Sigma J Eng Nat Sci, Vol. 43, No. 5, pp. 1636–1650, October, 2025

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00156

Research Article

A creative combination of time series models for accurate forecasting of monthly rainfall: A case study in Northeast India

D KARTHIKA[1], K KARTHIKEYAN[1,*]

[1]Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, Vellore, Tamil Nadu, 632014, India

ARTICLE INFO

Article history
Received: 08 May 2024
Revised: 24 June 2024
Accepted: 03 November 2024

Keywords:
ANN; Combined Forecasts; Rainfall Forecasting; SARIMA; Time Series Analysis

ABSTRACT

Rainfall forecasting is a complex and critical problem faced by many meteorologists. Classical forecasting models are struggle to capture the seasonal variations and long-term trends in rainfall data. So, it is essential to develop a more robust method to rainfall forecasting. The Seasonal Autoregressive Integrated Moving Average (SARIMA) and Holt-Winters Additive (HWA) models are used for parallel hybridization, with optimal weights determined by the variance-covariance matrix method. We evaluated the proposed model using monthly rainfall data from Jan 1990 to Dec 2017 of Northeast (NE) India. This region is divided into five divisions based on rainfall pattern that are West Bengal and Sikkim (WBS), Arunachal Pradesh (AP), Assam and Meghalaya (AM), Gangetic West Bengal (GWB), and Nagaland, Manipur, Mizoram, and Tripura (NMMT). The developed model is performed better than the classical models like SARIMA, HWA, Exponential Smoothing (ETS), Holt model, and FeedForward Neural Network (FFNN) across all regions. In the WBS region, it achieved an RMSE of 0.0798, an MAE of 0.0453, an MSE of 0.0063, an sMAPE of 0.3939, a correlation of 0.9414 between actual and predicted values, and an NSE of 0.8855. These results have significant implications for flood and drought management, climate change adaptation, and agricultural planning, particularly in the context of increasing climate variability.

Cite this article as: Karthika D, Karthikeyan K. A creative combination of time series models for accurate forecasting of monthly rainfall: A case study in Northeast India. Sigma J Eng Nat Sci 2025;43(5):1636–1650.

INTRODUCTION

Predicting rainfall precipitation is the most challenging research because of its time and spatial variance. So, precipitation prediction plays a significant role in society. The rainfall precipitation variance affects agricultural production, industry, and hydroelectric power generation, causing severe strain on the economy. The scarceness of monsoon rainfall rigorously affects large parts of the country [1,2]. Rainfall forecasting is important for minimizing drought and flood damages, improving farming production, and establishing appropriate irrigation plans [3].

Classical forecasting models estimate rainfall based solely on previous records, resulting in insufficient accuracy. To mitigate this issue, statistical approaches have been proposed. Statistical modeling is a significant approach

***Corresponding author.**
*E-mail address: k.karthikeyan@vit.ac.in

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

for testing, predicting, and deciding about hydrological cycle components. Over the past few decades, most of the acadamics have employed statistical techniques to address hydrological challeges [4–7].

The main factors influencing rainfall forecasts include monthly, annual, and seasonal cycles. In this article, we chose monthly rainfall data. Basically, in time series forecasting two types of models are used to predict: univariate model, and multivariate model. The multivariate time series models [6,8] are based on multiple parameters, including historical data, temperature, wind, and other influencing factors. Similarly, the univariate time series models [5,9–11] are based on single time series data for making short and long-term forecasts. To investigate the impact of univariate and multivariate time series models on the forecasting of rainfall, several forecasting techniques have been developed by many researchers. For accurate forecasting, some researchers concentrated on multivariate models. But it has some limitations: the availability of the data, and the complex structures of the model. So, our primary aim and objective of this study are to forecast rainfall precipitation using univariate time series models.

Considering the numerous monthly and seasonal patterns in model parameters and estimates, univariate time series models can be extremely beneficial for demand forecasting. In our research, we analyze various univariate forecasting models used by researchers worldwide to address hydrological issues and select suitable models for hybridization to improve forecast accuracy.

The ARIMA is a univariate time series model that is commonly used for forecasting. Box and Jenkins introduced the ARIMA model in the early 1970s. This model is also useful for identifying patterns in quasi time series. The ARIMA model is used to build the pure Seasonal ARIMA (p, d, q) (P, D, Q)s model [12]. In SARIMA model, the components P, D, and Q represent the suitable “seasonal autoregressive, integrated, and moving average components and p, d, and q represent autoregressive, integrated, and moving average components” [13,14]. The SARIMA model is commonly utilized in climatology, technology, financial statistics, and manufacturing prediction. Some authors have studied hydrological data prediction and used the SARIMA model for analyzing precipitation patterns in different regions [15–17].

Holt [18] modified exponentially weighted moving averages to provide the trend and seasonal variation. exponential smoothing is a common method for forecasting seasonal data. A Modified version of the exponential model is called the Holt-Winters model. it is a statistical forecasting model in univariate time series techniques. This model is useful for seasonal time series data, and it is predicted well [5,7].

The Artificial Neural Network (ANN) model is inspired by the human brain which is connected by neurons. It uses artificial neurons connected in layers. This model helps in predicting the nonlinear part of rainfall data. So, many researchers used ANN models to forecast rainfall data [19–21].

Similarly, A univariate time series model for examining the level, trend, and seasonal elements of time series data is the exponential smoothing model. A modified form of the weighted moving average that takes trend and seasonal fluctuations into consideration is the Holt model [18].

Accordingly, we can integrate forecasts produced from the most precise forecasting systems for various forecasting sources and ranges to reduce their susceptibility to weather changes as well as other seasonal patterns. Bates and Granger, [22] introduced the concept of the combined forecast by adding weights to attach the individual forecasting method. Newbold and Granger,[23] compared the performance of the combined forecasts in terms of the ratio of average squared forecast errors. Averaging distinct forecasts, and combining forecasts helps reduce errors [24]. This is especially helpful when we are unsure about the best forecasting technique to apply. Winkler and Makridakis, [25] and others have made some significant contributions to combined forecasts of univariate time series models [26,27]. The combined forecast is a very productive way to achieve greater accuracy with minimum effort and time. We employ a novel strategy for this type of research. To employ the strategy, we used 28 years of monthly rainfall data (Jan 1990-Dec 2017) in NE India. NE India has a peculiar climate which means that the region receives rainfall during summer compared to other parts of India. We propose a novel parallel hybrid model using SARIMA and HWA models to forecast the rainfall precipitation in NE India. This approach assembles the unique modeling strengths of SARIMA and HWA. In terms of popular error metrics, the proposed model is compared with SARIMA, HWA, ETS, ANN, and Holt methods predicting accuracy in NE India.

Many existing forecasting models have high computational complexity for large scale applications and the scalability is still another significant problem. These are the limitations of existing models.

Our developed model overcomes this limitation by parallel processing, considering univariate dataset, improved forecasting accuracy.

In addition to addressing the significant limitations of the existing models, enhancements in our methodology pave the way for future study and implementation in the metrological department.

Objectives of the Study

The primary objective of this study is to develop a parallel hybrid model combining Seasonal Autoregressive Integrated Moving Average (SARIMA) and Holt-Winters Additive (HWA) methods for improved rainfall forecasting in Northeast India.

The secondary objective is to compare the efficacy of the presented parallel hybrid model with existing forecasting models to demonstrate its effectiveness and advantages.

The remaining article is organized as follows: The next section discusses the study area. Section 3 describes the data and methodologies of the SARIMA, HWA, ANN, and our proposed hybrid approach. Section 4 presents the empirical research evidence, which is followed by conclusions in Section 5.

