· Article ·

Advances in Polar Science

doi:10.3724/SP.J.1085.2011.00192

September 2011 Vol.22 No.3 192–198

Improved significance testing of wavelet power spectrum near data boundaries as applied to polar research

ZHANG ZhiHua[1*] & John C Moore[2,3,4*]

[1]College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China;

[2]State Key Laboratory of Earth Surface Processes and Resource Ecology/College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China;

[3]Arctic Centre, University of Lapland, Rovaniemi, Finland;

[4]Department of Earth Sciences, Uppsala University, Sweden

Received March 7, 2011; accepted August 25, 2011

Abstract When one applies the wavelet transform to analyze finite-length time series, discontinuities at the data boundaries will distort its wavelet power spectrum in some regions which are defined as a wavelength-dependent cone of influence (COI). In the COI, significance tests are unreliable. At the same time, as many time series are short and noisy, the COI is a serious limitation in wavelet analysis of time series. In this paper, we will give a method to reduce boundary effects and discover significant frequencies in the COI. After that, we will apply our method to analyze Greenland winter temperature and Baltic sea ice. The new method makes use of line removal and odd extension of the time series. This causes the derivative of the series to be continuous (unlike the case for other padding methods). This will give the most reasonable padding methodology if the time series being analyzed has red noise characteristics.

Keywords **Wavelet power spectrum, significance testing, Greenland winter temperature, Baltic sea ice**

Citation: Zhang Z H, Moore J C. Improved significance testing of wavelet power spectrum near data boundaries as applied to polar research. Adv Polar Sci, 2011, 22: 192–198, doi: 10.3724/SP.J.1085.2011.00192

0 Introduction

In the early 1980s, Morlet introduced the wavelet transform and applied it in geophysics. After two decades of fast development, the wavelet method has become routine[1]. The continuous wavelet transform possesses the ability to construct a time-frequency representation of a time series that offers very good time and frequency localization, so wavelet transforms can analyze localized intermittent periodicities of geophysical time series. However, physical interpretation of the signals requires plausible statistical significance testing. For many geophysical time series, an appropriate background noise or null hypothesis is red noise, so in order to distinguish their intrinsic feature from red noise, one needs to use a significance test.

In practice, since one does not know the information of the time series in the past or in the future, the time series is usually padded with zeroes and then its wavelet power spectrum is computed. After that, a significance test may find significant regions. However, discontinuities at data boundaries will distort the wavelet power spectrum in some regions which are defined as a

*Corresponding authors(Zhang Zhihua, email: zhangzh@bnu.edu.cn; John C Moore, email: john.moore@bnu.edu.cn)

wavelength-dependent cone of influence (COI), so the significance test will be unreliable in the COI. Moreover, since many geophysical time series are short and noisy, the COI will be a serious limitation in wavelet analysis of time series. Many scientists have presented various methods to deal with boundary effects in wavelet analysis: Foster[2] and Frick et al.[3] constructed a new wavelet near the gap to deal with the wavelet transform of unevenly sampled time series, while Sweldens[4] presented the "lifting scheme" algorithm of constructing new wavelet bases. Johnson[5] modified the traditional wavelet transform by adapting wavelets through renormalization and modifying the shape of the analyzing window. Unfortunately, their methods cannot be applied to do significance tests and extract instinct features in the COI of wavelet power spectrum. In our paper, instead of defining a new wavelet, we suggest a simple approach to the improvement to significance testing of wavelet power spectrum near data boundaries. In Section 1, we will review the cone of influence and significance test of wavelet power spectrum. In Section 2, we give a method to reduce boundary effect. In Section 3, we apply our method to research Greenland winter temperature and Baltic sea ice.

