Geosfera Indonesia
p-ISSN 2598-9723, e-ISSN 2614-8528
available online at : https://jurnal.unej.ac.id/index.php/GEOSI
Vol. 7 No. 3, December 2022, 236-263

https://doi.org/10.19184/geosi.v7i3.34487

Research Article

Simulation of Rainfall Using Two Statistical Data Driven Models: A Study on Santhal Pargana Division of Jharkhand State, India

Shrinwantu Raha*, Shasanka Kumar Gayen
Department of Geography, Cooch Behar Panchanan Barma University, Cooch Behar, West Bengal, 736101, India
*Corresponding Author, Email address: shrinwanturaha1@gmail.com

ARTICLE INFO

Article history

Received :
14 October 2022

Revised :
17 November 2022

Accepted :
29 November 2022

Published :
24 December 2022

ABSTRACT

Although the variability and prediction of rainfall is an essential issue of the Santhal Pargana Division of the Jharkhand State but the issue is still far from its' conclusive statement till date. Therefore, this study aimed to simulate the monthly rainfall from 1901 to 2020 using an eight-step procedure. After downloading the monthly rainfall for the Santhal Pargana Division from 1901 to 2020, the TBATS and Naive models were used to simulate the rainfall. The accuracy assessment of each model was done by using the MASE, MAE, RMSE, ME, and R. For the Naïve model, the Godda station was noticed with a comparatively high combined error. The lowest combined error was found for the Pakur station in case of Naïve models. Similar result was also obtained for the TBATS model. The TBATS was found with comparatively higher accuracy, as the combined error was less for the TBATS. The spatial assessment for the standardized rainfall varied from 84.419 mm. to 149.225 mm. For the Naïve predicted model, the rainfall was marked in between 8.133 mm. to 67.059 mm. For the TBATS fitted model, the rainfall fluctuated from the 37.127 mm. to 62.993 mm. Dumka station was noticed with comparatively low rainfall (i.e.,37.127 mm.). Deoghar and Jamtara stations were marked with a moderate rainfall. Remaining stations were marked with higher amount of rainfall for the TBATS fitted model. The Wilcoxon test proved that each model was significant at 95% confidence interval. The result produced in this research is fruitful enough to be utilized for agricultural planning in the Santhal Pargana Division of the Jharkhand state, India.

Keywords : TBATS model; Naive model; simulation; accuracy

1. Introduction

Rainfall is one of the most crucial meteorological factors that affects our life in a significant way (Barrera-Animas et al., 2022; Shmuel et al., 2022). Evaluation of changes in the rainfall within the context of climate change have a substantial impact on a region's socioeconomic situation (Gobiet et al., 2014; Mukherjee et al., 2018). In a country like India, simulation of rainfall is required for the improvement of the agricultural planning and management (Bhatt & Mall, 2015). Variations in rainfall have substantial impact on the nation's agricultural, drinking water supply, energy industry, and population's way of life (Cosgrove & Loucks, 2015; Mora et al., 2018). Rainfall at the monsoon season is essential for both agriculture during the dry months and irrigation during the wet months (Biemans et al., 2019; Ferrant et al., 2017). The "Indian summer monsoon," which is an essential part of the Asian monsoon system, is renowned for having "a wide range of variability on daily and decadal, time

periods." (Menon et al., 2013; Singh et al., 2019). The monsoonal system is an essential part of the hydro-geological cycle of the planet and plays a crucial role in the functioning of the global circulation and ecological systems. It may be one of the primary and essential factors affecting South and East Asia's climate (Liu & Liu, 2019; Liu et al., 2019). According to Azad et al. (2022); Dubache et al. (2019); Mie Sein et al. (2015) crop productivity, hydropower output, foliage, water management, and biodiversity of these sections are impacted by the intensity and length of the monsoonal rains. Additionally, "India's inter-annual variability of the monsoonal rainfall causes widespread droughts and floods, having a significant impact on the nation's food grain and economy" (Mohapatra et al., 2017). Therefore, for the integrated management of agriculture in India, analysis of rainfall is urgently needed. Now a day, hydrologists and meteorologists are more concerned with the forecasting and simulation of rainfall. Forecasting rainfall in the context of time series modelling is challenging since the amount of rainfall fluctuation has a substantial influence on the nation's economy (Rahman & Lateh, 2017; Saha et al., 2020).

There are various simulation models are already developed in the field of civil engineering. The ARIMA model, is indeed the most commonly used model (Raha & Gayen, 2020). Due to its' univariate design and linearity assumption, the model is unable to consider the geographical interrelationships exists in the time series model. As a result, the ARIMA model is inappropriate when there are spatial connections between distinct time delays and the time measurement data isn't entirely linear (Saha et al., 2020). Different statistical models, including the bilinear model (Granger & Andersen, 1978), ARCH (Autoregressive Conditional Heteroskedasticity) model (Engle, 1982), GARCH model (Bollerslev, 1986), and the TAR model were developed to deal with nonlinear patterns (Tong & Lim, 1980). However, when working with data from the actual world, it is not always possible to identify certain specific assumptions that these models consider. In a real-world situation, inexpensive and effective reproducible models are required. As a result, some modified version of ARIMA model has been developed, which includes the Naïve models and Trigonometric Exponential Smoothing State Space model with Box-Cox transformation, ARMA errors, trend and seasonal component (TBATS) model (De Livera et al., 2011; Livera, 2010). The ARIMA (Autoregressive Integrated Moving Average), ARCH, GARCH and TAR models face the problems of overfitting. But the TBATS and Naïve models have an inbuilt capacity to avoid it. Additionally, the ARIMA, ARCH, GARCH and TAR models are suitable for only one seasonal cycle. But the TBATS and Naïve models are applicable for the complex seasonal pattern, which includes more than two seasonal cycles.

Some machine learning and data driven models are also popped up to deal with complex seasonality and non-linearity of the time series data. Some of the popular techniques in the Machine Learning Family are ANN, SVM and ANFIS techniques. With a mean R^2 value of 0.95, the SVM model outperforms the others when Samantaray et al., (2020) used ANN, SVM, and ANFIS to simulate rainfall in Bolangir District, Odisha. Liu et al., (2019) found the ANN with higher accuracy in case of rainfall forecasting than the Numerical Weather Prediction Models (NWPM). In addition to these, many academicians and researchers throughout the world apply machine learning and deep learning models to forecast the precipitation. For example, Chaturvedi (2015) used multilayer back propagation neural network model to simulate the rainfall data of Chennai, India. Here, the dataset of 1978 to 2009 was used as the training set. Using some model evaluation parameters the model was found with sufficient accuracy. Azad et al. (2015) simulated the rainfall data of India by using ANN, PSD and MRA decomposition technique. Here, the ANN was found with comparatively higher accuracy.

A comparison was done between the Support Vector Machine (SVM), Adaptive Neural Fuzzy Inference System (ANFIS) and Artificial Neural Network (ANN) to simulate the rainfall data of Yaoziang Province, China by Zhang et al. (2016). Here, the ANN was found with comparatively higher accuracy. Similarly, Bagirov et al. (2017); Balan et al. (2019); Chao et al. (2018); Mohd et al. (2020); Nourani et al. (2019); Scheuerer (2014); Vuyyuru et al. (2021); Wang et al. (2022); Xiang et al. (2018) used ANN, ANFIS, SVM to simulate the rainfall in different locations and in most of the cases, the ANN was found with comparatively low RMSE, MSE, MAPE value. However, a number of researchers

across the world have challenged the "Black-box character" of deep learning-based time series forecasting models. For instance, Dang (2020) created an original IDS architecture after seeing the Black-box characteristics of machine learning models. Similarly, Adadi & Berrada, (2018); Casado-Vara et al., (2021); Doorn, (2021); Mehdiyev & Fettke, (2021); Xu et al., (2022) argue that the nature machine learning models are really unpredictable and therefore, are difficult to interpret. Instead of machine learning, the statistical data driven models such as TBATS and Naïve have enormous flexibility. Although, there are enormous flexibility in the two above mentioned models, the researchers have not widely used them to simulate the long-term time series of hydro-meteorological variables. Therefore, simulation of rainfall using TBATS (Trigonometric Exponential Smoothing State Space model with Box-Cox transformation, ARMA errors, trend and seasonal component) and Naïve models are noble attempts.

The Asian monsoon is the dominant factor which controls the life and livelihoods of the Santhal Pargana Division. These section experiences hard terrain, moderate slope and moderate rainfall. Geologically, these portions are dominated by the Granite and Gneiss, which is unable to store water. But still, agriculture is the principle livelihood in these portions. Therefore, here, agricultural activity entirely depends on the average rainfall. Hence, for the stability of the livelihood, the predictive spatio-temporal simulation of rainfall is urgently needed. The upcoming pattern of rainfall in coming days are portrayed by the simulation, which helps to remove the uncertainty of the rainfall pattern. But till date, no prominent research works are found on the predictive pattern of rainfall in these sections. Therefore, the focus of the research are (1) to simulate the rainfall of Santhal Pargana Division using two statistical data driven models, and (2) to evaluate each model geo-spatially in order to comprehend the geographical distribution of the rainfall.

2. Study Area

One of the key divisions of the Jharkhand state is Santhal Pargana Division. It is comprised with six districts such as, Godda, Deoghar, Dumka, Jamtara, Pakur and Sahibganj (Figure 1). Godda is famous for the Rajmahal coalfield in Lalmatia. Agriculture is the chief source of income, and the three main crops are the paddy, wheat, and maize. Godda was listed by the Indian government as one of the 250 most underdeveloped districts in the nation on 2006 (out of a total of 640). According to the Census of India (2011), the Godda was found with 1,311,382 population. Almost, 12392.03 Acre irrigated area is present in the Godda district and approximately 82.16% of the district's population can read and write (literate) (Sharma & Singh, 2021). The Deoghar is situated in the western portion of the Santhal Parganas with a population of 1,491,879 people in accordance with the Census of India (Government of India, 2011). The area is made up of a number of groups of rocky hills covered in forest and connected by a network of long ridges and depressions. Highland crops are grown on the gently rolling highlands. The district is 247 metres above mean sea level on average. Hill ranges including Phuljari (750 m), Teror (670 m), and the Degaria are present (575 m) here. The district's overall slope declines towards the southeast. The Chhota-Nagpur granite gneiss of the Archean age dominates the district's geology, with areas of alluvium, sandstone, and shale from the Gondwana formations.

The Ajay and the Paltro are two significant rivers that pass through the area. These rivers create a network of numerous tributaries that make up the dissected plateau with a Dome-filled landscape. Dumka is famous for several pilgrimages and it had a total population of 1,321,442, according to the Census of India (2011) (Government of India, 2011). Dumka had an educational attainment of 62.54% and a ratio of men to women equals to 974 females with every 1000 males (Government of India, 2011). Granite, Gneiss, Gondwana and Rajmahal traps are the major constituents in the formation of geology of the Dumka District. An undulating highland region may be found in Santhal Parganas' southeast in

the Jamtara district. The Barakar, which comes from the Deoghar province in the western countryside, drains the region, briefly serves as a boundary between Jharkhand and West Bengal, and then departs into W.B. to the east, separates it from the Chotta Nagpur plateau and the Ajay. With approximately total population of 7,91,042 people, Jamtara has an area of 1811 sq km. In The sections of the lower heights of the Chottanagpur plateau; this district is located. More than 64% workers are engaged in the primary sector. The district's gender ratio was 954 people for every 1,000 men (Government of India, 2011). The district's overall literacy rate was 64.59%, and there were 437 people per square kilometre, according to the most recent Census. When Pakur first appeared, it was a collection of ponds and orchards surrounded by dense forest and stony terrain beneath the Rajmahal Hills. According to the latest Census, the Pakur had a total population of 900422 persons. About 63% people are engaged into the primary sector. The average population density of the district was 498. The literacy rate of the district was 48.82%. Total irrigated area of the district was 6,81,90 Square km. The rice, rapeseed, wheat and maize are the major crops in these sections of the study area. Sahibganj district is divided into two subdivisions: the first is Sahibganj subdivision and the another is Rajmahal subdivision. Sahibganj district has 1,150,567 people living there as of the 2011 census. The most dominant rock in the area was the Rajmahal Trap. Alluvium and Laterite are the district's other geological formations. The district's northern and eastern boundaries contain alluvium, which is primarily made of clay and sand.

Overall, the Santhal Pargana Division of the Jharkhand state has a tropical wet and dry climate. Mainly, summer, rainy, autumn, winter and spring seasons are noticed in these sections of the study area. The summer continues from the April to June. Average diurnal high temperature of these section is 37ºC and average daily low temperature is 25ºC. The southwest monsoon, which occurs in the state from June to October, is responsible for the annual rainfall, ranges of around 1,000 mm to 1500mm (40 in) in the western sections. The months of July and August see over half of the yearly precipitation. November through February are considered the winter months. Therefore, from the above analysis the following details are prominent about the Santhal Pargana Division:

1. These sections receive moderate rainfall annually.
2. Agriculture is the principle livelihood in this study area.
3. These sections are dominated by the Granite and Gneiss; therefore, these portions are unable to store the ground-water.
4. Seasonal variation of rainfall is prominent in these portions.

