

Lee Taesam (Orcid ID: 0000-0001-5110-5388)

Ouarda Taha BMJ (Orcid ID: 0000-0002-0969-063X)

Multivariate Nonstationary Oscillation Simulation of Climate Indices with Empirical Mode Decomposition

Taesam Lee¹ and Taha B. M. J. Ouarda²

¹Department of Civil Engineering, ERI, Gyeongsang National University, Jinju, South Korea

²Canada Research Chair in Statistical Hydro-Climatology, National Institute for Scientific Research, INRS-ETE, 490 de la Couronne, Quebec (QC), Canada G1K 9A9

Corresponding Author: Taesam Lee, Department of Civil Engineering, Gyeongsang National University, ERI, 501 Jinju-daero, Jinju, Gyeongsangnam-do, 660-701, South Korea, E-mail: tae3lee@gnu.ac.kr, Tel: +82 55 772 1797

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2018WR023892

Abstract

The objective of the current study is to build a stochastic model to simulate climate indices that are teleconnected with the hydrologic regimes of large-scale water resources systems such as the Great Lakes system. Climate indices generally contain nonstationary oscillations (NSOs). We adopted a stochastic simulation model based on Empirical Mode Decomposition (EMD). The procedure for the model is to decompose the observed series and then to simulate the decomposed components with the NSO resampling (NSOR) technique. Because the model has only been previously applied to single variables, a multivariate version of NSOR (M-NSOR) is developed to consider the links between the climate indices and to reproduce the NSO process. The proposed M-NSOR model is tested in a simulation study on the Rössler system. The simulation results indicate that the M-NSOR model reproduces the significant oscillatory behaviors of the system as well as the marginal statistical characteristics. Subsequently, the M-NSOR model is applied to three climate indices (i.e., Arctic Oscillation, El Niño-Southern Oscillation, and Pacific Decadal Oscillation) for the annual and winter datasets. The results of the proposed model are compared to those of the Contemporaneous Shifting Mean and Contemporaneous Autoregressive Moving Average (CSM-CARMA) model. The results indicate that the proposed M-NSOR model is superior to the CSM-CARMA model for reproducing the NSO process, while the other basic statistics are comparatively well preserved in both cases. The current study concludes that the proposed M-NSOR model can be a good alternative to simulate NSO processes and their teleconnections with climate indices.

Keywords: Atlantic Oscillation, Climate Indices, ENSO, Multivariate Simulation, Nonstationary Oscillation, Pacific Decadal Oscillation

1. Introduction

A number of climate indices have been developed to describe the state and the changes of regional climate systems such as the Arctic Oscillation (AO) [Lorenz, 1951; Thompson and Wallace, 1998], the El Niño-Southern Oscillation (ENSO) [Tsonis et al., 2007; Modarres and Ouarda, 2013], and the Pacific Decadal Oscillation (PDO) [Mantua et al., 1997] indices. These indices are related to each other as well as with hydro-meteorological variables in adjacent or remote regions and such relations are called ‘teleconnections’ [Wallace and Gutzler, 1981; Rodionov and Assel, 2000; Alexander et al., 2002; Chiew and McMahon, 2002; Jevrejeva et al., 2003; Rodionov and Assel, 2003; Rogers and Coleman, 2003; Hanson et al., 2004; An et al., 2007; Ouachani et al., 2013; Niranjana Kumar et al., 2016]. For instance, Jevrejeva et al. [2003] examined the ice conditions in the Baltic Sea using climate indices such as North Atlantic Oscillation and AO with wavelet approach and found that the significant phase-linkages between ice conditions and atmospheric circulation exist.

The teleconnections between climate indices and hydro-meteorological variables (e.g., air temperature, precipitation, wind speed, and streamflow) have been reported in the literature [Serreze et al., 1998; Rodionov and Assel, 2003; Ghanbari and Bravo, 2008; Naizghi and Ouarda, 2017]. Rodionov and Assel [2003] found that a substantial difference in the large-scale atmospheric circulation associated with the PDO and ENSO leads to an abnormally mild winter in the Great Lakes region. In another study, Ghanbari and Bravo [2008] presented significant relations between the levels of the Great Lakes and the PDO in interdecadal frequency ranges, as well as these levels and the AO in interannual frequency ranges. Barton and Ramírez [2004] showed that the decadal climate variability in the form of PDO provides a significant impact on seasonal water supply in the Columbia River Basin. Hamlet and Lettenmaier [1999] employed the ENSO and PDO climate signals to predict the

streamflow of the Columbia River Basin and showed that the climate information can be helpful to increase the lead time and forecasting skill. Those two climate indices were also employed to assess their influences on water supply in the Zayandeh-rood River Basin, Iran [Ashouria *et al.*, 2008]. The correlation analysis indicates that the inflow volume has a direct relation while the Southern Oscillation Index presents an inverse relation. Ouarda *et al.* [2014] showed that the sea surface temperature anomalies Associated to ENSO warm phase (El Niño) and cold phase (La Niña) change the weather and precipitation patterns over the Arabian Peninsula.

To analyze the impact of climate variability on a hydrological system (such as the Great Lakes system), it might be useful to simulate climate indices while preserving the teleconnections. This approach would eventually help water resources managers to develop plans and mitigate policies. However, it is known that climate indices commonly contain irregular nonstationary oscillation (NSO) processes, also called nonperiodic and quasi-periodic oscillations [Kwon *et al.*, 2007; Tsonis *et al.*, 2007; Keenlyside *et al.*, 2008]. These NSO processes are the reflection of external and internal forcings such as solar activity [Stevens and North, 1996] and oceanic circulation [Rahmstorf, 2003]. Therefore, a traditional time series model, such as an autoregressive moving average (ARMA), is unsuitable for simulating climate indices due to these irregular and oscillatory behaviors.

To tackle the limitation of traditional time series techniques, a number of stochastic simulation models have been developed with the purpose of modeling the oscillatory behaviors of climate indices [Sveinsson *et al.*, 2003; Ahn and Kim, 2005; Kwon *et al.*, 2007]. Sveinsson *et al.* [2003] considered the long-term variations of a climate index as sudden mean shifts and applied a shifting mean model (SM) originally developed by Salas and Boes [1980]. This model was expanded to contemporaneous SM (CSM) and further mixed with

contemporaneous ARMA (CARMA) for a multivariate case [Sveinsson, 2002; Salas *et al.*, 2009]. The mixed model, called CSM-CARMA, conserves the lag zero correlation between the underlying processes of the two different models.

Ahn and Kim [2005] used an autoregressive conditional heteroscedasticity model to express the nonlinear property of the Southern Oscillation index. *Kwon et al.* [2007] devised an autoregressive type model assisted by wavelet analysis and reproduced a nonstationary process. *Modarres and Ouarda* [2014] modeled the link between climate oscillation indices and drought occurrence using a generalized autoregressive conditional heteroscedasticity (M-GARCH) model.

A number of advanced stochastic models using state-of-the-art statistical techniques such as copula [*Khedun et al.*, 2014], wavelets [*Erkyihun et al.*, 2016], and independent component analysis [*Westra and Sharma*, 2009] have been developed to model the dependence between the oscillations in climate indices and hydrological variables. *Khedun et al.* [2014] developed a copula-based model to analyze the dependence structure between climate indices (ENSO and PDO) and average monthly seasonal precipitation. *Erkyihun et al.* [2016] presented a model generating stochastic streamflow projections conditioned on quasi-oscillatory climate indices (PDO and Atlantic Multi-decadal Oscillation) where the climate indices are simulated with wavelet-based k-nearest neighbor resampling. *Westra and Sharma* [2009] used a multivariate model based on independent component analysis to capture the principal variability of reservoir inflows. These models reproduce hydroclimatological variability well and relate climatic signals to hydrological variables. However, smoothly varying NSO processes in climate indices remain difficult to reproduce and there is a strong need to develop multivariate models that can mimic the NSOs embedded in more than one climate index.

Meanwhile, *Lee and Ouarda* [2012a] proposed a stochastic simulation model to adequately reproduce the smoothly varying NSO processes embedded in observed data. The proposed model employs a cutting-edge decomposition technique, called Empirical Mode Decomposition (EMD) and nonparametric time series models (i.e., k-nearest neighbor resampling (KNNR) and block bootstrapping). The EMD approach has since received increasing interest in the field of hydro-climate modeling [*Lee and Ouarda*, 2012a; *Basha et al.*, 2015; *Durocher et al.*, 2015].

The objective of the present study is to extend the work of *Lee and Ouarda* [2012a] to the multivariate framework in order to simulate climate indices. The primary focus of the proposed model is to preserve the teleconnections between the simulated climate indices and to reproduce both the NSO processes and the key statistics (e.g., mean, variance, and skewness).

The present paper is organized as follows. The background material concerning EMD analysis and the relevant spectral analysis is introduced in section 2. The proposed EMD-based model is presented in section 3. Section 4 presents the results of the simulation study. Section 5 includes a description of the employed dataset. In section 5, the proposed model is applied to annual climate indices. Section 7 discusses the critical points associated to the proposed model. Finally, a summary of the obtained results is presented in section 8 along with the conclusions of the study.

2. Background

2.1. Empirical mode decomposition (EMD)

Empirical mode decomposition (EMD) analysis is a method that extracts oscillation components from a targeted time series and changes them to intrinsic mode functions (IMFs)

to yield physically meaningful explanations of the time series. For an extracted component to be an IMF, it must have symmetry about the local zero mean and the same number of zero crossings and extrema [Huang *et al.*, 1998]. The procedure to obtain IMFs from a time series variable, x , is summarized below:

- (1) Identify all local extrema and connect all local maxima (minima) with a cubic spline as the upper (lower) envelope.
- (2) Obtain the first component, h , by finding the difference between the data (x) and the local mean (m) of the upper and lower envelopes, m , as $h = x - m$.
- (3) Substitute x by h and repeat steps (1) and (2) until a certain criterion is met, such that the number of zero crossings and extrema are equal or at most differ by one and stay the same consecutively with steps (1) and (2) repeated [Huang and Wu, 2008].
- (4) Designate the final h as the j^{th} IMF, c_j , and the residue as r_j (i.e., $r_j = r_{j-1} - c_j$ where $r_0 = x$).
- (5) Repeat steps (1)-(4) by treating the residue r_j as the original data (i.e., x) until the final residue becomes a monotonic function. The final residue (r_n) becomes c_{n+1} .
- (6) Finally, express the original time series at time t , $x(t)$, as the sum of the estimated IMFs:

$$x(t) = \sum_{j=1}^{n+1} c_j(t) \quad (1)$$

where $t=1, \dots, N$ and N denotes the record length. Note that a lower order component has a higher frequency level and vice versa. Additionally, c_{n+1} is not an IMF because it represents the overall trend of the time series $x(t)$ [Huang *et al.*, 1998].

In general, it has been observed that a signal of a similar frequency resides in different IMFs [Huang *et al.*, 1998]. Such a phenomenon is termed mode mixing. To avoid its

occurrence, *Wu and Huang* [2009] developed a procedure labeled *ensemble* EMD (EEMD) by adding white noise in the original data and taking the ensemble means of each IMF. *Wu and Huang* [2004] established a statistical significance test for IMF components derived from white noise. The fundamental function of the significant test is that when the IMF energy of the target data with a certain mean period is higher than the energy of white noise with a certain confidence interval, the corresponding IMF is considered statistically significant at the given level. However, *Lee and Ouarda* [2012a] presented some drawbacks of this significance test including unreliable test results and the inability to assess the signal produced by autocorrelated red noise. Therefore, they recommended selecting the significant components subjectively with the help of the significance test and a visual presentation of IMF components.

2.2. Hilbert transform (Hilbert spectral analysis)

The Hilbert transform [*Hilbert*, 1953] is thought of as the convolution of $x(t)$, a time series data, with the function of $g(t)=1/(\pi t)$. The Hilbert transform of $x(t)$ (i.e., the complex conjugate $\bar{x}(t)$) is presented with the Cauchy principal value (PV) from the non-integrable property of $g(t)$ [*Kanwal*, 1996] as:

$$\begin{aligned}\bar{x}(t) &= PV \int_{-\infty}^{\infty} x(\tau) g(t-\tau) d\tau = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \\ &= -\frac{1}{\pi} \lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^{\infty} \frac{x(t+\tau) - x(t-\tau)}{\tau} d\tau\end{aligned}\quad (2)$$

The analytical signal $z(t)$ derived from $x(t)$ and $\bar{x}(t)$ is represented as:

$$z(t) = x(t) + i\bar{x}(t) = a(t)e^{i\theta(t)}\quad (3)$$

where $a(t) = \sqrt{x^2(t) + \bar{x}^2(t)}$ is the instantaneous amplitude, $\theta(t) = \tan^{-1}(x(t)/\bar{x}(t))$ is the instantaneous phase function, and $i = \sqrt{-1}$. The instantaneous frequency is defined

as $\omega(t) = d\theta(t)/dt$. At a given time t , the instantaneous frequency ω and amplitude a are simultaneously calculated. The frequency-time distribution of the amplitude is designated as the Hilbert amplitude spectrum, $H(\omega, t)$. The marginal Hilbert spectrum is a measure of the total energy contribution from each frequency over the entire data span, expressed as:

$$M(\omega) = \int_0^N H(\omega, t) dt \quad (4)$$

where N is the total data length. This provides a quantitative way to describe the time-frequency-energy. The instantaneous energy (IE) provides information about the time variation of the energy over all frequency levels. It is expressed as:

$$IE(t) = \int_{\omega} H(\omega, t) d\omega \quad (5)$$

3. Model Description

The proposed model for the current study is an extension of the work of *Lee and Ouarda* [2012a] into the multivariate framework, and is denoted as Multivariate Nonstationary Oscillation Resampling (M-NSOR). It simulates the variables of interest (i.e., climate indices) while preserving the teleconnections between the variables and reproducing the NSO processes.

The rationale for the critical points of the original NSOR simulation procedure is as follows: (1) We employ block bootstrapping since the extracted IMF time series from EMD are oscillatory with a certain range for the phase and frequency. These oscillatory characteristics can be simulated by bootstrapping the extracted IMF series from observations. (2) Instead of bootstrapping the extracted IMF series itself, their change rates are bootstrapped to simulate totally new values and to avoid the discontinuity between the blocks. (3) Random block length is employed so that the change time for the phase and frequency is

not fixed and nonstationary oscillations can be properly simulated. (4) The Poisson distribution is applied since it is one of the most flexible discrete distributions and its parameter is estimated with the variance inflation factor employing the lagged autocorrelations of the selected components derived by *Wilks* [1997]. Note that the detailed distribution selection procedure cannot be applicable since a real block number dataset does not exist.

The procedure is briefly described below and schematically represented in Figure 1.

(1) Decompose the time series of each considered variable into a finite number of IMFs.

These IMFs are denoted as $c_j^s(t)$ where $s=1, \dots, S$ and $j = 1, \dots, n+1$, S is the total number of considered variables. Note that n is the number of IMFs and the last component c_{n+1} represents the overall trend.

(2) Select significant IMF components for each variable using the significance test [*Wu and Huang, 2004*] and visual illustration of the components showing significantly large up and down variations compared to the original time series [*Lee and Ouarda, 2010; 2011*].

(3) Fit stochastic time series models according to the nature of the components determined in step (2). In the current study, significant IMF components are modeled using M-NSOR (discussed later) and the residuals are modeled using an order-1 multivariate autoregressive (MAR(1)) model.

(4) Simulate the IMF components using the fitted models (M-NSOR and MAR(1)).

The simulated data are presented as $C_j^s(t)$ where j and s have the same meaning as in step (1).

(5) Sum-up the simulated IMFs from each model and for each variable of interest.

The M-NSOR was developed to model the significant IMFs which illustrate the cross-dependencies of climate indices at different frequency levels. Note that it is these cross-dependencies that are referred to as teleconnections in the introduction section. When only one variable is simulated (i.e., $S=1$), the M-NSOR model becomes equivalent to the NSOR model. In this case, a certain j^{th} IMF is selected in only one variable while the IMF of the other variables is noted as insignificant.

The simulation procedure of M-NSOR for the selected j^{th} IMF component is described below. Note that the subscript j is omitted for the simulated and observed data of the selected IMF components $C_j^s(t)$ and $c_j^s(t)$ during the following description for simplicity (i.e., $C_j^s(t) = C^s(t)$). The following simulation procedure is applied to all selected IMF components that were determined to be significant from step (2) above.

Assume that the current status is at time $t-1$ and its simulated data values ($C_j^s(t-1)$, $s=1, \dots, S$) are known. Now, the IMF component at time t is simulated with the following procedure:

- i. Simulate a block length, L_B , using Poisson distribution formula as given below:

$$L_B \sim \frac{e^{-\tau} \tau^{l-1}}{(l-1)!} \quad l = 1, 2, \dots \quad (6)$$

Note that the parameter (τ) is the mean of L_B . This parameter is selected using a variance inflation factor (VIF) as suggested by *Wilks* [1997] and *Lee and Ouarda* [2011].

