Geometric and topological approaches to significance testing in wavelet analysis

J. A. Schulte[1], C. Duffy[2], and R. G. Najjar[1]

[1] Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

[2] Department of Civil Engineering, The Pennsylvania State University, University Park, Pennsylvania

Correspondence to: J. A. Schulte (jas6367@psu.edu)

Abstract

Geometric and topological methods are applied to significance testing in the wavelet domain. A geometric test was developed for assigning significance to pointwise significance patches in local wavelet spectra, contiguous regions of significant wavelet power coefficients with respect to some noise model. This geometric significance test was found to produce results similar to an existing areawise significance test while being more computationally flexible and efficient. The geometric significance test can be readily applied to pointwise significance patches at various pointwise significance levels in wavelet power and coherence spectra. The geometric test determined that features in wavelet power of the North Atlantic Oscillation (NAO) are indistinguishable from a red-noise background, suggesting that the NAO is a stochastic, unpredictable process, which could render difficult the future projections of the NAO under a changing global system. The geometric test did, however, identify features in the wavelet power spectrum of an El Niño index (Niño 3.4) as distinguishable from a red-noise background. A topological analysis of pointwise significance patches determined that holes, deficits in pointwise significance embedded in significance patches, are capable of identifying important structures, some of which are undetected by the geometric and areawise tests. The application of the topological methods to ideal time series and to the time series of the Niño 3.4 and NAO indices showed that the areawise and geometric tests perform similarly in ideal and geophysical settings, while the topological methods showed that the Niño 3.4 time series contains numerous phase-coherent oscillations that could be interacting nonlinearly.

1. Introduction

Time series are often complex, composed of oscillations and trends. The goal of researchers is to decide whether the embedded structures in the time series are stochastic or deterministic. Such decisions can be made using Fourier analysis, with the assumption that the underlying time series is stationary (Jenkins and Watts, 1968). In many cases, however, the stationary assumption is not satisfied, making Fourier analysis an inappropriate tool for feature extraction. For non-stationary time series, wavelet analysis (Meyers, 1993; Torrence and Compo, 1998) can be used for decomposing a time series into both frequency and time components, allowing the extraction of transient features and dominant modes of variability. Once embedded structures in time series have been identified, a natural question arises: what physical mechanisms are responsible for the detected modes of variability? Linkages between the modes of variability and possible physical mechanisms can be obtained using wavelet coherence (Grinsted et al., 2004), a bivariate tool for detecting common oscillations between two time series. Together, wavelet power and coherence analyses have proven useful in climate science (Velasco and Mendoza, 2007; Muller et al., 2008), hydrology (Zhang et al., 2006; Ozger et al., 2009; Labat, 2008; Labat, 2010), atmospheric science (Terradellas et al., 2005; Schimanke et al., 2011), and oceanography (Lee and Lwiza, 2008).

The application of wavelet analysis alone is not sufficient for feature extraction of time series; indeed, random fluctuations can produce large values of spectral power or coherence related to the underlying process (e.g., red-noise) and not necessarily the time series. In Fourier analysis, one chooses a suitable noise model and assesses the significance of features relative to some analytically or empirically derived threshold. In climate science, for example, one often compares the sample power spectrum of a time series to that of a theoretical red-noise spectrum (Hasselman, 1976; Torrence and Compo, 1998). Statistical significance testing is also necessary in the wavelet domain. Torrence and Compo (1998) were the first to assess the significance of features in wavelet power spectra using discrete red-noise background spectra. Grinsted et al. (2004), using Monte Carlo methods, extended significance testing to wavelet coherence using surrogate red-noise time series. The (pointwise) significance tests developed by Torrence and Compo (2010) and Grinsted et al. (2004), however, have multiple-testing problems, given the large number of wavelet coefficients being tested simultaneously (Maraun and Kurths, 2004). Suppose, for example, that a pointwise significance test was applied to M wavelet power coefficients at the 5% significance level. Then, on average, there will be $0.05M$ false positive results, which would make the pointwise test permissive for large M. Maraun et al. (2007) addressed these problems by developing an areawise test that sorts through contiguous regions of pointwise significance called significance patches based on their area and geometry, minimizing spurious results, and thus giving researchers more insight into the time series in question. According to the areawise test, the larger the pointwise significance patch, the less likely it was generated from a stochastic fluctuation.

In this study, significance testing in the wavelet domain is improved through the following: (1) the development of a flexible and computationally efficient geometric test capable of minimizing spurious results from the pointwise test by associating p-values to individual patches in wavelet-power and wavelet-coherence spectra; and (2) the application of topological methods that can further distinguish spurious patches from true structures that can reveal information about time series undetected by current methods. Given the deficiencies of pointwise significance testing, there is a need to improve current methods of evaluating significance of features in the wavelet domain. The areawise test, though a substantial improvement from the pointwise test has one drawback: finding the significance level of the areawise test requires a complicated root-finding algorithm, making *p*-values for the areawise test difficult to obtain, as it would require the repeated application of a root-finding algorithm (see Sect. 4.1 for details).

The remainder of the paper is organized as follows. A brief overview of wavelet analysis is presented in Sect. 2. In Sect. 3, the pointwise and areawise tests are discussed briefly. The development of the geometric test is presented in Sect. 4. In Sect. 5, ideas inspired by persistence homology (Edelsbrunner, 2010) are used to show that holes, voids of pointwise significance surrounded by regions of pointwise significance, can distinguish important structures from trivial structures, linking the geometric and topological tests. Using ideas from Sect. 4 and Sect. 5, the application of a local geometric test is presented in Sect. 6. The new methods are applied to time series of two idealized cases, which provide important benchmarks for the methods, and to indices of two prominent climate modes, El Niño/Southern Oscillation and the North Atlantic Oscillation (NAO), to illustrate, in a geophysical setting, the insights afforded by the methods.

2. Definitions

In wavelet analysis, a time series is decomposed into frequency and time components by convolving the time series with a wavelet function satisfying certain conditions. There are many different kinds of wavelet functions but the most widely used is the Morlet wavelet, a sine wave damped by a Gaussian envelope expressed as

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

where ψ_0 is the Morlet wavelet, ω_0 is the dimensionless frequency, and $\eta = s \cdot t$, where s is the wavelet scale, and t is time (Torrence and Compo, 1998; Grinsted et al., 2004). The wavelet transform of a discrete time series x_n (n = 1, ..., N) is given by

$$W_n^X(s) = \sqrt{\frac{\delta t}{s}} \sum_{n'=1}^{N} x_{n'} \psi_0[(n' - n)\frac{\delta t}{s}], \tag{2}$$

where δt is a uniform time step determined from the time series and $|W_n^X(s)|^2$ is the wavelet power of a time series at scale s and time index n (Torrence and Compo, 1998; Grinsted et al., 2004). Note that for the Morlet wavelet with ω_0 = 6 the wavelet scale and the Fourier period λ are approximately equal ($\lambda \approx 1.03s$).

3. Existing significance testing methods

3.1 Pointwise significance testing

For climatic time series, the significance of wavelet power can be tested against a theoretical red-noise background (Torrence and Compo, 1998). For a first-order autoregressive (Markov) process

$$X_n = \alpha X_{n-1} + w_n \tag{3}$$

with lag-1 autocorrelation coefficient α, Gaussian white noise w_n, and $X_0 = 0$, the normalized theoretical red-noise power spectrum is given by

$$P_f = \frac{1- \alpha^2}{1+ \alpha^2 - 2\alpha \cos(2\pi f/N)}, \tag{4}$$

where $f = 0,\ldots, N/2$ is the frequency index (Gilman et al., 1963). To obtain, for example, the 5% pointwise significance level one must multiple Eq. (4) by the 95% 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 discrete Fourier red-noise spectrum has been shown by Torrence and Compo (1998) to be adequate in estimating the significance of local wavelet power and is thus used in this paper to estimate pointwise significance. The parameter α can be estimated using standards methods such as the Burg's and the Yule-Walker methods (Kay, 1988; Hayes, 1996).

Monthly time series and normalized wavelet power spectra for the NAO index (Hurrell et al., 1995, https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-station-based) and the Niño 3.4 index (Trenberth 1997, http://www.cgd.ucar.edu/cas/catalog/climind/Nino_3_3.4_indices.html) are shown in Figs. 1 and 2. The Niño 3.4 index data were converted to anomalies by subtracting the mean monthly values for each month from the monthly values. Note that the normalized wavelet power is the wavelet power at every time and period divided by the variance of the time series, which allows different wavelet power spectra to be readily compared. Another important feature of the wavelet power spectrum is the cone of influence, the region in which edge effects become important, or more precisely, the *e*-folding time of the autocorrelation for wavelet power at each scale, where the *e*-folding time is defined by Torrence and Compo (1998) as the point at which the wavelet power for a discontinuity at the edge drops by a factor of e^{-2}. The wavelet power spectrum of the NAO index reveals numerous time periods of enhanced variance at an array of time scales, though no preferred timescale is evident. For the Niño 3.4 index, the wavelet power spectrum detects statistically significant variance in the 16-64 month period band for the period 1960-2010. Another interesting feature emerges (labeled *H* in Fig. 2b): regions of no pointwise significance surrounded by regions of pointwise significance. These "holes" will turn out to be important structures in wavelet power spectra and are discussed thoroughly in Sect. 5.

3.2 Areawise significance testing

The idea behind the Maraun et al. (2007) areawise test (hereafter simply the "areawise test") is that correlations between adjacent wavelet coefficients arising from the reproducing kernel (see Appendix A) produce continuous regions of pointwise significance 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. Let (t, s) denote the location of a wavelet coefficient at scale s and time t. The correlation, $C(t, s, t', s')$, between any two wavelet coefficients located at (t, s) and (t', s') obtained from the wavelet transformation of a Gaussian white process is given by the reproducing kernel moved to t and stretched to s (Maraun et al., 2007), i.e.

$$C(t,s,t',s') = \sqrt{\frac{2s's}{(s')^2 + s^2}} \exp\left\{i\omega_0 \frac{s'+s}{(s')^2+s^2}(t'-t)\right\}$$

$$\times \exp\left\{-\frac{1}{2}\frac{(t'-t)^2 + \omega_0^2 (s'-s)^2}{(s')^2 + s^2}\right\} \tag{5}$$

(Maraun and Kurths, 2004). Thus, for significance patches generated from random fluctuations, the typical patch area is the area of the reproducing kernel. The test can be described more formally as follows: Let P_{pw} be the set of all pointwise significance values and define a critical area $P_{crit}(t,s)$ as the subset of the time-scale domain for which the reproducing kernel K (corresponding to the analyzing wavelet), dilated and translated to time t and scale s, exceeds the threshold of a critical level K_{crit}. Mathematically, $P_{crit}(t,s)$ is given by

$$P_{crit}(t,s) = \{(t',s') : K(t,s;t',s') > K_{crit}\}. \tag{6}$$

It is noted that critical area of the areawise test is not area of significance patches but the area of the reproducing kernel at some critical level and at some scale. For a patch of pointwise significant values, a point inside the patch is said to be areawise significant if the reproducing kernel dilated according to the scale in question entirely fits into the patch, i.e.

$$P_{aw} = \cup_{P_{crit}(t,s) \subset P_{pw}} P_{crit}(t,s), \tag{7}$$

where P_{aw} is the subset of pointwise significant values consisting of additionally areawise significant wavelet power coefficients. According to the areawise test, entire significance patches need not be areawise significant, just portions or subsets of them. That is, it is only those points that fit inside the kernel that are deemed areawise significant. The critical area is related to significance level of the areawise test by the following equation:

$$1 - \alpha_{aw} = 1 - \langle \frac{A_{aw}}{A_{pw}} \rangle, \tag{8}$$

where $1 - \alpha_{aw}$ is the significance level of the areawise test, A_{aw} is the area of the areawise significance patch, A_{pw} is the area of the pointwise significance patch, and $\langle \frac{A_{aw}}{A_{pw}} \rangle$ is the average ratio between the areas of area wise-significant patches and pointwise significance patches. It turns out that the calculation of α_{aw} is non-trivial, involving a root-finding algorithm that solves the equation $f(P_{crit}) - \alpha_{aw} = 0$ (see Sect. 4).

