Water Resources Systems—Water Availability and Global Change (Proceedings of symposium HS02a held during IUGG2003 at Sapporo, July 2003). IAHS Publ. no. 280, 2003.

Streamflow generation using a multivariate hybrid time series model

SEUNG YUP RIEU, YOUNG-OH KIM
School of Civil, Urban, Geosystem Engineering, Seoul National University, San 56-1 Shillim-dong, Gwanak-gu, Seoul, Korea Republic
rieu@snu.ac.kr

DONG RYUL LEE
Water Resources and Environmental Research Division, Korea Institute of Construction Technology, 2311 Daewha-dong, Ilsan-gu, Koyang-shi, Kyonggi-do, Korea Republic

Abstract The long-term water resources plan in Korea is usually established on the basis of the so-called water balance analysis that estimates the water availability over an entire basin. The analysis generally uses only a single 30-year historical record of natural streamflow. This study suggested that the water balance analysis should include the hydrological uncertainty in its results. For this purpose, a multivariate streamflow generation model called the multivariate contemporaneous PAR(1)NT-hybrid model was proposed and applied to a multi-site monthly streamflow generation problem for the Han River basin in Korea. The proposed model was then compared with a multivariate PAR(1) and a disaggregation model. This study showed that the proposed model reproduces cross-correlations at various lags better than the traditional parametric generation models.

Key words bootstrap; generation; Han River, Korea; hybrid; multivariate; stochastic process

INTRODUCTION

The long-term water resources plan in Korea is established every 10 years on the basis of a water balance analysis that estimates the water availability over an entire basin (KICT, 1999). The analysis generally uses only a single 30-year historical record of natural streamflow; the estimation of the corresponding 30-year water supply reliability is often controversial. It has been recommended recently that the reliability should be estimated based on more streamflow data sets. Therefore, this study selected an appropriate time series generation model to reproduce a number of synthetic streamflow series. The selected generation model should be able to preserve various statistical characteristics of a historical series, especially correlation structures in space and in time.

One can simply use a traditional parametric model such as ARMA (Auto-Regressive Moving Average) for the streamflow generation but this type of generation model assumes a normal distribution of the data that is seldom satisfied in hydrology. Furthermore, a ARMA(p,q) model cannot reproduce autocorrelations at higher lags than lag-p. Recently, Srinivas & Srinivasan (2001a,b) proposed a hybrid approach that combines the traditional parametric model with the post-blackening method proposed by Davison & Hinkely (1997). They reported that their hybrid model with the moving block resampling of the residual series can preserve basic statistics and the correlation structure of the historical data. The major advantage of the nonparametric approach in hydrological time series modelling is that the historical data do not need to be transformed to satisfy the assumption of normality. Furthermore, the hybrid time series model can preserve various correlation structures of the original data by the proper selection of the length of the moving blocks, even though the data have a complex dependence structure.

This study selected the hybrid model as a streamflow generation tool, extended it to a multi-site streamflow generation problem, and attempted to test if the hybrid model is able to reproduce the cross-correlation as well as the autocorrelation.

MULTIVARIATE CONTEMPORANEOUS PAR(1)NT-HYBRID MODEL

The MCPAR(1)NT (Multivariate Contemporaneous Periodic AutoRegressive(1) No Transformation)-hybrid model proposed in this study uses a simple multivariate contemporaneous PAR(1) (Salas *et al.*, 1980) as a parametric constituent of the model and the residual resampling scheme based on the moving block bootstrap as a nonparametric constituent. The modelling steps are as follows:

(a) Remove the non-stationarity of a time series, if non-stationary characteristics such as trend and cyclicity exist in the series, by a preliminary data analysis.

(b) Standardize the series $\mathbf{Y}_{\nu,\tau}$ to remove the periodicity:

$$\mathbf{Y}_{\nu,\tau} = \boldsymbol{\mu}_\tau + \boldsymbol{\sigma}_\tau \mathbf{Z}_{\nu,\tau} \tag{1}$$

where $\boldsymbol{\mu}_\tau$ and $\boldsymbol{\sigma}_\tau$ are the $(n \times 1)$ vectors representing periodic mean and standard deviation of season τ, respectively, $\mathbf{Y}_{\nu,\tau}$ is a $(n \times 1)$ vector of the original seasonal data and $\mathbf{Z}_{\nu,\tau}$ is the standardized data.

