Analytics, Modelling, and Optimization Journal (AMOJ) **2025, 1(1): 1-10**
https://www.longmanpublishers.com/

Summer Monsoon Rainfall Forecasting over Gangetic West Bengal, India Using Linear and Neural Regression and Intrinsic Mode Functions

Vidhya Sagar

Jadavpur University, Jadavpur, Kolkata, West Bengal 700032, India

Abstract: It is demonstrated that the rainfall data from the South West Monsoon in India's sixth meteorological subdivision, which includes Gangetic West Bengal, can be broken down into eight empirical time series, or intrinsic mode functions. As a result, the first empirical mode is identified as a nonlinear component of the data, whereas the following modes are identified as linear components. While the linear portion is logically described using the straightforward regression method, the nonlinear portion is modeled using the Neural Network-based Generalized Regression Neural Network model technique. As confirmed, the various intrinsic modes have a strong correlation with pertinent atmospheric characteristics, such as the sunspot cycle, El Nino, and the quasi-biennial oscillation. It has been found that the suggested model accounts for about 75% of the interannual variability (IAV) of the Gangetic West Bengal rainfall series. Independent analysis of the actual data confirms that the model is effective in statistically predicting the region's South West Monsoon rainfall. The actual rainfall for 2012 and 2013 was 93.19 cm and 115.20 cm, respectively, within one standard deviation of the mean rainfall, although the statistical predictions of SWM rainfall for GWB were 108.71 cm and 126.21 cm, respectively.
Keywords: South West Monsoon (SWM) rainfall; Intrinsic Mode Function (IMF); Generalized Regression Neural Network (GRNN); Quasi-biennial Oscillation (QBO); Inter annual variability (IAV).

1. Introduction

A significant portion of India's yearly rainfall, as well as the meteorological subdivision number 6 that covers the Gangetic West Bengal (GWB) region, is caused by the summer monsoon, also known as the South West Monsoon (SWM), which consists of the rainfalls of the months of June, July, August, and September. The features of SWM rainfall have a significant impact on agriculture and the economy. Scientists are primarily concerned with time series at various spatial and temporal scales [1], [2], and [3]. The Indian meteorological sub-divisions, including GWB, are depicted in Figure 1.

Actually, since ancient times, attempts have been made to comprehend the relationships between SWM rains and other meteorological and global phenomena. For instance, there are [4] connections between the worldwide Sea Surface Temperature (SST) data and the Indian monsoonal rainfall data. The alternative method only models historical data for a year ahead of time with a negligible error band in order to get a forecast that is unrelated to the phenomena.
Quasi-biennial Oscillation (QBO) [5], tidal forcing [6], El-Nino Southern Oscillation (ENSO) [7], [8], Sunspot Cycle [9], and intra-seasonal periodicities [10], [7] are some of the studies that use Fourier decomposition to elucidate the periods latent in the data.

The traditional approach to rainfall data analysis relies on a stationary random process with Gaussian characteristics [11], [3]. According to research on auto-correlation and power-spectral density, the SWM rainfall series of GWB are too weak to be modeled as linear time series [12]. It's unclear, nevertheless, what specific nonlinear model form should be applied.
On the other hand, determining coherent zones and the nature of the components can be accomplished through the breakdown of SWM rains upon application of primary components and comprehension of the component nature [11], [12], [13].
The causes should be inherent in the rainfall time series. In this regard, it is possible to refer to the works of [14], [15], wherein rainfall is appropriately empirically modeled using sufficient data series, even when the reasons are unknown.
Over the past ten years, it has been noted that the majority of the interannual variation (IAV) can be effectively explained by a nonlinear model with variable frequency harmonic terms. It has also been suggested that, if the basic data is not white noise, rainfall data can be broken down into hierarchical Integral Mode Functions (IMFs) as signals [16].

Iyenger and Raghukant have studied the method at the level of all of India [15]. There have been some recent studies on this strategy in other nations, such as Australia and the Caspian catchment area [19], [20], respectively.

However, the GWB (meteorological subdivision No. 6) region in the southern plain of the Indian state of West Bengal has not gotten much attention for the variability study and forecasting exercise. The region is well-known for its major contributions to agriculture, manufacturing, and civilization. This research examines a novel representation of the GWB's SWM rainfall in terms of narrow band IMF data. For modeling and forecasting, these time series are easier to understand than the original data.

The paper is organized as follows. We would start by talking about the Empirical Modes, also known as the IMFs of GWB. A forecasting technique would then be introduced. We would talk about the architecture of a Generalized Regression Neural Network (GRNN) in conjunction with Multiple Linear Regression analysis. Finally, a performance study of the model would be presented.

