

# Empirical Mode Decomposition: Improvement and Application

Peel, M.C.<sup>1</sup>, G.G.S. Pegram<sup>2</sup> and T.A. McMahon<sup>1</sup>

<sup>1</sup> Department of Civil and Environmental Engineering, The University of Melbourne, Victoria

<sup>2</sup> Civil Engineering, University of KwaZulu-Natal, South Africa

Email: mpeel@unimelb.edu.au

**Keywords:** Empirical Mode Decomposition, Hydroclimatology, Rational spline, Variability

## EXTENDED ABSTRACT

Empirical mode decomposition (EMD), a relatively new form of time series decomposition, has the feature of not assuming a time series is linear or stationary (like Fourier analysis). In hydroclimatology, where most variables exhibit non-linear and non-stationary behaviour, this feature is particularly useful, allowing more meaningful quantification of the proportion of variance in a time series due to fluctuations at different time scales than previous spectral analysis techniques. However, in its original form the EMD algorithm relies on cubic spline interpolation, which is suspected of inflating the variance of the resultant Intrinsic Mode Functions (IMFs) and residual (trend). In this paper a potential improvement to the EMD algorithm is briefly outlined and its effect on an example is assessed.

Rational splines are proposed as a potential improvement to the EMD algorithm. The rational splines are used in place of cubic splines for interpolating the maxima and minima envelopes within the EMD sifting process. Rational splines possess a parameter that enables the user to select the level of tension in the spline interpolation, ranging from cubic through to virtually straight line. By varying this tension parameter between 0 (cubic) and 50 (virtually linear) a set of EMD decompositions were generated for the annual precipitation record of Melbourne, Australia, as an example for initial testing of the idea.

The performance of the different decomposition sets was assessed based on the number of IMFs, the number of sifts required to extract each IMF, the total variance of the IMFs and the residual relative to the observed variance and the mutual orthogonality of the IMFs and residual. For the Melbourne annual precipitation time series, the cubic spline based EMD provided the least number of IMFs and the least number of sifts, indicating an efficient decomposition. However, the cubic based EMD had the highest variance sum for the IMFs

and residual relative to the observed variance (+24%) and the IMFs and residual were not orthogonal. As the rational spline tension parameter was increased the IMF and residual variance sum decreased and approached parity with the observed variance at a tension value equal to 10. For the Melbourne annual precipitation time series, a rational spline based EMD with tension parameter set to 5 appears to give the optimal decomposition in terms of the number of IMFs, number of sifts total variance and the orthogonality of the IMFs and residual.

This preliminary analysis shows that as the tension parameter increases the amount of variance associated with the IMFs and residual relative to the observed variance reduces as conjectured. This reduction in variance is also seen in the covariance terms, which improves the orthogonality of the EMD decomposition. Increasing the tension parameter has the desired effect of reducing the variance and covariance of the resultant IMFs and residual by progressively eliminating the well-known tendency for the cubic spline to over- and under-shoot.

This example also highlights the benefit of the rational spline variable tension parameter. This parameter can be modified to give decompositions that satisfy requirements of orthogonality, number of IMFs, number of sifts, etc, rather than the cubic spline based EMD which only provided a single decomposition. Before the introduction of rational splines to EMD, in cases where the cubic decomposition is poor, there was no other alternative, whereas now a range of reasonable decompositions is possible.

This initial application demonstrates that the variance inflation is greatly reduced, but at the expense of a progressive increase in IMFs and sifting. The rational spline based EMD demonstrates useful features that will be further tested on a larger dataset.

## 1. INTRODUCTION

Huang et al (1998) introduced Empirical Mode Decomposition (EMD) as the first part of a two part process for spectral analysis of non-linear and non-stationary time series. In stage one the time series is adaptively decomposed into a set of intrinsic mode functions (IMFs) and a residual, using the EMD algorithm, followed, in stage two, by a Hilbert transform of the IMFs. This two stage process has become known as the Hilbert Huang Transform (HHT) and is being increasingly used across a range of fields including hydrology and climatology (Gloersen & Huang, 2003; Sinclair and Pegram, 2005; Peel & McMahon, 2006).

The motivation of Huang et al (1998) to introduce the HHT was dissatisfaction with traditional forms of spectral analysis, like Fourier and Wavelets, when applied to non-linear and non-stationary data. For example, the assumption behind Fourier analysis is that a time series can be decomposed into a set of linear, stationary and harmonic components. However, as the degree of nonlinearity and non-stationarity in the time series increases, the number of harmonics required to describe the time series increases. Many of the resultant components are physically meaningless harmonics. Attempting to remove the non-stationarity, by de-trending the data prior to Fourier analysis, requires an assumption about the nature of the trend (linear, step change, other), which is usually unknown.

