

Multiscale characterization and prediction of monsoon rainfall in India using Hilbert–Huang transform and time-dependent intrinsic correlation analysis

S. Adarsh^{1,2} · M. Janga Reddy¹

Received: 4 January 2017 / Accepted: 14 July 2017 / Published online: 25 July 2017
© Springer-Verlag GmbH Austria 2017

Abstract In this paper, the Hilbert–Huang transform (HHT) approach is used for the multiscale characterization of All India Summer Monsoon Rainfall (AISMR) time series and monsoon rainfall time series from five homogeneous regions in India. The study employs the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) for multiscale decomposition of monsoon rainfall in India and uses the Normalized Hilbert Transform and Direct Quadrature (NHT-DQ) scheme for the time–frequency characterization. The cross-correlation analysis between orthogonal modes of All India monthly monsoon rainfall time series and that of five climate indices such as Quasi Biennial Oscillation (QBO), El Niño Southern Oscillation (ENSO), Sunspot Number (SN), Atlantic Multi Decadal Oscillation (AMO), and Equatorial Indian Ocean Oscillation (EQUINOO) in the time domain showed that the links of different climate indices with monsoon rainfall are expressed well only for few low-frequency modes and for the trend component. Furthermore, this paper investigated the hydro-climatic teleconnection of ISMR in multiple time scales using the HHT-based running correlation analysis technique called time-dependent intrinsic correlation (TDIC). The results showed that both the strength and nature of association between different climate indices and ISMR vary with time scale. Stemming from this finding, a methodology employing

Multivariate extension of EMD and Stepwise Linear Regression (MEMD-SLR) is proposed for prediction of monsoon rainfall in India. The proposed MEMD-SLR method clearly exhibited superior performance over the IMD operational forecast, M5 Model Tree (MT), and multiple linear regression methods in ISMR predictions and displayed excellent predictive skill during 1989–2012 including the four extreme events that have occurred during this period.

1 Introduction

From several decades, characterization and forecasting of Indian Summer Monsoon Rainfall (ISMR) have received a lot of attention by meteorologists and hydrologists worldwide. In the past, several studies attempted to understand the teleconnection between ISMR and different global climatic oscillations. Walker (1923) was the first who established the teleconnections between Indian summer monsoon and El Niño Southern Oscillation (ENSO); and later on, many researchers investigated this link (Shukla and Paolino 1983; Mooley and Parthasarathy 1983; Parthasarathy and Pant 1984; Krishna Kumar et al. 1999; Gadgil et al. 2004; Maity and Nagesh Kumar 2006a, b). The link of ISMR with Quasi Biennial Oscillation (QBO) (Rao and Lakhole 1978; Vijayakumar and Kulkarni 1995; Claud and Pascal 2007), tidal forcing (Campbell et al. 1983), solar indices such as sunspot cycle, group sunspot number (SN), solar irradiance and sunspot number (Bhalme and Jadhav 1984; Bhattacharya and Narasimha 2007; Azad 2011), Eurasian snow depth (Hahn and Shukla 1976; Kripalani and Kulkarni 1999; Mangain et al. 2010), etc were few earlier attempts in this direction. Such

Responsible Editor: John T. Fasullo.

✉ M. Janga Reddy
mjreddy@civil.iitb.ac.in

¹ Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

² TKM College of Engineering, Kollam 691005, Kerala, India

investigations helped to find the potential inputs for rainfall forecasting models (Iyengar and Raghunath 2005; Singh and Borah 2012). Most of the past studies noted that ISMR time series is non-linear and non-stationary series. More specifically, the changes in statistical moments or covariance refer to the non-stationarity of time series, while the features such as non-normality, asymmetric cycles, bimodality, non-linear relationship between lagged variables, variations of prediction performance over the state-space, time irreversibility, and sensitivity to initial conditions refer to the non-linearity of the time series (Fan and Yao 2003). Although many natural phenomena can be approximated by linear systems, they also have the tendency to be non-linear whenever their variations become finite in amplitude (Huang et al. 1998). The non-linearity induced in the ISMR time series may be due to the non-linearity of the process or the observational non-linearity. Intra-wave frequency modulations may present in rainfall time series data of India, which is a hall mark of classical non-linear oscillators (Dhanya and Nagesh Kumar 2010). Moreover, based on multiscale decomposition of ISMR data sets, Iyengar and Raghunath (2005) proved that the separated components possess bimodality and stated that this behavior indicates strong non-linearity in the dynamics of the process behind ISMR.

Applications of the most popular Fourier transforms for the spectral analysis of hydro-meteorological time series are often constrained by the requirements of linearity and stationarity of the data sets, the use of trigonometric basis functions, etc. Wavelet analysis evolved as a solution to these problems and has been widely used to analyze multiscale behavior of non-stationary hydro-meteorological time series data (Anctil and Coulibali 2004; Labat 2005; Massei et al. 2007). In the past, several studies employed wavelets to analyze the ISMR time series and established its association with global climate oscillations (Torrence and Webster 1999; Narasimha and Kailas 2001; Azad et al. 2008; Narasimha and Bhattacharyya 2010). Even though wavelet transforms successfully address the issue of non-stationarity of data sets when the data are non-linear in characteristic, but the performance may not be appealing. It is quite rare to see linearity in hydro-meteorological observations particularly when the data set is too short or observational manipulations are present (Franceschini and Tsai 2010). Moreover, their application demands 'a priori' selection of proper wavelet function and setting of appropriate level of decomposition. To analyze the non-linear and non-stationary time series data, Huang et al. (1998) proposed an alternative spectral analysis technique namely Hilbert–Huang transform (HHT), which integrates a data adaptive multiscale decomposition process, namely empirical mode decomposition (EMD) and the Hilbert transform (HT). The EMD is a data adaptive operation

which decomposes a time series into a set of zero-mean component (called intrinsic mode function, IMF) and a final residue, based on spline fitting through the extrema of the time series signal. The IMFs obtained are then subjected to HT to examine the spectral properties of the time series. In recent past, the method gained popularity to analyze the hydro-meteorological time series signals (Duffy 2004; Huang et al. 2009a; Kuai and Tsai 2012; Massei and Fournier 2012; Adarsh and Janga Reddy 2016; Janga Reddy and Adarsh 2016).

Iyengar and Raghunath (2005) applied EMD for the decomposition of ISMR time series from eight regions in India. Based on the periodicity of IMFs, the possible association of ISMR with QBO, ENSO, sunspot cycles, and tidal forcing was hypothesized, and the resulted IMFs were used for forecasting of monsoon rainfall using the artificial neural networks (ANN). However, a quantitative assessment to prove the links of monsoon rainfall with that of different climate indices in multiple time scales remained as an open problem in their study. Establishing the link between the different climate indices and rainfall based on periodicity alone can convey only limited information on the variability of such series owing to their multiscale behavior. Performing a running correlation analysis in a multiscale framework which also accounts the non-stationarity of the time series can be a viable approach to solve this problem. The Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)-based multiscale decomposition can be used as a useful mean to investigate the links between five different climate indices such as QBO, ENSO, Sunspot Number (SN), Atlantic Multi Decadal Oscillation (AMO), and Equatorial Indian Ocean Oscillation (EQUINOO) with monsoon rainfall. Chen et al. (2010) proposed EMD-based time-dependent intrinsic correlation (TDIC) and it was successfully applied for teleconnection studies recently (Huang and Schmitt 2014; Ismail et al. 2015; Adarsh and Janga Reddy 2016). This study proposes the application of TDIC method to investigate the hydro-climatic teleconnection of ISMR by quantitatively establishing its linkage with different climate indices. Few studies also proved that the information of teleconnected hydro-climatic variables can be useful for prediction of Indian monsoon rainfall (Maity and Nagesh Kumar 2006a, b; Nagesh Kumar et al. 2007; Kashid and Maity 2012). Understanding the scale-specific association between the climate index series and monsoon rainfall may help for improved prediction of monsoon rainfall, and none of the past studies accounted the scale-specific information for the prediction of ISMR. In this context, the present study proposes an alternative method for ISMR prediction employing multiscale decomposition of climate indices.

The specific objectives of the present study include: (i) to quantitatively investigate the hydro-climatic

teleconnections between monsoon rainfall and five different climate indices in multiple time scales by employing the TDIC analysis; (ii) to propose an improved framework for ISMR prediction by accounting the multiscale association between the teleconnected variables.

The rest of the paper is organized as follows. First, the methodology followed for investigating the multiscale teleconnection of ISMR with different climate indices, and details of procedures used for prediction of ISMR are presented in Sect. 2. Then, the details of the study area and data sets are presented in Sect. 3. Thereafter, application of methodology to case study and discussion of results are presented in Sect. 4. The results section is organized in three sub-sections: first subsection presents the application of CEEMDAN algorithm for multiscale decomposition of ISMR time series and its corresponding results; second subsection presents the results of CEEMDAN-based TDIC analysis to investigate the association of a specified climate index with ISMR in different time scales, applied for five different cases; subsequently, in third subsection, the practical utility of the multiscale teleconnection study is demonstrated for prediction of AISMR by employing the proposed MEMD-SLR method. Finally, the key conclusions drawn from the study are presented in Sect. 5.

2 Methods

2.1 Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)

The scale separation by EMD is purely data dependent and no user specified selection of basis functions or levels (like ‘dyadic’ scales in discrete wavelets) is necessary to proceed with this technique. However, sometimes, ‘scale mixing’ problem (in which a decomposed IMF contains a mixture of drastically different periodic scales) may exist on its implementation. The noise assisted and ensemble averaged variant of EMD namely the Complete Ensemble EMD with Adaptive Noise (CEEMDAN) may alleviate this problem (Torres et al. 2011).

In CEEMDAN method, noise series is added at each stage of the decomposition to result in a unique residue in each mode from the residue of previous mode (or the true signal, for the first mode) and currently generated IMF. To illustrate the functioning details, the flowchart of the CEEMDAN algorithm is given in Fig. 1.

2.2 Investigating the hydro-climatic teleconnection of ISMR using TDIC

A cross-correlation analysis between the oscillatory components of monsoon rainfall and that of different climate

indices can provide preliminary information on the hydro-climatic teleconnections in multiple time scales (Janga Reddy and Adarsh 2016). As the hydro-climatic time series possess multiscale behavior, a scale-dependent running correlation analysis is more appropriate to investigate the hydro-climatic teleconnections. Most of the running correlation methods proposed earlier involve the estimation of correlation coefficient of the data subsets by fixing suitable sliding window size (Papadimitriou et al. 2006; Rodo and Rodriguez-Arias 2006; Scafetta 2014). However, the selection of appropriate window size is a challenging problem while applying such techniques, and a data adaptive selection of window size can be a solution to this problem. The TDIC analysis as proposed by Chen et al. (2010) fixes the window size adaptively based on instantaneous frequency computed by HHT. TDIC method can be applied between a typical IMF of rainfall and the corresponding IMF of climate index to capture the association between the rainfall and climate index at specific time scale. The key steps involved in the process are given below:

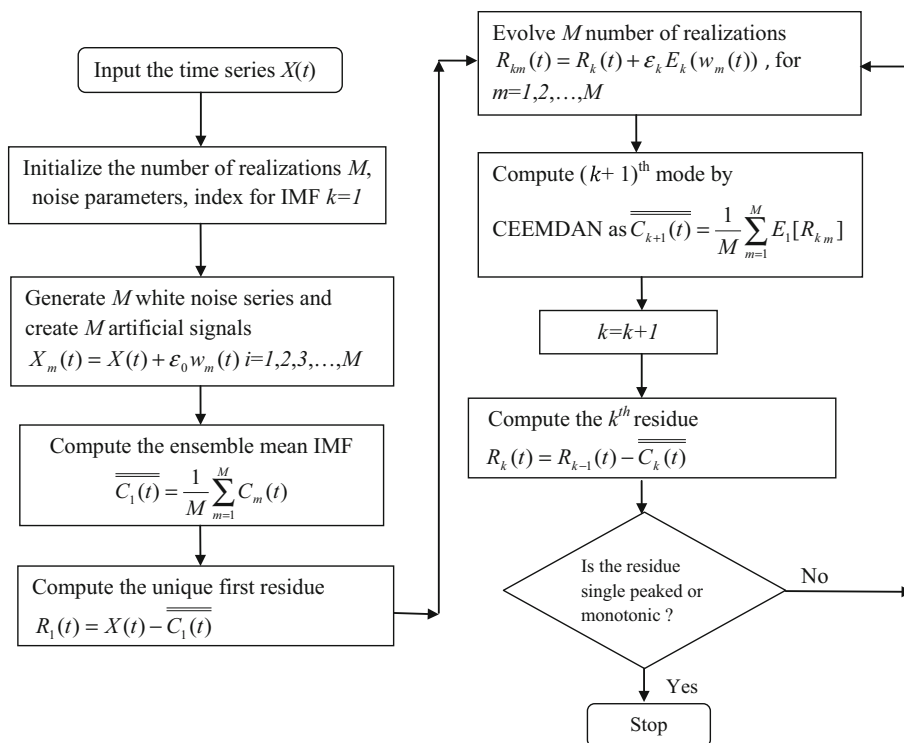
1. Apply HHT on the selected IMF pairs to obtain instantaneous frequencies (and hence instantaneous periods)
2. Fix the minimum sliding window size (t_d) as maximum instantaneous period between the two signals at the current position t_k , i.e., $t_d = \max(T_{1,i}(t_k), T_{2,i}(t_k))$, where $T_{1,i}$ and $T_{2,i}$ are instantaneous periods.
3. The sliding window is fixed as $t_w^n = [t_k - \frac{nt_d}{2} : t_k + \frac{nt_d}{2}]$,

where n is any positive number (a multiplication factor for minimum sliding window size) and normally n is selected as 1 (Huang and Schmitt 2014).