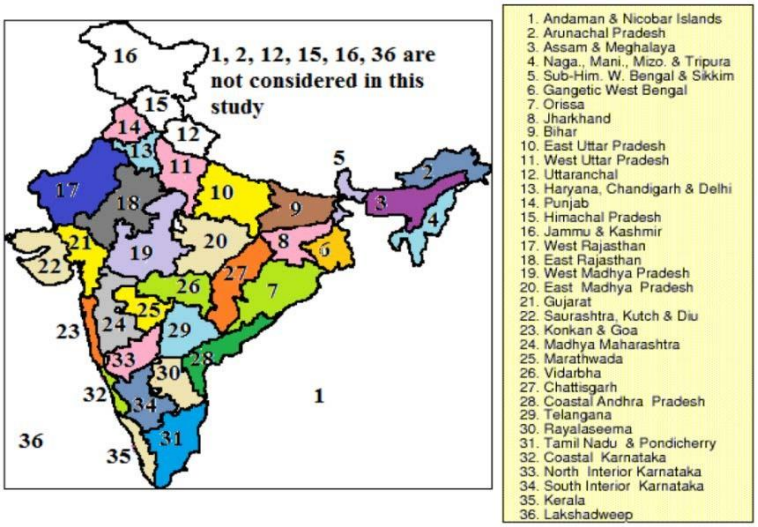

Figure 1. Meteorological Subdivisions including Gangetic West Bengal (GWB)

2. Rainfall data

The Indian Institute of Tropical Meteorology's (IITM) website, www.tropmet.res.in, is where the GWB's rainfall data are gathered. For a thorough analysis, the SWM rainfall data for the years 1871–2013—which is the total of the monthly rainfall values for June, July, August, and September—is chosen. For an initial notion, the GWB SWM data are shown in Figure 2. Table 1 displays some of the data's fundamental statistics, including the Climatic Normal (mR) and Climatic Deviation about the Normal (σR).

Table 1. SWM rainfall data (1871-2001)

Region	Area (Sq. Km.)	Mean (cm)(m_R)	Std. dev. (σ_R)	Skewness	Kurtosis
GWB	44300	106.93	10.2734	-0.4425	2.9845

* GWB: Gangetic West Bengal

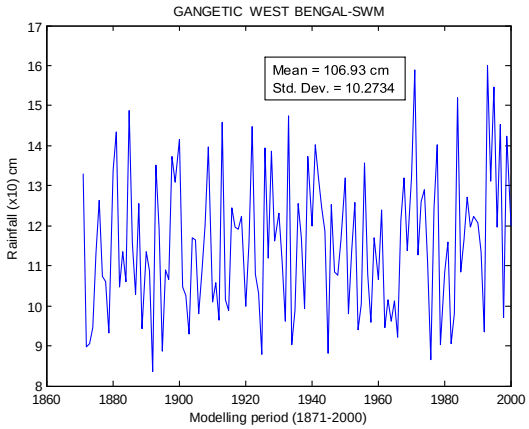

Figure 2. SWM rainfall of GWB for modeling period (1871-2000)

3. Intrinsic Mode Functions (IMFs)

The SWM rainfall time series is broken down into a number of empirical modes, or Intrinsic Mode Functions (IMFs) series. As Huang et al. [17] have shown, these IMF data identify prominent periods and amplitudes. According to their findings, these series have a decreasing order of importance correlation with SWM data but are uncorrelated with one another at zero latency.

Eight IMF series in all—IMF1, IMF2,..., IMF8—are extracted hierarchically until the sieved data shows no indications of oscillations. Every IMF has distinct center periods and is a narrow band process. The eight IMFs are long-term climate trends, center-line drifts, specific frequency or period, and long-term non-stationary features. Eight IMFs of SWM rainfall are shown for the time series in Figures 3–10. The latest two IMFs, IMF7 and IMF8, are seen to be consistently positive and to be a slowly fluctuating mode around the long-term average [17]. One way to conceptualize the process is as the typical or climatic component of the IAV of the monsoon rains that appears. To a better degree of accuracy, the total of all IMF series at a given point in time equals the original SWM data series.

