

# Complexity and periodicity of daily mean temperature and dew-point across India

RAJDEEP RAY<sup>1,\*</sup>, MOFAZZAL HOSSAIN KHONDEKAR<sup>2</sup>, KOUSHIK GHOSH<sup>3</sup>  
and ANUP KUMAR BHATTACHARJEE<sup>4</sup>

<sup>1</sup>Department of Electronics and Communication Engineering, Dr. B C Roy Engineering College, Durgapur 713 206, India.

<sup>2</sup>Department of Applied Electronics and Instrumentation Engineering, Dr. B C Roy Engineering College, Durgapur 713 206, India.

<sup>3</sup>Department of Mathematics, University Institute of Technology, University of Burdwan, Burdwan 713 104, India.

<sup>4</sup>Department of Electronics and Communication Engineering, National Institute of Technology, Durgapur 713 209, India.

\*Corresponding author. e-mail: ray.rajdeep78@gmail.com

MS received 18 August 2017; revised 17 January 2019; accepted 4 February 2019; published online 31 May 2019

The complexity of temperature and dew-point fluctuations across India are being investigated and analyzed with the help of recurrence plots (RP) and recurrence quantification analysis (RQA). The results firmly state that both data sets are non-linear, non-stationary and deterministic. Hilbert–Huang transform and an efficient peak detection algorithm (integral method) have been used to detect the underlying periodicity (above 95% CL) within these two signals. The nature of the complexity and the derived periods of these two weather variables show that there is a significant impact of the local geographical topology and global atmospheric fluctuations on the overall pattern and fluctuation of the temperature and dew-point profile across India.

**Keywords.** Indian climate; recurrence plot; Hilbert–Huang transform; empirical mode decomposition.

## 1. Introduction

Weather parameters such as temperature, dew-point, rainfall, atmospheric pressure, and wind speed are complicated in nature (Byakatonda *et al.* 2018; Reddy *et al.* 2018). The time series of these parameters need to be analyzed carefully as the dynamics of processes generating these time series are a priori unknown. Analysis to reveal the dynamics of the processes and to predict their future behaviour precisely is a matter of thorough study. These time series usually contain linear or non-linear, stationary or non-stationary, deterministic

or stochastic components, yet it is a common practice to analyze such processes using suboptimal, but mathematically tractable methods, which may lead towards a flawed verdict about the processes. Appropriate analyzing techniques are essential and can significantly improve the quality of the results revealing the important information which remained buried earlier while using conventional linear methods. There are numerous data analysis techniques to obtain insights into the complex processes observed in nature (Blasius *et al.* 1999; Glass 2001; Marvel 2001). The large size of India, its latitudinal extent, the presence of the Himalayas in the

north, the Indian Ocean, Arabian Sea and Bay of Bengal in the south and Thar desert in west have resulted in great variations in the distribution of weather variables in the subcontinent of India and hence in this paper an initiative have been taken to learn the dynamical aspects of the Indian weather.

The key meteorological parameters which influence the weather of a geographical region are (a) temperature, (b) pressure, (c) wind speed, (d) dew-point or humidity, (e) rainfall; (f) haze and contents in the air, and (g) solar-terrestrial radiation. Mere measurement of these meteorological parameters by the instruments (however sophisticated they might be) is of no importance to the human civilization unless proper analysis and investigation of these measured or recorded parameters are made. Human civilization got its momentum when the technology of transport and communication (ground, aerial or marine traffic) started to flourish and the maintenance of these cannot be made without proper analysis of the meteorological parameters. The analysis of these parameters are also important as far as the aspects like atmosphere pollution, non-conventional energy generation, agronomy, forest conservation, urban and rural planning, and irrigation management are concerned. The dynamics of the variation of the air temperature and humidity play the crucial role to assess the chemical reaction associated with the air pollutants.

The outcome of the analysis of the temporal or spatial data series of the combination of a number of the meteorological parameters is vital for weather prediction, modelling and climatology. This research work which involves the analysis of the two important meteorological parameters, temperature and dew-point recorded at various location across India, is an effort to unearth the complex dynamics and the embedded periodicities of the two parameters. The outcome of this analysis may further contribute as the necessary inputs to researchers, meteorologists who are working in various missions discussed above. Keeping these in view, two important parameters temperature and dew-point have been chosen as the subjects of investigation in this work.

The daily mean temperature and dew-point data (from 9 October, 1996 to 1 February, 2013) collected in the form of time series from seven weather stations (Kolkata, Chennai, New Delhi, Mumbai, Bhopal, Agartala, Ahmedabad) (Ray *et al.* 2015) distributed across India (NOAA 2014) have been taken for investigation as they have an effect on

Temperature–Humidity Index (THI), which is a measure of ‘discomfort’ (Steadman 1984). The length of the time series data sets of the daily mean temperature and dew-point for every station is the total number of days counting from 9 October, 1996 to 1 February, 2013 and it is 5960. The summary statistics of the time series signals used in this work are presented in table 1.

The dew point is the temperature below which the water vapour in air at constant barometric pressure condenses into liquid water at the same rate at which it evaporates. The dew-point is associated with relative humidity. A high relative humidity indicates that the dew point is closer to the current air temperature, whereas when it is much lower than the air temperature, the relative humidity is very low. Temperature of a place also varies at different times of a day and at different months and seasons of a year.

The variations of terrestrial temperature may be due to several factors. The processes of absorption, scattering and transmission of the incoming solar radiation directly affect the distribution of the temperature throughout the atmosphere (Ahmed *et al.* 2014). The angle of the sun’s rays and length of daylight determine the amount of insolation received in a certain place. Near-infrared radiation, which represents nearly half the radiation emitted by the Sun, is absorbed mainly by two gases in the atmosphere–water vapour and (to a lesser extent) carbon dioxide. Therefore, temperature and dew-point may have an inter-dependent relationship. It is a quest to the climate scientist to ascertain the character of this relationship for which statistical methodologies have been used. In this paper the authors devoted their effort in this search only for the weather prevailing in India.

