Int. J. Econ. Environ. Geol. Vol. 13 (1) 37-41, 2022
Journal home page: www.econ-environ-geol.org
Open Access
ISSN: 2223-957X

Modeling and Forecasting of Rainfall Time Series. A Case Study for Pakistan

Tayyab Raza Fraz

Department of Statistics, University of Karachi, Pakistan

*Email: tayyab.fraz@uok.edu.pk

Received: *09 November, 2021* ***Accepted:*** *27February, 2022*

Abstract: The change of weather conditions is considered as the major problem, particularly for developing country like Pakistan. Machine learning and artificial neural network models have become attractive forecast techniques for rainfall as compared to traditional statistical methods in the last few years. The behavioral pattern in rainfall (mm) annually by 1901 to 2020 is studied. Moreover, forecasts of three models based on past observations are evaluated. Fundamentally, different techniques are used for model development. Three modeling techniques include a traditional linear time series ARMA model, an emerging nonlinear threshold technique SETAR model, and influential machine learning technique NAR model. Evaluation of forecast performance is based on three forecast error criteria namely MSE, RMSE, and MAPE. Results indicate that the rainfall (mm) will slightly increase in the coming ten years i.e. 2021 to 2030. Furthermore, the findings also reveal that the NAR model is a suitable and appropriate model to forecast the rainfall which outperforms the ARMA as well as the SETAR model.

Keywords: Rainfall, forecasts, nonlinear, threshold, machine learning.

Introduction

Rainfall is one of the foremost factors which is responsible for many of the significant effects across the world. Rainfall does not only have a significant impact on the environmental factors, but also has a decisive relation with the macroeconomic financial factors. Rainfall has wide-ranging impacts on vital economic indicators i.e. agricultural yields and human capital (Damania et al., 2020). Statistics associated with co-movements of rainfall and other financial indicators i.e. gross domestic product (GDP) growth etc. indicate that there is an impact of rainfall and economic growth. Agriculture is exclusively reliant on rainfall. Irregular rainfall mostly destroys crops, farms and damage property. It is essential to find a forecast model that can predict rainfall accurately. These predictions will be used as an early warning which certainly reduces risks to life, property, and agricultural farms. However, forecasting rainfall is tricky and challenging mainly due to atmospheric processes. Mostly the climate is diverse in Pakistan, which is an agricultural, agronomic and industrialized country. Recently, Alam et al. (2021) stated that agriculture depends on the occurrence and quantity of rainfall. Drastic climate changes cause floods and glaciers to melt. Recent research in this domain mostly focused to find the correlations among the weather features. Nevertheless, requirement of rainfall data training, algorithms, and forecasts are also necessary. Boochabun et al. (2004) used dynamic harmonic regressions while many others (Nwokike et al., 2020; Ramirez et al., 2005; Hung et al., 2009; Khalili et al., 2016) used models to forecast rainfall data. In this study, three different techniques are used to model and forecast the yearly rainfall data. Techniques included linear time series ARMA model, nonlinear threshold SETAR model, and machine learning NAR model. Root mean square error, mean absolute error and mean absolute percentage error i.e. MAE, MSE, and RMSE, respectively are used to evaluate the forecast performance. Nwokike et al. (2020) revealed that the SARIMA model has low forecast error as compared to the Seasonal ANN model based on RMSE, MSE, forecast error (FE), and Mean forecast error (MFE). They used monthly rainfall data from 2006 to 2016 from Umuahia, Abia state of Nigeria. Olatayo and Taiwo (2014) studied the fluctuations in annual rainfall in Ibadan south west, Nigeria, from 1982 to 2012. It was revealed that the forecast performance of the fuzzy time series (FTS) model out performs the ARIMA and Theil's regression model and based on MAE and RMSE. Momani and Naill (2009) used the ARIMA model to estimate the monthly rainfall data from 1922 to 1999, and forecasted monthly rainfall for the upcoming 10 years. Wang et al. (2013) studied empirical mode decomposition i.e. EEMD on a Yellow River in China from 1919 to 1975. It was found that PSO–SVM–EEMD based ANN model improves the forecasting ability based on RMSE, average absolute relative error AARE, and Nash–Sutcliffe efficiency (NSE). Somvanshi et al. (2006) revealed that the ANN model is superior to ARIMA using mean annual rainfall data from 1901 to 2003 of Hyderabad, India, using the last 10 years' mean annual rainfall to evaluate the forecast performance. According to the literature, the results do not favor any individual model.

