

Ensemble Empirical Mode Decomposition of monthly climatic indices relevant to Australian hydroclimatology

M.C. Peel^a, R. Srikanthan^b, T.A. McMahon^a and D.J. Karoly^c

^a *Department of Infrastructure Engineering, University of Melbourne, Victoria*

^b *Water Division, Bureau of Meteorology, Melbourne, Victoria*

^c *School of Earth Sciences, University of Melbourne, Victoria*

Email: mpeel@unimelb.edu.au

Abstract: Application of the spectral analysis technique Ensemble Empirical Mode Decomposition (EEMD) to hydroclimatic time series produces an orthogonal set of decreasing frequency Intrinsic Mode Functions (IMFs) and a residual (or trend). As the EEMD algorithm is locally adaptive it is suitable for decomposing time series that appear non-stationary in terms of mean and variance: characteristics that are not uncommon in hydroclimatic series. EEMD utilises ensemble averaging to overcome the mode mixing problem observed in traditional Empirical Mode Decomposition (EMD). Here EEMD is briefly outlined and issues relating to the calculation of the Hilbert spectrum from EEMD output are discussed. EEMD is then applied to three climatic indices relevant to Australian hydroclimatic variability, namely the Southern Oscillation Index, Indian Ocean Dipole mode index and the Tasman Sea Index. Analysis of these indices reveals EEMD is capable of identifying physical process signals. For example, the known spectral structure of the SOI is identified by the EEMD algorithm. It was observed during analysis of the Indian Ocean Dipole mode index that EMD post processing may introduce end effects previously minimised by the EEMD ensemble average. These effects may inflate existing features, or potentially introduce low frequency artefacts, though this danger can be mitigated through careful examination of the complete Hilbert spectrum. EEMD and Hilbert spectrum analysis of hydroclimatic data provides insight into long-term trends, dominant frequencies and changes in energy with time for time series that are usually problematic to analyse with traditional spectral analysis techniques. The spectral fingerprints of the climatic indices assessed here are used in a companion paper that looks for those fingerprints within Australian precipitation and temperature data (Srikanthan et al., 2011).

Keywords: *Ensemble Empirical Mode Decomposition (EEMD), Southern Oscillation Index (SOI), Australian hydroclimatic variability and trend*

1. INTRODUCTION

Australian hydroclimatology is characterised by high inter-annual variability of streamflow and, to a lesser extent, precipitation relative to other continents that experience similar climate types (Peel *et al.*, 2004). Physical processes that influence hydroclimatic variability operate across a range of temporal scales and are embedded within the non-linear chaotic noise of the climate system. For example, locations influenced by the El Niño-Southern Oscillation exhibit higher inter-annual variability of precipitation (Nicholls, 1988; Peel *et al.*, 2002). Assessing the contribution of different physical processes to inter-annual hydroclimatic variability requires identifying any process signal from chaotic noisy data. In this paper we apply Ensemble Empirical Mode Decomposition (EEMD, Wu & Huang, 2009), a recently developed spectral analysis technique, to three climatic indices relevant to Australian hydroclimatology to investigate their dominant frequencies, long-term trends and changes in variance with time. Features identified using EEMD are effectively the spectral fingerprint of the climatic indices. If an ocean/atmosphere fluctuation, represented by a climatic index, contributes to Australian hydroclimatic variability, then we expect to see evidence of the spectral fingerprint for that index within Australian hydroclimatic time series. In a companion paper by the same authors (Srikanthan *et al.*, 2011) the influence of ocean/atmosphere fluctuations on Australian monthly precipitation and temperature station data is explored using the spectral analysis results presented here. Information regarding the contribution of ocean/atmosphere fluctuations to Australian hydroclimatic variability is potentially useful for developing seasonal forecasts and climate change impact assessments. Following this brief introduction, EEMD is outlined, the climatic indices introduced and results from EEMD analysis of the climatic indices are presented and discussed before conclusions are drawn.

