Dear Dr. Bevacqua,

I appreciate your review of this paper and the obvious effort you took. While I don't agree with some of your comments, I do respect them and will describe here in detail why I disagree and how I have modified the paper to make my points more clear.

In the following discussion, your original comments are in italics, my response is in normal text, and any additions to the text will be in quotes. Because some of your comments are lengthy, I have edited them here (indicated with …), but I will address all points.

Cheers,

Don Chambers

1. Pag 1836, line 12
The signal on which you run the EMD is built only trough noise. Explain better this, please. You should also explain why this experiment is interesting for the target of your paper.

I tried to explain the motivation in the following paragraph, but I can see now that could be confusing. Thus, I have substantially revised this section to better explain my motivation by bringing that discussion into this paragraph and expanding it. I also believe this will answer several of your other questions, notable Comments 6, 7, and 8.

"Moreover, Wu and Huang (2004) have shown that EMD behaves as a low-pass filter. If one runs random noise with a normal Gaussian distribution through the process, low-frequency signals will be seen in the resulting IMFs. They found there is roughly a doubling of the average period with each subsequent mode. This is a significant issue. It means that any random (or near-random, high-frequency) signal will propagate into low-frequency signals in the recovered IMFs. Wu and Huang (2009) proposed a method to quantify the uncertainty caused by this behavior by computing an ensemble mean of IMFs, starting from the same time-series but with different amounts of added random noise.

However, this method ignores that all geophysical time-series have an underlying real signal that has high variance and little serial correlation; i.e., a high-frequency, near-random signal. This "signal" will also be filtered by the EMD process and will likely appear as a quasi-stationary oscillation in higher order IMFs that is not real. Although adding multiple realizations of random noise to a time-series will account for uncertainty in the IMFs from random error in the measurement, it will not account for the shifting of high-frequency signal to low-frequency signal in the recovered IMFs from the high-frequency signal. One of the assumptions in EMD processing is this is captured in the lowest IMFs. Our testing indicates it is not.

To demonstrate the potential size of this problem, we ran EMD on a monthly-resolution time-series that is 150-years long with randomly-correlated values that have a standard deviation of 60 mm. We used 60 mm because this is the standard deviation of residual monthly sea level at San Francisco after fitting and removing a quadratic function plus annual and semiannual sinusoids, so is representative of high-frequency sea level at a typical site, although some sites can have significantly higher variability. We ran another case using a colored noise model that exactly reproduces the autocovariance of the San Francisco tide gauge residuals. The results were nearly identical to the ones shown with the random residuals, so we choose to use random values as they are faster to compute for the several thousand simulations we plan to run. The EMD finds IMFs that have quasi-period fluctuations of nearly 60-years and amplitudes as large as 10 mm (Figure 1); fluctuations at quasi-30-year periods are the same magnitude."

2. Pag 1839, line 7
I suggest to insert the value of the correlation of SOI and PDO also before you have worked on them.

We have added that information in the revision:

"Secondly, because the two indices are slightly correlated (-0.21, $p < 0.001$) due to similar interannual (< ten year) variations…"

3. Pag 1840, line 11
Where will you note that none accurately captures the input seasonal variation?

We note it right there. We do not show it, because it is beyond the scope of the paper and most authors use EMD to extract low-frequency variations, not seasonal. This is discussed in the introduction. Thus, we feel showing this is irrelevant.

4. When you compute the correlation between the "best IMF" and the simulated oscillation for the 1000 simulations, it should be interesting for the reader seeing an histogram (for the case 1) to have a better idea of the distribution of this parameter (with also a mean value with error).

We have added a new figure (Figure 4 in revision) and added the statistics. The new text reads:

"Figure 4 shows the histogram of computed correlations. Note that the correlations were not the same in every simulation. The 13-year oscillation had a mean correlation of 0.66 (standard deviation = 0.09), the 55-year had a mean of 0.52 (standard deviation = 0.11), while the mean correlation of the 80-year signal was 0.74 (standard deviation = 0.09)."

5.
Pag 1842, line 12
Please, explein better the following part, it is not clear: "We isolated this signal by looking at the autocorrelation of the remaining IMFs uncorrelated with either PDO or ENSO. The IMF with an autocorrelation greater than 0.9 at a lag of 1 year was selected."

