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.

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. A widely used method is wavelet analysis, which has proven useful in numerous geophysical investigations (Higuchi et al., 2003; Olsen et al., 2012; Meyers et al., 1993; Lee and Lwiza, 2008; Whitney, 2010; Wilson et al., 2014 Labat, 2004; Labat, 2008; Grinsted et al., 2004; Velasco and Mendoza, 2008, Schulte et al., 2016).

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 suffer from a binary decision because one must choose both a pointwise significance level together with an 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. 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. The data used in applications of the significance tests are described in Section 2 and a brief description of wavelet analysis is provided in Section 3. In Section 4, a review of existing statistical testing procedures is presented. The sensitivity of the geometric test to the chosen pointwise significance level is quantified in Section 5. The topological properties of red noise are analyzed in Section 6 and the cumulative areawise test is developed in Section 7. A comparison of the test in terms of statistical power to the existing geometric test is provided in Section 8. Applications of the test to prominent climate indices are presented in Section 9 and are followed by concluding remarks in Section 10.

2. Data

The Niño 3.4 index data from 1900-2014 were obtained from the National Center for Atmospheric Research. This index quantifies the strength of the El-Niño/Southern Oscillation (ENSO) and is defined as sea surface temperature (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 were obtained from University of Washington (http://research.jisao.washington.edu/pdo/PDO.latest) and describe detrended SST variability in the North Pacific poleward of 20°N latitude (Mantua and Hare, 2002).

3. Wavelet Analysis

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 is the Morlet wavelet, the focus of this paper, which 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. Throughout the paper, ω_0 = 6. The wavelet power spectrum is given by $|W_n(s)|^2$.

4. Existing Significance Tests

4.1 Pointwise Significance Test

Consider a first-order autoregressive (Markov) process

$$x_n = \rho x_{n-1} + w_n, \quad (3)$$

where ρ is the lag-1 autocorrelation coefficient and w_n is Gaussian white noise with $x_0 = 0$. The normalized theoretical power spectrum of the process is given by

$$p_f = \frac{1-\rho^2}{1+\rho^2-2\rho\cos(2\pi f/N)}, \quad (4)$$

where f = 0, … , 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. (4) 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. Such clusters will be referred to as patches.

4.2 Areawise and Geometric Significance Tests

The areawise test developed by Maraun et al. (2007) takes advantage of how correlations between adjacent wavelet coefficients arising from the reproducing kernel produce patches that resemble the reproducing kernel. For 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 greater area are more statistically significant. The estimation of the critical level of the test involves a root-finding algorithm that is computational inefficient. To remedy the computational drawback, Schulte et al. (2015) developed a geometric test that makes use of a normalized area. The normalized area allows patches at different scales to be compared simultaneously. The estimation of the critical level of the test is achieved simply through 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.

4.3 *Application of Existing Significance Tests*

Shown in Figure 2 is the wavelet power spectrum of the Niño 3.4 index. Large 5% patches were found and the largest was located in the time period 1950-2014 and in the period band 16-32 months. The large patches after 1950 were also found to be 5% geometrically significant (thick contours) and subsets of the patches were also found to be 5% areawise significant (blue shading).

Both the areawise and geometric tests identified few patches in the period band of 2 to 4 months as statistically significant. For the wavelet power spectrum of the PDO index, a large patch centered at a period of 512 months extending from 1910 to 1980 was detected. Most of the patches, however, were located at periods less than 8 months, time scales not typically associated with the PDO. A few patches were identified as areawise and geometrically significant and such patches were in the 2 to 8 month period band.

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

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

$$r = \frac{N_{\alpha_1,\alpha_2}}{N_{\alpha_1}}. \quad (5)$$

The quantity N_{α_1,α_2} is the number of geometrically significant patches at the pointwise significance level α_1 that are also geometrically significant at the pointwise significance level α_2. N_{α_1} is the number of patches at α_1 that are geometrically significant at the level 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 ρ =0.5 under four scenarios. 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 for 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}.

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 is 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 $\bar{r} = 0.22$. 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. Persistent Topology

6.1 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 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

$$[x] = \{y \in P: y \sim x\}. \tag{6}$$

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

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

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.

6.2 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 the number of patches at a particular pointwise significance level. Shown in Figure 5 is the ensemble mean of the number of patches as a function of α. The curve was obtained by 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. The number of patches reached minima at $\alpha = 0.01$ and $\alpha = 0.99$ and a maximum at $\alpha = 0.18$.

To understand more fully the curve shown in Figure 5, the persistent homology of patches generated from red-noise processes of length 150 was computed as α varied from 0.01 to 0.99. Barcodes representing the evolution of patches (homology classes) in the wavelet power spectra were also computed. In each case, ρ = 0.5, but the results are identical for other autocorrelation coefficients. Shown in Figure 6 is a barcode corresponding to a typical wavelet power spectrum of a red-noise process. 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 how 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 $\alpha = 0.7$, at which point all patches have merged into a single patch. To show that the distribution of persistence indices 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 spectrum were computed (Figure 7). The resulting distribution indicates that persistence indices are typically 0.01 and relatively few patches live longer than 0.6. Overall, the distribution characterizes patches generated from red-noise processes as short-lived.

7. Development of the cumulative areawise test

7.1 Geometric Pathways

The first step of the cumulative areawise test is to define the geometric evolution of a patch across a finite set of pointwise significance levels. The notion of evolution will be made precise by introducing the concept of a geometric pathway, which is defined as a collection $\mathcal{P}$ of L patches at the corresponding pointwise significance levels $\alpha_1 < \alpha_2 < \ldots < \alpha_L$ such that

$$P_1 \subset P_2 \subset P_3 \subset \cdots \subset P_L \tag{8}$$

and

$$g_1 < g_2 < g_3 \ldots < g_L, \tag{9}$$

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 (9) are guaranteed to hold for any nested sequence (8) (Appendix A). The length of a pathway will be given by L the number of elements in the pathway. The interval $I = [\alpha_{min}, \alpha_{max}]$ will be called the computation interval and the discrete spacing between adjacent pointwise significance levels, $\Delta\alpha$, will be be referred to as the resolution.

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.

