Empirical mode decomposition using rational splines: an application to rainfall time series

BY G. G. S. PEGRAM 1,*, M. C. PEEL 2 AND T. A. McMAHON 2

1 Civil Engineering, University of KwaZulu-Natal, Durban 4000, Republic of South Africa
2 Department of Civil and Environmental Engineering, The University of Melbourne, Victoria 3010, Australia

Empirical mode decomposition (EMD), a relatively new form of time-series decomposition, has the feature of not assuming that a time series is linear or stationary, as is implicitly done in Fourier analysis. In natural time series such as records of rainfall, streamflow, temperature, etc., where most variables exhibit nonlinear 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 of the extrema, which often inflate the variance of the resultant components (intrinsic mode functions) and residual. In this paper, a suggested improvement to the EMD algorithm, using rational splines and flexible treatment of the end conditions, is outlined and the consequent effect on three exemplary annual rainfall time series is assessed as a proof of concept. It is recognized that many more examples are needed for the fine tuning of the ideas before it can be widely recommended and this work is currently underway.

Keywords: empirical mode decomposition; end conditions; rational spline; efficiency; economy; appropriateness

1. Introduction

Huang et al. (1998) introduced empirical mode decomposition (EMD) as the first part of a two-stage process for analysis of nonlinear and non-stationary time series. In stage 1, the time series is adaptively decomposed into a set of intrinsic mode functions (IMFs) and a residual, using the EMD algorithm, followed, in stage 2, 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 & Pegram 2005; Peel & McMahon 2006; Fauchereau et al. 2008).

The motivation of Huang et al. (1998) to introduce the HHT was their dissatisfaction with traditional forms of spectral analysis, such as Fourier and wavelets, when applied to nonlinear 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. The application of wavelets to non-stationary time series demands the choice of selective-basis functions, requiring considerable skill in their selection and application. In the application of Fourier series methods, as the degree of nonlinearity and non-stationarity in the time series increases, so does the number of harmonics required to describe the time series. Many of the resultant components are physically meaningless harmonics. Attempting to remove the non-stationarity, by detrending the data prior to Fourier analysis, requires an assumption about the nature of the trend (linear, step change and other), which is usually unknown.

The Fourier spectrum of a time series reveals the relationship between variance and frequency averaged over the set taken as a whole, whereas the wavelet and the Hilbert spectra reveal the relationship between variance and frequency in intervals over time. Huang et al. (1998) noted that the Hilbert spectrum provides much sharper detail than the wavelet spectrum (based on the widely used Morlet wavelet) due to harmonics in the latter. 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 that in the application of wavelets, there is a need to choose a time-invariant filter (wavelet function) from a suite of possible contenders, which is incapable of adapting to local variations, unlike EMD (Flandrin & Gonçalvès 2004). However, Olhede & Walden (2004) noted 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. These discussions are extensions of the comprehensive comparison between EMD, Fourier and wavelet treatments of a wide variety of univariate time series in Huang et al. (1998).

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 not prescriptive but adapts to the local conditions that are present within the time series, so in a sense allows the dataset to ‘speak for itself’.

In this paper a modification to the EMD algorithm, the first part of the HHT process, is described in some detail and applied to three different annual rainfall time series as examples to demonstrate the impact of this change on the previous version of EMD. The paper expands on a conference paper published by Peel et al. (2007), where the ideas appearing here were outlined. There is naturally some commonality of text and ideas, but this paper includes much more detail, in particular the intricacies of the mathematical treatment of the end conditions which makes these ideas possible. Following this introduction, §2 presents the argument for change, §3 details the rational splines, §4 presents the algorithm developed for their endings, §5 presents the results of the experiments and §6 concludes the paper.

2. EMD background and the problem

The basic idea embodied in the EMD analysis, as introduced by Huang et al. (1998), is to allow for an adaptive and unsupervised representation of the intrinsic components of linear and nonlinear signals based purely on the properties
observed in the data without appealing to the concept of stationarity. As Huang et al. (1998) pointed out in their abstract: ‘This decomposition method is adaptive and therefore highly efficient. Since the decomposition is based on the local characteristic time scale of the data, it is applicable to nonlinear and non-stationary processes’.

To develop this idea further, the concept of stationarity can only be justified in the context of a theoretical hypothesis about the statistical properties of a population from which a sample is drawn. Few sequences of observations of natural phenomena are long enough to test the hypothesis of stationarity and hydroclimatological phenomena are often patently non-stationary. The EMD algorithm copes with stationarity (or the lack of it) by ignoring the concept, embracing non-stationarity as a practical reality. For a fuller discussion of the genesis of these ideas, see the introduction of Huang et al. (1998), who also heuristically demonstrate the implicit close orthogonality of the sequences of IMFs defined by the EMD algorithm. In Huang et al. (1998), a series of synthetic examples of time series (acoustic, financial and surface waves, as well as artificial functional forms) was included, demonstrating the methodology and economy of EMD in decomposing time series into a minimal set of functions that are quasi-orthogonal and in appropriate cases capturing underlying physically identifiable components embedded in noise.

