International Seminar on Hydrology of Extremes 1-3 December, 1988
Roorkee (India)

FORECASTING OF RIVER FLOWS USING A CARMA MODEL

Nattawuth Udayasen
Electricity Generating Authority of Thailand
Bangkok, Thailand

Edward McBean
Department of Civil Engineering
University of Waterloo
Waterloo, Ontario, Canada

SYNOPSIS

The Contemporaneous Autoregressive Moving Average (CARMA) model is used to forecast flows of the Ping and Nan Rivers in Thailand. Streamflow forecasting is executed for one year ahead. The CARMA model provides only limited information for forecasting streamflow conditions for this time-frame.

INTRODUCTION

Natural streamflows into reservoirs are not deterministic. Hence, it is necessary to consider their random behaviour in the design and operation of large scale water resources projects. Some information about the nature of the random behaviour is available from the historical record of hydrologic flow measurements. Unfortunately, the history of flows in itself in many ways is inadequate for future planning purposes. The historical set of flows at a specific location represents only a finite sample out of the infinite number of potential flow sequences that could occur at that location in the future. Planning a water resources project utilizing only the historical flow record would be a very precarious practice since it is unlikely that the same historical flow sequence will recur during any specified period of time in the future.

To account for the effects of random streamflows in the planning process, it has been customary to develop models that yield alternative potential flow sequences which are statistically indistinguishable from the historical time series.

The intent of the paper is to indicate the results of using one of the types of the model, namely the Contemporaneous Autoregressive Moving Average (CARMA) model, to forecast river flows for two rivers in Thailand.

REVIEW OF LITERATURE

Sudler [1927] was the first to synthesize flow sequences. Fifty annual flow values were written on cards, the cards shuffled and dealt 20 times to develop a 1000 year flow record. Sudler's work generated flow different from the historical data but did not provide flows more extreme than the 50 annual flow values in his deck. In addition, the auto-correlation structure of the historic data was destroyed.

Thomas and Fiering [1962] developed a seasonal model which expresses dependence between flows for successive months of the year by a Markov process. Since the research of Thomas and Fiering, published work involving synthetic hydrology has been profuse. A notable paper is that of Benson and Matalas [1967] who discussed issues such as operational bias and the importance of synthetic flows conforming to a particular distribution.

Intervention analysis by Box and Tiao [1975] and Hipel et al. [1977] provides a means to statistically describe intervention effects and also furnishes a stochastic model that can be used for applications such as simulation and forecasting.

Moss and Bryson [1974] have found that the seasonal geophysical time series exhibit autocorrelation structures which depend not only on the time lag between observations but also the season of the year. In an effort to account for these characteristics, a family of Periodic Autoregressive (PAR) models is introduced. Certain PAR models are recommended for forecasting monthly river flows by Noakes et al. [1985].

In general, water resources systems investigations may involve data generation and forecasting of both model inputs and model outputs. Model inputs are usually hydrologic variables, while model outputs are related to water uses such as irrigation, hydropower and water supply. Sets of related series defined at several points along a line, over an area or space are multivariate time series [Salas, et al. 1980]. Multivariate time series may also have been viewed as single or multiple series at a given site, having statistically distinguishable properties at various seasons of the year. Good reviews of multivariate modelling of water resources time series can be found in Salas et al. [1985] and Hipel [1986].

Previously, it was common for hydrologists to specify the exact form of the multivariate model even before examining the data. For example, often the model was assumed to be multivariate autoregressive (AR) lag one model (denoted as AR(1)) or a multivariate autoregressive moving average (ARMA with one AR parameter and one MA parameter (denoted as ARMA(1,1)). Such a procedure clearly leads to the possibility that the model will not fit the data very well. This led to the finding that hydrologic sequences generated from these models were inadequate [Finzi et al. 1975]. To overcome this problem, Ledolter [1978] suggested the use of the general multivariate class of ARMA models.

There are two principal disadvantages to the use of general multivariate ARMA models in hydrology:

1. they are very complicated (the number of parameters increases exponentially with the dimensionality of the model), and

2. an important feature is still being omitted, namely that the physical structure of the system imposes restrictions on the model.

In response to the first disadvantage of general multivariate ARMA model, Salas et al. [1980] proposed the use of a multivariate ARMA model which was restricted to have diagonal parameter matrices (i.e., the contemporaneous ARMA model), arguing that this would reduce the number of parameters to be estimated.

The CARMA model can, in many situations, cope with the second disadvantage of the general multivariate ARMA model as well. A further advantage of the CARMA model is that it can handle the case of time series with unequal sample sizes, a situation commonly encountered in practice [Hipel et al. 1985]. As is shown by Camacho [1984], it is possible to extend the simulation technique given by McLeod and Hipel [1978] to cover the case of the CARMA model. As illustrated by practical applications, the CARMA model can be efficiently used to model hydrological and other environmental time series [Camacho et al. 1985].

CASE STUDY HYDROLOGIC REGIME

The Bhumipol and Sirikit reservoirs are located in northern Thailand. The climate of these catchment areas is influenced by the southwest monsoon, which usually starts in May and finishes in October. The depression storms from the South China Sea in September and October produce peak streamflows into the reservoirs.

