Interactive comment on "Geometric and topological approaches to significance testing in wavelet analysis" *by* J. A. Schulte et al.

Anonymous Referee #1

Received and published: 16 October 2014

Review of "Geometric and topological approaches to significance testing in wavelet analysis" by J. A. Schulte, C. Duffy, and R. G. Najjar Nonlin. Processes Geophys. Discuss., 1, 1331–1363, 2014 www.nonlin-processes-geophys-discuss.net/1/1331/2014/ doi:10.5194/npgd-1-1331-2014

The present study will be an important addition to the significance testing of wavelet spectra. It provides what may be a useful alternative to the existing "areawise" test and also provides a new topological approach. Before publication, however, several unresolved issues need to be addressed. First, the consequences of a major assumption made in the method need to be discussed. Second, the authors justify the need for this test by the "multiple testing problem" but then they do not show how their test improves upon the pointwise test that was the motivating issue. Third, they also need to improve

the figures and their discussion of them. Fourth, the authors should justify their climate examples that use short segments from available long climate series or use longer ones. Additionally, many minor issues of clarity need to be addressed. Putting some of this work in the context of other climate studies would also be helpful and increase the importance of this study for climate science.

General comments

I. The proposed geometric test suffers from a binary decision of a pointwise threshold "significance" or not. The authors showed some sensitivity to that threshold. The authors should at least discuss an alternative test, very similar in spirit, which does not use binary assessments: the false discovery rate (Wilks 2006).

II. The authors need to justify the "coordinate system" being scale index. The width of the analyzing wavelet changes as a function of scale. I expect that the significance of a wide patch at a small scale to be different from a wide patch at a large scale. The calculation may be "simpler", but it could also be the wrong area to assess. This is my single biggest concern with this testing procedure. The authors need to show that using the actual coordination system would result in the same distribution of chi.

III. The areawise test based on the reproducing kernel of Maraun et al. 2007 is strictly limited to Gaussian white noise. The present authors are also making use of the reproducing kernel in their equation 10 and subsequent steps. a) How does what should be a changing kernel as a function of the noise alter the effectiveness or sensitivity of both the areawise and this geometric test? This is particularly relevant for the comparison testing in Section 4.2 and may be an additional strength of the geometric test if it is less sensitive to the form of the noise and the error in the kernel.

b) The comparison testing of Section 4.2 needs to be performed on different AR1 noises from an AR1 parameter of 0 to nearly 1. The reproducing kernel of Maraun et al. 2007 becomes less and less relevant as the auto-correlation increases, but the area of random significant patches could continue to grow as the AR1 parameter is

increased.

IV. The authors need to better motivate including convexity in the testing procedure.

V. The present method still seems to suffer from the multiple testing problem. If I have 20 patches and find 2 that are geometrically significant, how is the probability that both were still the result of the noise process addressed?

VI. Recent work (Hanna et al. 2014) claimed to detect a trend in the variance of the NAO. The present study's wavelet analysis of the NAO and new statistical significance testing procedure would be the ideal place to evaluate that claim. The authors should comment on any significant changes in variance detected.

VII. Can the authors provide some computer code or pseudo-code for how to implement their procedure?

Specific comments

1. pg 1332 line 19. The introductory juxtaposition of "random" and "meaningful" does not make sense. I think what is meant is stochastic or deterministic. Random structures are meaningful. The assessment is essential to understand the predictability of the system.

2. line 1333 line 15. Some reference for the "climate science" procedure of comparing spectra to red noise should be given.

3. pg 1334 line 5. The use of the phraseology "Moreover, the areawise,..." is confusing. The term areawise has not been introduced or defined.

4. pg 1334 line 13. "holes" has not been defined. The meaning here is unclear. Please clarify.

5. pg 1335 line 20. What is meant by "Another interesting feature emerges: periods of reduced pointwise significance surrounded by regions of pointwise significance." I don't see anything like this indicated on the figure.

6. pg 1336 line 4, "reproducing kernel" should be defined before its importance is discussed.

7. pg 1336 lines 8 and 9. The structure of the vaguely referenced equation relating the reproducing kernel to the correlation between wavelet coefficients is important to the argument and explanation here. The equation should be reproduced and then cited.

8. pg 1336 line 9. Is the area "given by" the reproducing kernel, or is the typical patch area the area of the reproducing kernel?

9. pg 1336 line 9 and following. What is meant precisely by area in this context should be defined. Particularly because the subsequent area of the geometric test is different.

10. pg 1336 line. Is the test for "any" reproducing kernel or the reproducing kernel corresponding to the analyzing wavelet?

11. page 1336 line20. One is not assessing the significance of the wavelet coefficients. One is assessing the wavelet spectrum or the coefficients squared.

12. page 1337 line 4-10. The discussion of the illustration needs improvement. Please provide the reader some specific examples so that they know what they are looking at. At what time and scale are some of these features seen? Why is Figure 1 plotted so differently from Figure 2? What are the red noise parameters being used?

13. equation 7. It may seem pedantic, but please include how this discrete equation 7 follows from Green's theorem, which applies to integrals (perhaps in a small appendix or provide a reference).

14. equation 7. The variable n is not defined.

15. equations 8 and 9. Provide a reference for this definition of a centroid. Doesn't it have a fundamental problem when polygons intersect, such as we see in Fig. 2 at a scale of 5 years around 1990?

16. pg 1340 line 3. Why is it "...noted that all holes..." When would holes be relevant

for this procedure? Please clarify or remove.

