Regional Hydrological Impacts of Climatic Change—Hydroclimatic Variability
(Proceedings of symposium S6 held during the Seventh IAHS Scientific Assembly at Foz do Iguaçu, Brazil, April 2005). IAHS Publ. 296, 2005.

Identification of oscillations in historical global streamflow data using empirical mode decomposition

FRANCIS H. S. CHIEW[1], MURRAY C. PEEL[1], GNANATHIKKAM E. AMIRTHANATHAN[1] & GEOFFREY G. S. PEGRAM[2]

1 *Department of Civil and Environmental Engineering, University of Melbourne, Victoria 3010, Australia*
f.chiew@civenv.unimelb.edu.au

2 *Civil Engineering, University of KwaZulu-Natal, Durban 4041, South Africa*

Abstract The hydroclimatic variability that occurs over various time scales has implications for the management of land and water resources systems. This paper describes the use of the empirical mode decomposition (EMD) method to identify oscillations/variability in annual streamflow time series. The EMD method is also applied to bootstrap replicates of the original time series to test the statistical significance of the identified oscillations. The EMD is a relatively new technique and has several advantages over spectral time series analysis techniques. It is also relatively easy to use as the intrinsic oscillations are automatically and adaptively selected from the time series. The EMD analysis is applied to annual streamflow time series from 20 catchments across the world. The results indicate that statistically significant oscillations of 11–14 and 20–25 years are observed in some, but not all, of the time series.

Key words cycles; empirical mode decomposition; global; hydroclimatic; interdecadal; oscillations; streamflow; variability

INTRODUCTION

Hydroclimatic variability occurs over various time scales (seasonal, inter-annual, 3–7 year oscillations associated with El Niño/Southern Oscillation (ENSO), interdecadal, and "climate change"). The management of land and water resources involves designing and operating to cope with this variability. For example, all water resources projects take into account seasonal and inter-annual variability, some authorities use ENSO-based seasonal forecasts for operational water management, and most are now concerned with the potential impacts of climate change on hydrology and water resources.

This paper uses the empirical mode decomposition (EMD) method to identify oscillations in historical annual streamflow time series for 20 unimpaired catchments from different parts of the world. In EMD analysis, a time series is decomposed into a set of intrinsic mode functions (IMFs) that are mutually independent. The decomposition is based on the direct extraction of energy (variance) associated with various intrinsic time scales that are automatically and adaptively selected from the time series. The EMD is a relatively new technique that is able to deal with both nonlinear and non-stationary data, and has several advantages over other spectral analysis techniques.

The two objectives of this paper are to demonstrate the application of the EMD method on annual streamflow data and to make some general observations about oscillations in historical global streamflow data.

EMPIRICAL MODE DECOMPOSTION

The EMD method was developed by Huang *et al.* (1998) for adaptively decomposing signals, and is described in detail in Huang *et al.* (1998, 1999, 2003). The decomposition is based on the assumption that a time series is formed by the linear superposition of different natural oscillations in a system. The decomposed oscillations resulting from the EMD analysis are called intrinsic mode functions (IMFs), with the first IMF capturing the highest frequency oscillation and the last IMF capturing the lowest frequency oscillation. Unlike other spectral analysis methods, EMD can easily handle amplitude and frequency modulations in the natural oscillations (without the need to add harmonics), as well as non-stationarity (trend) in the data.

Obtaining the IMFs

The EMD method is described here using an artificial time series as an example (see Fig. 1, labelled "combined"). The 99 years of artificial data are made up of three components: a 7-year sine wave cycle, a 22-year sine wave cycle, and a step jump in the mean (90 in the first 50 years and 110 in the last 49 years).

There are four steps involved in obtaining the first estimate of an IMF, as illustrated in Fig. 2. The first step is to identify all the local extrema (see Fig. 2(a)). The second step is to connect all the local maxima with a cubic spline and all the local minima with another cubic spline (see Fig. 2(b)). The third step is to construct the mean of the upper and lower envelopes fitted by the splines (see Fig. 2(c)). The fourth step is to subtract the mean (Fig. 2(c)) from the original time series data to obtain the first estimate of the first IMF (Fig. 2(d)).