The time series generated from natural systems generally exhibit non-linear, non-stationary and complex behaviour (Marwan and Kurths 2004). Climatic behaviour is one such natural aspect which is also supposed to exhibit complexity. A system with complexity is one where different intermingling components are found to interact with each other. Climate system definitely suits by this definition. The very identification of the complex behaviour of the climate can bring improvement in weather predictions and understanding. The temperature and the dew-point time series (NOAA 2014) of seven weather stations distributed across India are analyzed here in the context of its non-linear, non-stationary and complex behaviour (Marwan and Kurths 2004). In this

Table 1. Summary statistics of the daily mean temperature and dew point of the seven weather stations.

	Kolkata	Chennai	New Delhi	Mumbai	Bhopal	Agartala	Ahmedabad
<i>Daily mean temperature (°F)</i>							
Mean	79.35	82.92	76.28	81.67	77.52	76.75	80.93
Median	82.51	86.38	80.10	82.16	78.13	80.07	82.59
SD	7.19	4.15	12.51	3.08	8.63	7.04	7.80
Variance	51.75	17.22	156.52	9.51	74.52	49.56	60.86
Maximum	92.11	94.30	99.80	89.13	98.08	88.64	95.84
Minimum	58.37	73.99	47.19	68.13	54.79	53.13	60.68
Range	33.74	20.31	52.61	21	43.29	35.51	35.16
Skewness	−0.85	0.03	−0.39	−0.91	0.00	−0.89	−0.40
Kurtosis	−0.52	−1.04	−1.16	1.14	−0.63	−0.50	−0.81
<i>Daily mean dew point (°F)</i>							
Mean	69.82	72.38	61.05	68.56	57.50	69.66	60.62
Median	73.32	72.92	59.16	70.49	55.45	72.75	59.77
SD	8.75	2.75	11.27	6.79	11.31	8.19	11.78
Variance	76.51	7.55	127.10	46.04	123.83	67.15	138.86
Maximum	80.57	79.69	80.32	79.19	75.72	80.16	81.29
Minimum	45.36	62.75	31.86	44.30	29.97	46.98	25.59
Range	35.21	16.94	48.46	34.89	45.75	33.18	55.70
Skewness	−0.64	−0.68	0.13	−0.48	0.05	−0.67	−0.13
Kurtosis	−1.02	0.15	−1.29	−0.96	−1.30	−0.95	−1.25

framework recurrence plot (Eckmann *et al.* 1987) and recurrence quantification analysis (RQA) (Marwan *et al.* 2007) methodologies have been implemented to study the stationarity/non-stationarity, nature of cyclicity, determinism/stochastism, stability, chaosity, complexity and regularity of the temperature and dew-point of India (Ray *et al.* 2017).

In order to find out the temporal regularities within the time series the underlying periodicity needs to be detected, which reveals how the periodic patterns evolve within the time series (Dergachev and Geel 2004). The primary reason which makes the search for periodicity an area of fundamental research is to get an idea about the possible cause of their dynamics. Investigation and comparison of the patterns of these periodicities and their evolution helped to reveal the mutual relationship between the different meteorological and climatological parameters. Identification of the patterns of these periodicities of these parameters also provides necessary contribution in climatological forecasting and in decision making of various public welfare activities throughout the year of a particular location. Hilbert–Huang transform (Huang *et al.* 1998) using empirical mode decomposition method (Sole *et al.* 2007) has been employed to find the underlying periodicities within the signals under investigation, i.e., daily

mean temperature and dew-point of seven weather station across India. The intrinsic mode functions yield instantaneous frequencies as functions of time that give sharp identifications of embedded structures. Among the obtained instantaneous periods (1/frequency), the significant periods having more than 95% confidence level are picked by an efficient peak detection algorithm (integral method) (Azzini *et al.* 2004). These periods of the temperature and dew-point are compared to trace the presence of statistical association between them.

## 2. Theory

### 2.1 Recurrence plot (RP) and recurrence quantification analysis (RQA)

Recurrence plots can help to find out a first characterization of the data or to find transitions and interrelations. These techniques are applicable to any time series as they do not require *a priori* assumptions on the statistical properties or mathematical structure of the time series. Recurrence is a fundamental property of dynamical systems. After some time, the system will reach a state that is arbitrarily close to the former state and pass through a similar evolution. Recurrence plots visualize such recurrent behaviour of

dynamical systems. The degree of recurrence and other important measures of complexity are then estimated using recurrence quantification analysis (RQA).

### 2.1.1 Recurrence plot

Recurrence plot (RP) is a graphical tool introduced by Eckmann *et al.* (1987) in order to extract qualitative characteristics of a time series. It exhibits characteristic patterns for typical dynamical behaviour of a signal or time series (Khondekar *et al.* 2012). For example, a collection of single recurrence points, homogeneously and irregularly distributed over the whole plot, reveals stochastic behaviour of the signal.

It shows all the instances when the phase space trajectory (Thomasson *et al.* 2001; Marwan *et al.* 2002) of the system visits almost the same region in the phase space. The recurrence of a state  $i$  at a different time  $j$  or the  $d$ -dimensional phase space trajectory is depicted within a two-dimensional squared matrix with black and white dots, where the black dots represent a recurrence and both axes represent time. Such an RP can be mathematically expressed as

$$R_{i,j}^{d,\varepsilon_i} = H(\varepsilon_i - \|s_i - s_j\|), \quad s_i \in \mathbb{R}^d, \quad i, j = 1, \dots, N \quad (1)$$

where  $R_{i,j}$  is the recurrence plot,  $N$  is the number of considered states  $S_i$ ,  $\varepsilon_i$  is a threshold distance,  $\mathbb{R}$  is set of real numbers,  $\|s_i - s_j\|$  is the norm and  $H(*)$  is the Heaviside step function.

