant changes in monthly rainfall have been discovered in the plot over the years and in the years to come. Monthly rainfall increased from March to September, indicating more rain in the pre-monsoon (March-May) and monsoon (June-September) seasons. Papalaskaris et al. [17] reported a similar pattern when estimating rainfall over Bangladesh. Excessive rain will result in major floods, putting crops at risk and causing waterlogging in the city. On the other side, a similar falling (December-January) rainfall trend was observed in October-November, followed by an increase in February, indicating a lower rainfall and dryer crop season.

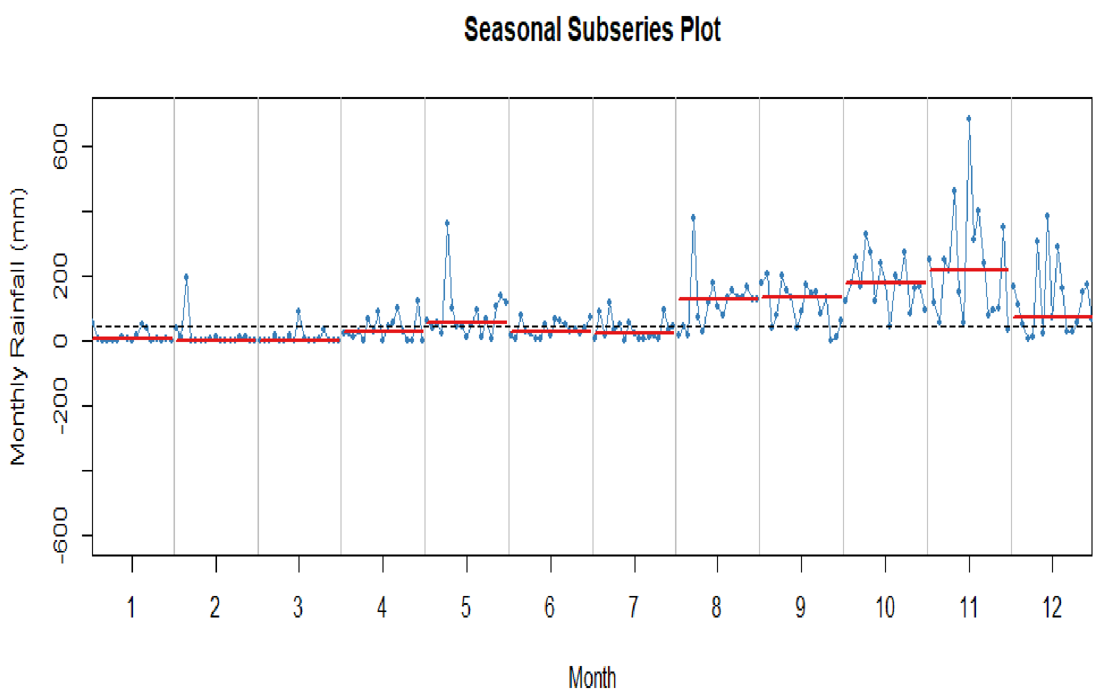

Fig. 6. Seasonal Subseries Plot.

4.1. Comparison of three models

A statistical model is the use of statistics to build a representation of the data and then conduct analysis to infer any relationships between variables or discover insights. Machine learning, on the other hand, is the use of mathematical or statistical models to obtain a general understanding of the data to make predictions. Still, many in the industry use these terms interchangeably. While some may not see any harm in this, a true data scientist must understand the distinction between the two.

1. ARIMA Model

Our data is given under a seasonal data set based on the strength and seasonal test, so it is regarded as stationary data. Six types of ARIMA models are used in this study and the best method out of six ARIMA models is chosen based on the AIC value. The capacity of the selected ARIMA model for precipitation and temperature (maximum and lowest) to evaluate the relative quality of statistical model for a given dataset is examined using AIC criterion. The Akaike Information Criterion (AIC) is a constant estimate plus the distance between the unknown true likelihood function of the data and the fitted likelihood function of the model, with a lower AIC indicating that the model is closer to the truth. In other words, AIC calculates the amount of information lost by a particular model, with the lower the amount of information lost, the higher the model's quality.