Sifting to obtain an IMF

Each IMF must satisfy two conditions: (i) the number of extrema (maxima + minima) and the number of zero crossings must be equal or differ by one; and (ii) the mean of the upper and lower envelopes (ΔM – see Fig. 2(c)) must be close to zero.

To satisfy the above conditions, a sifting process is carried out where the four steps illustrated in Fig. 2 are repeated, each time using the latest estimate of the IMF as the input data. The iteration process eliminates the riding waves in the IMF and smoothes the uneven amplitudes. Sufficient iterations should be carried out in the sifting process to ensure that the above conditions are met, but too many iterations will lead to over smoothing, which results in a loss of information about the natural oscillations.

Various criteria for terminating the sifting process have been suggested (see Huang *et al.*, 1999, 2003; Quek *et al.*, 2003; Rilling *et al.*, 2003). However, our analysis with various annual streamflow time series indicates that satisfying condition (i) (number

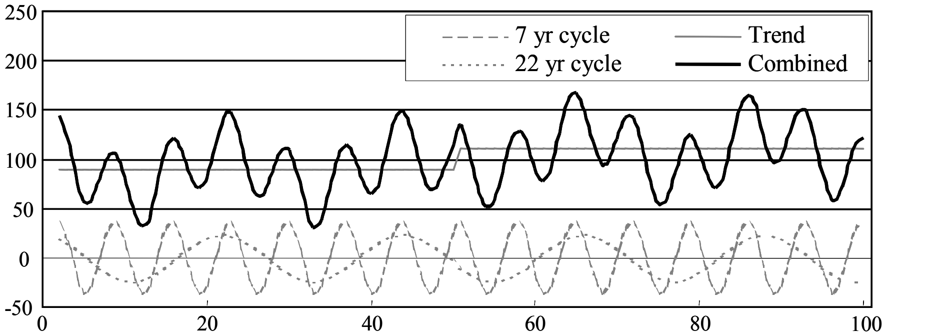

Fig. 1 Artificial time series made up of three components (7-year cycle, 22-year cycle and step jump in mean).