Historical flow records from April 1952 to March 1986 show that the average annual inflow of the Ping River into Bhumipol reservoir is 195.7 m^3/sec while the average annual inflow of the Nan River into the Sirikit reservoir is 185.7 m^3/sec. Flow patterns of the Ping River and Nan River are presented in Figures 1 and 2, respectively.

Various stochastic models are available for use in streamflow generation and forecasting. One such model is the autoregressive moving average (ARMA) model.

In general, for seasonal river flow data, seasonal differencing is usually required, to reduce a seemingly non-stationary or unstable series to an apparently stationary stable series that is easier to model. The procedure of simplifying the series of data by seasonal differencing before modelling is often recommended in the literature (e.g. Pandit and Wu, [1984]).

If the model is used for forecasting purposes, differencing has the desirable property of preserving the seasonal wave pattern in the eventual forecasting function. For simulation purposes, seasonal differencing means that the process is nonstationary and therefore, by definition, the simulated data are not restricted to any mean level within each season.

Accordingly, it is appropriate to first transform the data to remove seasonality. Then a suitable stationary non-seasonal model can be fit to the data. The procedure recommended by McBean and Hipel [1976] is to initially choose a suitable transformation technique to remove seasonality. Then, the proper ARMA model can be selected by utilizing the three stage modelling procedure of identification, estimation, and diagnostic checking.

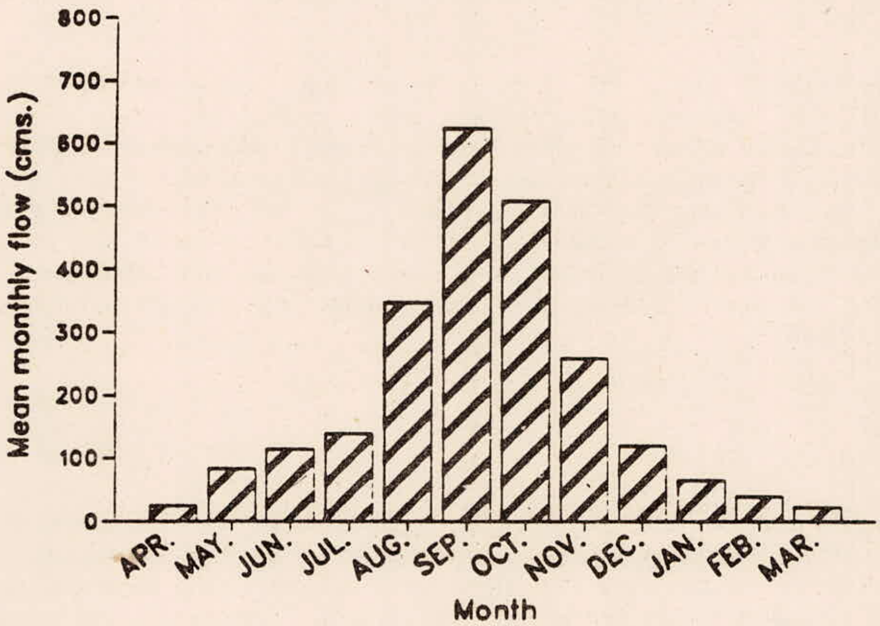

Figure 1 Flow Pattern of the Ping River at Bhumipol Dam

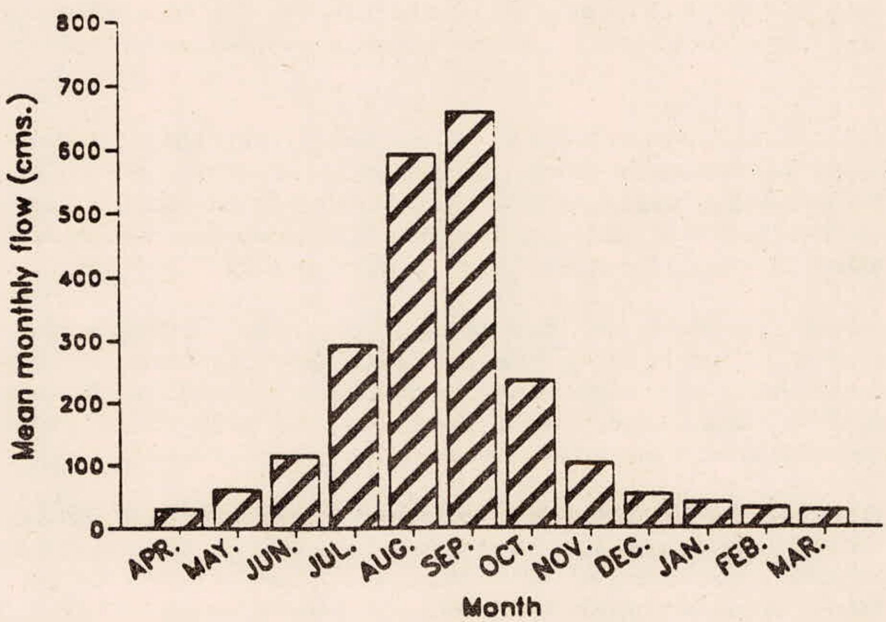

Figure 2 Flow Pattern of the Nan River at Sirikit Dam

In this study, after the parameter estimates have been obtained [Camacho et al. 1985, 1987a, 1987b], the cross correlations of the residuals of the two time series were calculated. The cross correlations at lag zero show values significantly different from zero. This indicates that the CARMA model could also be employed to model these hydrological time series.