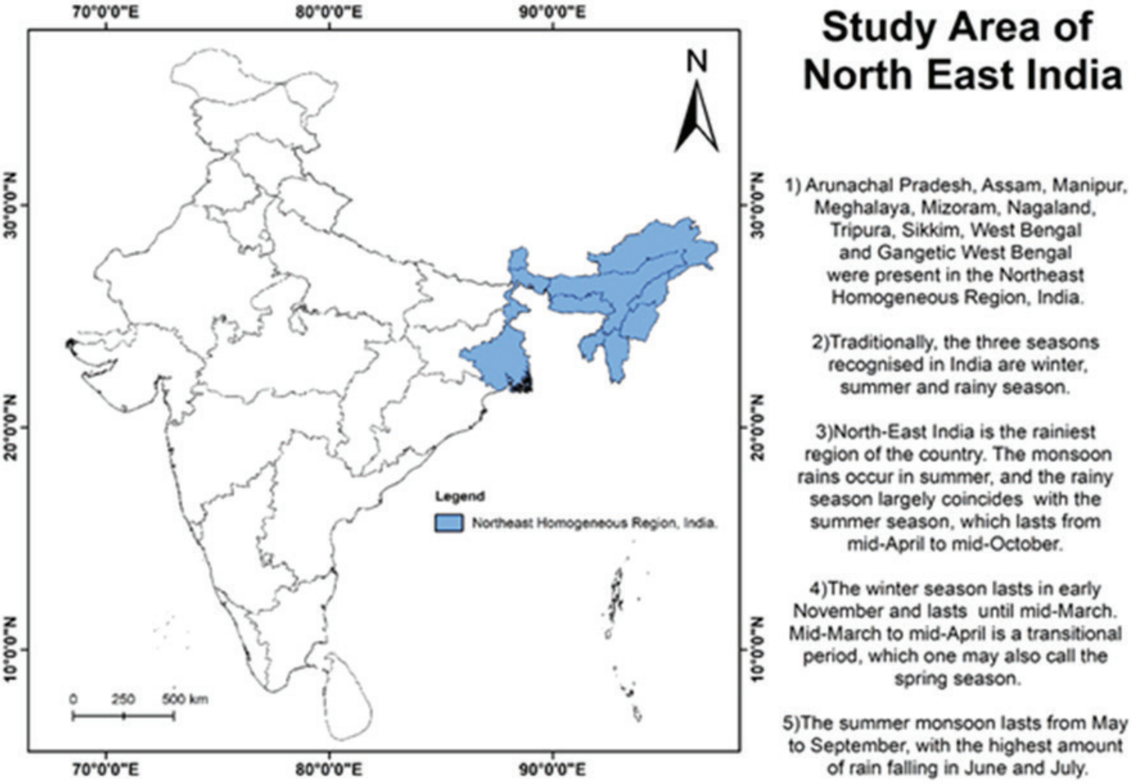
STUDY AREA

Northeast (NE) India is the easternmost region of the country. In this study, we collected monthly rainfall data for the period 1990 to 2017 for the homogeneous region of NE India (Figure 1). Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Tripura, Sikkim, West Bengal, and Gangetic West Bengal were present in the Northeast Homogeneous Region. Based on the homogenous properties of rainfall patterns, these states are combined as NE India. Further, the subdivisions are made to analyze the monthly rainfall data series. The NE region of India shares its borders with various countries. This region shares an international border with several neighbouring countries. It is bounded by China in the north, Myanmar in the east, Nepal in the west and Tibet in the northeast, Bangladesh in the southwest, and Bhutan in the northwest. It has a total area of 262,230 square kilometres which is 8% of the total area of India [2]. This region has a humid subtropical climate. That is, it has mild summers, strong monsoons, and extreme cold. It has the largest area of rainforest in the rest of India. Some parts of Northeast India receive an annual rainfall of 2000 mm. Arunachal Pradesh and Sikkim have an alpine climate with cold, frigid winters and warm summers [28]. Cherrapunji in Meghalaya receives rainfall throughout the year. The average annual rainfall of this place is 11,777mm. It is one of the rainiest regions of the world. 90% of the total rainfall of this region falls during the southwest monsoon. April to October is the rainy season and June and July are the heaviest rainy months [6].

DATA AND METHODS

Data

The monthly average rainfall precipitation of NE India, data from January 1990 to December 2017, have been collected from the Open Government Data Platform India (OGD) https://data.gov.in/rainfall-india. The 28 years of time series data were analyzed using the software R. To picture data, elementary statistics like Mean, Standard Deviation (SD), Minimum, and Maximum has calculated. Rainfall time series data were separated into training and testing datasets. Seventy percent of the data, from January 1990 to June 2009, were used as the training dataset, while



Study Area of North East India

1) Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Tripura, Sikkim, West Bengal and Gangetic West Bengal were present in the Northeast Homogeneous Region, India.

2)Traditionally, the three seasons recognised in India are winter, summer and rainy season.

3)North-East India is the rainiest region of the country. The monsoon rains occur in summer, and the rainy season largely coincides with the summer season, which lasts from mid-April to mid-October.

4)The winter season lasts in early November and lasts until mid-March. Mid-March to mid-April is a transitional period, which one may also call the spring season.

5)The summer monsoon lasts from May to September, with the highest amount of rain falling in June and July.

Figure 1. Study Area of Northeast India.

Table 1. Statistical Analysis for Northeast India rainfall data

Rainfall Data	Minimum	Maximum	Mean	SD	CV
a)WBS					
Actual Data	0	916.6	227.5658	233.8672	102.769
Training Data	0	916.6	235.6372	240.7008	102.148
Testing Data	0	777.9	209.049	217.3862	103.988
b)AP					
Actual Data	0.6	857.2	220.1321	193.7594	88.019
Training Data	1.4	857.2	221.4679	196.3434	88.655
Testing Data	0.6	724.9	217.0676	188.6143	86.891
c)AM					
Actual Data	0.2	848.4	203.9932	189.0011	92.650
Training Data	0.2	848.4	208.1372	190.8406	91.689
Testing Data	0.4	729.7	194.4863	185.2894	95.271
d)GWB					
Actual Data	0	633.1	132.3313	141.1562	106.668
Training Data	0	610.1	137.2491	142.7586	104.014
Testing Data	0	633.1	121.049	137.4342	113.536
e)NMMT					
Actual Data	0	619.4	179.8923	158.1471	88.26
Training Data	0	619.4	186.4244	161.7294	86.753
Testing Data	0.1	543.3	164.9069	149.2949	90.532

a) WBS- West Bengal and Sikkim, b) AP- Arunachal Pradesh, c) AM- Assam and Meghalaya, d) GWB- Gangetic West Bengal, e) NMMT- Nagaland, Manipur, Mizoram, and Tripura. SD is short form of standard deviation and CV is coefficient of variation.

thirty percent of the data, from July 2009 to December 2017, were used as the testing dataset. NE India has separated into five regions based on similarity characterization of rainfall precipitation. They are West Bengal and Sikkim (WBS), Arunachal Pradesh (AP), Assam and Meghalaya (AM), Gangetic West Bengal (GWB), Nagaland, Manipur, Mizoram, and Tripura (NMMT). The statistical analysis of the NE India rainfall data region-wise is in Table 1.

Decomposing Time Series

It is used to find the trend and seasonal components in time series data. It indicates basic components like an irregular, trend, and seasonal components. In Figure 2, The first row indicates the observed time series, the second row indicates the estimated irregular or residual component, the third row indicates the estimated seasonal component, and the final row indicates the estimated trend component [29]. The observed time series shows up and down pattern that indicates the rainfall data series remains seasonal. According to Figure 2, NE India's rainfall time series data have a seasonal component, with regular increasing and decreasing trends occurring annually. By Figure 2, WB, and Sikkim, from 2000 to 2015, almost the same trend in rainfall precipitation. Arunachal Pradesh met a downward trend from the year 2014 to 2017. GWB and AP met an almost downscale trend in 2010. In 2014, Nagaland, Manipur, Mizoram, Tripura, Assam, and Meghalaya had an increasing trend in rainfall precipitation.

MATERTIALS AND METHODS

Normalizing the Data

The data must be pre-processed before implementing the forecasting models. In data-driven modeling methodologies, data pre-processing is often employed to reduce any anomalies, incomplete data, or inaccurate data [8]. For the whole rainfall dataset, the equation as follows is applied to normalize the time series data:

$$Norm(X) = \frac{X - Minimum(X)}{Maximum(x) - Minimum(X)} \quad (1)$$

Here, X is time series data, and Norm(X) is normalized X. we further use this value of rainfall data to compute the models and predict.