- ii. Estimate distances between the observed data and the current status of the simulated data. The distances are given by:

$$D_j = \mathbf{E}(t, j)^T \mathbf{\Phi}^{-1} \mathbf{E}(t, j) \quad (7)$$

and

$$\mathbf{E}(t, j) = \begin{bmatrix} C^1(t-1) - c^1(j) \\ \vdots \\ C^S(t-1) - c^S(j) \\ \Delta C^1(t-1) - \Delta c^1(j) \\ \vdots \\ \Delta C^S(t-1) - \Delta c^S(j) \end{bmatrix} \quad (8)$$

where $j = 2, \dots, (N - L_B)$ and T implies the transpose of a vector. Additionally, $\Phi = \text{diag}(\sigma_c^2(1), \dots, \sigma_c^2(S), \sigma_{\Delta c}^2(1), \dots, \sigma_{\Delta c}^2(S))$ where $\sigma_c^2(s)$ and $\sigma_{\Delta c}^2(s)$ are the variances of the s^{th} component (c^s) and of the change rate of the s^{th} component (Δc^s), respectively.

Δc^s and ΔC^s are change rates of the observed and simulated data, respectively, of the s^{th} component such that $\Delta c(t) = c(t) - c(t-1)$.

iii. Arrange the distances in ascending order, select the first k distances, and store their time indices.

iv. Randomly select one of the stored k time indices with the weighting probability given by:

$$w_i = \frac{1/i}{\sum_{j=1}^k 1/j}, \quad i = 1, \dots, k$$

v. Assume the selected time index to be p , and take the change rates of the L_B successors (i.e., $\{\Delta c^s(p+1), \dots, \Delta c^s(p+L_B)\}$ for $s = 1, \dots, S$).

vi. Obtain the simulated data with length L_B using:

$$C^s(t+l-1) = C^s(t+l-2) + \Delta c^s(p+l) \quad (9)$$

for $l = 1, \dots, L_B$ and $s = 1, \dots, S$ where $C^s(0)$ is obtained by bootstrapping one of the observed data.

vii. Repeat steps i.-vi. until the required data are simulated.

A simple example of the proposed M-NSOR is presented in the Supplementary Material.

4. Simulation Study

4.1. Application Methodology

In order to validate the proposed M-NSOR model, one of the most famous and chaotic nonlinear dynamic systems, the Rössler attractor is used [Rössler, 1976; 1995]. The attractor is designed to behave similarly to the Lorenz attractor [Lorenz, 1963] but with a better understanding of the chaotic flow. Similar applications have been made for the univariate framework of the NSOR model [Lee and Ouarda, 2012a].

The attractor is represented by three nonlinear ordinary differential equations $\dot{x} = -y - z$, $\dot{y} = x + \alpha y$, and $\dot{z} = \beta + xz - \delta z$. Here, $(x, y, z) \in \mathfrak{R}^3$ and $(\alpha, \beta, \delta) \in \mathfrak{R}^3$ are dynamical variables and parameters respectively and \dot{x} represents the derivative of the variable x . In the current study, the parameter set $[\alpha, \beta, \delta] = [1/5, 1/5, 7/2]$ is applied as in Huang *et al.* [1998] and Kijewski-Correa and Kareem [2007], $N=500$ and with the initial state of $[x_0, y_0, z_0] = [-3, 3, 1]$.

Lee and Ouarda [2012a] selected the x variable of the Rössler attractor since it oscillates within a limited range, but the oscillations are chaotic such that the general assumption is met with the NSOR model. In the current study, we apply the x and y variables to test the multivariate version of the NSOR model (M-NSOR) since both x and y variables present similar oscillation behaviors as shown in the top panel of Figure 2 (thin line) with limited range and frequency. The z variable is not considered in the current simulation due to its totally different oscillatory behavior and low coherence with the other two variables. In addition, we added the random component for each variable as $x+\varepsilon$ and $y+\varepsilon$ where $\varepsilon \sim N(0, \sigma_\varepsilon)$ and $\sigma_\varepsilon=1/2$ is employed after testing a number of values to include its oscillatory behavior and randomness.

4.2. Results

For M-NSOR modeling, the Rössler series in the top panel of Figure 2 (thin line) was decomposed with EMD as IMFs. The IMFs of the X variable are shown in Figure 3. The oscillations in a lower frequency are captured in a lower number and vice versa. The overall long-term trend is presented in the last panel of Figure 3. Note that the scale of the y coordinate for each component is different. For example, even if the oscillatory behavior is clear in c_6 and c_7 , these magnitudes range only from -1 to 1 implying that the significance of these IMFs is not always strong enough.

The result of the significance test [Wu and Huang, 2004] for the IMFs of the Rössler X variable is shown in Figure 4 and indicates that the three IMFs c_4 , c_5 , and c_8 can be considered as real oscillatory components and are distinguishable from random noise. However, this significance test is not always reliable due to its documented drawbacks such as the no consideration for serially correlated series (i.e. white noise) [Wu and Huang, 2004]. Therefore, subjective judgement might be used with the additional help of the visual presentation of IMF components [Lee and Ouarda, 2012a].

For example, the result of the significance test is shown in Figure S1 of the supplementary material for the IMFs of the Rössler Y variable in Figure S2. The results show that c_6 and c_7 as well as c_4 , c_5 , and c_8 are significant even with the similar oscillation range to the X variable (Figure S2). In order to coincide with the X variable and the relatively small magnitude of the normalized energy in Figure S1, the sum of c_4 , c_5 , and c_8 is modeled with the M-NSOR for both the X and the Y variables. The summed components (i.e. $c_4 + c_5 + c_8$) are overlaid on the simulated Rössler series shown in the top panel of Figure 2 with thick solid (X variable) and dotted (Y variable) lines. The summed series illustrates well the oscillatory behavior of the original series.

Here, the summed series of the X and Y variables is modeled with the M-NSOR model and its residual is modeled with the MAR (1) model. From the model, one hundred series with the same record length of the original Rössler series ($N=500$) are generated. One example of the generated series is presented in the bottom panel of Figure 2. The figure indicates that the oscillation and randomness in the original series (top panel) is well reproduced in the generated series.

The statistical characteristics of the marginal distribution of the generated series are tested with the kernel density estimate [Salas and Lee, 2010] shown in Figure 5. The generated kernel density estimates (thin grey dotted lines) follow well the estimate of the original series (thick solid line). This indicates that the generated series from the M-NSOR model reproduces well the statistical behavior of the original series shown in the top panel of Figure 2.

The frequency domain characteristics of the M-NSOR generated data are tested with wavelet analysis. In Figure 6, global wavelet spectra of the 100 generated series from the M-NSOR model (dotted lines) are presented as well as the original Rössler series for X (top panel) and Y (bottom panel). The simulated Rössler series for both the X and Y variables have a strong spectral power during the period between $2^6=64$ and $2^7=128$ shown with a thick solid line in Figure 6, and in the top panel of Figure 2 for the original domain. This oscillatory characteristic is reproduced well with the generated series from the M-NSOR model as indicated in the thin dotted lines of Figure 6.

The cross coherence in the frequency domain for the generated series from the M-NSOR model as well as the Rössler series are presented in the bottom and top panels of Figure 7, respectively. The high coherence of the Rössler series during the period between $2^6=64$ and $2^7=128$ is reproduced well in the M-NSOR generated series. Overall, the results

indicate that the oscillation characteristics of the Rössler system are represented well by the M-NSOR model and reproduce the marginal statistics appropriately.

5. Data Description of case study

Three climate indices (AO, PDO, and ENSO) are selected for the case study to evaluate the proposed M-NSOR model. These indices are known to be teleconnected to each other and with hydro-climatological variables such as in the Great Lakes system [*The International Upper Great Lakes Study Board* 2012]. Therefore, these indices will be employed to simulate the net basin supply (NBS) of the Great Lakes system. A brief description of each of these climate indices is given in the following paragraphs.

The AO refers to opposing atmospheric pressure patterns in northern middle latitudes (approximately 45 degrees North) and high latitudes (polar regions) ranging from weeks to decades [*Thompson and Wallace*, 1998]. Among the various AO indices, the zonal averaged version produced by *Li and Wang* [2003] was employed in this study because it explains a large portion of the variance of the sea level pressure over the Atlantic region for all time scales such as monthly, seasonal, and annual and provides the strongest correlation pattern with surface air temperature [*Li and Wang*, 2003]. The dataset ranging from 1873 to 2009 was downloaded from <http://www.lasg.ac.cn/staff/ljp/data-NAM-SAM-NAO/NAM-AO.htm>. This record length was set since the current model development was made to analyze the hydrological changes of the Great Lakes system and the outputs of the current work must be compatible with the previous study [*Ehsanzadeh et al.*, 2013].

The PDO index represents the leading principal component of sea-surface temperature anomalies in the North Pacific Ocean, polewards of 20°N. Among a number of PDO indices, the most commonly used one, developed by *Mantua and Zhang* and their colleagues [*Mantua*

et al., 1997; Zhang *et al.*, 1997], was employed in the current study with the dataset ranging from 1900 to 2009. It was downloaded from <http://jisao.washington.edu/pdo/PDO.latest>.

The ENSO is a climatic pattern occurring across the tropical Pacific Ocean, causing climate variability in 3~7 year periods [Alexander *et al.*, 2002]. Among the various ENSO indices [Trenberth, 1997], the ENSO index, developed by Wolter and Timlin [1993], was employed in the current study because it includes the fundamental tropical atmospheric bridges. The dataset, ranging between 1950-2009, was downloaded from <http://www.esrl.noaa.gov/psd/people/klaus.wolter/MEI/>.

The annual (average of the monthly values from January to December) and winter season (average of January, February, and March; JFM) values of these climate indices were used in the current study. Because the annual and winter dataset might have different characteristics, separate models were used in each case and corresponding results are reported in sections 6 and supplementary material.

6. Annual Climate Indices

6.1. Application Methodology

The observed datasets of the annual climate indices are presented in the top panel of Figure 8. All climate indices shared long-term oscillation [Lee *et al.*, 2013], though the long-term pattern of the ENSO index is not obvious due to its relatively short record length. An EEMD on the PDO and AO indices was performed to investigate both the long-term and short-term oscillations. Note that, in the current study, we employed EEMD instead of EMD to avoid mode mixing as discussed in the background section. The decomposed IMFs from the AO and PDO indices are presented in Figure 9 and Figure 10, respectively. As presumed, the 5th IMF components of the two indices (c_5 in Figure 9 and Figure 10) varied in a similar fashion.

This behavior of sharing the same long-term oscillation between PDO and other climate indices such as the Southern Oscillation Index and Cold Tongue Index was previously reported by *Mantua et al.* [1997] and further analyzed by *Lee et al.* [2013].

The cross-correlations between the annual AO and PDO indices with each IMF component, as well as the observed data, are presented in the second row of Table 1. It is evident that there is no significant correlation with the observed data (-0.012). However, the correlations of each IMF component behaved differently. The 4th and 5th IMF components (c_4 and c_5), capturing the low-frequency oscillation, are significantly positively correlated, while the other components (c_1 , c_2 , and c_3), capturing high frequency oscillations, showed negative correlation.

Individual modeling of these indices does not take into account the teleconnection between them. Therefore, to preserve the cross-correlation at different frequency levels, the M-NSOR model for low-frequency components and the Multivariate AutoRegressive (1) model (MAR(1)) for high-frequency components was applied in the following way.

- (1) Assuming that AO and PDO share the same long-term NSO, the M-NSOR model was applied to the sum of the 4th and 5th IMF components of the two indices.
- (2) The MAR(1) model was fit to the residuals of the AO and PDO indices ($c_1 + c_2 + c_3 + c_6$). Note that the overall trend component, c_6 , was included in the MAR(1) model to be used as the overall mean, as in *Lee and Ouarda* [2012a].

To simulate the ENSO index, a simple linear regression model was fitted to the ENSO index, treating the PDO index as predictor, due to the following reasons: (i) the ENSO and PDO are highly correlated (as much as 0.73) as shown in Figure 11; (ii) no significant nonlinearity was observed between the ENSO and PDO; and (iii) a similar EEMD analysis and modeling, as in the case of AO and PDO indices, is not applicable for the ENSO index due to its relatively short record length. Note that graphical assessment with scatterplot and nonlinear

measurement based on mutual information [Khan *et al.*, 2006] indicate that not much significant nonlinear relationship exists between ENSO and PDO (result not shown).

To compare the proposed model, the existing CSM-CARMA model was applied. The CSM model was applied for the AO and PDO indices and the CARMA model was applied for the ENSO index as it was not possible to fit the CSM model to the ENSO index. The model was implemented through the software package Stochastic Analysis Modeling and Simulation (SAMS-2009) downloaded from <http://www.sams.colostate.edu/>. The brief description of the SM model is shown in the Appendix. Further details regarding the CARMA and SM models can be obtained in Sveinsson *et al.* [2003].

6.2. Results

To evaluate the performance of the proposed model, we simulated a dataset containing the three climate indices each of which had a length of 50,000. Then, the dataset was divided into 500 sequences, each having a record length of 100, to obtain a comparable record length to the observed data. Note that the record lengths of the observed AO and PDO indices are 137 and 110, respectively, and the length of ENSO is 60. This long dataset was divided into 500 sequences as $[\mathbf{x}_1, \dots, \mathbf{x}_{100}]$, $[\mathbf{x}_{101}, \dots, \mathbf{x}_{200}]$, \dots , $[\mathbf{x}_{49901}, \dots, \mathbf{x}_{50000}]$ where $[\mathbf{x}_1, \dots, \mathbf{x}_{100}] = [x^1_1, \dots, x^1_{100}; x^2_1, \dots, x^2_{100}; x^3_1, \dots, x^3_{100}]$. Note that the superscript indicates the number of each climate variable as ENSO, PDO, and AO, respectively. The separated sequence contains the temporal dependency from M-NSOR and MAR(1) modeling (see the model description in section 3) as well as the spatial dependency due to the multivariate framework of the models. One of the divided 500 sequences is illustrated in Figure 8 for the M-NSOR model (middle panel) and the CSM-CARMA model (bottom panel).

Figure 8 indicates that climate indices simulated from the M-NSOR model (middle panel) show a long-term oscillation pattern similar to that of the observed climate indices (top

panel). In contrast, each climate index simulated from the CSM-CARMA model (bottom panel) shows a different oscillation pattern. Comparing the two plots (middle and bottom panels) shows that the long-term NSO process of the climate indices is well preserved by the M-NSOR model compared with the CSM-CARMA model. This is because the latter treats the long-term pattern as the shifting mean process which consequently leads to the preservation of only the lag zero cross-correlation.

The scatter plot between the observed climate indices is shown in Figure 11 and for the simulated data of the M-NSOR model and the CSM-CARMA model in Figure 12 and Figure 13, respectively. Also, the bottom row of Table 2 presents the correlation between the climate indices. Note that the correlation at the historic AO column (the second column, 0.17) indicates the correlation between AO and ENSO; the ENSO column (third column, 0.17) shows the correlation between ENSO and AO; the PDO column (the fourth column, 0.73) does the correlation between PDO and ENSO. The same presentation is applied to the model outputs.

Not much strong dependency is present between the AO and PDO climate indices (the correlation is 0.17 as shown in Table 2) due to the offset of a positive correlation in the low frequency level and a negative correlation in the high frequency level as mentioned earlier (see Table 1). The same implication can be drawn regarding the relation between AO and ENSO in Figure 11 (the correlation is 0.17 as shown in Table 2). In contrast, ENSO and PDO have a significant linear correlation and its correlation is 0.73 (as shown in Table 2). The simulated data from the M-NSOR and CSM-CARMA models reproduce well these dependencies as shown in Figure 12 and Figure 13, respectively and also in Table 2. A stronger correlation between the ENSO and PDO indices is observed in the simulated data of the M-NSOR model (0.7 as shown in Table 2) than for that of the CSM-CARMA model (0.58 as shown in Table 2).

The key statistics of the simulated data from the M-NSOR model and the CSM-CARMA model were estimated using the divided 500 sequences and are presented in boxplots in Figure 14 and Figure 15, respectively and also in Table 2. For the boxplot, the end lines of the box are the 25th and 75th percentiles; the cross lines above the box on the whisker correspond to the 90th percentile and the maximum; and the cross lines below the box on the whisker represent the 10th percentile and the minimum. The circle represents the statistics of the observed data.

A slight overestimation of the skewness of the AO is observed with both models. The observed skewness is -0.22 and the skewness of the simulated data from both models is almost zero as shown in the fifth row of Table 2. The lag-1 correlation of the PDO index is underestimated by the CSM-CARMA model. Cross-correlations between climate indices are estimated for the M-NSOR and CSM-CARMA models as shown in Figure 16 and Table 2. The cross-correlation between AO and ENSO for both models is slightly underestimated, but for ENSO-PDO the underestimation is observed only in the case of the CSM-CARMA model. This correlation of the observed data is 0.73 and the ones of the simulated data from M-NSOR and CSM-CARMA are 0.7 (0.06) and 0.58(0.09) respectively as shown in the bottom row of Table 2. Note that the bracket values indicate the standard deviation of the estimated statistics from the 500 simulated series. In general, the results illustrate that both models preserve fairly well the observed key statistics and cross-correlations. Also, the marginal distributions of the observed climate indices are well reproduced with both models, M-NSOR and CSM-CARMA as indicated from the kernel density estimates shown in Figure S3 and Figure S4 of the supplementary material, respectively.