Using the one-dimensional time series  $x_i$  of a single observable variable (temperature or dew-point in this case), it is possible to reconstruct a phase space trajectory  $s_i$  using Taken's time delay method (Takens 1981) as,  $s_i = (x_i, x_{i+\tau}, \dots, x_{i+(d-1)\tau})$ , where  $d$  is the embedding dimension and  $\tau$  is the time delay (index based). The embedding dimension  $d$  is estimated by using the method based on false nearest neighbours (Kennell *et al.* 1992). The basic philosophy behind the false nearest neighbours method is that on decreasing the embedding dimension  $d$ , increasing number of neighbouring points will be projected in the phase space, even they may not be a real neighbour to the actual point. These points are known as false nearest neighbours (FNN). To achieve the minimal amount of embedding, the embedding dimension  $d$

is to be chosen in such a way that all the FNNs vanish (Kantz and Schreiber 1997).

The time delay is linearly proportional to embedding lag  $\tau$  and defined as equal to  $\tau$  times the sampling rate of the data series. Since the sampling rate of the observation in present work is 1 day for each time series, the time delay equals to embedding lag  $\tau$ . The optimal value of  $\tau$  is determined using the theory of mutual information (Roulston 1999). The best possible value of  $\tau$  is that value for which the mutual information between any observation and its  $\tau$ -delayed reconstruction in the phase space have lowest local minimum. The estimated values of  $d$  and  $\tau$  are shown at the bottom right corner of table 2 for temperature and dew-point data series for all the seven weather stations. Using these values of  $d$  and  $\tau$ , the corresponding recurrence plots of temperatures and dew-points of these stations are obtained as depicted in table 2.

### 2.1.2 Recurrence quantification analysis (RQA)

A quantitative analysis of the recurrence plots that measure the dynamical transitions in complex systems based on the recurrence point density, diagonal and vertical structures in the RP is known as recurrence quantification analysis (RQA) (Marwan *et al.* 2002, 2007; Zbilut and Jr. Webber 1992; Jr. Webber and Zbilut 1994; Khondekar *et al.* 2013). The parameters which are generally calculated from RP are recurrence rate (REC), determinism (DET), the longest diagonal line segment in the plot (LMAX) and entropy (ENTR).

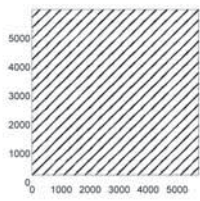
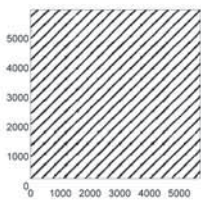
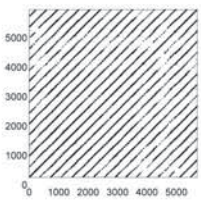
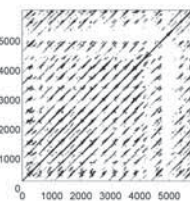
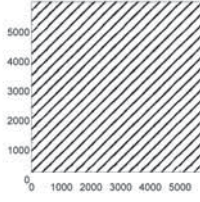
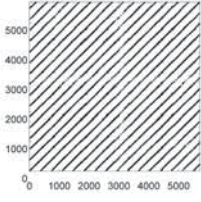
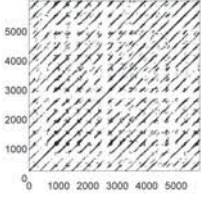
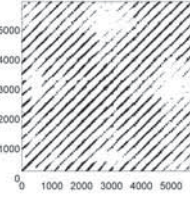
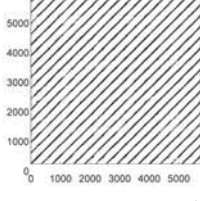
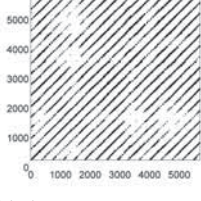
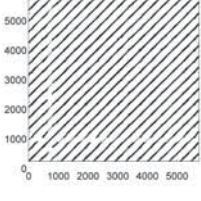
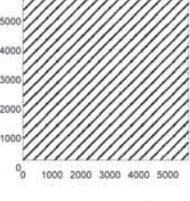
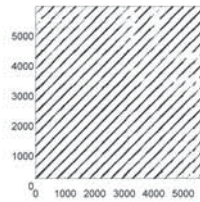
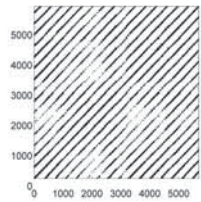
The recurrence rate is a measure of recurrences, or density of recurrence points in the RP. This rate gives the mean probability of recurrences in the system. The recurrence rate represents the fraction of recurrent points with respect to the total number of possible recurrences. It is a density measure of the RP. The expression for REC is as below

$$REC(\varepsilon_i) = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}(\varepsilon_i). \quad (2)$$

Determinism is a measure for predictability of the system. The determinism could also be explained as the percentage of recurrent points forming line segments, which are parallel to the line of identity (LOI). The measure of determinism (DET) ranges from 0 to 1. Numbers near zero indicate randomness while those approaching



Table 2. Recurrence plots of temperature and dew point of the seven weather stations and the corresponding embedding dimension ( $d$ ) and embedding lag ( $\tau$ ).

Temperature		Dew point		Temperature		Dew point	
Kolkata				Chennai			
							
New Delhi				Mumbai			
							
Bhopal				Agartala			
							
Ahmedabad							
		Temperature		Dew Point			
		( $d$ )	( $\tau$ )	( $d$ )	( $\tau$ )		
Kolkata		5	55	5	61		
Chennai		5	64	5	42		
New Delhi		4	63	5	63		
Mumbai		5	40	5	63		
Bhopal		5	53	5	65		
Agartala		5	55	5	61		
Ahmedabad		5	56	5	67		

one indicate the presence of a strong deterministic component in the signal. If a variable  $D_{ij}$  be defined as

$$D_{i,j} = \begin{cases} 1, & \text{if } (i, j) \text{ and } (i+1, j+1) \text{ or } (i-1, j-1) \text{ are recurrent} \\ 0, & \text{otherwise} \end{cases}$$

Then

$$DET = \frac{\sum_{i,j=1}^N D_{i,j}}{\sum_{i,j=1}^N R_{i,j}}. \quad (3)$$

The  $LMAX$  is simply the length of the longest diagonal line segment in the plot, excluding the main diagonal line of identity (LOI) where  $i = j$ . For the shorter  $LMAX$ , the signal can be judged as more chaotic and less stable while for longer

$LMAX$ , the signal can be evaluated as a more regular one with greater stability.