To illustrate the importance of the areawise significance test, the test was applied to the wavelet power spectra of the NAO and Niño 3.4 index time series (Figs. 3 and 4). Numerous 5% pointwise significance patches in the Niño 3.4 wavelet power spectrum were found to contain areawise-significant subsets, suggesting that these patches were less likely to be an artifact of multiple testing. For example, as indicated by the thick red contours, there are three areawise-significant regions located at a period of approximately 48 months, one at 1890, one at 1905, and a third one at 1985. Many more areawise-significant regions were identified at periods less than 8 months, especially before 1955. The wavelet power spectrum of the NAO index also contained pointwise significance patches with areawise-significant subsets, all at periods less than 8 months. However, it will be shown in Sect. 4 that they all may be artifacts of multiple testing, resulting from the large number of patches to which the areawise test was applied.

4. Geometric significance testing

4.1 Development

A disadvantage of the areawise test is the complexity of the α_{aw} calculation, which involves a root-finding algorithm. It is therefore desirable to construct an alternative test whose significance level is easy to calculate, readily allowing the following: (1) the application of the test to patches at various pointwise significance levels; (2) the adjustments of the significance level of the test; (3) the application of the test to wavelet power spectra obtained using other analyzing wavelets; and (4) the implementation of *p*–value adjustment procedures to control the family-wise error rates and false discovery rates.

The development of a geometric significance test will require ideas from basic geometry and set theory. In wavelet analysis, the wavelet power is computed at a discrete set of time coordinates T with elements t_i for $i = 1, \ldots, N$ and at a discrete set of scales S whose elements s_j (j =1,…,J) are given by

$$s_j = s_{min} s^{j\delta j} \tag{9}$$

and

$$J = \delta j^{-1} log_2 \left(\frac{N\delta t}{s_{min}}\right), \tag{10}$$

with δt a time step and s_{min} the smallest resolvable scale (Torrence and Compo, 1998). Note that the maximum value of δj for which adequate sampling can be achieved depends on the wavelet function, being approximately equal to 0.5 for the Morlet wavelet. For the geometric test, a patch will be considered to be a polygon with vertices $v_k = (x_k, y_k)$ for k = 0,…, m-1, where x_k and y_k are, respectively, elements from T and S and m-1 is the number of vertices. It is worth noting that not all patches are closed in the sense that some are located near the edges of the wavelet domain. To remedy this problem, semi-enclosed patches are artificially closed by connecting the two vertices located on the boundary of the wavelet domain with a line segment.

Perhaps the most fundamental property of a pointwise significance patch is its area, which can be calculated using the following special case of Green's Theorem:

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

where $y_0 = y_m$, $x_0 = x_m$ (Worboys and Duckham, 2004). For significance patches containing holes, the total area of the holes is subtracted from the area the significance patch would have if it did not contain the holes.

What will be of particular interest is the normalized area of a significance patch, not its absolute area. To compute the normalized area, the centroid of a significance patch will need to be calculated using the following formulas (Worboys and Duckham, 2004):

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

and

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

where C_t and C_s are the time and scale coordinates, respectively, of the centroid. Recall that the centroid is the area-weighted location of a polygon. If A_R is the area of the reproducing kernel dilated or contracted (at a certain critical level) to (C_t,C_s), then the normalized area of a significance patch is given by

$$A_n = \frac{A}{A_R}, \tag{14}$$

and allows one to compare sizes of significance patches across all scales simultaneously. Two idealized pointwise significance patches with equal normalized area are shown in Figs. 5a and 5b.

The idea of the geometric significance test is to generate a null distribution of A_n and use the null distribution to compute the significance of patches in the wavelet domain. In climate science, a suitable null hypothesis is red-noise so that A_n will be computed for a large ensemble of patches generated from red-noise processes. Using the null distribution of A_n, one can assign to each patch in the wavelet domain a probability p that the patch was not generated from a random stochastic fluctuation. It is noted that the null distribution of A_n depends on the choice of null hypothesis (not shown), with, for red-noise processes, A_n increasing with increasing lag-1 autocorrelation coefficients.

The calculation of the geometric significance level $1 - \alpha_g$, unlike the calculation of $1 - \alpha_{aw}$, is straightforward: for the areawise test one needs to compute α_{aw} as a function of P_{crit}, whereas for the geometric test α_g is no longer a function P_{crit}. Moreover, the estimation of P_{crit} involves a root-finding algorithm that solves the equation $f(P_{crit}) - \alpha_{aw} = 0$, where $f(P_{crit})$ is estimated using Monte Carlo simulations. Thus, the application of the areawise test to pointwise significance patches for M different values of α_{aw} would require M Monte Carlo ensembles, making p-values for the test difficult to obtain. For the geometric test, only a single Monte Carlo ensemble is needed, as a single choice of P_{crit} is needed to generate a null distribution, from which any desired value of α_g can be obtained. In fact, while the choice of P_{crit} impacts the mean value of the null distribution, the geometric significance of a significance patch is left unchanged, as the significance is relative to a distribution of χ under some noise model (Appendix B).

The elimination of the P_{crit} dependence from the calculation of the geometric significance level allows the geometric test to be readily performed on patches of various pointwise significance levels. For the areawise test, a new P_{crit} must be estimated for each pointwise significance level since A_{pw}, on average, will change depending on if the pointwise significance level $1-\alpha_p$ is increased (patches shrink) or is decreased (patches grow). For the geometric test, there is no need to find a new P_{crit} —simply compute a new null distribution based solely on the information of the pointwise significance patches at some pointwise significance level $1-\alpha_p$.

Another advantage of eliminating the P_{crit} dependence is that the geometric test can be readily applied to wavelet coherence, partial wavelet coherence (Ng, 2012), multiple wavelet coherence, and cross-wavelet spectra. The application of the geometric test to significance patches in the aforementioned wavelet spectra only requires a single Monte Carlo ensemble to generate a null distribution, eliminating the calculation of a new P_{crit} for each wavelet spectra and for each value of α_g. For the areawise test, a new P_{crit} must be estimated for each value of α_{aw} and for each wavelet spectra, making the areawise test difficult to implement in practical applications.

It may happen that a pointwise significance patch is so large that individual oscillations embedded in the patch cannot be detected by the geometric test. However, there are two solutions to this localization problem: the first solution is to increase the significance level of the pointwise test, allowing large patches to separate, and then perform the geometric test on the smaller patches. The second solution is to examine other properties of significance patches that may indicate the presence of multiple periodicities that form large significance patches from the merging of several smaller patches. The second solution will be addressed thoroughly in Sect. 5.

Another situation that may arise in practice is the application of the geometric test to patches located both inside and outside the cone of influence (COI). In the case of the pointwise significance test, the edge effects only influence those wavelet power coefficients that lie inside the COI; however, for the geometric test, the significance of the entire patch will be impacted even if the patch only partially lies inside the COI. The reason is that the COI will act to decrease the size of significance patches through the reduction of wavelet power in the COI and subsequently the total area of the patch. One should thus be cautious when interpreting the results of the geometric test for patches near the COI.

4.2 Multiple testing

If the geometric test was performed on K significance patches at the α_{geo} level, then, on average, one can expect $\alpha_{geo}K$ false positive results, which would make the geometric test permissive for large K. It is therefore necessary to reduce the number of false positive results. There are various ways to reduce the number of false positives, including the Walker test, Bonferroni correction, and other counting procedures (Wilks, 2006). Recently, methods for controlling the false discovery rate (FDR) have been developed, where the FDR is the expected proportion of rejected local null hypotheses that are actually true (Benjamini and Hochberg, 1995). In particular, Benjamini and Hochberg (1995) developed a method for controlling the FDR based on the number of local hypotheses being tested and the degree to which the local hypotheses were rejected, contrasting with other procedures that ignore the confidence with which the local tests reject the local hypotheses (Wilks, 2006). Moreover, the method has proven to have high statistical power, especially when only a small fraction of the K local tests correspond to false null hypotheses (Wilks, 2006). The procedure will therefore be used to control the false discovery rate of the geometric test, which will facilitate the interpretation of results.

Suppose that K local hypotheses were tested, where, in the present case, the local hypotheses refer to the testing of each patch individually under the assumption that the results of the individual tests are independent. A global geometric test can be performed at the α_{global} level as follows: Let $p_{(l)}$ denote the lth smallest of K local p-values; then, under the assumption that the K local tests are independent, the FDR can be controlled at the q-level by rejecting those local tests for which $p_{(l)}$ is no greater than

$$p_{FDR=} \max_{r=1,\ldots,K}\left[p_{(r)}: p_{(r)} \leq q(r/K)\right] \tag{15}$$

$$\max_{r=1,\ldots,K}\left[p_{(r)}: p_{(r)} \leq \alpha_{global}(r/K)\right] \tag{16}$$

so that the FDR level is equivalent to the global test level. According to the procedure, any local test resulting in a p-value less than or equal to the largest p-value for which Eq. (16) is satisfied is deemed significant. If no such local p-values exist, then none are deemed significant and, therefore, the global test hypothesis cannot be rejected. The global geometric test will thus only deem those significant patches with p-values satisfying Eq. (16) as significant. Throughout the paper $q = \alpha_{global}$ will be set to 0.05.

4.3 Comparisons with the areawise test

With a formal geometric significance test now developed, it is useful to compare the areawise and geometric significance tests, where comparisons will be made using an empirically derived quantity. Let N_{sig} be the number of pointwise significance patches in a given wavelet power spectrum, N_a the number of patches containing an areawise-significant region, N_g the number of geometrically significance patches, and N_{ag} the number patches that are both geometrically significant and that contain areawise-significant regions. The quantity

$$I_{sim} = \frac{N_{sig} - N_a - N_g + 2N_{ag}}{N_{sig}} \quad (17)$$

then measures the similarity between the two tests. The interpretation of I_{sim} is as follows: if I_{sim} = 1 then all patches containing areawise-significant regions are also geometrically significant and all patches which do not contain areawise-significant regions are also not geometrically significant. On the other hand, for values of I_{sim} less than 1 some patches containing areawise-significant regions may not be geometrically significant, with the converse also being true.

