Multivariate Shifting Mean Plus Persistence Model for Simulating the Great Lakes Net Basin Supplies

Óli Grétar Blöndal Sveinsson[1]
The National Power Company (Landsvirkjun), Reykjavík Iceland

Jose D. Salas
Water Resources, Hydrologic and Environmental Sciences Division, Civil Engineering Department, Colorado State University, Fort Collins

Abstract. The focus of this paper is to develop a multivariate model for modeling the annual net basin supplies (NBS) of the Great Lakes. Not all NBS series show similar behavior. For example, a feature that is apparent in some but not all NBS series is a sudden shifting pattern. In this paper previous studies of univariate shifting mean models are expanded to develop multivariate contemporaneous shifting mean models. These multivariate models are further mixed with CARMA models in such a way, that the lag zero correlation in space is conserved between the underlying processes of the different models. The full contemporaneous shifting mean CARMA models are successfully applied for modeling jointly the whole Great Lakes system, preserving the spatial correlation at lag zero between different lakes, and preserving other important statistical characteristics of the individual lakes.

1. Introduction

The Great Lakes System is one of the major lake systems in the world. It involves a series of five interconnected lakes (Superior, Michigan-Huron, St. Clair, Erie, and Ontario) that are subject to inter-basin flows and net basin supplies (NBS). Lake St. Clair is small compared to the other four lakes but being the middle lake, it is strategically located. Lakes Superior and Ontario have been regulated for the past several decades while the intermediate lakes are not regulated, although modifications in the connecting channels have caused some effect on the lake outflows (Quinn, 1985). Regulation of the two lakes depends on the expected NBS. In addition, the regulation of Lake Ontario, being the furthest downstream lake of the system, depends on the characteristics of the entire system, such as the expected NBS for all the lakes, the corresponding lake levels, and outflows. Thus the analysis, modeling and simulation of the NBS series for the various lakes have been of interest not only for testing alternative regulation plans but for re-evaluating the capacity of exist-

[1] The National Power Company
Research and Surveying Division
Háaleitisbraut 68
103 Reykjavík
Iceland
Tel: (+354) 515-9180
e-mail: olis@lv.is

ing waterworks, re-examining the performance of existing water systems, and assessing the capacity of new water resources systems.

Several studies have been made in the past for analyzing and modeling the NBS series of the entire Great Lakes system based on stochastic techniques. The NBS time series show complex patterns that are reflected in some of the statistical characteristics such as the mean, variance, persistence, high flow and low-flow statistics, short and long memory, and shifting level behavior – (Rassam et al., 1992). Some studies have been made attempting to understand and model some of the stochastic features of the NBS series. For example, Buchberger (1994) used a conceptual analysis based on water balance of the lakes to derive covariance properties of the annual NBS series.

Direct and indirect modeling schemes (Salas and Fernandez, 1989) have been proposed and applied for modeling monthly and quarter monthly NBS series (Yevjevich, 1975; Loucks,1989; Buchberger, 1992; Rassam et al., 1992). Direct modeling schemes imply using (for instance) monthly data and building a model to simulate monthly data directly at this time scale. For example, Yevjevich (1975) used a multivariate autoregressive (AR) model after seasonally standardizing the NBS series. The drawback with this type of modeling scheme is that while the monthly statistics are generally well preserved, statistics at higher time scales (for example years) are generally underestimated. Likewise, other statistics related to low frequency components such as random apparent shifts in the series are not represented. On the other hand, indirect modeling schemes imply modeling and generating monthly NBS in two or more steps (stages), that is firstly the time series is modeled at a higher time scale such as years so as to reproduce key annual statistics, subsequently annual NBS series generated from such a model are then disaggregated into smaller time scales such as months in such a way as to reproduce monthly statistics. For example, Rassam et al. (1992) employed one direct and two indirect modeling and generation schemes by using the so-called SPIGOT computer package (Grygier and Stedinger, 1990). The two indirect approaches included, a CARMA(1, 1) model with temporal disaggregation, and a mixture of multivariate AR(1) and shifting mean model with temporal disaggregation. The shifting mean model was included for generating the annual NBS series of Lakes Erie and Ontario because it was capable of reproducing the relevant statistics related to lake levels and outflows better than the other alternatives. The study suggested the need of further developing multisite shifting mean models.

In this paper we develop a multivariate modeling framework using shifting mean models. Two contemporaneous models are developed, namely: the contemporaneous shifting mean model plus AR(1) persistence dubbed as CSMAR(1) and a mixture of contemporaneous shifting mean and a contemporaneous ARMA, dubbed CSMAR(1)-CARMA. The CSMAR(1) model is based on the single site plus persistence model, SMAR(1), suggested by Sveinsson (2006). In addition, simpler versions of the models assuming no direct AR(1) persistence are included. The various models are illustrated and compared using the annual NBS data of the Great Lakes system.