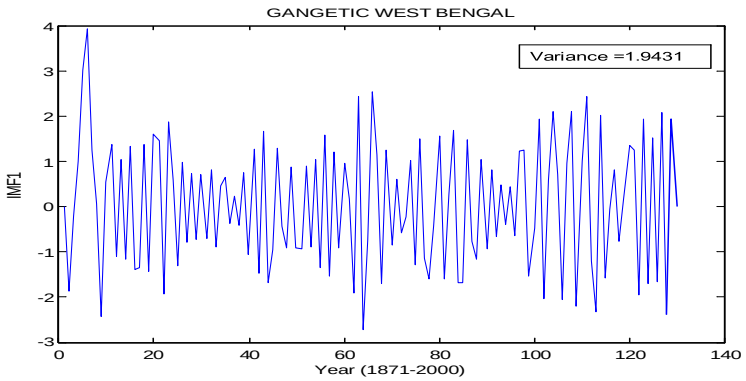

Figure 3. First Intrinsic Mode Function (IMF_1) of SWM rainfall of GWB

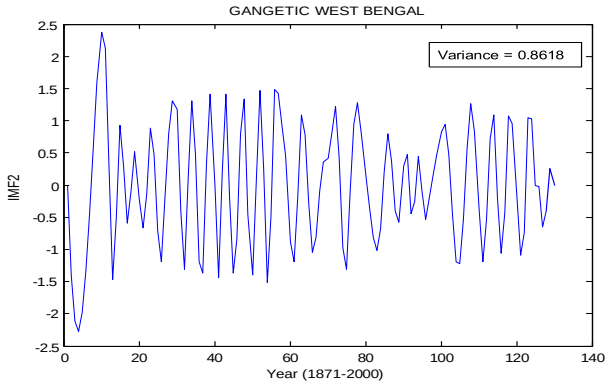

Figure 4. Second Intrinsic Mode Function (IMF_2) of SWM rainfall of GWB

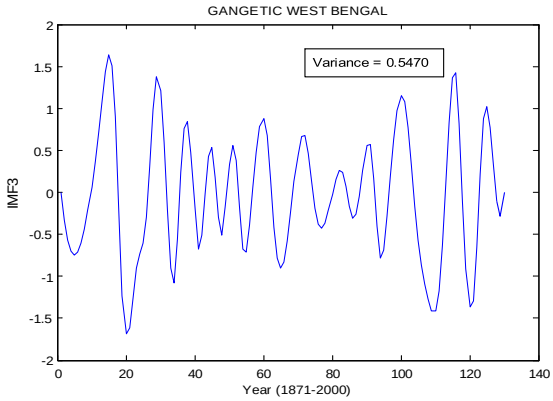

Figure 5. Third Intrinsic Mode Function (IMF_3) of SWM rainfall of GWB

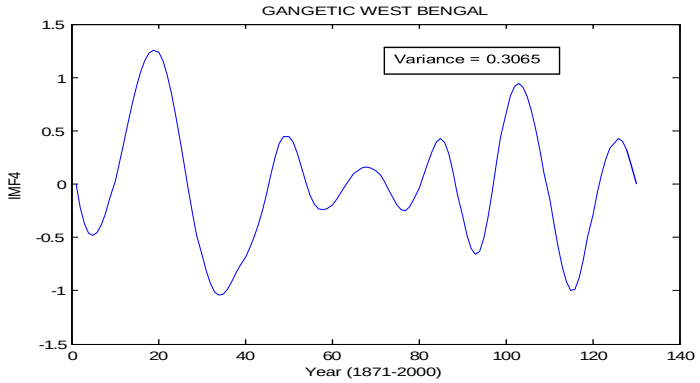

Figure 6. Fourth Intrinsic Mode Function (IMF_4) of SWM rainfall of GWB

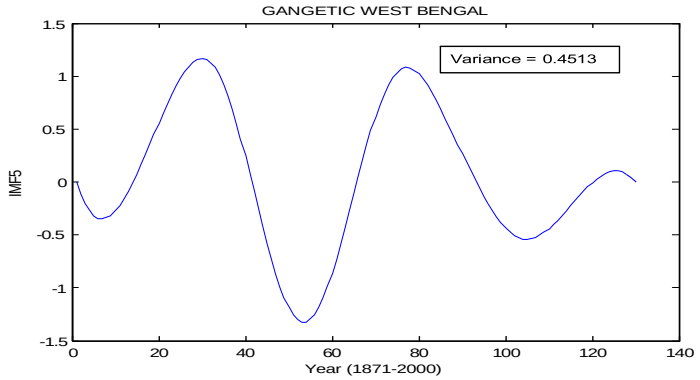

Figure 7. Fifth Intrinsic Mode Function (IMF_5) of SWM rainfall of GWB

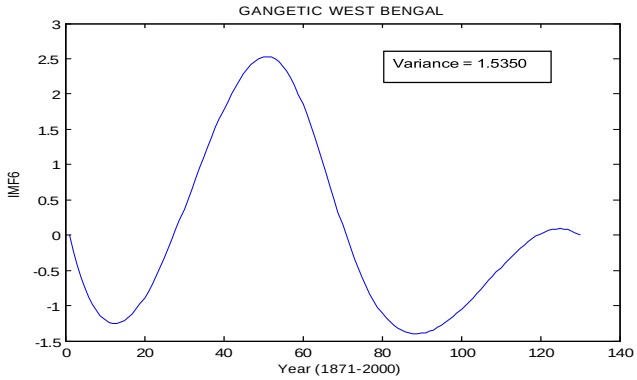

Figure 8. Sixth Intrinsic Mode Function (IMF_6) of SWM rainfall of GWB

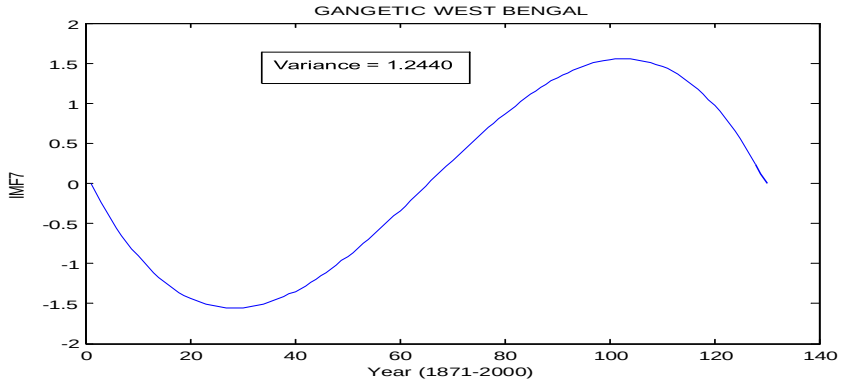

Figure 9. Seventh Intrinsic Mode Function (IMF_7) of SWM rainfall of GWB

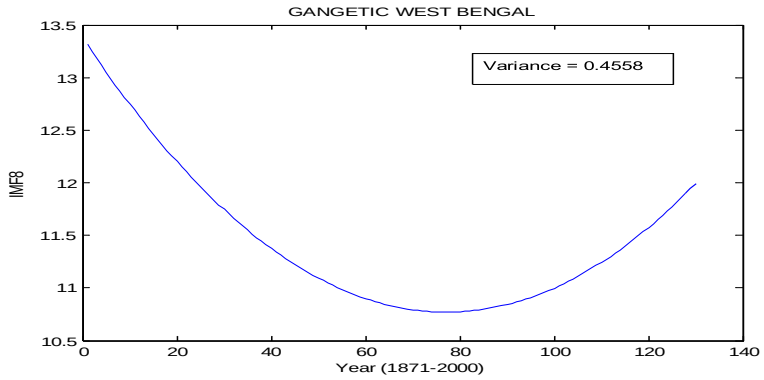

Figure10. Eighth Intrinsic Mode Function (IMF_8) of SWM rainfall of GWB

Table 2. Central period (T) of the IMF's in years and % variance contributed to IAV

Region	IMF1		IMF2		IMF3		IMF4		IMF5		IMF6		IMF7		IMF8	
	T	IAV%	T	IAV%	T	IAV%	T	IAV%	T	IAV%	T	IAV%	T	IAV%	T	IAV%
GWB	2.71	62.82	5.91	28.07	13.00	17.69	21.66	4.17	43.33	14.59	65	4.94	>100	16.94	Not detectable	6.20

*GWB: Gangetic West Bengal

The IMF series' contributions are calculated via temporal averaging, which projects each IMF's proportional share of the overall rainfall variability. A narrow band process is indicated by the slowly changing amplitudes and frequencies seen in all IMFs (Figures 3–10). By counting the zeros and extrema in an IMF series, one can determine the primary period of oscillation. The central period and each IMF's proportional contribution to IAV are shown in Table 2. With an average duration of 2.71 years, IMF1 is found to be a dominant mode, accounting for 62.82% of IAV. With a dominance term of 5.91 years, IMF2 is the second most prominent mode, accounting for 28.07% of IAV. As demonstrated at the All India level, these two modes are closely related to the El Niño–Southern Oscillation (ENSO) and Quasi-biennial Oscillation (QBO) phenomena, respectively [16]. Similarly, IMF3, which contributes 17.69% of IAV over a period of roughly 13 years, can be linked to the sunspot cycle, which lasts roughly 11 years, which is consistent with Bhalme and Jadav's findings [9]. According to Campbell et al. [6], the IMF4's central period, which lasts roughly 21 years, is closely related to the tidal forcing of roughly 19 years of the Indian monsoon quasi-cycle. An extended 43-year period may ensue, according to the IMF5 [17].