1 Cone of influence and significance test of wavelet power spectrum

In geophysics, due to the similar resolution in both time and frequency, the most popular wavelet is the Morlet wavelet[1] whose representation is

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

where ω_0 is the nondimensional frequency. In application, one often takes ω_0=6. Throughout this paper, the wavelet we use is the Morlet wavelet. Let f be a continuous time series. Then the wavelet transform of f is defined as[6]

$$W_f(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \overline{\psi_0(\frac{t-b}{a})} \mathrm{d}t \quad (a>0, b \in R) \tag{2}$$

where the overbar denotes a complex conjugate. We call $|W_f(a,b)|^2$ the wavelet power spectrum of f. For the discrete time series $\{x_n\}_0^{N-1}$ of time step δt, its wavelet transform is defined as

$$W_n(s) = \sqrt{\frac{\delta t}{s}} \sum_{n'=0}^{N-1} x'_n \overline{\psi_0[\frac{(n'-n)\delta t}{s}]} \tag{3}$$

We call $|W_n(s)|^2$ wavelet power spectrum of $\{x_n\}_0^{N-1}$. The wavelet transform is very useful for time series analysis where smooth, continuous variations in wavelet amplitude are expected.

Liu et al.[7] believed that there existed some bias in wavelet power spectrum and adjusted it by dividing by the corresponding scale. But the significant regions in the wavelet power spectrum will not be affected by using Liu et al's rectification algorithm if both the signal and noise background are treated the same.

1.1 Cone of influence

In the real world, since we do not have information on the time series in the past or in the future, we only can deal with the time series of finite length from T_1 to T_2. In order to compute its wavelet transform, the simplest algorithm is to pad the time series with zeroes and then use the following approximation

$$\begin{aligned} W_f(a,b) &= \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \overline{\psi_0(\frac{t-b}{a})} \mathrm{d}t \\ &\approx \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} F(t) \overline{\psi_0(\frac{t-b}{a})} \mathrm{d}t = W_F(a,b) \end{aligned} \tag{4}$$

where

$$F(t) = \begin{cases} f(t), & t \in [T_1, T_2], \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

In this formula, we use the wavelet transform of F to approximate to that of f. Because of discontinuities at the endpoints, errors will occur at the beginning and end of the wavelet transform. The COI is the region of the wavelet transform in which edge effects become important and is defined usually as the e-folding time for the autocorrelation of wavelet transform at each scale[1]. So the values of the wavelet transform (or the wavelet power spectrum) obtained in the COI are distorted.

1.2 Background noise and significance test

For many geophysical phenomena, the plausible null hypothesis to test against is red noise[8]. A simple model for red noise is the univariate lag-1 autoregressive [AR(1)] process as follows:

$$x_0 = 0, \quad x_{n+1} = \lambda x_n + \in_{n+1}, \quad n = 1, 2, 3, \ldots, N-1 \tag{6}$$

where λ is a constant and ε_n is independent Gaussian white noise with mean 0 and variance σ^2. Torrence and Compo[1] showed that, for the Morlet wavelet, the wavelet power spectrum $|W_n(s)|^2$ of an AR(1) process satisfies that

$$\frac{|W_n(s)|^2}{\widetilde{\sigma}^2} \text{ is distributed as } \frac{1}{2}P_k\chi_2^2 \tag{7}$$

where $\widetilde{\sigma}^2$ is the variance of time series, χ_2^2 is the distribution of the sum of the squares of two independent standard normal random variables,

$$P_k = \frac{1-\lambda^2}{1+\lambda^2-2\lambda\cos(2\pi k/N)} \tag{8}$$

and k is the Fourier frequency corresponding to the wavelet scale s. For the Morlet wavelet (see formula (1)), $k = \dfrac{N(\omega_0+\sqrt{2+\omega_0^2})}{4\pi s}$. If we take ω_0=6, then $k = \dfrac{N}{1.03s}$.

For a typical geophysical time series, If a region G is such that the values of the wavelet power spectrum on G are all outside the α% confidence interval of the distribution for wavelet power spectrum of an AR(1) red noise, then we call G a significant region. In geophysical research, one always takes α =90 or 95.