In the application of the EMD algorithm, the possibly nonlinear signal that may exhibit varying amplitude and local frequency modulation is linearly decomposed into a finite number of (zero mean) frequency and amplitude-modulated signals, as well as a residual function, which exhibits at most three extrema, is a monotonic trend or is simply a constant. Although EMD is a relatively new data analysis technique, its power and simplicity have encouraged its application in a myriad of fields. It is beyond the scope of this paper to give a complete review of the applications; however, a few interesting examples are cited here to give the reader a feeling for the broad scope of applications. Chiew et al. (2005) examined the one-dimensional EMD of several annual streamflow time series to search for significant trends in the data, using bootstrapping to test the statistical significance of identified trends. The technique has been used extensively in the analysis of ocean wave data (Hwang et al. 1999; Huang et al. 2003) as well as in the analysis of polar ice cover (Gloersen & Huang 2003). EMD has also been applied in the analysis of seismological data by Zhang et al. (2003) and has even been used to diagnose heart rate fluctuations (Balocchi et al. 2004).

(a) Computing the one-dimensional EMD

The EMD algorithm extracts the oscillatory mode that exhibits the highest local frequency from the data (‘detail’ in the wavelet context or the residue of a high-pass filter in Fourier analysis), leaving the remainder as a ‘residual’ (‘approximation’ in wavelet analysis). Successive applications of the algorithm on the sequence of residuals produce a complete decomposition of the data. The final residual is a constant, a monotone trend or a curve with at most three extrema.

The EMD of a one-dimensional dataset $z(k)$ given on a sequence of points $\{k\}$ is obtained using the following procedure.
(i) Set $r_0(k) = z(k)$ and set $i = 1$.

(ii) Identify all of the extrema (maxima and minima) in $r_{i-1}(k)$.

(iii) Compute a maximal envelope, $\max_{i-1}(k)$, by interpolating between the maxima found in step (ii). Similarly compute the minimal envelope, $\min_{i-1}(k)$. Cubic splines (as suggested by Huang et al. (1998)) were previously recommended to be the most appropriate interpolation method for deriving these envelopes in one dimension (Rilling et al. 2003); in this paper, we generalize these bounding functions to rational splines which include cubic splines as special cases.

(iv) Compute the mean value function $m_{i-1}(k)$ of the maximal and minimal envelopes

$$m_{i-1}(k) = \frac{[\max_{i-1}(k) + \min_{i-1}(k)]}{2}.$$ 

(v) The estimate of the IMF is computed from $\text{IMF}_i(k) = r_{i-1}(k) - m_{i-1}(k)$.

(vi) Each IMF is supposed to oscillate about a zero mean and in practice it is necessary to perform a ‘sifting’ process by iterating steps (ii)–(v) (setting $r_{i-1} = \text{IMF}_i$ before each iteration) until this is achieved. Successive IMFs will typically exhibit longer mean periodicities than their predecessors.

(vii) Once it is found that the IMF$_i$ has a mean value that is sufficiently close to zero over all $k$ (defined by a stopping criterion within some predefined tolerance $\varepsilon$, discussed in §2a(ii)) the residual $r_i(k) = r_{i-1}(k) - \text{IMF}_i(k)$ is computed.

(viii) If the residual $r_i(k)$ is a constant, a trend or has no more than three extrema, then stop; else increment $i$ and return to step (ii).

(i) **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 spline fitting used in the EMD sifting process to extract the IMFs. In Huang et al. (1998), and in the subsequent EMD–HHT papers, the spline function that fitted to the extrema in the EMD sifting process was a cubic spline. However, cubic splines are known to exhibit over- and undershoot problems (see figure 1 for an example of a section of the first sift of IMF1 for the annual precipitation record for Melbourne Regional Office, Australia). Since the mean of the two splines is calculated in step (iv) of the above algorithm, over- and undershooting in the spline-fitting procedure will transfer into the mean value function $m_{i-1}(k)$ and therefore the estimate of the IMF in step (v). The potential for IMFs to be corrupted by the cubic spline overshooting 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, p. 988) mentioned 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. Section 3 introduces rational splines as a more flexible
tool to be used in the sequel for determining if an improvement can be made, but first it is necessary to specify what has to be improved, so that criteria for evaluation can be defined.

Before applying and testing an alternative spline methodology, it is important to know how to assess the impact on the IMFs of the alternative, relative to the cubic spline original. Although Huang et al. (2003) suggested that there is no objective way to determine which set of IMFs are the best, there are three measures that can be used to indicate which set of IMFs might be better than others: total number of sifts, total number of IMFs and mutual orthogonality of the components of the decomposition, the IMFs and the residual.

(ii) Number of sifts

As mentioned in the algorithm outlined above, 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 given by the simultaneous satisfaction of two conditions: that (i) the number of extrema (sum of maxima and minima) and the number of zero crossings must be equal or differ by one and (ii) the means, at all sample points, of the two spline functions fitted through the local maxima and the local minima, respectively, must be close to zero. 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.

The first concern is that 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 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 tradeoff between the number of sifts and the second concern, which is that oversifting 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 minimizing the difference between residuals in successive sifts to below a predetermined level. However, this criterion does 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 led Huang et al. (2003) to propose an alternative stopping rule, used here, where sifting is conducted until IMF condition (i) is satisfied (number of extrema = number of zero crossings + 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 (i) was not satisfied, then the successive sift count would restart at the next sift where IMF condition (i) was satisfied. Huang et al. (2003) found that IMFs produced using this sifting stopping criterion satisfied condition (i) were consistently orthogonal and were not oversifted.

A potential source of difference between the spline specifications 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 oversifting concerns as well as the amount of computation.

(iii) Number of IMFs

The key purpose of the EMD method, as claimed by Huang et al. (1998) in their abstract, is that ‘any complicated dataset can be decomposed into a finite and often small number of ‘IMFs’ that admit well-behaved Hilbert transform’.