Table 1
Accuracy level of ARIMA model.

ARIMA (p,d,q) Model	AIC value
M1 (1,0,1)	1475.799
M2 (1,0,2)	**1454.979**
M3 (0,0,2)	1472.167
M4 (2,0,1)	1455.255
M5 (2,0,2)	1455.879
M6 –auto ARIMA	1463.207

Fig. 6 displays the Ljung-Box test and the ACF plot of model residuals. From Fig.6 it can be concluded that this model is acceptable for forecasting as its residuals represent the behaviour of white noise and are uncorrelated to each other.

2. ETS model

ETS stands for Error Trend Seasonality. The ETS stands for exponential smoothing state space models that effectively fit the data (A, Ad, M). The parameters that were utilised to create these models, which were chosen in order to produce data that appeared to be reasonably realistic. The method clearly has a high success rate in determining whether the errors are additive or multiplicative. The optimum result is obtained in ETS model when the Trend is treated as Additive series and Error and Seasonality are treated as Multiplicative series. After a residual check, ACF diagram shown in Fig. 7 demonstrates that the majority of sample autocorrelation coefficients of residuals from the fitted ETS state space models are within the model's bounds, implying that the residuals are white noise and the models are appropriate. The test results reveal that there are no autocorrelations in the in-sample forecast errors, as well as the distribution of forecast errors, confirming the evidence of no autocorrelations. This shows that the simple exponential smoothing method can be used to estimate rainfall with reasonable accuracy.

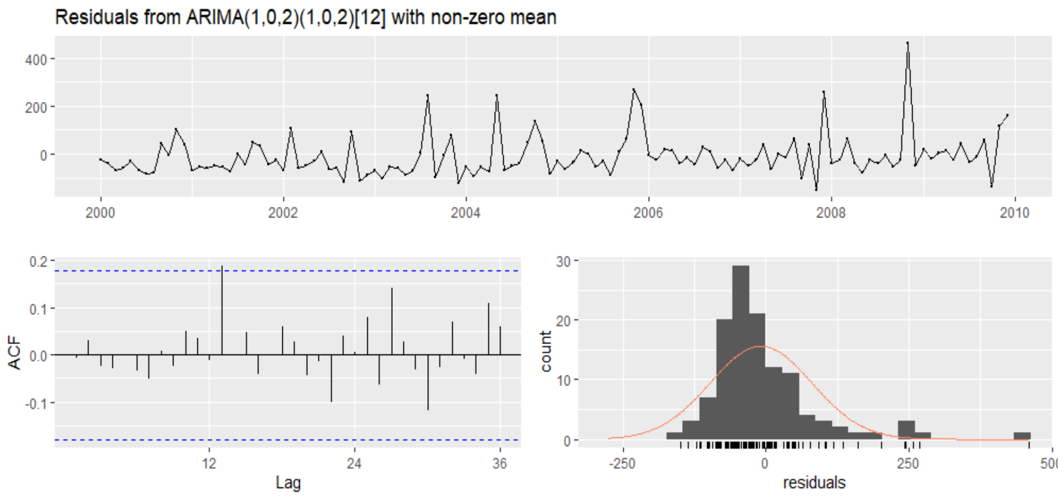

Fig. 7. Residual check on ARIMA model.

3. Holts-Winters Model

Holt-Winters model is also known as Triple Exponential smoothing. Here the given observed data is decomposed into seasonal, level and trend. The exponential weighted moving average of all three components is then blended and result is obtained. Prediction by this model (Fig.8) is also similar to the previous model. And there is a sign of little improvement in low magnitude rainfall. But there is no proper estimation of peak rainfall reported in the monsoon months.