The long-term lagged correlations (LLC) were estimated up to lag-80 and presented in Figure 17. As discussed, the lag-1 correlation of the simulated PDO data with the CSM-CARMA model (0.26, see the eighth row of Table 2) underestimates the one of the observed

data (0.53) while the M-NSOR model (0.52) well reproduces these observed statistics. Furthermore, a significant difference can be seen between the LLC of the simulated data from the M-NSOR and CSM-CARMA. The LLC for the M-NSOR model presents a long-term oscillation similar to the observed data especially for the AO and PDO data (See the top and bottom of the left panels in Figure 17, respectively). Since the ENSO data contains the shorter record than the other climate indices, this behavior of the LLC is rather vague. The observed LLC of the PDO is well reproduced at the M-NSOR model while the LLC of the AO is slightly offset after the 40th lag. This reproduction of the long-term oscillation is the major advantage for the NSOR modeling as indicated by *Lee and Ouarda* [2012a]. This advantage is inherited to the multivariate version of the NSOR model (M-NSOR). We further discussed why this reproduction of the long-term oscillation is critical in the discussion section of the current study.

Meanwhile, the CSM-ARMA model cannot reproduce the long-term oscillation behavior of the observed data. This is not surprising because the original SM model [*Sveinsson et al.*, 2003] has the same lagged correlation structure of the simple autoregressive moving average model (i.e. ARMA(1,1)) in that the lagged correlation is decreased exponentially along with lags. This correlation structure of the SM model is inherited to the multivariate version of the SM model (CSM). This long-term LLC behavior of the CSM model is well presented in the right panels of Figure 17. Therefore, it is obvious that the CSM model cannot reproduce the long-term oscillation of the observed climate indices while the M-NSOR model has this capability.

The Hilbert marginal spectrum (see Eq.(4)) was estimated for the PDO index as shown in Figure 18 as well as for the AO and ENSO indices in Figures S5 and S6 of the supplementary material, respectively with a thick dark line for the spectrum of observed data and a thin gray line for the spectrum of the simulated data. Note that the intermittency of the

observed Hilbert marginal spectrum, particularly in the ENSO index (see, thick line in Figure S6), exists because the Hilbert transform is evaluated from the IMF components whose estimates may not exist for all frequencies. The observational spectra of the PDO index (Figure 18) as well as the AO index in Figure S5 illustrate high amplitudes in low frequencies (1×10^{-2} - 9×10^{-1}). The figures indicate that the M-NSOR model reproduces the high observational Hilbert spectrum in low frequencies better than the CSM-CARMA model. Because the record length of the ENSO index is short, the relevant low frequency spectrum of the ENSO indices cannot be observed (Figure S6). However, if the ENSO is consistently correlated with the PDO, as observed for the period of 1900-1950, a similar observation, as in the cases of AO and PDO, is possible.

The global wavelet spectra (GWP) of the three climate indices were estimated and shown in Figure 19 and Figure 20 for the proposed M-NSOR model and the CSM-CARMA model, respectively. The observed global wavelet spectra indicate that there are high amplitudes in relatively large periods around 2^5 and 2^6 . These high amplitudes in the large periods are well reproduced from the M-NSOR model as in Figure 19 while the CSM-CARMA model cannot reproduce these high amplitudes of the AO and PDO (see Figure 20). This result of the global wavelet spectra is well matched with the result of the Hilbert spectra as mentioned earlier (Figure 18).

The peaks of the GWP are located around 2^5 - 2^7 with a slight deviation for each climate index as shown in Figure 19. The proposed M-NSOR model reproduces the overall GWP peak period but not each specific detail since the long-term oscillation components are simulated together with the summed up long-term EMD components. A separate NSOR modeling can be carried out to preserve the individual peak. However, its interconnection for the long-term oscillations cannot be preserved in individual modeling.

The short-term GWP in ENSO around 5-6 years was not reproduced by both models since the current study focuses on the long-term oscillation. More elaborate modeling can be carried out to reproduce this oscillation. Also, the GWP ENSO peak in the period of $2^6 = 64$ year is regarded as a monotonic trend since the record length of ENSO is only 60 years.

7. Discussion

The simulated climate indices were used to assess the future aspects of hydro-meteorological variables by interconnecting these indices to hydro-meteorological variables such as in the Great Lakes system. The long-term behavior of hydro-meteorological variables (e.g. precipitation and streamflow) is important for the design of hydrologic structures that are intended to mitigate hydro-meteorological extremes (droughts and floods) . For example, when a future projection of hydro-meteorological variables by Global Climate Models (GCMs) predicts an increase of hydro-meteorological variables and subsequently higher chance of floods, hydrologic structures must be redesigned to adapt to this increase. In general, GCM predictions provide information concerning a monotonic increase or decrease. In contrast, the assumption we make in the current study is that the climate system that affects the Great Lakes region evolves in the long-term in an oscillatory mode rather than a monotonic increase or decrease. The main capability of the stochastic simulation model developed in the current study is to mimic this long-term oscillation in the observed climate indices.

The results of the LLC (long-term lagged correlation) for the annual climate indices in Figure 17 indicates that the M-NSOR model reproduces the observed LLC behavior especially for the AO and PDO while the CSM-CARMA model has an ARMA(1,1) structure with an exponentially decreasing magnitude along with the lags. Furthermore, the global

wavelet spectra of the annual AO and PDO indices shown in Figure 19, indicate that the M-NSOR model reproduces the low frequency behavior of the observed data better than the CSM-CARMA model shown in Figure 20. Frequency is the reciprocal of the duration of one cycle (i.e. $\text{frequency} = \text{one cycle} / \text{duration}$). The low frequency behavior reflects the long-term oscillation evolution. In other words, a better preservation of the low frequency behavior of the observed data implies a better performance of the long-term oscillation as in the M-NSOR model. Finally, the good reproduction of the long-term oscillation of the observed data by the M-NSOR model indicates that the M-NSOR model and its simulated data can be used as a viable alternative for the future assessment of the hydro-meteorological variables in the Great Lakes systems compared to the monotonic increase (or decrease) predicted by GCM outputs.

8. Summary and Conclusions

Climate indices possess long-term oscillations and are interconnected to each other and to hydro-meteorological variables in adjacent or remote regions. In the current study, we developed a simulation model based on the EMD and NSOR techniques in a multivariate framework (called M-NSOR) to reproduce the long-term oscillation patterns and their interconnections.

The proposed model was tested using the nonlinear chaotic Rössler system. The results indicate that the proposed M-NSOR model reproduces well the oscillatory behavior of the system and its marginal statistical characteristics. The proposed M-NSOR model is also applied to the three climate indices and compared to the existing nonstationary simulation model (CSM-CARMA) using spectral analysis, key statistics, and cross-correlation. The performance of the M-NSOR model is comparable to the CSM-CARMA model for key-statistics and cross-correlation. The spectral analysis (i.e., Hilbert marginal spectrum and wavelet analysis) and the LLC result reveals that the M-NSOR model reproduces well the

long-term oscillation embedded in the climate indices whereas the CSM-CARMA model does not have this capability. The overall results demonstrate that the M-NSOR model reproduces better the observed long-term oscillation pattern.

The simulated data can be further employed to study the variability of hydrological variables such as the NBS of the Great Lakes system [*The International Upper Great Lakes Study Board* 2012] or as input to stochastic optimal management models, such as *Ouarda and Labadie* [2001] and *Sreekanth and Datta* [2014]. The study of hydrological variables employing the simulated climate indices as covariables is crucial for assessing the future evolution of these variables and for a better understanding of climate impacts on water resources.

The proposed model can also be applied to more complex hydrological systems such as the 20 streamflow stations of the Upper Colorado River Basin. A strategical modeling scheme is required to apply the developed M-NSOR model to such a complex hydrological system. We also want to note that the proposed scheme of the separate modeling over different combinations can be potentially problematic when a highly nonlinear system is combined in target variables and care must be taken in such cases. Further research is required to build an approach that simulates the multivariate NSO series without resorting to separate modeling over different combinations.

Appendix: Description of the shifting mean model (SM)

The SM model, developed by *Salas and Boes* [1980], was devised to model an abrupt shift process. *Sveinsson et al.* [2003] further improved this model and applied it to hydro-meteorological datasets. Note that the physical background of the SM process is a sudden systematic shift from volcanic activities or regulatory (human) effects rather than random processes.

The basic formula of the SM is presented as follows for a sequence of a random variable, $x(t)$. A general definition of the SM model is given by:

$$x(t) = u(t) + z(t)$$

(A.1)

where $u(t)$ is a sequence of independent identically distributed (iid) variables with mean μ_U and variance σ_u^2 . $z(t)$ is a sequence with mean zero and variance σ_z^2 and is expressed as follows:

$$z(t) = \sum_{i=1}^t m_i I_{(s_{i-1}, s_i]}(t)$$

(A.2)

where m_i is a real-valued random variable with a zero mean and variance σ_m^2 representing the different mean levels over the time span. $S_i = N_1 + N_2 + \dots + N_i$ with $S_0 = 0$ and N_i is the span at a certain noise level, i , and a positive geometric random variable with its parameter p as $E(N) = 1/p$. $I_{(a,b]}(t)$ is the indicator function which takes the value 1 if $t \in (a, b]$, otherwise it takes the value 0. The sequences m_i , N_i , and $x(t)$ are assumed to be mutually independent.

The correlation structure is expressed as follows:

$$\rho_X(\square) = \frac{\sigma_m^2}{\sigma_u^2 + \sigma_m^2} (1 - p)^\square, h=1,2,\dots$$

(A.3)

The parameters are estimated in terms of $\hat{\mu}_X, \hat{\sigma}_X^2, \hat{\rho}_X(\square)$ for $h=1, 2$, where the hat indicates the method of moments estimates of the corresponding parameters. These estimates are [Salas, 1993]:

$$\hat{p} = 1 - \frac{\hat{\rho}_X(2)}{\hat{\rho}_X(1)}$$

(A.4)

$$\hat{\sigma}_M^2 = \hat{\sigma}_X^2 \frac{\hat{\rho}_X(1)}{\hat{\rho}_X(2)}$$

(A.5)

$$\hat{\mu}_U = \hat{\mu}_X$$

(A.6)

$$\hat{\sigma}_U^2 = \hat{\sigma}_X^2 - \hat{\sigma}_M^2$$

(A.7)

Acknowledgements

The authors are grateful to the International Joint Commission for the Management of the Great Lakes for its financial support. The first author also acknowledges that this work was partly supported by a National Research Foundation of Korea (NRF) grant (NRF-2018R1A2B6001799) funded by the Korean Government (MEST). All the produced data from the current study is available in the Mendeley website as <https://data.mendeley.com/datasets/rjdyw7vhh2/1>. The authors wish also to thank the editor, Dr. Martyn Clark, and as well as the associate editor and three anonymous reviewers whose comments helped to improve considerably the quality of the manuscript.

Notations:

x : time series variable

h : candidate of an IMF component

c_j : j^{th} component of IMF

m : local mean of the upper and lower envelopes

r_j : residual of the j^{th} IMF

$H(\omega, t)$: Hilbert amplitude spectrum at frequency ω and time t

$M(\omega)$: Marginal Hilbert Spectrum

$IE(t)$: Instantaneous Energy

c_j^s and C_j^s : j^{th} component of IMF for the s^{th} variable and its random variable, respectively

L_B : block length generated from Poisson distribution with the parameter τ

$\Xi(t, j)$: vector whose elements are the distances between the current status at time t and j^{th} observed value of IMF component for all the considered S variables as well as their change rates

Φ : Diagonal matrix whose diagonal elements are the variances of IMF components for all the S variables as well as their change rates

w_i : weighting probability

AO: Arctic Oscillation

ARMA: AutoRegressive Moving Average

CSM: Contemporaneous SM

EEMD: Ensemble EMD

EMD: Empirical Mode Decomposition

ENSO: El Niño-Southern Oscillation

IMF: Intrinsic Mode Function

KNNR: k nearest neighbor resampling

MAR(1): lag-1 Multivariate AutoRegressive model

M-NSOR: Multivariate NSOR

NSOR: Nonstationary Oscillation Resampling

PDO: Pacific Decadal Oscillation

SM: Shifting Mean Model

VIF: Variance Inflation Factor

References

- Ahn, J. H., and H. S. Kim (2005), Nonlinear modeling of El Nino/southern oscillation index, *Journal of Hydrologic Engineering*, 10(1), 8-15.
- Alexander, M. A., I. Blade, M. Newman, J. R. Lanzante, N. C. Lau, and J. D. Scott (2002), The atmospheric bridge: The influence of ENSO teleconnections on air-sea interaction over the global oceans, *Journal of Climate*, 15(16), 2205-2231.
- An, S. I., J. S. Kug, A. Timmermann, I. S. Kang, and O. Timm (2007), The influence of ENSO on the generation of decadal variability in the North Pacific, *Journal of Climate*, 20(4), 667-680.
- Ashouria, H., A. Abrishamchib, H. Moradkhanic, and M. Tajrishyd (2008), Assessment of Interannual and Interdecadal Climate Variability Effects on Water Supply in Zayandeh-rood River Basin, Iran, in *In The first International Conference on Water Resources and Climate Change in the MENA Region*, edited, Muscat, Oman.
- Barton, S. B., and J. A. Ramirez (2004), Effects of El Niño Southern Oscillation and Pacific Interdecadal Oscillation on water supply in the Columbia River Basin, *Journal of Water Resources Planning and Management*, 130(4), 281-289.
- Basha, G., T. B. M. J. Ouarda, and P. R. Marpu (2015), Long-term projections of temperature, precipitation and soil moisture using non-stationary oscillation processes over the UAE region, *International Journal of Climatology*, 35(15), 4606-4618.
- Chiew, F. H. S., and T. A. McMahon (2002), Global ENSO-streamflow teleconnection, streamflow forecasting and interannual variability, *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 47(3), 505-522.

Durocher, M., T. S. Lee, T. B. M. J. Ouarda, and F. Chebana (2015), Hybrid signal detection approach for hydro-meteorological variables combining EMD and cross-wavelet analysis, *International Journal of Climatology*.

Ehsanzadeh, E., H. M. Saley, T. B. M. J. Ouarda, D. H. Burn, A. Pietroniro, O. Seidou, C. Charron, and D. Lee (2013), Analysis of changes in the Great Lakes hydro-climatic variables, *Journal of Great Lakes Research*, 39(3), 383-394.

Erkyihun, S. T., B. Rajagopalan, E. Zagona, U. Lall, and K. Nowak (2016), Wavelet-based time series bootstrap model for multidecadal streamflow simulation using climate indicators, *Water Resources Research*, 52(5), 4061-4077.

Ghanbari, R. N., and H. R. Bravo (2008), Coherence between atmospheric teleconnections, Great Lakes water levels, and regional climate, *Advances in Water Resources*, 31(10), 1284-1298.

Grinsted, A., J. C. Moore, and S. Jevrejeva (2004), Application of the cross wavelet transform and wavelet coherence to geophysical times series, *Nonlinear Processes in Geophysics*, 11(5-6), 561-566.

Hamlet, A. F., and D. P. Lettenmaier (1999), Columbia river streamflow forecasting based on ENSO and PDO climate signals, *Journal of Water Resources Planning and Management*, 125(6), 333-341.

Hanson, R. T., M. W. Newhouse, and M. D. Dettinger (2004), A methodology to assess relations between climatic variability and variations in hydrologic time series in the southwestern United States, *Journal of Hydrology*, 287(1-4), 252-269.

Hilbert, D. (1953), *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*, 282 pp., Chelsea Pub. Co., New York.

Huang, N. E., Z. Shen, S. R. Long, M. L. C. Wu, H. H. Shih, Q. N. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu (1998), The empirical mode decomposition and the Hilbert

spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences*, 454(1971), 903-995.

Huang, N. E., and Z. H. Wu (2008), A review on Hilbert-Huang transform: method and its applications to geophysical studies, *Reviews of Geophysics*, 46(2), RG2006.

Jevrejeva, S., J. C. Moore, and A. Grinsted (2003), Influence of the Arctic Oscillation and El Niño-Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach, *Journal of Geophysical Research D: Atmospheres*, 108(21), ACL 10-11 - ACL 10-11.

Kanwal, R. P. (1996), *Linear Integral Equations*, 332 pp., Birkhäuser Boston.

Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornbluh, and E. Roeckner (2008), Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, 453(7191), 84-88.

Khan, S., A. R. Ganguly, S. Bandyopadhyay, S. Saigal, D. J. Erickson Iii, V. Protopopescu, and G. Ostrouchov (2006), Nonlinear statistics reveals stronger tie between ENSO and the tropical hydrological cycle, *Geophysical Research Letters*, 33(24).