7.2 Test Construction

One can associate to each geometric pathway a test statistic, which will be the total sum of normalized areas

$$\gamma = \sum_{j=1}^{L} g_j. \tag{10}$$

The calculation of the critical level for the test can be computed using Monte Carlo methods by first fixing I and $\Delta\alpha$. Secondly, one generates red-noise processes with the same autocorrelation coefficients as the input time series and calculates synthetic wavelet power spectra corresponding to each red-noise process. The final step is to compute γ for every pathway. The calculation results in a null distribution from which the desired critical level of the test can be obtained. The critical level corresponding to the 5% significance level of the test, as an example, is the 95th percentile of the null distribution.

A null distribution calculated for a red-noise process with ρ = 0.5 is shown in Figure 8. In the experiments, $\alpha_{min} = 0.02$ and $\alpha_{max} = 0.82$ with $\Delta\alpha = 0.02$. The distribution of γ is generally similar to the shape of the distribution for the persistence indices for H_0 (Figure 7), where the smallest values of γ are preferred. It turns out that the distribution of γ 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 95th percentiles only differing by 1.0.

Associated with each element of the geometric pathway is the quantity

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

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

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

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

$$p_1 < p_2 < p_3 \ldots < p_L \tag{13}$$

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).

7.3 Application to Ideal Pathways

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

$$X_1 \subset X_2 \subset X_3 \subset X_4. \tag{14}$$

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

$$\gamma_1^x = g_1^x + g_2^x + g_3^x + g_4^x \tag{15}$$

$$\gamma_2^x = g_2^x + g_3^x + g_4^x, \tag{16}$$

$$\gamma_3^x = g_3^x + g_4^x, \tag{17}$$

and

$$\gamma_4^x = g_4^x \tag{18}$$

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

$$X_{sig} = X_2 = X_1 \cup X_2. \tag{19}$$

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

$$Y_{sig} = X_2 = Y_1 \cup Y_2. \tag{20}$$

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

$$Z_1 \subset Z_2 \subset Z_3 \subset Z_4 \subset Z_5. \tag{21}$$

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

$$\gamma_1^z = g_1^z + g_2^z + g_3^z + g_4^z + g_5^z, \tag{22}$$

$$\gamma_2^z = g_2^z + g_3^z + g_4^z + g_5^z, \tag{23}$$

$$\gamma_3^z = g_3^z + g_4^z + g_5^z, \tag{24}$$

$$\gamma_4^z = g_4^z + g_5^z, \tag{25}$$

and

$$\gamma_5^z = g_5^z. \tag{26}$$

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

$$X_2 = X_{sig} \cup Y_{sig}. \tag{27}$$

8. Comparison with the geometric test

8.1 True Positive 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

$$x(t) = A\sin(2\pi f t) + w(t) \tag{28}$$

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 cumulative 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 spectrum 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 is indicated by dotted lines in Figure 10, where the theoretical patch is a rectangle of fixed width extending from $t = 0$ to $t = 500$. In all experiments, $\alpha_{max} = 0.18$ and $\Delta\alpha = 0.02$, but implications of other choices are discussed at the end of the section.

Let P_{geo} be the union of all pointwise significance patches at α that are geometrically significant at the α_{geo} level and let P_{theory} be the theoretical patch. Then

$$r_a = \frac{A_{P_{geo} \cap P_{theory}}}{A_{P_{theory}}} \tag{29}$$

represents the areal fraction of P_{theory} detected by the geometric test. In Equation (29), $A_{P_{geo} \cap P_{theory}}$ 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 patch to be significant. A similar construction can be made for the cumulative areawise test by replacing P_{geo} with P_c. Figure 9a illustrates the procedure for the cumulative areawise test when f = 0.8, A = 1.0 and the signal-to-noise ratio (defined below) σ = 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 σ to determine if the tests truly perform differently. The signal-to-noise ratio is defined as

$$\sigma = 10\log\left(\frac{p_{signal}}{p_{noise}}\right), \tag{30}$$

where

$$p_{signal} = \frac{A^2}{2}, \tag{31}$$

p_{noise} is the average power of the Gaussian white noise, and σ is measured in decibels (dB). 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 cumulative areawise significance level (denoted by α_c, herreafter) was set to 0.05, α_{geo} = 0.05, and α = 0.01, 0.05, 0.1. The value of σ was varied from 5 dB to 5 dB. The results are shown in Figure 11a. 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. Differences between the test performances became larger as σ was increased and the cumulative areawise test outperformed the geometric test regardless of the chosen pointwise significance levels when $\sigma \geq -2.5$ dB. A second experiment was conducted using α_c = 0.01 and α_{geo} = 0.01 (Figure 11b). The results were found to be similar to that of the first experiment except that r_a was generally smaller for both tests. The result is consistent with how the significance levels of the tests were increased. The results indicate that the cumulative 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 3, the performance of the geometric test depended strongly on the chosen pointwise significance level, especially when the signal power was high.

Additional experiments were performed using different values of f. True positive detection, for a fixed σ, was generally found to increase for larger f, though the cumulative areawise test was still found to detect more true positives. Additionally, it was found that true positive detection changed little if $\Delta\alpha$ was set to a value less than 0.02. Setting $\Delta\alpha$ to be greater than 0.03 was

generally found to result in a decrease in true positive detection. On the other hand, true positive detection increased dramatically as α_{max} increased but with the caveat that the areas of spurious patches found outside the theoretical patch were found to be larger.

8.2 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

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

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

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

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

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

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 $\alpha_c = 0.01$.

9. Climate applications

The cumulative areawise test was applied to the wavelet power spectra of the PDO and Niño 3.4 indices 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 Niño 3.4 index indicates potential predictive capabilities (Figure 12a). 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 cumulative 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 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 12a show that neither the decadal nor the multi-decadal variability exceeded a red-noise background so modulations would be difficult to predict.

The wavelet power spectrum of the PDO index is shown in Figure 12b. There is enhanced variance at multi-decadal time scales but the variance does not exceed a red-noise background. Cumulative 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).

10. Conclusions

A cumulative 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 two significance levels. 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. 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 cumulative areawise test can be applied to wavelet power spectra obtained using other analyzing wavelets such as the Paul and Dog wavelets. Moreover, the results presented for the Morlet wavelet were found to be generally similar to those for the Paul and Dog wavelets. It is recommended, however, that different null distributions be calculated for the different analyzing wavelets.