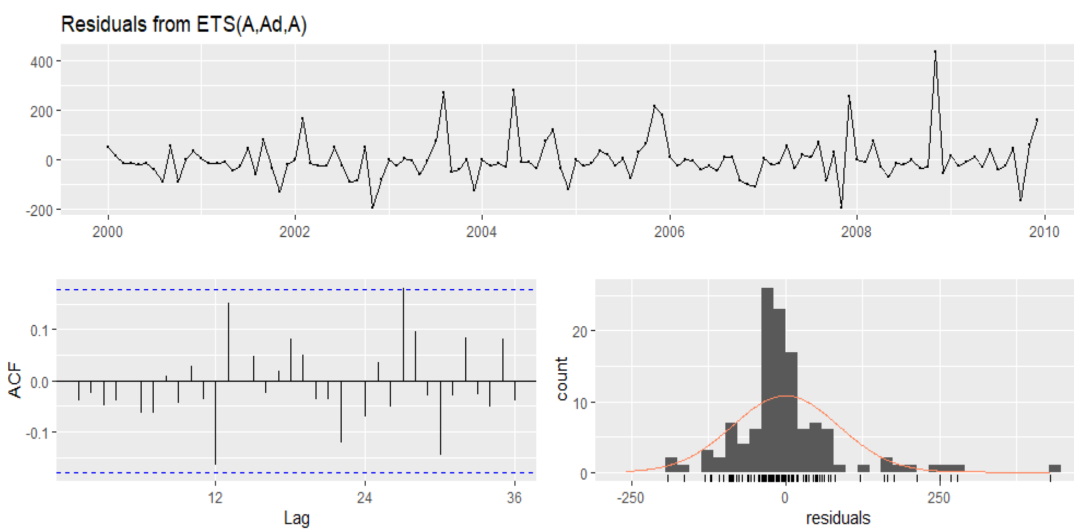

Fig. 8. Residual check on ETS model.

The selected model is compared with actual data set and it is shown in Fig.9. The green line represents the actual data ranges from 2000 to 2016. The other models ARIMA, ETS and HW are plotted with training data ranges from 2000 to 2009. By comparing actual data with model data, all the models are almost fit the same value with actual data. Based on the accuracy, HW Model doing better in both training and test set compared to ARIMA Model and ETS model.

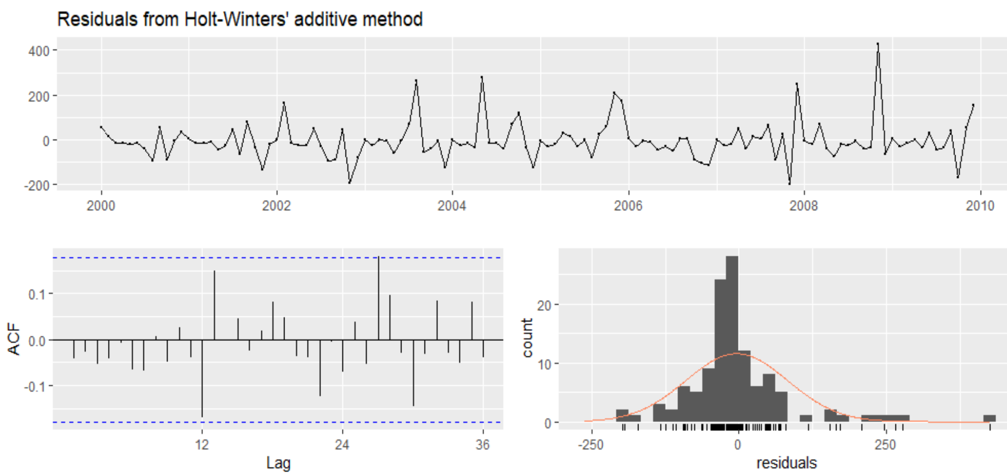

Fig. 9. Residual check on HW model.