As the agriculture, is totally dependent on the predictive pattern of rainfall therefore, an honest effort is needed to understand the micro level pattern of rainfall in these sections. But till date, such research work is missing. Therefore, the focus of this research is to simulate the rainfall using the TBATS and Naïve models.

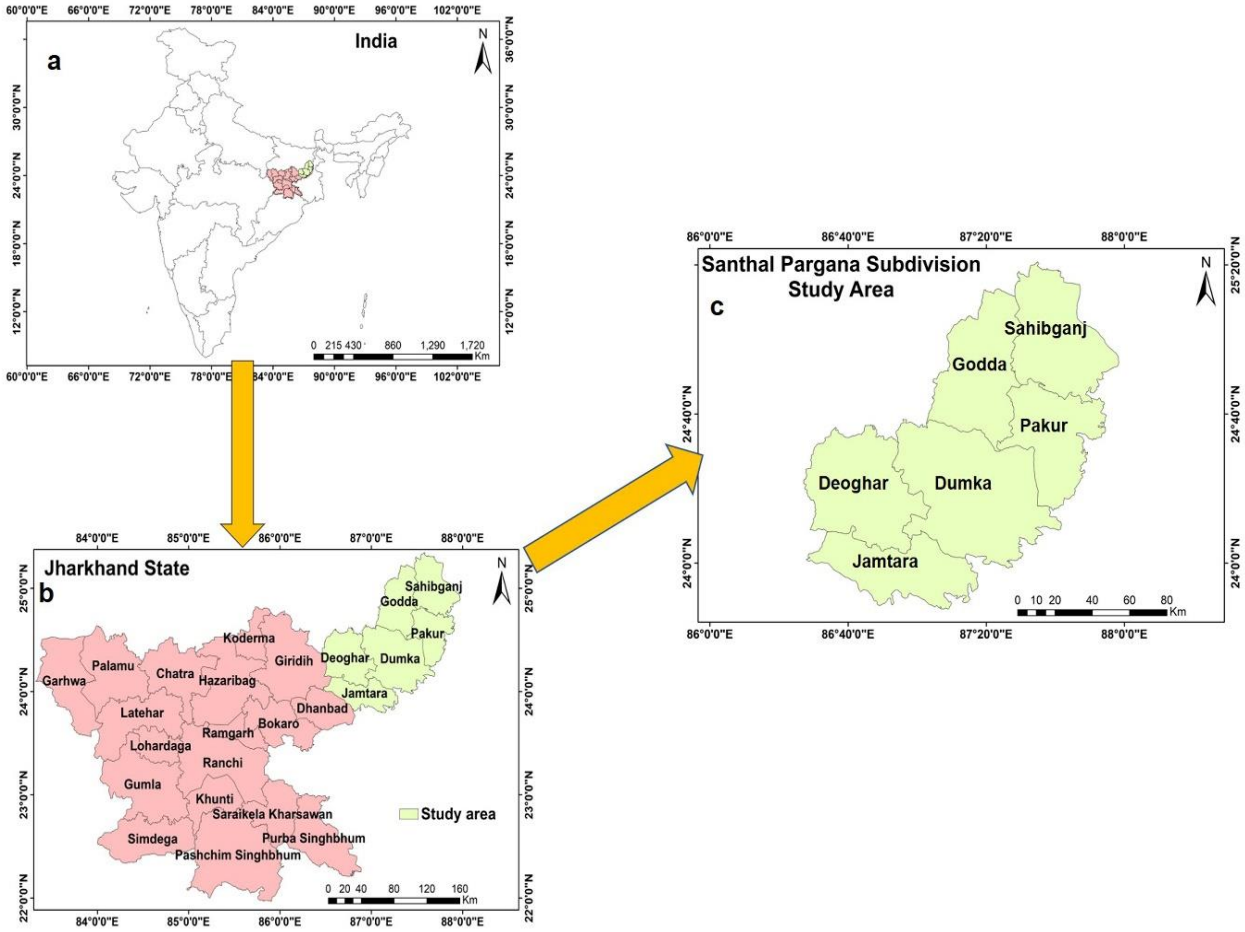

Figure 1. Study area

3. Methods

The methodological framework was portrayed in the Figure 2. This research used a 8-step methodology, which was displayed as follows:

3.1 First Step-Collection of the Rainfall Data

The rainfall data was collected from the India WRIS portal at the first step (India-WRIS, 2021). The most crucial elements of managing water resources are represented by water information in the public domain, an initiative of India's WRIS Project with the goal of disseminating data in the public domain. Planning and the strategy for managing water resources both require knowledge on the state of the water resources. Given the difficulties facing the water resources sector, the Government of India decided to create a centralised platform that would serve as a national repository for water resources and related data, with administrative granularity extending to smaller state-level units of government as well as hydrological levels like basins and sub-basins (Das, 2019; Shah et al., 2021). This entails gathering data from all different sources, standardizing it, and storing the complete gamete of data on a national scale. Consequently, delivering an operational system with greater visibility and a global standard platform on the public domain for all types of users (Prakash & Mishra, 2022). This portal has a feature called Water Data Online that enables users to view rainfall distribution for various states and basins throughout India using a variety of frequency, and aggregation types. To see trends of rainfall

data (average, total, min, max aggregation type) for a specific region of focus, the user can choose a time range (start and stop). Daily, monthly, or annual time steps are all acceptable. The portal displays the total and mean normal rainfall, which are based on the daily data from the Indian Meteorological Department (IMD) during 1901 to 2020. This study uses monthly rainfall information from 1901 to 2020.

3.2 Second Step-Checking of Partial Auto Correlation Function (PACF)

The Partial Autocorrelation was checked for the downloaded rainfall dataset at the second stage. If the decreasing autocorrelation value with increasing time steps were found, then the calculation moved into the next step, otherwise it was stopped. The formula of PACF was as follows (Silva, 2021):

$$PACF = \frac{covariance(\left[M_i \middle| M_{(i-1)}, M_{(i-2)} \dots \dots \dots M_{(i-k+1)}\right], \left[M_{(i-k)} \middle| M_{(i-1)}, M_{(i-2)} \dots \dots \dots M_{(i-k+1)}\right])}{\sigma\left[M_i \middle| M_{(i-1)}, M_{(i-2)} \dots \dots \dots M_{(i-k+1)}\right] \times \sigma\left[M_{(i-k)} \middle| M_{(i-1)}, M_{(i-2)} \dots \dots \dots M_{(i-k+1)}\right]} \quad (1)$$

Where, *M* is the downloaded rainfall series; *k* is the time series lag.

3.3 Third Step-Standardize the Rainfall Dataset

In the third step, it was essential to standardize the rainfall dataset. Standardization is required to minimize the randomness of the rainfall dataset (Yaseen et al., 2021):

$$Rainfall_{norm}(Y_i) = \frac{Rainfall_{observed} - Rainfall_{min}}{Rainfall_{max} - Rainfall_{min}} \quad (2)$$

$Rainfall_{observed}$ represents the observed rainfall data (downloaded monthly data from India WRIS), $Rainfall_{min}$ is the minimum rainfall, $Rainfall_{max}$ is the maximum rainfall, and $Rainfall_{norm}$ is the value corresponding normalized numeric of rainfall.

3.4 Fourth Step-Checking of Outliers

The fourth stage was marked by the two-sided Grubb's test (Grubbs, 1969; Stefansky, 2012) to identify outliers in the univariate rainfall data set, which has a roughly normal distribution (Urbanczyk-Wochniak et al., 2003):

$$G = \frac{max|Y_i - \bar{Y}|}{s_y} \quad (3)$$

Where, G is the Grubb's test; Y is the observed rainfall dataset and $\bar{Y}$ is the mean of the standardized rainfall dataset. The data set for rainfall has a standard deviation of s_y. If outlier value is more than 1 and beyond the 0 to 1 range, it was removed from the dataset and the missing value was replaced by the preceding and forward values in the dataset.

3.5 Fifth Step-Partition of data

For a large dataset, data partition and consecutive validation are essential for the effective simulation (Li et al., 2017; Yaseen et al., 2021). In the fifth stage, the data was partitioned into two parts. 70% of the dataset was deemed the training set, while the remaining 30% was deemed the test set (Chang et al., 2021).

3.6 Sixth Step- Simulation using Naïve and TBATS models

In the sixth step, the rainfall data was simulated using the Naïve and TBATS models. In case of the Naïve forecasts, all simulated values were set to its' last observation. For a highly seasonal data, seasonal Naïve models perform incredibly well (Hyndman & Athanasopoulos, 2018). The seasonal Naïve model was expressed here as follows:

$$Y_{T+h|T} = Y_{T+h-m(k+1)} \quad (4)$$

Where, *m* is the period, which is seasonal and *k* is the part of *(h-1)/m* (i.e., the number of years in the forecasted period that have been completed before time *T+h*), *T* is the observed duration of time and *h* is the simulated time step.

TBATS is an extended form of the exponential smoothing, which includes the Box-Cox procedure to transform the dataset, ARMA model for residuals and the Trigonometric Seasonal method. Although its' computation cost is very high but it can deal with the multi-seasonal data with or without too many parameters. The model can work also under the high frequency data (Hyndman & Athanasopoulos, 2018). The model was illustrated as follows (Alizadeh et al., 2021; De Livera et al., 2011):

$$Y_t = p_{t-1} + a_{t-1} + c_t^{(1)} + c_t^{(2)} + f_t, \quad (5)$$

$$l_t = p_{t-1} + a_{t-1} + \alpha f_t, \quad (6)$$

$$a_t = a_{t-1} + \beta f_t, \quad (7)$$

$$c_t^{(1)} = c_{t-m_1}^{(1)} + \gamma_1 f_t, \quad (8)$$

$$c_t^{(1)} = c_{t-m_2}^{(1)} + \gamma_2 f_t \quad (9)$$

Where, seasonal cycles are the m_1 and m_2 . f_t is the white noise which denotes errors of the prediction process. The 'level' and 'trend' components are the p_t and a_t respectively; t is the time stamp. $c_t^{(i)}$ is the i-th component at the t time, and those i-th component is seasonal (Hyndman & Athanasopoulos, 2018). The coefficients $\alpha, \beta, \gamma_1, \gamma_2$ are the parameters of smoothing, and the initial state variables are $l_0, b_0, \left\{c_{1-m_1}^{(1)}, \dots\dots\dots, c_0^{(i)}\right\}, and\ \left\{c_{1-m_2}^{(2)}, \dots\dots\dots, c_0^{(2)}\right\}$. These are also known as the seeds.

By extending those basic models of exponential smoothing (eq. s 4,5, 6 and 7) De Livera et al. (2011) formulated the following descriptions for the TBATS:

$$Y_t^{\omega} = \begin{cases} \frac{Y_t^{\omega}-1}{\omega} \\ logY_t, \end{cases}, \omega \neq 0, \omega = 0, \quad (10)$$

$$Y_t^{\omega} = p_{t-1} + \emptyset a_{t-1} + \sum_{i=1}^{T} c_{t-m_i}^{i} + f_t, \quad (11)$$

$$l_t = p_{t-1} + \emptyset a_{t-1} + \alpha \times f_t, \quad (12)$$

$$a_t = (1-\emptyset)a + a_{t-1} + \beta \times f_t, \quad (13)$$

$$c_t^{(1)} = c_{t-m_i}^{(1)} + \gamma_i f_t, \quad (14)$$

$$c_t^{(1)} = c_{t-m_i}^{(1)} + \gamma_i f_t, \quad (15)$$

$$f_t = \sum_{i=1}^{p'} \vartheta_i f_{t-i} + \sum_{i=1}^{q'} \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (16)$$

Where, $m_1, \dots\dots\dots, m_T$ denote the seasonal periods. The local level is denoted as p_t in period t, a is the long-run trend in the entire period *t*, $c_t^{(i)}$ represents the i-th seasonal component in entire period t, the $ARMA\ (p',\ q')$ process, ε_t is the Gaussian process and it has mean value zero and a constant variance , which is represented by δ^2. The smoothing parameters are given by α, β, γ_i for $i = 1, \dots\dots\dots, T$. The trigonometric representation of above series is as follows (Abdelgawad et al., 2015):

$$c_t^{(1)} = \sum_{i=1}^{T} c_{j,t}^{(i)}, \quad (17)$$

$$c_t^{(1)} = c_{j,t-1}^{(i)} cos\lambda_j^{(i)} + c_{j,t-1}^{*(i)} sin\lambda_j^{(i)} + \gamma_1^{(i)} f_t, \quad (18)$$

$$c_{j,t-1}^{*(i)} = -c_{j,t-1}^{(i)} cos\lambda_j^{(i)} + c_{j,t-1}^{*(i)} cos\lambda_j^{(i)} + \gamma_2^{(i)} f_t, \quad (19)$$

Where, $\gamma_1^{(i)}$ and $\gamma_2^{(i)}$ are the parameters, which represent smoothing and $\lambda_j^{(i)} = 2\pi j/m_i$. The level (stochastic) of i-th seasonal component is represented by $c_{j,t}^{(i)}$, and the growth (slow and stochastic) of the seasonal component is needed to describe the change in the seasonal component over time t by $c_{j,t}^{*(i)}$. The symbol k_i stands for the frequency of a i-th seasonal component. For the trigonometric seasonal fluctuation; the following equation was used (De Livera et al., 2011; Livera, 2010):

$$Y_t^{\omega} = p_{t-1} + \emptyset a_{t-1} + \sum_{i=1}^{T} c_{t-1}^{(i)} + f_t \quad (20)$$

The class is designated by TBATS. The TBATS is represented as $(\omega, \emptyset, p, q, \{m_1, k_1\}, \{m_2, k_2\}, \ldots\ldots\ldots\ldots\ldots\ldots, \{m_T, k_T\})$.

