

New significance test methods for Fourier analysis of geophysical time series

Z. Zhang¹ and J. Moore^{2,3,4}

¹College of Global Change and Earth System Science, Beijing Normal University, China

²State Key Laboratory of Earth Surface Processes and Resource Ecology/College of Global Change and Earth System Science, Beijing Normal University, China

³Arctic Centre, University of Lapland, Rovaniemi, Finland

⁴Department of Earth Sciences, Uppsala University, Sweden

Received: 20 April 2011 – Revised: 2 September 2011 – Accepted: 6 September 2011 – Published: 27 September 2011

Abstract. When one applies the discrete Fourier transform to analyze finite-length time series, discontinuities at the data boundaries will distort its Fourier power spectrum. In this paper, based on a rigid statistics framework, we present a new significance test method which can extract the intrinsic feature of a geophysical time series very well. We show the difference in significance level compared with traditional Fourier tests by analyzing the Arctic Oscillation (AO) and the Nino3.4 time series. In the AO, we find significant peaks at about 2.8, 4.3, and 5.7 yr periods and in Nino3.4 at about 12 yr period in tests against red noise. These peaks are not significant in traditional tests.

1 Introduction

Continuous and Discrete Fourier transform has been used for numerous studies in geophysics for a long time (Jenkins and Watts, 1968; Chatfield, 1989; Zar, 1999). However, when one applies the discrete Fourier transform to analyze finite-length time series, Fourier power spectrum obtained is distorted in the high frequency domain and the low frequency domain. The reason why Fourier power spectrum is distorted in the high frequency domain is that the time series is discontinuous at the data boundaries. The reason why Fourier power spectrum is distorted in the low frequency domain is that the time series is of finite length. So, when one uses the discrete Fourier transform to compute the Fourier power spectrum, only the middle part of Fourier power spectrum obtained is real. In order to reduce boundary effect,

two algorithms have been tried previously (e.g. Oppenheim and Schaffer, 1989). The first algorithm is to pad the time series on each endpoint with zeroes, but it artificially creates discontinuities at the endpoints. The second algorithm uses some window (e.g. “cosine damping”) to preprocess the time series. Although it overcomes the boundary effect, at the same time, we lose almost all the information at the endpoints.

Statistical significance tests are used to extract intrinsic features from Fourier power spectrum of geophysical time series against a null hypothesis of some noise model. Traditional significance tests on Fourier power spectra consist of the following three steps.

Step 1. Use the discrete Fourier transform to compute the Fourier power spectrum of geophysical time series.

Step 2. Choose a suitable null hypothesis. For many geophysical phenomena, an appropriate noise model is red noise, that is, a first order autoregressive process. However, choice of noise model is crucial to reliable significance testing (Mann and Lees, 1996).

Step 3. Extract intrinsic features. If a peak in the Fourier power spectrum of geophysical time series is above this background noise spectrum at some arbitrary level, say 95 % confidence level, then it is a true feature of geophysical time series with 95 % confidence level.

Since in reality, we only deal with finite-length geophysical time series, discontinuities at the data boundaries will distort the Fourier power spectrum. Therefore, significance tests are unreliable. The purpose of this paper is to present a new and more reliable significance test method for the Fourier power spectra.



Correspondence to: Z. Zhang
(zhangzh@bnu.edu.cn)

This paper is organized as follows: in Sect. 2, we explain why Fourier power spectrum obtained by discrete Fourier transform is distorted in the high frequency domain, so traditional significance tests are unreliable. In Sect. 3, we present a new method to decompose the time series into two parts and introduce the concept of modified Fourier power spectrum as a basis for new significance tests. In Sect. 4, since for many geophysical time series, an appropriate background noise or null hypothesis is red noise, we will give the modified Fourier power spectrum of red noise in a rigid statistics framework. Finally, in Sect. 5, we will apply our significance tests to analyze Arctic Oscillation and Nino3.4 indices which are typical examples of climatic time series with, respectively, little and considerable autocorrelation, to demonstrate the strikingly different significance test results.

2 Distortion of Fourier power spectrum in the high frequency domain

The continuous Fourier transform of the continuous time series $f(t)$, $t \in (-\infty, \infty)$ is defined by

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-2\pi i\omega t} dt, \quad \omega \in (-\infty, \infty).$$

and $|\hat{f}(\omega)|^2$ is called its Fourier power spectrum.

In practice, time series have finite duration, so we can only deal with the time series from 0 to T . In order to compute the Fourier transform, the time series is usually padded with zeroes. In this way, the continuous Fourier transform of f is approximated by

$$\hat{f}(\omega) \approx \int_{-\infty}^{\infty} F(t)e^{-2\pi i\omega t} dt, \tag{1}$$

where F is a discontinuous function at $t = 0$ and $t = T$

$$F(t) = \begin{cases} f(t), & 0 \leq t \leq T, \\ 0, & t < 0, t > T. \end{cases}$$

Because of the discontinuity at the boundary points $t = 0$ and $t = T$, the Fourier transform obtained is distorted in the high frequency domain.

In fact, it is also impossible to obtain the value of f on each point in $[0, T]$. Now assume that we are given the discretized version of the continuous time series f sampled as $\{f(n\delta t)\}_{n=0}^{N-1}$ with a step of δt , where $T = N\delta t$. So, \hat{f} at the point $\frac{k}{N\delta t}$ is approximated by the following sum:

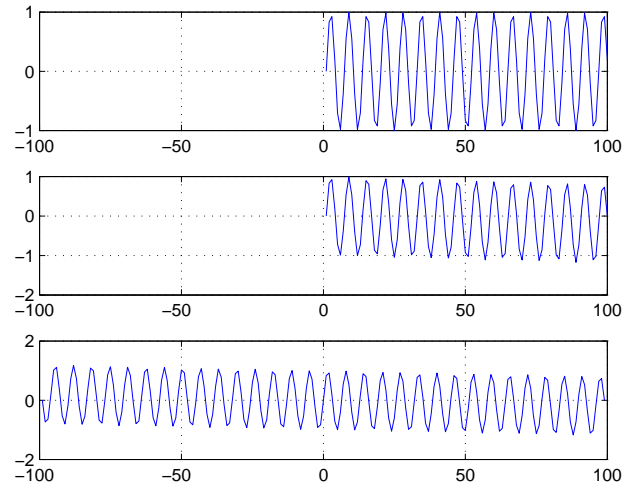


Fig. 1. (Top) A simple time series: $f(t) = \sin(\frac{\pi t}{3.2})$. (Middle) The adjusted time series obtained by removing the line which connect two endpoints of the original time series. (Bottom) We do the odd extension for the adjusted time series

$$\begin{aligned} \hat{f}(\frac{k}{N\delta t}) &\approx \int_0^T f(t)e^{-i\frac{2k\pi}{N\delta t}t} dt \\ &= \sum_{n=0}^{N-1} \int_{n\delta t}^{(n+1)\delta t} f(t)e^{-i\frac{2k\pi}{N\delta t}t} dt \\ &\approx \delta t \sum_{n=0}^{N-1} f(n\delta t)e^{-i\frac{2\pi kn}{N}}, \quad k = 0, 1, \dots, N-1. \end{aligned}$$

From here, we see that $\{\hat{f}(\frac{k}{N\delta t})\}_{k=0}^{N-1}$ can be computed approximately by the discrete Fourier transform of the discrete time series $\{f(n\delta t)\}_{n=0}^{N-1}$. However, by Eq. (1), we know that the Fourier transform obtained is distorted in the high frequency domain.