2. Contemporaneous SMAR(1) : CSMAR(1)

The CSMAR(1) model is a contemporaneous SMAR(1) model that can be used to model multiple time series that are correlated in space. For detailed description of the SMAR(1) model refer to Sveinsson (2002) or Sveinsson et al. (2006). If $\mathbf{X}_t = \left[X_t^{(1)} \quad X_t^{(2)} \ldots X_t^{(n)}\right]^{\mathrm{T}}$ is a column vector of observations at time t for n different sites, where each site is assumed to follow a SMAR(1) process, then the CSMAR(1) process can be expressed as

$$\mathbf{X}_t = \mathbf{Y}_t + \mathbf{Z}_t \tag{1}$$

where $\mathbf{Y}_t$ and $\mathbf{Z}_t$ are column vectors defined in the same way as $\mathbf{X}_t$. For a single site the noise level process $\{Z_t\}$ can be written as

$$Z_t = \sum_{i=1}^{t} M_i \; I_{(S_{i-1},S_i]}(t) \tag{2}$$

Where $\{M_t\}_{i=1}^{\infty} \sim iid\ \mathrm{N}\left(0, \sigma_M^2 = \sigma_Z^2\right)$, $S_i = N_1 + N_2 + \cdots + N_i$ with $S_0 = 0$, and $I_{(a,b)}(t)$ is the indicator function equal to one if $t \in (a,b)$ and zero otherwise. The $\{N_t\}_{i=1}^{\infty}$ is a discrete, stationary, delayed-renewal sequence on the positive integers, with $\{N_t\}_{i=1}^{\infty} \sim iid$ Positive Geometric(p) (Sveinsson et al., 2003 and 2005). The cross covariance function (CCVF) of $\{\mathbf{X}_t\}$ at lag h is denoted by

$$\mathbf{C}_{\mathbf{X}}(h) = E[(\mathbf{X}_{t+h} - \mu_{\mathbf{X}})(\mathbf{X}_t - \mu_{\mathbf{X}})^{\mathrm{T}}] = \begin{bmatrix} c_{\mathbf{X}}^{11}(h) & \cdots & c_{\mathbf{X}}^{1n}(h) \\ \vdots & \ddots & \vdots \\ c_{\mathbf{X}}^{n1}(h) & \cdots & c_{\mathbf{X}}^{nn}(h) \end{bmatrix} \tag{3}$$

where $c_{\mathbf{X}}^{ij}(h) = E[(X_{t+h}^{(i)} - \mu_X^{(i)})(X_t^{(j)} - \mu_X^{(j)})]$ is the CCVF at lag h between site i and site j, and $\mu_{\mathbf{X}}$ is the mean vector of $\mathbf{X}_t$. In the CSMAR(1) model the following assumptions are made about the independent sequences $\{\mathbf{Y}_t\}$ and $\{\mathbf{Z}_t\}$:

1. The sequences $\{Y_t^{(1)}\},\{Y_t^{(2)}\},\ldots,\{Y_t^{(n)}\}$ are modeled by a contemporaneous AR(1), CAR(1), process given by $\mathbf{Y}_t - \mu_{\mathbf{Y}} = \Phi(\mathbf{Y}_{t-1} - \mu_{\mathbf{Y}}) + \boldsymbol{\varepsilon}_t$ where Φ is a diagonal $n \times n$ matrix, and $\{\boldsymbol{\varepsilon}_t\} \sim iid\ \mathrm{MVN}\left(\mathbf{0}, \mathbf{C}_{\boldsymbol{\varepsilon}}(0)\right)$. The CCVF of the processes in the CAR(1) model are given by

$$\mathbf{C}_{\boldsymbol{\varepsilon}}(0) = \mathbf{C}_{\mathbf{Y}}(0) - \Phi \mathbf{C}_{\mathbf{Y}}^{\mathrm{T}}(1) \tag{4}$$

$$\mathbf{C}_{\mathbf{Y}}(h) = \Phi^h \mathbf{C}_{\mathbf{Y}}(0) \quad h = 0,1,\ldots \tag{5}$$

   Thus $\mathbf{C}_{\mathbf{Y}}(h)$ has the same decaying structure, with respect to h, in space as in time. That is, $c_{\mathbf{Y}}^{ij}(h) = (\phi^{ii})^h c_{\mathbf{Y}}^{ij}(0)$ for $i,j \in \{1,2,\ldots,n\}$, where ϕ^{ii} is the ith row and ith column element of Φ.

2. The sequences $\{M_i^{(1)}\},\{M_i^{(2)}\},\ldots,\{M_i^{(n)}\}$ are correlated in space only at lag zero. That is, $\{\mathbf{M}_i\} \sim iid \ \text{MVN}(\mathbf{0},\mathbf{C}_\mathbf{M}(0))$. It can be shown that a necessary and sufficient condition for $\{\mathbf{Z}_t\}$ to be stationary in the covariance is that $N_1, N_2, \ldots$ is a common sequence for all sites. In that case the covariance function of $\mathbf{Z}_t$ at lag h is given by
$$\mathbf{C}_\mathbf{Z}(h) = (1-p)^h \mathbf{C}_\mathbf{M}(0) \quad h = 0,1,\ldots \tag{6}$$

The condition that $\{N_t\}_{i=1}^{\infty}$ is a common sequence for all sites may also be supported in practice, if the shifts in the means are thought of being caused by changes in natural processes, such as changes in climate. In such cases it should be expected that time series of the same hydrologic variable within a geographic region would all exhibit shifts at the same times. Thus, in general the CSMAR(1) model should not be applied for multivariate analysis of time series if it is clear that shifts in different time series do not coincide in time. Such cases can come up if a shift in a time series is caused by a construction of a dam or other man made constructions, where the construction does not affect the other time series being analyzed. Note that if $\mathbf{M}_t$ is assumed uncorrelated in space then the condition for stationarity that $\{N_t\}_{i=1}^{\infty}$ is a common sequence for all sites is not necessary any more.