3.7 Seventh Step- Spatial assessments

The seventh step was marked with the spatial assessments of the observed, fitted and predicted models. For this, at first the mean centres were estimated for each district (Figure 2 and Figure 8). Here, it was assumed that the total rainfall of a particular district was concentrated in the middle point of a particular district. To estimate the exact mean point within every district following procedures were implemented with three phases:

a) First, two tangents were drawn for the *X*-axis (*OX*) and *Y*-axis (*OY*) respectively. Now, AA', BB', CC', DD', EE', FF' were drawn respectively for the Jamtara, Deoghar, Dumka, Godda, Pakur and Sahibganj districts. AA', BB', CC', DD', EE', FF' were drawn, which were parallel to the *X*-axis.
b) Simultaneously, for the Jamtara, Deoghar, Dumka, Godda, Pakur and Sahibganj districts respectively, AA'', BB'', CC'', DD'', EE'', FF'' were also drawn, which were parallel to the *Y*-axis.
c) The meeting point of AA'' and AA' was the point *A*, the meeting point of BB'' and BB' was the point *B*, the meeting point of CC'' and CC' was the point *C*, the meeting point of DD'' and DD' was the point *D*, the intersection point of EE'' and EE' was the point *E* and the insecting point of FF'' and FF' was the point *F*. *A*, *B*, *C*, *D*, *E* and *F* are the mean centers of the Jamtara, Deoghar, Dumka, Godda, Pakur and Sahibganj districts, which were considered as the stations.
d) For the spatial evaluations in the ARCMAP 10.5 platform, the Inverse Distance Weightage (IDW) approach was used after computing mean centre.

3.8 Eighth Step- Accuracy assessments and significance test

3.8.1 Accuracy assessments

The seventh step was marked by the accuracy assessments of the different models. Here, Mean Absolute Scaled Error (MASE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Error (ME) and Correlation Coefficient (R) were used to judge the accuracy of every model. Descriptions of different errors were discussed as follows:

3.8.1.1 Mean Absolute Scaled Error (MASE)

For the time series (seasonal), the MASE was estimated by utilizing the following formula (Franses, 2016):

$$MASE = mean(\frac{|e_j|}{\frac{1}{T-m}\sum_{t=m+1}^{T}|Y_t - Y_{t-m}|}) \quad (21)$$

Where, $|e_j|$ is the prediction error for a specific time and j is quantity of forecasts. The (Y_j) is the standardized rainfall dataset (actual or the observed value), and F_j is the forecasted or predicted dataset; here $e_j = Y_j - F_j$. Here, seasonal period or seasonality is represented by m. T is the number of time or the total time frame considered in this research.

3.8.1.2 Mean Absolute Error (MAE)

MAE is popular measurement of accuracy; can be expressed as follows (Wang & Lu, 2018):

$$MAE = \frac{\sum_{i=1}^{T}|Y_i - F_i|}{T} = \frac{\sum_{i=1}^{T}|e_i|}{T} \quad (22)$$

Here, T is the total time frame; Y_i is actual or observed value; F_i is modelled or fitted value for the TBATS and Naïve model. e is the error term which represents $Y_i - F_i$. MAE takes the modulas of the numerator.

3.8.1.3 Root Mean Squared Error (RMSE)

RMSE is the standard deviation of the actual or observed value and the modelled or predicted dataset. RMSE represents how spread out the residuals are. The RMSE is represented as follows (Chai & Draxler, 2014; Raha & Gayen, 2020):

$$RMSE = \frac{\sqrt{\sum_{t=1}^{T}(Y_i - F_i)^2}}{\sqrt{T}} \quad (23)$$

Here, T is the total time frame; Y_i is the standardized rainfall dataset (actual or observed model); F_i is the fitted value for the TBATS and Naïve model.

3.8.1.4 Mean Error (ME)

The ME can be expressed as follows (Lee et al., 2015):

$$ME = \frac{\sum_{i=1}^{T}(Y_i - F_i)}{T} \quad (24)$$

Here, T is the total time frame; Y_i is the standardized rainfall dataset (actual or observed value); F_i is the fitted model for the TBATS and Naïve model.

3.8.1.5 Correlation Coefficient (R)

The Pearson correlation coefficient was determined using the following formula (Mukaka, 2012; Raha & Gayen, 2020):

$$r = \frac{\sum(Y_i - \bar{Y})(M_i - \bar{M})}{\sqrt{\sum(Y_i - \bar{Y})^2}\sqrt{\sum(M_i - \bar{M})^2}} \quad (25)$$

Where, r is correlation coefficient between two dataset; Y_i is the standardized rainfall dataset (observed model), $\bar{Y}$ is the mean of the standardized rainfall dataset; M_i is the fitted or predicted rainfall dataset run by the Naïve or TBATS model; $\bar{M}$ is the mean of the fitted or predicted rainfall dataset run by the naïve or TBATS model.

3.8.2 Significance test by Wilcoxon Signed Rank Test:

The significance test was done by the Wilcoxon Signed Rank Test. This test is especially applicable for a paired or single framed data. For each station, each model was tested at 95% confidence interval using this test. After computing the total number of data points $(Y_1, Y_2, \ldots\ldots\ldots, Y_n)$; total dataset

was sorted and the rank $(R_1, R_2, \dots\dots\dots, R_n)$ for respective data points were assigned. Therefore, the equation of Wilcoxon Signed Rank Test was expressed as follows (Bauer, 1972; Myles et al., 2013):

$$M = \sum_{i=1}^{T} sgn(Y_i)R_i \quad (26)$$

Where, M is the signed rank test; Y_i represents the standardized rainfall dataset (observed model) and R_i is the rank at each of the data point.

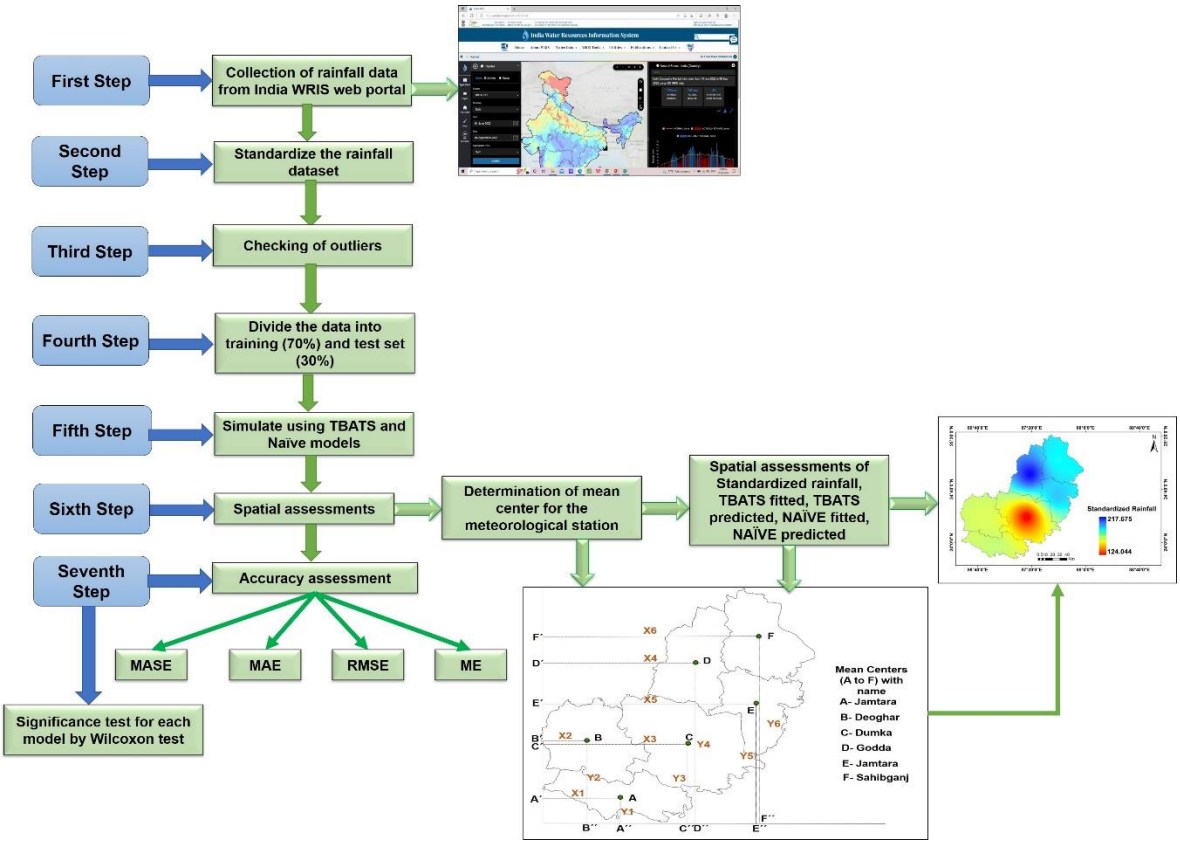

Figure 2. Methods

4. Results

4.1 Descriptive Analysis of The Downloaded Rainfall Dataset

Station wise mean and standard deviation of rainfall were portrayed in the Table 1. Dumka station was noticed with the highest mean (123.662 mm.) and standard deviation of rainfall (149.929). The coefficient of variation of rainfall was high at the Dumka and Godda meteorological stations. The Jamtara station was marked with a comparatively low mean rainfall (115.337) and standard deviation of rainfall (138.419). The lowest variation of rainfall was identified for the Pakur station (Table 1). Overall, for all station, the CV of (~Coefficient of Variation of) rainfall varies from 117% to 124% (Table 1). For all meteorological station, decreasing PACF with increasing lags were found (Figure 3a to 3f). For every station, only one data point is identified as outliers by Grubbs outlier test (Figure 4). For the Jamtara (Figure 4a), Deoghar (Figure 4b), Dumka (Figure 4c), Godda (Figure 4d) and Pakur (Figure 4e) stations 1197th row was identified as the outlier, which is 1.0. For the Sahibganj station

(Figure 4f), 1137th row was marked as the outlier. For all cases, the outliers were significant at 95% confidence interval.

Table 1. Station wise mean rainfall (mm.), standard deviation of rainfall and coefficient of variation of rainfall

Estimated mean centres	Name of the mean centres (considered as stations)	Longitude	Latitude	Mean rainfall (mm)	Standard deviation of rainfall	Coefficient of variation of rainfall
A	Jamtara	86.8171	23.9505	115.337	138.419	120.012
B	Deoghar	86.5948	24.4852	111.042	137.855	124.146
C	Dumka	87.2722	24.5846	123.662	149.929	121.240
D	Godda	87.3140	24.8619	118.458	144.796	122.234
E	Pakur	86.6376	24.5802	118.392	138.758	117.202
F	Sahibganj	87.6454	25.2381	121.485	147.705	121.582

