Intrinsic feature extraction in the COI of wavelet power spectra of climatic signals

Zhihua Zhang
College of Global Change and Earth System Science
Beijing Normal University
Beijing, China
Email: zhangzh@bnu.edu.cn

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

Arctic Centre, University of Lapland, Finland
Dept of Earth Sciences, Uppsala University, Sweden

Abstract— **Since the wavelet power spectra are distorted at data boundaries (the cone of influence, COI), using traditional methods, one cannot judge whether there is a significant region in COI or not. In this paper, with the help of a first-order autoregressive (AR1) extension and using our simple and rigorous method, we can obtain realistic significant regions and intrinsic feature in the COI of wavelet power spectra. We verify our method using the 300 year record of ice extent in the Baltic Sea.**

Keywords-wavelet power spectrum; AR1 extension; feature extraction

I. INTRODUCTION

In the past twenty years, wavelet analysis has been widely applied in many branches of science and engineering [2-3, 6-7]. The continuous wavelet transform possesses the ability to construct a time-frequency representation of a signal that offers very good time and frequency localization, so wavelet transforms can analyze localized intermittent periodicities of potentially great interest in climatic signals. In climatic signal analysis, one needs to distinguish intrinsic feature from background noise. The cone of influence (COI) is the region where the wavelet power spectra are distorted because of the influence of the end points of finite length signals. Based on the traditional methods, it is difficult to extract intrinsic feature in these region of the wavelet power spectra of climatic signals. In this study, we will tackle this issue. With the help of the first-order autoregressive (AR1) extension and using our simple and rigorous analytical methods, we can obtain realistic levels of significance in wavelet power spectra near data boundaries.

II. WAVELET TRANSFORM AND CONE OF INFLUENCE

Let f be a continuous signal. Then the wavelet transform of f is defined as [2-3, 6]

$$W_f(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t)\overline{\psi\left(\frac{t-b}{a}\right)}\,dt \quad (a>0, b\in R) \quad (1)$$

where ψ is called a wavelet and $\left|W_f(a,b)\right|^2$ is called the wavelet power spectrum of f.

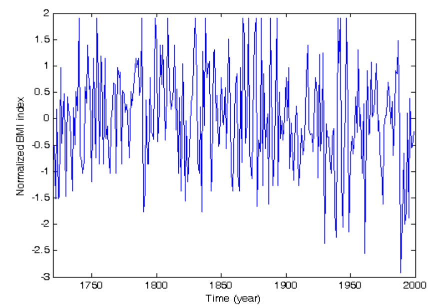

Figure 1. The normalized winter BMI index (the annual maximum ice extent in the Baltic Sea)

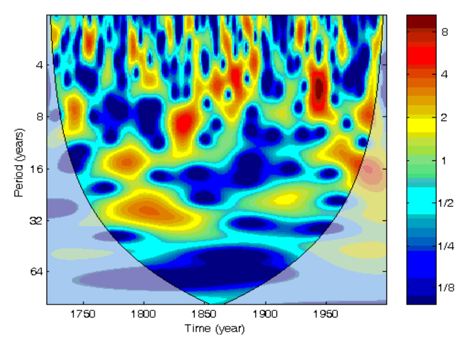

Figure 2. Wavelet power spectrum of BMI indices obtained by "zero padding". The cone of influence (COI) is shown as a lighter shade.

For the discrete signal $\{x_n\}$, its wavelet transform is defined as [2,3,6]

$$W_n(s) = \sqrt{\frac{\delta t}{s}}\sum_{n'} x_{n'}\overline{\psi\left(\frac{(n-n')\delta t}{s}\right)} \quad (2)$$

We call $|W_n(s)|^2$ the wavelet power spectrum of $\{x_n\}$.