The cumulative areawise test can also be applied to wavelet coherence, wavelet partial coherence, and multiple wavelet coherence spectra. In these cases, the critical levels of the pointwise test would need to be calculated using Monte Carlo methods. The implementation of the cumulative areawise test, however, is exactly the same as for wavelet power spectra. It is noted that different null distributions for the cumulative areawise test should be used for each, as the cumulative areas of patches in coherence spectra may differ from those found in wavelet power spectra.

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 PDO is limited and that the multi-decadal variability of the 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.

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

Appendix A

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

$$A = A_1 + A_2, \tag{A1}$$

$$A = r_1 A + r_2 A, \tag{A2}$$

and

$$r_2 = 1 - r_1, \tag{A3}$$

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

$$\frac{1}{A}\iint_P sdsdt = \frac{1}{A}\iint_{P_1} sdsdt + \frac{1}{A}\iint_{P_2} sdsdt \tag{A4}$$

$$= \frac{1}{A}\iint_P sdsdt = \frac{r_1}{A_1}\iint_{P_1} sdsdt + \frac{r_2}{A_2}\iint_{P_2} sdsdt \tag{A5}$$

or

$$C^s = r_1 C_1^s + r_2 C_2^s, \tag{A6}$$

so that

$$\frac{C^s - r_2\, C_2^s}{r_1} = C_1^s, \tag{A7}$$

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

$$C^S - r_2\, C_2^s > 0 \tag{A8}$$

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

$$A^N = \frac{A}{(C^s)^2} \tag{A9}$$

and

$$A_1^N = \frac{A_1}{(C_1^s)^2}. \tag{A10}$$

Thus,

$$r_{norm} = \frac{A_1^N}{A^N} = \frac{r_1^3 (C^s)^2}{(C^s - (1 - r_1) C_2^s)^2}. \tag{A11}$$

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.

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.

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.

Lipschutz, S.: Theory and Problems of General Topology. Schaum's Outline. McGraw-Hill. pp. 238, 1965.

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.

Schulte, J. A, Najjar, R. G., and Li, M: Climate mode impacts on streamflow in the mid-Atlantic region of the United States, 5, 80–99, 2016.

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.

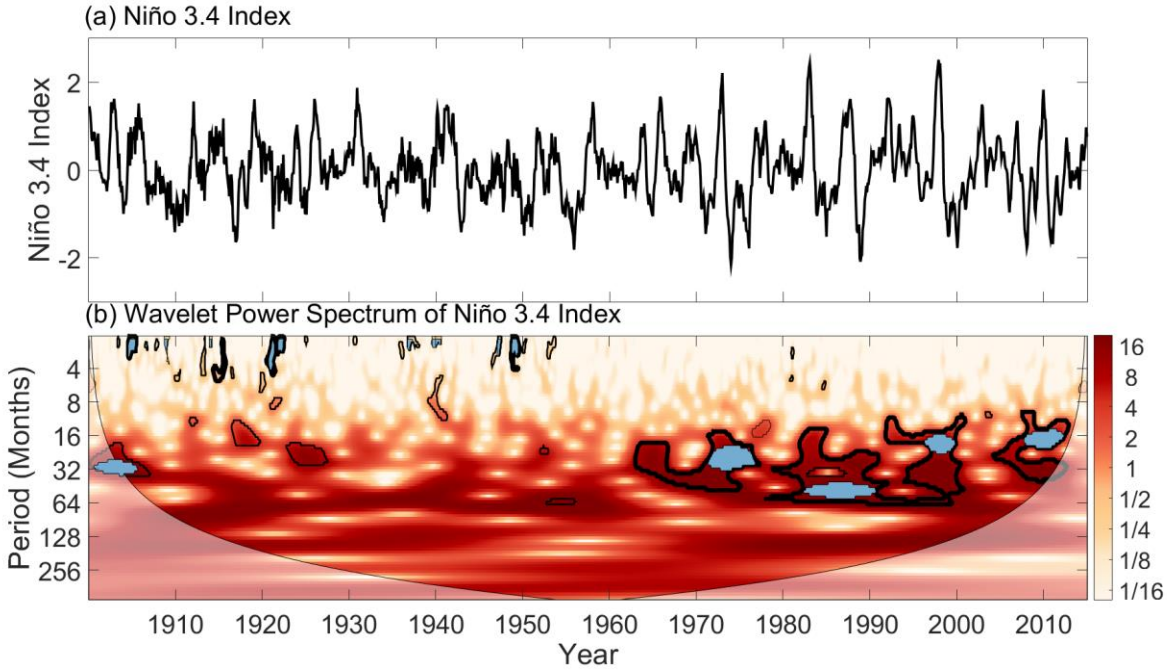

Figure 1. Wavelet power spectra of the Niño 3.4 index. Thin black contours enclose regions of 5% pointwise significance and thick contours indicate those patches that are geometrically significant at the 5% level. Light blue shading represents 5% areawise significant subsets of the patches. Light shading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