I can see this was a little confusing. I've modified the text thusly:

"In addition, we found in nearly every case (99%) the EMD computed one to two IMFs with a periodic signal that did not correlate highly with either PDO or ENSO, but had a low-frequency. Because this was not always contained in a single IMF between the two prescribed periodic fluctuations, we had to adapt a method to search for it. We isolated this signal by looking at the autocorrelation of the IMFs after removing those correlated with PDO or ENSO, as well as the last mode. To find the mode with the longest-period fluctuation, we examined the autocorrelation at a 1-year lag. Only IMFs with an autocorrelation greater than 0.9 at a lag of 1-year were examined, and if two existed, the one with the higher autocorrelation was selected."

6. Pag 1836, from line 20
You cite (Wu et al., 2004) saying what they do in their paper, but I know that they do an other thing. I know that they propose a test useful when you analyze a signal in which is present some noise. The test is useful to identify the IMFs due to noise ("non significative IMFs"), in such a way to not consider them for a physic discussion about the intrinsic oscillations present in the signal. Perhaps do you talk about the work present in the other reference you cite in line 20, or about (Wu et al., 2007)? This confused me because you propose your approach as alternative also to their works, but actually I have a comment exactly on the test of (Wu et al., 2004), in particular I don't understand why you don't apply the test (and so I write what follow in 8).

Thanks for pointing out the citation was wrong. You are correct; it should be Wu and Huang (2009), for their Ensemble EMD. I have corrected this in the text (see reply for Point 1). As far as why I don't compute the confidence tests of modes described in Wu and Huang (2004), I don't need to, since I know what the signal should be and can compare the IMFs to it. This is a much more stringent test than the statistical testing of Wu and Huang (2004), which assumes high-frequency signal is properly captured in low IMFs and does not distort higher IMFs. I will discuss this more in my response to Comment 8.

7. You decide to use the random noise to represent high-frequency variability. You chose a noise with a variance to match the variance of the difference between the original data and the model. This signal is actually due both to noise part and some other signal with appreciable characteristic frequencies (one way to appreciate these is, for example, applying the EMD on this). So using the noise to represent this "high frequency variability", you actually represent only the noise of this (and you should say this).

Here is where we begin to disagree. Using the random values is not just representing noise in this simulation. It is representing the magnitude of high-frequency variability that is real signal in a simple, repeatable simulation. Real signals will of course have serial correlations in the data, but for tide gauge measurements they are not large after a few months. I actually did test with a colored-noise model that reproduces the autocovariance of the San Francisco tide gauge residuals quite well. The results were no different than the random values, so I chose to only discuss those as they are more easily reproduced.

I added a comment regarding this in the section I added to your Comment 1:

"We ran another case using a colored noise model that exactly reproduces the autocovariance of the San Francisco tide gauge residuals. The results were nearly identical to the ones shown with the random residuals, so we choose to use random values as they are faster to compute for the thousands of simulations we plan to run."

Moreover, one might expect that by averaging 1000 different IMFs computed from different randomly simulated residuals added to the base signal, that the mean would be zero. This is NOT the case, and what this simulation clearly demonstrates. This is exactly the type of error that can still remain in the Ensemble EMD method, which is what motivated this paper.

8. Performing your experiment, in any of 1000 run, I don't understand why you don't apply the noise test (Wu et al., 2004), that give you the possibility to isolate, and not consider, the part of signal due to noise ("non significative signal"). I know that clearly in the assumption that you represent the "high frequency signal" with noise, all the noise is significative (because you insert it!), nevertheless in this way you discuss also about IMFs due to noise. The crucial point of this, is that performing the EMD on a generic real signal you can apply the test and so avoid to consider the part of signal due to noise. The problem of non apply the test could be that, if you find a "problematic" IMF, you are finding a "problem" in a IMF that could be not actually significative (i.e. due to noise), so in a IMF that is actually due to a part of signal that you can avoid to consider. I observe that you don't discuss about the first IMFs, and usually applying the test you discover that IMFs due to noise are the first but it's not absolutely a rule; so in any of the 1000 simulations, if you find a "problem" in one IMF, before say that this is a real "problem" you should ascertain that is not due to noise, applying the test.

There is no reason to conduct the significance testing of Wu and Huang (2004) in this experiment, as I know what the answer should be and can compare the IMFs to it. This is a much more stringent test than the statistical testing of Wu and Huang (2004), which assumes high-frequency signal is properly captured in low IMFs and does not distort higher IMFs. This is the whole point of the experiment – to determine if the EMD method can find input, known signals in the presence of high-frequency variability with realistic variance. Although I have not performed the significance testing, I suspect it will say the lowest high-frequency modes are not significant, but the higher, low-frequency ones are. Just based on the spread of the ensemble, they appear significant. But they are wrong!