Notwithstanding the objective set out in the previous paragraph on sifting, there seems to be no satisfactory objective function to determine which set of IMFs is ‘best’. This is a frequent problem encountered in modelling time series, even when they are ‘stationary’; the problem is exacerbated in the context of nonlinearity and non-stationarity. Nevertheless, it seems reasonable to suggest that the EMD version that can decompose a time series into the smallest number of IMFs is more likely to identify any meaningful underlying signal(s) in the observations than a version that requires more IMFs. This desirable quality was exemplified in the examples presented by Huang et al. (1998).

(iv) Orthogonality of IMFs

In describing the orthogonality of IMFs, Huang et al. (1998) suggested that, although the conventional definition of orthogonality is applied globally to pairs of functions, the meaning in EMD applies only locally. They go on to say, in discussing their eqn (6.5): ‘For some special data, the neighbouring components could certainly have sections of data carrying the same frequency at different time durations. But locally, any two components should be orthogonal for all practical purposes’. When assessing the quality of a set of IMFs, Huang et al. (2003) used an orthogonality index to reject IMF sets that were non-orthogonal;

Proc. R. Soc. A (2008)
although not strictly required by Huang et al. (1998) in the nonlinear case, IMFs and the residual component are generally expected to be reasonably orthogonal to each other and our experience is that they are.

Here we introduce the measure of orthogonality that will be used in the following sections. In stationary series, 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. Explicitly

\[
\text{var}(X) = \sum \{\text{var}(\text{IMF}_i) + \text{var}(\text{Res})\} \\
+ 2\left\{\sum \text{cov}(\text{IMF}_i, \text{IMF}_k) + \sum \text{cov}(\text{IMF}_i, \text{Res})\right\}, \quad (2.1)
\]

where \(\text{var}(X)\) is the variance of the observed data; IMF, and Res are the \(i\)th IMF and the residual; and \(\text{cov}(\ )\) is covariance. Note that the concept of (co)variance is somewhat violated in the context of non-stationary series, particularly where a trend is present, so this measure should be treated as a guide, not a prescription; it should work well for IMFs which by definition have zero mean.

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

In summary, the spline methodology that (i) leads to the smallest number of sifts, (ii) needs the smallest number of IMFs in the decomposition, and (iii) provides the smallest IMF and residual covariance terms will herein be considered the best method.

3. Why rational splines?

In EMD, the core idea is the fitting of splines to extrema in the process of extracting the IMFs and the residual in the decomposition of a one-dimensional discrete sample of a (possibly nonlinear and non-stationary) stochastic process. Before getting into the detail required to explain the recommendations made in this paper, it is prudent, nay necessary, to distinguish between the various coordinate conventions used here, in order to obviate an overelaborate notation, so we now dispense with the \(z, m\) and \(r\) notation used in introducing the EMD algorithm. Most of the discourse concerns the bounding extremum functions that are characterized by nodes at local extrema of the underlying series, which in turn may be the original series or a residual series, depending on the stage in the decomposition process. These original/residual series are defined at the sample locations of the original series, so can be thought of as discrete points \((X_j, Y_j)\), \(j=1, 2, ..., N\), where \(N\) is the length of the original series and the intervals between the \(X_j\) are (usually) equal. We will confine the use of upper-case letters to refer to these series.
By contrast, the extremum bounding functions are only defined by the points (nodes) where the underlying Y-values are local extrema, so without resorting to complicated subscript notation, we shall call these nodes \((x_k, y_k), k = 1, 2, \ldots, n\), where \(n\) is the number of points defining the bounding function \((n < N)\) and the intervals \(h_k = x_{k+1} - x_k\) are not necessarily equal in length. Of course, the set of locations \(\{x_k\}\) of the nodes of the upper or lower bounding extremum function must be a subset of the parent observations’ locations \(\{X_j\}\) and the first and last abcissae in each series must match, because they define the length of the original series.

Previous applications of EMD employed cubic splines, a very good introduction to which is given by Press et al. (1992, p. 113). For the end conditions, Chiew et al. (2005) and Peel et al. (2005) devised a procedure that guaranteed a zero gradient at the positions of the first and last observations (even if the spline bounded the points, i.e. did not collocate with them) resulting in ‘flat’ ends. The second derivatives were then evaluated at all extremum locations on the \(x\)-axis at \(x_k, k = 1, 2, \ldots, n\) and were passed to a conventional cubic spline interpolator. It was found in those applications that the cubic spline tended to over- (under) shoot the extrema as shown in figure 1, resulting in an artificial increase in the total variance, computed as the sum of the variances of the IMFs and the residual.

In this generalization of the spline-fitting procedure, we keep the desirable flat end property, but tighten up the splines by using rational splines with a pole parameter \(p\), which controls the tautness of the spline interpolants. The rational spline has the property that it defaults to the cubic spline used earlier (by setting \(p = 0\)) and with \(p\) ‘large’, one is able to tauten the spline to be as close as desired to a polygon (linear) interpolation. This facility helps to explore the effect of the pole parameter \(p\) on important features of the spline which are defined as (i) the appearance of the spline curves (overshoots and wild endings), (ii) the tendency of the extended envelope of the extrema to bound the end data points (which may not be part of the set of extrema) in addition to the quantifiable measures introduced in the previous section, (iii) the behaviour of the recomposed variance of the decomposition compared to that of the original data, (iv) the number of IMFs in the decomposition, and (v) the number of sifts necessary to extract each IMF. These are the criteria used to judge the most appropriate choice of spline defined by \(p\).