The Box-Jenkins procedures of model building was applied to the two series of mean monthly flows, the Ping River and the Nan River. The historical streamflow data base is 34 years (from April 1952 to March 1986) reflecting data obtained from the Electric Generation Authority of Thailand (EGAT).

The basic tools for identification are a plot of the original time series, and plots of the autocorrelation (ACF) and partial autocorrelation (PACF) functions. The plot of the original time series indicates seasonality.

Figure 3 is a plot of mean monthly flows, and Figures 4 and 5 are plots of the ACF and PACF of the sample time series for the Ping River. Examination of these plots indicates a strong seasonality. The ACF and PACF are significant at lag one month, and twelve months, and the cycle is observed to repeat every twelve lags. Therefore, it is assumed that deseasonalization or differencing is necessary.

DESEASONALIZATION

To remove seasonality, the given data can be standardized by subtracting the monthly mean of the time series from each of the monthly observations and then dividing the result by the monthly stardard deviation.

Let $Z_{r,j}$ denote the observation in the r_{th} year ($r = 1,2,\ldots, \bar{N}$ where $\bar{N} = N/S$, and S is the seasonal period) and the j^{th} season ($j = 1,2,\ldots, S$). For the j^{th} season

$$E\left[Z_{r,j}^{(\lambda)}\right] = \mu_j \qquad (1)$$

$$VAR\left[Z_{r,j}^{(\lambda)}\right] = \sigma_j^2 \qquad (2)$$

where $Z_{r,j}^{\lambda}$ indicates a possible Box-Cox [Box and Cox, 1984] transformation.

The time series:

$$\chi_{r,j} = \frac{Z_{r,j}^{(\lambda)} - \mu_j}{\sigma_j} \qquad (3)$$

is called the deseasonalized series, and it is assumed that $\chi_{r,j}$ can be modelled by an ARMA model [McLeod and Hipel, 1978]

$$\phi(B)\ \chi_t = \theta(B)\ \alpha_t \qquad (4)$$

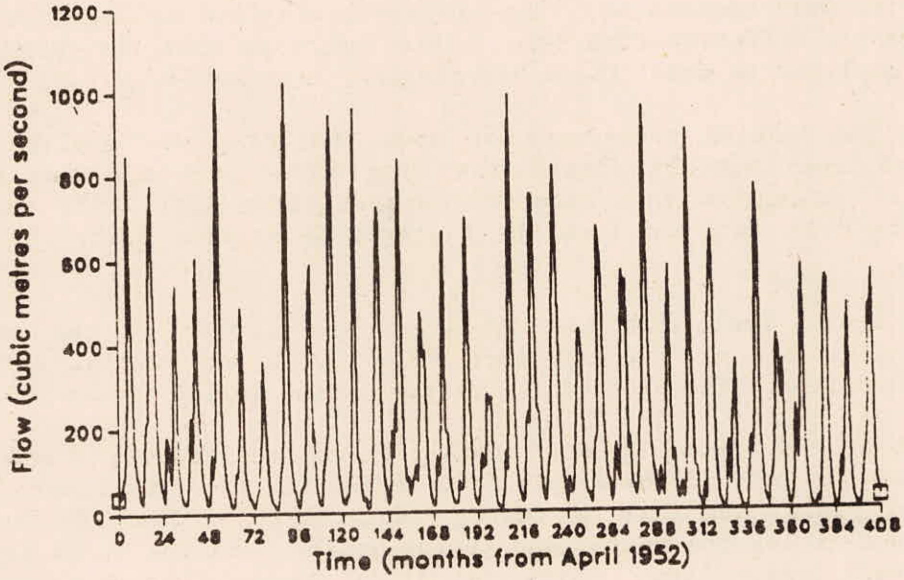

Figure 3 Mean Monthly Flow of the Ping River at Bhumipol Dam

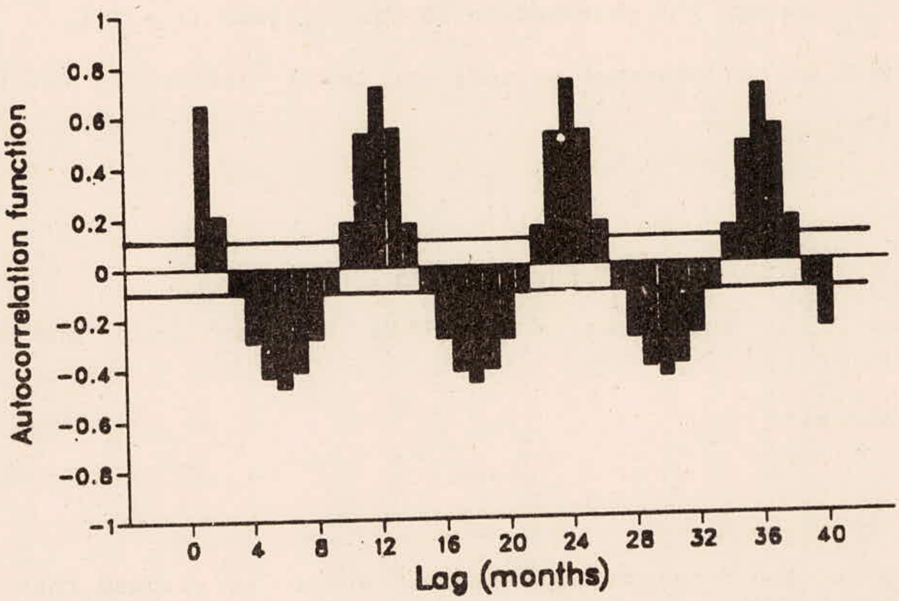

Figure 4 Autocorrelation Function for Historical Flow of the Ping River at Bhumipol Dam