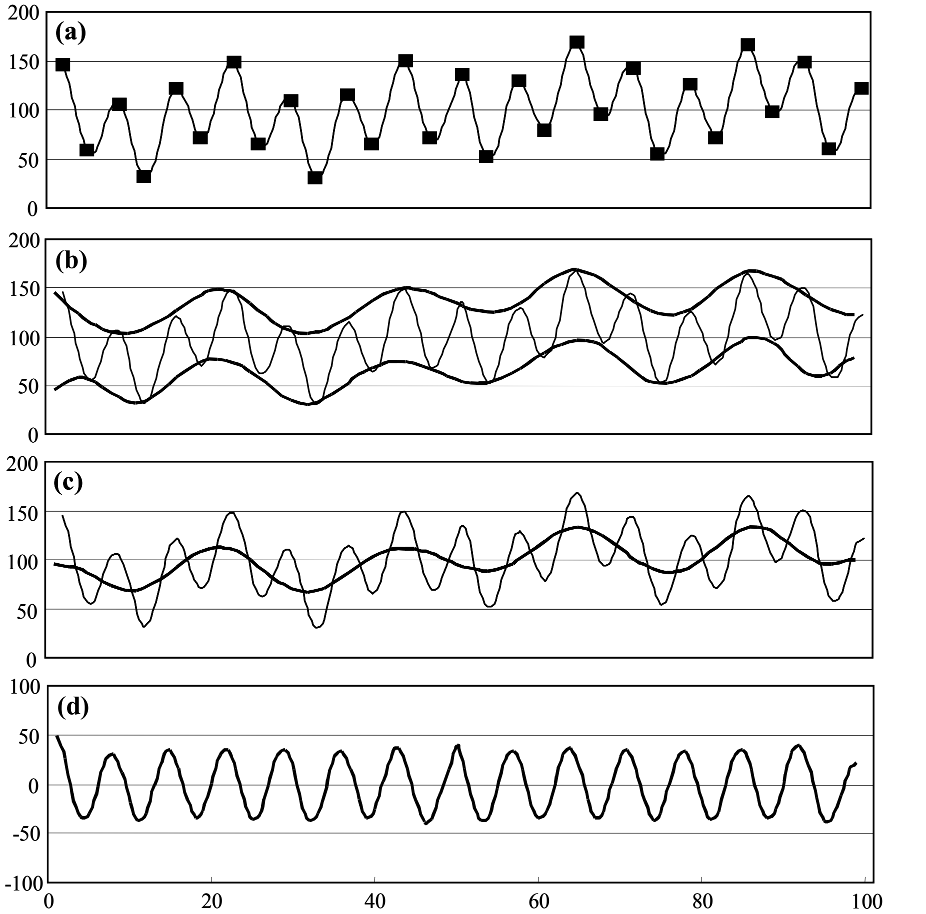

Fig. 2 Steps to obtain the first estimate of the first IMF.

of extrema and number of zero crossings must be equal or differ by one), and ensuring that the number of extrema and number of zero crossings remain the same for five successive iterations, generally resulted in a very small ΔM and very small differences between ΔM in the final successive iterations. This termination criterion, which is used in the following EMD analysis, is consistent with the observations and recommendation of Huang *et al*. (2003).

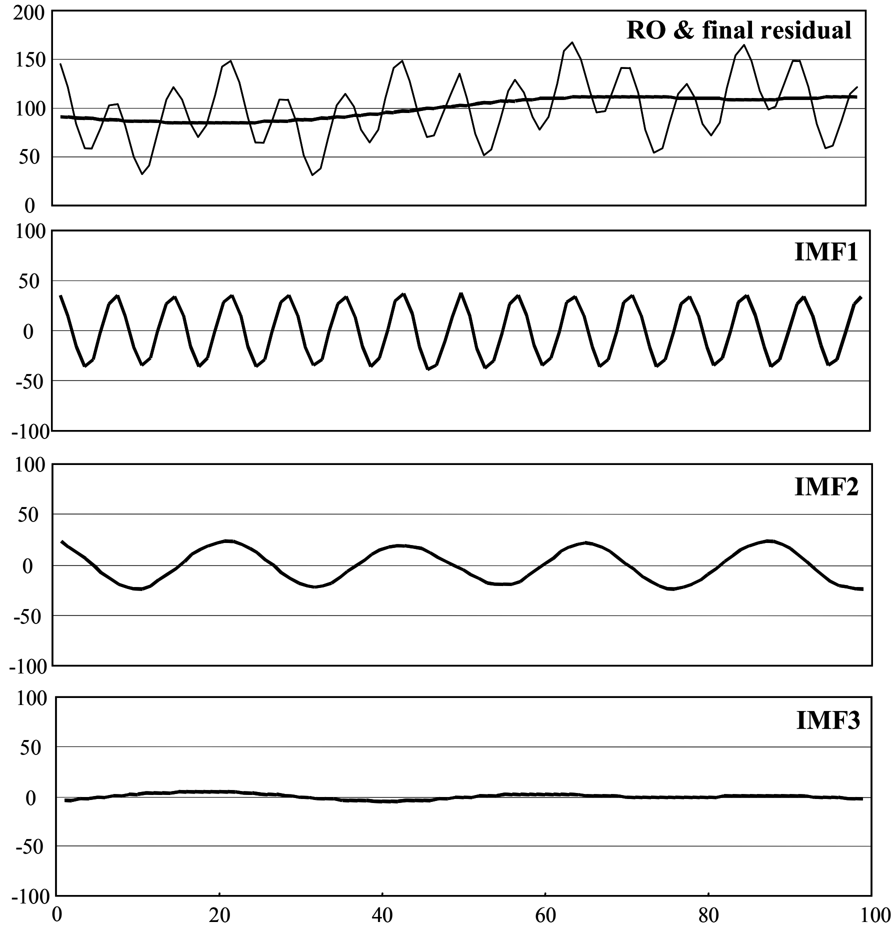

Fig. 3 IMFs and final residual from EMD analysis of artificial time series.