2.1. Parameter Estimation for the CSMAR(1) model

The parameter estimation procedure is relatively simple for the CSMAR(1) model. First the CSMAR(1) model is uncoupled into univariate SMAR(1) models. If the common p is not known, then $p^{(i)}$ is first estimated at each site i (Sveinsson, 2002; and Sveinsson et al., 2006). The common p can then be estimated as a weighted average of the $\hat{p}^{(i)}$ s
$$\hat{p} = \frac{1}{n^{(1)} + n^{(2)} + \cdots + n^{(n)}} \sum_{i=1}^{n} n^{(i)} \hat{p}^{(i)} \tag{7}$$
Given $\hat{p}$ the parameters of each univariate model are reestimated.

After estimating the parameters of the univariate SMAR(1) models, what remains is estimating the non-diagonal elements of $\mathbf{C}_\varepsilon(0)$ and $\mathbf{C}_\mathbf{M}(0)$. Using Eqs (1) and (5)-(6), and the independence of $\{\mathbf{Y}_t\}$ and $\{\mathbf{Z}_t\}$ it follows that
$$\mathbf{C}_\mathbf{X}(h) = \Phi^h \mathbf{C}_\mathbf{Y}(0) + (1-p)^h \mathbf{C}_\mathbf{M}(0) \quad h = 0,1,\ldots \tag{8}$$
Estimates of $\mathbf{C}_\mathbf{M}(0)$ and $\mathbf{C}_\mathbf{Y}(0)$ are obtained by solving Eq (8) with $h = 0$ and $h = 1$ for $\mathbf{C}_\mathbf{M}(0)$ and $\mathbf{C}_\mathbf{Y}(0)$. It follows that
$$\hat{\mathbf{C}}_\mathbf{M}(0) = [\hat{\Phi} - (1-\hat{p})\mathbf{I}]^{-1}(\hat{\Phi}\,\hat{\mathbf{C}}_\mathbf{X}(0) - \hat{\mathbf{C}}_\mathbf{X}(1)) \tag{9}$$
$$\hat{\mathbf{C}}_\mathbf{Y}(0) = \hat{\mathbf{C}}_\mathbf{X}(0) - \hat{\mathbf{C}}_\mathbf{M}(0) \tag{10}$$
Finally using Eqs (4) and (5), $\mathbf{C}_\varepsilon(0)$ is estimated from
$$\hat{\mathbf{C}}_\varepsilon(0) = \hat{\mathbf{C}}_\mathbf{Y}(0) - \hat{\Phi}\,\hat{\mathbf{C}}_\mathbf{Y}^{\mathrm{T}}(0)\,\hat{\Phi}^{\mathrm{T}} \tag{11}$$

2.2. Problems Arising in Parameter Estimation

It is required that $\hat{\mathbf{C}}_{\mathbf{M}}(0)$ and $\hat{\mathbf{C}}_{\boldsymbol{\varepsilon}}(0)$ are symmetric matrixes. In order for $\hat{\mathbf{C}}_{\mathbf{M}}(0)$ in Eq (10) to be symmetric a complex relationship is needed between $\hat{c}^{ij}_{\mathbf{X}}(1)$ and $\hat{c}^{ji}_{\mathbf{X}}(1)$ (Sveinsson, 2002), this is unlikely to be followed by the data. Thus an adjustment is needed to make $\hat{\mathbf{C}}_{\mathbf{M}}(0)$ symmetric, and the simplest such adjustment is to replace all $\hat{c}^{ij}_{\mathbf{M}}(0)$ and $\hat{c}^{ji}_{\mathbf{M}}(0)$ with their respective averages. If $\hat{\mathbf{C}}_{\mathbf{M}}(0)$ is symmetric then no further adjustment is needed in the estimation of $\hat{\mathbf{C}}_{\boldsymbol{\varepsilon}}(0)$.

3. CSMAR(1)-CARMA(*p*, *q*)

Analyzes of multiple time series of different hydrologic variables may require mixing of models. For example shifts in time series of one hydrologic variable may not be present in a time series of another hydrologic variable. Or, if different geographic locations are used for analysis of a single hydrologic variable, then characteristics of the corresponding times series may be dependent on their geographic location. In such cases mixing of multiple CSMAR(1) models and other time series models, such as CARMA(p, q), may be desirable. In this section we will formulate a mixture of one CSMAR(1) model with one CARMA(p, q) model, where the lag zero cross correlation function (CCF) in space is preserved between the CARMA(p, q) model and the CAR(1) component of the CSMAR(1) model.