4. Let IMF1 and IMF2 are two IMFs of nearly the same mean period pertaining to two different time series. The TDIC of the pair of IMFs can be found out from $R_i(t_k^n) = \text{Corr}(\text{IMF}_{1,i}(t_w^n), \text{IMF}_{2,i}(t_w^n))$ at any t_k , where Corr is the correlation coefficient of two time series
5. Perform Student’s t test to investigate statistical significance of correlation coefficients obtained in the previous step.
6. Repeat the above two steps iteratively till the boundary of the sliding window exceeds the end points of the time series.

After computing the TDIC matrix, the TDIC plot is prepared. The horizontal axis of the TDIC plot is the time axis corresponding to the center position of the sliding window, and the vertical axis is the size of the sliding window. The TDIC plot will be triangular in shape and the correlation at the apex point will be the correlation coefficient between the series considering the entire data

Fig. 1 Flowchart of CEEMDAN algorithm. The operator $E_k(\cdot)$ produces k th the mode obtained by EMD



length, if the data length is chosen as the maximum window size (Chen et al. 2010).

2.3 MEMD-SLR approach for rainfall prediction

In this study, the multivariate extension of EMD (MEMD) proposed by Rehman and Mandic (2010) is used for decomposition of multivariate data set, as MEMD performs the decomposition of all variables in a single step and ensures equal number of modes (Huang et al. 2016). The SLR method is used for modeling individual components, as it facilitates the inclusion or exclusion of a particular variable at a specific time scale based on its statistical significance. The theoretical details of MEMD are presented in the “Appendix 1” and the details on SLR can be found in Draper and Smith (1998). The flowchart depicting the key steps involved in MEMD-SLR approach for rainfall prediction is presented in Fig. 2.

3 Study area and data sets

In this study, the monsoon rainfall data pertaining to different regions of India are analyzed. Indian Institute of Tropical Meteorology (IITM) (<http://www.tropmet.res.in>) located at Pune have identified 36 meteorological subdivisions in India. In addition, these subdivisions are grouped into eight regions such as All India (AI), Homogeneous India (HOI), Core Monsoon (CMI), Western Central India

(WCI), Central North East India (CNEI), North East India (NEI), North West India (NWI), and Peninsular India (PI) by the IITM Pune. Out of eight regions, the first three are overlapping regions. All India (AI) considers all the 36 meteorological subdivisions except for the some hilly regions in northern part of India, while the other regions are formed based on similarity in rainfall characteristics. The five non-overlapping regions considered for the study are shown in Fig. 3. The monthly rainfall data of all the regions for the period 1871–2012 were collected from IITM Pune. The monsoon rainfall for each region is computed by adding the monthly rainfall values of monsoon period (i.e., June, July, August, and September months).

To study and explore the hydro-climatic teleconnections of Indian monsoon rainfall, data of five different climate indices were collected and used in this study. The monthly data of QBO were obtained from the website of National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (<http://www.esrl.noaa.gov/psd/data/correlation/qbo.data>) for the period 1950–2012. The intensities of El Niño Southern Oscillation (ENSO) are generally assessed on the basis of the average Sea Surface Temperature (SST) over different Niño regions in the Pacific Ocean within specific latitudes and longitudes, and it has been found that summer monsoon rainfall over India is best correlated with temperature anomaly from Niño 3.4 region, which overlaps between Niño 3 and Niño 4. The SST anomaly data corresponding to the Niño 3.4 region (120°W–170°W, 5°S–5°N) called as Oceanic Niño Index

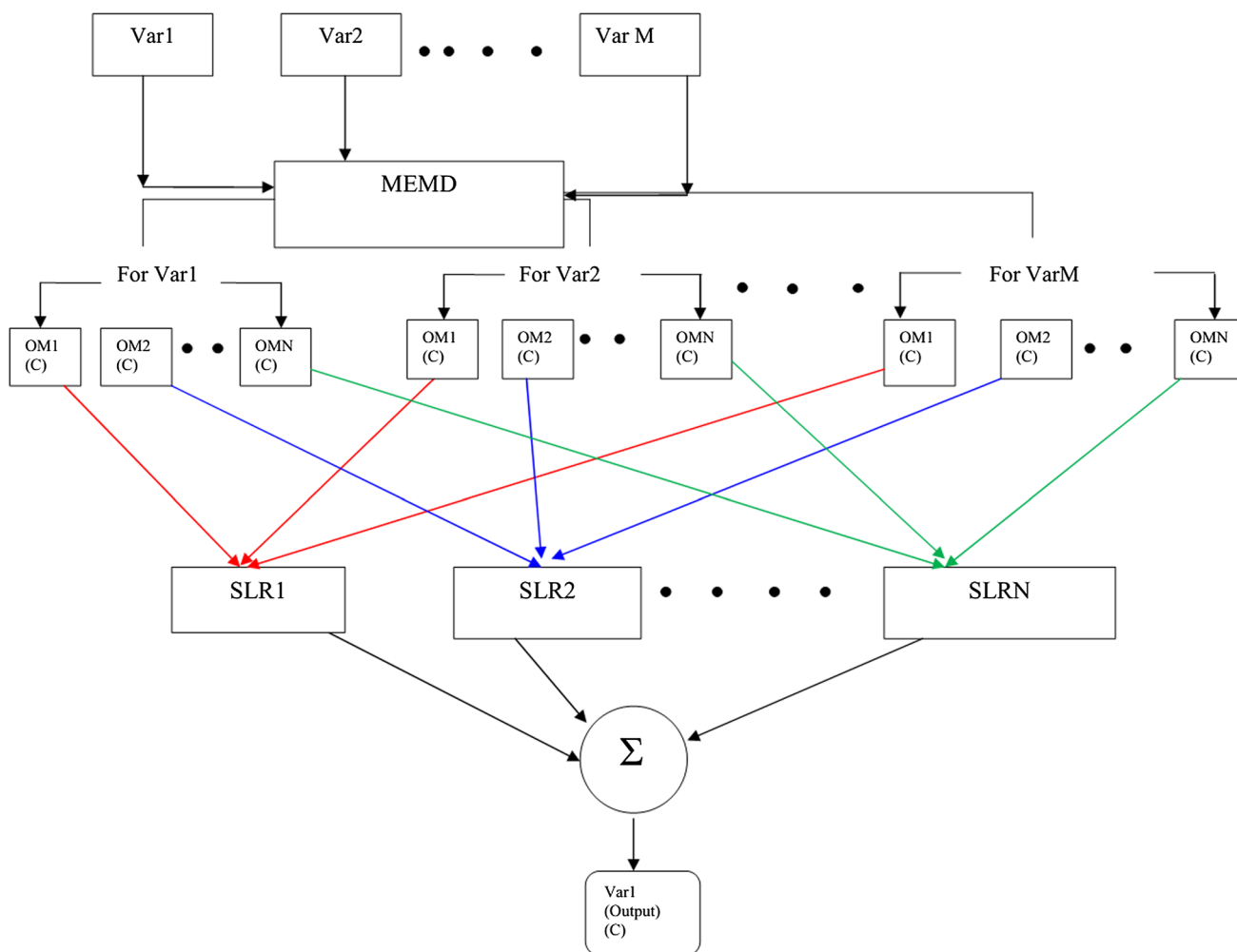


Fig. 2 Flowchart of MEMD-SLR procedure, describing the steps of the model development for rainfall prediction (*Var* variable, *C* calibration, *OM* orthogonal mode, *SLR* stepwise linear regression, *MEMD*

multivariate empirical mode decomposition). Here, the first variable is considered as output and rest of the variables as inputs

(ONI) were obtained from NOAA National Weather Service Climate Prediction Center (<http://www.cpc.ncep.noaa.gov/data/indices/>) for the period 1950–2012 and used as the ENSO index. The sunspot number records are obtained from the solar physics group at NASA’s Marshall Space Flight Centre (http://solarscience.msfc.nasa.gov/greenwch/spot_num.txt) for the same period and used in the present study. The relationship of AMO with monsoon rainfall was also investigated in few studies (Goswami et al. 2006; Lu et al. 2006; Zhang and Delworth 2006; Feng and Hu 2008). The data of monthly AMO indices were obtained from NOAA National Weather Service Climate Prediction Center (<http://www.esrl.noaa.gov/psd/data/timeseries/AMO/>) for the period 1950–2012. The relation between the EQUINOO and Indian monsoon was studied extensively by various researchers (Gadgil et al. 2004; Maity and Nagesh Kumar 2006a, b; Nagesh Kumar et al. 2007; Kashid and Maity 2012). The zonal wind component for

the region 60°–90°E, 2.5°S–2.5°N) was obtained from the National Centre for Environmental prediction (NCEP) (<http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html>) for the same period. The negative of the anomaly of the zonal component of surface wind in the equatorial Indian ocean region (60°–90°E, 2.5°S–2.5°N) is considered as EQUINOO index (Maity and Nagesh Kumar 2006a). It is to be noted that the data from 1950 to 2012 have been considered for the present analysis, as the data for most of the climate indices were available from 1950 onwards.

4 Results

In this section, first, the results of multiscale decomposition of AISMR and monsoon rainfall of five non-overlapping regions by employing the CEEMDAN algorithm are

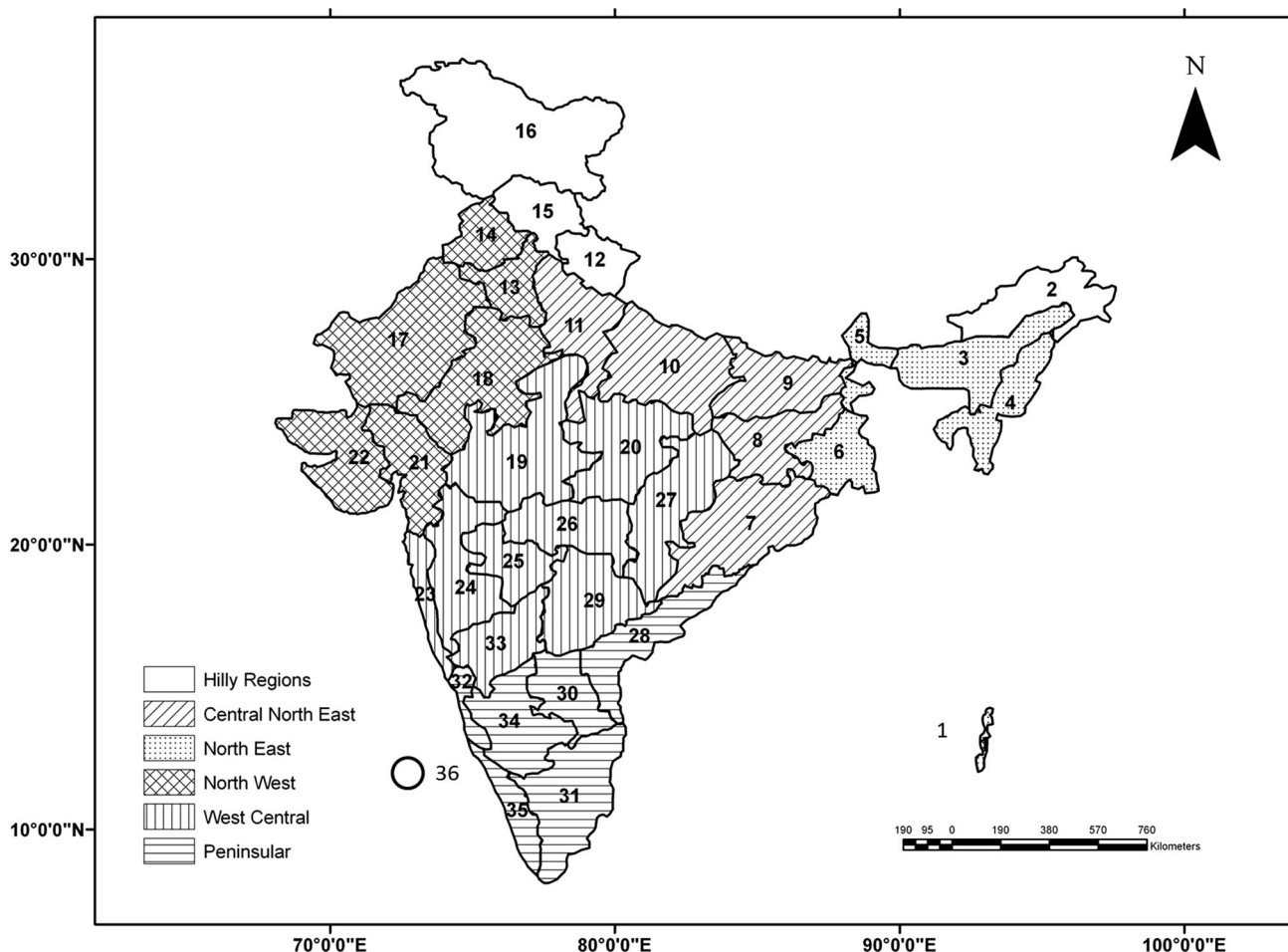


Fig. 3 Location map showing five non-overlapping regions in India (1 Andaman Nicobar Islands, 2 Arunachal Pradesh, 3 Assam & Meghalaya, 4 Nagaland, Manipur, Mizoram & Tripura, 5 Sub Himalaya, West Bengal & Assam, 6 Gangetic West Bengal, 7 Orissa, 8 Jharkhand, 9 Bihar, 10 East Uttar Pradesh, 11 West Uttar Pradesh, 12 Uttaranchal, 13 Haryana, Chandigarh & Delhi, 14 Punjab, 15 Himachal Pradesh, 16 Jammu & Kashmir, 17 West Rajasthan, 18 East

Rajasthan, 19 West Madhya Pradesh, 20 East Madhya Pradesh, 21 Gujarat, 22 Saurashtra, Kutch & Diu, 23 Konkan & Goa, 24 Madhya Maharashtra, 25 Marathwada, 26 Vidarbha, 27 Chhattisgarh, 28 Coastal Andhra Pradesh, 29 Telangana, 30 Rayalaseema, 31 Tamil Nadu & Pondicherry, 32 Coastal Karnataka, 33 North Interior Karnataka, 34 South Interior Karnataka, 35 Kerala, 36 Lakshadweep)

presented, and then, the results of spectral analysis of IMF components obtained from the decomposition are discussed. Subsequently, the orthogonal modes of monthly AISMR time series are compared with the modes of five climatic indices series in the time domain, and the inter-relationships between them are investigated by the TDIC analysis. Finally, the results of CEEMDAN-SLR for monsoon rainfall prediction are presented.

