

Earth's Future



RESEARCH ARTICLE

10.1029/2022EF003049

Key Points:

- Stochastic modeling of nonstationary hydrological time series was proposed for future flood management strategies
- Autoregressive with a trend, shifting mean level, and nonstationary oscillation modes were tested
- Stochastic models were validated with performance statistics, and the series were adequately fitted with the oscillation model

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

T. Lee,
tae3lee@gnu.ac.kr

Citation:

Lee, T., & Ouarda, T. B. M. J. (2023). Trends, shifting, or oscillations? Stochastic modeling of nonstationary time series for future water-related risk management. *Earth's Future*, *11*, e2022EF003049. <https://doi.org/10.1029/2022EF003049>

Received 12 JUL 2022
Accepted 21 MAR 2023

Author Contributions:

Conceptualization: Taha B. M. J. Ouarda
Data curation: Taesam Lee
Funding acquisition: Taha B. M. J. Ouarda
Methodology: Taesam Lee, Taha B. M. J. Ouarda
Project Administration: Taha B. M. J. Ouarda
Software: Taesam Lee
Supervision: Taha B. M. J. Ouarda
Validation: Taesam Lee
Visualization: Taesam Lee
Writing – original draft: Taesam Lee
Writing – review & editing: Taha B. M. J. Ouarda

© 2023. The Authors. Earth's Future published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Trends, Shifting, or Oscillations? Stochastic Modeling of Nonstationary Time Series for Future Water-Related Risk Management

Taesam Lee¹  and Taha B. M. J. Ouarda² 

¹Department of Civil Engineering, Gyeongsang National University, 501 Jinju-daero, Jinju 52828, South Korea, ²Canada Research Chair in Statistical Hydro-Climatology, Institut national de la recherche scientifique, Centre Eau Terre, 490 de la Couronne, Quebec, QC G1K 9A9, Canada

Abstract Hydrological time series often present nonstationarities such as trends, shifts, or oscillations due to anthropogenic effects and hydroclimatological variations, including global climate change. For water managers, it is crucial to recognize and define the nonstationarities in hydrological records. The nonstationarities must be appropriately modeled and stochastically simulated according to the characteristics of observed records to evaluate the adequacy of flood risk mitigation measures and future water resources management strategies. Therefore, in the current study, three approaches were suggested to address stochastically nonstationary behaviors, especially in the long-term variability of hydrological variables: as an overall trend, shifting mean, or as a long-term oscillation. To represent these options for hydrological variables, the autoregressive model with an overall trend, shifting mean level (SML), and empirical mode decomposition with nonstationary oscillation resampling (EMD-NSOR) were employed in the hydrological series of the net basin supply in the Lake Champlain-River Richelieu basin, where the International Joint Committee recently managed and significant flood damage from long consistent high flows occurred. The detailed results indicate that the EMD-NSOR model can be an appropriate option by reproducing long-term dependence statistics and generating manageable scenarios, while the SML model does not properly reproduce the observed long-term dependence, that are critical to simulate sustainable flood events. The trend model produces too many risks for floods in the future but no risk for droughts. The overall results conclude that the nonstationarities in hydrological series should be carefully handled in stochastic simulation models to appropriately manage future water-related risks.

Plain Language Summary The current study explores and develops three possible approaches to simulate nonstationarities in hydrometeorological variables for future management strategies of water-related disasters, such as floods. These include the simulation as trends, shifting mean models, and oscillatory signals. Depending on the characteristics of the series, the three stochastic simulation models are tested, and conclusions are formulated.

1. Introduction

The time series of hydroclimatological variables are modeled and stochastically simulated to provide alternatives for future water resources management (Hipel & McLeod, 1996; Lall & Sharma, 1996). In a hydroclimatological time series, diverse nonstationarities were included, such as trends, change points, or oscillations, especially in recent years due to anthropogenic effects and climate change (Du et al., 2020; Ganguli & Coulibaly, 2017; Hui et al., 2018; Zhou, 2020).

Different stochastic simulation models have been adopted to fit those nonstationarities. For example, a number of hydroclimatological variables present trend in their time series due to climatic, environmental, and anthropogenic effects (Khaliq et al., 2009; Myronidis et al., 2018; Sakiur Rahman et al., 2018). Subsequently, the efforts for detecting their trends and modeling the trends have been performed (Fathian & Vaheddoost, 2021; Pishgah Hadiyan et al., 2022; Razavi & Vogel, 2018).

Furthermore, Salas and Boes (1980) developed the shifting mean level (SML) model for hydrological time series and Sveinsson et al. (2003) illustrated that SML model can be a good candidate to describe the alternating mean levels of hydroclimatological variables, including the Pacific Decadal Oscillation index and mean annual flows of

the Niger River. The SML model was also adopted to stochastically model and simulate the Great Lakes system (Fagherazzi et al., 2007, 2011).

In addition, Kwon et al. (2007) employed wavelet analysis (Foufoula-Georgiou & Kumar, 1994; Labat, 2005) and autoregressive moving average (ARMA) models to simulate the selected signals of hydroclimatological variables. Even if its application was successful in simulating rainfall and temperature series, its wavelet components were not easy to select, and the ARMA model does not have feature the ability to reproduce oscillatory signals extracted from wavelet analysis. Also, Lee and Ouarda (2012) proposed a model that extracts signals from a time series with empirical mode decomposition (EMD) and simulates the combined significant signals with the nonstationary oscillation resampling technique (NSOR). The EMD results from a finite number of signals from the time series and the NSOR has the ability to reproduce the smoothed variability of the extracted signals. Further model development has been made to simulate time series with a multivariate framework (Lee & Ouarda, 2019).

Furthermore, deep learning models have been popularly employed in the literature, especially for the prediction of hydroclimatological variables by modeling nonlinear temporal relations in hydroclimatological variables (Kumar et al., 2019; Le et al., 2019; Qi et al., 2019). Among others, long short-term memory (LSTM), abbreviated as LSTM (Hochreiter & Schmidhuber, 1997), was employed as one of the most popular temporal deep learning models. Lee et al. (2020) recently employed the LSTM model to simulate hydroclimatological variables and proved that the LSTM model can be a good alternative to reproduce the long-term dependence structure of a historical time series since the LSTM model was originally invented to reproduce the long-term memory as presented in the name itself (i.e., LSTM). Likewise, there are a number of alternative models to illustrate the nonlinearities and nonstationarity embedded in hydroclimatological series (Khaliq et al., 2009; Myronidis et al., 2018; Sakiur Rahman et al., 2018; Zirulia et al., 2021).

Nonetheless, it is often not clear which type of nonstationarity is included in the target hydroclimatological time series, and its impact on water management is significant according to the definition of nonstationarity. Meanwhile, the record-breaking floods of 2011 in Lake Champlain (LC) in Vermont and New York, U.S. and the Richelieu River (RR) in the province of Quebec, Canada prompted the U.S. and Canadian governments to work on identifying how flood forecasting, preparedness and mitigation can be improved in the Lake Champlain and Richelieu River (LCRR) basin (Saad et al., 2016). The International Joint Commission (IJC) submitted the LCRR Plan of Study (PoS) to the Governments of Canada and the United States in 2013. The 2013 PoS presented the work required in the LCRR basin to analyze potential floodplain management solutions, to build a new flood forecasting system for the LCRR basin, and to suggest structural and non-structural flood mitigation and prevention measures. To perform the objectives suggested in the PoS, it is important to adequately generate synthetic net basin supply (NBS) series of the LCRR basin. These series are crucial to evaluate the adequacy of flood risk mitigation measures and management strategies under many potential hydrological scenarios that could occur in the future (Riboust & Brissette, 2015).

To simulate synthetic NBS series, an appropriate time series model must be selected according to the characteristics of the observed NBS series, including the stochastically nonstationary behavior in that the series presents oscillatory long-term variability (note that this has been discussed later). Furthermore, it has been requested that the simulated series reproduce the historical levels of LC and RR flows by the LCRR IJC technical work group.

Therefore, the objectives of the current study are (a) to apply feasible simulation models that can produce the long-term nonstationary variability of the hydrological series; (b) to compare the simulation models according to the reproduction of the water-related risks for water management; and (c) to provide all feasible future scenarios to test structural and non-structural flood mitigation and prevention measures according to the simulated scenarios.

The remainder of this paper is organized as follows. In Section 2, the mathematical background of the employed SML model and empirical mode decomposition with nonstationary oscillation resampling (EMD-NSOR) are reviewed. Section 3 presents study area and data description. Section 4 reports the results for the simulated data with the SML and EMD-NSOR models as well as the AR with trend. In Section 5, the comparison result of the tested three different models were discussed followed by future scenarios. Section 6 provides a summary and conclusions.

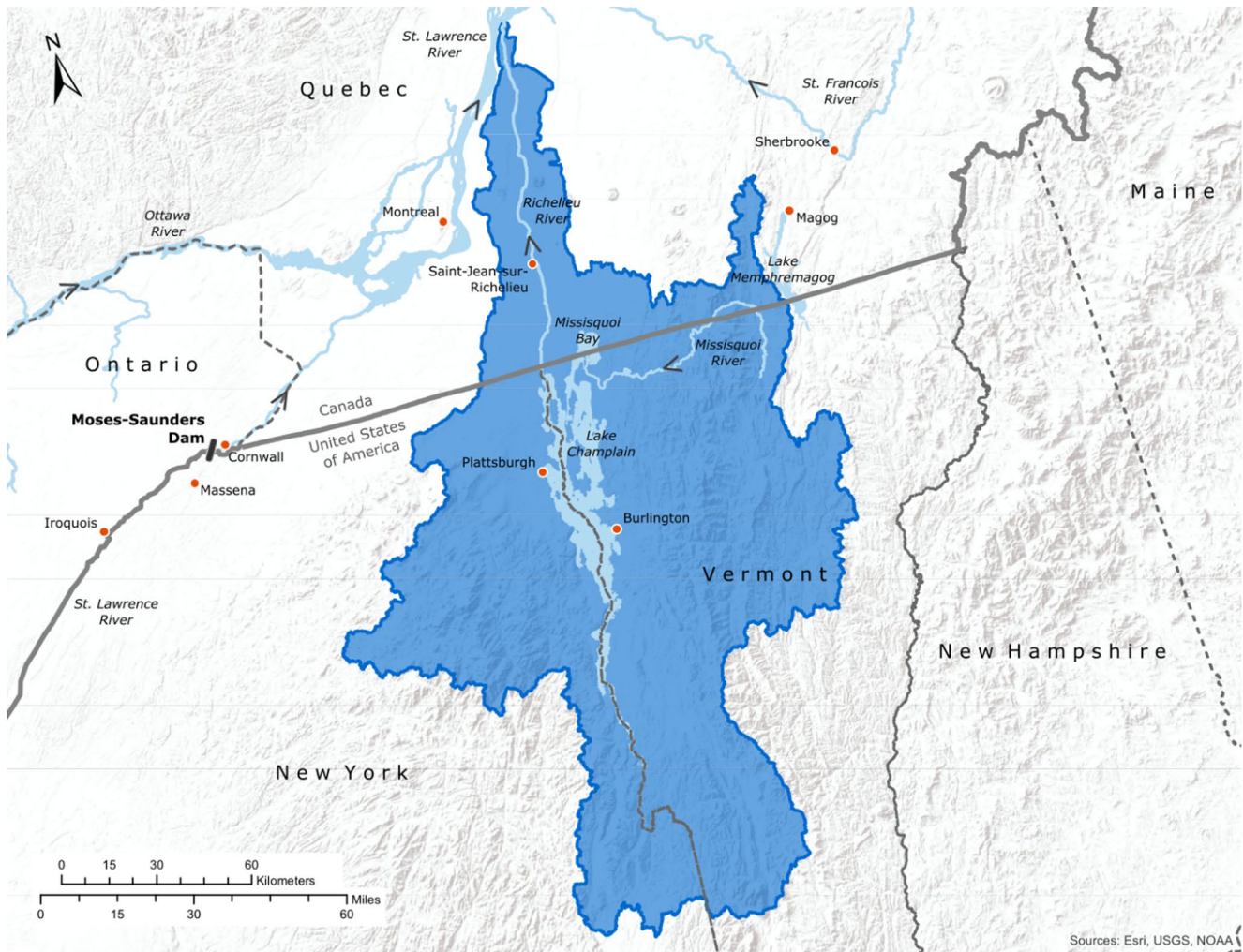


Figure 1. Map of the Lake Champlain and Richelieu River (LCRR) basin. Note that dark blue presents the whole area of the LCRR basin, light blue inside the dark blue presents Lake Champlain (LC), and the northward line from LC illustrates the Richelieu River. The map was provided by the International Joint Commission.

2. Mathematical Backgrounds

As discussed, the major objective of the current study is to build a stochastic simulation model that reflects the statistical characteristics of the NBS for the LCRR system shown in Figure 1 and to simulate a long-term series with the model. To build a relevant time series model, the NBS time series was observed, in Figure 2. Ouarda and Charron (2019) performed frequency analysis for the RR flood flows and Champlain level and they found that abrupt change points were detected in the NBS series. The overall variation in the NBS series can be viewed in following three different ways: as the overall linear trend (top panel), shifting mean process (middle panel), and long-term oscillation (bottom panel), as seen in Figure 2.

Modeling the overall variation is important in the current NBS simulation because this series should be used in testing mitigation plans for all possible hydrological events, such as the 2011 flood event in the LCRR basin (Saad et al., 2016) and water level declination of Great Lakes in 2014 (Watras et al., 2014). The overall trend, in the top panel of Figure 2, was modeled with a linear trend model and AR model for its residual. The shifting mean process in the middle panel of Figure 2 was modeled with the SML by Sveinsson et al. (2003). Additionally, the long-term oscillation feature in the bottom panel of Figure 2 was modeled with the EMD-NSOR model (Lee & Ouarda, 2012, 2019).