I've added a paragraph after the discussion of the results for Case 1 to try to highlight the probably misinterpretation of an analysis of the EMD results for Case 1. Note, Figure 5 is the old Figure 4.

“More importantly, consider the interpretation of the results from this simplistic simulation in terms of longer-term climate change if only the EMD results (Figure 5) were analyzed. Based only on the returned IMFs, one could easily argue that there was no significant low-frequency variation in the sea level before 1950, then a rather dramatic rise in the 1970s, followed by a return to normal condition. In fact, there were equally large sea level shifts in the early part of the simulated record that were lost due to the way the EMD method partitions the real signal. “

9. Comment on “Case 2” You study if it is possible to reproduce each simulated signal trough one IMF. It should be observed a conceptual difference that exist between case 1 and 2. Actually, in principle, you can reproduce each simulated signal trough one IMF (for each signal) only in the case 1, because sinusoids respond to the definition of a IMF (Huang et al., 1998). In the case 2, instead, because of ENSO/PDO doesn’t respond to IMF’s definition, you know already in principle that you can’t capture this signal trough a single IMF. In principle, you should need at least two IMF (the sum of two IMF doesn’t have to respond to the definition of IMF) to reproduce that signal.

So a part of the signal of ENSO/PDO is diffused (necessarily) in other IMFs and we can expect this before performing the EMD.

I don’t disagree with anything you say here. I point out what I wrote in the introduction:

“However, there are some potential pitfalls that we believe have not been fully addressed in previous papers utilizing the method. First and foremost, EMD is a purely mathematical deconstruction of the data, with no regard to intrinsic covariance of the signals or physics. Second, it assumes that IMFs are comprised of fluctuating signals where the magnitude of nearby peaks and troughs are balanced to create a zero mean – an assumption not based on any physical requirement, as real observations can have quite large ranges in magnitudes, especially sea level data affected by climate signals and synoptic storm events.”

I stand by all of that, which is in close agreement to what you state. EMD is a mathematical deconvolution of the data, and no individual mode has any physical meaning. But many scientists are trying to analyze individual IMFs for climatic signals. I give a list of some I know about in the beginning section. Thus, I believe my analysis of whether EMD can extract physically meaningful signals in a single mode is justified. I also point out that although you state that in Case 1 you should be able to extract the physical modes in a single EMD, I demonstrate that you cannot.

You say (pag 1843, lines 1-4): “We know of none that find multiple modes that add up to correlate with an ENSO index. Thus, we argue it is more relevant to quantify if EMD can extract physically meaningful climate modes than whether it can extract modes with interannual and multi-decadal variability”. Performing the decomposition with other techniques you obtain different results, clearly we know that each techniques work in a different way. I agree that it’s very important thata technique give you modes that have *physical meaning. But with EMD, actually mathematically you already know before performing the analysis that you can't obtain this mode in a unique IMF (the same for PDO). After a decomposition, for sure if you retain for some reason that physically this signal is a "unique signal" you have to sum the IMF that give you the signal (clearly if you know already what you want to build, after performing EMD), but EMD can't say this to us (see for example Alberti et al., 2014. NOTE: the citation of this reference should be interesting to give the reader the awareness, although this "critical" paper, that EMD is a delicate tool but useful when used in the right way).*

Thanks for sending me that reference. I had not seen it in my literature review. In that case, though, you and your colleagues were using the EMD as a type of low-pass filter by adding up the higher modes that did not pass the significance testing. I have no qualm with that. I have added a comment on low-pass filtering with EMD in the conclusions:

"EMD is a quick and relatively easy tool to identify possible multidecadal fluctuations in a sea level record. However, it should not be used solely to quantify magnitude and phase, nor should analysis of climatic signals be based on a single IMF. One should also be cautious in interpreting acceleration computed from the final IMF, especially in light of the significant errors found in the early and later parts of the low-frequency IMFs (Figures 5, 6, and 7). Where EMD has shown to be useful has been in low-pass filtering data to reduce the impact of high-frequency variability and noise (e.g., Alberti et al., 2014). In that case, the sum of the higher IMFs are used as the low-pass filter."