(c) Pre-whiten the series at each site with an univariate PAR(1)NT-hybrid model:

$$\mathbf{Z}_{\nu,\tau} = \mathbf{A}_{1,\tau}\mathbf{Z}_{\nu,\tau-1} + \boldsymbol{\varepsilon}_{\nu,\tau} \tag{2}$$

where $\mathbf{A}_{1,\tau}$ is an $(n \times n)$ matrix of lag–1 autoregressive coefficient and $\boldsymbol{\varepsilon}_{\nu,\tau}$ is an $(n \times 1)$ vector of residuals. Equation (2) can be re-written as follows:

$$\begin{bmatrix} z_{\nu,\tau}^{(1)} \\ z_{\nu,\tau}^{(2)} \\ \vdots \\ z_{\nu,\tau}^{(n)} \end{bmatrix} = \begin{bmatrix} a_{\nu,\tau}^{(1)} & 0 & \cdots & 0 \\ 0 & a_{\nu,\tau}^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{\nu,\tau}^{(n)} \end{bmatrix} \begin{bmatrix} z_{\nu,\tau}^{(1)} \\ z_{\nu,\tau}^{(2)} \\ \vdots \\ z_{\nu,\tau}^{(n)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{\nu,\tau}^{(1)} \\ \varepsilon_{\nu,\tau}^{(2)} \\ \vdots \\ \varepsilon_{\nu,\tau}^{(n)} \end{bmatrix}$$

where n is the number of site and $a_{\nu,\tau}^{(n)}$ is a lag–1 autoregressive coefficient for site n and season τ.

The autoregressive coefficient can be estimated by:

$$a_{\nu,\tau}^{(n)} = \gamma_{1,\tau}^{(n)} = \left(\sum \left(x_{\nu,\tau} - \bar{x}_{\nu,\tau}\right)\left(x_{\nu,\tau-1} - \bar{x}_{\nu,\tau-1}\right)\right) / \left(s_\tau \cdot s_{\tau-1}\right) / N$$

where $\gamma_{1,\tau}^{(n)}$ is a lag–1 autocorrelation coefficient for site n and season τ, s_τ is the standard deviation for season τ, and N is the length of the time series.

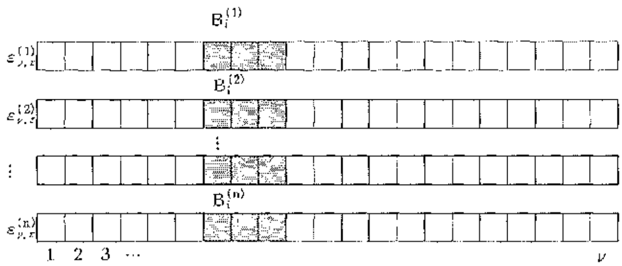

Fig. 1 Resampling scheme of the MCPAR(1)NT-hybrid model. The simulated innovations are collected randomly from the moving blocks, $B_i^{(n)}$ at the same time step across each site.

(d) Generate the innovations (or residuals) by moving block bootstrap. To maintain the cross dependence between sites, the length of moving blocks should be same at each site. Figure 1 shows the method of obtaining the simulated innovations for the multi-site and multi-season data:

$$\boldsymbol{\varepsilon}_{\nu,\tau} = \mathbf{Z}_{\nu,\tau} - \mathbf{A}_{1,\tau}\mathbf{Z}_{\nu,\tau-1} \tag{3}$$

(e) Generate the series $\mathbf{Z}_{\nu,\tau}$ from equation (2) using the simulated innovation $\boldsymbol{\varepsilon}_{\nu,\tau}$ and then generate the series $\mathbf{Y}_{\nu,\tau}$ by the inverse standardization of equation (1).