Khedun, C. P., A. K. Mishra, V. P. Singh, and J. R. Giardino (2014), A copula-based precipitation forecasting model: Investigating the interdecadal modulation of ENSO's impacts on monthly precipitation, *Water Resources Research*, 50(1), 580-600.

Kwon, H. H., U. Lall, and A. F. Khalil (2007), Stochastic simulation model for nonstationary time series using an autoregressive wavelet decomposition: Applications to rainfall and temperature, *Water Resources Research*, 43(5), W05407 05401-05415.

Lee, T., and T. B. M. J. Ouarda (2010), Long-term prediction of precipitation and hydrologic extremes with nonstationary oscillation processes, *Journal of Geophysical Research D: Atmospheres*, 115(13).

Lee, T., and T. B. M. J. Ouarda (2011), Prediction of climate nonstationary oscillation processes with empirical mode decomposition, *Journal of Geophysical Research D: Atmospheres*, 116(6).

Lee, T., and T. B. M. J. Ouarda (2012a), Stochastic simulation of nonstationary oscillation hydroclimatic processes using empirical mode decomposition, *Water Resources Research*, 48(2).

Lee, T., and T. B. M. J. Ouarda (2012b), An EMD and PCA hybrid approach for separating noise from signal, and signal in climate change detection, *International Journal of Climatology*, 32(4), 624-634.

Lee, T., T. B. M. J. Ouarda, and J. Li (2013), An orchestrated climate song from the Pacific and Atlantic Oceans and its implication on climatological processes, *International Journal of Climatology*, 33(4), 1015-1020.

Li, J. P., and J. L. X. L. Wang (2003), A modified zonal index and its physical sense, *Geophysical Research Letters*, 30(12), 34.31- 34.34.

Lorenz, E. N. (1951), Seasonal and Irregular Variations of the Northern Hemisphere Sea-level Pressure Profile, *Journal of Meteorology*, 8(1), 52-59.

Lorenz, E. N. (1963), Deterministic nonperiodic flow, *Journal of Atmospheric Science*, 20, 130-141.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis (1997), A Pacific interdecadal climate oscillation with impacts on salmon production, *Bulletin of the American Meteorological Society*, 78(6), 1069-1079.

Modarres, R., and T. B. M. J. Ouarda (2013), Testing and modelling the volatility change in ENSO, *Atmosphere - Ocean*, 51(5), 561-570.

Modarres, R., and T. B. M. J. Ouarda (2014), A generalized conditional heteroscedastic model for temperature downscaling, *Climate Dynamics*, 43(9-10), 2629-2649.

Naizghi, M. S., and T. B. M. J. Ouarda (2017), Teleconnections and analysis of long-term wind speed variability in the UAE, *International Journal of Climatology*, 37(1), 230-248.

Niranjan Kumar, K., T. B. M. J. Ouarda, S. Sandeep, and R. S. Ajayamohan (2016), Wintertime precipitation variability over the Arabian Peninsula and its relationship with ENSO in the CAM4 simulations, *Climate Dynamics*, 47(7-8), 2443-2454.

Ouachani, R., Z. Bargaoui, and T. Ouarda (2013), Power of teleconnection patterns on precipitation and streamflow variability of upper Medjerda Basin, *International Journal of Climatology*, 33(1), 58-76.

Ouarda, T. B. M. J., and J. W. Labadie (2001), Chance-constrained optimal control for multireservoir system optimization and risk analysis, *Stochastic Environmental Research and Risk Assessment*, 15(3), 185-204.

Ouarda, T. B. M. J., C. Charron, K. Niranjan Kumar, P. R. Marpu, H. Ghedira, A. Molini, and I. Khayal (2014), Evolution of the rainfall regime in the united arab emirates, *Journal of Hydrology*, 514, 258-270.

Rahmstorf, S. (2003), Thermohaline circulation: The current climate, *Nature*, 421(6924), 699-699.

Rodionov, S., and R. Assel (2000), Atmospheric teleconnection patterns and severity of winters in the laurentian Great Lakes basin, *Atmosphere-Ocean*, 38(4), 601-635.

Rodionov, S., and R. A. Assel (2003), Winter severity in the Great Lakes region: a tale of two oscillations, *Climate Research*, 24(1), 19-31.

Rogers, J. C., and J. S. M. Coleman (2003), Interactions between the Atlantic Multidecadal Oscillation, El Nino/La Nina, and the PNA in winter Mississippi valley stream flow, *Geophysical Research Letters*, 30(10).

Rössler, O. E. (1976), An equation for continuous chaos, *Physics Letters A*, 57(5), 397-398.

Rössler, O. E. (1995), An Introduction to Chaos, *International Journal of Intelligent Systems*, 10(1), 5-13.

Salas, J. D., and D. C. Boes (1980), Shifting level modeling of hydrologic series, *Advances in Water Resources*, 3, 59-63.

Salas, J. D. (1993), Analysis and Modeling of Hydrologic Time Series, in *Handbook of Hydrology*, edited by D. R. Maidment, pp. 19.11-19.72, McGraw-Hill, New York.

Salas, J. D., O. Sveinsson, T. S. Lee, W. Lane, and D. Frevert (2009), Developments on Stochastic Analysis, Modeling, and Simulation (SAMS 2009), in *World Environmental and Water Resources Congress 2009: Great Rivers*, edited, pp. 4840-4849, ASCE, Kansas City, MO.

Salas, J. D., and T. Lee (2010), Nonparametric Simulation of Single-Site Seasonal Streamflows, *Journal of Hydrologic Engineering*, 15(4), 284-296.

Serreze, M. C., M. P. Clark, D. L. McGinnis, and D. A. Robinson (1998), Characteristics of snowfall over the eastern half of the United States and relationships with principal modes of low-frequency atmospheric variability, *Journal of Climate*, 11(2), 234-250.

Sreekanth, J., and B. Datta (2014), Stochastic and Robust Multi-Objective Optimal Management of Pumping from Coastal Aquifers Under Parameter Uncertainty, *Water Resources Management*, 28(7), 2005-2019.

Stevens, M. J., and G. R. North (1996), Detection of the climate response to the solar cycle, *Journal of the Atmospheric Sciences*, 53(18), 2594-2608.

Sveinsson, O. G. B. (2002), Modeling of Stationary and Non-stationary Hydrological Processes, Doctoral thesis, 208 pp, Colorado State University, Fort Collins.

Sveinsson, O. G. B., J. D. Salas, D. C. Boes, and R. A. Pielke (2003), Modeling the dynamics of long-term variability of hydroclimatic processes, *Journal of Hydrometeorology*, 4(3), 489-505.

The International Upper Great Lakes Study Board (2012), Hydroclimatic Conditions: Past, Present and Future, 236 pp, Ottawa, ON.

Thompson, D. W. J., and J. M. Wallace (1998), The Arctic Oscillation signature in the wintertime geopotential height and temperature fields, *Geophysical Research Letters*, 25(9), 1297-1300.

Trenberth, K. E. (1997), The definition of El Nino, *Bulletin of the American Meteorological Society*, 78(12), 2771-2777.

Tsonis, A. A., J. B. Elsner, and D.-Z. Sun (2007), The Role of El Niño—Southern Oscillation in Regulating its Background State, in *Nonlinear Dynamics in Geosciences*, edited, pp. 537-555, Springer New York.

Wallace, J. M., and D. S. Gutzler (1981), Teleconnections in the geopotential height field during the Northern Hemisphere winter, *Monthly Weather Review*, 109(4), 784-812.

Westra, S., and A. Sharma (2009), Probabilistic estimation of multivariate streamflow using independent component analysis and climate information, *Journal of Hydrometeorology*, 10(6), 1479-1492.

Wilks, D. S. (1997), Resampling Hypothesis Tests for Autocorrelated Fields, *Journal of Climate*, 10(1), 65-82.

Wolter, K., and M. S. Timlin (1993), Monitoring ENSO in COADS with a seasonally adjusted principal component index, paper presented at Proc. of the 17th Climate Diagnostics Workshop, NOAA/NMC/CAC, NSSL, Oklahoma Clim. Survey, CIMMS and the School of Meteor., Univ. of Oklahoma, Norman, OK.

Accepted Article

Wu, Z. H., and N. E. Huang (2004), A study of the characteristics of white noise using the empirical mode decomposition method, *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences*, 460(2046), 1597-1611.

Wu, Z. H., and N. E. Huang (2009), Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method, *Advances in Adaptive Data Analysis*, 1(1), 1-41.

Zhang, Y., J. M. Wallace, and D. S. Battisti (1997), ENSO-like interdecadal variability: 1900-93, *Journal of Climate*, 10(5), 1004-1020.

Table 1. Cross-correlation between the oscillation IMFs of the AO and POD indices for annual and winter data. Note that the approximate 95 percent confidence interval ($\pm 2/\sqrt{N}$) is ± 0.19 .

	Obs.	C ₁	C ₂	C ₃	C ₄	C ₅
Annual	-0.012	-0.187	-0.285	-0.275	0.438	0.894
Winter	-0.12	-0.134	-0.286	-0.604	-0.316	0.873

Accepted Article

Table 2. Key statistics and correlation of the observed and modeled data for the annual climate indices. Note that the values inside the parentheses indicate the standard deviation of the estimated statistics from 500 simulated series while the values in front of the parentheses present the mean of the estimated statistics.

	Obs.			M-NSOR			CSM-CARMA		
	AO	ENSO	PDO	AO	ENSO	PDO	AO	ENSO	PDO
Mean	0.01	0.05	0.03	0.08(0.14)	0.13(0.11)	0.01(0.14)	0(0.14)	0.05(0.11)	0.05(0.14)
Std	0.76	0.77	0.77	0.81(0.07)	0.74(0.06)	0.78(0.08)	0.73(0.07)	0.73(0.06)	0.76(0.08)
Skew	-0.22	-0.06	0.08	0.02(0.23)	0.02(0.24)	0.03(0.24)	-0.05(0.23)	-0.04(0.24)	0(0.24)
Max	1.94	1.66	1.99	2.09(0.36)	1.99(0.34)	1.94(0.37)	1.75(0.36)	1.82(0.34)	1.96(0.37)
Min	-1.98	-1.61	-1.95	-1.88(0.35)	-1.7(0.32)	-1.87(0.36)	-1.88(0.35)	-1.77(0.32)	-1.83(0.36)
Auto1 [†]	0.43	0.29	0.53	0.45(0.1)	0.25(0.11)	0.52(0.09)	0.38(0.38)	0.35(0.35)	0.26(0.26)
Auto2	-0.31	-0.15	0.27	0.29(0.12)	0.17(0.11)	0.35(0.12)	0.29(0.29)	0.25(0.25)	0.04(0.04)
Corr. [‡]	0.17	0.17	0.73	0.04(0.12)	0.06(0.12)	0.7(0.06)	-0.03(0.18)	0.18(0.13)	0.58(0.09)

[†]Auto1 and Auto2 indicate the autocorrelations of the lag-1 and lag-2, respectively.

[‡]Corr. indicates the correlations between the climate indices. The AO column presents the correlation between AO-ENSO (e.g. 0.04 for M-NSOR), the ENSO column between ENSO-PDO (e.g. 0.06 for M-NSOR), and the PDO column between PDO-AO (e.g. 0.07 for M-NSOR).

Accepted

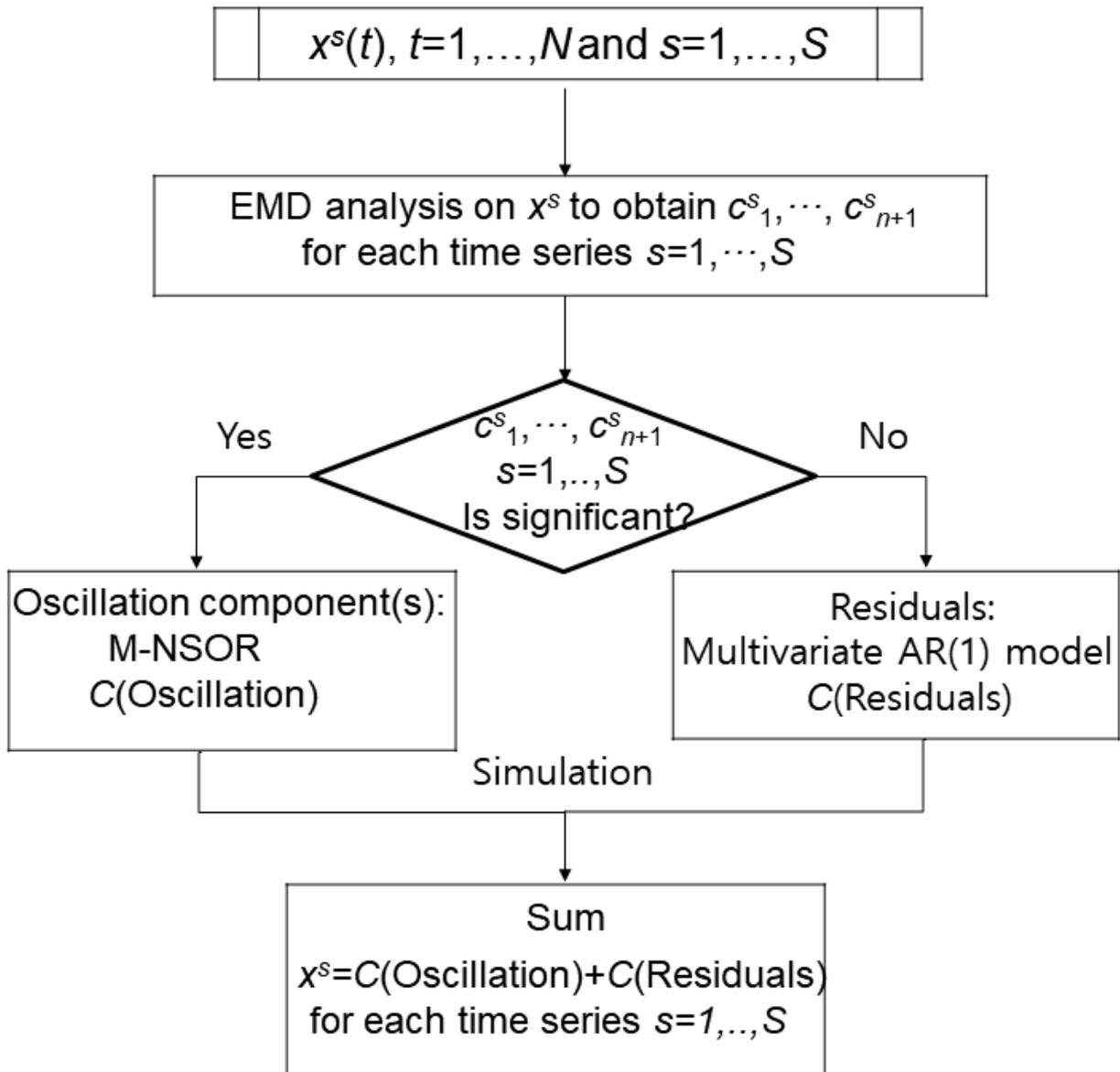


Figure 1. Overall process of the proposed model M-NSOR.

