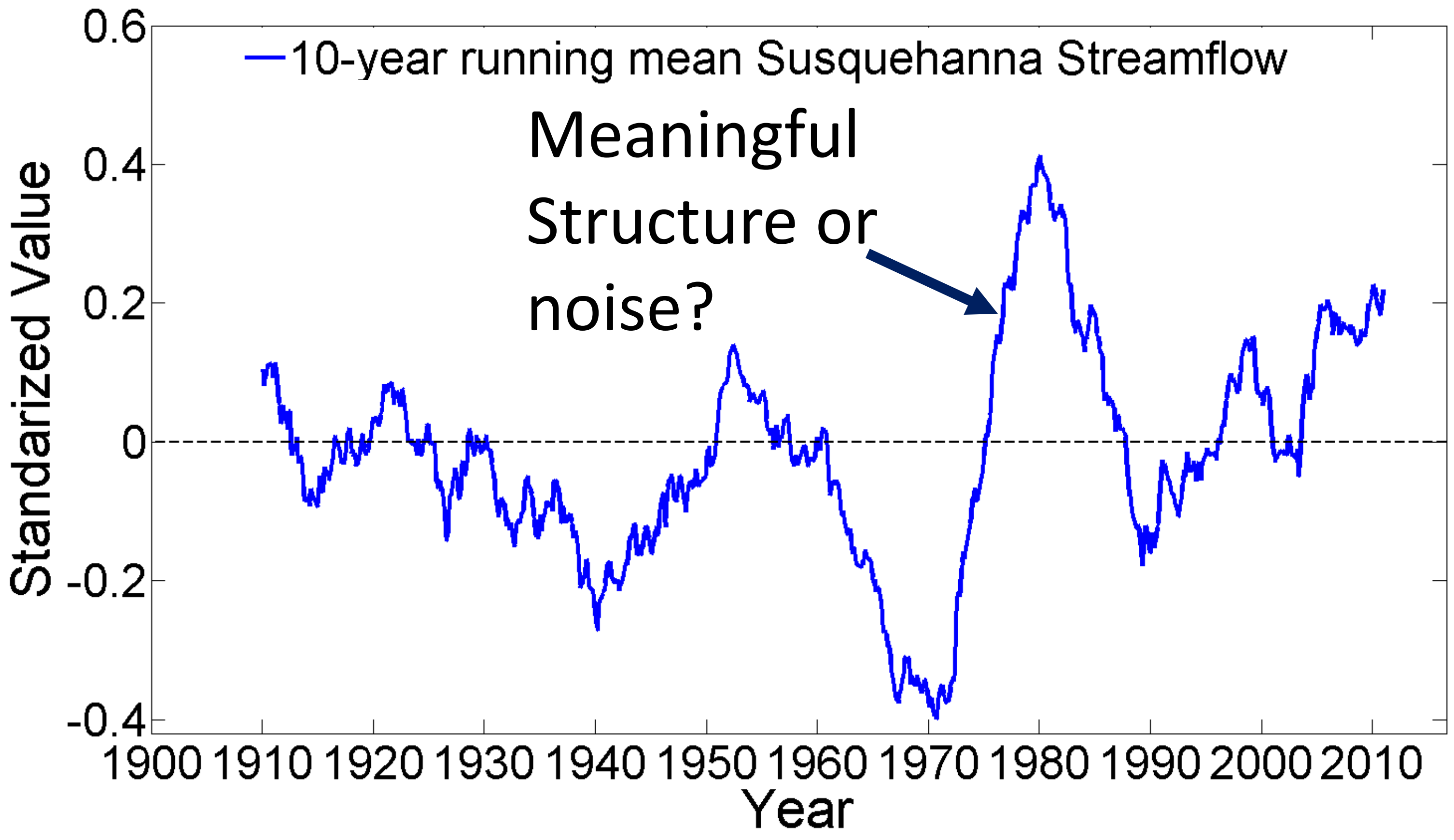
Advances In Wavelet Analysis: Significance Testing Methods and Pitfalls

Justin Schulte
Department of Meteorology
410 Walker Building
jas6367@psu.edu

Outline

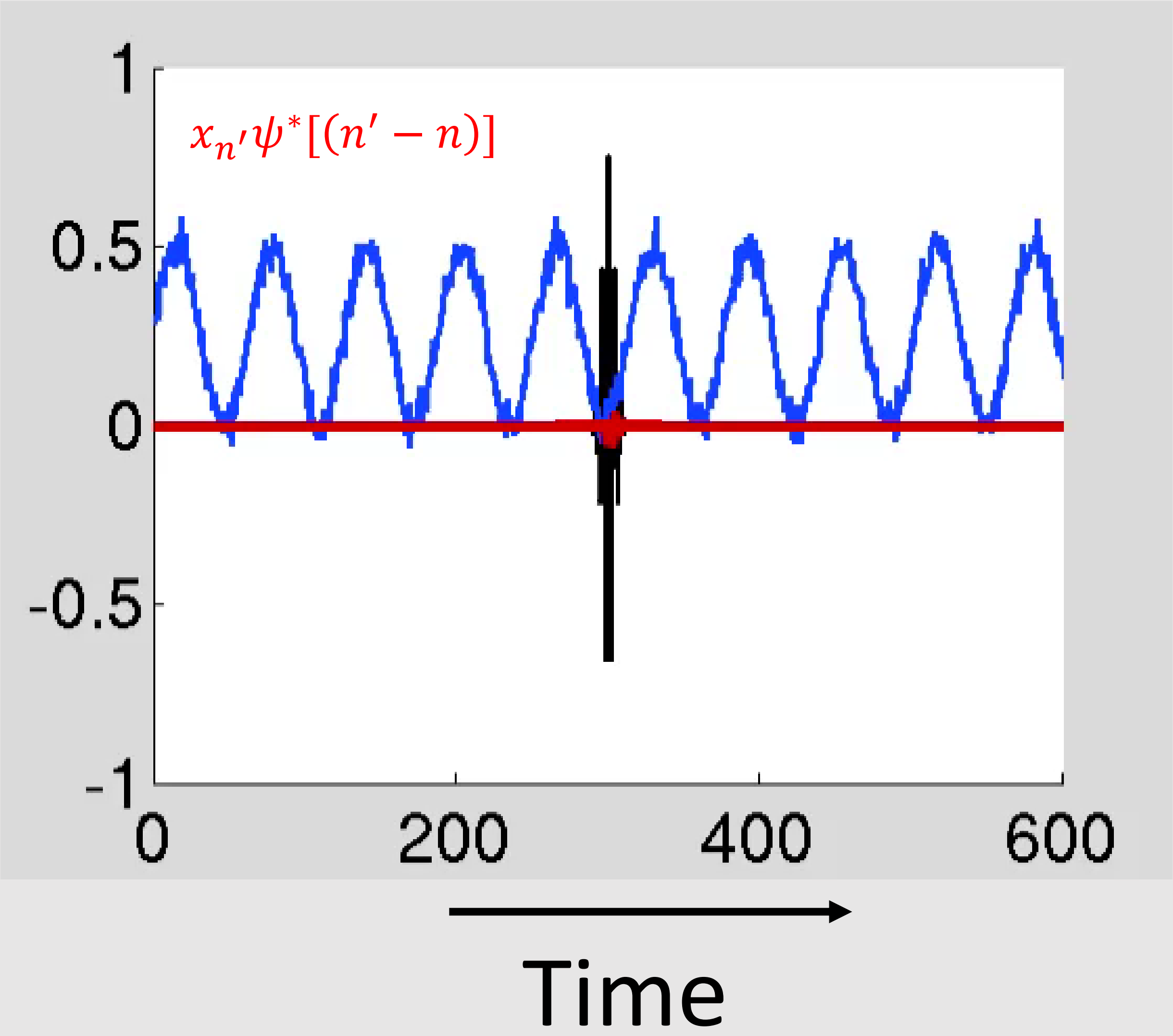
- Brief Introduction to Wavelet Analysis
- Significance Testing Methods
  - Pointwise Significance Testing
  - Areawise Significance Testing
  - Geometric Significance Testing
  - Topological Methods*
- Future Work

Why wavelet Analysis?



- **Extract non-stationary information from a time series**
- **Geophysical time series are non-stationary**
- **Unlike Fourier Analysis, there is no need to select an averaging window**
- **Represent the variance of a time series as a function of time and frequency**

Application of Wavelet Transform



$$\psi(\eta) = \pi^{-1/4} e^{i\omega\eta} e^{-\frac{1}{2}\eta^2}$$

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

$$|W_n(s)|^2 = \text{Wavelet Power}$$

s = wavelet scale
n = local time index
δt = time step
$X = (x_n\ n = 1,\ldots,N)$
Asterisk denotes the complex conjugate

