A Decomposed MAR(1) Model For Generating Multi-site Multi-season Flows

Chavalit Chaleeraktrakoon and Sakdiporn Yaklay
Department of Civil Engineering, Faculty of Engineering Thammasat University
Klong Luang, Pathumthani 12121, Thailand
E-mail cchava@engr.tu.ac.th

Abstract : This paper proposes a decomposed MAR(1) model for generating simultaneously seasonal flows at several sites. The proposed DMAR(1) model generally applies an MAR(1) model to each sub-matrix of standardized principal components of normalized flows. The model can reproduce adequately the basic statistical properties and drought related statistics of historic flows at monthly and annual levels. In this paper, the model has been shown its performances of sufficiently describing the flow statistics at tri-monthly time scales with three monthly-flow series [P.12 (Ping River), Y.1C (Yom River) and SK (Nan River)]. It can be thus concluded that the DMAR(1) model is viable for the multi-site multi-season problem. In addition, the uncertainties of the basic flow statistics are analyzed.

1. Introduction

Generation of multi-site multi-season flows is usually necessary for design and operational studies of water resource systems (e.g., design of reservoir storage capacities, estimation of optimal operating policies, and calculation of long-term deficit risks in water supply and hydroelectric generation). A number of stochastic models have been proposed in water resource literature in order to generate such flows. In particular, the multivariate autoregressive (MAR) models {e.g., [1] and [2]}, and the multivariate autoregressive and moving average models {e.g., [3] and [4]} have been widely used. Generally, these models can preserve the basic statistical properties (mean, variance, and the first few lags of temporal and spatial correlations) of the historic flow data considered. Unfortunately, the accuracy of the models is limited to the time scale of the data used. That is, the inference and deduction based on these models must be restricted to the particular aggregation level of the flow data from which these models are developed. For example, if observed monthly flows are used, the monthly basic properties will be preserved but the historic statistics at successive aggregation (e.g., seasonal and annual) levels are not reproduced. The inadequate representation of the aggregate flow properties may cause an inaccurate estimation of system responses; such as storage capacities, operating policies, and deficit risks; and consequently lead to serious errors in the design and planning of the reservoir system {[5] and [6]}.

To reproduce the basic flow statistics at various time intervals, full disaggregation models {e.g., [7] and [8]} have been proposed. The models disaggregate higher-level (e.g., annual) flows, generated using a suitable stochastic model (e.g., ARMA), to lower-level (seasonal) flows. The historic basic statistics of the higher-level (aggregate) flows are assured by the stochastic model used, while the statistical properties of the lower-level (disaggregate) flows, and between the flows at the two different levels are guaranteed by the disaggregation structure. However, when dealing with the multi-site flow generation, the application of the full disaggregation models are limited to the number of sites considered. The number of parameters required in these disaggregation models may exceed easily the number of available information because, in addition to the basic statistical properties of each individual flow series, the spatial correlation structure between the flow processes at different locations must be taken into consideration. Several condensed disaggregation models {e.g., [9], [10] and [11]} have been thus proposed for simulating

RECEIVED 24 June, 2005
ACCEPTED 11 January, 2006

simultaneously multi-site multi-season flows. In general, these models can preserve the basic statistical properties at the two particular levels (e.g., monthly and annual time intervals) considered. Unfortunately, the generated monthly flows do not preserve the observed basic properties at the intermediate (partial) aggregation levels (e.g., bimonthly, tri-monthly, quarterly and semi-annual time scales).

This paper thus proposes a decomposed MAR of order one, DMAR(1), model for generating simultaneously multi-site multi-season flows. The DMAR(1) model generally fits a traditional MAR(1) model [1] to each considered matrix of standardized principal components (SPC) of observed flow data [12]. The model has been previously shown to be capable of preserving the observed basic and drought related statistics at considered and total aggregation time levels (monthly and annual time intervals). Its assessment in the present paper has indicated that the proposed model can reproduce adequately these flow characteristics of the observed flows at partial (tri-monthly) aggregation time intervals. It can be hence concluded that the DMAR(1) model is feasible for the multi-site multi-season problem.