17. pg 1340 line 14. Do patches of equal area "need to be distinguished"? I would think that they should have the same significance that would depend on how often they occur.

18. pg 1340 line 17. I don't understand why two patches with the same normalized significant area, regardless of shape, don't have the same significance. The authors need to better motivate in what context this difference in geometry matters. One could simply be testing the area without regard to this shape issue. If there was no reliance on the reproducing kernel (which becomes less and less relevant for strongly auto-correlated noise), the test should be on the distributions of A and nothing else.

19. pg 1340 line 20. Is the null distribution of chi independent of the form of the null hypothesis noise? If not, then the dependence should be explicitly stated here.

20. pg 1343 line 2. The "large number" should be stated. Their length in time should also be stated. Does the length matter?

21. pg 1343 line 10. How should one interpret the differences between the two tests? If a patch is areawise significant but not geometrically significant in particular, seems to possibly point to a substantial issue in the testing procedures, a problem with including convexity in the geometric procedure, or a problem with the reproducing kernel approach in both tests. These discrepancies need to be addressed and discussed in depth, as it goes to the heart of the point of this paper.

22. pg 1343 line 17. Why is significance level of 0.9 being used (I think you mean 0.1). Why not 0.05, as it more common and reduces the risk of the Type I error more?

23. pg 1343 lines 25-27. The multiple testing problem is still not resolved, at least in this discussion and application of the test. In Fig. 1, I see more than 20 patches and 3 are geometrically significant. Couldn't I have gotten that result from chance at the 10% level of the test?

24. pg 1343 line 27. I don't see any obvious seasonality in the wavelet power spectrum as shown. The time-averages of the wavelet power for each season would help to make the "variability" point. It is currently not supported by the figure.

25. pg 1344 line 3 and following. I don't see a period of 32 months or of 12 months plotted on the figure 2. It only goes to a period of 7 months.

26. pg 1344 line 26. The definition and method of calculating a "hole" needs to be given.

27. pg 1345 line 1. What is the sensitivity of the shape and amplitude of Fig. 5 to the choice of autocorrelation. Would 0.9 and 0.1 be different or the same? The authors are making generalizations based on just one parameter setting.

28. pg 1345 line 11. Why is an 80% significance level used here? 90% was used earlier. Both have a larger risk of a Type I error than the traditional 95%. (note that the nomenclature should actually be 20%, 10%, and 5% when the "significance" is being considered rather than the error bar).

29. pg 1345 and following. What null hypothesis is being compared in this simple test of white noise and a sinusoid. How different were the amplitudes? Is this actually a general result or specific to the parameters chosen for the series? A similar lack of specificity and detail applies to the rest of the discussion through page 1349. The results only have theoretical implications if they are generalizable. From the present discussion, this cannot be assessed.

30. pg 1349 line 9. Couldn't the same effect be found in the linear AR2 model for some choices of its two parameters? I don't think that nonlinearity needs to be invoked to see this behavior of "holes".

31. pg 1349 line 19. What is meant by "phase coherent oscillations"?

32. Is there a sensitivity to the dj used in the wavelet analysis? If so, this should stated. If dj is somehow intrinsic to the wavelet function, this should be referenced or shown. I

do not know of any support for this idea. It is a tunable parameter as far as I know.

33. Fig. 1. I recommend some other symbol or method to indicate the geometrically significant patches. Stippling or hatching them would help them better stand out. The x makes me think that these patches have been eliminated, rather than highlighted.

34. Fig. 1. The "I_sim" indicated on the figure should be defined in the caption.

35. Fig. 1, Somewhere the actual wavelet spectral values should be shown to get a sense of how the regions passing the pointwise test compare to those not passing.

36. Fig. 1. I don't see in the text where "normalized" has been defined.

37. The red noise equation being used should be shown and how the parameters are fit should be stated. Some discussion of why one is testing against discrete red noise compared to continuous red noise should be given. The spectra are not the same.

38. Why are Fig. 1 and Fig 2 plotted so differently? Also, please double check that the time series in Fig 2a is monthly resolution. It does not appear to be monthly. It looks like it has been smoothed.

39. Why are such short time series considered? The NAO extends back to the early 1800s. Nino3.4 goes back to 1850 in several datasets. I would think that the longest possible record would help in defining the distribution of areas. It would also push out the cone of influence.

40. pg 1350 line 10. No one has shown that "spurious results" are "ubiquitous" in wavelet spectra and neither has this paper. In contrast, Maraun et al. 2007 showed (Appendix C) that the sensitivity of pointwise and areawise tests depends on the signal to noise of the series. As exemplified in the discussion of Fig. 1, this geometric test still has the multiple testing problem.

Technical corrections

1. pg 1338 "would have it did not contain" needs an "if". 2. Fig. 2. I think something is

wrong with the y-axis as given. Nino3.4 should not have so much power at periods of 5 months and the cone-of-influence for monthly data should be at much longer scales than 7 months.

References cited:

D. S. Wilks, 2006: On "Field Significance" and the False Discovery Rate. J. Appl. Meteor. Climatol., 45, 1181–1189. doi: http://dx.doi.org/10.1175/JAM2404.1

Hanna, E., Cropper, T. E., Jones, P. D., Scaife, A. A. and Allan, R. (2014), Recent seasonal asymmetric changes in the NAO (a marked summer decline and increased winter variability) and associated changes in the AO and Greenland Blocking Index. Int. J. Climatol.. doi: 10.1002/joc.4157

_____

Interactive comment on Nonlin. Processes Geophys. Discuss., 1, 1331, 2014.