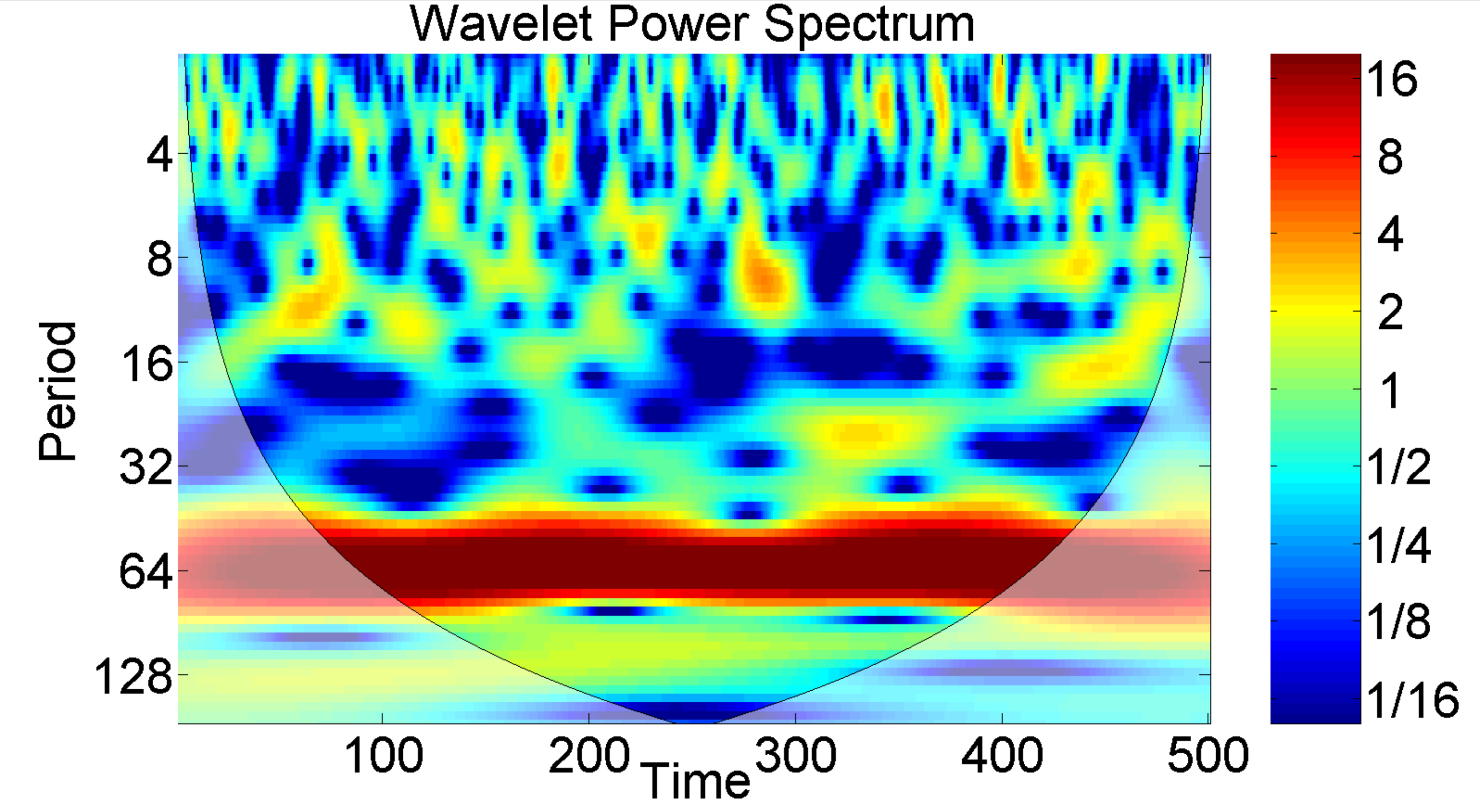
Application of Wavelet Transform

Color → Normalized Variance

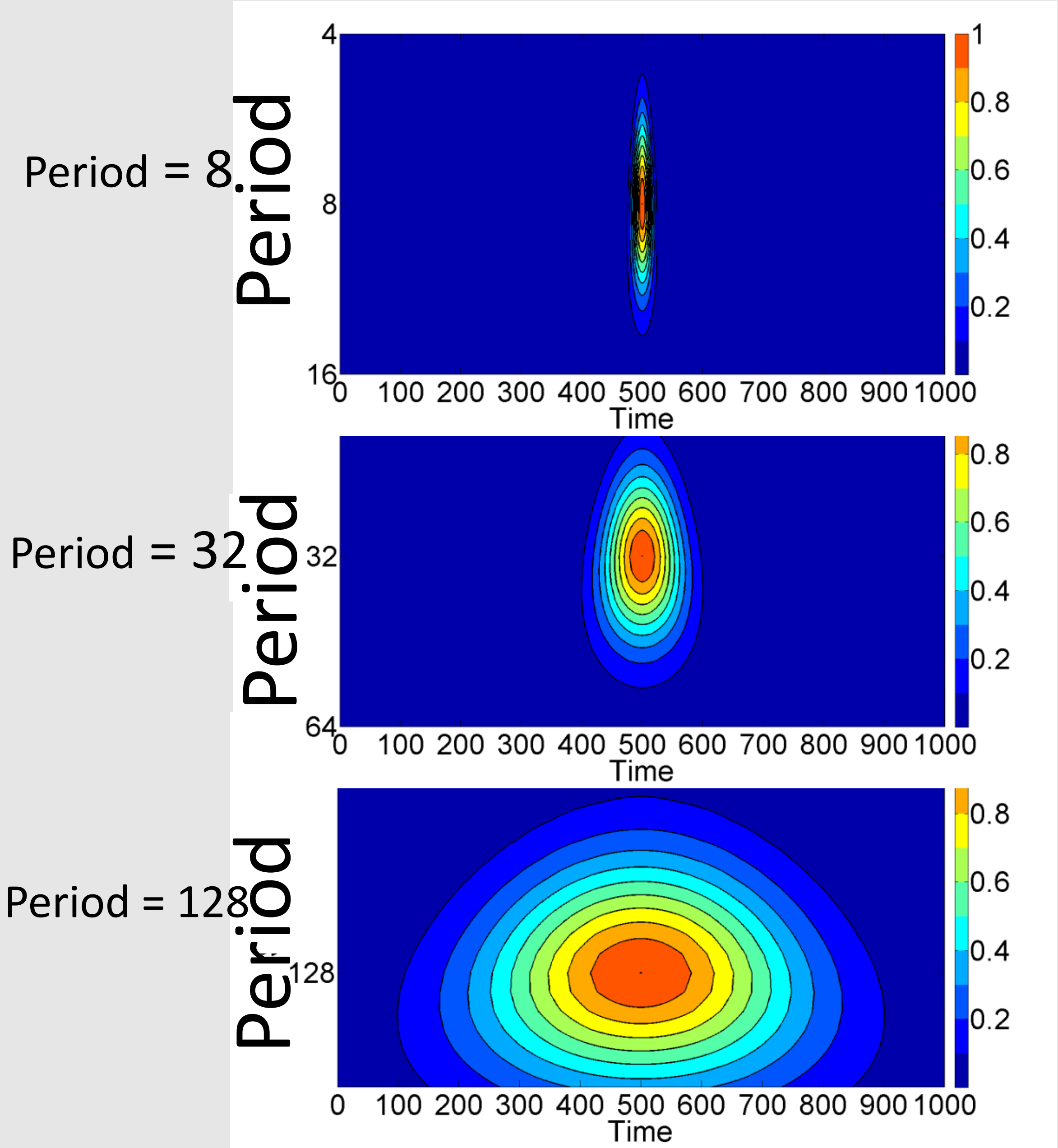
Light Shading → Cone of Influence

Time → Horizontal Axis

Period → Vertical Axis, Increasing Downwards



Reproducing Kernel of Morlet Wavelet

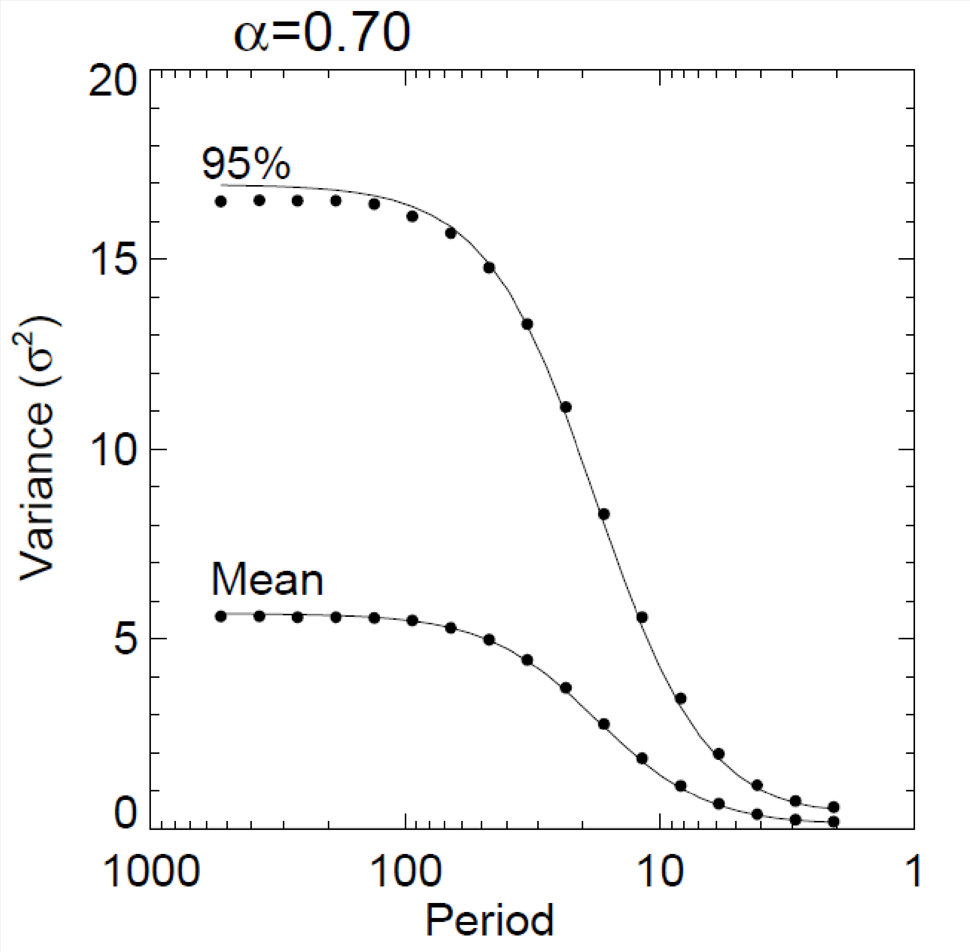


- Represents the intrinsic correlations between adjacent wavelet coefficients
- Represents time-scale uncertainty
- Expands linearly in time direction with scale
- Contours represent critical levels of the reproducing kernel
- Area of reproducing kernel is the area of region enclosed by a given contour
- More wavelet coefficients become correlated at large scales

Pointwise Significance Testing

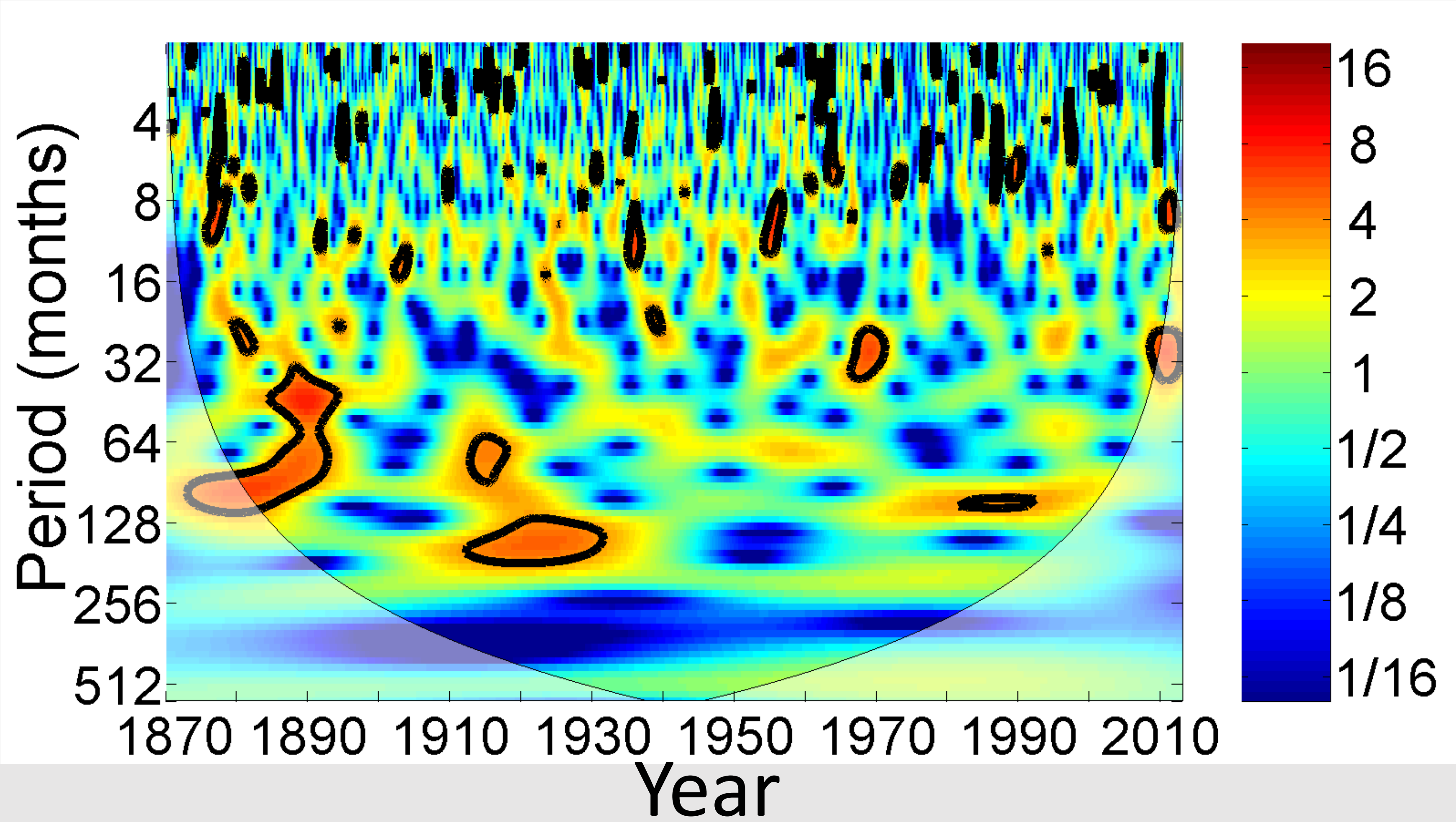
- Need to determine if the wavelet power (or coherence estimate) at each scale and time exceeds a suitable noise background
- Red-noise is an appropriate noise model for climatic time series
- Torrence and Compo (1998) were the first to put wavelet analysis into a statistical framework