Materials and Methods

Annual data of rainfall (mm) of Pakistan from 1901 to 2020 is used in this study. Data were taken from the World Bank Group website. Firstly, the descriptive statistics of rainfall are calculated. Augmented – dickey-fuller (ADF) and Phillips-Perron (PP) stationary tests are applied to confirm the presence of

any unit root in annual average rainfall data. After that, data is split into two parts. For estimation, data used from 1901 to 2004, while the remaining part from 2005 to 2020 is used to compare the out-of-sample forecast performance based on three forecast error criteria, namely mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). Lastly, the forecasts from all three models are also compared with the projections from the World Bank climate knowledge portal. Ten years forecast from 2021 to 2030 are also found. Moreover, these one-year-ahead forecasts from 2021 to 2030 are also compared with the World Bank climate projections. The three modeling techniques are as follows:

Linear Time-series ARMA Model

It calculates the future values of time-series data depending on the information of preceding lags without ignoring the residuals. The auto-regressive moving average ARMA (p,q) model can be written as if y_t is stochastic process:

$$y_t = \mu + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots \emptyset_p y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} + u_t \quad 1$$

Where y_t and u_t is ~ WN (0, δ^2)

Non-linear Threshold SETAR Model

The regime that occurs at a time can be determined by an observable variable relative to a threshold value. Self-exciting threshold autoregressive (SETAR) model follows the threshold variable that is chosen to be the lagged value of the time series itself.

$$y_t = x_t \gamma^i + \delta^i \epsilon_t \; if \; \gamma_{i-1} < z_t < r_j \quad 2$$

Here x_t is the column vector of parameters and z_t is an exogenous threshold variable, where non-trivial thresholds distribute the domain of threshold into k regimes.

Machine learning NAR Model

It divides the time series data into training (70%), testing (15%), and validation (15%). Furthermore, it is a discrete model consisting of an input layer, input delay, hidden layer, an output layer, and output delay. The ML nonlinear autoregressive model can be written as:

$$\hat{y} = g\left(y(t-1) + y(t-2) + \cdots + y(t-i)\right) + u_t \quad 3$$

Here $is\,forecasted$ from "y" time series. g is unknown function.

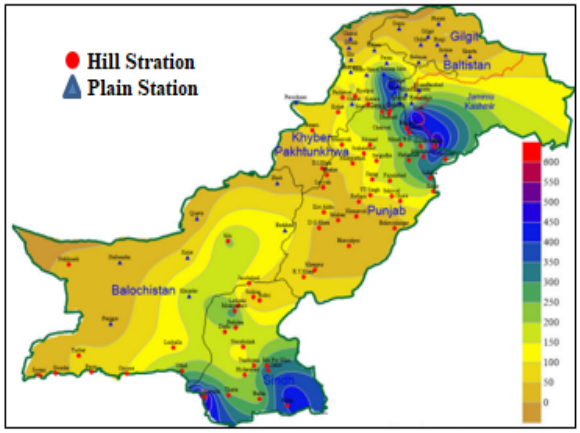

Fig. 1 Rainfall map of Pakistan (2020). Source: Pakistan Meteorological Department, CDPC, Pakistan

Results and Discussion

Graphically, the rainfall data seems to be stationary but shows a non-linear trend (Fig 2). Firstly, a logarithm is taken on the time series rainfall data to stabilize the variance of the data. According to descriptive statistics, the average rainfall in mm is expected to be 5.65 mm (Table 1). Maximum rainfall is expected to be slightly more than 6 mm annually, while the minimum is expected to be 5.25 mm per annum. There is no unit root present in the rainfall data. These findings are according to two stationary tests, namely ADF and PP tests (Table 2). ARMA (4,4) and SETAR AR (4) models are selected based on Akaike and Shawrtz info criteria (Table 3). Forecast performance of traditional ARMA, nonlinear threshold SETAR, and machine learning NAR models are reported in Table 4. Also, the NAR model is selected on the basis of the lowest mean square error (MSE) with ten hidden layers with delay value 2. According to the forecast evaluation criteria, namely RMSE, MAE, and MAPE, the NAR model is found to be the best forecast model. Furthermore, this indicates that the machine learning technique is much easier and accurate as compared to other techniques included in this study. Moreover, the forecasts estimated from all the models are also compared with the projections of the World Bank (Table 3 and Fig 3).

