Stochastic simulation of nonstationary oscillation hydroclimatic processes using empirical mode decomposition

T. Lee[1] and T. B. M. J. Ouarda[2,3]

Received 13 March 2011; revised 31 October 2011; accepted 29 December 2011; published 15 February 2012.

[1] Nonstationary oscillation (NSO) processes are observed in a number of hydroclimatic data series. Stochastic simulation models are useful to study the impacts of the climatic variations induced by NSO processes into hydroclimatic regimes. Reproducing NSO processes in a stochastic time series model is, however, a difficult task because of the complexity of the nonstationary behaviors. In the current study, a novel stochastic simulation technique that reproduces the NSO processes embedded in hydroclimatic data series is presented. The proposed model reproduces NSO processes by utilizing empirical mode decomposition (EMD) and nonparametric simulation techniques (i.e., *k*-nearest-neighbor resampling and block bootstrapping). The model was first tested with synthetic data sets from trigonometric functions and the Rössler system. The North Atlantic Oscillation (NAO) index was then examined as a real case study. This NAO index was then employed as an exogenous variable for the stochastic simulation of streamflows at the Romaine River in the province of Quebec, Canada. The results of the application to the synthetic data sets and the real-world case studies indicate that the proposed model preserves well the NSO processes along with the key statistical characteristics of the observations. It was concluded that the proposed model possesses a reasonable simulation capacity and a high potential as a stochastic model, especially for hydroclimatic data sets that embed NSO processes.

Citation: Lee, T., and T. B. M. J. Ouarda (2012), Stochastic simulation of nonstationary oscillation hydroclimatic processes using empirical mode decomposition, *Water Resour. Res.*, *48*, W02514, doi:10.1029/2011WR010660.

1. Introduction

[2] Stochastic simulation models of climatological and hydrological variables have been widely developed as an alternative to unavailable long series of historic data and to assess the associated risk in agricultural, environmental, and water resources studies. A number of time series models have emanated from autoregressive moving average (ARMA) models with the assumption of stationarity or periodic stationarity [*Salas et al.*, 1980; *Stedinger and Taylor*, 1982a, 1982b; *Stedinger et al.*, 1985; *Bartolini et al.*, 1988; *Garrido and Garcia*, 1992; *Salas*, 1993; *Chebaane et al.*, 1995; *Katz and Parlange*, 1996; *Gamiz-Fortis et al.*, 2008]. ARMA type models generally require data transformation. It has been known that modeling in a transformed domain may cause some bias in reproducing the key statistics in the real domain (e.g., variance, skewness).

[3] Nonparametric time series simulation techniques have been developed to overcome some of the drawbacks of ARMA type models with the same stationarity and periodic stationarity assumption [*Lall et al.*, 1996; *Lall and Sharma*, 1996; *Ouarda et al.*, 1997; *Rajagopalan and Lall*, 1999; *Lee et al.*, 2010; *Salas and Lee*, 2010]. Unlike parametric models, nonparametric techniques do not make any assumption concerning the underlying distribution. There is also no need for parameterization in order to reproduce observational statistics.

[4] During the previous decades, as sufficiently long records of hydroclimatological data have been recorded, experts started realizing the structure of long-term variations in climate and hydrologic data sets [*Ghil and Vautard*, 1991; *Dettinger et al.*, 1995; *Lee and Ouarda*, 2011]. These long-term variations are commonly nonstationary or quasiperiodic. Researchers have focused then on the development of time series models that mimic these nonstationary or quasiperiodic variations [*Sveinsson et al.*, 2003; *Ahn and Kim*, 2005; *Kwon et al.*, 2007].

[5] *Sveinsson et al.* [2003] considered the long-term variations as sudden mean shifts and applied the shifting mean level (SML) model, originally developed by *Salas and Boes* [1980]. SML was fitted to the Pacific Decadal Oscillation (PDO) index resulting in a fairly good reproduction of the statistical characteristics of the index. *Ahn and Kim* [2005] employed an autoregressive conditional heteroscedasticity model to describe the nonlinear properties of the Southern Oscillation Index (SOI). *Kwon et al.* [2007] devised an autoregressive type model assisted by wavelet analysis and applied it to rainfall and temperature series. The NINO 3.4 series was also tested by *Kwon et al.* [2007]

[1]Department of Civil Engineering, ERI, Gyeongsang National University, Jinju, South Korea.
[2]Masdar Institute of Science and Technology, Abu Dhabi, United Arab Emirates.
[3]Canada Research Chair on the Estimation of Hydrometeorological Variables, INRS-ETE, Quebec, Canada.

0043-1397/12/2011WR010660

and the results indicated that nonstationary processes embedded in the series were fairly well reproduced. All these models, however, do not take into account the slowly varying nonstationary oscillation (NSO) processes. Specifically, the modeling of long-term NSO processes still needs to be carried out.

[6] Meanwhile, *Lee and Ouarda* [2010, 2011] proposed a NSO model to predict the future evolution of hydroclimatological variables. The proposed model employs a novel decomposition technique, called empirical mode decomposition (EMD), and nonparametric time series models (i.e., k-nearest-neighbor resampling (KNNR) and block bootstrapping). The results presented by *Lee and Ouarda* [2010, 2011] revealed that the future long-term oscillation process is relatively well predicted with the proposed model.

[7] The objective of the current study is to revise the model of *Lee and Ouarda* [2010, 2011] and to demonstrate that it can be employed as a stochastic simulation model for hydroclimatological variables. The simulation performance of this model is tested with synthetic data sets and real-world case studies. Furthermore, using teleconnection results, a synthetic climate index showing long-term oscillations is employed in simulating the hydrological variable whose observed data do not reveal long-term oscillations because of the relatively short record length. It is shown that this procedure is efficient when the index and the hydrological variable are significantly correlated.

[8] The paper is organized as follows. The theoretical background and the proposed model are described in sections 2 and 3, respectively. To validate the model, synthetic data sets from trigonometric functions and the Rössler system are used in section 4. The results of the application of the proposed model on the real-world case studies representing a low-frequency climatic index and a hydrologic variable are illustrated in section 5. Finally, the summary and conclusions are presented in section 6.

2. Background

2.1. Hilbert-Huang Transformation

[9] The Hilbert-Huang transformation (HHT) consists of Hilbert spectral analysis (HSA) and EMD. The Hilbert spectral analysis allows a way to express the nonstationarity of time series data by evaluating the instantaneous frequency and instantaneous amplitude through the Hilbert transform (HT). To apply the HT, a purely oscillatory function with a zero reference level is required. Motivated by the requirements of purely oscillatory functions, the EMD method was developed by *Huang et al.* [1998]. The HSA and EMD have been applied in the field of hydroclimatology during the last few years [*Xie et al.*, 2002; *Chiew et al.*, 2005; *Huang and Wu*, 2008; *McMahon et al.*, 2008; *Pegram et al.*, 2008; *Peel et al.*, 2009; *Lee and Ouarda*, 2010, 2011]. *Barnes* [2007] showed that the instantaneous frequency has the disadvantage of not being robust in the presence of noise and fluctuations. However, *Han and van der Baan* [2011] showed that this disadvantage was not inherited in EMD. The methodology of the HT and EMD is presented in this section.

2.2. EMD Analysis and Sifting Process

[10] EMD is devised to find the different intrinsic modes of oscillations in any data set with different frequencies. An intrinsic mode of oscillation is known as intrinsic mode function (IMF) when it satisfies the following conditions: (1) the number of extrema must be equal to the number of zero crossings or differ from it at most by one, and (2) the mean value of the two envelopes determined by the local maxima and minima must be zero. When it satisfies these conditions, an IMF is a pure oscillatory mode that bears amplitude and frequency modulations. In EMD, any complex data set can be decomposed into a finite number of IMFs through the sifting process. The sifting process to obtain IMFs from a time series $x(t)$ where $t = 1, \ldots, N$ is as follows.

[11] 1. Identify all of the local extrema and connect all local maxima (minima) with a smoothing technique to obtain the upper (lower) envelope. A cubic spline [*Press et al.*, 2002] has been commonly used [*Huang and Wu*, 2008].

[12] 2. Obtain the first component, h, by finding the difference between the data and the local mean of the upper and lower envelopes m as $h = x - m$.

[13] 3. Substitute x by h and repeat steps 1 and 2 until a certain criterion in which the component (h) is guaranteed to retain enough physical sense of both amplitude and frequency modulations is met [*Huang and Wu*, 2008].

[14] 4. Assign the final h as the jth IMF, c_j, and the residue is r_j (i.e., $r_j = r_{j-1} - c_j$ where $r_0 = x$).

[15] 5. Repeat steps 1–4 by treating the residue r_j as the original data until the final residue becomes a monotonic function. The final residue (r_n) becomes c_{n+1}.

[16] Finally, the original time series, $x(t)$, is presented as the summation of the estimated IMFs as

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

Lower-order components have a higher frequency level and vice versa. For example, the highest frequency level component is c_1 and the lowest frequency level component is c_{n+1}.

[17] To examine whether an IMF obtained from EMD contains a true signal or just a white noise component, *Wu and Huang* [2004, 2005] developed a statistical significance test for IMFs. The numerical experiments of *Wu and Huang* [2004] reveal that

$$\ln E_j + \ln T_j = 0, \tag{2}$$

where E_j and T_j are the spectral energy and the mean oscillation period calculated from the jth IMF, respectively. This test compares the spectral energy and the mean period relation between the IMFs of the original signal and the white noise. If the IMF energy of the observed data with a certain mean period is located above the confidence level, the corresponding IMF is considered statistically significant at the given level.

[18] The ensemble EMD (EEMD) approach was also devised by *Wu and Huang* [2009] by perturbing the observed data and ensembling each IMF component of the perturbed data in order to reduce the chance of mode mixing and preserve the dyadic property. EEMD is employed in the current study.