Lets assume that there are total of n sites, of which n_1 sites follow a CSMAR(1) model and the remaining n_2 sites follow a CARMA(p, q) model. The model of the n sites can be presented by Eq (1), where the first n_1 elements of $\mathbf{X}_t$ represent the CSMAR(1) model and the remaining n_2 elements of $\mathbf{X}_t$ represent the CARMA(p, q) model

$$\begin{bmatrix} X_t^{(1)} \\ \vdots \\ X_t^{(n_1)} \\ X_t^{(n_1+1)} \\ \vdots \\ X_t^{(n)} \end{bmatrix} = \begin{bmatrix} Y_t^{(1)} \\ \vdots \\ Y_t^{(n_1)} \\ Y_t^{(n_1+1)} \\ \vdots \\ Y_t^{(n)} \end{bmatrix} + \begin{bmatrix} Z_t^{(1)} \\ \vdots \\ Z_t^{(n_1)} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{12}$$

In general the whole vector $\mathbf{Y}_t$ can be looked at as being modeled by a CARMA(p, q) model

$$\mathbf{Y}_t - \mu_{\mathbf{Y}} = \sum_{j=1}^{p} \Phi_j (\mathbf{Y}_{t-j} - \mu_{\mathbf{Y}}) + \boldsymbol{\varepsilon}_t - \sum_{j=1}^{q} \Theta_j \, \boldsymbol{\varepsilon}_{t-j} \tag{13}$$

where $\{\boldsymbol{\varepsilon}_t\} \sim iid\ \mathrm{MVN}\big(\mathbf{0}, \mathbf{C}_{\boldsymbol{\varepsilon}}(0)\big)$, and the parameters $\Phi_1, \Phi_2, \ldots, \Phi_p$ and $\Theta_1, \Theta_2, \ldots, \Theta_q$ are diagonal $n \times n$ matrixes. Each of the first n_1 elements of $\mathbf{Y}_t$ is an AR(1) process, and each of the remaining n_2 elements of $\mathbf{Y}_t$ follows some ARMA(p, q) process. That is, $Y_t^{(i)}$ is an ARMA(p_i, q_i) process, $i = 1, 2, \ldots, n$, where the p_i s can be different and the q_i s can be different. The p and the q of the CARMA(p, q) model are $p = \max(p_1, p_2, \ldots, p_n)$ and $q = \max(q_1, q_2, \ldots, q_n)$.

The parameter matrixes of the CARMA(p, q) are diagonal, thus estimation of parameters of the CSMAR(1)-CARMA model can be done in a similar way as for the CSMAR(1) model, where Eq (13) is uncoupled into univariate model. For the CSMAR(1) portion of Eq (13), parameters are estimated using procedures in section 2.1. For estimation of each of the univariate ARMA(p_i, q_i), $i = n_1 + 1, n_1 + 2, \ldots, n$, models refer to Salas (1993; Hipel and McLeod (1994); and Brockwell and Davis (1996). Hipel and McLeod (1994) also give a joint multivariate estimation algorithm for estimation of the parameters of the CARMA(p, q) model. The algorithm to estimate $\mathbf{C}_{\boldsymbol{\varepsilon}}(0)$ is simple, but a necessary condition is that the CARMA(p, q) is causal. This is equivalent to requiring each of the estimated univariate ARMA(p, q) models to be causal (often a common requirement in estimation procedures for ARMA models). Causality implies that $\mathbf{Y}_t$ can be written out as an infinite moving average model. As a result $\mathbf{C}_{\boldsymbol{\varepsilon}}(0)$ is estimated from

$$\mathbf{C}_{\mathbf{Y}}(0) = \sum_{j=0}^{\infty} \Psi_j \mathbf{C}_{\boldsymbol{\varepsilon}}(0) \Psi_j^{\mathrm{T}} \tag{14}$$

where Ψ_j are matrixes with absolutely summable elements given by $\Psi_0 = I$ and $\Psi_j = -\Theta_j + \sum_{k=1}^{p} \Phi_k \Psi_j^{\mathrm{T}}$, where $\Psi_j = \mathbf{0}$ for $j < 0$ and $\Theta_j = \mathbf{0}$ for $j > q$. For detailed information refer to Sveinsson (2002), where exact equations are given for estimating the elements of $\mathbf{C}_{\boldsymbol{\varepsilon}}(0)$.

4. The Special Case : The CSM-1-CARMA Model

In the special case with $\phi = 0$ (no persistence in the Y_t process) the CSMAR(1) model in section 2 reduces to a contemporaneous SM-1 model, dubbed here as CSM-1. Thus, the sequences $\{Y_t^{(1)}\}, \{Y_t^{(2)}\}, \ldots, \{Y_t^{(n)}\}$ are correlated in space at lag 0 only, and independent in time, with $\{\mathbf{Y}_t\} \sim iid\ \mathrm{MVN}(\mu_{\mathbf{Y}}, \mathbf{C}_{\mathbf{Y}}(0))$. The properties of the $\{M_i^{(1)}\}, \{M_i^{(2)}\}, \ldots, \{M_i^{(n)}\}$ do not change.