In Figure 3, the overall procedure of the current study is presented. At first, the observed data was obtained and transformed into the normal domain. Then, the tested models as AR with Trend, SML, and EMD-NSOR were fitted to the observations followed by simulating the series for the current period. The model performance and comparison were made with the backtransformed simulation data. Finally, the future extension was made with the tested models.

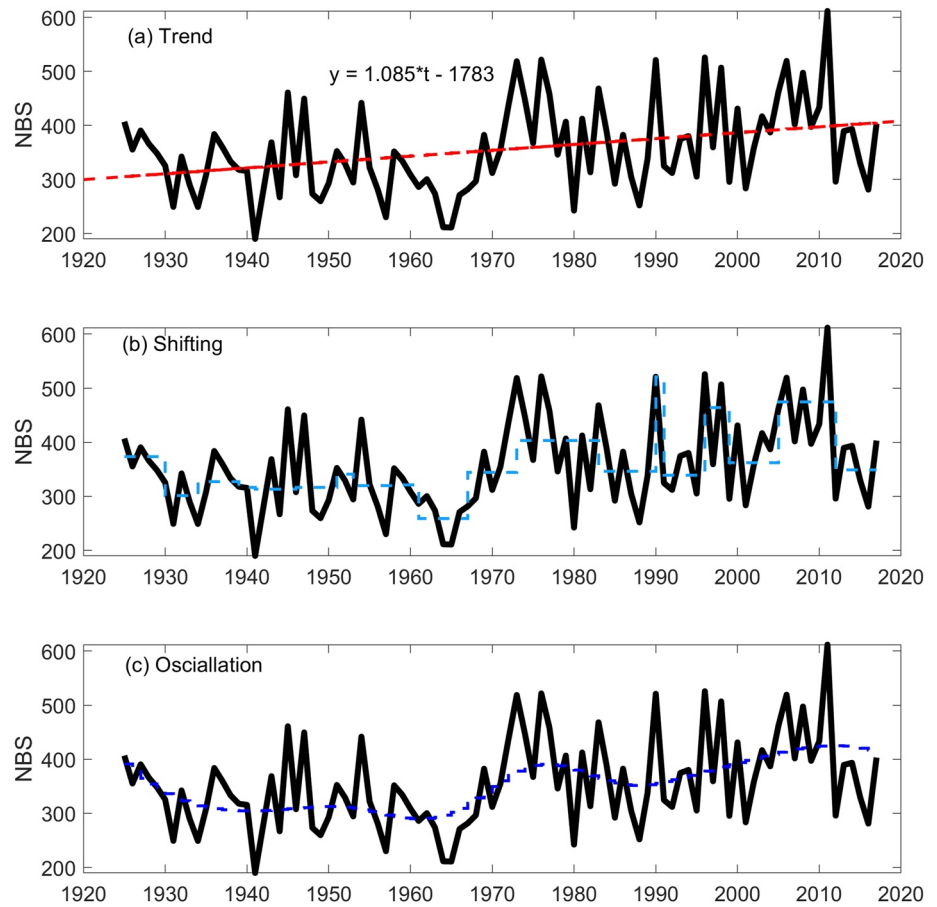


Figure 2. Time series of the observed annual net basin supply (NBS) (unit: m^3/s) of Lake Champlain and Richelieu River (thick solid line) viewed as the overall linear trend (top panel), shifting mean process (middle panel), and long-term oscillation (bottom panel).

2.1. Autoregressive Model

A number of time series models have been tested in the literature to simulate time series of hydroclimatological variables (Lall et al., 1996; Laux et al., 2011; Monbet & Marteau, 2004; Salas, 1993; Tao & Delleur, 1976). The traditional autoregressive (AR) model has been commonly used to consider random noise and time-dependent structures (Brockwell & Davis, 2003). For a hydroclimatological variable at time t (referring a year), Y_t , the AR model of order p (AR(p)) can be explained as follows:

$$Y_t = \sum_{j=1}^p \varphi_j Y_{t-j} + \varepsilon_t \quad (1)$$

where Y_t is the time-dependent variable with zero mean, and ε_t is the time-independent component as $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, and φ_j is the parameter of the model for the j th lagged variable (i.e., Y_{t-j}). The target time series should have the condition that there is no long-term trend or cyclicity and the time series variable Y_t is assumed to be normally distributed. For example, the AR(1) model can be simply written as follows:

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t \quad (2)$$

In other words, the AR model assumes stationarity so that the marginal and joint probability distribution of the variable Y_t is consistent as follows:

$$F(Y_t) = F(Y_{t+k}) \quad \text{for any } t \text{ and } k \quad (3)$$

$$F(Y_t, Y_{t+k}) = F(Y_u, Y_{u+k}) \quad \text{for any } t, u, \text{ and } k \quad (4)$$

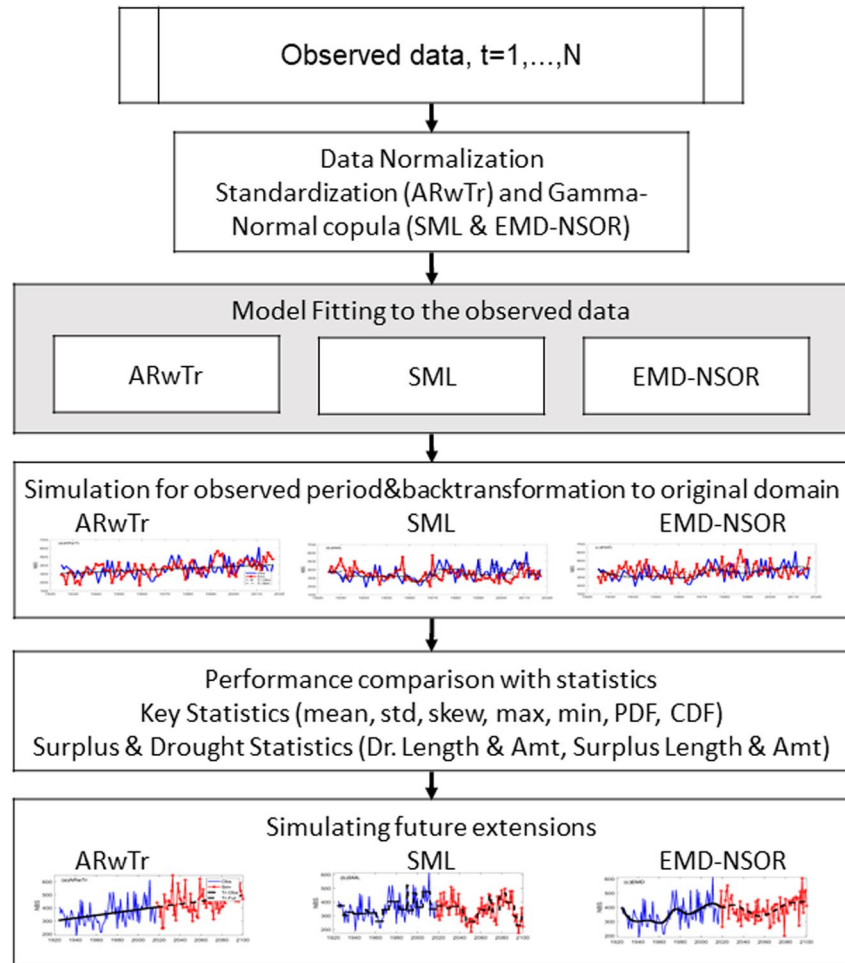


Figure 3. Overall procedure of the current study with the tested models as ARwTr, SML and empirical mode decomposition with nonstationary oscillation resampling (EMD-NSOR).

Another feasible way to consider the time series of the annual NBS would be that the time series presents the overall trend, in Figure 2. It can be identified and modeled with a linear trend as follows:

$$Y_t = \alpha_1 t + \alpha_0 \quad (5)$$

where α_0 and α_1 are parameters and “ t ” represents time (i.e., year). The residual can be modeled as a time series model such as AR(p) in Equation 1 (Hipel & McLeod, 1996). The model of AR(1) with Trend is referred to as ARwTr in the current study.

2.2. Shifting Mean Level Model

Salas and Boes (1980) devised the SML to model an abrupt shift process, and Sveinsson et al. (2003) further improved this model and applied it to hydroclimatological variables. Note that the physical background of the SML is a sudden systematic shift from human effects or natural phenomena such as volcanic activities rather than random processes.

The basic formula of the SML (Salas & Boes, 1980) is presented for a sequence of a random variable Y_t as follows.

$$Y_t = U_t + Z_t \quad (6)$$

where U_t is a sequence of independent identically distributed 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:

$$Z(t) = \sum_{i=1}^t m_i I_{(s_{i-1}, s_i]}(t) \quad (7)$$

where m_i is a random variable with a zero mean and standard deviation of σ_m . This represents the various mean levels over the time expansion. $S_i = N_1 + N_2 + \dots + N_i$ with $S_0 = 0$ and N_i is the timespan at a certain level, i , and a random variable with a geometric distribution and its parameter p as the following:

$$E(N) = \frac{1}{p} \quad (8)$$

Note that p is the probability that a mean level changes, and its reverse is the expected length that its mean level stays. Additionally, $I_{(a,b)}(t)$ 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 U_i are assumed to be mutually independent of each other. The correlation structure of the sequence Y_i (Salas & Boes, 1980) is expressed as follows:

$$\rho_Y(h) = \frac{\sigma_m^2}{\sigma_u^2 + \sigma_m^2} (1-p)^h, \quad h = 1, 2, \dots \quad (9)$$

where $\rho_Y(h)$ is the lagged correlation for the order h . This model has been tested, and its performance for reproducing long-term variability is well studied in the literature (Lee & Ouarda, 2012; Ouarda & Charron, 2019; Sveinsson et al., 2003).

2.3. EMD-NSOR Model

2.3.1. EMD

EMD is invented to extract the different intrinsic modes of oscillations at separate frequencies in a time series. An intrinsic mode of oscillation is also called an intrinsic mode function (IMF) when it meets the following conditions: the number of extrema must be the same as the number of zero crossings or differ from it by no more than one, and the mean value of the two envelopes of the local maxima and minima must be zero. A time sequence can be decomposed into a finite number of IMFs through the shifting process in EMD. The sifting process is performed to obtain IMFs from a time series, $y(t)$, where $t = 1, \dots, N$ as follows:

1. Estimate all of the local extrema and link all local maxima (minima) with a smoothing technique to obtain the upper and lower envelopes with cubic spline (Press et al., 2002) in general (Huang & Wu, 2008).
2. Acquire the first component, h , by finding the difference between the time series and the local mean of the upper and lower envelopes, m , as $h = y - m$.
3. Substitute y by h and repeat steps 1 and 2 until a certain criterion is met such that component (h) surely retains a sufficient physical sense of both amplitude and frequency modulations (Huang & Wu, 2008).
4. Assign the final h as the l th IMF, c_l , and the residue is r_l (i.e., $r_l = r_{l-1} - c_l$ where $r_0 = x$).
5. Repeat steps 1–4 by handling the residue r_l as the original data until the final residue (r_n) is a monotonic function, defined as c_{n+1} .

Finally, the original time sequence, $x(t)$, is presented as the summation of the estimated IMFs c_j , where $j = 1, \dots, n + 1$ (Huang & Wu, 2008) as

$$y(t) = \sum_{j=1}^{n+1} c_j(t) \quad (10)$$

Components with lower order have a higher frequency level and vice versa. Note that the lowest frequency level component is c_{n+1} , while the highest frequency level component is c_1 .

Z. Wu and Huang (2004) and Z. H. Wu and Huang (2005) developed a statistical significance test for IMFs to examine whether an IMF obtained from EMD includes a true signal or just a noise component. This test compares the mean period and the spectral energy between the IMF of the original signal and that of the white noise. When the IMF energy of the observed data with a mean period is placed above the confidence level, it indicates that the corresponding IMF is considered to be statistically significant at the given level.

2.3.2. NSOR

The NSOR model (Lee & Ouarda, 2012, 2019) is based on two nonparametric techniques: K-nearest neighbor resampling and block bootstrapping. The stochastic simulation procedure of the NSOR model is briefly described in the following.

Assume a time series of a certain IMF component $c(t)$ where $t = 1, \dots, N$. The superscripts H and G are employed to illustrate the historical and generated data, respectively.

1. Choose an initial value $c^G(0)$ from the N observations, supposing that each observation has an equal probability of being selected.
2. Simulate a block length, L_B , from a discrete distribution (here, Poisson as in Lee & Ouarda, 2010).
3. Calculate the distances as:

$$D_j = \sqrt{\alpha_1 \{c^G(t-1) - c^H(j-1)\}^2 + \alpha_2 \{\Delta c^G(t-1) - \Delta c^H(j-1)\}^2} \quad (11)$$

where α_1 and α_2 are the inverse variances of each component c and its change rates Δc , respectively.

4. Arrange the distances in ascending order and choose the first k values. Then, randomly choose one among the k values with the weighting probability of $w_i = (1/i) / \sum_{j=1}^k 1/j$, where $i = 1, \dots, k$.
5. Calculate the generated data with length L_B with the following:

$$c^G(l) = c^G(l-1) + \Delta c^H(\bar{p} + l) \quad \text{for } l = 1, \dots, L_B$$

where \bar{p} is the selected time index in step 4.