\((a)\) Description of the rational spline interpolation and its specialization

As a justification for the selection of rational splines as the interpolants to explore in the application of EMD, we quote Spațh (1995, p. 342), who observed that ‘It is our conviction that [...] rational spline interpolants cannot be surpassed for their effectiveness and efficiency’. Rational splines behave like exponential splines but are computationally more economical. This is attractive in the EMD applications where large sets of hydrological and climatological time series are decomposed.

To introduce the ideas, the following explanation, largely based on the development and notation of Spațh (1995), is offered in some detail, because then the treatment of the end conditions can be developed and thus be used by others, if desired. Figure 2 shows the part of a set of separately identified extremum data
(nodes) that is to be interpolated. Because splines are piece-wise continuous through the nodes, each segment (say the \(k\)th segment between point \(k\) and \(k+1\)) is fitted by a different function (the parameters change locally) of the same type, the choice here being a rational polynomial. The proviso is that where they meet from either side at a point, the values of the fitting function match the value of the data point and, in addition, they have matching first and second derivatives.

In more detail, given a set of discrete data \((X_j, Y_j), j = 1, 2, \ldots, N\), the purpose is to fit spline interpolation segments, to either their local maxima or minima at \(\{x_k\}\), in intervals \((x_k, x_{k+1})\) \(k = 1, 2, \ldots, n-1\), where the individual segment functions are given by rational splines of the following form:

\[
s_k(x) = A_k u + B_k t + C_k u^3/(1 + pt) + D_k t^3/(1 + pu),
\]

(3.1)

where

\[
t = (x - x_k)/h_k = (x - x_k)/(x_k + 1 - x_k)
\]

\[
u = 1 - t
\]

and \(p\) is the pole parameter controlling spline tension.

The pole parameter \((p)\) controls the spline tension in the following way. The last two terms of equation (3.1) have denominators \((1 + pt)\) and \((1 + pu)\). If these go to zero anywhere in the interval as \(t\) and \(u\) go from 0 to 1, then \(s_k(x)\) is undefined (goes to infinity) at those points, called poles. This will not happen if \(p > -1\). When \(p\) is set to \(-1\) the denominators become \(u\) and \(t\) respectively, reducing the power of the corresponding variable to 2, so that equation (3.1) reduces to a quadratic spline interpolant, which usually fluctuates more wildly than even the cubic, so will not be used in this application. The larger the value of \(p\), the closer will the poles be relative to the interval (and symmetrically sited either side of the interval) and the last two terms in \(s_k(x)\) will become correspondingly smaller, reducing to nothing as \(p \to \infty\), so that in the limit, the spline becomes a straight line (polygonal) interpolator. The pole parameter \(p\) is thus a ‘tautness’ parameter. When multiplied by the cubic polynomial in the numerator \(D_k t^3\) of equation (3.1), the combined term becomes a flattened version of the cubic; the same applies to the \(C_k u^3\) term.

The second derivatives of \(s_k(x)\) with respect to \(x\) evaluated from equation (3.1) at the ends of the interval \(k\) are

\[
s_k(x)'' = q \cdot C_k/h_k^2 \quad \text{and} \quad s_{k+1}(x)'' = q \cdot D_k/h_k^2,
\]

where we set \(q = 2(p^2 + 3p + 3)\).

Figure 2. Sketch of spline function segments meeting at data points with matching values and first and second derivatives.
These values of the second derivatives of the spline functions \(s_k''\) and \(s_{k+1}''\) meeting at the point \(x\) are assumed to be equal to the unknown second derivatives of the data: \(y_k''\) and \(y_{k+1}''\), respectively.

These, together with the other interpolation conditions of matching values and first derivatives

\[
s_{k-1}(x) = s_k(x),
\]

and

\[
s_{k-1}'(x) = s_k'(x) \quad \text{for} \quad k = 2, 3, \ldots, n - 1,
\]
yield the \(k\)th spline segment coefficients after some algebra (see Späth 1995),

\[
C_k = y_k''h_k^2/q \quad D_k = y_{k+1}''h_k^2/q \quad A_k = y_k - C_k \quad B_k = y_{k+1} - D_k,
\]

and, in addition, the system of equations relating the \(y''\) estimates for \(k=2, 3, \ldots, n-1:\)

\[
h_{k-1}y_{k-1}'' + (2 + p)(h_{k-1} + h_k)y_k'' + h_ky_{k+1}'' = (d_k - d_{k-1})q,
\]

where

\[
d_k = (y_{k+1} - y_k)h_k.
\]

When \(p=0\) (so \(q=6\)), this set reduces to a system of equations for elaborating cubic splines. As \(p \to -1\), the system produces a set of quadratic splines. As \(p \to \infty\), it is seen that in equation (3.2) \(C_k\) and \(D_k\) \(\to 0\), so that equation (3.1) reduces to a set of linear interpolants between the points.

In conventional spline applications, the end values of the second derivatives \(y_1''\) and \(y_n''\) have to be specified to enable a solution to be obtained from the \(n-2\) equation (3.3). If they are set to zero, the so-called ‘natural spline’ will result. Other alternatives are to make them linear combinations of the internal second derivatives; Späth (1995, p. 91) lists several possibilities. The following section uses the above notation and ideas to deal with endings where there is not necessarily a node to fix the spline.