APPLICATION

The proposed model was applied to monthly streamflow series of three sub-basins of the Han River basin. The official codes of the sub-basins are basin numbers 1007, 1013 and 1015. Data exist for the 30-year period from 1967 to 1996. The preliminary data analysis suggested that the Han River sub-basin data did not show any evidence of long-term persistence.

To test the performance of the proposed model this study compared three generation models: MPAR(1) (Multivariate Periodic AR(1)), Lane's condensed disaggregation, and the MCPAR(1)-hybrid models. Each model generated 300 sets of 30-year monthly series for the Han River sub-basins. SAMS2000 (Salas *et al.*, 2000) was used for calibrating the MPAR(1) and the Lane's condensed disaggregation models. The disaggregation model generated annual series at each site and then disaggregated the annual series to monthly series using Lane's approach (Salas *et al.*, 1980). For the parametric models such as the MPAR(1) and the Lane's condensed disaggregation models, the original data were transformed with a lognormal distribution on a monthly basis to meet the normality assumption. This transformation may cause distortion of the required cross-correlation structures of the original data.

RESULTS

Table 1 shows that there is little difference in periodic means between models but the MCPAR(1)NT-hybrid model preserves the periodic variance and the skewness coefficient better than the other two models.

Table 1 Relative biases in periodic mean, variance, and skewness coefficient. The underlined relative bias indicates the minimum value between models.

Month	MCPAR(1)NT-hybrid: Mean	Variance	Skewness	MPAR(1): Mean	Variance	Skewness	Disaggregation: Mean	Variance	Skewness
1	0.01	0.00	−0.33	0.00	0.01	−0.07	0.00	0.01	−0.05
2	0.02	0.00	−0.20	−0.01	−0.08	−0.19	0.00	−0.04	−0.18
3	0.00	−0.04	−0.25	−0.01	−0.11	−0.50	0.00	−0.08	−0.46
4	0.01	−0.02	−0.10	−0.01	−0.09	−0.11	−0.01	−0.07	−0.09
5	0.03	−0.01	−0.24	0.00	−0.05	−0.29	0.00	−0.05	−0.27
6	−0.01	−0.07	−0.18	0.01	−0.02	−0.23	0.01	0.00	−0.22
7	0.00	0.01	−0.09	0.01	0.05	1.29	0.00	0.03	1.19
8	0.00	−0.06	−0.12	0.00	−0.03	−0.15	0.00	−0.02	−0.13
9	0.02	−0.04	−0.19	0.02	−0.03	−0.29	0.00	−0.05	−0.28
10	0.02	−0.02	−0.13	0.00	−0.07	−0.19	−0.01	−0.07	−0.17
11	0.00	0.00	−0.10	0.00	0.00	−0.05	0.00	0.01	0.01
12	−0.01	−0.04	−0.28	0.00	−0.01	−0.37	0.00	−0.02	−0.38

Figure 2 compares the cross-correlation coefficients at various lags. Since the parametric models such as the MPAR(1) and the disaggregation models are able to