6. Repeat steps 2–5 until all the required data are simulated.

2.3.3. Overall Procedure of Stochastic Simulation With EMD-NSOR

The proposed EMD-NSOR model can be briefly described as follows:

1. Decompose the target time series $y(t)$ into a finite number of IMFs with EMD.
2. Select the significant components from the significance test (Wu & Huang, 2004).
3. Apply the NSOR model to the selected significant IMF components and the stochastic simulation models to the residuals.
4. Simulate the selected IMFs and residuals with the NSOR model and the fitted stochastic simulation model, respectively.
5. Estimate the final simulation series by summing up the simulated components.

2.4. Performance Statistics

In order to compare the performances of the tested models in the current study, a number of statistics were estimated including key statistics and surplus-drought statistics. The key statistics applied are mean, standard deviation, skewness, lag-1 correlation, maximum and minimum as well as empirical cumulative distribution function (CDF) and kernel density estimate (Chen & Hsu, 2004; Lall et al., 1993) for probability density function (PDF). The drought statistics were also estimated as drought length and amount, surplus length and amount, and storage capacity as well as the Hurst Coefficient (Hurst, 1951). These drought statistics can be estimated with the sequence peak algorithm as the following.

With the water demand WD_t at time t , surplus and drought is defined with the deviation from the water demand as R_t . In general, the mean value of the observation is applied for the water demand (i.e., $WD_t = \mu_y$) (Yevjevich, 1967). Therefore,

$$R_t = y_t - \mu_y \quad (12)$$

The length of the continuous R_t values that are greater than zero is surplus length and its summation is the surplus amount. Also, the length of the continuous values R_t smaller than zero is drought length and its summation is the drought amount. Furthermore, the storage capacity can be estimated with the sequence peak algorithm (Loucks et al., 1981) as:

$$S_t = \begin{cases} S_{t-1} + R_t, & \text{if } S_t > 0 \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

The storage capacity is the maximum value of S_r . The maximum range (R_n , unit: m^3/s) is calculated as

$$R_n = \max(S_r) - \min(S_r) \quad (14)$$

Also, the Hurst Coefficient (Hurst, 1951) is estimated by

$$\text{HC} = \log\left(\frac{R_n}{S_n}\right) / \log\left(\frac{N}{2}\right) \quad (15)$$

where S_n is the standard deviation of the accumulated shortage for R_n and N is the record length.

A trend test was also applied in the current study to detect the trend in the NBS time series. Among others, the Mann-Kendall test was applied (Kendall, 1955; Mann, 1945) in the current study due to its popularity. The Mann-Kendall test is a rank correlation test between the rank of the values and the ordered values as

$$S_k = \sum_{h=1}^{N-1} \sum_{t=h+1}^N \text{sign}(Z_t - Z_h) \quad (16)$$

The standardized value of S_k as is assumed to be normally distributed and tested with two-sided significance test (Yue & Pilon, 2004).

3. Study Area and Data Description

3.1. Study Area

The LCRR basin spans 23,900 square kilometers, with approximately 84% of the basin in northeastern New York and northwestern Vermont in the US and 16% in the province of Quebec in Canada, in Figure 1. The water supplies to the lake or river are commonly referred to as NBS and are estimated with both component-based and residual-based methods (Croley & Lee, 1993). Here, the component-based NBS series is used due to its accuracy and popularity in the literature (Fagherazzi et al., 2011; Ouarda & Charron, 2019).

During the spring of 2011, as mentioned, the LCRR basin experienced the devastating flood recorded over the past 100 years, with damages estimated at more than US\$82 million. Diverse simulation series for all possible hydrologic scenarios are requested to check feasible mitigation plans for the LCRR system. Figure 2 presents the observed time series of the annual NBS (thick black line). It shows that the time series does not seem to have high autocorrelation, but the values are relatively lower during 1925–1970 compared to the later part of the series. Ouarda and Charron (2019) detected a significant change point approximately 1970 for this series, and it also presented a significant trend based on the modified Mann-Kendall trend test (Yue & Pilon, 2004).

3.2. Data Normalization

Standardization of datasets has been commonly employed for time series modeling to meet the normality condition of traditional time series models and many other machine learning algorithms by subtracting the mean and dividing by the standard deviation so that the result variable (Y_t) becomes standard normal with zero mean and unit variance.

In some cases, the standardization might not be appropriate when the original variable does not follow normally distributed data (e.g., skewed data). To prevent this case, the normal copula transformation can be applied (Jeong & Lee, 2015) as the following:

$$Y_t = F_{\Phi}^{-1}[F_Z(Z; \theta)] \quad (17)$$

where F_Z is the CDF of the Z variable, such as gamma, and F_{Φ}^{-1} is its inverse function of the standard normal distribution. From this transformation, the output variable has a standard normal distribution. Gamma distribution was applied for a marginal distribution of the original variable in the current study since the distribution has been popularly employed for hydroclimatological variables and well fitted to positively skewed nonnegative variables such as precipitation, flood, and drought data (Hanson & Vogel, 2008; Nadarajah, 2007, 2009).

The Gamma CDF (F_Z) can be fitted to the hydrological data (here, NBS) by estimating the parameter of the Gamma distribution. In principle, this Gamma-Gaussian transformation with the normal copula can be described

differently and it has been widely adopted in the hydroclimatological field. For example, quantile mapping for hydrological variables with the Gamma marginal distribution is equivalent to this Gamma-Gaussian copula transformation (Cannon et al., 2015; Maraun, 2013) and in developing standardized precipitation index (SPI) (McKee et al., 1993; Serinaldi et al., 2009), this transformation can be applied even though its approximation form was proposed in the original study of the SPI.

Finally, the back-transformation must be performed to have its original form as:

$$Z_t = F_Z^{-1}[F_\Phi(Y; \theta)] \quad (18)$$

4. Results of Simulated Data

4.1. AR Model With Trend (ARwTr)

In the panel (a) of Figure 2, the overall trend was fitted after the trend was found to be significant with the Mann-Kendall test (a rejection of the null hypothesis with the significance level of 0.05, $p = 0.0027$). The residual is modeled with the AR(1) model since not much significant autocorrelation remains except for the lag-2 autocorrelation function (ACF) (data not shown). The same record length and 200 series were simulated, and one of the series is depicted in the top panel of Figure 4.

Normal copula standardization as in Equation 17 was not feasible for this model since its trend results in some infinite values are in the back-transformation. Instead, the typical standardization in Equation was applied. As a result, the skewness and extrema underestimate the observed values, as shown in Figure 5, while the other statistics as mean, standard deviation, lag-1 correlation are reproduced well. The drought and surplus statistics show some improvement in Figure 6 compared to the AR model presented in Figure S4 in the Supporting Information S1. However, it still presents some underestimation.

The simple AR(1) model has also been tested, and the result is shown in Text S1 in the Supporting Information S1 as well as Figures S1, S3, and S4 in the Supporting Information S1. The result indicates that no better performance can be found from the original AR(1) model than the ARwTr.

4.2. SML Model

As mentioned, the shifting mean level model assumes that the mean level persisted for a period and is changing randomly. Following these statistical characteristics, the model can reproduce the mean level shifting in the annual NBS, as shown in the middle panel of Figure 4. Note that the mean levels from the observation in the middle panel of Figure 4 were selected from the visual inspection. In contrast, the simulated mean levels with the standard normal distribution and the geometric distribution (Z_t in Equation 7) were taken from the simulated data. The simulated data (see the red thick solid line with circles) show that its mean level is changing.

The basic statistics of the observation are reproduced fairly from the simulated data of the SML model in Figure 7, even though slight underestimation can be observed in the skewness and maxima. The drought and surplus statistics as well as the storage capacity of the observed data are reproduced fairly well with the SML model (see Figure 8) compared to the ARwTr and AR(1) models. However, the results still show underestimation for these statistics. Note that the lagged dependence structure of the SML model is exponentially decreased as shown in Equation 9 and the SML model reproduces only the short-term memory of the applied time series (Brockwell & Davis, 2003).

4.3. EMD-NSOR Model

To apply the EMD-NSOR model, the EMD procedure must be applied to the target time series so that the target time series are split into a few IMFs, as shown in Figure 9. Note that (a) the standardized data with the normal copula approach were applied to preserve the marginal distribution of the observed data; (b) long-term oscillation can be seen in the IMFs of c_3 , c_4 , and c_5 ; and (c) the last component (c_6) is not an IMF but the overall trend. Among the IMFs, one should select the significant IMFs. The significance test was applied to the IMFs, and the result is shown in Figure 10. Each asterisk represents each IMF, and an IMF farther away from the lines indicates that its corresponding IMF is distinguishable from random noise. The result indicates that c_5 and c_6 are

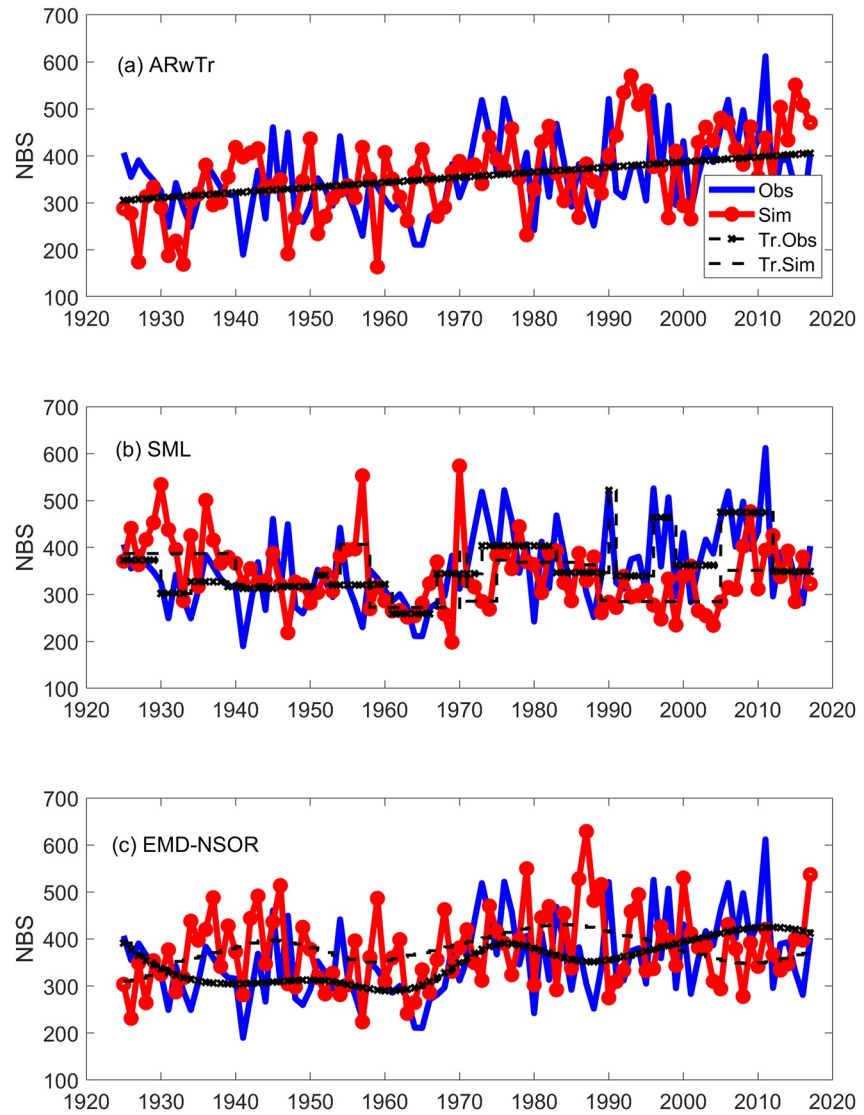


Figure 4. Time series of the observed annual net basin supply (NBS) (unit: m^3/s) of Lake Champlain and Richelieu River (LCRR) and an example of the simulated series with (a) the AR(1) with trend referred to as ARwTr; (b) the SML model; (c) the empirical mode decomposition with nonstationary oscillation resampling (EMD-NSOR) model. Note that (1) the blue solid line presents the observed time series, while the red solid line with a circle shows the simulated data; (2) the black dashed line with x markers presents the trend for the observed data, and the black dashed line presents the trend of the simulated data; and (3) the trend indicates the overall trend at AR(1) with the overall trend, shifting mean levels at the SML model and the summation of the oscillatory components.

the significant IMFs. However, Lee and Ouarda (2010) argued that this significance test has some drawbacks, such that no red noise signal (i.e., existing of the underlying time-dependent structure) is considered and thereby suggested a subjective judgment with inspecting several combinations visually and selecting the best combination that presents the best reproduction of the key and drought statistics as well as the best reproduction of the observed long-term oscillations. We tested several combinations in the present study and concluded that the summation of the IMFs of c_4 , c_5 , and c_6 should be used since c_5 and c_6 must be included based on the significance test, and c_4 must also be included since it contains relatively long-term oscillation and its modulation is close to the other selected IMFs. The other combinations did not lead to better performances in reproducing the long-term oscillation structure of the observed data.