3 Our decomposition method

In order to develop a new significance test to extract the intrinsic features from time series, in this section, we will decompose the time series and introduce the concept of modified Fourier power spectrum.

3.1 Continuous time series

Let f be a continuous time series $f(t)$ on $[0, T]$ (see Fig. 1, Top), we divide f into two parts

$$f(t) = u(t) + v(t) \quad (0 \leq t \leq T)$$

where $u(t)$ is a linear function such that $u(0) = f(0)$, $u(T) = f(T)$. Denote $u(t) = at + b$. Then $a = \frac{f(T) - f(0)}{T}$, $b = f(0)$, and $v(0) = v(T) = 0$ (see Fig. 1, Middle). So we have

$$v(t) = f(t) - \frac{f(T) - f(0)}{T}t - f(0) \quad (0 \leq t \leq T).$$

Now we make the odd extension of v (see Fig. 1, Bottom):

$$v^\circ(t) = \begin{cases} v(t), & 0 \leq t \leq T, \\ -v(-t), & -T \leq t < 0. \end{cases}$$

Again we make the $2T$ -periodic extension, we get v^* :

$$v^*(t + 2nT) = v^\circ(t) \quad (t \in [-T, T], n \in \mathbb{Z}).$$

By the following proposition, we can see that the above approach can remove the discontinuities at the data boundary very well. The proof of this proposition can be found in Appendix A at the end of this paper.

Proposition 1. If the derivative of f is continuous on $[0, T]$, then the derivative of v^* is continuous on $(-\infty, \infty)$.

Finally, we expand $v^*(t)$ into the Fourier series

$$v^*(t) = \sum_{k=-\infty}^{\infty} c_n e^{i \frac{k\pi t}{T}}$$

where the Fourier coefficient c_n is computed by

$$c_n = \frac{1}{2T} \int_{-T}^T v^*(t) e^{-i \frac{n\pi t}{T}} dt \quad (2)$$

Comparing $f(t)$ with $v^*(t)$, we find that in $v^*(t)$, discontinuities on data boundary are removed. We call $|c_n|^2$ the modified Fourier power spectrum of continuous time series $f(t)$.

Now we begin to simplify the formula (2):

$$c_n = \frac{1}{2T} \int_{-T}^T v^*(t) \cos \frac{n\pi t}{T} dt - \frac{i}{2T} \int_{-T}^T v^*(t) \sin \frac{n\pi t}{T} dt$$

Since v^* is odd and $\cos \frac{n\pi t}{T}$ is even, the first term on the right-hand side of the above formula is zero. Again since $\sin \frac{n\pi t}{T}$ is odd, we have

$$c_n = -\frac{i}{T} \int_0^T v^*(t) \sin \frac{n\pi t}{T} dt. \quad (3)$$

3.2 Discrete time series

In practice, we are only given the samples of some continuous time series f on $[0, T]$, i.e.

$$x_k = f(k\delta t), \quad k = 0, 1, 2, \dots, N \quad (T = N\delta t)$$

As for the continuous case, we decompose the time series into

$$x_k = u_k + v_k \quad (k = 0, \dots, N)$$

such that $u_k = ak + b$ ($k = 0, \dots, N$) and $u_0 = x_0, u_N = x_N$. So we have

$$a = \frac{x_N - x_0}{N}, \quad b = x_0,$$

and

$$u_k = \frac{k}{N}(x_N - x_0) + x_0 \quad (k = 0, \dots, N).$$

This implies that

$$x_k = \frac{k}{N}(x_N - x_0) + x_0 + v_k \quad \text{and} \quad v_0 = v_N = 0.$$

We make the odd extension of the sequence $\{v_k\}$:

$$v_k^\circ = \begin{cases} v_k, & n = 0, \dots, N, \\ -v_{-k}, & n = -1, \dots, -N + 1, \end{cases}$$

and then make the $2N$ -periodic extension

$$v_{k+2nN}^* = v_k^\circ \quad k = 0, \pm 1, \dots, \pm(N-1), N, \quad n \in \mathbb{Z}.$$

So $\{v_k^*\}$ is a $2N$ -periodic odd sequence. Moreover, $\{v_k^*\}$ are the samples of $v^*(t)$ which has been defined in Sect. 3.1. In other words,

$$v_k^* = v^*(k\delta t), \quad k = 0, 1, 2, \dots, N$$

Numerically integrating Eq. (3), noticing that $T = N\delta t$, $v^*(0) = 0$, and $v^*(T) = 0$, we obtain the relation between $\{v_k^*\}$ and the Fourier coefficients $\{c_n\}$ as follows:

$$c_n = -\frac{i}{T} \int_0^T v^*(t) \sin \frac{n\pi t}{T} dt \approx -\frac{i}{N} \sum_{k=1}^{N-1} v_k^* \sin \frac{kn\pi}{N}. \quad (4)$$

So we can define the modified Fourier power spectrum of discrete time series $\{x_n\}_0^N$ as

$$\left| \frac{1}{N} \sum_{k=1}^{N-1} v_k^* \sin \frac{kn\pi}{N} \right|^2. \quad (5)$$

4 The modified Fourier power spectrum of red noise

For many geophysical phenomena, an appropriate background is red noise (Mann and Lees, 1996; Ghil et al., 2002). A simple model for red noise is the univariate lag-1 autoregressive [AR(1)] process $\{x_n\}_0^N$:

$$x_0 = 0, \quad x_n = \lambda x_{n-1} + Z_n \quad (n = 1, \dots, N),$$

where λ is a constant and called the AR(1) coefficient, $|\lambda| < 1$, and $\{Z_n\}$ are independent Gaussian random variables with mean 0 and variance σ^2 .

If $\{x_n\}_1^N$ are known, one can use the following formula (Brockwell and Davis, 1991) to estimate the AR(1) coefficient λ and noise variance σ^2 .

$$\lambda = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} (x_i - \bar{x})(x_{i+1} - \bar{x})}{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2},$$

$$\sigma^2 = \frac{1 - \lambda^2}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

where $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$.