4.1. Parameter Estimation for the CSM-1 model

The parameter estimation procedure for the CSM-1 model follows the same steps as the parameter estimation procedure for the CSMAR(1) model in section 2.1. That is, first the CSM-1 is coupled into univariate SM-1 models and the parameters are estimated at each site using procedures in Sveinsson et al. (2006). Then the common p for all sites is estimated as a weighted average of the estimated $p^{(i)}$ of the univariate SM-1 models (refer to Eq (7)), and given $\hat{p}$ the parameters of the univariate SM-1 models are reestimated. What remains is estimating the non-diagonal elements of $\mathbf{C_Y}(0)$ and $\mathbf{C_M}(0)$. The $\mathbf{C_M}(0)$ is estimated from

$$\hat{\mathbf{C}}_{\mathbf{M}}(0) = (1-\hat{p})^{-1}\hat{\mathbf{C}}_{\mathbf{X}}(1) \tag{15}$$

where if necessary $\mathbf{C_M}(0)$ is made symmetric by replacing $\hat{c}^{ij}_{\mathbf{M}}(0)$ and $\hat{c}^{ji}_{\mathbf{M}}(0)$ with their respective averages. Then $\mathbf{C_Y}(0)$ is estimated from

$$\hat{\mathbf{C}}_{\mathbf{Y}}(0) = \hat{\mathbf{C}}_{\mathbf{X}}(0) - \hat{\mathbf{C}}_{\mathbf{M}}(0) \tag{16}$$

4.2 Parameter Estimation for the CSM-1-CARMA model

The CSM-1-CARMA follows the same concept as the CSMAR(1)-CARMA model in section 3. Given the CSM-1 model then parameters of the CSM-1-CARMA model are estimated using the procedures for estimation of the CSMAR(1)-CARMA parameters in section 3, where each of the elements of $\{\mathbf{Y}_t\}$ corresponding to the CSM-1 process is looked at as being modeled by an ARMA(0, 0) process.

5. The Great Lakes System

The intent here is to fit a multivariate model to the annual net basin supplies (NBS) of the lakes in the Great Lakes system using the procedures presented in this paper. The data were obtained from Hydro-Quebec, and span the period 1900-1999 for lakes Erie, Michigan-Huron, Ontario, and Superior, and the period 1900-1989 for Lake St. Clair. Data post 1989 for Lake St. Clair were still preliminary, and hence are not used in this study. The annual NBS time series of the Great Lakes and their ACFs can be seen in Fig. 1. The data for Lake Superior and Lake Michigan-Huron in do not seem to exhibit any sudden shifts, and in addition the ACFs of the data do not have shapes that are expected of the SMAR(1) model. On the other hand, the data for the other lakes in appear to be characterized by sudden shifts. Furthermore, the cross correlation coefficients at lag zero are significant across all lakes (not shown). Thus, contemporaneous models could be used to preserve the lag zero cross correlation coefficient between different lakes.

The sample mean, standard deviation, skewness, Hurst slope K, storage capacity SC, and the longest drought length DL and the corresponding drought magnitude DM based on demand level $d = \hat{\mu}_X$ of the Great Lakes data are shown in Table 1.

Table 1. Sample statistics of the Great Lakes NBS time series. Fitted ACFs for the CSMAR(1) and the CSM-1 models are also shown.

Statistic	Erie	Ontario	St. Clair	Michigan-H.	Superior
$\hat{\mu}_X$ [cms]	574.1	1033	121.7	3177	2043
$\hat{\sigma}_X$ [cms]	265.4	241.6	63.34	737.0	478.8
$\hat{\gamma}_X$ [cms]	0.138	0.491	0.311	-0.091	0.033
K	0.787	0.786	0.847	0.713	0.654
SC [cms]	5506	5083	1529	11978	4755
DL	8	11	9	8	5
DM [cms]	1720	2034	659.3	6029	3850

