Summary

The author is grateful for the constructive comments provided by the reviewers. Substantial modifications have been made to the revised manuscript. The overall length of the manuscript has been reduced by nearly five pages. Some subsections containing redundant information have been deleted. Sentences throughout the manuscript have been simplified and numerous paragraphs have been consolidated. The reduction in text volume allows the cumulative areawise to be better emphasized and also enhances the overall readability of the manuscript. Additionally, the formalization of the persistent homology method enhances the readability of Section 3. Provided below are the reviewer comments in bold text and the responses to the comments.

Reviewer 1

General comments:

1) The manuscript is lengthy and poorly constructed. The linkage between sections and subsections is weak. Some materials have been repeated again and again in the manuscript, making it very boring. Simplification to the manuscript is strongly recommended to enhance its readability.

To reduce the length of the manuscript, subsections 4.2 and 4.3 were deleted, as the material in those subsections was repeated later in the manuscript. The length of the manuscript was also reduced my consolidating numerous paragraphs throughout the manuscript and deleting unimportant information. The overall structure of the manuscript has also been modified. The main changes are in Section 2, which is now Section 4 (see comment 20).

2) The main focus of the manuscript is on the cumulative areawise test. The author should put more effort to highlight it. Including a higher proportion of text for introducing this new test may help. The ratio of the summary of existing significant test to the new test is about 1:1 now. The author is advised to increase the proportion for the new test, at least to a ratio of 1:2.

The author agrees that too much emphasis was placed on the existing procedures. To put more emphasis on the cumulative areawise test, the description of the existing procedures has been shortened. For example, text composing paragraphs on pages 1233 through 1235 have been consolidated and rewritten, which resulted in Section 2 (now Section 4) being much shorter.

3) Some materials do not contribute much to the understanding of the test. It gives a feeling that the author tries to insert everything he knows. The author is suggested to make good use of the citation concept. Readers are expected to refer to previous publications for details of some less important information.

Some materials and corresponding figures have been deleted. For example, panel b of Figure 5 has been deleted, as it was shown in Schulte et al., 2015. The reader is now referred to the paper. The text describing Figure 5a has also been removed. Lines 12-15 defining a hole has been removed, as it was found to contribute nothing to the understanding of the cumulative areawise test. The details of the geometric test have also been removed. The reader is now referred to Schulte et al., 2015 for more details of the testing procedure.

4) Sentences are tedious. The author should try to keep the sentences simple but precise.

Throughout the manuscript sentences have been simplified and made more precise.

5) The inclusion of four different climatic oscillation indices as examples does not seem necessary. The author should try demonstrating the techniques using one or two examples. Alternatively, the author may also demonstrate the test using other wavelet techniques, e.g. wavelet coherence (also refer to other comments).

The cumulative areawise test is now demonstrated using only the PDO and Nino 3.4 indices.

6) The Nonlinear Processes in Geophysics is a journal for the publication of researchers on nonlinear processes in geophysical applications. Therefore, the geophysical applications should not be only an example.

Physical interpretation of results has been largely excluded from the manuscript. As suggested by Reviewer 2, the main focus of the manuscript should be on the development of the test, not its application. The examples used in the paper are to provide important benchmarks for further application of the new testing procedure.

Specific comments:

1) Page 1228, lines 10-15: The examples used and their results are not the most important message of the paper. The sentences "The new testing procedure was applied . . . was found in the 2-7 year period band for the Nino 3.4 index" is suggested to be removed or simplified to one sentence, e.g. "The new testing procedure is demonstrated by applying to various climatic oscillation indices".

Lines 10-15 have been removed to allow more emphasis on the cumulative areawise test.

2) Page 1228, line 17: First paragraph of introduction does not seem necessary. It contains too much information about wavelet applications. The main focus of this manuscript should be on the significant test. The author should give one or two sentence brief introduction about wavelet and then connect it to the second paragraph.

The first paragraph of Section 1 has been substantially shortened and now contains three sentences.

3) Page 1229, lines 13-14: The sentence "In geophysical applications, for example, red noise is typically chosen as the null hypothesis." can be removed, as this piece of information appears in section 2.2.

Lines 13-14 on page 1229 have been removed.

4) Page 1229, line 19-21: The sentence "Despite the insights gained . . . simply due to multiple testing" can be reformed to "Despite the insights gained from the statistical procedure, Maraun and Kurths (2004) showed that it can lead to many spurious results due to multiple testing."

The suggested change has been adopted.

5) Page 1229, line 23-27: The summary on the areawise test developed by Maraun etal. (2007) can be more precise. The author may refer to the abstract of Manraun et al. (2007).

Lines 23-27 on page 1229 have been made more precise.

6) Page 1229, line 24-27: Please remove the sentence "though dramatically reduce the number of spurious results".

Lines 24-27 on page 1229 have been removed.

7) Page 1230, lines 3-10: This paragraph can be simplified and merge with the precise paragraph. Emphasizing the difference between the areawise test and geometric test should be good enough, as areawise test has just been introduced. The sentence "Like the areawise test, . . . allows patches at different periods to be compared simultaneously" does not seem necessary.

The paragraph corresponding to lines 3-10 has been merged with the previous paragraph. Some text in lines 3-10 has been deleted has as well.

8) Page 1230, line 17: Could real be a better word than present in "In the present case"?

In an effort to shorten the paragraph corresponding to lines 11-28 on page 1230, Line 17 and other lines have been deleted.

9) Page 1230, lines 11-28: This paragraph is supposed to state clearly the objective of the manuscript. However, it is poorly written and the objective is ambiguous. Putting the last few sentences "This test has the important feature that the significance of the wavelet power . . . a consistent statistical construction" at the end of this paragraph does not seem appropriate.

The paragraph corresponding to lines 11-28 on page 1230 has been completely rewritten. The objectives of the paper are now clearly stated at the end of the paragraph. Furthermore, sentence structure has been simplified. The last few sentences "This test has the important feature that the significance of the wavelet power . . . a consistent statistical construction" have been deleted.

10) Page 1231 lines 2-5: The author may consider deleting "including the sensitivity of the geometric test. . . to the development of the new testing procedure"

Lines 2-5 on page 1231 have been deleted.

11) Page 1231 line 11: Why is wavelet analysis under Section 2? It is not a significant test.

The introduction to wavelet analysis has been moved to its own section (now Section 3).

12) Page 1231 lines 12-18: Is there any special reason to include a long paragraph introducing Morlet, Paul and Dog wavelets? It is understood that the cumulative areawise test is demonstrated using different wavelets in section 4, but their results do not seem to be different. The author is advised to pick one for demonstration purpose.

The cumulative areawise test is now applied using only the Morlet wavelet. A paragraph in the conclusion/discussion section has been added that briefly summarizes the results for the other analyzing wavelets. Text corresponding to the discussion of results for the Dog and Paul wavelets has been deleted throughout the manuscript.

13) Page 1231 lines 12-18: The author actually may consider removing the introduction of wavelet analysis. The readers should already have some basic knowledge about wavelet analysis before reading a paper related to its significant test.

The introduction of wavelet analysis is still present in the revised manuscript (now Section 3). The brief introduction will provide a quick review for those familiar with wavelet analysis and useful references for those seeking a better understanding of wavelet analysis. Nevertheless, the introduction section has been substantially shorted.

14) Page 1233 lines 8-9: The sentence "In spectral analysis, . . . against a noise background" can be removed.

Lines 8-9 on page 1233 have been deleted.

15) Page 1234 lines 3-23: This paragraph basically introduces the example and data used. The author should include a section introducing the data used before section 2. Including all these in section 2 makes the manuscript very messy. Please refer to Grinsted et al. (2004).

A subsection describing the data has been added and is now Section 2 of the revised manuscript.

16) Page 1234 line 24: To simplify the manuscript and give it a better structure, the author should consider introduce all existing significant tests first and then demonstrate them all together using one or two example. Some comparisons can be easily made as well.

The existing testing procedures are now introduced first in Section 2 (now Section 4). The application of the tests is now the final subsection of Section 4 in the revised manuscript.

17) Page 1234 line 24: If there is no special reason to include four examples, the author should consider use one or two examples to demonstrate all the significant tests. Actually, using idealized examples may also be a possible way of demonstration.

The techniques is now demonstrated using two examples, one for the PDO index and one for the Nino 3.4 index.

18) Page 1235: Why is areawise test by Maraun et al. (2007) left out in section 2? It is a little bit weird, as the author did introduce it in introduction.

The author agrees. A new subsection (now subsection 4.2) has been included in Section 2 (now Section 4) that describes briefly the areawise test. The geometric test is also mentioned briefly to enhance readability and to decrease the length of the manuscript (see also comment 3).

19) Page 1237 line 4: What is the purpose of including a sensitivity test for geometric test corresponding to different pointwise significant level? Are these results previously been

documented? If not, it shouldn't be put in Section 2, which is supposed to be a summary of existing significant tests.

The results of the sensitivity test for the geometric test are presented to quantify the binary problem of the test. The results have been moved to a new section (now Section 5). The inclusion of this section is motivated by how the quantification of the binary problem of the geometric test has not been documented.

20) Pages 1231-1238: The author may consider to reconstruct Section 2 by first give a summary on pointwise test, and then introduce areawise test, with emphasis on its improvement to pointwise test. Further, geometric test may be introduced as a simplified version of areawise test. And then conclude the section with special stress on the binary decision problem suffered by areawise and geometric test and demonstrations of different significant test.

Section 2 (now Section 4) has been reconstructed with the suggested structure as described in the response to comment 16.

21) Page 1241 lines 17-18: It is confusing to refer cumulative areawise test as areawise test, as readers may mess it up with the areawise test developed by Maraun et al. (2007).

The areawise test is now referred to as the cumulative areawise test throughout the manuscript.

Other comments:

In author's previous paper (Schulte et al. 2015), it was mentioned that the geometric test has an advantage of applying to other wavelet applications, e.g. wavelet coherence (Grinsted et al. 2012), partial wavelet coherence and multiple wavelet coherence (Ng and Chan 2012). Is the new cumulative areawise test also applicable to these wavelet applications? If yes, it would be good to include this piece of information in this manuscript as well. Also, the authors may consider demonstrating the cumulative areawise using wavelet coherence, which should be of great interest to many readers.

A paragraph has been added in summary/discussion section (now Section 10) that describes the application of the procedure to wavelet coherence, partial coherence, and multiple coherence.

Reviewer 2

First of all, the manuscript has two distinct goals, one on the methodology and the other on its application. This distracts the audience from reading the manuscript.

Text describing the physical interpretation of results has been shorted to put less emphasis on the applications. However, the author feels that geophysical examples in a journal aimed at geophysical research is important because the inclusion sets important benchmarks for further application of the method.

The second critical point is that the manuscript has an imbalance in section structure.

Substantial reconstructing of sections has been made to balance the text volume among the sections. Please see the response to the comment below.

Section 1, 2, and 3 have understandable text volume, but section 4 is very long (with 6 subsections). Section 5 and 6, in contract, are very short. I cannot read section 4 so easily (i.e., without repeating from the beginning of the section) for its tiring construction.

The length of Section 4 has been dramatically reduced. Subsections 4.2, 4.4 and 4.5 have been deleted. Text from the deleted subsections have been consolidated and shortened. Subsection 4.6 has been made into a new section (Section 8) and split into two subsections. The new Section 8 describes the statistical properties of the cumulative areawise test and is thus separate from the development section, which now Section 7. The result of the text consolidation and the splitting of Section 4 in the original manuscript is shorter sections with text volume more balanced with the other sections in the revised manuscript.

The manuscript is organized as follows. Here are my comments on each section.

Section 1 addresses the wavelet estimator among other spectral estimators, and discusses the problem or importance of the test for statistical significance in the wavelet spectrum. The goal of the manuscript is not clearly set and it is difficult to follow the strategy of the method development in the manuscript.

The objectives of the manuscript are now clearly stated in Section 1. The clearly stated objectives will help the reader follow the overall strategy of the method development.

Section 2 reviews the significance tests such as pointwise and geometric tests. This section essentially overlaps with the author's recent paper (NPG, 22, 139-156, 2015). I propose to delete this section. The contents are already published by the author. Also, Figure 1 distracts the audience from concentrating on the new method with homology.

The review of the existing tests has been substantially shortened to reduce the amount of overlap with worked published in Schulte et al., 2015. A complete removal of the section seems inappropriate because knowledge of the existing procedures is necessary to understand the development of the cumulative areawise test.

Section 3 finally (on page 12) presents the method of the homology with an application to red-colored noise. Nevertheless, the exact or quantitative definition is not given, so it is unclear to the readers how the algorithm is constructed to evaluate the persistent topology. This section needs a lot more explanations with equations and definitions. As the concept of the homology is not quantitatively defined, I do not follow the homology method.

Some equations and formal definitions are now presented in Section 3 (now Section 6). However, a full mathematical treatment of persistent homology is beyond the scope of the paper. A full treatment would require the introduction of concepts from general topology, group theory, and algebraic topology, which would substantially increase the length of the manuscript. Nevertheless, the formal definitions added are adequate for a basic understanding of persistent homology applied in this paper. The reader is referred to cited works for details of persistent homology because Reviewer 1 found the paper too long and encouraged the use of citations

throughout the manuscript. It is also noted that Figure 4 has been changed to reflect the changes made throughout Section 6 in the revised manuscript.

Section 4 is hard to read. It is too long (13 pages and 6 subsections). The subsections are: 4.1 Geometric pathways, 4.2 Pointwise significance level selection: maximization method, 4.3 Application to ideal pathways, 4.4 The null distribution, 4.5 Computational remarks, 4.6 Comparison with geometric test. This structure is not understandable, and I do not see what the author wants to say in this section.

Section 4 (now Section 7) has been substantially shortened. Subsections 4.4 and 4.5 have been removed entirely. Moreover, subsection 4.6 has made into its own section (now Section 8). The reason for making subsection 4.6 into a new section is that the subsection examines the properties of the cumulative areawise test and therefore does not belong in the development section.

Section 5 (Climate applications) is a small section with only 1.5 pages (35 lines), and presents an application of the developed method. The text volume is too small and I do not see any necessity or reason to add this section into the manuscript. Delete the section.

The author feels that this section is important because the intended audience of the paper is geophysical researchers. The application of the testing procedure provides important benchmarks for further applications of the method. The author could not find a way of increasing the text volume of the Section 5 (now Section 9) without compromising the flow of the paper.

Section 6 (Conclusions) is merely summarizing the manuscript and does not discuss the method in depth. For examples, what is the limit of the method? Also, I do not appreciate to state that a Matlab software is available without presenting the algorithm in this manuscript. Delete the sentence.

Section 6 (now Section 10) has been expanded by inserting a paragraph describing the limitations of the test. Also, included is a paragraph describing the application of the procedure to wavelet coherence and global wavelet spectra and discussion of its generalizations to higher dimensions.

The algorithm was presented throughout the manuscript and therefore stating the availability of Matlab software seems appropriate.

References:

Grinsted, A., Moore, J. C., and Jevrejeva, S.: Application of the cross wavelet transform and wavelet coherence to geophys- ical time series, Nonlin. Processes Geophys., 11, 561–566, doi:10.5194/npg-11-561-2004, 2004.

Maraun, D., Kurths, J., and Holschneider, M.: Nonstationary Gaussian Processes in Wavelet Domain: Synthesis, Estima- tion, and Significance Testing, Phys. Rev. E, 75, 016707, doi:10.1103/PhysRevE.75.016707, 2007.

Ng, E. K.W. and Chan, J. C. L.: Geophysical Applications of PartialWavelet Coherence and Multiple Wavelet Coherence, J. Atmos. Oceanic Technol., 29, 1845–1853, 2012.

Schulte, J. A., Duffy, C., and Najjar, R. G.: Geometric and topological approaches to significance testing in wavelet analysis, Nonlin. Processes Geophys., 22, 139–156, doi:10.5194/npg-22- 139-2015, 2015.