Many statistical tests assume that the probability density function (pdf) is close to normal. Before one uses the wavelet transform to analyze typical geophysical time series, one often transforms the original time series such that the pdf of the transformed data is normal[9−10]. A practical way of doing this is by taking the inverse normal cumulative distribution function (cdf). The main reason for doing it is because otherwise the normally distributed red noise null hypothesis we use is wrong. Further the AR(1) estimators can be quite poor at dealing with non-normal data. So, if the data are not normalized, then the obtained significant regions will be misleading simply because the null-hypothesis can be trivially rejected.

2 Reducing the boundary effect

In order to reduce the data boundary effect and obtain the better wavelet power spectrum in the COI, several algorithms have been tried previously. The first algorithm[11] is to pad the time series on each endpoint with zeroes. The algorithm artificially creates discontinuities at the endpoints. The second algorithm is "cosine damping"[11] which preprocesses the time series $f(t), t \in [T_1, T_2]$ as follow

$$g(t) = f(t) \times [1-\cos^2(\frac{t-T_1}{T_2-T_1}\pi)] \tag{9}$$

It is clear that at the endpoints, $g(T_1) = g(T_2)$=0. Although this method overcomes the boundary effect, at the same time, we lose almost all the information from the wavelet power spectrum at the endpoints. This is often an unacceptable artifact. The third algorithm is even symmetric extension[11]. This algorithm guarantees that the signal is continuous at the endpoints. However, Meyers et al.[11] indicated that it has the disadvantage of artificially creating discontinuities in the first derivative at the endpoints.

Now, we will give a new extension algorithm to remove boundary effect in wavelet power spectra. The advantage of our algorithm is that the first derivatives of extended time series at the endpoints are still continuous. Let a function $f(t)$ be defined on $[T_1, T_2]$.

Step 1. We divide into two parts. The first part is a segment line joining two points $(T_1, f(T_1))$ and $(T_2, f(T_2))$

$$u(t) = \frac{f(T_2)-f(T_1)}{T_2-T_1}(t-T_1)+f(T_1) \tag{10}$$

The second part is

$$v(t) = f(t) - u(t) \tag{11}$$

From here, we know that

$$v(T_1) = v(T_2) = 0 \tag{12}$$

Step 2. We do the odd extension for v:

$$v_{odd}(t) = v(t),\ t \in [T_1, T_2] \tag{13}$$

$$V_{odd}(t) = -v(2T_1-t),\ t \in [2T_1-T_2, T_1] \tag{14}$$

Step 3. We do the periodic extension with period $2(T_2-T_1)$ for $v_{odd}(t)$, i. e.,

$$v^*(t) = v_{odd}(t),\ t \in [2T_1-T_2, T_2] \tag{15}$$

$$v^*(t+2k(T_2-T_1)) = v^*(t),$$
$$t \in [2T_1-T_2, T_2],\ k = \pm1, \pm2, \ldots \tag{16}$$

Step 4. We do the wavelet transform for v^* as

$$W_{v^*}(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} v^*(t)\overline{\psi_0(\frac{t-b}{a})}\mathrm{d}t \tag{17}$$

where ψ_0 is the Morlet wavelet.

Below we give two propositions on our method.

Proposition 1. In (17), for any polynomial $h(t) = ct + d$ of degree one, we have $W_h(a,b) \approx 0$.

Proof. The Morlet wavelet ψ_0 has one vanishing moment approximately and the error is very small, then $\int_{-\infty}^{\infty} \psi(t)\mathrm{d}t \approx 0$ and $\int_{-\infty}^{\infty} t\psi(t)\mathrm{d}t \approx 0$. So

$$W_h(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} (ct+d)\overline{\psi_0(\frac{t-b}{a})}\mathrm{d}t = \sqrt{a} \int_{-\infty}^{\infty} [c(at+b)+d]\overline{\psi_0(t)}\mathrm{d}t \approx 0 \qquad (18)$$

Remark 1. For a wavelet with vanishing moments up to some power p, the corresponding wavelet transform of a polynomial of order p is zero.

