

Ensemble empirical mode decomposition of Australian monthly rainfall and temperature data

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

^a *Climate and Water Division, Bureau of Meteorology, Melbourne,* ^b *Department of Infrastructure Engineering, The University of Melbourne, Victoria,* ^c *School of Earth Sciences, University of Melbourne, Victoria*

Email: r.srikanthan@bom.gov.au

Abstract: Empirical Mode Decomposition (EMD), developed by Huang *et al.* (1998), is a form of adaptive time series decomposition. Traditional forms of spectral analysis, like Fourier, assume that a time series (either linear or nonlinear) can be decomposed into a set of linear components. However, as the degree of non-periodic behaviour and non-stationarity in a time series increases, the set of linear components describing that time series increases substantially when using Fourier techniques. In the physical sciences, time series are often non-periodic, more stochastic and even non-stationary, so Fourier based spectral analysis techniques often produce large sets of physically meaningless harmonics when applied to these problems (Huang *et al.* 1998). In contrast the EMD method does not assume a time series is linear or stationary prior to analysis. EMD adaptively decomposes a time series into a set of independent intrinsic mode functions (IMFs) and a residual component. When the IMFs and residual are summed together they form the original time series.

An inconvenient feature of EMD is mode mixing, where a fluctuation of given frequency may split across two IMFs. The data driven adaptive iterative nature of the EMD algorithm means mode mixing is difficult to avoid without subjectively deciding on the likely nature of any signal to be extracted prior to analysis. Mode mixing between IMFs is problematic when investigating the physical significance of IMFs, as an expected physical signal may be present but split across IMFs. Wu and Huang (2009) proposed Ensemble EMD (EEMD), a noise assisted data analysis method, as a way of overcoming the mode mixing problem. In EEMD, an ensemble of EMD trials is obtained by adding finite amplitude normally distributed white noise to the time series prior to each EMD run. The IMFs and residual from each trial are grouped by IMF order into ensembles and the IMF and residual ensemble averages form the EEMD result. Since the white noise is different for each EMD trial the noise cancels out during averaging as the ensemble size increases. However, the noise serves the useful purpose of changing the ordering of local maxima and minima within the time series, thus provoking a different EMD outcome in each trial. Wu and Huang (2009) believe EEMD provides more physically meaningful IMFs and residual than traditional EMD.

The EEMD is carried out using monthly rainfall and temperature data from 44 Australian rainfall and ten climate stations respectively. The results from the analysis are presented in terms of dominant frequencies, trends, changes in variance with time (Hilbert spectrum) and relationships between climate indices and rainfall and temperature. For the monthly rainfall data, almost all the stations exhibited an annual cycle while a number of stations showed cycles with periods about 3 and 20 years. The mean monthly temperature data only showed an annual cycle. However, the residuals from the EEMD analysis of the temperature data showed a steadily increasing trend for all ten sites.

Keywords: *Ensemble Empirical Mode Decomposition (EMD), Hilbert Spectrum, intrinsic mode function, residual*

1. INTRODUCTION

Empirical Mode Decomposition (EMD), developed by Huang *et al.* (1998), is a form of adaptive time series decomposition. Traditional forms of spectral analysis, like Fourier, assume that a time series (either linear or nonlinear) can be decomposed into a set of linear components. However, as the degree of non-periodic behaviour and non-stationarity in a time series increases, the set of linear components describing that time series increases substantially when using Fourier techniques. In the physical sciences, time series are often non-periodic, more stochastic and even non-stationary, so Fourier based spectral analysis techniques often produce large sets of physically meaningless harmonics when applied to these problems (Huang *et al.* 1998). In contrast the EMD method does not assume a time series is linear or stationary prior to analysis, it lets the data speak for themselves. EMD adaptively decomposes a time series into a set of independent intrinsic mode functions (IMFs) and a residual component. When the IMFs and residual are summed together they form the original time series.