The parameters μ_j and σ_j for each season can be estimated by

$$\hat{\mu} = \frac{1}{N} \sum_{r=1}^{\bar{N}} Z_{r,j}^{(\lambda)} \tag{5}$$

and

$$\hat{\sigma}_j = \frac{1}{N} \sum_{r=1}^{\bar{N}} (Z_{r,j}^{\lambda} - \hat{\mu})^2 \tag{6}$$

The three stage modelling procedure is then repeated on the deseasonalized series.

The resulting ACF exhibited characteristics of attenuation at lag 6 months; however, the PACF is significant at lag one and lag two months (see Udayasen [1988] for further details). From these observations it seems reasonable to suspect that a model of order greater than one will provide a better fit to the data.

MODEL FIT TO DESEASONALIZED SERIES

Following the Box-Jenkins modelling procedure [Box and Jenkins, 1976], the proper model for both deseasonalized series is the ARMA (2,0) model.

Table 1 Parameter Estimates for an ARMA (2,0) Model Fit to the Deseasonalized Ping River and Nan River

Parameters	Estimates	Standard Error
1. The Ping River		
ϕ_1	0.4862	0.0490
ϕ_2	0.1415	0.0490
ϕ_α^2	0.6643	
2. The Nan River		
ϕ_1	0.4920	0.0490
ϕ_2	0.1484	0.0490
σ_α^2	0.6527	

The parameter estimates for an ARMA (2,0) process with $\lambda = 0$ are displayed in Table 1 and the finite difference equations are shown in Equations 7 and 8.

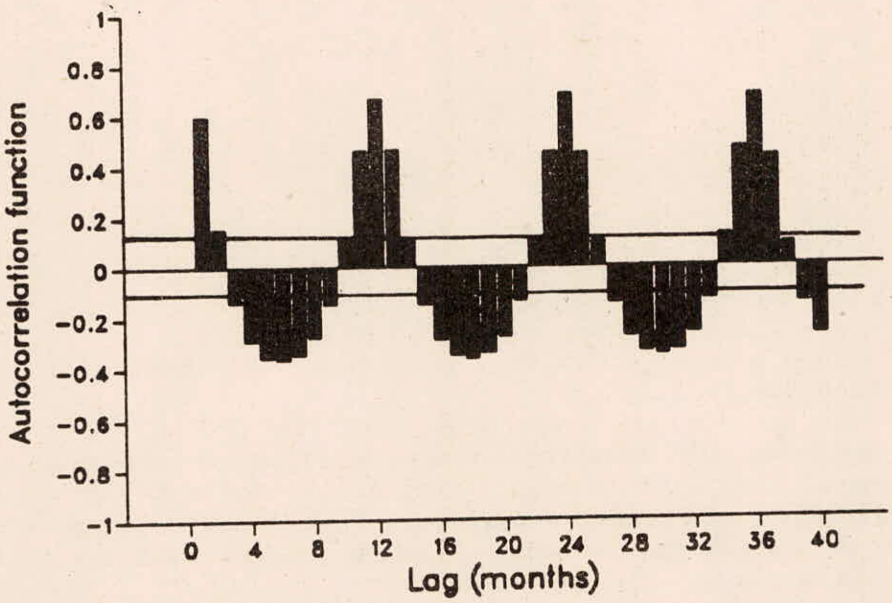

Figure 5 Autocorrelation Function for Historical Flow of the Nan River at Sirikit Dam

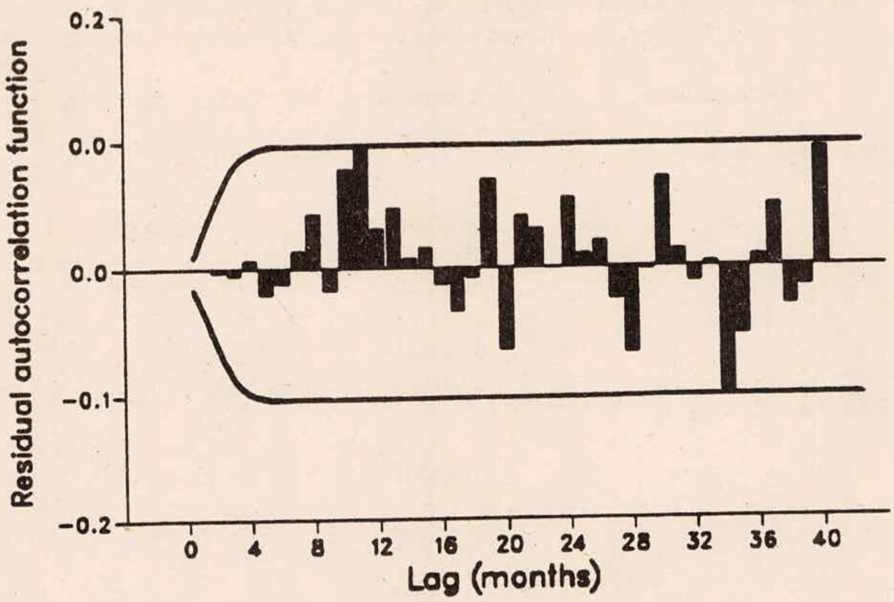

Figure 6 Residual Autocorrelation Function for Deseasonalized Flow of the Ping River at Bhumipol Dam