SARIMA Model

The SARIMA model is the best linear model for univariate data analysis and forecasting. it has been used to

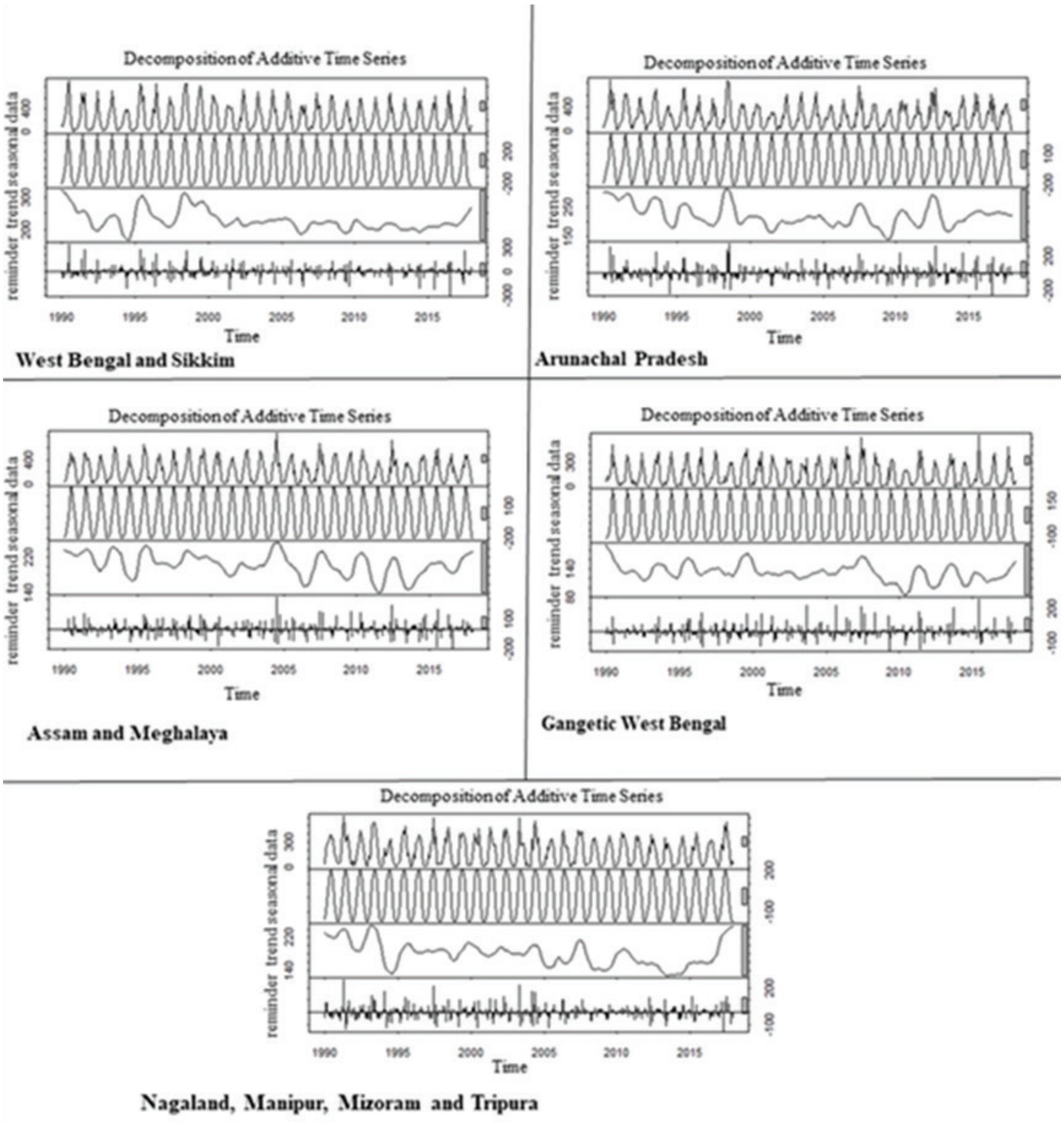


Figure 2. Decomposition of Time Series.

forecast and simulate the behaviour of univariate time series data [4,30]. The SARIMA modeling is based on correlational approaches, which must be utilized to represent components that are not obvious in the available data. The steps involved in predicting rainfall precipitation in NE using SARIMA are stationarity check, identification, and selection. The stationarity of time series data has been checked by Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) values. Also, the statistical test, Augmented Dickey–Fuller (ADF), has been applied.

its general form of ARIMA (p, d, q) (P, D, Q) s is as follows [29].

$$\varphi(B)\varphi(B^s)(1-B)^d(1-B^s)^D(Z_{t-\mu}) = \theta(B)\theta(B^s)\in_t \qquad (2)$$

Here, $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p$

(The p order of the AR term),

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$

(The q order of the MA term),

$\varphi(B^s) = 1 - \varphi_1 B^s - \varphi_2 B^{2s} - \ldots - \varphi_P B^{Ps}$

(The P order of the seasonal AR term),

$\theta(B^s\,) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \ldots - \theta_Q B^{Qs}$

(The Q order of the seasonal MA term)

and $\in(t) \sim WN(0, \sigma^2)$, the difference d: quasi number, s is the absolute value which is always higher than one. Here μ is 0, if d or D is greater than 0. The ACF and PACF functions are used to estimate the model's order. The parameterization process is computed by maximum likelihood approaches after the model order has been determined by the Akaike information criteria, Bayesian information criteria, and other model selection criteria. The best fitted

SARIMA models for these regions are selected by the minimum value of RMSE, MAE, sMAPE, and MSE.

Holt-Winter Additive Method (HWA)

This approach uses a weighted moving average. It is used to evaluate seasonality, trend, and level [31]. The equations for Holt-Winters' additive approach for a series $\{X_t\}$ with period m are:

$$A_t = \gamma(X_t - S_{t-m}) + (1-\gamma)(A_{t-1} + B_{t-1}) \tag{3}$$

$$B_t = \alpha(A_t - A_{t-1}) + (1-\alpha)B_{t-1} \tag{4}$$

$$C_t = \beta(X_t - A_t) + (1-\beta)C_{t-m} \tag{5}$$

$$F_n(h) = A_n + hB_n + C_{n+h-m.} \tag{6}$$

In this case, the smoothing estimates of level, trend, and seasonality at time t are represented by the variables A_t, B_t, and C_t, respectively. The smoothing parameters are γ, α, and β. They are employed to distinguish between the effects of recent and old observational data. In Eq. (6), The time series data's level at time n is represented by A_n and its trend at time n is represented by B_n. The trend is represented by hB_n, and the seasonal effect is represented by C_{n+h-m} (for yearly data m = 12, C_{n+h-12} is the expected seasonal in the corresponding month of the previous year). For point forecasting and model evaluation, we utilize the R software's "forecast" package.

Feed Forward Neural Networks (FFNN)

It is a type of ANN which is very good at stimulating the nonlinear patterns generally present in time series data from the real world [19]. Layers of synthetic neurons that mimic the neuronal connections seen in the human brain compose these models. The standard form of this model is NNAR (p, k); here 'p' stands for the number of lag inputs and 'k' for the number of hidden layers. The seasonal variant of FFNN model is indicated as NNAR (p, P, k). its working flow is single forward direction there is no recurrent or backward connections. There are no connections between neurons in the same layer [20]. A single hidden layer is typically included in NNAR models to reduce overfitting and make training and interpretation easier. This model is trained using "neuralnet" package in R. Based on the data it will optimizes the model parameters for better predicting by automatically choosing the sizes of the lag and hidden layer.

Proposed Methodology

The idea of integrating forecasts was initially proposed by Bates and Granger [22], who added weights to each forecasting technique. By integrating the predicted values from various models, they developed a linear hybrid model. A hybrid model's output is a linear combination of the results that each model predicted, with the proper weights for each model determined by different kinds of techniques. Simple summing of predicted values, proportional MSE, and proportional forecast squared errors are a few ways to compute these weights [23, 26]. The time series models were recently merged by Najafabadipour et al. [27] utilizing particular weights that were obtained from the least squares approach. Existing univariate rainfall forecasting methods have various drawbacks, as was previously mentioned. Whenever rainfall estimates are created with these techniques, such errors may compromise the reliability of distribution systems. We introduce a hybrid forecasting model designed especially for Northeast India to increase the precision of rainfall predictions. The parallel hybrid structure's fundamental structure is as follows:

$$f_{combined,t} = \varphi\left(w_1\widetilde{f_{1,t}}, w_2\widetilde{f_{2,t}}, \ldots . w_n\widetilde{f_{n,t}}\right)\ t = 1,2, \ldots .T \tag{7}$$

Here, The entire combination variable is denoted by φ, and the adjusted predicted value of every single model at time t is indicated by $w_i\widetilde{f_{i,t}}$ (i = 1, 2, …, n) [11,32].

Each predicted value is multiplied by its appropriate weight to determine the final forecast, which is obtained after applying the original data to each individual model. The parallel hybridization method, which combines a number of linear models, is used in this work to merge several forecasting models. The weighted forecasts from each separate model are added together to produce the final forecasts.