In the bottom panel of Figure 4, the summed selected IMFs (thick black dotted line with cross markers) of the observed data are presented. The long-term variation in the observed time series is captured well. The selected

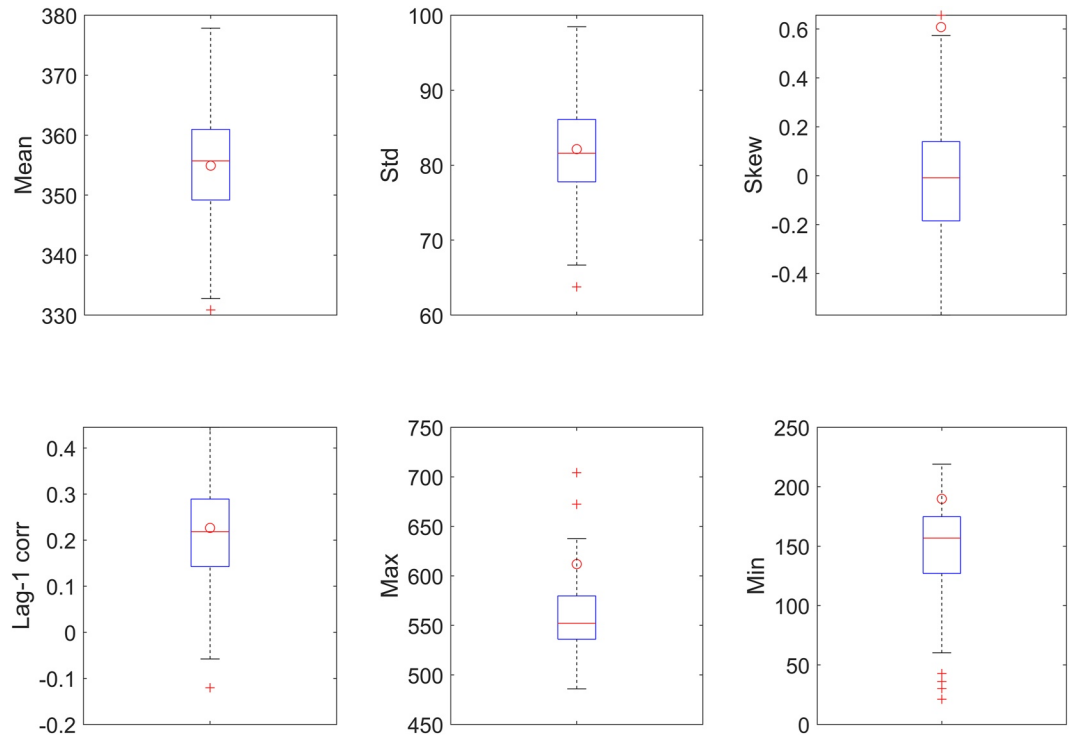


Figure 5. Boxplots of basic statistics of simulated series (93 years 200 series) with the ARwTr for the net basin supply (unit: m^3/s) of Lake Champlain and Richelieu River as well as its observation (circle).

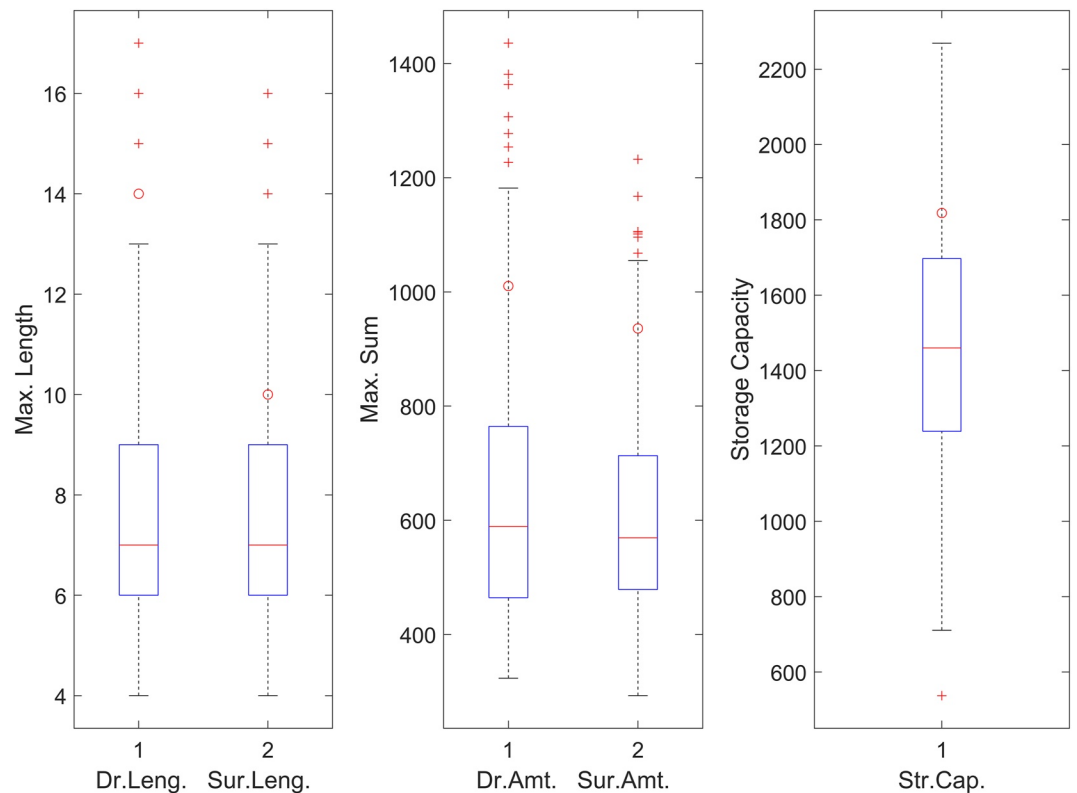


Figure 6. Boxplots of drought and surplus statistics of simulated series (93 years 200 series) with the ARwTr model for the annual net basin supply (unit: m^3/s) of Lake Champlain and Richelieu River as well as its observation (circle).

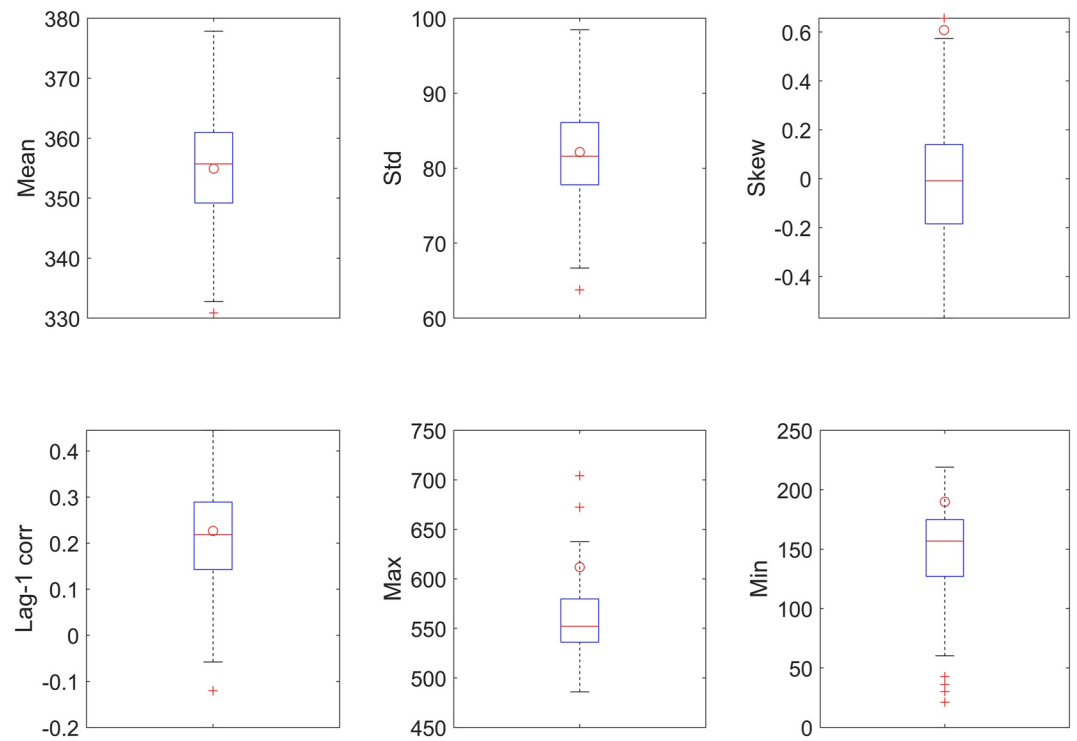


Figure 7. Boxplots of standard statistics of the simulated series (93 years 200 series) with the SML model for the net basin supply (unit: m^3/s) of Lake Champlain and Richelieu River as well as its observation (circle).

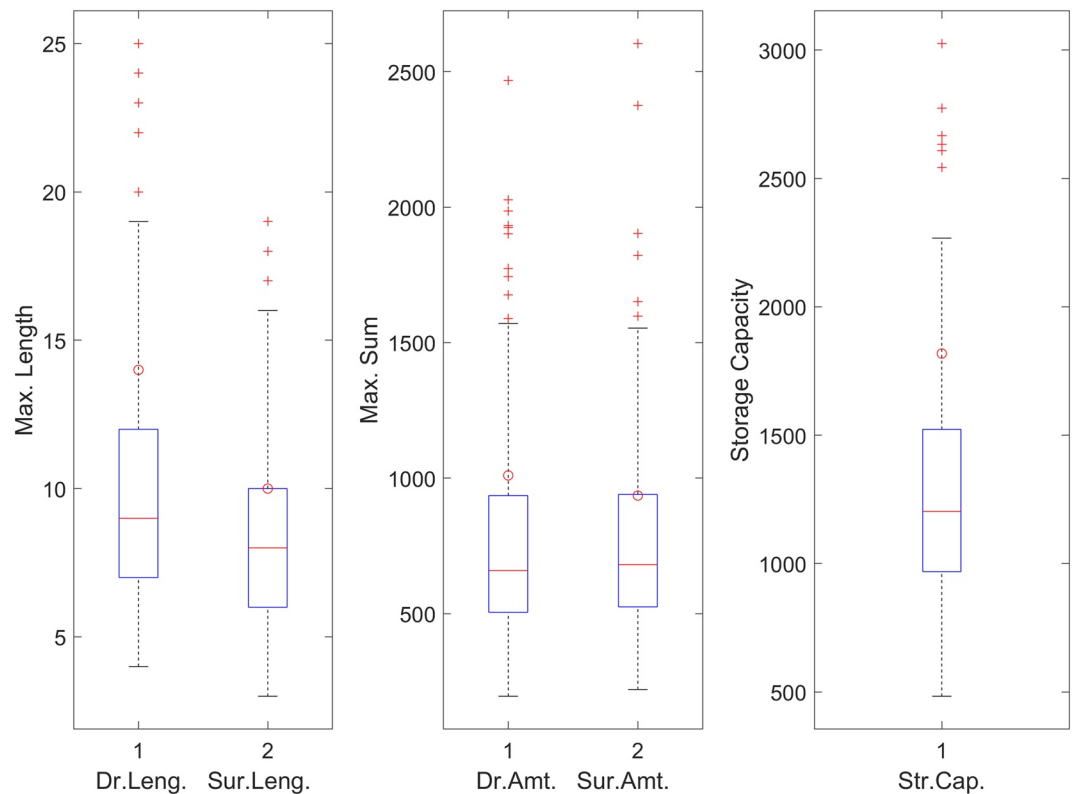


Figure 8. Boxplots of drought and surplus statistics of simulated series (93 years and 200 series) with the SML model for the net basin supply (unit: m^3/s) of Lake Champlain and Richelieu River as well as its observation (circle).

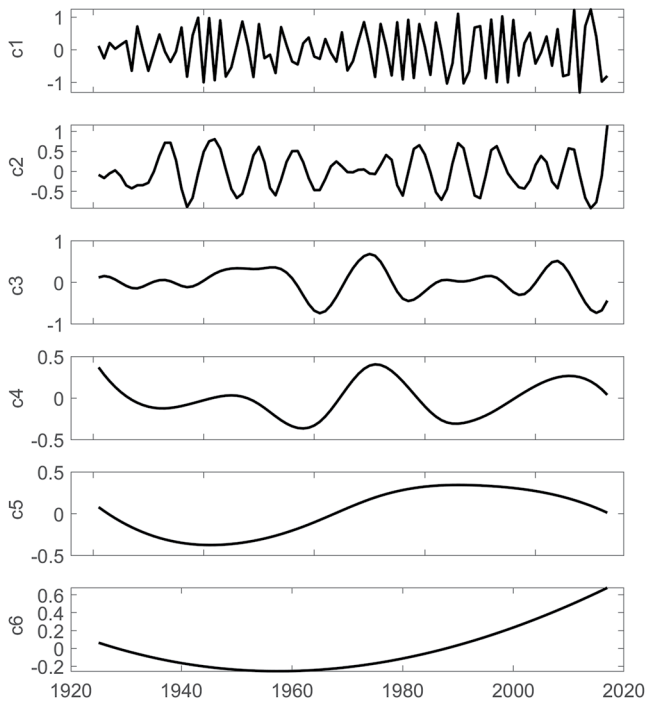


Figure 9. Time series of the extracted components from empirical mode decomposition (c_1 – c_6) from the standardized net basin supply data. Note that (1) the long-term oscillation pattern can be observed in c_4 and c_5 and (2) each y-axis has a different scale to present the variability of each component appropriately.

IMF was modeled with the NSOR model, while the summed residual IMF series was modeled with the AR(1) model to avoid losing any underlying time-dependent structure. The simulated series of the selected IMFs with the NSOR model is presented in the bottom panel of Figure 4, shown with the thin black dotted line. The plot shows that the NSOR model reproduces the long-term variability of the observed time series well.

Figure 11 shows that the basic statistics are reproduced well, similar to the other models, since the normal copula approach was also applied in this model. As shown in Figure 12, the drought and surplus statistics are preserved best among all the tested models even if a slight underestimation can still be observed. The best reproduction of these statistics is from the reproducibility of the long-term oscillatory behavior in the observed data.