ACCEPTED

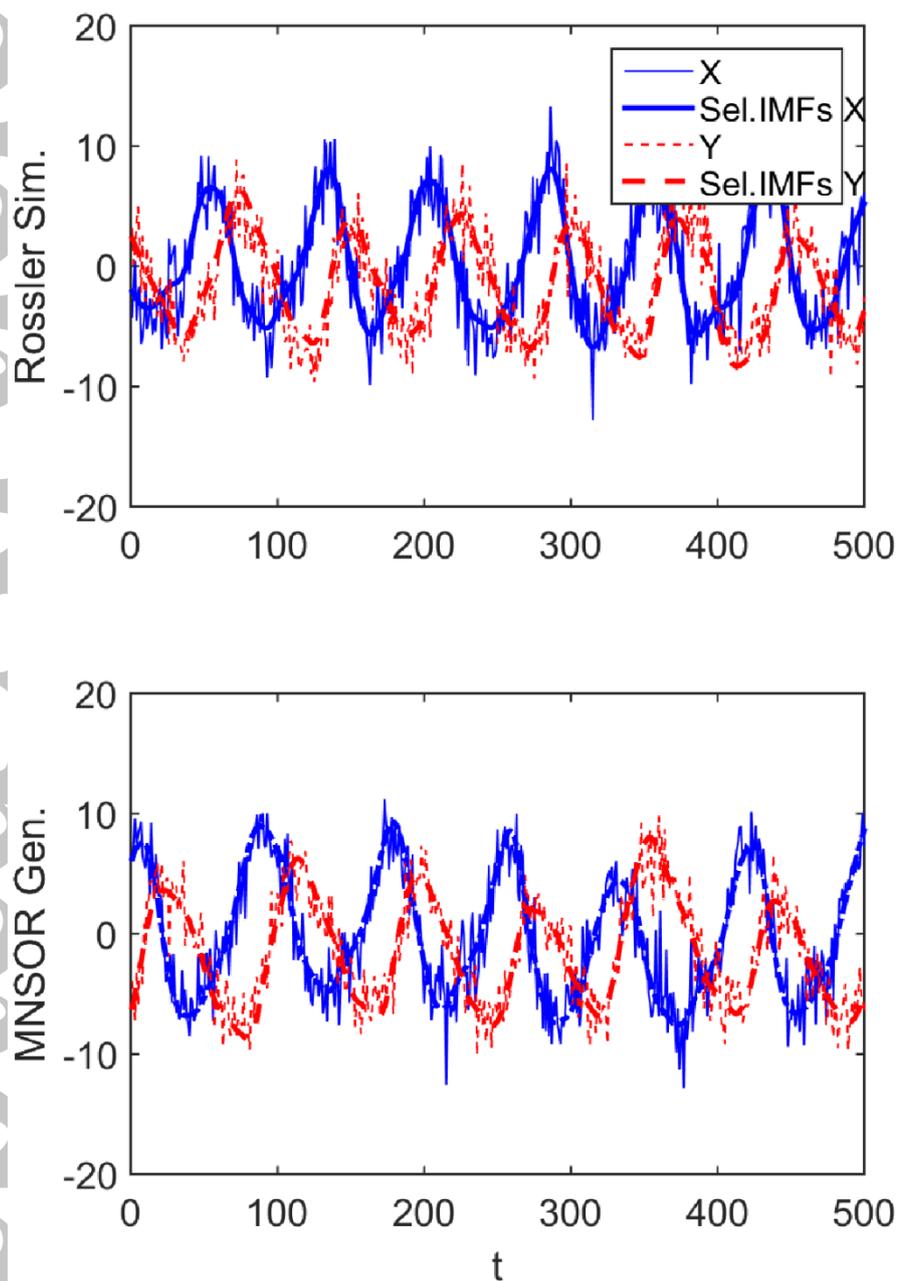


Figure 2. Time series of the simulated X (thin solid line) and Y (thin dotted line) variables from Rössler attractor (top panel) and the generated variables with the proposed M-NSOR model (bottom panel). The thick lines indicate the sum of the selected IMF components from EMD as $c_4+c_5+c_8$.

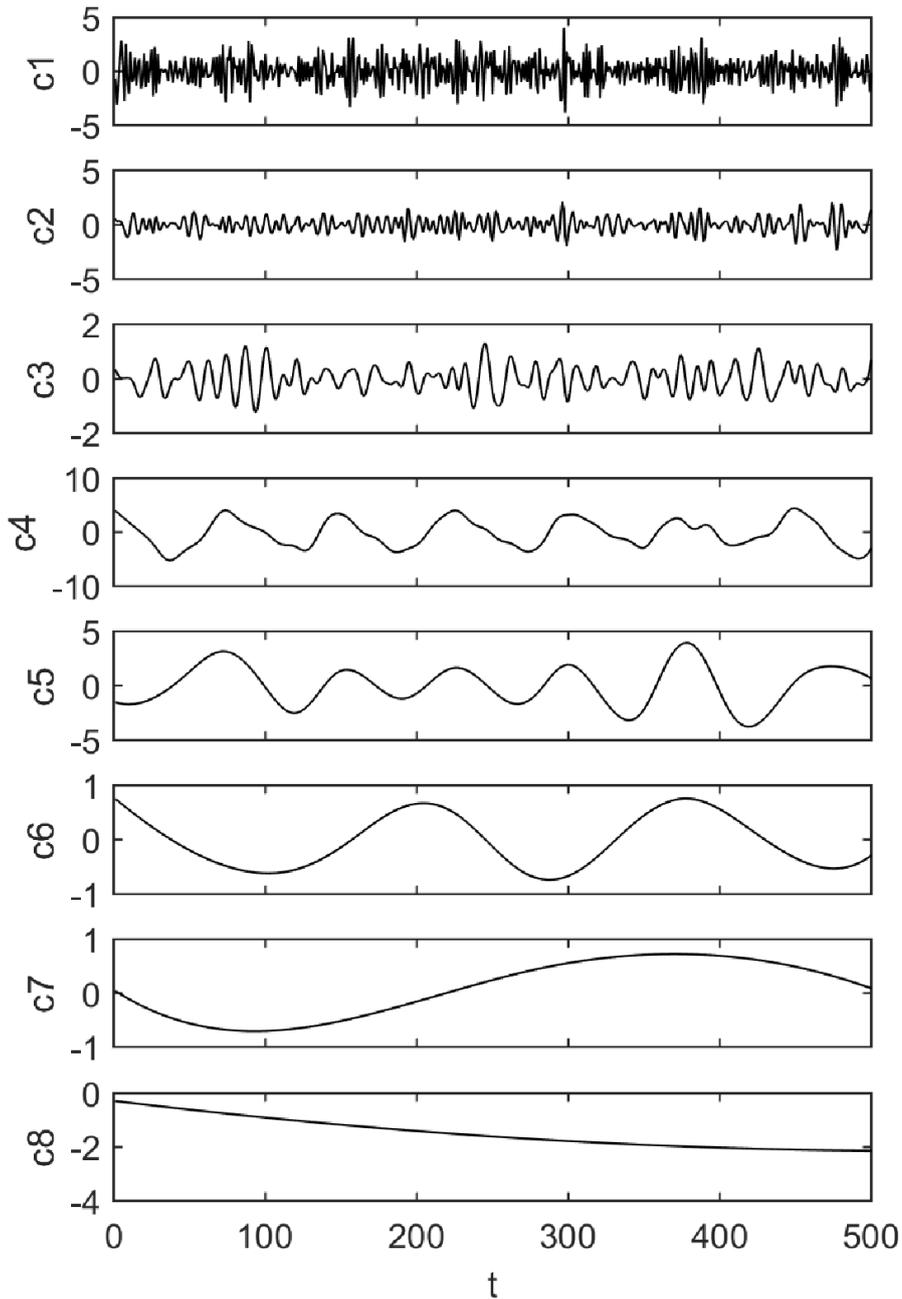


Figure 3. IMFs for the simulated X variable from Rössler attractor. Note that the extracted IMFs are ordered as they are extracted from the procedure in the section 2.1; a component with a lower number (e.g. c_1 or c_2) presents a higher frequency and vice versa; and the last component (c_8) represents the overall trend.

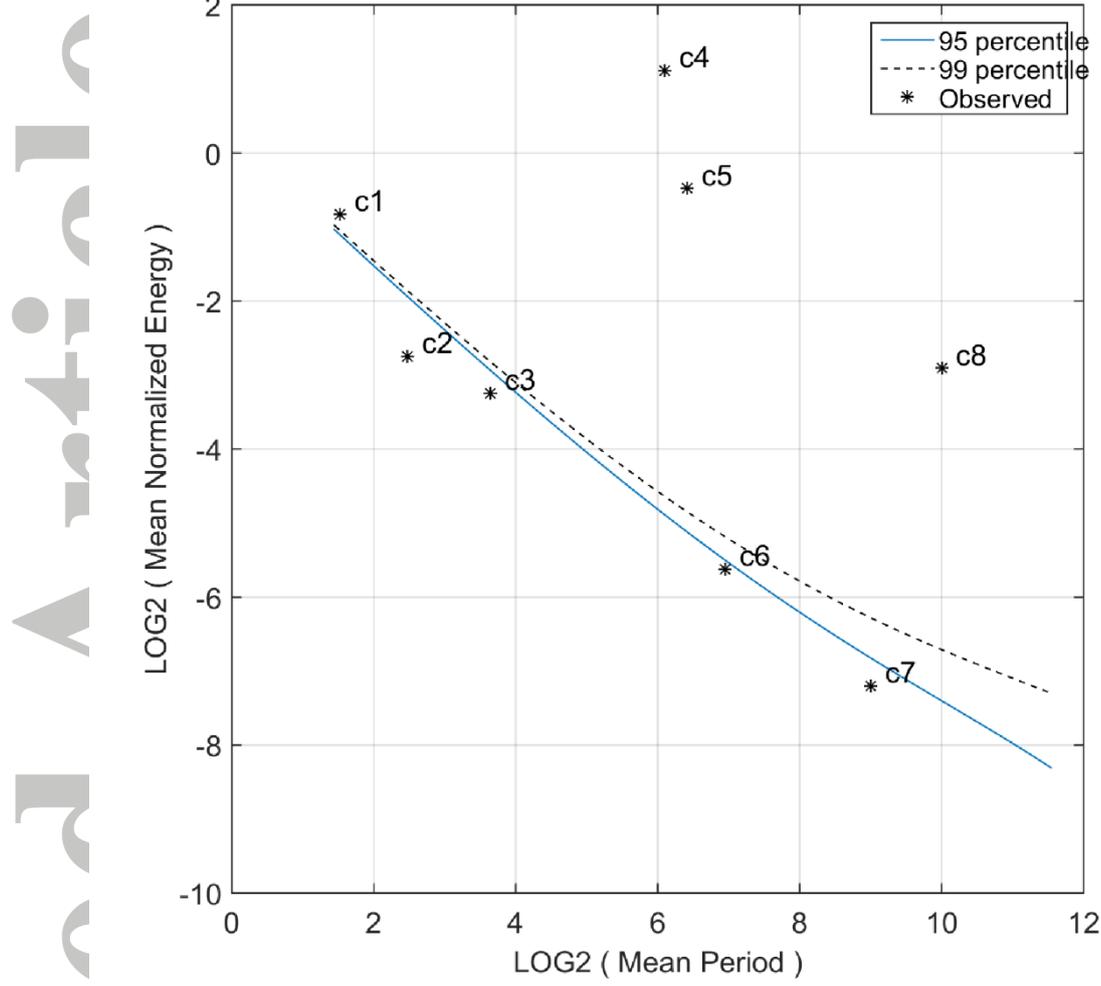


Figure 4. Significance test of the simulated X variable from Rössler attractor with 95% (solid line) and 99% (dotted line) confidence limits. Each point (*) below the lines indicates that the hypothesis that the corresponding IMF of the observed series is not distinguishable from the corresponding IMF of a random noise series cannot be rejected with the confidence levels (95% and 99% respectively). Notice that c_4 , c_5 , and c_8 are significant while c_4 has the highest mean normal energy and the first component c_1 is generally considered as random component in EMD analysis and neglected in the test [Wu and Huang, 2004; Lee and Ouarda, 2012b].

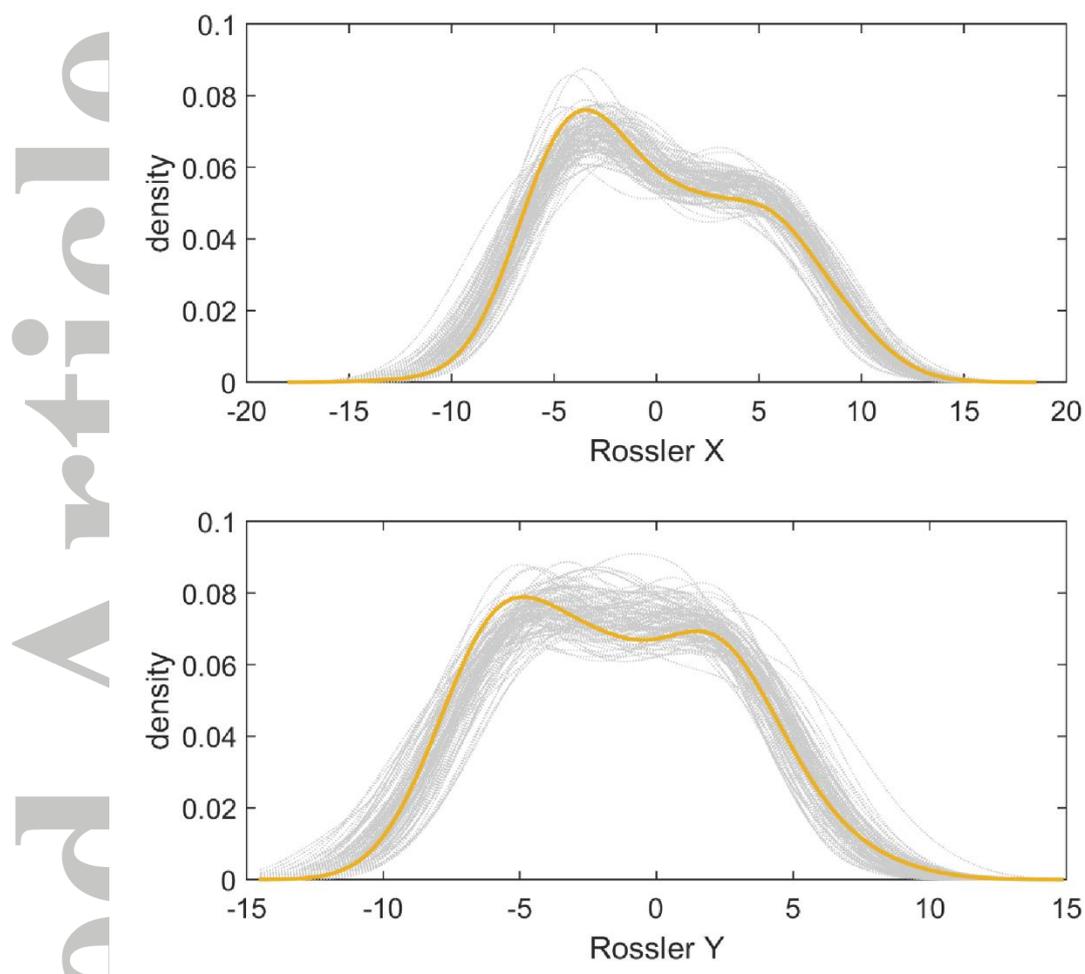


Figure 5. Kernel densities for the X (top panel) and Y (bottom panel) variables from the Rössler attractor (thick line) and the ones from 100 generation series (thin lines) with the M-NSOR model.

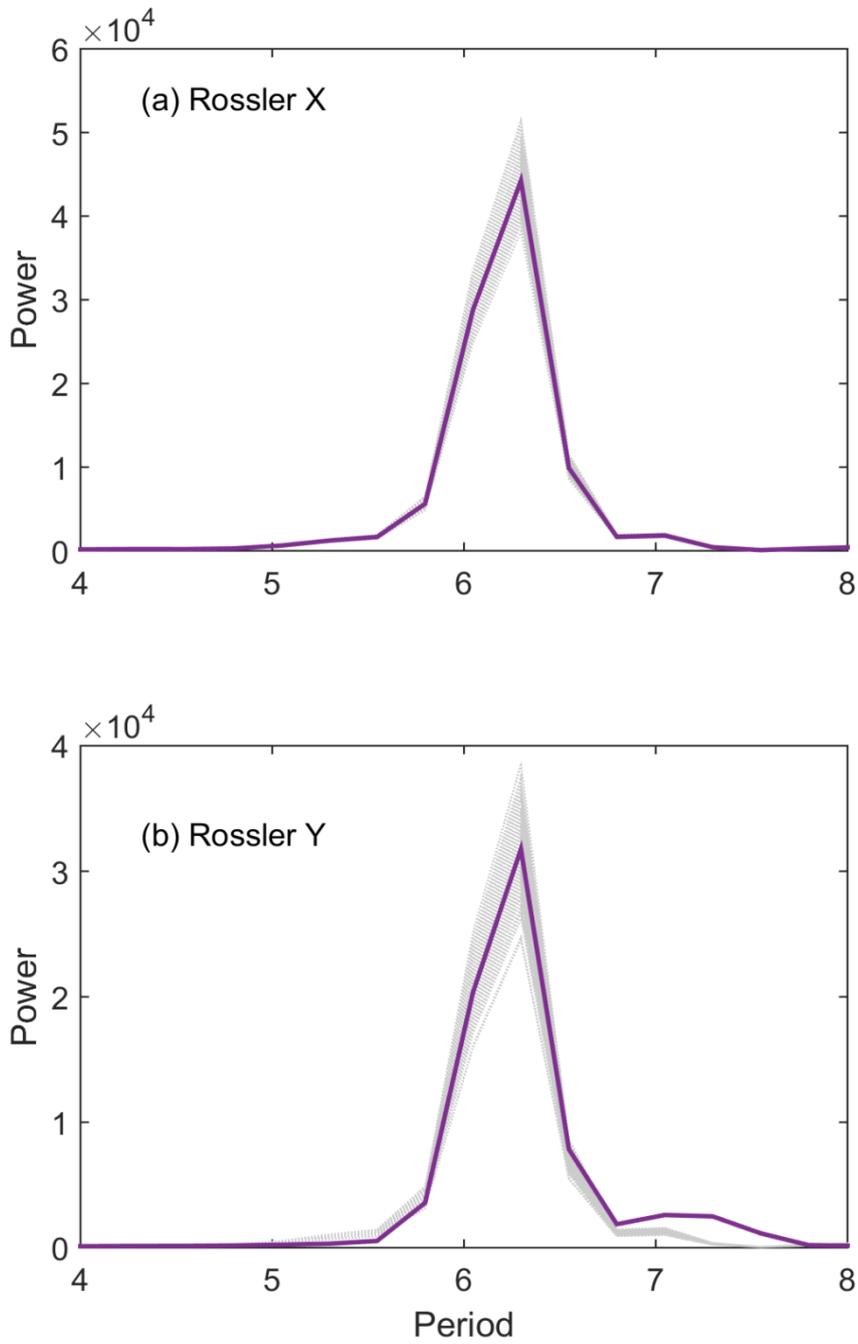


Figure 6. Global wavelet spectra of the simulated X and Y variables from Rössler attractor (thick solid line) and one hundred simulated series (thin dotted lines) with the M-NSOR model.