Holt-Winters Additive (HWA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) are two univariate models that we unified in this study. Both the HWA and SARIMA models were used to generate forecasts for rainfall data in five different regions, and the precision of each model was evaluated. SARIMA outperformed HWA in most cases.

Thus, we merged the models as SARIMA-HWA or, in certain situations, as HWA-SARIMA. Eq. 8 yields the combined prediction model, and Eq. 9 computes the forecasting error. The variance-covariance matrix approach is used to calculate the weights. Figure 3 represents the flow chart of the proposed methodology. The following equation is then used to combine the results of the two prediction models:

$$Y_t = \textstyle\sum_{i=1}^{2} w_i f_{it}\ \ (t = 1,2, \ldots .n) \quad (w_1 + w_2 = 1)$$

$$= w_1 * forecasts\ from\ first\ forecast\ model + w_2 * forecasts\ from\ second\ forecast\ model \tag{8}$$

$$E_{combined\ t} = X_t - Y_t$$

$$= \sum_{i=1}^{n} w_i X_t - (w_1 f_{1,t} + w_2 f_{2,t})$$

$$= w_1(X_t - w_1 f_{1,t}) + w_2(X_t - w_2 f_{2,t}) \tag{9}$$

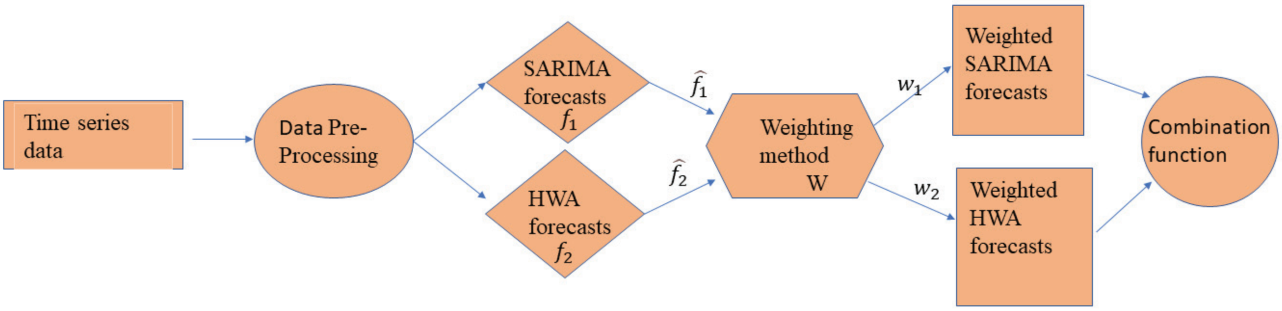


Figure 3. Flow chart of the Proposed Methodology.

Here w_1 and w_2 are the weights of first model's forecast, and weights of second model's forecast, respectively. The predicted value for the ith model is f_{it}. The following is how the weights are determined:

$$w_1 = \frac{var(e_2)-cov(e_1,e_2)}{var(e_1)+var(e_2)-2cov(e_1,e_2)} \tag{10}$$

$$w_2 = \frac{var(e_1)-cov(e_1,e_2)}{var(e_1)+var(e_2)-2cov(e_1,e_2)} \tag{11}$$

SARIMA and HWA error values are e_1 and e_2, respectively. We calculate $var(e_1)$, $var(e_2)$, and $cov(e_1, e_2)$, to determine w_1 and w_2. Then, we derive w_1 and w_2 using Eq.10 and 11.

Performance Statistics for Model Evaluation

This research examines the efficiency of the fitted models in terms of the statistical indicators, like RMSE, MAE, MSE, sMAPE, NSE and Correlation Coefficient (R)

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} \tag{12}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|X_i - Y_i| \tag{13}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2 \tag{14}$$

$$sMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|X_i - Y_i|}{|X_i|+|Y_i|/2} * 100 \tag{15}$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}(X_i-Y_i)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right] \tag{16}$$

$$Correlation\ Coefficient(R) = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\ \sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \tag{17}$$

Here, X_i and Y_i stand for real and predicted values, respectively. The set of data obtained is indicated by n. Prediction accuracy is indicated by RMSE, MSE values close to zero, and R and NSE values around one. When using real time series data to analyze and validate a hydrological model, NSE is the most widely utilized efficiency criteria [33].

RESULTS AND DISCUSSION

As discussed earlier, the normalized rainfall data has been used to model the proposed methodology. This methodology aims to combine univariate models to accurately forecast rainfall. The univariate models used for forecasting rainfall in Northeast India include SARIMA, Holt-Winters Additive (HWA), Feedforward Neural Networks (FFNN), Exponential Smoothing State Space Model (ETS), and the Holt Model When it comes to rainfall forecasting, each of these models is important. The performance of each forecasting model determines whether it should be hybridized.

SARIMA Model

The steps involved in forecasting rainfall precipitation using the SARIMA model include checking for stationarity, model identification, and selection. We conducted forecasts for each region of Northeast India following these steps. To determine the stationarity of each region's rainfall data, we used the Augmented Dickey-Fuller (ADF) test with a p-value threshold of 0.05 and a lag order of 6. All regions of Northeast India had p-values less than 0.05, indicating that the normalized time series data for each region is stationary and the detailed results are presented in Table 2. The normalized time series data of each region of NE India is stationary. The SARIMA model was discovered and calculated using R software. The p, q and P, Q of the model is derived using the maximum likelihood approach. The Akaike information criterion (AIC) utilized to evaluate the model. The model with the lowest AIC, RMSE, MAE, MSE, and sMAPE is among the models chosen as the best-fitted model.

From Table 3, the ARIMA (1,0,1)(2,1,1)12 model was selected as the best fit for the data in the WBS region of

Northeast India, with a minimum AIC value of -367.4906. This model has been used to forecast rainfall precipitation for the next 60 months, accounting for seasonal variance. Furthermore, ARIMA (1,0,1) (1,1,1)12 was determined as the most suitable model for rainfall forecasting in AP, with an AIC = -320.87. The ARIMA (1,0,1) (1,0,1)12 model was determined for AM, with An AIC = -415.86, and ARIMA (1,0,0) (1,1,0) 12 was chosen for GWB, with an AIC = -306.16. For NMMT, the ARIMA (1,0,1) (1,0,1) 12 model was effective, with an AIC value of -300.55. In Northeast India, these models were shown to be the most effective for predicting rainfall.

Table 4 presents the prediction accuracy of the selected SARIMA models for each region. The results of the SARIMA model for the WBS region are as follows: RMSE = 0.0855, MAE = 0.0539, MSE = 0.0073, and sMAPE = 0.4611. The true and projected values had a 0.9322 correlation, with a hydrological value of 0.8686 for the model. In the AP region, the selected model provides: RMSE = 0.1248, MAE = 0.0879, MSE = 0.0155, sMAPE = 0.6727, a correlation = 0.8327, and a hydrological model value of 0.6894. For AM, the SARIMA model gives RMSE = 0.1052, MAE = 0.0748, MSE = 0.0110, sMAPE = 0.5606, a correlation = 0.8784, and a hydrological model value of 0.7655. In GWB, the selected model results : RMSE = 0.1257, MAE = 0.0758, MSE = 0.0158, sMAPE = 0.7533, a correlation = 0.8187, and a hydrological model value of 0.6549. For NMMT, the SARIMA model yields RMSE = 0.0992, MAE = 0.0737, MSE = 0.0098, sMAPE = 0.6144, a correlation = 0.7753, and a hydrological model value of 0.9135.

The forecasting of monthly average rainfall for all regions of Northeast India was conducted after creating the appropriate time series models. The monthly data from January 1990 to June 2009 were used for model validation, while the data from July 2009 to December 2017 were used for testing the forecasts. Using the chosen SARIMA models, Figure 4 shows the rainfall forecasting values for all of Northeast India's areas from July 2009 to December 2022. The blue line represents the predicted values, while the black line represents the actual values.

HWA Model

The Holt-Winters Additive (HWA) model was described in the technique section. The steps required in making predictions using the HWA model and analyzing the outcomes will be covered in detail in this section. Given that this model is a parametric approach, its parameters must be assigned weights. A higher weight means that the model gives the most recent observed data a higher priority when making predictions.

Using the best-fitting HWA models for all of Northeast India's areas, Figure 5 displays the rainfall predicted values from July 2009 to December 2022. The blue line represents the predicted values, while the black line represents the actual values. It is evident that the HWA model predictions in Northeast India follow the same seasonality as the historical data. To evaluate the performance of these prediction models, statistical measures such as RMSE, MAE, MSE, sMAPE, correlation coefficient R, and Nash-Sutcliffe Efficiency (NSE) are summarized in Table 4.