The Fourier spectrum of a time series reveals the relationship between variance and frequency over the set taken as a whole, whereas the Wavelet and Hilbert spectrums reveal the relationship between variance and frequency over time. Huang et al (1998) note that the Hilbert spectrum provides much sharper detail than the Wavelet spectrum (based on the widely used Morlet wavelet) due to harmonics in the Wavelet spectrum. By contrast the local nature of the HHT does not smear energy over a range of frequencies and time, as in Wavelets. One of the main differences between EMD and Wavelets is the choice of a time invariant filter (wavelet function) in Wavelets, which is incapable of adapting to local variations, unlike EMD (Flandrin & Gonçalvès, 2004). However, Olhede & Walden (2004) note that Wavelets can provide comparable results to HHT, when appropriately chosen wavelet functions are used in combination with the Hilbert spectrum for non-stationary time series.

In the context of the physical sciences, the HHT method is ideally suited since it does not require assumptions that time series are linear or stationary

(usually it is not known if they are not). Furthermore, HHT does not require a detailed knowledge of wavelet functions, in order to select an appropriate one for the task at hand, rather the EMD algorithm is adaptive to the local conditions that present within the time series.

In this paper an improvement to the EMD algorithm, the first part of the HHT process, is outlined and applied to demonstrate the impact of this change on the previous version of EMD.

## 2. THE PROBLEM

In the original EMD paper of Huang et al (1998) five areas of future research for improving the HHT methodology were outlined. The first area was the spline fitting used in the EMD sifting process to extract the IMFs. This sifting process consists of 4 steps:

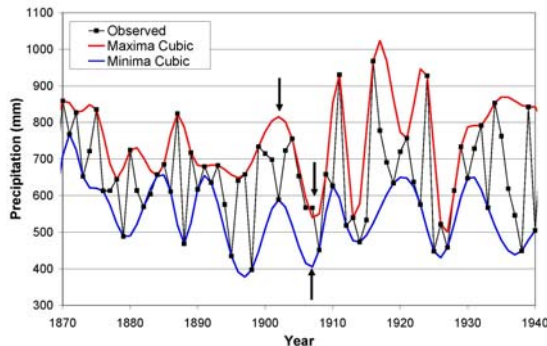
1. Identify the local extrema, both maxima and minima in the input time series;
2. Fit splines to the sequences of maxima and minima;
3. Take the mean of the two splines;
4. Subtract the mean of the splines from the input time series of step 1.

The difference from step 4 represents a residual and is an estimate of an IMF. If the input time series to step 1 were the original observed data, then the product of step 4 would be the first estimate of the first IMF. If the output from step 4 did not satisfy the requirements for an IMF, then it would form the input to step 1 and the sifting process would be repeated until a residual was found that satisfied the requirements for an IMF.

In Huang et al (1998), and in subsequent EMD/HHT papers, the spline fitted to the extrema in the EMD sifting process has been a cubic spline. However, cubic splines are known to exhibit over- and under-shoot problems (see Figure 1). Since the mean of the two splines is taken as step 3 of the sifting process, over- and under-shooting in the spline fitting procedure will transfer into the mean and therefore the estimate of the IMF. The potential for IMFs to be corrupted by the cubic spline over-shooting problem may be amplified by the iterative nature of the sifting process.

The impact of using alternative interpolation techniques to cubic splines in EMD has not been adequately assessed. Huang et al (1998, page 988) mention that trials with taut splines, where a spline

tension parameter can be adjusted, showed only marginal improvement. Rilling et al (2003) suggested that cubic splines were preferred to linear or polynomial interpolation based on some unreported experiments.



**Figure 1.** Section of the annual precipitation record for Melbourne Regional Office with maxima and minima cubic splines for the first sift of the first IMF. Over- and under-shoots are indicated by arrows.

### 3. ASSESSING ALTERNATIVES

Before applying and testing an alternative spline methodology, it is important to know how to assess the impact of the alternative, relative to the cubic spline original, on the IMFs. Although there is no objective way to determine which set of IMFs are the best (Huang et al., 2003) there are three measures that can be used to inform which set of IMFs might be better.