APPLIED OPERATIONAL HYDROLOGY

The finite difference equation for:

1. The Ping River is:

$$(1 - 0.4862B - 0.1415B^2)[\ell n(Z_t - \hat{\mu})] = \alpha_t \quad (7)$$

2. The Nan River is:

$$(1 - 0.4920B - 0.1484B^2)[\ell n(Z_t - \hat{\mu})] = \alpha_t \quad (8)$$

The residual autocorrelation function for the Ping River is presented in Figure 6.

APPLIED CARMA MODEL

To elucidate the relationship between these two time series, the cross correlation functions of the residuals have been estimated. At lag zero, a significant difference of cross correlation has been found (Figure 7). This indicates that the CARMA model is adequate to model these time series (Camacho [1985]).

Using the parameter estimation developed by Camacho [1985] the parameter estimates for the CARMA process are shown in Table 2.

COMPUTATIONAL AND COMPUTER PROGRAMME

The computations for the analysis of the historical streamflows and the Box-Jenkins procedure [Box and Jenkins, 1976] were performed using the McLeod and Hipel time series package [McLeod and Hipel, 1978]. The parameter estimates for the CARMA model were obtained using Camacho's CARMA parameters estimation programme [Camacho, 1984].

Table 2 Parameter Estimates for an CARMA Model Fit to the Deseasonalized Ping River and Nan River

Parameters	Estimates	Standard Error
1. The Ping River		
ϕ_1	0.5053	0.0465
ϕ_2	0.1184	0.0465
σ_α^2	0.6645	
2. The Nan River		
ϕ_1	0.5320	0.0465
ϕ_2	0.1186	0.0465
σ_α^2	0.6536	

The streamflow generation of a time series from a given CARMA model was programmed in FORTRAN 77. This programme generates the time series:

$$Z_{t,} \; i = 1,2,\ldots, LW \text{ as}$$

$$Z_t = \sum_{j=1}^{IP} ARPS_j Z_{i-j} + PMAC + \alpha_i - \sum_{j=1}^{IQ} PMAS_j \alpha_{i-j} \qquad (9)$$

where

ARPS = vector of length IP containing the AR parameters of the model;

PMAC = over all MA parameters;

PMAS = vector of length IQ containing the MA parameters of the model;

α_i = the white noise series;

LM = Length of the time series to be generated; and

Z_t = generated time series.

APPLIED OPERATIONAL HYDROLOGY

The streamflow forecasting of a time series from a given CARMA model was programmed in FORTRAN 77. This programme computes the time series using a fitted CARMA parameter. The model is of the form:

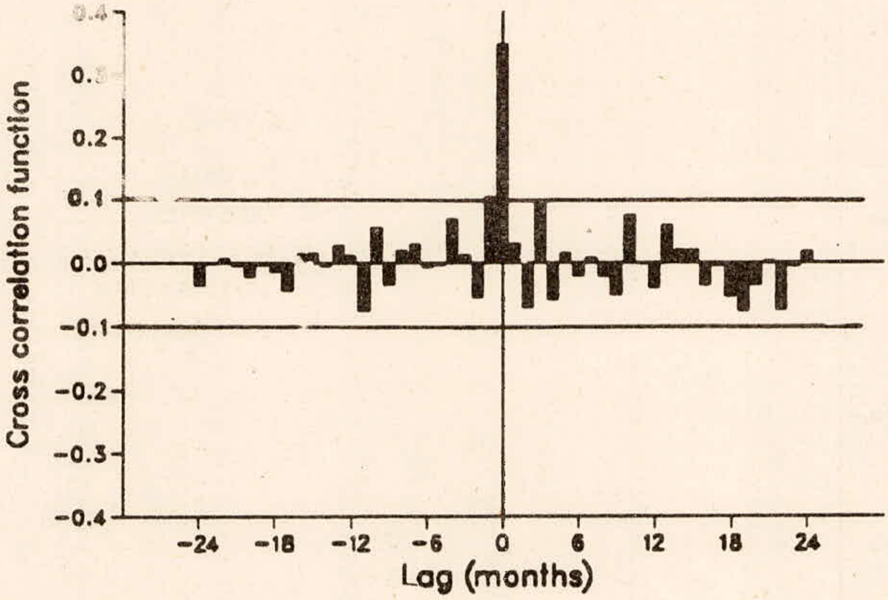

Figure 7 Cross Correlation Function for Deseasonalized Flows of the Ping and the Nan Rivers