4.1 Multiscale decomposition of monsoon rainfall time series

The CEEMDAN algorithm is invoked to decompose all the six time series to a set of orthogonal modes (OM), where each mode is associated with specific time scale. The results of CEEMDAN-based decomposition of AISMR and

other time series are presented in Fig. 4. To run the CEEMDAN algorithm, a noise standard deviation of 0.2 and 500 realizations was selected, considering the suggestions from the past studies (Torres et al. 2011; Antico et al. 2014). The CEEMDAN method decomposes the time series data into five different IMFs and a residue except for the rainfall time series of NEI. The mean period of the time series can be approximated by dividing the number of data samples by half the number of zero crossings (Barnhart and Eichinger 2011). The mean period of different modes along with the percentage variability explained by different modes is presented in Table 1.

The mean periods and the percentage variability explained by different modes show close matching with those reported by Iyengar and Raghukanth (2005). Identification of similar periodicities in different modes of ISMR

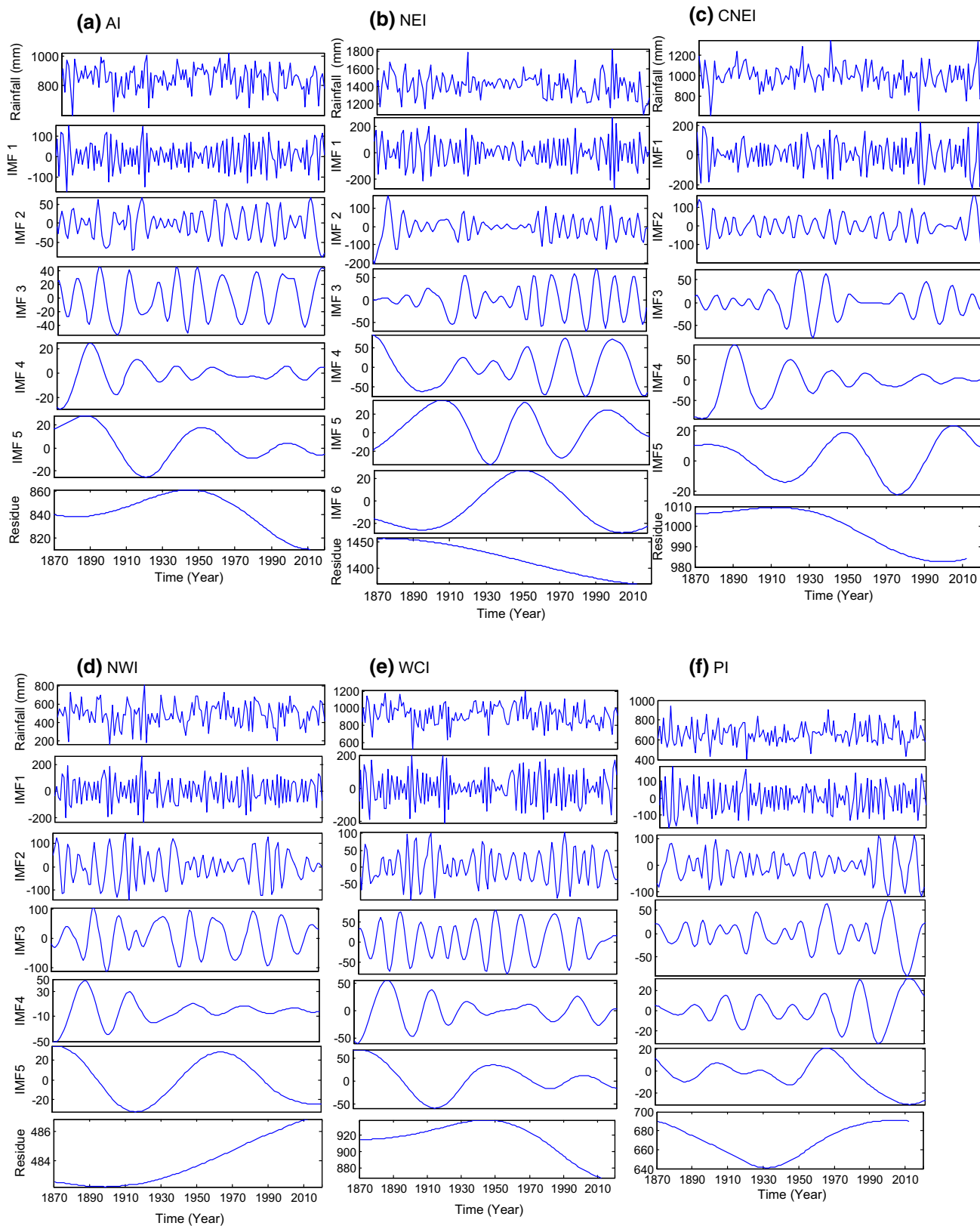


Fig. 4 Orthogonal modes of rainfall time series of different regions: **a** All India (AI), **b** North East India (NEI), **c** Central North East India (CNEI), **d** North West India (NWI), **e** Western Central India (WCI), and **f** Peninsular India (PI)

Table 1 Mean periods (in years) and percentage variability explained (VE) by different modes of rainfall time series from different regions of India

Mode	AI		NEI		CNEI		NWI		WCI		PI	
	Mean period (years)	VE (%)	Mean period (years)	VE (%)	Mean period (years)	VE (%)	Mean period (years)	VE (%)	Mean period (years)	VE (%)	Mean period (years)	VE (%)
IMF1	2.73	64.67	2.9	55.92	3.0	62.01	2.8	54.78	2.6	60.96	2.73	63.16
IMF2	5.68	17.13	5.7	19.48	6.4	23.15	5.7	23.55	5.7	13.63	6.17	21.29
IMF3	12.91	11.57	10.9	5.52	10.1	4.44	14.2	17.05	11.8	11.87	11.83	7.80
IMF4	28.40	1.47	28.4	10.72	20.3	8.71	28.4	1.95	28.4	3.75	20.28	2.13
IMF5	71.0	3.28	47.3	2.09	47.3	1.02	71.0	2.66	47.3	7.39	47.33	2.18
IMF6/ Residue	142.0	1.89	71.0	1.94	71.0	0.67	142.0	0.01	142.0	2.40	142.0	3.45
Residue			142.0	4.33								

and the modes of the different climate indices suggests the preliminary evidence of possible teleconnections of ISMR with different global climate indices. IMF1 possesses an average period of 2.6–3.02 years in different regions (Table 1). The periodicity of biennial oscillation is 2–3 years and observing such a mode in ISMR time series suggests the possible link between the two. The ENSO has 3–7 year periodicity (Kripalani and Kulkarni 1997a, b; Iyengar and Raghukanth 2005) and the average period of second mode obtained by the decomposition of ISMR from different region varies between 5.68 and 6.17. Hence, a possible link of ISMR and ENSO could be assessed. The IMF3 has a mean periodicity between 10.9 and 11.9 years in all the regions except NWI. It is well understood that sunspot time series have a mean periodicity of 11 years (Usoskin and Mursula 2003; Claud et al. 2008; Barnhart and Eichinger 2011); hence, a preliminary notion on the link between ISMR and sunspot cycle could be established. The mean period of fourth mode (IMF4) is found to be between 23 and 29 years, which could be possibly linked with the tidal forcing of similar periodicity (Campbell et al. 1983; Iyengar and Raghukanth 2005). Campbell et al. (1983) examined the precipitation records of northern India for the period 1895–1975 using the eigenvector analysis. By comparing its dominant frequencies of the precipitation data with that of soli-lunar tidal potential at the latitude of northern India, the study hypothesized that the tidal effects modulate the advance of monsoon front and proposed a method for prediction of June rainfall one year in advance using the information on tidal frequencies. The IMFs having multidecadal periodicity of more than 60 years (IMF5) may represent a possible association of AMO with the Indian monsoon rainfall (Goswami et al. 2006). Moreover, a plot between the IMF and the respective climate index time series can be prepared to explain the link

between climate indices and the different modes obtained by the multiscale decomposition. As the annual data sets for the period 1871–2012 are available only for the SN time series and AMO time series, hence only the plot of IMF3 and IMF5 is considered for analysis. The comparison of respective IMF and climate index time series is presented in Fig. 5. The visual comparison between the IMF and time series shows striking similarity and similar evolution in most of the time duration. This further confirms the distinct possibility of the link of SN and AMO with the ISMR.

4.2 Investigating hydro-climatic teleconnection of AISMR

The ‘hydro-climatic teleconnection’ refers to the association of hydrologic variables with large-scale atmospheric/oceanic oscillations from different parts of the world. In the teleconnection analysis using the multiscale decomposition, the comparison of periodicity of oscillatory modes of rainfall with that of climatic variable is a standard approach to establish the links between the two variables (Iyengar and Raghukanth 2005; Massei et al. 2007; Massei and Fournier 2012). The five climate indices considered in the present study include QBO, ENSO, SN, AMO, and EQUINOO.

As a first step, the multiscale decomposition of AI monthly monsoon rainfall time series and all the five climate indices for the period 1950–2012 is performed using the CEEMDAN algorithm. The mean period of the modes obtained from the decomposition of different climate indices and AISMR are presented in Table 2. Table 2 shows that different climate indices possess nearly the same period of evolution as that of AISMR particularly for lower modes. For higher modes, the periodicities do not match and this may be because only limited number of cycles is present in

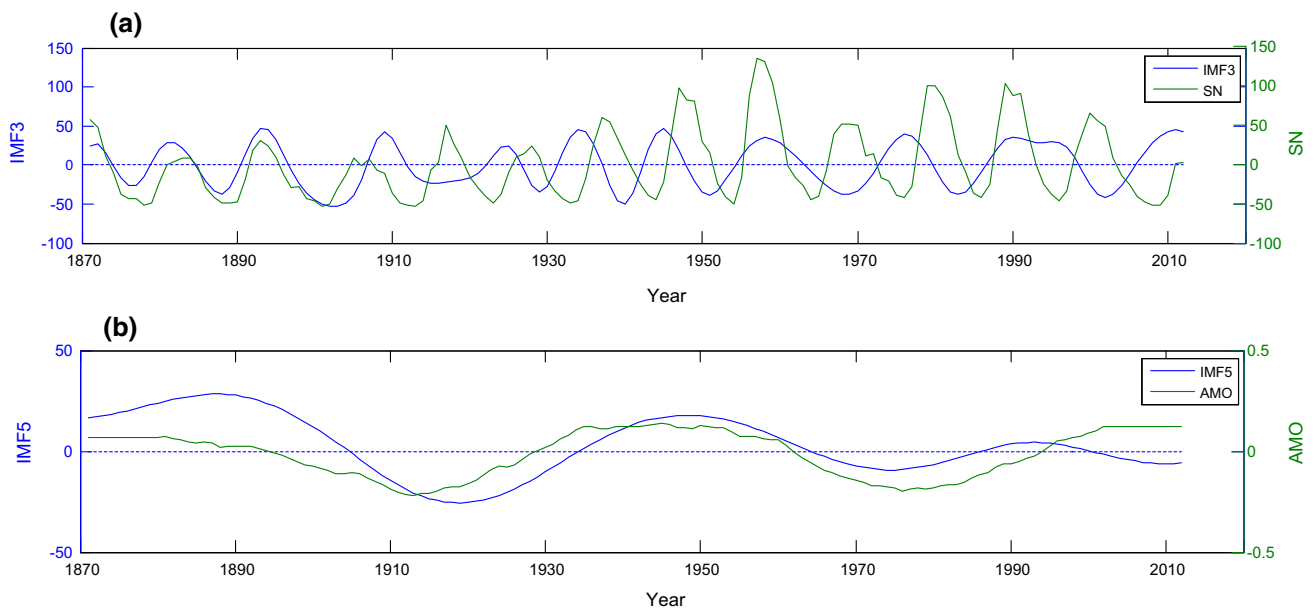


Fig. 5 Comparison of IMF of rainfall and the zero-mean climate index time series **a** IMF3 with the SN time series, **b** IMF5 with the AMO time series

Table 2 Time period of modes of AI monthly monsoon rainfall and different climate indices for 1950–2012

Mode	Rainfall AI	Climate indices				
		QBO	ENSO	SN	AMO	EQUINOQ
IMF1	3.231	3.150	3.600	3.111	3.360	3.150
IMF2	6.300	9.333	8.400	6.146	7.636	6.811
IMF3	13.263	18.000	15.750	11.455	14.000	14.000
IMF4	28.000	31.500	28.000	42.000	28.000	25.200
IMF5	42.000	63.000	50.400	126.000	84.000	50.400
IMF6	84.000	126.000	126.000	126.000	126.000	126.000
Residue	252.000	252.000	252.000	252.000	252.000	252.000

IMFs at larger scales. Furthermore, a cross-correlation analysis between the modes of AISMR time series with that of different climate indices is performed and the corresponding results are presented in Table 3. From Table 3, it can be noticed that there exists a very high degree of correlation only between the trend components of the monsoon rainfall time series and different climate indices. Also few higher order modes show a reasonable correlation, as highlighted in bold numbers in Table 3, for example, IMF6 of monsoon rainfall and IMF5/IMF6 of SN time series; IMF6 of monsoon rainfall with IMF6 of QBO, etc. This is an important observation as it can be concluded that the relationship between climate indices and monsoon rainfall shows a better agreement in the low-frequency part of their spectra. To understand the pattern of evolution in the time domain (i.e., how the changes occur above and below the mean value over the time domain), the plots of trend components of monsoon rainfall with the trend component of all five climate indices are prepared and presented in Fig. 6. A

comparison of the trends of different indices with that of rainfall shows that different indices have very good agreement with rainfall. In addition, the changes above the mean are found to be occurring in similar manner. Furthermore, it is noticed that the zero crossing occurs more or less at the same time instant in all cases. This also establishes the strong long-term association of the different climate indices with Indian monsoon rainfall.