### 4. The end conditions

What sets the following treatment apart from conventional spline interpolation is that the spline segments at the ends of the series do not necessarily collocate with the endpoints. The reason is that the spline curves are fitted to extrema and the endpoints may not qualify to be included. Rules must be set to define how to treat the various end conditions that may occur. What we have decided to do here is to ignore the endpoints to start off with and project the last spline segment at each end to a mirror of the penultimate extremum points, each as far from the end of the series as is the penultimate extremum point. There is then a choice as to how to specify the end conditions controlled either by setting the first or second derivatives at the mirrored point; we choose to set the second derivative to one of a selection of preset values. It is this treatment that is detailed in this section.

The algorithm is described by examining two cases at the start of the series at \(k=1\); the argument carries over naturally to \(k=n\). The first case is when \(Y_1\) is identified as an extremum \(y_1\), in which case \(s_1(x)\) collocates with the point \((x_1, y_1)\); this point is defined as an extremum if, after the first fit of the splines in
a sifting process, it lies on or outside the envelope defined by the upper and lower splines. The second case is where the point \((x_1, Y_1)\) is identified as lying within the bounds of the spline envelope defined by the extrema internal to the endpoints and is ignored in the definition of the bounding splines. In both the cases, a virtual point \((x_0, y_0)\) is identified where \(x_0 = 2x_1 - x_2\) and \(y_0 = y_2\), so that it is a reflection of the extremum point \((x_2, y_2)\) about \(x_1\). This choice does not necessarily imply that the spline segment from \(x_0\) to \(x_2\) is symmetrical about \(x_1\), but more on that later.

It is necessary to be able to exercise the option of either setting \(y_0'' = 0\) (a natural spline at \(x_0\)) or setting \(y_0'' = y_2''\). The latter is designed to provide as much symmetry (reflection) about the endpoints at \(x_1\) or \(x_n\) as possible. It is noted here that if \(p = 0\) (so that the splines are all cubic), then the latter choice \((y_0'' = y_2'')\) does in fact produce pure reflection at \(x_1\), because with the symmetry of the zeroth and second derivatives, all odd-order functions disappear in the spline definition and \(y_1'' = y_0'' = y_2''\), with \(y_1''\) implicitly equal to zero. In the cases where \(p \neq 0\), \(y_1''\) is not necessarily zero, although it is usually close to zero, resulting in a very flat appearance at the ends. These ideas will become clearer after the following development.

The application of the algorithm proceeds with chosen values of the pole parameter \(p\) and the end condition parameter \(r\) (which takes on a value of 0 or 1, the latter choice demanding fully reflective ends) and is done in two independent steps, first to the upper bounding extrema and then to the lower bounding extrema. We will treat only the upper bound because the lower bound case is a reflection of the description.

The first operation is to fit the splines to all the upper bounding extrema between \(x_0\) and \(x_n + 1\), the latter being the virtual point at the upper end of the sequence of observations; this first operation omits the data points at the ends, \(x_1\) and \(x_n\). If either one or both of the endpoints in question (at 1 or \(n\)) is found to lie above the spline fitted to the upper extrema and the virtual points, then it is included as an extremum and the splines are recalculated including the point(s) in question. On the other hand, if both the endpoints \((x_1, Y_1)\) and \((x_n, Y_n)\) lie within the bounds of (closer to the \(x\)-axis than) the last spline segments, that solution is adopted as the defining spline. We proceed with the description of the calculation of the spline at the lower end. The simpler case of \(s_1(x_1)\) collocating with the point \((x_1, y_1)\) is described first, followed by the case where the endpoint is omitted from the calculation as in the first operation described above.

(a) The case where \(s_1(x_1) = y_1\)

In the case where the spline curve goes through the endpoint \((x_1, y_1)\), an extra equation labelled \(k = 1\) in the set of equation (3.3) is developed. It is

\[
h_0 y''_0 + (2 + p)(h_0 + h_1)y''_1 + h_1 y''_2 = (d_1 - d_0)q.\]

Here, we set \(y''_0 = r \cdot y''_2\) (we can choose the value of the reflection coefficient \(r\) to suit the application), then in terms of the symmetry conditions we have \(h_0 = h_1\) and \(d_0 = -d_1\), (the last relation as the points at \(x_0\), \(x_1\) and \(x_2\) are the vertices of an isosceles triangle). The equation can be written in terms of ‘known’ values.
Equation (3.1) can immediately be defined without modification for the additional $k=1$; proceed as usual to calculate the spline interpolant starting from $x_1$, but now with $k=1$ equation.

The obvious modification of equation (4.1a) also applies for $k=n$ when $s_{k-1}(x_n) = y_n$.

\[
2(2 + p)h_1 y_1'' + (1 + r)h_1 y_2'' = 2d_1 q. \tag{4.1a}
\]
The case where $s_1(x_1)$ misses $y_1$

When the spline segment between $x_0$ and $x_2$ misses $y_1$, a new condition in the end interval occurs. In this case a sketch will help to convey the ideas. In figure 3, the symmetry of location of points 0 and 2 is maintained but point 1 on the boundary has been omitted because it does not lie on the spline, so does not influence the equation. Again, we have the choice of a natural end condition or a reflection of the spline curve around the vertical line through $x_1$; the first condition is obtained by setting $r=0$ in the equation $y_0'' = r \cdot y_2''$, the second condition by setting $r=1$. Figure 3 shows the case when $r=0$, the natural spline condition, the values of $y_0$ and $y_2$ are the same, but the curvature at point 2 is not copied at point 0.