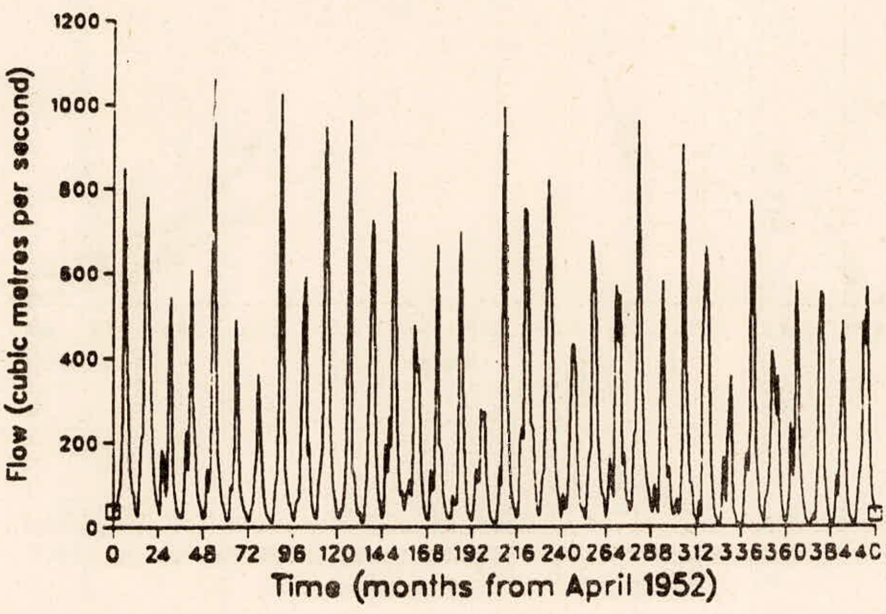

Figure 8 Mean Monthly Flow Into Bhumipol Reservoir Recorded From April 1952 to March 1986

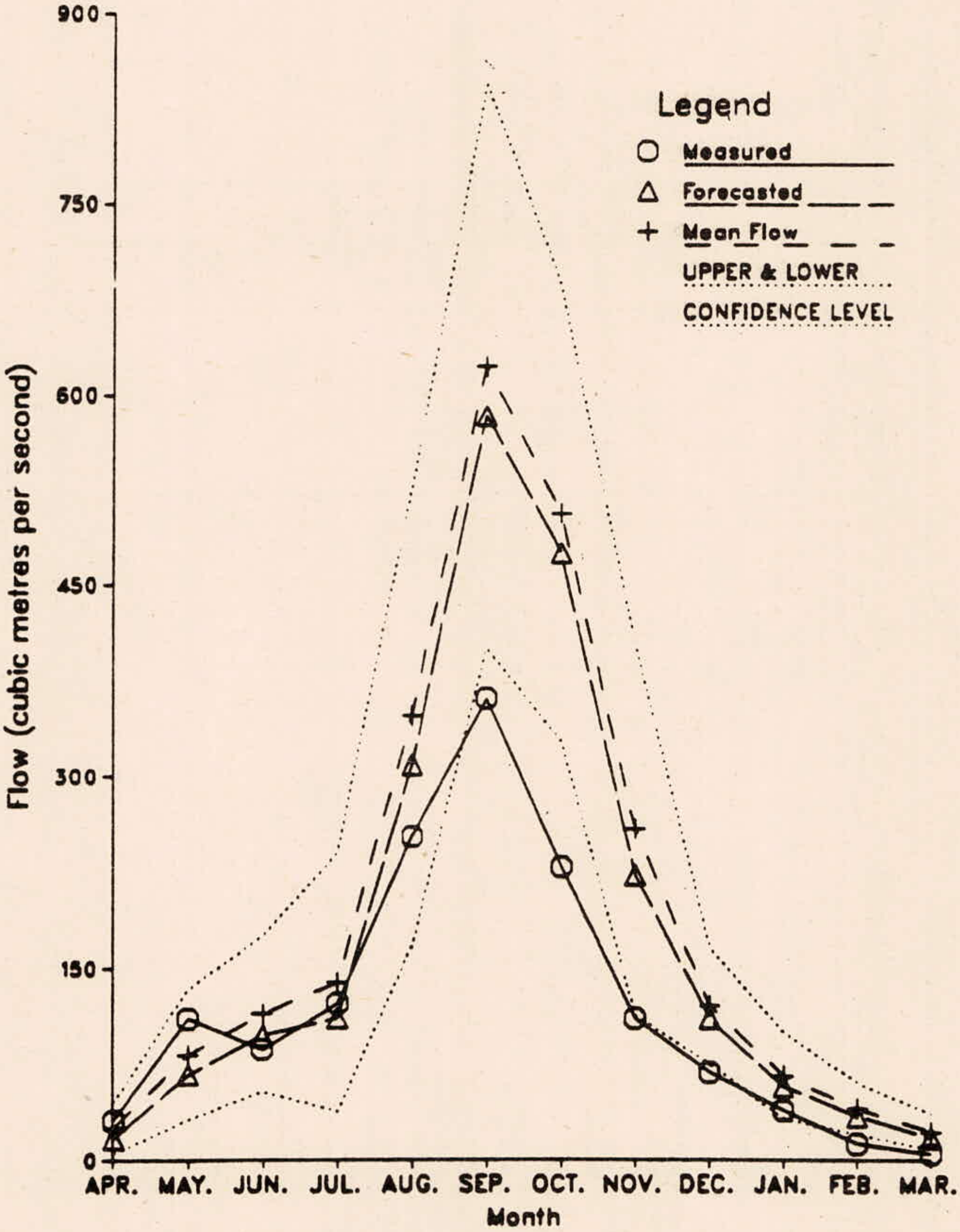

Figure 9 The Ping River: Historical, Forecast and Measured Mean Monthly Flows (From April 1986 to March 1987)