2. DMAR(1) Model

Let $X = \left[x_{\nu\tau}^{s}\right]$ be an $p \times n\omega$ matrix of multi-site seasonal-flow data for year ν, season τ and site s (p = the total length of available flow records in years, ω = the number of seasons and n = the number of considered sites). If the seasonal flow data X do not conform to normal distribution, the flow records X shall be normalized into the $p \times n\omega$ matrix of Y using lognormal transformation by

$$Y = \ln(X - A) \tag{1}$$

where $A = \left[a_{\tau}^{j}\right]$ is the $p \times n\omega$ matrix of lower bounds for three-parameter lognormal distribution. The lower bound parameter a_{τ}^{s} is estimated as [13]

$$a_{\tau}^{s} = \frac{\mathbf{Max}\left(x_{\nu\tau}^{s}\right)\mathbf{Min}\left(x_{\nu\tau}^{s}\right) - \left[\mathbf{Med}\left(x_{\nu\tau}^{s}\right)\right]^2}{\mathbf{Max}\left(x_{\nu\tau}^{s}\right) + \mathbf{Min}\left(x_{\nu\tau}^{s}\right) - 2\,\mathbf{Med}\left(x_{\nu\tau}^{s}\right)} \tag{2}$$

in which $\mathbf{Max}(.)$, $\mathbf{Min}(.)$ and $\mathbf{Med}(.)$ are, respectively the maximum, minimum and median of $x_{\nu\tau}^{s}$ for ν = 1, 2, ..., p.

The proposed model has aimed to describe the within-the-year and over-the-year correlations (R and R_1) of Y. One simple solution is to apply an MAR(1) model [1] to Y directly. Unfortunately, the total number of parameters of the application is often large for the small number of sites considered or even excessive for the large number of the locations. To overcome the referred problem, the multi-site and multi-season flow matrix Y is, hence, transformed to its $p \times r$ (r = the rank of Y) matrix Z of SPC using the well-known singular value decomposition (SVD) technique of matrices {[14], [15] and [16]} as

$$Z = \tilde{Y}\Delta\left(\Theta^{1/2}\right)^{-1} \quad \text{for } r \le n\omega, \tag{3}$$
$$Z = \sqrt{p}\,\Pi \quad \text{for } r \le p.$$

in which $\tilde{Y}$ is the $p \times n\omega$ standardized flow matrix of Y [$\tilde{Y} = (Y - \bar{Y})/S_y$ in which $\bar{Y}$ and S_y = the matrices of means and standard deviations of Y], Θ is the $r \times r$ diagonal matrix of eigenvalues, Δ is the $n\omega \times r$ matrix of associated eigenvectors of R, and Π is the $p \times r$ matrix of corresponding eigenvectors of a row-product matrix R'. The correlation matrix R is preserved through the following characteristic equation:

$$R\Delta = \Delta\Theta \tag{4}$$

The adequate reproduction of the correlation matrix R_1 is, hence, achieved by describing three $(r/3)\times(r/3)$ block-diagonal sub-matrices of M_1 (M_1 = the lag-one correlation matrix of Z). That is, the matrix Z is separated equally (if possible) into three $p\times(r/3)$ partitioned matrices of successive SPC Z^j for j = 1, 2 and 3 $\left(Z = Z^1 : Z^2 : Z^3\right)$. Otherwise, Z^1 and Z^2 will contain the first two sets of $(r/3)$ consecutive SPC while Z^3 will include the rest. Each matrix Z^j is then fitted using an MAR(1) model [1] as

$$Z_v^j = Z_{v-1}^j \Phi^j + U_v^j \tag{5}$$

where Z_v^j is the $1\times(r/3)$ row vector of Z^j, Φ^j is the $(r/3)\times(r/3)$ matrix of parameters, and U_v^j is the $1\times(r/3)$ row vector of residuals with zero means and G^j variance covariance properties. The parameter matrix Φ^j is estimated by

$$\Phi^j = \left(M^j\right)^{-1} M_1^j = I^j M_1^j = M_1^j \tag{6}$$

in which M^j and M_1^j= the $r/3\times r/3$ lag-zero and lag-one correlation matrices of Z^j, and I^j= the identity matrix of size $r/3\times r/3$. The variance covariance estimates G^j of maximum likelihood method are expressed as [17]

$$G^j = (1/p)\left(U^j\right)^T U^j \tag{7}$$

Notice that the MAR(1) model of Z^j [DMAR(1)] is feasible since the matrix Z^j has the number of rows p which is always greater than the number of columns $r/3$. The total number of parameters (k) for the decomposed model is $(2/3)r^2$, excluding $\bar{Y}$ and S_y, while the degrees of freedom (df) left for calculating G^j are $p-(r/3)$ [11].