Cumulative Areawise Testing in Wavelet Analysis and its Application to Geophysical Time Series

Corresponding Author: Justin A. Schulte

Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

Abstract

Statistical significance testing in wavelet analysis was improved through the development of a cumulative areawise test. The test was developed to eliminate the selection of two significance levels that an existing geometric test requires for implementation. The selection of two significance levels was found to make the test sensitive to the chosen pointwise significance level, which may preclude further scientific investigation. A set of experiments determined that the cumulative areawise test has greater statistical power than the geometric test in most cases, especially when the signal-to-noise ratio is high. The number of false positives identified by the tests was found to be similar if the respective significance levels were set to 0.05. The new testing procedure was applied to the time series of the Atlantic Multi-decadal Oscillation (AMO), North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO), and Niño 3.4 index. The testing procedure determined that the NAO, PDO, and AMO are consistent with red-noise processes, whereas significant power was found in the 2-7 year period band for the Niño 3.4 index.

1. Introduction

In many research fields, it is of interest to understand the behavior of time series in order to achieve a deeper understanding of physical mechanisms or relationships. Such a task can be formidable given that time series are composed of oscillations, non-stationarities, and noise. Fortunately, many tools have been developed to extract information from time series, including singular spectrum analysis (Vautard et al., 1992), Fourier analysis (Jenkins and Watts, 1968), and wavelet analysis (Meyers et al., 1993; Torrence and Compo, 1998). ~~The goal of each of these tools is to assess whether deterministic features are embedded in a time series. Fourier analysis, as an example, is a method by which a time series is decomposed into frequency components so that embedded oscillations can be detected. However, the underlying assumption in Fourier analysis is that time series are stationary. The limitation can be circumvented by using a windowed Fourier analysis but with the caveat that the window width is fixed, which can lead to poor resolution at low frequencies (Lau and Weng, 1995). Wavelet analysis, relaxing the assumption of stationarity, offers an alternative method to Fourier analysis in which the window width is no longer fixed, minimizing aliasing (Meyers et al., 1993; Torrence and Compo, 1998).~~ Wavelet analysis has been demonstrated to be useful in the understanding of the North Atlantic Oscillation (NAO; Higuchi et al., 2003; Olsen et al., 2012), applications to oceanographic problems (Meyers et al., 1993; Lee

and Lwiza, 2008; Whitney, 2010; Wilson et al., 2014), assessments of historical hydroclimate variability (Labat, 2004; Labat, 2008), and many other geophysical applications (Grinsted et al., 2004; Velasco and Mendoza, 2008).

~~When using any time series extraction procedure it is important to assess the statistical significance of the computed test statistic against some null hypothesis. In geophysical applications, for example, red noise is typically chosen as the null hypothesis. Torrence and Compo (1998) were the first to apply wavelet analysis in a statistical framework using pointwise significance testing, allowing deterministic features to be distinguished from stochastic features. In a pointwise significance test, one tests each estimated wavelet power coefficient against a stationary theoretical red-noise background spectrum. Despite the insights gained from the statistical procedure, it has many deficiencies, as noted by Maruan and Kurths (2004), who showed that it can lead to many spurious results simply due to multiple testing. Addressing the multiple-testing problem, Maraun et al. (2007) developed an areawise test that assesses the significance of so-called pointwise significance patches, contiguous regions of pointwise significance in a wavelet power spectrum. The areawise test, though dramatically reducing the number of spurious results, is computationally inefficient, involving a root-finding algorithm to estimate a critical area of the reproducing kernel corresponding to the desired significance level of the test. Furthermore, the critical area needs to be computed for different analyzing wavelets and for their associated parameters, such as the central frequency for the Morlet wavelet and the order in the case of the Paul wavelet.~~

~~A simpler procedure for addressing multiple testing problems is the geometric test developed by Schulte et al. (2015). Like the areawise test, the test statistic for the procedure is based on patch area, or more specifically, the normalized area of the patch, which allows patches at different periods to be compared simultaneously. The calculation of the critical level for the geometric test is much simpler than that for the areawise test, involving the computation of the normalized area for a large ensemble of patches under a null hypothesis that results in a null distribution from which the desired critical level can be calculated.~~

When using any time series extraction procedure it is important to assess the significance of the computed test statistic against some null hypothesis. In geophysical applications, for example, red noise is typically chosen as the null hypothesis. Torrence and Compo (1998) were the first to apply wavelet analysis in a statistical framework using pointwise significance testing, allowing deterministic features to be distinguished from stochastic features. In a pointwise significance test, one tests each estimated wavelet power coefficient against a stationary theoretical red-noise background spectrum. Despite the insights gained from the statistical procedure, Maraun and Kurths (2004) showed that it can lead to many spurious results simply due to multiple testing. Addressing the multiple-testing problem, Maraun et al. (2007) developed an areawise test that decides whether a pointwise significant result is a deterministic feature distinguishable from typical stochastic fluctuations by using basic properties of the continuous wavelet transform. A simpler procedure for addressing multiple testing problems is the geometric test developed by Schulte et al. (2015). The calculation of the critical level for the geometric test is much simpler

than that for the areawise test because it is calculated using a basic Monte Carlo procedure that generates a null distribution of the test statistic.

Both the geometric and areawise tests~~, however,~~ suffer from a binary decision because~~:~~ one must choose both a pointwise significance level together with an ~~and~~ areawise or geometric significance level. The problem with such a statistical construction is that the outcomes of the testing procedure may depend on the chosen pointwise significance level. For an ideal test, there is a single significance level that is chosen and the results of the testing procedure depend only on that significance level ~~so that a test statistic, for example, that is 1% significant is guaranteed to be 5% significant. In the present case, however, there is no such guarantee: a 1% geometrically significant patch at one pointwise significance level may not be 5% significant at another pointwise significance level.~~ In such cases, the statistical significance of patches is ambiguous and may preclude further scientific investigation. ~~This sensitivity problem underscores the need to develop a computationally efficient testing procedure free of binary decisions. The approach taken here will consider the areas of patches over all pointwise significance levels, and hence the method is called the *cumulative* areawise test.~~

~~This test has the important feature that the significance of the wavelet power coefficients is a monotonic increasing function of the pointwise significance level. In other words, a wavelet power coefficient, for example, that is 1% significant under the new procedure will be guaranteed to be 5% significant, a consistent statistical construction.~~

Thus, the objectives of this paper are the following:

1) Quantify how the binary decision of the geometric test can lead to ambiguity in interpreting results;

2) Understand and quantify the evolution of pointwise significant regions under a changing pointwise significance level using persistent topology;

3) Design a statistical test whose application only requires the choice of a single significance level.

Motivated by Objectives 1 and 2, the approach to achieve Objective 3 will be to consider the areas of pointwise significant regions over all pointwise significance levels, and hence the method will be called the *cumulative* areawise test.

The paper is organized as follows: in Section 2, a brief description of wavelet analysis is provided together with a discussion of existing statistical testing procedures, including the sensitivity of the geometric test to the chosen pointwise significance, motivating the construction of the cumulative areawise test. Before proceeding to the development of the new testing procedure, the topological properties of red noise are analyzed in Section 3. In Section 4, the cumulative areawise test is developed and is followed by a comparison of the test in terms of statistical power to the existing geometric test. Applications of the test to prominent climate indices are presented in Section 5, followed by concluding remarks in Section 6.

2. Data

The Nino 3.4 index data from 1900-2014 was obtained from the Climate Prediction Center. The Niño 3.4 index is an oceanic metric for quantifying the strength of the El-Nino/Southern Oscillation (ENSO) and is defined as SST anomalies in the Equatorial Pacific in the region bounded by 120°W-170°W and 5°S-5°N (Trenberth, 1998). The Pacific Decadal Oscillation Index data was obtained from University of Washington and index describes detrended sea surface temperature (SST) variability in the North Pacific poleward of 20°N latitude (Mantua and Hare, 2002)

~~**2. Existing Wavelet Analysis Significance Tests**~~

3 ~~2.1~~ Wavelet Analysis

The wavelet transform of a time series is defined as the convolution of the time series with a wavelet function ψ_0. The wavelet transform of a time series x_n (n = 1, ... , N) with a wavelet function ψ_0 is given by

$$W_n(s) = \sqrt{\frac{\delta t}{s}} \sum_{n'=1}^{N} x_{n'} \psi_0[(n'-n)\frac{\delta t}{s}], \quad (1)$$

where s is the wavelet scale, δt is a time step determined by the data, and N is the length of the time series. There are many kinds of wavelets, but perhaps the most common are the Morlet, Paul, and Dog wavelets. This paper will focus on the ~~For geophysical applications,~~ the Morlet wavelet, which ~~is often used and~~ is given by

$$\psi_0(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-\frac{1}{2}\eta^2}, \quad (2)$$

where ω_0 is the dimensionless frequency, $\eta = s \cdot t$, t is time, and the wavelet scale is related to the Fourier period by $\lambda = 1.03s$ if ω_0 = 6. This particular wavelet balances both frequency and time-localizations. Thoughout the paper, ω_0 = 6. ~~The Paul wavelet, which is also a complex wavelet, is more localized in time, less localized in frequency space, and is given by~~

~~$$\psi_0(\eta) = \frac{(2i)^m m!}{\sqrt{\pi(2m)!}} (1 - i\eta)^{-(m-1)}, \quad (3)$$~~

~~where m is the order of the Paul wavelet, which controls the localization properties of the analyzing wavelet. In this case, the Fourier period is related to the wavelet scale by the following equation:~~

~~$$\lambda = \frac{4\pi s}{2m+1}. \quad (4)$$~~

~~If even more time-localization is desired, one can use the Dog wavelet, a real wavelet given by~~

~~$$\psi_0(\eta) = \frac{(-1)^{m+1}}{\sqrt{\Gamma\left(m+\frac{1}{2}\right)}} \frac{d^m}{d\eta^m} e^{-\eta^2/2}, \quad (5)$$~~

~~where m represents the order of the derivative and Γ is the gamma function. For the Dog wavelet, the Fourier period is related to the wavelet scale by the equation~~

$\lambda = \frac{2\pi s}{\sqrt{m+\frac{1}{2}}}$. (6)

In this paper, the Paul wavelet is used with $m = 4$ and the Dog wavelet is used with $m = 2$.

The wavelet power is given by

$$|W_n(s)|^2 \quad (7)$$

and represents the wavelet power spectrum of the time series. Inherent in the wavelet transform are edge effects due to the finite time series. In particular, in a wavelet power spectrum there exists a region called the cone of influence, which is defined as the *e*-folding time of the autocorrelation for wavelet power at each scale. The *e*-folding time is defined as the point at which the wavelet power for a discontinuity at the edge drops by a factor of e^{-2} (Torrence and Compo, 1998).

3. Existing Significance Tests

42.12 Pointwise Significance Test

In spectral analysis, it is important to assess the statistical significance of spectral power against a noise background. In geophysical applications of wavelet analysis, one often tests each individual wavelet power coefficient against a stationary red-noise background to determine their statistical significance (Torrence and Compo, 1998). For a first-order autoregressive (Markov) process

$$x_n = \rho x_{n-1} + w_n, \quad (8)$$

where ρ is the lag-1 autocorrelation coefficient and w_n is Gaussian white noise with $x_0 = 0$, the normalized theoretical stationary red-noise power spectrum is given by

$$p_f = \frac{1-\rho^2}{1+\rho^2-2\rho\cos(2\pi f/N)}, \quad (9)$$

where $f = 0, \ldots, N/2$ is the frequency index (Gilman et al., 1963). To obtain, for example, the 5% pointwise significance level ($\alpha = 0.05$), one must multiply Eq. (9) by the 95th percentile of a chi-square distribution with two degrees of freedom and divide the result by 2 to remove the degree-of-freedom factor (Torrence and Compo, 1998). The result of the so-called pointwise testing procedure is a subset of wavelet power coefficients whose values exceed the specified background noise spectrum. Recall from Section 1 that significant wavelet power coefficients often occur in clusters or contiguous regions of pointwise significance called pointwise significance patches (referred to as patches, hereafter).

Consider the time series of the Atlantic Multi-decadal Oscillation (AMO), North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO), and Niño 3.4 indices shown in Figure 1. The PDO index describes detrended sea surface temperature (SST) variability in the North Pacific poleward of 20°N latitude (Mantua and Hare, 2002) and the AMO index captures the detrended SST variability in the Atlantic Ocean basin (Kerr, 2002). As shown in Figures 1a and 1b, the PDO and AMO indices exhibited multi-decadal variability, with periods of 20-60 years, from, respectively, 1856 to 2014 and 1900 to 2014. The reason for the low-frequency variability of the

~~PDO is subject to debate. Some studies suggest it is the reddened response to white noise atmospheric forcing, whereas other studies hypothesize that it is also the integrated response of the El Niño/Southern Oscillation (ENSO) signal (Newman et al., 2004). The NAO index, an atmospheric index, quantifies the difference in sea level pressure of the Icelandic Low and the Azores High and is related to the strength and position of the jet stream across the North Atlantic (Hurrell et al., 2003). As shown in Figure 1b, the NAO mainly operated on time scales of months and season, and the raw time series is rather noisy. The Niño 3.4 index is an oceanic metric for quantifying the strength of ENSO and is defined as SST anomalies in the Equatorial Pacific in the region bounded by 120°W 170°W and 5°S 5°N (Trenberth, 1998). The Niño 3.4 index time series exhibited variability on an array of time scales, especially in the 2 7 year period band. Various physical interpretations for the 2 7 year oscillation have been proposed, including the Unified Oscillator, Delayed Oscillator, and the Recharge Oscillator (Wang et al., 2004).~~

~~Shown in Figure 2 are the wavelet power spectra of the AMO, NAO, Niño 3.4, and PDO indices. The wavelet power spectrum of the AMO detected enhanced variance at a period of 512 months, as indicated by the thin contour that encloses a region of 5% pointwise significance. All of the other patches are located at periods less than 32 months. The wavelet power spectrum of the NAO indicated that the NAO exhibited enhanced variability on an array of time scales. For example, there is a patch located at a period of 64 months and 1910. Like the wavelet power spectrum of the AMO, numerous patches were also found at periods less than 32 months. Large regions of enhanced variance were found in the wavelet power spectrum of the Niño 3.4 index. The largest of these regions was located in the time period 1950 2014 and the period band 16 32 months, consistent with how the ENSO varies with periods of 2 7 years. In the same patch, there are holes as described by Schulte et al. (2015) that may indicate the presence of nonlinearities. Holes are defined formally as follows. For the closed unit interval $I = [0\ 1]$, let $f: I \rightarrow P$ be a continuous closed path in a significance patch P. A patch is said to contain a hole if there exists a path that cannot be continuously deformed into a point, where the feature obstructing such a deformation is the hole. Two patches in the same period band were also identified from 1870 to 1890. For the wavelet power of the PDO index, a large patch centered at a period of 512 months extending from 1910 to 2013 was detected. Most of the patches, however, were located at periods less than 8 months, time scales not typically associated with the PDO.~~

~~42~~.~~23~~ Areawise and Geometric Significance Tests

3.2 Areawise significance testing

The areawise test developed by Maraun et al. (2007) takes of advantage of how correlations between adjacent wavelet coefficients arising from the reproducing kernel produce continuous regions of pointwise significance (patches) that resemble the reproducing kernel. The reproducing kernel for a given analyzing wavelet represents the time-scale uncertainty, which is related to the scale and time localization properties of the analyzing wavelet. For significance patches generated from random fluctuations, the typical patch area is the area of the reproducing kernel. The areawise test assesses the significance of patches based on their area, where patches with areas greater than that of the reproducing kernel have more statistical significance. The estimation of the critical level of the test involves a root-find algorithm that is computational inefficient. To remedy the

Formatted: Indent: First line: 0.5"

computational drawback, Schulte et al. (2015) developed a geometric test that make uses of a normalized area. The normalized area allows patches at different scales to be compared simultaneously. The estimation of critical level of test is achieved simply by Monte Carlo methods by generating a large ensemble of patches under a null hypothesis to create a null distribution from which the critical level of the test can be obtained.