It is to be noted that in the present correlation analysis, the correlation between modes is computed by considering the entire time span. However, there is a possibility that the IMFs of climate indices and rainfall may show strong positive (or negative) correlation for shorter time spells. Also at some of the time spells, such correlation can be negative, and at some other spells, it may be positive. To illustrate this aspect, the modes of ENSO and AISMR are plotted and presented in Fig. 7.

From Fig. 7, it can be noticed that the correlation between the two is negative in most of the lower modes and

Table 3 Correlation coefficients between orthogonal components of monthly monsoon rainfall (1950–2012) of AI and orthogonal components of different climate indices [italic letters indicate significant correlation (p value <0.05) and bold letters indicates moderate-to-strong correlation, say $> \pm 0.5$]

Mode of monsoon rainfall	Climate index	Mode of climate index						
		IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	Residue
	QBO							
IMF1		-0.085	0.044	-0.03	-0.012	-0.013	-0.018	-0.002
IMF2		0.015	0.019	0.078	0.015	0.039	0.018	0.027
IMF3		0.019	-0.007	<i>-0.407</i>	0.12	-0.048	-0.061	-0.024
IMF4		0.016	0.042	-0.133	-0.115	0.563	<i>0.217</i>	0.156
IMF5		-0.009	0.01	0.026	-0.084	<i>-0.395</i>	<i>-0.264</i>	-0.111
IMF6		-0.031	-0.016	0.049	-0.023	-0.027	0.879	0.15
Residue		0.01	0.035	0.096	0.035	-0.064	-0.181	0.999
	ENSO							
IMF1		-0.031	-0.022	0.022	0.000	-0.038	-0.014	0.000
IMF2		-0.114	-0.347	-0.165	-0.018	0.038	-0.008	-0.025
IMF3		0.065	-0.084	-0.399	-0.183	0.085	-0.013	0.015
IMF4		-0.025	0.035	-0.104	-0.100	-0.055	0.170	-0.134
IMF5		0.016	-0.005	0.020	0.094	-0.154	-0.188	0.088
IMF6		0.035	-0.032	0.038	-0.110	-0.202	0.945	-0.142
Residue		-0.003	-0.014	0.078	-0.097	-0.147	0.119	-0.992
	SN							
IMF1		-0.010	-0.018	-0.041	-0.017	0.014	-0.002	-0.001
IMF2		0.018	-0.014	-0.032	-0.053	-0.017	-0.007	0.025
IMF3		-0.009	0.013	0.068	0.157	-0.042	-0.042	-0.019
IMF4		-0.017	0.039	-0.102	0.073	<i>0.329</i>	<i>0.273</i>	0.151
IMF5		0.030	-0.028	0.129	-0.171	<i>-0.354</i>	<i>-0.328</i>	-0.104
IMF6		0.015	0.003	0.117	-0.079	0.653	0.810	0.173
Residue		-0.008	0.009	0.127	-0.058	<i>0.238</i>	0.175	0.996
	AMO							
IMF1		0.085	0.020	-0.023	0.010	0.032	0.010	0.001
IMF2		-0.065	-0.044	0.017	-0.045	-0.015	-0.006	-0.027
IMF3		-0.063	-0.072	0.074	0.124	0.054	0.054	0.018
IMF4		0.024	0.154	0.059	<i>-0.206</i>	-0.003	-0.173	-0.142
IMF5		-0.007	-0.030	0.122	-0.014	<i>-0.401</i>	0.129	0.102
IMF6		-0.032	0.024	0.136	0.007	<i>0.418</i>	-0.026	-0.181
Residue		0.022	0.077	-0.027	0.023	<i>0.320</i>	-0.559	-0.992
	EQUINOO							
IMF1		<i>-0.349</i>	-0.068	-0.013	-0.017	-0.028	0.000	0.000
IMF2		-0.036	-0.154	-0.032	-0.023	0.010	-0.003	-0.024
IMF3		-0.006	-0.094	-0.151	0.011	-0.069	-0.037	0.011
IMF4		0.012	0.051	-0.166	<i>-0.256</i>	0.082	0.020	-0.126
IMF5		-0.035	-0.071	0.021	0.034	-0.041	0.004	0.079
IMF6		-0.047	0.026	-0.037	-0.132	<i>0.208</i>	<i>-0.496</i>	-0.157
Residue		0.027	0.028	0.023	-0.098	<i>0.207</i>	0.141	-0.985

the residue, but the two are positively correlated for the shorter time spells ~1960–65 in IMF2, ~1978–82 in IMF2 and IMF3, ~1994–96 in IMF3, etc. This implies that

the correlation coefficient of two data series on the whole domain alone may not reveal the possible relationship between them particularly when the processes are

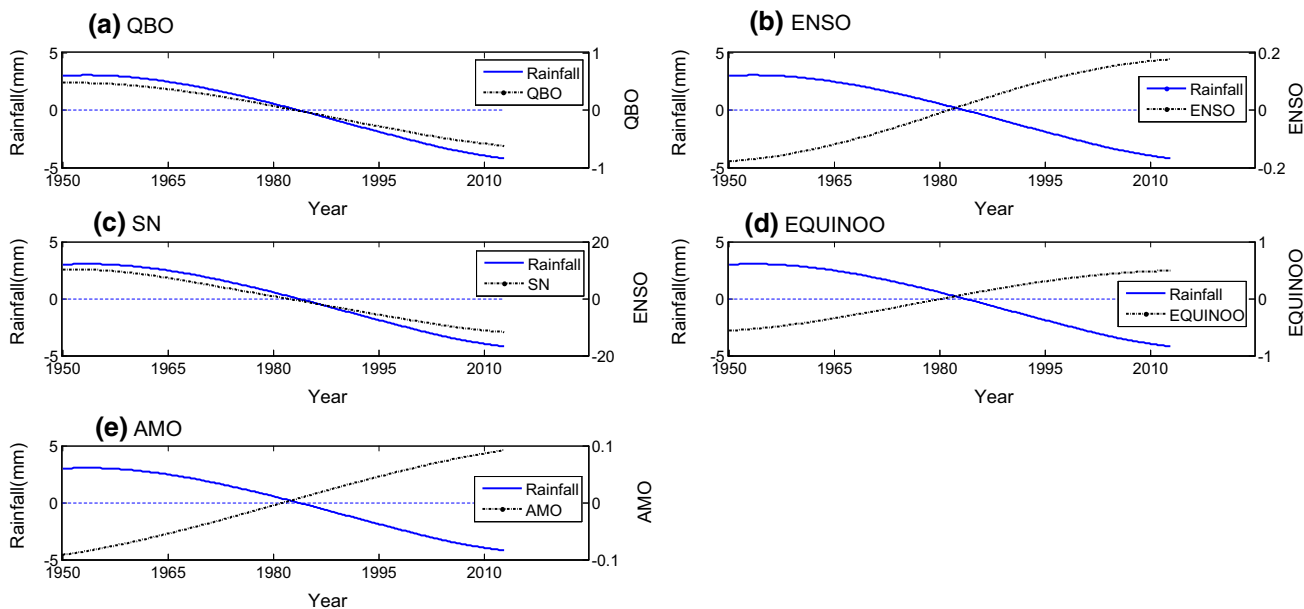


Fig. 6 Trend components of climate indices and monsoon rainfall of AI during 1950–2012

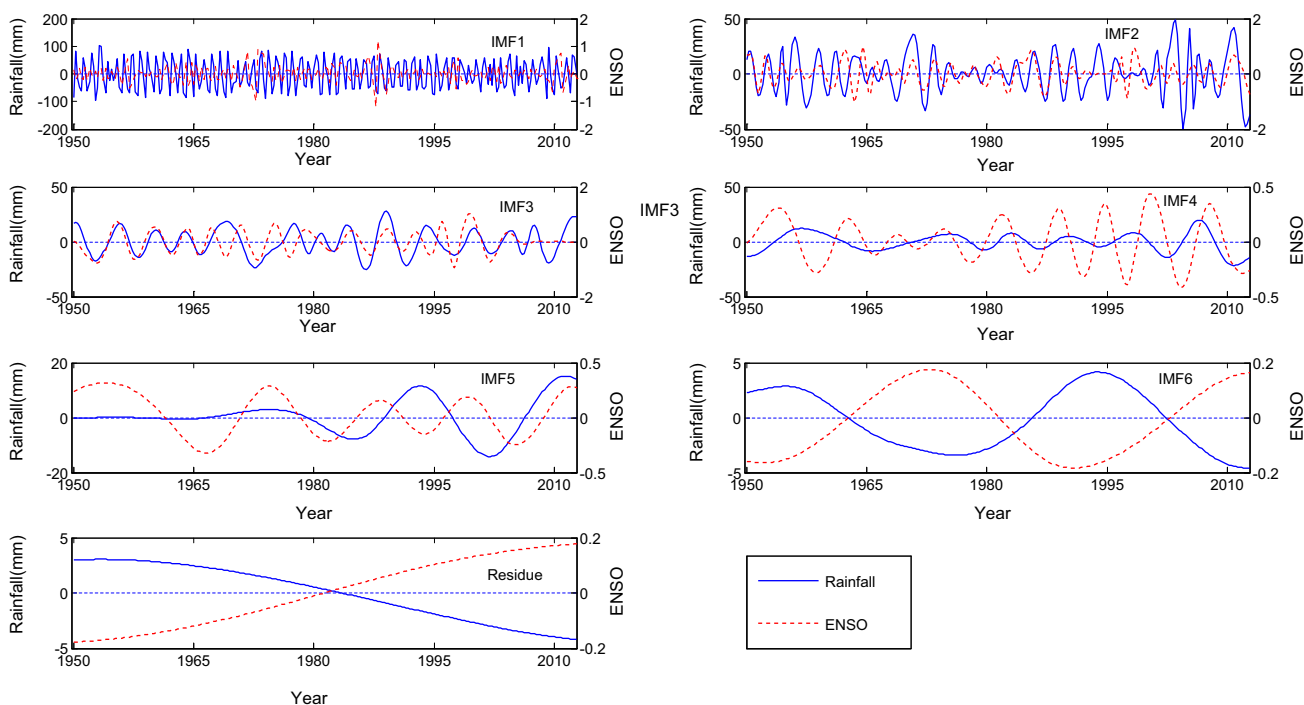


Fig. 7 Plot of modes of ENSO and AISMR time series

intermittent or non-stationary or contain drift and trends. In addition, it is noticed that the relation between IMF4 of ENSO and rainfall is very weak for most of the time periods. A strong positive correlation between ENSO and AISMR is observed in IMF6 in most of the time periods and in IMF5 for the period ~1987–2007. The relation in IMF5 was noticed with a phase shift for the period ~1997–2007.

Hence, it can be concluded that for the complex relation between rainfall and climate indices, certain processes involved may correlate with each other in one scale but not in others. In addition, the correlation between rainfall and climate indices might have changed from strong positive correlation to strong negative correlation in some of the scales (i.e., for IMFs). This evidence is clear on considering the much investigated relation between ENSO and

ISMR by the comparison of their IMFs. The two events are negatively correlated for the residue component (correlation coefficient of -0.992), but a strong positive correlation for the IMF6 (correlation coefficient of 0.945) (Table 3). This is noteworthy, because negative correlation at one scale would counteract positive correlation at another scale, resulting low overall correlation between rainfall and climate indices. In this case, the overall correlation between ENSO and monsoon rainfall series is only -0.21 , while the correlation values between rainfall series and QBO, SN, AMO, and EQUINOX are 0.01 , -0.001 , 0.013 , and -0.33 , respectively. It is to be noted that above correlation values are quite low and the correlation coefficient depicts only the linear association between the different variables with ISMR, while the true association between them might be of non-linear in characteristics. However, the climate forcing of lower periodicity may last for shorter time spells and they might have significant impact on the hydrological processes of the regime. To identify such local associations, a correlation analysis considering shorter time spells can be adopted. Therefore, an in-depth intrinsic correlation analysis of IMFs may give more insight to the linkages between climate indices and Indian monsoon rainfall. Therefore, a time-dependent intrinsic correlation analysis (TDIC) of Indian monsoon rainfall time series is performed.

4.3 Time-dependent intrinsic correlation analysis of AISMR and climate indices

The TDIC was calculated among the different pairs of IMFs of climate index and monsoon rainfall, and the TDIC plots are prepared for the first four IMFs, as the periodicities of monsoon rainfall and different climate indices matches reasonably well only for the first four IMFs, and the TDIC analysis of higher order modes (IMF5 onwards) rainfall and climate indices does not pass the student t test in most of the time scales for most of the time spells.