2. ENSEMBLE EMPIRICAL MODE DECOMPOSITION (EEMD)

2.1. Background

Recently, Wu & Huang (2009) proposed Ensemble Empirical Mode Decomposition (EEMD) as an improvement over the Empirical Mode Decomposition (EMD, Huang *et al.*, 1998) spectral analysis technique. The locally adaptive nature of both the EMD and EEMD algorithms means they are suitable for application to non-linear and or non-stationary time series. Traditional spectral analysis techniques, like Fourier and Wavelet transforms, can produce less informative results when applied to such time series.

Wu & Huang (2009) discuss their rationale for seeking to improve EMD, which is not repeated here. EEMD is a noise assisted data analysis technique in which an ensemble of EMD trials are obtained for a time series. Prior to each EMD run, finite amplitude normally distributed white noise is added to the series. The EMD algorithm then adaptively decomposes the series into a set of intrinsic mode functions (IMFs) and a residual (trend). An IMF is a zero-mean fluctuation whose frequency and amplitude may vary with time. The highest frequency IMF is extracted first with subsequent IMFs having progressively lower frequency fluctuations. Once all IMFs are extracted the residual represents any trend within the series, which may be an unresolved low frequency fluctuation with an average period longer than the period of record or a linear or non-linear trend. The EMD results from each ensemble member are grouped by IMF order and the ensemble average of each IMF group and residual forms the EEMD result. The added noise cancels out in the ensemble average.

Peel *et al.* (2011) tested the performance of EEMD on synthetic time series constructed with fluctuations of known frequency and trend (linear, ramp, step or parabolic) to test the ability of EEMD to successfully identify the embedded components. Initial EEMD results did not consistently perform well. However, by combining non-orthogonal IMFs, following the suggestion of Wu & Huang (2009), EEMD's performance became more consistent and much improved. The orthogonality index based IMF combining rules developed in Peel *et al.* (2011) are used in this paper. The EEMD parameter settings recommended by Peel *et al.* (2011): ensemble size = 200; standard deviation of white noise added as a fraction of observed standard deviation = 0.4; and rational spline tension = 0.5 (Pegram *et al.*, 2008); are also used in this paper.

2.2. Hilbert Spectrum analysis

The Hilbert spectrum of a time series reveals how variance at a given frequency changes with time. For example in Figure 1a three synthetic components (pure signal, no noise, from Peel *et al.*, 2011) are shown, which by construction contain amplitude and frequency modulations with time. The Hilbert spectrum of these three components (Figure 1b) shows the changes in frequency over time, but also the changes in variance over time with low energy coloured blue and high energy coloured red. In Figure 1b, and all subsequent Hilbert spectrum plots, a 5x5 Gaussian filter is applied to the raw Hilbert spectrum results to assist data visualisation.

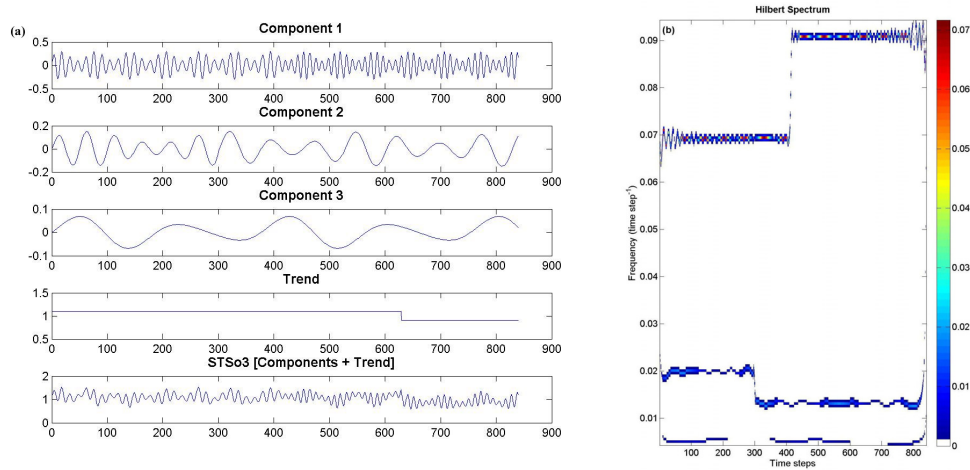


Figure 1. Synthetic time series (STSo3) (a) components, trend and sum of components and trend and (b) Hilbert spectrum of components.