3. Assessment of the DMAR(1) Model

Three series of natural monthly flows at stations P.12 (the Ping River), Y.1C (the Yom River) and SK (the Nan River) in the Chao Phraya River Basin were considered. The total period of the flow records is 46 years (1955 – 2000). Figure 1 presents the locations of the selected gauging stations.

The DMAR(1) model [(1)-(7)] was applied using the chosen flow-data set. One hundred samples of synthetic monthly flows with the same size as that of the historic sample considered were generated simultaneously. The basic flow characteristics (e.g., mean, standard deviation, coefficient of skewness, and season-to-season correlation and cross-correlation coefficients), and the related drought statistics (i.e., average and maximum drought durations, and average and greatest drought magnitudes against a truncation level b in which $b = 0.75\bar{X}_\tau^s, \bar{X}_\tau^s$, and $1.5\bar{X}_\tau^s$) of every synthetic flow sample were computed at partial (e.g., tri-monthly) time scales. Then, for each considered observed property h, the average value μ and standard deviation σ of $\hat{h}$ ($\hat{h}$= the generated statistics of h) were estimated from those of the one-hundred generated samples. They were used to construct the assessment interval $[\mu\pm\sigma]$ of $\hat{h}$ {[11] and [18]} for assessing the MAR(1) performance of reproducing h. For instance, if the historic flow characteristic h is contained in the interval, it will be inferred that the model can adequately preserve the flow property h.

In the following, the results of evaluating the DMAR(1) model are presented. Figure 2 shows the historic and generated means, standard deviations, skewness coefficients, and season-to-season correlation and cross-correlation coefficients of historic tri-monthly flows for station P.12. It appears that the model reproduces the observed flow statistics adequately. Figure 3 illustrates the historic and generated drought statistics (average and longest drought durations, and average and greatest drought magnitudes) of P.12 tri-monthly flows at the few different threshold levels of $0.75\bar{X}_\tau^s$, $\bar{X}_\tau^s$ and $1.5\bar{X}_\tau^s$. It is evident that the model is able to preserve the drought properties satisfactorily.

4. Uncertainties of Basic Statistics of the Model

Although the DMAR(1) model has been reduced many parameters, one may wish to know the variations of basic statistics of the model. Hence, it was applied to the previously selected set of the monthly flows from stations P.12, Y.1C and SK (see Figure 1). Their one-standard-deviation intervals of basic flow properties at the considered monthly level were then constructed for various sample sizes of 100, 75, 50 and 25 years.

Figure 4 presents the one-standard-deviation intervals of cross-correlation (P.12 and Y.1C) of all considered sample sizes [DMAR(1)]. The figure demonstrates that the uncertainties of the flow statistics generally vary with sample sizes considered, as expected. The intervals tend to be large for short sample size. Moreover, it should be noted that the basic correlations (month-to-month correlation and lag-zero cross-correlation) are usually preserved, like the results of [12]. Only the historic correlations in March are slightly out of the assessed intervals.

5. Summary and Conclusions

Generation of seasonal flows at multiple sites simultaneously that can preserve adequately the basic and drought properties at various aggregation levels is fundamental to a success in design and operational studies of water resource systems. The present paper thus proposes an DMAR(1) model for generating multi-site seasonal flows concurrently. The proposed model [12] uses the technique of SVD to transform the normalized flow data Y into its SPC Z, and fits an MAR(1) model to three equal sub-matrices of consecutive SPC. The model is basically equivalent to the MAR(1) model of Y, but has fewer parameters. Moreover, previous work [12] has demonstrated that the DMAR(1) model can preserve the flow characteristics at considered (monthly) and total aggregate (annual) levels.

In this study, the performances of the DMAR(1) model for reproducing the referred flow statistics at intermediate (bi-monthly, tri-monthly, quarterly and semi-annual) time intervals are assessed. Hence, the proposed model has been used to generate simultaneously three sequences of 29-year (1969 – 1997) monthly flows [P.12 (the Ping River), Y.1C (the Yom River), and SK (the Nan River)]. Results of the assessment based on the criterion of one-standard-deviation interval have indicated that the DMAR(1) is able to preserve the basic flow characteristics and drought related statistics at the partial (e.g., tri-monthly) levels adequately. Thus, it can be concluded that the DMAR(1) model is feasible for the multi-site multi-season problem.