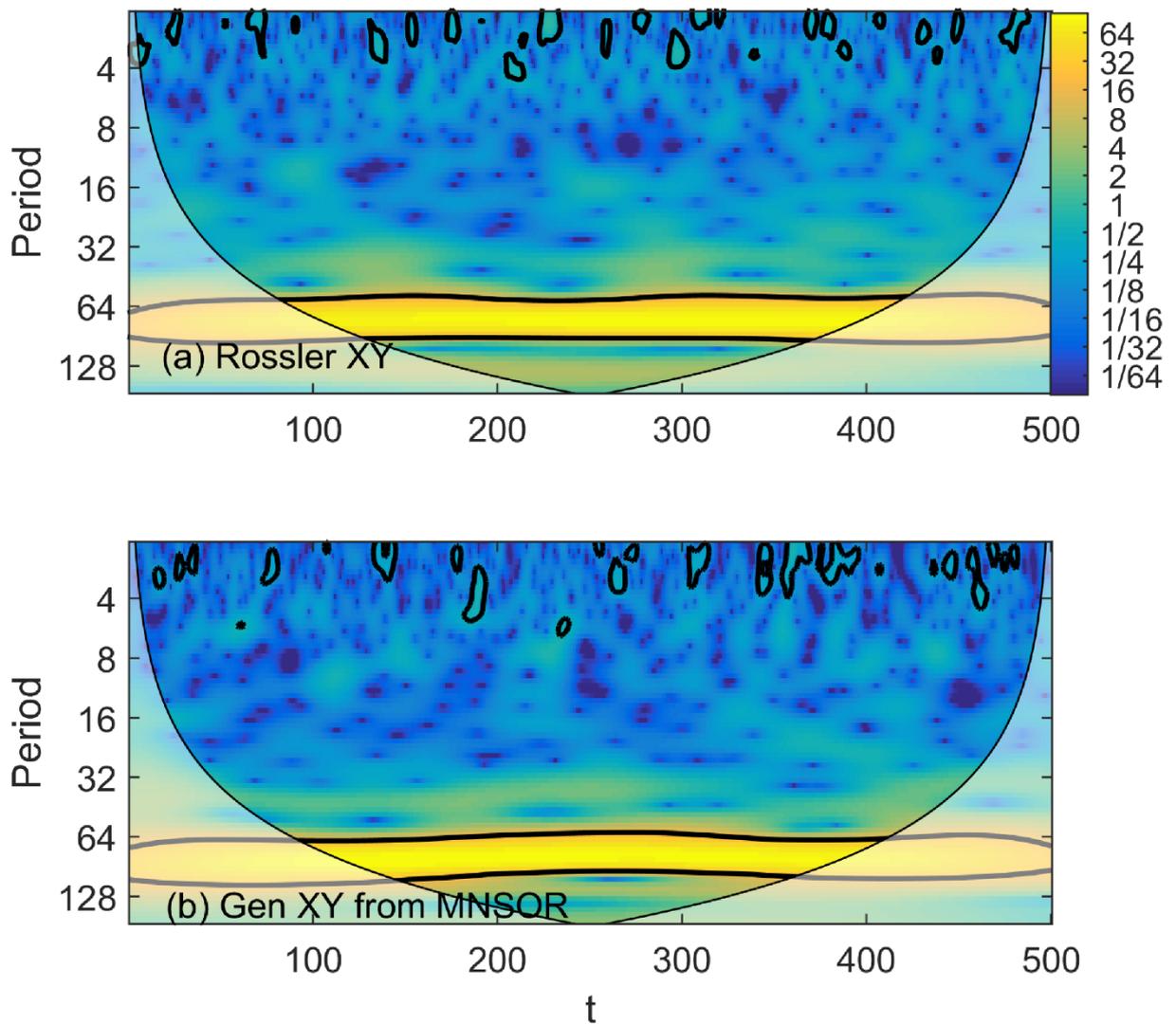


Figure 7. Cross wavelet spectra [Grinsted *et al.*, 2004] of the simulated X and Y variables from Rössler attractor (top panel) and one sample (bottom panel) with the M-NSOR model. Note that the 5% significance level against red noise is shown as a thick contour.

Accepted

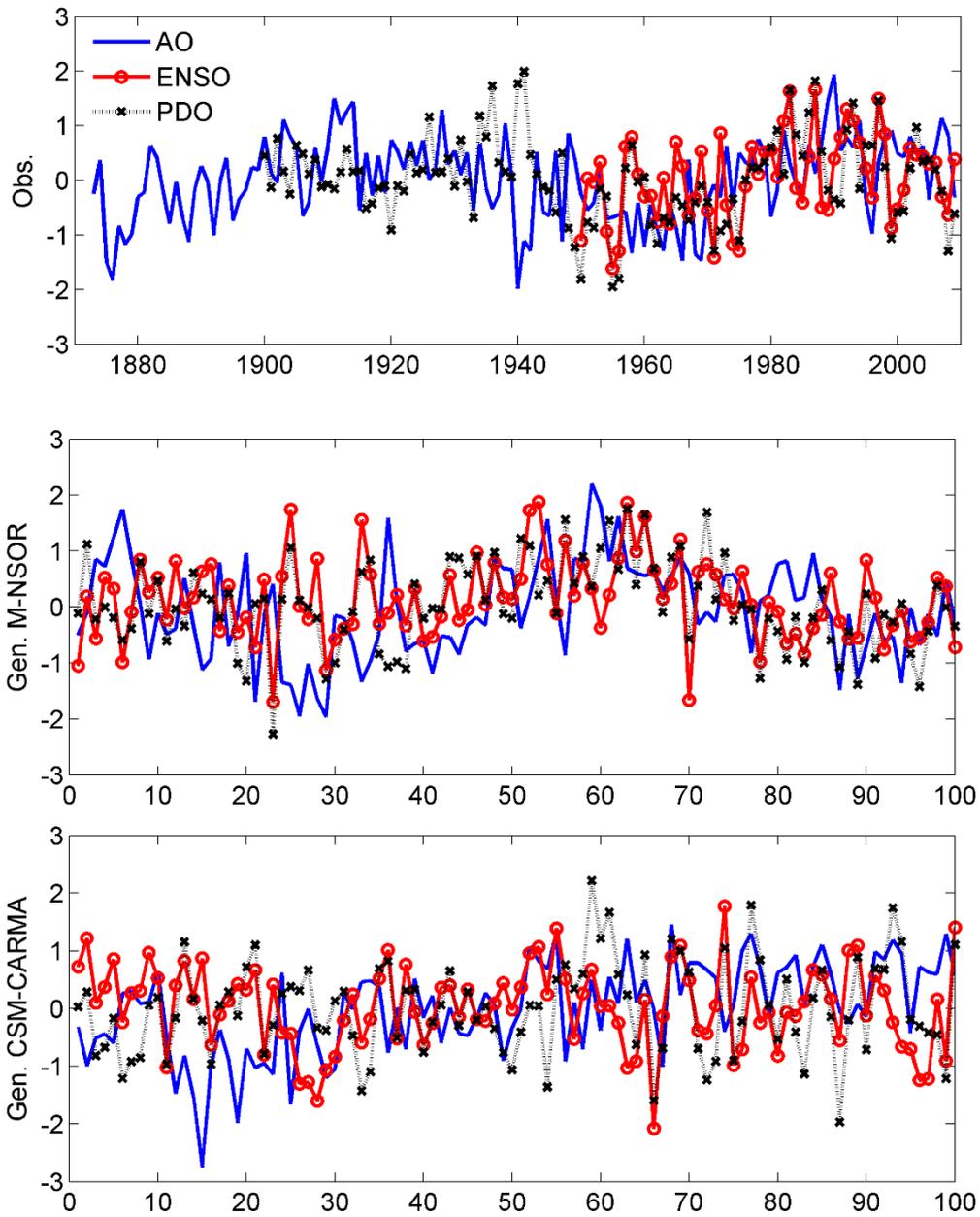


Figure 8. Time series of the observed (top panel) and simulated annual climate indices with $N=100$ for the M-NSOR (middle panel) and CSM-CARMA models (bottom panel). Note that the record lengths of the observed AO and PDO indices are 137 years and 110 years, respectively and the length of ENSO is 60 years. The lengths of the generated data are 100 for both models (see the middle and bottom panels).

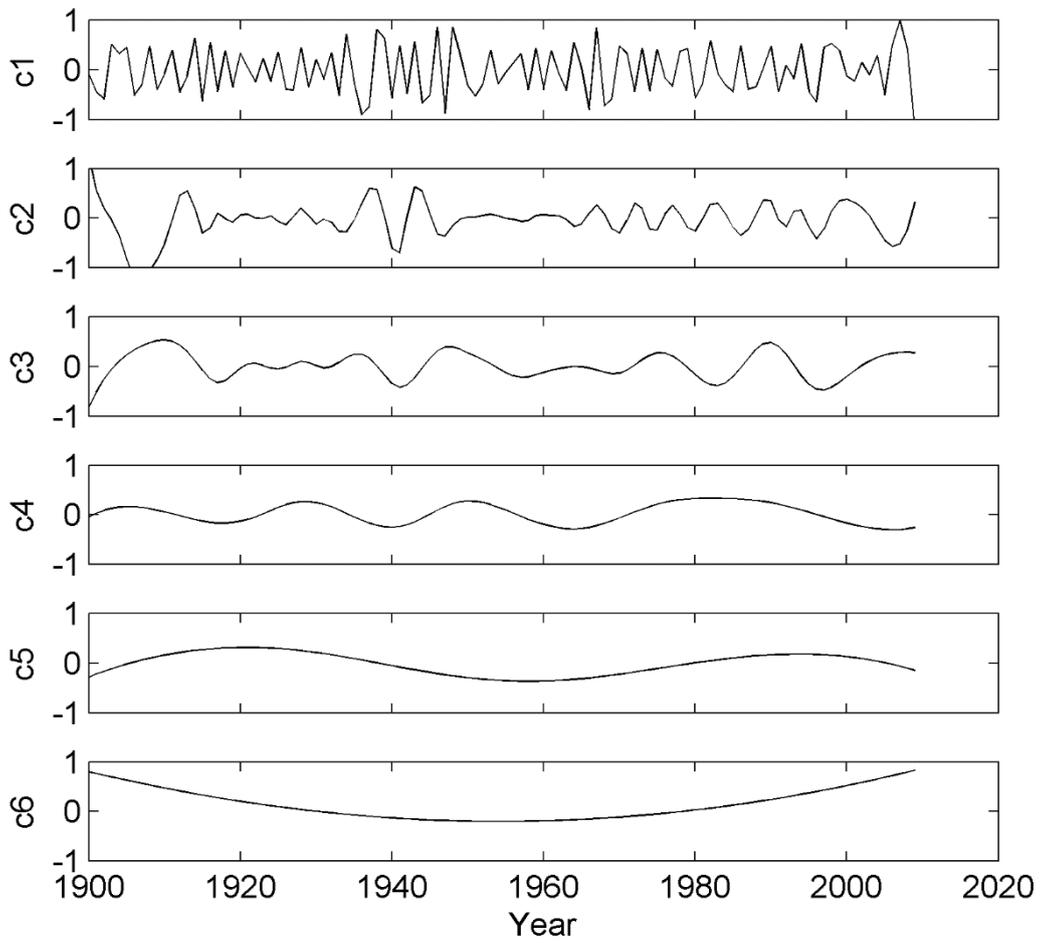


Figure 9. IMFs for the observed annual AO index. Note that the extracted IMFs are ordered as they are extracted from the procedure in the section 2.1; a component with a lower number (e.g. c_1 or c_2) presents a higher frequency and vice versa; and the last component (c_6) represents the overall trend.

Accepted

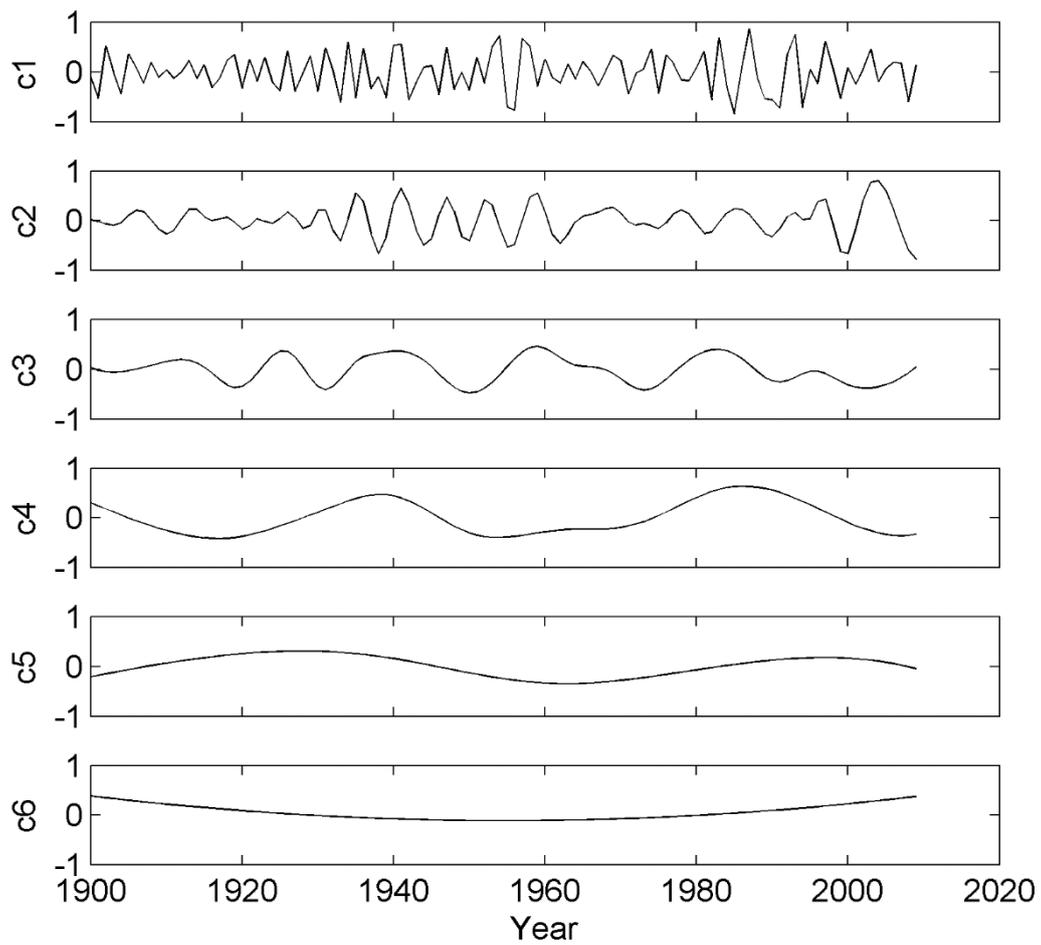


Figure 10. IMFs for the observed annual PDO index. Note that the extracted IMFs are ordered as they are extracted from the procedure in the section 2.1; a component with a lower number presents a higher frequency vice versa; and the last component (c_6) represents the overall trend.