2.3. Post processing EEMD output for Hilbert Spectrum analysis

A key motivation for the original EMD was to obtain components (IMFs) that are amenable to the Hilbert transform for Hilbert spectrum analysis (eg: Figure 1b). The required IMF properties are: (1) zero mean; and (2) the number of local extrema equals the number of zero crossings ± 1 (Huang *et al.*, 1998). The Hilbert spectrum of IMFs that do not satisfy these conditions may contain significant aliasing (Wu & Huang, 2009).

Ensemble average IMFs from the EEMD algorithm may not satisfy the IMF rules, in particular, the second rule regarding the number of local extrema being equal to the number of zero crossings ± 1 . Wu & Huang (2009) recommended EEMD outputs that don't satisfy the second rule be post-processed using EMD to ensure they do. The post processing methodology suggested by Wu & Huang (2009) to ensure suitability for Hilbert spectrum analysis is summarised below for a set of EEMD IMFs (x_1, x_2, \dots, x_n):

1. Begin at the first IMF (x_1) and assess whether each IMF satisfies the second IMF rule,
2. When an IMF does not satisfy the rule (eg: x_3),
 - a. Combine the IMF and subsequent IMF together (eg: $x_3 + x_4 = y_3$),
 - b. EMD the combined IMFs (eg: y_3),
 - c. Use the first EMD IMF (eg: y_{3_1}) as a replacement for the EEMD IMF (eg: x_3),
 - d. Add the EMD remainder (eg: $y_3 - y_{3_1}$) to the next EEMD IMF (eg: x_5) to form a new combination (eg: y_4),
 - e. Repeat steps 2b – 2d until the last EEMD IMF (x_n) has been combined and EMD applied.

In analysis not shown here, application of post processing to EEMD results of synthetic data from Peel *et al.* (2011) raised several issues. Should post processing begin at the first IMF that does not satisfy the second IMF rule, or are some cases where the rule is broken acceptable? For example, a set of EEMD results may identify the synthetic components well, but not all IMFs satisfy the second IMF rule. During IMF post processing reproduction of the synthetic components can become degraded. In some cases post processed EEMD results are less orthogonal than raw EEMD results. Should the user adopt raw EEMD or post processed EEMD results for Hilbert spectrum analysis? The answer is not obvious, particularly as Wu & Huang (2009) note that post processing EEMD may not provide a good solution in every case.

To assist in automatic application of EEMD we suggest two qualifications of the first post processing step, based on analysis of the synthetic data of Peel *et al.* (2011), as given below:

1. Begin at the first IMF (x_1) and assess whether each IMF satisfies the second IMF rule subject to,
 - a. The IMF has $> 1\%$ of the total variance of the EEMD results,
 - b. The number of local extrema - number of zero crossings > 10 .

These two qualifications mean IMFs that: (a) contribute little overall variance; and (b) are almost IMFs, don't trigger post processing of the EEMD results. We observed (not shown) that these two qualifications generally improved the subsequent Hilbert spectrum when the raw EEMD results had high orthogonality index (non-orthogonal IMFs, Peel *et al.*, 2009) and IMFs that clearly did not meet the second IMF rule. However, in cases where the raw EEMD results have low orthogonality index and the IMFs satisfy, or almost satisfy, the second IMF rule, post processing may result in a Hilbert spectrum that is either similar or degraded relative

to the Hilbert spectrum of the raw EEMD results. Therefore, in the following EEMD analysis of climatic indices we ran EEMD and calculated the orthogonality index of the results. We then applied post processing, using the two qualifications, to the raw EEMD results and calculated the orthogonality index of the post processed results. The results with the lowest orthogonality index, raw or post processed EEMD, were adopted for Hilbert spectrum analysis.

3. EEMD ANALYSIS OF CLIMATIC INDICES

The Australian Bureau of Meteorology provided the monthly climatic indices data in the following analyses.