Turning to the set-up of the equations again, $x_0 = 2x_1 - x_2$ (and $y_0'' = r \cdot y_2''$ as above), but there is no extra equation generated; however, equation (3.3) needs to be modified for the case $k=2$. It becomes

$$(h_0 + h_1) y_0'' + (2 + p)(h_0 + h_1 + h_2) y_2'' + h_2 y_3''$$

$$= [(y_2 - y_0)/(h_2 - (y_2 - y_0)/(x_2 - x_0))],$$

(b) The case where $s_1(x_1)$ misses $y_1$

Table 1. Ottapalam: number of sifts required to identify each IMF as a function of $p$.

| $p$ | IMF1 | IMF2 | IMF3 | IMF4 | IMF5 | IMF6 | total |
|-----|------|------|------|------|------|------|-------|
| 0   | 7    | 8    | 7    | —    | —    | —    | 22    |
| 0.5 | 7    | 7    | 7    | —    | —    | —    | 21    |
| 1   | 7    | 8    | 7    | —    | —    | —    | 22    |
| 2   | 7    | 8    | 7    | —    | —    | —    | 22    |
| 5   | 7    | 8    | 7    | 6    | —    | —    | 28    |
| 10  | 7    | 8    | 8    | 7    | —    | —    | 30    |
| 20  | 7    | 8    | 8    | 7    | —    | —    | 30    |
| 50  | 7    | 8    | 8    | 8    | —    | —    | 31    |

Table 2. Ottapalam: IMF and residual variance as a percentage of the observed sequence for each $p$. (Note the relatively very large variance for IMF2 (italicized) computed with the cubic spline ($p=0$), demonstrated dramatically in figure 5.)

| $p$ | IMF1 | IMF2 | IMF3 | IMF4 | IMF5 | residual | total |
|-----|------|------|------|------|------|----------|-------|
| 0   | 0.692 | 1.172 | 0.625 | —    | —    | 0.365    | 2.854 |
| 0.5 | 0.669 | 0.267 | 0.198 | —    | —    | 0.142    | 1.276 |
| 1   | 0.653 | 0.263 | 0.206 | —    | —    | 0.151    | 1.273 |
| 2   | 0.634 | 0.253 | 0.201 | —    | —    | 0.140    | 1.228 |
| 5   | 0.605 | 0.222 | 0.147 | 0.011| —    | 0.079    | 1.064 |
| 10  | 0.588 | 0.224 | 0.148 | 0.013| —    | 0.069    | 1.042 |
| 20  | 0.577 | 0.237 | 0.148 | 0.012| —    | 0.062    | 1.036 |
| 50  | 0.568 | 0.238 | 0.144 | 0.010| —    | 0.057    | 1.016 |

(Proc. R. Soc. A (2008))
Because $y_2 = y_0$, the equation reduces to

$$[(2 + p)(2h_1 + h_2) + 2s \cdot h_1]y''_2 + h_2y'''_2 = d_2 q.$$  \hspace{1cm} (4.1b)

To be able to compute the spline interpolant in the interval $(x_1, x_2)$ at the $X_j$ points, it would have been computationally convenient to be able to evaluate a surrogate endpoint as $y_1 = s_1(x_1)$ and $y''_1 = s_1(x_1)''$ as was done in the case of the cubic spline and proceed by using the conventional computation procedure involving the coefficients of equation (3.2) in interval 1. Unfortunately, when $p \neq 0$ the support over $(x_1, x_2)$ for this short-cut procedure does not provide an affine transformation of the equation spanning $(x_0, x_2)$. The work-around
necessitates computing the spline segment spanning \((x_0, x_2)\), but starting from \(x_1\) \((t = u = \text{one-half})\), which requires an unconventional dedicated spline interpolation algorithm. The following modifications of equation (3.2) are needed in the case of \(k = 1\): 

\[
D_1 = \frac{y''_2(2h_k)^2}{q}, \quad C_1 = r \cdot D_1, \quad B_1 = y_2 - D_1, \quad A_1 = y_2 - C_1. \quad (4.2)
\]

The obvious modification of this set of equations is obtained for the upper end of the sequence, in the case where the spline segment between \(x_{n-1}\) and \(x_{n+1}\) misses \(Y_n\). This concludes the development of the algorithm, which is applied to three example time series in §5.

**Table 4. Bundarra: IMF and residual variance as a percentage of the observed sequence for each \(p\).**

| PAR | IMF1 | IMF2 | IMF3 | IMF4 | IMF5 | IMF6 | residual | total |
|-----|------|------|------|------|------|------|----------|-------|
| 0   | 0.652| 0.215| 0.057| —    | —    | —    | 0.076    | 1.000 |
| 0.5 | 0.629| 0.216| 0.046| 0.069| —    | —    | 0.037    | 0.997 |
| 1   | 0.615| 0.213| 0.040| 0.065| —    | —    | 0.039    | 0.972 |
| 2   | 0.599| 0.211| 0.035| 0.055| —    | —    | 0.040    | 0.940 |
| 5   | 0.581| 0.203| 0.022| 0.080| 0.021| —    | 0.022    | 0.928 |
| 10  | 0.574| 0.198| 0.017| 0.023| 0.040| 0.04 | 0.030    | 0.920 |
| 20  | 0.571| 0.194| 0.019| 0.017| 0.038| 0.0045| 0.085    | 0.928 |
| 50  | 0.569| 0.193| 0.020| 0.052| —    | —    | 0.053    | 0.887 |

**Figure 6.** The first sift of IMF1 for the Winthrop series, for a range of the rational spline tension parameter \(p\). Since the plot shows the first sift of IMF1, the observed record is the one being treated to the EMD sifting process and hence the splines all go through the same extrema (triangles, observed; lines: red, cubic; PAR (orange, 0.5; lime, 1; bright green, 2; turquoise, 5; light blue, 10; blue, 20; black, 50)).
5. Three example time series

Three example time series were selected to demonstrate the types of problem encountered with EMD in practice and how they can be treated using rational splines for interpolation of extrema. The set comprises a bad, an average and a good example of series from the points of view of overshoot and covariance between the components of the decomposition. The three series are annual rainfall totals from three gauges from around the world: Ottapalam (N=63 years, India), Bundarra (N=74 years, Australia) and Winthrop (N=73 years, USA).