The DMAR(1) model is further analyzed the uncertainties of basic flow properties. The analysis was performed by applying the model to generate concurrently the above-mentioned set of the flow records for several sample sizes (100-, 75-, 50- and 25-year). The analysis results have indicated that the uncertainties of the flow statistics are great for limited sample size.

6. Acknowledgments

Financial support by TRF -- PDF24/2542 – is appreciated. The authors are also thankful Royal Irrigation Department (Thailand) for providing monthly flow records used in this study.

References

[1] Matalas, N. C. (1967). "Mathematical assessment of synthetic hydrology." *Water Resources Research*, 3(4), 937-945.

[2] Salas, J. D., and Pegram, G. G. S. (1977). "A seasonal multivariate multilag autoregressive model in hydrology." *Proc. of the third Int. Symp. of Theoretical and Applied Hydrology,* Colorado State University, Fort Collins, Colorado.

[3] O'Connell, P. E. (1974)."Stochastic modeling of long-term persistence in stream-flow sequences." *Ph.D. Thesis*, Imperial College, University of London

[4] Haltiner, J. P., and Salas, J. D. (1988). "Development and testing of a multivariate, seasonal ARMA(1,1) model." *J. Hydrol.,* 104, 247-272.

[5] Bras, R. L., and Rodriguez-Iturbe, I. (1985). *Random Functions and Hydrology*, Addison-Wesley Pub. Co., Reading, Mass

[6] Oliveira, G. C., Kelman, J., Pereira, M. V. F., and Stedinger, J. R. (1988). "A representation of spatial cross correlations in large stochastic seasonal streamflow models." *Water Resources Research*, 24(5), 781-785.

[7] Valencia, D. R., and Schaake, J. C. (1973). "Disaggregation Processes in Stochastic Hydrology." *Water Resources Research*, 9(3), 580-585.

[8] Tao, P. C., and Delleur, J. W. (1976). "Multistation multiyear synthesis of hydrologic time series by disaggregation." *Water Resources Research*, 12(60, 1303-1312.

[9] Lane, W. L. (1982). "Correlated parameter estimates for disaggregation schemes." in: *Statistical Analysis of Rainfall and Runoff*, edited by V. P. Singh, Water Resources Publications, Littleton, Colorado, 505-530.

[10] Grygier, J.C., and Stedinger, J.R. (1988). "Condensed Disaggregation Procedure and Conservation Correlations for Stochastic Hydrology." *Water Resources Research*, 24(10), 1574-1584.

[11] Santos, E. G., and Salas, J. D. (1992). "Stepwise Disaggregation Scheme for Synthetic Hydrology." *J. Hydraul. Eng.*, ASCE, 118(5), 765-784.

[12] Chaleeraktrakoon, C. (1999). "Stochastic Procedure for Genarating Seasonal Flows." *J. Hydrol. Eng.*, ASCE, 4(4), 337-343.

[13] Hoshi, K., J. R. Stedinger, and S. Burges, "Estimation of Log Normal Quantiles: Monte Carlo Results and First-Order Approximations, " *J. Hydro., vol.* 71, no. 1/2, pp. 1-30, 1984.

[14] Jöreskog, K. G., Klovan, J. E., and Reyment, R. A. (1976). "*Geological Factor Analysis*." Elsevier Scientific Pub. Co., New York.

[15] Green, P. E. (1978). *Analyzing Multivariate Data*, The Dryden Press, Hinsdale, Illinois

[16] Cavadias, G. S. (1986). "A Multivariate Seasonal Model for Streamflow Simulation." in: *La recherche en Hydrologie au Quebec*, edited by V-T-V. Nguyen, and Y. Faucher, Les Presses de l'Universite du Quebec, 35, 58-77.

17] Stedinger, J. R., Lettenmaier, D. P., and Vogel, R. M. (1985a) "Multisite ARMA(1,1) and disaggregation models for annual streamflow generation." *Water Resources Research*, 21(5), 665-675.

[18] Loucks, D. P., Stedinger, J. R., and Haith, D. A. (1981). *Water Resource Systems Planning and Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey, 559 pages.