IMFs from the EMD method

Once the sifting process has terminated and an IMF has been defined (e.g. IMF1), a residual (R1) is obtained by subtracting the IMF from the original data (R0). The residual is then used as the input data, and the above steps are repeated to obtain each subsequent IMF. This process of extracting IMFs from the original time series is carried out until the residual does not contain a complete cycle. This final residual represents the trend in the data, which may be an incomplete cycle with a longer period than the length of record, or it may be a monotonic trend. The sum of the IMFs and the final residual is the original time series.

The IMFs and the final residual obtained from EMD analysis of the artificial data in Fig. 1 are shown in Fig. 3. IMF1 is a 7-year cycle, IMF2 is a 22-year cycle, IMF3 is a 34-year oscillation (noise/error from the analysis representing a small proportion of the total variance), and the final residual shows the trend in the data (the step jump in mean of the artificial data).

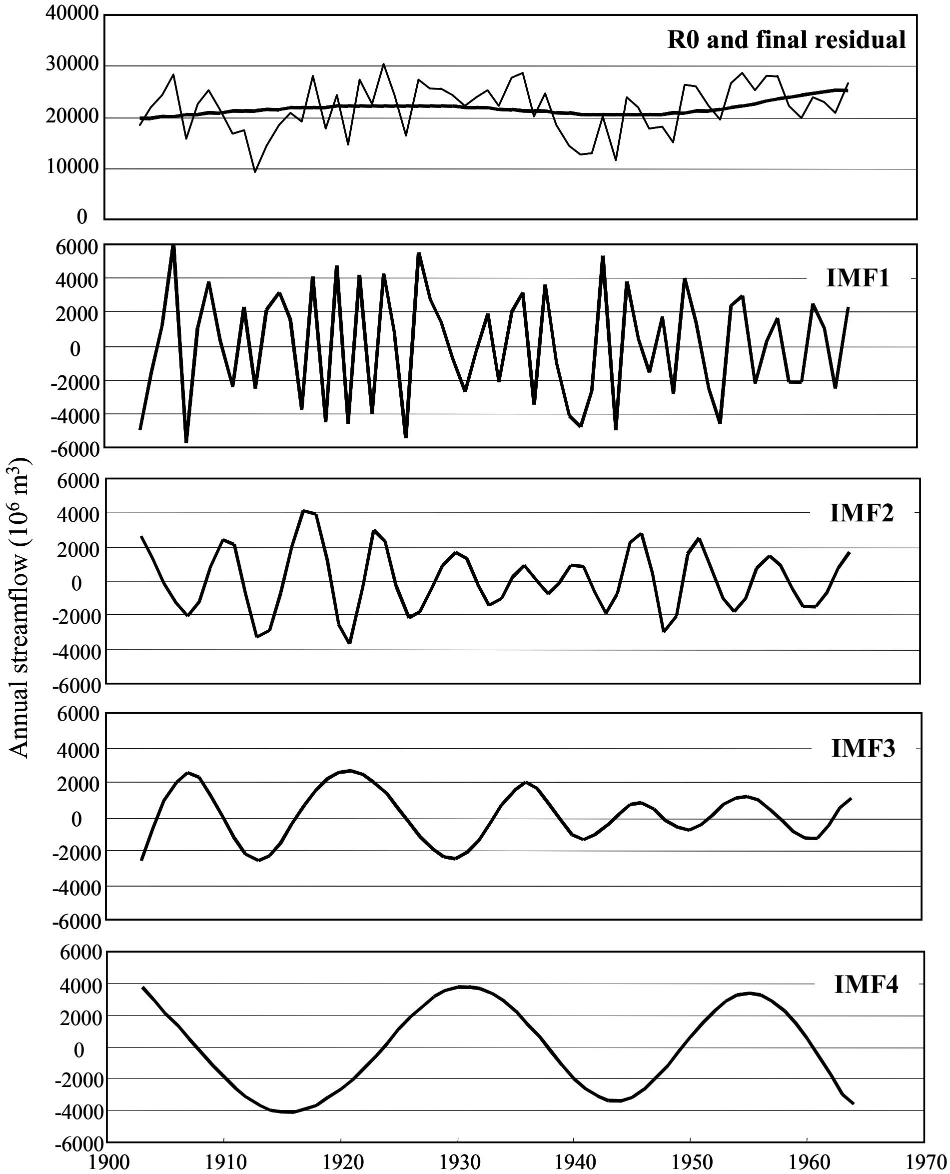

Fig. 4 IMFs and final residual from EMD analysis of annual streamflow time series for Senegal River at Dagana.