Figure 4 shows the first sift of IMF2 for the Ottapalam series, for a range of the rational spline tension parameter \( p \) (in all applications, the end reflection parameter \( r \) is set to 1). Since we are plotting the first sift of the IMF2, the ‘observed’ record is now the ‘1st residual’ (observed—IMF1). Because the 1st residual will be different for each setting of \( p \), IMF1 will be different in each case. This means that not all the \( p \) envelopes go through the extrema on the same plot. However, to provide a visual comparison, the 1st residual with \( p = 0 \) (cubic splines) is used as the ‘data’. On this point, it will be seen from figure 4 that the envelopes of the maxima and minima at the ends of the series come closer to each other as \( p \) increases. In addition, all flatten to a zero gradient at the ends, with the exception of the minimum at the left of the series; note that the spline for \( p = 0 \) goes through the point at the end, the case where \( s_1(x_1) = Y_1 \). This continues as the splines get tighter (the ends are not flat), but the points are not shown, even though their positions change owing to the influence of the first IMF.

Table 1 lists the number of IMFs, and sifts for each, as a function of \( p \). There is generally an increase of the number of sifts and IMFs as \( p \) increases. Table 2 lists, for the Ottapalam series, the variance of the components of the EMD analysis as a percentage of the variance of the observed sequence, for each \( p \). Note the relatively very large variance for IMF2 computed with the cubic spline (\( p = 0 \)), demonstrated dramatically in figure 4. Note also the drop in total variance to a respectable 1.064, when \( p \) increments from 2 to 5, but with a concomitant increase of the number of IMFs from 3 to 4.

Figure 4 shows a gross case of overshooting by a spline. On the other hand, figure 5 shows relatively good behaviour in the first sift of IMF1 for the Bundarra series, for a range of the rational spline tension parameter \( p \). Since the plot shows the first sift of IMF1, the observed record is now the one being treated to the EMD process, hence the splines all go through the same extrema. Here is an example of mild over- and undershoot which shows good behaviour of the endings. Table 3 shows that there is an increase in the number of IMFs (and concomitant sifts) as \( p \) increases to 5 and again, the sum of the variances of the components reduces with an increase in \( p \). The sum of the variances is below 100% because in all cases except the cubic spline (\( p = 0 \)), the sum of the covariances is positive, so that the variances must sum to less than 100% by equation (2.1) (table 4).

In the final set of results, figure 6 shows the fitting of the extremal envelope splines to IMF1 of the Winthrop series. It seems that even the cubic splines are near to polygon fits and there are no apparent difficulties. Tables 5 and 6 show similar trends to their earlier counterparts with the outcomes for \( p = 5 \) giving the smallest variance sum and the lowest number of sifts of those sets with five IMFs.
6. Conclusions

This paper demonstrates that there is a sensible alternative to fitting cubic splines to extrema in the sifting process associated with EMD. The candidate spline model used here, the rational spline, was explained in full and its benefits explored and reported. In particular, the treatment of the endings of the extremal splines was systematized and explained in detail. This treatment allows some flexibility in the practical application of EMD, in an attempt to militate against the undesirable properties of cubic splines, which are a special case of rational splines.

It is not wise to draw general conclusions from a set of only three disparate series. This was not the purpose of the paper. However, the choice of spline envelope type defined by the tension parameter \( p \) clearly influences the final number of IMFs, the number of sifts required to identify the IMFs and the degree of orthogonality of the resultant IMFs and residual. The cubic setting (\( p = 0 \)) generally results in a low number of IMFs and sifts, but higher covariance terms (less orthogonal) relative to the tauter spline settings. The benefit of the rational spline method is that this interplay between spline tension and IMF characteristics can be investigated, whereas the original cubic method provides a single output. From the three sets it is clear that there are some better-behaved combinations than others and they come from setting the tension parameter \( p \) to values in the region between 1 and 5, giving sensible fits and pleasing curves, besides the numerical criteria.

Table 5. Winthrop: number of sifts required to identify each IMF as a function of \( p \).

| PAR | IMF1 | IMF2 | IMF3 | IMF4 | IMF5 | total |
|-----|------|------|------|------|------|-------|
| 0   | 11   | 7    | 9    | —    | —    | 27    |
| 0.5 | 10   | 7    | 9    | —    | —    | 26    |
| 1   | 10   | 9    | 8    | 8    | 7    | 42    |
| 2   | 10   | 8    | 8    | 7    | 7    | 40    |
| 5   | 10   | 7    | 7    | 8    | 7    | 39    |
| 10  | 9    | 7    | 8    | —    | —    | 24    |
| 20  | 9    | 7    | 16   | 8    | —    | 40    |
| 50  | 10   | 7    | 13   | 7    | —    | 37    |

Table 6. Winthrop: IMF and residual variance as a percentage of the observed sequence for each \( p \).