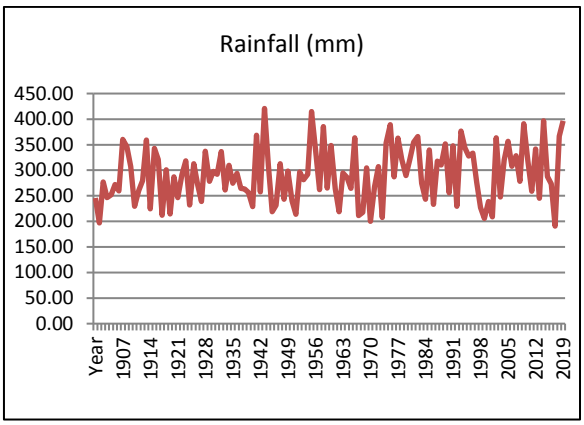

Fig. 2 Pakistan rainfall (mm).

Table 1. Descriptive statistics.

Mean	Median	Max.	Min.	Std. Dev.	Skew	Kurtosis	Sum	Sum Sq. Dev.	Obs.
5.657	5.660	6.042	5.250	0.186	-0.085	2.191	678.918	4.142	120

Table 2. Unit root tests

Unit root tests	**Test Statistic value**	**P-value**
Augmented Dickey-Fuller	-11.452	<0.01
Phillips-Perron	-11.4462	<0.01

Table 3. Forecast evaluation.

Forecast Models	**Log likelihood**	**AIC**	**BIC**
ARMA	38.6513	-0.61468	-0.54346
SETAR	33.0159	-0.53476	-0.48728

Table 4. Forecast evaluation.

Forecast	**RMSE**	**MAE**	**MAPE**
WORLD BANK PROJECTIONS	63.988	54.316	16.863
ARMA	67.684	53.193	16.047
SETAR	66.963	52.488	16.397
NAR	56.229	46.516	15.762

Table 5. Estimated forecasts of rainfall (mm) annually.

Year	**Annual Mean**	**World Bank**	**ARMA**	**SETAR**	**NAR**
2005	320.2	271.1	302.77	296.15	297.19
2006	356.73	296.9	269.53	305.90	299.80
2007	308.04	270.62	284.38	268.73	317.20
2008	328.47	302.58	291.41	295.17	307.82
2009	278.03	261.24	296.95	278.03	325.96
2010	391.04	284.77	261.79	270.95	304.02
2011	317.7	255.77	282.53	282.13	294.20
2012	258.89	275.08	287.23	278.69	300.59
2013	341.38	290.29	299.66	288.54	307.02
2014	245.45	285.82	247.45	268.46	300.39
2015	396.76	271.21	278.59	280.32	301.98
2016	287.45	276.88	297.14	294.09	306.22
2017	273.09	294.64	289.55	277.19	280.19
2018	190.75	266.52	261.77	297.36	313.67
2019	366.53	302.17	251.22	268.91	303.50
2020	396.83	290.45	297.14	287.22	358.35
2021	-	284.38	294.46	290.52	350.78
2022	-	306.51	296.40	308.65	259.06
2023	-	282.02	246.97	276.45	350.30
2024	-	285.63	278.76	272.77	274.35
2025	-	278.81	293.45	287.50	315.82
2026	-	301.71	293.51	284.58	309.09
2027	-	249.33	257.77	289.92	307.92
2028	-	249.82	281.32	290.58	297.81
2029	-	281.65	292.37	288.01	309.99
2030	-	274.66	292.03	288.51	294.77

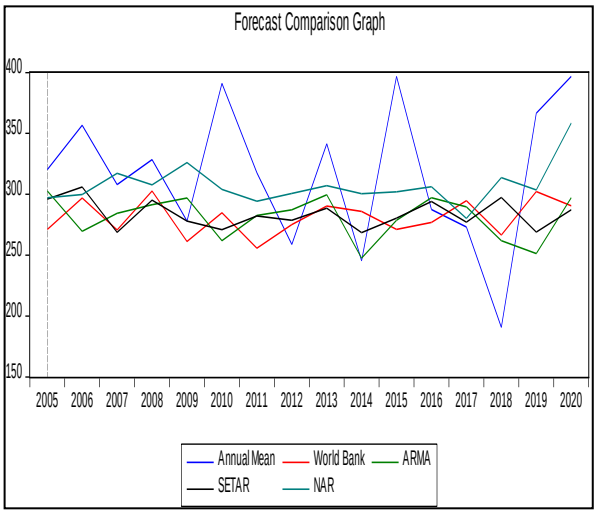

Fig. 3 Forecasts comparison graph.