Red-noise background spectrum

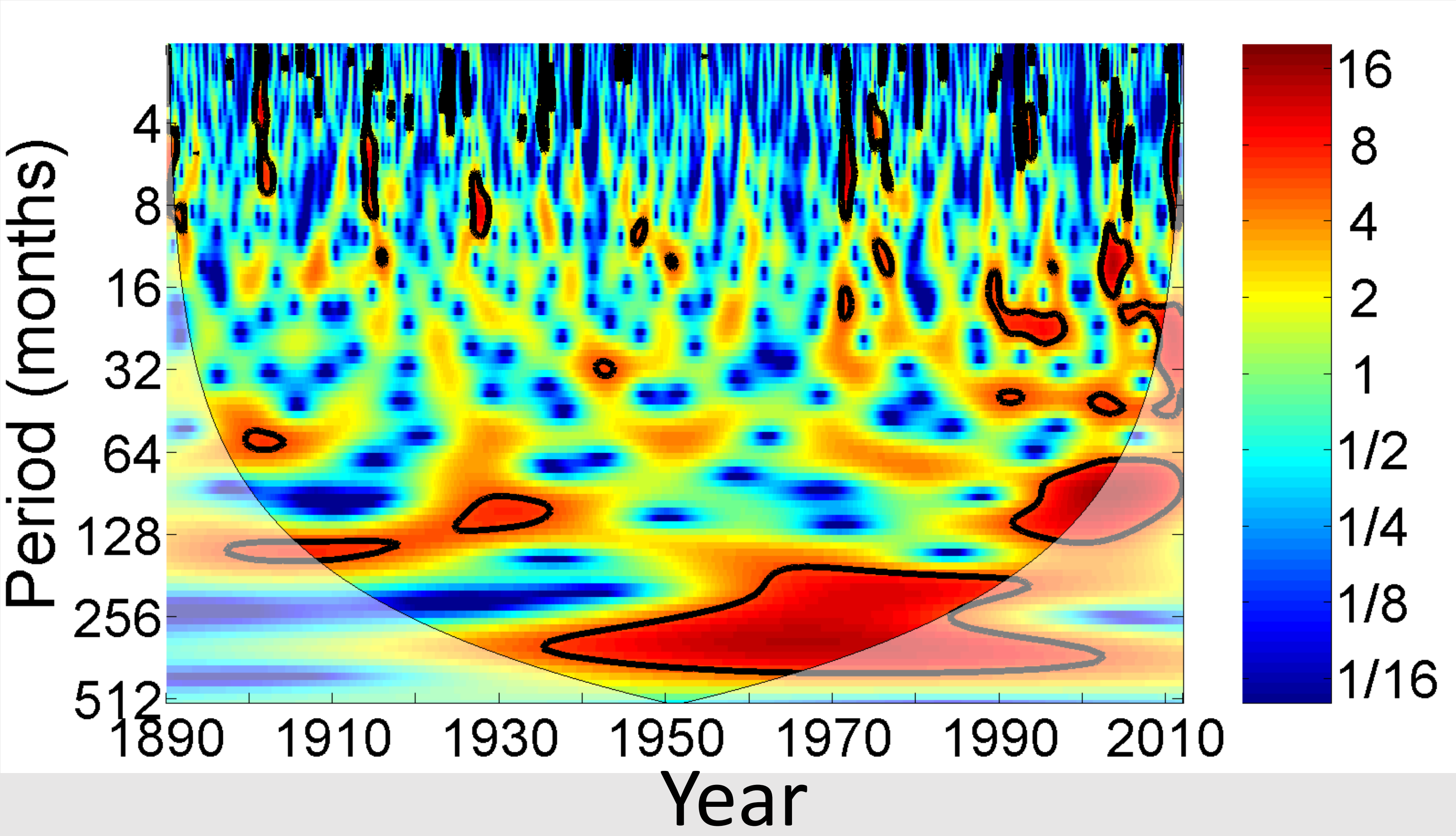


Torrence and Compo (1998)