To better compare the similarity between the two tests, distributions of I_{sim} were constructed by generating 1000 synthetic wavelet power spectra of red-noise processes with fixed autocorrelation coefficients and length N = 1000 (arbitrary units) and computing I_{sim} for each of the synthetic wavelet power spectra. The experiment was performed for red-noise processes with different lag-1 autocorrelation coefficients to determine if I_{sim} depends on the AR1 model. The results are shown Fig. 6a. With a mean value of 0.90, a strong agreement was found between the areawise and geometric tests, differences arising from the fact that the areawise test is a local test, finding significant regions within patches, whereas the geometric test assigns a significance value to entire patches (see discussion below). Since I_{sim} was often less than 1.0, some patches containing areawise-significant regions were not found to be geometrically significant, and, conversely, some patches were geometrically significant without containing areawise-significant regions.

The quantity $r_{neg} = N_g/N_a$, which measures the ratio of false positive results between both tests, was also computed for case when both the geometric and areawise test levels were set to 0.05 (Fig. 6b). In this case, the mean value of r_{neg} was found to range from 1.0 to 2 and the median value was found to be generally greater than 1.0, ranging from 1 to 1.8. No dependence on the lag-1 autocorrelation coefficients was identified. The results indicate that the geometric test is generally less conservative than the areawise test for a given wavelet power spectrum. The lack of conservativeness, however, can be remedied by controlling the FDR of the geometric test at the q = 0.05 level. Fig. 6b shows r_{adj}, the ratio of false positive results between the areawise tests and the geometric test but with FDR controlled for the geometric test. As indicated in Fig. 6b, by controlling the FDR the geometric test is much more conservative than the areawise test, resulting in fewer false positive results, with a typical value of r_{adj} ranging from 0.02 to 0.05.

To explain the differences between the areawise and geometric tests, it will be necessary to consider the convexity of a patch, the degree to which a polygon or point set lacks concavities. The reason for considering convexity is illustrated by considering the two significance patches shown Fig. 5, which have equal values of A_n but different geometries: one is convex (i.e., has no concavities, Fig. 5a) and the other is not convex (Fig. 5b). Suppose that the areawise test was performed on the two patches at the α_{aw} level. For the convex patch shown Fig. 5a, the reproducing kernel is capable of fitting entirely inside the patch but is unable to fit inside the non-convex patch as a result of the concavity. Thus, although having equal area, the two patches differ in their areawise significance, where the difference in significance is related to their geometry. Thus, $p_{aw} = g(\mathcal{C}, A; H_0)$ for some function g, where p_{aw} is the areawise test p-value associated with a patch calculated under the null hypothesis H_0 and $\mathcal{C}$ is the convexity of the patch, which is now formally defined.

Rigorously, convexity is defined as follows: Let x and y be any two points in a set Z; then the set Z is convex if for all t the line segment

$$[x, y] = \{tx + (1 - t)y: 0 \le t \le 1\} \quad (18)$$

is in Z (Ziegler, 1995). Equivalently, a set is convex if it contains any line segment joining any pair of points in Z. Under this definition, for example, patches with thin bridges as described by Maraun et al. (2007) are not convex.

To quantify convexity, another idea from set theory, the convex hull, will be needed, which for a point set Z is defined as the intersection of all convex sets containing Z (Ziegler, 1995). In other words, it is the smallest convex set containing Z constructed from the intersection of all convex sets containing Z. Mathematically, the convex hull of a point set Z is expressed as

$$\mathrm{conv}(Z) = \bigcap\{Z' \subseteq \mathbb{R}^2: Z \subseteq Z', Z' \text{ convex}\}. \quad (19)$$

In practical applications, the convex hull of a set can be easily computed using existing algorithms (Barber et al., 1996). It is noted that all holes are ignored in the computation of the convex hull because the computation of the convex hull assumes that there are no holes in the polygon. A patch containing a hole can never have a smallest convex set containing the set because holes allow line segments to leave the patch regardless of the size of the convex hull.

A metric for convexity will now be defined using the area of a significance patch together with the area of its convex hull as follows: If A_k is the area of the convex hull of a significance patch whose area is A, then the convexity is

$$\mathcal{C} = \frac{\mathrm{A}}{A_k}, \quad (20)$$

where $0 \le \mathcal{C} \le 1$. High values of $\mathcal{C}$ correspond to significance patches with relatively small concavities, whereas small values of $\mathcal{C}$ correspond to patches with relatively large concavities, as in the case of significance patches with thin bridges.

According to the areawise test, patches with smaller values of $\mathcal{C}$ are less likely to be areawise significant so that it is expected that patches deemed significant by the areawise test will be primarily convex. To test this hypothesis, 10,000 patches arising from red-noise processes with different lag-1 autocorrelation coefficients were generated and the convexity of those patches deemed areawise significant at the $\alpha_{aw} = 0.05$ level was calculated. The results in Fig. 6c show the mean convexity as a function of the lag-1 autocorrelation coefficients, together with the 95%

confidence bound. The mean convexity of the patches was found to be approximately 0.8, regardless of the lag-1 autocorrelation coefficient. An identical experiment was also performed for geometrically significant patches but with the convexity of patches that are geometrically significant at the $\alpha_{geo} = 0.05$ being computed. In contrast to areawise-significant patches, patches that were found to be geometrically significant, on average, had lower convexity, the reason for which is that the calculation of α_{geo} makes no assumption about convexity. The *p*-value for the geometric test is thus $p_{geo} = f(A; H_0)$ for some function *f*, contrasting with p_{aw} that depends on convexity. The results of the experiments are consistent with Figs. 5a and 5b, where both the ideal patches have the same geometric significance but the ideal patch in Fig. 5b has a larger p_{aw} so that $p_{aw} > p_{geo}$.

Convexity cannot fully explain the differences between p_{aw} and p_{geo} for a given patch. More generally, $p_{aw} = g(\mathcal{C}, A, S_1, \ldots, S_R; H_0)$, where S_1 to S_R are shape parameters of the patch, such as aspect ratio and symmetry. Consider, for example, a convex patch whose length in the time direction is long with respect to the reproducing kernel (at some critical level) but thin in the scale direction with respect to the reproducing kernel. Such a patch would be deemed insignificant by the areawise test, though it may have an area much larger than the critical area of the areawise test. Asymmetry with respect to the scale axis, as another example, may also result in a patch being deemed insignificant by the areawise test if, for example, the width of the patch in the scale direction decreases with time. If the normalized areas of such patches are larger than the critical level of the geometric test, the patches will be geometrically significant, though may not be areawise significant if the reproducing kernel is unable to fit inside the narrow portion of the patch. The above arguments suggest that $f(A; H_0) \neq g(\mathcal{C}, A, S_1, \ldots, S_R; H_0)$ and thus the significance of patches as determined by the geometric and areawise tests need not be equal.

4.4 Geometric significance testing of climatic data

For climatic time series, significance is often tested against a red-noise background and therefore it is reasonable to expect that the areawise and geometric tests behave similarly when applied to climatic time series. As such, the areawise and geometric tests were applied to the NAO and Niño 3.4 time series. For the wavelet power spectrum of the NAO index time series (see Fig. 3), not a single patch was found to be geometrically significant after controlling the FDR at the 0.05 level, suggesting the NAO index time series is composed of stochastic fluctuations. In fact, the NAO has already been shown to be consistent with a first-order Markov process (Feldstein, 2002). Recent work by Hanna et al. (2014) claimed that the NAO variability has increased over the past 30 years; however, the results from this analysis suggest that such changes cannot be distinguished from stochastic fluctuations, which could render difficult projections of future changes of the NAO.

The wavelet power spectrum of the Niño 3.4 index (see Fig. 4) was found to contain numerous geometrically significant patches in the period band 16-64 months, especially after 1960. The 5% pointwise significance patch extending from 1980 to 2000, as an example, was found to be significant, as well as the patch centered at 2008. The significance patch centered at 1985 and at a period of 32 months, however, is so large that individual oscillations could not be identified. To remedy the problem, the geometric significance was applied to 1% (α_p = 0.01) pointwise significance patches with q = 0.05, resulting in 1% pointwise significance patches at 1970, 1995, and 2007 being deemed significant, all of which also contained areawise-significant regions. Patches located at a period less than 8 months were also found to be geometrically significant, though only before 1955.

5. Topological significance testing

5.1 Topological significance testing of ideal time series

Topology is a branch of mathematics concerned with properties of spaces that remain unchanged after continuous deformations. So far only geometric aspects of significance patches have been discussed. Area of a significance patch, as an example, is a geometric property in the sense that stretching the patch in both the scale and time direction would increase its area. There are properties, however, that would be unaffected by stretching the significance patch. As a motivating example, consider the significance patches shown in Fig. 4 corresponding to the wavelet power spectrum of the Niño 3.4 index (see Fig. 2), where there is a hole or void of pointwise significance located within a significance patch at 1985. This feature is topological, as the hole would remain under a continuous deformation such as stretching. A more formal definition of a hole will require some notions from topology. Let I = [0,1] be the closed unit interval. Then a path from a point a to a point b in a significance patch P is a continuous function $f: I \to P$ with $f(0) = a$ and $f(1) = b$, where in the case that $f(0) = f(1) = c$ the path is said to be closed (Hatcher, 2002). Note that a point is a special kind of closed path called the constant path. A patch will be said to contain a hole if there exists a path in the significance patch such that it cannot be continuously deformed into a point, where the feature obstructing the path from such a deformation is a hole. The definition is consistent with notions of simply-connectedness in topology (Hatcher, 2002). Figure 4 shows an example of a closed path (blue curve) in a patch that cannot be contracted to a point because it surrounds a hole located in the patch.

For a patch with a hole, there will be two boundaries, an external boundary and an internal boundary representing the boundary between the hole and the patch. Thus, if a patch contains an internal boundary or contour it will contain a hole, whereas a patch without a hole will contain no internal contours. In practical applications, the existence of a hole can be determined by orienting external contours in the clockwise direction and internal contours in the counter-clockwise direction, a procedure automatically implemented by the Matlab contour routine. The number of counter-clockwise oriented contours is thus the number of holes in the wavelet power spectrum at a given pointwise significance level.

To begin the topological analysis, the topology of time series with known structures will be analyzed. Given the importance of red-noise processes in the spectral analysis of climatic time series, the topology of patches generated from red-noise processes is first considered to determine if pointwise significance patches can be distinguished from those generated from red-noise processes solely based on their topology. To answer this question, 10,000 wavelet power spectra of red-noise processes were generated and the number of holes (denoted by N_h hereafter) at a finite set of pointwise significance levels was computed for each wavelet power spectra (Fig. 7). It was found that N_h is not a random function of the pointwise significance level, as indicated by the 95% confidence bounds. Most importantly, for pointwise significance levels less than 10%, few patches contained holes, suggesting that holes are an uncommon feature of significance patches generated from red-noise processes (Table 1) and therefore can be used to distinguish spurious patches from important structures. It also noted that neither the shape nor the amplitude of the curve in Fig. 7 depends on the lag-1 autocorrelation coefficient of the red-noise process. Table 1 also suggests that patches containing more than a single hole are unlikely to be the result of red-noise, even for a modest pointwise significance level of 20%. For pointwise significance levels of 1% and 5%, no more than a single hole was identified in a given patch.