In the WBS region, the HWA model produces an RMSE value of 0.0983, an MAE value of 0.0792, an MSE value of 0.0096, an sMAPE value of 0.9981, a correlation coefficient of 0.9360, and a hydrological model value of 0.8262. For AP, the selected model yields an RMSE value of 0.1224, an MAE value of 0.0845, an MSE value of 0.0149, an sMAPE value

Table 2. ADF test for monthly rainfall precipitation data

Rainfall Data	Test	Calculated Value	Lag order	P value	Comment
WBS	ADF	-17.024	6	0.01	Stationary
AP	ADF	-12.88	6	0.01	Stationary
AM	ADF	-15.011	6	0.01	Stationary
GWB	ADF	-15.248	6	0.01	Stationary
NMMT	ADF	-13.314	6	0.01	Stationary

Table 3. Selected SARIMA model for monthly rainfall of each region of NE India

Region	SARIMA model	AIC	RMSE
WBS	ARIMA (1,0,1) (2,1,1) 12	-367.4906	0.0855
AP	ARIMA (1,0,1) (1,1,1) 12	-320.87	0.1248
AM	ARIMA (1,0,1) (1,0,1) 12	- 415.86	0.1052
GWB	ARIMA (1,0,0) (1,1,0) 12	-306.16.	0.1257
NMMT	ARIMA (1,0,1) (1,0,1)12	-300.55	0.0992

Table 4. Comparison of error measures for Best fitted models of monthly rainfall data series

Rainfall Data	RMSE	MAE	MSE	sMAPE	NSE	R
West Bengal and Sikkim (WBS)						
HWA-SARIMA	0.0798	0.0453	0.0063	0.3939	0.8855	0.9414
HWA	0.0983	0.0792	0.0096	0.9981	0.8262	0.9360
SARIMA	0.0855	0.0539	0.0073	0.4611	0.8686	0.9322
FFNN	0.0944	0.0688	0.0089	0.6474	0.8397	0.9196
ETS	0.0861	0.0613	0.0074	0.8972	0.8666	0.9370
Holt	0.5373	0.4461	0.2887	1.6844	-4.4490	0.0240
Arunachal Pradesh (AP)						
HWA-SARIMA	0.1206	0.0791	0.0145	0.3765	0.6986	0.8399
SARIMA	0.1248	0.0879	0.0155	0.6727	0.6894	0.8327
HWA	0.1224	0.0845	0.0149	0.4539	0.6770	0.8329
FFNN	0.1710	0.1239	0.0292	0.6104	0.3825	0.6481
ETS	0.1628	0.1060	0.0265	0.5230	0.4402	0.6866
Holt Model	0.2263	0.1909	0.0512	0.8546	-0.0817	0.1213
Assam and Meghalaya (AM)						
SARIMA-HWA	0.0961	0.0644	0.0092	0.4284	0.8046	0.8977
SARIMA	0.1052	0.0748	0.0110	0.5606	0.7655	0.8784
HWA	0.0983	0.0732	0.0096	0.8173	0.7955	0.8978
FFNN	0.1058	0.0789	0.0112	0.6411	0.7627	0.8688
ETS	0.1017	0.0746	0.0103	0.5441	0.7843	0.8934
Holt Model	0.2281	0.0520	0.1959	1.0569	-0.0139	0.0452
Gangetic WestBengal (GWB)						
SARIMA-HWA	0.0169	0.0572	0.0114	0.6630	0.7503	0.8681
SARIMA	0.1257	0.0758	0.0158	0.7533	0.6549	0.8187
HWA	0.1122	0.1933	0.0126	1.1094	0.7247	0.8583
FFNN	0.1299	0.0907	0.0168	0.7851	0.6314	0.8075
ETS	0.1186	0.0798	0.0140	0.7479	0.6164	0.8469
Holt	0.3635	0.3330	0.1321	1.2120	-2.6012	-0.0575
Nagaland, Manipur, Mizoram, and Tripura (NMMT)						
HWA-SARIMA	0.0834	0.0496	0.0069	0.4926	0.9382	0.8793
HWA	0.1137	0.0864	0.0129	0.8889	0.9186	0.8287
SARIMA	0.0992	0.0737	0.0098	0.6144	0.9135	0.7753
FFNN	0.1420	0.0956	0.0201	0.6802	0.6493	0.8156
ETS	0.1016	0.0757	0.0103	0.7368	0.8203	0.9094
Holt	0.4333	0.3655	0.1878	1.0205	-2.4605	-0.0839

of 0.4539, a correlation coefficient of 0.8329, and an NSE value of 0.6770. In AM, the HWA model gives an RMSE value of 0.0983, an MAE value of 0.0732, an MSE value of 0.0096, an sMAPE value of 0.8173, a correlation coefficient of 0.8978, and a hydrological model value of 0.7955. In GWB, the HWA model produces an RMSE value of 0.1122, an MAE value of 0.1933, an MSE value of 0.0126, an sMAPE value of 1.1094, a correlation coefficient of 0.8583, and an NSE value of 0.7247. For the NMMT region, the HWA model results in an RMSE value of 0.1137, an MAE value of 0.0864, an MSE value of 0.0129, an sMAPE value of 0.8889, a correlation coefficient of 0.8287, and an NSE value of 0.9186.

FFNN Model

In the methodology section, the Feedforward Neural Network (FFNN) methodology was explained. This part details the processing steps for prediction using the FFNN

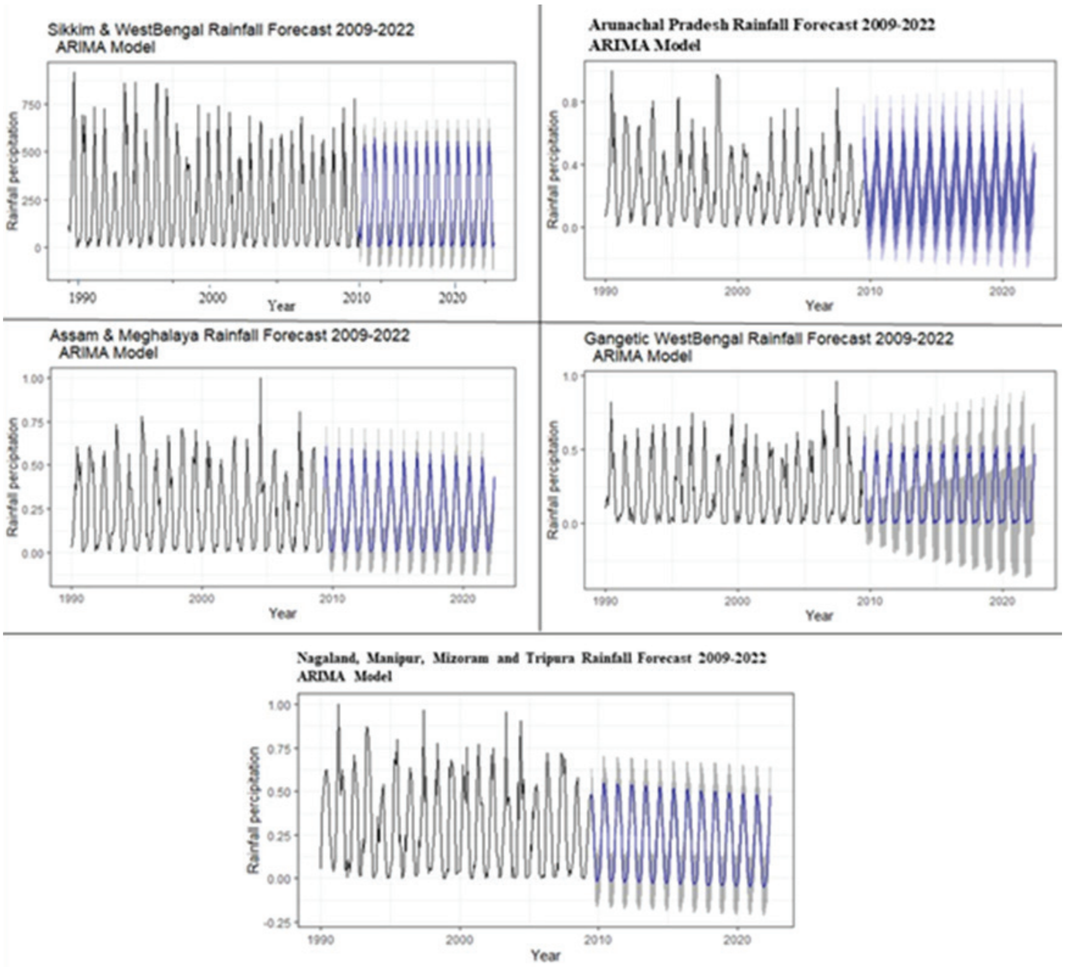


Figure 4. Forecast from best fitted SARIMA model for rainfall data series of Northeast India.

model, as well as the analysis of the results produced [31]. The artificial neural network model serves as an alternative to the SARIMA model for time series forecasting, capturing non-linear patterns in the data. In this study, we used a single hidden layer FFNN with one output node. Figure 6 displays the rainfall forecasting values for all regions of Northeast India. From Figure 6, it is evident that the FFNN model's predictions for each region follow the same seasonality as the historical data. The RMSE, MAE, MSE, sMAPE, correlation coefficient R, and Nash-Sutcliffe Efficiency (NSE) values for each region are summarized in Table 4.