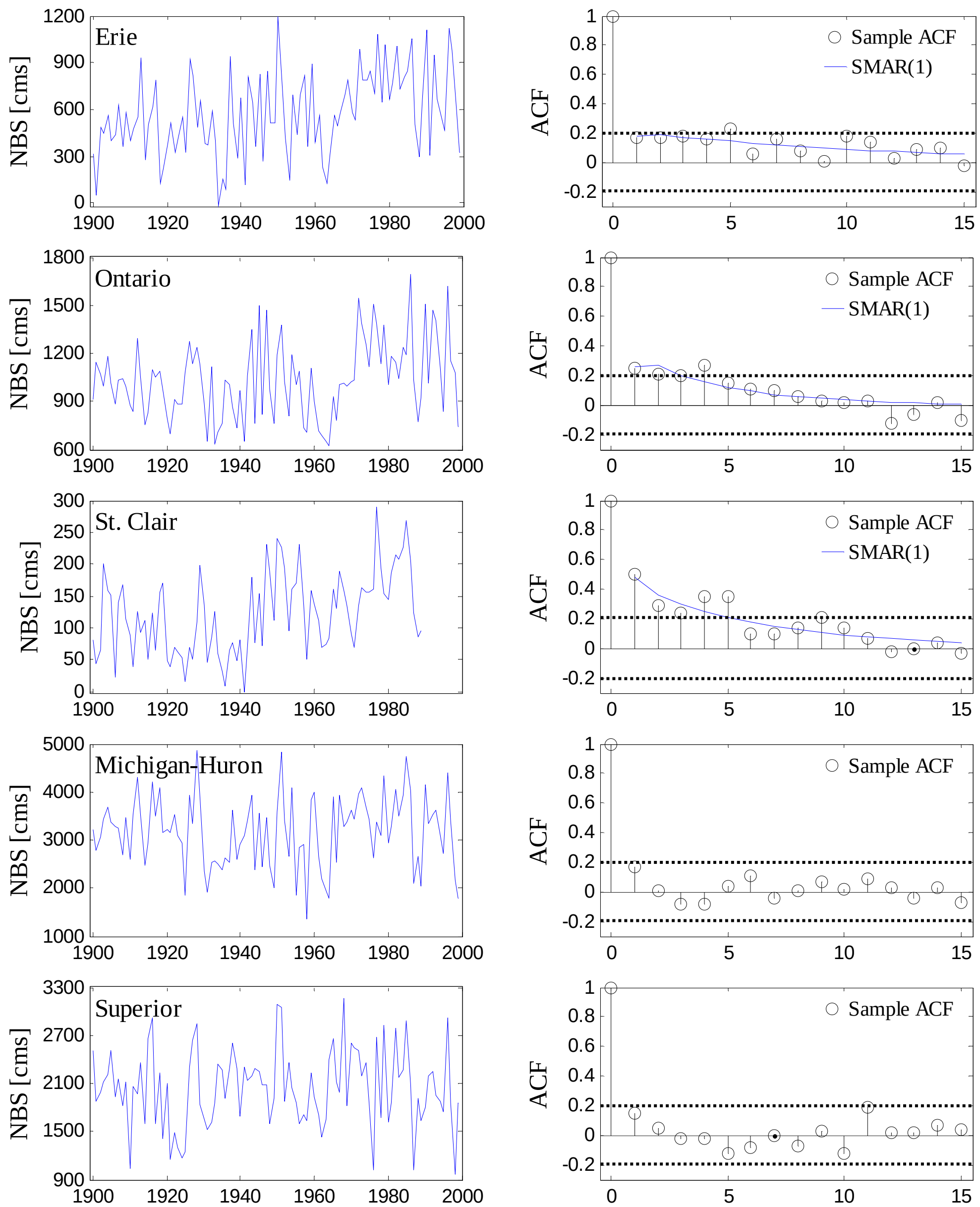

Figure 1. Time series of NBS of the Great Lakes System and the corresponding ACFs.

5.1 Fitting a CSMAR(1)-CARMA(p, q) model to the Great Lakes

We will attempt to fit a mixture of CSMAR(1) and a CARMA(p, q) model to the data, where the lakes Erie, Ontario, and St. Clair will be modeled by a CSMAR(1) model, and the lakes Michigan-Huron and Superior will be modeled by CARMA(p, q) model. The ACF and the partial ACF (not shown) of lakes Michigan-Huron and Superior in Fig. 1 suggests a CARMA(0, 0) model (a bivariate normal model) or a CARMA(1, 0) model. Lake St. Clair has 10 year shorter record than the other lakes. Thus, the models are fitted to sample series of different lengths. The approach used here preserves the variance of the full records and the lag zero correlation between concurrent records. For further information refer to Sveinsson (2004).

For the purpose of this study a CARMA(0, 0) model was selected for modeling lakes Michigan-Huron and Superior. The CSMAR(1)-CARMA and the CSM(1)-CARMA models were both fitted to the data, but only the estimated parameters of the CSMAR(1)-CARMA model are shown in Table 2. The fitted ACFs used in parameter estimation of the univariate SMAR(1) models are shown in Fig. 1.

To analyze how capable the fitted models are in preserving the sample statistics used in the fitting procedures, 1,000 realizations of the same lengths as the historical records were generated for the full models in Table 2. All the

Table 2. Parameters of the fitted CSMAR(1)-CARMA model.