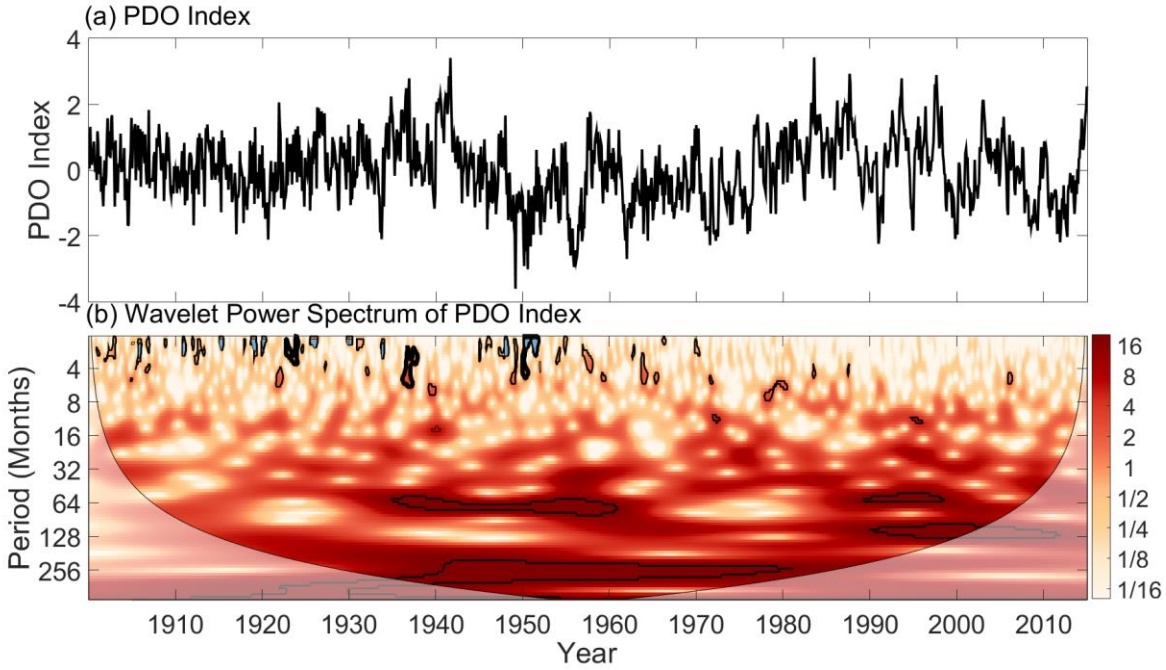

Figure 2. Same as Figure 1 but for the PDO index.

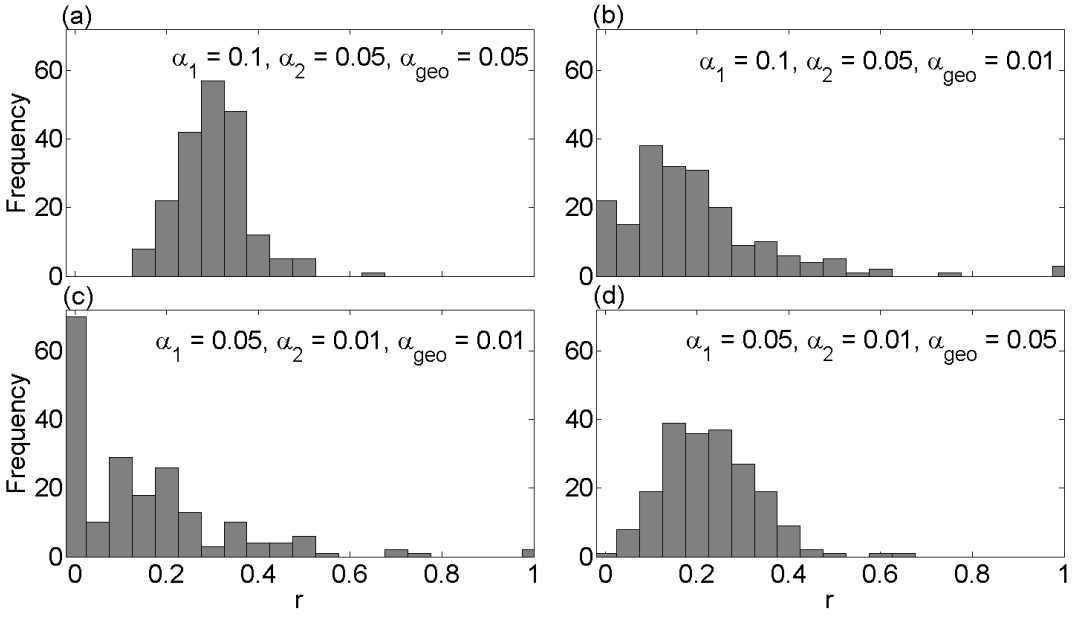

Figure 3. (a) A histogram of $r = \frac{N_{\alpha_1,\alpha_2}}{N_{\alpha_1}}$ for $\alpha_1 = 0.1$, $\alpha_2 = 0.05$, and $\alpha_{geo} = 0.05$ obtained from the generation of 300 wavelet power spectra of red-noise processes of length 1000 with lag-1 autocorrelation coefficients equal to 0.5. (b) Same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

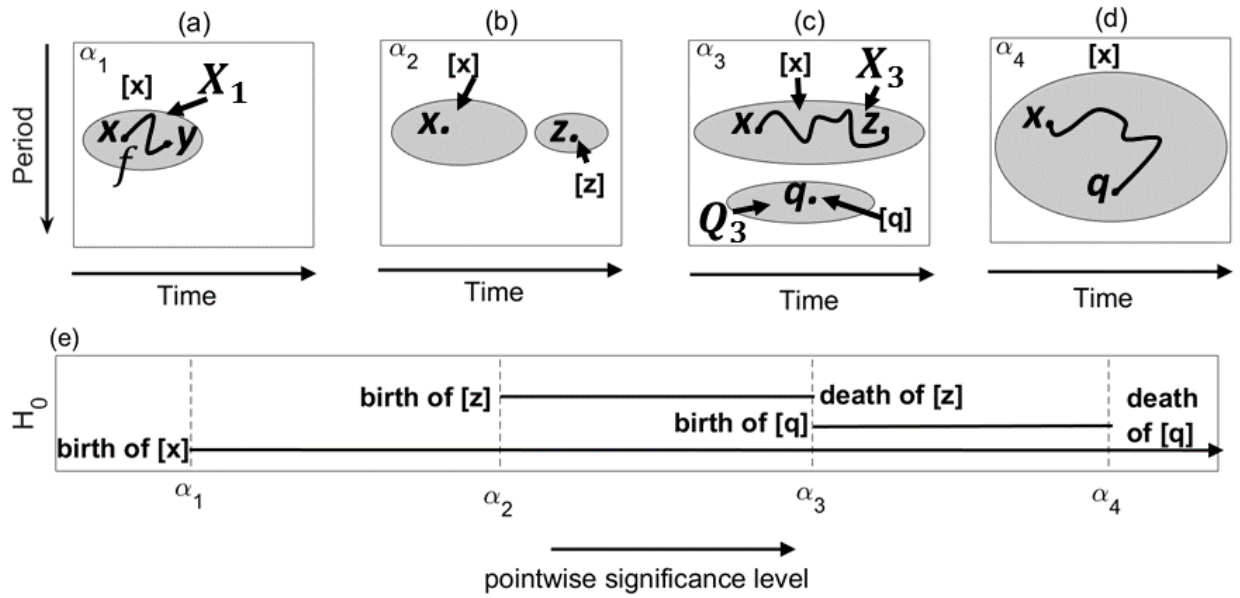

Figure 4. (a) - (d) The topological evolution of patches across four pointwise significance levels. (e) The barcode showing the birth and death of patches throughout the evolution process. Horizontal lines with arrows indicate those patches that never die.