2.3. Hilbert Transform (Hilbert Spectral Analysis)

[19] The HT [*Frederick*, 2009] is considered as the convolution of $x(t)$, a time series data, with the function $g(t) =$

$1/(\pi t)$. Because of the nonintegrable property of $g(t)$, the HT of $x(t)$ (i.e., the complex conjugate $y(t)$) is presented with the Cauchy principal value (PV) [*Kanwal*, 1996] as

$$\begin{aligned} y(t) &= \mathrm{PV}\int_{-\infty}^{\infty} x(\tau)g(t-\tau)d\tau = \frac{1}{\pi}\mathrm{PV}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau \\ &= -\frac{1}{\pi}\lim_{\varepsilon\to 0}\int_{\varepsilon}^{\infty}\frac{x(t+\tau)-x(t-\tau)}{\tau}d\tau. \end{aligned} \tag{3}$$

The analytical signal $z(t)$ is derived from $x(t)$ and $y(t)$ as

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta}(t), \tag{4}$$

where $a(t) = \sqrt{x^2(t)+y^2(t)}$ is the instantaneous amplitude, $\theta(t) = \tan^{-1}(x(t)/y(t))$ 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 $\omega(t)$ and amplitude $a(t)$ 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, denoted as

$$M(\omega) = \int_0^N H(\omega, t)dt, \tag{5}$$

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

$$\mathrm{IE}(t) = \int_{\omega} H(\omega, t)d\omega. \tag{6}$$

3. Description of the Proposed Model

[20] A modification of the NSO resampling model (NSOR) with EMD for the stochastic simulation is presented in the current study, following the previous prediction work of *Lee and Ouarda* [2010, 2011]. The current simulation study differs from the previous work in several aspects including the work objectives and the initial condition. Since the previous work was focused on the extension of the current series to obtain the future evolution, the initial condition was considered as the most recent observed value. On the other hand, the objective of the current study is to build a simulation model that reproduces the statistical characteristics of the observed data. Therefore, the initial condition should be randomly selected. Here, the initial condition is randomly selected from one of the observations (i.e., bootstrapping). After generating a longer sample, the warm-up period is deleted to avoid any potential initial bias.

[21] The overall procedure of the proposed model and the specific algorithm of NSOR are presented in this section

3.1. Overall Stochastic Simulation Procedure

[22] The proposed model is briefly described as follows: (1) Decompose the time series $x(t)$ into a finite number of IMFs. (2) Find the significant components among them by using the EMD significance test [*Wu and Huang*, 2004]. (3) Fit stochastic time series models to the selected significant IMF components and the residuals (i.e., excluding the significant components from the observed time series, $x(t)$). (4) Simulate each IMF component as well as the residuals from the fitted models. (5) Sum up the simulated components.

[23] The procedure of stochastic simulation utilizes the change rate of an oscillatory time series, defined as $\Delta c(t)/\Delta t$, instead of the direct modeling component $c(t)$. This avoids abrupt changes while reproducing a smooth oscillation and easily adapts to the previous conditions. Since observations are discretely measured with the same interval (e.g., seconds, hours, and years), i.e., $\Delta t = 1$, we therefore have

$$\begin{aligned} \Delta c(t)/\Delta t &= \{c(t) - c(t-1)\}/(t-(t-1)) \\ &= c(t) - c(t-1) = \Delta c(t). \end{aligned} \tag{7}$$

[24] Three different types of models are used for simulating the NSO components, the trend component, and the residuals. First, the significant oscillation components are modeled with the NSOR model as explained in section 3.2. Second, the last trend component should be modeled only when the component is proven to be highly significant. This is because the trend component might be induced by the inability of the EMD procedure to efficiently capture the IMF components [*Lee and Ouarda*, 2010], and also because the current trend may be defined only for the time span of the observed data [*Wu et al.*, 2007; *Huang and Wu*, 2008]. Third, the residuals of the data excluded from the significant components are treated as either random noise or autocorrelated noise according to their time dependency structure. A proper stochastic simulation model for the residuals should be fitted, such as bootstrapping [*Lall and Sharma*, 1996; *Vogel and Shallcross*, 1996], ARMA [*Salas*, 1993; *Brockwell and Davis*, 2003], etc. In the current study, the KNNR model [*Lall and Sharma*, 1996] is used for the residuals similarly to *Lee and Ouarda* [2010, 2011].

3.2. NSOR

[25] The NSOR model is based on two nonparametric techniques (i.e., KNNR and bloc bootstrapping). A brief description of the stochastic simulation procedure of the NSOR model is presented here. Suppose that we have a sequence of a certain IMF component $c(t)$ where $t = 1, \ldots, N$. The superscripts H and G are used to represent the historical and generated data, respectively. The NSOR simulation procedure is as follows.

[26] 1. Select an initial value $c^G(0)$ from the N historical observations assuming that each observation has equal probability of being chosen.

[27] 2. Randomly generate a block length, L_B, from a discrete distribution (e.g., Poisson or Geometric). A Poisson distribution is generally selected [*Lee*, 2008; *Lee and Ouarda*, 2010, 2011] because the distribution shape is close to a normal distribution centered on the large mean (here L_B). The employed Poisson distribution with parameter τ is $L_B \sim e^{\tau}\tau^{k-1}/(k-1)!$ where k is a positive integer value. For the estimation of the parameter τ, the method proposed by *Wilks* [1997] and *Lee and Ouarda* [2011] is used. It is

estimated with a function of the variance inflation factor (VIF), and the record length, N, by assuming that the data follows a second-order autoregressive model. *Lee and Ouarda* [2011] tested this method and checked its feasibility to the NSOR model.

[28] 3. Estimate the distances as follows:

$$D_j = \sqrt{\alpha_1\{c^G(t-1) - c^H(j-1)\}^2 + \alpha_2\left(\Delta c^G(t-1) - \Delta c^H(j-1)\right)^2}, \tag{8}$$

where α_1 and α_2 are the inverse variances of the component data c and the change rates Δc respectively, and $j = 2, \ldots, N - L_B$. The last L_B elements are excluded to ensure that the selected points are followed by records at least L_B in length.

[29] 4. Arrange the distances in ascending order and select the first k values. Next, out of these k values, randomly select one with the weighting probability given as $w_i = (1/i)/\sum_{j=1}^{k} 1/j$ where $i = 1, \ldots, k$. Assign the corresponding time index of the selected one as $\tilde{p}$. Here the number of nearest neighbors (k) is estimated using the heuristic approach (i.e., $k = \sqrt{N}$) proposed by *Lall and Sharma* [1996].

[30] 5. Obtain the generated data with length L_b using

$$c^G(l) = c^G(l-1) + \Delta c^H(\tilde{p} + l) \qquad l = 1, \ldots, L_B,$$

where $\tilde{p}$ is the selected time index from step 4.

[31] 6. Repeat steps 2–5 until all the required data are generated.

[32] Note that in the applications carried out in the present study, one tenth of the simulated records are additionally generated and deleted as part of the warm-up period.

4. Model Validation

[33] A total of three synthetic data sets (i.e., two data sets representing combinations of trigonometric functions and the Rössler system) were employed in order to validate the performance of the proposed model.

4.1. Validation With the Combination of Trigonometric Functions

4.1.1. Methodology

[34] A synthetic data set is obtained from the combination of trigonometric functions as

$$f(t) = \cos(0.05t) + \sin(0.02t) + 0.5\varepsilon(t), \tag{9}$$

where $\varepsilon(t)$ is the random component which follows a standard normal distribution. The employed data set realized from equation (9) with the record length $N = 1000$ is illustrated by circles in Figure 1 (top). The three terms were intentionally parameterized so that two oscillation components and one random component are appropriately mixed. The sine component, $\sin(0.02t)$, has about 300 time units per cycle while the cosine component, $\cos(0.05t)$, has about 100 time units per cycle. The sine component has a longer time cycle than the cosine component. The time series obtained by the combination of these two components (thick solid line in Figure 1, top) illustrates the oscillation pattern of the time series $f(t)$. The random component is included in order to perturbate the combination of the trigonometric functions.

[35] A second trigonometric function including an overall trend has been tested with two highly different modulations as:

$$f(t) = \cos(\pi t/180) + \sin(10\pi t/180) + 0.2\varepsilon(t) + 2 \times 10^{-6} t^2. \tag{10}$$

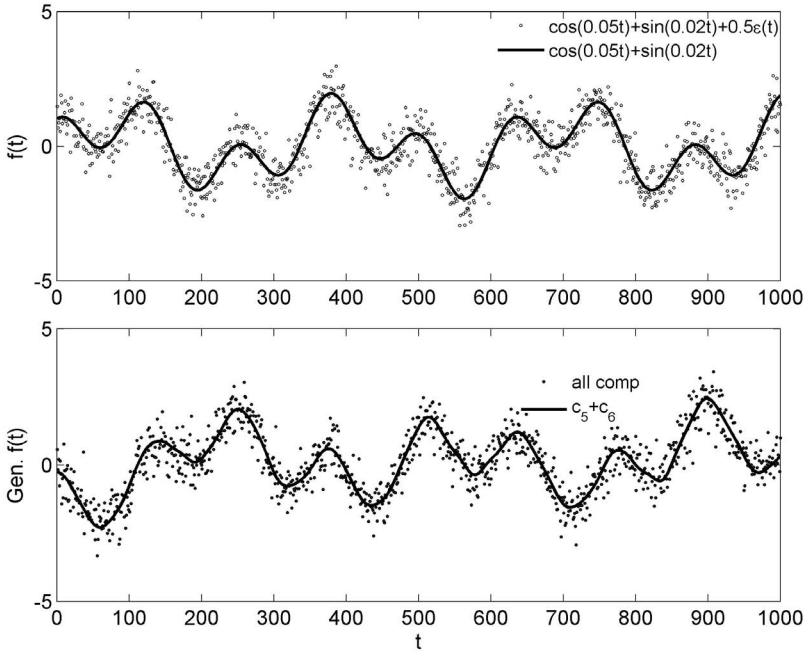

Figure 1. Time series of (top) the realized data from $f(t)$ in equation (9) and (bottom) an example of the generated data from the proposed nonstationary oscillation resampling (NSOR) model.