An inconvenient feature of EMD is mode mixing, where a fluctuation of given frequency may split across two IMFs (Peel, et al., 2011a). The data driven adaptive iterative nature of the EMD algorithm means mode mixing is difficult to avoid without subjectively deciding on the likely nature of any signal to be extracted prior to analysis. Mode mixing between IMFs is problematic when investigating the physical significance of IMFs, as an expected physical signal may be present but split across IMFs. Wu and Huang (2009) proposed Ensemble EMD (EEMD), a noise assisted data analysis method, as a way of overcoming the mode mixing problem. In EEMD, an ensemble of EMD trials is obtained by adding finite amplitude normally distributed white noise to the time series prior to each EMD run. The IMFs and residual from each trial are grouped by IMF order into ensembles and the IMF and residual ensemble averages form the EEMD result. Since the white noise is different for each EMD trial the noise cancels out during averaging as the ensemble size increases. However, the noise serves the useful purpose of changing the ordering of local maxima and minima within the time series, thus provoking a different EMD outcome in each trial. Wu and Huang (2009) believe EEMD provides more physically meaningful IMFs and residual than traditional EMD.

An earlier study carried out on annual rainfall data from 44 rainfall stations using EMD did not result in any identifiable cycles in the data (Srikanthan et al., 2006). Here EEMD is carried out using monthly rainfall and temperature data from 44 Australian rainfall and ten climate stations respectively. From the results of the analysis, dominant periods of the IMFs and the trends in the residual are presented along with the changes in variance with time (Hilbert spectrum) and relationships between climate indices and rainfall and temperature are discussed.

2. AUSTRALIAN RAINFALL AND TEMPERATURE DATA

Monthly rainfall data from 44 stations and monthly mean temperature from ten stations located in various parts of Australia are used in the study. Locations of the rainfall stations and the temperature stations are shown in Figure 1. The rainfall data length varies from 69 to 143 years and the mean annual rainfall of the stations varies from 164 to 1550 mm. The temperature data length varies from 62 to 155 years.

3. EEMD ANALYSIS OF RAINFALL AND TEMPERATURE DATA

EEMD is a noise assisted data analysis technique in which an ensemble of EMD runs is obtained for a time series. The following brief description of EEMD is reproduced from the companion paper of Peel et al. (2011b). Prior to each EMD run, finite amplitude normally distributed white noise is added to the series. The EMD algorithm (Peel et al. 2009) 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.

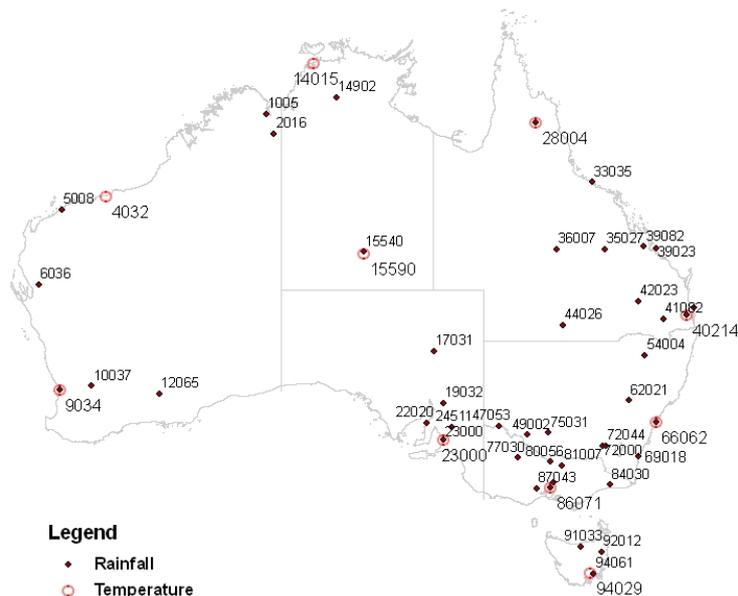


Figure 1. Location of rainfall and temperature stations.

3.1. Rainfall

EEMD analysis was first carried out on the monthly rainfall data resulting in IMFs and a residual. In order to identify potential dominant frequencies the variance of the IMFs is plotted against their average period along with IMF results from EMD of 200 white noise sequences of the same length and variance as the rainfall series. These plots are not presented in the paper due to lack of space. The average period of the IMFs are presented in Table 1. Apart from the annual cycle, there is no consistent pattern in the periods of the IMFs.