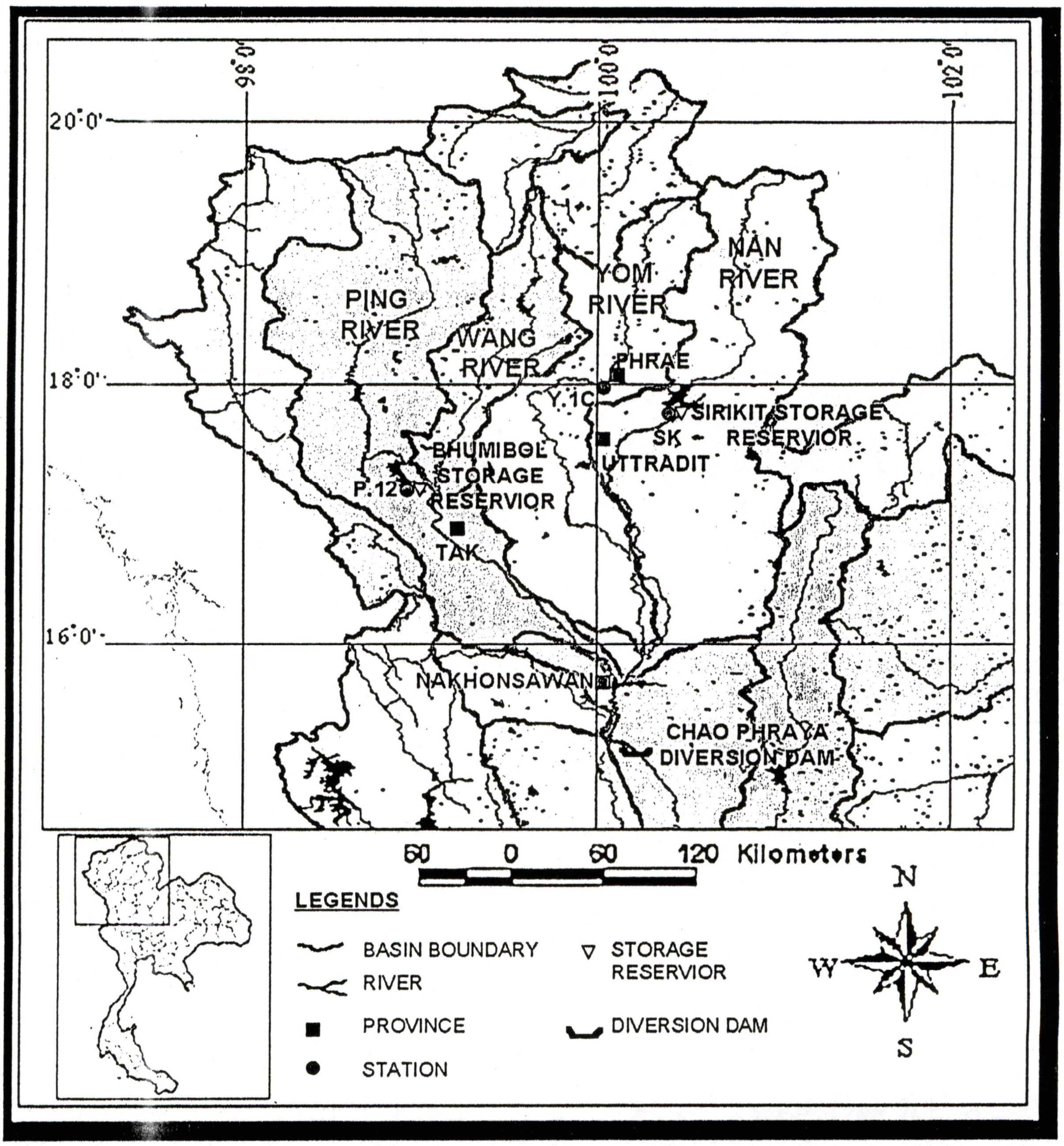

Figure 1. Locations of the selected gauging stations.