Denote the discrete Fourier transform of an AR(1) process $\{x_n\}_0^{N-1}$ by $\{f_n\}_0^{N-1}$. Torrence and Compo (1998) showed that the Fourier power spectrum $|f_n|^2$ is distributed as

$$\frac{\tilde{\sigma}^2(1 - \lambda^2)}{2N(1 + \lambda^2 - 2\lambda \cos \frac{2\pi n}{N})} \chi_2^2 \tag{6}$$

where $\tilde{\sigma}^2$ is the variance of time series and χ_2^2 is the chi-square distribution with two degrees of freedom. In traditional significance tests, one uses this distribution to extract the intrinsic features of geophysical time series.

In order to satisfy the needs of our new significance tests, we will give the distribution for the modified Fourier power spectrum of AR(1) process in a rigid statistics framework.

From the definition of an AR(1) process, we have $x_0 = 0$ and $x_1 = \lambda x_0 + Z_1 = Z_1$. Furthermore,

$$x_2 = \lambda x_1 + Z_2 = \lambda Z_1 + Z_2$$

and

$$x_3 = \lambda^2 Z_1 + \lambda Z_2 + Z_3.$$

In general, we have

$$x_k = \sum_{l=1}^k \lambda^{k-l} Z_l.$$

Based on our decomposition method for discrete time series in Sect. 3.2, we have

$$x_k = \alpha_k + \beta_k \quad (k = 0, 1, \dots, N), \quad \text{where} \quad \alpha_k = \frac{k}{N} x_N.$$

Therefore,

$$\beta_k = x_k - \alpha_k = \sum_{l=1}^k \lambda^{k-l} Z_l - \frac{k}{N} \sum_{l=1}^N \lambda^{N-l} Z_l.$$

By Eq. (5), the modified Fourier power spectrum of AR(1) process is $|Y_n|^2$, where

$$Y_n = \frac{1}{N} \sum_{k=1}^{N-1} \beta_k \sin \frac{nk\pi}{N}.$$

So

$$\begin{aligned} NY_n &= \sum_{k=1}^{N-1} \sum_{l=1}^k \lambda^{k-l} Z_l \sin \frac{nk\pi}{N} \\ &\quad - \sum_{k=1}^{N-1} \sum_{l=1}^N \frac{k}{N} \lambda^{N-l} Z_l \sin \frac{nk\pi}{N} \\ &=: I_n - J_n. \end{aligned} \tag{7}$$

It is easy to check the following formulas:

$$\sum_{k=1}^{N-1} \left(\sum_{l=1}^k a_{k,l} \right) = \sum_{l=1}^{N-1} \left(\sum_{k=l}^{N-1} a_{k,l} \right)$$

and

$$\sum_{k=1}^{N-1} \left(\sum_{l=1}^N b_{k,l} \right) = \sum_{l=1}^N \left(\sum_{k=1}^{N-1} b_{k,l} \right).$$

Furthermore, we deduce that

$$\begin{aligned} I_n &= \sum_{l=1}^{N-1} \left(\sum_{k=l}^{N-1} \lambda^{k-l} Z_l \sin \frac{nk\pi}{N} \right) \\ &= \sum_{l=1}^{N-1} \lambda^{-l} Z_l \left(\sum_{k=l}^{N-1} \lambda^k \sin \frac{nk\pi}{N} \right), \end{aligned}$$

$$J_n = \left(\sum_{l=1}^N \frac{1}{N} \lambda^{N-l} Z_l \right) \cdot \left(\sum_{k=1}^{N-1} k \sin \frac{nk\pi}{N} \right).$$

By $\sin \frac{nN\pi}{N} = 0$, we can rewrite I_n as

$$I_n = \sum_{l=1}^N \lambda^{-l} Z_l \left(\sum_{k=l}^N \lambda^k \sin \frac{nk\pi}{N} \right)$$

From this and Eq. (6), we get

$$NY_n = \sum_{l=1}^N \Lambda_{n,l} Z_l, \tag{8}$$

where

$$\Lambda_{n,l} = \lambda^{-l} c_{n,l} - \frac{\lambda^{N-l}}{N} d_n,$$

and

$$c_{n,l} = \sum_{k=l}^N \lambda^k \sin \frac{nk\pi}{N}, \quad d_n = \sum_{k=1}^{N-1} k \sin \frac{nk\pi}{N}.$$

From this, we deduce that

$$N^2 Y_n^2 = \sum_{l=1}^N \Lambda_{n,l}^2 Z_l^2 + \sum_{\substack{l,l'=1 \\ l \neq l'}}^N \Lambda_{n,l} \Lambda_{n,l'} Z_l Z_{l'},$$

and the mathematical expectation

$$N^2 E[Y_n^2] = \sum_{l=1}^N \Lambda_{n,l}^2 E[Z_l^2] + \sum_{\substack{l,l'=1 \\ l \neq l'}}^N \Lambda_{n,l} \Lambda_{n,l'} E[Z_l Z_{l'}]. \quad (9)$$

Noticing that $\{Z_l\}_1^N$ are independent Gaussian random variables with mean 0 and variance σ^2 , we get

$$E[Z_l^2] = \sigma^2 \quad (l = 1, \dots, N),$$

$$E[Z_l Z_k] = (E[Z_l]) \cdot (E[Z_k]) = 0 \quad (l, k = 1, \dots, N; l \neq k).$$

By Eq. (9), we have

$$N^2 E[Y_n^2] = \sigma^2 \sum_{l=1}^N \Lambda_{n,l}^2. \quad (10)$$

For $c_{n,l}$, we have

$$\begin{aligned} c_{n,l} &= \sum_{k=l}^N \lambda^k \sin \frac{nk\pi}{N} = I_m \left(\sum_{k=l}^N \lambda^k e^{i \frac{nk\pi}{N}} \right) \\ &= I_m \left(\sum_{k=l}^N \left(\lambda e^{i \frac{n\pi}{N}} \right)^k \right) \\ &= I_m \left(-\frac{\lambda^{N+1} (-1)^n e^{i \frac{n\pi}{N}} - \lambda^l e^{i \frac{n\pi}{N}}}{1 - \lambda e^{i \frac{n\pi}{N}}} \right) \\ &= \frac{\lambda^l \sin \frac{n\pi}{N} + \lambda^{N+1} (-1)^{n+1} \sin \frac{n\pi}{N} - \lambda^{l+1} \sin \frac{n(l-1)\pi}{N}}{1 + \lambda^2 - 2\lambda \cos \frac{n\pi}{N}}. \end{aligned} \quad (11)$$