3.1. Southern Oscillation Index

The El Niño/Southern Oscillation (ENSO) is a tropical Pacific ocean/atmosphere phenomenon that fluctuates every 3-7 years. Troup's Southern Oscillation index (SOI), which represents fluctuations in sea level pressure between Darwin and Tahiti, is shown in Figure 2a from 1/1950 – 9/2010. The EEMD residual (trend), also shown in Figure 2a, indicates the trend in SOI decreased from +4 in 1950 to a minimum of -4.6 in 1988 and increased to +3 in 2010. In Figure 2b the variance and average period of the SOI EEMD IMF results are plotted. Also shown are EMD IMF results for 200 white noise series of the same length and variance as the original SOI series with added white noise. If the SOI is well described by white noise, the EEMD IMFs from SOI are expected to plot within, or near, the white noise results. SOI IMFs that plot above the white noise cloud indicate they contain more variance than expected from a white noise process and may be dominant frequencies (a signal). For monthly SOI, IMFs 4, 5 and, to a lesser extent, 6 all plot above the white noise cloud. These IMFs have average periods of 2.3, 4.5 and 10 years respectively.

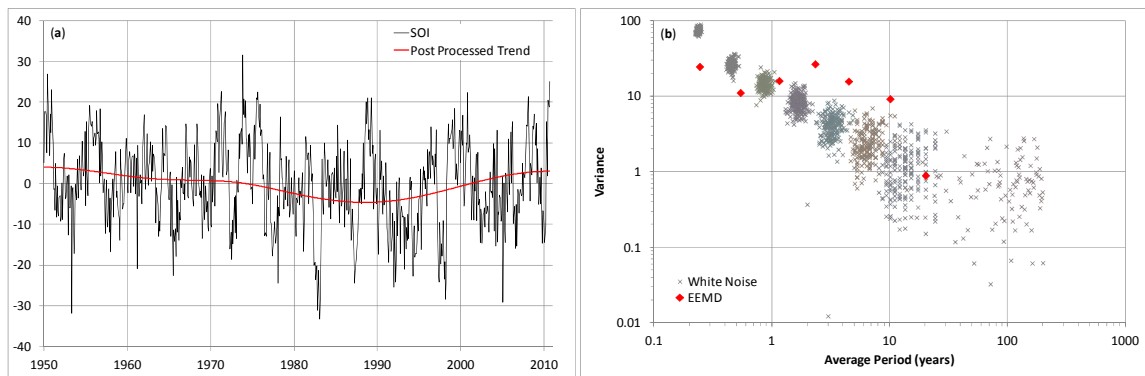


Figure 2. EEMD results for monthly SOI (a) monthly data and EEMD trend (b) Variance versus Average Period for SOI and white noise.