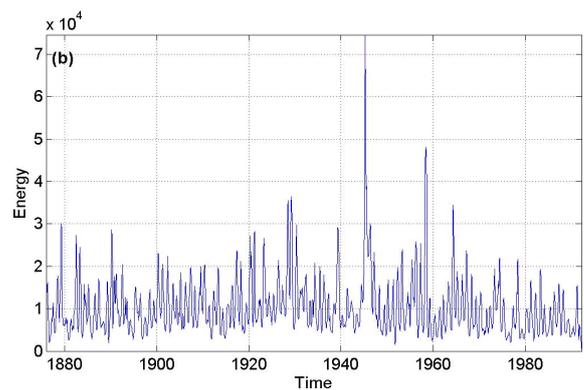
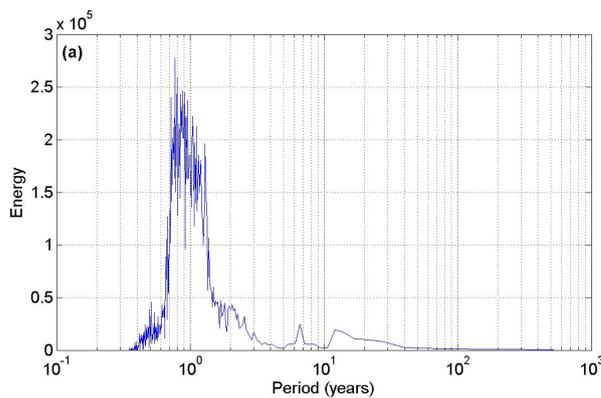
Besides plots of the variance of the IMFs against their average period, dominant frequencies were also assessed through the Hilbert's spectrum (Huang *et al.*, 1998). The marginal Hilbert spectrum with respect to frequency and time for three stations are shown in Figure 2. Only an annual cycle is noticeable for Perth (9034). Emerald (35027) shows a fluctuation of period 3 years (may be related to ENSO) and another about 20 years in addition to the annual cycle. Moruya (69018) shows two fluctuations of period about 10 and 20 years in addition to the annual cycle. From the plots of energy versus time, high energy (variance) was observed in around 1950 for Perth, 1920 and 1950 for Emerald and 1920 and 1980 for Moruya. Figure 3 shows a time series plot of the residual and monthly rainfall for Emerald. There is no noticeable trend in the residual.

3.2. Temperature

The average periods of the IMFs for the temperature data are presented in Table 2. The annual cycle is clearly represented by IMF2. The marginal Hilbert spectrum for Perth (9034) is shown in Figure 4. This only shows an annual cycle. The same result was found for the other nine sites as well. A time series plot of the residual and mean monthly temperature for Perth is shown in Figure 5. The residual plot shows a steady increase in temperature over time resulting in about 1°C over the last 100 years. For the remaining nine sites, the residuals show an increase of 0.13°C (Adelaide) to 1.68°C (Port Headland) over the last 100 years.

Table 1. Average period (months) of IMFs of monthly rainfall. IMFs that are significantly different from white noise at the 99% level of significance are shown in bold (Wu and Huang, 2004).

Station #	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6	Station #	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6
1005	3.6	5.9	12.1	18.9	39.7	102.3	41082	2.9	5.4	10.9	19.7	37.2	74.4
2016	3.5	5.9	12.0	19.4	43.4	236.5	42023	3.0	8.4	23.6	76.3	161.1	414.3
5008	3.4	5.7	13.7	63.9	179.8	719.0	44026	2.9	5.4	10.5	24.0	87.9	351.6
6036	3.1	5.7	12.3	46.0	127.7	383.0	47053	2.9	7.3	21.3	70.0	263.7	685.6
9034	3.0	11.8	25.9	77.6	214.8	465.3	49002	2.8	4.9	8.9	16.6	31.6	62.0
10037	2.9	5.7	11.5	24.8	73.8	152.0	54004	2.9	5.2	9.6	15.7	32.2	65.1
12065	3.0	7.8	18.5	39.0	124.2	341.5	62021	2.9	7.2	22.8	82.9	323.2	1077.3
14902	3.4	12.0	18.8	39.0	92.8	336.3	66062	3.0	5.7	11.2	18.8	33.5	63.0
15540	3.1	8.0	21.4	60.8	124.3	303.8	69018	3.0	5.3	12.0	31.0	63.7	141.3
17031	3.1	5.5	12.9	40.1	120.2	250.5	72000	2.9	5.3	9.9	18.7	39.3	71.2
19032	2.9	5.1	9.7	16.8	32.5	53.7	72044	2.9	7.6	24.2	63.7	186.8	400.3
22020	2.9	5.5	11.2	20.3	37.9	71.9	75031	2.8	5.1	9.3	15.8	31.3	58.0
23000	2.9	5.7	11.8	22.1	46.1	78.2	77030	2.9	5.5	12.7	45.2	248.3	596.0
24511	2.9	8.8	19.1	49.5	173.3	520.0	80056	2.8	5.2	9.8	18.8	33.3	60.1
28004	3.3	6.0	12.0	18.7	49.2	207.3	81007	2.9	8.0	26.8	75.6	233.8	514.4
33035	3.2	5.8	12.0	22.4	85.1	248.2	84030	2.9	5.5	11.7	41.4	109.5	306.6
35027	3.0	5.8	12.2	44.5	164.3	657.0	86071	2.8	5.0	9.7	16.3	28.9	56.6
36007	3.1	5.7	12.4	41.3	123.9	297.4	86117	2.8	5.0	8.9	16.1	29.0	66.1
39023	3.1	5.8	11.8	25.5	70.5	264.5	87043	2.9	5.0	9.5	17.5	32.0	67.9
39082	3.1	5.7	12.0	48.4	152.8	496.5	91033	2.9	5.5	12.0	21.1	41.6	72.5
40043	2.9	4.9	9.7	16.6	33.2	73.7	92012	2.9	5.0	9.3	17.5	30.9	58.7
40214	2.9	5.6	12.0	42.8	191.3	542.0	94061	3.0	5.5	10.4	18.6	36.3	77.4