#### 3.1. Number of Sifts

As mentioned in the previous section sifting is the method used to extract the IMFs. As described in Peel et al (2005) the original definition of an IMF from Huang et al (1998) was that (1) the number of extrema (sum of maxima and minima) and the number of zero crossings must be equal or differ by one, and (2) the mean of the cubic splines must be equal to zero at all points. Since sifting is a recursive process, a sifting stopping rule is required. This stopping rule needs to halt the process when the residual satisfies the definition of an IMF, while also avoiding the following two concerns.

Firstly, the sifting process reveals hidden riding waves in the input time series, which when identified change the number of extrema during sifting. If the process is halted prior to all of the riding waves being identified, then the residual may only temporarily satisfy the definition of an IMF, and subsequent sifting would reveal a significantly different IMF. This first concern can

be addressed by sifting the residual until it is stable for several consecutive sifts. However, there is a trade off between the number of sifts and the second concern, which is that over-sifting tends to produce smooth amplitude IMFs where any physically meaningful amplitude variation has been sifted away.

The sifting stopping rule used by Huang et al (1998), in their trials with taut splines and cubic splines, was a convergence criterion based on minimising the difference between residuals in successive sifts to below a predetermined level. However, this criterion did not explicitly take into account the two IMF conditions, so the predetermined level could be obtained without the two IMF conditions being satisfied (Huang et al 2003). This lead Huang et al (2003) to propose an alternate stopping rule, used here, where sifting is conducted until IMF condition (1) is satisfied (number of extrema = number of zero crossings  $\pm 1$ ). At this sift, and each subsequent sift, the number of extrema and zero crossings are recorded and compared to those of the previous sift. When the number of extrema and zero crossings remain constant for five successive sifts, the sifting is stopped and the residual is designated as an IMF. If after any sift IMF condition (1) was not satisfied then the successive sift count would restart at the next sift where IMF condition (1) was satisfied. Huang et al (2003) found that IMFs produced using this sifting stopping criterion satisfied condition (i), were consistently orthogonal and were not over-sifted.

A potential source of difference between the spline methodologies is the number of sifts required to define the IMFs. The method that requires the least sifts would be considered the better method, since this would reduce any over-sifting concerns.

#### 3.2. Number of IMFs

The premise of EMD is to identify any intrinsic oscillatory modes within a time series and extract them as IMFs.

As noted previously, there is no objective way to determine which set of IMFs is best. However, in general the EMD version that can decompose a time series into the smallest number of IMFs is likely to provide more physically realistic IMFs than a version that requires more IMFs.

#### 3.3. Orthogonal IMFs

When assessing the quality of a set of IMFs, Huang et al (2003) used an orthogonal index to reject IMF sets that were non-orthogonal.

Although not strictly required by Huang et al (1998) in the non-linear case, IMFs and the residual component are generally expected to be orthogonal to each other.

Ideally if the IMFs and residual are orthogonal the observed variance will equal the sum of the variances of the IMFs and residual, since the covariance terms will be zero, following

$$\text{Var}(X) = \Sigma \text{Var}(\text{IMFs} + R) + 2\Sigma \text{Cov}(\text{IMFs}+R), \quad (1)$$

where  $\text{Var}(X)$  is the variance of the observed data,  $\text{Var}(\text{IMFs} + R)$  is the variance of the IMFs and residual and  $\text{Cov}()$  is covariance of the IMFs and residual.

Chiew et al (2005) and Peel et al. (2005) noted in applying EMD that the sum of the IMF and residual variances were on average 15% to 20% larger than the variance of the original time series, which indicates non-trivial covariance terms between the IMFs and residual. Peel et al (2005) suggested the sum of the variances being greater than the variance of the sum was due to a combination of rounding errors, the non-linearity of the original time series and the introduction of variance by the cubic spline, particularly at the end points. Another potential cause is the cubic spline over- and under-shooting throughout the entire record.

Therefore the spline methodology that provides the smallest IMF and residual covariance terms (giving the more orthogonal IMFs and residual) would be considered the better method.