Now we compute d_n . Since $\sum_{k=1}^{N-1} e^{ikt} = \frac{1-e^{iNt}}{1-e^{it}} - 1$, we have

$$\begin{aligned} \sum_{k=1}^{N-1} k e^{ikt} &= \frac{1}{i} \frac{d}{dt} \left(\frac{1-e^{iNt}}{1-e^{it}} - 1 \right) \\ &= \frac{-N e^{iNt} + (N-1) e^{i(N+1)t} + e^{it}}{(1-e^{it})^2} \\ &= -\frac{1+(N-1)e^{iNt} - N e^{i(N-1)t}}{4 \sin^2 \frac{t}{2}}, \end{aligned}$$

and so

$$I_m \left\{ \sum_{k=1}^{N-1} k e^{ikt} \right\} = -\frac{(N-1) \sin Nt - N \sin((N-1)t)}{4 \sin^2 \frac{t}{2}}.$$

Finally, we have

$$\begin{aligned} d_n &= \sum_{k=1}^{N-1} k \sin \frac{nk\pi}{N} \\ &= I_m \left\{ \sum_{k=1}^{N-1} k e^{i \frac{nk\pi}{N}} \right\} = \frac{N \sin \frac{(N-1)n\pi}{2N}}{4 \sin^2 \frac{n\pi}{2N}}. \end{aligned} \quad (12)$$

By Eqs. (8)–(12), we know that Y_n is a Gaussian random variable with mean 0 and variance

$$\begin{aligned} \text{var}(Y_n) &= E[(Y_n - E[Y_n])^2] = E[Y_n^2] \\ &= \frac{\sigma^2}{N^2} \sum_{l=1}^N \left(\lambda^{-l} c_{n,l} - \frac{\lambda^{N-l}}{N} d_n \right)^2, \end{aligned} \quad (13)$$

where

$$\begin{aligned} c_{n,l} &= \frac{\lambda^l \sin \frac{n\pi}{N} + \lambda^{N+1} (-1)^{n+1} \sin \frac{n\pi}{N} - \lambda^{l+1} \sin \frac{n(l-1)\pi}{N}}{1 + \lambda^2 - 2\lambda \cos \frac{n\pi}{N}} \\ d_n &= \frac{N \sin \frac{(N-1)n\pi}{2N}}{4 \sin^2 \frac{n\pi}{2N}}. \end{aligned} \quad (14)$$

In the other words, Y_n is distributed as

$$\frac{\sigma}{N} \sqrt{\sum_{l=1}^N \left(\lambda^{-l} c_{n,l} - \frac{\lambda^{N-l}}{N} d_n \right)^2} \mathbf{Z}$$

where \mathbf{Z} is a Gaussian random variable with mean 0 and variance 1.

Since the modified Fourier power spectrum of AR(1) process is $|Y_n|^2$, we have

Theorem 1. The modified Fourier power spectrum of an AR(1) process is distributed as

$$\frac{\sigma^2}{N^2} \sum_{l=1}^N \left(\lambda^{-l} c_{n,l} - \frac{\lambda^{N-l}}{N} d_n \right)^2 \mathbf{Z}^2. \quad (15)$$

where \mathbf{Z} is a Gaussian random variable with mean 0 and variance 1, and $c_{n,l}$, d_n are stated in Eq. (14).

In Appendix B, we will simplify the formula (15) further.

5 Applications

In this section, based on Theorem 1 (or Theorem 2 in Appendix B), we will use our significance test to extract intrinsic features of time series from the Fourier power spectrum in the high-frequency domain against a red noise Null Hypothesis. First of all, we consider an artificial example. We embed

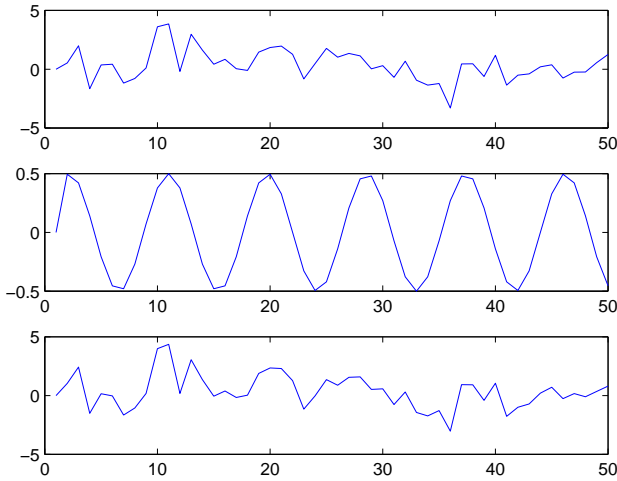


Fig. 2. (Top) AR(1) process with AR(1) coefficient 0.3 and noise variance 1. (Middle) a pure sinusoid of period 8.8. (Bottom) The sum of top signal and middle signal

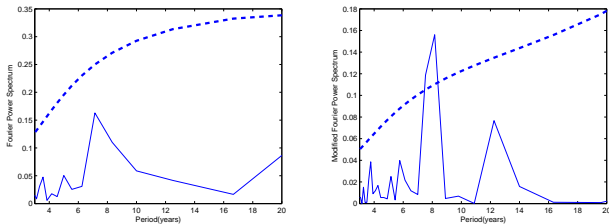


Fig. 3. (Left) Fourier power spectrum is computed by discrete Fourier transform. (Right) Modified Fourier power spectrum is computed by our method. The dashed line is 99 % confidence red noise spectrum .

a pure sinusoid of period 8.8 within an AR(1) process with AR(1) coefficient 0.3 and noise variance 1 (see Fig. 2). We will do significance tests by using traditional Fourier method (Fig. 3, Left) and our method (Fig. 3, Right). From Fig. 3, we see that our method discovers this pure sinusoid while the traditional method does not.

Below we examine the winter Arctic Oscillation(AO) indices (December–February 1851–1997) of Thompson and Wallace (1998). The AO is a key aspect of climate variability in the Northern Hemisphere. The AO index is defined as the leading empirical orthogonal function (EOF) of Northern Hemisphere sea level pressure anomalies pole ward of 20° N (Thompson and Wallace, 1998), and characterized by an exchange of atmospheric mass between the Arctic and middle latitudes. We first make a significance test for the Fourier power spectrum of the AO index based on the traditional method with a red noise Null Hypothesis. In Fig. 4 (Top), we can see that only two peaks are well above the background