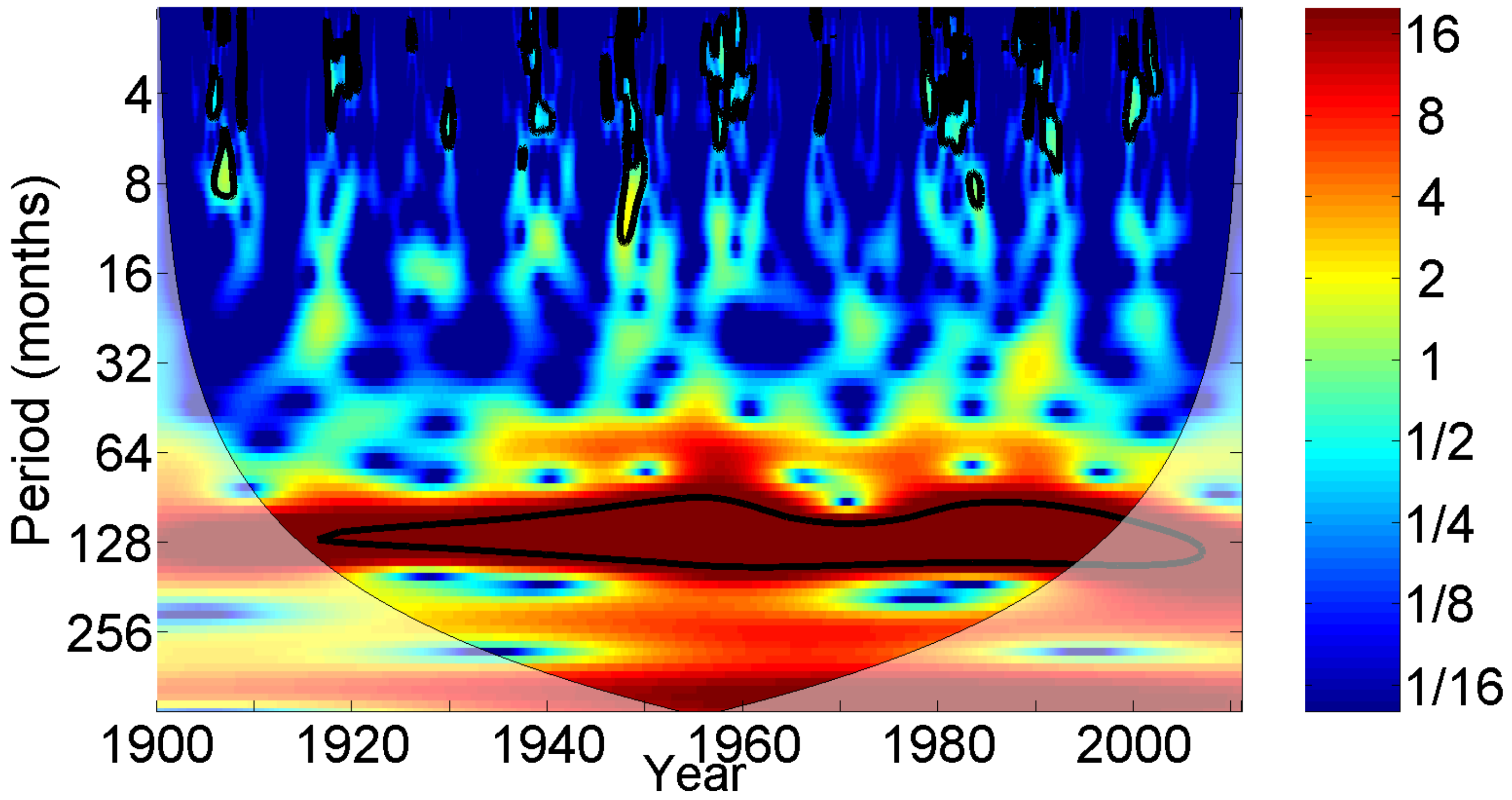
NAO Index 1870-2013



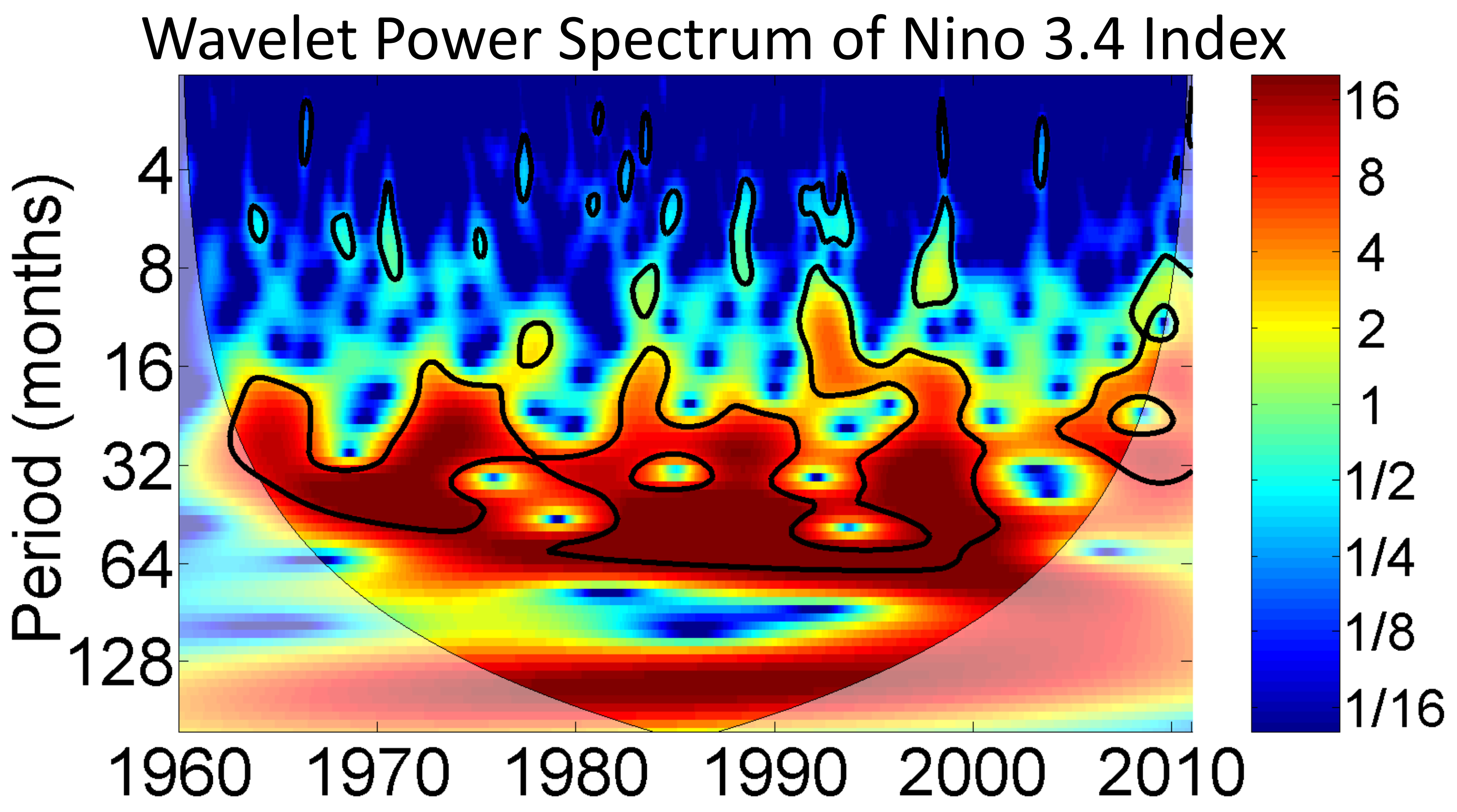
Susquehanna Streamflow 1890-2010



Sunspot Number 1900-2010

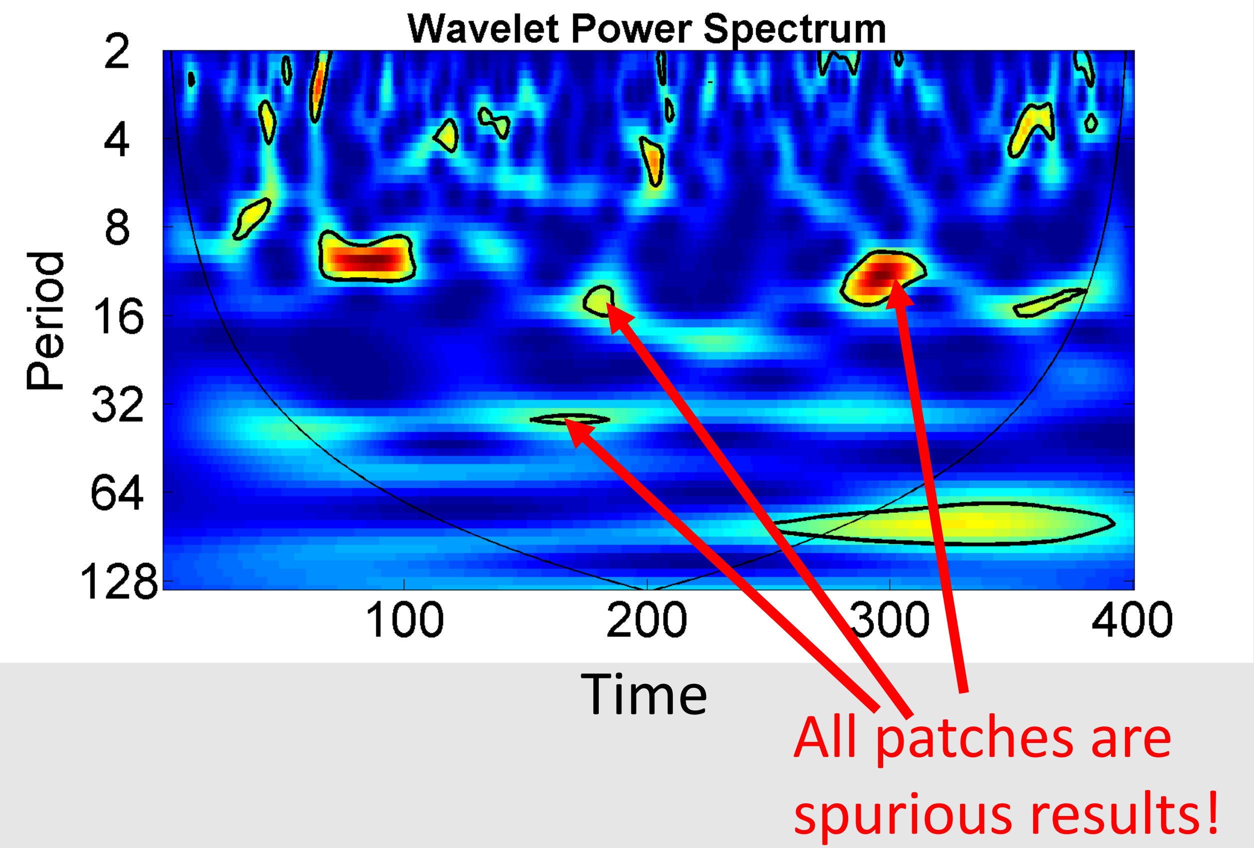


Nino 3.4 Index 1960-2010



Deficiencies of Pointwise Significance Testing

Wavelet Power Spectrum of a white noise process

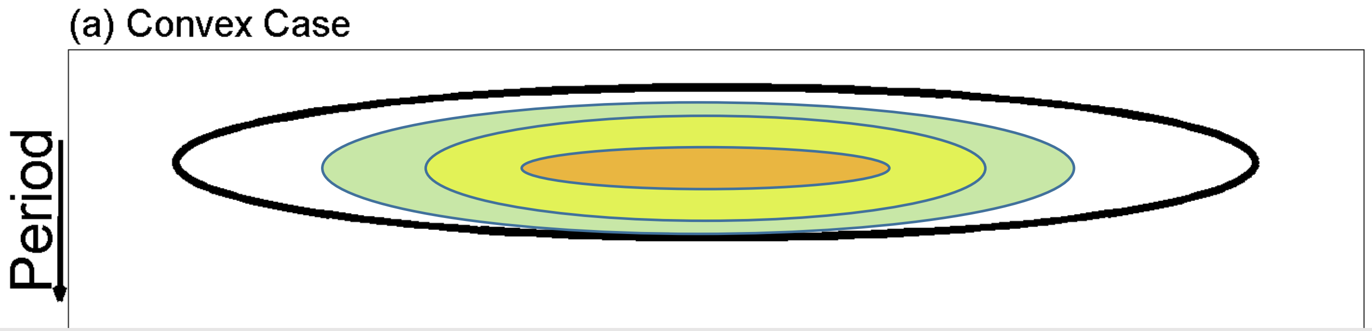


- **The large number of wavelet coefficients being tested simultaneously results in multiple testing issues**
- **Adjacent wavelet coefficients are correlated so spurious results occur in clusters or significance patches**
- **Power spectra of white noise processes contain numerous spurious results**
- **Which patches are artifacts of multiple testing and which are meaningful structures?**

Areawise Significance Testing

- Maraun and Kurths (2004) recognized the limitations of pointwise significance testing
- Maraun et al. (2007) developed an areawise test that reduces the number of spurious results by sorting through patches based on geometry and size
- A point inside a patch is said to be areawise significant if any reproducing kernel dilated (at a certain critical level) according to the scale in question entirely fits into the patch (Maraun et al. 2007)
- $\alpha_{aw} = \langle \frac{\text{Area of Areawise Significant Regions}}{\text{Area of Pointwise Significant Regions}} \rangle = = \langle \frac{A_{aw}}{A_{pw}} \rangle$

Application of the Areawise Test

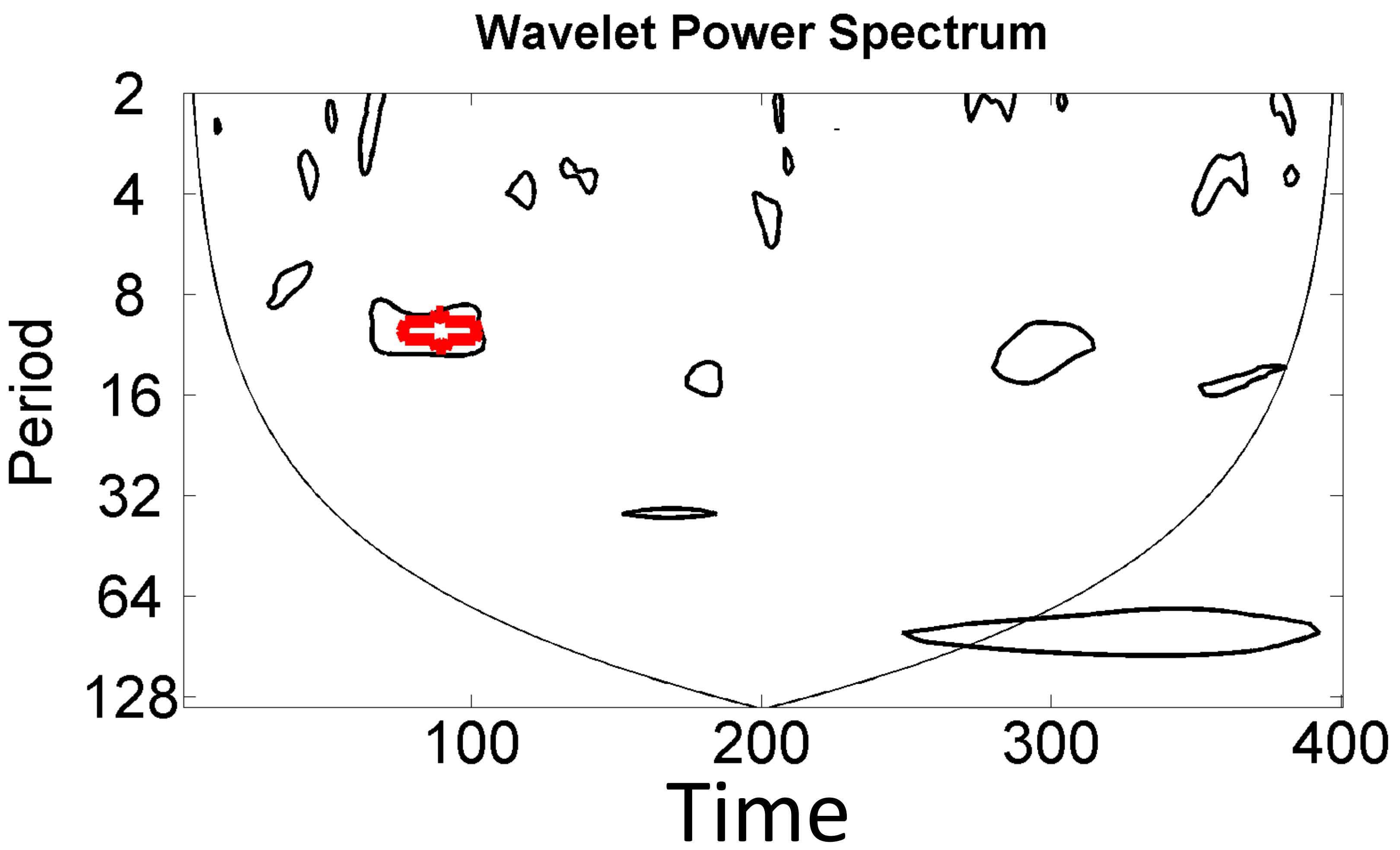


Green = 95% Areawise Significance
Yellow = 90% Areawise Significance
Orange = 85% Areawise Significance

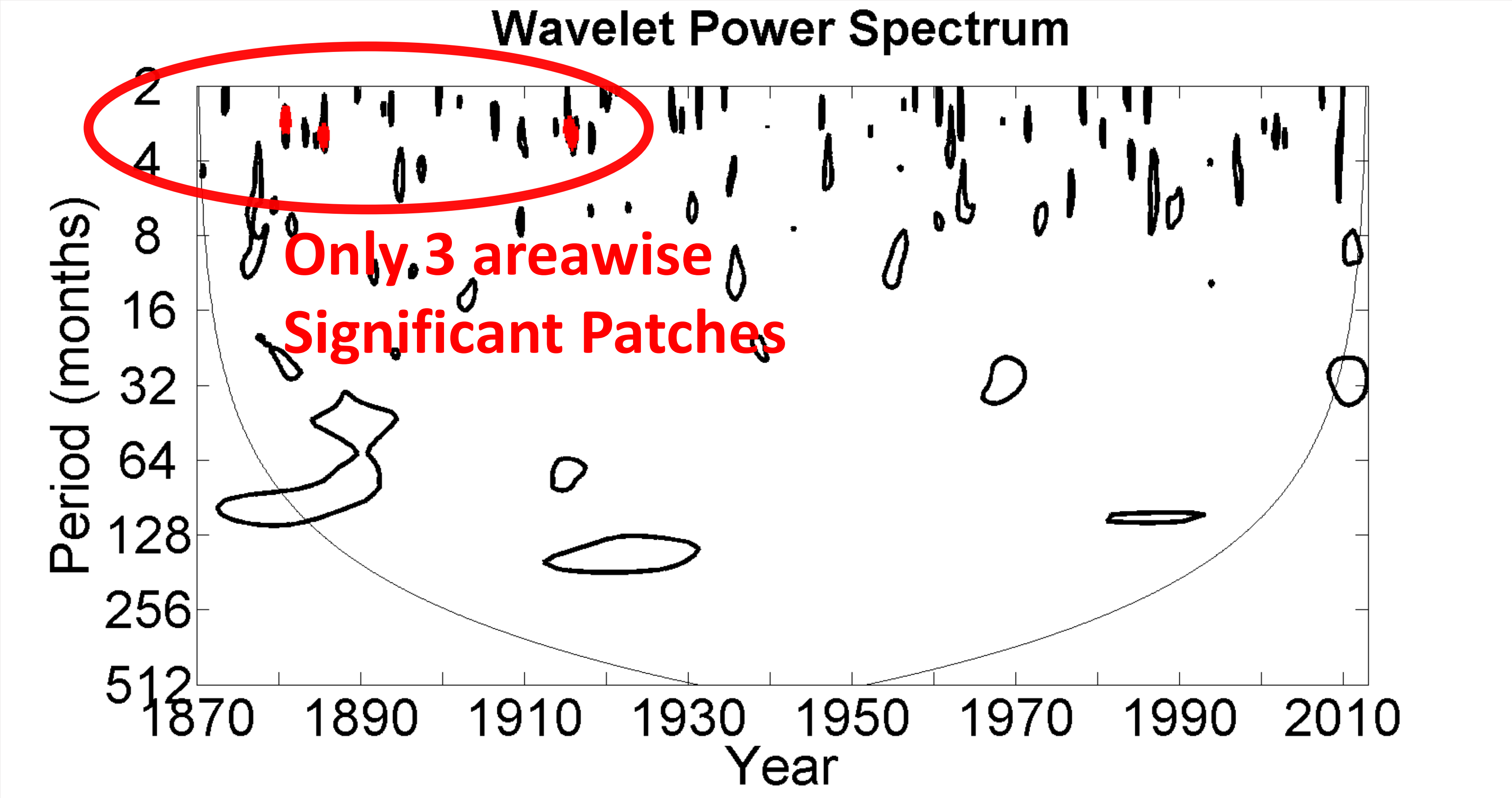
- For the convex case, all three reproducing kernels fit inside the significance patch
- For the non-convex case, only one reproducing kernel corresponding to the 85% significance level fits inside

$$\alpha_{aw} = \left\langle \frac{A_{aw}}{A_{pw}} \right\rangle > \left\langle \frac{A_{aw}}{A_{pw}} \right\rangle > \left\langle \frac{A_{aw}}{A_{pw}} \right\rangle$$