In the implementation of TDIC analysis, first, the spectral analysis of IMFs is to be performed and the instantaneous frequencies (and the reciprocal of which gives the instantaneous period) are to be estimated. Sometimes, the traditional HT may lead to instantaneous frequencies that are of less physical meaning (such as negative frequency) or it may show mathematical incorrectness as stated in two well-known theorems such as Bedrosian theorem (Bedrosian 1963) and Nuttall theorem (Nuttall 1966). The restriction of the Bedrosian theorem (Bedrosian 1963) can be surpassed through normalization of the resulting IMFs and restriction imposed by Nuttall (1966) theorem can be circumvented by the direct quadrature (DQ) method proposed by Huang et al. (2009b). In this study, the Normalized Hilbert Transform and DQ

(NHT-DQ) scheme are used for determining the instantaneous frequency, and the details of this scheme can be found elsewhere (Huang and Wu 2008; Huang et al. 2009a, b). Using the instantaneous periods, the TDIC algorithm is invoked and the TDIC plots are prepared for different cases. To illustrate the applicability of TDIC analysis, first, the relationship between ENSO and rainfall is considered, and the TDIC plots for different IMFs are presented in Fig. 8. The bottom contour of the triangular plots depicts the instantaneous frequency and hence a shift of the plots to larger time scales can be noticed in higher order IMFs (of low-frequency modes). The white spaces in the plots represents that such correlations fails to satisfy the Student's t test and hence not statistically significant. It is to be noted that in the teleconnection studies, the appended monthly data sets of monsoon period from different years are considered, while HHT is one such tool recommended even for the spectral analysis of data sets with irregular periodicity (Huang et al. 2009a; Cong and Chetouani 2009; Huang 2013; Rahman et al. 2015). In this case, the main aim is to examine the association between the variables in multiple time scales whatever be the scale (periodicity) associated with them. However, it is important to perform the appropriate time scale conversion, while periodicity becomes the central focus of the discussion. ENSO shows a long range negative correlation with AISMR for IMF2 ($\sim 8.4 \times 4/12 = 3$ year periodicity) and IMF3 ($\sim 15.7 \times 4/12 = 5$ year periodicity), so it can be inferred that both modes may be contributed by same physical processes like westerly wind bursts or oscillatory patterns such as Madden-Julian oscillations (MJO) which vary in intraseasonal scales (Wang and Picaut 2004), even though the exact physical mechanism behind such transition in the nature of correlation could not be adduced from the present analysis at this stage. However, localized positive correlation of ENSO with AISMR is observed in different modes in different shorter time spells, for example, IMF2 in $\sim 1964-66$, IMF3 in $\sim 1979-82$, IMF4 in $1995-1998$, etc. It is well known that the strongest El Niño of the century (1997-98) triggered an Indian Ocean Dipole (IOD) mode and which resulted in above-average rainfall during the period 1997-98 (Kumar et al. 2006). From the TDIC analysis (Fig. 8), it is found that for IMF4, the correlation between the AISMR and ENSO time series is strongly positive during 1995-98. This matches with the results presented by Chen et al. (2010) who established the relation between IOD and ENSO using the TDIC analysis. Moreover, it is observed that there are frequent alterations in nature of correlation between ENSO and AISMR in the high-frequency mode (IMF1). Such dynamics (transition from positive correlation to negative and vice versa) are more apparent for smaller window size (i.e., in the lower part of the TDIC plot). In addition, it is the well understood

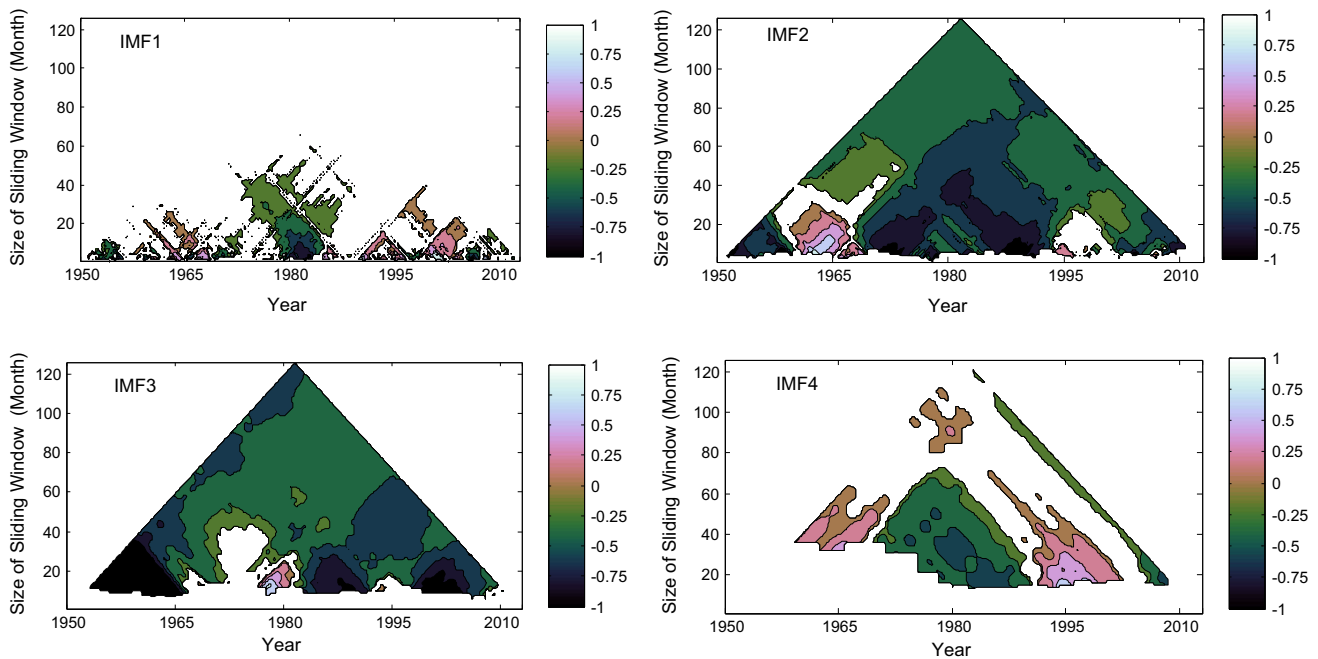


Fig. 8 TDIC plots between IMFs of ENSO and monsoon rainfall. The white space of the TDIC plot means that the correlation coefficient is not significant at 5% significance level

that in 2001–2002, Indian monsoon weakened due to the effect of ENSO, and it is observed that IMF2 and IMF3 shows very strong negative correlation with rainfall. Even though the multiscale relation between ENSO and ISMR is revealed, more investigation is needed to elucidate the exact physical processes behind such relationships and the reasons for the changes in correlation from positive to negative at different time spells.

The TDIC plots of QBO, SN, AMO, and EQUINOO links are presented in Figs. 9, 10, 11, and 12. Figure 9 shows that IMF1 of QBO is negatively correlated with that of AISMR in almost all scales. A direct correlation between QBO and ISMR was observed in the period ~1967–1994 for the IMF2; but in the recent past (~1995–2010), the IMF2 also shows an anti-correlating behavior. For IMF3, the relation between QBO and AISMR is primarily long range negative correlation. However, some significant direct relation between the two events was noticed during short-term spells 1953–55, 1978–80, and 1997–98. From Fig. 10, it is noticed that the high-frequency modes of SN (IMF1 and IMF2) show statistical significance only up to window size of ~60 months. Overall, the correlation between the SN and AISMR is weak and shows that very rich dynamics (i.e., frequent alterations in correlations) in pattern are noticed in different TDIC plots of SN time series.

From Fig. 11, a weak positive correlation is observed between AMO and AISMR in different modes. Interestingly, all the first four modes show a strong direct

correlation with monsoon rainfall in the recent past (~1995–2010). More specifically, it is observed that during 1997–98, there exists a strong correlation between the two in all modes, which supports the earlier findings of the modulation of Indian monsoon by AMO and its link with ENSO (Goswami et al. 2006; Dong et al. 2006). From Fig. 12, it is noticed that most of the modes of EQUINOO show an anti-correlation with the monsoon rainfall, both at shorter or longer scales, but very localized direct correlations are observed between the two at lower time scales of less than 2 years (for e.g., IMF3 in 1950s, IMF1 in 2000s, etc.). The significant negative correlation of IMF1 of EQUINOO and AISMR is quite different from that of other indices considered in this study. In short at a particular time spell, the relation of different climate indices with monsoon rainfall may be quite different; for example, during 1950–1960, for scale ranges <20 months, the correlation between the different modes of AMO and monsoon rainfall, during ~1960–1970 period, for scale ranges 20–40 months, the correlation between the different modes of EQUINOO and monsoon rainfall, etc. This also infers a possible interconnection between the different climate oscillations (such as QBO, ENSO, and AMO) and it can be inferred that their joint influence may decide the fate of monsoon rainfall during different periods. Such linkages need to be investigated further to corroborate more strong inferences. Moreover, the high (positive or negative) correlation between the IMFs of rainfall and climate oscillation at a particular time spell infers that they might be

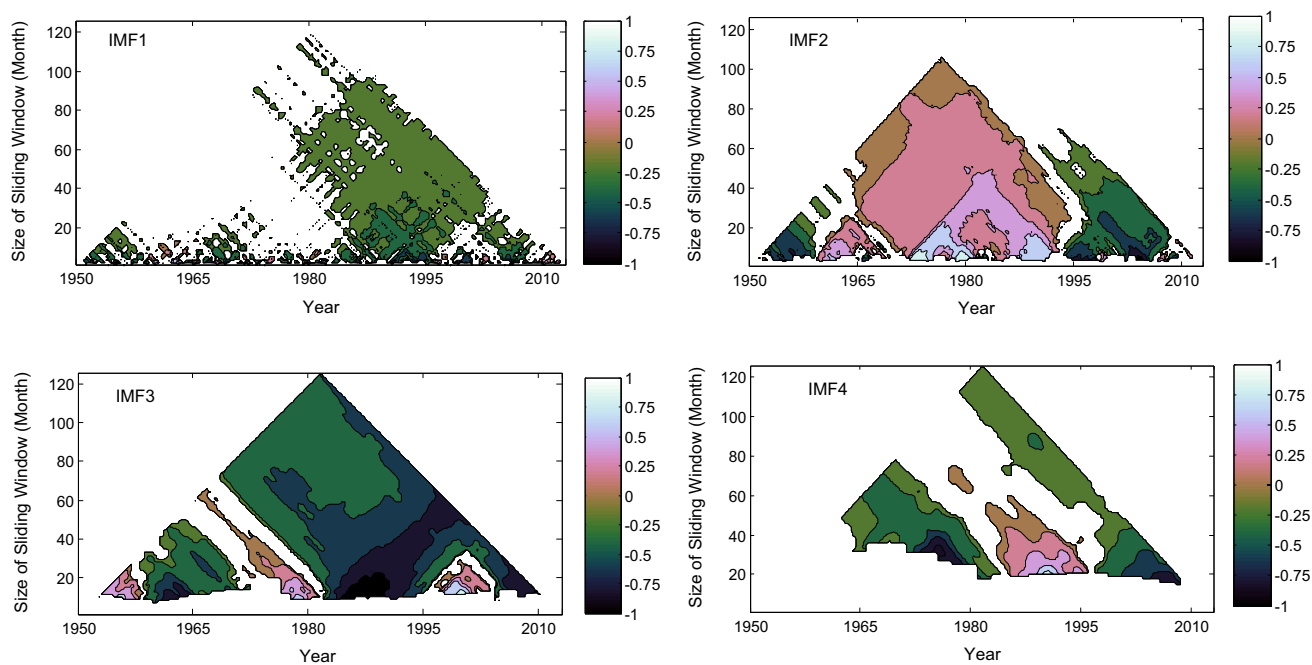


Fig. 9 TDIC plots between IMFs of QBO and monsoon rainfall. The *white space* of the TDIC plot means that the correlation coefficient is not significant at 5% significance level

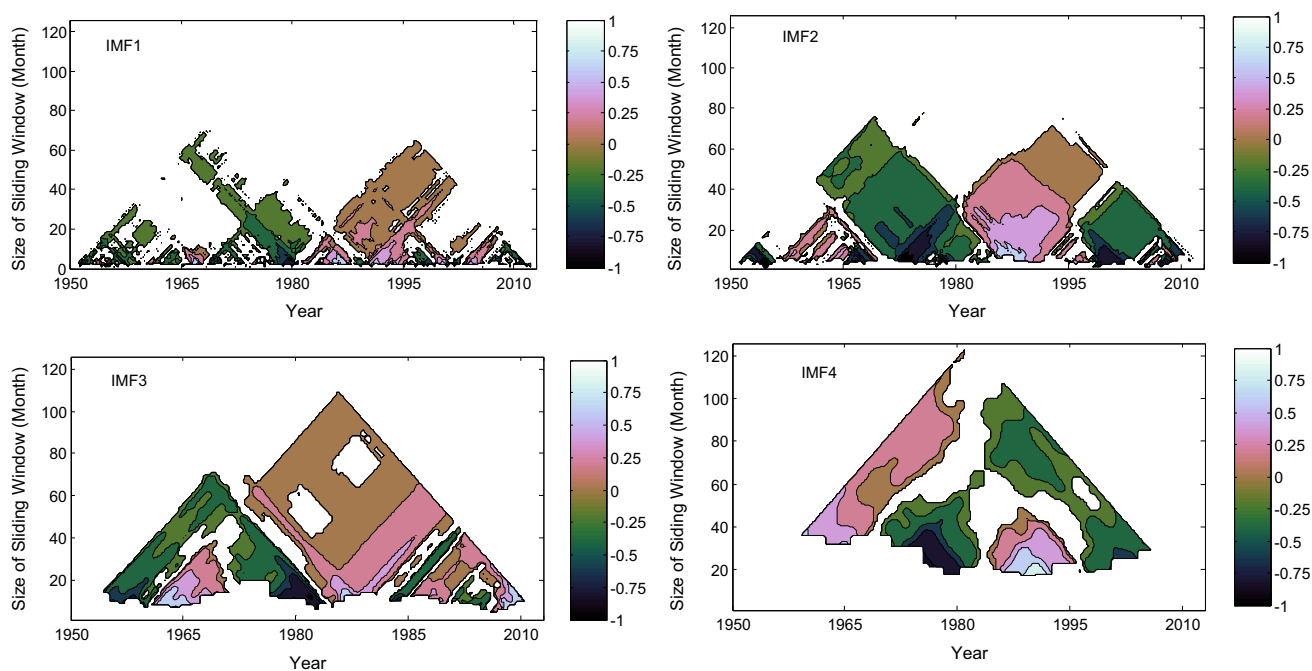


Fig. 10 TDIC plots between IMFs of SN and monsoon rainfall. The *white space* of the TDIC plot means that the correlation coefficient is not significant at 5% significance level

contributed to the same physical processes. Identifying such physical processes and investigating their link with climate indices are challenging task, but it is important for complete understanding of the characteristics of monsoon rainfall. It is to be noted that the present study considered

the TDIC analysis of IMF of monsoon rainfall and one climate index at a time, but the correlation among the different climate indices is not attempted in this study. For performing such analyses, TDIC is a useful method and may give more insight in identifying the physical