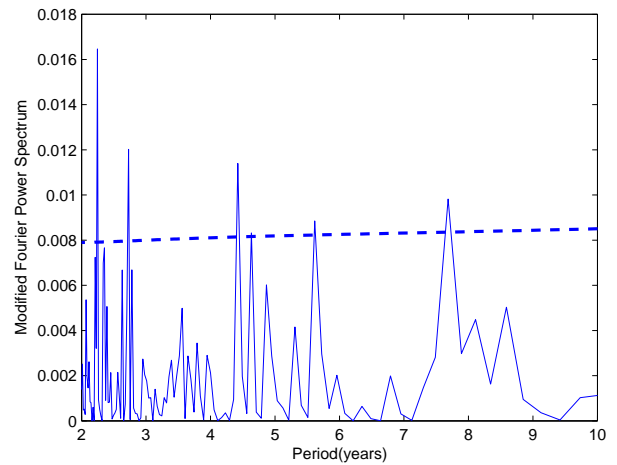
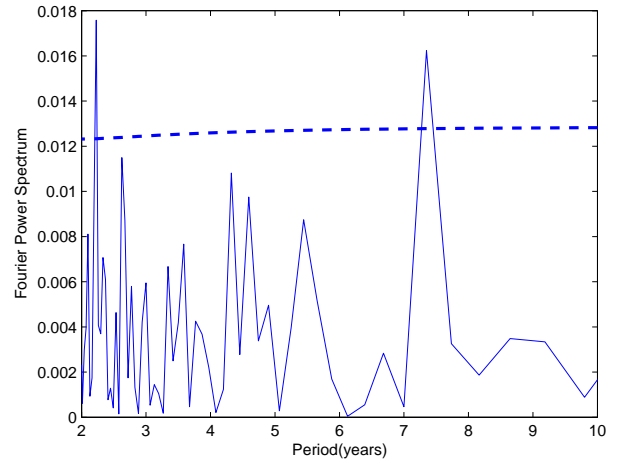


Fig. 4. (Top) Fourier power spectrum of AO indices is computed by discrete Fourier transform. (Bottom) Modified Fourier power spectrum of AO indices is computed by our method. The dashed line is 95 % confidence red noise spectrum .

spectrum. When one applies the discrete Fourier transform to analyze a finite-length AR(1) process, the data boundary will distort its Fourier power spectrum, especially in the high frequency domain. Now, based on Theorem 1 (or Theorem 2 in Appendix B), we use our significance test. Since discontinuities at the data boundaries are removed, we find more significant peaks at about 2.8, 4.3 and 5.7 yr periods well above the background spectrum (see Fig. 4, Bottom). These peaks are not significant in traditional tests.

Nino3.4 is used as a measure of the amplitude of the El Nino Southern Oscillation (ENSO). The Nino3.4 index is defined as the average monthly sea surface temperature anomaly in the region bounded by 5° N to 5° S from 170° W to 120° W. Unlike the winter AO annual index, the monthly Nino3.4 time series has a high autocorrelation and hence the significant level shows a strong trend with frequency.

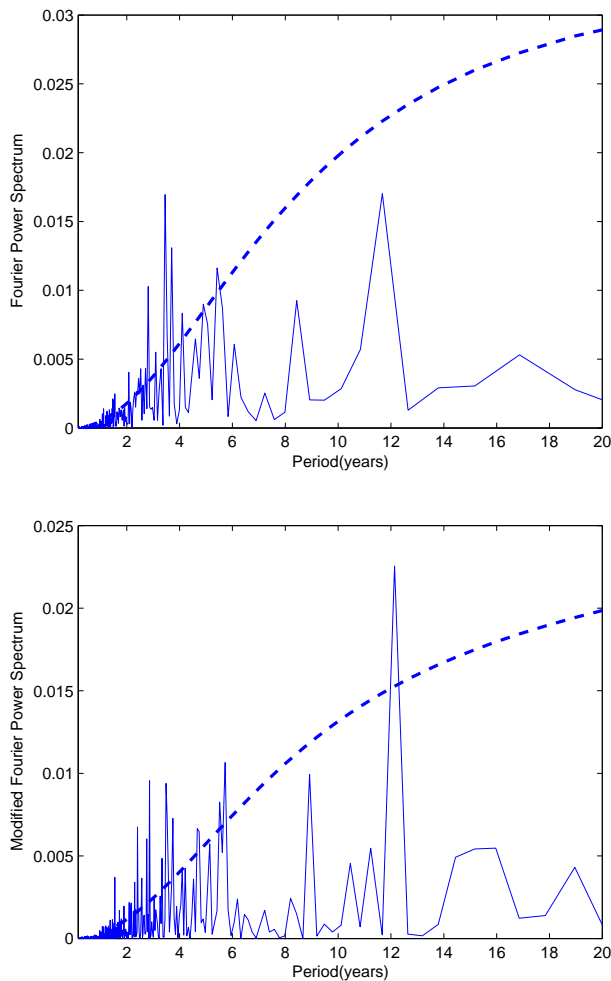


Fig. 5. (Top) Fourier power spectrum of Nino3.4 indices is computed by discrete Fourier transform. (Bottom) Modified Fourier power spectrum of Nino3.4 indices is computed by our method. The dashed line is 95 % confidence red noise spectrum .

Figure 5 shows the Fourier and modified Fourier power spectra, and we find a significant peak at about 12 yr period in a test against red noise. This peak is not significant in traditional tests. The main reason for the increased significance is the removal of boundary effect in our new significance test. Notice also that the modified Fourier power spectra has higher resolution than traditional spectrum due to the odd extension made to the original time series. This high resolution may result in sharper spectral peaks at any period which is especially noticeable around 12 yr in Fig. 5.

The periodicities in the AO time series have been examined using wavelet coherence methods (Jevrejeva et al., 2003), and Monte Carlo Singular Spectrum Analysis (MCSSA, Jevrejeva et al., 2004). These papers also discuss the relationship between the AO and Nino3 time series (Nino3 is closely related to Nino3.4 being defined as the monthly Sea

Surface Temperatures (SST) averaged over the eastern half of the tropical Pacific (5°S – 5°N , 90° – 150°W). The MCSSA and wavelet coherence analyses revealed that periodic signals in the AO and Nino3 time series vary in over time, e.g. AO exhibits significant power in the 3.5–5.7 yr band between 1935–1950, and power in the 12–14 yr band in the second half of the record. In contrast the Fourier spectrum analysis shows a simple result representing the whole time series studied.

In general the strongest signals detected with MCSSA and wavelets were at 2.2, 3.5, 5.7 and a broad band centered on 14 yr periods. Jevrejeva et al. (2003) discuss the phase differences between these signals in various climatic time series as a function of their geographic location. They found that the signals in the 2.2–5.7 yr bands have short phase lags of about 3 months between their location of origin (the tropical Pacific, and the polar regions, implying a rapid transportation mechanism likely via the stratosphere as proposed by Baldwin and Dunkerton (2001)). In contrast the 13.9 yr signal has a phase lag of about 0.7 yr between the Nino3.4 and AO series, and a further year between the atmospheric record recorded in the Southern Oscillation Index and Nino3.4. These much slower transport patterns can also be tracked in global sea surface temperature fields (Jevrejeva et al., 2004), and points to a oceanic transport mechanism from tropics to both polar regions.