If  $N_l$  is the number of diagonal lines and  $l_i$  be the length of the  $i^{\text{th}}$  diagonal line

$$LMAX = \max(l_i), \quad (4)$$

where  $i = 1, \dots, N_l$

The entropy ( $ENTR$ ) refers to the Shannon entropy of the frequency distribution of the diagonal line lengths. According to several authors, the basic idea is that information (Shannon) entropy of the random processes is abundantly supplied with the qualitative and quantitative data on the object

Table 3. *RQA of temperature and dew point series for seven stations.*

Stations	Temperature				Dew Point			
	%REC	DET	LMAX	ENTR	%REC	DET	LMAX	ENTR
Kolkata	19.75	1.00	38.76	3.54	22.80	0.99	26.76	3.27
Chennai	6.96	0.99	10.28	2.87	1.39	0.93	4.93	1.78
New Delhi	46.15	1.00	76.80	3.67	31.52	1.00	37.10	3.62
Mumbai	3.45	0.96	2.15	5.95	35.02	1.00	47.63	3.59
Bhopal	26.01	1.00	41.88	3.50	18.71	0.99	12.68	2.98
Agartala	21.22	1.00	40.71	3.61	21.19	0.99	27.08	3.37
Ahmedabad	19.79	1.00	37.21	3.58	42.81	1.00	43.37	3.52

under research. The entropy of a system is given by

$$ENTR = - \sum_{l=1}^{N_l} p(l) \ln p(l) \quad (5)$$

where  $p(l)$  is the distribution of diagonal line lengths,  $ENTR$  is a measure of signal complexity and is calibrated in units of bits/bin. Signals having periodic behaviour yield high entropy, whereas chaotic signal will have low entropy.

The values of the four distinct variables of recurrence quantification analysis for daily mean temperature and dew-point data of different stations have been estimated and shown in table 3.

## 2.2 Search for underlying periods using Hilbert–Huang transform (HHT)

Most of the geophysical and astrophysical signals are not stationary nor do they have linearity. Though Fourier transform (FT) remains one of the most popular spectral methods to find the underlying periods in a dataset when analyzing the geophysical and astrophysical signals FT is not at all suitable to detect the underlying periodicities within the signal (Huang *et al.* 2001). This results in a trade-off between time and frequency resolution for non-stationary signals and create spurious harmonics for non-linear signals. The Fourier representation of a signal  $x(t)$

$$x(t) = \text{Re} \left( \sum_{n=0}^{N-1} a_j e^{i\omega_j(t)t} \right), \text{ where} \\ j = 0, 1, 2, \dots, N-1 \quad (6)$$

Hilbert–Huang transform (HHT), an alternative spectral method, avoids the linearity and stationarity constraints of Fourier analysis and hence it has been used here to search the relevant

periods of the temperature and dew-point time series. The HHT (Huang *et al.* 1998; Khondekar *et al.* 2010) defines instantaneous frequency as the time derivative of phase, illuminating previously inaccessible spectral details in transient signals. Non-linear signals become frequency modulations rather than a series of fitted sinusoids, eliminating artificial harmonics in the resulting spectrogram. It has been tested and validated carefully by many researchers, but only empirically. In all the cases studied, the HHT gave results much sharper than those from any of the traditional analysis methods in time-frequency-energy representations. Additionally, the HHT revealed true physical meanings in many of the data examined. Powerful as it is, the method is entirely empirical. The HHT consists of two parts: empirical mode decomposition (EMD) and Hilbert spectral analysis (HSA).

### 2.2.1 Empirical mode decomposition (EMD)

#### 2.2.1.1 Defining the instantaneous frequency

The Hilbert transform  $H(t)$  of an arbitrary time series,  $x(t)$  is given by

$$H(t) = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau \quad (7)$$

in which the  $PV$  indicates the principal value of the singular integral. The analytic signal  $\kappa(t)$  is defined as

$$\kappa(t) = x(t) + iH(t) = a(t)e^{i\phi(t)} \quad (8)$$

where  $a(t) = \{x^2(t) + H^2(t)\}^{1/2}$  and  $\phi(t) = \arctan \left[ \frac{H(t)}{x(t)} \right]$

In equation (7), the polar coordinate expression further clarifies the local nature of this representation: it is the best local fit of an amplitude and

phase-varying trigonometric function to  $x(t)$ . Here  $a(t)$  is the instantaneous amplitude, and  $\phi(t)$  is the phase function. The instantaneous frequency of  $x(t)$  is defined by

$$\omega = \frac{d\phi(t)}{dt}. \quad (9)$$

### 2.2.1.2 Decomposition into intrinsic mode functions

This definition of instantaneous frequency is only meaningful for mono-component signals with a zero mean. Otherwise, the instantaneous frequency does not reflect the actual frequency content of the signal. The real advantage of the Hilbert transform became obvious only after Huang *et al.* (1998) introduced the empirical mode decomposition method. Huang *et al.* (1998) proposed a class of functions designated as intrinsic mode functions (IMF) for the instantaneous frequency to make sense. The decomposition steps are as follows

- Step 1: Identify all the extremas of the dataset  $x(t)$ ,
- Step 2: Connect all the local maxima and minima by the cubic spline method to form the upper and lower envelopes, respectively,
- Step 3: Estimate the mean values  $m_1(t)$  by averaging the upper and lower envelope,
- Step 4: Estimate the differences between the data and the mean values to get the first component

$$h_1(t) = x(t) - m_1(t) \quad (10)$$

- Step 5: If the first component does not satisfy the condition of an IMF,  $h_1(t)$  is being assigned as the new data set. The Step 1 to 4 are repeated until the first component is an IMF,
- Step 6: If the first IMF component is called  $c_1(t)$  then estimate the residue  $r_1(t) = x(t) - c_1(t)$
- Step 7: Continue Step 1 to 4 until the residue  $rn(t)$  is smaller than a predetermined value or becomes a monotonic function that no more IMF can be extracted. In this way 11 IMFs has been obtained both for temperature and dew-point time series.