5. Model Comparison and Future Strategies

5.1. Comparison of Nonstationary Models

In the current study, a number of stochastic simulation models were applied to simulate the annual NBS for the LCRR system. To be a comparable stochastic simulation model, it must preserve the basic statistics and reproduce the long-term persistency.

The long-term variability of the observed time series cannot be reproduced with the traditional time series model (AR(1)) as well as with the trend (i.e., AR(1) with trend). The LSTM model was also not able to reproduce it and resulted in only minimal improvement. The SML model treats the long-term

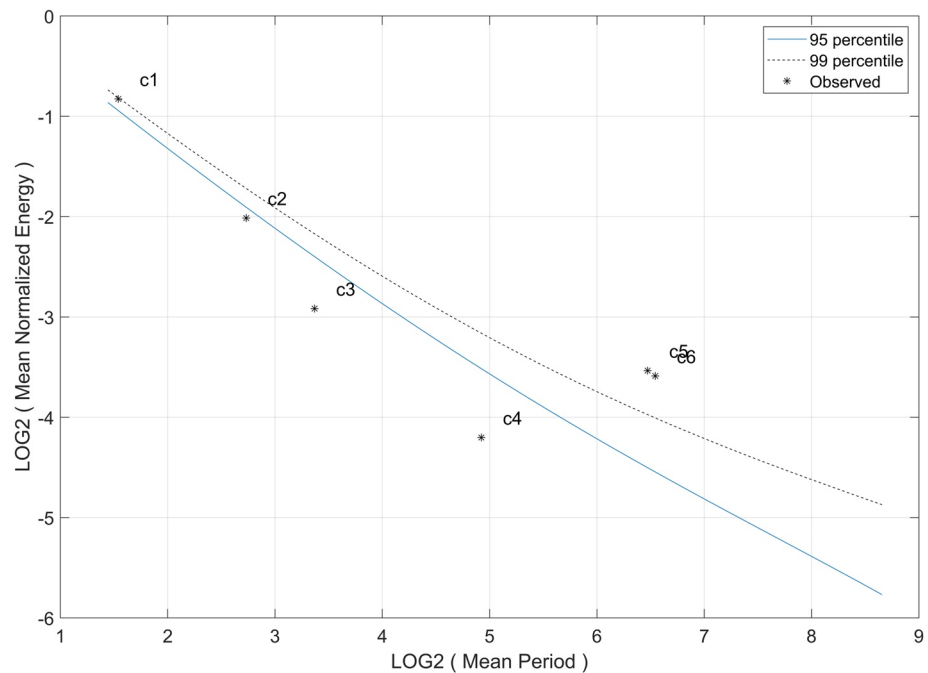


Figure 10. Significance test of the annual net basin supply data with the 95th (solid line) and 99th (dotted line) percentiles. The asterisks (*) farther away from the lines indicate that the corresponding intrinsic mode function (IMF) of the observed series is different from the corresponding IMF of random noise. Note that (1) c_5 and c_6 are significant, while c_1 , c_2 , c_3 , and c_4 can be considered insignificant signals random component (Wu & Huang, 2004) and (2) this significance test still does not consider a red noise signal as underlying short-term correlations, and further visual inspection and subjective selection must be included to find the proper selection of components (Lee & Ouarda, 2010, 2011).

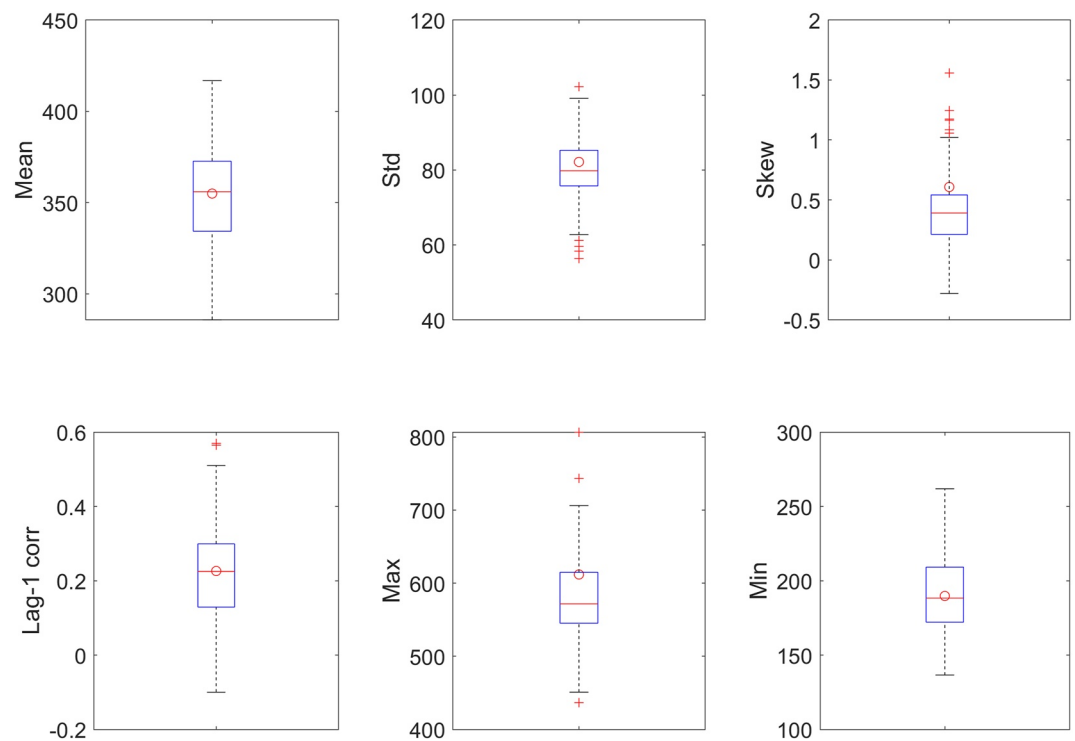


Figure 11. Boxplots of basic statistics of simulated series (93 years 200 series) with the empirical mode decomposition with nonstationary oscillation resampling model for the net basin supply (unit: m^3/s) of Lake Champlain and Richelieu River as well as its observation (circle).

variability as shifts of the mean level and leads to substantial improvement. The EMD-NSOR model mimics the long-term variability with long-term oscillations and preserves the long-term variability best even though a slight underestimation can still be observed.

Since the SML and EMD-NSOR models are comparable and it is worthwhile to investigate the availability of the SML because the model was applied once to the NBS of the current system and to propose a comparable model, these two models were further tested. To do this work, further comparisons were made by focusing on these two models with the statistics of kernel density, CDF, spectral analysis, and Hurst coefficient.

In Figure 13, the CDF and kernel density for the simulated series and observed series are presented. Note that a kernel density is the PDF that is nonparametrically driven (Lall et al., 1993; Salas & Lee, 2010). The result indicates that both models reproduce the observed CDF and density well. In particular, the tail behavior of the observed CDF is well reproduced.

In the panel (d) of Figure 13, it seems that the simulated series with the EMD-NSOR model presents two types of distributions, with one having a median of $300 \text{ m}^3/\text{s}$ and the other having a median of approximately $400 \text{ m}^3/\text{s}$, as shown in the two representative examples (see the thick black lines). This might result from the long-term persistent characteristics of the simulated series from the EMD-NSOR model. The simulated series can be either high or low for an extended period of time. As shown in Figure 4, the simulated series is higher during the 1960–1990 period and lower for the rest of the series and can hence be either high or low for a long time. The distribution of this series seems to be similar to the one in the left one among two representative distributions shown with the thick black lines in the panel (d) of Figure 13. Meanwhile, the series from the SML does not present this behavior. This lack of presentation might be because the estimated probability parameter “ p ” in Equation 8 for the annual NBS is 0.184. Therefore, the average number of sequences with the same mean is 5.438 ($1/0.184$) for the shifting mean process. The relatively high value of the probability parameter (i.e., $p = 0.184$) is due to the lower values of ACFs shown in Figure S2 in the Supporting Information S1. The parameters of the SML model are estimated with the first four or five ACFs and Equation 9. This indicates that the average length of a sequence with the same

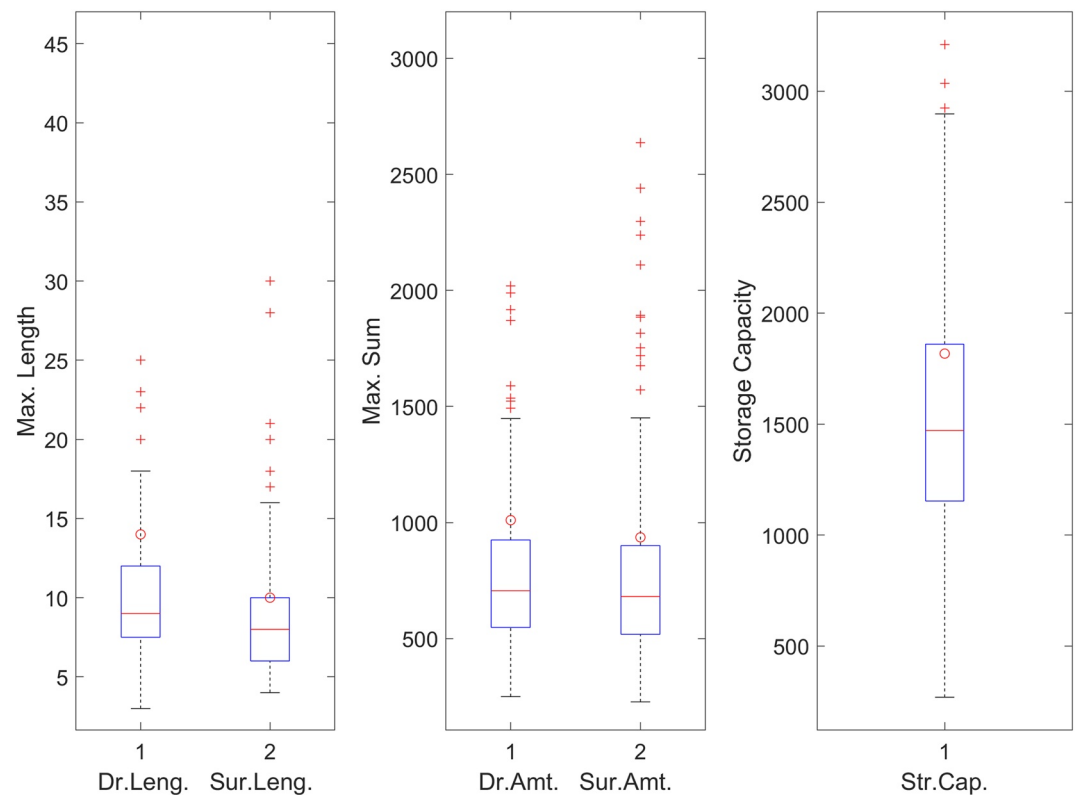


Figure 12. Boxplots of drought and surplus statistics of simulated series (93 years and 200 series) with the empirical mode decomposition with nonstationary oscillation resampling model for the net basin supply of Lake Champlain and Richelieu River as well as its observation (circle).

mean simulated from the SML model is not more than 6 years. It illustrates that the simulated series from the SML does not have as much as of a long-term persistency as the ones from the EMD-NSOR model.

In Figure 14, the estimated spectral density is presented along with the frequency ($f = 1/T$, where T represents time [here, years]). Note that a spectral density with lower frequency (f) indicates long-term persistency (i.e., larger T years). In detail, the EMD-NSOR model better preserves the spectral density for the lower frequency (f), representing long-term persistence even though the difference is not substantial.

We further estimated the Hurst coefficient and the maximum range (R_n) of the difference of a current NBS from the mean, as shown in Figure 15 (see its caption for details). The Hurst coefficient is the measure that Hurst (1951) suggested, for which a random and autocorrelation process has a value of 0.5, while most geophysical time series present a value of approximately 0.73. Here, the observed annual NBS has a Hurst coefficient value slightly larger than 0.8. It has been a challenging task to reproduce the observed Hurst coefficient because a number of stochastic simulation models with long-term dependence structures cannot reproduce the observed value. The median of the Hurst coefficient for the 200 simulated series from the SML model reaches up to 0.75, which indicates that the model can reproduce this statistic for most geophysical time series (approximately 0.73 as mentioned above). However, the observed annual NBS presents a higher Hurst coefficient value than the average, even if its ACFs are not very high. This high Hurst coefficient might result from the long-term persistency, as shown in the historical time series, and it is observed in the ACFs for the large lags between 40 and 44 (see Figure S2 in the Supporting Information S1).

Meanwhile, the EMD-NSOR model reproduces this statistic well, in the top panel of Figure 15. The median of the Hurst coefficient of the simulated series from the EMD-NSOR model is approximately 0.8 and is very close to the observed value. This indicates that the modeling of the long-term variability as long-term oscillation (i.e., EMD-NSOR) might be appropriate for the annual NBS of the LCRR basin.