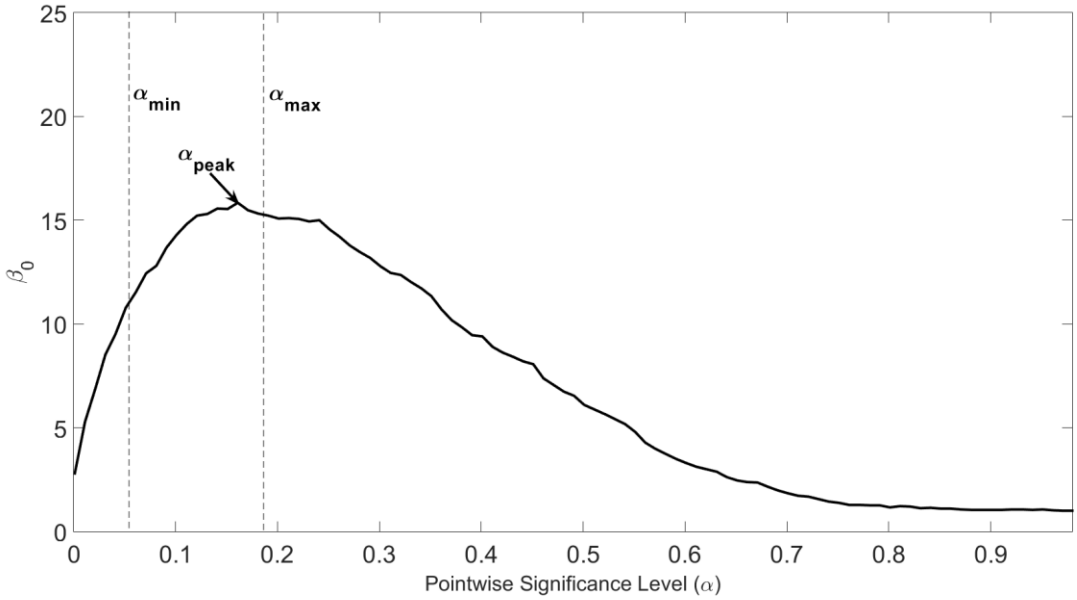

Figure 5. The mean number of patches as a function of α. The curve was obtained by generating 100 wavelet power spectra of red-noise processes of length 300 and computing β_0 for each of the wavelet power spectra at each pointwise significance level. The quantities α_{min} and α_{max} are the lower and upper bounds of the computation interval for the cumulative areawise test and α_{peak} is the pointwise significance level for which the number of patches reaches a maximum within the computation interval.

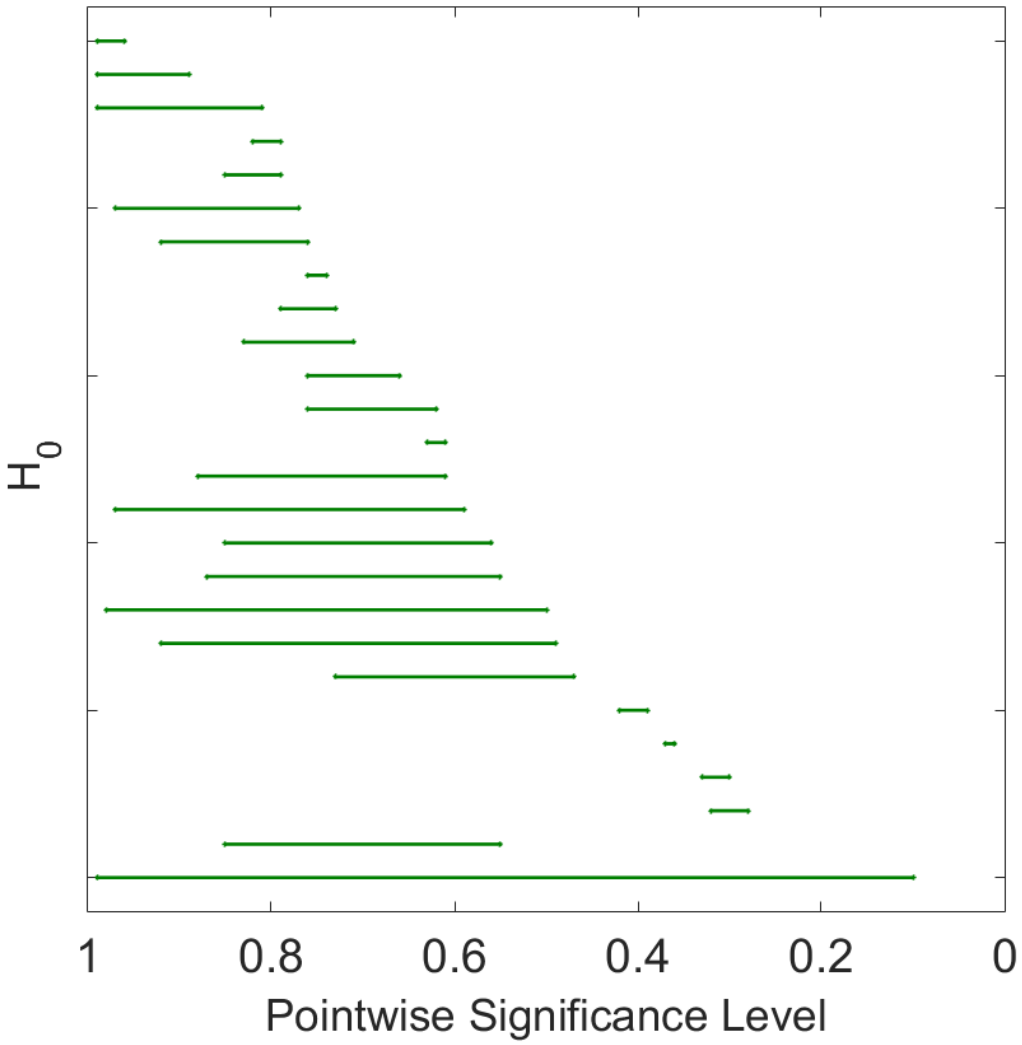

Figure 6. Example barcode for H_0 corresponding to a wavelet power spectrum of a red-noise process with length 150 and $\rho = 0.5$.