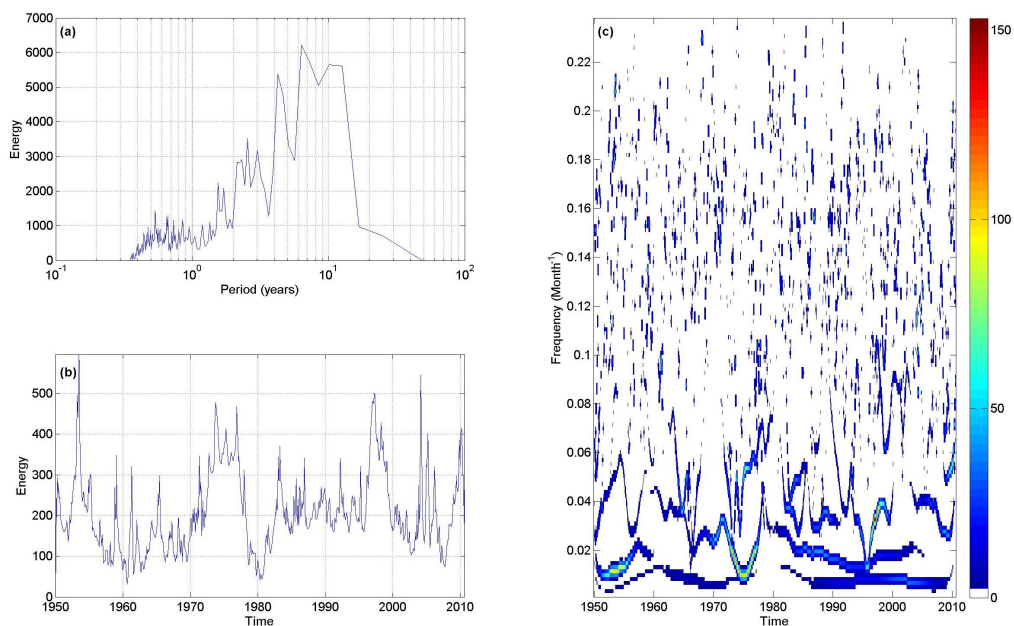


Figure 3. EEMD Hilbert spectrum results for monthly SOI data (a) marginal distribution with respect to period; (b) marginal distribution with respect to time and (c) full spectrum.

Figure 2 presented the results in terms of average period. Application of the Hilbert transform to the IMFs allows the instantaneous frequency at each time step to be calculated and this information is summarised in the Hilbert spectrum (Figure 3). The energy associated with IMFs 4, 5 and 6 are now seen in Figure 3a and between frequencies 0.035 – 0.008 in Figure 3c. The dominant frequencies in the monthly SOI series are observed at 2.5, 4, 6 and 10 years. Changes in SOI energy with time are shown in Figure 3b and two periods of sustained high energy (>300) from 9/1972 – 4/1977 and 7/1996 – 3/1999 are apparent. These two periods coincide with regime shifts within the Pacific ocean/atmosphere system that have been associated with changes in the Pacific Decadal Oscillation (Mantua and Hare, 2002). In summary, the EEMD algorithm appears capable of identifying the expected structure within the SOI series.

3.2. Indian Ocean Dipole mode index

The Indian Ocean Dipole (IOD) is a tropical Indian ocean/atmosphere phenomenon (Saji *et al.*, 1999). The Dipole Mode Index (DMI), the difference in sea surface temperature between the western and eastern tropical Indian Ocean (Saji *et al.*, 1999), is shown in Figure 4a from 1/1950 – 9/2010. Also shown in Figure 4a is the EEMD residual (trend), which indicates DMI has little trend other than a slightly higher period during the 1970s. In Figure 4b the variance and average period of the DMI EEMD IMF results are plotted along with EMD IMF results for 200 white noise series of the same length and variance as the original DMI series with added white noise. Dominant frequency DMI IMFs that plot above the white noise cloud are IMFs 3, 4 and, to a lesser extent, 6. These IMFs have average periods of 1.3, 3.2 and 11 years respectively.

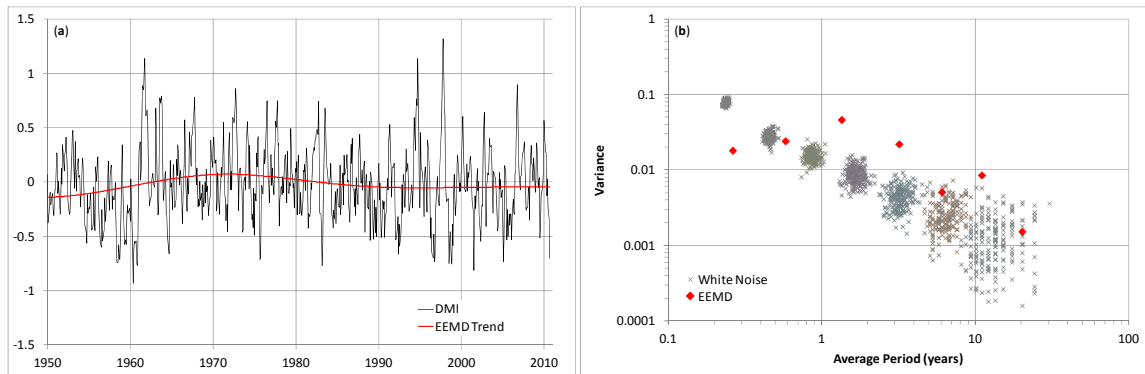


Figure 4. EEMD results for monthly DMI (a) monthly data and EEMD trend (b) Variance versus Average Period for DMI and white noise.