The presence of six quasi-cycle modes, as suggested by Narashima and Kailash [7] using wavelet analysis of Indian monsoon rainfall, is thus represented by IMF6, which is thus recognized as commonly associated to the extended mode of 65 years providing approximately 5% of IAV. There is a fragmented, slowly shifting, long-term deterministic shift of monsoon in the GWB that is causing changes in SWM rainfall, according to the IMF7 and IMF8 contributing over periods of 100 years and above. In India's monsoon rainfall, this portion goes unnoticed [15].

4. IMF statistics

The correlation matrix of the time series must be constructed in order to comprehend the statistical relationship between the IMFs and the data. Table 3 displays the correlation matrix (7x7) between the seven variable IMFs and the SWM GWB data. The correlation values between the data and a select few IMFs—IMF1, IMF2, and IMF3—are known to be statistically significant and incredibly valuable. Additionally, because the IMFs are uncorrelated, the modest influence of sample size is disregarded, and the sum of the IMFs' variances is expected to be almost equal to the total variance of the data. This allows us to pinpoint the intrinsic modes that make up the data's overall behavior.

5. Forecasting strategy

One way to think of forecasting is as a step-by-step extension of the data series over the upcoming years. Modeling in this sense refers to an equation that, with the least amount of error, closely resembles the data. Talyor's series expansion makes the exercise straightforward for simple functions having an analytic form; but, because rainfall data is so unpredictable, it is impossible to fit a basic linear function to the data series. Data decomposition into IMF series appears to be a different method for predicting a region's monsoonal rainfall using the decomposed IMF series, which is undoubtedly easier than utilizing the original data.

With the exception of a few sites [3], [12], and [16], SWM rainfall is a Gaussian random variable in the data examined thus far. Currently, the eight hierarchical IMFs have the ability to forecast SWM rainfall that incorporates IAV. It is anticipated that the first IMF will be higher and more arbitrary than the others since it contains the information with the highest frequency end. The histogram (Figure 11), which rules out the notion of Gaussianness, clearly illustrates the new bi-modality feature that appears for the first IMF. The demonstrated bi-modality rules out the presence of a linear auto-regressive representation and suggests significant non-linearity in the process dynamics. The Chi-square test with designed 14 intervals supports it. With observed Chi-square values of 25.67 at 13 degrees of freedom, the test indicates that IMF1 is non-Gaussian at the 5% level.

Table 3. Correlation matrix of IMFs of GWB

	Data	IMF_1	IMF_2	IMF_3	IMF_4	IMF_5	IMF_6	IMF_7
Data	1.0000	0.6728*	0.4665*	0.3578*	0.1159	-0.0411	0.0465	0.0426
IMF_1		1.0000	-0.1059	-0.1243	-0.0125	-0.0284	-0.0445	-0.0035
IMF_2			1.0000	0.1337	0.0271	0.0057	0.0143	+0.0134
IMF_3				1.0000	-0.0297	0.0080	0.0081	-0.0286
IMF_4					1.0000	-0.0032	-0.2570	-0.0031
IMF_5						1.0000	-0.4380*	-0.1249
IMF_6							1.0000	-0.4817
IMF_7								1.0000

*Significant at 5% level.

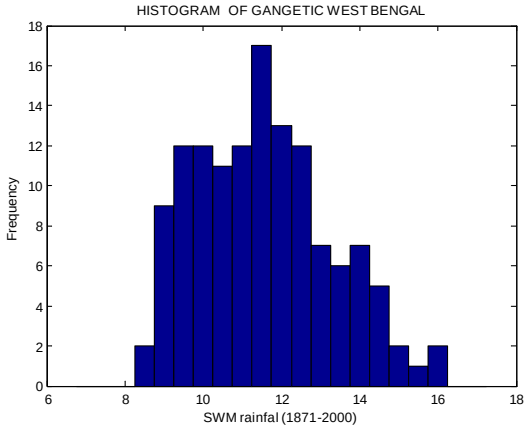

Figure 11. Histogram of IMF_1 showing bi-modality.