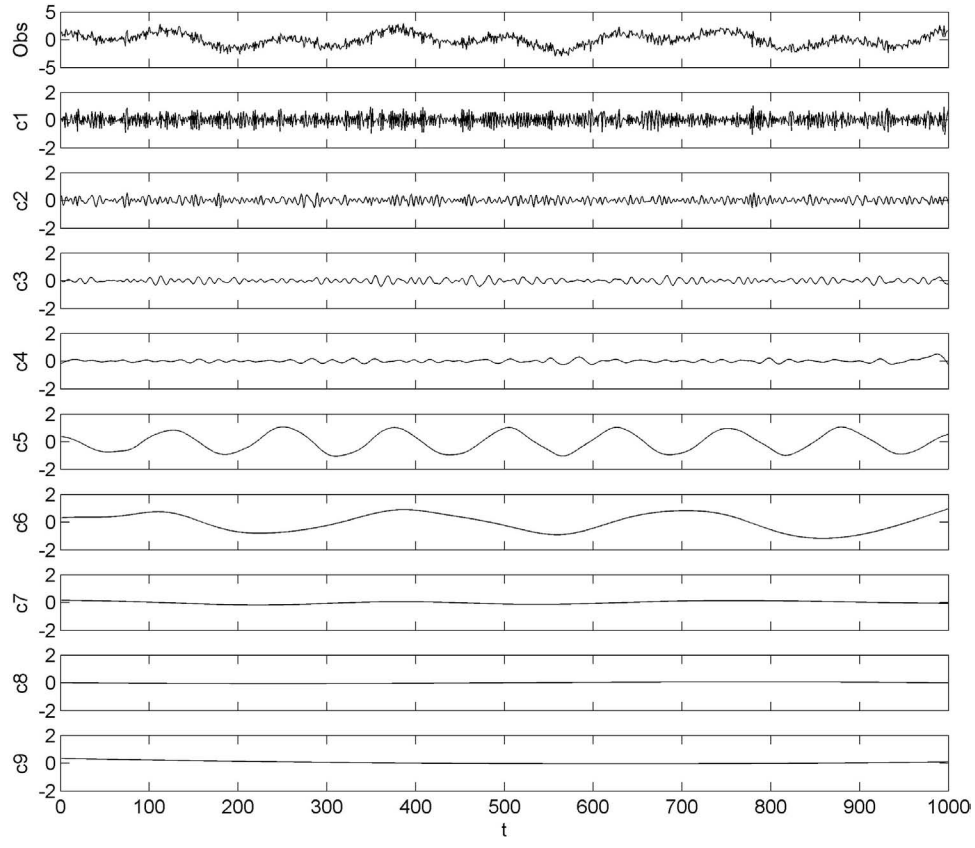

Figure 2. Realized sequence from $f(t)$ in equation (9) and the extracted intrinsic mode function (IMF) components. Note that most of the cosine and sine signals in $f(t)$ are sifted in c_5 and c_6.

This represents the case where long-term oscillations (e.g., seasonal cycle) are modulated by short-term oscillations (e.g., diurnal cycle) while a significant trend exists. This trigonometric function illustrates hence a case that is commonly encountered in hydrologic and climatic series. The same procedure for the previous case (i.e., equation (9)) is taken in order to reproduce the oscillations and the trend. Here, the historical trend extracted from EMD analysis is used in simulation instead of fitting a polynomial regression. Note that it is expected that the second-order polynomial model leads to the best fit for equation (10).

4.1.2. Results

[36] Nine IMF components were extracted from the realized data of $f(t)$ in equation (9) using EMD analysis as shown in Figure 2. The fifth and sixth components contain most of the information about $\cos(0.05t)$ and $\sin(0.02t)$, respectively. The other components might be the random part of $f(t)$ or the inability of the EMD to efficiently capture the IMF components, as discussed in section 3.

[37] The significance test of the IMF components [*Wu and Huang*, 2004, 2005] reflects well the importance of each component as shown in Figure 3. In Figure 3, the points (asterisks) show the energy magnitude with a certain mean period of each component. If the point is located above a line indicating a certain significance level (e.g., solid and dotted lines for 95% and 99% significance levels, respectively), the hypothesis that the component is induced by random noise is rejected with the corresponding significance level. The 5th and 6th components are highly significant while the 7th, 8th, and 9th components are less than moderately significant. The last three components

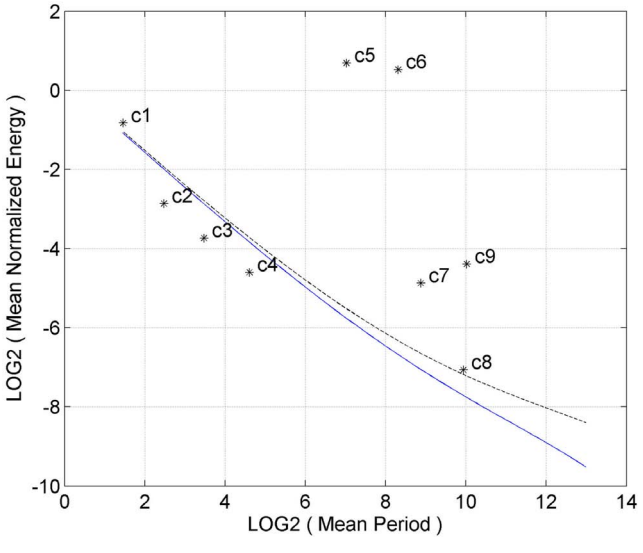

Figure 3. Significance test of the IMFs in Figure 2. The points (asterisks) below the lines indicate the acceptance of the hypothesis that the corresponding IMF of the target series is not distinguishable from the corresponding IMF of a random noise series with 95th (solid line) and 99th (dotted line) confidence levels.

(i.e., c_7, c_8, and c_9), however, present a very small variability, as shown in Figure 2, with almost no contribution to the variability of the original time series (first panel in Figure 2). Therefore, these components are not included in the NSOR modeling. The shortcomings of the IMF significance test suggested by *Wu and Huang* [2004, 2005] are discussed by *Lee and Ouarda* [2010]. In the latter, it is shown that subjective judgment through visual illustration (e.g., Figure 2) is also required for the selection of NSOR modeling components. Only the fifth and sixth components are selected for NSOR modeling while the sum of the other components is treated as residual. The residuals were modeled with KNNR as it can preserve nonlinear serial relations as well as linear relations [*Lall and Sharma*, 1996; *Prairie et al.*, 2005, 2006; *Salas and Lee*, 2010].

[38] Employing the modeling scheme presented above, 200 sequences were generated with the same record length as the observed data (i.e., $N = 1000$). A sample sequence is presented in Figure 1 (bottom). The overall oscillation pattern induced from the combination of the two trigonometric functions as well as the variability of the random component is reproduced well in the generated data.

[39] The basic statistics (i.e., mean, standard deviation, and skewness) of the 200 generated sequences were estimated and compared to the statistics of the observed data as presented in the three box plots in Figure 4 (top). In these box plots, boxes display the interquartile range (IQR) and whiskers extend to the extrema with horizontal lines at the 5th and 95th percentiles of the serial correlations of the 200 generated sequences. The horizontal lines inside the boxes depict the median of the data. Also, the value of the statistic corresponding to the observed data is represented with a circle. The box plots of the three basic statistics indicate that the proposed model reproduces well the statistical characteristics of the observed data.

[40] The serial correlations were compared for the observed and generated data from the proposed model as shown in Figure 4 (bottom). The solid line represents the observed correlations while the dash-dotted line and the two dotted lines illustrate the mean and the 10th and 90th percentiles, respectively. The observed serial correlations are located within the 10th and 90th percentiles of the generated series indicating that the proposed model characterizes well the dependence structure of the observed data as well as the long-term oscillation property. Overall, the proposed NSOR model reproduces well the key characteristics of the test function. The marginal Hilbert spectrum of the historical data (see equation (5)) is well preserved in the simulated data, which indicates that the NSOR model has a strong capability to reproduce the spectral characteristics of the historical data (data not shown).

[41] The realized data set for the second trigonometric function (see equation (10)) is shown in Figure 5 (top), and one example of the simulated data is presented in Figure 5 (bottom). The results show that the combination of the two highly different frequency oscillations and the trend component are reproduced well in the simulated data using the proposed NSOR approach. The marginal Hilbert spectra of the realized data and the 200 generated sequences are illustrated in Figure 6 with a thick black line and with thin gray lines, respectively. Figure 6 indicates that the marginal spectrum of the realized data is well preserved in the simulated data. Also, the results indicate that the simulations for equation (10) have a similar reproduction capability for

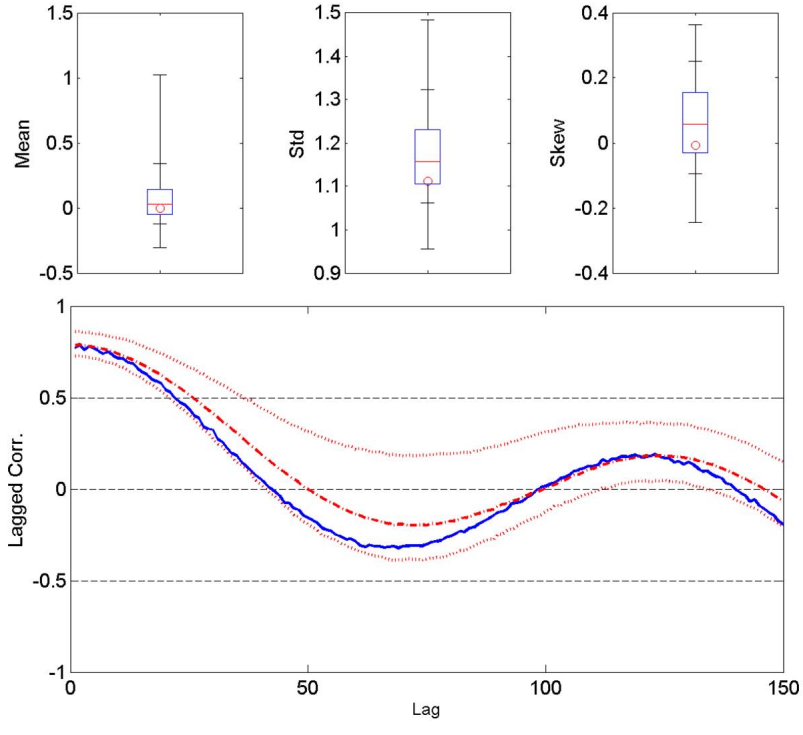

Figure 4. (top) Basic statistics of the generated (box plots) and observed (circles) data; (bottom) serial correlations of the observed (solid line) and generated data (a dashed line for median and two dotted lines for the 5th and 95th percentiles).