### 2.2.2 The Hilbert spectral analysis

After the decomposition step, the data are reduced to several IMF components. By performing the

Hilbert transform on each IMF components, the data can be expressed as follows

$$x(t) = \text{Re} \left( \sum_{j=1}^n a_j(t) e^{i(\int \omega_j(t) dt)} \right) + r_n. \quad (11)$$

Here the residue  $rn(t)$  is omitted because it is either a mean trend or it might be constant having value smaller than a predetermined threshold. Comparing to the Fourier expansion as given by equation (6), then equation (11) is a generalized Fourier expansion. In equation (6),  $a_j$  and  $\omega_j$  are constants, whereas in equation (11) both  $a_j(t)$  and  $\omega_j(t)$  of each components are continuous functions of time. This time-varying amplitudes and instantaneous frequencies have not only greatly improved the efficiency of the expansion, but also enabled the expansion to accommodate non-linear and non-stationary data. Furthermore, equation (11) enables us to represent the amplitude and the instantaneous frequency as functions of time. In other words, the amplitude can be contoured on the frequency–time plane. This frequency–time distribution of the amplitude is designated as the Hilbert amplitude spectrum,  $H(\omega, t)$ , or simply Hilbert spectrum.

The marginal average power spectrum  $h(\omega)$  is defined as

$$h(\omega) = \frac{1}{T} \int_0^T H^2(\omega, t) dt. \quad (12)$$

This marginal spectrum gives the measure of average power contribution from each frequency value. The frequencies and hence the corresponding periods (1/frequency) at which  $h(\omega)$  are significantly high for each IMF can now be detected.

### 2.2.3 Algorithm for the search of prominent peaks

The periods obtained in the last section for each IMF are sorted out serially. Their corresponding powers are sorted out accordingly. Now the local peaks among the periods for each IMF are detected. If  $N$  number of peaks are being identified for an IMF, then let each peaks are being denoted by  $p[n]$  and their corresponding powers are being denoted by  $x[n]$  where  $n = 1, 2, \dots, N$ . Confidence levels (CL) of all the peaks are calculated and the peaks having CL above or equal to 95% are selected for further processing. The

integral method of sorting (Sboner *et al.* 2004) is then implemented on these selected peaks to detect the most authenticate peak/peaks for a particular IMF. The algorithm of this sorting is as follows

Step 1: Integral -  $\sum_{t_n=1}^N |x(n)| = A_1$

Step 2: Discrete Derivative -  $\frac{x(n+1)-x(n)}{t(n+1)-t(n)} = \Delta_1 x(n)$

Step 3: Maximum localisation -  $\tau = \operatorname{argmax}(\Delta x(n))$

Step 4: Local Integral -  $\sum_{t_n=\tau-u}^{t_n=\tau+u} |x(n)| = A_2$  where  $u = 1 \text{ or } 2$

Step 5: Score -  $\frac{A_2}{A_1} = S_0$   
A vector of  $S_0$  is obtained for all peaks of each IMF.

Step 6: Dataset stratification

The vector of  $S_0$  is sorted as per their magnitude.

Now the most relevant or prominent peak is the one which has highest score among those peaks having confidence level above or equals to 95%.

### 3. Results and discussion

The recurrence threshold  $\varepsilon$  is a crucial parameter in the RP analysis. In this work recurrence threshold  $\varepsilon$  is being chosen as 5% of the maximal phase space diameter (Schinkel *et al.* 2008).

From the recurrence plots of temperature and dew-point time series for all stations the followings has been observed.

- a) The occurrences of disruptions or white bands in all the RPs in table 2 indicate non-stationarity in the process. This non-stationarity is due to the time varying nature of the driving forces behind the temperature and dew-point fluctuations.
- b) Periodic patterns like short or long diagonal lines indicate cyclicities in the process. The time distance between the lines corresponds to the period. The RPs for Chennai and Mumbai are with short diagonal lines compared to that of the RPs of other stations, which claim a quasi-periodic temperature and dew-point dynamics in these two locations compared to purely periodic dynamics of the other stations. It is to be noted that these two stations with quasi-periodic dynamics are very close to sea (2 km) than the other six stations.
- c) The presence of diagonal lines parallel to the LOI (line of identity) declares that the evolution of states of the dynamics is similar at different times. This demands that the processes are deterministic in nature and hence forecasting of temperature as well as dew-point can be made with sufficient reliability.
- d) Presence of single isolated points indicates some heavy fluctuations in the process which may be due to sudden changes in dynamics like natural calamity.
- e) In case of New Delhi the vertical, horizontal lines or clusters says that there may be some states that either do not change or change slowly for some time; an indication for laminar states. It is to be noted that the distance between the sea and New Delhi is maximum (943 km) compared to the other six stations.
- f) The RP's of Kolkata and Agartala are very similar, may be for the reason that both are from eastern India. On the other hand, there are many similarities in RP's of Ahmedabad, New Delhi and Bhopal.
- g) There are disruptions in RP's of both Mumbai and Chennai, which signifies that the signature of non-stationarity in temperature and dew-point is much stronger than that of other locations. One more interesting observation is that the level of disruptions are high for dew-point of Chennai and temperature of Mumbai, but low for dew-point of Mumbai and temperature of Chennai. Further investigation is required to reveal the causality between temperature and dew-point at these two stations.

From table 3 it is found that the average values of % REC for temperature and dew-point across India are 20.48 and 22.77%, respectively, which are though not close to 0.00%, but far away from 100.00% indicating the presence of a fair amount irregularity in their profiles. The values of % REC for the temperature of Mumbai and Chennai, which are in close proximity to the sea, are very low whereas that for New Delhi, which is close to the Himalayan range, is very high when compared to the average value of % REC of India. In fact, the % REC values for the temperature data are decreasing as the vicinity of the stations towards the sea increases. So it can be inferred that the regularity in the daily temperature behaviour is affected by the sea. This is not true in case of the % REC of the daily dew-point series. Ahmedabad, which



is very close to Thar desert, exhibit maximum regularity in its dew-point fluctuation whereas Chennai, which is close to sea, show poor regularity. Again, the dew-point profile of Mumbai, which is close to sea, is very regular having high value of % REC. So as far as the values of % REC of the dew-point profiles are concerned, the relationship between the vicinity to the sea or desert and the regularity of the time series is not a simple one but complex.