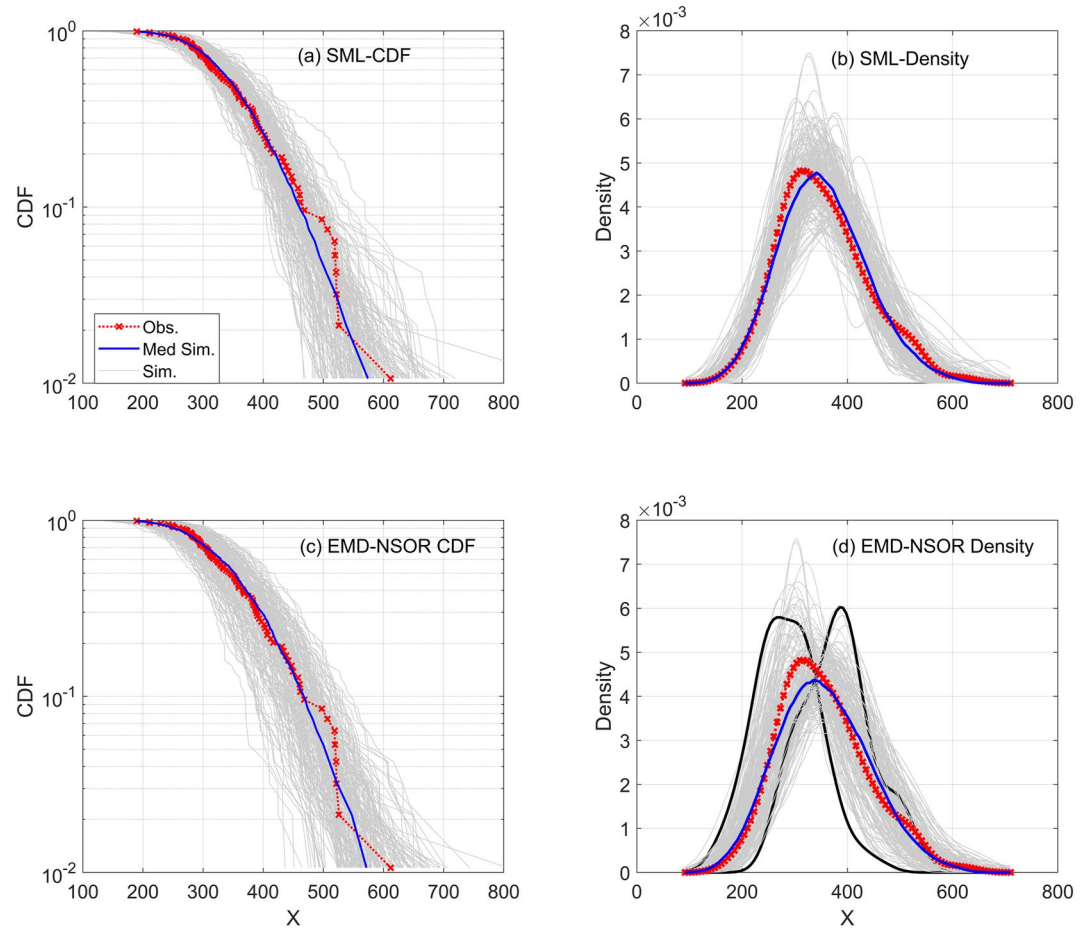


Figure 13. Spaghetti plots of the cumulative distribution function (CDF, left panels) and kernel density (right panels) for the 200 simulated net basin supply (NBS) series (gray lines) with SML (top panels) and empirical mode decomposition with nonstationary oscillation resampling (bottom panels). Note that (1) the median distribution of the simulated 200 NBS is presented with the solid blue line to compare itself with the observed one (the red dotted line with cross marker); and (2) in Panel (d), two examples of the simulated NBS among 200 are emphasized with thick black lines.

Note that we also tested the LSTM deep learning model further. The description of the model and its result can be found in Text S2 in the Supporting Information S1. Not much critical improvement is seen from the result of the LSTM model, in Figures S6 and S7 in the Supporting Information S1.

The annual NBS for the LCRR shows a relatively lower correlation, especially in small lags, but it presents long-term persistency. The results conclude that the EMD-NSOR model can be a comparable alternative to stochastically simulate the series of the annual NBS for the LCRR even if the series has a very unusual dependent structure to reproduce. Note that all the additional statistics shown in this section were also estimated for the other models, such as AR(1), AR(1) with Trend, and LSTM. No model can outperform the EMD-NSOR model for these additional statistics (results not shown).

5.2. Future Strategies for Water-Related Risks

The provided results showed that the EMD-NSOR model provides reasonable simulation scenarios for handling water-related risks compared to the trend and SML models. In addition, a future extension with three tested simulation models was performed, and their implications for the management of water-related risks were studied. In Figure 16, each example of the extended time series up to 2,100 is illustrated for the ARwTr, SML, and EMD-NSOR at the top, middle, and bottom panels, respectively. Note that the initial condition was set as the

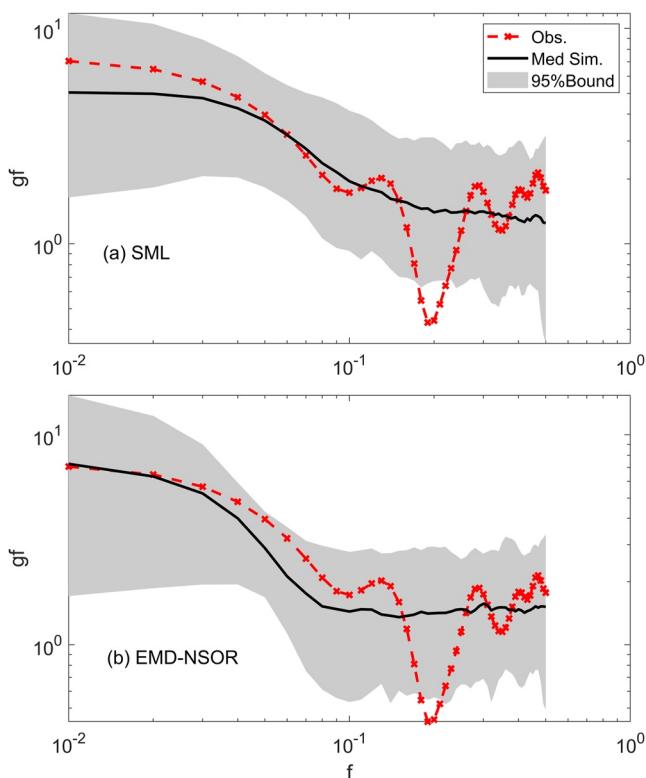


Figure 14. Spectral density (gf) versus frequency ($1/T$, here T is year) of the 95% bound for the 200 simulated net basin supply (NBS) series with SML (top panel) and empirical mode decomposition with nonstationary oscillation resampling (bottom panel) as well as the observed (red dashed line with x marker, $-x$) and the median of the simulated ones (black dashed line) for the annual NBS of the Lake Champlain and Richelieu River basin.

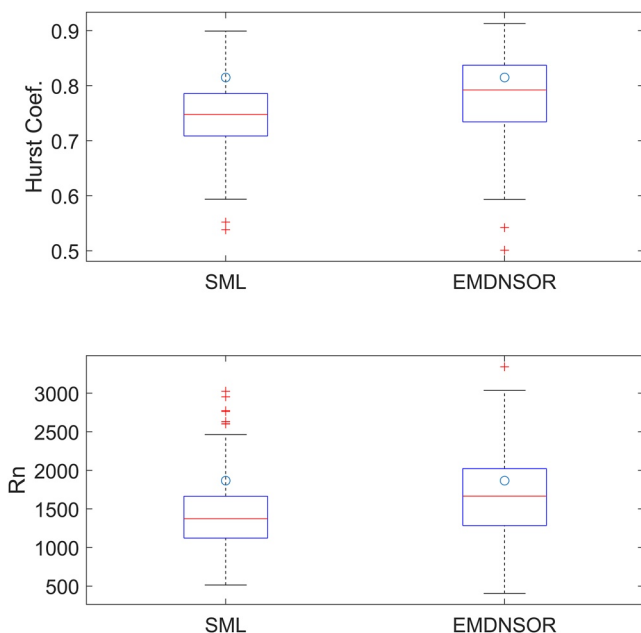


Figure 15. Boxplots of the Hurst coefficient and the maximum range (R_n , unit: m^3/s) of the difference from the mean for the 200 simulated series with the SML and empirical mode decomposition with nonstationary oscillation resampling (EMD-NSOR) models as well as its observation (circle).

NBS value of 2017 for the ARwTr and EMD-NSOR models, while no initial condition was employed for the SML due to its model characteristic.

The time series for the ARwTr at the top panel of Figure 16 illustrated the constant increase in the mean and resulted in an approximate mean of $450 m^3/s$ for the future simulation, while it was approximately $400 m^3/s$ during the observed period. This time series trend might lead to the over-estimation of the flood risk and the underestimation of the drought risk. Significant physical evidence is requested to follow this simulation result. Key statistics and drought and surplus statistics for the simulated future series are presented in Figures 17 and 18, respectively. As expected, the mean (see the first box plot of the left top panel in Figure 17) is very high compared to the observed (see the circle) as well as the maximum and minimum shown in Figure 5 for the key statistics of the observed period. The drought length and amount are much lower than the observations shown in the first boxplots of the panels in Figure 18, while the surplus length and amount are much higher than the observations. The storage capacity is much higher than the observation. Note that the storage capacity is estimated by accumulating the difference between the current NBS (y_t) and its demand (here, the observed mean μ_y), as $S_t = S_{t-1} + (y_t - \mu_y)$. Therefore, the accumulated surplus should be relatively large in the increasing trend, especially for the future extension period. This scenario might lead to too high a risk for a surplus event (i.e., flood) and too low risk for drought.

The future scenario of the SML is randomly simulated without considering the initial condition of the last stage. Even if the mean level stays for a few sequences and varies afterward, the overall structure of the SML is stationary since the parameters of the SML are not varied. Therefore, the key and drought statistics are not much varied in the future extension, in Figures 17 and 18, compared to the observation period in Figures 7 and 8.

Meanwhile, the initial condition of the annual NBS for the EMD-NSOR model is relatively higher in the last 20 years, in the bottom panel of Figure 16. Therefore, the high phase continues up to 2,040, followed by the 30-year low phase, and then the 20-year high phase is followed again. This oscillation results in the future extension up to 2,100 with the EMD-NSOR model containing longer high phases than low phases. Subsequently, the mean of the annual NBS for the future extension (see Figure 17) might be larger than the observed period shown in Figure 11, and the minimum of the future period is larger than that of the observed period. As a result, the drought characteristics for the future extension with the EMD-NSOR model illustrate that the drought length and amount are biased to the lower values, in the last boxplots of the top panels of Figure 18, and the surplus length and amount are biased to the higher values. For the same reason, the storage capacity is higher than that observed. The magnitude of the bias difference from the observation for the EMD-NSOR model is smaller than that for the ARwTr model since the future extension with the EMD-NSOR model oscillates high and low phases rather than one direction increase such as the ARwTr.

According to the model structures, the simulated series of the future extension have distinct characteristics. Unfortunately, it is very difficult to select a singular best model for a water manager as long as substantial physical evidence exists for a target water management system. Therefore, a water manager should prepare the system following the simulated future series and the economic budget allowed. Similar to the flood risk management system in the LCRR basin, surplus statistics are critical to apply. Even if the increas-

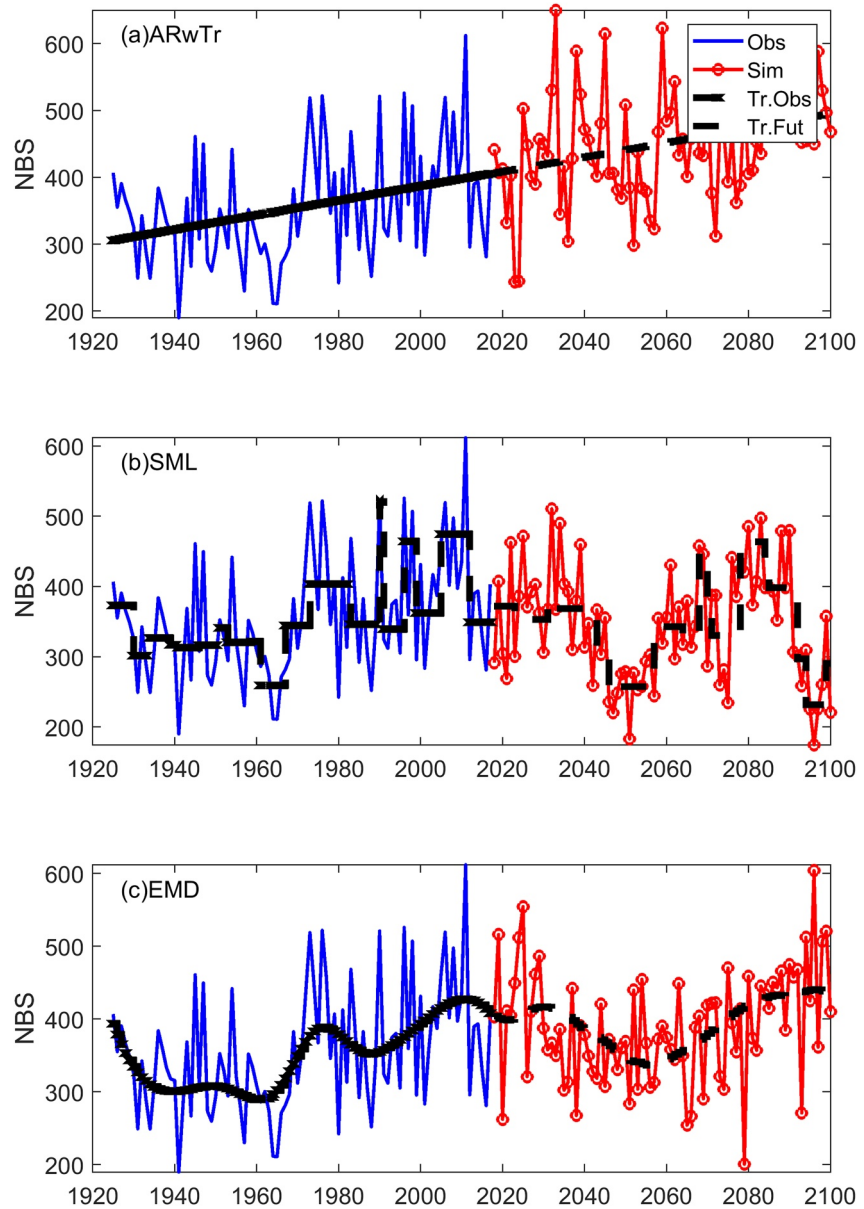


Figure 16. Time series of the net basin supply (NBS) series (unit: m^3/s) for the observed (solid black line) and the simulated series for the future extension (2018–2100 year) from three tested models as the trend, shifting mean, and summed selected IMFs. Note that the simulated example of the future extension is presented with the red thick solid line with a circle, while the trend, shifting mean, and oscillation are depicted with the black thick dotted line.

ing trend might result in the worst scenario from the point view of floods, it is often difficult for a water manager to accept the trend scenarios. Instead, the EMD-NSOR scenario can be a reasonable alternative that provides oscillatory scenarios and a substantial increase in flood risk.