| PAR | IMF1 | IMF2  | IMF3  | IMF4  | IMF5  | residual | total   |
|-----|------|-------|-------|-------|-------|----------|---------|
| 0   | 0.594| 0.326 | 0.100 | —     | —     | 0.231    | 1.251   |
| 0.5 | 0.568| 0.293 | 0.102 | —     | —     | 0.261    | 1.223   |
| 1   | 0.555| 0.284 | 0.139 | 0.160 | 0.006 | 0.027    | 1.170   |
| 2   | 0.539| 0.271 | 0.136 | 0.153 | 0.006 | 0.034    | 1.138   |
| 5   | 0.517| 0.252 | 0.127 | 0.117 | 0.012 | 0.036    | 1.061   |
| 10  | 0.504| 0.239 | 0.124 | —     | —     | 0.250    | 1.116   |
| 20  | 0.495| 0.216 | 0.067 | 0.160 | —     | 0.210    | 1.149   |
| 50  | 0.485| 0.218 | 0.066 | 0.145 | —     | 0.206    | 1.119   |
We think rational spline-based EMD is good and intend to test it on a larger dataset in a future paper so that we may offer some useful guidelines for helping to automate the EMD procedure, reducing the amount of subjectivity in what should be an impartial objective process of time-series analysis.

This research was financially supported by Australian Research Council (ARC) grant no. DP0773016. The discussions with John Fenton about exponential and taut splines are gratefully acknowledged.

References

Balocchi, R., Menicucci, D., Santarcangelo, E., Sebastiani, L., Gemignani, A., Gellarducci, B. & Varanini, M. 2004 Deriving the respiratory sinus arrhythmia from the heartbeat time series using empirical mode decomposition. Chaos Solitons Fract. 20, 171–177. (doi:10.1016/S0960-0779(03)00441-7)

Chiew, F. H. S., Peel, M. C., Amirthanathan, G. E. & Pegram, G. G. S. 2005 Identification of oscillations in historical global streamflow data using empirical mode decomposition. In Seventh IAHS Scientific Assembly, Foz do Iguacu, Brazil, pp. 53–62. IAHS Publisher 296.

Fauchereau, N., Sinclair, S. & Pegram, G. 2008 2-D empirical mode decomposition on the sphere, application to the spatial scales of surface temperature variations. Hydrol. Earth Syst. Sci. Discuss. 5, 405–435. (sRefID:1812-2166/hessd/2008-5-405)

Flandrin, P. & Gonçalves, P. 2004 Empirical mode decomposition as data-driven wavelet-like expansions. Int. J. Wavelets, Multiresolution Inf. Process. 2, 477–496. (doi:10.1142/S0219691304000561)

Gloersen, P. & Huang, N. 2003 Comparison of interannual intrinsic modes in hemispheric sea ice covers and other geophysical parameters. IEEE Trans. Geosci. Remote Sens. 41, 1062–1074. (doi:10.1109/TGRS.2003.811814)

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q. N., Yen, C., Tung, C. C. & Liu, H. H. 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. A 454, 903–995. (doi:10.1098/rspa.1998.0193)

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. R. Soc. A 459, 2317–2345. (doi:10.1098/rspa.2003.1123)

Hwang, P. A., Huang, N. E. & Wang, D. W. 1999 A note on analyzing nonlinear and nonstationary ocean wave data. Appl. Ocean Res. 25, 187–193. (doi:10.1016/j.apor.2003.11.001)

Olheide, S. & Walden, A. T. 2004 The Hilbert spectrum via wavelet projections. Proc. R. Soc. A 460, 955–975. (doi:10.1098/rspa.2003.1199)

Peel, M. C. & McMahon, T. A. 2006 Recent frequency component changes in interannual climate variability. Geophys. Res. Lett. 33, L16810. (doi:10.1029/2006GL025670)

Peel, M. C., Amirthanathan, G. E., Pegram, G. G. S., McMahon, T. A. & Chiew, F. H. S. 2005 Issues with the application of empirical mode decomposition analysis. In Proc. MODSIM 2005 International Congress on Modelling and Simulation, Melbourne, Australia, 12–15 December 2005 (eds A. Zerger & R. M. Argent), pp. 1681–1687. Modelling and Simulation Society of Australia and New Zealand.

Peel, M. C., Pegram, G. G. S. & McMahon, T. A. 2007 Empirical mode decomposition: improvement and application. In Proc. MODSIM 2007 International Congress on Modelling and Simulation, Christchurch, New Zealand, 10–13 December 2007 (eds L. Oxley & D. Kulasiri), pp. 2996–3002. Modelling and Simulation Society of Australia and New Zealand.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 1992 Numerical recipes in C. Cambridge, UK: Cambridge University Press.

Proc. R. Soc. A (2008)
Rilling, G., Flandrin, P. & Gonçalvès, P. 2003 On empirical mode decomposition and its algorithms. In Proc. IEEE-EURASIP Workshop on Non-linear Signal and Image Processing NSIP-03, 8–11 June, Grado, Italy.

Sinclair, S. & Pegram, G. G. S. 2005 Empirical mode decomposition in 2-D space and time: a tool for space-time rainfall analysis and nowcasting. Hydrol. Earth Syst. Sci. 9, 127–137. (www.copernicus.org/EGU/hess/hess/9/127/SRef-ID:1607-7938/hess/2005-9-127)

Späth, H. 1995 One dimensional spline interpolation algorithms. Wellesley, MA: AK Peters.

Zhang, R. R., Ma, S. & Hartzell, S. 2003 Signatures of the seismic source in EMD-based characterization of the 1994 Northridge, California, earthquake recordings. Bull. Seismol. Soc. Am. 93, 501–518. (doi:10.1785/0120010285)