Proposition 2. Let a function $f(t)$ be defined on $[T_1, T_2]$ and $v^*(t)$ be defined in Step 3. If the derivative $f'(t)$ is continuous on $[T_1, T_2]$, then the derivative $v^{*'}(t)$ is continuous on $(-\infty, \infty)$.

Proof. Since f' is continuous on the interval $[T_1, T_2]$, from the construction 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 T_1 and T_2. By the similarity of arguments, now we consider the point T_1.

First we prove that the derivative of v^* exists at T_1. By Step 2, $v_{odd}(t) = v(t)$, $t \in [T_1, T_2]$. We consider the right derivative of v_{odd} at the endpoint T_1,

$$\lim_{t\to T_1+} \frac{v_{odd}(t) - v_{odd}(T_1)}{t - T_1} = \lim_{t\to T_1+} \frac{v(t) - v(T_1)}{t - T_1} = v'(T_1) \qquad (19)$$

By Step 2 and Step 1,

$$v_{odd}(t) = -v(2T_1 - t), \quad t \in [2T_1 - T_2, T_1] \qquad (20)$$

and $v_{odd}(T_1) = v(T_1) = 0$. Considering the left derivative of v_{odd} at the endpoint T_1, we obtain that

$$\lim_{t\to T_1-} \frac{v_{odd}(t) - v_{odd}(T_1)}{t - T_1} = \lim_{t\to T_1-} \frac{v(T_1) - v(2T_1 - t)}{T_1 - (2T_1 - t)} \qquad (21)$$

Let $t' = 2T_1 - t$. Then, for $t \to T_1$, $t < T_1$, we have $t' \to T_1$, $t' > T_1$. So we obtain that

$$\lim_{t\to T_1-} \frac{v_{odd}(t) - v_{odd}(T_1)}{t - T_1} = \lim_{t'\to T_1+} \frac{v(T_1) - v(t')}{T_1 - t'} = v'(T_1) \qquad (22)$$

Combining (19) with (22), we know that the derivative of v_{odd} exists at the endpoint T_1 and

$$v'_{odd}(T_1) = v'(T_1) \qquad (23)$$

By Step 3, $v^*(t) = v_{odd}(t)$, $t \in [2T_1 - T_2, T_2]$ and $2T_1 - T_2 < T_1 < T_2$, we know that the derivative of v^* exists at T_1.

Next we prove that the derivative of v^* is continuous at T_1. From $v_{odd}(t) = -v(2T_1 - t), t \in [2T_1 - T_2, T_1]$, we have

$$v'_{odd}(t) = v'(2T_1 - t), \quad t \in [2T_1 - T_2, T_1] \qquad (24)$$

This implies that

$$\lim_{t\to T_1-} v'_{odd}(t) = \lim_{t\to T_1-} v'(2T_1 - t) = \lim_{t\to T_1+} v'(t) = v'(T_1) = v'_{odd}(T_1) \qquad (25)$$

On the other hand, by $v_{odd}(t) = v(t), t \in [T_1, T_2]$, we have

$$\lim_{t\to T_1+} v'_{odd}(t) = \lim_{t\to T_1+} v'(t) = v'(T_1) = v'_{odd}(T_1) \qquad (26)$$

By (25) and (26), we get $\lim_{t\to T_1-} v'_{odd}(t) = v'_{odd}(T_1)$, i.e., $v'_{odd}(t)$ is continuous at $t = T_1$. By Step 3, we know that the derivative of $v^*(t)$ is continuous at T_1. Proposition 2 is proved.

From Proposition 1, we know that although we remove a linear function from the original signal, the corresponding wavelet power spectrum will not be changed. By Proposition 2, our algorithm "linear removal+odd extension" guarantees that the derivatives of the time series are continuous at the endpoints. So our algorithm can overcome boundary effect better than "zero-padding" and "even extension (even-padding)".

