Annals of Agri-Bio Research **29** (1): 89-92, 2024 **Impact Score : 0.28 (Scopus)**

Time Series Modelling to Predict Rainfall

RAHUL GROVER, SIDDHARTHA SHARMA, BALJIT SINGH WALIA AND CHADETRIK ROUT*

Department of Civil Engineering, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala-133 207 (Haryana), India.
**(e-mail: chadetrikrout@gmail.com; Mobile : 89016 28070)*

(Received: August 1, 2023; Accepted: September 22, 2023)

ABSTRACT

The present study was aimed at developing a rainfall forecasting methodology using historical records of rainfall data. Time series modelling was carried out using IMD rainfall data over the Udaipur region of Rajasthan, India, from 1988-2020 to make a forecast for the year 2021. Auto-Regression Integrated Moving Average (ARIMA) model was used to make a forecast. Further, the forecast rainfall data were used for spatial distribution and spatial estimation of rainfall using the interpolation method of Kriging. The error in estimation was quantified with the R^2 value. The methodology used in this study was found adequate and quite efficient to be relied upon for making rainfall forecast in the future.

Key words: ACF, ARIMA, Kriging interpolation, spatial interpolation, time series modelling

INTRODUCTION

All types of water that fall on the ground from the atmosphere are referred to as precipitation. Rainfall, snowfall, hail, frost and dew are the typical manifestations (Ahmad *et al.*, 2017). The first two of these contribute a considerable amount of water. Unless otherwise specified, the term "rainfall" refers to the most common type of precipitation that causes stream flow, particularly the flood flow in the majority of rivers in India (Das and Scaringi, 2021). Rainfall's intensity changes throughout time and space (White *et al.*, 2017). It is unnecessary to elaborate on differences in rainfall amounts between regions of the country at a given time or changes in rainfall at a location throughout the year. Many hydrological issues, including floods and droughts, are caused by this volatility. According to Li *et al.* (2022), a variety of climatic parameters, including temperature, radiation, relative humidity, wind speed, etc., affect a location's net rainfall as well as how much falls.

According to Mohapatra *et al.* (2021), the north eastern part of India experienced rainfall of 150 mm/month in the pre-monsoon months (March, April and May). Pre-monsoon rainfall was comparatively less in the west and south of India. The Himalayas and north India are affected by western disturbances that bring heavy rainfall to the hilly region in the winter months (December, January and February) (Midhuna *et al.*, 2020). Due to induced low-pressure systems brought on by the western disturbances, other regions of north and central India see a modest quantity of rainfall at the same time (Ray *et al.*, 2019). The entire nation receives roughly 75% (1070.4 mm) of the total annual rainfall (1426.2 mm) during the south-west monsoon (June, July, August and September), with significant regional and temporal variability (Mohapatra *et al.*, 2021). Since the predominant westerly winds are parallel to the Western Ghats, the west coast experiences heavy rainfall mostly as a result of orographic lifting (Sumesh *et al.*, 2022). Similar conditions are seen in the Eastern Ghats during the north-east monsoon (October, November and December), along with more precipitation (Mohapatra *et al.*, 2021).

Rainfall is a natural phenomenon with wide-ranging and important effects on the global economy, environment, industry and society. Correct rainfall estimation is helpful in resolving concerns with the environment and water resources (Voget *et al.*, 2020). So, especially during dry climatic periods, rainfall forecasting is crucial for the planning and management of water resource systems. However, rainfall assessment and forecasting are not always easy. Realizing the above facts, the current study was aimed at developing a mechanism for forecasting rainfall based on historical rainfall data.

MATERIALS AND METHODS

In this study, a time series model was developed for forecasting, and a geo-statistical technique was used for interpolating the spatial concentration of rainfall in the Udaipur region of Rajasthan, situated between 23°46′ to 25°5′ on North latitude and 73°9′ to 74°35′ on East Longitude. ARIMA process was used in the cases where data showed evidence of distortion and De-trend. ARIMA approach consisted of three steps (Rout *et al.*, 2022):

(i) Identification: Investigated the characteristics and statistics of a time series and attempts to relate them to those of specific models.

(ii) Estimation: Using the available data, the parameters of the tentatively identified model(s) were estimated.

(iii): Diagnostic checking: The estimated model(s) and residuals of the fitted model(s) were investigated to determine whether the model(s) made sense and were consistent with the assumptions.