~~To determine if the results from the pointwise test are artifacts of multiple testing, a geometric test was applied to the patches located in the wavelet power spectra (Schulte et al., 2015). The test statistic for the geometric test is given by a normalized area~~

$$~~A_R = \frac{A}{A_R},~~ \quad ~~(10)$$

~~where A is the area of the patch and A_R is the area of the reproducing kernel dilated and translated according to the centroid of the patch. Regarding patches as polygons with vertices (x_k, y_k) with $k = 1, \dots, m$ 1, the area of the patch is determined by a simple formula given by~~

$$~~A = \frac{1}{2}\left|\sum_{k=0}^{m-1}(x_k\, y_{k+1} - x_{k+1} y_k)\right|,~~ \quad ~~(11)$$

~~where $y_0 = y_m$ and $x_0 = x_m$. Simiarly, the centroids of the polygons are given by~~

$$~~C_t = \frac{1}{6A}\sum_{k=0}^{m-1}(x_k + y_{k+1})\,(x_k y_{k+1} - x_{k+1} y_k)~~ \quad ~~(12)$$

~~and~~

$$~~C_s = \frac{1}{6A}\sum_{k=0}^{m-1}(y_k + x_{k+1})\,(x_k y_{k+1} - x_{k+1} y_k),~~ \quad ~~(13)$$

~~where C_t and C_s are the time and scale coordinates, respectively, of the centroid (Schulte et al., 2015).~~

~~To determine the critical level of test, a large ensemble of patches under a noise model is generated and A_R is computed for each patch, resulting in a null distribution from which the desired critical level of the test can be obtained.~~ **4.3 Application of Existing Significance Tests**

Shown in Figure 2 are the wavelet power spectra of Niño 3.4, and PDO indices. Large regions of enhanced variance were found in the wavelet power spectrum of the Niño 3.4 index. The largest of these regions was located in the time period 1950-2014 and the period band 16-32 months, consistent with how the ENSO varies with periods of 2-7 years. In the same patch, there are holes as described by Schulte et al. (2015) that may indicate the presence of nonlinearities. Holes are defined formally as follows. For the closed unit interval $I = [0\ 1]$, let $f: I \to P$ be a continuous closed path in a significance patch P. A patch is said to contain a hole if there exists a path that cannot be continuously deformed into a point, where the feature obstructing such a deformation is the hole. Two patches in the same period band were also identified from 1870 to 1890. For the wavelet power of the PDO index, a large patch centered at a period of 512 months extending from 1910 to 2013 was detected. Most of the patches, however, were located at periods less than 8 months, time scales not typically associated with the PDO.

The application of the geometric test to the wavelet power spectra of the ~~AMO, NAO,~~ PDO~~,~~ and Niño 3.4 time series is also shown in Figure 2. ~~For the AMO index 5% geometrically significant patches were identified in the period band 2-16 months prior to 1880, after which no patches were identified as geometrically significant. Note that the large patch centered around the period of 512 months was not found to be geometrically significant at the 5% level, suggesting that the multi-decadal variability is stochastic. A few geometrically significant patches were identified in the wavelet power spectrum of the NAO: one centered at a period of 64 months and 1910, a second one centered at a period of 8 months and 1910, and several others centered at a period of 4 months throughout the record length.~~ Many geometrically significant patches were identified in the wavelet power spectrum of the Niño 3.4 index. For example, a rather large geometrically significant patch is located in the 16-64 month period band from 1950 to 2014. There are also many geometrically significant patches located in the 2-8 month period band from 1900 and 1950. Numerous geometrically significant patches were identified in the wavelet power spectrum for the PDO index, all of which were located in the 2-8 month period band. There was also a large patch centered on a period of 512 months, but the patch was not found to be geometrically significant~~, suggesting that, like the AMO,~~ the multi-decadal variability arose from stochastic processes.

5.~~2.4~~ Sensitivity of the Geometric Test to the Chosen Pointwise Significance Level

To show that the geometric test is sensitive to the pointwise significance level chosen, it will be useful to compute the quantity

$$r = \frac{N_{\alpha_1,\alpha_2}}{N_{\alpha_1}}, \tag{14}$$

where N_{α_1,α_2} is the number of geometrically significant patches at pointwise significance level α_1 that are also geometrically significant at pointwise significance level α_2 and N_{α_1} is the number of patches at α_1 that are geometrically significant, where the geometric significance level for all cases is fixed α_{geo}. In the ideal situation, r = 1, indicating that geometrically significant patches never lose their geometric significance as the pointwise significance level is increased. This case, however, is optimistic, as the calculation of geometric significance is rather stochastic. To demonstrate the stochastic nature of the geometric test, r was computed for 1000 wavelet power spectra of red-noise processes with lengths 1000 and lag-1 autocorrelation coefficients equal to 0.5 under four different scenarios, the first of which (Scenario 1) is the case in which $\alpha_1 = 0.1$, $\alpha_2 = 0.05$, and $\alpha_{geo} = 0.05$ (Figure 3a). With the mean of r (denoted by $\bar{r}$ hereafter) being 0.3, it can hardly be expected that a geometrically significant patch at $\alpha_1 = 0.1$ to remain significant when the pointwise significance level is changed to $\alpha_2 = 0.05$, at least in the case of red-noise processes. Scenario 2, shown in Figure 3b, is the same as Scenario 1 except that $\alpha_{geo} = 0.01$. In this case, $\bar{r} = 0.15$, suggesting that the geometric test is even more sensitive to the chosen pointwise significance level for smaller α_{geo}. Also note that, unlike the distribution shown in Figure 3a, the distribution is skewed, favoring lower values and supporting the idea that the geometric test is more sensitive to the chosen pointwise significance level for $\alpha_{geo} = 0.01$.

In Scenario 3, $\alpha_1 = 0.05$ and $\alpha_2 = 0.01$, with $\alpha_{geo} = 0.01$. The distribution shown in Figure 3c is even more skewed than that corresponding to Scenario 2, with $\bar{r} = 0.05$. Also note that in many cases $r = 0$, indicating that there are patches that are not geometrically significant for both $\alpha_1 = 0.05$ and $\alpha_2 = 0.01$. The reason is that some patches existed at $\alpha_1 = 0.05$ but did not exist at $\alpha_2 = 0.01$ so that their normalized areas are zero.

Scenario 4 was similar to Scenario 3 except that $\alpha_{geo} = 0.05$. Although Scenarios 3 and 4 used the same pointwise significance levels, the results differ, with $\bar{r} = 0.22$, suggesting that the geometric test is less sensitive for larger α_{geo}. The results are similar to that of Scenarios 1 and 2, where increasing the pointwise significance level increased the sensitivity of the geometric test to the chosen pointwise significance level.

6.~~3~~. Persistent Topology

6~~3~~.1 Persistent Homology

~~Before developing the cumulative areawise test, it will be necessary to understand the topology of features found in a typical wavelet power spectrum. It will be especially important to understand how the features evolve as the pointwise significance level is increased or decreased. Such information can be obtained using persistent homology, a tool in applied algebraic topology (Edelsbrunner, 2004). Persistent homology will provide a formal setting for calculating the lifetimes of patches and holes, where lifetimes describe when features first appear and when they disappear. For example, a patch appearing at the 5% pointwise significance level and vanishing at the 1% pointwise significance level would have a lifetime of 4. In this paper, the intuitive foundation of persistent homology will be described rather than giving a formal mathematical exposition.~~

~~Suppose that A is a patch at the pointwise significance level $\alpha = \alpha_1$ as shown in Figure 4a. One can increase the size of the patch A by increasing α to α_2, which lowers the threshold for significance, resulting in the new geometric realization A' of A shown in Figure 4a. The evolution of the patch can be monitored using a barcode (Ghrist, 2008), which is a collection of horizontal lines representing the birth and death of features. Following the convention of persistent homology, the y-axes of barcodes will be denoted by H_0 for patches. In algebraic topology, H_0 are called homology groups and measure the path-connectedness (Appendix A) of sets. The patch A, being created at α_1, results in the line segment beginning at α_1 in Figure 4c. Furthermore, the patch neither vanishes nor merges with another patch at α_2 so that the horizontal line continues to α_2. Note that a new patch B is created at α_2 because the pointwise significance test is less stringent. The creation of the patch results in a new line starting from α_2. A more complicated situation occurs at α_3, where the patches A' and B merge and result in a new patch C. In this case, it is unclear if C is a geometric realization of A' or B. It will therefore be necessary to use the so-called Elder rule from persistent homology (Edelsbrunner and Harer, 2009). According to this rule, the oldest patch will continue to live and the younger patch will die entering the merger point. In the present case, A' is the older patch because it first appears as A at α_1 and B is the younger patch, being created after A at α_2. Therefore, according to the Elder rule, the horizontal line corresponding to A in the barcode continues to α_4, but the line corresponding to B terminates at α_3, as it dies~~

~~entering α_3. Also note the creation of a new patch D at α_3 and the corresponding line segment in the barcode. Another merger occurs at α_4 and the Elder rule determines that the line segment for D ends and the horizontal line for A continues, where the arrow indicates that A never dies.~~

~~Persistent indices of patches can also be obtained, which will be defined as the difference between the pointwise significance level for which the features die entering and the level at which the features were born. For patches that never die, their persistent indices, by convention, will be set to infinity.~~

Before developing the cumulative areawise test, it will be necessary to understand the topology of features found in a typical wavelet power spectrum. It will be especially important to understand how the features evolve as the pointwise significance level is increased or decreased. Such information can be obtained using persistent homology, a tool in applied algebraic topology (Edelsbrunner, 2004). Persistent homology will provide a formal setting for quantifying the evolution of patches. Some formal definitions will be given below, but the reader is referred to Edelsbrunner (2004) for a more detailed description of persistent homology.

A pointwise significance patch will be defined formally as follows. A path in a set X is defined as a continuous function $f: [0\ 1] \rightarrow X$ (Lipschutz, 1965). A set X is said to be path-connected if any two points x and y in X can be joined by a path. The path-component of a set X is the maximal path-connected subset of a set. Intuitively, one can think of a path-component as an isolated piece of the set. In the present setting, patches are path-connected components because they represent isolated pieces of the set consisting of all wavelet power coefficients that are pointwise significant.

Denote by P the set of all pointwise wavelet power coefficients that are significant at the α level. Then two points $x, y \in P$ will be called homologous (written $x \sim y$) at α if there exists a path $f: [0\ 1] \rightarrow P$ such that $f(0) = x$ and $f(1) = y$ (Figure 4a). The definition implies two points are homologous when they can be joined by a continuous path. The set of all points that are homologous to x form an equivalence class called a homology class that is denoted by

$$[x] = \{y \in P: y \sim x\}. \quad (6)$$

The set of all homology classes of P will be denoted by $H_0(P)$, where $H_0(P)$ is called the 0-dimensional homology group (Hatcher, 2002). Each member of a homologous class is homologous but no two points from distinct homology classes are homologous. The homology classes form a partition of P into path-connected components and therefore patches at a given pointwise significance level can be regarded formally as homology classes. Mathematically, we have the quotient

$$H_0(P) = {}^{P}/_{\sim} = \{[x]: x \in P\} \quad (7)$$

and the fundamental theorem of equivalence classes (Lipschutz, 1965) says that $H_0(P)$ forms a partition of P.

The number of equivalence classes, β_0, can change as α is increased or decreased. A homology class at α_2 will be said to be born at α_2 if it did not exists at α_1, for every $\alpha_1 < \alpha_2$. The

homology class $[z]$ shown in Figure 4b, for example, was born at α_2. Suppose that a homology class $[x]$ is born at α_1 and $[y]$ is born at α_2 for $\alpha_1 < \alpha_2$. Then $[x]$ will be said to be older than $[y]$.

Homology classes can also die. The death of a homology class will simply mean that two classes have merged so that two points that are not homologous at α_1 become homologous at α_2. To see this, consider the homology classes $[x]$ and $[z]$ at α_2 shown in Figure 4b. They both represent different homology classes because the point x cannot be connected to z by a path. At α_3, on the other hand, $x \sim z$ or $z \sim x$ so that x is member of $[z]$ or z is a member of $[x]$. The result is a reduction in the number of homology classes. When homology classes merge, it will be necessary to use the Elder rule (Edelsbrunner, 2004) from persistent homology to determine which classes die from a merger and which ones live. The Elder rule states that when two classes merge, the older class will continue to live. Therefore, according to the Elder rule, the class $[x]$ will live after the merger with the class $[z]$ at α_3 and $[z]$ will die. The reason $[x]$ lives is because it was born at α_1 and [z] born was born at α_2 so that $[x]$ is older. The lifetime or persistence index of a homology class will be defined as the difference between the pointwise significance level at which it dies and the one at which it was born. If a homology class never dies, then its persistence index, by convention, will be set to infinity.

The evolution of a homology class can be monitored using a barcode (Ghrist, 2008), which is a collection of horizontal lines representing the birth and death of homology classes. Following the convention of persistent homology, the *y*-axes of barcodes will be denoted by H_0 and the *x*-axes will be the pointwise significance level. In the barcode, the birth of a homology class will begin a horizontal line segment at the pointwise significance level at which it was born. The line segment will terminate at the pointwise significance level at which it dies.

An example barcode is shown in Figure 4e for the evolution of homology classes shown in Figures 4a through 4d. The homology class $[x]$ was born at α_1 so that a horizontal line begins at α_1. The patch does not merge with another patch at α_2 so that the horizontal line continues through α_2. The homology class $[z]$ is born at α_2 and the birth of the homology class results in a new line starting from α_2. The merger of the homology classes $[x]$ and $[z]$ at α_2 results in the death of $[z]$. According to the Elder rule, the horizontal line corresponding to $[x]$ in the barcode continues through α_3, but the line corresponding to $[z]$ terminates at α_3. Also note the birth of a new homology class $[q]$ at α_3 and the corresponding beginning of the line segment. Another merger occurs at α_4 because $x \sim q$ and the Elder rule determines that the line segment for $[q]$ ends and the horizontal line for $[x]$ continues. The arrow indicates that $[x]$ never dies.

63.2 Persistent homology of red-noise

To understand the topology of patches generated from red-noise processes, it is useful to use Monte Carlo methods to determine, for example, the number of patches at a particular pointwise significance level, or similarly, the number of holes. Shown in Figure 5a is the ensemble mean of the number of patches (denoted by β_0, hereafter) as a function of α for the Moret, Paul, and Dog wavelets obtained from generating 100 wavelet power spectra of red-noise processes of

length 300 and computing β_0 for each of the wavelet power spectra at each pointwise significance level.

For the the Morlet ~~and Paul wavelets,~~ the number of patches reached minima at $\alpha = 0.01$ and $\alpha = 0.99$ and maxima at $\alpha = 0.18$. The number of holes, as noted by Schulte et al. (2015), peake at $\alpha = 0.99$ and approach zero as α tends to unity. ~~The minima in the number of patches for the Dog wavelet was the same as the Morlet and Paul wavelets, but the maximum occurred at $\alpha = 0.5$. Perhaps more interesting are the number of holes for all three wavelets: very few holes existed at low pointwise significance levels and the number increased rapidly until $\alpha = 0.9$, at which point β_0 no longer increased.~~

To understand more fully the curves shown in Figure 5a, the persistent homology of patches generated from red-noise processes of length 150 was computed as α varies from 0.01 to 0.99, and barcodes representing the evolution of patches in the wavelet power spectra were computed. In each case the lag-1 autocorrelation coefficients were set to 0.5, but the results are identical for other autocorrelation coefficients. Shown in Figure 6a is a barcode corresponding to a wavelet power spectrum obtained using the Morlet wavelet. Recalling that the beginning of the line segment represents the birth of patches, the barcode indicates that a few patches were born at α = 0.02. As α increases to α = 0.3 more patches are born, consistent with the fact that more spurious results occur for larger pointwise significance levels. Note that, for $\alpha > 0.2$, patches begin to die, representing the merger of smaller patches into larger patches. The merging process occurs until α = 0.7, at which point all patches have merged into a single patch, which is called the essential class. To show that the distribution of the lifetimes for patches generated from red-noise processes is not random, 100 wavelet power spectra of red-noise processes were generated and the persistence indices for all patches in each wavelet power spectra were computed (Figure 6b). The resulting distribution indicates that lifetimes of patches is typically 0.1 and relatively few patches live longer than 0.6. Overall, the distribution characterizes patches generated from red-noise processes as short-lived.