A simple algorithm for assessing the significance of holes is therefore developed. To find the significance of holes, plot the centroids of holes at a finite set of pointwise significance levels and project the centroids onto the wavelet domain, resulting in a topological wavelet diagram. The number of holes contained in a patch should also be computed, as patches with more holes are less likely to result from red-noise. In accordance with Fig. 7 and Table 1, regions in the wavelet domain where holes exist below the 20% pointwise significance level will be considered regions with significant topological features.

With a method for assessing the significance of holes, it is reasonable to analyze different ideal time series, both linear and nonlinear, to determine what types of time series produce holes in significance patches. Perhaps the simplest case is a single sinusoid with additive white noise (not shown), where the time series power spectrum in tested against a white-noise background spectrum. In this case, no evidence was found that a single sine wave, regardless of amplitude and signal-to-noise ratio, is capable of generating holes in 5% pointwise significance patches. A similar experiment was repeated but the power spectra of the sine waves were tested against red-noise spectra. The results also indicated that a single sine wave is incapable of producing holes in 5% pointwise significance patches, implying holes arise from a richer structure embedded in time series. Thus, two more complex cases are considered.

To derive the Case 1 time series, first consider the nonlinear system

$$X_{out}(t) = \ bX_{in}(t) + \ \gamma X_{in}^2(t), \tag{21}$$

where $X_{in}(t)$ is the input into the system, $X_{out}(t)$ is the output of the system, b is a linear coefficient, and γ is a nonlinear coefficient. The output from this system will be quadratically phased coupled (King, 1996), where quadratic phase coupling indicates that for frequencies f_1, f_2, and f_3 and corresponding phases ϕ_1, ϕ_2, and ϕ_3 the sum rules $f_1 + f_2 = f_3$ and $\phi_1 + \phi_2 = \phi_3$ are satisfied. In Case 1, $X_{in} = cos\, 2\pi f t$ so that

$$X_{out}(t) = \frac{\gamma}{2} + b\cos 2\pi f t - \frac{\gamma}{2}\cos 4\pi f t, \qquad (22)$$

indicating that the output contains an additional frequency component at the harmonic $2f$ (harmonic generation) and the mean value of the output has shifted (rectification) with respect to the input. Figures 8a and 8b show the time series of X_{out} and the significance of the wavelet power for the case when $f = 1/64 = 1/\lambda_1$, b = 1, $\phi_1 = \pi/2$, $\phi_2 = \pi/3$, and γ = 0.25 (arbitrary units) and with Gaussian white noise added to the output. In this case, the significance of the wavelet power was tested against a red-noise background spectrum. Figure 8 shows numerous pointwise significance patches, all of which are spurious except for the one at $\lambda_1 = 64$. The areawise and geometric test correctly identified the pointwise significance patch at $\lambda_1 = 64$ to be significant but deemed a spurious patch as significant at time 140 and at $\lambda = 3$. It is noted that the geometric test only deemed the 1% pointwise significance patch at $\lambda_1 = 64$ as significant. Also note that the pointwise significance test was unable to detect the harmonic with period $\lambda_2 = 32$ using a red-noise background spectrum.

It should be noted, however, that if the parameter γ were increased to a value greater than 1, the oscillation with period $\lambda_2 = 32$ would become more prominent. In fact, it was found that for $\gamma \geq 1$ the areawise and geometric tests perform better (not shown), correctly identifying the oscillation with period $\lambda_2 = 32$, with the result also depending on the noise level of the white noise. Case 1 thus only serves as an illustrative example of a situation that may arise when a wavelet analysis is applied to a geophysical (often noisy) time series.

To extract more information from the wavelet power spectrum, the centroids of holes were plotted as a function of the pointwise significance level (Fig. 8c). Figure 8c shows that holes only existed at pointwise significance levels of at most 15% and 20% and therefore not all nonlinear time series can generate holes at the 5% pointwise significance level, suggesting that the relative difference between the primary frequency components or the resulting frequency combinations is important, as discussed below. The amplitudes of the coefficients b and γ, as well as the signal-to-noise ratio of the Gaussian white noise, turn out to be also important, which is discussed below.

Case 2 is the quadratically phase-coupled time series

$$X(t) = a\cos(2\pi f_1 t + \phi_1) + b\cos(2\pi f_2 t + \phi_2) +$$

$$\gamma\cos[2\pi(f_1 + f_2)t + \phi_1 + \phi_2], \qquad (23)$$

which consists of three frequency components: $f_1 = 1/20 = 1/\lambda_1$, $f_2 = 1/30 = 1/\lambda_2$, and $f_1 + f_2 = 1/12 = 1/\lambda_3$, and γ is assumed to be 0.5. It is noted that Case 1 is a special case of Case 2. Like Case 1, wavelet power was also tested against a red-noise background. Unlike the significance patches in Fig. 8c corresponding to Case 1, holes have appeared in 5% pointwise significance patches between periods $\lambda_1 = 20$ and $\lambda_2 = 30$ (Fig. 9b). Moreover, the 5% pointwise significance patch containing the hole (labeled P_1) was found to be geometrically significant but was not found to contain an areawise-significant subset. It is also worth noting that the areawise and geometric tests failed to detect a significant periodicity at $\lambda_1 = 20$ despite the fact that it is known to exist by construction. Figure 9c shows that a few holes existed at low pointwise significant levels ($\leq$ 20%), though only one was found at the 5% pointwise significance level (light red shading). However, if one applies the pointwise significance test to the wavelet power at the 20% significance level a feature emerges that can hardly be produced from red-noise (see Table 1), namely a large 20% significance patch (light blue shading) containing four holes located in the period band 20-30. One can thus have confidence that the feature is significant. Furthermore, by constructing a patch topologically unlike those generated from red-noise, significant wavelet power extending from time 20 to 300, undetected by the pointwise, areawise, and geometric tests, has been recovered, whereas only applying the 5% pointwise test would result in two patches that are seemingly indistinguishable from red-noise (labeled P_2 and P_3), with only one at $\lambda_2 = 30$ being geometrically significant.

The ability of the pointwise, areawise, and geometric tests to detect significant structures inevitably depends on the parameters a, b, γ, f_1, and f_2. In fact, Maruan et al. (2007) has already determined that the pointwise test and areawise test are sensitive to the signal-to-noise level. It was hypothesized that the results of the topological method also depend on the parameters a, b, γ, f_1, and f_2. To test the hypothesis, several experiments were performed, the first of which investigated the relationship between f_1, f_2, and the number of holes. The experiment is described below.

Though both ideal time series contain a quadratic nonlinearity, the nonlinear interaction in Case 2 contained oscillations with nearby frequency components, allowing the formation of holes, whereas for Case 1 no significant holes appeared in significance patches. It appears that the presence of holes depends on the relative location of two oscillations in the frequency domain, and thus it is reasonable to suspect that there exists a critical frequency difference Δf_{crit}, measuring the maximum frequency difference for which holes will appear in a wavelet power spectrum. An empirically derived Δf_{crit} was determined by generating a large ensemble of time series of the form

$$x(t) = \cos 2\pi f_1 t + \cos 2\pi f_2 t + w(t), \tag{24}$$

where $f_2 > f_1 > 0$ were generated at random, $w(t)$ is additive white noise, and all the time series were of a fixed length. The signal-to-noise ratio was fixed to 20 and each wavelet power spectrum was tested against a red-noise background spectrum. Figure 10 shows the mean value of N_h as a function of $\Delta r = (f_2 - f_1)/f_2$, the relative fractional change. For $\Delta r = 0.5$, holes never appeared, whereas for $\Delta r = 0.3$ holes appeared frequently. There is therefore a preferred frequency combination for which holes are more likely to appear. It was estimated that the upper critical value of Δr is $\Delta r_{crit} = 0.45$. Using the definition of Δr, one can write $\Delta f_{crit} = 0.45 f_2$ and therefore the critical frequency difference is a function of f_2.

It turns out that even if the above experiment (not shown) was repeated using white-noise rather than red-noise background spectra Δr_{crit} would still be equal to 0.45, though more holes were found to appear at signal-to-noise ratios less than 2. It was expected, however, that Δr_{crit} also depends on the amplitudes of the cosines in Eq. 24. Thus, a third experiment was conducted in which the amplitudes of the cosines were allowed to vary from 1 to 50 and f_1 and f_2 were allowed to vary from 0 to 0.5. The experiment was repeated for signal-to-noise ratios from 1 to 20. The results from the experiments (not shown) indicate that for red-noise background spectra and for a signal-to-noise ratio of 20 that Δr_{crit} = 0.53, contrasting with the case for white-noise background spectra where Δr_{crit} was found to be 0.51.

The empirical results shown in Fig. 10 have theoretical implications. Suppose that a time series contained two oscillations of equal amplitude such that frequency components of the two oscillations were such that $f_2 = 2f_1$. Furthermore, suppose that the wavelet power of the oscillations were computed and the significance was tested against a red-noise or white-noise background spectrum. In this case, Δr = 0.45 and therefore holes will almost never appear in 5% pointwise significance patches, making the detection of quadratic phase coupling using topological methods more difficult in the case of self-interactions. More generally, suppose that a single sinusoid $X_{in}(t) = cos2\pi ft$ is passed through the nonlinear system

$$X_{out}(t) = bX_{in}(t) + \gamma X_{in}^{2n}(t), \tag{25}$$

where, after using the power-reduction for a cosine (Beyer, 1987), the output is given by

$$X_{out}(t) = b\cos 2\pi t + \frac{\gamma}{2^{2n}}\binom{2n}{n} + \frac{\gamma}{2^{2n-1}}\sum_{k=0}^{n-1}\binom{2n}{k}\cos 4\pi f(n-k)t\ , \tag{26}$$

where *n* is a positive integer and $\binom{n}{q}$ is a binomial coefficient. For the cosines in the summation, the frequency difference between any two cosines is

$$\Delta f = 4\pi f(n-p) - 4\pi f(n-m) = 4\pi f(m\text{-}p), \tag{27}$$

where $0 \le p < m \le n\text{-}1$. Thus,

$$\Delta r = (f_2 - f_1)/f_2 = \frac{4\pi f(m-p)}{4\pi f(n-p)} = \frac{m-p}{n-p}. \tag{28}$$

Using the fact that holes can only appear between oscillation pairs with $\Delta r \le 0.53$ for a red-noise background spectrum, one can show that for large *n* more holes are able to appear in wavelet power spectra, with the likelihood of holes appearing depending on b and γ, with larger values of b and γ producing more holes. In this case, holes can form in the wavelet spectrum since, for example, if m = 6 and p = 5 with n = 10 the condition $\Delta r \le$ 0.53 will be satisfied. The result also holds if the order of the nonlinear interaction was odd and if the cosine function $X_{in}(t)$ was replaced by a sine function. For an odd order nonlinear interaction, however, $\Delta r = (2m-2p)/(2n+1-2p)$, where $0 \le p < m \le n$.