The results of the Fourier significance analysis presented in Figs. 4 and 5 show similar features as described in the wavelet and MCSSA analyses. There are a broad band of signals from 2.2–5.7 yr that are associated with ENSO variability in the Pacific region – with the 5.7 yr signal detected as significant in our new test exactly coinciding with features in the MCSSA (Jevrejeva et al., 2004) and wavelet analysis (Jevrejeva et al., 2003). These signals propagate globally as a forced response to the slow warming of tropical oceans. We find a 12 yr peak in the Nino3.4 series but the same period is significant in AO only in the second half of that time series (and only in our new significance test not the traditional one). It is likely that this 12 yr signal corresponds to the 13.9 yr period peak discussed above. The difference in periodicity of the peak is perhaps that the MCSSA approach is a time domain analysis, and does not rely on sinusoidal basis functions. The MCSSA analysis also reveals that the signal is growing over time, especially in the 20th century while it is insignificant in the 19th century. There are also some differences in the periodicity of the high frequency signals in Nino3.4 between methods and probably reflect choice of parameters such as the embedding dimension in the MCSSA analysis (Ghil et al., 2002) – which may lead to a 2.8 yr signal being resolved as a 2.2 yr and 3.5 yr signals. We also note that in the traditional significance test finds a peak at 5.4 yr in Nino3.4 (Fig. 5), while the new test reveals 2 peaks at 5.5 and 5.7 yr.

6 Conclusions

We present a novel way of modifying the traditional Fourier transform to remove the impact of discontinuity at the data boundaries. We show that when the modified Fourier decomposition is made, we can calculate significance levels against a red noise Null Hypothesis in a rigid statistics framework. Tests on real time series show important differences in significance levels compared with traditional methods. This suggests that the Fourier spectra produced commonly in geophysical or climatic time series are suspect. In many time series that may reflect on-going change, such as climatic series, the method we suggest has clear advantages over traditional Fourier analysis since the information at the most recent times, i.e. the data boundary is preserved. It is worth noticing that in many time series of practical importance such as the AO or Nino3.4, the differences in significance levels include multi-year and decadal frequencies. Such multi-year and decadal variability is the target of considerable current research, and therefore establishing the presence of significant oscillatory power in these bands is highly relevant.

Appendix A

Proof of Proposition 1

Since the derivative of f is continuous on the interval $[0, T]$, from the definition of v^* , we know that in order to prove the continuity of the derivative of v^* on $(-\infty, \infty)$, we only need to prove that the derivative of v^* is continuous at the points 0 and T . Because these two arguments are similar, we only prove that the derivative of v^* is continuous at 0.

First we prove that the derivative of $v^*(t)$ exists at 0.

By $v^\circ(t) = v(t), t \in [0, T]$. We consider the right derivative of v° at the endpoint 0,

$$\lim_{\substack{t \rightarrow 0 \\ t > 0}} \frac{v^\circ(t) - v^\circ(0)}{t - 0} = \lim_{\substack{t \rightarrow 0 \\ t > 0}} \frac{v(t) - v(0)}{t - 0} = v'(0). \tag{A1}$$

By $v^\circ(t) = -v(-t), t \in [-T, 0]$ and $v^\circ(0) = v(0) = 0$, we obtain that the left derivative of v° at the endpoint 0 is

$$\lim_{\substack{t \rightarrow 0 \\ t < 0}} \frac{v^\circ(t) - v^\circ(0)}{t - 0} = \lim_{\substack{t \rightarrow 0 \\ t < 0}} \frac{v(0) - v(-t)}{0 - (-t)}.$$

Let $t' = -t$. Then, for $t \rightarrow 0, t < 0$, we have $t' \rightarrow 0, t' > 0$. So we obtain that

$$\lim_{\substack{t \rightarrow 0 \\ t < 0}} \frac{v^\circ(t) - v^\circ(0)}{t - 0} = \lim_{\substack{t' \rightarrow 0 \\ t' > 0}} \frac{v(0) - v(t')}{0 - t'} = v'(0). \tag{A2}$$

Combining Eq. (A1) with Eq. (A2), we know that the derivative of v° exists at the endpoint 0 and

$$(v^\circ)'(0) = v'(0). \tag{A3}$$

By $v^*(t) = v^\circ(t), t \in [-T, T]$, we know that the derivative of v^* exists at 0.

Next we prove that the derivative of v^* is continuous at 0.

From $v^\circ(t) = -v(-t), t \in [-T, 0]$, we have

$$(v^\circ)'(t) = v'(-t), \quad t \in (-T, 0).$$

This implies by Eq. (A3) that

$$\lim_{\substack{t \rightarrow 0 \\ t < 0}} (v^\circ)'(t) = \lim_{\substack{t \rightarrow 0 \\ t < 0}} v'(-t) = \lim_{\substack{t \rightarrow 0 \\ t < 0}} v'(t) = v'(0) = (v^\circ)'(0).$$

On the other hand, by $v^\circ(t) = v(t), t \in [0, T]$, we have

$$\lim_{\substack{t \rightarrow 0 \\ t > 0}} (v^\circ)'(t) = \lim_{\substack{t \rightarrow 0 \\ t > 0}} v'(t) = v'(0) = (v^\circ)'(0).$$

From this, we get $\lim_{t \rightarrow 0} (v^\circ)'(t) = (v^\circ)'(0)$, i.e., the derivative of $v^\circ(t)$ is continuous at $t = 0$. Therefore, we know that the derivative of $v^*(t)$ is continuous at 0. Proposition 1 is proved.

Appendix B

Further simplification of Formula (15)

We will simplify the sum in Eq. (15). Denote

$$J = \frac{\sigma^2}{N^2} \sum_{l=1}^N \left(\lambda^{-l} c_{n,l} - \frac{\lambda^{N-l}}{N} d_n \right)^2. \tag{B1}$$

From this, we get

$$\begin{aligned} \sigma^{-2} N^2 J &= \sum_{l=1}^N \lambda^{-2l} c_{n,l}^2 + \frac{\lambda^{2N}}{N^2} \left(\sum_{l=1}^N \lambda^{-2l} \right) d_n^2 \\ &\quad - \frac{2\lambda^N}{N} \left(\sum_{l=1}^N \lambda^{-2l} c_{n,l} \right) d_n \\ &=: H_1 + H_2 - H_3. \end{aligned} \tag{B2}$$

From Eq. (11), we see that

$$\begin{aligned} &\lambda^{-2l} c_{n,l}^2 (1 + \lambda^2 - 2\lambda \cos \frac{n\pi}{N})^2 \\ &= \sin^2 \frac{n\pi}{N} + \lambda^{-2l} \lambda^{2N+2} \sin^2 \frac{n\pi}{N} + \lambda^2 \sin^2 \frac{n(l-1)\pi}{N} \\ &\quad + 2(-1)^{n+1} \lambda^{N+1} \sin \frac{n\pi}{N} (\lambda^{-l} \sin \frac{n\pi}{N}) \\ &\quad - 2\lambda \sin \frac{n\pi}{N} \sin \frac{n(l-1)\pi}{N} \\ &\quad + 2(-1)^n \lambda^{N+2} \sin \frac{n\pi}{N} (\lambda^{-l} \sin \frac{n(l-1)\pi}{N}). \end{aligned}$$