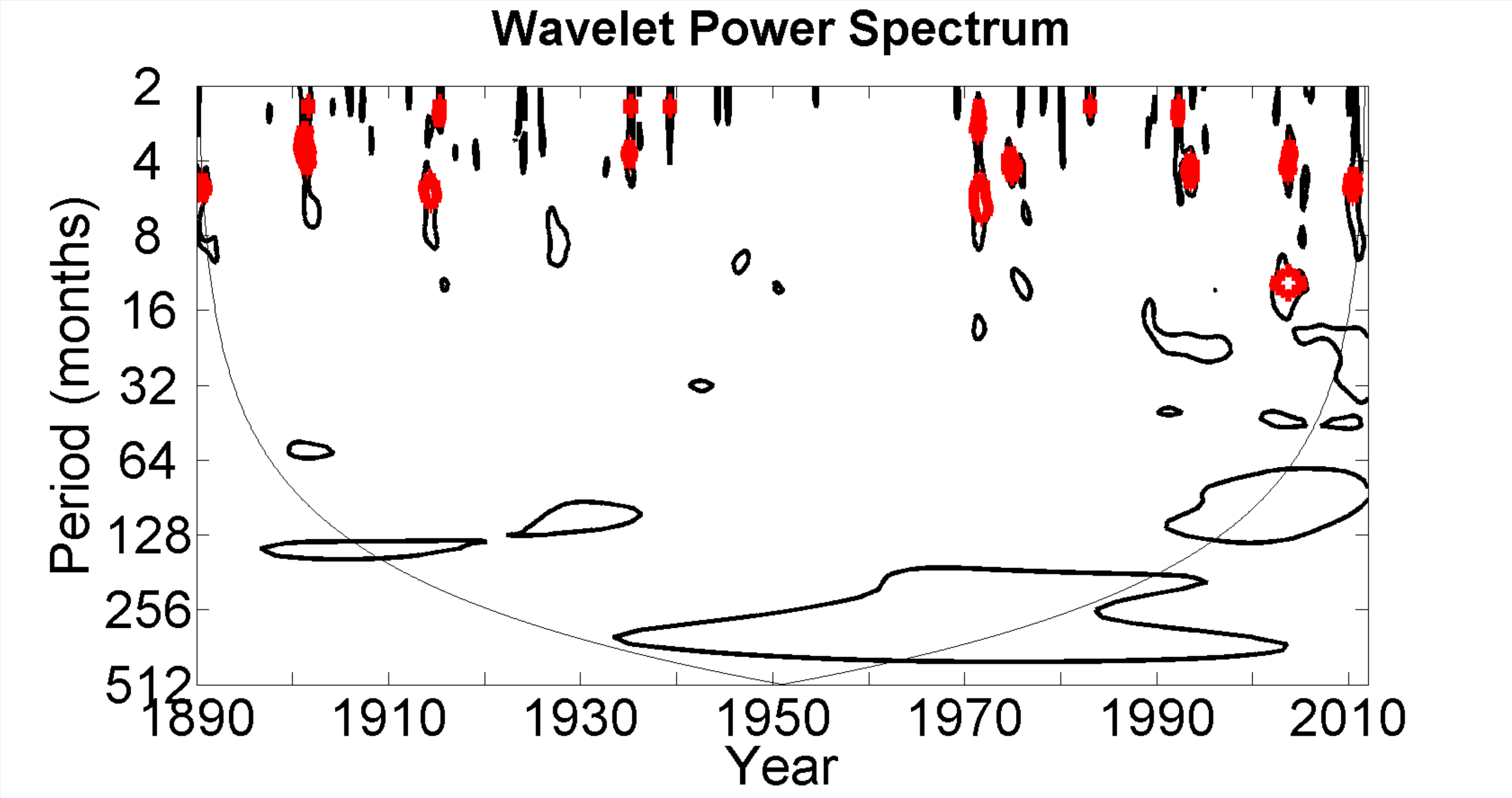
Application of Areawise Test for White Noise Process



Application of Areawise Test for NAO Index



Application of Areawise Test for Susquehanna Streamflow



Deficiencies of Areawise Testing

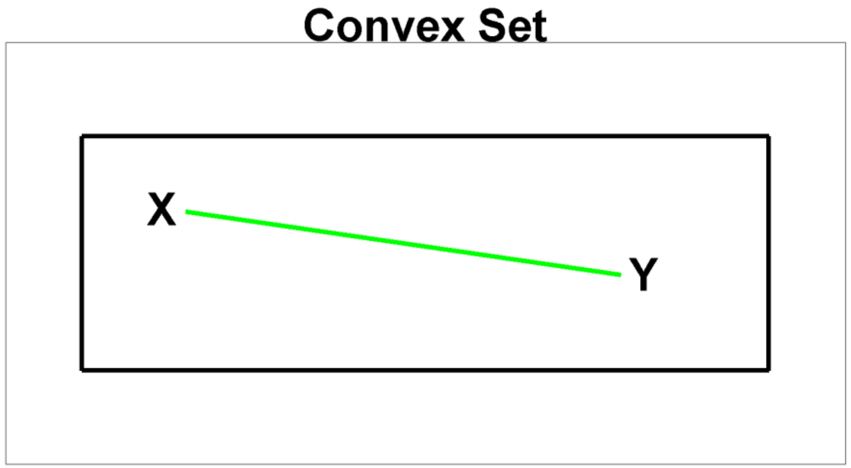
- The significance level of the test is difficult to calculate
- A new significance level must be calculated for different pointwise significance levels, wavelet spectra, and Mother wavelets
- The calculation of each significance level involves a stochastic root-finding algorithm
- The areawise test assumes that the geometry of a patch is locally consistent with the reproducing kernel
- Very long but thin patches may spuriously be deemed insignificant
- Very wide (in scale) and short patches may spuriously be deemed insignificant

Geometric Significance Testing

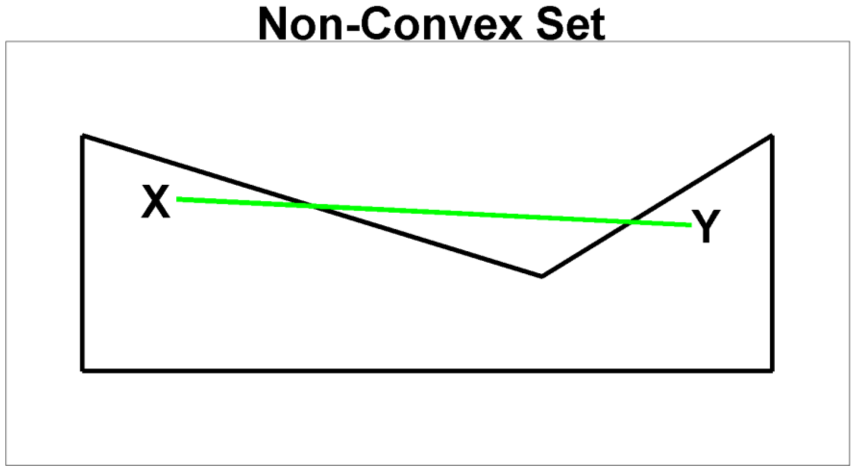
- Schulte et al. (2014) developed a geometric test that sorts through patches based on area and geometry
- The assumption about the geometry of the patch is relaxed
- The algorithm is more computationally efficient than the areawise test
- *p*-values are easily obtained
- The test can be readily applied to patches at various pointwise significance levels and to patches in various wavelet spectra
- The test can be applied to wavelet spectra obtained using different mother wavelets.