5.2 Topological significance testing of climatic time series

With a better understanding of the origins of holes contained in significance patches, the wavelet power spectra shown in Figs. 1 and 2 are now analyzed more closely. Shown in Fig. 11a is the topological wavelet diagram corresponding to the wavelet power spectrum of the Niño 3.4 index, which shows the existence of numerous holes at low ($\leq 20\%$) pointwise significance levels, indicating that these patches are significant features (see Table 1). For example, the rather large patch extending from 1960 to 2013 in the period band 16 to 64 months contains a hole located at 1985 and at a period of 32 months that existed at the 5% pointwise significance level. In the same patch, three more holes existed at the 10% pointwise significance level, one located at 1975 and at a period of 48 months, a second one located at 1995 and at a period of 64 months, and a third one located at 2008 and at a period of 24 months. According to Table 1, three holes in a single 10% pointwise significance patch under the null hypothesis of red-noise is extremely unlikely, if not impossible. On can thus conclude with high confidence that the patch was not generated from a random stochastic fluctuation. Moreover, the discussion in Sect. 5.1 suggests that at the very least phase-coherent oscillations were likely present in the Niño 3.4 time series, where phase coherency implies that two oscillations have a stable relative phase relationship but are not necessarily interacting nonlinearly.

The wavelet topological diagram (Fig. 11b) corresponding to the wavelet power spectrum of the NAO is less interesting, containing few holes at high pointwise significance levels. At 1875, however, a patch contained holes at the 10% pointwise significance level, suggesting that the patch is a significant feature.

7. Summary and Discussion

A geometric significance test was developed for more rigorously assessing the significance of features in the wavelet domain. The geometric test, although related to the existing areawise test, was found to be more flexible in the sense that *p*-values could be readily calculated, involving a single Monte Carlo ensemble. Another strength of the geometric test is that the false discovery rate can be controlled at a desire level, minimizing the number of false rejections of the null hypothesis. On the other hand, the geometric test had the disadvantage of being less local than the areawise test.

It is noted that the geometric test was only applied to patches arising from the convolution of the Morlet wavelet with a time series. The results presented in this paper are not valid for wavelet power spectra obtained using other analyzing wavelets, the reason for which is that each wavelet function has different time- and scale-localization properties that inevitably impact the geometry of patches. For example, patches found in the wavelet power spectrum obtained using a Paul wavelet are elongated in the scale direction relative to those obtained using a Morlet wavelet with $\omega_0 = 6$, resulting in nearby patches at different scales merging together. The merging of patches at different scales will alter their geometry with respect to the relatively thin (in scale) patches obtained using the Morlet wavelet.

One disadvantage of the geometric and areawise tests is that they require a binary decision in which pointwise and geometric significance levels must be chosen. The binary decision can be circumvented by applying a *p*-value adjustment procedure to the wavelet power coefficients directly. For example, one could apply the Benjamini and Hochberg (1995) procedure to the wavelet power coefficients or a modified version of the procedure developed by Benjamini and Yekutieli (2002), which is valid for any dependency structure among the local test statistics. The latter procedure would seem most appropriate given the autocorrelation structure of wavelet power coefficients; however, it is noted that the procedure has less statistical power than the original procedure valid for independent local test statistics, though Wilks (2006) found the Benjamini and Hochberg (1995) procedure to remain powerful even when the assumption of independence is violated.

The topology of significant patches was also analyzed. Holes in significant patches, a topological notion, were capable of distinguishing spurious patches from true structures. The holes were identified as arising from phase-coherent oscillations with nearby frequency components and may indicate the existence of a nonlinear interaction. Patches arising from different analyzing wavelets can differ topologically. For the Paul wavelet, the shrinking of patches in time, for example, was found, after a preliminary investigation, to reduce the number of holes in wavelet power spectra. The reduction in the number of holes can be attributed to the tearing of a patch in the time direction. The results, however, require further investigation and are a subject of future work.

The new methods introduced in this paper were applied to the NAO and Niño 3.4 indices, two well-known but contrasting time series. For the Nino 3.4 index, the methods detected geometrically significant structures as well as topological structures unlike that of red-noise, which provide evidence of some predictability of El Niño/Southern Oscillation, which has become of increasing importance in climate science given that its future state is uncertain under a changing global climate system (Latif and Keenlyside, 2008). For the NAO index, the new methods were unable to detect features that are distinguishable from background noise, suggesting that the NAO is a stochastic process with little predictability. The methods developed in this paper will give researchers the tools needed for a better understanding of features found in wavelet power spectra.

Appendix A

Let $F(s, t)$ be the continuous wavelet transform of a function $f(t)$ such that

$$F(s, t) = \iint K(s, t;\ s', t'')F(s', t'')ds'dt''. \qquad \text{(A1)}$$

Then the reproducing kernel is given by

$$K = \frac{1}{C_\psi \sqrt{s s'^{5/2}}} \int \left[\psi\left(\frac{t' - t''}{s'}\right) \psi^* \left(\frac{t - t'}{s}\right) \right] dt', \qquad \text{(A2)}$$

where

$$C_\psi = \int_0^\infty \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty, \qquad \text{(A3)}$$

and $\Psi(\omega)$ is the Fourier transform of ψ, and the asterisk denotes the complex conjugate. The reproducing kernel captures the structure of wavelet coefficients whereby the wavelet coefficient at any point contains information about a nearby wavelet coefficient weighted by K (Tropea, 2007).

Appendix B

Let $A^N_{patch}(C_t, C_s)$ be the test statistic associated with a significance patch whose centroid is (C_t, C_s) and let $A^N_{\alpha_g}$ be the value of the test statistic corresponding to the $1-\alpha_g$ significance level of the geometric test. Writing

$$A^N_{\alpha_g} = \frac{A_{\alpha g}}{A_R} \tag{B1}$$

and

$$A^N_{patch}(C_t, C_s) = \frac{A_{patch}}{A_R}, \tag{B2}$$

it follows that

$$\frac{A^N_{patch}(C_t, C_s)}{A^N_{\alpha_g}} = \frac{A_{patch}}{A_{\alpha g}}, \tag{B3}$$

where is A_{patch} the area of the significance patch and is the A_{α_g} the area of a typical patch under the null hypothesis corresponding to the $1-\alpha_g$ significance level. Since Eq. (B3) no longer contains A_R, the relationship between $A^N_{patch}(C_t, C_s)$ and $A^N_{\alpha_g}$ no longer depends on P_{crit}.

Appendix C

Recall that Green's Theorem in the plane states that

$$\int_C (Pdx + Qdy) = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA\,, \qquad \text{(C1)}$$

where *C* is a positively oriented, piecewise smooth curve, bounding a region *D*, $\boldsymbol{F} = \langle P, Q \rangle$ is a vector field on *D*, and *x* and *y* are the usual Cartesian coordinates (Baxandall and Liebeck, 2008). Note that if one sets

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 1, \qquad \text{(C2)}$$

then the right-hand side of Eq. (C1) can be used to calculate the area of a region *D*. Thus, let $Q = x/2$ and $P = -y/2$ so that

$$\frac{1}{2}\int_C (xdy - ydx) = A(D), \qquad \text{(C3)}$$

where $A(D)$ denotes the area of *D*. Let (x_0, y_0),…, (x_{m-1}, y_{m-1}) be *m*-1 vertices of a polygon. If C_0 is a line segment from (x_0, y_0) to (x_1, y_1), then

$$\int_{C_0} (xdy - ydx) = x_0\, y_1 - x_1 y_0. \qquad \text{(C4)}$$

More generally, denote by C_k the segment from (x_k, y_k) to (x_{k+1}, y_{k+1}), recalling that $x_m = x_0$ and $y_m = y_0$. Since *C* = $C_0 \cup C_1, \ldots, \cup\, C_{m-1}$, we can write

$$A(D) = \frac{1}{2}\int_C (xdy - ydx)$$

$$= \frac{1}{2}\int_{C_0} (xdy - ydx) + \frac{1}{2}\int_{C_1} (xdy - ydx) + \cdots + \frac{1}{2}\int_{C_{m-1}} (xdy - ydx) \qquad \text{(C5)}$$

and thus

$$A(D) = \frac{1}{2}(x_0 y_1 - x_1 y_0) + \frac{1}{2}(x_1 y_2 - x_2 y_1) + \cdots + \frac{1}{2}(x_{m-1} y_0 - x_0 y_{m-1})$$

$$= \frac{1}{2}\sum_{k=0}^{m-1}(x_k\, y_{k+1} - x_{k+1} y_k). \qquad \text{(C6)}$$

References

Benjamini, Y. and Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Power Approach to Multiple Testing, J. Roy. Stat. Soc., 57, 289-300, 1995.

Benjamini, Y. and Yekutieli, D.: The Control of the False Discovery Rate in Multiple Testing under Dependency, Ann. Stat., 29, 1165–1188, 2001.

Barber, C. B., Dobkin, D. P., and Huhdanpaa, H.: Quickhull Algorithm for Convex Hulls. ACM T. Math. Software, 22, 469-483, 1996.

Baxandall, P. and Liebeck, H.: Vector Calculus, Dover Publications, INC., Mineloa, New York, 550 pp., 2008.

Beyer, W. H.: CRC Standard Mathematical Tables. 28th Ed. CRC Press, Boca Raton, Florida, 674 pp., 1987.

Edelsbrunner, H. and Harer, J.: Persistent Homology – A Survey, Cotemp. Math., 12, 1-26, 2010.

Feldstein, S. B.: The Time Scale, Power Spectra, and Climate Noise Properties of Teleconnection Patterns, J. Climatol., 13, 4430-4440, 2000.

Gilman, D. L., Fuglister, F. J., and Mitchell J. M. Jr.: On the Power Spectrum of “Red Noise”, J. Atmos. Sci., 20, 182–184, 1963.

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.

Hasselmann, K.: Stochastic Climate Models Part I. Theory, Tellus, 28, 473-485, 1976.

Hatcher, A.: Algebraic Topology, Cambridge University Press, New York, 544 pp., 2001.

Hayes, M. H.: Statistical Digital Signal Processing and Modeling, John Wiley & Sons, New York, 624, 1996.

Hanna, E., Cropper, T. E., Jones, P. D., Scaife, A. A., and Allan, R.: Recent Seasonal Asymmetric Changes in the NAO (a Marked Summer Decline and Increased Winter variability) and Associated Changes in the AO and Greenland Blocking Index, Int. J. Climatol., 2014.

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.

Kay, S. M.: Modern Spectral Estimation: Theory and Application, Prentice Hall, Englewood Cliffs, NJ, 560 pp., 1988.

King, T.: Quantifying Nonlinearity and Geometry in Time Series of Climate. Quat. Sci. Rev., 15, 247–266, 1996.

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 Nino/Southern Oscillation Response to Global Warming, Proc. Natl. Acad. Sci. USA, 106, 20578–20583, 2008.

Lee, Y. J. and Lwiza, K. M. M.: Factors Driving Bottom Salinity Variability in the Chesapeake Bay. Cont. Shelf Res., 28, 1352-1362, 2008.

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.

Müller, W. A., Frankignoul, C., and Chouaib, N.: Observed Decadal Tropical Pacific–North Atlantic Teleconnections. Geophys. Res. Lett., 35, doi:10.1029/2008GL035901, 2008.

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

Özger, M., Mishra, A. K., and Singh, V. P.: Low Frequency Drought Variability Associated with Climate Indices. J. Hydrol., 364, 152–162, 2009.