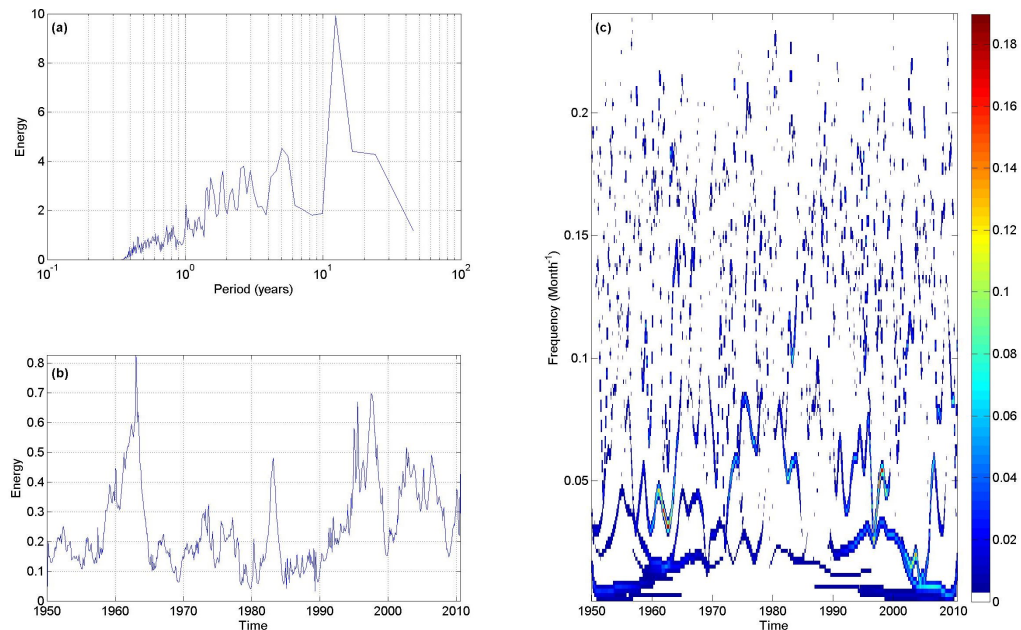


Figure 5. EEMD Hilbert spectrum results for monthly DMI data (a) marginal distribution with respect to period; (b) marginal distribution with respect to time and (c) full spectrum.

Hilbert spectrum results for DMI are shown in Figure 5. A well defined peak in energy at 12 years is observed in Figure 5a and is spread across frequencies below 0.05 in Figure 5c. Interestingly the energy associated with IMFs 3, 4 and 6 in Figure 4b are not uniformly observed in Figure 5a. The influence of IMF 6 at the 12 year period is apparent, but IMFs 3 and 4 appear less dominant than Figure 4b would suggest. A possible explanation for this is seen in Figures 5c where a period of moderate energy (light blue) persists at a constant ~ 12 year period for several years during the mid 2000s. This extended period may be an EMD post processing end of record effect that artificially inflates the 12 year period (Figure 5a) relative to the other dominant frequencies. Changes in DMI energy with time are shown in Figure 5b with three periods of sustained high energy observed: the early 1960's (peak 1963); early 1980's (peak 1983); and a sustained period since the mid 1990's (peak 1997). The last period coincides with a regime shift within the Pacific system and the second may be related to the strong El Niño of 1982/83. The cause of the high energy period in the early 1960's remains for further research.

3.3. Tasman Sea index

The Tasman Sea Index (TSI), a metric based on sea surface temperatures adjacent to south eastern Australia (Murphy and Timbal, 2008), is shown in Figure 6a from 1/1950 – 9/2010. Also shown in Figure 6a is the EEMD residual (trend), which indicates TSI has increased over the period of record, particularly after the mid 1990s. In Figure 6b the variance and average period of the TSI EEMD IMF results are plotted along with EMD IMF results for 200 white noise series of the same length and variance as the original TSI series with added white noise. Dominant frequency TSI IMFs that plot above the white noise cloud are IMFs 3 and 4. These IMFs have average periods of approximately 3 and 12 years respectively.