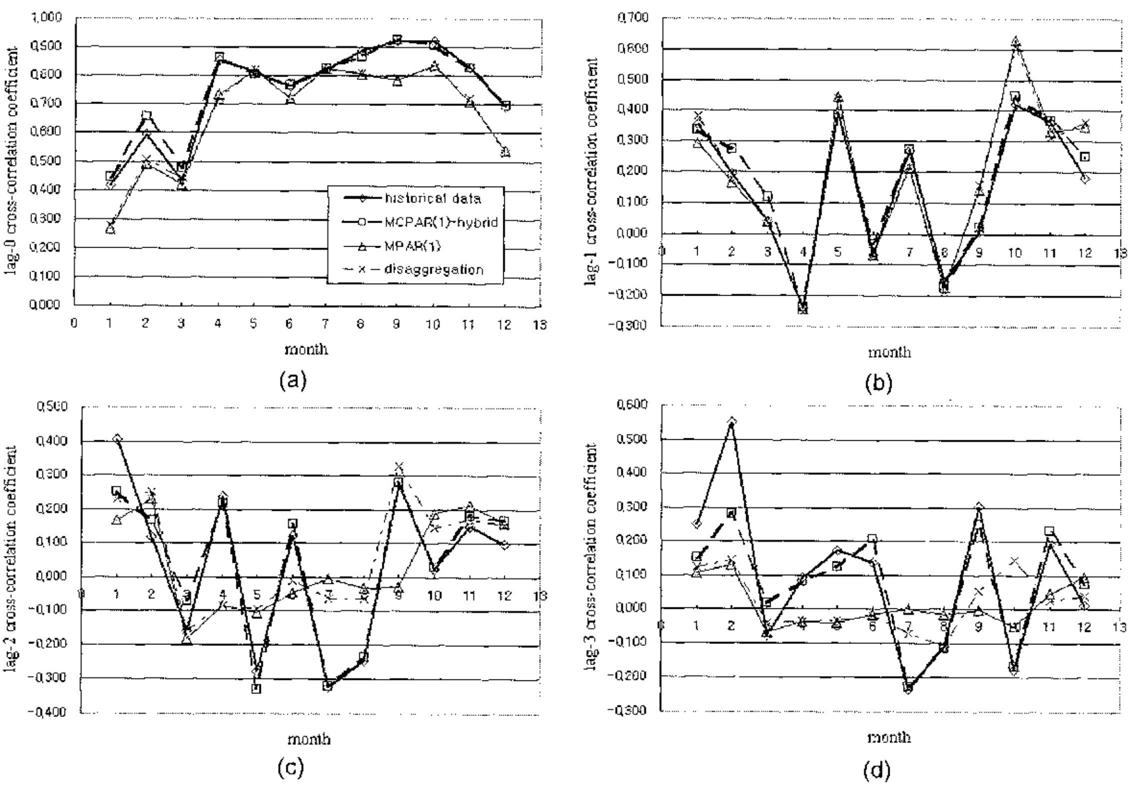

Fig. 2 Comparison of the historical and the generated periodic cross-correlation coefficients at (a) lag–0, (b) lag–1, (c) lag–2, and (d) lag–3.

preserve only the lag–0 and lag–1 cross-correlation structures due to their inherent modelling characteristic, they performed very poor at higher lags (Fig. 2(c) and (d)) than lag–1. As expected, all three models are very competitive at lag–1 (Fig. 2(b)) but the proposed hybrid model is superior to the other two models at a couple of months even in Fig. 2(b). The similar result is shown in Fig. 2(a). This is because the hybrid effect with no transformation improves performance of the parametric constituent although the proposed hybrid model uses a simple multivariate contemporaneous PAR(1) model that can only preserve the lag–0 dependence in space.

Acknowledgements This research was supported by a grant (code no. 1-5-1) from the Sustainable Water Resources Research Centre of the Twenty-first Century Frontier Research Programme of Korea.

REFERENCES

Box, G. E. P. & Jenkins, G. M. (1976) *Time Series Analysis Forecasting and Control* (2nd edn). Holden-Day, San Francisco, California, USA.

Davison, A. C. & Hinkley, D. V. (1997) *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge, Massachusetts, USA.

Korea Institute of Construction Technology (1999) *A Study on the Optimal Water Resources Planning (III)*. Korea Water Resources Corporation, Korea Republic (in Korean).

Salas, J. D., Delleur, J. W., Yevjevich, V. & Lane, W. L. (1980) *Applied Modeling of Hydrologic Time Series*. Water Resources Publications, Littleton, Colorado, USA.

Salas, J. D., Saada, N., Chung, C. H., Lane, W. L. & Frevert, D. K. (2000) *Stochastic Analysis, Modeling, and Simulation (SAMS) Version 2000 User's Manual*. Water Resources, Hydrologic and Environmental Research Center, Fort Collins, Colorado, USA.

Srinivas, V. V. & Srinivasan, K. (2001a) Post-blackening approach for modeling periodic streamflow. *J. Hydrol.* **241**, 221–269.

Srinivas, V. V. & Srinivasan, K. (2001b) A hybrid stochastic model for multiseason streamflow simulation. *Water Resour. Res.* **37**, 2537–2549.