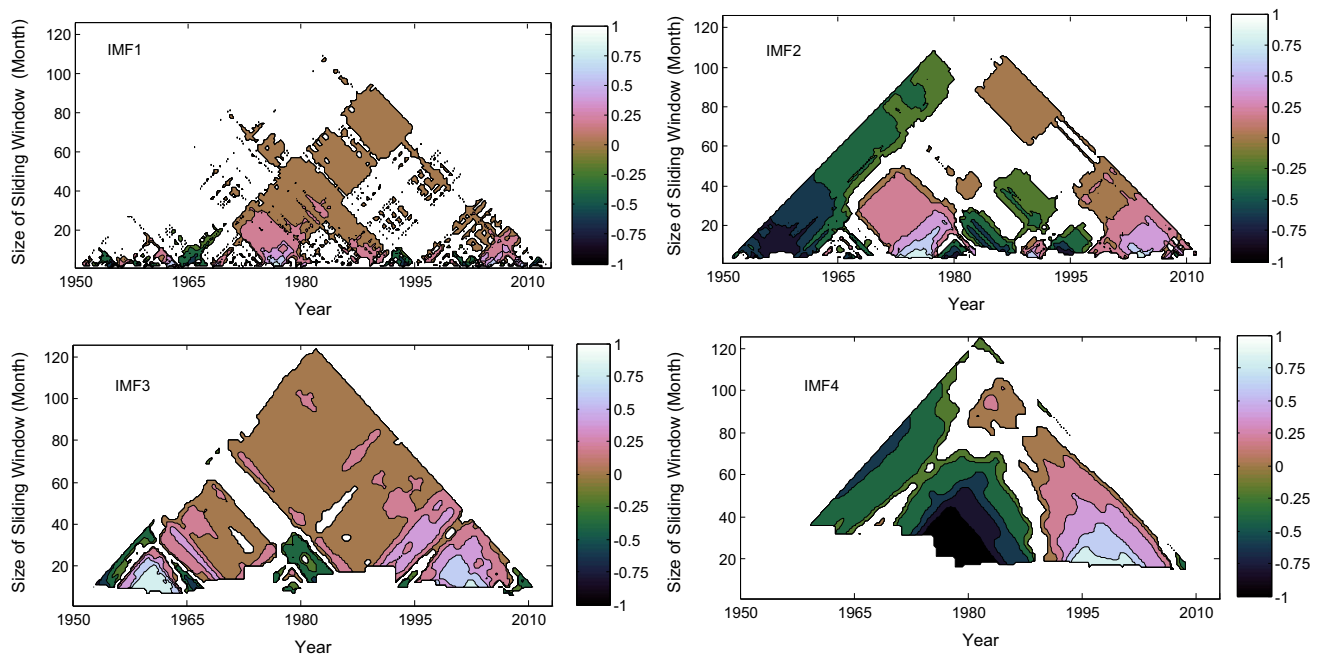


Fig. 11 TDIC plots between IMFs of AMO and monsoon rainfall. The white space of the TDIC plot means that the correlation coefficient is not significant at 5% significance level

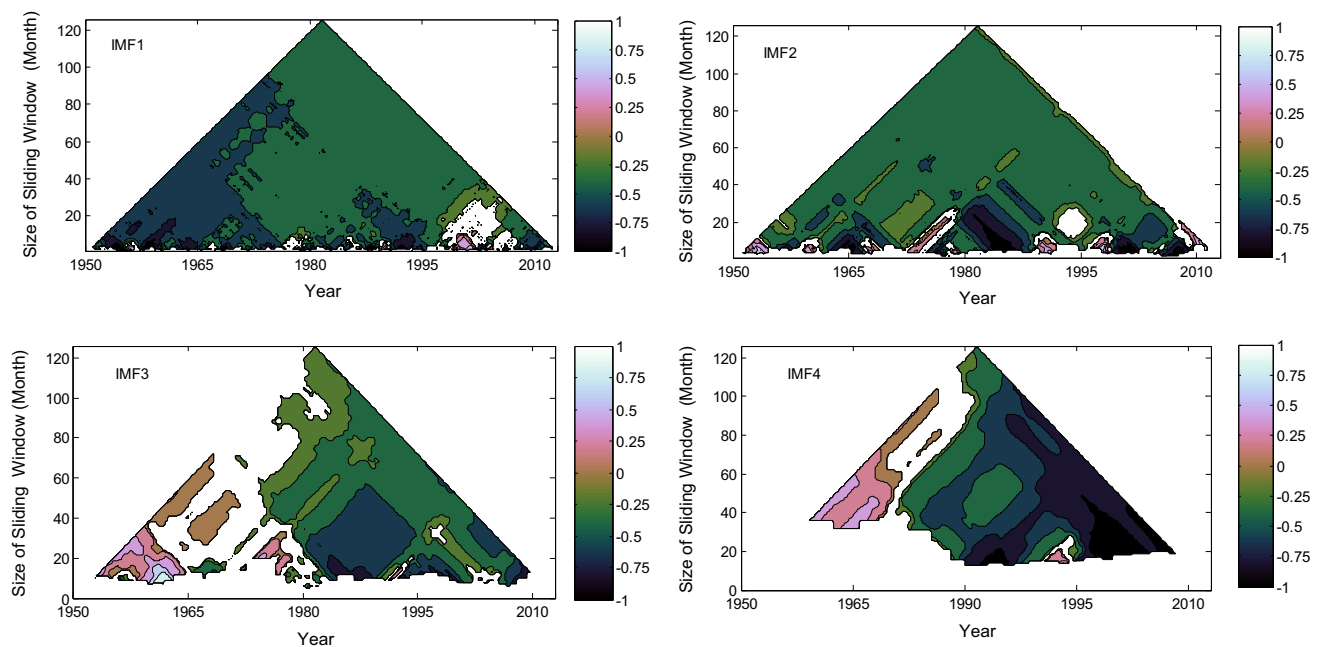


Fig. 12 TDIC plots between IMFs of EQUINOO and monsoon rainfall. The white space of the TDIC plot means that the correlation coefficient is not significant at 5% significance level

mechanisms responsible to such teleconnections. Techniques similar to TDIC (such as wavelet coherency analysis) involve the complex step of appropriate selection of mother wavelet function, which may alter the correlation plots and subsequent interpretations (Grinsted et al. 2004), but TDIC can provide a unique set of plots which decipher

the ‘true’ association between hydro-climatic variables in multiple time scales. Moreover, the wavelet coherency demands smoothing operation, which may alter the quality of representations either in time or frequency domains (Liu 1994; Grinsted et al. 2004). TDIC also solves the complex problem of selection of appropriate scaling window in the

running correlation analysis exercise, and thus, TDIC can be a viable alternative for investigating hydro-climatic teleconnections of ISMR.

4.4 Prediction of AISMR based on multiscale decomposition of climate indices

In the past, various studies have noted that the information of large-scale climate oscillations (or climate indices) can be used as inputs for improved prediction of ISMR (Gadgil et al. 2004; Maity and Nagesh Kumar 2006a, b; Kashid and Maity 2012). In most of such studies, along with the lagged values of rainfall (which can be considered as signatures of physical factors influencing the process), the lagged values of climate indices were used as inputs to predict Indian monsoon rainfall. The HHT-TDIC-based analysis for detecting hydro-climatic teleconnections proved that the nature and strength of association between the rainfall and climatic indices vary over the time scales. It is hypothesized that capturing such information may improve the prediction capabilities of ISMR. For testing this hypothesis, prediction of AISMR time series for the period 1950–2012 is performed using MEMD-SLR approach (whose details are presented in Sect. 2.3). As details of the overall framework for prediction of AISMR are presented in Fig. 2, where MEMD is used for the decomposition of the multivariate data set (comprising lagged inputs of rainfall and climate oscillations) and the stepwise linear regression (SLR) is used for building the regression models for each of the orthogonal components resulting from the decomposition.

The model calibration stage of the MEMD-SLR approach involves the following steps:

- (i) Select candidate predictor variables and appropriate lags by correlation analysis.
- (ii) Perform the decomposition operation of the climate indices and rainfall data using MEMD.
- (iii) Prepare separate models to predict each component of rainfall using SLR.
- (iv) Recombine the predicted orthogonal modes (OM) to get the rainfall value.

For modeling AISMR, the annual series of rainfall and the five climate indices for different monsoon months (JJAS) are considered. In this exercise, first, the past 5 years lagged values of rainfall of the particular month are included as inputs following the suggestions from the past studies (Sahai et al. 2000; Kashid and Maity 2012; Singh and Borah 2012). The correlation of each of the monthly values of climatic indices of the present year (say, for January–May months) with the given month (say June) is computed and the one which gives highest correlation is chosen as model input. Thus, for the prediction of June rainfall, the model is of the form:

$$R_{t\text{June}} = f(R_{\text{June}(t-1)}, \dots, R_{\text{June}(t-5)}, \text{QBO}_{\text{May}}, \text{SN}_{\text{May}}, \text{EQ}_{\text{March}}, \text{AMO}_{\text{January}}, \text{EN}_{\text{May}}).$$

Similarly, the models are considered for prediction of rainfall in July, August, and September months. Then, the seasonal rainfall (monsoon season) can be calculated as follows:

$$R_{\text{monsoon}} = R_{\text{June}} + R_{\text{July}} + R_{\text{August}} + R_{\text{September}}.$$

The data set for 1950–1988 is used for model calibration and that for the rest of the period (1989–2012) is used for validation. For modeling rainfall of a month (say, June), the multivariate data set comprising all input parameters is first decomposed by MEMD method. The maximum and minimum threshold parameters of MEMD are fixed as 0.075 and 0.75, which ensure globally small fluctuations in mean while accounting locally large fluctuations during the ‘sifting’; and the fraction for controlling the sifting iteration is chosen as 0.075, after following the suggestions from past studies (Rilling et al. 2003; Hu and Si 2013). The decomposition resulted in seven IMFs and residue. The SLR models are developed for individual modes of rainfall. For each model, the regression coefficients, the statistical significance (at 5% significance level) was decided based on p value statistics and regression coefficients that fail to meet the criteria are brought to zero, i.e., it can be assessed that such variable is not influential at the respective time scale. The same procedure is repeated for the rainfall of the remaining months (July, August, and September) and the decomposition resulted in six modes for July month series, and seven modes for August and September month series.

To illustrate this procedure, the model of IMF1 and IMF4 is provided below:

$$\text{IMF1}_{\text{Rainfall}} = -2.029 \text{IMF1}_{R_{t-1}} - 2.759 \text{IMF1}_{R_{t-2}} - 2.539 \text{IMF1}_{R_{t-3}} - 1.5971 \text{IMF1}_{R_{t-4}}$$

$$\text{IMF4}_{\text{Rainfall}} = 1.447 \text{IMF4}_{R_{t-1}} - 1.301 \text{IMF4}_{R_{t-2}} + 0.309 \text{IMF4}_{R_{t-3}} - 1.271 \text{IMF4}_{\text{QBO}_{\text{May}}} + 0.062 \text{IMF4}_{\text{EQ}_{\text{March}}}$$

From the above expression, it is clear that none of the climate index is influential at the timescale of IMF1 (as none of the oscillatory mode of climate index is present in the expression). At the time scale of IMF4, only QBO and EQ are influential and similar observations can be made for other oscillatory modes also.

Final summation of the predicted OMs provides the rainfall of June month. Based on the developed models, the OMs of validation data are obtained and the summation of the predicted OMs gives the rainfall during those periods. Similar procedure is followed to predict the rainfall of July, August, and September months. The predictions for the individual months are aggregated to get the seasonal

(monsoon) prediction. To examine the efficacy of the proposed approach, the prediction of rainfall is also performed using the M5 Model Tree technique and multiple linear regression (MLR) for a comparative analysis. Here, M5 model trees (Quinlan 1992) are chosen (from the available different data-driven paradigms for regression) because of its simplicity and superior performance for hydrological predictions as reported in the past studies (Singh et al. 2010; Jothiprakash and Kote 2011). The performance of models is evaluated by estimating the correlation coefficient (R), Mean Square Skill Score (MSSS) (Murphy 1988), and Root Mean Square Error (RMSE) statistics both for the calibration and validation data sets. For validation period (1989–2012), the actual data and IMD operational forecasts are collected from the website of IMD (http://imdpune.gov.in/Clim_Pred_LRF_New/Home/LRF_Perform_89-2014.png). The performance of models for validation period is assessed by computing the different performance measures, which involves comparison of the predictions by different methods (MEMD-SLR, MT, MLR, and IMD operational forecast) with the actual rainfall. The performance evaluation statistics for seasonal (monsoon) rainfall during the calibration period (1950–1988) and validation period are presented in Table 4. For visual illustration, scatter plot and time series plots of monsoon rainfall prediction by different methods are shown in Fig. 13 for the validation period (1989–2012).

For the calibration data, the results show remarkable improvement in performance of MEMD-SLR model in terms of higher correlation (0.973) and smaller error statistics when compared with MT/MLR methods. Recalling that the ENSEMBLE project Multi Model Ensemble (MME) hindcast of All India Rainfall Index (AIRI) has a skill of 0.63 for the period from 1960 to 1988, Physically motivated Empirical model (P-E model) displayed a skill of 0.77 (Wang et al. 2015), and it is clearly evident that model calibration is acceptable for prediction.