The ETS and Holt models were forecasted using R software. Performance statistics for each model are shown in Table 4. The Holt model struggled with seasonal volatility when applied to this rainfall time series data. The HWA model outperformed the FFNN model in terms of error metrics, making it the best option. In comparison to the other models in the table, the SARIMA and HWA models exhibit superior performance. Although FFNN is also suitable for rainfall forecasting, it is less effective than SARIMA and HWA. Therefore, the ETS andHolt models were disregarded. We will proceed with combining the SARIMA and HWA models.

Proposed Methodology

To combine the selected forecasting models, we used the variance-covariance matrix method to calculate the weights for each model. The weights for each forecasting model were determined using Eq. 10 and Eq. 11. Subsequently, forecasts and error values for the proposed model were computed using Eq. 8 and Eq. 9. Figure 7 shows the forecasted values of all regions of NE India using the proposed method. Blue line denotes the predicted values and the red line denotes the actual values.

The HWA- SARIMA model was chosen as effective model for WBS region. Since, RMSE value of 0.0798, MAE of 0.0453, MSE of 0.0063, sMAPE of 0.3939, a correlation between actual and predicted values of 0.9414, and an NSE value of 0.8855. For AP, the HWA-SARIMA model produced an RMSE of 0.1206, MAE of 0.0791, MSE of 0.0145, sMAPE of 0.3765, NSE of 0.6986, and a correlation value of 0.8399. In AM, the SARIMA-HWA model resulted in an RMSE of 0.0961, MAE of 0.0644, MSE of 0.0092, sMAPE of

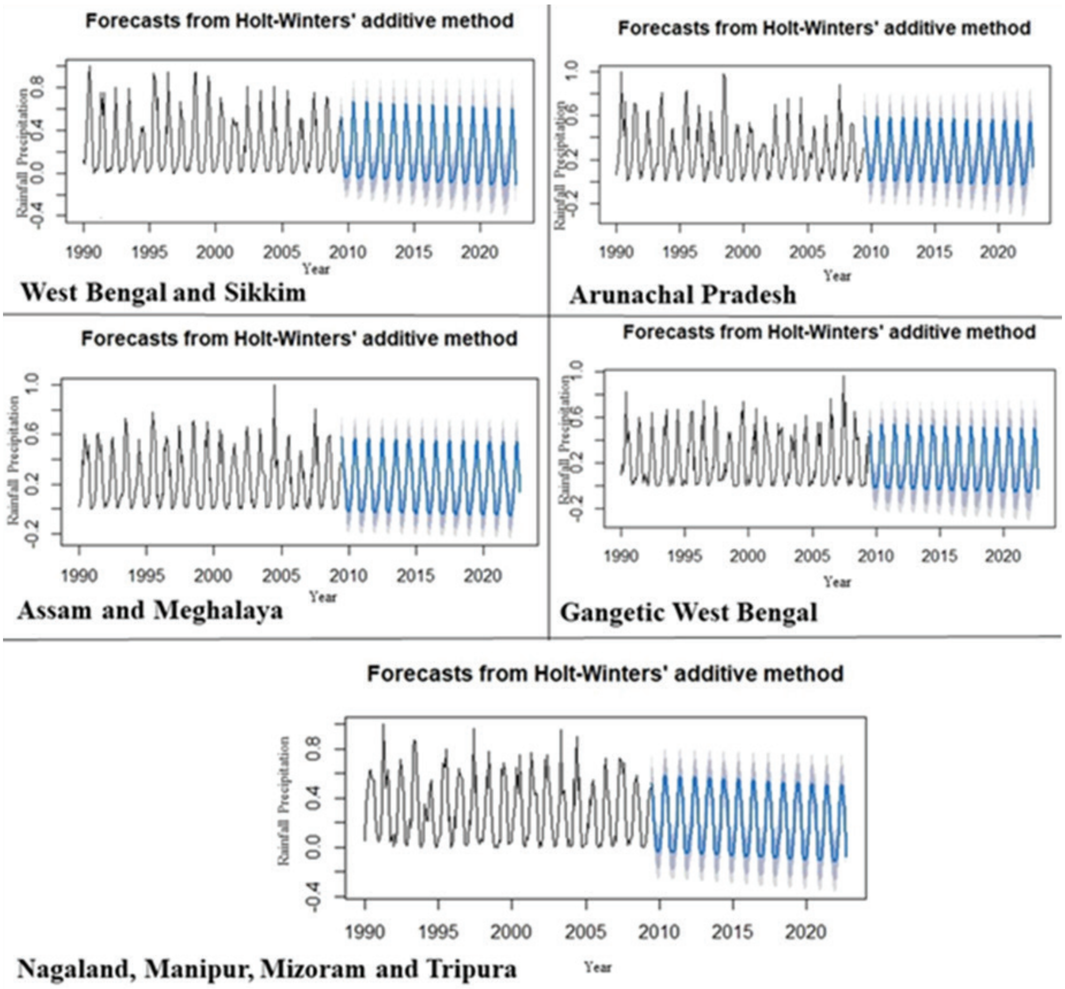


Figure 5. Forecast from best fitted Holt Winter's Additive Model for rainfall data series of Northeast India.

0.4284, NSE of 0.8046, and a correlation value of 0.8977. For GWB, the SARIMA-HWA model gave an RMSE of 0.0169, MAE of 0.0572, MSE of 0.0114, sMAPE of 0.6630, NSE of 0.7503, and a correlation value of 0.8681. In NMMT, the HWA-SARIMA model achieved an RMSE of 0.0834, MAE of 0.0496, MSE of 0.0069, sMAPE of 0.4926, NSE of 0.9382, and a correlation value of 0.8793.

These results indicate that the proposed methodology is more accurate than the existing forecasting models. The proposed approach is highly effective for rainfall forecasting. Table 4 demonstrates that the combined model outperforms the individual models and benchmark methods such as the ETS and Holt models, confirming that combining the best individual forecasting models yields superior performance.

Using monthly rainfall precipitation data from 1990 to 2017, we employed R software to forecast rainfall from 2018 to 2022. We proposed a hybrid statistical model for each region of Northeast India and compared it with individual models, including the ANN model. The results indicate that the proposed hybrid model outperformed the individual forecasting models. Consequently, we used this model to forecast rainfall precipitation in Northeast India for the next 60 months (see Figure 7).