Definitions

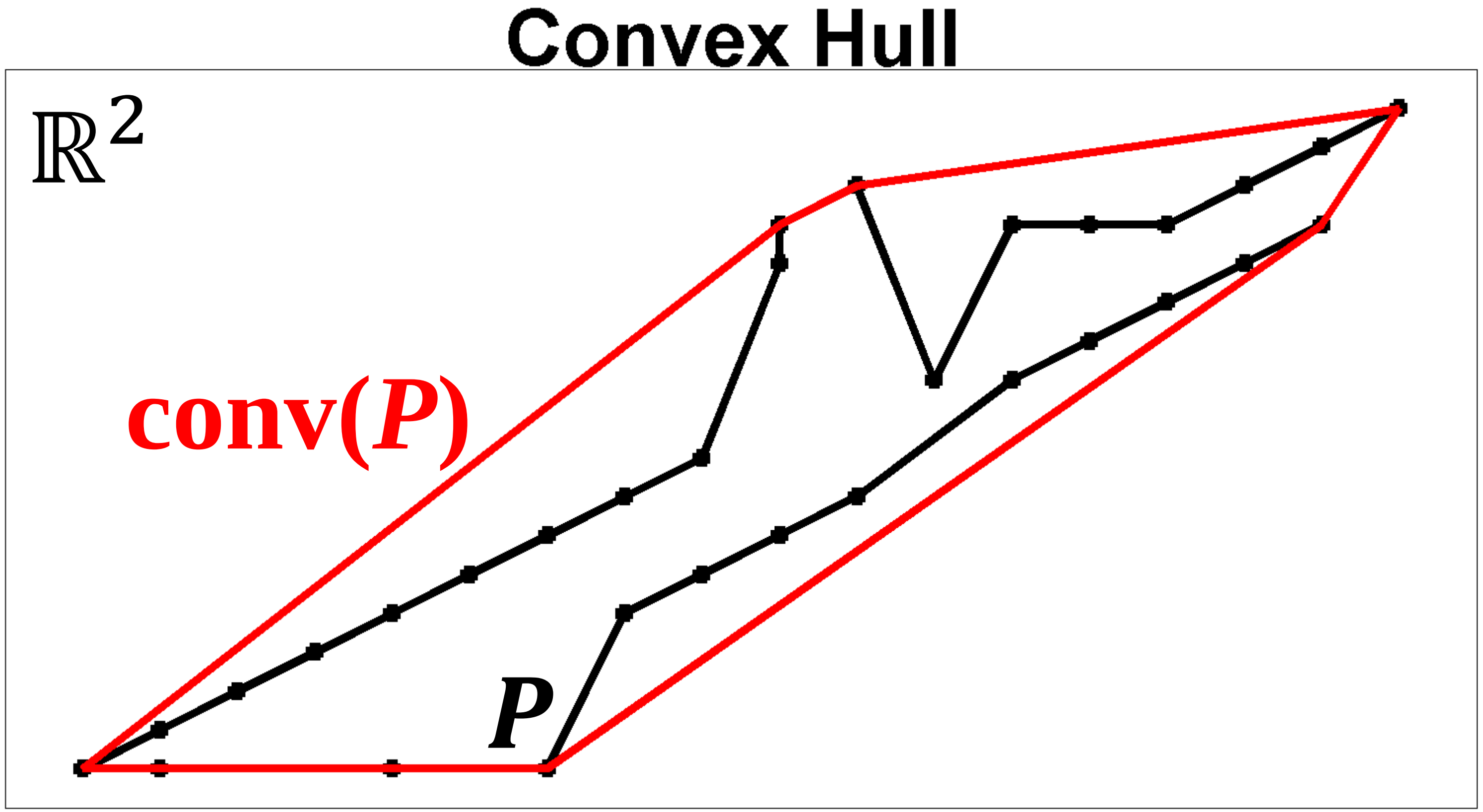
Convex



Non-Convex Set

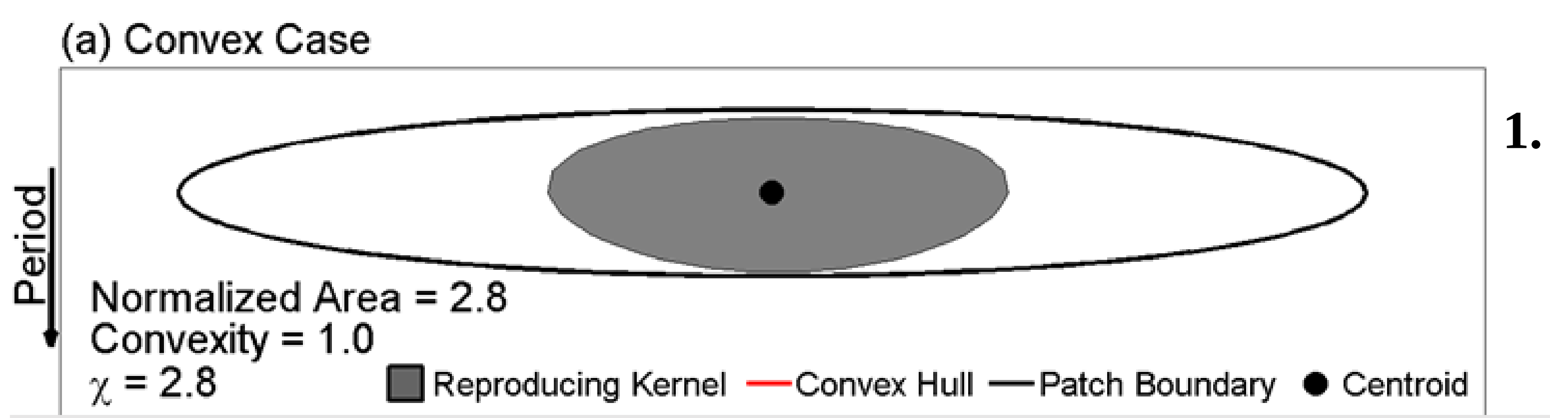


Definitions



$$\mathbf{conv}(\boldsymbol{P}) = \bigcap \{ \boldsymbol{P}' \subseteq \mathbb{R}^2 : \boldsymbol{P} \subseteq \boldsymbol{P}', \boldsymbol{P}' \; \boldsymbol{convex} \}$$

Algorithm



1. **Normalized Area** $= \dfrac{\textbf{Area of Patch}}{\textbf{Area of Reproducing Kernel}}$

↓

2. Compute Convex Hull

↓

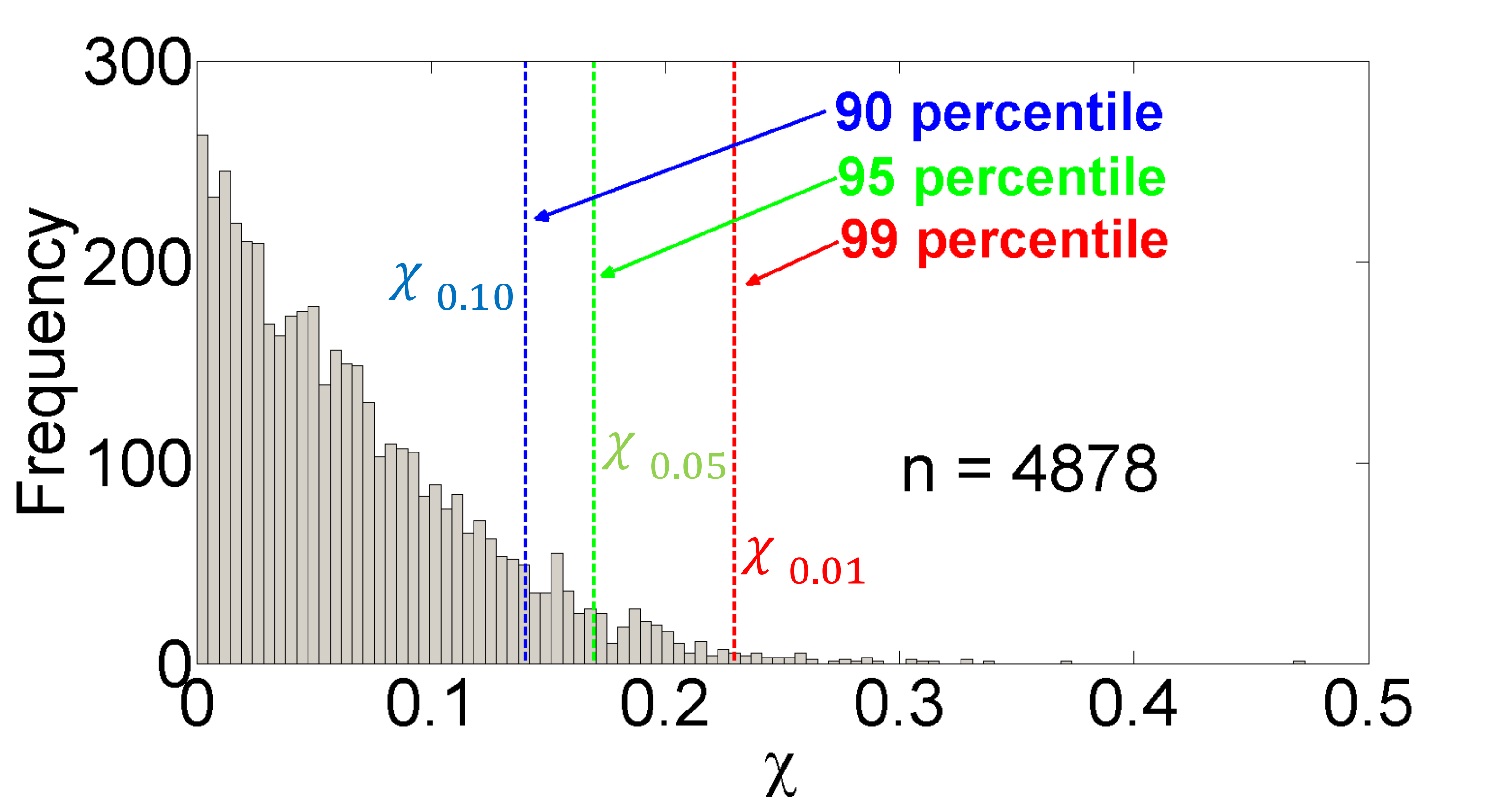
3. **Convexity** $= \dfrac{\textbf{Area of Patch}}{\textbf{Area of Convex Hull}}$

↓

4. χ **= (Convexity)(Normalized Area)**

Schulte et al. (2014)

Null Distribution



1. Generate a null distribution of χ

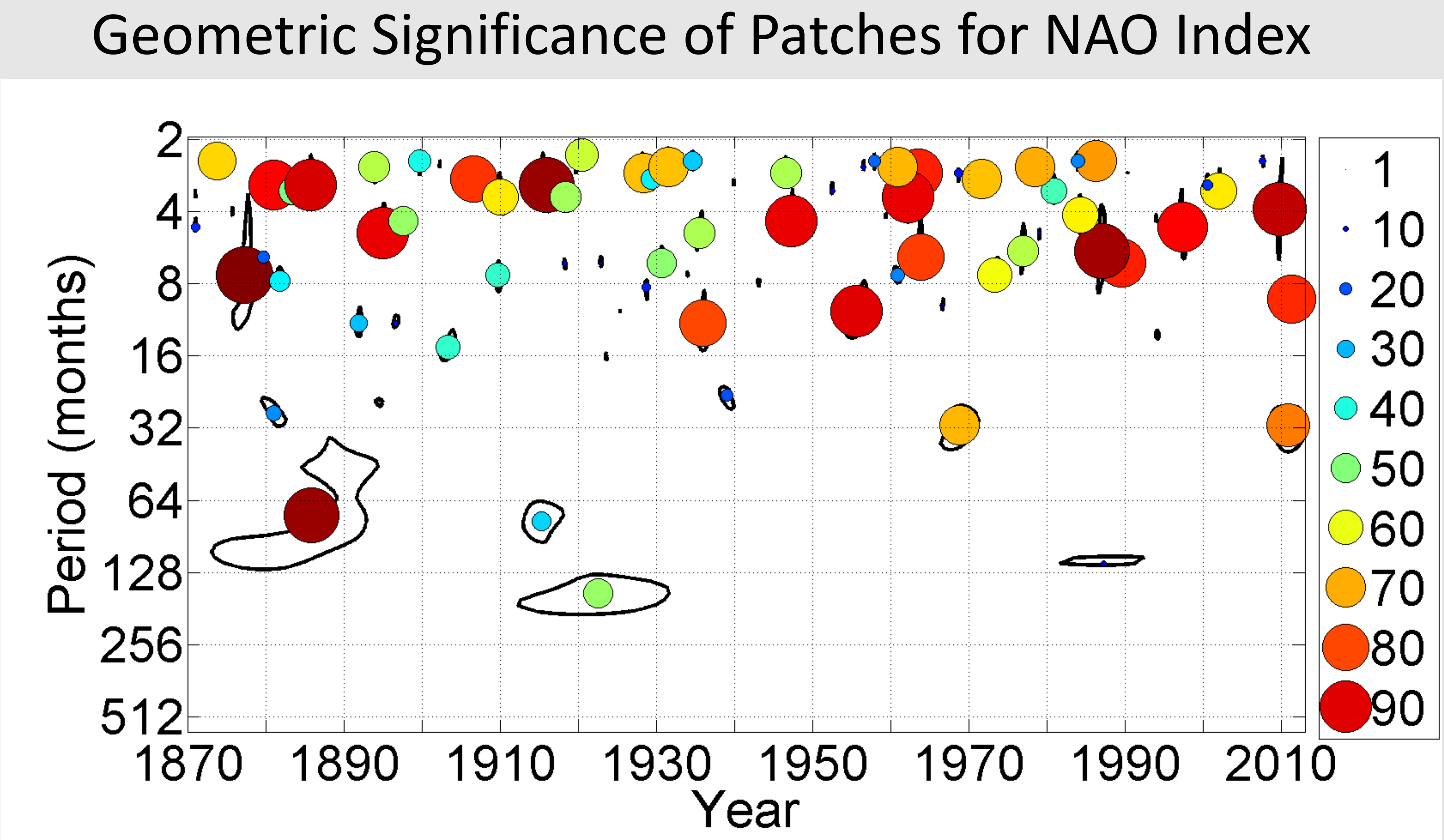
2. Find the value of χ_{α} corresponding to the 100(1-α)th percentile of the distribution