The tabulated performance statistics (in Table 4) for predictions by the four methods during the validation period (1989–2012) show that highest correlation skill is displayed by the MEMD-SLR model (0.798), while that of IMD forecast is only -0.12 . It is to be noted that the correlation for predictions for this period (1989–2012) is

0.51 based on the P-E model proposed by Wang et al. (2015). The MSSS is 0.37 for the proposed method, while that is negative for other models, which is an indicative of poor prediction skills. Here, the RMSE of predictions are also computed, which show that the RMSE is least for MEMD-SLR model (55.66) which shows an improvement of 44% over the MT approach and 47% over the IMD operational forecasts. The results presented in Table 4 clearly indicate superior performance of MEMD-SLR model for prediction of ISMR over the other methods during the validation period, which infers that MEMD-SLR has better generalization capabilities over the IMD operational forecasts, and MT and MLR methods.

The time series plot (Fig. 13a) clearly depicts close matching of predictions by MEMD-SLR model with that of observed values, whereas larger deviations can be noticed in the predictions by MT, MLR, and IMD operational forecasts. In the scatter plot (Fig. 13b), the closeness of points towards the ideal fit line clearly indicates the superiority of the proposed MEMD-SLR method. The plots further show that the highest rainfall magnitude (1005.7 in the year 1994) is predicted best by MEMD-SLR model (957.71) when compared with other methods (805.78 mm, 781.64 mm, and 818.8 mm, respectively by MT, MLR, and IMD operational forecast). Similarly, the lowest two rainfall record of 2009 (694.2) is well predicted by MEMD-SLR model (684.34) followed by the other methods (825.91 mm, 820.69 mm, and 827.7 mm by MT, MLR, and IMD forecast, respectively). This also shows that the deviations of predictions of extreme values from the actual values are the least for the MEMD-SLR method. To enable a better comparison, four critical years as referred by Wang et al. (2015) are considered, for which the observed rainfall and the rainfall predicted by different methods are presented in the form of a bar graph in Fig. 14.

Figure 14 shows that the rainfall in all the critical years is well predicted by the proposed method. Apart from the extreme rainfall years of 1994 and 2009, the rainfall for the critical years 2002 and 2004 is also predicted well by the proposed method. The actual rainfall values of these years are 720.9 mm in 2002 and 765.4 mm in 2004, which are predicted as 679.97 mm and 785.61 mm, respectively, by

Table 4 Performance evaluation of rainfall predictions by different methods during calibration (1950–1988) and validation period (1989–2012)

Performance evaluation criteria	Calibration			Validation			
	MEMD-SLR	MT	MLR	MEMD-SLR	MT	MLR	IMD
R	0.973	0.819	0.691	0.798	-0.064	-0.082	-0.120
MSSS	0.943	0.469	0.464	0.371	-0.309	-0.734	-0.361
RMSE	21.612	65.937	66.235	55.668	80.282	92.402	81.853

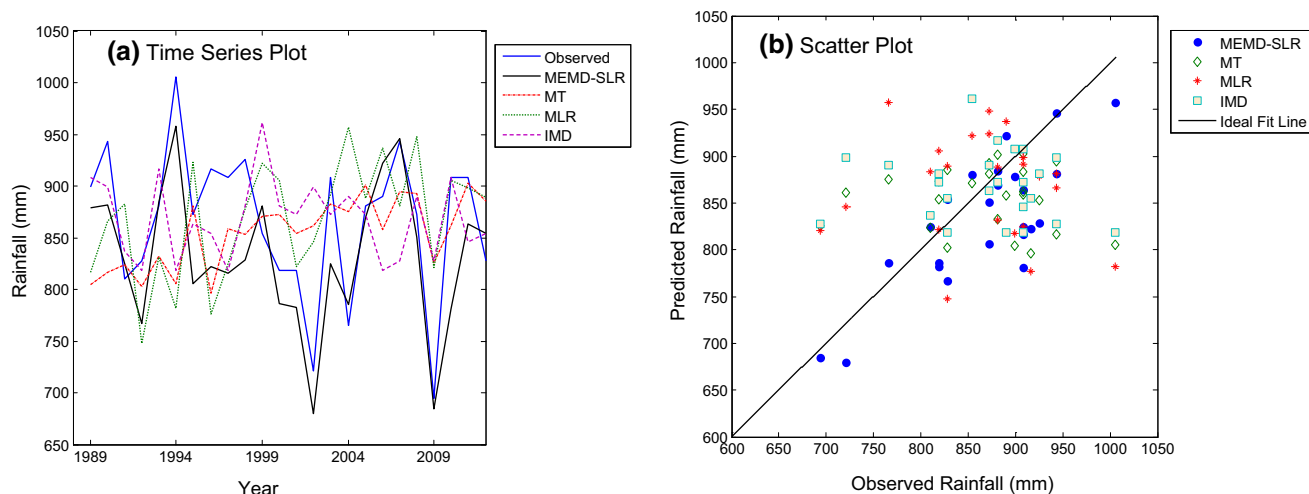
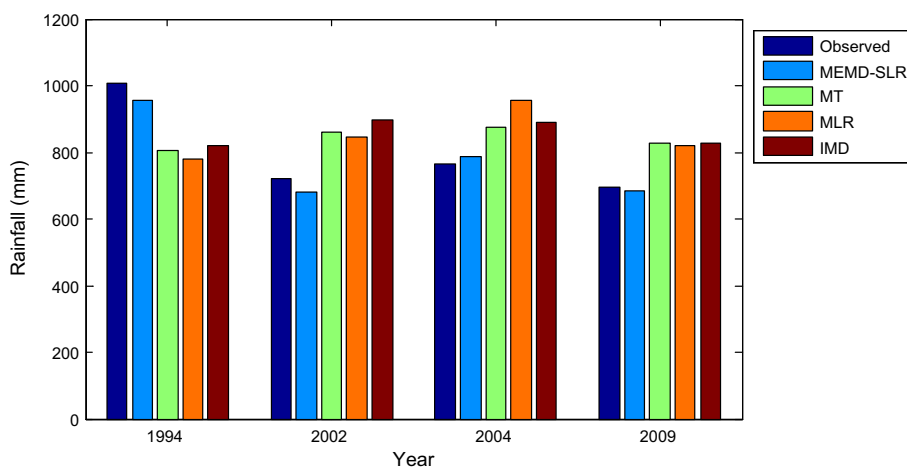


Fig. 13 a Time series plot and b scatter plot of AISMR predictions by different methods for validation period (1989–2012)

Fig. 14 Comparison of rainfall predictions by different methods for four critical years (1994, 2002, 2004, and 2009)



MEMD-SLR method. Furthermore, to check the consistency of predictions by the different methods, standard deviation of actual data (1989–2012) is compared with that obtained by the four methods. The value of standard deviation for observed data was 71.67 mm, while standard deviation values for predicted series by the four methods are 69.02 mm, 33.58 mm, 55.8 mm, and 37.48 mm, which clearly show that the deviations of predictions by MEMD-SLR method are in good agreement with that for observed data.

Moreover, by considering the data set for the years 1989–2005, the correlation skill of predictions is computed, which are found to be 0.81, -0.35 , -0.32 , and -0.22 by the four methods MEMD-SLR, MT, MLR, and IMD forecasts. As per Wang et al. (2015), the correlation skills of ENSEMBLE models are 0.09 for 1989–2005, Asia-Pacific Economic Cooperation (APEC) climate center (APCC) Climate Prediction and its application to Society (CliPAS) models' skill is 0.24, while that of P-E

model is 0.77. In addition, corresponding MSSS skills for ENSEMBLE and CliPAS are -1.32 (1989–2005) and -1.36 (1989–2005), while that of MEMD-SLR, MT, MLR, and IMD forecasts are 0.299, -0.66 , -0.66 , and -0.55 , respectively. These results also clearly indicate that better prediction capabilities of MEMD-SLR model. The presented modeling strategy considers the lagged values of climatic indices as inputs. Hence, this can be treated as medium range forecast with 1 month lead time which has its own importance for planning of agricultural activities and judicious management of available water resources. Overall, maximum correlation for predictions for this period (1989–2012) is 0.51 based on the P-E model and displayed a skill of 0.64 for the 92-year (1921–2012) retrospective forecast (Wang et al. 2015). While the previous studies reported a maximum correlation skill of ~ 0.5 , from statistical and dynamic models for shorter periods (DelSole and Shukla 2009, 2012; Wang et al. 2015) the forecasting skill of the proposed method showed

substantial improvement. In addition, IMD uses six potential predictors for the operational forecasts, while the study by Wang et al. (2015) contributed four complementary predictors for ISMR prediction, which exhibited superiority in prediction skill for the recent decades under a global warming scenario. The present study considers five established climatic indices whose periodic properties are quite similar with the complementary predictors used by Wang et al. (2015) for ISMR prediction. MEMD is quite successful in information capturing from these predictors in different periodic time scales and the SLR is capable to identify the relevant inputs at each of these time scales. This facilitates to retain the potential input and omit the less significant input at different time scales, which cannot be achieved through the conventional modeling methods. This may be the reason behind the superior performance of the proposed methodology. Overall, the proposed strategy involving the ‘decomposition and exclusion’ is found to be a promising modeling practice for ISMR prediction. Substantial improvement in prediction skills based on the use of time scale information from the five climate indices by employing the MEMD-SLR method is a real-value addition for the complex problem of ISMR prediction. It is to be noted that the methodology proposed is a general one, so in future studies, the relative predictive capabilities can be examined systematically by extensive analysis by considering different combinations of climatic indices (those influence the ISMR) as predictor variables.

5 Conclusions

In this study, the multiscale teleconnection between ISMR and different climatic indices is investigated using the HHT-based TDIC analysis. Eventually a MEMD-SLR approach is proposed for effective prediction of ISMR. Specific conclusions of the study are:

- The cross-correlation between orthogonal modes proven that the link between climatic indices and all India monsoon rainfall is expressed well mainly for low-frequency modes and the trend component.
- The TDIC analysis between oscillatory modes of El Niño Southern Oscillation (ENSO) and Indian summer monsoon rainfall successfully captured the overall negative correlation and the localized direct correlations between them.
- The study inferred the existence of strong long range negative correlation between EQUINO and AISMR and positive correlation of different modes of AMO with monsoon rainfall along with the respective short-term counter correlations.

- The association between the climatic oscillations and ISMR varies with time scales and it differs in both nature and strength of the association.
- The proposed MEMD-SLR method facilitates the selection of significant climate indices that are responsible for the variability of rainfall at different time scales, which eventually lead to significant improvement in rainfall forecasts over the IMD operational forecast, MT, and MLR methods for prediction of ISMR. The proposed method performed better than the other methods in predicting the extreme rainfall during the critical years (1994, 2002, 2004, and 2009) and displayed better predictive skill than the dynamical methods and physically motivated empirical (P-E) methods for rainfall prediction during the period 1989–2005.

6 Appendix I

6.1 Multivariate empirical mode decomposition (MEMD)

Multivariate extension of EMD (MEMD) (Rehman and Mandic 2010) decomposes multiple time series simultaneously after identifying the common scales inherent in different time series of concern. A brief description of the MEMD algorithm is presented below:

In this method, multiple envelopes are produced by taking projections of the multiple inputs along different directions in an m -dimensional space.

Assuming $V(t) = \{v_1(t), v_2(t), \dots, v_m(t)\}$ being the m vectors as a function of time t and $X^{\alpha_k} = \{x_1^k, x_2^k, \dots, x_m^k\}$ denoting the direction vector along different directions given by angles $\alpha_k = \{\alpha_1^k, \alpha_2^k, \dots, \alpha_{m-1}^k\}$ in a direction set X ($k = 1, 2, 3, \dots, K$; K is the total number of directions). It can be noted that the rotational modes appear as the counterparts of the oscillatory modes in EMD or its variants. The IMFs of m temporal data sets can be obtained by the following algorithm:

1. Generate a suitable set of direction vectors by sampling on a $(m - 1)$ unit hypersphere
2. Calculate the projection $p^{\alpha_k}(t)$ of the data sets $V(t)$ along the direction vector X^{α_k} for all k
3. Find temporal instants $t_i^{\alpha_k}$ corresponding to the maxima of projection for all k
4. Interpolate $[t_i^{\alpha_k}, V(t_i^{\alpha_k})]$ to obtain multivariate envelope curves $e^{\alpha_k(t)}$ for all k
5. The mean of envelope curves ($M(t)$) is calculated by
$$M(t) = \frac{1}{K} \sum_{k=1}^K e^{\alpha_k(t)}$$

6. Extract the 'detail' $D(t)$ using $D(t) = V(t) - M(t)$. If $D(t)$ fulfills the stoppage criterion (Huang et al. 1998), apply the above procedure to $V(t) - D(t)$, otherwise apply it to $D(t)$.

For the generation of direction vectors, Hammersley sampling sequence was used (Huang et al. 2016).