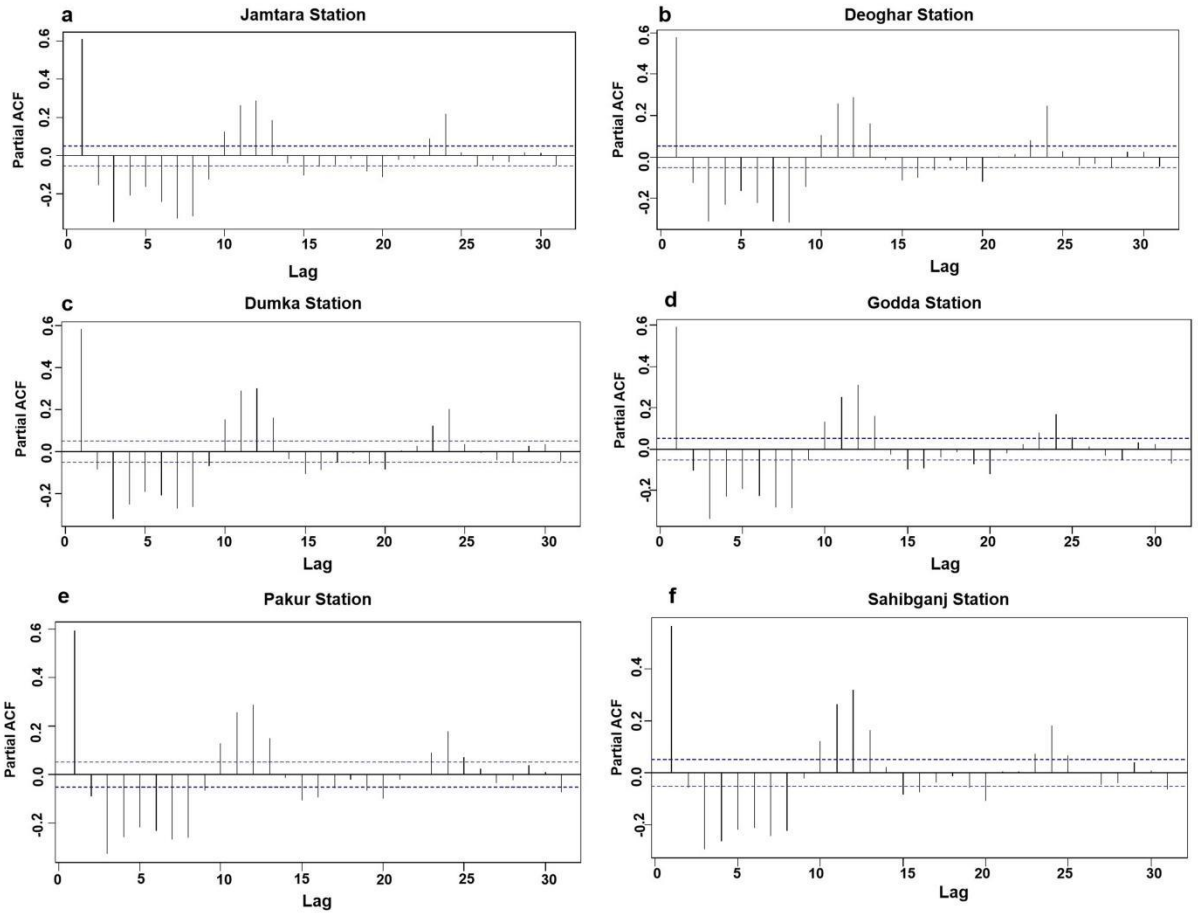

Figure 3. Partial Autocorrelation at several stations in the Santhal Pargana Division: (a) Jamtara Station; (b) Deoghar station; (c) Dumka station; (d) Godda station; (e) Pakur station; (f) Sahibganj station

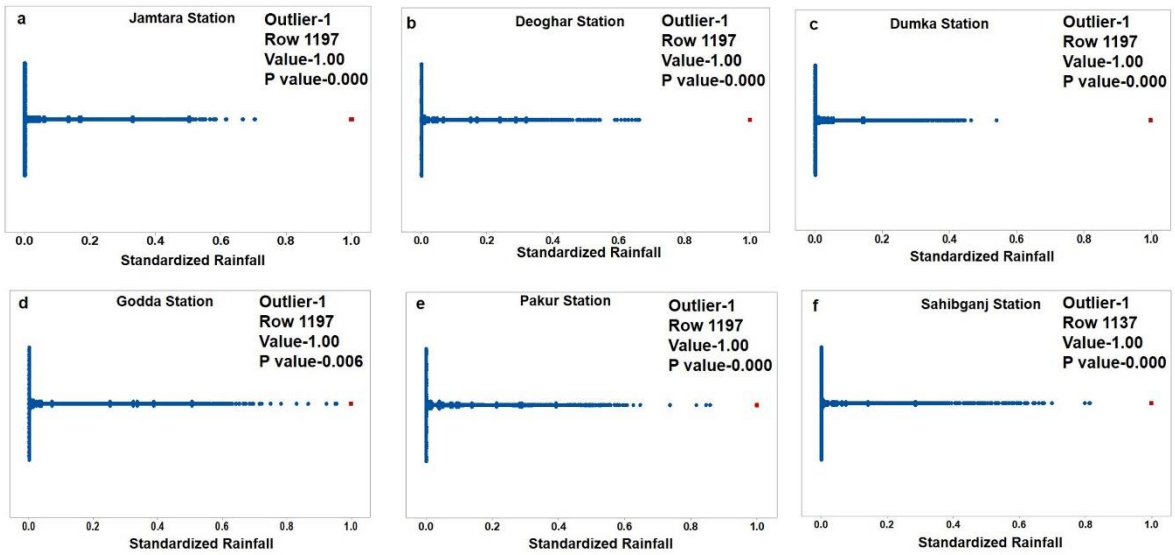

Figure 4. Schematic diagramme of outlier plot for different meteorological stations (a) Jamtara Station (b) Deoghar station (c) Dumka station (d) Godda station (e) Pakur station f. Sahibganj station

4.2 Station Wise Comparative Assessments of Observed and Simulated Models

For all stations, the observed model (standardized rainfall) and simulated models were fitted with each other entirely (Figure 5a, 5b, 5c and 5d; Figure 6a, 6b, 6c, 6d; Figure 7a,7b,7c,7d). For the Jamtara station, the sum of the standardized rainfall was 168.7262mm. For the Naïve fitted model, the rainfall was 116.8509 mm. In case of naïve predicted model, 60.2492 mm. rainfall was obtained. For the TBATS fitted and TBATS predicted models the rainfall was marked as 116.3617 and 50.6951 mm. respectively (Figure 5a and Figure 5b). The best fitted TBATS model for the Jamtara station was specified in the Table 2 (1, {0,0}, -, {<12,4>}). The α, γ_1, and γ_2 values were 0.0056, -0.0006, and 0.002 respectively. The autocorrelation function (ACF1) is 0.0421 and the sigma value is 0.0661.

For the Deoghar station, the Naïve fitted and Naïve predicted model-based rainfall values were obtained as 111.8012 mm. and 63.6812 mm. respectively. The observed standardized rainfall value was 164.6511 mm. For the TBATS fitted and TBATS predicted models the rainfall was obtained as 112.8640 mm. and 50.5391 mm. respectively (Figure 5c and 5d). The best fitted TBATS model for the Deoghar station was specified in the Table 2 (1, {0,0}, -, {<12,4>}). The α, γ_1, and γ_2 were 0.0050, -0.0009, and 0.0020 respectively. The autocorrelation function (ACF1) is 0.0094 and the sigma value is 0.0708.

Table 2. Model information of TBATS

Station Name	(ω, p, q, ø)	$\boldsymbol{\alpha, \gamma_1, \gamma_2}$	ACF1	Sigma
Jamtara	[1, {0,0}, -, {<12,4>}]	0.0056, -0.0006, 0.002	0.0421	0.0661
Deoghar	[1, {0,0}, -, {<12,4>}]	0.0050, -0.0009, 0.0020	0.0094	0.0708
Dumka	[1, {0,0}, -, {<12,4>}]	0.0072, -0.0009, 0.0020	0.0381	0.0487
Godda	[1, {0,0}, -, {<12,5>}]	0.0099, 0.0013, 0.0024	-0.0023	0.0905
Pakur	[1, {0,0}, -, {<12,4>}]	0.0103, -0.0002, 0.002	0.0034	0.0787
Sahibganj	[1, {0,0}, -, {<12,5>}]	0.0110, 0.0013, 0.0020	-0.0272	0.0804

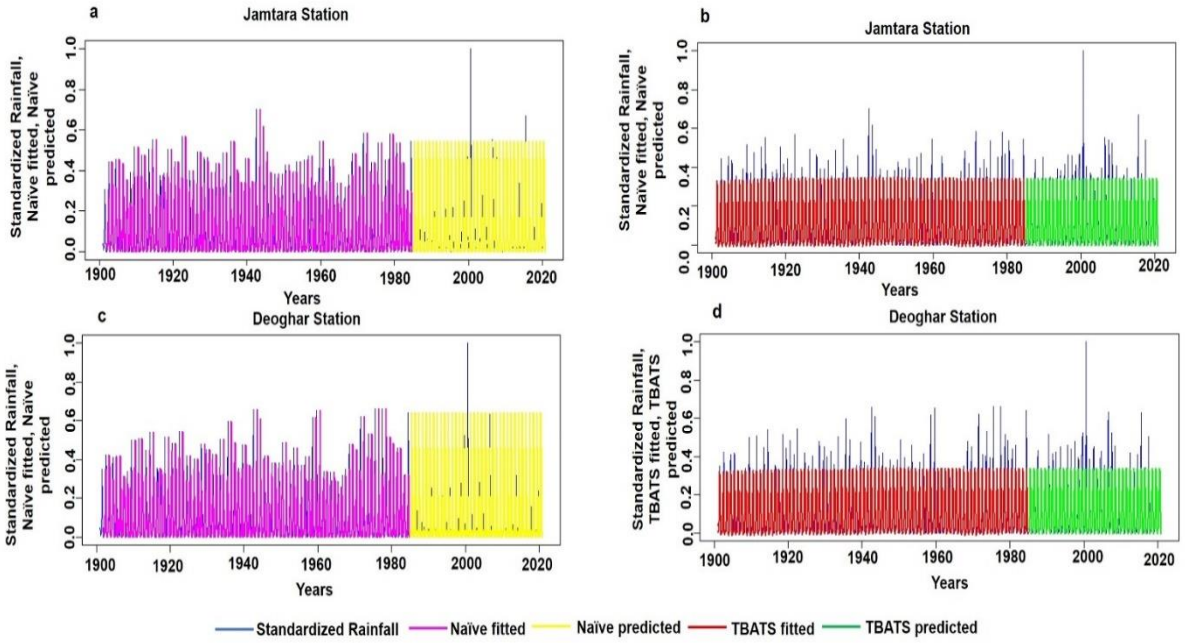

Figure 5. Observed and simulated models at Jamtara and Deoghar station

For the Dumka station, the observed rainfall was obtained as 124.0347. Simulation of rainfall by the Naïve fitted and Naïve predicted rainfall were obtained as 84.4127 and 41.2499. The simulated rainfall value for the TBATS fitted and TBATS predicted models were marked as 84.1075 mm. and 37.1244 mm. respectively (Figure 6a and 6b). The best fitted TBATS model for the Dumka station was specified in the Table 2 (1, {0,0}, -, {<12,4>}). The α, γ_1, and γ_2 were 0.0072**,** -0.0009, and 0.0020 respectively. The autocorrelation function (ACF1) is 0.0381 and the sigma value is 0.0487.

The observed standardized rainfall for the Godda station was 217.6787. The Naïve fitted and Naïve predicted rainfall for the Godda station were identified as 149.2579 mm. and 67.0600 mm. respectively. The TBATS fitted and TBATS predicted rainfall were identified as 149.2651 mm. and 62.9941 mm. respectively for the Godda meteorological station (Figure 6c and 6d). The best fitted TBATS model for the Godda station was specified in the Table 2 (1, {0,0}, -, {<12,5>}). The α, γ_1, and γ_2 were 0.0099, 0.0013, and 0.0024 respectively. The autocorrelation function (ACF1) is -0.0023 and the sigma value is 0.0905.

For the Pakur station, the sum of the standardized rainfall was 193.6738 mm. The Naïve fitted and Naïve predicted model based simulated rainfall was obtained as 133.9177 mm. and 61.1255 mm. For the TBATS fitted and TBATS predicted models, the simulated rainfall values were obtained as 133.9457 mm. and 59.8102 mm. respectively (Figure 7a and 7b). The best fitted TBATS model for the Pakur station was specified in the Table 2 (1, {0,0}, -, {<12,4>}). The α, γ_1, and γ_2 were 0.0103, -0.0002, and 0.002 respectively. The autocorrelation function (ACF1) is 0.0034 and the sigma value is 0.0787.