~~A typical barcode corresponding to a wavelet power spectrum of a red-noise process whose wavelet power spectrum was obtained using the Paul wavelet is shown in Figure 6c. The barcode is similar to that of the Morlet wavelet, with many patches being born before $\alpha = 0.3$ and generally merging for $\alpha > 0.3$. The distribution of persistence indices shown in Figure 6d obtained from Monte Carlo methods as described above suggests that the lifetime patches are typically longer than that generated from the Morlet wavelet; the distribution has smaller negative skewness, and there are more persistence indices within the range 0.6 – 0.85.~~

~~A barcode corresponding to a wavelet power spectrum of a red-noise process obtained using a Dog wavelet is shown in Figure 6c. Consistent with Figure 6, the barcode differs from that of the other wavelets, with many patches merging for $\alpha > 0.4$. In fact, unlike the case for the Morlet wavelet, many patches were found to have merged after $\alpha > 0.8$. The merging for large α suggests that patches generated using the Dog wavelet tend to be smaller so that it takes longer to form the essential class. In other words, larger patches represent a larger fractional area of the wavelet domain so that fewer of them can reside in the wavelet domain, whereas a greater number of small patches can exist in a wavelet domain of equal size. The results for the Monte Carlo methods~~

~~shown in Figure 6f show that the distribution of persistence indices is indeed different. Compared to the Morlet and Paul wavelets, fewer patches were found to have persistence indices less than 0.15, and persistence indices are more uniformly distributed in the range 0.2 – 0.7.~~

The non-random evolution of patches and holes for red-noise processes suggests that a test can be developed that uses the information of patches at many pointwise significance levels. One would expect that patches arising from signals would behave differently, breaking apart less frequently when the pointwise significance level is increased. In the extreme case that a patch is the result of taking the wavelet transform of a pure sinusoid, there would be a single patch at all pointwise significance levels and that patch would not break apart. Suppose, however, that white noise was added to the sinusoid such that the patch breaks apart for small α. In this case, the area of the patch would be smaller with respect to the pure sinusoid case, the difference in area arising from the splitting of the patch. On the other hand, if α were increased, then the patch would become a single patch by merging, and the area will be closer to that of the pure case. These facts suggests that a test can be constructed whose test statistic is calculated over a set of pointwise significance levels, capturing the behavior of patches as they evolve. In particular, the test should make use of how the evolution of patches and holes under the null hypothesis of red noise is not random and how the evolution of patches arising from signals may differ from that of the null hypothesis.

74. Development of the cumulative areawise test

74.1 Geometric Pathways

~~Unlike the geometric test that assesses the significance of patches at a single pointwise significance level, the cumulative areawise test (referred to as the areawise test, hereafter) will assess the significance of a patch as it evolves under a changing pointwise significance level. The goal of the method is to remove the binary decision from which the geometric test suffers. The idea behind the test statistic will be that a patch that is consistently geometrically significant for different pointwise significance levels is more significant than a patch that is only geometrically significant at a single pointwise significance level. In other words, the geometric test simply assesses the significance of patches at a single pointwise significance level but does not take into account that the patch could have been significant at that particular pointwise significance level by chance (Figure 3).~~

The first step of the cumulative areawise test is to select a finite set of pointwise significance levels that remains fixed throughout the testing procedure. Although there are infinite number of pointwise levels to choose from, it will be shown through empirical arguments that the selection need only be limited to a finite set of pointwise significance levels. Furthermore, it will be shown that one needs only to limit the test to a certain range of pointwise significance levels, making the test more computationally feasible. The selection of pointwise significance levels will be discussed in depth in Sections 4.2. and 4.5. For the theoretical development of the testing procedure, the pointwise significance levels will be left unspecified.

The evolution of a patch under a changing pointwise significance level will be made precise by introducing the notion of a geometric pathway. A geometric pathway will be defined as a

collection $\mathcal{P}$ of r patches at the corresponding pointwise significance levels $\alpha_1 < \alpha_2 < \ldots < \alpha_r$ such that

$$P_1 \subset P_2 \subset P_3 \subset \cdots \subset P_{L_{\mathcal{P}}} \tag{15}$$

and

$$g_1 < g_2 < g_3 \ldots < g_{L_{\mathcal{P}}}, \tag{16}$$

where each g_j is a normalized area corresponding to the patch P_j. For this testing procedure, the normalized area will be calculated by dividing the patch area by the scale coordinate of the centroid squared. The inequalities (16) are guaranteed to hold for any nested sequence (15) (Appendix B). Viewing the α_j's as time parameters, one can think of the patch being in its initial configuration at α_1 and at its final configuration at $\alpha_{L_{\mathcal{P}}}$. The length of a pathway will be given by $L_{\mathcal{P}}$, the number of elements in the pathway. If the computation of the geometric pathways is limited to an interval $I = [\alpha_{min}, \alpha_{max}]$, then all pathways will end at the same pointwise significance level but need not begin at the same pointwise significance level. The reason why all pathways end at α_{max} is because, once a pathway is generated, it can never die, as elements of $\mathcal{P}$ grow in size relative to its initial element.

There is a close relationship between geometric pathways and persistent homology. The birth of homology classes also signifies the creation of a geometric pathway. In contrast, the death of homology classes does not indicate the termination of a geometric pathway. According to Equation (9), once the first element of the pathway is created the pathway cannot terminate because elements grow relative to the first element.

The number of geometric pathways that are computed in a given wavelet power spectrum is related to $\Delta\alpha$ and the persistent homology of patches quantified in Section 6.2. To see this, suppose geometric pathways were calculated at the resolution $\Delta\alpha = \alpha_3 - \alpha_1$ starting at $\alpha_1 = \alpha_{min}$ and ending at $\alpha_3 = \alpha_{max}$ as shown in Figure 4. At this resolution, two pathways would be created, $X_1 \subset X_3$ and Q_3. If the point z had not become homologous to the point x at α_3, then an additional pathway corresponding to $[z]$ would have been calculated because it would still be a path-connected component (i.e., a patch) distinct from X_3 and Q_3. The argument suggests that only geometric pathways comprised of patches with lifetimes greater than or equal to $\Delta\alpha$ will be detected. A natural questions thus arises: how small should be $\Delta\alpha$? It should certainly be made small enough to adequately capture the birth and merging of patches. The distribution of persistent indices shown in Figure 7 suggests that $\Delta\alpha = 0.01$ because most persistent indices are at that value. However, the discussion in Section 8 will suggest a coarser resolution may be used without altering the statistical properties of the test.

~~To illustrate the idea of a pathway, it is perhaps best to consider an ideal case (Figure 7). Consider, for example, three pathways X, Y, and Z whose lengths are, respectively, $r_x = 4$, $r_y = 4$, and $r_z = 5$. The pathway X can be written explicitly as~~

~~$$X_1 \subset X_2 \subset X_3 \subset X_4, \quad (17)$$~~

~~indicating that the patch exists at $\alpha_1^* = \alpha_2$, $\alpha_2^* = \alpha_3$, $\alpha_3^* = \alpha_4$, and $\alpha_4^* = \alpha_5 = \alpha_{max}$. There also exists another pathway Y such that $X_2 = Y_2$, $X_3 = Y_3$, and $X_4 = Y_4$ but whose initial element Y_1 is distinct from X. The pathways are thus distinct until α_3, at which point the pathways merge, resulting in the remaining elements being identical (Figure 7b). The third pathway Z, on the other hand, shares only one element with X and Y, merging at α_5, so that Z represents a distinct pathway. Unlike the other pathways, the pathway Z begins at $\alpha_1 = \alpha_{min}$, implying that the pathway exists over the entire interval I.~~

~~The development of the areawise test will require the calculation of a test statistic for all pathways in the wavelet domain that end at the same pointwise significance level for a selected interval I. The test statistic used in this procedure will be the total sum of normalized areas~~

~~$$\gamma = \sum_{j=1}^{r} g_j, \quad (18)$$~~

~~one for each pathway.~~

~~The calculation of the critical level for the test can be computed using Monte Carlo methods as follows: (1) fix I and a partition of the interval with uniform spacing and generate red-noise processes with the same autocorrelation coefficients as the input time series; (2) for each red-noise process generate synthetic wavelet power spectra and all corresponding pathways; and (3) for every pathway, compute γ, resulting in a null distribution from which the desired critical level of test can be obtained. The critical level corresponding to the 5% significance level of the test, as an example, is the 95% percentile of the null distribution.~~

7.2 Test Construction

Associated with each element of the geometric pathway is the quantity

$$\gamma_j = \sum_{i=0}^{L-j} g_{L-i}, \quad (11)$$

which represents the cumulative sum of the last $L-j+1$ elements of the pathway. One can calculate a p-value for every pathway element using Equation (11) by comparing each γ_j to the null distribution. Mathematically, for a null distribution γ_{null} the p-values are given by

$$p_j = \Pr(\gamma_{null} \geq \gamma_j). \quad (12)$$

The pathway element P_j will be said to be cumulative areawise significant at the α_c significance level if $p_j < \alpha_c$. The union of all cumulative areawise significant pathway elements will be the output of the testing procedure.

The p_j satisfy

Formatted: Justified

Formatted: Indent: First line: 0"

Formatted: Font: Bold

Formatted: Indent: First line: 0", Space After: 10 pt, Line spacing: Multiple 1.15 li

$$p_1 < p_2 < p_3 \ldots < p_L \quad (13)$$

because $\gamma_i > \gamma_j$ for $i > j$. The inequality (13) and nested sequence (8) together show that the cumulative areawise significance of wavelet power coefficient is a monotonic function of the pointwise significance. To see this, denote by x_j a wavelet power coefficient of the patch P_j in a geometric pathway. If p_j^{pw} is the p-value of x_j associated with the pointwise test, then $p_i^{pw} > p_j^{pw}$ for $i > j$. Let F be a function assigning to every p_j^{pw} a p_j. The function F is everywhere monotonically increasing because $p_i^{pw} > p_j^{pw}$ implies that $F(p_i^{pw}) = p_i > p_j > F(p_j^{pw})$ for $i > j$ by inequality (13). This monotonicity property is not shared by the areawise or geometric tests, where there is no one-to-one function between the pointwise significance p-values and p-values for the areawise or geometric tests. In other words, wavelet power coefficients of different pointwise significance can have identical areawise or geometric significance. The monotonicity property also implies that each p_j is only a function of p_j^{pw} and thus it has been shown that the cumulative areawise test is free of a binary decision (Objective 3).

~~**4.2 Pointwise Significance Level Selection: Maximization Method**~~

~~It will often happen that pathways under consideration are of different lengths so that it is not obvious what pointwise significance patches to report after the implementation of the testing procedure. The problem can be circumvented, however, using the following procedure: Let~~

$$\gamma_j = \sum_{i=0}^{r-j} g_{r-i} \quad (19)$$

~~be the cumulative sum associated with the jth element P_j of a pathway with length r, and let γ_{crit} be the critical level of the test; then, the appropriate pointwise significance level to use for a pathway is determined by the following rule:~~

$$\gamma_{\max} = \max_{j=1,2,\ldots,r} \gamma_j > \gamma_{crit}, \quad (20)$$

~~where the statistic satisfying the above rule is denoted by γ_{max}. The element of the pathway corresponding to $\gamma_{\max}$ is the output of the testing procedure. Note that the output elements need not be located at the same pointwise significance levels, contrasting with the pointwise and geometric tests. If, for a given pathway the statistic γ_{max} does not exist, then the pathway is deemed insignificant. It is also important to note that using the inequality (20) guarantees that the appropriate number of pathways will be deemed insignificant if the null hypothesis is known to be true, the reason for which is that if M pathways have statistics γ such that $\gamma > \gamma_{crit}$, then there also must exist M statistics γ_{max} satisfying the inequality (20). In other words, the sums γ must have crossed the critical level at some point in the pathways, or, otherwise, they would have not been deemed significant.~~

74.3 Application to Ideal Pathways

~~The testing procedure is now demonstrated using an ideal case. For the pathway X shown in Figure 7 the test statistic at α_5 is equal to the area of a single patch in the pathway, i.e.~~

~~$\gamma_x = A_4^x$, (21)~~

~~where A_j^x denotes the area of a pathway element at α_j^x. The test statistic at α_1^x, on the other hand, is the sum of four areas~~

~~$\gamma_x = A_1^x + A_2^x + A_3^x + A_4^x$ (22)~~

~~so that the test statistic reaches its maximum value at the smallest pointwise significance level for which the pathway exists. In fact, all pathways will satisfy this property. The evolution of the patch can be viewed as a shape changing with time as shown in Figure 7b. Viewing α_5 as a time parameter, one can say that X merges with a large patch at time α_5 and first appears at time α_2. The evolution of the test statistic is shown in Figure 7c, in which the test statistic increases with decreasing α_j such that its maximum value is attained at α_2 .~~

~~Now consider the pathway Y, whose maximal test statistic is similar to that of X except that A_2^y differs from A_2^x. According to Figure 7b, the pathways evolved identically from α_3 to α_5, merging at α_3. As shown in Figure 7c, the evolution of γ_y only differs slightly from that of γ_x and also represents a significant pathway. For both X and Y, the condition $\gamma_{max} > \gamma_{crit}$ is satisfied at α_3 so that a single patch at α_3 will be the output of the testing procedure, not the two patches representing distinct elements of X and Y at α_2. As will be shown in Sect. 4.6, the overall effect of this merging process is to enhance the detection of signals.~~

~~For the pathway Z, the maximum value of the test statistic can be decomposed into five summands, i.e.~~

~~$\gamma_z = A_1^z + A_2^z + A_3^z + A_4^z + A_5^z$. (23)~~

~~The pathway Z, being longer than X and Y, therefore has an additional pointwise significance level to allow the test statistic to exceed the critical level of the test. However, the elements of the pathway have smaller normalized areas so that the pathway is not significant. The length of a pathway is thus not the only factor influencing the significance of a pathway, as the size of elements is also important.~~

To illustrate the testing procedure, it is perhaps best to consider an ideal case (Figure 9). Consider the pathway X, which can be written explicitly as

$$X_1 \subset X_2 \subset X_3 \subset X_4. \quad (14)$$

The patch exists at $\alpha_1^x = \alpha_2$, $\alpha_2^x = \alpha_3$, $\alpha_3^x = \alpha_4$, and $\alpha_4^x = \alpha_5 = \alpha_{max}$. The test statistics, using Equation (11), for the geometric pathway are

$$\gamma_1^x = g_1^x + g_2^x + g_3^x + g_4^x \quad (15)$$

$$\gamma_2^x = g_2^x + g_3^x + g_4^x, \quad (16)$$

$$\gamma_3^x = g_3^x + g_4^x, \quad (17)$$

and

$$\gamma_4^x = g_4^x \quad (18)$$

where g_j^x denotes the normalized area of a pathway element at α_j^x. According to Figure 9b, both X_1 and X_2 are cumulative areawise pathway elements because $\gamma_1^x, \gamma_2^x > \gamma_{crit}$. The output of the testing procedure is therefore given by

$$X_{sig} = X_2 = X_1 \cup X_2. \quad (19)$$

A similar results holds for the pathway *Y*, where the output of the testing procedure is

$$Y_{sig} = X_2 = Y_1 \cup Y_2. \quad (20)$$

The pathway *Z* shown in Figure 9a can be written as

$$Z_1 \subset Z_2 \subset Z_3 \subset Z_4 \subset Z_5. \quad (21)$$

The test statistics associated with each of the five pathway elements are

$$\gamma_1^z = g_1^z + g_2^z + g_3^z + g_4^z + g_5^z, \quad (22)$$

$$\gamma_2^z = g_2^z + g_3^z + g_4^z + g_5^z, \quad (23)$$

$$\gamma_3^z = g_3^z + g_4^z + g_5^z, \quad (24)$$

$$\gamma_4^z = g_4^z + g_5^z, \quad (25)$$

and

$$\gamma_5^z = g_5^z. \quad (26)$$

As shown in Figure 9b, none of the test statistics exceed γ_{crit} and therefore the pathway elements are not cumulative areawise significant. The total output of the testing procedure in this case will be

$$X_2 = X_{sig} \cup Y_{sig}. \quad (27)$$

~~4.4 The Null Distribution~~

~~Recall from Section 3.2 that, for patches generated from red noise, the merging of patches is not random, with typical lifetimes of patches following a non-uniform distribution, favoring shorter lifetimes. Therefore, if the test statistic is proportional to the lifetime of patches, one can expect the test statistic to follow a similar distribution to that of persistence indices, where the smallest values of γ are preferred. To test this hypothesis, 1000 wavelet power spectra of red-noise processes with fixed autocorrelation coefficients were generated and the cumulative area of all pathways in each wavelet power spectra was computed. In the experiments, $\alpha_{min} = 0.02$, $\alpha_{max} = 0.82$, and $\Delta\alpha$, the discrete spacing between adjacent pointwise significance levels, is set to 0.02 to make calculations less computationally expensive. The results for the Morlet, Paul, and Dog wavelets are shown in Figure 8.~~

The distribution of γ for the Morlet wavelet is generally similar to the shape of the distribution for the persistence indices for H_0 (Figure 6b), where the smallest values of γ are preferred. It turns out that the distribution of γ for the Morlet wavelet can be well described by an exponential distribution. Using the method of maximum likelihood (Weerahandi, 2003), a theoretical exponential distribution was fitted to the empirical distribution, where the empirical distribution was found to be best described by an exponential distribution with mean 6.5. To show that the theoretical distribution models the empirical distribution, the percentiles of a theoretical exponential distribution with mean 6.5 were plotted as a function of the percentiles of the empirical distribution (Figure 8b). The linear relationship between the percentiles shown in Figure 8b indicates that the theoretical distribution well models the empirical distribution, with the 95% percentiles only differing by 1.0.