Accepted

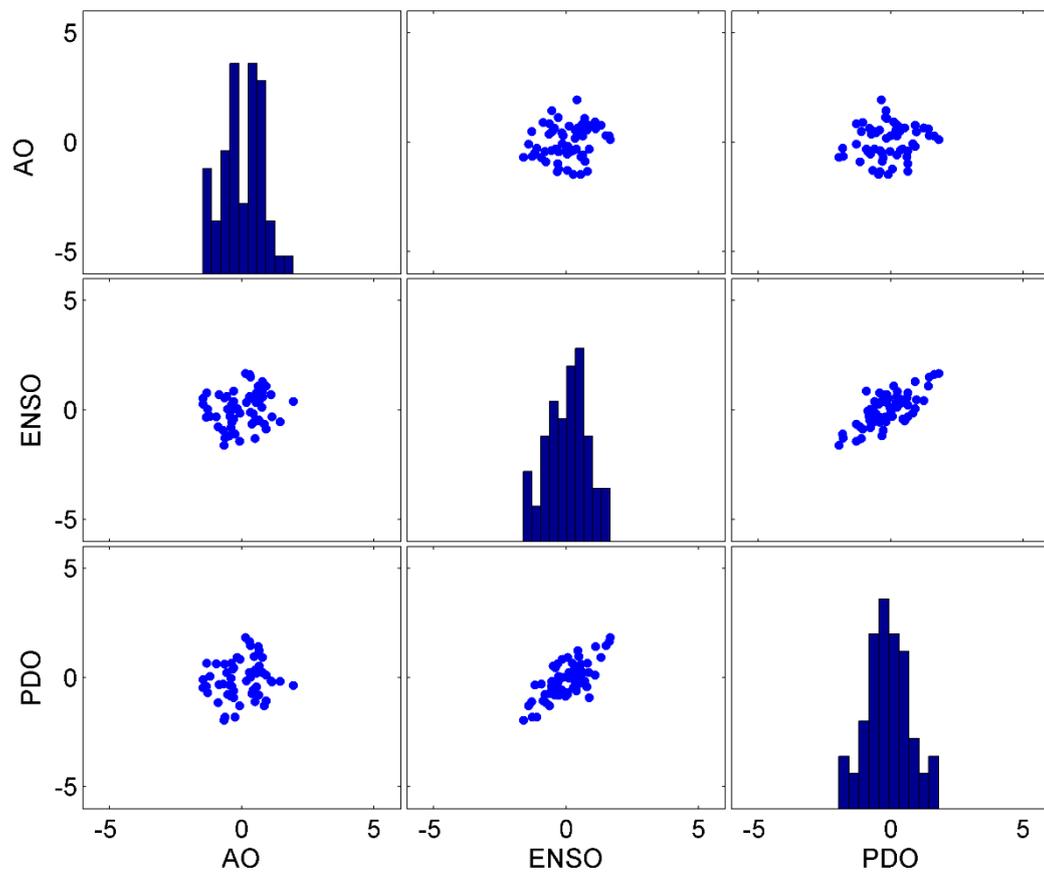


Figure 11. Histograms and scatterplots between the observed annual climate indices (AO, ENSO, and PDO).

Accept

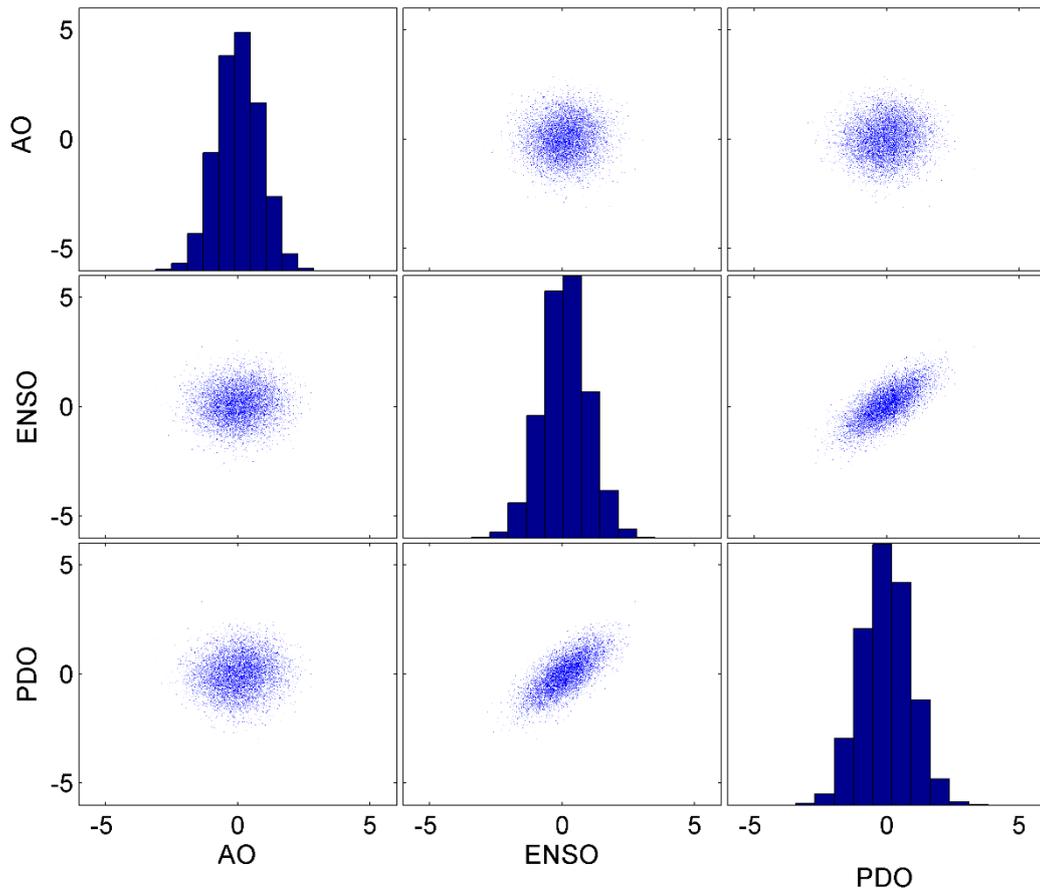


Figure 12. Histograms and scatterplots between the simulated annual climate indices from the M-NSOR model.

Accept

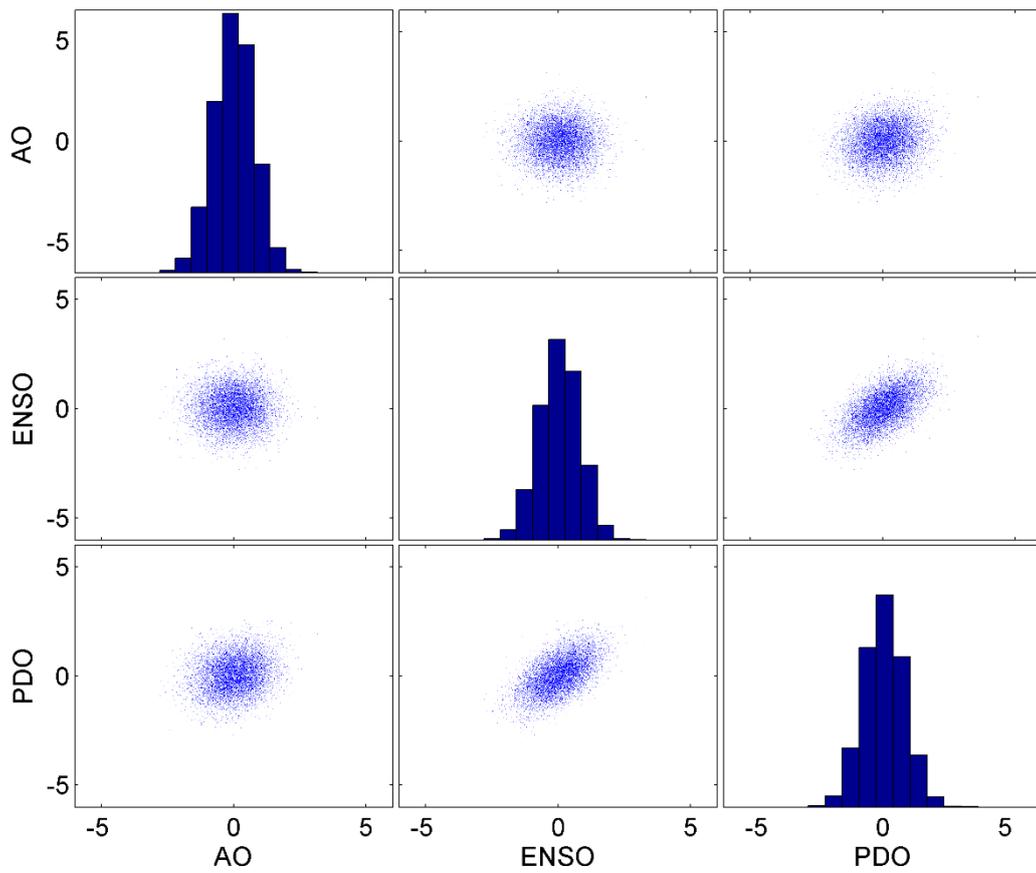


Figure 13. Histograms and scatterplots between the simulated annual climate indices from the CSM-CARMA model.

Accept

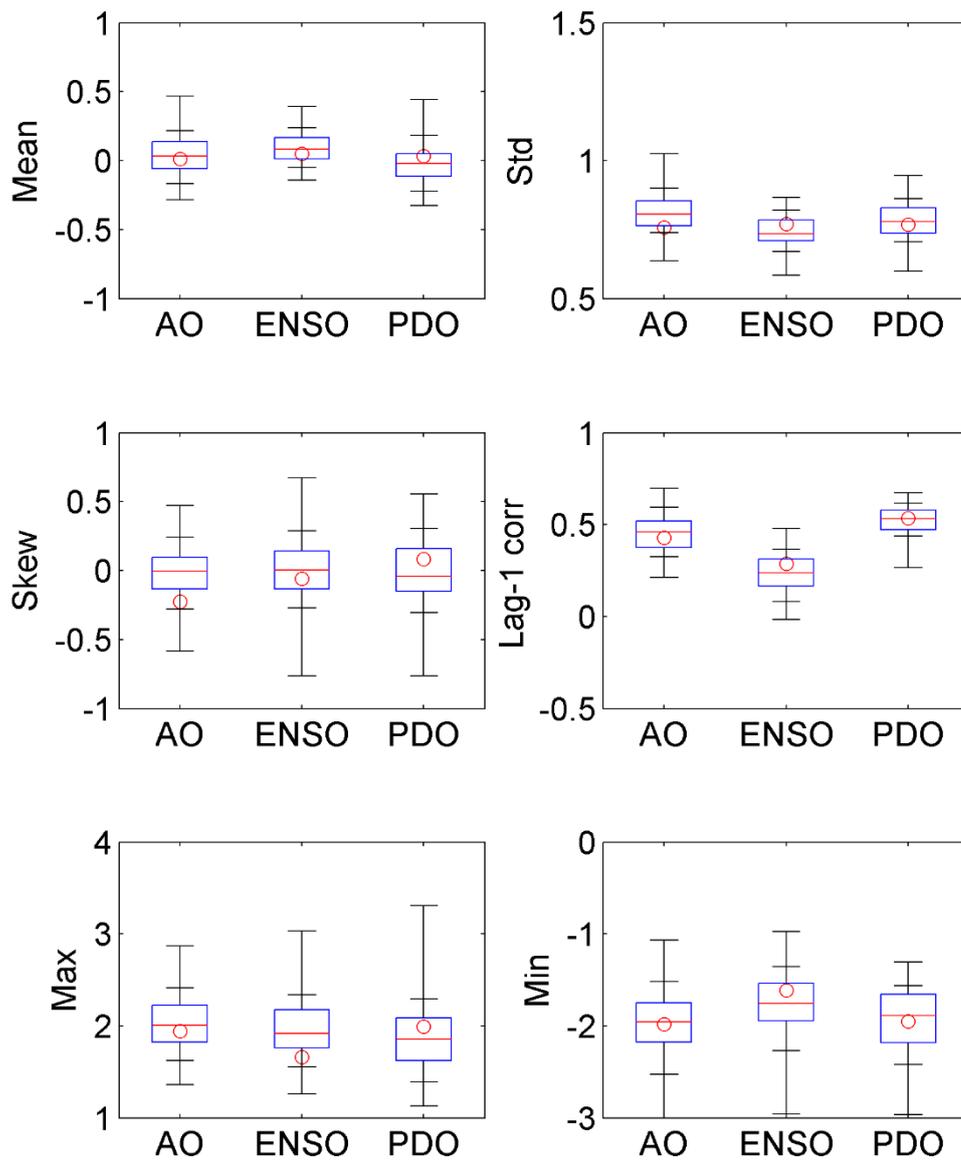


Figure 14. Key statistics of the annual climate indices for the observed (circle) and simulated data from the M-NSOR model (boxplot). For the boxplot, the end lines of the box are the 25th and 75th percentiles, the cross lines above the box on the whisker correspond to the 90th percentile and the maximum, and the cross lines below the box on the whisker represent the 10th percentile and the minimum. The circle represents the statistics of the observed data.

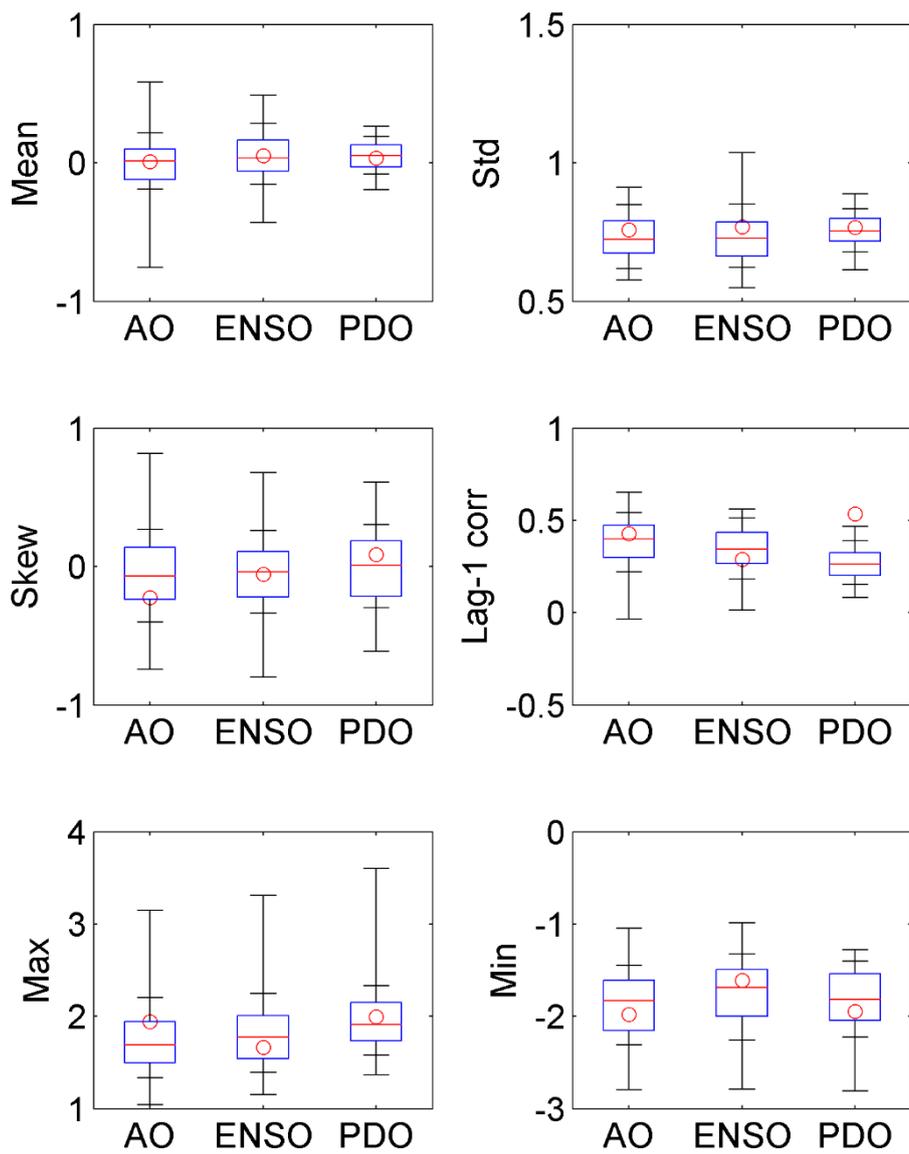


Figure 15. Same as Figure 14 but for simulated data from the CSM-CARMA model.

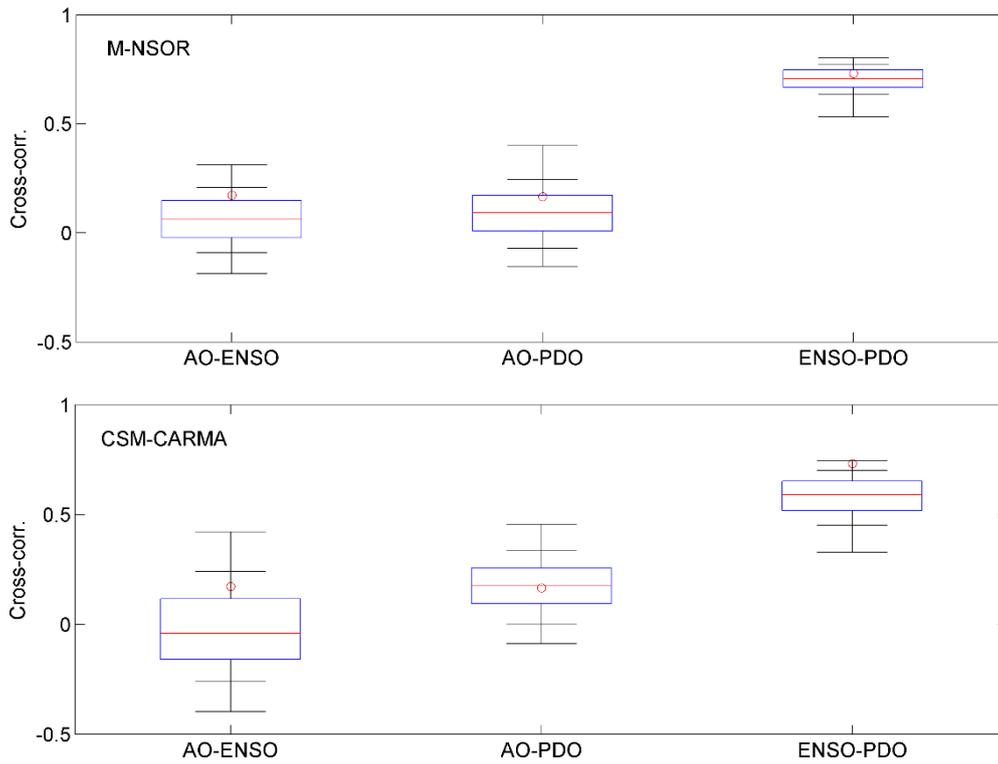


Figure 16. Cross-correlation of the annual climate indices for the observed (circles) and simulated data from the M-NSOR model (see the top panel) and the CSM-CARMA model (see the bottom panel). See Figure 14 for further information about a boxplot.