For convenience, denote

$$g_n(\mu, \nu, \zeta) = \sum_{l=1}^N \lambda^{-\mu l} \sin \frac{n(l-\nu)\pi}{N} \sin \frac{n(l-\zeta)\pi}{N},$$

$$h_n(\mu, \nu) = \sum_{l=1}^N \lambda^{-\mu l} \sin \frac{n(l-\nu)\pi}{N}. \tag{B3}$$

Then we have

$$\begin{aligned} H_1 &= (1 + \lambda^2 - 2\lambda \cos \frac{n\pi}{N})^{-2} (g_n(0, 0, 0) \\ &+ \frac{\lambda^{2N+2} - \lambda^2}{\lambda^2 - 1} \sin^2 \frac{n\pi}{N} + \lambda^2 g_n(0, 1, 1) \\ &+ 2(-1)^{n+1} \lambda^{N+1} \sin \frac{n\pi}{N} h_n(1, 0) - 2\lambda g_n(0, 0, 1) \\ &+ 2(-1)^n \lambda^{N+2} \sin \frac{n\pi}{N} h_n(1, 1)). \end{aligned} \tag{B4}$$

By Eqs. (B2) and (12), we get

$$H_2 = \frac{\lambda^{2N}}{N^2} \left(\sum_{l=1}^N \lambda^{-2l} \right) d_n^2 = \frac{(\lambda^{2N} - 1) \sin^2 \frac{(N-1)n\pi}{N}}{16(\lambda^2 - 1) \sin^4 \frac{n\pi}{2N}}. \tag{B5}$$

By (11), we get

$$\begin{aligned} &\left(\sum_{l=1}^N \lambda^{-2l} c_{n,l} \right) (1 + \lambda^2 - 2\lambda \cos \frac{n\pi}{N}) \\ &= \sum_{l=1}^N \lambda^{-l} \sin \frac{nl\pi}{N} + (-1)^{n+1} \lambda^{N+1} \sin \frac{n\pi}{N} \left(\sum_{l=1}^N \lambda^{-2l} \right) \\ &- \sum_{l=1}^N \lambda^{-l+1} \sin \frac{n(l-1)\pi}{N}. \\ &= h_n(1, 0) + (-1)^{n+1} \frac{\lambda^{N+1} - \lambda^{-N+1}}{\lambda^2 - 1} \sin \frac{n\pi}{N} - \lambda h_n(1, 1). \end{aligned}$$

By Eqs. (12) and (B2), we get

$$\begin{aligned} H_3 &= \frac{\lambda^N \sin \frac{(N-1)n\pi}{N}}{2(1 + \lambda^2 - 2\lambda \cos \frac{n\pi}{N}) \sin^2 \frac{n\pi}{2N}} (h_n(1, 0) \\ &+ (-1)^{n+1} \frac{\lambda^{N+1} - \lambda^{-N+1}}{\lambda^2 - 1} \sin \frac{n\pi}{N} \\ &- \lambda h_n(1, 1)t). \end{aligned} \tag{B6}$$

Finally, we compute $g_n(\mu, \nu, \zeta)$ and $h_n(\mu, \nu)$ in Eq. (B3). First we consider $g_n(\mu, \nu, \zeta)$. We see that

$$\begin{aligned} -2g_n(\mu, \nu, \zeta) &= \sum_{l=1}^N \lambda^{-\mu l} \left(\cos \frac{(2l-n\nu-n\zeta)\pi}{N} - \cos \frac{n(\zeta-\nu)\pi}{N} \right) \\ &= \operatorname{Re} \left\{ e^{-i \frac{n(\nu+\zeta)\pi}{N}} \sum_{l=1}^N (\lambda^{-\mu} e^{i \frac{2n\pi}{N}})^l \right\} \\ &- \cos \frac{n(\zeta-\nu)\pi}{N} \sum_{l=1}^N \lambda^{-\mu l} \\ &=: S_1 - S_2, \end{aligned} \tag{B7}$$

For S_2 , it is clear that

$$S_2 = \frac{\lambda^{-\mu} - \lambda^{-\mu(N+1)}}{1 - \lambda^{-\mu}} \cos \frac{n(\zeta - \nu)\pi}{N}.$$

For S_1 , since $e^{2\pi i n} = 1$, we have

$$\sum_{l=1}^N (\lambda^{-\mu} e^{i \frac{2n\pi}{N}})^l = \frac{\lambda^{-\mu} e^{i \frac{2n\pi}{N}} (1 - \lambda^{-\mu N})}{1 - \lambda^{-\mu} e^{i \frac{2n\pi}{N}}}$$

Furthermore, we get

$$\begin{aligned} S_1 &= \lambda^{-\mu} (1 - \lambda^{-\mu N}) \operatorname{Re} \left\{ \frac{e^{i \frac{2n\pi - n(\nu+\zeta)\pi}{N}}}{1 - \lambda^{-\mu} e^{i \frac{2n\pi}{N}}} \right\} \\ &= \lambda^{-\mu} (1 - \lambda^{-\mu N}) \operatorname{Re} \left\{ \frac{e^{i \frac{2n\pi - n(\nu+\zeta)\pi}{N}} (1 - \lambda^{-\mu} e^{-i \frac{2n\pi}{N}})}{1 + \lambda^{-2\mu} - 2\lambda^{-\mu} \cos \frac{2n\pi}{N}} \right\} \\ &= \frac{\lambda^{-\mu} (1 - \lambda^{-\mu N}) (\cos \frac{2n\pi - n(\nu+\zeta)\pi}{N} - \lambda^{-\mu} \cos \frac{n(\nu+\zeta)\pi}{N})}{1 + \lambda^{-2\mu} - 2\lambda^{-\mu} \cos \frac{2n\pi}{N}} \end{aligned} \tag{B8}$$

From this and Eq. (B7), we have

$$\begin{aligned} g_n(\mu, \nu, \zeta) &= - \frac{\lambda^{-\mu} (1 - \lambda^{-\mu N}) (\cos \frac{2n\pi - n(\nu+\zeta)\pi}{N} - \lambda^{-\mu} \cos \frac{n(\nu+\zeta)\pi}{N})}{2 + 2\lambda^{-2\mu} - 4\lambda^{-\mu} \cos \frac{2n\pi}{N}} \\ &+ \frac{\lambda^{-\mu} - \lambda^{-\mu(N+1)}}{2 - 2\lambda^{-\mu}} \cos \frac{n(\zeta - \nu)\pi}{N} \end{aligned} \tag{B9}$$