Parameter	Erie	Ontario	St. Clair	Michigan-H.	Superior
$\hat{\varphi}_X$	-0.1170	-0.0162	0.1262		
$\hat{p}$	0.1574	0.1574	0.1574		
$\hat{\mu}_Y$ [cms]	574.1	1033	121.7	3177	2043
$\hat{\mathbf{C}}_{\mathbf{M}}(0)$ [cms]	22401	18016	5519		
	18016	18121	3973		
	5519	3973	2002		
$\hat{\mathbf{C}}_{\mathbf{Y}}(0)$ [cms]	48058	26654	3445	103113	37944
	26654	40238	4323	114183	31060
	3445	4323	2010	20964	7431
	103113	114183	20964	543228	195286
	37944	31060	7431	195286	229280
$\hat{\mathbf{C}}_{\boldsymbol{\varepsilon}}(0)$ [cms]	47400	26603	3496	103113	37944
	26603	40228	4331	114183	31060
	3496	4331	1978	20964	7431
	103113	114183	20964	543228	195286
	37944	31060	7431	195286	229280

lakes have historical records of the same length, n = 100, except Lake St. Clair, which has a record length n = 90. Thus the generated records for Lake St. Clair were truncated to match the length of the historical record. The average sample statistics of the 1,000 generated realizations are shown in Table 3. Comparing with the historical sample statistics in Table 1, the mean and the standard deviation are well preserved for all lakes. Comparing the storage related statistics *K*, *SC*, *DL*, and *DM* in Table 3 with the corresponding historical statistics in Table 1, it can be said that they are in general relatively well preserved. Comparing the storage related statistics of the two fitted models, then the CSMAR(1)-CARMA and the CSM-1-CARMA give very similar results. A reason for the similarity may be that the ϕ parameters are close to zero in the CSMAR(1) part of the CSMAR(1)-CARMA model.

In Table 4 the lag 0 and lag 1 historical CCF matrixes are shown along with the corresponding CCF matrixes based on the 1,000 realizations of length

n. Comparing the CCF matrixes based on the generated sequences with the historical CCF matrixes, then as expected the lag 0 CCF is very well preserved between all stations. The lag 1 CCF for the CSMAR(1) part of the model (the upper 3 × 3 submatrix of $\hat{\rho}_{\mathbf{X}}(1)$) may not be exactly preserved due to the adjustments to $\hat{\mathbf{C}}_{\mathbf{M}}(0)$ and $\hat{\mathbf{C}}_{\boldsymbol{\varepsilon}}(0)$ to make them symmetric, but in general the off-diagonal averages of $\hat{\rho}^{ij}_{\mathbf{X}}(1)$ and $\hat{\rho}^{ji}_{\mathbf{X}}(1)$ should be relatively well preserved. The values of the lag 1 CCF in Table 4 support this. Note that any lag 1 CCF including Lake Michigan-Huron or Lake Superior is not expected to be preserved. Comparing the results among the two different models (results not shown for the CSM(1)-CARMA model), then again both models give similar results.

Table 3. Average sample statistics of 1,000 generated NBS series of the Great Lakes of the same lengths as the historical records.

Statistic	Erie	Ontario	St. Clair	Michigan-H.	Superior
			CSMAR(1)–CARMA Model		
$\hat{\mu}_X$ [cms]	573.3	1033	121.9	3175	2042
$\hat{\sigma}_X$ [cms]	261.2	237.1	61.33	732.2	474.9
$\hat{\gamma}_X$ [cms]	-0.014	0.006	-0.011	-0.010	0.004
K	0.727	0.730	0.779	0.616	0.617
SC [cms]	4877	4484	1290	8434	5475
DL	8.660	8.794	10.87	5.931	5.986
DM [cms]	2379	2206	760.4	4044	2624
			CSM(1)–CARMA Model		
$\hat{\mu}_X$ [cms]	573.3	1033	121.9	3175	2041
$\hat{\sigma}_X$ [cms]	261.3	237.0	61.31	732.1	474.9
$\hat{\gamma}_X$ [cms]	-0.013	0.008	-0.006	-0.010	0.004
K	0.728	0.729	0.779	0.616	0.617
SC [cms]	4890	4490	1300	8433	5476
DL	8.612	8.707	10.80	5.934	5.991
DM [cms]	2385	2184	752.5	4036	2624

6. Summary and Final Remarks

In this paper a multivariate shifting mean modeling framework was developed. More precisely, a contemporaneous version of the univariate shifting mean autoregressive AR(1) model, SMAR(1), in Sveinsson et al. (2006), was developed and dubbed as CSMAR(1). In addition, a general contemporaneous model mixing CSMAR(1) and CARMA models was developed for modeling of systems, where some of the sites exhibit sudden shifting patterns while others do not. This model was dubbed as CSMAR(1)-CARMA. The special cases of the above models assuming no direct AR(1) persistence in the CSMAR(1) model were also developed. The special cases were, dubbed as CSM-1 and CSM-1-CARMA. A necessary condition for stationarity of the CSMAR(1) is that the sequence of the mean level lengths is common for all sites, that is that shifts at different sites coincide in time. The above models are capable of preserving the lag zero cross correlation in space between different sites. In addition, for sites modeled by the CSMAR(1) or the CSM-1 models, some characteristics related to the lag one cross correlation in space are also preserved.

Table 4. Historical and generated cross correlation function (CCF) matrixes. The generated CCF is the average of 1,000 generated NBS series of the same lengths as the historical records.