The high values of DET for all the stations claim that both the temperature and dew-point profiles across India are deterministic in nature with periodic or quasi-periodic variation.

The average LINEMAX values for temperature are found to be more than that of the dew-point across India signifying that the former is more regular and stable. Though the overall temperature profile is more regular and stable, the span of its LINEMAX values (2.1481–76.7977) is much wider than that of dew-point (4.9281–47.628) exhibiting more variant fluctuations in regularity and stability of temperature profile compared to the dew-point across India. Interestingly Mumbai, which is on west coast along Arabian Sea, exhibits minimum LINEMAX for temperature and maximum LINEMAX for its dew-point profile, which says that though its temperature profile is considerably chaotic with less stability, its dew-point profile is regular with more stability. On the contrary, the dew-point of another coastal station, Chennai, which is on east coast along Bay of Bengal, has LINEMAX that is minimum of all the other seven stations claiming highly chaotic and unstable profile. From this observation, will it be correct to say that if Arabian Sea makes the temperature fluctuation chaotic then the same is done for dew-point by Bay of Bengal. The LINEMAX for both temperature and dew-point of New Delhi are much higher suggesting strong periodicity and maximum stability whereas moderate values of LMAX for rest four stations, which indicates a probable presence of both chaos and periodicity (may be a periodic chaos) in the system.

As shown in table 3, the ENTR of temperature and dew-point of Chennai, located on the east coast, are minimum establishing the claim that the fluctuations of these two weather variables are not only chaotic in behaviour but also have less complex structure and hence few information is required to reveal their complete dynamics. On the contrary, the entropy of temperature and dew-point of Mumbai, located on the

west coast, are maximum, which express that the fluctuations of these two weather variables are more complex in structure with deterministic periodic behaviour and hence more information is required to expose their complete dynamics.

Another noticeable observation is that the RQA parameters magnitudes for Kolkata (22.57°N), Ahmedabad (23.04°N), Bhopal (23.26°N) and Agartala (23.83°N), which are in the vicinity of Tropic of Cancer (23.5°N) are very close whereas those for Mumbai (19.02°N), Chennai (13.06°N) and New Delhi (28.34°N), which are away from Tropic of Cancer (23.5°N) exhibit extremities. As per Koeppen-Geiger classification, New Delhi is in *Cwa* and Chennai is in *As* and Mumbai is in *Aw*. In case of temperature, most of the RQA parameters are maximum for New Delhi, whereas for most of the parameters for dew-point, Chennai exhibits minimum and Mumbai exhibits maximum. Noticeably the RQA parameters of Kolkata for both temperature and dew-point are close to their average values with respect to entire India.

The frequencies of occurrence of periodic pattern within signals are detected using HHT and EMD. The periods in each IMF for both temperature and dew-point with highest score and having confidence level (CL) above and equals to 95% for different stations are given in table 4. From table 4 on comparison between the periodicities of temperature and dew-point for each station it is obvious that significant long-period variations are more prevalent in the temperature series than in the precipitation series for all stations except Agartala.

From table 4 it is found that the distribution of the periods spans over a long range with small periods as low as less than a month to long periods as high as more than 2500 years for temperature and 750 years for dew-point with maximum number of periods between 1 month and 1 year as depicted in figure 1. It means that both the temperature and dew-point variations are influenced by the oscillations of other atmospheric phenomena and climate fluctuations having periodicities ranging between less than a months to more than few thousand years.

About 21% of the total periods of temperature across India are below one month with an average of 10.4 days whereas 18% of the total dew-point periods are below one month with an average of 12.5 days. These periods of 10.4 days for temperature and 12.5 days for dew-point resembles with the periodicity of the oscillation of frontal system

Table 4. Significant periods for different stations.

Periods with highest scores and CL above or equal to 95% (d: day, m: month, y: year)						
Kolkata	Chennai	New Delhi	Mumbai	Bhopal	Agartala	Ahmedabad
<i>Temperature</i>						
10d (1)	7d (6)	22d (1)	6d (6)	10d (6)	10d (1)	11d (1)
11.5d (11)	11.4d (1)	1m 10d (6)	11d (1)	10d (1)	12d (11)	12d (11)
3m 19d (7)	12d (11)	1m 23d (11)	11d (11)	11d (11)	1m 21d (5)	1m 13d (6)
6m (6)	3y (4)	9m 7d (4)	7m 10d (7)	2m 12d (5)	4m 13d (6)	5m 8d (5)
1y (5)	5y (5)	1y (9)	1y 1m (5)	6m (4)	7m (7)	1y (4)
3y (8)	5y (8)	1y 4m (8)	6y (4)	1y 4m (7)	1y 3m (4)	2y 7m (7)
4y (4)	6y (7)	1y 8m (10)	18y (8)	7y (3)	5y (3)	3y 2m (3)
7y 6m (3)	13y (3)	4y (5)	19y 6m (3)	9y (10)	9y (10)	5y 3m (8)
21y (10)	18y (9)	6.5y (3)	31y (9)	14y (9)	14y (9)	10y 8m (9)
34y (9)	144y (2)	13y (7)	44y (2)	14y (8)	18y (8)	15y 6m (10)
71y (2)	2551y (10)	26y (2)	261y (10)	50y (2)	30y (2)	101y 5m (2)
<i>Dew point</i>						
11d (11)	7d (6)	10d (11)	NIL (7)	10d (11)	11d (11)	5d (11)
16d (1)	11d (1)	18d (1)	12d (11)	19d (1)	16d (1)	1m 6d (1)
1.5m (6)	12d (11)	2.5m (5)	17d (1)	5m (6)	1m 6d (6)	3m 4d (6)
5m 21d (7)	5m (7)	2m 18d (6)	1m 19d (6)	5m 13d (5)	4m (5)	8m (7)
8m (5)	1y 5m (5)	1y 1m (4)	10m (5)	6m (4)	8m (4)	10m 14d (9)
11m 6d (4)	12y 5m (4)	1y 3m (8)	1y 1m (4)	4y 6m (7)	1y 5m (7)	1y (4)
3y (8)	15y (8)	3y 10m (10)	7y 10m (3)	5y 6m (3)	3y 10m (8)	1y 1m (10)
7y 6m (10)	19y 1m(3)	4y 10m (7)	12y 2m (9)	6y 5m (9)	7y 2m(3)	1y 1m(5)
7y 9m (3)	33y 8m (9)	8y 7m (3)	35y 3m (2)	6y 10m (8)	43y 3m(2)	1y 5m (8)
16y 7m (9)	49y 7m (2)	10y 7m (2)	45y 9m (8)	11y 5m (2)	53y 9m (10)	4y 3m (3)
31y 5m (2)	60y 3m (10)	18y 9m (9)	787y 2m (10)	25y 9m (10)	97y 12m (9)	10y 7m (2)