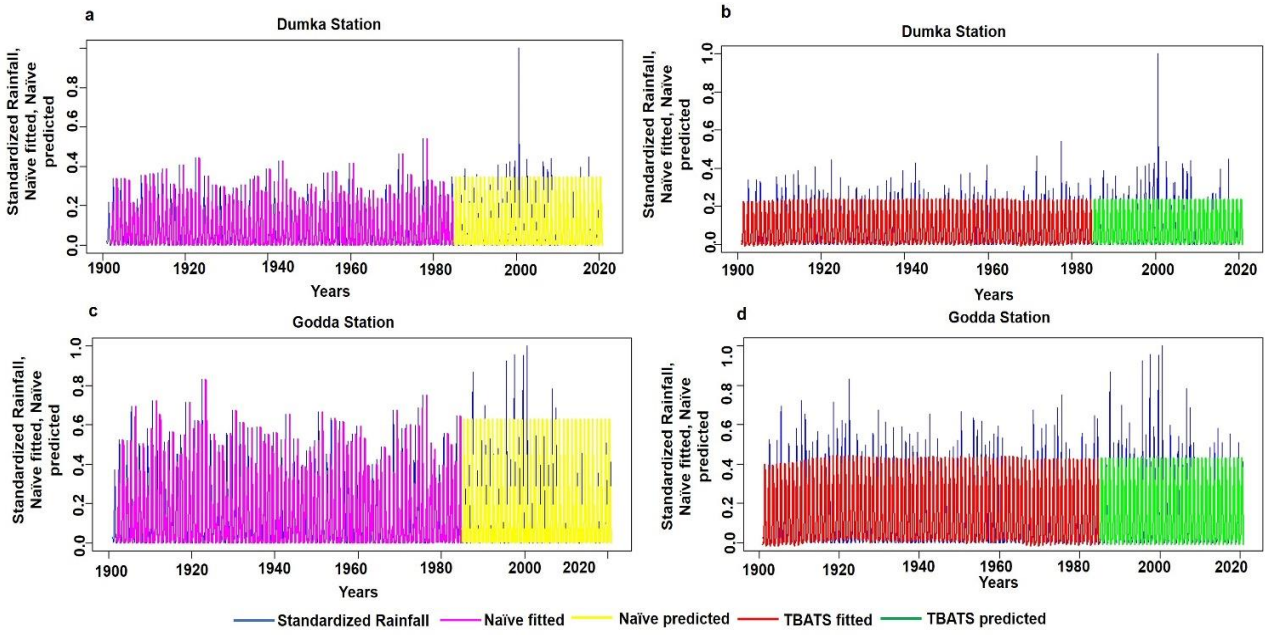

Figure 6. Observed and simulated models at Dumka and Godda station

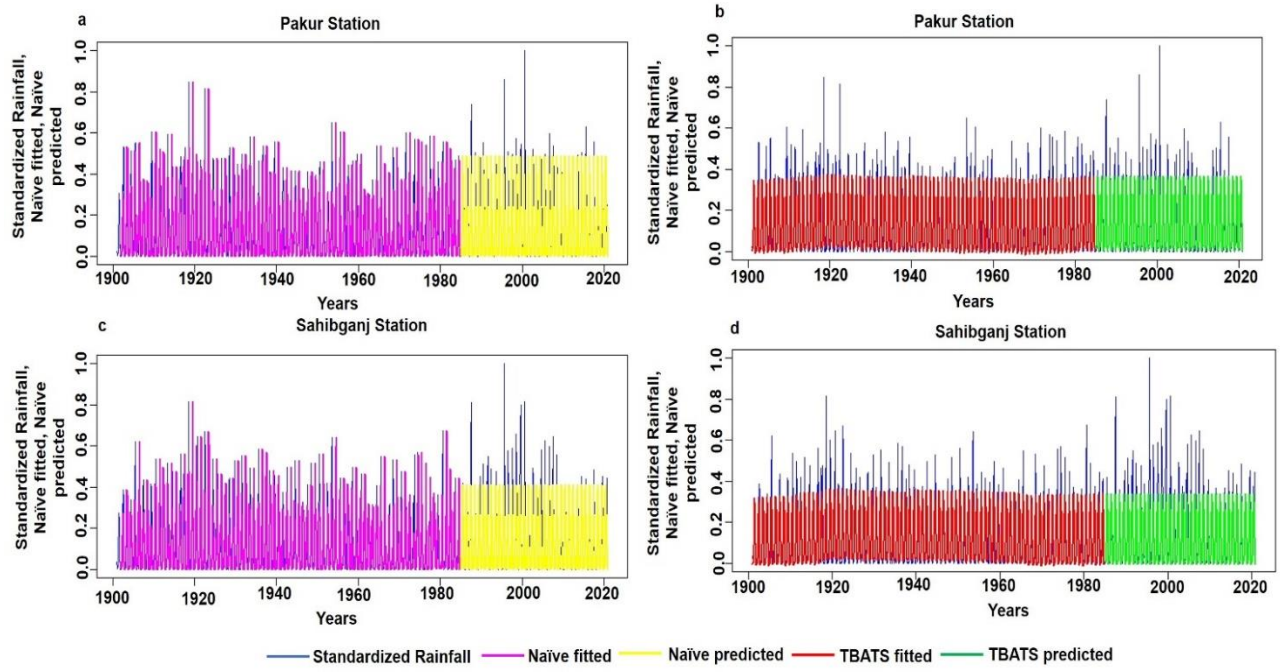

Figure 7. Observed and simulated models at the Pakur and Sahibganj station

The standardized rainfall total for the Sahibganj station was 186.281 mm.. The Naïve fitted and Naïve predicted model based simulated rainfall were identified as 124.3170 mm. and 60.5400 mm.. For the TBATS fitted and TBATS predicted models, the simulated rainfall values were marked as 125.1150 mm. and 51.4930 mm. respectively (Figure 7c and Figure 7d). The best fitted TBATS model for the Sahibganj station was specified in the Table 2 (1, {0,0}, -, {<12,5>}). The α, γ_1, and γ_2 were 0.0110, 0.0013, and 0.0020 respectively. The autocorrelation function (ACF1) is -0.0272 and the sigma value is 0.0804.

4.3 Station Wise Correlation of Each Models

The correlation matrix of each models for each station were portrayed at Table 3. For the Jamtara station, the Naïve fitted and TBATS fitted models were noticed with highest correlation coefficient value (i.e., 0.896). Highest negative correlation coefficient (-0.278) was found for the TBATS fitted and TBATS predicted models. Other models were noticed low to moderate positive or negative correlation coefficient for the Jamtara station. For the Deoghar station, Naïve predicted and TBATS predicted models were marked with the highest positive correlation coefficient (i.e., 0.878). The highest negative correlation coefficient (-0.249) was found for the TBATS fitted and TBATS predicted models. For the other models, the correlation coefficient value varied from 0.3 to 0.6.

For the Godda station, Naïve fitted and Tbats fitted and Naïve mean and TBATS predicted models were identified with comparatively high correlation coefficient value (i.e., 0.882 and 0.932 respectively). The highest negative correlation coefficient was found for the TBATS fitted and TBATS predicted models. Moderate correlation coefficient value (positive or negative; +-0.3<r<0.6) was marked for all other models, which were described in Table 3. Similarly, for the Pakur station, the Naïve predicted and TBATS predicted models were observed with the highest correlation coefficient value (i.e., 0.932). The highest negative correlation coefficient was found for the TBATS fitted and TBATS predicted models (i.e., -0.310). TBATS fitted and Naïve fitted models were marked with comparatively high (>0.8) correlation coefficient value. Other models were noticed with low to moderate correlation coefficient value (+-0.3<r<0.6). Table 3 entails the detailed measurement of observed and simulated models for the Pakur station.

For the Sahibganj station, the Naïve fitted and TBATS fitted and as well as the Naïve predicted and TBATS predicted models were noticed with comparatively higher correlation coefficient value (i.e., 0.969 and 0.866). The lowest correlation coefficient value was marked for the Naïve predicted and TBATS fitted models (i.e., -0.295). Other models are marked with moderate to high correlation coefficient value (positive or negative stretches from 0.3 to 0.6). The detailed measurement for the observed and simulated models were portrayed in the Table 3.

Table 3. Station wise correlation matrix of every models

Station Jamtara	STN Rainfall (Observed model)	Naïve Fitted	Naïve Predicted	TBATS Fitted	TBATS Predicted
STN_Rainfall (Observed model)	1.000				
Naïve Fitted	0.575	1.000			
Naïve Mean	0.337	-0.217	1.000		
TBATS Fitted	0.656	0.896	-0.235	1.000	
TBATS Predicted	0.365	-0.257	0.914	-0.278	1.000

Station Deoghar	**STN Rainfall (Observed model)**	**Naïve Fitted**	**Naïve Predicted**	**TBATS Fitted**	**TBATS Predicted**
STN_Rainfall (Observed model)	1.000				
Naïve Fitted	0.560	1.000			
Naïve Mean	0.313	-0.201	1.000		
TBATS Fitted	0.577	0.780	-0.203	1.000	
TBATS Predicted	0.356	-0.247	0.878	-0.249	1.000
Station Dumka	**STN Rainfall (Observed model)**	**Naïve Fitted**	**Naïve Predicted**	**TBATS Fitted**	**TBATS Predicted**
STN_Rainfall (Observed model)	1.000				
Naïve Fitted	0.530	1.000			
Naïve Mean	0.357	-0.236	1.000		
TBATS Fitted	0.611	0.890	-0.258	1.000	
TBATS Predicted	0.386	-0.264	0.927	-0.288	1.0000
Station Godda	**STN_Rainfall (Observed model)**	**Naïve Fitted**	**Naïve Predicted**	**TBATS Fitted**	**TBATS Predicted**
STN_Rainfall (Observed model)	1.000				
Naïve Fitted	0.545	1.000			
Naïve predicted	0.314	-0.228	1.000		
TBATS Fitted	0.628	0.885	-0.253	1.000	
TBATS predicted	0.381	-0.249	0.879	-0.278	1.000
Station Pakur	**STN_Rainfall (Observed model)**	**Naïve Fitted**	**Naïve Predicted**	**TBATS Fitted**	**TBATS Predicted**
STN_Rainfall (Observed model)	1.0000				
Naïve Fitted	0.542	1.000			
Naïve Predicted	0.328	-0.253	1.000		
TBATS Fitted	0.624	0.882	-0.282	1.000	
TBATS Predicted	0.357	-0.277	0.932	-0.310	1.000
Station Sahibganj	**STN_Rainfall (Observed model)**	**Naïve Fitted**	**Naïve Predicted**	**TBATS Fitted**	**TBATS Predicted**
STN_Rainfall (Observed model)	1.000				
Naïve Fitted	0.491	1.000			
Naïve Predicted	0.386	-0.258	1.000		
TBATS Fitted	0.574	0.866	-0.295	1.000	
TBATS Predicted	0.413	-0.254	0.969	-0.289	1.000

4.4 Spatial Assessments

Within the study area, the standardized rainfall varied from 124.044 mm. to 217.675 mm. The Dumka station was noticed with comparatively low rainfall. On the contrary, the Godda station was marked with comparatively higher rainfall. Remaining stations were marked with moderate to high amount of standardized rainfall (Figure 9a). For the Naïve fitted model, the standardized rainfall fluctuates from 84.419 mm. to 149.225 mm. The Naïve fitted value was comparatively low for the Dumka station and high for the Godda station. Remaining stations were noticed with moderate to high standardized rainfall (Figure 9b). The Deoghar station was noticed with comparatively low rainfall for the Naïve predicted model. The Godda, Deoghar and Sahibganj stations were identified with comparatively high rainfall for the Naïve predicted model. Overall, for the Naïve predicted model, the rainfall varied between 8.133 mm. to 67.059 mm. (Figure 9c). For the TBATS fitted model, the rainfall fluctuated from the 37.127 mm. to 62.993 mm. Dumka station was noticed with comparatively low rainfall (i.e.,37.127 mm.). Deoghar and Jamtara stations were marked with moderate rainfall. Remaining stations i.e., Godda, Sahibganj and Pakur stations were marked with comparatively high rainfall for the TBATS fitted model (Figure 9d). Similar, spatial pattern was noticed for the other stations (Figure 9e).