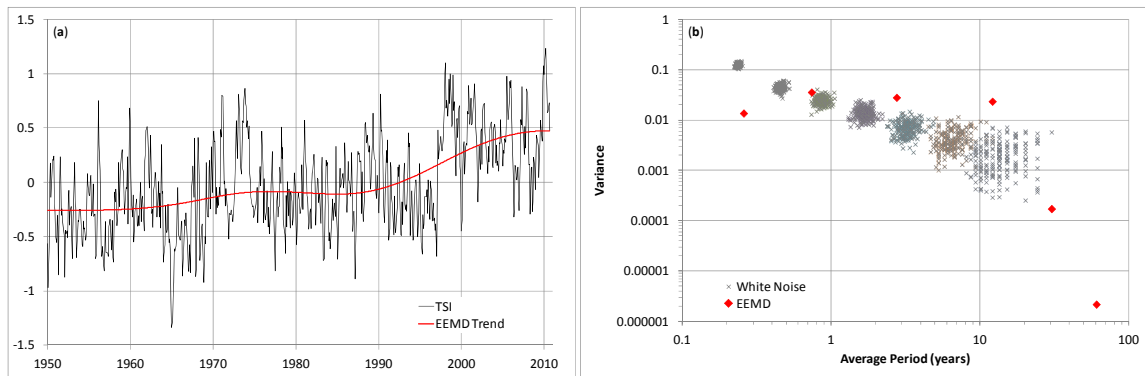


Figure 6. EEMD results for monthly TSI (a) monthly data and EEMD trend (b) Variance versus Average Period for TSI and white noise.

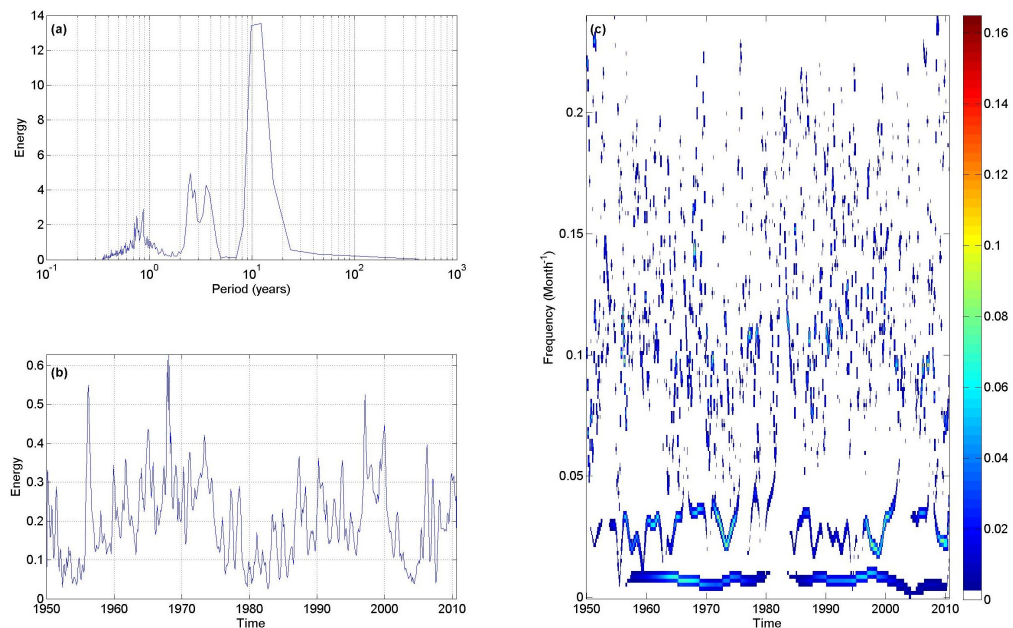


Figure 7. EEMD Hilbert spectrum results for monthly TSI data (a) marginal distribution with respect to period; (b) marginal distribution with respect to time and (c) full spectrum.