The results for the Paul wavelet are generally similar except that the mean value of the test statistic was found to be smaller with a value of 6.2. The distribution of the test statistic, unlike that for the Morlet wavelet, was found to be poorly modeled by an exponential distribution, with the empirical distribution generally having larger values than those predicted by a theoretical exponential distribution; this difference was more pronounced for larger γ.

The results for the Dog wavelet are shown in Figure 8c. The null distribution differs from that for the Morlet and Paul wavelets, with the mean value of the distribution being 3.4. This result is perhaps not surprising; Figure 6 shows that the lifetimes of patches generated using the Dog wavelet have larger persistence indices attributed to how patches die at larger α compared to that of the Morlet and Paul wavelets, resulting from the lack of merging with the essential class. The delayed merging implies that the area of patches must be smaller relative to that of the Morlet and Paul wavelets for large α to prevent merging with other patches. The effect is to generate a smaller cumulative sum. The null distribution, unlike that for the Morlet wavelet, was not found to be well-described by a theoretical exponential distribution.

As will be shown in Section 4.6, the implementation of the proposed testing procedure will allow small patches that are seemingly indistinguishable from noise to emerge, allowing the recovery of a signal that has been contaminated by noise.

4.5 Computational Remarks

An important parameter in the areawise test is $\Delta\alpha$, the spacing between pointwise significance levels. It is critical that $\Delta\alpha$ is chosen to be small enough to sample the merging of patches so that, for example, the null distribution can be representative of the null hypothesis. If $\Delta\alpha$ is set too large, then a pathway may be missed entirely if it is born between two adjacent pointwise significance levels. On the other hand, if $\Delta\alpha$ is too small, then the test will become computationally expensive, requiring the calculation of more normalized areas. The distribution of the persistence indices for H_0 suggests that $\Delta\alpha = 0.01$ because the modes of the distributions are 0.01. However, it will be shown in Section 4.6 that the results of the areawise test do not differ if one chooses $\Delta\alpha = 0.02$.

Another important parameter is α_{max}. It would be computationally inefficient if the testing procedure had to be performed on pathways whose end points are, for example, $\alpha_{max} = 0.99$.

~~Moreover, as shown in Figure 5b, the number of holes increases for increasing α, which would increase the computational costs associated with the calculation of patch areas for large α. The reason for the increase in computational costs is that the areas of the holes need to be subtracted from the area the patches would have if they did not contain holes. To circumvent the problem, properties of the areawise test can be assessed at peaks of the curves shown in Figure 5a so that results for other end points can be inferred. For the Morlet wavelet, the peak occurs at α = 0.18 so that $\alpha_{max} = 0.18$. To the left of the maximum, the average number of patches are equal to that of some pointwise significance level to the right of the maximum. However, patches to the right of the maximum have larger areas so that choosing $\alpha_{max} > 0.18$ would result in the testing procedure detecting larger patches as significant. The same reasoning holds for the Paul and Dog wavelets except that the maximum of the curve for the Dog wavelet is $\alpha_{max} = 0.5$. The sensitivity of the testing procedure to α_{max} is discussed in Section 4.6.~~

~~Figure 6 also suggests that α_{max} need not be any larger than 0.7 for the Morlet wavelet because all patches have merged with the essential class before that point, at least for red noise processes. Thus, for $\alpha_{max} > 0.7$, patches arising from signals cannot be distinguished from those generated from noise because all patches have merged into a single, large patch. A similar argument holds for the Paul wavelet, but, for the Dog wavelet, all patches merge at a larger α_{max}.~~

7. ~~4.6~~ Comparison with geometric test

7.1 True Positve Detection

With the cumulative areawise test now developed, it will be useful to assess the statistical power of the test relative to that of the geometric test. The first aspect of the assessment will be to quantify how well both tests detect true positive results. To do so, let

$$x(t) = A\sin(2\pi ft) + w(t) \tag{24}$$

be a sinusoid with amplitude A, frequency f, and additive Gaussian white noise $w(t)$. The goal will be to evaluate the ability of both tests to detect true positives within a particular period band. A theoretical patch to which the ability of the geometric and areawise tests were compared was constructed as follows: (1) the time series $x(t)$ for all $t \in [0, 500]$ was generated but with no additive white noise; (2) the wavelet power spectra of $x(t)$ was computed and the 5% pointwise significance test was performed on the wavelet power spectrum; and (3) the width of the significance patch in the wavelet power spectrum was calculated at t = 250 where edge effects are negligible. The theoretical patch ~~calculated using the Morlet wavelet~~ is indicated by dotted lines in Figure 9, where the theoretical patch is a rectangle of fixed width extending from t = 0 to t = 500. In all experiments, α_{max} = 0.18 and $\Delta\alpha = 0.02$, but implications of other choices are discussed at the end of the section.

To assess the ability of the tests to detect true positives, the area of patches deemed significant by the tests were compared to the total area of the theoretical patch. More specifically, if P_{geo} is the union of all pointwise significance patches at α that are geometrically significant at the α_{geo} level and P_{theory} is the theoretical patch, then

$$r_a = \frac{A_{P_{geo} \cap P_{theory}}}{A_{P_{thoery}}} \tag{25}$$

represents the areal fraction of P_{theory} detected by the geometric test, where $A_{P_{geo} \cap P_{theory}}$ denotes the area of $P_{geo} \cap P_{theory}$ and $A_{P_{thoery}}$ denotes the area of P_{theory}. If r_a = 1, then the test detected all of the true positive results that are known by construction. Small values of r_a indicate that the tests performed poorly, detecting only a fraction of the theoretical patches to be significant. A similar construction can be made for the areawise test by replacing P_{geo} with P_c. Figure 9a illustrates the procedure for the cumulative areawise test when f = 0.8, A = 0.8, and the signal-to-noise ratio (defined below) equals 1.0. As indicated by the thick black contours, the cumulative areawise test was able to detect 30% of the true positives comprising the theoretical patch, whereas Figure 9b shows that the geometric test was only able to detect 20% of the true positives. It will be necessary to compute N = 1000 values of r_a for different values of f and signal-to-noise ratios of the Gaussian white noise to determine if the tests truly perform differently. In the following experiments, the signal-to-noise ratio is defined as

$$\sigma = 10\log\left(\frac{p_{signal}}{p_{noise}}\right), \tag{26}$$

where

$$p_{signal} = \frac{A^2}{2}, \tag{27}$$

p_{noise} is the average power of the Gaussian white noise, and σ is measured in decibels (DB). In the discussion below the results for the Morlet are presented first~~; results for the Paul and Dog wavelets are discussed at the end of the section.~~ It is also noted that because σ and A do not vary independently there is no need to perform different experiments for different values of A. For the experiments, A was set to 1.0.

In the first experiment, the areawise significance (denoted by α_c, herreafter) was set to 0.05, α_{geo} = 0.05, and α = 0.01, 0.05, 0.1. The signal-to-noise ratio for the Gaussian white noise was varied from -5 DB to 5 DB. The results are shown in Figure 10. For both tests, the ability to detect true positives increased with increasing signal-to-noise level. At low signal-to-noise ratios, the tests performed similarly, detecting on average 10% of true positives. ~~On the other hand,~~ D~~d~~ifferences between the test performances became larger as the signal-to-noise ratio was increased and~~; in fact,~~ the areawise test outperformed the geometric test regardless of the chosen pointwise significance levels when $\sigma \geq -2.5$ DB. The results indicate that the areawise test is particularly useful in low noise situations but one can expect the test to detect more true positives even in high noise conditions. In agreement with Figure 3a, ~~It also worth noting that~~ the performance of the geometric test depended strongly on the chosen pointwise significance level, especially when the signal power was high.

A second experiment was conducted where α_c = 0.01 and α_{geo} = 0.01. The same pointwise significance levels as the first experiment were chosen and the range of signal-to-noise ratios was also the same. The results are presented in Figure 11. Note that the shape of the curves for both

tests are the same as the first experiment, where greater true positive detection is preferred for large signal-to-noise ratios. However, both tests detected fewer true positives, consistent with how the significance levels of the tests were increased. The relationships between the tests, like for experiment 1, depended on the signal-to-noise ratio, ~~;~~ F~~f~~or low signal-to-ratios, the performance of the tests are similar, whereas for high signal-to-noise ratios differences between r_a for the areawise test and ~~and r_a for the~~ geometric increases. All the results were found to be statistically significant.

Additional experiments were performed using different values of f to determine if the frequency at which patches are located affects the performances of the areawise and geometric tests. True positive detection, for a fixed σ, was generally found to increase for larger f, though the areawise test was still found to detect more true positives. The behavioral assessments of both tests for $f < 0.8$ required the use of synthetic time series whose lengths were greater than 500, as patches lengthened in the time direction for lower frequencies and the COI impacts became more pronounced. The array of experiments concluded that the areawise test should be the preferred method for signal detection.

7.2 False Positive Detection

The false positive detection of both tests depends on the topology of patches. The number of false positives produced by the geometric test performed at the pointwise significance level α on average will be

$$N_{geo}(\alpha, \alpha_{geo}) = \alpha_{geo}\beta_0(\alpha). \quad (32)$$

For the cumulative areawise test, the number of false positives produced will be on average

$$N_c(\alpha_{peak}, \alpha_c) = \alpha_c\beta_0(\alpha_{peak}), \quad (33)$$

where α_{peak} satisfies $\alpha_{min} \leq \alpha_{peak} \leq \alpha_{max}$ and denotes the pointwise significance level for which β_0 locally reaches a maximum (Figure 5). If $\alpha_c = \alpha_{geo}$, then the ratio of false positives for both tests is

$$r_{false} = \frac{\beta_0(\alpha)}{\beta_0(\alpha_{peak})}. \quad (34)$$

Thus, if $\alpha = \alpha_{peak}$ both tests on average will have the same number of false positive results. On the other hand, $r_{false} < 1$ if $\alpha_{max} < \alpha < 0.18$. According to Figure 5, $N_{geo}(0.05, 0.05)$ is approximately 11 and $N_c(0.18,0.05) = 15$ so that $r_{false} = 0.73$ and therefore one can expect 36% more false positives. However, this calculation is an overestimate because the output of the cumulative areawise test is the union of pathway elements as shown in Figure 7 and discussed in Section 7. In fact, an experiment was conducted by generating 1000 wavelet power spectra of red-noise process with $\rho = 0.5$ and lengths equal to 1000. The ratio r_{false} for $\alpha_{max} = 0.18$, $\alpha_{min} = 0.02$, $\Delta\alpha = 0.02$, $\alpha_c = 0.05$, $\alpha_{geo} = 0.05$, $\alpha = 0.05$ was computed for each wavelet power spectra. The mean value of r_{false} was found to be $r_{false} = 0.82$, slightly higher than the theoretical value. The result implies that one can expect the cumulative areawise test to produce 22% more false positive results. Confidence in results can be enhanced if $\alpha_c = 0.01$ without much loss in

true positive detection, as shown by the comparison of the curves for $\alpha_c = 0.01$ and $\alpha_c = 0.05$ in Figures 8 and 9. An experiment similar to the previous experiment with $\alpha_c = 0.01$ showed that one can expect 50% more false positives for the geometric test. The reduction in false positive detection together with relatively high true positive detection suggests that the cumulative areawise test is reliable when $\alpha_c = 0.01$.

~~Another set of experiments were conducted to evaluate how the areawise test detects false positives relative to the geometric test. In the first experiment, 1000 wavelet power spectra arising from red-noise processes of length 1000 with equal lag-1 autocorrelation coefficients were generated. For each wavelet power spectra, the geometric and areawise tests were applied at the 0.05 level and the pointwise significance level was set to 0.05. The results were found to be independent of the chosen lag-1 autocorrelation coefficient, so only the results for the case when the lag-1 autocorrelation coefficient was set to 0.5 are presented. The ratio between the number of false positive results for the geometric and areawise tests for each wavelet power spectra was then computed. The average ratio was found to be 0.08, implying that one can expect more false positive results for the areawise test relative to the geometric test. However, in most cases, the increase in the number of false positive results with respect to the geometric test was small compared to the increase in true positive detection (Figures 10 and 11), suggesting that the areawise test had greater overall statistical power. Note that spurious patches arising from the areawise test will have generally larger areas so that results from the test should be interpreted carefully.~~

~~The above experiment was repeated for different pointwise significance levels. The experiments indicated that the difference between false positive detection between both tests decreased as α increased, the reason for which is that the number of false positive results for the geometric test for a fixed α_{geo} will increase until $\alpha = 0.18$, the point at which the number of patches generally peaks (Figure 5a). For $\alpha > 0.18$, the number of spurious patches may decrease, but the areas of the spurious patches resulting from the geometric test will be larger. A similar argument would hold if α_c and α_{geo} were increased or decreased by the same amount.~~

~~All the above experiments were performed with the Paul and Dog wavelets and the theoretical 5% pointwise significance patch was adjusted to account for the different scale- and time-localization properties of the wavelets, where the corresponding theoretical patches were found to be wider in scale compared to the Morlet wavelet. The results from the experiments were qualitatively similar to that of the Morlet wavelet, as the areawise test detected more true positives than the geometric test for both the Paul and Dog wavelets. Like for the Morlet wavelet, the differences in performances were more pronounced for larger σ. The array of experiments provides evidence that the areawise test has greater statistical power than the geometric test regardless of the chosen analyzing wavelet.~~

~~All the above experiments were conducted for different values of $\Delta\alpha$. For $\Delta\alpha < 0.02$, the results of the experiments were virtually identical. On the other hand, if $\Delta\alpha > 0.03$, the percentage of true positives detected by the areawise test decreased. The results suggest that $\Delta\alpha$ should be chosen to be no larger than 0.02 to ensure that true positive detection is maximized.~~

Formatted: Indent: First line: 0"

~~The parameter α_{max} was found to strongly influence the performance of the areawise test. For $\alpha_{max} > 0.18$ and $\alpha_e = 0.05$, the number of true positives detected increased and the number of false positives decreased relative to the geometric test with $\alpha_{geo} = 0.05$. However, the normalized areas of the spurious patches resulting from the areawise test with $\alpha_e = 0.05$ or $\alpha_e = 0.01$ were found to be 2-10 times larger than those resulting from the geometric test, making spurious features appear significant. The problem was remedied by decreasing α_e, but this adjustment was found to be the same as increasing α_{max}. For $\alpha_{max} < 0.18$, the true and false positive detection of the areawise test for $\alpha_e = 0.05$ approached that of the geometric test until $\alpha_{max} = 0.05$, at which point they were approximately equal.~~

5. Climate applications

To determine if any features in the wavelet power spectra of the AMO, NAO, Niño 3.4, and PDO time series are deterministic or stochastic, the cumulative areawise test was performed on geometric pathways in the wavelet power spectra at the 0.01 level. A red-noise background spectrum was used for each, with $\alpha_{max} = 0.18$, $\alpha_{min} = 0.02$, and $\Delta\alpha = 0.02$. ~~The wavelet power spectrum for the AMO index is shown in Figure 12a. Although the AMO is usually characterized by its multi-decadal variability, no areawise significant wavelet power coefficients were detected at periods greater than 24 months, suggesting that the variability results from stochastic processes. In contrast, from 1860 to 1940, areawise significant wavelet power coefficients were detected, primarily in the 2-16 month period band. After 1940, no areawise significant wavelet power coefficients were identified at any time scale.~~

~~The wavelet power spectrum of the NAO index is shown in Figure 12b. Only a few regions of areawise significance were detected: one at a period of 2 months and 1900, a second one at a period of 6 months and 1880, and a third one located at a period of 100 months and 1870. The lack of areawise significance suggests that the NAO is a red-noise process with no preferred time scale, consistent with Feldstein (2002).~~

The wavelet power spectrum for the Niño 3.4 index, on the other hand, does indicate potential predictive capabilities (Figure 12c). There are two notable features, one extending from 1870 to 1920 in the 16-64 month period band and another one extending from 1960 to 2014 in the 8-64 month period band. Perhaps just as interesting is the deficit in areawise significance from 1920 to 1960 in the 8-64 month period band. The deficit could be the result of the 2-7 year mode being modulated by a decadal ENSO mode, a nonlinear paradigm (Timmermann, 2002). Such a modulation would imply that the behavior of the 2-7 year mode is determined by the phase of the decadal mode, where, for example, more extreme El Niño phases would be favored if the decadal mode is in a positive regime. On the other hand, results shown in Figure 11c show that neither the decadal nor the multi-decadal variability exceed a red-noise background so modulations would be difficult to predict.