Figures within the bracket indicates the IMF from which the period is obtained.

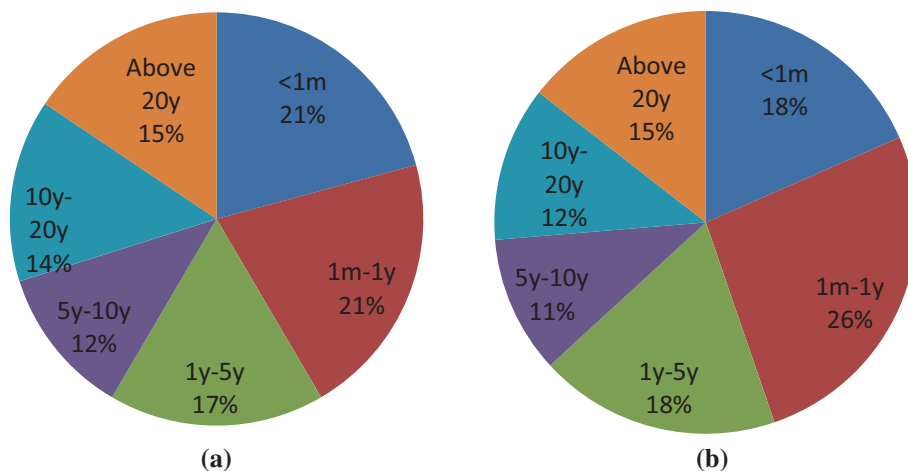


Figure 1. Distribution of periods for (a) temperature and (b) dew point.

having periods ranging between 4.5 and 15.5 days and Hadley oscillation having periods between 10 and 15 days (Goswami and Shukla 1983).

The periods which are above 1 month but less than 1 year are 21 and 26% of the total periods for temperature and dew-point, respectively with their corresponding averages of 5 months 25 days (~6 months) and 5 months 16 days (~6 months)

exhibiting the presence of semi-annual oscillations (SAO) in them. Besides this, Madden–Julian oscillation, which has the periodicity in between 1 month and 1 year (more specifically 30–60 days) may also have impact on the temperature and dew-point profile causing.

In between 1 and 5 years periodicity, temperature has a representation of 17%, whereas

dew-point has 18% of representation with average of 2.8 years and 2.41 years, respectively. This range of periodicity is also there for El Niño, La Niña and ENSO phenomena whose cyclicities varies between 2 and 5 years. The periodicities of Indian Ocean Dipole (IOD) (also known as India Nina) have also been found to be within 2–5 years by many authors.

About 11–12% of the total number of periods is in between 5 and 10 years for both the temperature and dew-point with averages of 6.75 and 7.2 years, respectively. Similar type of periodicity is observed in the atmospheric phenomena like North Atlantic Oscillation (NAO) with periods of 7.7 years (da Costa and de Verde 2002).

About 14% of the total periods of temperature are between 10 and 20 years with an average of 15.2 years, whereas in case of dew-point nearly 12% of the periods are in this category with an average period of 14.1 years.

The high value periods (over 20 years) are also there in the temperature and dew-point profile with a 15% presence. Of these high periods nearly 48% are concentrated in between 30 and 50 years with an average of 38.9 years, which mostly bear a resemblance to Bruckner Cycle (Bruckner 1890; Raspopov *et al.* 2000).

It is quite evident from the above discussion that the temperature and dew-point fluctuations across India are not only affected by the local geographical topology but also heavily influenced by the terrestrial climatic and atmospheric oscillations.

The analysis and interpretation which are being made in this paper are based on the results of operation of the two statistical tools (a) RP and RQA and (b) HHT. Of these two, the RP and RQA, which are used for the exploring the complexity of the temperature and dew-point time series are based on mere visual inspection of the plots. The authors consider this as a limitation of this work where the RP is being used to assess the intricate and weighty aspect like complexity of the time series. To explore the complexity of the meteorological parameters more statistically sound methodologies can be used in future.

## Acknowledgements

The authors express their heartfelt gratitude to the learned reviewers for their valuable comments advised to improve the quality of the paper.