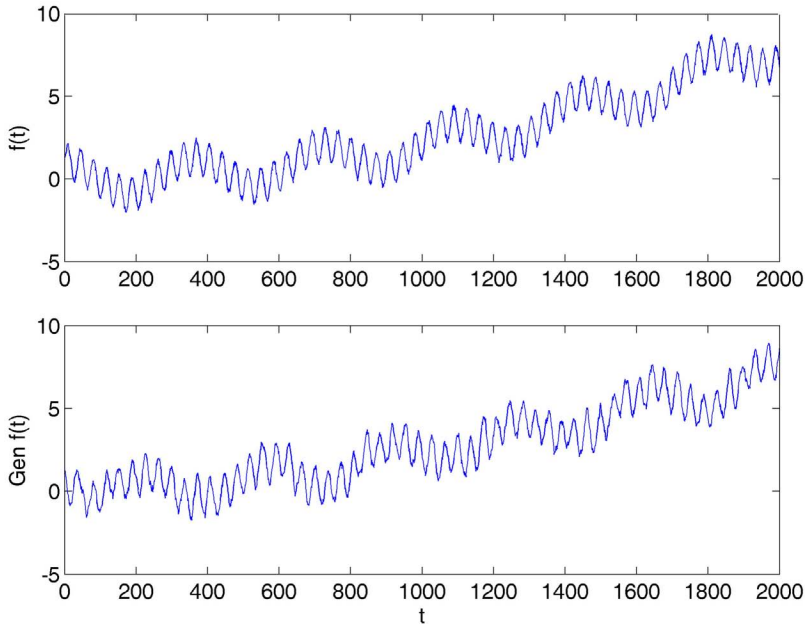

Figure 5. Time series of (top) the realized data from equation (10) and (bottom) an example of the generated data from the proposed NSOR model.

statistical characteristics as in the case of equation (9) (data not shown).

4.2. Validation With the Rössler System

4.2.1. Methodology

[42] One of the most famous and chaotic nonlinear dynamic systems, the Rössler attractor [*Rössler*, 1976, 1995], was selected in order to test the performance of the suggested stochastic simulation model. This attractor was intended to behave similarly to the Lorenz attractor [*Lorenz*, 1963] but with a better qualitative understanding of the chaotic flow. The system is represented by three nonlinear ordinary differential equations as $\dot{p} = -q - s$, $\dot{q} = p + \alpha q$, and $\dot{s} = \beta + ps - \delta s$. Here $(p, q, s) \in \Re^3$ and $(\alpha, \beta, \delta) \in \Re^3$ are dynamical variables and parameters, respectively, and $\dot{p}$ represents the derivative of the variable x. *Huang et al.* [1998] and *Kijewski-Correa and Kareem* [2007] showed that the nonlinear oscillatory characteristics of this system are well described with EMD and HT. This attractor is selected since it oscillates within a fixed range but the oscillations are chaotic, which is the general assumption with which the NSOR model was applied [*Lee and Ouarda*, 2010, 2011].

4.2.2. Application and Results

[43] The system was realized with the same parameter set (i.e., $[\alpha, \beta, \delta] = [1/5, 1/5, 7/2]$) as in the work of *Huang et al.* [1998] and *Kijewski-Correa and Kareem* [2007], $N = 500$ and with the initial state of $[p_0, q_0, s_0] = [-3, 3, 1]$. Among the three variables p, q, and s, we used the p variable as $p + \varepsilon$ by perturbating the system with a random component. Here ε is normally distributed with zero mean and the same variance as p. Through the significance test and its implication [*Wu and Huang*, 2004, 2005; *Lee and Ouarda*, 2011], the combination of the fourth and fifth IMF components among eight IMFs was selected to fit the NSOR model. The other components were considered as the residuals and modeled with the KNNR model. From the fitted model, 200 sequences were generated with the same record length as the observed data.

[44] Hilbert spectrum analysis is applied to the realized data and the generated sequences to validate the simulation performance of the proposed model in the spectral domain. The results are presented in Figures 7, 8, and 9. The results for the observations and one example of the generated sequences are presented in Figures 7 and 8, respectively, for the time series, the time-frequency spectrogram of $H(\omega, t)$, and the marginal Hilbert spectrum of equation (5).

[45] The overall variability of the generated time series (Figure 8a) is similar to the variability of the realized data

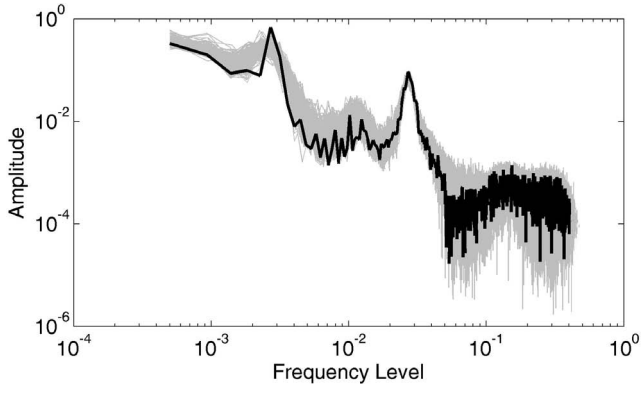

Figure 6. Marginal Hilbert spectrum for the realized data of the test function in equation (10) (thick black line) and 200 generated sequences from the NSOR model (thin gray lines). The first peak of the frequency level ($\sim 2 \times 10^{-3}$) indicates the short-term oscillation in equation (10), while the second peak of the frequency level ($\sim 2 \times 10^{-2}$) shows the long-term oscillation.

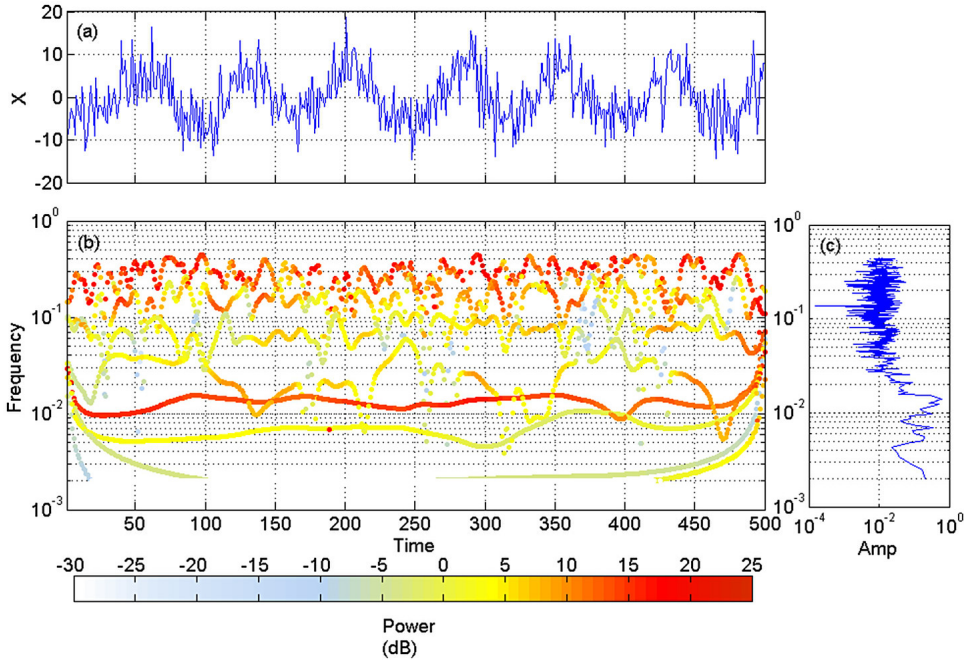

Figure 7. Hilbert transform of the realized Rössler attractor for (a) time series, (b) time-frequency spectrogram of $H(\omega, t)$, and (c) marginal Hilbert spectrum of equation (5).

(Figure 7a). For the spectrogram of the realized data shown in Figure 7b, there are two frequency regions, i.e., 10^{-1} to 5×10^{-1} and 3×10^{-2} to 8×10^{-3} where the spectrum magnitude is consistently high over the entire time span. As shown in Figure 7c, the amplitude of the marginal Hilbert spectrum for the lower-frequency region (i.e., 3×10^{-2} to 8×10^{-3}) is also high, while for the higher-frequency region (i.e., 10^{-1} to 5×10^{-1}) it is low. Note that the components that produce low spectrum amplitude in the high-frequency level (i.e., 10^{-1} to 5×10^{-1}) are