(a) Perth

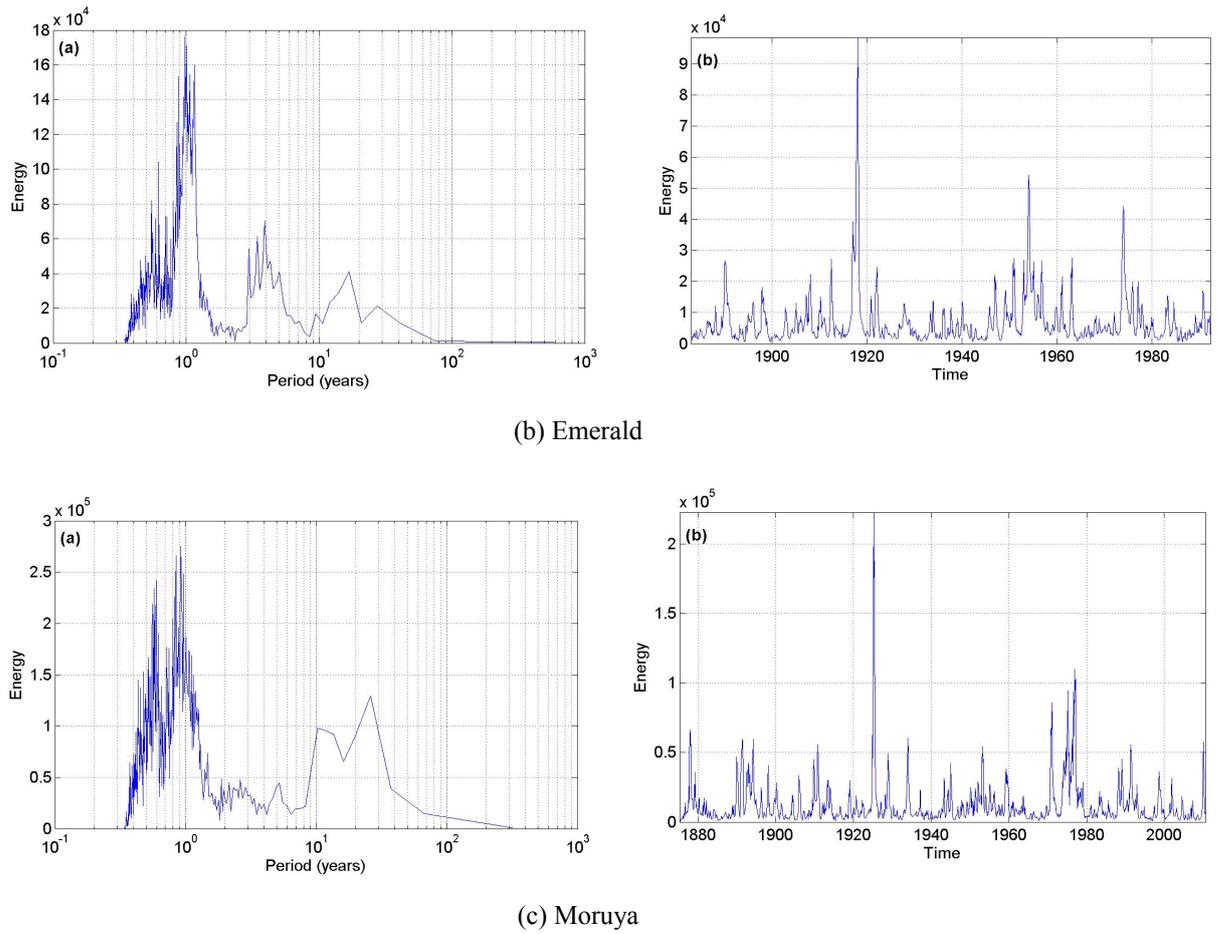


Figure 2. Marginal Hilbert spectrum plots of energy versus period and energy versus time for Perth, Emerald and Moruya.

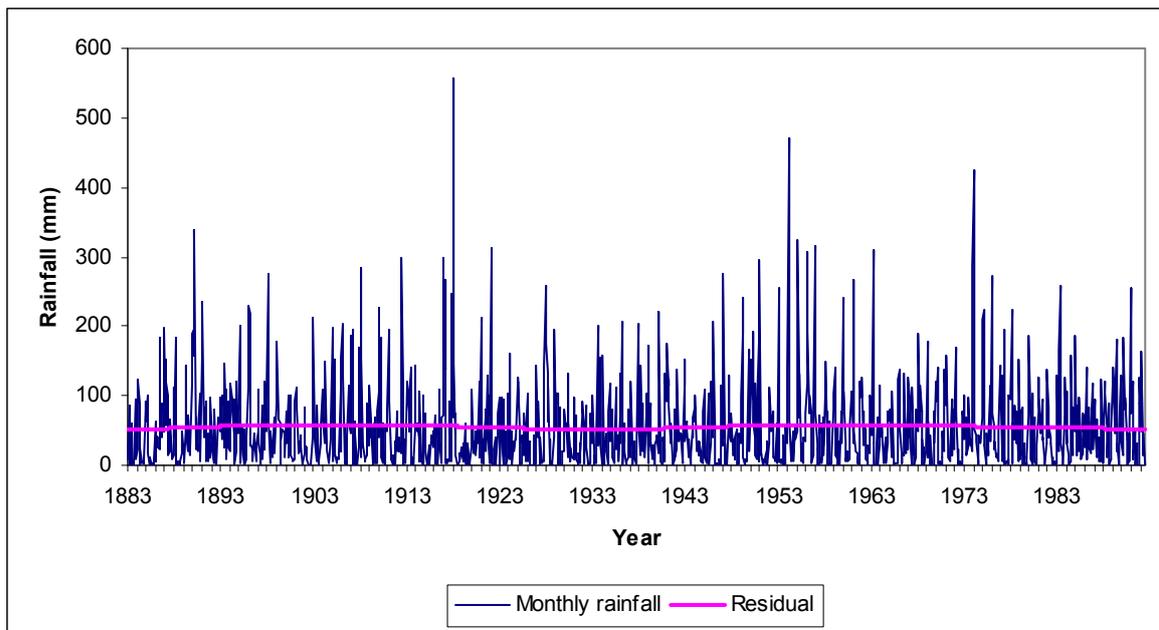


Figure 3. Plot of monthly rainfall and EEMD residual for Emerald.

Table 2. Average period (months) of IMFs of monthly mean temperature. IMFs that are significantly different from white noise at the 99% level of significance are shown in bold (Wu and Huang, 2004).

Station #	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	Station #	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5
4032	3.1	11.9	36.3	124.0	297.6	28004	3.1	11.9	34.9	139.8	396.0
9034	2.8	11.9	37.4	134.1	456.0	40214	2.9	11.9	40.3	158.4	475.2
14015	3.4	11.9	24.0	60.0	168.0	66062	2.9	12.0	39.2	130.3	729.6
15590	3.0	11.9	32.5	72.0	331.2	86071	2.8	12.0	27.6	67.6	218.8
23000	2.9	11.9	31.5	71.2	315.4	94029	2.9	11.9	41.0	134.9	404.6