3 Application

We will examine Southern Greenland winter temperature index and Baltic Sea ice index with the help of our algorithm, i.e., we will make a significance test against the null hypothesis of climate noise and determine significant regions near data boundaries.

3.1 Southern Greenland winter temperature index

The isotopic ratio $\delta^{18}O$ measured in ice cores can be used as a temperature proxy because of the temperature dependent fractionation of oxygen isotopes, that takes place while moisture travels from its evaporation area to the Greenland ice sheet. Vinther et al.[12] analyzed ice core data from 7 different drill sites in the southern half of Greenland in order to get Southern Greenland winter temperature indices. We apply our method to analyze the wavelet power spectrum of this time series. So, first of all, we transform the original data such that the pdf of the transformed data is normal (Figure 1). In order to show the performance of our method, we consider the middle part of the full temperature indices. The wavelet power spectrum of the middle part can be computed by using full length data (Figure 2a). Now we only use the middle part to compute its wavelet power spectrum by "zero-padding", "even extension", and our algorithm (Figures 2b–2d). In the COI, the significant region indicated by our method is more closer to the real significant region than that by "zero-padding" or "even extension".

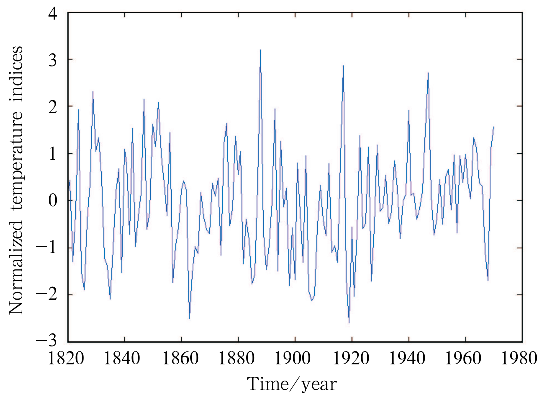

Figure 1 Thenormalized Southern Greenland winter temperature indices.