Statistic	Erie	Ontario	St. Clair	Michigan-H.	Superior
	Historical CCF Matrixes				
$\hat{\rho}_{\mathbf{X}}(0)$	1	0.697	0.549	0.527	0.299
	0.697	1	0.569	0.641	0.269
	0.549	0.569	1	0.452	0.245
	0.527	0.641	0.452	1	0.553
	0.299	0.269	0.245	0.553	1
$\hat{\rho}_{\mathbf{X}}(1)$	0.173	0.198	0.144	0.030	0.156
	0.220	0.250	0.084	0.151	0.255
	0.393	0.380	0.504	0.196	0.240
	0.181	0.129	0.027	0.168	0.322
	0.006	-0.014	0.036	-0.066	0.153
	CSMAR(1)–CARMA: average generated CCF Matrixes				
$\hat{\rho}_{\mathbf{X}}(0)$	1	0.692	0.521	0.536	0.301
	0.692	1	0.537	0.652	0.271
	0.521	0.537	1	0.461	0.251
	0.536	0.652	0.461	1	0.552
	0.301	0.271	0.251	0.552	1
$\hat{\rho}_{\mathbf{X}}(1)$	0.142	0.164	0.226	-0.065	-0.036
	0.204	0.204	0.189	-0.018	-0.005
	0.275	0.232	0.417	0.052	0.029
	-0.002	-0.004	-0.002	-0.009	0.000
	-0.001	0.003	-0.001	-0.003	-0.004

Historical records of some of the lakes in the Great Lakes system show evidence of sudden shifts in addition to autocorrelation, while records for other lakes do not indicate such behavior. The proposed models, where applied for modeling jointly the Great Lakes system as a whole, with lakes Erie, Ontario, and St. Clair modeled by contemporaneous shifting mean models, and lakes Michigan-Huron and Superior modeled by a CARMA(0,0) model. The models were capable of preserving the lag zero spatial correlation between different lakes, in addition to preserving other important statistical characteristics of the individual lakes.

As a general conclusion, the proposed mixture models mixing contemporaneous shifting mean models and contemporaneous ARMA models appear to be robust and seem to have a wide range of applicability for modeling of hydro-climatic and geophysical systems.

Acknowledgements. Support from the Colorado Agricultural Experiment Station project on ``Predictability of Extreme Hydrologic Events Related to Colorado's Agriculture", National Science Foundation grant CMS-9625685 on ``Uncertainty and Risk Analysis Under Extreme Hydrologic Events", and support from Hydro-Quebec are gratefully acknowledged.

References

Brockwell, P. J. and R. A. Davis, 1996: *Introduction to Time Series and Forecasting*. Springer Texts in Statistics. Springer, first edition.

Buchberger, S., 1992: *Modeling and forecasting Great Lakes monthly net basin supplies*. Technical report, U.S. Army Corps of Engineers, Detroit, MI. 48pp.

Buchberger, S., 1994: Modeling and forecasting of Great Lakes annual net basin supplies. *Water Resources Research*, **30**(10):2725-2735.

Grygier, J. C. and J. R. Stedinger, 1990: *SPIGOT a Synthetic Streamflow Generation Software Package*, Technical Description, Version 2.6. Cornell University, Ithaca, New York.

Hipel, K. W. and A. I. McLeod, 1994: *Time Series Modelling of Water Resources and Environmental Systems*. Number 45 in Developments in Water Science. Elsevier, first edition.

Loucks, E., 1989: *Modeling the Great Lakes Hydrologic-Hydraulic System*. PhD thesis, University of Wisconsin.

Quinn, F. H., 1985: Temporal effects of St. Clair river dredging on lakes St. Clair and Erie water levels and connecting channel flow. *Journal of Great Lakes Research*, **11**(3):400-403.

Rassam, J.-C., Faherazzi, L. D., Bobée, B., Mathier, L., Roy, R., and L. Carballada, 1992: *Beauharnois-Les Cedres spillway: Design flood study with stochastic approach*. Final report to the experts committee, Hydro-Québec, Montreal, Quebec, Canada. 105pp.

Salas, J. D., 1993: *Analysis and Modeling of Hydrologic Time Series*, chapter 19. Handbook of Hydrology. McGraw-Hill.

Sveinsson, O. G. B., 2002: *Modeling of stationary and non-stationary hydrologic processes*. PhD thesis, Colorado State University, Fort Collins, Co.

Sveinsson, O. G. B., 2004: *Unequal record lengths in SAMS*. Unpublished manuscript (please contact author).

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

Sveinsson, O. G. B., Salas, J. D., and D. C. Boes, 2005: Prediction of extreme events in Hydrologic Processes that exhibit abrupt shifting patterns. *Journal of Hydrologic Engineering*, **10**(4):315-326.

Sveinsson, O. G. B., Salas, J. D., and V. Fortin, 2006: Shifting mean plus persistence model for simulating the great lakes net basin supplies. *Journal of Hydrologic Engineering*, submitted.

Yevjevich, V., 1975: *Generation of hydrologic samples-case study of the Great Lakes*. Hydrology Paper 72, Colorado State University, Fort Collins, CO. 39pp.