The auto regressive component describes the dependency among successive observation sets. The value of "a" is the number of auto regressive components in ARIMA (a, b, c) model. The value of "a" is 0 if there was no relation between two consecutive observation sets. When the value of a=1, there was a relationship between observations at lag 1 and the correlation coefficient φ_t at t=1. When the value of a=2, there was a relation between two observation sets at lag 2 and φ_2. Thus, "a" was the number of correlations needed to model the relationship.

For a = 2 ARIMA (2,0,0) model will be:

$$Yt = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t$$

Where,

Yt = Observation at time t

φ_t = Magnitude of the relationship between successive observations and

ε_t = Random shock at time t.

For a non-stationary series, the mean or the variance or both were not constant. The non-stationary trend was removed by differencing the time series. Determining the order of differencing was an important step needed to stationeries the series before fitting an ARIMA model (Dimri *et al.*, 2020). The order of differencing was unity, i.e. b=1, the time series was differenced once, and when b=2, it was differenced twice. The first difference eliminated linear trend, the second eliminated quadratic trend, and so on. If a series exhibited a long-term trend or lacked a tendency to return to its mean value i.e. if its auto correlations were strong even with 10 or more lags, a higher order of differencing was required (Dimri *et al.*, 2020).

In ARIMA (a, b, c), the value "c" showed the number of moving average components. When "c" was zero there were no moving average components. When c = 1, there was a relationship between the current observation and the random shock at lag 1. Similarly, for c = 2, there was a relationship between current observation and the random shock at lag 2.

Fig. 1 depicts the converted rainfall data as a time series. The sample ACF helped in identifying the relationship between data points of the study area's time series dataset (Fig. 2). The sample ACF plot was used to comprehend the autocorrelation structure of the time series after it had been transformed with the help of differencing (Fig. 3).

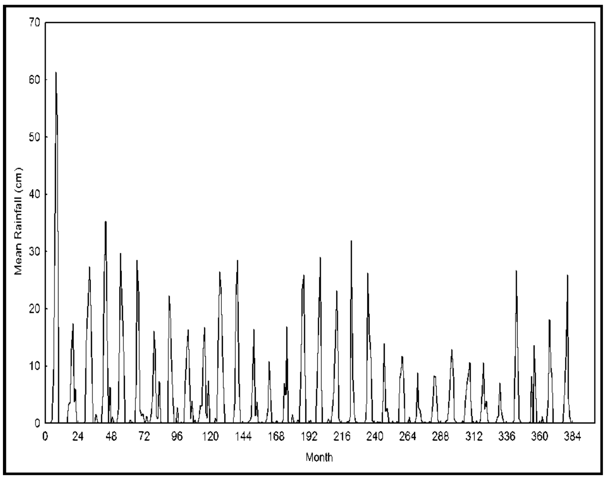

Fig. 1. Time series plot for mean monthly rainfall in cm.

In this study, the Kriging method was used for interpolation. Kriging is a method based on spatial autocorrelation (Ghalhari and Roudbari, 2018; Hamzehpour *et al.*, 2020). It is an improvement over other methods used for spatial interpolation because forecasting

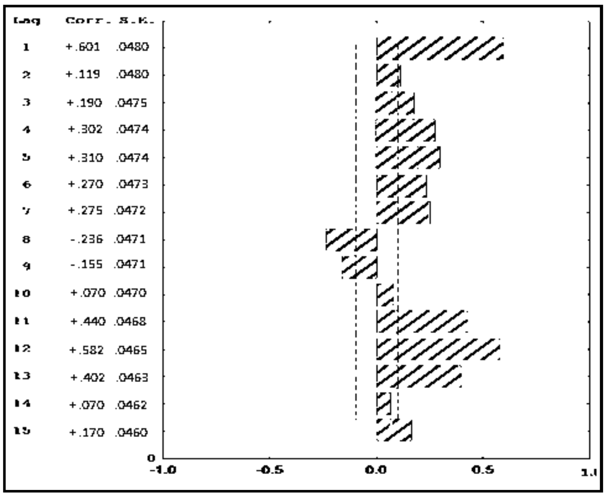

Fig. 2. Sample ACF of time series for Udaipur.