The wavelet power spectrum of the PDO index is shown in Figure 12d. Like the AMO index, there is enhanced variance at multi-decadal time scales but the variance does not exceed a red-noise background. Areawise-significant regions, however, were detected in the 2-8 month

period band from 1900 to 1960. The results indicate that the PDO is a red-noise process, consistent with prior work showing that the PDO results from the oceanic integration of atmospheric white-noise stochastic forcing (Newmann et al., 2003).

6. Conclusions

An areawise test was developed for assessing the significance of features in wavelet power spectra. The test was generally found to have greater statistical power than the geometric test except possibly under high-noise situations, in which case the tests were found to perform similarly. The main advantage of the new testing procedure is that the results are no longer dependent on the chosen pointwise significance level. In contrast, the geometric test results were found to be very sensitive to the chosen pointwise significance level, making it difficult for researchers to decide what patches are significant and what patches are not significant. In particular, the cumulative areawise test was found to detect more true positives relative to the geometric test for some common pointwise and geometric significance levels. The large increase in true-positive detection of the cumulative areawise test was also accompanied by a small increase in false-positive detection compared to the geometric test performed at the 5% level, resulting in the areawise test having greater statistical power. The difference between the tests, however, was found to decrease for low signal-to-noise ratios, indicating that there are still deterministic features that are, in principle, indistinguishable from background noise.

The cumulative areawise test applied in this paper was limited to two-dimensional wavelet power spectra. The method, however, may also be applied to global power spectra obtained by time averaging wavelet power at each scale. In this 1-dimensional case, geometric pathways would be a nested sequence of arcs. Each member of the nested sequence would be a portion of a global peak that lies above the critical level of the test. Additionally, the 1-dimesional test may also prove useful for global coherence (Schulte et al., 2015), which measures the coherence between two time series as a function of wavelet scale. More generally, one can construct an n-dimensional cumulative areawise test where the test statistics would be the cumulative sum of n-dimensional volumes corresponding to a nested sequence of n-dimensional geometric objects.

A potential drawback of the cumulative areawise test is that it may become computationally expensive for very long time series. As the length of the time series increases, the number of geometric pathways that need to be calculated also increases. The increase in the number of geometric pathways was found to be nonlinear (not shown), meaning a small change in the time series length yielded a larger change in the number of geometric pathways. Another limitation is that the test relies on the selection of several parameters. One needs to select I and $\Delta\alpha$. Fortunately, the results of procedure were found to change little if $\Delta\alpha$ was smaller than 0.02.

The results from the climate-mode analysis suggest that the predictability of the ~~AMO,~~ PDO, and ~~NAO~~ is limited and that the multi-decadal variability of the ~~AMO~~ and PDO is the result of a stochastic process. The Niño 3.4 index, by contrast, was found to have deterministic features, implying that future states of ENSO may be predictable. Such predictability is important given that ENSO has regional- to global-scale impacts on precipitation and temperature. The ability to

predict future changes of regional climate thus, to some extent, depends on the ability to predict ENSO. However, currently, the future state of ENSO as determined by climate models is uncertain, with some climate models suggesting large changes and others indicating no change at all (Latif and Keenlyside, 2008).

A Matlab software package written by the author to implement the cumulative areawise test is available at justinschulte.com. .

Appendix A

A path in a set X is defined as a continuous function $f: I \rightarrow X$. A set X is said to be path-connected if any two points x and y in X can be joined by a path. The path-component of a topological space X is the maximal path-connected subset of a set. Intuitively, one can think of a path-component as the largest isolated piece of the set. For example, the set could be the disjoint union of a square and a disc, in which case both the square and the disc are path-components.

Appendix B

Let P_1 and P_2 be two subsets of a patch P with area A such that $P = P_1 \cup P_2$. Let A_1 and A_2 denote the areas of P_1 and P_2, respectively. One can thus write

$$A = A_1 + A_2, \tag{B1}$$

$$A = r_1 A + r_2 A, \tag{B2}$$

and

$$r_2 = 1 - r_1, \tag{B3}$$

where $r_1, r_2 \in [0,1]$. The centroid of P can be written as

$$\frac{1}{A}\iint_P sdsdt = \frac{1}{A}\iint_{P_1} sdsdt + \frac{1}{A}\iint_{P_2} sdsdt \tag{B4}$$

$$= \frac{1}{A}\iint_P sdsdt = \frac{r_1}{A_1}\iint_{P_1} sdsdt + \frac{r_2}{A_2}\iint_{P_2} sdsdt \tag{B5}$$

or

$$C^s = r_1 C_1^s + r_2 C_2^s, \tag{B6}$$

so that

$$\frac{C^s - r_2\, C_2^s}{r_1} = C_1^s, \tag{B7}$$

where C_1^s and C_2^s are the scale coordinates of the centroids for P_1 and P_2. The equation implies that

$$C^S - r_2\, C_2^s > 0 \tag{B8}$$

because C_1^s is always positive. The normalized areas of P and P_1 are given by

$$A^N = \frac{A}{(C^s)^2} \tag{B9}$$

and

$$A_1^N = \frac{A_1}{(C_1^s)^2}. \tag{B10}$$

Thus,

$$r_{norm} = \frac{A_1^N}{A^N} = \frac{r_1^3 (C^s)^2}{(C^s - (1 - r_1) C_2^s)^2}. \quad \text{(B11)}$$

At $r_1 = 0$, $r_{norm} = 0$ because P_1 has no area. At $r_1 = 1$, $r_{norm} = 1$ because $A_1 = A$. Moreover, the function is monotonically increasing for $r_1 = [0\ 1]$ so that $r_{norm} \leq 1$. The same arguments hold for P_2 except that r_{norm} decreases monotonically.

Acknowledgements: Support for this research was provided by the National Science Foundation Physical Oceanography Program (award number 0961423) and the Hudson River Foundation (award number GF/02/14). It is a pleasure to thank R. G. Najjar and N. Higson for the helpful advice, which resulted in an improved manuscript.

References

Edelsbrunner, H. and Harer, J.: Computational topology: an introduction, Amer. Math. Soc., Providence, Rhode Island, 241 pp., 2009.

Edelsbrunner, H. and Harer, J.: Persistent homology-a survey, Cotemp. Math., 12, 1-26, 2010.

Efron, B.: Bootstrap methods: another look at the jackknife, Ann. Statist., 7, 1–26, 1979.

Ghrist, R.: Barcodes: the persistent topology of data, Bull. Amer. Math. Soc., 45, 61-75, 2008.

Grinsted, A., Moore, J. C. and Jevrejeva, S.: Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Process. Geophys., 11 , 561–566, 2004.

Higuchi, K., Huang, J., Shabbar, A.: A wavelet characterization of the North Atlantic Oscillation variation and its relationship to the North Atlantic sea surface temperature. Int. J. Climatol., 19, 1119-1129, 1999.

Hurrell, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M. (Eds.): The North Atlantic Oscillation: Climatic Significance and Environmental Impact. Geophys. Monogr. Ser., 134, American Geophysical Union, 279 pp., 2003.

Jenkins, G. W., Watts, D. G.: Spectral Analysis and its Applications. Holden-Day, San Francisco, California, 541 pp., 1968.

Kerr, R. A.: A North Atlantic climate pacemaker for the centuries. Science, 288, 1984-1985, 2000.

Labat, D.: Wavelet analysis of the annual discharge records of the world's largest rivers, Adv. Water Resour., 31, 109-117, 2008.

Labat, D.: Cross wavelet analyses of annual continental freshwater discharge and selected climate indices, J. Hydrol., 385, 269-278, 2010.

Latif, M., Keenlyside, N. S.: El Niño/Southern Oscillation response to global warming, Proc. Natl. Acad. Sci. USA, 106, 20578–20583, 2008.

Lau, K. M. and Weng, H.: Climate signal detection using wavelet transform: how to make a time series sing. Bull. Amer. Meteor. Soc., 76, 2391–2402, 1995.

Lee, Y. J. and Lwiza, K. M. M.: Factors driving bottom salinity variability in the Chesapeake Bay. Cont. Shelf Res., 28, 1352-1362, 2008.

Mantua, N. J. and Hare, S. R.: The Pacific Decadal Oscillation. J. Oceanogr., 58, 35-44, 2002.

Maraun, D. and Kurths, J.: Cross wavelet analysis: significance testing and pitfalls, Nonlin. Processes Geophys., 11, 505-514, 2004.

Maraun, D., Kurths, J., and Holschneider, M.: Nonstationary Gaussian processes in wavelet domain: synthesis, estimation, and significance testing, Phys. Rev. E, 75, doi: 10.1103/PhysRevE.75.016707, 2007.

Meyers, S. D., Kelly, B. G., and O'Brien, J. J.: An introduction to wavelet analysis in oceanography and meteorology: With application to the dispersion of Yanai waves. Mon. Weather Rev., 121, 2858–2866, 1993.

Newman, M., Compo, G. P., Alexander, M. A.: ENSO-forced variability of the Pacific Decadal Oscillation. J. Climate, 16, 3853-3857, 2003.

Olsen, J., Anderson, J. N., Knudsen, M. F.: Variability of the North Atlantic Oscillation over the past 5,200 years, Nature Geosci., 5, 808-812, 2012.

Torrence, C. and Compo, G. P.: A practical guide to wavelet analysis, Bull. Amer. Meteor. Soc., 79, 61–78, 1998.

Trenberth, K. E.: The Definition of El Niño., Bull. Amer. Meteor. Soc., 78, 2771-2777, 1997.

Vautard, R., Yiou, P., Ghil, M.: Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, Physica D: Nonlinear Phenomena, 58, 95-126, 1992.

Velasco, V. M. and Mendoza, B.: Assessing the relationship between solar activity and some large scale climatic phenomena, Adv. Sp. Res., 42, 866–878, 2008.

Schulte, J. A., Duffy, C., and Najjar, R. G.: Geometric and topological approaches to significance testing in wavelet analysis, Nonlin. Processes Geophys., 22, 139-156, 2015.

Wang, C. and Picaut, J. (Eds. Wang, C., Xie, S.P. and Carton, J.A.): Understanding ENSO physics—A review, Earth's Climate AGU, Washington, D. C, 2004.

Weerahandi, S.: Exact Statistical Methods for Data Analysis, Springer, pp. 329, 2003.

Whitney, M. M.: A study on river discharge and salinity variability in the Middle Atlantic Bight and Long Island Sound, Cont. Shelf Res., 30, 305-318, 2010.

Wilson, M., Meyers S. D., Luther, M. E.: Synoptic volumetric variations and flushing of the Tampa Bay estuary, Clim. Dyn., 42, 1587-1594, 2014.

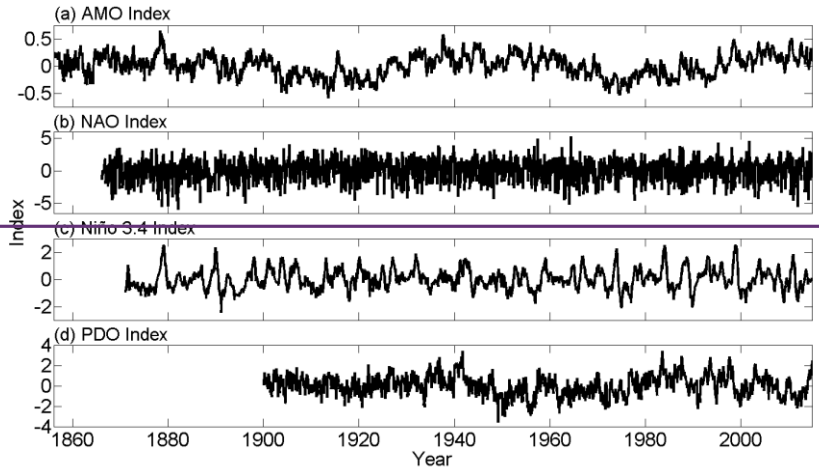
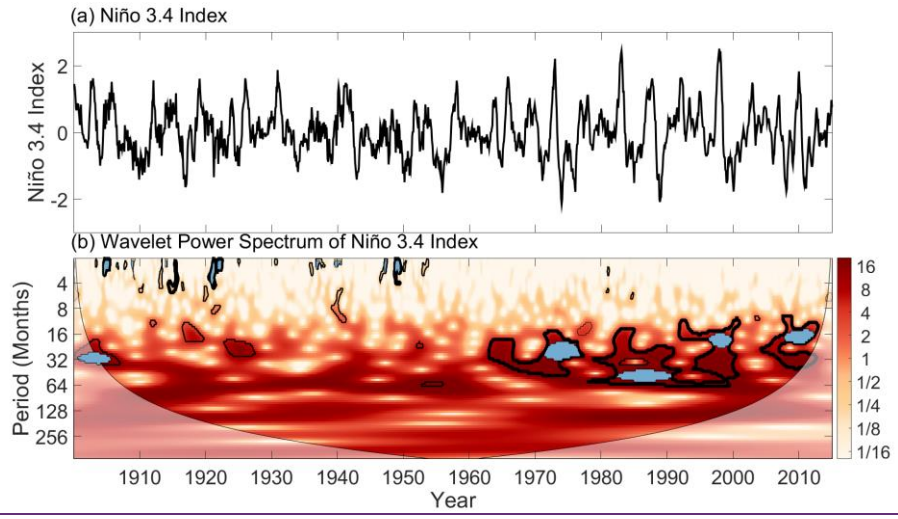


Figure 1. The monthly (a) AMO, (b) NAO, (c) Niño 3.4, and (d) PDO indices. Data sources are the Climate Prediction Center for the AMO index (http://www.esrl.noaa.gov/psd/data/climateindices/list/), National Center for Atmospheric Research for the NAO (https://climatedataguide.ucar.edu/data-type/climate-indices) and Niño 3.4 (http://www.cgd.ucar.edu/cas/catalog/climind/TNI_N34/) indices, and University of Washington for the PDO index (http://research.jisao.washington.edu/pdo/PDO.latest). The Niño 3.4 index was converted to monthly anomalies by subtracting off the mean annual cycle.



~~Figure 2. Wavelet power spectra of the (a) AMO, (b) NAO, (c) Niño 3.4, and (d) PDO indices. Thin black contours enclose regions of 5% pointwise significance and thick blue contours indicate those patches that are geometrically significant at the 5% level. Light shading represents the cone of influence (COI), the region in which edge effects cannot be ignored.~~

Figure 1. Wavelet power spectra of the Niño 3.4 index. Thin black contours enclose regions of 5% pointwise significance and thick contours indicate those patches that are geometrically significant at the 5% level. Light blue shading represents 5% areawise significant subsets of the patches. Light shading represents cone of influence (COI), the region in which edge effects cannot be ignored.