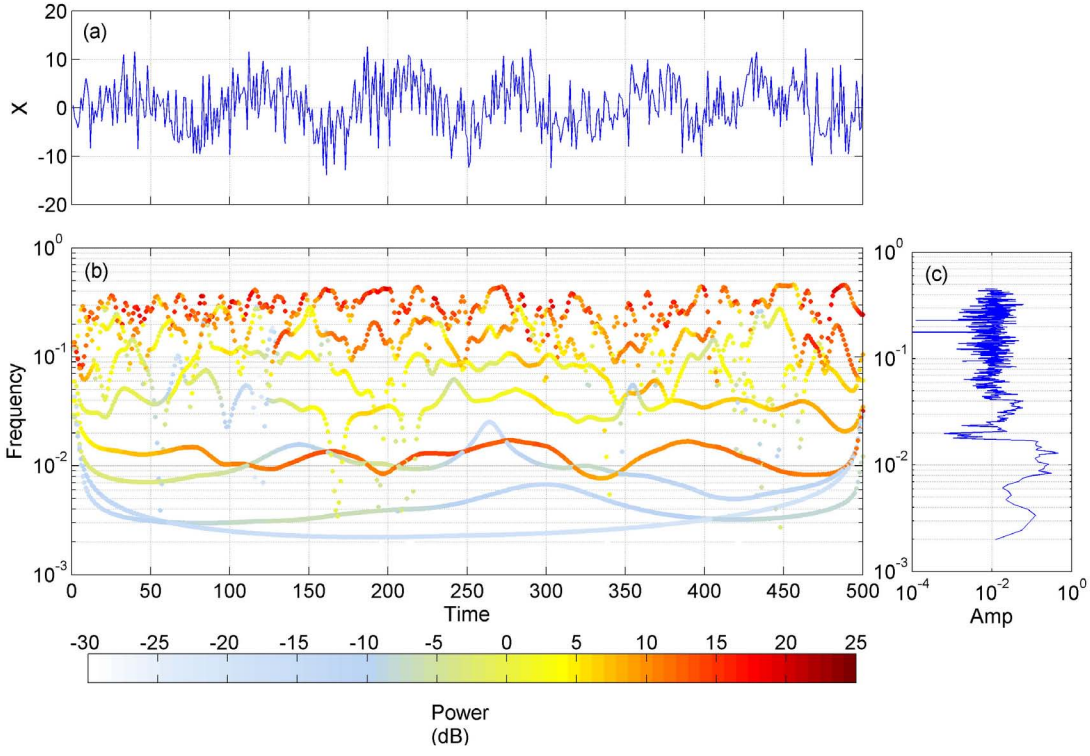

Figure 8. The same as Figure 7, except the employed data are an example of the generated sequences from the proposed NSOR model.

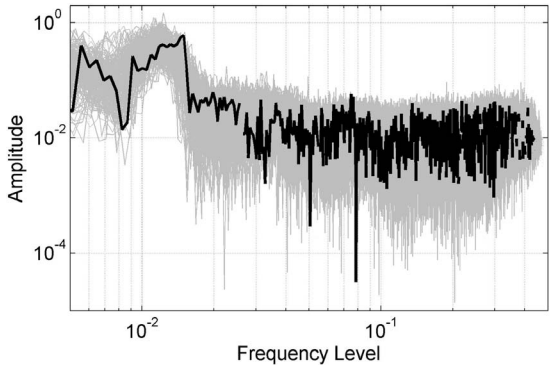

Figure 9. Marginal Hilbert spectrum for the realized data of Rössler attractor (thick black line) and 200 generated sequences from the NSOR model (thin gray lines). Note that at a certain frequency level, the corresponding amplitude might be zero since the amplitude is estimated from the instantaneous frequency of IMFs (e.g., frequency level 2.5×10^{-2}).

low-order IMFs (e.g., c_1 and c_2). Therefore, the frequency level of these components is highly sparse because of the randomness of the lower-order components.

[46] These spectral characteristics of the HT in the realized data are reproduced well in the HT of the example of the generated data as shown in Figures 8b and 8c. The consistent high amplitude of the realized data is also observed in the HT of the example of the generated data, shown in Figure 8b as well as the marginal Hilbert spectrum in Figure 8c.

[47] The marginal Hilbert spectra of the realized data and all the generated sequences are illustrated in Figure 9. The amplitude of the realized data resides within the variability of the amplitude of the generated sequences over the whole frequency level. This indicates that the generated sequences from the proposed model preserve well the spectral characteristics of the realized data. Also, the key statistics and serial correlations of the realized data are reproduced well in the generated sequences (data not shown).

5. Case Studies

5.1. North Atlantic Oscillation Index

5.1.1. Data Description

[48] The North Atlantic Oscillation (NAO) is a large-scale oscillation in atmospheric mass between the subtropical high (the Azores High) and the polar low (the Icelandic Low) in the North Atlantic region [*Rogers*, 1990]. A zonal average version of the NAO index was introduced by *Li and Wang* [2003] and employed in the current study instead of the instrumental record-based indices [*Hurrell*, 1995; *Jones et al.*, 1997; *Cullen et al.*, 2002]. This index, estimated from the difference of the monthly sea level pressure (SLP) at 35°N and 65°N over the longitudes of 80°W–30°E, is employed in the current study for the period 1873–2007 as shown in Figure 10. It has been proven that this index explains a large portion of the variance of the SLP over the North Atlantic region for all time scales and provides the strongest correlation pattern with surface air temperature among the NAO indices. The data was downloaded from the site of the State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (http://www.lasg.ac.cn/staff/ljp/data-NAM-SAM-NAO/NAO.htm).

5.1.2. Results

[49] On the basis of the IMF significance test, the fourth and fifth IMF components were selected among eight IMFs. The time series of the observed NAO index and the combination of the selected components are shown in Figure 10. Even if the serial correlations of the observed data are relatively small (Figure 11a), the pattern shows that the long-term oscillation does exist in the observed data. The serial correlations of the residuals, obtained by subtracting the selected components from the observations, do not show this oscillation pattern (Figure 11b). This indicates

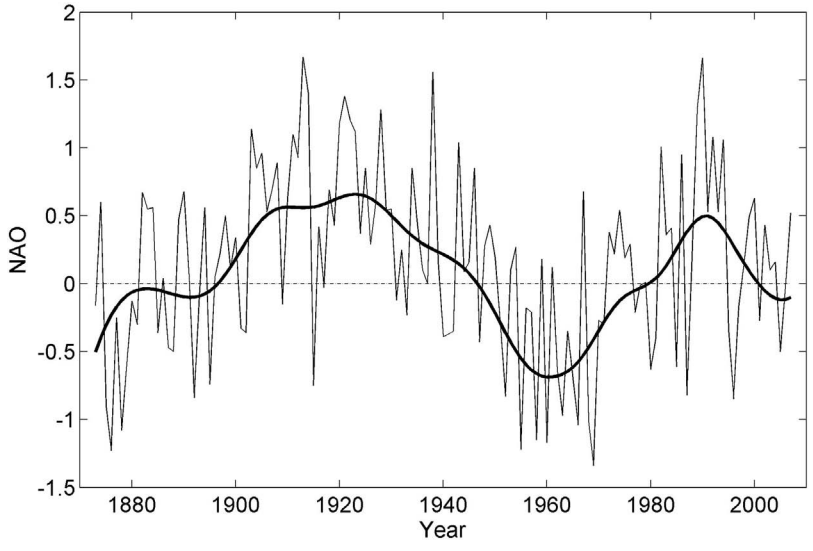

Figure 10. Time series of the North Atlantic Oscillation (NAO) index (thin solid line, years 1873–2007) and its selected components (thick solid line).

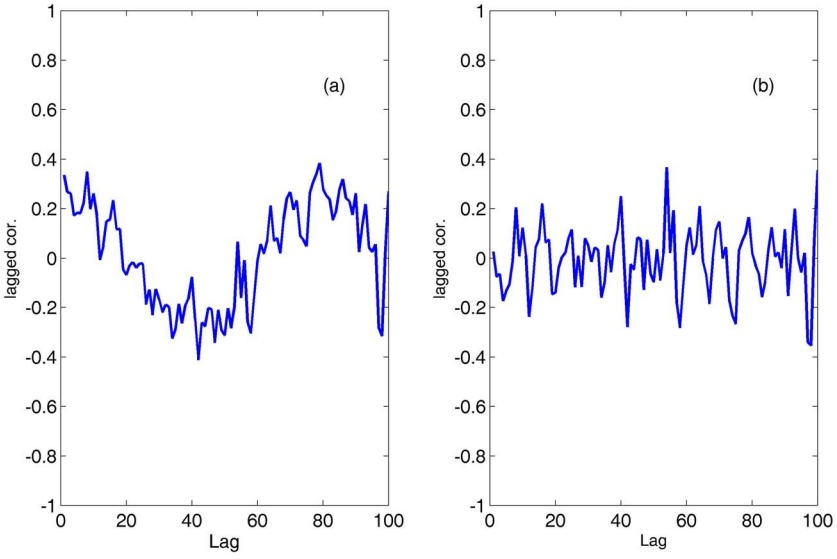

Figure 11. Serial correlations of (a) the observed data and (b) the residuals obtained by subtracting the selected IMF components from the observed data for the NAO index.

that the selected components contain the long-term NSO of the observed data as discussed by *Lee and Ouarda* [2010, 2011] and *Schlesinger and Ramankutty* [1994].

[50] As discussed in section 1, the SML model [*Salas and Boes*, 1980; *Sveinsson et al.*, 2003] is one of the commonly adopted alternatives to represent long-term nonstationary processes as abrupt shifts. The SML model was applied to the NAO index for the comparative study with the proposed NSOR model.

[51] The basic annual statistics of the annual NAO index are presented with box plots for the NSOR and SML model results in Figures 12 and 13, respectively. The key statistics

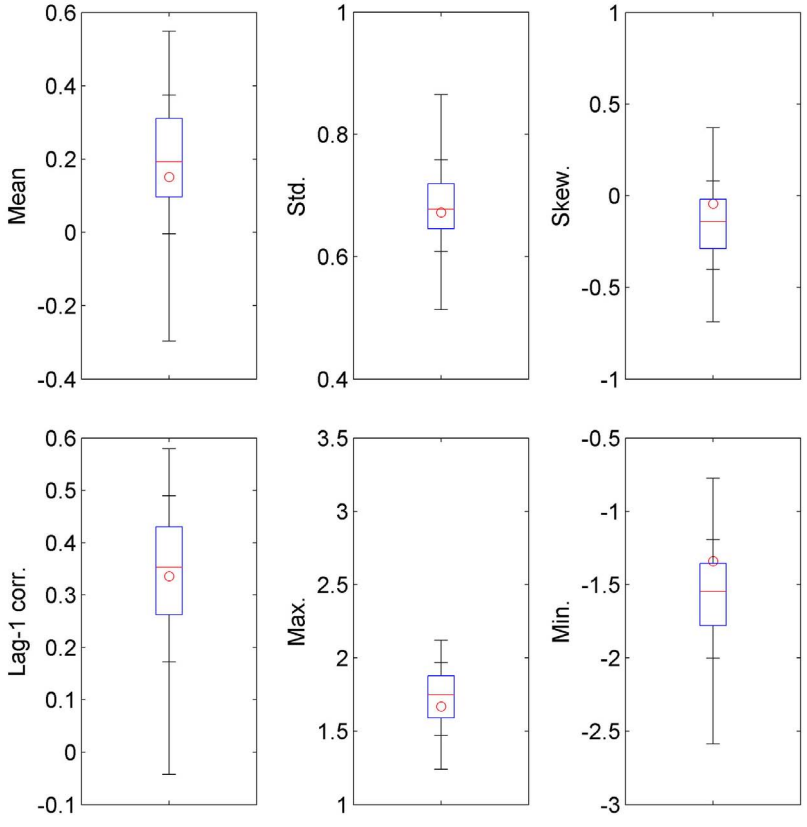

Figure 12. Basic statistics of the observed (box plots) and generated (circles) data from the NSOR model for the NAO index.