6. Summary and Conclusions

The nonstationarities in hydroclimatological time series, such as trends, shifts, and oscillations from human development and climatic long-term variations, are critical to consider for water-related risk management. In the current study, how stochastic models can take nonstationarities into account by simulating different scenarios with different nonstationarities, such as trends, shifts, and oscillations, is illustrated. Furthermore, nonstationary

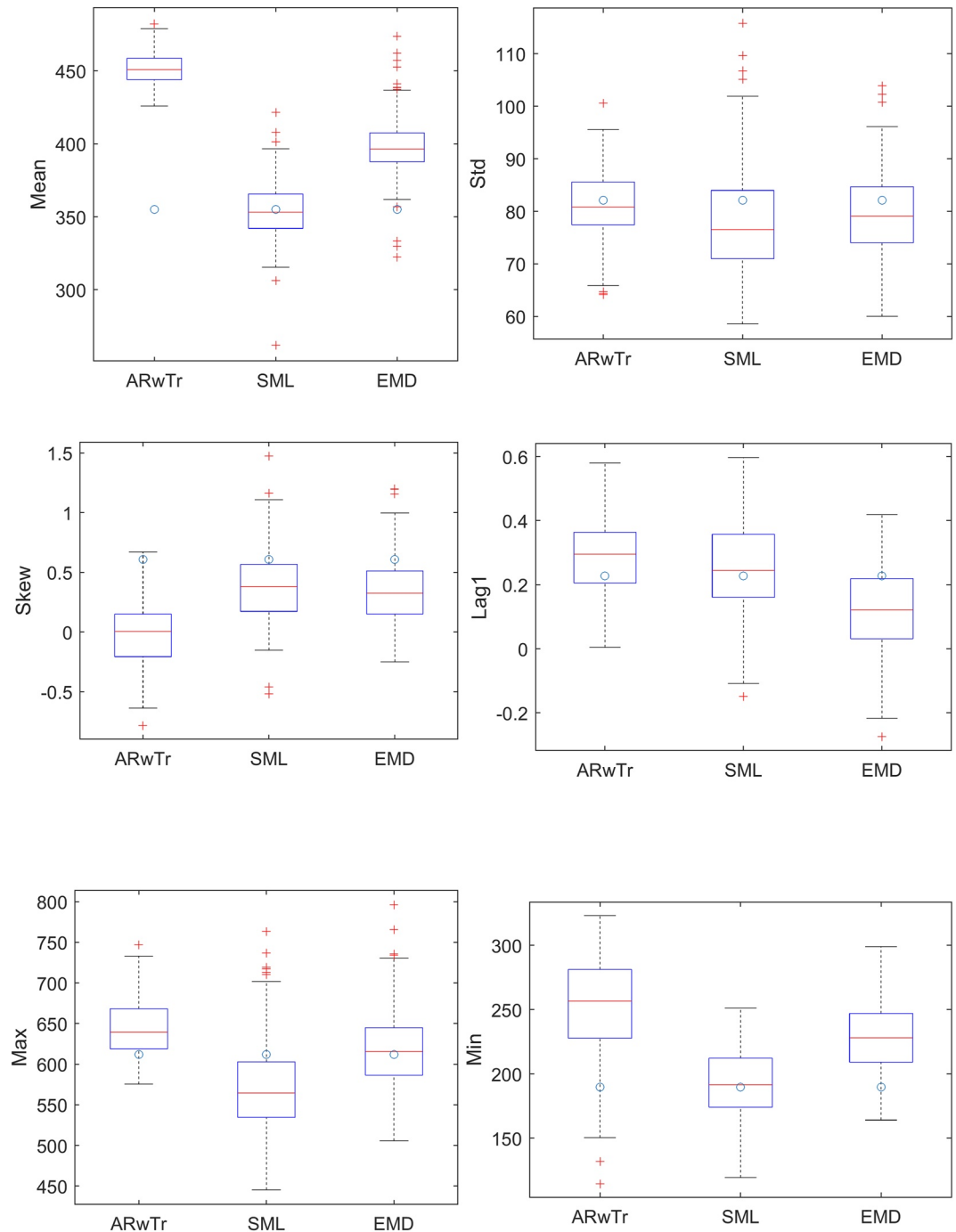


Figure 17. Boxplots of basic statistics for the observed (blue circle) and the simulated series from three tested models with the ARwTr, SML, and empirical mode decomposition with nonstationary oscillation resampling (EMD-NSOR) for the future extension (2018–2100 yr) of the annual net basin supply.

stochastic simulation models were applied to the annual NBS of the LRCC basin to test feasible alternatives for flood mitigation plans and other actions for the LCRR basin.

Since the SML has already been applied in the Great Lakes study for stochastic simulation (Fagherazzi et al., 2011), the feasibility of the SML model and the suggestion of a comparable model for the NBS of the LCRR system

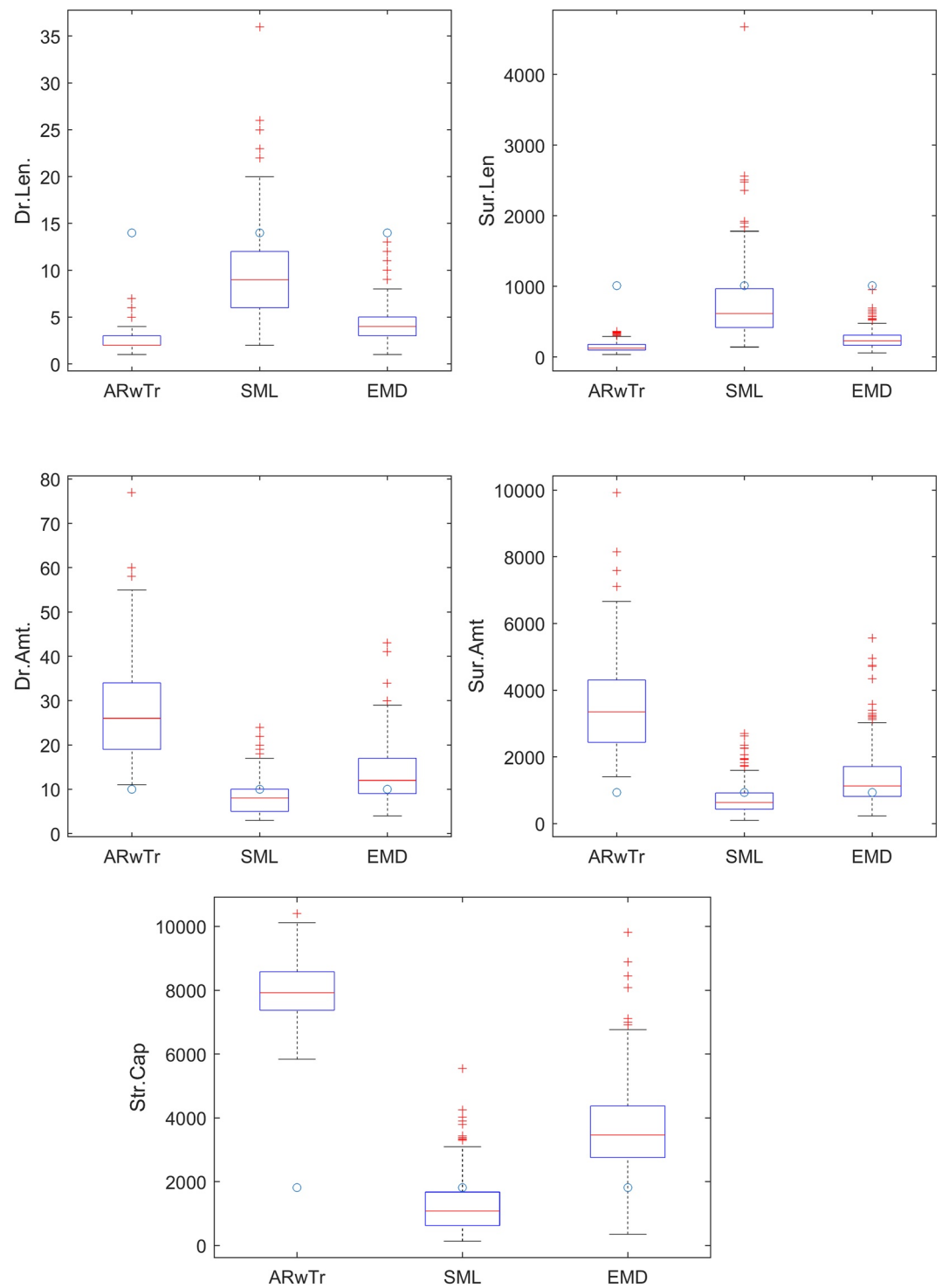


Figure 18. Boxplots of drought and surplus statistics for the future simulated series (2018–2100 yr) with three tested models (ARwTr, SML, and empirical mode decomposition with nonstationary oscillation resampling [EMD-NSOR]) for the net basin supply of Lake Champlain and Richelieu River as well as its observation (circle).

were requested from the water managers of the LRCC. The annual NBS series for the LCRR system was investigated, and we noticed that the series seems to have long-term variability. The long-term variability was treated as an overall trend, shifting mean, and long-term oscillation, and appropriate models were selected to mimic the variability accordingly through the ARwTr, SML, and EMD-NSOR. Two additional models were also tested: the

AR model for traditional time series modeling and LSTM for a recent deep learning model with the capability of long-term memory.

The results show that the AR, ARwTr, and LSTM models do not adequately reproduce the observed drought and surplus statistics as well as storage capacity, while the SML and EMD-NSOR models reproduce these statistics. Meticulous comparisons were made with the statistics of CDF, density, spectral density, and Hurst coefficient for these two models. The detailed comparison results indicate that the EMD-NSOR model better preserves the SML model, especially for the long-term variability of the observed NBS. The poor performance of the SML model might be due to the weak ACF of the observed NBS, especially for short lags. In contrast, the EMD-NSOR model reproduces the long-term variability well, as illustrated with the Hurst coefficient.

In addition, the future extension of the annual NBS up to 2,100 was performed with the tested stochastic simulation models. The results showed that the ARwTr presents too much risk for floods and is difficult to adopt for flood risk management due to its limited budget. The SML scenarios of the future extension are not much different from the observed period. Instead, the EMD-NSOR model produces higher flood risk by providing reasonable and acceptable scenarios to apply.

Overall, the provided results indicate that the EMD-NSOR model can be a good alternative for stochastic simulation of the annual NBS for the LCRR system by treating the long-term variability as long-term oscillations. The synthetic series simulated from the EMD-NSOR model can be used to check the appropriateness for feasible mitigation action plans and preparedness for future floods in the LCRR basin. Further stochastic simulation models can be adopted by incorporating climate variables for future timeframes such as sea surface temperature or climate indices. The behavior of the simulated series is highly dependent on the incorporated climate variables not on the model itself.

Abbreviations

ACF	Autocorrelation function
AR	Autoregressive
ARMA	Autoregressive moving average
ARwTr	AR(1) with trend
CCF	Cross-correlation function
CDF	Cumulative distribution function
EMD	Empirical mode decomposition
IJC	International Joint Commission
IMF	Intrinsic mode function
KLR	KNN-based local linear regression model
KNNR	K-nearest neighbor resampling
LCRR	Lake Champlain-Richelieu River
LSTM	Long short-term memory
NBS	Net basin supply
NSOR	Nonstationary oscillation resampling
PDO	Pacific decal oscillation
PoS	Plan of study
SML	Shifting mean level

Data Availability Statement

The simulation data were published in the Mendeley Data as Simulated NBS for LCRR and downloadable from <https://data.mendeley.com/datasets/5n4vr53hbp>. The net basin supply (NBS) observed data are available from the International Joint Commission (IJC, <https://www.ijc.org/en>) upon request. Software: MATLAB software of EMD was used in the link <https://kr.mathworks.com/discovery/empirical-mode-decomposition.html>. The shifting mean level model was performed with SAM2007, which is downloadable from <http://www.sams.colostate.edu/>. The MATLAB code (Gen_NSOR_sim) for the NSOR model is downloadable from the same website for the simulation data as <https://data.mendeley.com/datasets/5n4vr53hbp>.

Acknowledgments

The first author acknowledges that this research was partially supported by a grant (2022-MOIS63-001) of Cooperative Research Method and Safety Management Technology in National Disaster funded by Ministry of Interior and Safety (MOIS, Korea). This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (2023R1A2C1003850).