…
Besides, looking very crudely at fig. 5 seems that the sum of "unsimulated low frequency" IMF and the "PDO IMF" give a good approximation of the total PDO signal, except for first years (regarding this, however, I already said in you that it wasn't clear what you said in pag 1842, line 12).

But the change in the early part of the "PDO" IMF could lead to erroneous climatic interpretations. I added a small comment to the section where I discussed this, noted here in bold.

"As with the ENSO-mode, the mean PDO-mode IMF tracks the general periodicity of the PDO, although the amplitudes are on average too small. Again, the standard deviation suggests any single simulation would give a considerable range of amplitudes. We note that as with the Case 1 results, there is a tendency for an increasing amplitude in time for the mean IMF, **which could be misinterpreted as a sign of climate change; this is inconsistent with the true signal,** where the first two peaks in the given PDO signal are roughly the same magnitude."

According to me, it should be interesting to perform the same experiment using, instead of ENSO and PDO signal like simulated signal, some IMFs ("simulated IMFs") obtained performing EMD on an other signal. I suggest to do it. You could use also the ENSO and PDO to extract and define the "simulated IMFs". I think this procedure should be *interesting because in this case, like in case 1, the EMD could actually extract the "simulated IMFs" in a new IMFs from a theoretical point of view.*

I don't see how this would provide any more insight than the experiments I have already conducted and decline to add them.

10. Pag 1844, line 1 (About the case 3)
You say: "By enforcing an unrealistic balance of equal highs and lows, the method creates a low-frequency oscillation that does not exist." However I think that should be necessary comment the result of EMD's application to "case 3" comparing this with "case 3 without add the extreme event". I say this because also in "case 3 without add the extreme event" I expect that you will obtain some oscillation that "does not exist" (no prescribed oscillations), and this should be clarified.

I showed the result of the random-only case in Figure 1, but I did not reference it here. I've fixed that in the revision. The addition is shown in bold:

"Because the EMD method implicitly assumes local highs are balanced perfectly by nearby lows, it cannot handle an extreme event like this. By enforcing an unrealistic balance of equal highs and lows, the method creates a low-frequency oscillation that does not exist. **Although the random-only case (Figure 1) also produces low-frequency erroneous oscillations, the amplitudes are significantly less for the longer-period IMFs.** With a larger pulse, the magnitude of the error is even higher."

11. Period IMFs How do you obtain the periods of IMFs? From instantaneous frequency, from values peak-peak or?

I assume this refers to the statement:

"Notice the large, non-stationary oscillation with a period of about 10-years in IMF6."

Since I don't dwell on the period and just use it to describe a rough period, I don't feel it is necessary to describe the details as it would clutter the text. It is based on dividing the length of the time-series by the number of peaks.

Technical Comments

1. Pag 1837, line 7
Before introduce the cases, you should add that you will analyze three cases. After this talk about them.

Two lines above that we state:

“Thus, we propose to test the EMD process not on real observations where one does not know the underlying modes, but on three simulated data sets where the modes are prescribed.”

2. "Data and methods"
I suggest to present the three cases in a more schematic way, to give a more immediate vision to the reader … (You could use the same division in “Results and analysis”). I suggest also to insert the analytic expression for the third case using a Dirac’s delta to underline that is only one the point in which you insert the extreme value.

This is more a matter of writing styles than a technical problem. We prefer the text the way it is written.

3. Pag 1842, line 11
The sentence:
“In addition, we found in nearly every case (99 %) the EMD computed an IMF with a periodic signal between the ENSO and the PDO signal.”
should be: “In addition, we found in nearly every case (99 %) the EMD computed an IMF with a periodicity between the periodicity of the IMFs designed to describe ENSO and the PDO.”

That sentence has already been revised according to Comment 5 above:

““In addition, we found in nearly every case (99%) the EMD computed one to two IMFs with a periodic signal that did not correlate highly with either PDO or ENSO, but had a low-frequency.”

4. Figures In figures in which the average periods of IMFs are missed, I suggest to insert them.

I’m sorry, but I do not understand this request and so do not know how to respond.

5.
Pag 1847, line 10
The title of the follow reference is not correct.
Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, E. H., Zheng, Q., Tung, C. C., and Liu, H. H.: The Empirical Mode Decomposition and the Hilbert spectrum for non stationary time series analysis, P. R. Soc. London, 454, 903–995, 1998.

Thanks for catching that. One of the pitfalls of typing in reference is missing a word or two in the title.