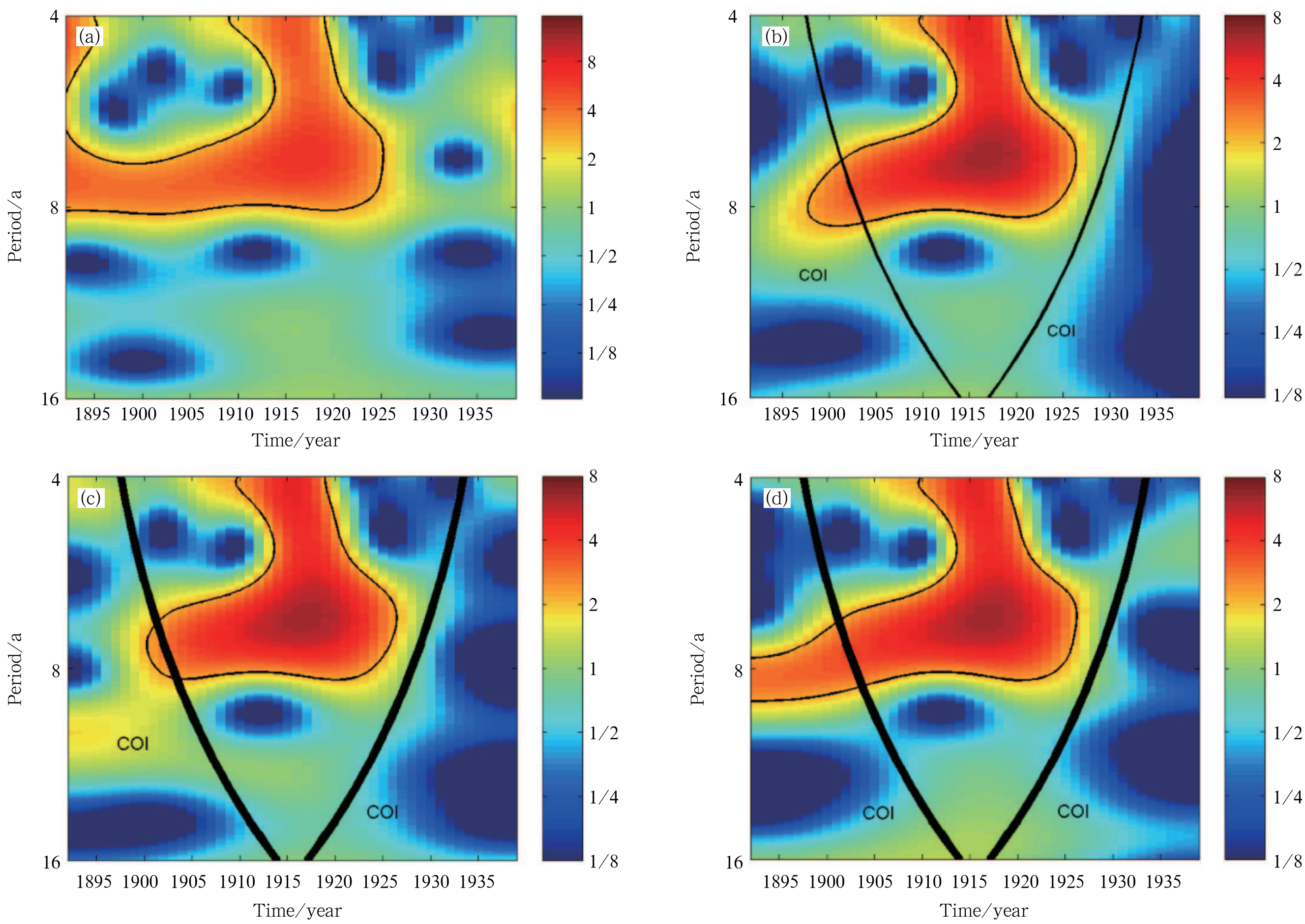

Figure 2 Wavelet power spectrum of Southern Greenland winter temperature indices is obtained by using: (a) full-length data, (b) "zero-padding" method, (c) "even-padding" method, and (d) our method. The black contour designates the 90% significance level against red noise, and COI is just the region below the thick line.

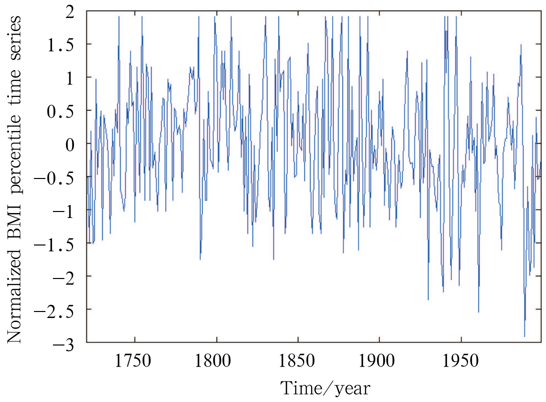

Figure 3 The normalized winter BMI percentile time series.

3.2 Baltic Sea ice index

The maximum extent of the Baltic Sea ice (BMI) is characterized by a bi-modal pdf with maximum likelihoods near 10 and 100% ice covered[13]. We transform the original data such that the pdf of the transformed data is normal (Figure 3). We take the 1765–1870 part of the full-length BMI indices. It is obvious that we can obtain the true wavelet power spectrum and significant region of this 105-year data by using full length data. The true wavelet power spectrum of the 105-year data is shown in Figure 4a. Now we use the 105-year data to compute the wavelet power spectrum by using three approaches: "zero-padding", "even extension" and our method (Figures 4b–4d). If we compare the shapes of two significant regions on the top right corner of wavelet power spectrum, it is obvious that the significant region obtained by our extension algorithm is closer to true significant region than that obtained by "zero-padding" or "even extension" (Figure 4).