However, the usual run test of decadal variance about the median (counting to 7 and within the range 4-11 with 13 decades at 5% level) indicates that IMF1 is stationary. This section of IMF1 is regarded as non-linear. The remaining portion (Rj – IMF1) of the series, excluding the first IMF, is examined for stationarity and Gaussianness as described. The standard run test on decadal variance has confirmed the stationarity of this section, while the chi-square test has shown the Gaussianness.

Table 4. Regression coefficients of Equation (1)

Region	C_1	C_2	C_3	C_4	C_5	C_6	$\sigma_y(e)$	Correlation Coefficient (CC)
GWB	0.1196	1.5425	-1.6454	1.1161	0.4727	3.9072	0.5394	0.9144

* GWB: Gangetic West Bengal

Table 4 displays the correlation coefficient (CC) and standard deviation of the error $\sigma y(e)$ between the actual data and the model fitted. The regression coefficients are derived from the data series of 1871–2000, excluding the first 4 years because the regression equation contains up to yj-4 terms using the least squares method. The correlation is very significant, supporting the idea that yj is the linear component of GWB's SWM rainfall.

6. Generalized Regression Networks architecture connected to non linear IMF_1

The first IMF is a non-Gaussian, non-linear process that explains the majority of the IAV of monsoon rainfall. The Generalized Regression Neural Network (GRNN), a refined form of the Neural Network class of technology based on non-parametric regression and proposed by Specht [18], is used in the situation of unstructured complex problems.

6.1. Architecture of GRNN

Pattern neurons and summation neurons make up the two hidden layers of a GRNN model. Each GRNN pattern neuron performs the following computations: exp (-Dj2/2σ2), where Dj is the distance between training samples and σ is the smoothness parameter. The normal distribution is taken into account for each training sample. The pattern neuron's signals are weighted with the matching values of the training samples Yj before entering the denominator neuron. The signals entering the numerator have weights of one. Every point that GRNN predicts is influenced by each sample from the training data.

The author [18] demonstrated how GRNN can be used to model and extend function approximation, regression, prediction, and classification. Every training sample is supposed to represent the mean to a radial basis neuron. Figure 12 illustrates the use of a GRNN with a hidden layer following multiple experiments with the number of prior values of IMF1.

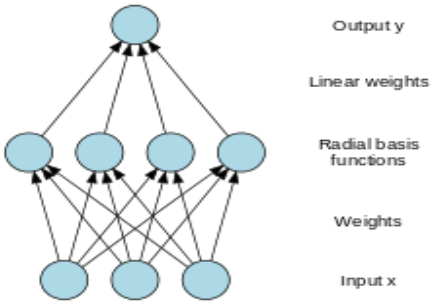

Figure 12. General Regression Neural Network with Radial Basis Functions

6.2. Results of IMF_1 with GRNN

The GRNN algorithms were computed using the MATLAB toolbox, and the training period was 1871–2000. The GRNN model can forecast IMF1 for the year (n +1) with the use of antecedent IMF1 data. Table 5 displays the correlation coefficient (CC) between the real IMF1 and the GRNN findings as well as the standard deviation σy(e) of the errors constructed on the training period data. The significant correlation (0.8062) between the simulated and actual IMF1 values indicates that GRNN is very adaptable in capturing the latent nonlinear structure. This method has the benefit of allowing for statistical characterization of the model's inaccuracy.

Table 5. Statistics of GRNN model for IMF: training period (1871–2000)

Region	$\sigma_y(e)$	Correlation Coefficient (CC)
GWB	2.2125	0.806153

* GWB: Gangetic West Bengal

7. Forecasting

It is possible to anticipate the rainfall value for the following year by extending the successful modeling of IMF1j and yj by one year. The preceding models are used to calculate n+1 first for yn+1 and subsequently for IMF1. The prognosis for Rn+1 is the result of adding the two values. Here, the forecast strategy's success is examined by taking into account the years 2001–2013, which were purposefully excluded from the modeling exercise. Table 6 shows the effectiveness of one-step-ahead predicting during the testing period (2001–2013) and the quality of modeling Rj over the training period (1875–2000).