3. α corresponds to the significance level of the test

Geometric Significance Test

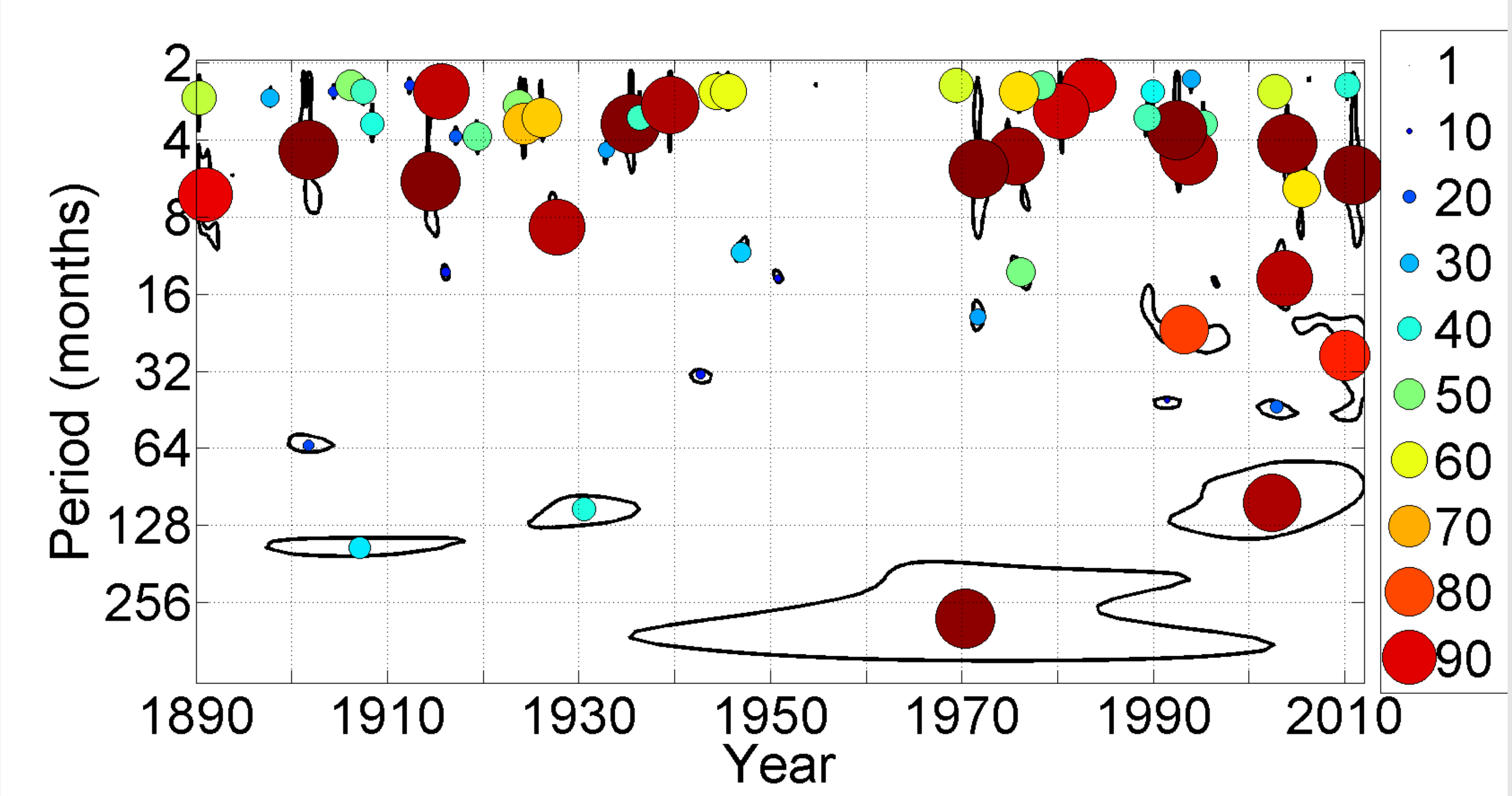
- *p*-value for the test is defined as follows: *p*-value = $P(\chi_{patch} \geq \chi_{\alpha}; H_0)$
- *p*-value for the test is the probability that we obtain the observed value of the test statistic or a value that is more extreme in the direction of the alternative hypothesis calculated when the null hypotheses is true.
- H_0: The patch was generated from a stochastic fluctuation ($\chi_{patch} \leq \chi_{\alpha}$)
- H_1: The patch was not generated from a stochastic fluctuation ($\chi_{patch} > \chi_{\alpha}$)
- Reject H_0 at the α significance level if *p*-value < α

Geometric Significance Testing



Geometric Significance Testing

Geometric Significance of Patches for Susquehanna Streamflow



Summary

- Wavelet analysis is a powerful for feature extraction of geophysical time series
- Pointwise significance testing can be used to determine what features exceed a suitable noise background
- Areawise testing can dramatically reduce the number of spurious results from the pointwise test
- Geometric testing can also reduce the number of spurious results, while also being more computationally efficient

Future Work

- Development of wavelet methods for non-Gaussian and non-linear time series (manuscript in preparation)
- Development of theory to better understand significance testing in wavelet analysis (in progress)
- Development of A Matlab software Package (some software available now)

References

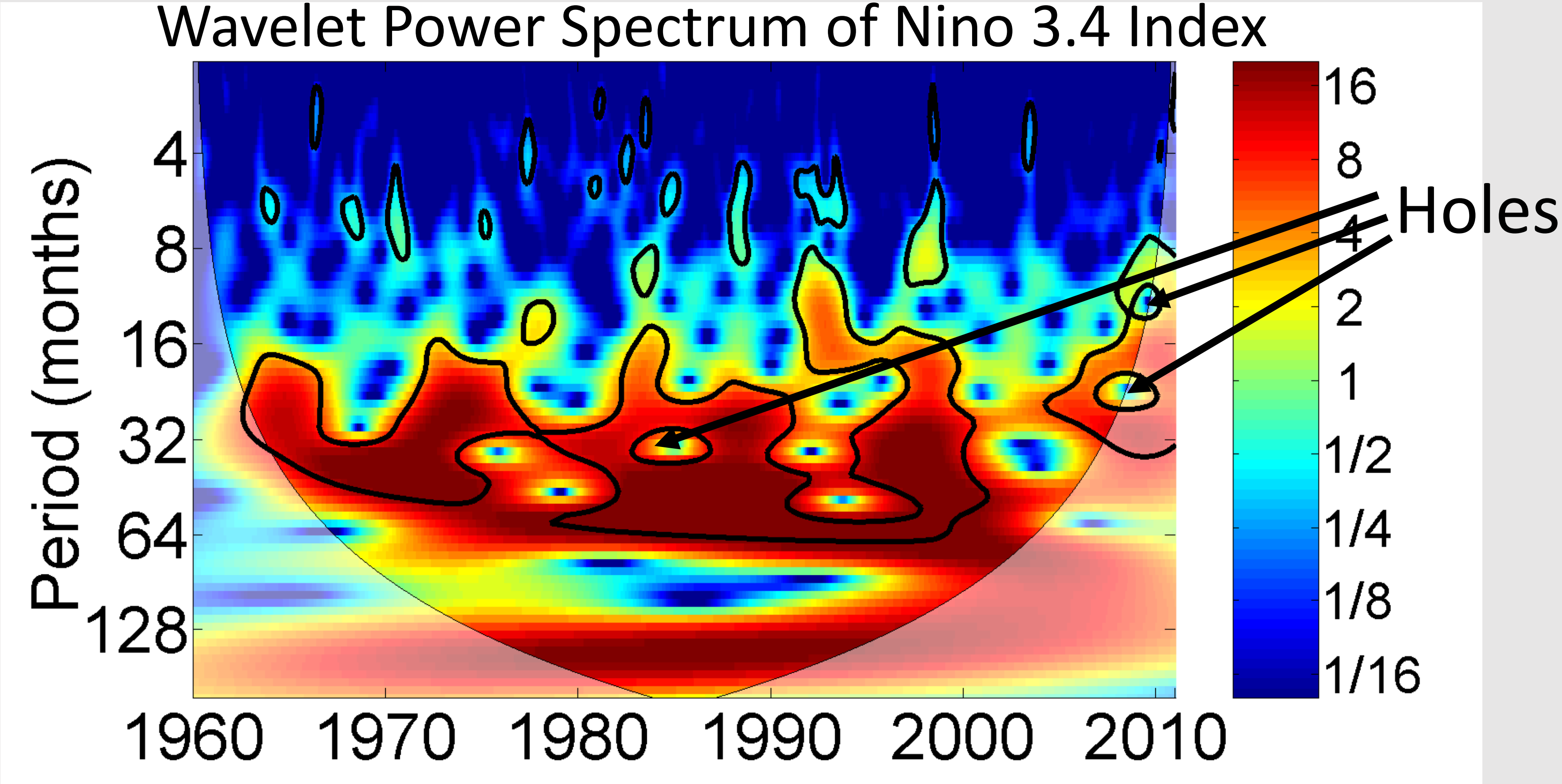
- Edelsbrunner, H. and Harer, J.: Persistent Homology – A Survey, Cotemp. Math., 12, 1-26, 2010.
- Maraun, D. and Kurths, J.: Cross Wavelet Analysis: Significance Testing and Pitfalls, Nonlin. Processes Geophys., 11, 505-514, 2004.
- Maraun, D., Kurths, J., and Holschneider, M.: Nonstationary Gaussian Processes in Wavelet Domain: Synthesis, Estimation, and Significance Testing, Phys. Rev. E, 75, doi: 10.1103/PhysRevE.75.016707, 2007.
- Schulte, J. A., Duffy, C., and Najjar, R. G.: Geometric and topological approaches to significance testing in wavelet analysis, Nonlin. Processes Geophys. Discuss., 1, 1331-1363, doi:10.5194/npgd-1-1331-2014, 2014.
- Torrence, C. and Compo, G. P.: A Practical Guide to Wavelet Analysis, Bull. Am. Meteorol. Soc., 79, 61–78, 1998.

Topological Significance Testing

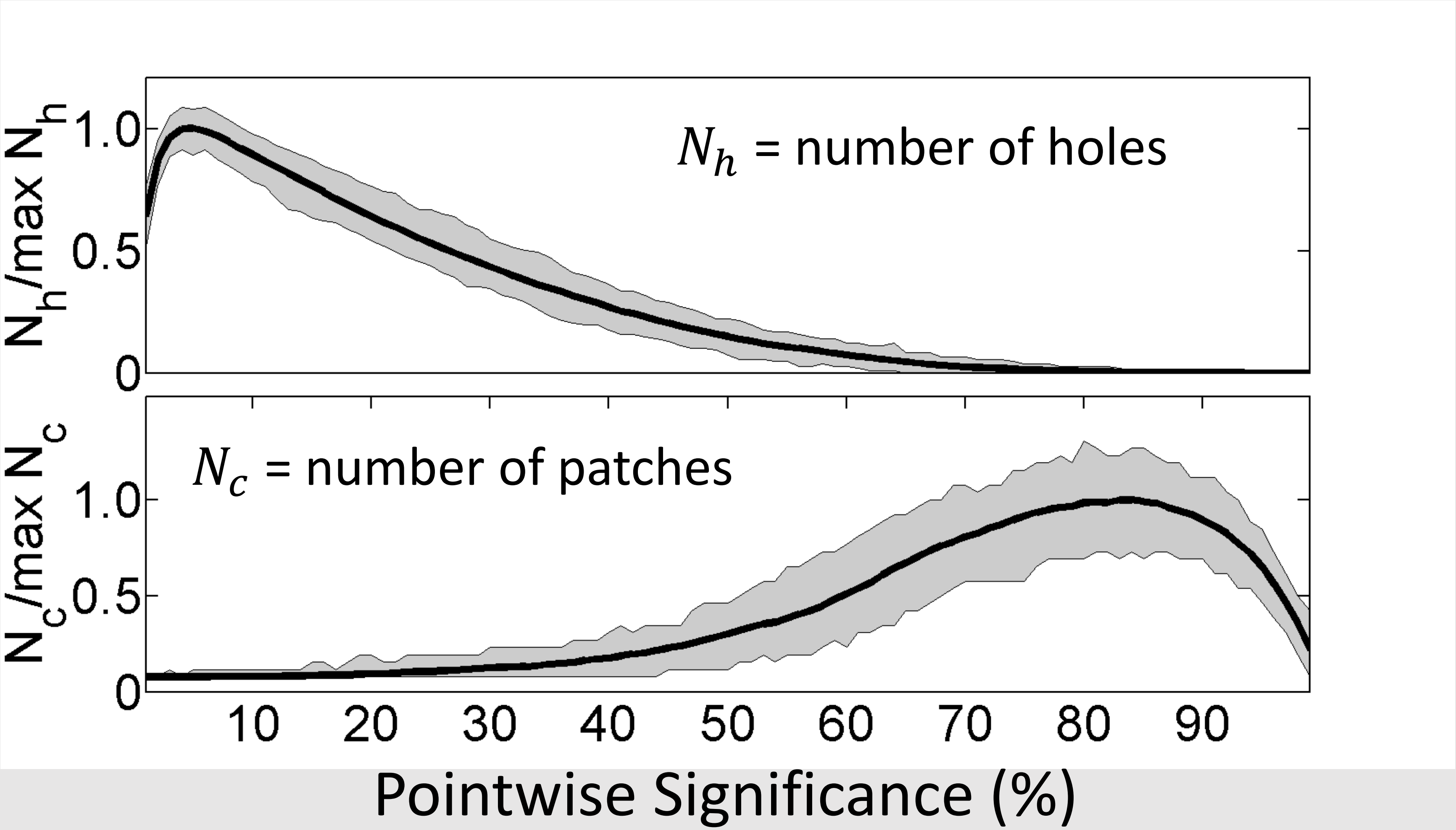
Topological Significance Testing

- The assumption about the geometry of patches is relaxed
- May reveal information about a time series undetected by the areawise and geometric tests
- Inspired by recent work in applied algebraic topology (Edelsbrunner, H. and Harer, 2010)
- Topological features remain after continuous deformations (homeomorphisms)
- Such features are called topological invariants
- An example of a topological feature is a hole, which is related to the simply-connectedness of a region