The selected model is compared with actual data set and it is shown in Fig.10. The green line represents the actual data ranges from 2000 to 2016. The other models ARIMA, ETS and HW are plotted with training data ranges from 2000 to 2009. By comparing actual data with model data, all the models are almost fit the same value with actual data. Based on the accuracy, HW Model doing better in both training and test set compared to ARIMA Model and ETS model.

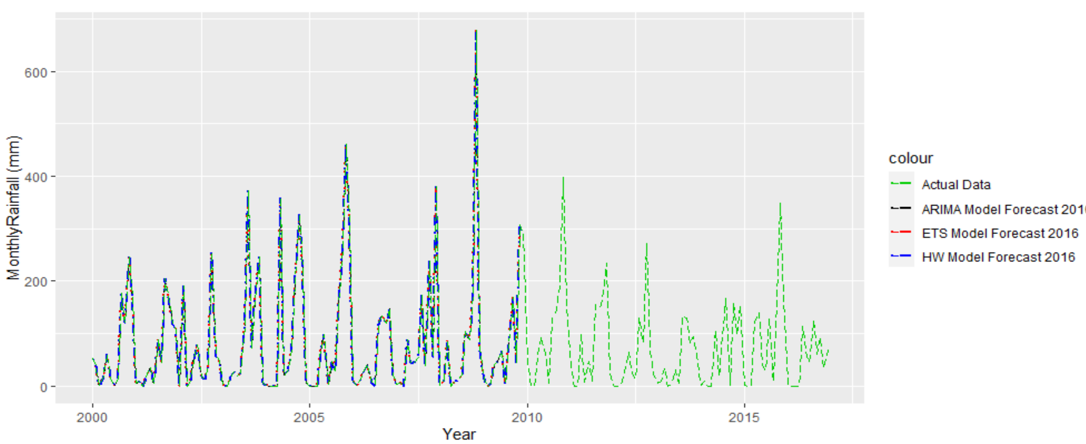

Fig. 10. Actual Data vs ARIMA, ETS and Holt-winters Forecasting.

Forecasting was done using three models, ARIMA, ETS and HW is shown in Fig. 11 to Fig. 13 respectively. The models show similar movement based on the plot with the lowest value of rainfall will occur beginning month of each year as well as it follows the seasonal rainfall pattern of our study area. By comparing the ETS and HW forecasting models, both the model predicts similar way and ARIMA model slightly differ than the other models. The performance of the model is evaluated with reference to Root Mean Squared Error (RMSE), AIC value and model fit.

The RMSE and AIC values for models are given in Table 2. Both the RMSE and AIC value reveal that HW model is outperforming the rest of the models. It can be seen from the Table II that the highest accuracy is reported for HW model followed by ETS and ARIMA model. HW model has better correlation with actual values. Hence, the results shows that the HW as well as ETS models are suitable to predict future rainfall and seasonal pattern of the rainfall in the study area. This prediction of rainfall using ML can be useful for a farmer who wants to know when is the best month to start planting, as well as for the government who needs to prepare some strategy to avoid rainy season floods and dry season drought. The most important thing is that this forecast is based only on the historical average, using meteorological data and some knowledge from climate experts to incorporate the more detailed forecast. The future work focus on the same data set will be applied in the recurrent neural network-based prediction and try to improve accurate results [3,17,18]. As a result, the additive Holt-Winters approach is recommended for future forecasting above the multiplicative Holt-Winters method. The anticipated values will aid disaster management in determining future rainfall patterns, whether drought or flooding is expected. Furthermore, it will assist farmers in making timely decisions on the seeding of crops, fruits, and dried fruits.

Table 2
Comparison of three Models.