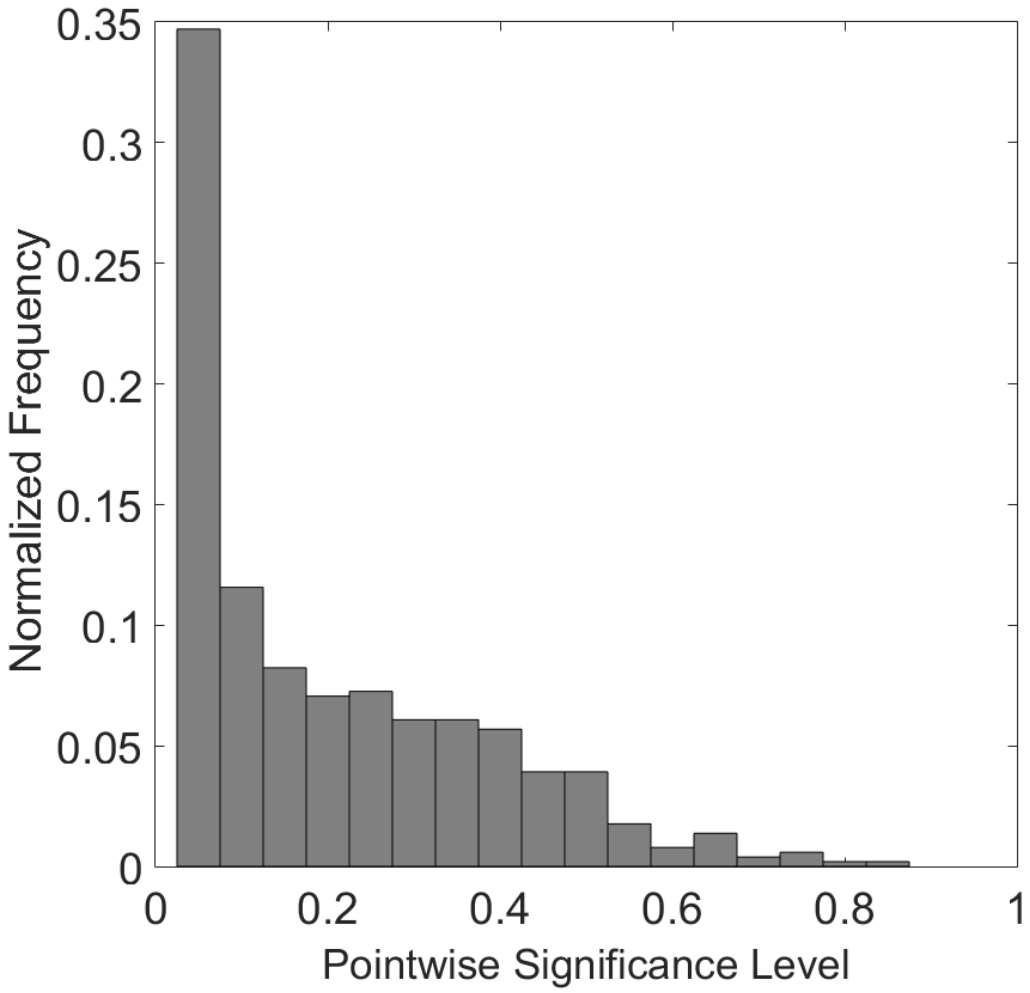

Figure 7. Distribution of persistence indices representing the lifetimes of patches. The distribution was obtained by generating 1000 wavelet power spectra of red-noise processes with lengths 500 and $\rho = 0.5$. Persistence indices equal to infinity are excluded from the distribution.

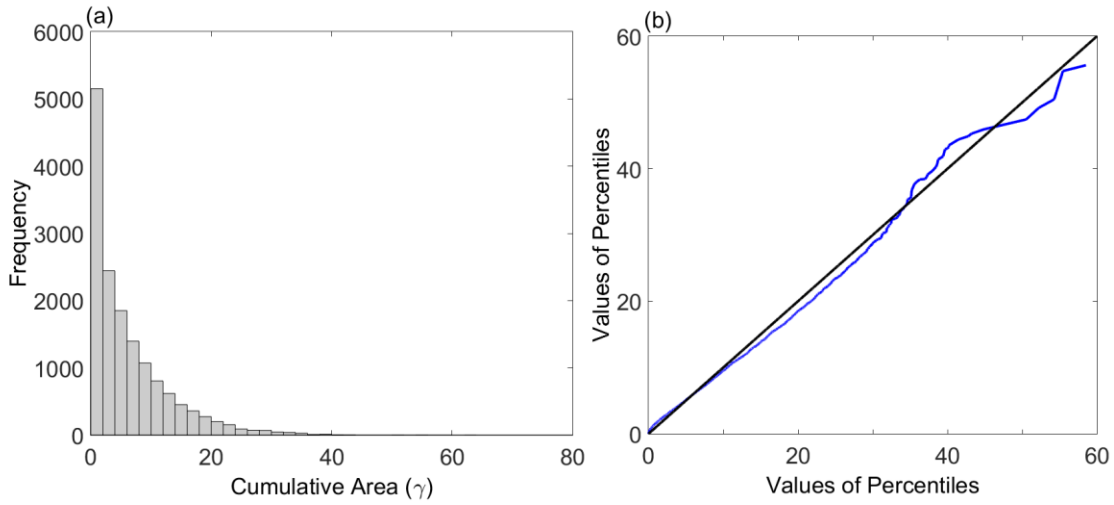

Figure 8. (a) Null distribution of γ obtained by generating 10,000 geometric pathways under the null hypothesis of red-noise, where the red-noise processes were of length 1000 and had lag-1 autocorrelation coefficients equal to 0.5. (b) Percentiles of a theoretical exponential distribution with mean 6.5 plotted as a function of the percentiles calculated from the distribution shown in (a).