Paluš, M. and Novotná, D.: Quasi-biennial Oscillations Extracted from the Monthly NAO Index and Temperature Records are Phase-synchronized. Nonlin. Process. Geophys., 13, 287–296, 2006.

Schimanke, S., Körper, J., Spangehl, T., and Cubasch, U.: Multi-decadal Variability of Sudden Stratospheric Warmings in an AOGCM. Geophys. Res. Lett., 38, doi: 10.1029/2010GL045756, 2011.

Terradellas, E., Soler, M. R., Ferreres, E., and Bravo, M.: Analysis of Oscillations in the Stable Atmospheric Boundary Layer Using Wavelet Methods. Boundary-Layer Meteorol., 114, 489–518, 2005.

Torrence, C. and Compo, G. P.: A Practical Guide to Wavelet Analysis, Bull. Am. Meteorol. Soc., 79, 61–78, 1998.

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

Tropea, C., Yarin, A. L., and Foss, J. F. (Eds.): Springer Handbook of Experimental Fluid Mechanics. Springer, Berlin, Germany, 1557, 2007.

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.

Wilks, D. S.: On “Field Significance” and the False Discovery Rate, J. Appl. Meteor. Climatol., 45, 1181-1189, 2006.

Worsby, F. M., Duckham, M.: GIS: A Computing Perspective, CRC Press, Boca Raton, FL, 448 pp., 2004.

Zhang, Q., Xu, C., Jiang, T., and Wu, Y.: Possible Influence of ENSO on Annual Maximum Streamflow of the Yangtze River, China, J. Hydrol., 333, 265–274, 2007.

Ziegler, G. M.: Lectures On Polytopes, Graduate Texts in Mathematics, 152 pp., Springer, New York, 370, 1995.

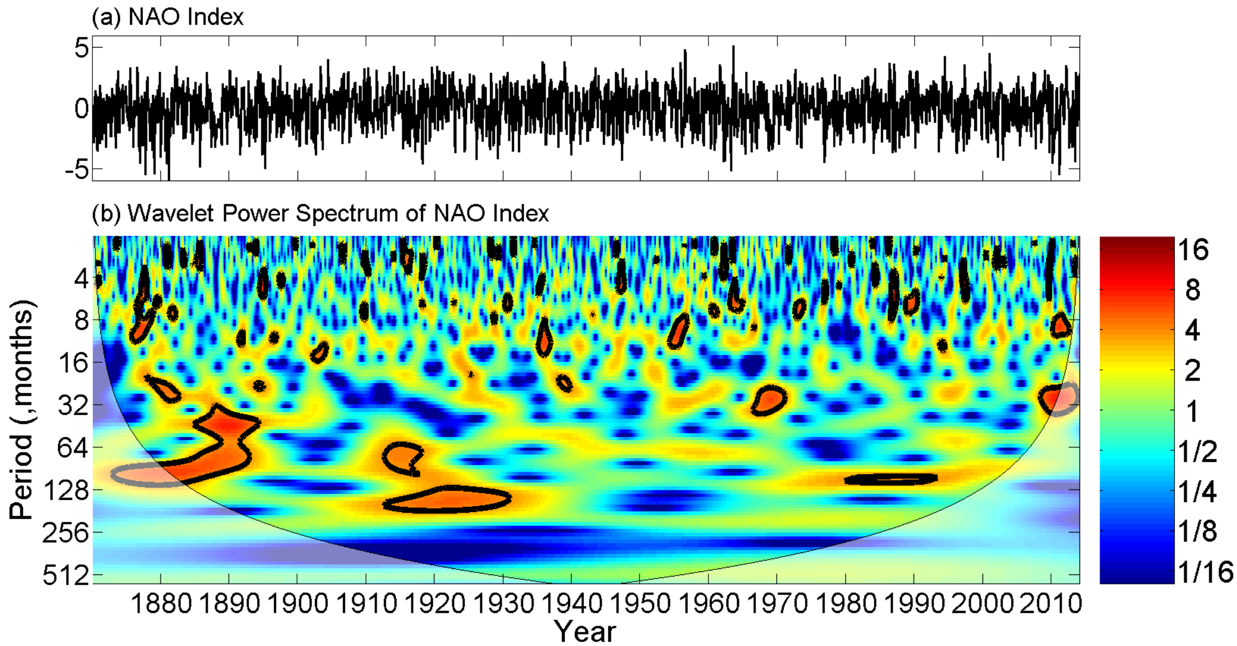

Figure 1. (a) The NAO index from 1870 to 2013. (b) The normalized wavelet power spectrum of the NAO index. Thick contours enclose regions of 5% pointwise significance. Light shading corresponds to the cone of influence, the region in which edge effects become important.

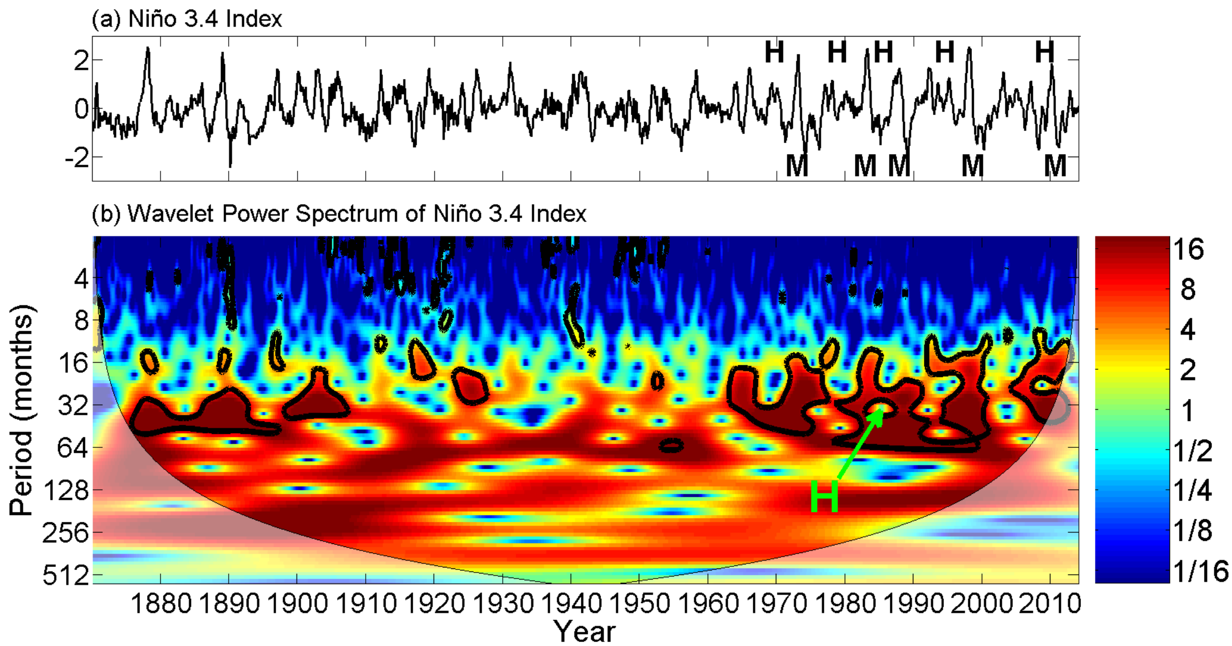

Figure 2. (a) The Niño 3.4 index time series from 1870 to 2013. Points labeled *M* indicate where the merging process occurred and points labeled *H* indicate where a hole was formed (see Sect. 5.2 for details). (b) Same as Fig. 1b except for the Niño 3.4 index for the period 1870-2013. *H* together with the arrow marks the location of a hole.

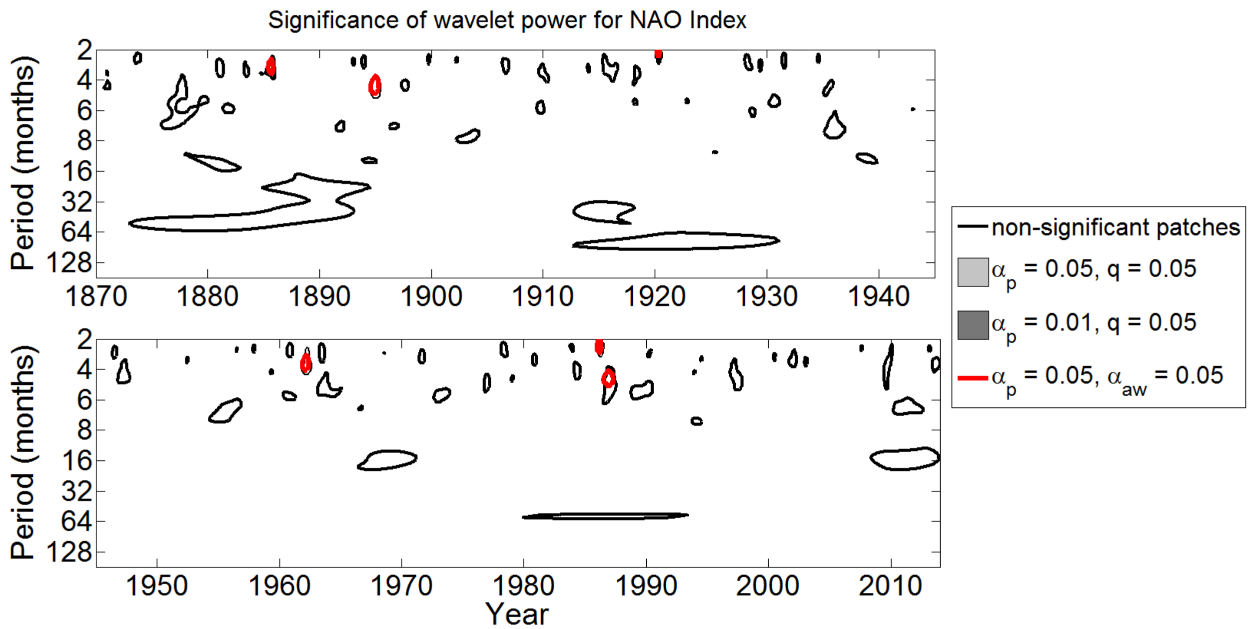

Figure 3. Significance of wavelet power for the NAO index mean monthly values for the period 1870-2013. Black contours enclose regions of 5% pointwise significance (see Sect. 3.1) and thick red contours are the 5% areawise-significant subsets (see Sect. 3.2). Light gray shading indicates those 5% pointwise significance patches that are geometrically significant at the $q = 0.05$ level and dark gray shading indicates those 1% pointwise significance patches that are geometrically significant at the $q = 0.05$ level.