Model	RMSE	MAE	AIC value
ARIMA	54.287	39.474	1454.979
ETS	49.158	37.460	1452.286
HOLT-WINTERS	48..670	36.751	1450.817

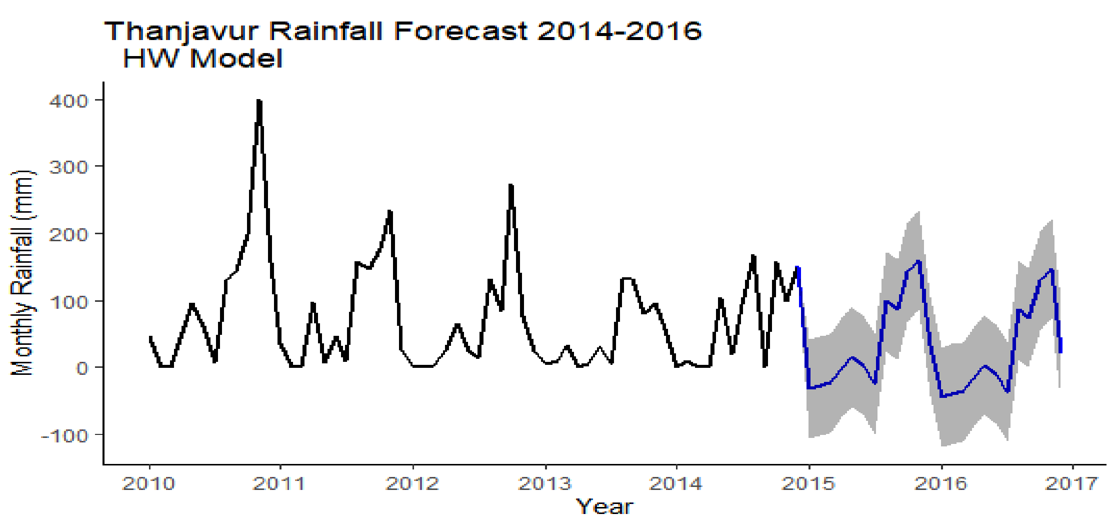

Fig. 11. Prediction of monthly rainfall using ARIMA model.

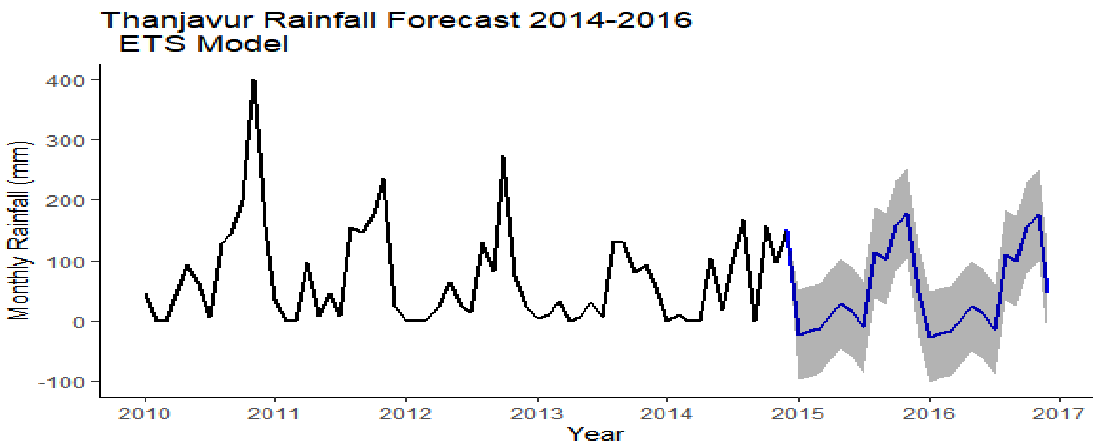

Fig. 12. Prediction of monthly rainfall using ETS model.