References

- Brockwell, P. J., & Davis, R. A. (2003). *Introduction to time series and forecasting* (2 ed., p. 456). Springer.
- Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, 28(17), 6938–6959. <https://doi.org/10.1175/jcli-d-14-00754.1>
- Chen, S. M., & Hsu, Y. S. (2004). Kernel density estimations for maximum of two independent random variables. *Journal of Nonparametric Statistics*, 16(6), 901–924. <https://doi.org/10.1080/1048525042000213012>
- Croley, T. E., II, & Lee, D. H. (1993). Evaluation of Great Lakes net basin supply forecasts. *JAWRA Journal of the American Water Resources Association*, 29(2), 267–282. <https://doi.org/10.1111/j.1752-1688.1993.tb03207.x>
- Du, S., Scussolini, P., Ward, P. J., Zhang, M., Wen, J., Wang, L., et al. (2020). Hard or soft flood adaptation? Advantages of a hybrid strategy for Shanghai. *Global Environmental Change*, 61, 102037. <https://doi.org/10.1016/j.gloenvcha.2020.102037>
- Fagherazzi, L., Fay, D., & Salas, J. (2007). Synthetic hydrology and climate change scenarios to improve multi-purpose complex water resource systems management. The Lake Ontario St Lawrence River Study of the International Canada and US Joint Commission. *Water Resources Management IV*, 103, 163–177.
- Fagherazzi, L., Salas, J. D., & Sveinsson, O. (2011). *Stochastic modeling and simulation of the Great Lakes system*. International Upper Great Lakes Quebec.
- Fathian, F., & Vaheddoost, B. (2021). Modeling the volatility changes in Lake Urmia water level time series. *Theoretical and Applied Climatology*, 143(1–2), 61–72. <https://doi.org/10.1007/s00704-020-03417-8>
- Foufoula-Georgiou, E., & Kumar, P. (Eds.). (1994). *Wavelets in geophysics* (p. 373). Academic Press.
- Ganguli, P., & Coulibaly, P. (2017). Does nonstationarity in rainfall require nonstationary intensity-duration-frequency curves? *Hydrology and Earth System Sciences*, 21(12), 6461–6483. <https://doi.org/10.5194/hess-21-6461-2017>
- Hanson, L. S., & Vogel, R. (2008). The probability distribution of daily rainfall in the United States. In Paper presented at World Environmental and Water Resources Congress 2008. ASCE Library, Honolulu, HI.
- Hipel, K. W., & McLeod, A. I. (1996). *Time series modelling of water resources and environmental systems*. Elsevier.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Huang, N. E., & Wu, Z. H. (2008). A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 46(2), RG2006. <https://doi.org/10.1029/2007rg000228>
- Hui, R., Herman, J., Lund, J., & Madani, K. (2018). Adaptive water infrastructure planning for nonstationary hydrology. *Advances in Water Resources*, 118, 83–94. <https://doi.org/10.1016/j.advwatres.2018.05.009>
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116(1), 770–799. <https://doi.org/10.1061/aceat.0006518>
- Jeong, C., & Lee, T. (2015). Copula-based modeling and stochastic simulation of seasonal intermittent streamflows for arid regions. *Journal of Hydro-Environment Research*, 9(4), 604–613. <https://doi.org/10.1016/j.jher.2014.06.001>
- Kendall, M. G. (1955). *Rank correlation methods* (2nd ed. ed.). Griffin.
- Khaliq, M. N., Ouarda, T. B. M. J., & Gachon, P. (2009). Identification of temporal trends in annual and seasonal low flows occurring in Canadian rivers: The effect of short- and long-term persistence. *Journal of Hydrology*, 369(1–2), 183–197. <https://doi.org/10.1016/j.jhydrol.2009.02.045>
- Kumar, D., Singh, A., Samui, P., & Jha, R. K. (2019). Forecasting monthly precipitation using sequential modelling. *Hydrological Sciences Journal*, 64(6), 690–700. <https://doi.org/10.1080/02626667.2019.1595624>
- Kwon, H. H., Lall, U., & Khalil, A. F. (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. <https://doi.org/10.1029/2006wr005258>
- Labat, D. (2005). Recent advances in wavelet analyses: Part 1. A review of concepts. *Journal of Hydrology*, 314(1–4), 275–288. <https://doi.org/10.1016/j.jhydrol.2005.04.003>
- Lall, U., Moon, Y. I., & Bosworth, K. (1993). Kernel flood frequency estimators - Bandwidth selection and kernel choice. *Water Resources Research*, 29(4), 1003–1015. <https://doi.org/10.1029/92wr02466>
- Lall, U., Sangoyomi, T., & Abarbanel, H. D. I. (1996). Nonlinear dynamics of the Great Salt Lake: Nonparametric short-term forecasting. *Water Resources Research*, 32(4), 975–985. <https://doi.org/10.1029/95wr03402>
- Lall, U., & Sharma, A. (1996). A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research*, 32(3), 679–693. <https://doi.org/10.1029/95wr02966>
- Laux, P., Vogl, S., Qiu, W., Knoche, H. R., & Kunstmann, H. (2011). Copula-based statistical refinement of precipitation in RCM simulations over complex terrain. *Hydrology and Earth System Sciences*, 15(7), 2401–2419. <https://doi.org/10.5194/hess-15-2401-2011>
- Le, X. H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of long short-term memory (LSTM) neural network for flood forecasting. *Water (Switzerland)*, 11(7), 1387. <https://doi.org/10.3390/w11071387>
- Lee, T., & Ouarda, T. B. M. J. (2010). Long-term prediction of precipitation and hydrologic extremes with nonstationary oscillation processes. *Journal of Geophysical Research*, 115(13), D13107. <https://doi.org/10.1029/2009jd012801>
- Lee, T., & Ouarda, T. B. M. J. (2011). Prediction of climate nonstationary oscillation processes with empirical mode decomposition. *Journal of Geophysical Research Atmospheres*, 116(6), D06107. <https://doi.org/10.1029/2010jd015142>
- Lee, T., & Ouarda, T. B. M. J. (2012). Stochastic simulation of nonstationary oscillation hydroclimatic processes using empirical mode decomposition. *Water Resources Research*, 48(2), W02514. <https://doi.org/10.1029/2011wr010660>
- Lee, T., & Ouarda, T. B. M. J. (2019). Multivariate nonstationary oscillation simulation of climate indices with empirical mode decomposition. *Water Resources Research*, 55(6), 5033–5052. <https://doi.org/10.1029/2018wr023892>
- Lee, T., Shin, J. Y., Kim, J. S., & Singh, V. P. (2020). Stochastic simulation on reproducing long-term memory of hydroclimatic variables using deep learning model. *Journal of Hydrology*, 582, 124540. <https://doi.org/10.1016/j.jhydrol.2019.124540>
- Loucks, D. P., Stendiger, J. R., & Haith, D. A. (1981). *Water resources systems planning and analysis* (559 pp.). Prentice-Hall.
- Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, 13(3), 245–259. <https://doi.org/10.2307/1907187>
- Maraun, D. (2013). Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate*, 26(2013), 2013–2014. <https://doi.org/10.1175/jcli-d-12-00821.1>
- McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In *Paper presented at proceedings of the eighth conference on applied climatology*. American Meteorological Society.
- Monbet, V., & Marteau, P. F. (2004). *Non parametric modelling of cyclo-stationary Markovian processes - Part I: Simulation of multivariate sea state processes* (pp. 138–144). International Society Offshore & Polar Engineers.

- Myronidis, D., Ioannou, K., Fotakis, D., & Dörflinger, G. (2018). Streamflow and hydrological drought trend analysis and forecasting in Cyprus. *Water Resources Management*, 32(5), 1759–1776. <https://doi.org/10.1007/s11269-018-1902-z>
- Nadarajah, S. (2007). A bivariate gamma model for drought. *Water Resources Research*, 43(8), W08501. <https://doi.org/10.1029/2006wr005641>
- Nadarajah, S. (2009). A bivariate distribution with gamma and beta marginals with application to drought data. *Journal of Applied Statistics*, 36(3), 277–301. <https://doi.org/10.1080/02664760802443996>
- Ouarda, T. B. M. J., & Charron, C. (2019). *Frequency analysis of Richelieu River flood flows, Lake Champlain flood level and NBS to the Richelieu River basin*. Final project report. INRS-ETE. 60 pp.
- Pishgah Hadiyan, P., Moeini, R., Ehsanzadeh, E., & Karvanpour, M. (2022). Trend analysis of water inflow into the dam reservoirs under future conditions predicted by dynamic NAR and NARX models. *Water Resources Management*, 36(8), 2703–2723. <https://doi.org/10.1007/s11269-022-03170-9>
- Press, W., Teukolsky, S., William, T., & Flannery, B. (2002). *Numerical recipes in C++* (2 ed.). Cambridge University Press.
- Qi, Y., Zhou, Z., Yang, L., Quan, Y., & Miao, Q. (2019). A decomposition-ensemble learning model based on LSTM neural network for daily reservoir inflow forecasting. *Water Resources Management*, 33(12), 4123–4139. <https://doi.org/10.1007/s11269-019-02345-1>
- Razavi, S., & Vogel, R. (2018). Prewhitening of hydroclimatic time series? Implications for inferred change and variability across time scales. *Journal of Hydrology*, 557, 109–115. <https://doi.org/10.1016/j.jhydrol.2017.11.053>
- Riboust, P., & Brissette, F. (2015). Climate change impacts and uncertainties on spring flooding of Lake Champlain and the Richelieu River. *Journal of the American Water Resources Association*, 51(3), 776–793. <https://doi.org/10.1111/jawr.12271>
- Saad, C., St-Hilaire, A., Gachon, P., & El Adlouni, S. (2016). The 2011 flood event in the Richelieu River basin: Causes, assessment and damages. *Canadian Water Resources Journal*, 41(1–2), 129–138. <https://doi.org/10.1080/07011784.2014.999825>
- Sakiur Rahman, A. T. M., Shakil Ahmed, M., Mohammad Adnan, H., Kamruzzaman, M., Abdul Khalek, M., Hasan Mazumder, Q., & Sarwar Jahan, C. (2018). Modeling the changes in water balance components of the highly irrigated western part of Bangladesh. *Hydrology and Earth System Sciences*, 22(8), 4213–4228. <https://doi.org/10.5194/hess-22-4213-2018>
- Salas, J. D. (1993). Analysis and modeling of hydrologic time series. In D. R. Maidment (Ed.), *Handbook of hydrology* (pp. 19.11–19.72). McGraw-Hill.
- Salas, J. D., & Boes, D. C. (1980). Shifting level modeling of hydrologic series. *Advances in Water Resources*, 3(2), 59–63. [https://doi.org/10.1016/0309-1708\(80\)90028-7](https://doi.org/10.1016/0309-1708(80)90028-7)
- Salas, J. D., & Lee, T. (2010). Nonparametric simulation of single-site seasonal streamflows. *Journal of Hydrologic Engineering*, 15(4), 284–296. [https://doi.org/10.1061/\(asce\)he.1943-5584.0000189](https://doi.org/10.1061/(asce)he.1943-5584.0000189)
- Serinaldi, F., Bonaccorso, B., Cancelliere, A., & Grimaldi, S. (2009). Probabilistic characterization of drought properties through copulas. *Physics and Chemistry of the Earth*, 34(10–12), 596–605. <https://doi.org/10.1016/j.pce.2008.09.004>
- Sveinsson, O. G., Salas, J. D., Boes, D. C., & Pielke, R. A., Sr. (2003). Modeling the dynamics of long-term variability of hydroclimatic processes. *Journal of Hydrometeorology*, 4(3), 489–505. [https://doi.org/10.1175/1525-7541\(2003\)004<0489:mtdolv>2.0.co;2](https://doi.org/10.1175/1525-7541(2003)004<0489:mtdolv>2.0.co;2)
- Tao, P. C., & Delleur, J. W. (1976). Multistation, multiyear synthesis of hydrologic time series by disaggregation. *Water Resources Research*, 12(6), 1303–1312. <https://doi.org/10.1029/wr012i006p01303>
- Watras, C. J., Read, J. S., Holman, K. D., Liu, Z., Song, Y. Y., Watras, A. J., et al. (2014). Decadal oscillation of lakes and aquifers in the upper Great Lakes region of North America: Hydroclimatic implications. *Geophysical Research Letters*, 41(2), 456–462. <https://doi.org/10.1002/2013gl058679>
- Wu, Z., & Huang, N. E. (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. <https://doi.org/10.1098/rspa.2003.1221>
- Wu, Z. H., & Huang, N. E. (2005). Statistical significance test of intrinsic mode functions. In N. E. Huang & S. P. Shen (Eds.), *Hilbert-Huang transform and its applications* (pp. 125–148). World Scientific Publishing Company.
- Yevjevich, V. (1967). An objective approach to definitions and investigations of continental hydrologic droughts. *Hydrology Papers*, 23, 1–18.
- Yue, S., & Pilon, P. (2004). A comparison of the power of the t test, Mann-Kendall and bootstrap tests for trend detection. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 49(1), 21–37. <https://doi.org/10.1623/hysj.49.1.21.53996>
- Zhou, Y. (2020). Exploring multidecadal changes in climate and reservoir storage for assessing nonstationarity in flood peaks and risks worldwide by an integrated frequency analysis approach. *Water Research*, 185, 116265. <https://doi.org/10.1016/j.watres.2020.116265>
- Zirulia, A., Brancale, M., Barbagli, A., Guastaldi, E., & Colonna, T. (2021). Hydrological changes: Are they present at local scales? *Rendiconti Lincei*, 32(2), 295–309. <https://doi.org/10.1007/s12210-021-00983-5>