Significance of oscillations in time series data

Figure 4 shows results from the EMD analysis of 62 years of annual streamflow time series for the Senegal River at Dagana. The EMD analysis identified oscillations with average periods of 3.1 (IMF1), 6.1 (IMF2), 11.6 (IMF3) and 26.5 years (IMF4).

Results from EMD analysis can be summarized in a ln$T$ *vs* ln$E$ plot (ln period/-oscillation *vs* ln energy/variance, see Wu & Huang, 2004), with a data point for each IMF (Fig. 5). The energy for each IMF is defined as the proportion of variance in the original data accounted for by the IMF. The earlier IMFs have higher energies, and the sum of energies from all the IMFs and the final residual is equal to one if the IMFs are mutually independent (orthogonal). The choice of sifting termination criteria also

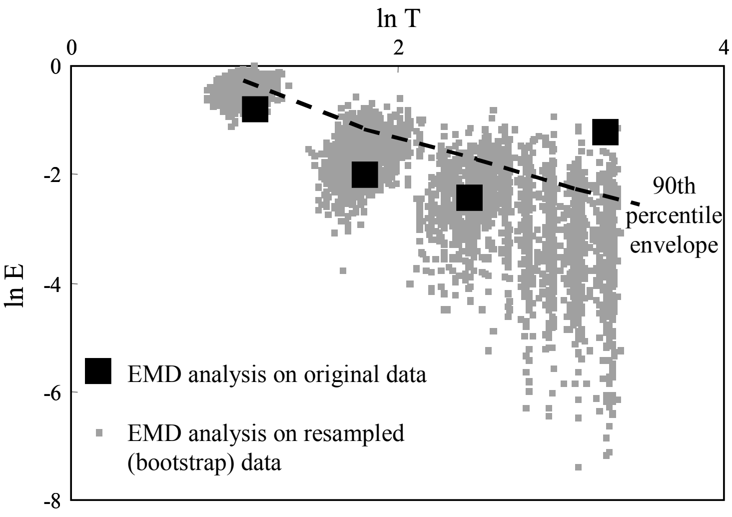

Fig. 5 ln*T* (period) *vs* ln*E* (energy) plot from EMD analysis of annual streamflow time series for Senegal River at Dagana.

influences the level of IMF independence and the criteria described above and used here balances the trade off between IMF independence and over sifting the IMFs.

The statistical significance of the variance accounted for by each IMF (and therefore the statistical significance of the period of oscillation) can be estimated by comparing the variance accounted for by the IMF in the data and the variance accounted for by the IMF in random (white noise) data. Envelopes for various statistical significance levels are found in Wu & Huang (2004) for uniformly distributed white noise data. However, because the annual streamflow data are not uniformly distributed, and the IMFs from the EMD analysis are not consistently orthogonal, a bootstrapping method (see Efron & Tibshirani, 1993) is used here to estimate the statistical significance of the oscillations identified in the EMD analysis. In the bootstrapping method, 10 000 replicates of the same length as the original time series are generated, by resampling from the original data with replacement. The EMD analysis is then carried out on the 10 000 replicate time series, giving [10 000 × number of IMFs] points in the ln*T* *vs* ln*E* plot. The IMFs from the original time series are then compared against the IMFs from the bootstrap replicates (resampled original data with replacement) to estimate the statistical significance of the observed IMFs/oscillations in the original data (Fig. 5). For example, the bootstrap results indicate that the 26.5 year oscillation in the Senegal River annual streamflow data is statistically significant at $\alpha = 0.1$ (90% level) (Fig. 5).