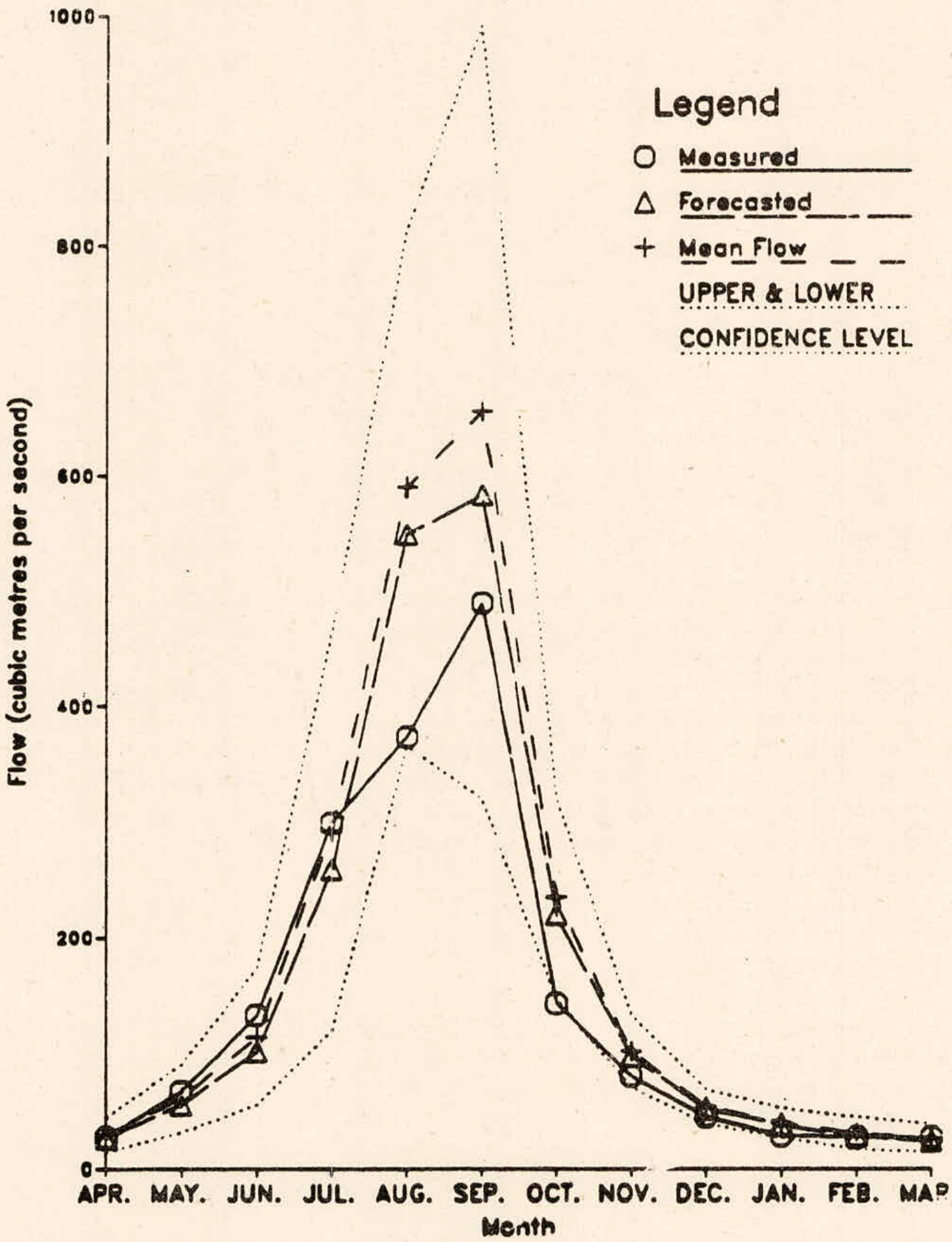

Figure 10 The Nan River: Historical, Forecast and Measured Mean Monthly Flows (From April 1986 to March 1987)

$$\nabla^d Z_t = \theta_o + \sum_{i=1}^{p} \sigma \nabla^d Z_{t-1} + \alpha_t - \sum_{i=1}^{q} \theta_i \alpha_{t-1} \quad (10)$$

where estimates of the parameters $\theta_0, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q$ are the inputs PMAC, ARPS and PMAS, respectively.

The computational technique is described in Box and Jenkins 1976.

FORECASTING

The contemporaneous autoregressive moving average (CARMA) model was selected as the streamflow generation model for the Ping and Nan Rivers. Simulated flows were generated for the 408 month period. A plot of the historical monthly mean and simulated flows for the Ping River is presented in Figure 8. The results indicate that the simulated flows closely resemble the historical flow data.

Streamflow forecasting was executed for a one year period (April 1986 to March 1987). The comparisons among the historical, forecast and measured monthly mean flows are presented in Figures 9 and 10 together with the upper and lower confidence level bounds. The results clearly illustrate that the forecast and measured monthly mean flows are less than the historical mean flows. They indicate that the model predicted low flows for this period, but the measured monthly mean flows were lower than the model predicted, especially for the Ping River (Figure 9). In September and October 1986 the measured values were less than the lower confidence level.