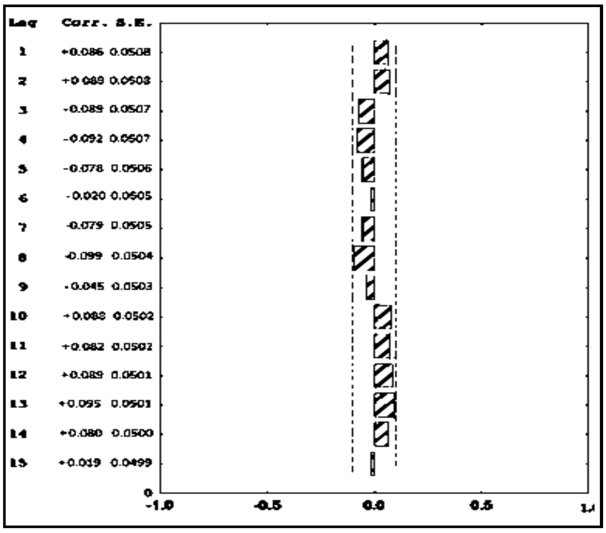

Fig. 3. Sample ACF plot of transformed time series.

estimates tend to be less biased as they are accompanied by forecasting standard errors which provide quantification of the uncertainty in the predicted value. Such uncertain information is an important tool in the Gaussian process. There are several theoretical semi-variogram models to choose from, including circular, spherical, exponential, linear and Gaussian models. The RMS error between the semi-variance values obtained from experimentally observed semi-variance and the theoretical model workout semi-variance values can be used to identify the best semi-variogram model (Rout *et al.*, 2022).

Building analysis data set of the total 31 selected co-ordinate points of Udaipur (Rajasthan area), 10 were retained for data validation, and the remaining 21 were utilized as input data for performing spatial interpolation using the Kriging technique.

Conducting semi-variogram analysis: This was accomplished by generating semi-variograms for Gaussian, spherical and exponential models, and then selecting the model with the lowest root mean square error as the best suited.

Kriging for spatial prediction: The variogram's estimated parameters were supplied into the Gaussian process as input.

Cross-validation: The interpolated values were cross-validated, and the relational coefficient (R^2 value) between the interpolated value and the observed data was calculated.

RESULTS AND DISCUSSION

Differentiating and logarithmic transformations were used to erase the observed trend in the time series. The differencing sequence, auto-regressive component and moving average component were determined and found as a = 0, b = 1 and c = 1. For Udaipur, determination of these ARIMA parameters, ARIMA (0, 1, 1) model was determined to be correct for the Rajasthan region data. So, ARIMA (0, 1, 1) model was used to anticipate rainfall amounts at all 31 locations. The coefficient of determination (R^2=0.782) was estimated by comparing the forecast values of rainfall with the IMD observed data (indicating a good fit in Fig. 4). Forecasts were made for all 31 sites for each month from January-December, 2021 (Table 1) and the spatial distribution of rainfall data of the study area has been shown in Fig. 5.

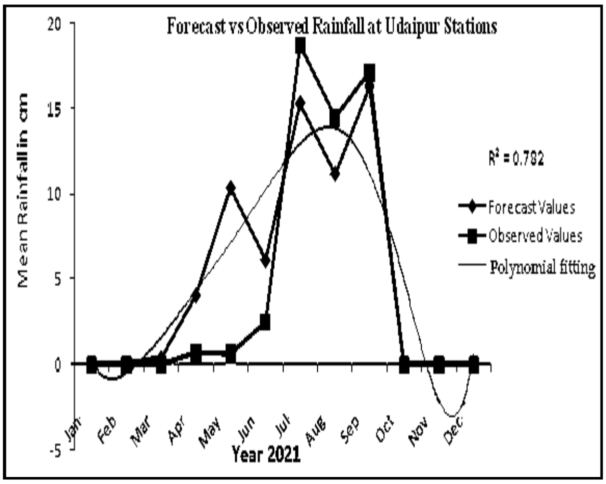

Fig. 4. The IMD observed rainfall vs. forecast rainfall of time series analysis.

On the validation sites, a comparison of the IMD observed rainfall and the Kriging interpolated rainfall was performed for each month from January-December 2021. In this

Table 1. Forecast values of rainfall (cm) for 2021 using time series modelling

Time	Forecast values	Time	Forecast values
Jan-21	0.03	Jul-21	15.3
Feb-21	0.151	Aug-21	11.21
Mar-21	0.399	Sep-21	16.3
Apr-21	4.06	Oct-21	0.037
May-21	10.358	Nov-21	0.069
Jun-21	6.127	Dec-21	0.072