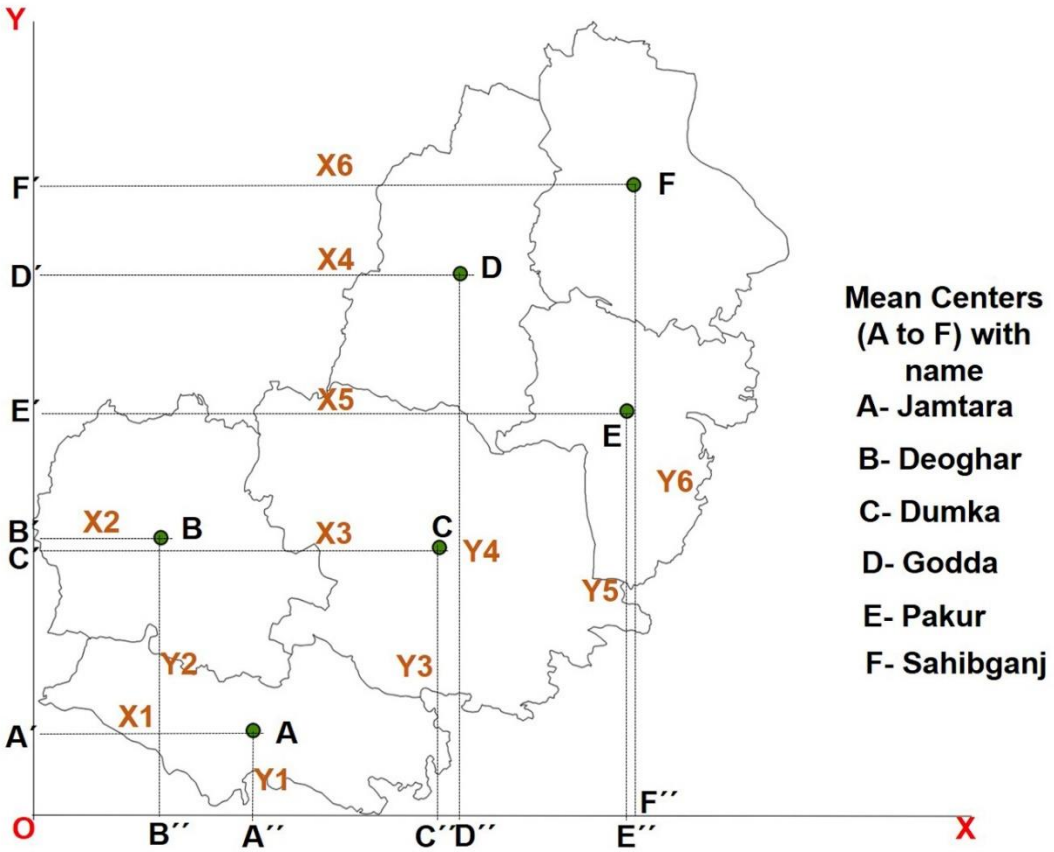

Figure 8. Assessment of mean centre for the meteorological station

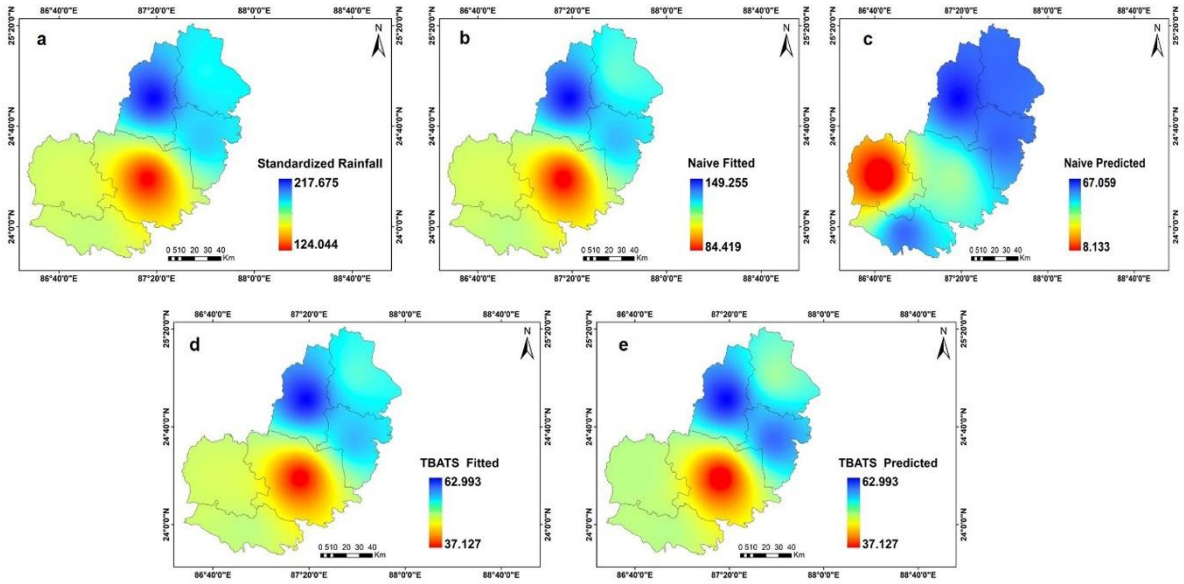

Figure 9. Spatial assessments of different observed and simulated models : a. Standardized rainfall (observed) b. Naïve fitted c. Naive predicted d. TBATS fitted e. TBATS predicted

4.5 Station Wise Comparative Assessments of Accuracy of Each Model

For the Naïve model, the Godda station was noticed with a comparatively high combined error. The lowest combined error was found for the Pakur station in case of Naïve models. By combining all stations, for the Naïve models, the highest combined error was found for the Godda station and the lowest error was marked for the Pakur station. For the Naïve model, the Mean Error (ME) was comparatively high for the Godda (0.0005), Jamtara (0.0005) and Sahibganj (0.0005) stations. For the Naïve model the Mean Error (ME) was comparatively low for the Deoghar station. For the Naïve model, the RMSE was comparatively low for the Pakur station (0.0004) and it was comparatively high for the Godda (0.1313), Deoghar (0.1013) and Sahibganj station (0.1157). The Mean Absolute Error (MAE) was comparatively high for the Godda, Sahibganj and Pakur meteorological stations and comparatively low for the Dumka (0.044) and Jamtara stations (0.0596). MASE was equal for all the stations in case of the Naïve models. R was comparatively high for the Deoghar and Dumka station and it was comparatively low for the Pakur station. For the Naïve model, the MASE was comparatively high (6.000) rather than the Mean Error (ME), which was comparatively low (0.0024) (Figure 10). The detailed calculation for the Naïve model is portrayed in Table 4 and Figure 10.

For the TBATS model, the ME was relatively high (0.0028) for the Deoghar station and low for the Sahibganj station (0.0009). RMSE was comparatively high for the Godda (0.0905) and Sahibganj station and it was low for the Pakur station (0.0017). The MAE (0.0577) was comparatively high for the Godda station and low (0.0306) for the Dumka station. MASE was comparatively high for the Pakur (0.7123) and Sahibganj stations (0.7124) and comparatively low for the Dumka (0.6966), Godda (0.6977) and Deoghar stations (0.6967). Combinedly, the Godda (0.8478) and Sahibganj stations (0.8444) were noticed with the highest combined error and the Pakur (0.7668) and Dumka (0.7773) stations were marked with the lowest combined error. Combinedly, TBATS model had achieved relatively low error (4.8614) rather than the Naïve model (6.9097). MASE was relatively high for the TBATS (4.2173) and ME was comparatively low for the TBATS model (0.0107) (Figure 11). The correlation coefficient (R) is comparatively high for the Jamtara and Godda station and comparatively low for the Deoghar station. The detailed calculation was portrayed in Table 4 and Figure 11. The combined error was comparatively high (1.3522) for the Godda station and it was comparatively low

(1.2574) for the Pakur station. At the 95% confidence interval, all models are significant at 95% confidence interval (Table 5).

Table 4. Station wise comparative study of accuracy of Naïve and TBATS models

Station name	ME Naive	RMSE Naive	MAE Naive	MASE Naive	R Naive	SUM
Deoghar	0.0002	0.1013	0.0618	1	0.4435	1.6068
Dumka	0.0003	0.0712	0.044	1	0.4435	1.559
Godda	0.0005	0.1313	0.0827	1	0.4295	1.644
Jamtara	0.0005	0.0962	0.0596	1	0.4365	1.5928
Sahibganj	0.0005	0.1157	0.0713	1	0.4385	1.626
Pakur	0.0004	0.0004	0.0717	1	0.435	1.5075
Sum (Aggregated)	**0.0024**	**0.5161**	**0.3911**	**6**	**2.6265**	**9.5361**
Station name	**ME TBATS**	**RMSE TBATS**	**MAE TBATS**	**MASE TBATS**	**R TBATS**	**SUM**
Deoghar	0.0028	0.0708	0.0433	0.6967	0.4671	1.2807
Dumka	0.0014	0.0487	0.0306	0.6966	0.4985	1.2758
Godda	0.0018	0.0905	0.0577	0.6977	0.5045	1.3522
Jamtara	0.0021	0.0661	0.0418	0.7016	0.5105	1.3221
Sahibganj	0.0009	0.0804	0.0508	0.7123	0.4935	1.3379
Pakur	0.0017	0.0017	0.0511	0.7124	0.4905	1.2574
Sum (Aggregated)	**0.0107**	**0.3582**	**0.2753**	**4.2173**	**2.9646**	**7.8261**

Table 5. Station wise significance test of each model

Station name	Name of the model	Wilcoxon Rank test of each model and P-value**
Jamtara	Observed model	V = 861328, p-value < 2.2e-16
	Naïve fitted	V = 414505, p-value < 2.2e-16
	Naïve predicted	V = 52650, p-value < 2.2e-16
	TBATS fitted	V = 501582, p-value < 2.2e-16
	TBATS predicted	V = 92862, p-value < 2.2e-16
Deoghar	Observed model	V = 849556, p-value < 2.2e-16
	Naïve fitted	V = 392055, p-value < 2.2e-16
	Naïve predicted	V = 52650, p-value < 2.2e-16
	TBATS fitted	V = 402753, p-value < 2.2e-16
	TBATS predicted	V = 92862, p-value < 2.2e-16
Dumka	Observed model	V = 860016, p-value < 2.2e-16
	Naïve fitted	V = 411778, p-value < 2.2e-16
	Naïve predicted	V = 64980, p-value < 2.2e-16
	TBATS fitted	V = 502879, p-value < 2.2e-16
	TBATS predicted	V = 92862, p-value < 2.2e-16
Godda	Observed model	V = 848253, p-value < 2.2e-16
	Naïve fitted	V = 408156, p-value < 2.2e-16
	Naïve predicted	V = 78606, p-value < 2.2e-16
	TBATS fitted	V = 499607, p-value < 2.2e-16
	TBATS predicted	V = 90900, p-value < 2.2e-16
Pakur	Observed model	V = 846951, p-value < 2.2e-16
	Naïve fitted	V = 401856, p-value < 2.2e-16
	Naïve predicted	V = 78606, p-value < 2.2e-16
	TBATS fitted	V = 503470, p-value < 2.2e-16
	TBATS predicted	V = 93528, p-value < 2.2e-16
Sahibganj	Observed model	V = 832695; p-value < 2.2e-16
	Naïve fitted	V = 399171; p-value < 2.2e-16
	Naïve predicted	V = 78606; p-value < 2.2e-16
	TBATS fitted	V = 498594; p-value < 2.2e-16
	TBATS predicted	V = 90900; p-value < 2.2e-16

Note : **95% confidence interval

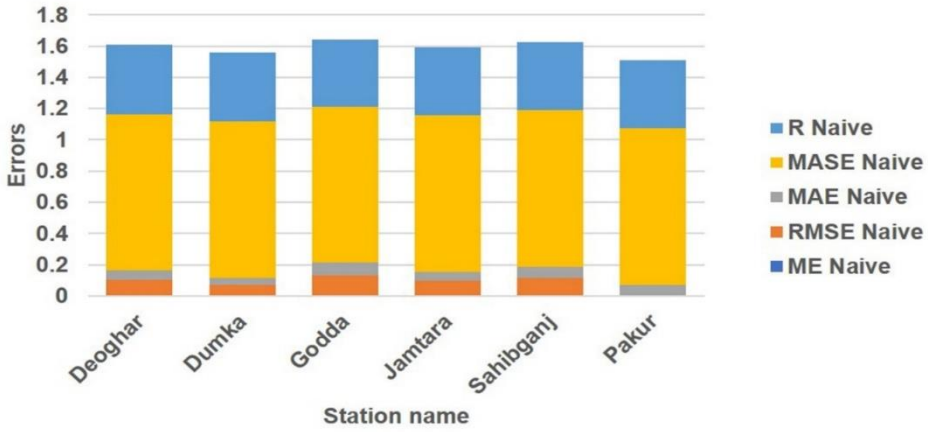

Figure 10. Accuracy assessment for the Naïve models

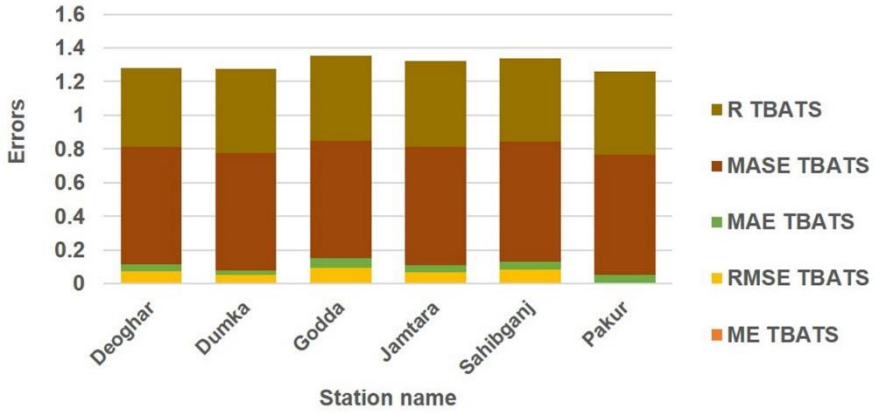

Figure 11. Accuracy assessments for the TBATS model

5. Discussion

The predictive pattern of rainfall in the Santhal Pargana Division of the Jharkhand state, India is consistent with the rainfall patterns of both Jharkhand and across all India. Potential causes of such tendency may be closely related to geo-environmental factors like global climate change (Kayet et al., 2022). The association between the India's monsoon patterns and ENSO changed around 1990 and reached its' peak positivity in the past ten years and therefore it demonstrates a recent deterioration in the relationship between ENSO and the monsoon (Bhardwaj et al., 2020). Walker circulation shifts southward, as a result, global warming, and greater surface temperatures become prominent in these sections and those are the real cause of such a relationship (Kundu & Mondal, 2019). Therefore, a single ENSO event is unable to explain the rainfall pattern in the context of Santhal Pargana. From a wider viewpoint, the crippling of the easterly and south-easterly jet stream and the exhilaration of the equatorial ocean are two local events that typically influence the increasing tendency of rainfall at eastern portions of the Jharkhand (Yadav, 2022). The study of rainfall in these sections is devoted to local level variability of rainfall at six meteorological stations. As a result, the impact of ENSO and the

effects of other climatological phenomena were not evaluated here. The local level aspects, such as the topography, altitude, gradient, etc., and modifications in agricultural property due to the introduction of irrigated agriculture, deforestation, and growing urban expansion are what this research had fundamentally implied (Kundu & Mondal, 2019). The Santhal Pargana has crystalline bedrock that were once part of a granitic environment and was thinly covered by a worn mantle (Mandal & Ray, 2015). These sections have dispersed shallow rupture zones, which are unable to hold enough groundwater (Das et al., 2022). According to Upadhayay et al. (2019), weathered fracture zone restricts the amount of groundwater in the Santhal Pargana division. The fracture zone exhibits secondary porosity, which denotes the development of a thick profile in the porous material. The hard, cemented rock is getting weathered continuously and as a result the secondary porosity is noticed. Generally, the Santhal Pargana division has experienced moderate rainfall (Shree & Kumar, 2018).

