Trends analysis of precipitation data over the tropical South-West Indian Ocean (SWIO) basin using the Ensemble Empirical Mode Decomposition (EEMD) method

K. BOODHOO, M. R. LOLLCHUND* and A. F. DILMAHAMOD**
Department of Chemistry, Faculty of Science, University of Mauritius
*Department of Physics, Faculty of Science, University of Mauritius
**Department of Oceanography, University of Cape Town, South Africa

(Received 24 July 2014, Accepted 18 November 2014)
*e mail: r.lollchund@uom.ac.mu

ABSTRACT. In this paper, we propose the use of the Ensemble Empirical Mode Decomposition (EEMD) method in the analysis of trends in climate data. As compared to existing traditional methods, EEMD is simple, fast and reliable. It works by decomposing the time-series data into intrinsic mode functions until a residual component is obtained which represents the trend in the data. The dataset considered consists of satellite precipitation estimates (SPE) obtained from the Tropical Rainfall Measuring Mission (TRMM) for the tropical South-West Indian Ocean (SWIO) basin recorded during the periods January 1998 to December 2013. The SWIO basin spans from the latitudes 5° S to 35° S and the longitudes 30° E to 70° E and comprises of part of the east coast of Africa and some small island developing states (SIDS) such as Comoros, Madagascar, Mauritius and Reunion Island. The EEMD analysis is carried out for summer, winter and yearly time series of the SPE data. The results from the study are presented in terms of intrinsic mode functions (IMFs) and the trends. The analysis reveals that in summer, there is a tendency to have an increase in the amount of rainfall, whereas in winter, from 1998 to 2004 there has been an initial increase of 0.0022 mm/hr/year and from there onwards till 2013 a decrease of 0.00052 mm/hr/year was noted.

Key words – Precipitation, Trend analysis, Ensemble Empirical Mode Decomposition (EEMD), Intrinsic Mode Function (IMF), South West Indian Ocean (SWIO) basin.

1. Introduction

In the study of global, regional as well as local climate change impact scenarios, it is important to understand the trends in the climatic variables. In literature, most of the related works have relied on the least squares best fit lines (Lunagaria et al., 2012; Tong et al., 2014) method while few others have employed...
statistical models (von Storch & Zwiers, 1999). The traditional stochastic methods include the probability distribution method (give amplitude information only), auto- and cross-correlation function (show time-independence only) and the power-density spectrum (contains frequency-dependent information only) (Goodwin, 2008). However, climate data are complex, non-linear and non-stationary as they are constantly changing with position and time. Analysis tools for non-linear time series normally employ Fourier transforms. However, owing to the global nature of the transforms, the final results may lead to inconclusive interpretations (Mak, 1995). Even with the use of wavelet analysis, which was developed especially to deal with non-stationary time-series data, confusing and contradicting results were obtained when applied to climate (Oh et al., 2003). Therefore, an appropriate non-stationary time-series data analysis technique must be employed.

Huang et al. (1998) developed the Empirical Mode Decomposition (EMD) method as an effective tool for analyzing the non-stationary properties of time series data. The method allows the breaking down of a signal into several intrinsic mode functions (IMFs) with different oscillation periods and a residual component without leaving the time domain. The process is useful for analyzing natural signals, especially for nonlinear and non-stationary signals, compared to Fourier and Wavelet Transforms. EMD has successfully been applied in a number of scientific and engineering fields and even in finance. In the meteorological field for instance, Molla et al. (2006) used EMD to analyze the trend in rainfall data collected from Agricultural experimental farms in Jharkhand, India from 1989 to 2004. Coughlin & Tung (2005) applied the method on data provided by the National Center for the Environmental Prediction to demonstrate that dynamical variables in the zonally averaged troposphere and lower stratosphere contain only five oscillation modes and a trend. In spite of the many advantages of the EMD method, it has the inconvenient feature of mode mixing between the IMFs (Sharma & Kaur, 2014). Hence, the variation of a given frequency may split across two IMFs which is a problem when investigating the physical significances of the latter (Wu et al., 2014). Recently, Wu & Huang (2009) proposed the Ensemble Empirical Mode Decomposition (EEMD), a noise-assisted method, to overcome the problem of mode mixing. Here, noise is superimposed with the time-series in order to bring in the ordering of local maxima and minima which are the basis of finding IMFs in EMD. It should be noted that the IMFs and residual obtained using EEMD have more physical meaning than their counterparts in EMD (Wu & Huang, 2009; Sharma & Kaur, 2014; Wu et al., 2014).