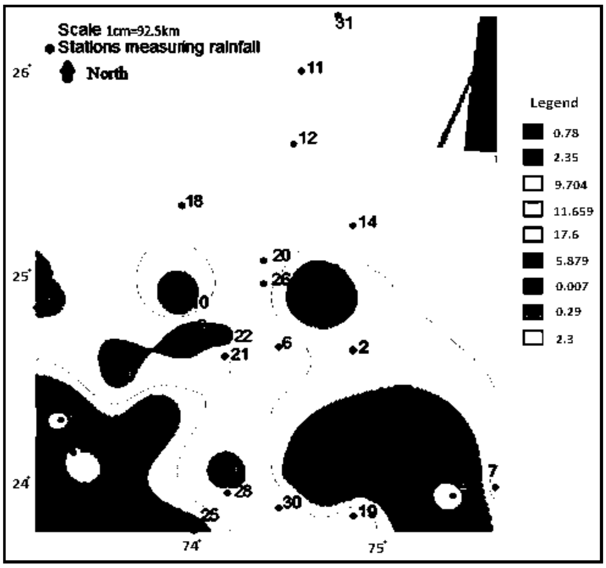

Fig. 5. Spatial distribution of rainfall (cm).

section, the coefficient of determination R^2 was utilized to demonstrate the quality of fit between the interpolated rainfall values using Kriging and the IMD observed data. The R^2 values lied between (0.62 to 0.724) indicated that estimated values of rainfall at the validation locations used ordinary Kriging fit adequately.

CONCLUSION

The spatial interpolation and time series modelling estimation techniques were simple to use, and the integration of their results was a good strategy. The IMD observed rainfall data and time series forecasts of the data were compared. The R^2 values were observed to fall between 0.610-0.724. R^2 values for spatial interpolation estimations were found to range between 0.611-0.750. The R^2 values achieved in this investigation were sufficient for rainfall forecasting. The methods used in this study should be able to help with future forecasts and be relied upon to estimate rainfall at other sites without ground monitoring.

REFERENCES

Ahmad, L., Kanth, R. H., Parvaze, S. and Mahdi, S. S. (2017). Measurement of precipitation. In: *Experimental Agrometeorology: A Practical Manual.* pp. 55-66. Springer, Cham. *https://doi.org/10.1007/978-3-319-69185-5_9.*

Das, S. and Scaringi, G. (2021). River flooding in a changing climate: Rainfall-discharge trends, controlling factors and susceptibility mapping for the Mahi catchment, Western India. *Nat. Hazards* **109**: 2439-2459.

Dimri, T., Ahmad, S. and Sharif, M. (2020). Time series analysis of climate variables using seasonal ARIMA approach. *J. Earth Syst. Sci.* **129**: 01-16.

Ghalhari, G. F. and Roudbari, A. D. (2018). An investigation on thermal patterns in Iran based on spatial auto correlation. *Theor. Appl. Climatol.* **131**: 865-876.

Hamzehpour, N., Rahmati, M. and Roohzad, B. (2020). Spatial prediction of soil electrical conductivity using soil axillary data, soft data derived from general linear model and error measurement. *Desert* **25**: 87-99.

Li, X., Zeng, Q., Lei, G. and Sun, G. (2022). Effects of meteorological factors on waterbird functional diversity and community composition in Liaohe Estuary, China. *Int. J. Environ. Res. Public Health* **19**: 5392.

Midhuna, T. M., Kumar, P. and Dimri, A. P. (2020). A new western disturbance index for the Indian winter monsoon. *J. Earth Syst. Sci.* **129**: 01-14.

Mohapatra, G., Rakesh, V., Purwar, S. and Dimri, A. P. (2021). Spatio-temporal rainfall variability over different meteorological subdivisions in India: Analysis using different machine learning techniques. *Theor. Appl. Climatol.* **145**: 673-686.

Ray, K., Pandey, P., Pandey, C., Dimri, A. P. and Kishore, K. (2019). On the recent floods in India. *Curr. Sci.* **117**: 204-218.

Rout, C., Shukla, G., Beniwal, V., Singh, S. P. and Grover, R. (2022). GIS and time series modelling approach to predict tropospheric nitrogen dioxide. *Ann. Agri. Bio. Res.* **27**: 35-41.

Sumesh, R. K., Resmi, E. A., Unnikrishnan, C. K., Jash, D. and Padmalal, D. (2022). The extreme precipitation events of August 2018 and 2019 over southern Western Ghats, India: A microphysical analysis using *in situ* measurements. *Atmos. Res.* **277**: 106322.

Vogel, P., Knippertz, P., Fink, A. H., Schlueter, A. and Gneiting, T. (2020). Skill of global raw and post-processed ensemble predictions of rainfall in the tropics. *Weather Forecast* **35**: 2367-2385.

White, R. H., Battisti, D. S. and Skok, G. (2017). Tracking precipitation events in time and space in gridded observational data. *Geophys. Res. Lett.* **44**: 8637-8646.