Accepted

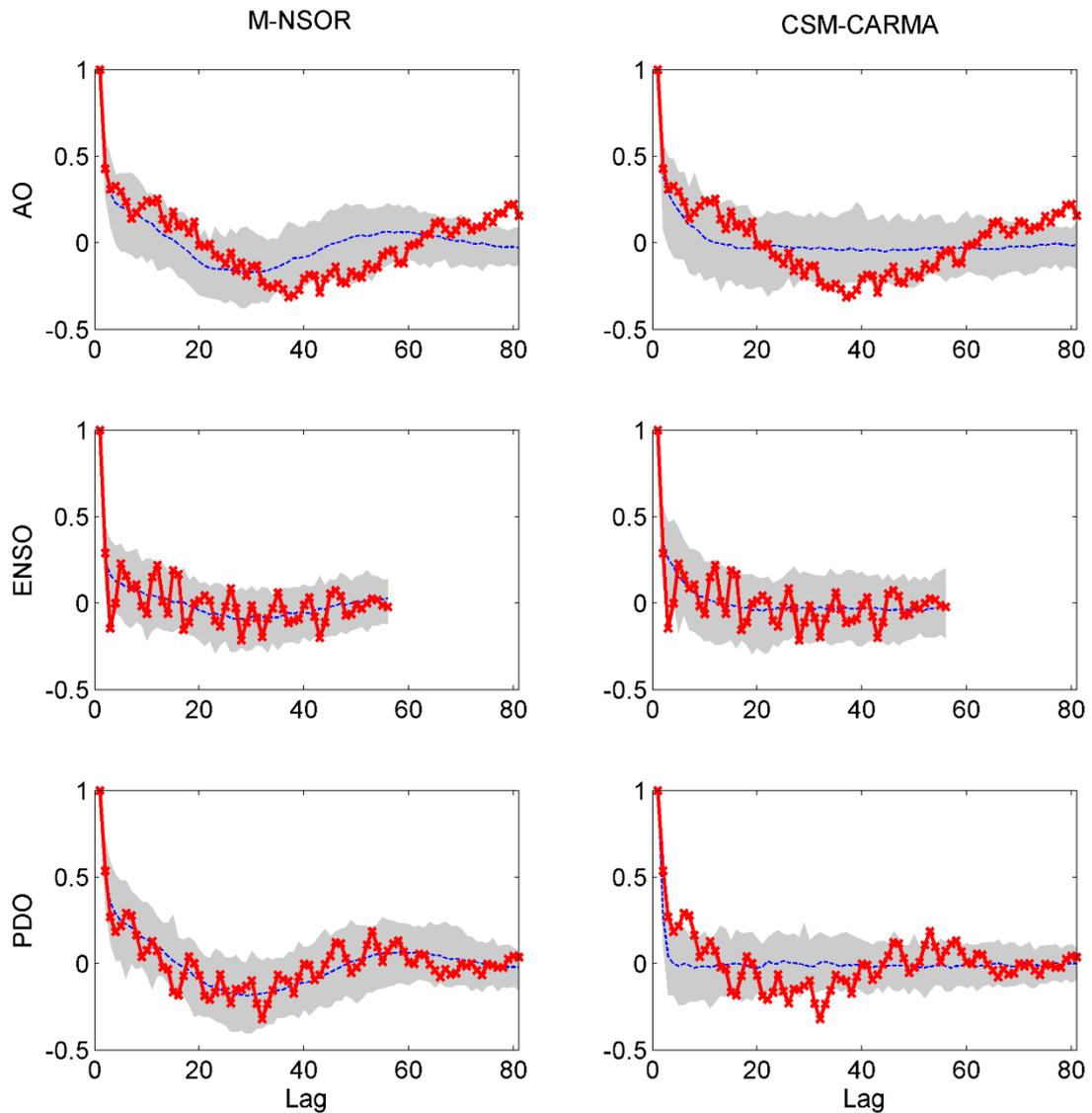


Figure 17. Lagged correlation of the annual climate indices for the observed (red thick solid line with cross marker) and simulated data from the M-NSOR model (see the left panels) and the CSM-CARMA model (see the right panels). Thin dotted blue line presents the mean of the estimated correlations from 500 series and grey shade depicts the 95 percentiles of the estimated statistics from 500 series while the thick red line with cross markers shows the observed one. Note that the lagged correlation for the ENSO data was estimated up to 55 lags because its record length is relatively shorter than the other climate indices.

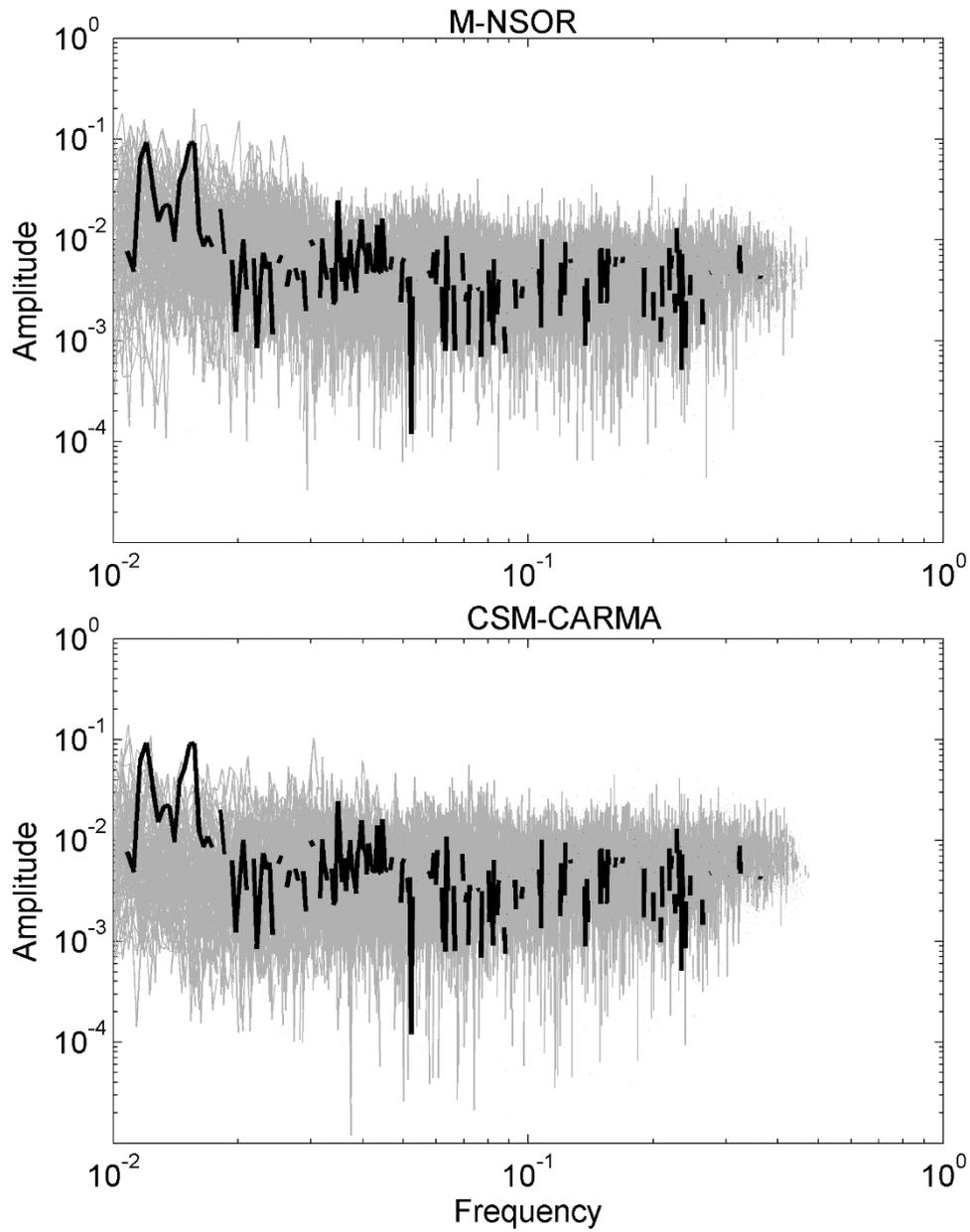


Figure 18. Hilbert marginal spectrum (see Eq. (4)) for the annual PDO index (thick solid line) and the simulated series (grey lines) from the M-NSOR model (top panel) and the CSM-CARMA model (bottom panel).

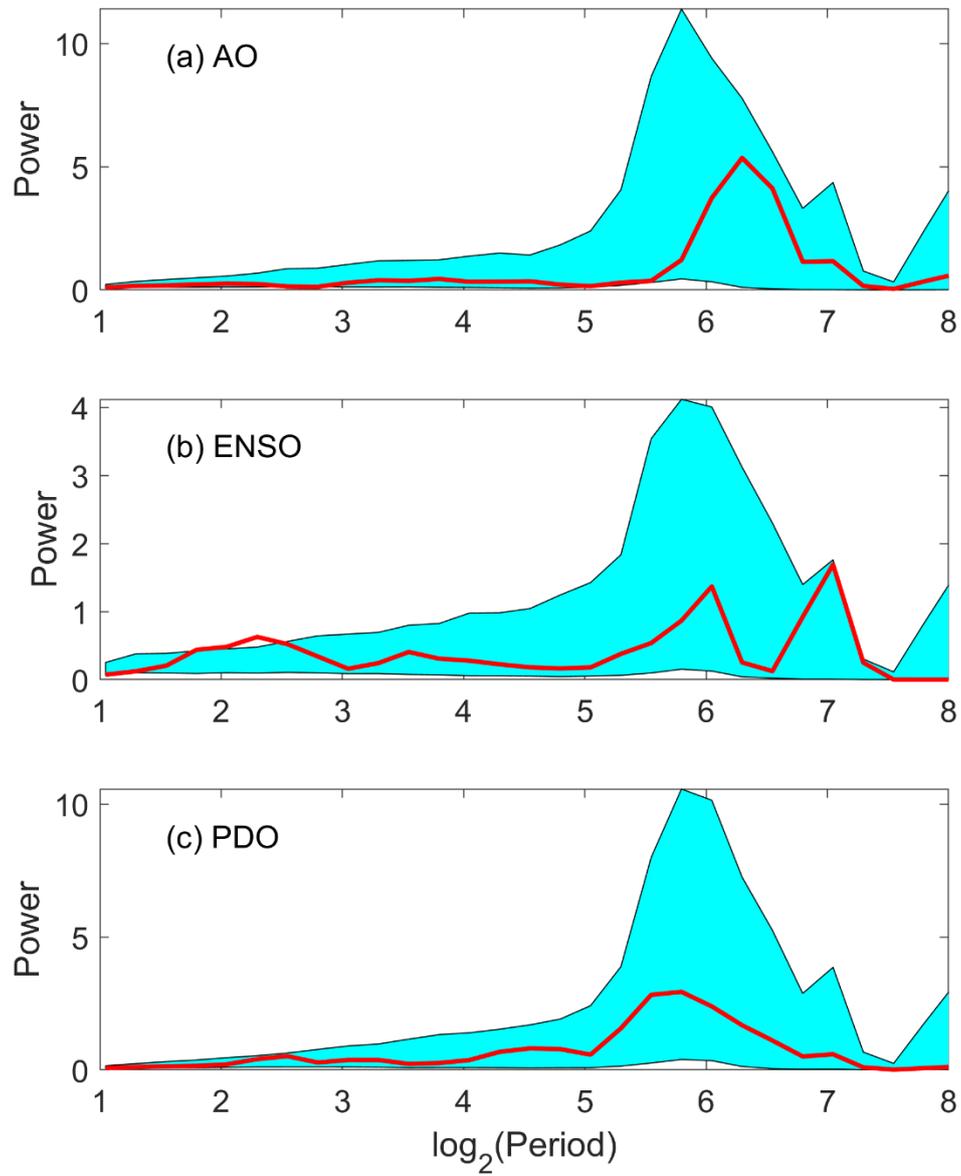


Figure 19. Global wavelet spectrum for the annual climate indices (thick solid line) and the simulated series (95percent interval) from the M-NSOR model.

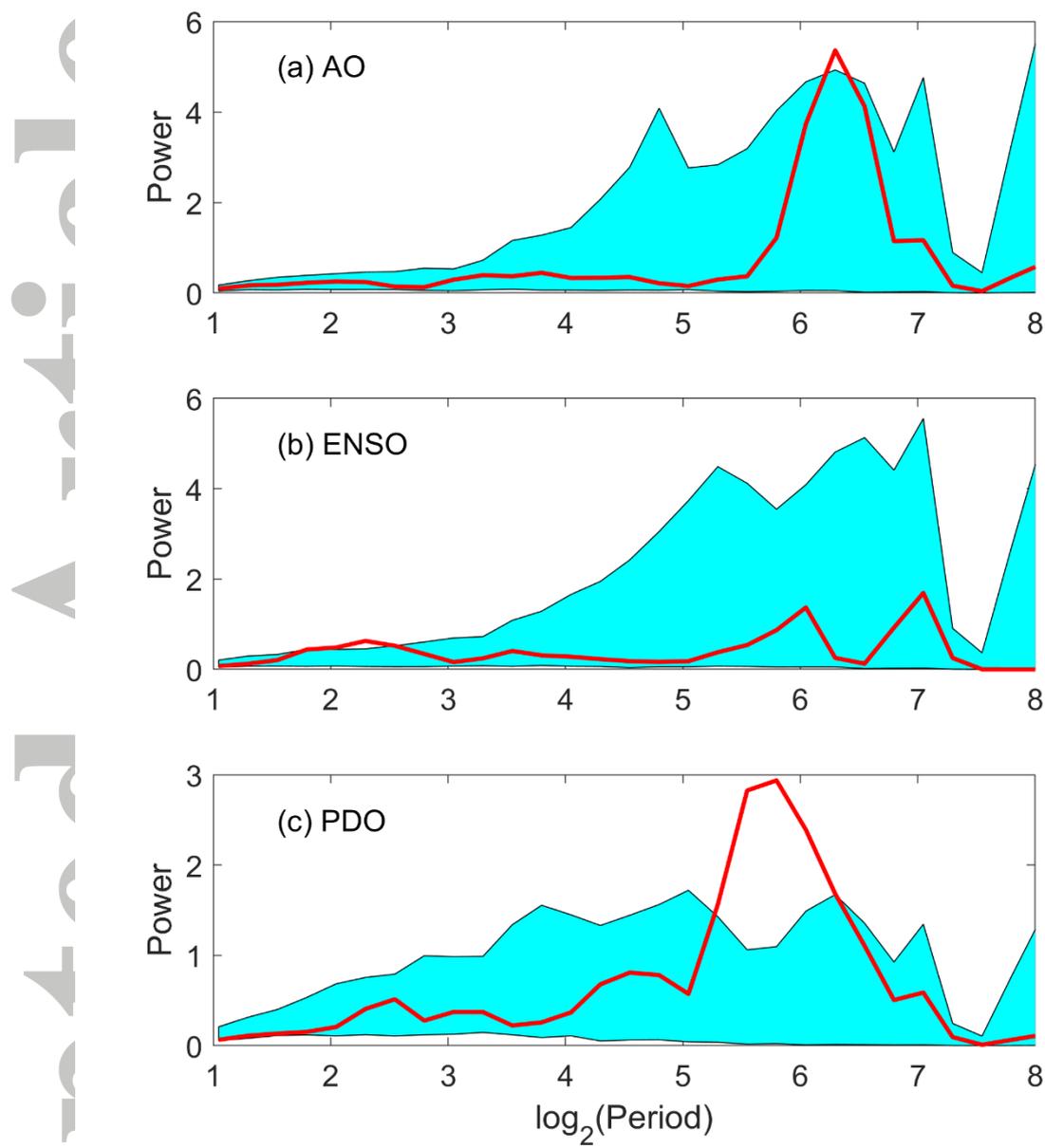


Figure 20. Global wavelet spectrum for the annual climate indices (thick solid line) and the simulated series (95percent interval) from the CSM-CARMA model.