In this study, an attempt is made to employ EEMD in the analysis of trends in time-series data for satellite precipitation estimates (SPE). Rainfall is most of the times the major source of freshwater. Nowadays, due to the current prevailing climate change, it has been observed that in some areas of the globe there are severe droughts over long periods of time and on the other hand there is serious flooding, leading in both cases to a scarcity in freshwater supply (MMS, 2014; MFR, 2014). The reason is that in the first scenario there will be a significant depletion of freshwater whereas in the second scenario the latter will get contaminated. In that regard, it is highly imperative to monitor and predict the trends in precipitation data in those affected regions. SPE data measured over the tropical South-West Indian-Ocean (SWIO) region during the period January 1998 to December 2013 are used. They are decomposed into several IMFs with different characteristic scales and a residue using EEMD and their trends and physical properties are analyzed. To the best of the authors’ knowledge, this study is the first to use such an approach in the tropical SWIO basin.

2. Data and methodology

2.1. Study area and data description

The study area is a region within the tropical SWIO basin which spans from the latitudes 5° S to 35° S and the longitudes 30° E to 70° E (Fig. 1). It comprises of a part of the east coast of Africa (Mozambique, Tanzania and South Africa) and some small island developing states (SIDS) such as Comoros, Madagascar, Mauritius, Mayotte and Reunion Island. The climate in this part of the world varies between tropical and subtropical with a cyclonic season of about 5 months starting from December to May. The mean annual rainfall is about 0.22 mm/hr. The wettest months are February and March. The driest month
is October. Usually, most of the rainfall occurs in summer months. The air temperature lies between 12 °C to 38 °C. The sea surface temperature can reach up to 28 °C. The SWIO region has two seasons: a warm humid summer extending from November to April and a relatively less humid winter from May to October (MFR, 2014; MMS, 2014).

Remotely sensed TRMM data used in the present study for the SWIO region are available at the Goddard Space Flight Center from the North America’s Space Agency (National Aeronautics and Space Administration - NASA). The monthly precipitation data (3B43 Products) spans from January 1998 to December 2013 and have a spatial resolution of 0.25°.

It is worthwhile mentioning that precipitation products from TRMM are being used as an alternative rainfall measurement data around the world. The precipitation radar on the satellite covers a global scale and has accurate calibration, downward viewing geometry, and lack of beam blockage (Yang & Nesbitt, 2014). However, these measurements are only estimates and are subject to uncertainties. Previous studies have tackled this issue by comparing TRMM-based precipitation to rain gauge measurements in different case studies (Yang and Nesbitt, 2014; Kneis et al., 2014; Haque et al., 2013; Huffman et al., 2006). For instance, Yang & Nesbitt (2014) showed that moderate and light rain events are likely to be missed by TRMM radar but these do not have significant influence on large-scale statistics of stratiform rain amount and convection. However, over the Mahanadi River in India, it has been revealed that even intense rainfall events are not registered by the satellite-based precipitation estimates (Kneis et al., 2014). Comparing high and low altitude stations data with TRMM, Haque et al. (2013) revealed better correlation of rain with low altitude stations, with the sensor failing to detect high rainfall intensity close to mountains. In general, the validation studies have shown that TRMM data compares fairly well with gage measurements on a monthly time scale compared to finer time scales (Haque et al., 2013; Huffman et al., 2006).

2.2. The EEMD method

As mentioned in the introduction, the EEMD method is an improvement of the original EMD method to overcome the problem of mode mixing. EEMD uses white noise along with an ensemble of EMD runs to decompose a time series (Wu & Huang, 2009). The paragraphs which follow briefly describe the EMD and EEMD algorithms.

EMD basically detaches non-linear oscillatory patterns of higher frequencies from those of lower frequencies in the data. The method is straightforward. For a given time series signal \( x(t) \), EMD represents the signal in a set of basis functions (simpler signals), which are termed as intrinsic mode functions (IMF), using local temporal and structural characteristics of the data. An IMF should satisfy the following properties: (1) zero mean; and (2) the number of local extrema equals the number of zero crossings ± 1 (Huang et al., 1998; Coughlin & Tung, 2005; Molla et al., 2006).

The IMFs are obtained by the following procedure:

1. Connect the maxima in \( x(t) \) using smooth lines and denote this curve as upper envelope, \( E_u(t) \). Similarly, connect the maxima in \( x(t) \) using smooth lines and denote this curve as lower envelope, \( E_l(t) \).
2. Compute the mean of these envelopes:

\[
m(t) = \frac{1}{2} [ E_u(t) + E_l(t) ]
\]

It should be noted that \( m(t) \), at any temporal location, is nearly equal to the lowest frequency component in the data.
3. Subtract the mean of the envelopes from \( x(t) \) and store the result in \( h_1(t) \). That is

\[
h_1(t) = x(t) - m(t)
\]

This results into isolating the higher frequency components from the lower frequency components in the original signal. Ideally \( h_1(t) \) is the first IMF.