In climatic signal processing, due to its similar resolution in both time and frequency, one often uses Morlet wavelet whose representation is

$$\psi(t) = \pi^{-1/4}e^{i6t}e^{-\frac{t^2}{2}} \quad (3)$$

In order to apply wavelet transform to finite-length signals, one often pads the end of signals with zeroes before doing the wavelet transform. Errors will occur at the wavelet power spectrum near data boundary. The cone of influence (COI) is

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

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 [2,3,6]. Moreover, many climatic signals are short and noisy, hence the COI is a serious limitation in wavelet analysis of climatic signals. Due to the distortion of the wavelet power spectrum in the COI, it will be difficult to extract intrinsic feature in the COI of wavelet power spectra of climatic signals. As an example, Figure 1 shows the normalized index for the maximum extent of the Baltic Sea ice (BMI index). Figure 2 shows the wavelet power spectra of BMI index obtained by "zero padding". The cone of influence (COI) is shown as a lighter shade.

III. BACKGROUND NOISE AND TRADITIONAL INSTINSIC FEATURE EXTRACTION MEHTOD

When one wishes to extract intrinsic feature of climatic signals from background noise, white noise is not a suitable model for background noise. A plausible noise model is red noise [4]. A simple model for red noise is the univariate lag-1 autoregressive process (in short, AR1) as follows:

$$x_0 = 0, \quad x_{n+1} = \lambda x_n + \varepsilon_n, \quad n = 1,2,\dots,N, \quad (3)$$

where λ is called AR1 coefficient and ε_n is the independence Gaussian white noise with mean 0 and variance a^2.

The wavelet power spectrum $W_n(s)$ of AR1 red noise satisfies[2,3,6]:

$$\frac{|W_n(s)|^2}{\sigma^2} \text{ is distributed as } \frac{1}{2}P_k\chi_2^2 \quad (4)$$

where σ^2 is the variance of the signal, k is the Fourier frequency corresponding to the wavelet scale s,

$$P_k = \frac{1-\lambda^2}{1+\lambda^2-2\lambda\cos(2\pi k/N)} \quad (5)$$

and χ_2^2 is the distribution of the sum of the squares of two independent standard normal random variables.

The probability density function of χ_2^2 is

$$f(t) = \frac{1}{2}e^{-\frac{t}{2}}, t > 0 \text{ and } f(t) = 0, t \le 0 \quad (6)$$

If, for some η, $\int_0^\eta f(t)dt = 95\%$, then we say the interval $[0,\eta]$ is called 95% confidence interval of χ_2^2 distribution.

For a typical climatic signals, if a region G is such that the values of the wavelet power spectrum on G are all outside the 95% confidence interval of the distribution of the wavelet power spectrum $W_n(s)$ of AR1 red noise, i.e.,

$$G = \left\{(n,s): \quad \frac{2|W_n(s)|^2}{\sigma^2 P_k} > \eta\right\} \quad (7)$$

then one calls G a significant region.

The wavelet power spectrum of climatic signals in significant region contains intrinsic features of climatic signals.

IV. INSTINSIC FEATURE EXTRACTION IN THE COI

In this section, we will give a method to distinguish the intrinsic features of the wavelet power spectrum in the COI from AR1 red noise. In order to achieve this purpose, we first approximate the signal $\{x_n\}_0^N$ by an AR1 process, in the other words; we will choose the suitable AR1 coefficient $\hat{\lambda}$ λ and the suitable noise variance $\hat{a}^2$ such that the corresponding AR1 process can approximate to $\{x_n\}_0^N$ very well. After that, we extend this signal by using an AR1 process with AR1 coefficient $\hat{\lambda}\lambda$ and noise variance $\hat{a}^2$. If this original signal is just an AR1 process, then so is the extended signal. Therefore, if we find some significant regions in the wavelet power spectrum of this extended signal, then the original signal must be different from an AR1 process and the wavelet power spectrum in significant regions then contain the intrinsic features. Although the constructed extension may be different from the behavior of the real signal outside the given time interval, we can still find the location of significant regions in the COI of its wavelet power spectrum.

In detail, our method is as follows.