Hilbert spectrum results for TSI are shown in Figure 7, with peaks in energy occurring at sub annual, 2.5-3.5 years and 10 – 12 years (Figure 7a and 7c). The latter two energy peaks are associated with IMFs 3 and 4. The dominant frequencies in the monthly TSI series are observed around 3 and 10 years. Although these peak energy periods are similar to those observed for SOI, they are more clearly pronounced in the TSI. Changes in TSI energy with time (Figure 7b) show a prolonged period of high energy in the 1960s/early 1970s and a shorter period in the late 1990s. Again the last period coincides with a known regime shift within the Pacific system and in this case a change in trend (Figure 6a). The cause of the prolonged first period of high energy during the 1960s – 70s remains for further research.

4. CONCLUSIONS

The recently developed spectral analysis technique Ensemble Empirical Mode Decomposition was outlined and issues regarding its application to the Hilbert spectrum discussed. Application of EEMD to three climatic indices relevant to Australian hydroclimatic variability revealed EEMD is capable of identifying physical process signals. For example, the known spectral structure of the SOI was identified by the EEMD algorithm. It was observed during analysis of the Indian Ocean Dipole mode index that EMD post processing may introduce end effects previously minimised by the EEMD ensemble average. These effects may inflate existing features, or potentially introduce low frequency artefacts, though this danger can be mitigated through careful examination of the complete Hilbert spectrum. EEMD and Hilbert spectrum analysis of hydroclimatic time series provide insights into long-term trends, dominant frequencies and changes in energy with time for series that are usually problematic to analyse with traditional spectral analysis techniques. The spectral fingerprints of the climatic indices assessed here are used in a companion paper that looks for those fingerprints within Australian precipitation and temperature data (Srikanthan *et al.*, 2011).

ACKNOWLEDGMENTS

Australian Research Council grant LP100100756, Melbourne Water and the Australian Bureau of Meteorology financially supported this research. Discussions with Prof Geoff Pegram are gratefully acknowledged.

REFERENCES

- Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., and Liu, H.H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Roy. Soc. London A*, 454, 903-995.
- Mantua, N.J., and Hare, S.R. (2002). The Pacific Decadal Oscillation. *Journal of Oceanography*, 58, 35-44.
- Murphy, B.F., and Timbal, B. (2008). A review of recent climate variability and climate change in southeastern Australia. *International Journal of Climatology*, 28, 859-879.
- Nicholls, N. (1988). El Niño–Southern Oscillation and rainfall variability. *Journal of Climate*, 1, 418–421.
- Peel, M.C., McMahon, T.A., and Finlayson, B.L. (2002). Variability of annual precipitation and its relationship to the El Niño–Southern Oscillation. *Journal of Climate*, 15(5), 545-551.
- Peel, M.C., McMahon, T.A., and Finlayson, B.L. (2004). Continental differences in the variability of annual runoff - update and reassessment. *Journal of Hydrology*, 295, 185-197.
- Peel, M.C., McMahon, T.A., and Pegram, G.G.S. (2009). Assessing the performance of rational spline based Empirical Mode Decomposition using a global annual precipitation dataset. *Proc. Roy. Soc. London A*, 465, 1919-1937.
- Peel, M.C., McMahon, T.A., Srikanthan, R., and Tan, K.S. (2011). Ensemble Empirical Mode Decomposition: Testing and objective automation. Proceedings of the 33rd Hydrology and Water Resources Symposium, Brisbane, Engineers Australia, pp: 702-709.
- Pegram, G.G.S, Peel, M.C., and McMahon, T.A. (2008). Empirical mode decomposition using rational splines: an application to rainfall time series. *Proc. Roy. Soc. London A*. 464: 1483-1501.
- Saji, N.H., Goswami, B.N., Vinayachandran, P.N., and Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature*, 401, 360-363.
- Srikanthan, R., Peel, M.C., McMahon, T.A., and Karoly, D.J. (2011). Ensemble Empirical Mode Decomposition of Australian monthly rainfall and temperature data. MODSIM 2011.
- Wu, Z. and Huang, N.E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method, *Advances in Adaptive Data Analysis*, 1(1), 1-41.