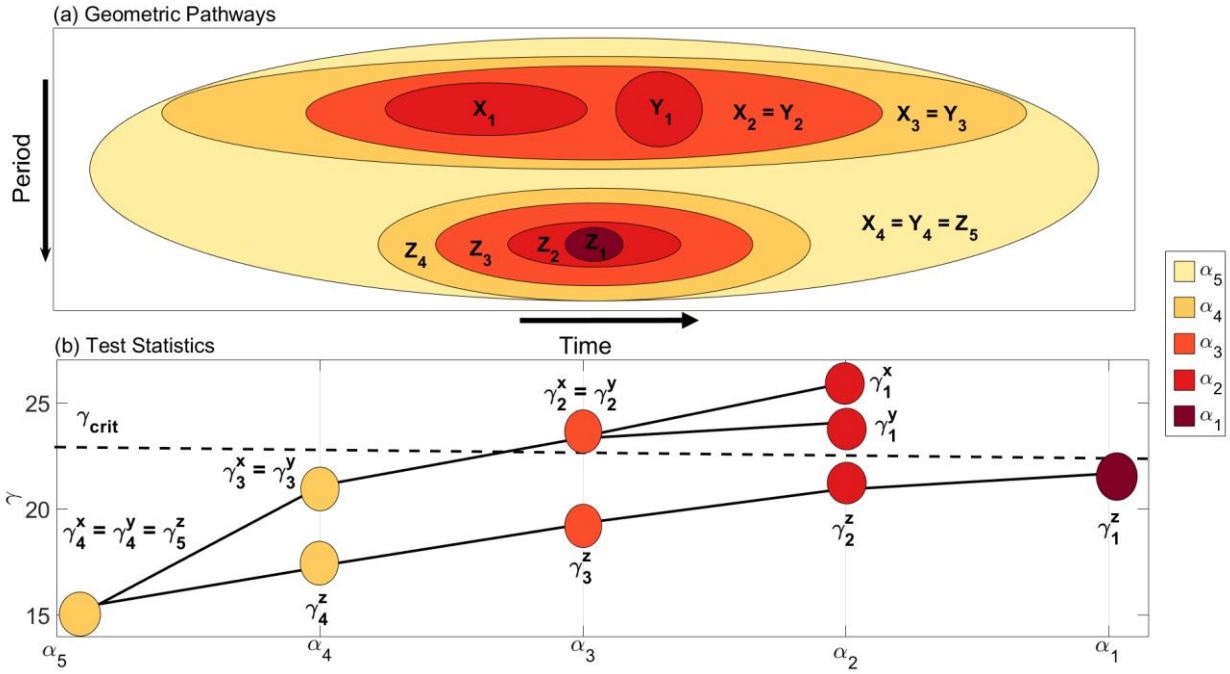

Figure 9. (a) Geometric pathway of three significance patches, X, Y, and Z in the interval $I = [\alpha_1, \alpha_5]$. (b) The geometric evolution of the pathways showing how Z_5 was created from the merging of X_3 and Z_4 as α changed from α_4 to α_5. (c) The cumulative areas of geometric pathway elements, where the summation begins at α_5, and the dotted line represents the critical level of the cumulative areawise test.

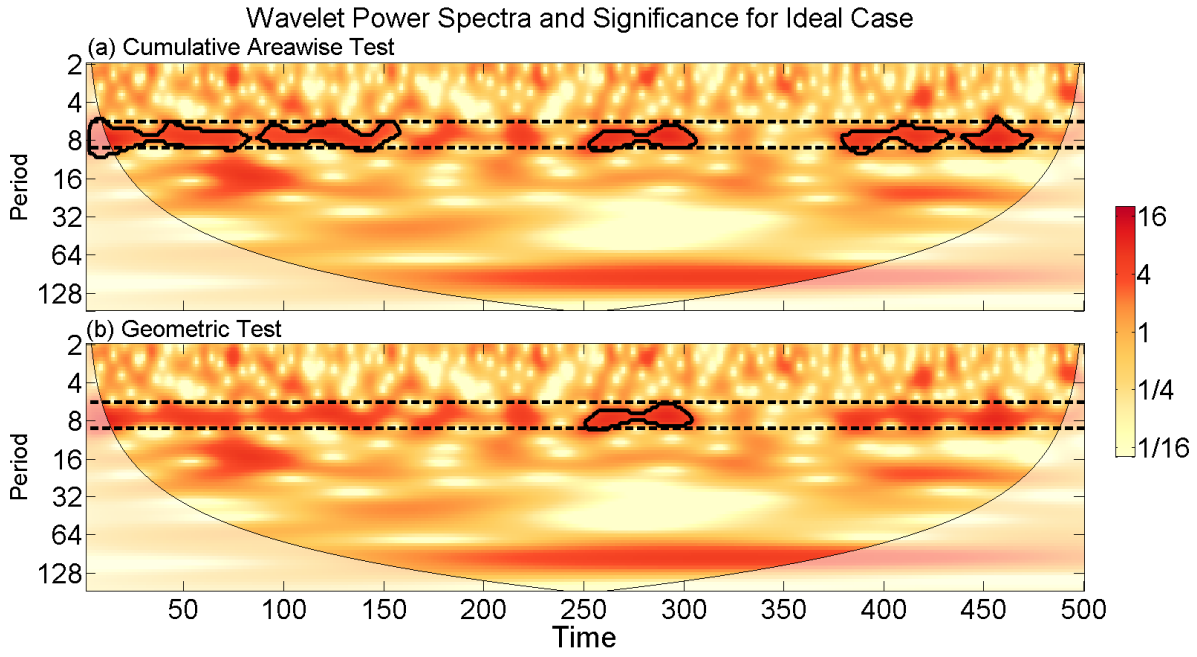

Figure 10. (a) Cumulative areawise test applied to a sinusoid with a frequency of 0.8 and amplitude equal to 0.8. Signal-to-noise ratio is 1.0. Contours represent patches that are elements of 5% significant pathways. Dotted lines represent the upper and lower boundaries of a theoretical patch obtained 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.