Next, we compute $h_n(\mu, \nu)$.

$$\begin{aligned} h_n(\mu, \nu) &= \sum_{l=1}^N \lambda^{-\mu l} \sin \frac{n(l-\nu)\pi}{N} \\ &= \operatorname{Im} \left\{ \sum_{l=1}^N \lambda^{-\mu l} e^{i \frac{n(l-\nu)\pi}{N}} \right\} \\ &= \operatorname{Im} \left\{ e^{-i \frac{n\nu\pi}{N}} \left(\frac{\lambda^{-\mu} e^{i \frac{n\pi}{N}} (1 - \lambda^{-\mu N} (-1)^n)}{1 - \lambda^{-\mu} e^{i \frac{n\pi}{N}}} \right) \right\} \\ &= \frac{1 - \lambda^{-\mu N} (-1)^n}{\lambda^\mu} \operatorname{Im} \left\{ \frac{e^{i \frac{n(1-\nu)\pi}{N}}}{1 - \lambda^{-\mu} e^{i \frac{n\pi}{N}}} \right\} \\ &= \frac{1 - \lambda^{-\mu N} (-1)^n}{\lambda^\mu} \operatorname{Im} \left\{ \frac{e^{i \frac{n(1-\nu)\pi}{N}} (1 - \lambda^{-\mu} e^{-i \frac{n\pi}{N}})}{1 + \lambda^{-2\mu} - 2\lambda^{-\mu} \cos \frac{n\pi}{N}} \right\} \\ &= \frac{\lambda^{-\mu} (1 - \lambda^{-\mu N} (-1)^n) (\sin \frac{n(1-\nu)\pi}{N} + \lambda^{-\mu} \sin \frac{n\nu\pi}{N})}{1 + \lambda^{-2\mu} - 2\lambda^{-\mu} \cos \frac{n\pi}{N}}. \end{aligned} \tag{B10}$$

By Eqs. (B1)–(B6), we get the following

Theorem 2. The modified Fourier power spectrum of an AR(1) process is distributed as

$$\frac{\sigma^2}{N^2} (H_1 + H_2 - H_3) \mathbf{Z}^2,$$

where \mathbf{Z} is a Gaussian random variable with mean 0 and variance 1, and

$$\begin{aligned} H_1 &= (1 + \lambda^2 - 2\lambda \cos \frac{n\pi}{N})^{-2} (g_n(0, 0, 0) \\ &+ \frac{\lambda^{2N+2} - \lambda^2}{\lambda^2 - 1} \sin^2 \frac{n\pi}{N} + \lambda^2 g_n(0, 1, 1) \\ &+ 2(-1)^{n+1} \lambda^{N+1} \sin \frac{n\pi}{N} h_n(1, 0) \\ &- 2\lambda g_n(0, 0, 1) + 2(-1)^n \lambda^{N+2} \sin \frac{n\pi}{N} h_n(1, 1)), \end{aligned}$$

$$H_2 = \frac{(\lambda^{2N} - 1) \sin^2 \frac{(N-1)n\pi}{N}}{16(\lambda^2 - 1) \sin^4 \frac{n\pi}{2N}},$$

$$\begin{aligned} H_3 &= \frac{\lambda^N \sin \frac{(N-1)n\pi}{N}}{2(1 + \lambda^2 - 2\lambda \cos \frac{n\pi}{N}) \sin^2 \frac{n\pi}{2N}} \times \\ &\left(h_n(1, 0) + \frac{(-1)^{n+1} (\lambda^{N+1} - \lambda^{-N+1}) \sin \frac{n\pi}{N}}{\lambda^2 - 1} - \lambda h_n(1, 1) \right). \end{aligned}$$

and $g_n(\mu, \nu, \zeta)$, $h_n(\mu, \nu)$ are stated in Eqs. (B9) and (B10), respectively.

Remark. When the length of geophysical time series is large, we suggest to use Theorem 2 instead of Theorem 1 to do a significance test.

Acknowledgements. We thank two anonymous referees for helpful comments and suggestions. This research was partially supported by Fundamental Research Funds for the Central Universities (Key Program), NSFC No. 41076125, 973 project No. 2010CB950504, 863 project No. 2010AA012305, Polar Climate and Environment Key Laboratory, and State Key Lab of Earth Surface Processes and Resource Ecology.

Edited by: J. Kurths

Reviewed by: two anonymous referees

References

- Baldwin, M. P. and T. J. Dunkerton, Stratospheric harbingers of anomalous weather regimes, *Science*, 294, 581–584, 2001.
- Brockwell, P. J. and Davis, R. A., *Time series: Theory and Methods*, Springer-Verlag, New York, 1991.
- Chatfield, C.: *The Analysis of Time Series: An Introduction*, 4th Edn., Chapman and Hall, 1989.
- Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P.: Advanced spectral methods for climatic time series, *Rev. Geophys.*, 40, 1003–1043, 2002.
- Gilman, D. L., Fuglister, F. J., and Mitchell Jr., J. M.: On the power spectrum of “red noise”, *J. Atmos. Sci.*, 20, 182–184, 1963.
- Jenkins, G. M. and Watts, D. G.: *Spectral Analysis and Its Applications*, Holden-Day, 1968.
- Jevrejeva, S., Moore, J., and Grinsted, A.: Influence of the Arctic Oscillation and El Nino-Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach, *J. Geophys. Res.*, 108, 4677–4687, 2003.
- Jevrejeva, S., Moore, J., and Grinsted, A.: Oceanic and atmospheric transport of multiyear El Nino-Southern Oscillation (ENSO) signatures to the polar regions, *Geophys. Res. Lett.*, 31, L24210/1–4, doi:10.1029/2004GL020871, 2004.
- Kaplan, A., Cane, M., Kushnir, Y., Clement, A., Blumenthal, M., and Rajagopalan, B.: Analyses of global sea surface temperature 1856–1991, *J. Geophys. Res.*, 103, 18567–18589, 1998.
- Mann, M. E. and Lees, J. M.: Robust Estimation of Background Noise and Signal Detection in Climatic Time Series, *Clim. Change*, 33, 409–445, 1996.
- Oppenheim, A. V. and Schaffer, R. W.: *Discrete-Time Signal Processing*, Prentice-Hall, 1989.
- Thompson, D. W. J. and Wallace, J. M.: The Arctic Oscillation signature in the winter geopotential height and temperature fields, *Geophys. Res. Lett.*, 25, 1297–1300, 1998.
- Torrence, C. and Compo, G. P.: A Practical Guide to Wavelet Analysis, *B. Am. Meteorol. Soc.*, 79, 61–78, 1998.
- Zar, J. H.: *Biostatistical Analysis*, Prentice hall, New Jersey, 1999.