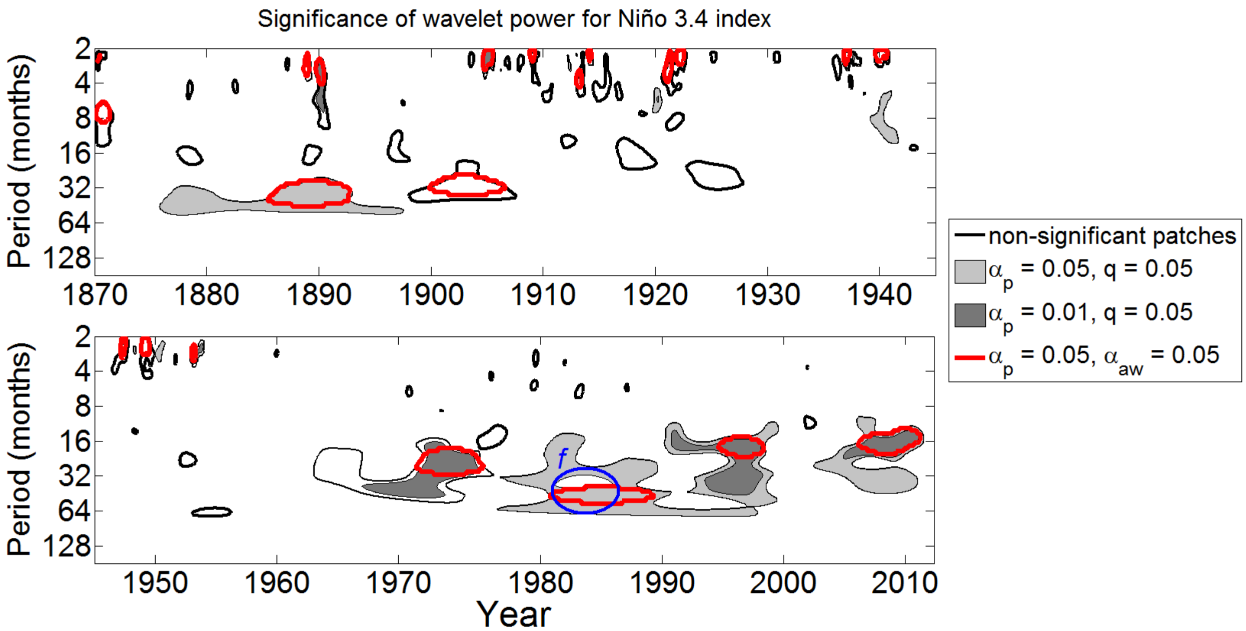

Figure 4. Same as Fig. 3 but for the Niño 3.4 for the period 1870-2013. The blue curve represents a closed path f that is not contractible to a point because it surrounds a hole (see Sect. 5.1 and Fig. 2).

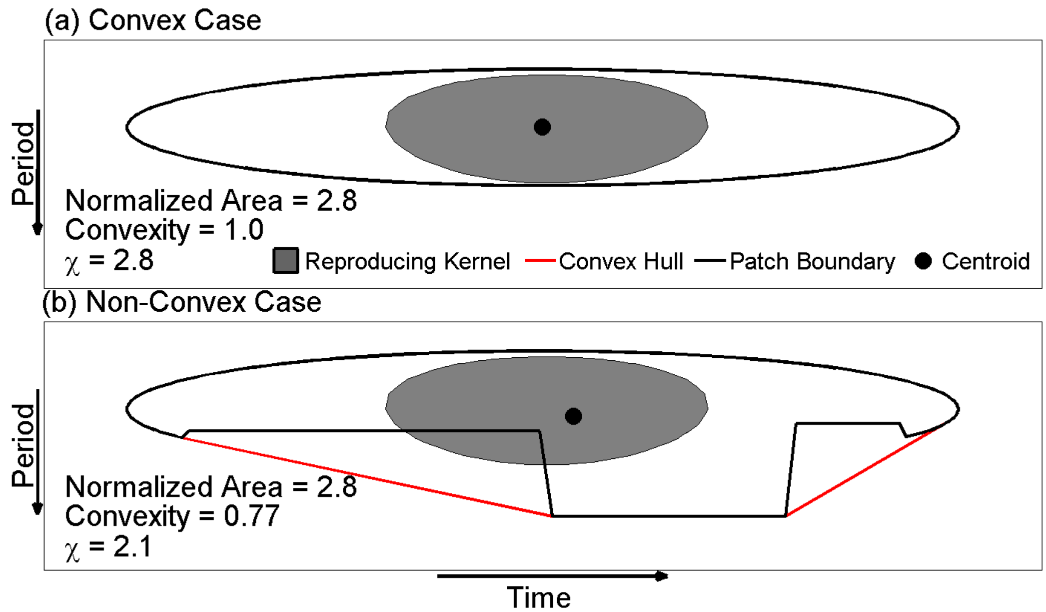

Figure 5. (a) An idealized convex pointwise significance patch whose boundary is indicated by the black contour and whose centroid is indicated by the black dot. For reference, the reproducing kernel associated with the areawise test is shown, which is indicated by gray shading. In this case, the reproducing kernel lies entirely inside the patch. The convexity, normalized area, and χ are displayed on the bottom left corner. (b) Same as (a) except the area of the convex hull (red curve) is not equal to the area of the patch and the reproducing kernel is unable to fit entirely inside the patch.

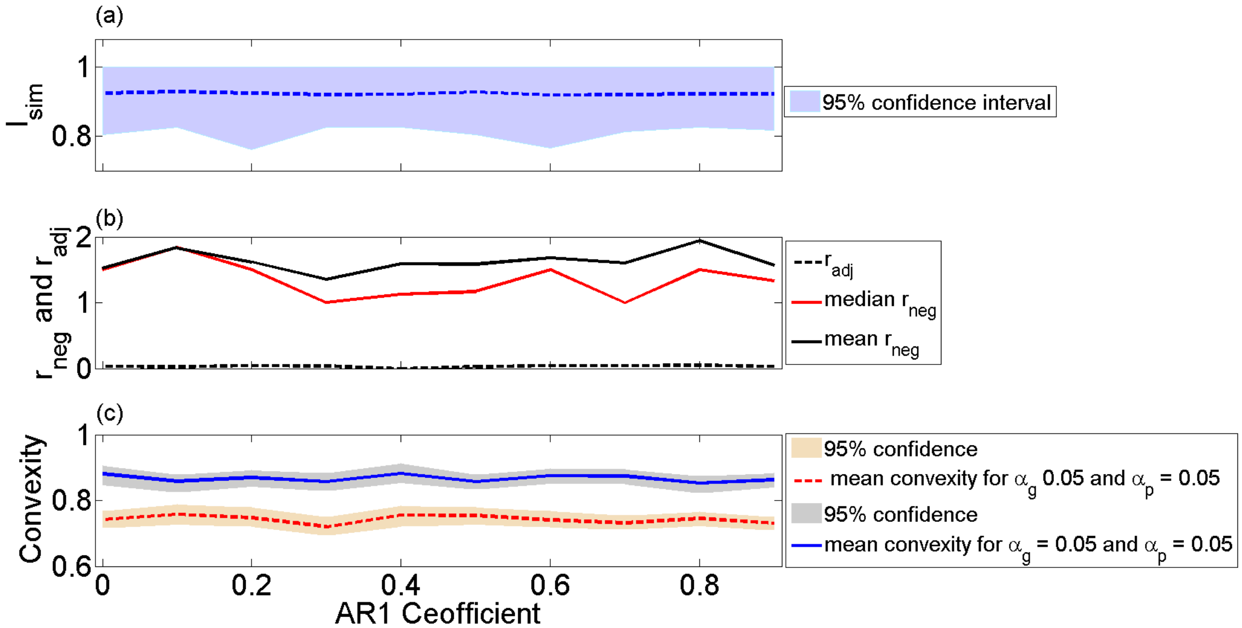

Figure 6. (a) Similarity index between the geometric and areawise tests for different lag-1 autocorrelation coefficients for red-noise processes (see text). (b) Same as (a) except for the ratio between the false positive results of the geometric and areawise tests. The dotted black line represents the ratio of false positive between the two tests when the false discovery rate of the geometric test is controlled at the 0.05 level. (c) Same as (a) but for the mean convexity of 5% pointwise significance patches that are geometrically significant at the 5% level and for the mean convexity of 5% pointwise significance patches that are areawise significant at the 5% level.

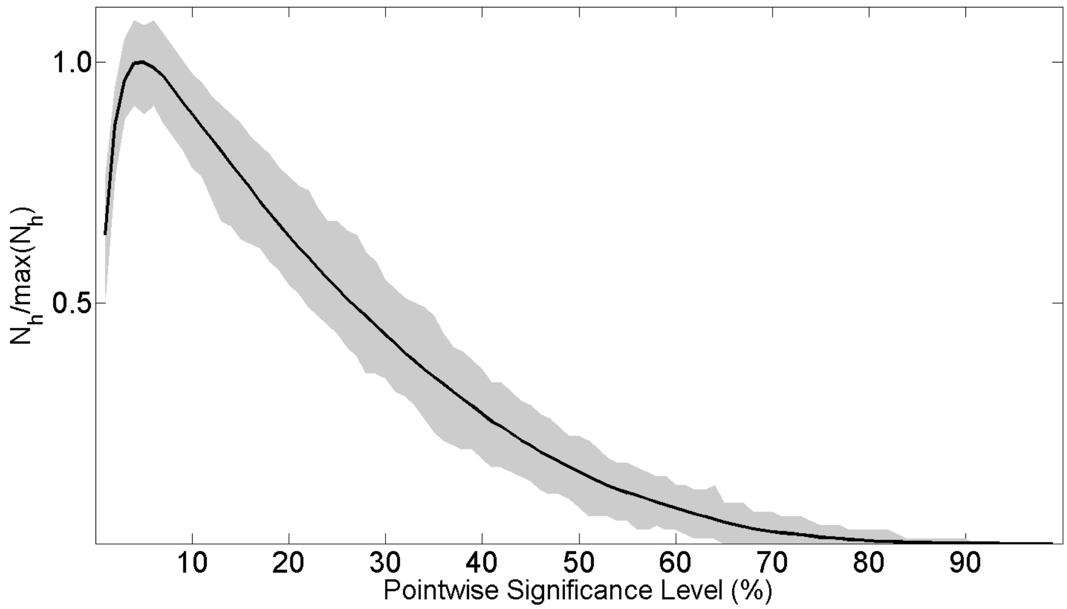

Figure 7. Normalized mean number of holes as a function of pointwise significance level. The number of holes was calculated by generating 10,000 synthetic wavelet power spectra of red-noise processes with fixed autocorrelation coefficients of 0.5 and computing the number of holes Gray shading represents the 95% confidence interval.