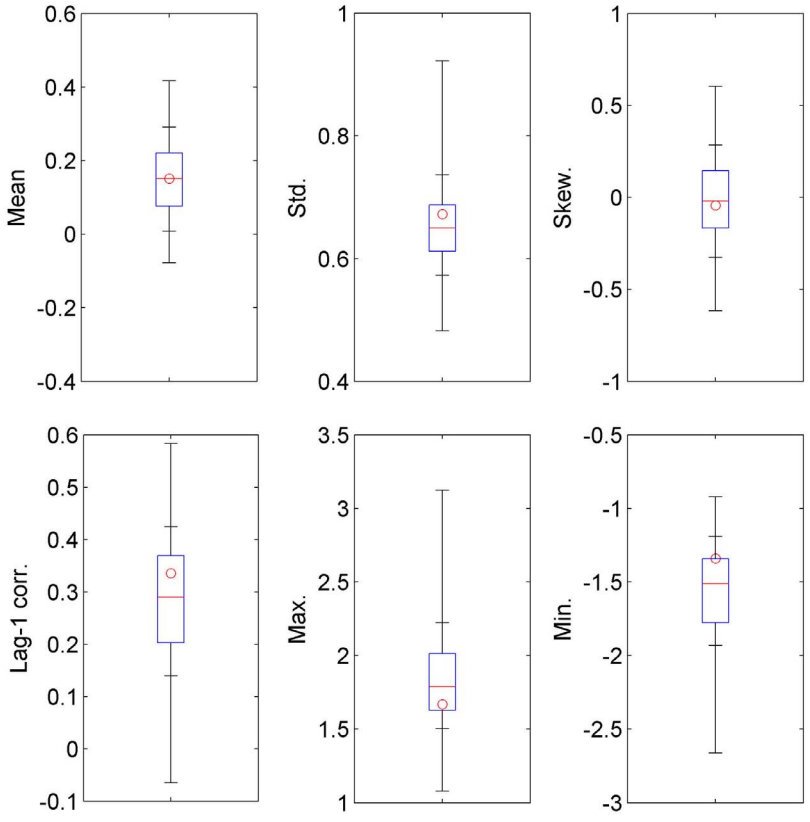

Figure 13. Basic statistics of the observed (box plots) and generated (circles) data from the shifting mean level (SML) model for the NAO index.

are well preserved in both models. The statistics of the mean, standard deviation, and lag 1 serial correlation are supposed to be preserved in the SML model since the model employs these statistics for parameter estimation. Even if none of these statistics are employed as parameters in the proposed NSOR model, it also preserves these statistics. Nonparametric models are intended to let observed data speak for themselves instead of parameterization with the assumption of a certain distribution such as the normal. Slight underestimation of the minimum is observed in both models. No significant difference can be found between the two models. The densities of the simulated data from the NSOR model and the SML model are compared using kernel density estimates [*Silverman*, 1986], as shown in Figure 14. The figure shows that both models reproduce the historical density well.

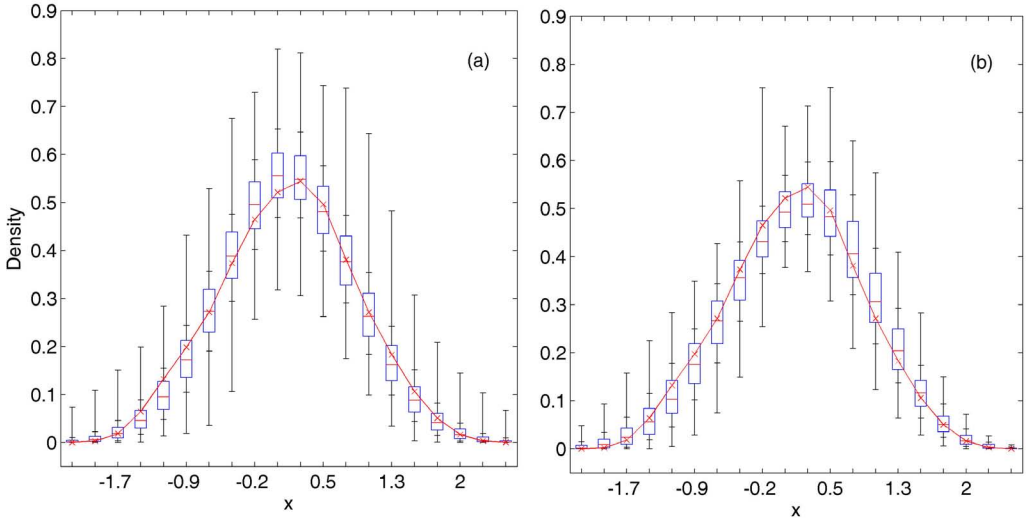

Figure 14. Densities of the observed (box plots) and generated (solid line with crosses) data from (a) the NSOR and (b) the SML models for the NAO index.

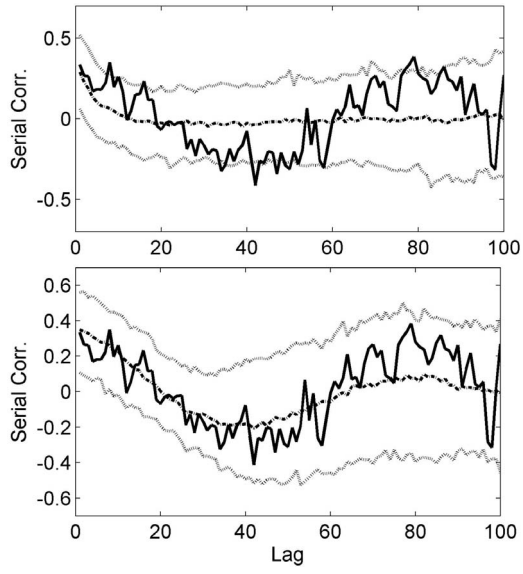

Figure 15. Serial correlations of the observed and the generated data for (top) the SML model and (bottom) the NSOR model (observed, solid line; mean value, dash-dotted line; 5th and 95th percentiles, dotted lines) for the NAO index.

[52] In Figure 15, the serial correlations of the observed data are represented with a solid line, while the ones of 200 generated sequences from the SML model (Figure 15, top) and the NSOR model (Figure 15, bottom) are represented with a dash-dotted line (mean) and two dotted lines (5th and 95th percentiles). The serial correlations of the generated sequences from the two models show significant differences. The correlations of the SML model decrease exponentially, while the correlations of the observed data (the segment line) show a long-term NSO pattern, as shown in Figure 15 (top). In contrast, this long-term NSO pattern of the serial correlations is well preserved in the NSOR model, as shown in Figure 15 (bottom). The long-term NSO presented in the serial correlations cannot be preserved through the SML model since this model has an exponentially decreasing correlation structure [*Sveinsson et al.*, 2003].

[53] A behavior similar to the serial correlation structure is also presented by the marginal Hilbert spectrum, as shown in Figure 16. The high spectrum amplitude of the observed data at the low-frequency level around 10^{-2} is underestimated from the generated sequences of the SML model (Figure 16a) while the same is preserved well in the sequences of the NSOR model (Figure 16b).

5.2. Romaine River Streamflows

5.2.1. Data Description and Methodology

[54] The Romaine River station is located at 50.35°N, 61.19°W in the province of Quebec, Canada with a drainage area of 13,000 km^2. Recently, Hydro-Quebec has indicated that it plans to build a hydroelectric complex with a capacity of 1500 megawatts (MW) on this river. The mean annual streamflows of the Romaine River station are simulated in association with the NAO index in order to illustrate the potential usefulness of the generated climate index in the simulation of the hydrological regime.

[55] Stochastic simulation of streamflows associated with a climatic system provides better insights for the future conditions of water availability. However, the relatively short record length (1957–2008) of the Romaine River observations, as shown in Figure 17a, hinders the capacity to observe and capture long-term variations such as shifting means and NSO processes. A stochastic simulation model of streamflow data incorporating the long-term variations of the appropriate climate index is a useful

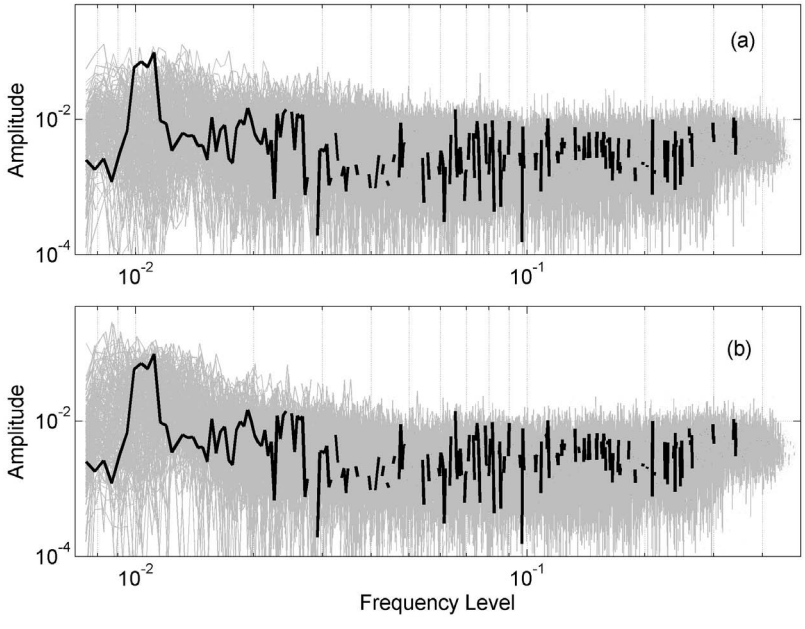

Figure 16. Marginal Hilbert spectrum for the observed data of the NAO index (thick black line) and 200 generated sequences (thin gray lines) from (a) the SML model and (b) the NSOR model for the NAO index.