4 Summary and conclusion

Discontinuities at the data boundaries cause the

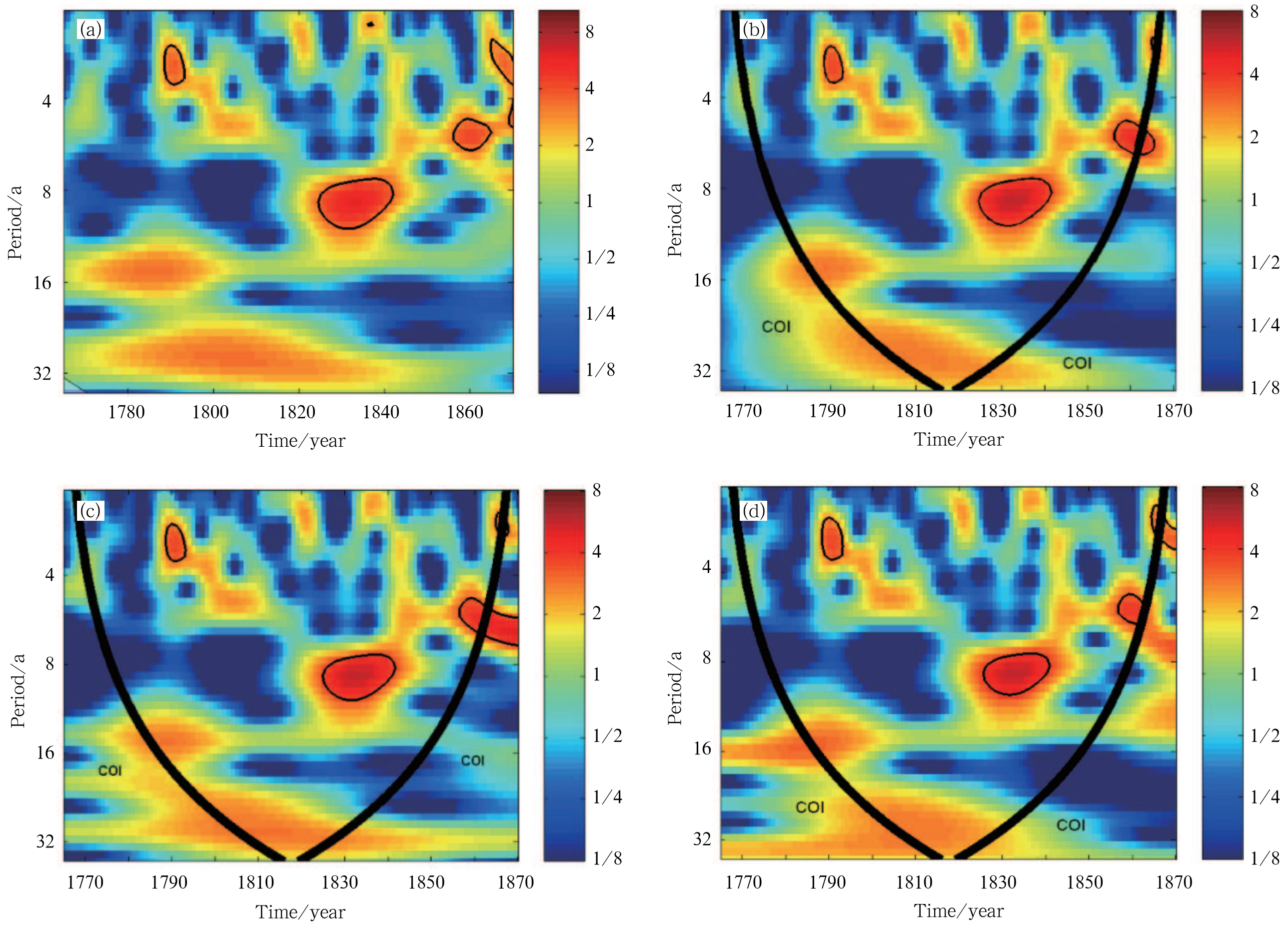

Figure 4 Wavelet power spectrum of BMI indices is obtained by using: (a) full-length data, (b) "zero-padding" method, (c) "even-padding" method, and (d) our method. The black contour designates the 95% significance level against red noise, and COI is just the region below the thick line.