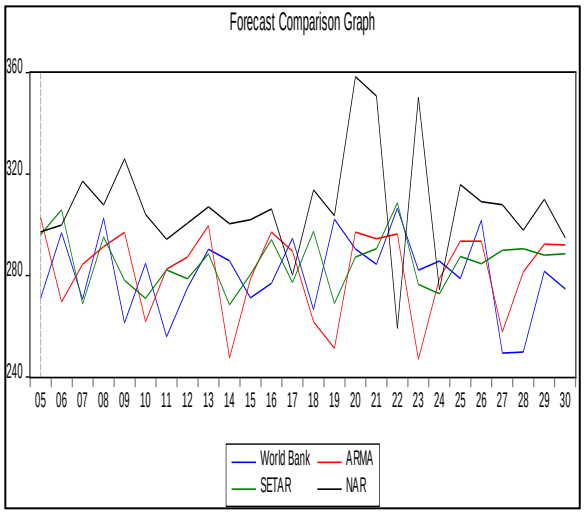

Fig. 4 Forecast comparison graph from 2021-2030.

Additionally, the forecasts of ten years are also estimated. It is not suitable to estimate long-term forecasts for climate-related data, therefore, only ten years i.e. 2021 to 2030 forecasts are estimated in this study (Table 5). Lastly, to endorse the forecast performance of all the time series models, the forecasts from 2021 to 2030 are also compared with World Bank projections (Table 4). Surprisingly, the nonlinear threshold SETAR model is found to be the most accurate forecast model as compared to ARMA and the NAR model. It is suggested to use more machine learning models to forecast the rainfall data, especially in the case of Pakistan.

Conclusion

Rainfall is the sole necessity for the maximum agricultural economy of Pakistan. It is a primary concern for policymakers of agricultural and water demand management. In this study, three forecast techniques, namely traditional ARMA, SETAR, and NAR models are compared for annual rainfall for Pakistan. Results revealed that the SETAR model is superior in annual rainfall (mm) forecasts as compared to the ARMA model. It is concluded that a nonlinear threshold model can be used, instead of a linear model. Furthermore, the forecast performance of the NAR model outperforms both ARMA and SETAR models based on MSE, RMSE, and MAE forecast evaluation criteria. Finally, the forecasts are compared with World Bank projections. The forecast performance of the NAR model outperforms other models. Lastly, the out-of-sample forecasts from 2021 to 2030 are also estimated. These forecasts would be beneficial for the agricultural policymakers and water-demanding management. It is recommended to use the NAR model to reduce the percentage errors for annual rainfall.

References

Alam, F., Salam, M., Khalil, N. A., Khan, M. (2021). Rainfall trend analysis and weather forecast accuracy in selected parts of Khyber Pakhtunkhwa, Pakistan. *SN Applied Sciences*, **3** (5), 1-14.

Boochabun, K., Tych, W., Chappell, N. A., Carling, P., Lorsirirat, K., Pa-Obsaeng, S. (2004). Statistical modelling of rainfall and river flow in Thailand. *Journal-Geological Society of India*, **64** (4), 503-516

Damania, R., Desbureaux, S., & Zaveri, E. (2020). Does rainfall matter for economic growth? Evidence from global sub-national data (1990–2014). *Journal of Environmental Economics and Management*, **102**, 102335.

Hung, N. Q., Babel, M. S., Weesakul, S.,Tripathi, N. (2009). An artificial neural network model for rainfall forecasting in Bangkok, Thailand. *Hydrology and Earth System Sciences*, **13** (8), 1413-1425.

Khalili, N., Khodashenas, S. R., Davary, K., Baygi, M. M., Karimaldini, F. (2016). Prediction of rainfall using artificial neural networks for synoptic station of Mashhad: a case study. *Arabian Journal of Geosciences*, **9** (13), 1-9.

Momani, P. E. N. M., & Naill, P. E. (2009). Time series analysis model for rainfall data in Jordan: Case study for using time series analysis. *American Journal of Environmental Sciences*, **5**(5), 599.

Nwokike, C. C., Offorha, B. C., Obubu, M., Ugoala, C. B., Ukomah, H. I. (2020). Comparing SANN and SARIMA for forecasting frequency of monthly rainfall in Umuahia. *Scientific African*, **10**, e00621.

Olatayo, T. O., Taiwo, A. I. (2014). Statistical modelling and prediction of rainfall time series data. *Global Journal of Computer Science and Technology*, **14** *(*1), 1-10.

Ramirez, M. C. V., de Campos Velho, H. F., Ferreira, N. J. (2005). Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. *Journal of hydrology*, **301** (1-4), 146-162.

Somvanshi, V. K., Pandey, O. P., Agrawal, P. K., Kalanker, N. V., Prakash, M. R., Chand, R. (2006). Modeling and prediction of rainfall using artificial neural network and ARIMA techniques. *J. Ind. Geophys. Union*, **10** (2), 141-151.

Wang, W. C., Xu, D. M., Chau, K. W., Chen, S. (2013). Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. *Journal of Hydro informatics*, **15** (4), 1377-1390.