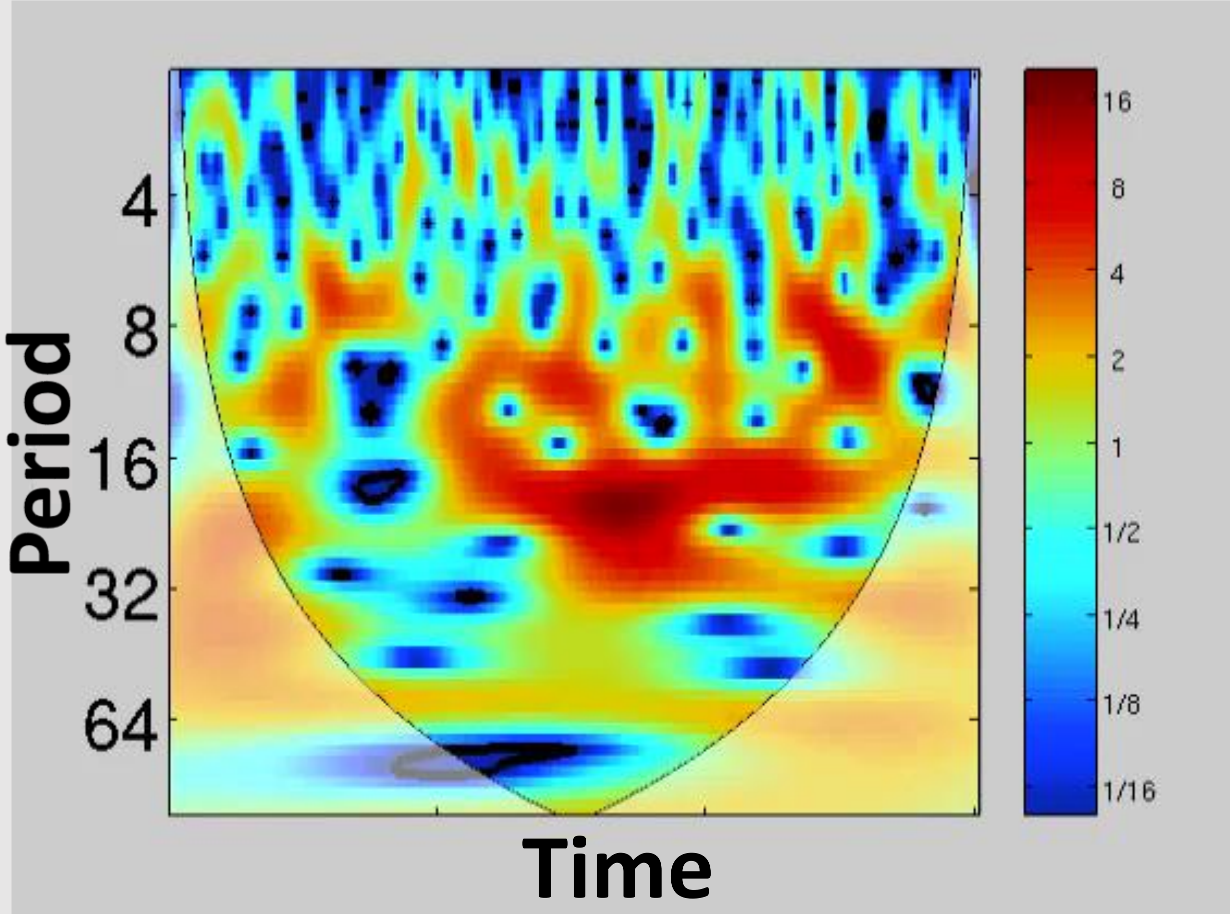
Topological Significance Testing



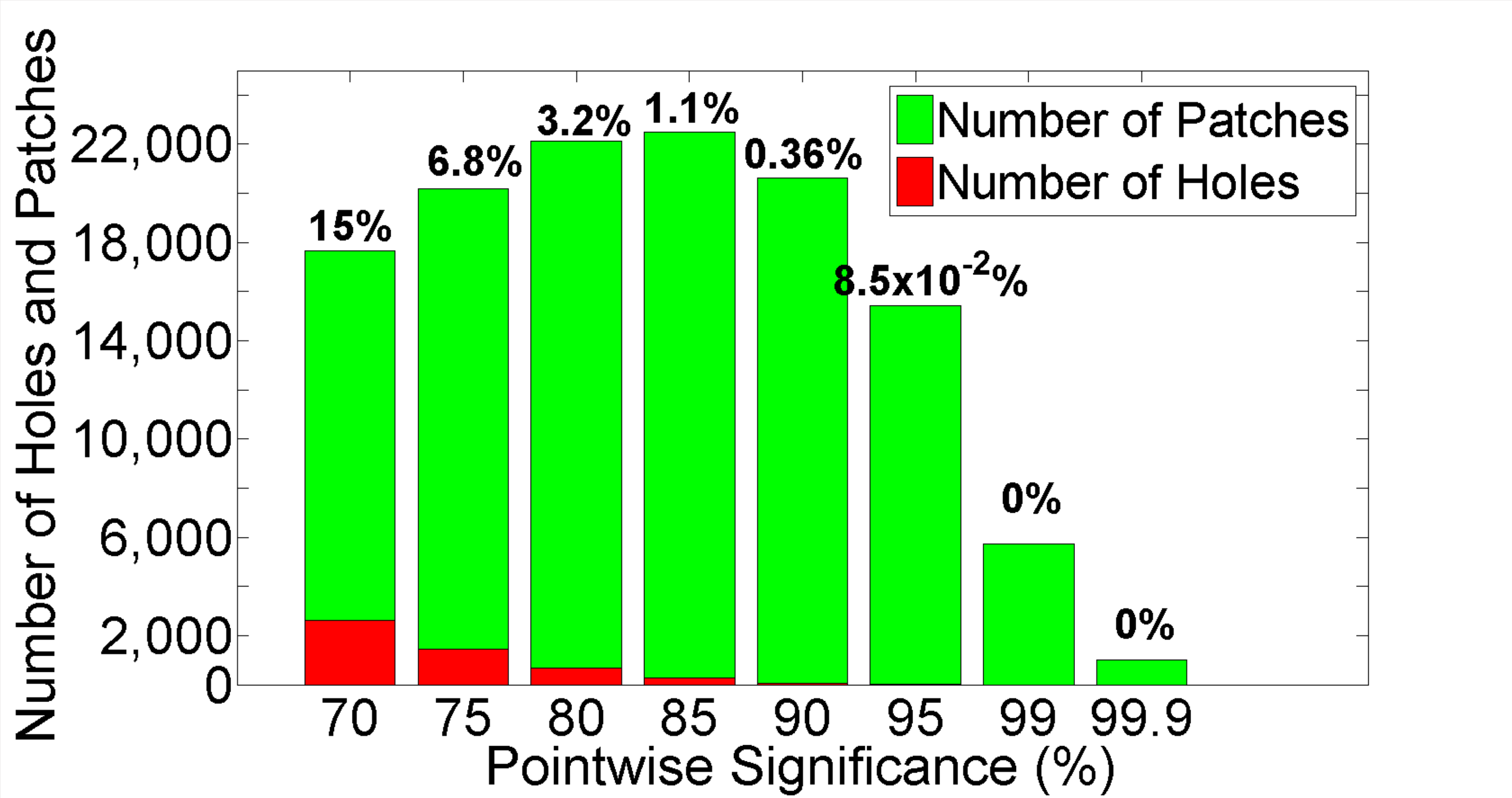
Topological Significance Testing



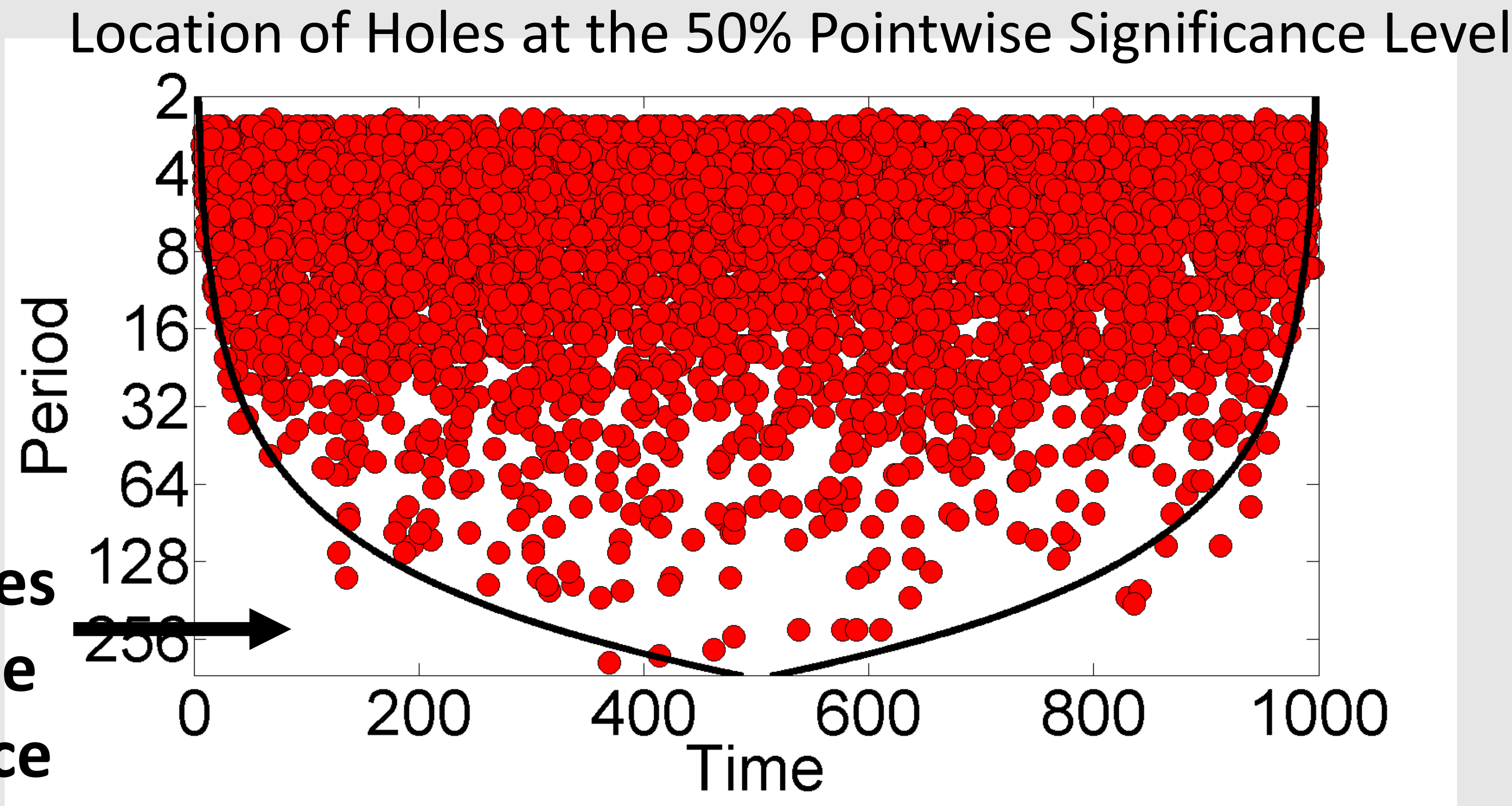
Topological Significance Testing



Topological Significance Testing



Topological Significance Testing



Fewer holes in the Cone of Influence

Topological Significance Testing