distortion of its wavelet power spectrum. The even-extension method guarantees that the time series is continuous at the endpoints. However, the first derivatives on the endpoints are still discontinuous. In contrast with the even-extension method, our "linear removal+odd extension" method has a continuous first derivative, so our algorithm can overcome boundary effect better than "even extension". The desirable quality of continuous time derivative means that our "linear removal + odd extension" is a good choice of padding for time series that have significant first order autoregressive components, or exhibit persistence (such as a Markov or fractional Gaussian process).

We tested the new method of "linear removal+odd extension" with two different real-world climate proxies. One of which was unusually non-Normal in its pdf. We extracted the middle parts of these times series and compared different methods of testing significance levels near the data boundaries with the known significant regions from the complete time series. It is clear that the new method produces more reliable estimations of the significant regions of the wavelet power spectrum than existing alternatives. We therefore commend this new method as worthwhile.

No method can forecast future behavior completely accurately, especially if the considered time series is related to climate which is non-stationary and potentially non-linear. As a general rule, we recommend using additional techniques to confirm possibly marginally significant features affected by the COI, such as singular spectrum analysis[14] or Granger causality[15] that may have different advantages and shortcomings.

Acknowledgments This research was partially supported by the National Key Science Program for Global Change Research (Grant no. 2010CB950504), the National High-Technology Research & Development Program of China (863 Program, Grant no. 2010AA012305), the National Natural Science Foundation of China(Grant no. 41076125), the Fundamental Research Funds for the Central Universities (Key Program), and the Polar Climate and Environment Key Laboratory.

References

1 Torrence C, Compo G P. A practical guide to wavelet analysis. Bull Am Meteorol Soc, 1998, 79: 61–78

2 Foster G. Wavelets for period analysis of unevenly sampled timeseries. Astron J, 1996, 112: 1709–1729

3 Frick P, Baliunas S L, Galyagin D, et al. Waveletanalysis of stellar chromospheric activity variations. Astrophys J, 1997, 483: 426–434

4 Sweldens W. The lifting scheme, a construction of second generationwavelets. SIAM J Math Anal, 1998, 29: 511–546

5 Johnson R W. Edge adapted wavelets, solar magnetic activity, andclimate Change. Astrophys Space Sci, 2010, 326: 181–189

6 Daubechies I. Ten Lectures on Wavelets, CBMS-Conference Lecture Notes, Philadelphia, SIAM, 1992

7 Liu Y, Liang X S, Weisberg R H. Rectification of the bias in thewavelet power spectrum. J Atmos Oceanic Technol, 2007, 24: 2093–2102

8 Mann M E, Lees J M. Robust estimation of background noise andsignal detection in climatic time series. Climate Change, 1996, 33: 409–445

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

10 Jevrejeva S, Moore J C, 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, 2003, 108(D21): 4677–4687

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

12 Vinther B M, Johnsen S J, Andersen K K, et al. NAO signal recorded in the stable isotopes of Greenland ice cores. Geophys Res Letters, 2003, 30: 1387–1390

13 Seina A, Gronvall H, Kalliosaari S, et al. Ice seasons 1996–2000 in Finnish sea areas /Jaatalvet 1996–2000 Suomenmerialueilla. Meri, Report Series of the Finnish Inst. of Marine Res., 2001, 43

14 Ghil M, et al. Advanced spectral methods for climatic time series. Rev Geophys, 2002, 40: 1003–1043

15 Granger C W J. Investigating causal relations by econometric modelsand cross-spectral methods. Econometrica, 1969, 37: 424–438