The results presented in this research are also easily comparable with other research activities done in other portions or blocks of the Jharkhand State. Shree & Kumar (2018) assessed the rainfall trends of the Ranchi District, Jharkhand using nonparametric measurements (i.e., MK test and Sen's slope). The maximum decrease of rainfall was found in the Monsoon and minimum decrease was found in the Winter. High variation of rainfall was noticed for the entire region, and these sections are prone to droughts and floods. Sharma & Singh (2021) analysed the seasonality of rainfall for the Jharkhand State, India over the last decade (1901-2002). Significant decreasing trend was obtained by the authors for the Deoghar, Dumka, Godda, Pakur and Sahibganj districts. Tirkey et al. (2018) looked at the rainfall distribution in the state of Jharkhand and discovered a drop of 26 to 270 mm to with an increasing rate upto 19 to 440 mm. Bahadur et al. (2020) assessed the rainfall trend at the Hazaribagh State, Jharkhand and found an increasing trend in the Summer rainfall. Gupta & Kumar (2018) evaluated the variance in rainfall in the Garhwa District of the state of Jharkhand and discovered that the rainfall during the Kharif season was of a consistent nature. Overall, a decreasing rainfall trend was obtained by the authors. Zamani et al., (2018) also illustrated the daily, seasonal and annual rainfall trends of the Jharkhand state, India and found overall decreasing tendency of rainfall. All of the above research works done in these sections of the Santhal Pargana Division, are supported by the present research work. With the help of the comparative assessment charts (Figure 5,6,7 and 8) it is proved that the rainfall in these portions is highly seasonal and consistent. The nature of PACF (Figure 3a to 3f) at each meteorological station of the study area also proved it. The spatial pattern (Figure 9a to 9e) also confirms that the Dumka and Sahibganj districts experience comparatively low and high rainfall respectively in the Santhal Pargana Division of the Jharkhand State. All of the research works done on these sections concentrated on the trends and tendency of the rainfall. Till date, hardly any research activities were found on the simulation and predictive pattern of the Santhal Pargana Division of the Jharkhand State. The study of rainfall in the Santhal Pargana Division of the State of Jharkhand is therefore accelerates a new dimension of applied climatology, and have an immense applicability to strategize the agricultural planning of the study area.

6. Conclusion

In this research an attempt was made to simulate the monthly rainfall during 1901 to 2020 using the Naive and TBATS model. To do it, an eight-step procedure was followed. After downloading the monthly rainfall for the Santhal Pargana Division, India during 1901 to 2020 (first step), the Partial Auto Correlation Function (PACF) was checked for different time lags (Second step). Here, the decreasing lags with increasing time steps were obtained, therefore, the experiment moved into the next step. Next, the rainfall dataset was standardized (0 to 1) (Third step) and the outliers were checked. If two or more significant outliers were found outside the range of the standardized dataset (0 to 1), then the outliers were removed and it was replaced by the mean of the preceding and forwarding values of the standardized rainfall dataset. The fourth step was marked by the division of the rainfall dataset to

training (70% of the total rainfall dataset) and test data (30% of the total rainfall dataset). Test dataset were also considered for the validation dataset. Next, the TBATS and Naive models were used to simulate the rainfall (Sixth step). The spatial assessments were done individually for the standardized rainfall (observed model), TBATS fitted, TBATS predicted, Naive fitted and Naive predicted models at the seventh step. Here, the Godda station was identified for the Naive model with a high combined error. On contrary, the Pakur station had the lowest error. A comparable outcome was also attained using the TBATS model. The standardized rainfall's spatial evaluation ranged from 84.419 mm to 149.225 mm. Rainfall for the Naive Predicted Model ranged from 8.133 mm to 67.059 mm. Rainfall varied for the TBATS fitted model, ranging from 37.127 mm to 62.993 mm. It was noted that the Dumka station had relatively little amount of rainfall (i.e.,37.127 mm.). A fair amount of rainfall was noticed for the Deoghar and Jamtara stations. The eighth step was marked by the accuracy assessment of each model by using the Mean Absolute Scaled Error (MASE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Correlation Coefficient (R) and Mean Error (ME). Due to the TBATS's lower cumulative error, it was discovered with a better degree of accuracy. Significance test for each model at each station was done using the Wilcoxon Test at the eighth step. For each station, all models were significant at 95% confidence interval. The result produced in this research is fruitful one as it can be easily utilized for the local level agricultural planning of the Santhal Pargana Division of the Jharkhand state, India.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgements

This research was supported by Department of Geography, Cooch Behar Panchanan Barma University, Cooch Behar, West Bengal, India.

References

Abdelgawad, H., Abdulazim, T., Abdulhai, B., Hadayeghi, A., & Harrett, W. (2015). Data imputation and nested seasonality time series modelling for permanent data collection stations: Methodology and application to Ontario. *Canadian Journal of Civil Engineering*, *42*(5), 287–302. https://doi.org/10.1139/cjce-2014-0087.

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052.

Alizadeh, A., Rajabi, A., Shabanlou, S., Yaghoubi, B., & Yosefvand, F. (2021). Simulation of rainfall and runoff time series using robust machine learning*. *Irrigation and Drainage*, *70*(1), 84–102. https://doi.org/10.1002/ird.2518.

Azad, Md. A. K., Islam, A. R. Md. T., Ayen, K., Rahman, Md. S., Shahid, S., & Mallick, J. (2022). Changes in monsoon precipitation patterns over Bangladesh and its teleconnections with global climate. *Theoretical and Applied Climatology*, *148*(3), 1261–1278. https://doi.org/10.1007/s00704-022-03996-8.

Azad, S., Debnath, S., & Rajeevan, M. (2015). Analysing Predictability in Indian Monsoon Rainfall: A Data Analytic Approach. *Environmental Processes*, *2*(4), 717–727. https://doi.org/10.1007/s40710-015-0108-0.

Bagirov, A. M., Mahmood, A., & Barton, A. (2017). Prediction of monthly rainfall in Victoria, Australia: Clusterwise linear regression approach. *Atmospheric Research*, *188*, 20–29. https://doi.org/10.1016/j.atmosres.2017.01.003.

Bahadur, R., Jaiswal, R. K., Nema, A. K., Gangwar, A., & Kumar, S. (2020). Trends Analysis of Rainfall and Temperature over Nagwan Watershed, Hazaribagh District, Jharkhand. *Current Journal of Applied Science and Technology*, 112–128. https://doi.org/10.9734/cjast/2020/v39i1930798.

Balan, S., M., Selvan, J. P., Bisht, H. R., Gadjil, Y. A., Khaladkar, I. R., & Lomte, V. M. (2019). Rainfall Prediction using Deep Learning on Highly Non-Linear Data. *International Journal of Research in Engineering, Science and Management*.

Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, *7*, 100204. https://doi.org/10.1016/j.mlwa.2021.100204.

Bauer, D. F. (1972). Constructing Confidence Sets Using Rank Statistics. *Journal of the American Statistical Association*, *67*(339), 687–690. https://doi.org/10.2307/2284469.

Bhardwaj, K., Shah, D., Aadhar, S., & Mishra, V. (2020). Propagation of Meteorological to Hydrological Droughts in India. *Journal of Geophysical Research: Atmospheres*, *125*(22), e2020JD033455. https://doi.org/10.1029/2020JD033455.

Bhatt, D., & Mall, R. K. (2015). Surface Water Resources, Climate Change and Simulation Modeling. *Aquatic Procedia*, *4*, 730–738. https://doi.org/10.1016/j.aqpro.2015.02.094.

Biemans, H., Siderius, C., Lutz, A. F., Nepal, S., Ahmad, B., Hassan, T., von Bloh, W., Wijngaard, R. R., Wester, P., Shrestha, A. B., & Immerzeel, W. W. (2019). Importance of snow and glacier meltwater for agriculture on the Indo-Gangetic Plain. *Nature Sustainability*, *2*(7), Article 7. https://doi.org/10.1038/s41893-019-0305-3.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1.

Box George E.P., J. G. M., & Reincel G.C., L. G. M. (n.d.). *Time Series Analysis: Forecasting and Control, 5th Edition | Wiley*. Wiley.Com.

Casado-Vara, R., Martin del Rey, A., Pérez-Palau, D., de-la-Fuente-Valentín, L., & Corchado, J. M. (2021). Web Traffic Time Series Forecasting Using LSTM Neural Networks with Distributed Asynchronous Training. *Mathematics*, *9*(4). https://doi.org/10.3390/math9040421.

Government of India. (2022). *Census of India*. Retrieved October 25, 2022, from https://censusindia.gov.in/census.website/data/census-tables.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014.

Chang, Y., Xiao, J., Li, X., Middel, A., Zhang, Y., Gu, Z., Wu, Y., & He, S. (2021). Exploring diurnal thermal variations in urban local climate zones with ECOSTRESS land surface temperature data. *Remote Sensing of Environment*, *263*, 112544. https://doi.org/10.1016/j.rse.2021.112544.

Chao, Z., Pu, F., Yin, Y., Han, B., & Chen, X. (2018). Research on Real-Time Local Rainfall Prediction Based on MEMS Sensors. *Journal of Sensors*, *2018*, e6184713. https://doi.org/10.1155/2018/6184713.

Chaturvedi, A. (2015). Rainfall Prediction using Back-Propagation Feed Forward Network. *International Journal of Computer Applications*, *119*(4), 1–5.

Cosgrove, W. J., & Loucks, D. P. (2015). Water management: Current and future challenges and research directions. *Water Resources Research*, *51*(6), 4823–4839. https://doi.org/10.1002/2014WR016869.

Dang, Q.-V. (2020). Understanding the Decision of Machine Learning Based Intrusion Detection Systems. In T. K. Dang, J. Küng, M. Takizawa, & T. M. Chung (Eds.), *Future Data and Security Engineering* (pp. 379–396). Springer International Publishing. https://doi.org/10.1007/978-3-030-63924-2_22.

Das, P., Maya, K., & Padmalal, D. (2022). Hydrogeochemistry of the Indian thermal springs: Current status. *Earth-Science Reviews*, *224*, 103890. https://doi.org/10.1016/j.earscirev.2021.103890.

Das, S. (2019). Four decades of water and sediment discharge records in Subarnarekha and Burhabalang basins: An approach towards trend analysis and abrupt change detection. *Sustainable Water Resources Management*, *5*(4), 1665–1676. https://doi.org/10.1007/s40899-019-00326-1.

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*, *106*(496), 1513–1527. https://doi.org/10.1198/jasa.2011.tm09771.

Doorn, N. (2021). Artificial intelligence in the water domain: Opportunities for responsible use. *Science of The Total Environment*, *755*, 142561. https://doi.org/10.1016/j.scitotenv.2020.142561.

Dubache, G., Ogwang, B. A., Ongoma, V., & Towfiqul Islam, A. R. Md. (2019). The effect of Indian Ocean on Ethiopian seasonal rainfall. *Meteorology and Atmospheric Physics*, *131*, 1753–1761. https://doi.org/10.1007/s00703-019-00667-8.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987–1007. https://doi.org/10.2307/1912773.

Ferrant, S., Selles, A., Le Page, M., Herrault, P.-A., Pelletier, C., Al-Bitar, A., Mermoz, S., Gascoin, S., Bouvet, A., Saqalli, M., Dewandel, B., Caballero, Y., Ahmed, S., Maréchal, J.-C., & Kerr, Y. (2017). Detection of Irrigated Crops from Sentinel-1 and Sentinel-2 Data to Estimate Seasonal Groundwater Use in South India. *Remote Sensing*, *9*(11). https://doi.org/10.3390/rs9111119.

Franses, P. H. (2016). A note on the Mean Absolute Scaled Error. *International Journal of Forecasting*, *32*(1), 20–22. https://doi.org/10.1016/j.ijforecast.2015.03.008.

Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., & Stoffel, M. (2014). 21st century climate change in the European Alps—A review. *Science of The Total Environment*, *493*, 1138–1151. https://doi.org/10.1016/j.scitotenv.2013.07.050.

Granger, C. W. J., & Andersen, A. (1978). On the invertibility of time series models. *Stochastic Processes and Their Applications*, *8*(1), 87–92. https://doi.org/10.1016/0304-4149(78)90069-8.

Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*, *11*(1), 1–21. https://doi.org/10.1080/00401706.1969.10490657.