References

- Adarsh S, Janga Reddy M (2016) Analysing the hydroclimatic teleconnections of summer monsoon rainfall in Kerala, India using multivariate empirical mode decomposition and time dependent intrinsic correlation. *IEEE Geosci Remote Sens Lett* 13(9):1221–1225
- Antil F, Coulibaly P (2004) Wavelet analysis of the inter-annual variability in southern Québec streamflow. *J Clim* 17:163–173
- Antico A, Schlotthauer G, Torres ME (2014) Analysis of hydroclimatic variability and trends using a novel empirical mode decomposition: application to Parana river basin. *J Geophys Res Atmos* 119(3):1219–1233. doi:10.1002/2013JD020420
- Azad S (2011) Extreme Indian monsoon rainfall years and the sunspot cycle. *Adv Sci Lett* 4(1):1–5
- Azad S, Narasimha R, Sett SK (2008) A wavelet based significance test for periodicities in Indian monsoon rainfall. *Int J Wavelets Multi-Resolut Inf Process* 6(2):291–304
- Barnhart BL, Eichinger WE (2011) Empirical mode decomposition applied to solar irradiance, global temperature, sunspot number and CO₂ concentration data. *J Atmos Solar Terr Phys* 73(2011):1771–1779
- Bedrosian E (1963) A product theorem for Hilbert transforms. *Proc IEEE Trans* 51:868–869
- Bhalme HN, Jadhav SK (1984) The double (Hale) sunspot cycle and floods and droughts in India. *Weather* 39:112
- Bhattacharya S, Narasimha R (2007) Regional differentiation in multi-decadal connections between Indian monsoon rainfall and solar activity. *J Geophys Res* 112:D24103. doi:10.1029/2006JD008353
- Campbell WH, Blechman JB, Bryson RA (1983) Long period tidal forcing of Indian monsoon rainfall: a hypothesis. *J Clim Appl Meteorol* 22:287–296
- Chen X, Wu Z, Huang NE (2010) The time-dependent intrinsic correlation based on the empirical mode decomposition. *Adv Adapt Data Anal* 2(2):233–265
- Claud C, Pascal T (2007) Revisiting the possible links between the quasi-biennial oscillation and the Indian summer monsoon using NCEP R-2 and CMAP fields. *J Clim* 20:773–787
- Claud C, Duchiron B, Terray P (2008) On associations between the 11-year solar cycle and the Indian summer monsoon system. *J Geophys Res* 113:D09105. doi:10.1029/2007JD008996
- Cong Z, Chetouani M (2009) Hilbert-Huang transform based physiological signals analysis for emotion recognition. In: *International symposium on in signal processing and information technology (ISSPIT)*, pp 334–339
- DelSole T, Shukla J (2009) Artificial skill due to predictor screening. *J Clim* 22:331–345
- DelSole T, Shukla J (2012) Climate models produce skillful predictions of Indian summer monsoon rainfall. *Geophys Res Lett* 39(L09703):28
- Dhanya CT, Nagesh Kumar D (2010) Nonlinear ensemble prediction of chaotic daily rainfall. *Adv Water Resour* 33(2010):327–347
- Dong B, Sutton RT, Scaife AA (2006) Multidecadal modulation of El Niño-Southern Oscillation (ENSO) variance by Atlantic Ocean sea surface temperatures. *Geophys Res Lett* 33:L08705. doi:10.1029/2006GL025766
- Draper NR, Smith H (1998) *Applied regression analysis*. Wiley, Hoboken, pp 307–312
- Duffy DG (2004) The application of Hilbert Huang transforms to meteorological datasets. *J Atmos Oceanic Technol* 21:599–611
- Fan J, Yao Q (2003) *Non-linear time series: non parametric and parametric methods*. Springer, New York
- Feng S, Hu Q (2008) How the North Atlantic multidecadal oscillation may have influenced the Indian summer monsoon during the past two millennia. *Geophys Res Lett* 35:L01707. doi:10.1029/2007GL032484
- Franceschini S, Tsai CW (2010) Application of Hilbert-Huang transform method for analyzing toxic concentrations in the Niagara river. *J Hydrol Eng* 15(2):90–96
- Gadgil S, Vinayachandran PN, Francis PA, Gadgil S (2004) Extremes of the Indian summer monsoon rainfall, ENSO and equatorial Indian Ocean oscillation. *Geophys Res Lett* 31:L12213. doi:10.1029/2004GL019733
- Goswami BN, Madhusoodanan MS, Neema CP, Sengupta D (2006) A physical mechanism for North Atlantic SST influence on the Indian summer monsoon. *Geophys Res Lett* 33(2):L02706. doi:10.1029/2005GL024803
- Grinsted A, Moore JC, Jevrejeva S (2004) Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Process Geophys* 11(5/6):561–566
- Hahn DG, Shukla J (1976) An apparent relationship between Eurasian snow cover and Indian monsoon rainfall. *J Atmos Sci* 33:2461–2462
- Hu W, Si BC (2013) Soil water prediction based on its scale-specific control using multivariate empirical mode decomposition. *Geoderma* 193–194:180–188
- Huang Y-X (2013) Hilbert-Huang transform in ocean turbulence. In: *EGU general assembly 2013, held 7–12 April 2013, Vienna, ID EGU2013-9900*
- Huang Y, Schmitt FG (2014) Time dependent intrinsic correlation analysis of temperature and dissolved oxygen time series using empirical mode decomposition. *J Mar Syst* 130(2014):90–100
- Huang NE, Wu Z (2008) A review on Hilbert Huang transform: method and its applications to geophysical studies. *Rev Geophys* 46(2):RG2006. doi:10.1029/2007RG000228
- Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond Ser A* 454:903–995
- Huang Y, Schmitt FG, Lu Z, Liu Y (2009a) Analysis of daily river flow fluctuations using empirical mode decomposition and arbitrary order Hilbert spectral analysis. *J Hydrol* 454:103–111
- Huang NE, Wu Z, Long SR, Arnold KC, Blank K, Liu TW (2009b) On instantaneous frequency. *Adv Adapt Data Anal* 1(2):177–229
- Huang G, Su Y, Kareem A, Liao H (2016) Time-frequency analysis of non-stationary process based on multivariate empirical mode decomposition. *J Eng Mech* 142. doi:10.1061/(ASCE)EM.1943-7889.0000975
- Ismail DKB, Lazure P, Puillat I (2015) Advanced spectral analysis and cross correlation based on empirical mode decomposition: application to the environmental time series. *Geosci Remote Sens Lett* 12(9):1968–1972
- Iyengar RN, Raghu Kanth TSG (2005) Intrinsic mode functions and a strategy for forecasting Indian monsoon rainfall. *Meteorol Atmos Phys* 90:17–36

- Janga Reddy M, Adarsh S (2016) Time frequency characterization of subdivisional scale seasonal rainfall in India using Hilbert Huang transform. *Stoch Environ Res Risk Assess* 30(4):1063–1085
- Jothiprakash V, Kote AS (2011) Effect of pruning and smoothing while using M5 model tree technique for reservoir inflow prediction. *J Hydrol Eng* 16(7):563–574
- Kashid SS, Maity R (2012) Prediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation patterns using genetic programming. *J Hydrol* 454(2012):26–41
- Kripalani RH, Kulkarni A (1997a) Rainfall variability over south East Asia—connections with Indian monsoon and ENSO extremes: new perspectives. *Int J Climatol* 17:1155–1168
- Kripalani RH, Kulkarni A (1997b) Climatic impacts of El Niño/La Niña on the Indian monsoon: a new perspective. *Weather* 52:39–46
- Kripalani RH, Kulkarni A (1999) Climatology and variability of historical Soviet snow depth data: some new perspectives in snow—Indian monsoon teleconnections. *Clim Dyn* 15:475–489
- Krishna Kumar KB, Rajagopalan B, Cane MA (1999) On the weakening relationship between the Indian monsoon and ENSO. *Science* 284:2156–2159
- Kuai KZ, Tsai CW (2012) Identification of varying time scales in sediment transport using the Hilbert-Huang transform method. *J Hydrol* 420–421:245–254
- Kumar KK, Rajagopalan B, Hoerling M, Bates G, Cane MA (2006) Unraveling the mystery of Indian monsoon failure during El Niño. *Science* 314:115–119
- Labat D (2005) Recent advances in wavelet analyses: Part 1. A review of concepts. *J Hydrol* 314:275–288
- Liu PC (1994) Wavelet spectrum analysis and ocean wind waves. In: Foufoula-Georgiou E, Praveen Kumar M (eds) *Wavelets in geophysics*, pp 151–166
- Lu R, Dong B, Ding H (2006) Impact of the Atlantic multidecadal oscillation on the Asian summer monsoon. *Geophys Res Lett* 33:L24701. doi:10.1029/2006GL027655
- Maity R, Nagesh Kumar D (2006a) Bayesian dynamic modeling for monthly Indian summer monsoon rainfall using El Niño–Southern Oscillation (ENSO) and Equatorial Indian Ocean Oscillation (EQUINO). *J Geophys Res* 111:D07104. doi:10.1029/2005JD006539
- Maity R, Nagesh Kumar D (2006b) Hydro-climatic association of monthly summer monsoon rainfall over India with large-scale atmospheric circulation from tropical Pacific Ocean and Indian Ocean. *Atmos Sci Lett* 7(4):101–107
- Mamgain A, Dash SK, Parth sarthi P (2010) Characteristics of Eurasian snow depth with respect to Indian summer monsoon rainfall. *Meteorol Atmos Phys* 110:71–83
- Massei N, Fournier M (2012) Assessing the expression of large scale climatic fluctuations in the hydrologic variability of daily Seine river flow (France) between 1950–2008 using Hilbert Huang Transform. *J Hydrol* 448(2012):119–128
- Massei N, Durand A, Deloffre J, Dupont J, Valdes D, Laignel B (2007) Investigating possible links between the North Atlantic oscillation and rainfall variability in north western France over the past 35 years. *J Geophys Res Atmos* 112:1–10
- Mooley DA, Parthasarathy B (1983) Indian summer monsoon and El Niño. *Pageoph* 121:339–352
- Murphy AH (1988) Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon Weather Rev* 116:2417–2424
- Nagesh Kumar D, Janga Reddy M, Maity R (2007) Regional rainfall forecasting using large scale climate teleconnections and artificial intelligence techniques. *J Intell Syst* 16(4):307–322
- Narasimha R, Bhattacharyya S (2010) A wavelet cross-spectral analysis of Solar-ENSO-rainfall connections in the Indian monsoons. *Appl Comput Harmon Anal* 28:285–295
- Narasimha R, Kailas SV (2001) A wavelet map of monsoon variability. *Proc Indian Natl Sci Acad* 67(3):327–341
- Nuttall AH (1966) On the quadrature approximation to the Hilbert transform of modulated signals. *Proc IEEE* 54:1458–1459
- Papadimitriou S, Sun J, Yu PS (2006) Local correlation tracking in time series. In: *Proceedings of IEEE sixth international conference on data mining*, 18–22 December 2006, Hong Kong, pp 456–465
- Parthasarathy B, Pant GB (1984) The spatial and temporal relationships between the Indian summer monsoon rainfall and the Southern oscillation. *Tellus* 36A:269–277
- Quinlan JR (1992) Learning with continuous classes. In: *Proceedings of Australian joint conference on artificial intelligence*. World Scientific Press, Singapore, pp 343–348
- Rahman Md. A, Chetty M, Bulach D, Wangikar PP (2015) Frequency decomposition based gene clustering. *neural information processing*. In: *Lecture notes in computer science*, pp 170–181
- Rao RKS, Lakhole NJ (1978) Quasi-biennial oscillation and summer southwest monsoon. *Indian J Meteorol Hydrol Geophys* 29:403–411
- Rehman N, Mandic DP (2010) Multivariate empirical mode decomposition. *Proc R Soc* 466:1291–1302
- Rilling G, Fladrin P, Goncalves P (2003) On empirical mode decomposition and its algorithms. In: *Proceedings of IEEE-EURASIP workshop on nonlinear signal and image processing NSIP-03*, Grado, pp 8–11
- Rodo X, Rodriguez-Arias MA (2006) A new method to detect transitory signatures and local time/space variability structures in the climate system: the scale-dependent correlation analysis. *Clim Dyn* 27:441–458
- Sahai AK, Soman MK, Satyan V (2000) All India summer monsoon rainfall prediction using an artificial neural network. *Clim Dyn* 16:291–302
- Scafetta N (2014) Multi-scale dynamical analysis (MSDA) of sea level records versus PDO, AMO, and NAO indexes. *Clim Dyn* 43:175–192
- Shukla J, Paolino DA (1983) The southern oscillation and long-range forecasting of the summer monsoon rainfall over India. *Monsoon Weather Rev* 111:1830–1837
- Singh P, Borah B (2012) Indian summer monsoon rainfall prediction using neural network. *Stoch Environ Res Risk Assess* 27(7):1585–1599
- Singh KK, Pal M, Singh VP (2010) Estimation of mean annual flood in Indian catchments using back-propagation neural network and M5 model tree. *Water Resour Manag* 24:2007–2019
- Torrence C, Webster PJ (1999) Inter-decadal changes in the ENSO-monsoon system. *J Clim* 12:2679–2690
- Torres ME, Colominas MA, Schlotthauer G, Fladrin P (2011) A complete ensemble empirical mode decomposition with adaptive noise. *IEEE Int Conf Acoust Speech Signal Process Prague* 22–27:4144–4147
- Usoskin IG, Mursula K (2003) Long-term solar cycle evolution: review of recent developments. *Sol Phys* 218:319–343
- Vijayakumar R, Kulkarni R (1995) The variability of the inter-annual oscillations of the Indian summer monsoon rainfall. *Adv Atmos Sci* 12(1):95–102
- Walker G (1923) Correlation in seasonal variations of weather, VII: a preliminary study of world weather. *Mem India Meteorol Dep* 24(4):75–131
- Wang C, Picaut J (2004) Understanding ENSO physics—a review. In: Wang C, Xie S-P, Carton JA (eds) *Earth’s climate: the ocean–atmosphere interaction*. Geophysical monograph series, vol 147. AGU, Washington, D.C., pp 21–48

- Wang B, Xiang B, Li J, Webster PJ, Rajeevan MN, Liu J, Kuung-Ja H (2015) Rethinking Indian monsoon rainfall prediction in the context of recent global warming. *Nat Commun* 6 (article number 7154). doi:[10.1038/ncomms8154](https://doi.org/10.1038/ncomms8154)
- Zhang R, Delworth TL (2006) Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. *Geophys Res Lett* 33:L17712. doi:[10.1029/2006GL02626](https://doi.org/10.1029/2006GL02626)