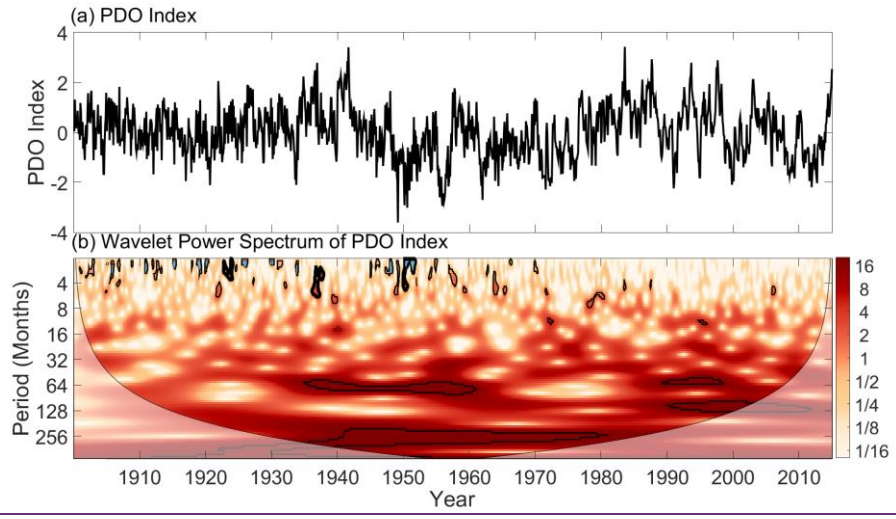


Figure 2. Same as Figure 1 but for the PDO index.

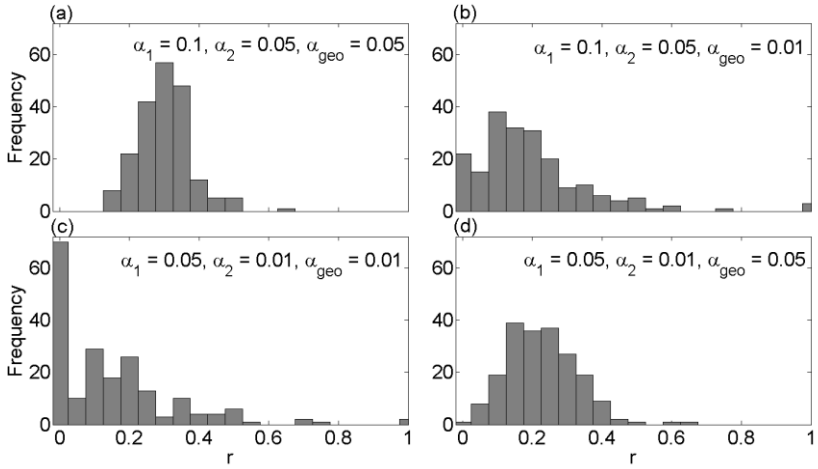


Figure 3. (a) A histogram of $r = \frac{N_{\alpha_1,\alpha_2}}{N_{\alpha_1}}$ for $\alpha_1 = 0.1$, $\alpha_2 = 0.05$, and $\alpha_{geo} = 0.05$ obtained from the generation of 300 wavelet power spectra of red-noise processes of length 1000 with lag-1 autocorrelation coefficients equal to 0.5. (b) Same as (a) but with $\alpha_{geo} = 0.01$. (c) Same as (a) but with $\alpha_1 = 0.05$ and $\alpha_2 = 0.01$, and $\alpha_{geo} = 0.01$. (d) Same as (c) but with $\alpha_{geo} = 0.05$.

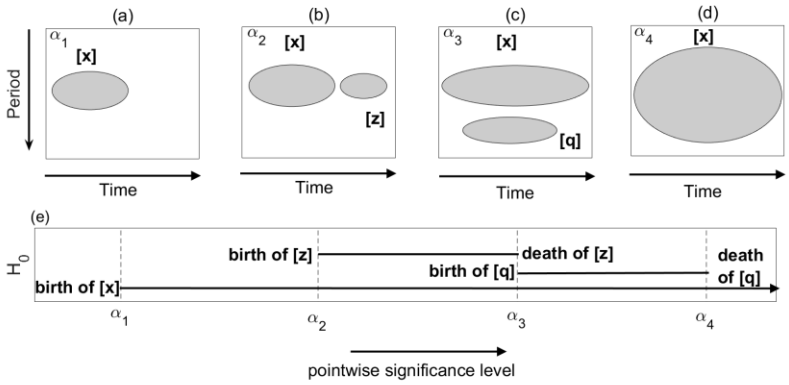
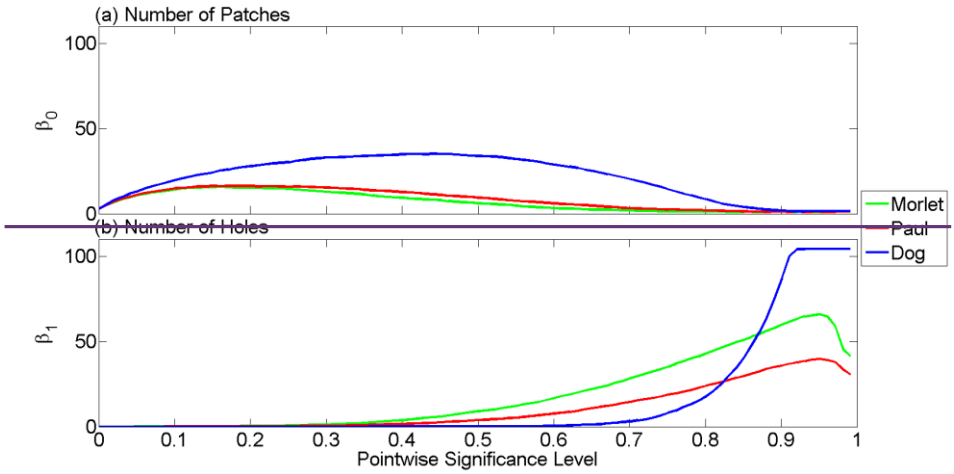


Figure 4. (a) - (d) The topological evolution of patches across four pointwise significance levels. (e) The barcode showing the birth and death of patches throughout the evolution process. Horizontal lines with arrows indicate those patches that never die, the so-called essential classes.



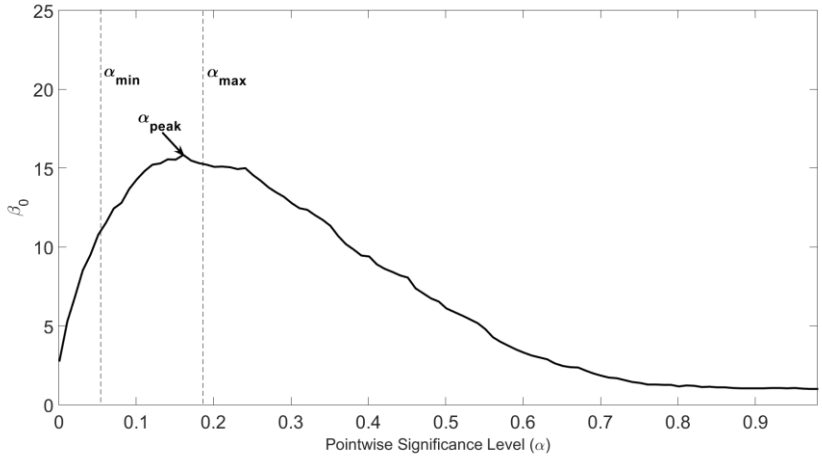
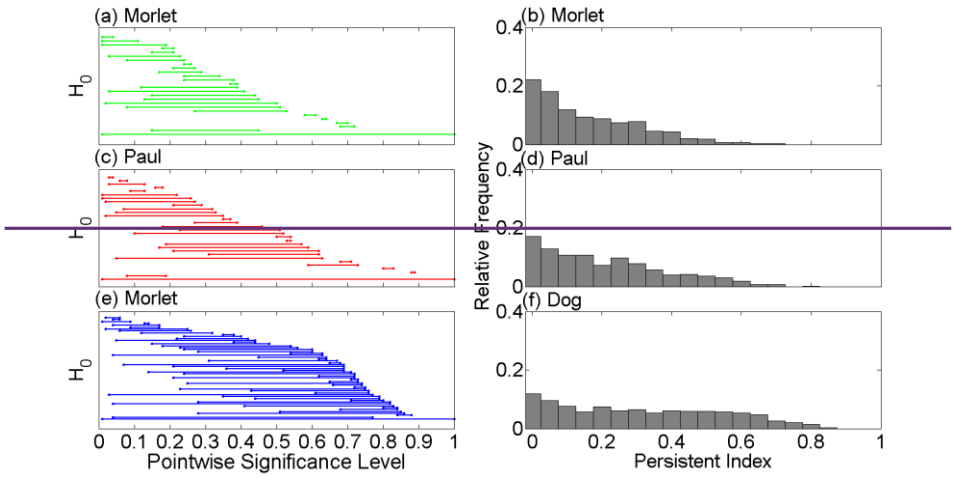
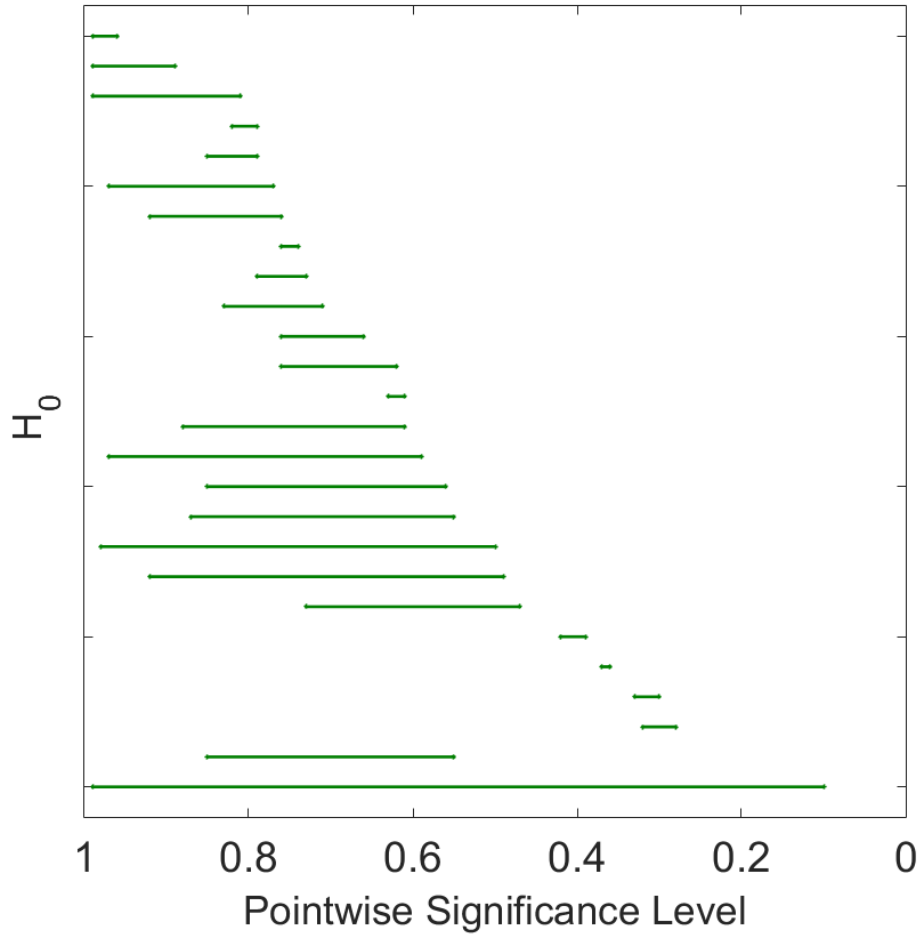


Figure 5. (a) Number of patches and (b) the number of holes for three analyzing wavelets as a function of α.


(a) Morlet
(b) Morlet
(c) Paul
(d) Paul
(e) Morlet
(f) Dog
H0
Relative Frequency
0.4
0.2
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Pointwise Significance Level
0 0.2 0.4 0.6 0.8 1
Persistent Index

1



~~Figure 6. (a) Example barcodes for H_0 corresponding to a wavelet power spectrum obtained from the (a) Morlet, (b) Paul, and (c) Dog wavelets. A red-noise process with length 150 and lag-1 autocorrelation coefficient equal to 0.5 was used to create the barcodes. Distribution of persistence indices representing the lifetimes of patches for the (b) Morlet, (d) Paul, and (f) Dog wavelets. Distribution was obtained by generating 1000 wavelet power spectra of red-noise processes with lengths 500 and lag-1 autocorrelation coefficients equal to 0.5. Essential classes have been removed from the distributions.~~

Figure 6. Example barcode for H_0 corresponding to a wavelet power spectrum of a red-noise process with length 150 and $\rho = 0.5$.

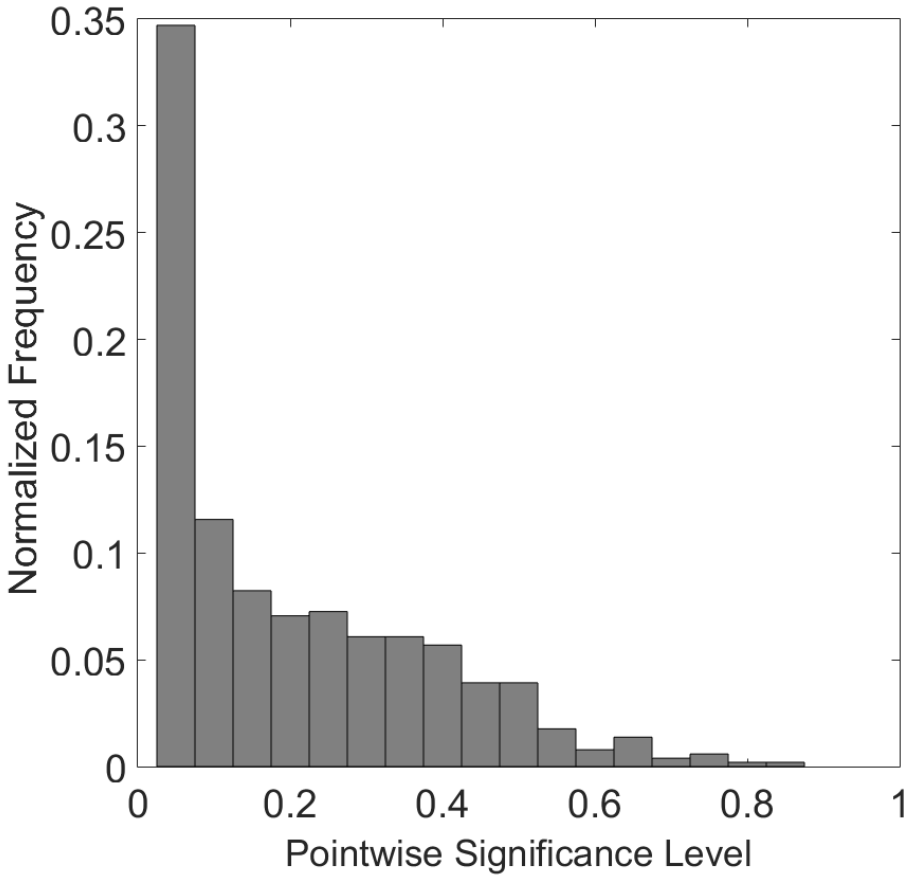


Figure 7. Distribution of persistence indices representing the lifetimes of patches. The distribution was obtained by generating 1000 wavelet power spectra of red-noise processes with lengths 500 and $\rho = 0.5$. Persistence indices equal to infinity are excluded from the distribution.