CONCLUSIONS

It can be concluded that the streamflow forecasting for one-year ahead by CARMA mode provides only limited information for conditions of Thailand for the one-year reservoir operation plan.

BIBLIOGRAPHY

[1] Benson, M.A., and N.C. Matalas, "Mathematical Assessment of Synthetic Hydrology", Water Resources Research, Vol. 3, No. 4, Fourth Quarter, 1967, pp. 931-935.

[2] Box, G.E.P., and Cox, D.R., "An Analysis of Transformation", Journal of Royal Statistic Society, Series B, Vol. 26, 1964, pp. 211-252.

[3] Box, G.E.P., and Jenkins, G.M., Time Series Analysis: Forecasting and Control", Holden-Day Inc., San Francisco, 1970.

[4] Box, G.E.P., and G.C. Tiao, "Intervention Analysis with Applications to Economic and Environment Problems", Journal of the American Statistic Association, Vol. 70, No. 2, March 1975, pp. 277-298.

[5] Camacho, F., "Contemporaneous ARMA Modelling with Applications", Ph.D. Thesis, Department of Statistical and Actuarial Sciences, The University of Western Ontario, London, Ontario, 1984.

[6] Camacho, F., A.I. McLeod, and K.W. Hipel, "Contemporaneous Autoregressive-Moving Average (CARMA) Modelling in Water Resources", Water Resources Bulletin, American Water Resources Association, Vol. 21, No. 4, August 1985, pp. 709-720.

[7] Camacho, F., McLeod, A.I. and Hipel, K.W., "Multivariate Contemporaneous ARMA Models with Hydrological Applications", Stochastic Hydrology and Hydraulics, Vol. 1, No. 2, 1987a, pp. 141-154.

[8] Camacho, F., McLeod, A.I. and Hipel, K.W., "Contemporaneous Bivariate Time Series", Biometrika, Vol. 74, No. 1, 1987b, pp. 103-113.

[9] Finzi, G., E. Todini, and J.R. Willis, "Comments Upon Multivariate Synthetic Hydrology", Water Resources Research, Vol. 11, December 1975, pp. 844-850.

[10] Hipel, K.W., "Stochastic Research in Multivariate Analysis", Keynote paper in Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium on "Multivariate Analysis of Hydrologic Processes". Edited by H.W. Shen, et al. Engineering Research Centre, Colorado State University, 1986. pp. 2-50.

[11] Hipel, K.W., A.I. McLeod, and W.K. Li, "Causal and Dynamic Relationships between Natural Phenomenon, in Time Series Analysis: Theory and Practice 6, O.D. Anderson. Proceedings of the International Conference on Hydrological, Geographical and Spatial Time Series, held at Toronto, August 1983. Published by Elsevier Science Publishers B.V., North Holland, Amsterdam, 1985, pp. 13-34.

[12] Hipel, K.W., A.I. McLeod, and E.A. McBean, "Stochastic Modelling of the Effects of Reservoir Operation", Journal of Hydrology, Vol. 32, 1977, pp. 97-113.

[13] Ledolter, J., "The Analysis of Multivariate Time Series Applied to Problem in Hydrology", Journal of Hydrology, Vol. 36, 1978, pp. 327-352.

[14] McBean, E.A., and K.W. Hipel, "Design of Lake Diefenbaker Hydroelectric", Technical Report 76-3, Water Resources Group, Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, May 1976.

[15] McLeod, A.I., and K.W. Hipel, "The McLeod-Hipel Package Users' Manual", Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, November 1978.

[16] Moss, M.E. and M.C. Bryson, "Autocorrelation Structure of Monthly Streamflows", Water Resources Research, Vol. 10, No. 4, August 1974, pp. 737-744.

[17] Noakes, D.J., A.I. McLeod, and K.W. Hipel, "Forecasting Monthly Riverflow Time Series", International Journal of Forecasting, Vol. 1, 1985, pp. 179-190.

[18] Pandit, S.M., and S-M. Wu, "Time Series and System Analysis with Applications", John Wiley and Sons, New York, 1983.

[19] Salas, J.D., J. Delleur, V. Yevjevich, and W. Lane, "Applied Modelling of Hydrology Series", Water Resources Publication, Littleton, Colorado, 1980.

[20] Salas, J.D., Tabios III, G.Q., and Bartolini, P., "Approaches to Multivariate Modelling of Water Resources Time Series", Water Resources Bulletin, Vol. 21, No. 4, 1985.

[21] Sudler, C.E., "Storage Required for the Regulation Streamflow", Trans. ASCE, Vol. 91, December 1927, pp. 622-660.

[22] Thomas, H.A., and M.B. Fiering, "Mathematical Synthesis of Streamflow Sequences for the Analysis of River Basins Simulation", in Maass et al., Design of Water Resource Systems, Chapter 12, Harvard University Press, Massachusetts, 1962.

[23] Udayasen, N., "Determination of Optimal Operation Policies for a Combined Hydro-Thermal Power Generating System", Ph.D. Thesis, Dept. of Civil Engineering, University of Waterloo, Waterloo, Ontario, Canada, 1988.