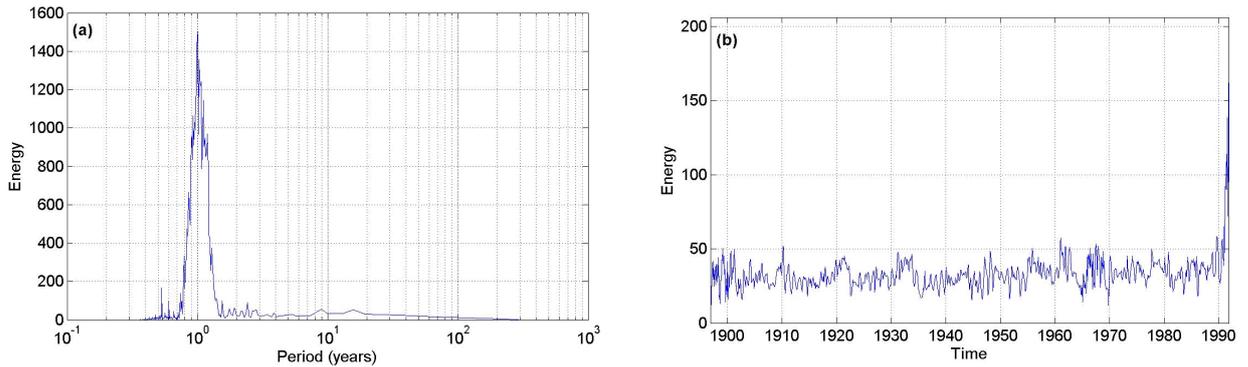


Figure 4. Marginal Hilbert spectrum plots of energy versus period and energy versus time for Perth.

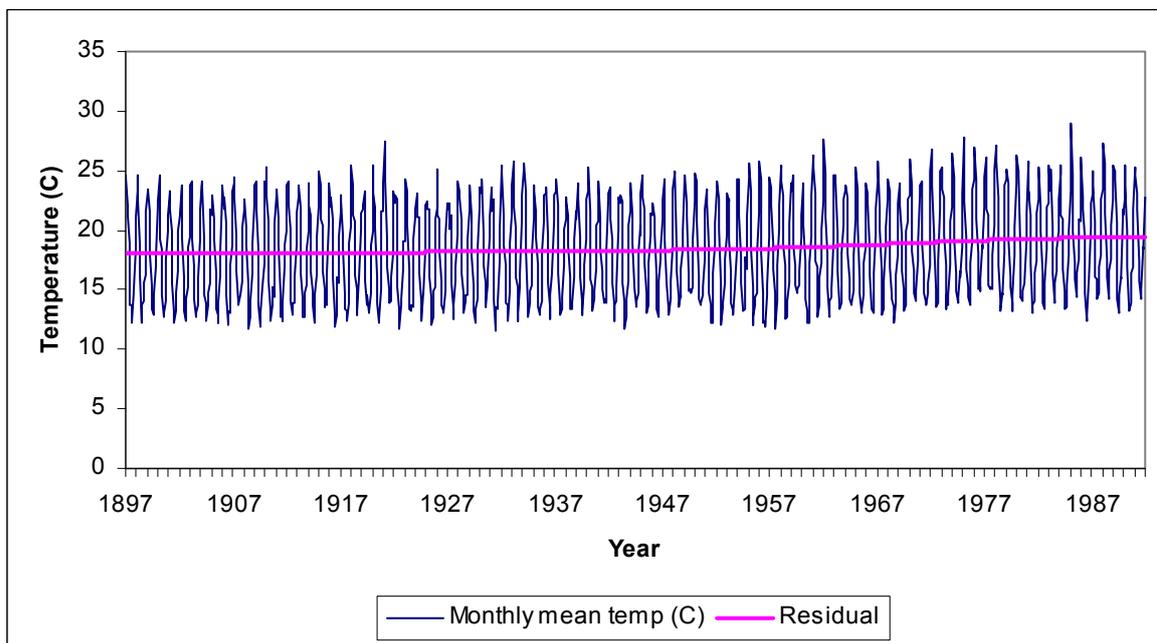


Figure 5. Plot of mean monthly temperature and EEMD residual for Perth.