Rainfall patterns in Northeast India are becoming more unpredictable and intense. It is a sign of larger shifts brought on by global climate change. Recent research indicates that climate change is causing extreme rainfall and other weather events. From a 28-year study period of rainfall data, the state of Assam receives most of its rainfall in July, with a 21% variation when looking at historical data. Meghalaya receives 71% of its rainfall during the monsoon period. This area has decreasing trend in rainfall precipitation. Arunachal Pradesh had the rainfall in July, with a 31 percent variation. The state of Nagaland receives most of the rains in the month of July with 30% variation in monsoon rainfall when looking at historical data. Twenty-five percent of the total rainfall has varied [34]. Manipur has a consistent pattern. Mizoram has shown a consistent trend for overall rainfall. But while analysing the monthly trend, it shows little variation with a significant decrease in the number of rainy days. Tripura has a decreasing trend,

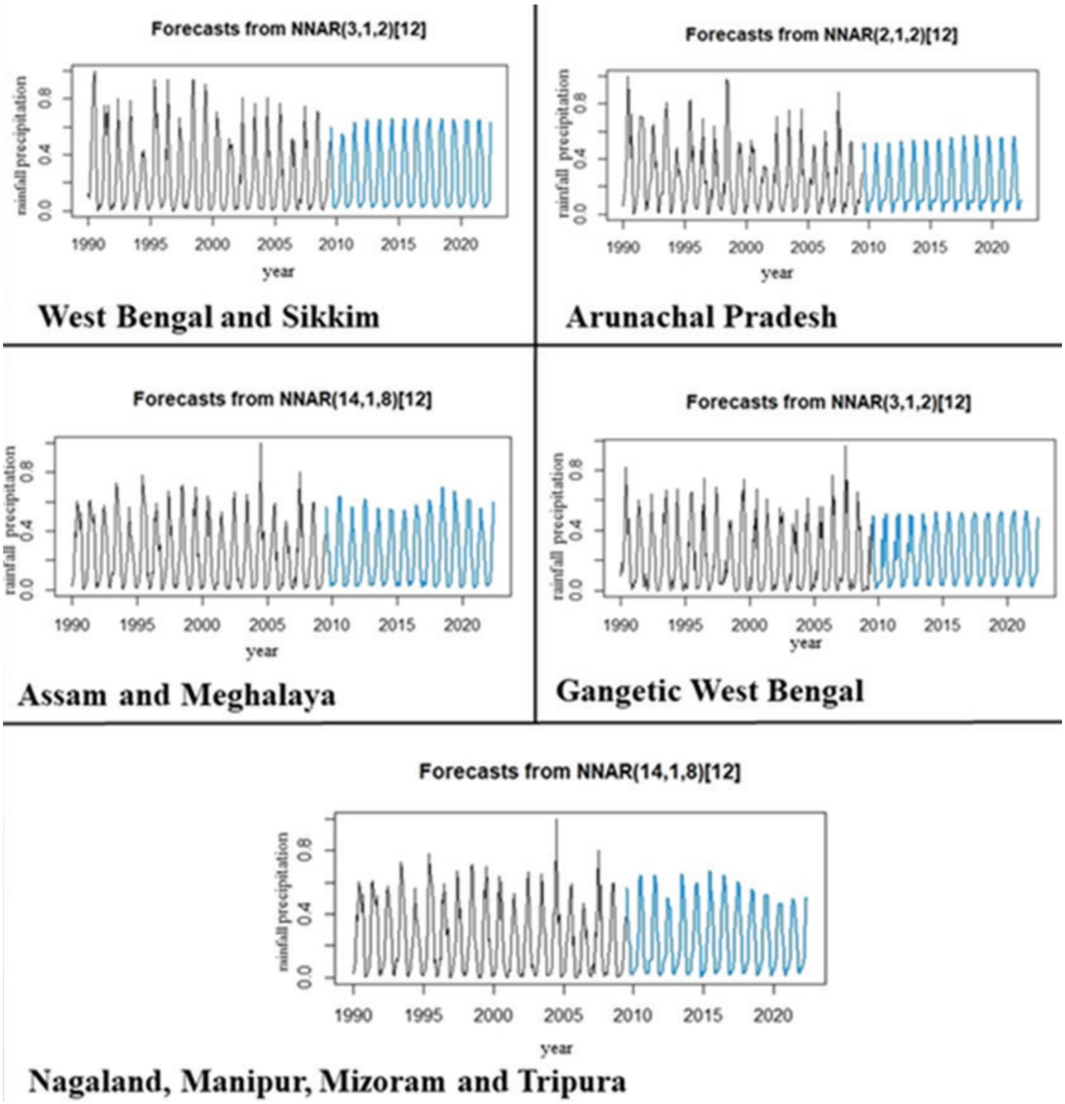


Figure 6. Forecast from Feed Forward Neural Network forecasting model rainfall data series of Northeast India.

even though it is not meaningful. Sikkim and West Bengal Monsoon rainfall increased slightly with a 23% variation. The month of July received the most rain. Gangetic West Bengal receives its highest rainfall in the month of June [35].

The choice of the preferred model based on univariate models may vary with different data sets. Therefore, it is crucial to evaluate all time series models for any location and hydrological factor to select the most suitable model for our needs. Our results indicate that statistical models are the most successful methods for rainfall prediction. They are very successful in identifying changes in rainfall time series components.

Applications of proposed model:

Flood Control: our proposed model is greatly beneficial in flood risk assessment and management. Experts can lessen the impact of floods on communities that are already at risk by implementing early warning systems. They can taking preventative action when periods of heavy rainfall are anticipated.

Drought Preparedness: On the other hand, precise forecasts of decreased precipitation can help with drought resilience.

Infrastructure and Urban Planning: Designing and building sturdy structures requires an understanding of shifting rainfall patterns.

Policy and Decision-Making: The results highlight how crucial it is to incorporate climate change issues into national and regional policies. The model's insights can be used by policymakers to create more comprehensive and proactive plans to tackle the long-term problems caused by climate change, especially in regions that are vulnerable to drought and flooding.

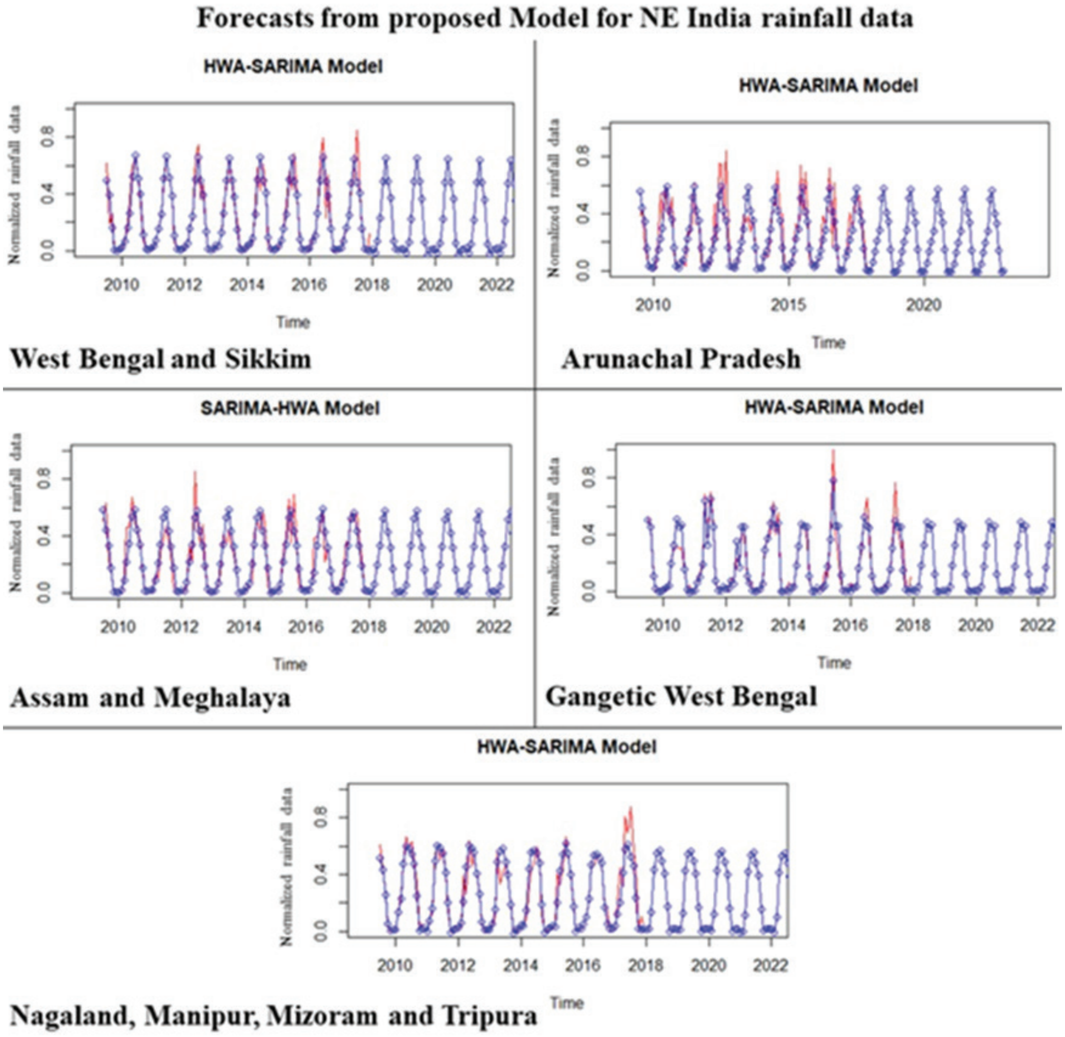


Figure 7. Forecast from combined statistical forecasting model rainfall data series of Northeast India.