EMD ANALYSIS OF GLOBAL STREAMFLOW DATA

Data

The annual streamflow data used in this paper are drawn from the global database of monthly streamflow data for over 1200 catchments described in Peel *et al.* (2001). The streamflow data are believed to be unregulated over the period of record in the

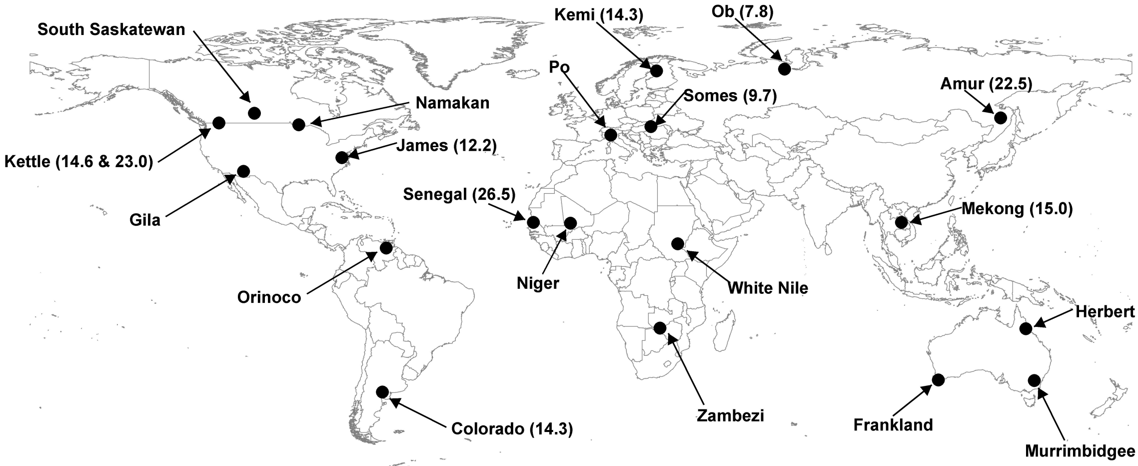

Fig. 6 Locations of streamflow gauging stations used in this study and average periods of statistically significant (at $\alpha = 0.1$) oscillations from EMD analysis.

Table 1 Streamflow gauging stations and summary of annual streamflow characteristics.

Flow gauging station	Catchment area (km^2)	Period of data (years)	*MAR** (mm)	*Cv**	r_1*
South Saskatchewan River at Saskatoon	141 000	1912–1967 (56)	62	0.38	0.43
Kettle River at Near Laurier	9 840	1930–1983 (54)	266	0.24	–0.02
Namakan River at Outlet of Lac La Croix	13 400	1923–1987 (65)	256	0.32	0.37
Gila River at Calva	29 700	1930–1983 (54)	9	1.10	0.03
James River at Cartersville	16 200	1925–1983 (59)	374	0.31	0.10
Orinoco River at Puente Angostura	836 000	1925–1988 (64)	1180	0.11	0.20
Colorado River at Pichi Mahuida	22 300	1919–1979 (61)	184	0.35	0.13
Senegal River at Dagana	268 000	1903–1964 (62)	81	0.23	0.30
Niger River at Dire	340 000	1924–1989 (66)	94	0.28	0.75
White Nile River at Malakal	1 080 000	1912–1981 (70)	28	0.19	0.75
Zambezi River at Victoria Falls	517 000	1924–1977 (54)	78	0.33	0.48
Po River at Piacenza	42 000	1924–1978 (55)	738	0.30	0.08
Kemi River at Taivalkoski	50 800	1911–1971 (61)	332	0.21	0.19
Somes River at Satu Mare	15 200	1925–1987 (63)	262	0.35	0.13
Ob River at Salekhard	2 950 000	1930–1983 (54)	134	0.16	0.40
Amur River at Komsomolsk	1 730 000	1933–1983 (51)	178	0.20	0.43
Mekong River at Mukdahan	391 000	1924–1986 (63)	651	0.14	0.38
Frankland River at Mt Franklan	5 800	1941–1902 (62)	28	0.63	–0.13
Herbert River at Gleneagle	5 300	1922–2001 (80)	199	0.88	0.05
Murrumbidgee River at Gundagai	21 100	1887–1954 (68)	154	0.56	0.31