Without loss of generalization, we assume that $\{x_n\}_0^N$ is a Normalized signal with mean 0 and variance 1. If necessary, we transform the original data such that the pdf of the transformed data is Normal

Step 1. Choose the suitable AR1 coefficient $\hat{\lambda}$ and the noise variance $\hat{a}^2$ such that we approximate to $\{x_n\}_0^N$ by AR1 process very well.

Step 2. We extend the signal $\{x_n\}_0^N$ at its right endpoint as follows:

$$x_{N+l} = \hat{\lambda}x_{N+l-1} + \zeta_l, \quad l = 1,2,\dots m \quad (8)$$

where ζ_l is a sample of the Gaussian white noise with mean 0 and variance $\hat{a}^2$.

Now we extend the signal $\{x_n\}_0^N$ at its left endpoint. We will construct a sample of AR1 process as follows:

$$y_0 = 0, \; y_l = \hat{\lambda}y_{l-1} + \xi_l, \quad l = 1,2,\dots m \quad (9)$$

where ξ_l is a sample of the Gaussian white noise with mean 0 and variance $\hat{a}^2$.

If $y_m \approx x_0$, then, we define

$$x_l = y_{m+l}, \quad l = -1,-2,\dots -m \quad (10)$$

If the formula $y_m \approx x_0$ does not hold, we need to construct another sample of AR1 process again and again until the formula $y_m \approx x_0$ holds.

Step 3. Compute the wavelet transform $W_n(s)$ for $\{x_n\}_{-m}^{N+m}$.

Step 4. With help of the distribution of the wavelet power spectrum $W_n(s)$ of AR1 red noise, noticing that the variance of the signal $\{x_n\}_0^N$ is 1, we define significance index as follows:

$$SIG_n(s) = \frac{2|W_n(s)|^2}{P_k\eta}, \; n = 1,2,\dots,N \quad (11)$$

where η is such that the interval $[0,\eta]$ is 95% confidence interval of χ_2^2 distribution, k is the Fourier frequency corresponding to the wavelet scale s, and

$$P_k = \frac{1-\hat{\lambda}^2}{1+\hat{\lambda}^2-2\hat{\lambda}\cos(2\pi k/N)} \quad (12)$$

Step 5. When we run from Step 1 to Step 4, each time, the extension is just a sample of an AR1 process, so $\mathrm{SIG}_n(s)$ obtained by us may depend on this extension. In order to avoid it, in practice, we need to repeat Steps 1 to 4 several times, and we obtain the significance indices $SIG_n^i(s), \; i = 1,2,\dots,L$.

Step 6. Let

$$\overline{SIGIF_n}(s) = \frac{1}{L}\sum_{i=1}^{L} SIG_n^i(s) \quad (13)$$

Then the significant region is

$$\Omega = \{(n, s): \quad \overline{SIGIF}_n(s) > 1\} \tag{14}$$

From this, we know that in the COI, the significant region of wavelet power spectrum is $\mathrm{COI} \cap \Omega$.

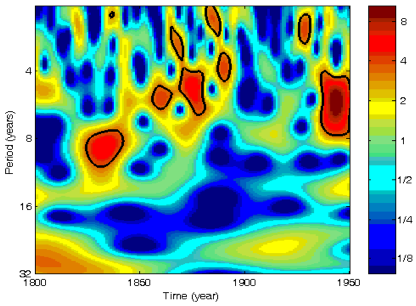

Figure 3. Real wavelet power spectrum of BMI indices. The significant regions are contoured by the thick black contour and contain intrinsic features of BMI indices.