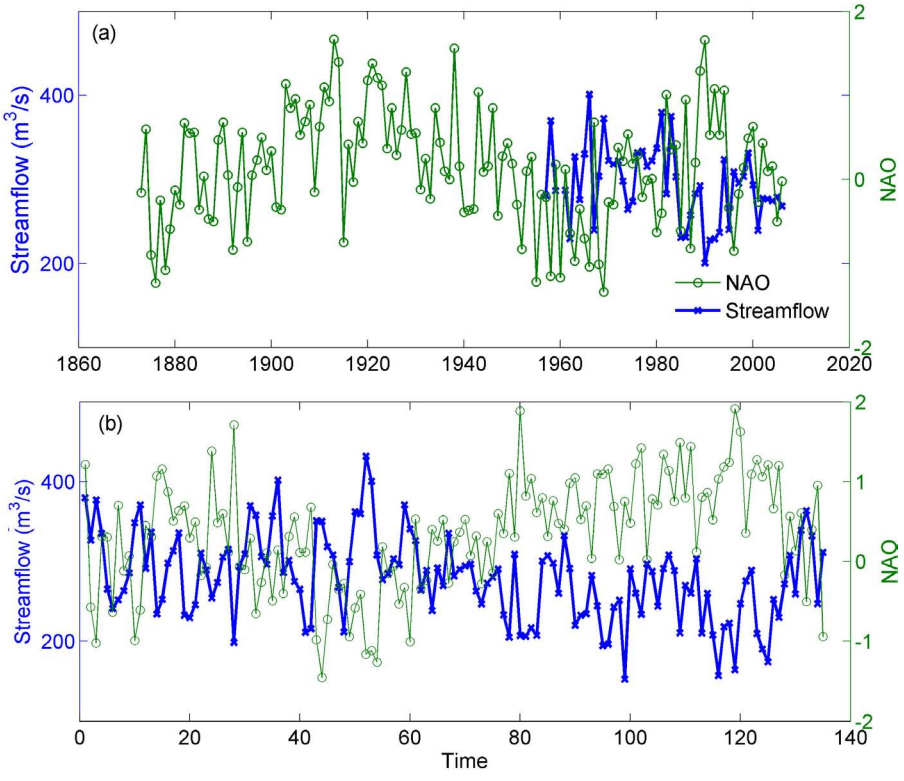

Figure 17. Time series of the NAO index and Romaine River streamflows for (a) observed data and (b) an example of the generated sequences.

alternative in this case. The connection between the NAO index and the Romaine river streamflows was explored by *Lee and Ouarda* [2010]. The opposite long-term phase between the Romaine River and the NAO index can be observed in Figure 17a. A significant negative cross correlation (−0.44) is shown between the NAO index and the Romaine River streamflow.

[56] Since no significant nonstationarities were observed from visual inspection and statistical testing, an autoregressive model with exogenous input (ARX) [*Hipel and McLeod*, 1996] was applied in simulating the Romaine River streamflows. The order 1 for the autoregressive (AR) term and the order 1 for the exogenous term are used. Since the record length is short and no long-memory structure is observed in the streamflow data, more complex ARX models were not considered. A nonlinear classical ARX model [*Nelles*, 2001] was also tested as a surrogate of the ARX model. The results were not satisfactory because of overfitting of the short record length of the streamflow station.

5.2.2. Results

[57] To test the performance of the proposed NSOR model, 200 sequences were generated with the same record length as the NAO index (135 years). The simulation records are longer than the historical data and enable the generated streamflow data to contain long-term patterns. The opposite long-term phases between the streamflow and the NAO index, embedded in the observations (Figure 17a), are reproduced well in the simulated data, as shown in Figure 17b. Note that during the period when only the NAO index is available (i.e., 1875–1957) the index is on a warming phase. The expected streamflow values during this period might be somewhat lower than normal because of the negative cross correlation with the NAO index. Therefore, the mean of the simulated streamflows is expected to be lower than the mean of the observed data if the record is extended up to the year 1875. The mean of the simulated data is lower than the observed one, as presented in Figure 18. This behavior was expected since the simulation of the streamflows employs the long-term oscillation of the NAO index. For the same reason, the minimum of the simulated data is also lower than the minimum of the observed data. Slight underestimation of the skewness and overestimation of the lag 1 correlation are observed in Figure 18. In conclusion, the ARX model of the Romaine River annual streamflows employing the simulated NAO index as an exogenous variable reproduces fairly well the key statistics of the observed streamflows as well as the long-term NSO process that is not observed in the historical data.

6. Summary and Conclusions

[58] The stochastic simulation NSOR model, based on EMD analysis and nonparametric techniques, for hydroclimatological data was presented in the current study. The modeling procedure can be summarized as (1) the observed time series is decomposed using EMD into IMFs, (2) the IMF significance test is performed, (3) each component is modeled and simulated according to its characteristics (i.e., NSOR, residuals, and overall trend), and (4) the simulated components are combined.

[59] Validation studies were carried out on the combination of trigonometric functions and the Rössler system. The results indicate that the proposed model reproduces well

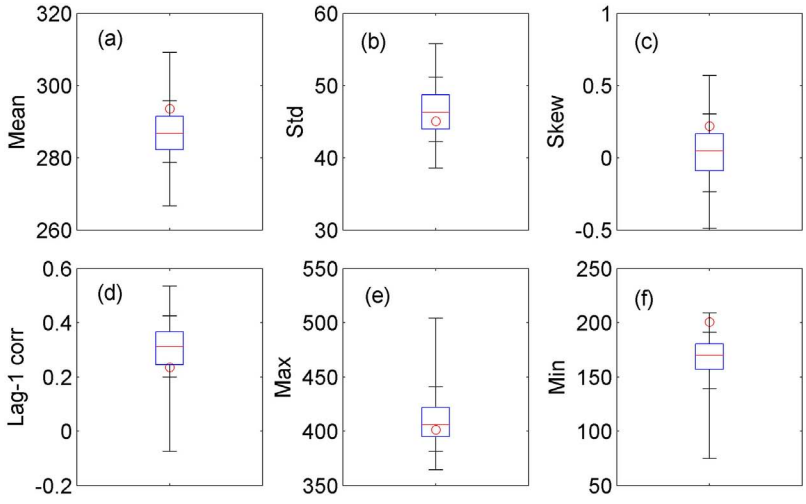

Figure 18. Basic statistics of the Romaine River streamflows. Note that the unit of the observed data is $m^3 s^{-1}$.

the important statistical characteristics and the NSO processes. In the real case study of the NAO index, the proposed model outperforms the existing SML model in reproducing NSO processes. It was shown that the hydrological variables (i.e., streamflows at the Romaine River) can be efficiently generated with the ARX model employing the NAO index as an exogenous variable. The unprecedented long-term NSO processes were illustrated in the simulated Romaine River streamflows assisted by the long records of the NAO index.

[60] From the results of the validation studies and the applications to case studies, we conclude that the proposed NSOR model based on EMD analysis represents a good alternative to simulate hydroclimatological variables containing NSO processes. The results indicate that the NSOR model is useful when a significant NSO process exists in the observed time series. Otherwise, it would be better and easier to apply other time series models, such as SML or ARMA. Furthermore, an elaborate process is still required to find significant components for the EMD process since the current significance test is not definitive yet. Employing the AR(1) time series instead of white noise in the significance test might be a good alternative to improve the selection capability of the significant IMF components.

Notation

$x(t)$ observation variable.
$\varepsilon(t)$ a white noise random process with zero mean.
t time index.
N record length.
m mean of the upper and lower envelopes in sifting process.
h difference between the mean envelopes and the data in the EMD analysis.
c_j extracted jth EMD component (IMF).
n number of IMFs not including the last trend component.
r_j jth residue variable.
E_j mean energy of jth IMF.
T_j mean oscillation period calculated from jth IMF.
PV Cauchy principal value.
$y(t)$ Hilbert transform of x (i.e., complex conjugate).
$a(t)$ instantaneous amplitude.
$\theta(t)$ instantaneous phase function.
$\omega(t)$ instantaneous frequency.
$H(\omega,t)$ Hilbert amplitude spectrum.
$M(\omega)$ marginal Hilbert spectrum.
IE(t) instantaneous energy.
Δc change rate of IMF.
k number of nearest neighbors in the KNNR model.
w_i weighted probability with $i = 1, \ldots, k$.
L_B block length.
D_j distance measurements in the NSOR model.

[61] **Acknowledgments.** The authors acknowledge funding from the National Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair Program. The authors wish also to thank the Editor, Praveen Kumar, and the Associate Editor as well as Upmanu Lall and two anonymous reviewers, whose comments helped considerably improve the quality of the paper.

References

Ahn, J. H., and H. S. Kim (2005), Nonlinear modeling of El Nino/Southern Oscillation index, *J. Hydrol. Eng.*, *10*(1), 8–15.

Barnes, A. E. (2007), A tutorial on complex seismic trace analysis, *Geophysics*, *72*(6), W33–W43.

Bartolini, P., J. D. Salas, and J. T. B. Obeysekera (1988), Multivariate periodic ARMA(1, 1) processes, *Water Resour. Res.*, *24*(8), 1237–1246.