In conclusion, the urgent necessity for adaptation measures is highlighted by the evidence of climate change represented in our rainfall projections. Our work supports sustainable development in the region by contributing to larger efforts to safeguard against the growing dangers connected with climate change by offering more accurate projections.

CONCLUSION

The fundamental purpose of this research is the design of a cohesive forecasting model. In order to enhance the accuracy of precipitation forecasts, our proposed model integrates the benefits of both the HWA and SARIMA frameworks. In this time series data, linear trends captured by the SARIMA model and the HWA model is proficient in detecting seasonal fluctuations. Finally, we have developed a combination method that optimizes overall forecasting efficacy by combining these models.

For rainfall forecasting, earlier studies, like [7], only used univariate models like HWA and ETS. But These models are not suitable for time series data that have seasonal variability, like Northeast India. According to our results, the HWA performed poorly in areas such as NMMT, with sMAPE values considerably greater than those derived from the suggested hybrid model.

Our suggested model is appropriate for application in the meteorological field due to its high accuracy. It makes easier to analyze and predict a range of hydrological data, such as groundwater levels and rainfall precipitation. Our model's accuracy and resilience make it appropriate for a variety of forecasting scenarios and useful in a wide range of areas. Hybrid model performs better than individual models and has applications in resource management, agricultural planning, and catastrophe preparedness. Its use could help reduce the hazards associated with extreme weather events like droughts and floods brought on by climate change throughout different parts of the world.

Future studies may investigate applying the parallel hybrid model to different places with diverse climatic circumstances to assess its wider application. Furthermore, introducing other factors such as temperature, wind speed, and humidity into the forecasting models could boost prediction accuracy by capturing more complicated interactions within the climate system.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Terzioğlu ZÖ, Kankal M, Yüksek Ö, Nemli MÖ, Akçay F. Analysis of the precipitation intensity values of various durations in Trabzon Province of Turkey by Şen's innovative trend method. Sigma J Eng Nat Sci 2019;37:241-250.

[2] Borah P, Hazarika S, Prakash A. Assessing the state of homogeneity, variability and trends in the rainfall time series from 1969 to 2017 and its significance for groundwater in north-east India. Nat Hazards 2022;111:585-617. [CrossRef]

[3] Nyatuame M, Agodzo SK. Stochastic ARIMA model for annual rainfall and maximum temperature forecasting over Tordzie watershed in Ghana. J Water Land Dev 2018;37:127-140. [CrossRef]

[4] Soltani S, Modarres R, Eslamian SS. The use of time series modeling for the determination of rainfall climates of Iran. Int J Climatol 2007;27:819-829. [CrossRef]

[5] Alonso Brito GR, Rivero Villaverde A, Lau Quan A, Ruíz Pérez ME. Comparison between SARIMA and Holt–Winters models for forecasting monthly streamflow in the western region of Cuba. SN Appl Sci 2021;3:671. [CrossRef]

[6] Lama A, Singh KN, Singh H, Shekhawat R, Mishra P, Gurung B. Forecasting monthly rainfall of Sub-Himalayan region of India using parametric and non-parametric modelling approaches. Model Earth Syst Environ 2022;8:837-845. [CrossRef]

[7] Puah YJ, Huang YF, Chua KC, Lee TS. River catchment rainfall series analysis using additive Holt–Winters method. J Earth Syst Sci 2016;125. [CrossRef]

[8] Mehdizadeh S, Fathian F, Safari MJS, Adamowski JF. Comparative assessment of time series and artificial intelligence models to estimate monthly streamflow: A local and external data analysis approach. J Hydrol (Amst) 2019;579. [CrossRef]

[9] Karthika D, Karthikeyan K. Estimation of electrical energy consumption in Tamil Nadu using univariate time-series analysis. Ann Optim Theory Pract 2021;4.

[10] Karthika D, Karthikeyan K. Analysis of mathematical models for rainfall prediction using seasonal rainfall data: A case study for Tamil Nadu, India. In: 1st International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT 2022); 2022. [CrossRef]

[11] Karthika D, Karthikeyan K. Performance of combined forecasting model for monthly rainfall precipitation. Adv Appl Stat 2023;90. [CrossRef]

[12] Narasimha Murthy KV, Saravana R, Vijaya Kumar K. Modeling and forecasting rainfall patterns of southwest monsoons in North-East India as a SARIMA process. Meteorol Atmos Phys 2017;130:99-106. [CrossRef]

[13] Eni D, Adeyeye FJ. Seasonal ARIMA modeling and forecasting of rainfall in Warri Town, Nigeria. J Geosci Environ Prot 2015;3:91-98. [CrossRef]

[14] Dastorani M, Mirzavand M, Dastorani MT, Sadatinejad SJ. Comparative study among different time series models applied to monthly rainfall forecasting in semi-arid climate condition. Nat Hazards 2016;81:1811-1827. [CrossRef]

[15] Mirzavand M, Ghazavi R. A stochastic modelling technique for groundwater level forecasting in an arid environment using time series methods. Water Resour Manag 2015;29:1315-1328. [CrossRef]

[16] Papacharalampous G, Tyralis H, Koutsoyiannis D. Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. Stoch Environ Res Risk Assess 2019;33. [CrossRef]

[17] AlSubih M, Kumari M, Mallick J, Ramakrishnan R, Islam S, Singh CK. Time series trend analysis of rainfall in last five decades and its quantification in Aseer Region of Saudi Arabia. Arab J Geosci 2021;14. [CrossRef]

[18] Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. Int J Forecast 2004;20. [CrossRef]

[19] Luk KC, Ball JE, Sharma A. An application of artificial neural networks for rainfall forecasting. Math Comput Model 2001;33. [CrossRef]

[20] Chattopadhyay S. Feed forward artificial neural network model to predict the average summer-monsoon rainfall in India. Acta Geophys 2007;55. [CrossRef]

[21] Guhathakurta P. Long lead monsoon rainfall prediction for meteorological sub-divisions of India using deterministic artificial neural network model. Meteorol Atmos Phys 2008;101. [CrossRef]

[22] Bates JM, Granger CWJ. The combination of forecasts. J Oper Res Soc 1969;20:451–468. [CrossRef]

[23] Newbold P, Granger CWJ. Experience with forecasting univariate time series and the combination of forecasts. J R Stat Soc Ser A 1974;137. [CrossRef]

[24] Clemen RT. Combining forecasts: A review and annotated bibliography. Int J Forecast 1989;5. [CrossRef]

[25] Winkler RL, Makridakis S. The combination of forecasts. J R Stat Soc Ser A 1983;146:150. [CrossRef]

[26] Caiado J. Performance of combined double seasonal univariate time series models for forecasting water demand. J Hydrol Eng 2010;15. [CrossRef]

[27] Najafabadipour A, Kamali G, Nezamabadi-pour H. The innovative combination of time series analysis methods for the forecasting of groundwater fluctuations. Water Resour 2022;49. [CrossRef]

[28] Dikshit KR, Dikshit JK. North-East India: land, people and economy. Dordrecht: Springer Netherlands; 2014. [CrossRef]

[29] Ray S, Das SS, Mishra P, Al Khatib AMG. Time series SARIMA modelling and forecasting of monthly rainfall and temperature in the South Asian countries. Earth Syst Environ 2021;5. [CrossRef]

[30] Valipour M. Long-term runoff study using SARIMA and ARIMA models in the United States. Meteorol Appl 2015;22. [CrossRef]

[31] Winters PR. Forecasting sales by exponentially weighted moving averages. Manage Sci 1960;6. [CrossRef]

[32] Hajirahimi Z, Khashei M. Hybrid structures in time series modeling and forecasting: A review. Eng Appl Artif Intell 2019;86. [CrossRef]

[33] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models. Part I – A discussion of principles. J Hydrol (Amst) 1970;10. [CrossRef]

[34] Das S, Tomar CS, Saha D, Shaw SO, Singh C. Trends in rainfall patterns over North-East India during 1961–2010. Int J Earth Atmos Sci 2015;2.

[35] Ravindranath NH, Rao S, Sharma N, Nair M, Gopalakrishnan R, Rao AS, et al. Climate change vulnerability profiles for North East India. Curr Sci 2011;101.