**MAR*: mean annual runoff; *Cv*: coefficient of variation (standard deviation divided by the mean) of annual runoff; r_1: lag-one serial correlation of annual runoff.

database. Annual streamflow time series from 20 catchments are used here to demonstrate the EMD method. The 20 catchments are larger than 5000 km^2, have relatively long and unbroken streamflow data, and are chosen to provide coverage across the world. The catchment locations are shown in Fig. 6 and a summary of the annual streamflow characteristics is given in Table 1.

Table 2 Oscillations identified in EMD analysis of annual streamflow time series.

Flow gauging station	Periods of oscillations				
South Saskatchewan River at Saskatoon	3.1	6.9	13.0	19.6	
Kettle River at Near Laurier	2.8	6.4	**14.6**	**23.0**	
Namakan River at Outlet of Lac La Croix	3.0	7.0	13.6	20.0	28.5
Gila River at Calva	3.2	6.4	12.5	19.6	
James River at Cartersville	2.9	5.6	**12.2**	18.3	
Orinoco River at Puente Angostura	3.2	7.2	15.3		
Colorado River at Pichi Mahuida	3.2	7.3	**14.3**	17.7	
Senegal River at Dagana	3.1	6.1	11.6	**26.5**	
Niger River at Dire	2.7	5.6	17.1	23.6	
White Nile River at Malakal	3.3	7.6	14.5	21.3	
Zambezi River at Victoria Falls	3.5	6.5	11.1		
Po River at Piacenza	2.7	5.6	10.2		
Kemi River at Taivalkoski	3.0	7.7	**14.3**		
Somes River at Satu Mare	2.8	5.6	**9.7**	19.3	
Ob River at Salekhard	2.8	**7.8**	14.6		
Amur River at Komsomolsk	2.9	7.2	11.8	**22.5**	
Mekong River at Mukdahan	3.2	6.8	**15.0**	19.7	
Frankland River at Mt Franklan	3.2	7.3	14.0	26.5	
Herbert River at Gleneagle	2.6	5.5	11.7	16.7	36.0
Murrumbidgee River at Gundagai	3.3	6.5	13.8	21.0	

Average oscillation periods for the first, second, third, fourth (most catchments) and 5th (two catchments) IMFs.
Statistically significant oscillations (at $\alpha = 0.1$) are highlighted in bold.

RESULTS

The EMD analysis is applied to the annual streamflow time series from the 20 catchments (on the original time series and the 10 000 bootstrap replicates) to identify oscillations in the data and test their statistical significance. The average oscillation period of the IMFs are shown in Table 2, with the statistically significant (at $\alpha = 0.1$) oscillations highlighted in bold. The statistically significant oscillations are also shown in Fig. 6.

DISCUSSION

The average oscillations of the first, second, third and fourth IMFs are about 3, 6–7, 11–15 and 20–25 years respectively. The first and second IMFs are likely to be related to El Niño/Southern Oscillation (ENSO). The first IMF is not statistically significant because it contains most of the high frequency noise in the annual time series (see Wu & Huang, 2004). It is possible that EMD analysis of monthly streamflow time series will identify statistically significant oscillations related to ENSO because the first IMF is likely to have an intra-annual period and oscillations related to ENSO would be contained in subsequent IMFs.

The third IMF (11- to 15-year oscillation) is statistically significant (at $\alpha = 0.1$) in six out of the 20 catchments and the fourth IMF (20–25 year oscillation) is statistically significant in three out of 15 catchments (five catchments do not have a fourth IMF).