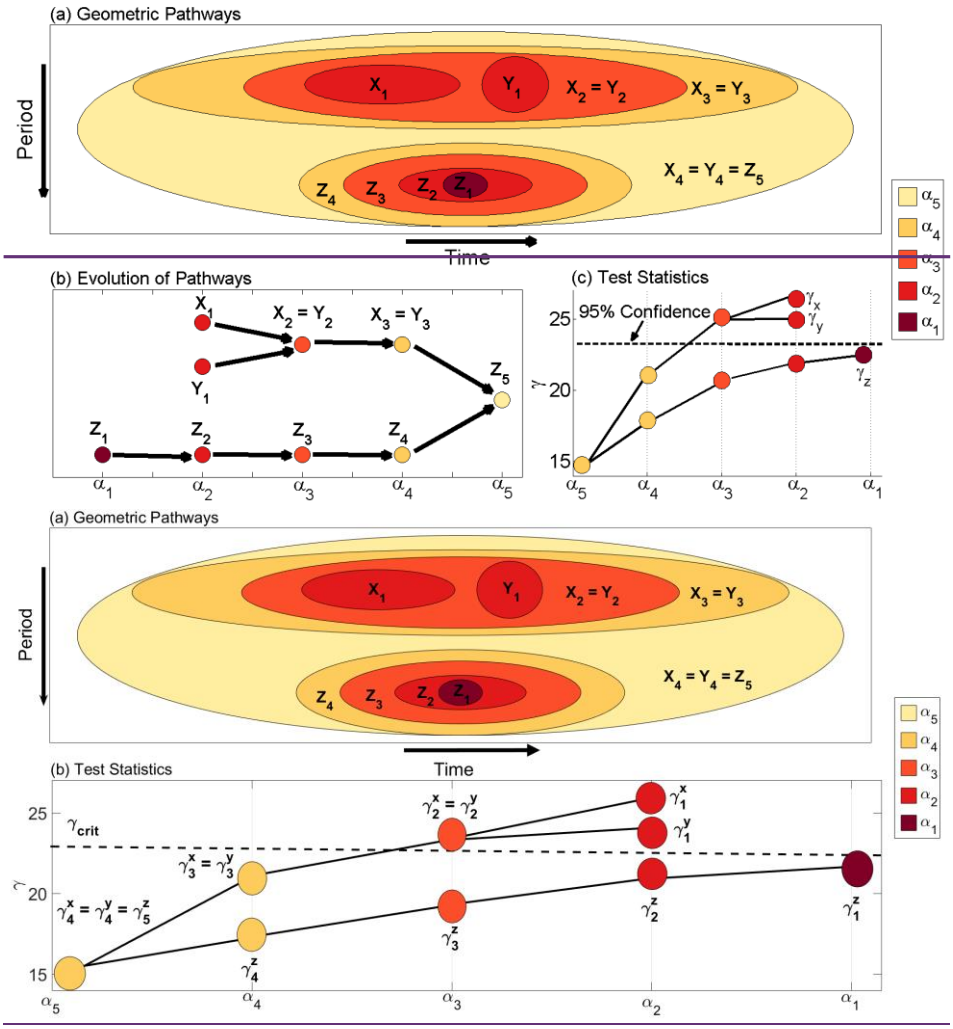
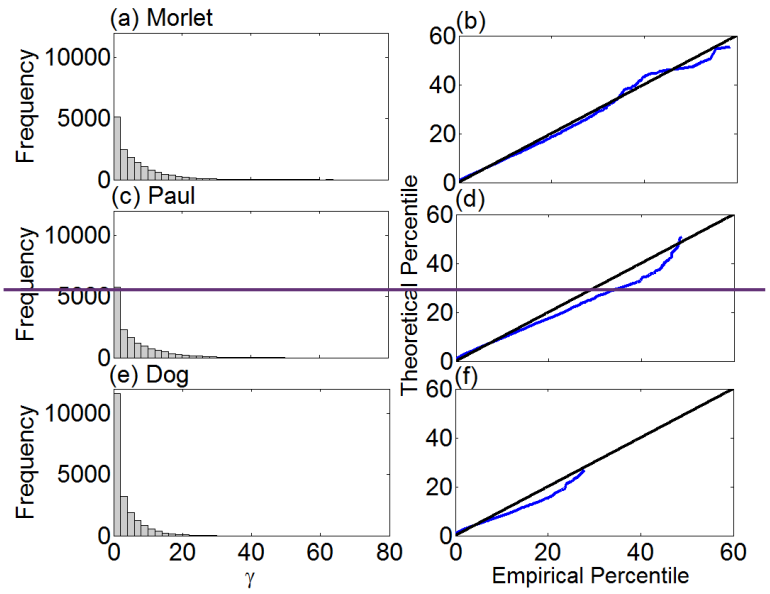


Figure 7. (a) Geometric pathway of three significance patches, X, Y, and Z in the interval $I = [\alpha_1, \alpha_5]$. (b) The geometric evolution of the pathways showing how Z_5 was created from the merging of X_3 and Z_4 as α changed from α_4 to α_5. (c) The cumulative areas at each step of pathway for each geometric pathway, where the summation begins at α_5, the initial point of the pathway, and the dotted line represents the critical level of cumulative areawise test.

1



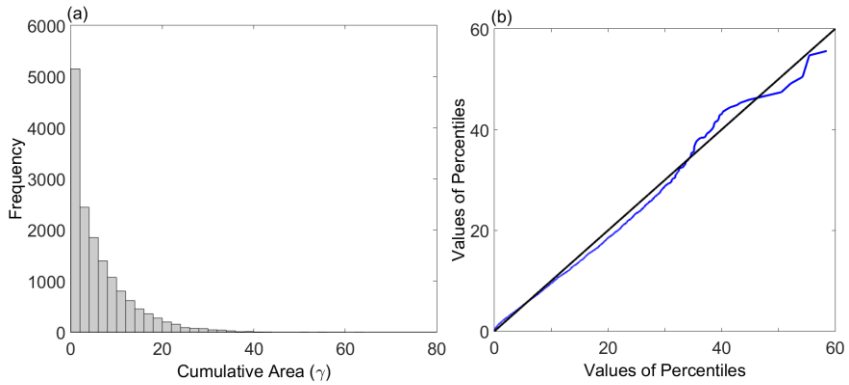


Figure 8. (a) Null distribution of γ for the Morlet wavelet obtained by generating 10,000 geometric pathways under the null hypothesis of red-noise, where the red-noise processes were of length 1000 and had lag-1 autocorrelation coefficients equal to 0.5. (b) Percentiles of a theoretical

exponential distribution with mean 6.5 plotted as a function of the percentiles calculated from the distribution shown in (a). ~~(c) – (d) Same as (a) – (b) but for the Paul wavelet. (e) – (f) Same as (a) – (b) except for the Dog wavelet.~~

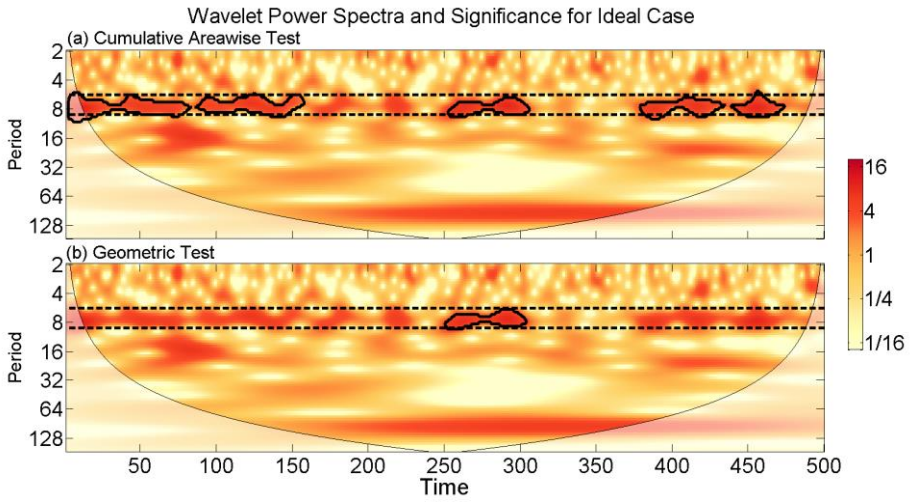
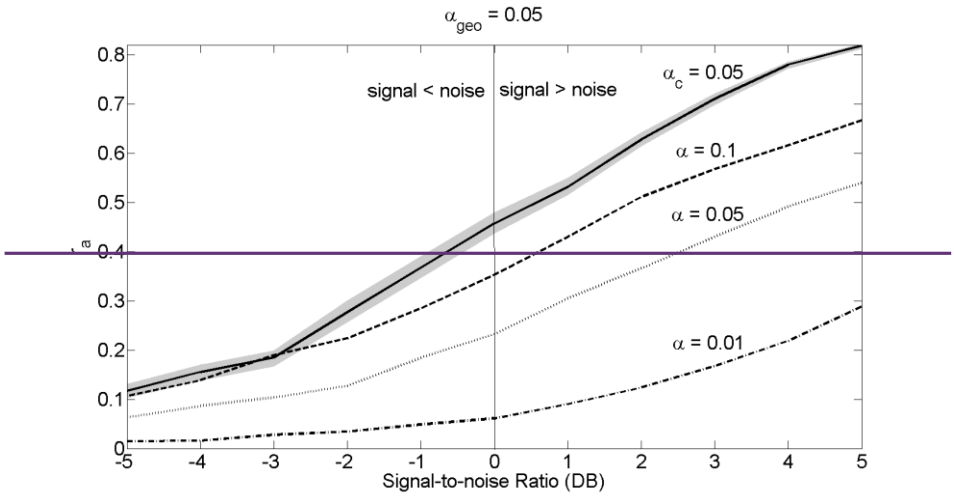


Figure 9. (a) Cumulative areawise test applied to a sinusoid with a frequency of 0.8 and amplitude equal to 0.8. Signal-to-noise ratio is 1.0. Contours represent patches that are elements of 5% significant pathways. Dotted lines represent the upper and lower boundaries of a theoretical patch obtained by generating the wavelet power spectrum of a pure sine wave and calculating the width of the patch at t = 250. (b) Same as (a) except for the geometric test with $\alpha = 0.05$ and $\alpha_{geo} = 0.05$. Contours represent patches that are geometrically significant.


α_{geo} = 0.05
signal < noise
signal > noise
α_c = 0.05
α = 0.1
α = 0.05
α = 0.01
r_a
Signal-to-noise Ratio (DB)

1

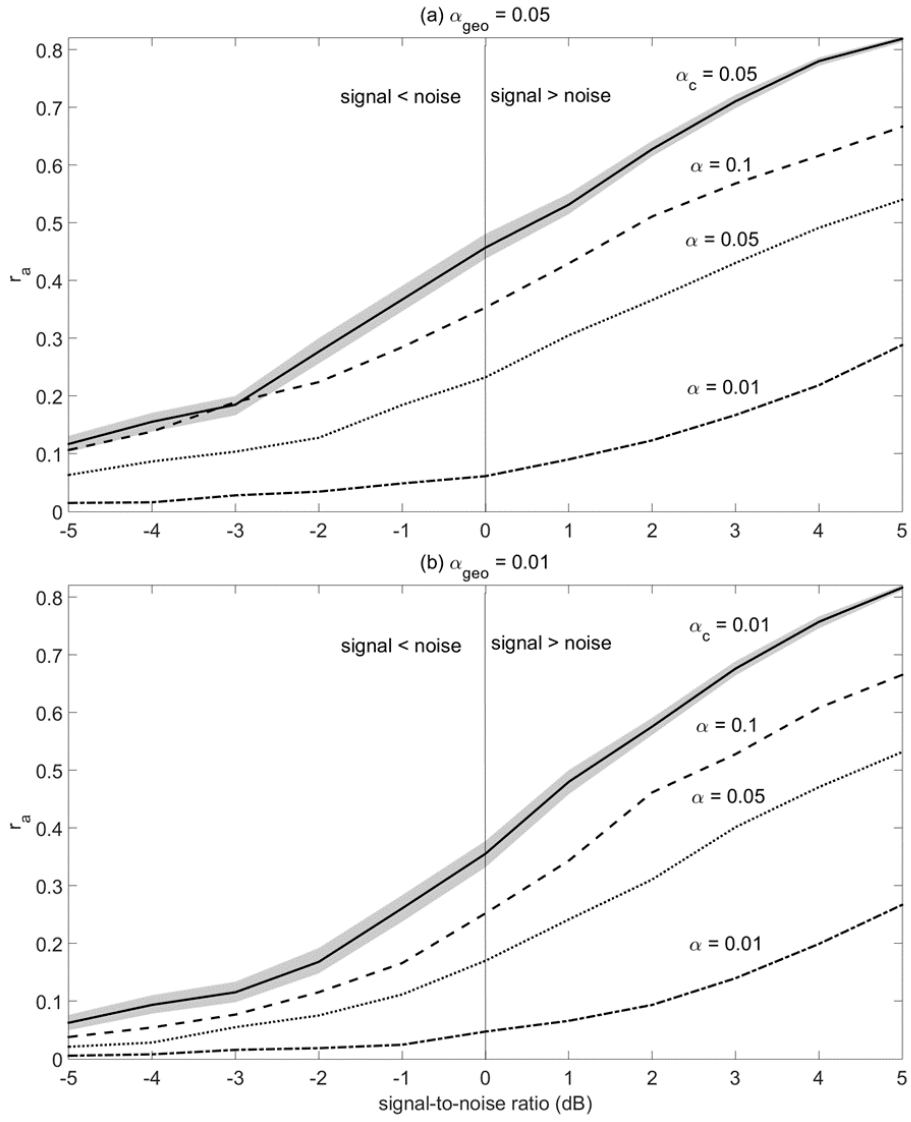


Figure 1~~0~~. The ensemble mean r_a as a function of the signal-to-noise to ratio for the areawise test with $\alpha_c = 0.05$ and the geometric test with $\alpha_{geo} = 0.05$. Gray shading represents the 95% confidence interval and all means for the geometric test are significantly different at the 5% level from the means for the areawise test except for those corresponding to the $\alpha = 0.01$ curve for

signal-to-noise ratios less than -1.5. The confidence intervals and statistical significance were obtained by the bootstrap method (Efron, 1979). The data for each signal-to-noise-ratio were sampled with replacement 1000 times to generate a distribution of bootstrap replicates, from which 95% confidence intervals were obtained. Two ensemble means were said to significantly different at the 5% level if their 95% confidence intervals did not intersect.

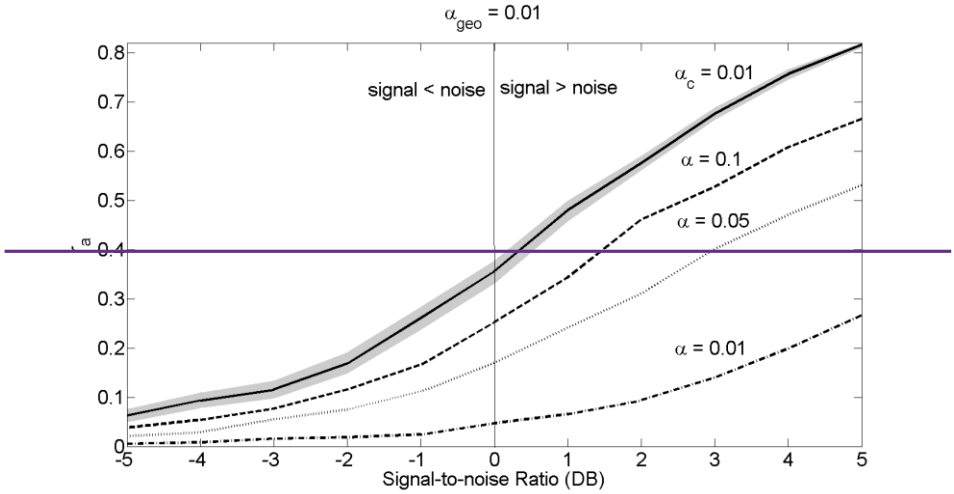


Figure 11. Same as Figure 10 except with $\alpha_c = 0.01$ and $\alpha_{geo} = 0.01$. All means for the geometric test are significantly different at the 5% level from the means for the areawise test

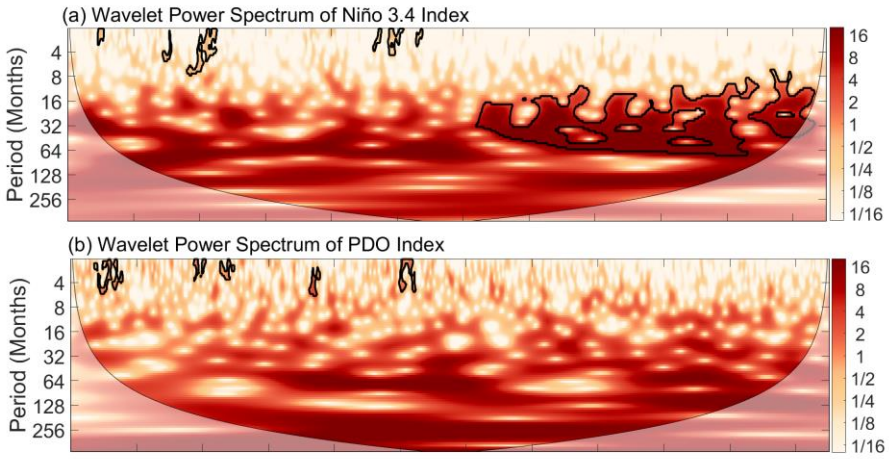


Figure 12. (a) The application of the cumulative areawise test to the AMO index (a) the NAO index, (c) the Niño 3.4 index, and (d) the PDO index. α_c was set to 0.01 in all cases and contours enclose regions of 1% areawise significance.