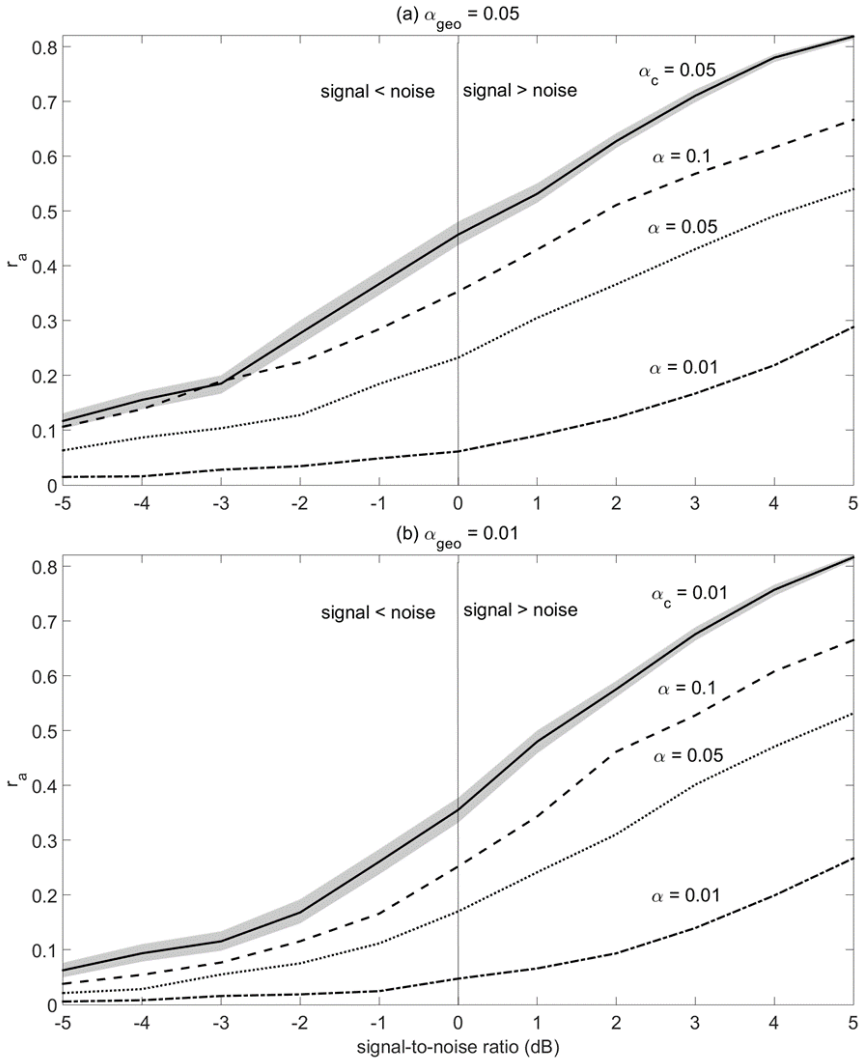

Figure 11. (a) The ensemble mean r_a as a function of the signal-to-noise to ratio for the areawise test with $\alpha_c = 0.05$ and the geometric test with $\alpha_{geo} = 0.05$. Gray shading represents the 95% confidence interval and all means for the geometric test are significantly different at the 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. (b) Same as (a) except with $\alpha_c = 0.01$ and $\alpha_{geo} = 0.01$. All means for the geometric test in panel (b) are significantly different at the 5% level from the means for the areawise test.

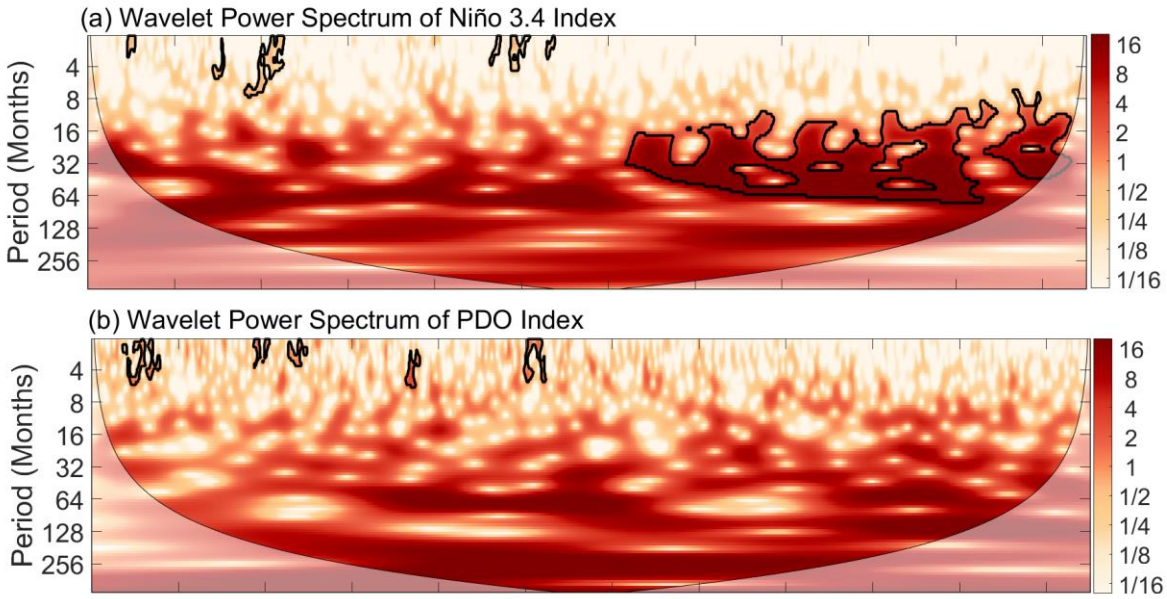

Figure 12. The application of the cumulative areawise test to the (a) Niño 3.4 index and (b) PDO index. Contours enclose regions of 1% cumulative areawise significance.