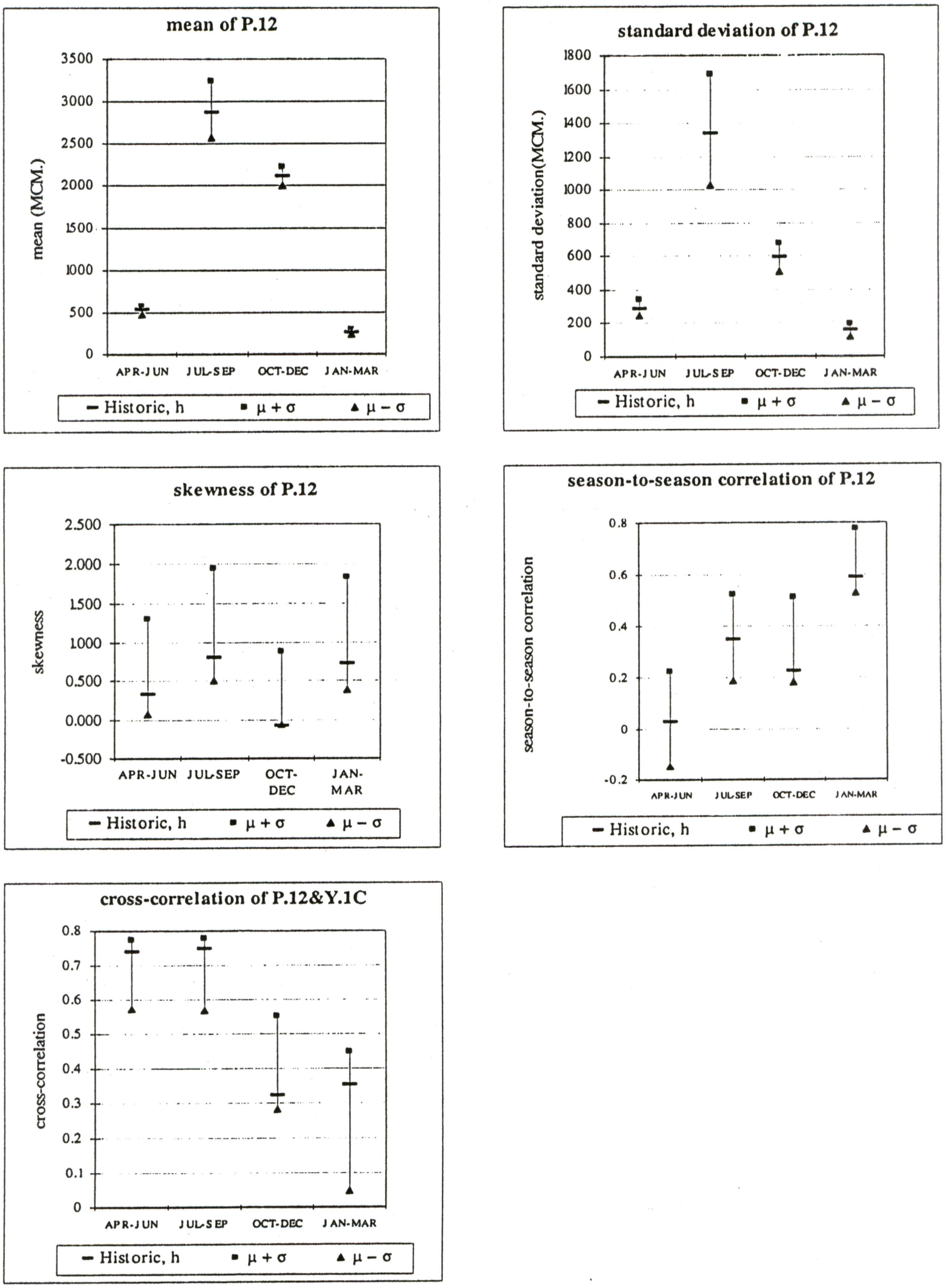

Figure 2. Historic and generated basic statistical properties of tri-monthly flows (P.12).