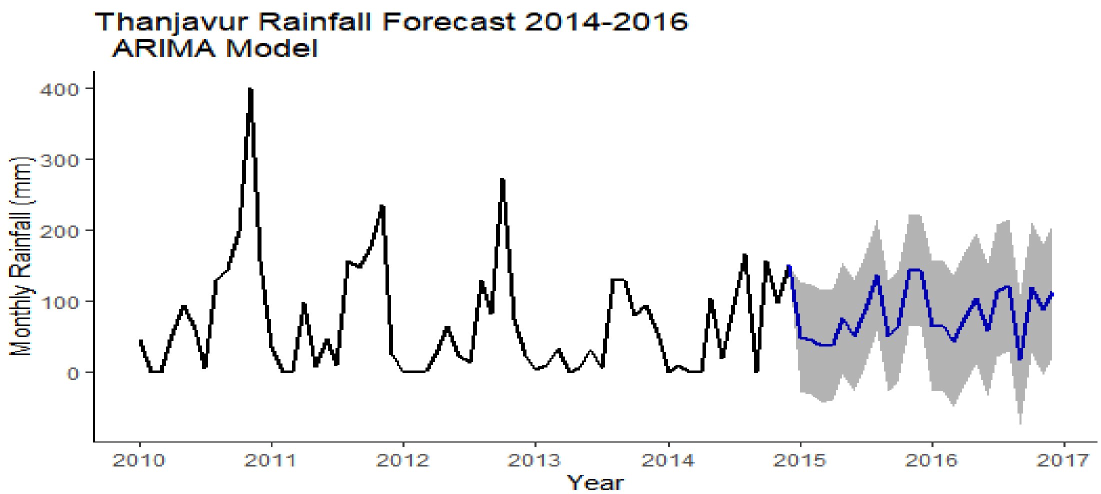

Fig. 13. Prediction of monthly rainfall using Holt-Winters model.

Given the fact that it does not rain much during the dry season, there is a nonsignificant positive relationship between rainfall and average temperature from November to January, indicating that a small increase in average temperature results in more rainfall. In any other month, there is no notable relationship. During the Pre-Monsoon and Post-Monsoon seasons, rainfall and temperature have a slight inverse relationship. Despite the fact that there is no significant yearly relationship, temperature fluctuates unfavourably during Rabi season and favourably during Kharif season.

5. Conclusion

In the present study, we have reported the time-series analysis and comparative study of machine learning models for the forecasting of rainfall at Thajnavur station of Tamilnadu. The dataset consists of monthly rainfall updates from January 2000 to December 2016. The time-series data is visualized by plotting time-series plot and correlation plots. For the timeseries forecasting of rainfall at Thanjavur station is carried out by building ARIMA, ETS and Holt-winters models.

The performance of the model is evaluated with reference to Root Mean Squared Error (RMSE), MAE and AIC value. The comparative analysis revealed that HW model accurately forecasts the rainfall with less error. Thus, derived model could be used to forecast monthly rainfall for the upcoming years. Research concludes that the imperative issue of accurate forecasting of rainfall can be handled by machine learning models. It is significant to mention that, while model forecasts cannot predict exact precipitation amounts, they can reveal the likely trend of future rains and provide information that can assist decision-makers in developing strategies in areas such as agriculture, where knowing the start and end of rainy seasons is critical, civil works planning, and the time to prepare of mitigation plans for natural hazards, such as flooding. Finally, it's worth noting that rational planning and complete management of water resources necessitate forecasting future events while keeping in mind that most forecasts are based on previous events.

References

[1] Hipni A, El-shafie A, Najah A, Karim OA, Hussain A, Mukhlisin M. Daily Forecasting of Dam Water Levels: Comparing a Support Vector Machine (SVM) Model With Adaptive Neuro Fuzzy Inference System (ANFIS). Water Resour Manag 2013;27:3803–23. https://doi.org/10.1007/s11269-013-0382-4.

[2] Najah A, El-Shafie A, Karim OA, Jaafar O. Integrated versus isolated scenario for prediction dissolved oxygen at progression of water quality monitoring stations. Hydrol Earth Syst Sci 2011;15:2693–708. https://doi.org/10.5194/hess-15-2693-2011.