V. APPLICATIONS

The Baltic Sea is a transition zone between the North Atlantic region and the continental area of Eurasia, leading to large inter-annual variability of ice conditions. The Baltic Sea is partly covered by ice every winter season, the maximum annual ice extent varies between 10–100% of the sea area, the length of ice season is 4–7 months, and the maximum annual thickness of ice is 50–120 cm. In this paper, we examine the connection ice conditions represented by maximum annual ice extent in the Baltic Sea (BMI indices) for the period 1720–2000 (see Figure 1). It is obvious that we can extract the significant regions in wavelet power spectrum of 1800-1950 BMI index by using full length data (Figure 3). In order to show the performance of our method, we assume that only the 1800-1950 part of the full-length BMI index is available. The wavelet power spectrum of the 1800-1950 BMI indices is shown in Figure 4. It is clear that outside COI, we can use traditional formula in Section III to obtain significant regions of wavelet power spectrum of the 1800-1950 BMI indices. However, due to the distortion of the wavelet power spectrum in the COI, traditional formula cannot be applied to deal with the COI of wavelet power spectrum. In the COI, we will use our method in Section IV to extract significant regions. In Figure 4, it is shown that there is a significant region in the 6 year band around 1930 in the COI of wavelet power spectrum. Comparing Figure 5 with Figure 3, we can see that our method can well extract realistic significance regions and intrinsic feature in the COI of wavelet power spectra.

VI. CONCLUSIONS

Since the wavelet power spectra are distorted near data boundaries, one cannot judge whether a significant region indicated near data boundaries is true or not. With the help of AR1 extension, we have shown that our simple and rigorous analytical methods of obtaining realistic levels of significance in wavelet power spectra close to data boundaries work well on time series that are far from ideal. In particular, our methods can deal with short and noisy geophysical time series effectively.

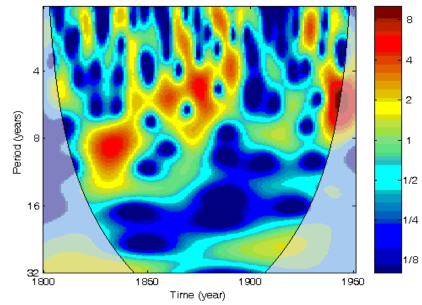

Figure 4. Wavelet power spectrum of 1800-1950 BMI indices obtained by ‘‘zero-padding’’ method. The cone of influence (COI) is shown as a lighter shade.

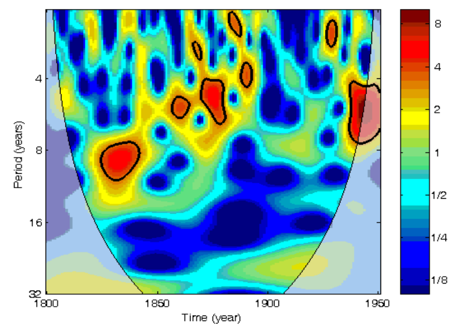

Figure 5. With the help of our method, the obtained significant regions are contoured by the thick black contour and contain intrinsic features of BMI indices. It is shown that there is a significant region in the 6 year band around 1930 in the COI of wavelet power spectrum

REFERENCES

[1] J. Moore, S. Jevrejeva, and A. Grinsted, Efficacy of geoengineering to limit 21st century sea-level rise. Proceedings of the National Academy of Sciences, vol. 107, pp. 15699-15703, 2010

[2] A. Grinsted, J. Moore, and S. Jevrejeva, Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlinear Processes in Geophysics vol.11, pp. 561-566, 2003

[3] S. Jevrejeva, J. Moore, and A. Grinsted, Influence of the Arctic Oscillation and El Nieno-Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach, J. Geophys. Res., vol. 108, pp. 4677, 2003.

[4] M. E. Mann and J. M. Lees, Robust estimation of background noise and signal detection in climatic time series. Climate Change, vol. 33, pp。409-445, 1996.

[5] A. Seina, H. Gronvall, S. Kalliosaari, and J. Vainio, Ice seasons 1996-2000 in Finnish sea areas/*Jaa*talvet 1996-2000 Suomen merialueilla, Meri, Report Series of the Finnish Institute of Marine Research, vol. 43, 2001

[6] C. Torrence, and G. P. Compo, A Practical Guide to Wavelet Analysis, Bull. Amer. Meteor. Soc., vol. 79, pp. 61-78, 1998

[7] Z. Zhang, A new method of constructions of non-tensor product wavelets, Acta Appl.Math., vol. 111, pp. 153–169, 2010.