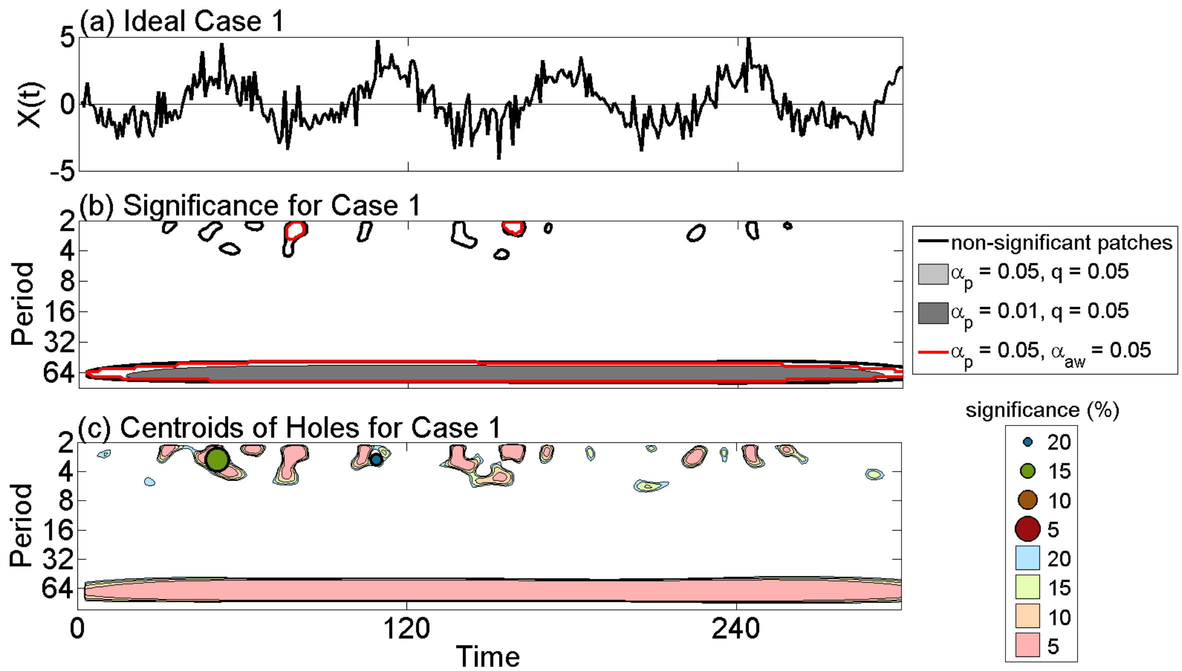

Figure 8. (a) Time series of Case 1, which results from passing a single sinusoidal input with period $\lambda = 64$ through Eq. (16). Gaussian additive white noise with a signal-to-noise of 2 was added to the output response. (b) The significance of wavelet power for Case 1 (see Fig. 3 for details). (c) Topological wavelet diagram corresponding to (b). Points are the centroids of the holes at a given pointwise significance level, where both the color and size of the dots indicate the pointwise significance level at which the hole existed. The shading of the patches corresponds to the pointwise significance level at which the wavelet power coefficient existed, with the color of the shading lighter than the dots for clarity.

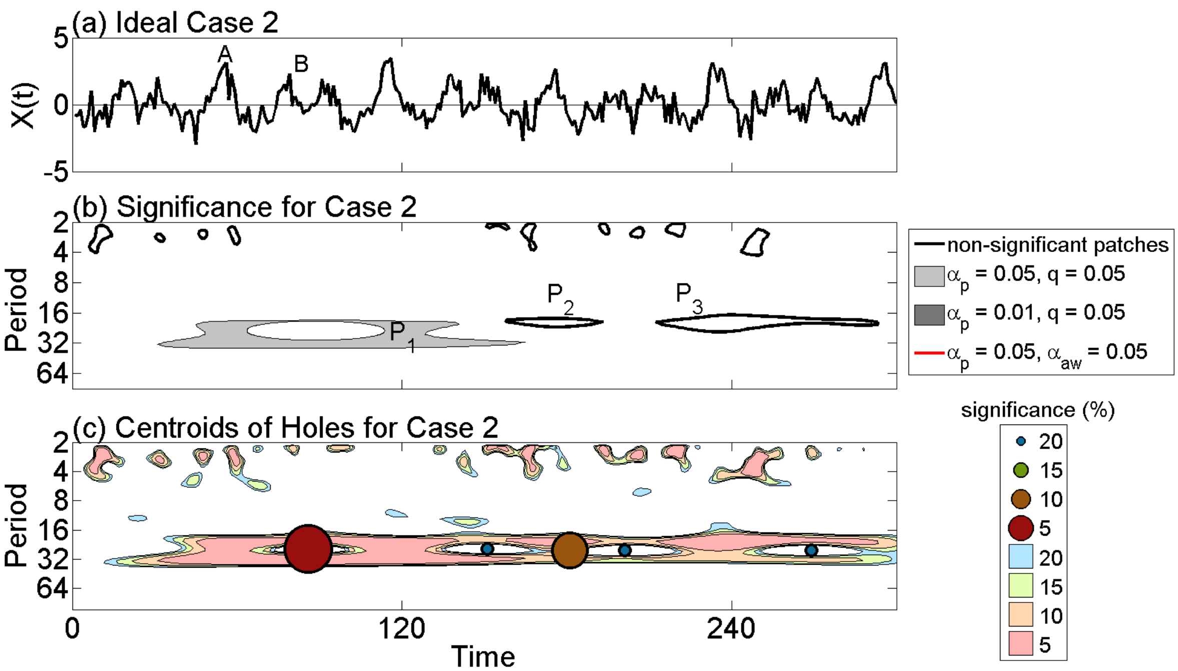

Figure 9. (a) Time series of Case 2. Gaussian additive white noise with a signal-to-noise ratio of 8 was added to the time series. At the point labeled *A*, two oscillations resonate, merging two pointwise significance patches in the wavelet domain. At the point labeled *B* no such resonance occurs and the two significance patches separate. (b) The significance of wavelet power (see Fig. 3 for details). The pointwise significance patch labeled P_1 contains a hole and the pointwise significance patches labeled P_2 and P_3 were falsely deemed insignificant by the geometric and areawise tests. (c) Same as Fig. 8c except for Case 2.

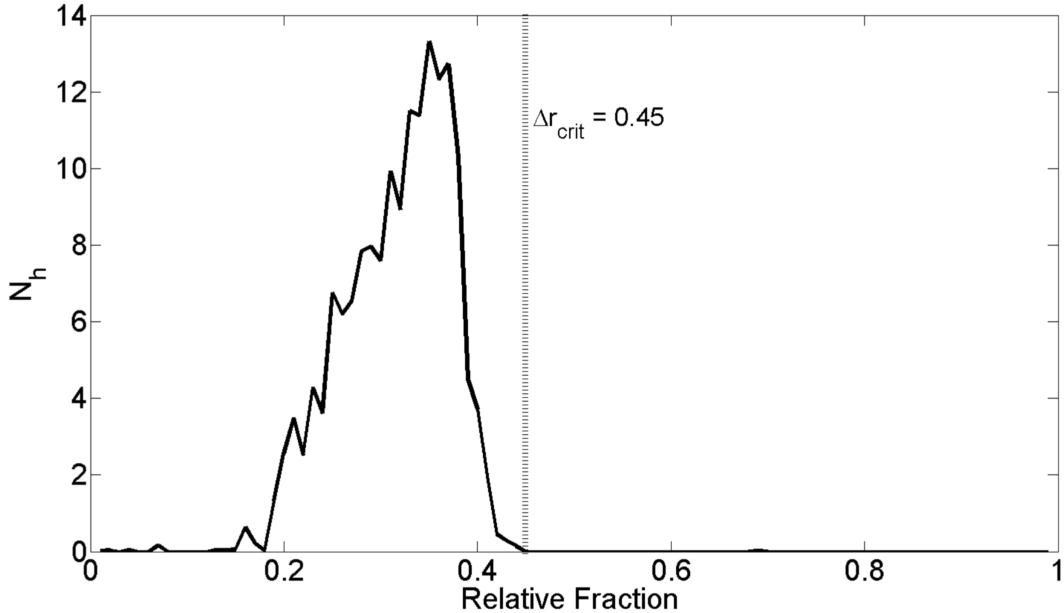

Figure 10. Mean number of holes found in 5% pointwise significance patches as a function of $\Delta r = (f_2 - f_1)/ f_2$ for a sum of two sinusoids with amplitudes equal to unity and frequency components f_1 and f_2 such that $f_2 > f_1 > 0$. Additive white noise with a signal-to-noise ratio of 30 was added to the sum of sinusoids. Pointwise significance was tested against a red-noise background. Dashed line represents the critical value of Δr, the value beyond which holes will rarely occur between oscillations of equal amplitude (set to unity) with frequencies f_1 and f_2.

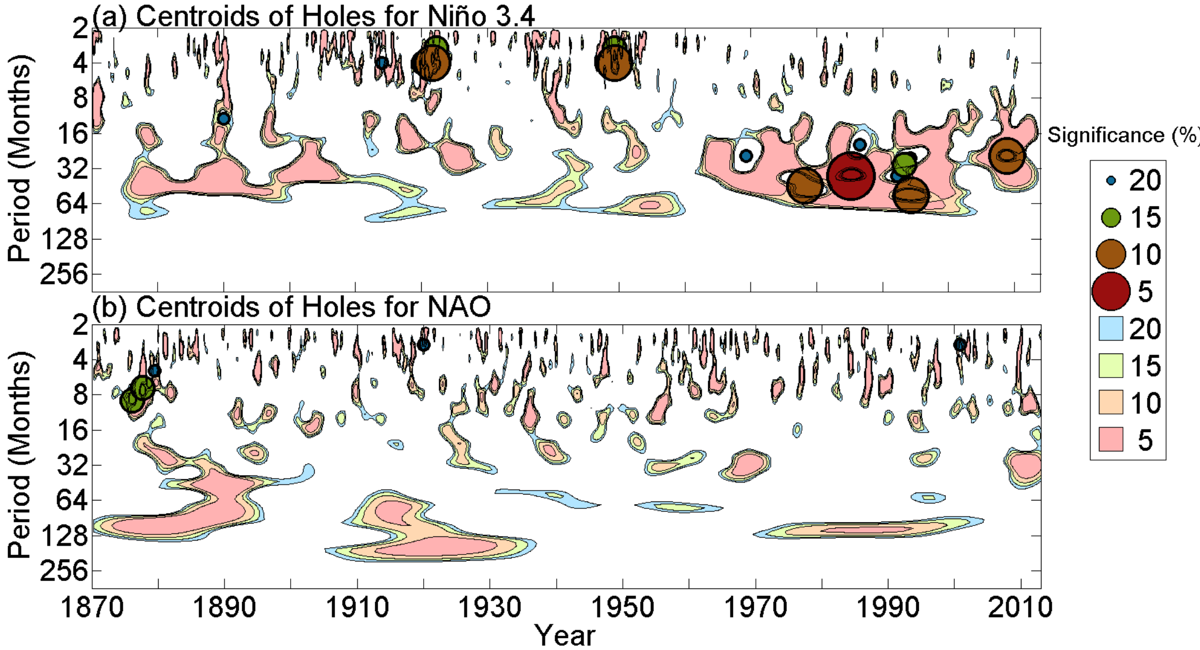

Figure 11. Same as Fig. 8c but for the mean monthly (a) Niño 3.4 and (b) NAO index anomalies for 1870-2013.

Table 1. Fraction of pointwise significance patches containing at least N_h holes as a function of the pointwise significance level calculated from an ensemble of 200,000 significance patches generated from red-noise processes with fixed autocorrelation coefficients equal to 0.5.

Significance level (%)	$\boldsymbol{N_h \geq 1}$	$\boldsymbol{N_h \geq 2}$	$\boldsymbol{N_h \geq 3}$	$\boldsymbol{N_h \geq 4}$
20	2.3×10^{-2}	2.6×10^{-3}	4.0×10^{-3}	0
15	1.0×10^{-2}	5.0×10^{-3}	1.0×10^{-3}	0
10	2.0×10^{-3}	1.0×10^{-3}	0	0
5	3.4×10^{-4}	0	0	0
1	0	0	0	0