Table 6. Performance of the modeling and forecasting strategy

Region	Modeling period (1871–2000)			Forecasting period (2001–2013)		
	$\sigma_m(e)$	CC_m	PP_m	$\sigma_f(e)$	CC_f	PP_f
GWB	2.06158	0.7060	0.7321	2.3562	0.8919	0.7499

* GWB: Gangetic West Bengal

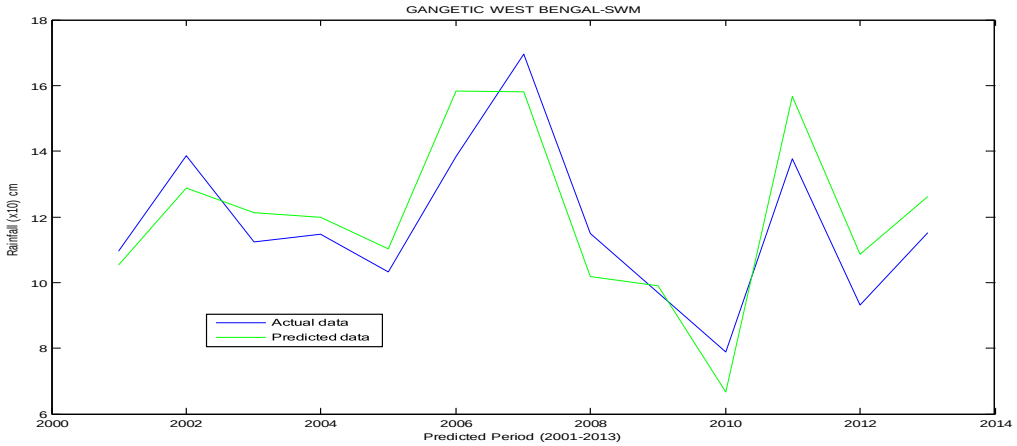

Figure 13. The actual SWM rainfall and Predicted SWM rainfall of GWB for the testing period (2001-2013)

.

The sample forecast may differ slightly from the actual observation because it is an expected value. Comprehensive numerical findings for the independent forecasts are shown in Table 7. The actual and forecasted rainfall data for the testing period (2001–2013) are detailed in Fig. 13.
It is demonstrated that, within certain bounds, the current method for predicting SWM rainfall one year in advance performs admirably. It should be mentioned that the sample forecast may not exactly match the actual observation because it is an expected value.

Table 7. Independent test forecasting (Gangetic West Bengal)

Year	Actual(x10)cm	Forecast(x10)cm)
2001	10.9499	10.5282
2002	13.8459	11.1995
2003	11.2329	14.3479
2004	11.4639	12.0564
2005	10.3210	12.7563
2006	13.8490	15.5299
2007	16.9519	16.9519
2008	11.5039	10.1950
2009	9.6949	9.8957
2010	7.8919	6.6618
2011	13.7800	15.6741
2012	9.3190	10.8719
2013	11.5200	13.5200

8. Performance of the model

Three statistical parameters are selected in order to calculate the model's performance. The first two are the correlation coefficient (CCm) and the Root Mean Square Error (RMSE) between the model's simulated values and the provided data. Additionally extracted is a statistic known as Performance Parameter [4], which is $PPm_d = 1 - (\sigma m2)/(\sigma 2)$, where $\sigma m2$ is the mean square error and $\sigma 2$ is the actual data variance. In an ideal model, both CCm and PPm would move toward unity, and $\sigma m2$ would be zero. According to Table 6, the current model's effectiveness is good for the testing period, and the correlation coefficient between the actual and predicted data is a respectably high 0.89. By loosening the restrictions on the forecasting exercise, the model parameters are kept constant for the entire thirteen-year period from 2001 to 2013, allowing the model to be verified for forecasting. However, the model parameters must be changed annually prior to forecasting. It is noted that even in the case of fewer than in ideal circumstances, the model's predictions are sufficient. The correlation coefficient (CCf) throughout the test period must be at least 0.6 in order to be considered significant for a sample size of N = 13 (2001–2013). Table 6 shows that CCf is significantly higher than 0.6.

9. Discussion

This research examines the IAV of GWB monsoon rainfall from a useful standpoint and highlights several intriguing features. Eight statistically nearly uncorrelated modes may be found in the seasonal SWM rainfall time series of GWB; the sum of these modes yields the original data. The climatic variance that persists across the entire data base is readily linked to the seventh and eighth modes. The six remaining empirical modes (IMFs) are narrow band random processes that are associated with specific, well-defined meteorological phenomena and have well-defined center periods. Strongly non-Gaussian, the first IMF that explains the most variability can be accurately predicted with GRNN algorithms. After subtracting the first IMF, the residual rainfall is suitable for a linear multiple regressive representation. A approach for rainfall forecasting has been created using two distinct representations. Other variability, such as intraannual, interseasonal, or intraseasonal variability that persists in monsoon rainfall, is not taken into consideration in this approach. The actual SWM rainfall for GWB in 2012 and 2013 was 93.19 cm and 115.20 cm, respectively, within one standard deviation of the mean rainfall, while the forecasted SWM rainfall for the same year was 108.71 cm and 126.21 cm. The first three IMFs have been found to have contributed about 90% of the variability among the first six IMFs. One could conclude that there is a significant likelihood of drought if those are both negative. These are quite positive and in line with [15] for situations like floods.