4. DISCUSSION AND CONCLUSIONS

EEMD analysis was carried out on monthly rainfall data from 44 stations and mean monthly temperature data from ten stations. For the monthly rainfall data, almost all the stations exhibited an annual cycle while a number of stations showed cycles with periods about 3 and 20 years. There were some varying fluctuations

observed in the rainfall residuals from the EEMD analysis with some showing no trend. A plot of the residuals from the EEMD analysis of monthly rainfall data from five sites is shown in Figure 6. The residuals are standardised to 100 with respect to the largest value for each site and expressed as a percentage in Figure 6. The mean monthly temperature data only showed an annual cycle. However, the residuals from the EEMD analysis showed a steadily increasing trend for all ten sites.

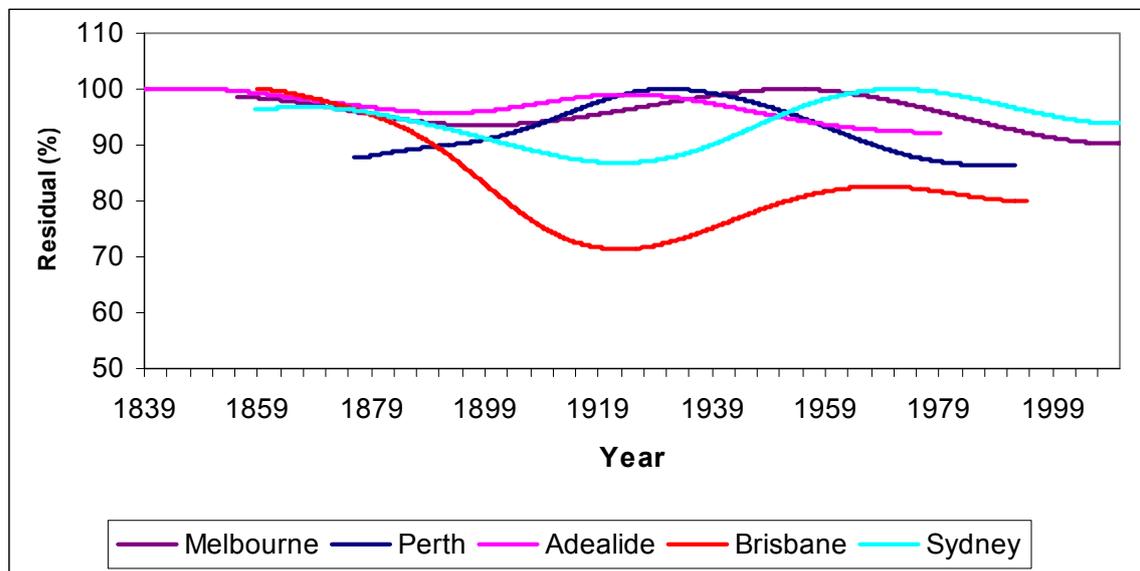


Figure 6. Plot of EEMD residuals for Perth, Adelaide, Brisbane, Sydney and Melbourne rainfall.

In order to determine any relationship between the rainfall and climate indices, dominant cycles in the climate indices were obtained. The climate indices SOI, Nino3, Nino34, Nino4, EMI, D20, DMI, WPI, EPI, TSI and B140 all have energy peaks for periods in the range ~2 to ~10 years. Since the periods of cycles for rainfall and the climate indices are in the same range, there may be some relationship between the two. Work is in progress to quantify the relationship between the IMFs of rainfall and climate indices.

ACKNOWLEDGMENTS

This research was financially supported by Australian Research Council grant LP100100756, Melbourne Water and the Australian Bureau of Meteorology. 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. & Liu, H.H. (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society London A*, 454, 903-995.
- 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. (2011a) Ensemble Empirical Mode Decomposition: Testing and objective automation. *Proceedings of the 33rd Hydrology and Water Resources Symposium, Brisbane, Engineers Australia*, pp: 702-709.
- Peel, M.C., Srikanthan, R., McMahon, T.A., and Karoly, D.J. (2011b) Ensemble Empirical Mode Decomposition of monthly climatic indices relevant to Australian hydroclimatology. MODSIM 2011.
- Srikanthan, R., Peel, M. C., Pegram, G.G.S. and McMahon, T.A. (2006) Stochastic modelling of annual rainfall data. *Hydrology and Water Resources Symposium, Engineers Australia*.
- Wu Z. and Huang N.E. (2004) A study of the characteristics of white noise using the empirical mode decomposition method. *Proc. Roy. Soc. London A*, 460: 1597-1611.
- 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.