- If \( h_n(t) \) does not satisfy the properties of an IMF, then it is treated as a “new” signal and is re-entered in the process of computing the IMFs. This iteration, known as the sifting process, is repeated until the IMF is obtained.
- Repeat the process until all IMFs, \( h_n(t) \), are obtained. At this final stage (say after \( n \) iterations), a constant or monotonic function \( r_n(t) \) remains which is termed as the residual. It represents any trend within the original series.

It is intuitive to observe that the original signal can be reconstructed by summing all IMFs and the residual using

\[
x(t) = \sum_{k=1}^{n} h_k(t) + r_n(t)
\]
The IMF components and the residual have physical significance.

The EEMD algorithm for the time series signal $x(t)$, can be described as follows (Wu & Huang, 2009):

- Generate $x'(t) = x(t) + w'(t)$, where $w'(t)$ $(i = 1, \ldots, N)$ are different realizations of white Gaussian noise.

- Each $x'(t)(i = 1, \ldots, N)$ is fully decomposed using the EMD algorithm described above to obtain their IMFs $[h_k^i(t)]$, where $k = 1, \ldots, n$.

- Assign $\overline{h_k(t)}$ as the $k$-th mode of $x(t)$, which is obtained as the average of the corresponding IMFs using

$$\overline{h_k(t)} = \frac{1}{N} \sum_{i=1}^{N} h_k^i(t)$$

It should also be mentioned that the white noise introduced in EEMD helps in the separation of different timescales in noisy data, but has no involvement in the final IMFs (Wu et al., 2014). The added noise cancels out in the ensemble average. The EEMD method is very effective in extracting signals contained in the data, and represents a major enhancement of the EMD method (Wu & Huang, 2009; Sharma & Kaur, 2014; Wu et al., 2014).

3. Results and discussion

In this section the SPE of the SWIO basin obtained for the period January 1998 to December 2013 are subject to trend analysis using EEMD. They are initially processed to generate mean mapping of the data over the study region. The grid size of the SPE data for each month is 160 along the longitude axis and 128 along the latitude axis. The time-series representations of the SPE for the study region are obtained by averaging the dataset values over the whole grid for each month.

3.1. EEMD analysis of yearly SPE

Fig. 2 shows the resulting mean SPE map which demonstrates the mean amount of precipitation estimates (in mm/hour) that has accumulated over the SWIO basin between January 1998 and December 2013.

It can be observed that the data is quite variable, irregular and depends on geographical location. A high amount of precipitation is estimated towards the North-Eastern part of the study area (close to the Indian Ocean equatorial belt) and also along the east coast of Madagascar. It is a major point of concern to conclude that over these particular years there has been a drastic fall in the mean SPE. For instance, Mauritius and Reunion Islands lie in the dark-blue environment, corresponding to the lowest value on the SPE scale.
**Fig. 4.** Time series of mean SPE data (mm/hour) over SWIO basin and its corresponding EEMD trend

**Fig. 5(a&b).** Seasonal mapping of mean SPE (mm/hour) over the SWIO basin for the period January 1998 to December 2013 (a) Summer Season (b) Winter Season

Fig. 3 displays the resulting time-series of the SPE data and its complete decomposition by EEMD in terms of IMFs and a residual. The minimum value of this time-series is 0.0320 mm/hour, maximum is 0.2923 mm/hour, and mean is 0.1166 mm/hour with a standard deviation of 0.0617 mm/hour. It can also be observed that the EEMD decomposition separates the time series into six IMFs and a trend. Relative to the signal data, the periods of the IMFs increase while their amplitudes decrease after each decomposition until the residual is obtained.

In order to analyze the IMFs, their mean periods \( \tau_\mu \) are computed using

\[
\tau_\mu = \frac{1}{L-1} \sum_{l=1}^{L} \rho'_l,
\]

where, \( L \) is the total number of maxima and \( \rho'_l \) is the sample length between \( l \)th and \((l+1)\)th maxima (Molla et al., 2011). Table 1 gives the total number of local maxima as well as the mean average periods (in terms of number of years) of the IMFs of the yearly SPE data.