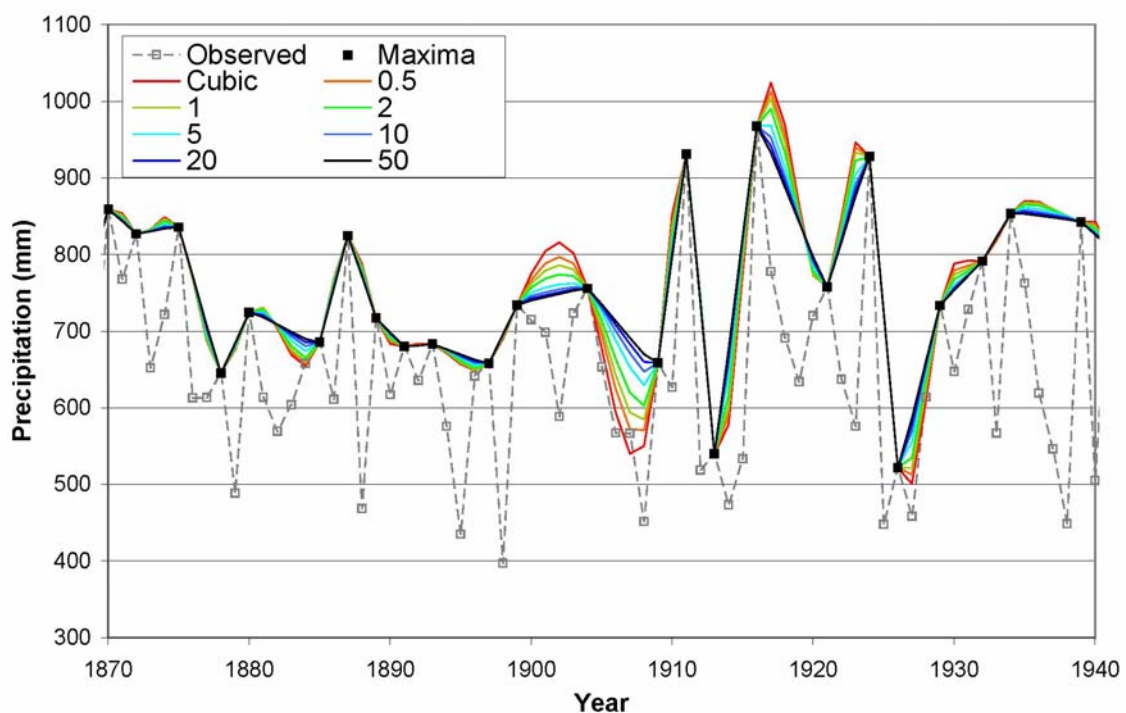
## 4. RATIONAL SPLINES

The alternative spline interpolation methodology introduced here is rational splines, which was proposed by the second author. A detailed explanation of the theory behind rational splines and their application to EMD is currently in preparation (Pegram et al., 2007). A brief summary of rational splines and an example of their use in EMD are provided below.

### 4.1. Rational Splines in brief

The impetus for using an alternative to cubic spline interpolation in the EMD algorithm was the over- and under-shooting often experienced with cubic splines. The rational spline has a variable tension parameter,  $P$ , which allows the tautness of the spline to be controlled. When  $P$  is set to 0, the rational spline defaults to a cubic spline. As  $P$  increases, the spline becomes less curved and more linear, until at values near 50 it becomes virtually linear between the extrema being interpolated. By varying  $P$ , the degree of over- and under-shooting can be controlled during the EMD process.

Relative to exponential splines (another tension based interpolation scheme), Späth (1995) notes that rational splines are superior in terms of effectiveness and efficiency of computation. The tension spline concept is different to the taut spline approach, in that a single tension parameter is used to stretch the spline, whereas taut splines involve the introduction of extra nodes or knots to remove unwanted oscillations in a cubic spline.



**Figure 2.** Section of the annual precipitation record for Melbourne Regional Office with maxima rational splines for a range of tension parameter values for the first sift of the first IMF.

## 4.2. Maxima spline example

In Figure 2 the rational spline has been applied to a section of the observed annual precipitation record for Melbourne. In the EMD sifting process, Figure 2 represents the first sift of the first IMF for a range of  $P$ . The observed record is shown along with the local maxima and several rational splines with  $P$  ranging from 0 (cubic), 0.5, 1, 2, 5, 10, 20 & 50 (nearly linear). The maximum envelope in the period 1900 – 1910 exhibits an over-shoot and an under-shoot in the cubic spline, which are progressively reduced as  $P$  is increased so that the rational spline becomes almost linear between the extrema. Figure 2 also shows that where the cubic spline performs well, the rational spline performs similarly across the range of tension parameter settings.

## 5. RESULTS

The rational spline based EMD was applied to the annual precipitation record for Melbourne, Victoria, Australia (1856 – 2006,  $N = 151$ ) in order to provide insight into how the rational spline alternative performed relative to the original cubic spline method. The Melbourne record was decomposed using rational spline based EMD for tension parameter settings of 0 (cubic), 0.5, 1, 2, 5, 10, 20 & 50. The performance was assessed using the previously discussed indicators of the number of sifts required to define the IMFs, the number of IMFs and whether the IMFs were orthogonal.