10. Conclusion

In the current paper, IAV of GWB has been examined from a novel perspective. It is known that the yearly SWM rainfall time series is broken down into eight statistically orthogonal modes; the total of these modes provides the original data with a high degree of accuracy. The following six empirical modes are linked to narrow-band random processes with defined central periods and significant meteorological phenomenon parameters, while the seventh and eighth modes are associated with the general climatic fluctuation. An intriguing method shows that the first mode, IMF1, which accounts for the most variability, is substantially non-Gaussian and is modeled using the GRNN technique, whereas the remaining portion of the rainfall may be represented using linear auto-regression. The rainfall prediction predicting exercise is finished for GWB by combining the two methods. The specific method is sufficiently broad, and attempts are underway to extend the analysis to other parts of India.

11. References

[1] Mooley DA, Parthasarathy B. Fluctuations in All-India summer monsoon rainfall during 1871–1978. Climate Change. 1978;6(3):287-301.
[2] Rupa Kumar K, Sahai AK, Krishna Kumar K, Patwardhan SK, Mishra PK, Revadekar J, Kamala K, Pant GB. High-resolution climate change scenarios for India for the 21st century. Current Science. 2006;90(3):334-345.
[3] Iyenger RN, Basak P. Regionalization of Indian monsoon rainfall and long term variability signals. Int. J. Climatol.1994;14:1095-1114.
[4] Sahai AK, Grimm AM, Satyan V, Pant GB. Long-lead prediction of Indian summer monsoon rainfall from global SST evolution. Climate Dyn. 2003:20: 855–863.
[5] Raja Rao KS, Lakhole NT. Quasi-biennial oscillation of summer southwest monsoon. Ind. J Meteorol. Hydrol. Geophys. 1978; 29: 403–411.
[6] Campbel WH, Blechman JB, Bryson RA. Long-period tidal forcing of Indian monsoon rainfall: a hypothesis. J. Clim. Appl. Meteorol. 1983; 22: 287–296.
[7] Narasimha R, Kailas SV. A wavelet map of monsoon variability. Proc. Ind. Nat. Sci. Acad. 2003;67(3): 327–341.
[8] Shukla J, Paolino DA. The southern oscillation and long-range forecasting of the summer monsoon rainfall over India. Mon. Wea. Rev. 1983;111: 1830–1837.
[9] Bhalme HN, Jadhav SK. The double (Hale) sunspot cycle and floods and droughts in India. Weather. 1984;39: 112-116.
[10] Hartmann DL, Michelsen ML. Intraseasonal periodicities in Indian rainfall. J. Atmos. Sci. 1989; 46(18): 2838–2862.
[11] Iyenger RN. Application of principal component analysis to understand variability of rainfall. Proc. Ind. Acad. Sc. (Earth Planet Sc.). 1991: 100(2): 105-126.
[12] Basak P. Variability of south west monsoon rainfall in West Bengal: An application of principal component analysis. Mausam. 2014; 65(4): 559-568.
[13] Basak P. Southwest monsoon rainfall in Assam: An application of principal component analysis. Mausam. 2017; 68(2): 357-366.

[14] Sahai AK, Soman MK, Satyan V. All India summer monsoon rainfall prediction using an artificial neural network. Climate Dyn.2000; 16:291–302.
[15] Iyengar RN, Raghu Kanth STG. Empirical modeling and forecasting of Indian monsoon rainfall. Curr. Sci. 2003; 85(8):1189–1201.
[16] Iyengar RN, Raghu Kanth STG. Intrinsic mode function and a strategy for forecasting Indian monsoon rainfall. Meteorol. Atmos. Phys. 2005; 90: 17-36.
[17] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N, Tung CC, Liu HH. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc Royal Soc London. 1998; A454: 903–995.
[18] Specht DF. A general regression neural network. IEEE Transaction on Neural Network. 2002; 2(6): 568-576.
[19] Zvarevashe W, Krishnanair S, Sivkumar V. Analysis of rainfall and temperature data using Ensemble Empirical Mode Decomposition. Data Science Journal. 2019; 18(1): 46.
[20] Sabzehee F, Nafisi V, Pour SI, Vishwakarma BD. Analysis of the precipitation climate signal using Empirical Mode Decomposition (EMD) over the Caspian Catchment Area. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2019; XLII-4/W18:923-929.