## References

- Ahmed S I, Rudra R, Dickinson T and Ahmed M 2014 Trend and periodicity of temperature time series in Ontario; *Am. J. Clim. Change* **3** 272–288.
- Azzini R, Ciocchetta F D I, Sboner D F A and Blanzieri A M E 2004 Simple methods of peak detection in time series microarray data; <http://www.camda.duke.edu/camda04/papers/days/thursday/azzini/paper.pdf>. Accessed 24 June 2014.
- Blasius B, Huppert A and Stone L 1999 Complex dynamics and phase synchronization in spatially extended ecological systems; *Nature* **399** 354–359.
- Bruckner E 1890 Klimaschwankungen seit 1700 nebst Bemerkungen über die Klimaschwankungen in der Diluvialzeit; *Geograph. Abh.* **4**(2) 153–484.
- Byakatonda J, Parida B P, Kenabatho P K and Moalafhi D B 2018 Analysis of rainfall and temperature time series to detect long-term climatic trends and variability over semi-arid Botswana; *J. Earth Syst. Sci.* **127**(25) 1–20.
- da Costa E D and de Verde A C 2002 The 7.7 year north Atlantic oscillation; *Q. J. R. Meteorol. Soc.* **128**(581) 797–818.
- Dergachev V A and Geel B V 2004 Large-scale periodicity of climate change during the Holocene; In: *Impact of the Environment on Human Migration in Eurasia* (eds) Scott E M, Alekseev A Y and Zaitseva G, Proceedings of the NATO Advanced Research Workshop, pp. 159–183.
- Eckmann J P, Kamphorst S O and Ruelle D 1987 Recurrence plots of dynamical systems; *Europhys. Lett.* **4**(9) 973–977.
- Glass L 2001 Synchronization and rhythmic processes in physiology; *Nature* **410** 277–284.
- Goswami B and Shukla J 1983 Quasi-periodic oscillations in symmetric general circulation model; *J. Atmos. Sci.* **40**(1) 20–37.
- Huang N E, Chern C C, Huang K, Salvino L W, Long S R and Fan K L 2001 A new spectral representation of earthquake data: Hilbert spectral analysis of Station TCU192, Chi-Chi, Taiwan, 21 September 1999; *Bull. Seismol. Soc. Am.* **91**(5) 1310–1338.
- Huang N E, Chern C C, Huang K, Salvino L W, Long S R and Fan K L 1998 The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis; *Proc. R. Soc. Lond.* **454** 903–995.
- Kantz H and Schreiber T 1997 *Non-linear time series analysis*; Cambridge University Press, Cambridge.
- Kennell M B, Brown R and Abarbane H 1992 Determining embedding dimension for phase-space reconstruction using a geometrical construction; *Phys. Rev. Lett. A* **45** 3403–3411.
- Khondekar H M, Ghosh D N, Ghosh K and Bhattacharya A K 2010 Application of signal processing to investigate the total active 8B solar neutrino flux signal from Sudbury Neutrino Observatory (SNO); *Int. J. Electron. Eng. Res.* **2**(3) 303–324.
- Khondekar H M, Ghosh D N, Ghosh K and Bhattacharya A K 2012 Non-linearity and chaos in 8B solar neutrino flux signals from sudbury neutrino observatory; *Fractals* **20**(1) 17–32.

- Khondekar H M, Ghosh D N, Ghosh K and Bhattacharya A K 2013 Complexity in solar irradiance from the earth radiation budget satellite; *IEEE Syst. J.* **9**(2) 487–494.
- Marvel K B 2001 Astrophysics: a stellar performance; *Nature* **411** 251–252.
- Marwan N and Kurths J 2004 Cross recurrence plots and their applications; In: *Mathematical Physics Research at the Cutting Edge* (eds.) Benton C V, Aboughantous C H, Cantcheff M B, Kurths J, Kryukov A A, Marwan N, Piotrowski E W, Sladkowski J and Varlamov V V, Nova Science Publishers, Hauppauge, pp. 101–139.
- Marwan N, Romano M C and Kurths J 2007 Recurrence plots for the analysis of complex systems; *Phys. Rep.* **438**(5–6) 237–329.
- Marwan N, Wessel N, Meyerfeldt U, Schirdewan A and Kurths J 2002 Recurrence plot based measures of complexity and its application to heart rate variability data; *Phys. Rev. E* **66**(2) 1–8.
- NOAA 2014 Climate data online; <http://www7.ncdc.noaa.gov/CDO/cdosubqueryrouter.cmd>. Accessed 20 Feb 2012.
- Raspopov O M, Shumilov O I, Kasatkina E A, Turunen E and Lindholm M 2000 35-year climatic Bruckner cycle-solar control of climate variability?; Proceeding of the 1st Solar and Space Weather Euroconference, Tenerife.
- Ray R, Khondekar M H, Ghosh K and Bhattacharjee A K 2015 Memory persistency and non-linearity in daily mean dew point; *Theor. Appl. Climatol.* **119**(3–4) 1–12.
- Ray R, Majumder P, Khondekar M H, Ghosh K and Bhattacharjee A K 2017 Self-organized criticality, causality and correlation of probability of recurrence between daily mean temperature and dew point across India; Third IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICR-CICN), pp. 98–103.
- Reddy B N K, Venkatesan R, Osuri K K, Mathew S, Kadiyam J and Joseph K J 2018 Comparison of AMSR-2 wind speed and sea surface temperature with moored buoy observations over the northern Indian ocean; *J. Earth Syst. Sci.* **127**; <https://doi.org/10.1007/s12040-017-0902-3>.
- Roulston M S 1999 Estimating the errors on measured entropy and mutual information; *Phys. D* **125**(3–4) 285–294.
- Sboner A *et al.* 2004 Simple data; In: *Methods of Microarray Data Analysis* (ed.) McConnell P, V. Povo (TN): Springer US, pp. 27–44.
- Schinkel S, Dimigen O and Marwan N 2008 Selection of recurrence threshold for signal detection; *Eur. Phys. J. Spec. Top.* **165**(1) 45–53.
- Sole J, Turiel A and Llebot J E 2007 Using empirical mode decomposition to correlate paleoclimatic time-series; *Nat. Hazards Earth Syst. Sci.* **7** 299–307.
- Steadman R G 1984 A universal scale of apparent temperature; *J. Clim. Appl. Met.* **23** 1674–1687.
- Takens F 1981 Detecting strange attractors in turbulence; In: *Dynamical Systems and Turbulence, Warwick 1980* (eds) Rand D A and Young L S, Springer, Berlin, pp. 366–387.
- Thomasson N, Hoeppepner T J, Webber Jr C L and Zbilut J P 2001 Recurrence quantification in epileptic EEGs; *Phys. Lett. A* **279**(1–2) 94–101.
- Webber Jr C L and Zbilut J P 1994 Dynamical assessment of physiological systems and states using recurrence plot strategies; *J. Appl. Physiol.* **76**(2) 965–973.
- Zbilut J P and Webber Jr C L 1992 Embeddings and delays as derived from quantification of recurrence plots; *Phys. Lett. A* **171**(3–4) 199–203.

Corresponding editor: A K SAHAI