As can be seen in Table 1, IMF 5 and IMF 6 have only one local maximum and therefore the mean period could not be determined from equation (5). The IMFs are capable of identifying the expected structure within the

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Period statistics of yearly mean SPE IMFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMF</td>
<td>1</td>
</tr>
<tr>
<td>No. of local maxima</td>
<td>57</td>
</tr>
<tr>
<td>Mean period (year)</td>
<td>0.2798</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Period statistics of seasonal mean SPE IMFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMF</td>
<td>Season</td>
</tr>
<tr>
<td>No. of local maxima</td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>4</td>
</tr>
<tr>
<td>Mean period (year)</td>
<td>Summer</td>
</tr>
<tr>
<td>Winter</td>
<td>3.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Statistical data for seasonal time-series of mean SPEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>Summer</td>
<td>0.1341</td>
</tr>
<tr>
<td>Winter</td>
<td>0.0654</td>
</tr>
</tbody>
</table>
SPE series. The climatic phenomena associated with the different periods in the IMFs can be related to Madden-Julian Oscillation and El Niño–Southern Oscillation (ENSO) (Washington & Preston, 2006; Yeh et al., 2009). However, discussions on the latter are beyond the scope of this work.

In Fig. 4, both the time-series representation of the SPE data and the resulting EEMD trend (residual) are depicted. The wettest periods lie in December to January months (summer season), while the driest periods lie in the months of June to July (winter season).

The highest level of mean SPE was noted in December 1998 with a value of 0.2935 mm/hour while the lowest mean SPE was in July 1999 with a value of 0.0320 mm/hour. It can also be seen that, generally, the trend line shows an average decrease of 0.0012 mm/hour/year in the mean SPE from January 1998 to December 2006 and a slight increase between January 2007 to December 2013 at a rate of $3.8 \times 10^{-4}$ mm/hour/year.

3.2. EEMD analysis of seasonal SPE

Figs. 5(a&b) displays the resulting mean SPE maps (mean amount of precipitation estimates) that has accumulated over the SWIO basin in both summer (October to March) and winter (April to September) seasons for the period January 1998 to December 2013. It can be deduced that the mean SPE in summer is higher in the eastern part of Africa, central and northern parts of Madagascar and north-east of SWIO basin. During winter the east coast of Madagascar as well as north-east region of SWIO recorded the highest SPE, whereas over Africa, the western part of Madagascar as well as the eastern region of SWIO recorded the lowest amount of mean SPE.

The quasi-permanent feature of the precipitation belt of the north-eastern part of SWIO in both summer and winter periods can be explained by the flow dynamics involved in the region. As explained by Kripalani & Kumar (2004), the south-westerly and equatorial westerly flows as well as the north-easterly winds from the Bay of Bengal all converge in this region. The resulting time-series graphs as well as the corresponding IMFs and residuals for both summer and winter mean SPEs are presented in Figs. 6(a&b).
Table 2 shows the period statistics of the seasonal mean SPE IMFs as computed using equation (5). Table 3 displays the statistical data computed for the summer and winter time-series mean SPEs. The low standard deviations (SD) obtained for both data suggest that there has not been significant change over the years 1998 to 2013 in the amount of rainfall.

In Fig. 7, both the time-series representation of the SPE data for summer and winter seasons and their resulting EEMD trends (residuals) are depicted. The positive trend in summer SPE time-series data suggests that the summer season is becoming relatively wetter. From 1998 to 2013, during the summer periods, the mean monthly SPE varied from 0.1341 mm/hr, occurring in 1999 to 0.1723 mm/hr, corresponding to the year 2006. The trendline of the mean variations of the SPE in summer over the 15 years gave a gradient of 0.0012 mm/hr/year. This can be explained by the fact that there have been more and more cyclones of high intensity visiting the region as well as more frequent heavy downpours due to climate change (MMS, 2014; SARUA, 2014). However, during the winter periods, the mean monthly SPE varied from 0.0654 mm/hr, occurring in 2013 to 0.0972 mm/hr, corresponding to the year 2012. Interestingly, the trendline of the mean SPE in winter over the 15 years indicated that from 1998 to 2004 the rate of change in the monthly rainfall was 0.0022 mm/hr/year whereas this value dropped significantly to -0.00052 mm/hr/year over the period 2009 to 2013. Between 2004 and 2009 the SPE has been approximately constant. The likely explanation for this is that the temperatures in winters have increased slightly over the region (as observed by MMS (2014) and MFR (2014). Overall, it can be observed that summer is much wetter than winter over the 15 years considered in this study.

4. Conclusion

In this work the EEMD method allows the decomposition of the time-series SPE data for the SWIO basin into several IMFs and a residual. The latter showed the trend in the data. EEMD has been very effective to demonstrate how the SPE data is evolving over the years. Therefore, it is a very useful tool for studying trends in climate data. The authors believe that such analysis will be helpful to the government and any organization in making an informed decision with regard to water resource management in the SWIO region.

References