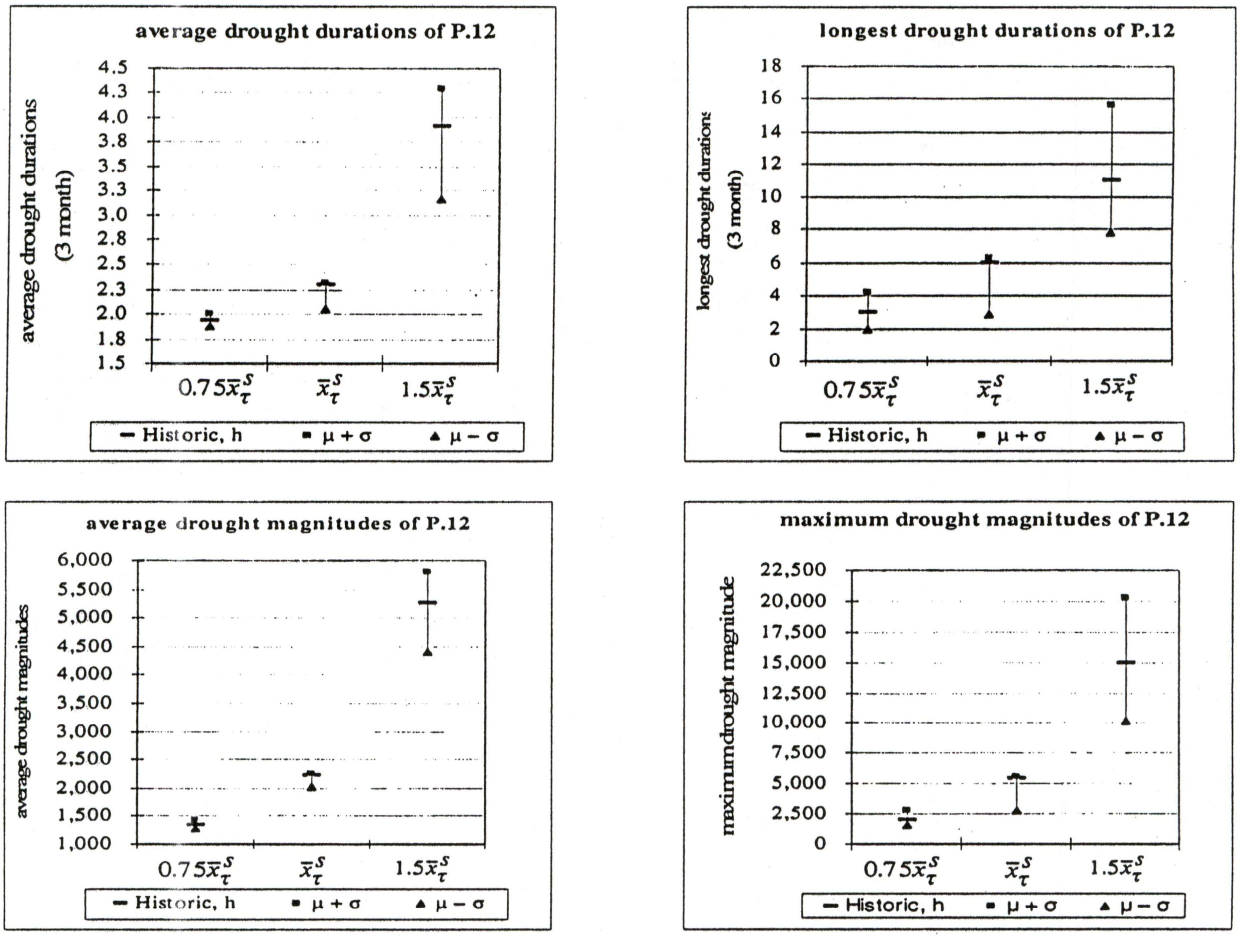

Figure 3 Historic and generated drought statistical properties of P.12 tri-monthly flows at few different threshold levels.

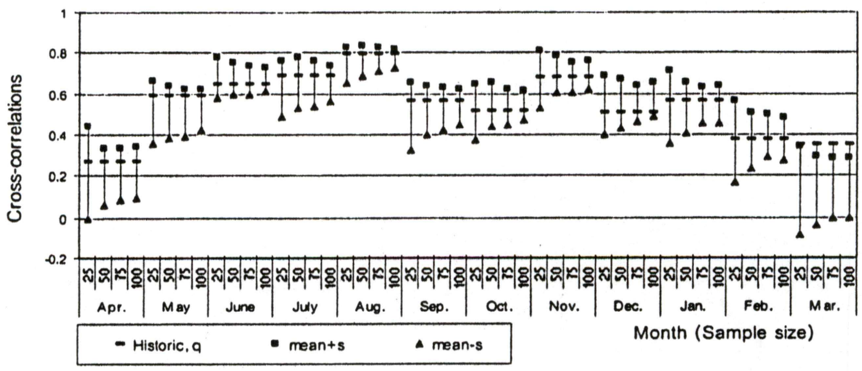

Figure 4 One-standard-deviation intervals of cross-correlations (P.12 and Y.1C) of all considered sample sizes [DMAR(1)].