### 5.1. Number of IMFs and sifts

In Table 1 the number of sifts required to identify each IMF for the different  $P$  settings are presented.

The first feature of note in Table 1 is that of the 8 tension settings the cubic spline ( $P = 0$ ) produces the least number of IMFs (4). As  $P$  increases, the number of IMFs required to decompose the time series increases, with  $P = 50$  requiring 8 IMFs.

The cubic spline also requires the least number of

**Table 1.** Number of sifts required to identify each IMF for each tension parameter setting.

Tension	IMF								Total
	1	2	3	4	5	6	7	8	
0	7	10	7	7					31
0.5	7	10	7	11	8				43
1	8	11	10	7	7				43
2	7	12	8	7	7				41
5	7	11	7	7	7				39
10	7	11	8	7	7	7			47
20	7	24	9	7	8	12			67
50	7	19	12	7	7	18	8	8	86

sifts to extract any given IMF. There is some indication in Table 1 that as  $P$  increases, the number of sifts required for each IMF increases, particularly for higher tension values.

The main influence of the tension parameter appears to be to increase the number of IMFs in the decomposition and to a lesser extent to increase the number of sifts needed to extract each IMF.

### 5.2. Orthogonal IMFs

The variance of the IMFs and residual as a percentage of the observed variance is shown in Table 2 for each  $P$  setting. The cubic spline based EMD has the highest IMF and residual variance total of all the settings, with 24% more variance than the observed record.  $P$  equal to 10 and 20 provide variance totals that are closest to the observed, with a general decrease in total variance as  $P$  increases.

The fact that the IMF and residual variance sum is 24% greater than observed for the cubic spline case indicates that the covariance terms between the IMFs and residual are non-trivial. In equation 1 the sum of the variances + twice the covariances equals the observed variance, indicating that for the cubic spline case the covariance terms sum to 12% of the observed variance. In terms of the orthogonality index of Huang et al (1998), this translates into a value of 0.12, which is greater than the acceptable limit of 0.1 indicated by Huang et al (2003). Therefore, the cubic spline IMFs and residual can not be considered to be orthogonal. For the other tension settings the orthogonality index values are 0.07 (tension = 1) or less. For  $P$  greater than 2, the orthogonality index values are less than 0.05, which is more acceptable.

In an analysis not shown increasing  $P$  to significantly higher values than 50 (100 & 300) produced similar results to those for  $P = 50$ . This result is expected, since the near linear interpolation achieved with  $P$  equal to 50 leaves little scope for higher tensions to produce different

**Table 2.** IMF and residual variance as a % the observed for each tension parameter setting.

Tension	IMF								Res.	Total
	1	2	3	4	5	6	7	8		
0	61	27	20	8.1					7.7	124
0.5	60	25	16	6.2	2.6				3.0	113
1	59	25	14	8.5	2.9				4.6	114
2	58	23	14	7.9	3.0				4.4	110
5	57	22	14	7.6	3.2				3.1	106
10	56	19	10	5.4	4.6	2.3			3.2	101
20	55	13	11	16	1.7	1.5			3.0	101
50	55	13	9.8	16	1.2	0.1	0.7	0.4	0.7	97

results.

Table 2 clearly shows that as the tension parameter increases the amount of variance associated with the IMFs and residual relative to the observed variance reduces. This reduction in variance is also seen in the covariance terms. Increasing  $P$  has the effect of progressively eliminating the cubic spline over- and under-shooting and thus reducing the variance and covariance of the resultant IMFs and residual.

## 6. DISCUSSION

Based on the performance indicators of the number of IMFs and the number of sifts for each IMF, the cubic spline based EMD outperforms the rational spline based EMD for the Melbourne annual precipitation time series. This is because in this case the residual is not decomposed into two components - the enhanced variance of the residual displays this effect. However, in terms of the total variance and orthogonality of the IMFs and residual, the cubic spline based EMD clearly underperforms the rational splined based EMD. For Melbourne annual precipitation a rational spline based EMD with tension parameter set to 5 appears to give the optimal (best qualitative compromise) decomposition in terms of the number of IMFs, number of sifts, total variance and the orthogonality of the IMFs and residual.