[3] Mahsin M, Akhter Y, Begum M. Modeling Rainfall in Dhaka Division of Bangladesh Using Time Series. J Math Model Appl 2012;1:67–73.

[4] Tektaş M. Weather Forecasting Using ANFIS and ARIMA MODELS. A Case Study for Istanbul. Environ Res Eng Manag 2010;1:5–10. https://doi.org/10.5755/j01.erem.51.1.58.

[5] Sciences E. Time Series Analysis Model for Rainfall Data in Jordan : Case Study for Using Time Series Analysis P . E . Naill M . Momani King Abdul Aziz University , Jeddah , Kingdom of Saudi Arabia. Am J Environ Sci 2009;5:599–604.

[6] Shamsnia SA, Shahidi N, Liaghat A, Sarraf A, Vahdat SF. Modeling of weather parameters using stochastic methods (ARIMA model)(case study: Abadeh Region, Iran). Int Conf Environ Ind Innov IPCBEE 2011;12:282–5.

[7] Suhartono, Faulina R, Lusia DA, Otok BW, Sutikno, Kuswanto H. Ensemble method based on ANFIS-ARIMA for rainfall prediction. ICSSBE 2012 - Proceedings, 2012 Int Conf Stat Sci Bus Eng "Empowering Decis Mak with Stat Sci 2012:240–3. https://doi.org/10.1109/ICSSBE.2012.6396564.

[8] Li G, Chang W, Yang H. A Novel Combined Prediction Model for Monthly Mean Precipitation with Error Correction Strategy. IEEE Access 2020;8:141432–45. https://doi.org/10.1109/ACCESS.2020.3013354.

[9] Vienna A. R Core Team R: A language and environment for statistical computing 2017.

[10] Hyndman [10] R. J. Forecasting functions for time series and linear models_. R package version 8.2. 2017.

[11] Mila FA, Parvin MT. Forecasting Area, Production and Yield of Onion in Bangladesh by Using ARIMA Model. Asian J Agric Extension, Econ Sociol 2019:1–12. https://doi.org/10.9734/ajaees/2019/v37i430274.

[12] Punia M, Joshi PK, Porwal MC. Decision tree classification of land use land cover for Delhi, India using IRS-P6 AWiFS data. Expert Syst Appl 2011;38:5577–83. https://doi.org/10.1016/j.eswa.2010.10.078.

[13] Burlando, P.; Rosso, R.; Cadavid, L.G.; Salas J. Forecasting of short-term rainfall using ARMA models. J Hydrol 1993:144: 193–211.

[14] Salas, J.D.; Obeysekera JT. ARMA model identification of hydrologic time series. Water Resour Manag 1982;18:1011–1021.

[15] Ridwan WM, Sapitang M, Aziz A, Kushiar KF, Ahmed AN, El-Shafie A. Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. Ain Shams Eng J 2021;12:1651–63. https://doi.org/10.1016/j.asej.2020.09.011.

[16] Valipour M. Ability of Box-Jenkins Models to Estimate of Reference Potential Evapotranspiration (A Case Study: Mehrabad Synoptic Station, Tehran, Iran). IOSR J Agric Vet Sci 2012;1:01–11. https://doi.org/10.9790/2380-0150111.

[17] Papalaskaris T, Panagiotidis T, Pantrakis A. Stochastic Monthly Rainfall Time Series Analysis, Modeling and Forecasting in Kavala City, Greece, North-Eastern Mediterranean Basin. Procedia Eng 2016;162:254–63. https://doi.org/10.1016/j.proeng.2016.11.054.

[18] Thakkar AK, Desai VR, Patel A, Potdar MB. Post-classification corrections in improving the classification of Land Use/Land Cover of arid region using RS and GIS: The case of Arjuni watershed, Gujarat, India. Egypt J Remote Sens Sp Sci 2017;20:79–89. https://doi.org/10.1016/j.ejrs.2016.11.006.