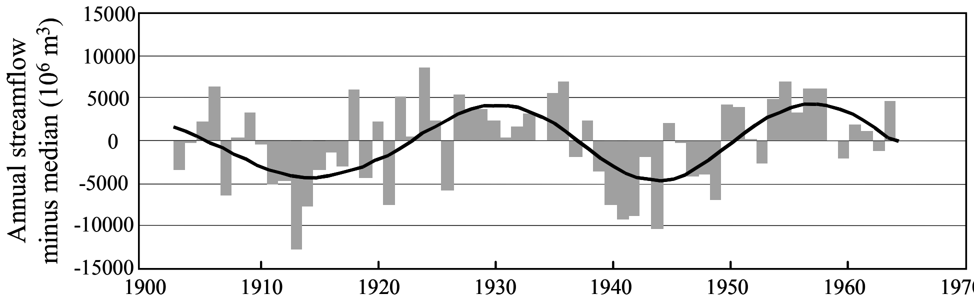

Fig. 7 Annual streamflow time series for Senegal River at Dagana (with curve showing the sum of the fourth IMF and final residual from EMD analysis).

The results are inconclusive but the number of statistically significant oscillations observed here is more than would be expected from a random sample. Oscillations with the above periods have also been reported in various studies of local rainfall and streamflow time series, and they have been attributed to the natural oscillations of the global atmosphere–ocean system and/or solar variability (see references in Burroughs (2003) and Peel *et al.* (2004)).

There are no obvious relationships between the statistically significant oscillations and catchment location, climatic zone, coefficient of variation of annual streamflow, or lag-one serial correlation of annual streamflow. However, the number of catchments used here is small, and further analysis with the entire global streamflow database (currently being conducted) may show some relationships. Nevertheless, any study with streamflow data will be limited by the homogeneity of the catchment conditions (e.g. land use change may have occurred over the period of record).

Although not all the streamflow time series show statistically significant oscillations, many data sets indicate that some interdecadal periods are considerably wetter/drier than others (Fig. 7). This interdecadal variability has implications for the management of water resources, in particular the security of supply in different decadal periods.

CONCLUSIONS

This paper demonstrates that the EMD analysis can be used to identify oscillations in annual streamflow time series. The EMD analysis can also be applied to bootstrap samples from the original time series to test the statistical significance of the identified oscillations. The EMD method is relatively easy to use as the intrinsic oscillations are automatically and adaptively selected from the time series.

The EMD analysis is applied to annual streamflow time series from 20 catchments across the world. The results indicate that statistically significant oscillations of 11–14 and 20–25 years are observed in some, but not all, of the time series.

Acknowledgements The authors wish to acknowledge the Australian Research Council (grant no. DP0449685) and Melbourne Water for providing some of the funding for this and related studies, and the Global Runoff Data Centre (Koblenz, Germany) for providing some of the data.

REFERENCES

Burroughs, W. J. (2003) *Weather Cycles—Real or Imaginary*. Cambridge University Press, Cambridge, UK.

Efron, B. & Tibshirani, R. J. (1993) *Introduction to the Bootstrap*. Chapman & Hall, New York, USA.

Huang, N. E., Shen, Z., Long, S. R, Wu, M. L., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C. & Lui, 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.

Huang, N. E., Shen, Z. & Long, S. R. (1999) A new view of nonlinear water waves: the Hilbert spectrum. *Ann. Rev. Fluid Mech.* **31**, 417–457.

Huang N. E., Wu, M. C., Long, S. R., Shen S. S. P., Qu, W., Gloersen, P. & Fan, K. L. (2003) A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proc. Roy. Soc. London A* **459**, 2317–2345.

Peel, M. C., McMahon, T. A., Finlayson, B. L. & Watson, F. G. R. (2001) Identification and explanation of continental differences in the variability of annual runoff. *J. Hydrol.* **250**, 224–240.

Peel, M. C., Pegram, G. G. S. & McMahon, T. A. (2004) Global analysis of runs of annual precipitation and runoff equal to or below the median: run length. *Int. J. Climatol.* **24**, 807–822.

Quek, S. T., Tua, P. S. & Wang, Q. (2003) Comparison of Hilbert-Huang, wavelet and Fourier transforms for selected applications. In: Mini Symposium on Hilbert-Huang Transform in Engineering Applications, November 2003, Newark, Delaware, USA.

Rilling, G., Flandrin, P. & Gonclaves, P. (2003) On empirical model decomposition and its algorithms. In: *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, NSIP-03.

Wu, Z. & 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.