Brockwell, P. J., and R. A. Davis (2003), *Introduction to Time Series and Forecasting*, 2nd ed., 456 pp., Springer, Berlin.

Chebaane, M., J. D. Salas, and D. C. Boes (1995), Product periodic autoregressive processes for modeling intermittent monthly streamflows, *Water Resour. Res.*, *31*(6), 1513–1518.

Chiew, F. H. S., M. C. Peel, G. E. Amirthanathan, and G. G. S. Pegram (2005), Identification of oscillations in historical global streamflow data using empirical mode decomposition, in *Regional Hydrological Impacts of Climate Change: Hydroclimatic Variability*, edited by S. Franks et al., pp. 53–62, IAHS Press, Wallingford, UK.

Cullen, H. M., A. Kaplan, P. A. Arkin, and P. B. Demenocal (2002), Impact of the North Atlantic Oscillation on Middle Eastern climate and streamflow, *Clim. Change*, *55*(3), 315–338.

Dettinger, M. D., M. Ghil, and C. L. Keppenne (1995), Interannual and interdecadal variability in United States surface air temperatures, 1910–87, *Clim. Change*, *31*(1), 35–66.

Frederick, W. K. (2009), *Hilbert Transforms*, 896 pp., Cambridge Univ. Press, Cambridge, U. K.

Gamiz-Fortis, S., D. Pozo-Vazquez, R. M. Trigo, and Y. Castro-Diez (2008), Quantifying the predictability of winter river flow in Iberia. Part II: Seasonal predictability, *J. Clim.*, *21*(11), 2503–2518.

Garrido, J., and J. A. Garcia (1992), Periodic signals in Spanish monthly precipitation data, *Theor. Appl. Climatol.*, *45*(2), 97–106.

Ghil, M., and R. Vautard (1991), Interdecadal oscillations and the warming trend in global temperature time series, *Nature*, *350*(6316), 324–327.

Han, J., and M. van der Baan (2011), Empirical mode decomposition and robust seismic attribute analysis, paper presented at Recovery 2011, Joint Annual Conventions, Canadian Society of Exploration Geophysicists (CSEG), Calgary, Alberta, Canada.

Hipel, K. W., and A. I. McLeod (1996), *Time Series Modelling of Water Resources and Environmental Systems*, Elsevier, Amsterdam.

Huang, N. E., and Z. H. Wu (2008), A review on Hilbert-Huang transform: Method and its applications to geophysical studies, *Rev. Geophys.*, *46*, RG2006, doi:10.1029/2007RG000228.

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, *Proc. R. Soc. London, Ser. A*, *454*(1971), 903–995.

Hurrell, J. W. (1995), Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation, *Science*, *269*(5224), 676–679.

Jones, P. D., T. Jonsson, and D. Wheeler (1997), Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland, *Int. J. Climatol.*, *17*(13), 1433–1450.

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

Katz, R. W., and M. B. Parlange (1996), Mixtures of stochastic processes: Application to statistical downscaling, *Clim. Res.*, *7*(2), 185–193.

Kijewski-Correa, I., and A. Kareem (2007), Performance of wavelet transform and empirical mode decomposition in extracting signals embedded in noise, *J. Eng. Mech.*, *133*(7), 849–852.

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 Resour. Res.*, *43*, W05407, doi:10.1029/2006WR005258.

Lall, U., and A. Sharma (1996), A nearest neighbor bootstrap for resampling hydrologic time series, *Water Resour. Res.*, *32*(3), 679–693.

Lall, U., B. Rajagopalan, and D. G. Tarboton (1996), A nonparametric wet/dry spell model for resampling daily precipitation, *Water Resour. Res.*, *32*(9), 2803–2823.

Lee, T., and T. B. M. J. Ouarda (2010), Long-term prediction of precipitation and hydrologic extremes with nonstationary oscillation processes, *J. Geophys. Res.*, *115*, D13107, doi:10.1029/2009JD012801.

Lee, T., and T. B. M. J. Ouarda (2011), Prediction of climate nonstationary oscillation processes with empirical mode decomposition, *J. Geophys. Res.*, *116*, D06107, doi:10.1029/2010JD015142.

Lee, T., J. D. Salas, and J. Prairie (2010), An enhanced nonparametric streamflow disaggregation model with genetic algorithm, *Water Resour. Res.*, *46*, W08545, doi:10.1029/2009WR007761.

Lee, T. S. (2008), Stochastic simulation of hydrologic data based on nonparametric approaches, PhD dissertation, 346 pp., Colo. State Univ., Fort Collins.

Li, J. P., and J. X. L. Wang (2003), A new North Atlantic Oscillation index and its variability, *Adv. Atmos. Sci.*, *20*(5), 661–676.

Lorenz, E. N. (1963), Deterministic nonperiodic flow, *J. Atmos. Sci.*, *20*, 130–141.

McMahon, T. A., A. S. Kiem, M. C. Peel, P. W. Jordan, and G. G. S. Pegram (2008), A new approach to stochastically generating six-monthly rainfall sequences based on empirical mode decomposition, *J. Hydrometeorol.*, *9*(6), 1377–1389.

Nelles, O. (2001), *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*, 785 pp., Springer, Heidelberg, Germany.

Ouarda, T. B. M. J., J. W. Labadie, and D. G. Fontane (1997), Indexed sequential hydrologic modeling for hydropower capacity estimation, *J. Am. Water Resour. Assoc.*, *33*(6), 1337–1349.

Peel, M. C., T. A. McMahon, and G. G. S. Pegram (2009), Assessing the performance of rational spline-based empirical mode decomposition using a global annual precipitation data set, *Proc. R. Soc. A*, *465*(2106), 1919–1937.

Pegram, G. G. S., M. C. Peel, and T. A. McMahon (2008), Empirical mode decomposition using rational splines: An application to rainfall time series, *Proc. R. Soc. London, Ser. A*, *464*(2094), 1483–1501.

Prairie, J. R., B. Rajagopalan, T. J. Fulp, and E. A. Zagona (2005), Statistical nonparametric model for natural salt estimation, *J. Environ. Eng.*, *131*(1), 130–138.

Prairie, J. R., B. Rajagopalan, T. J. Fulp, and E. A. Zagona (2006), Modified K-NN model for stochastic streamflow simulation, *J. Hydrol. Eng.*, *11*(4), 371–378.

Press, W., S. Teukolsky, T. William, and B. Flannery (2002), *Numerical Recipes in C++*, 2nd ed., Cambridge Univ. Press, Cambridge, U. K.

Rajagopalan, B., and U. Lall (1999), A k-nearest-neighbor simulator for daily precipitation and other weather variables, *Water Resour. Res.*, *35*(10), 3089–3101.

Rogers, J. C. (1990), Patterns of low frequency monthly sea level pressure variability (1899–1986) and associated wave cyclone frequencies, *J. Clim.*, *3*(12), 1364–1379.

Rössler, O. E. (1976), An equation for continuous chaos, *Phys. Lett. A*, *57*(5), 397–398.

Rössler, O. E. (1995), An introduction to chaos, *Int. J. Intell. Syst.*, *10*(1), 5–13.

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., and D. C. Boes (1980), Shifting level modeling of hydrologic series, *Adv. Water Resour.*, *3*, 59–63.

Salas, J. D., and T. Lee (2010), Non-parametric simulation of single site seasonal streamflows, *J. Hydrol. Eng.*, *15*(4), 284–296.

Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L. Lane (1980), *Applied Modeling of Hydrologic Time Series*, 484 pp., Water Resour. Publ., Littleton, Colo.

Schlesinger, M. E., and N. Ramankutty (1994), Low-frequency oscillation: Reply, *Nature*, *372*(6506), 508–509.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis: Monographs on Statistics and Applied Probability*, 175 pp., Chapman and Hall, London.

Stedinger, J. R., and M. R. Taylor (1982a), Synthetic streamflow generation: 1. Model verification and validation, *Water Resour. Res.*, *18*(4), 909–918.

Stedinger, J. R., and M. R. Taylor (1982b), Synthetic streamflow generation: 2. Effect of parameter uncertainty, *Water Resour. Res.*, *18*(4), 919–924.

Stedinger, J. R., D. P. Lettenmaier, and R. M. Vogel (1985), Multisite ARMA(1,1) and disaggregation models for annual streamflow generation, *Water Resour. Res.*, *21*(4), 497–509.

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, *J. Hydrometeorol.*, *4*(3), 489–505.

Vogel, R. M., and A. L. Shallcross (1996), The moving blocks bootstrap versus parametric time series models, *Water Resour. Res.*, *32*(6), 1875–1882.

Wilks, D. S. (1997), Resampling hypothesis tests for autocorrelated fields, *J. Clim.*, *10*(1), 65–82.

Wu, Z. H., and N. E. Huang (2004), A study of the characteristics of white noise using the empirical mode decomposition method, *Proc. R. Soc. London, Ser. A*, *460*(2046), 1597–1611.

Wu, Z. H., and N. E. Huang (2005), Statistical significance test of intrinsic mode functions, in *Hilbert-Huang Transform and Its Applications*, edited by N. E. Huang and S. P. Shen, pp. 125–148, World Sci., Hackensack, N. J.

Wu, Z. H., and N. E. Huang (2009), Ensemble empirical mode decomposition: A noise-assisted data analysis method, *Adv. Adaptive Data Anal.*, *1*(1), 1–41.

Wu, Z. H., N. E. Huang, S. R. Long, and C. K. Peng (2007), On the trend, detrending, and variability of nonlinear and nonstationary time series, *Proc. Natl. Acad. Sci. U. S. A.*, *104*(38), 14,889–14,894.

Xie, L., L. J. Pietrafesa, and K. J. Wu (2002), Interannual and decadal variability of landfalling tropical cyclones in the southeast coastal states of the United States, *Adv. Atmos. Sci.*, *19*(4), 677–686.

T. Lee, Department of Civil Engineering, ERI, Gyeongsang National University, 501 Jinju-daero, Jinju, Gyeongsangnam, 660-701, South Korea. (tae3lee@gnu.ac.kr)

T. B. M. J. Ouarda, Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, United Arab Emirates.