Gupta, C. K., & Kumar, R. (2018). Rainfall variation and trend analysis of Garhwa district, Jharkhand: An assessment of spatial and seasonal variability. *Journal of Pharmacognosy and Phytochemistry*, *7*(5), 2143–2145.

Hyndman, R. J., & Athanasopoulos, A. (2002). *Forecasting: Principles and Practice (2nd ed)*. Retrieved October 23, 2022, from https://otexts.com/fpp2/.

India-WRIS. (2002). Retrieved October 25, 2022, from https://indiawris.gov.in/wris/#/rainfall.

Kayet, N., Pathak, K., Chakrabarty, A., Kumar, S., Chowdary, V. M., & Singh, C. P. (2022). Risk assessment and prediction of forest health for effective geo-environmental planning and monitoring of mining affected forest area in hilltop region. *Geocarto International*, *37*(11), 3091–3115. https://doi.org/10.1080/10106049.2020.1849413.

Kundu, S. K., & Mondal, T. K. (2019). Analysis of long-term rainfall trends and change point in West Bengal, India. *Theoretical and Applied Climatology*, *138*(3), 1647–1666. https://doi.org/10.1007/s00704-019-02916-7.

Lee, D. K., In, J., & Lee, S. (2015). Standard deviation and standard error of the mean. *Korean Journal of Anesthesiology*, *68*(3), 220–223. https://doi.org/10.4097/kjae.2015.68.3.220.

Li, H., Zhang, L., Zhou, X., & Huang, B. (2017). Cost-sensitive sequential three-way decision modeling using a deep neural network. *International Journal of Approximate Reasoning*, *85*, 68–78. https://doi.org/10.1016/j.ijar.2017.03.008.

Liu, F., & Liu, R. (2019). China, the United States, and order transition in East Asia: An economy-security Nexus approach. *The Pacific Review*, *32*(6), 972–995. https://doi.org/10.1080/09512748.2018.1526205.

Liu, Q., Zou, Y., Liu, X., & Linge, N. (2019). A survey on rainfall forecasting using artificial neural network. *International Journal of Embedded Systems*, *11*(2), 240. https://doi.org/10.1504/IJES.2019.098300.

Livera, A. M. D. (2010). Automatic forecasting with a modified exponential smoothing state space framework. In *Monash Econometrics and Business Statistics Working Papers* (No. 10/10; Monash Econometrics and Business Statistics Working Papers). Monash University, Department of Econometrics and Business Statistics. https://ideas.repec.org/p/msh/ebswps/2010-10.html.

Mandal, A., & Ray, A. (2015). Petrological and geochemical studies of ultramafic–mafic rocks from the North Puruliya Shear Zone (eastern India). *Journal of Earth System Science*, *124*(8), 1781–1799. https://doi.org/10.1007/s12040-015-0634-1.

Mehdiyev, N., & Fettke, P. (2021). Explainable Artificial Intelligence for Process Mining: A General Overview and Application of a Novel Local Explanation Approach for Predictive Process Monitoring. In W. Pedrycz & S.-M. Chen (Eds.), *Interpretable Artificial Intelligence: A Perspective of Granular Computing* (pp. 1–28). Springer International Publishing. https://doi.org/10.1007/978-3-030-64949-4_1.

Menon, A., Levermann, A., Schewe, J., Lehmann, J., & Frieler, K. (2013). Consistent increase in Indian monsoon rainfall and its variability across CMIP-5 models. *Earth System Dynamics*, *4*(2), 287–300. https://doi.org/10.5194/esd-4-287-2013.

Mie Sein, Z. M., Islam, A. R. M. T., Maw, K. W., & Moya, T. B. (2015). Characterization of southwest monsoon onset over Myanmar. *Meteorology and Atmospheric Physics*, *127*(5), 587–603. https://doi.org/10.1007/s00703-015-0386-0.

Mohapatra, M., Srivastava, A. K., Balachandran, S., & Geetha, B. (2017). Inter-annual Variation and Trends in Tropical Cyclones and Monsoon Depressions Over the North Indian Ocean. In M. N. Rajeevan & S. Nayak (Eds.), *Observed Climate Variability and Change over the Indian Region* (pp. 89–106). Springer. https://doi.org/10.1007/978-981-10-2531-0_6.

Mohd, R., Butt, M. A., & Baba, M. Z. (2020). GWLM–NARX: Grey Wolf Levenberg–Marquardt-based neural network for rainfall prediction. *Data Technologies and Applications*, *54*(1), 85–102. https://doi.org/10.1108/DTA-08-2019-0130.

Mora, C., Spirandelli, D., Franklin, E. C., Lynham, J., Kantar, M. B., Miles, W., Smith, C. Z., Freel, K., Moy, J., Louis, L. V., Barba, E. W., Bettinger, K., Frazier, A. G., Colburn IX, J. F., Hanasaki, N., Hawkins, E., Hirabayashi, Y., Knorr, W., Little, C. M., … Hunter, C. L. (2018). Broad threat to humanity from cumulative climate hazards intensified by greenhouse gas emissions. *Nature Climate Change*, *8*(12), Article 12. https://doi.org/10.1038/s41558-018-0315-6.

Mukaka, M. M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal*, *24*(3). https://doi.org/10.4314/mmj.v24i3.

Mukherjee, S., Mishra, A., & Trenberth, K. E. (2018). Climate Change and Drought: A Perspective on Drought Indices. *Current Climate Change Reports*, *4*(2), 145–163. https://doi.org/10.1007/s40641-018-0098-x.

Myles, H., Wolfe, Douglas. A., & Checksen, E. (n.d.). *Nonparametric Statistical Methods, 3rd Edition | Wiley*. Wiley.Com.

Nourani, V., Uzelaltinbulat, S., Sadikoglu, F., & Behfar, N. (2019). Artificial Intelligence Based Ensemble Modeling for Multi-Station Prediction of Precipitation. *Atmosphere*, *10*(2). https://doi.org/10.3390/atmos10020080.

Prakash, V., & Mishra, V. (2022). Soil Moisture and Streamflow Data Assimilation for Streamflow Prediction in the Narmada River Basin. *Journal of Hydrometeorology*, *1*(aop). https://doi.org/10.1175/JHM-D-21-0139.1.

Raha, S., & Gayen, S. K. (2020). Simulation of meteorological drought using exponential smoothing models: A study on Bankura District, West Bengal, India. *SN Applied Sciences*, *2*(5), 909. https://doi.org/10.1007/s42452-020-2730-3.

Rahman, Md. R., & Lateh, H. (2017). Climate change in Bangladesh: A spatio-temporal analysis and simulation of recent temperature and rainfall data using GIS and time series analysis model. *Theoretical and Applied Climatology*, *128*(1), 27–41. https://doi.org/10.1007/s00704-015-1688-3.

Saha, A., Singh, K. N., Ray, M., & Rathod, S. (2020). A hybrid spatio-temporal modelling: An application to space-time rainfall forecasting. *Theoretical and Applied Climatology*, *142*(3), 1271–1282. https://doi.org/10.1007/s00704-020-03374-2.

Samantaray, S., Tripathy, O., Sahoo, A., & Ghose, D. K. (2020). Rainfall Forecasting Through ANN and SVM in Bolangir Watershed, India. In S. C. Satapathy, V. Bhateja, J. R. Mohanty, & S. K. Udgata (Eds.), *Smart Intelligent Computing and Applications* (pp. 767–774). Springer. https://doi.org/10.1007/978-981-13-9282-5_74.

Scheuerer, M. (2014). Probabilistic quantitative precipitation forecasting using Ensemble Model Output Statistics. *Quarterly Journal of the Royal Meteorological Society*, *140*(680), 1086–1096. https://doi.org/10.1002/qj.2183.

Shah, N. V., Patel, Y. S., & DBhangaonkar, P. (2021). Assessing Impact of Climate Change on Rainfall Patterns of Vadodara District, Gujarat, India. *Journal of Physics: Conference Series*, *1714*(1), 012046. https://doi.org/10.1088/1742-6596/1714/1/012046.

Sharma, T., & Singh, S. (2021). Feto-maternal outcome in previous one cesarean section: A retrospective observational study at a district hospital of Jharkhand. *International Journal of Reproduction, Contraception, Obstetrics and Gynecology*, *10*(10), 3834–3840.

Shmuel, A., Ziv, Y., & Heifetz, E. (2022). Machine-Learning-based evaluation of the time-lagged effect of meteorological factors on 10-hour dead fuel moisture content. *Forest Ecology and Management*, *505*, 119897. https://doi.org/10.1016/j.foreco.2021.119897.

Shree, S., & Kumar, M. (2018). Analysis of seasonal and annual rainfall trends for Ranchi district, Jharkhand, India. *Environmental Earth Sciences*, *77*(19), 693. https://doi.org/10.1007/s12665-018-7884-6.

Silva, L. da. (2021). *Modelagem geométrica do par roda-trilho com descrição por meio de NURBS*. [Text, Universidade de São Paulo]. https://doi.org/10.11606/D.3.2021.tde-25102021-145823.

Singh, D., Ghosh, S., Roxy, M. K., & McDermid, S. (2019). Indian summer monsoon: Extreme events, historical changes, and role of anthropogenic forcings. *WIREs Climate Change*, *10*(2), e571. https://doi.org/10.1002/wcc.571.

Stefansky, W. (2012). *Rejecting Outliers in Factorial Designs. Technometrics: 14*(2). https://www.tandfonline.com/doi/abs/10.1080/00401706.1972.10488930.

Tirkey, A. S., Ghosh, M., Pandey, A. C., & Shekhar, S. (2018). Assessment of climate extremes and its long term spatial variability over the Jharkhand state of India. *The Egyptian Journal of Remote Sensing and Space Science*, *21*(1), 49–63. https://doi.org/10.1016/j.ejrs.2016.12.007.

Tong, H., & Lim, K. S. (1980). Threshold Autoregression, Limit Cycles and Cyclical Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *42*(3), 245–292.

Upadhayay, U., Kumar, N., Kumar, R., & Kumari, P. (2019). Hydro-Geological Status of the Core and Buffer Zone of Beekay Steel Industries Limited, Adityapur Industrial Area, Saraikela, Kharsawan, Jharkhand. In *Wastewater Reuse and Watershed Management*. Apple Academic Press.

Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., & Selbig, J. (n.d.). *Parallel analysis of transcript and metabolic profiles: A new approach in systems biology | EMBO reports*. Retrieved October 25, 2022, from https://www.embopress.org/doi/full/10.1038/sj.embor.embor944.

Vuyyuru, V. Ankalu., Apparao, Giduturi., & Anuradha, S. (2021). Rainfall Prediction Using Machine Learning Based Ensemble Model. *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, 1–4. https://doi.org/10.1109/ISCON52037.2021.9702409.

Wang, W., & Lu, Y. (2018). Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conference Series: Materials Science and Engineering*, *324*(1), 012049. https://doi.org/10.1088/1757-899X/324/1/012049.

Wang, Y., Yuan, Z., Liu, H., Xing, Z., Ji, Y., Li, H., Fu, Q., & Mo, C. (2022). A new scheme for probabilistic forecasting with an ensemble model based on CEEMDAN and AM-MCMC and its application in precipitation forecasting. *Expert Systems with Applications*, *187*, 115872. https://doi.org/10.1016/j.eswa.2021.115872.

Xiang, Y., Gou, L., He, L., Xia, S., & Wang, W. (2018). A SVR–ANN combined model based on ensemble EMD for rainfall prediction. *Applied Soft Computing*, *73*, 874–883. https://doi.org/10.1016/j.asoc.2018.09.018.

Xu, C., Liao, Z., Li, C., Zhou, X., & Xie, R. (2022). Review on Interpretable Machine Learning in Smart Grid. *Energies*, *15*(12). https://doi.org/10.3390/en15124427.

Yadav, M. (2022). South Asian Monsoon Extremes and Climate Change. In P. Saxena, A. Shukla, & A. K. Gupta (Eds.), *Extremes in Atmospheric Processes and Phenomenon: Assessment, Impacts and Mitigation* (pp. 59–86). Springer Nature. https://doi.org/10.1007/978-981-16-7727-4_4.

Yaseen, Z. M., Ali, M., Sharafati, A., Al-Ansari, N., & Shahid, S. (2021). Forecasting standardized precipitation index using data intelligence models: Regional investigation of Bangladesh. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-82977-9.

Zamani, R., Mirabbasi, R., Nazeri, M., Meshram, S. G., & Ahmadi, F. (2018). Spatio-temporal analysis of daily, seasonal and annual precipitation concentration in Jharkhand state, India. *Stochastic Environmental Research and Risk Assessment*, *32*(4), 1085–1097. https://doi.org/10.1007/s00477-017-1447-3.

Zhang, T., Yang, S., Jiang, X., & Zhao, P. (2016). Seasonal–Interannual Variation and Prediction of Wet and Dry Season Rainfall over the Maritime Continent: Roles of ENSO and Monsoon Circulation. *Journal of Climate*, *29*(10), 3675–3695. https://doi.org/10.1175/JCLI-D-15-0222.1.