This example application of rational spline based EMD highlights the benefit of the variable tension parameter. This parameter can be modified to give decompositions that satisfy requirements of orthogonality, number of IMFs, number of sifts, etc, unlike the cubic spline based EMD which provides a single decomposition. In cases where the cubic decomposition was poor, there was no alternative, whereas now a set of rational decompositions is possible. Furthermore, with the cubic spline as a special case of rational splines, the cubic decomposition is not lost if rational splines are adopted.

In cases where the orthogonality of the IMFs and residual is critical, for example, when using EMD as a basis for stochastically generating hydroclimatic data (McMahon et al., 2007), the ability to vary the tension parameter to produce efficient and orthogonal decompositions will prove very useful.

This application demonstrates the positive aspects of rational spline based EMD; further testing of this method on a wider set of data is in process.

## 7. CONCLUSION

Artificial variance inflation in the IMFs and residual experienced with cubic spline based EMD, due to over- and under-shooting of the cubic splines, has been discussed. An alternative method of rational spline based EMD, which reduces the over- and under-shooting through a tension parameter, has been briefly outlined and demonstrated. Initial results indicate the variance inflation is greatly reduced, but at the expense of more IMFs and more sifting. The rational spline based EMD demonstrates useful features that will be further tested on a larger dataset.

## 8. ACKNOWLEDGEMENTS

This research was financially supported by Australian Research Council (ARC) grant DP0773016. Discussions with Prof John Fenton about exponential and taut splines are gratefully acknowledged.

## 9. REFERENCES

- Chiew, F.H.S., M.C. Peel, G.E. Amirthanathan & G.G.S. Pegram (2005), Identification of oscillations in historical global streamflow data using empirical mode decomposition. Seventh IAHS Scientific Assembly at Foz do Iguaçu, Brazil. IAHS Publ. 296, pp: 53-62.
- Flandrin, P. & P. Gonçalvès (2004), Empirical mode decomposition as data-driven wavelet-like expansions, *International Journal of Wavelets, Multiresolution and Information Processing*, 2(4), 477-496.
- Gloersen, P. & N. Huang (2003), Comparison of interannual intrinsic modes in hemispheric sea ice covers and other geophysical parameters, *IEEE Transactions on Geoscience and Remote Sensing*, 41(5), 1062-1074.
- Huang, N.E., Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung & H.H. Liu (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.
- Huang, N.E., M.C. Wu, S.R. Long, S.S.P. Shen, W. Qu, P. Gloersen & K.L. Fan (2003), A confidence limit for the empirical mode decomposition and Hilbert spectral

analysis, *Proceedings of the Royal Society London A* 459, 2317-2345.

McMahon, T.A., A.S. Kiem, M.C. Peel, P.W. Jordan, G.G.S. Pegram & R.J. Nathan (2007), A new approach to stochastically generating six-monthly rainfall sequences based on Empirical Mode Decomposition, in preparation.

Olhede, S. & A.T. Walden (2004), The Hilbert spectrum via wavelet projections, *Proceedings of the Royal Society London A*, 460, 955-975.

Peel M.C., G.E. Amirthanathan, G.G.S. Pegram, T.A. McMahon & F.H.S. Chiew (2005), Issues with the application of Empirical Mode Decomposition Analysis. In Zerger A and Argent RM (eds) MODSIM 2005 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand, December 2005, pp. 1681-1687. ISBN: 0-9758400-2-9.

Peel, M.C. & T.A. McMahon (2006), Recent frequency component changes in interannual climate variability, *Geophysical Research Letters*, 33, L16810, doi:10.1029/2006GL025670.

Pegram, G.G.S., M.C. Peel & T.A. McMahon (2007), Rational spline based Empirical Mode Decomposition, in preparation.

Rilling, G., P. Flandrin & P. Gonçalvès (2003), On Empirical Mode Decomposition and its algorithms, IEEE-EURASIP Workshop on Non-linear signal and Image Processing NSIP-03, Grado, Italy.

Sinclair, S. & G.G.S. Pegram (2005), Empirical Mode Decomposition in 2-D space and time: a tool for space-time rainfall analysis and nowcasting, *Hydrology and Earth System Sciences*, 9, 127-137, [www.copernicus.org/EGU/hess/hess/9/127/](http://www.copernicus.org/EGU/hess/hess/9/127/) SRef-ID: 1607-7938/hess/2005-9-127.

Späth, H. (1995), *One Dimensional Spline Interpolation Algorithms*, AK Peters, Massachusetts, ISBN 1-56881-016-4.