*Nordic Hydrology,* 17, 1986, 185-202
No part may be reproduced by any process without complete reference

# Multivariate Transfer Function-Noise Model of River Flow for Hydropower Operation

**Tryggvi Olason and W. Edgar Watt**

Department of Civil Engineering, Queen's University,
Kingston, Ontario, Canada

The formulation of multivariate autoregressive moving average (ARMA) time series models and their transfer function noise (TFN) form is described. Development of a multivariate TFN model is difficult if the multiple inputs are correlated. Various methods for developing a multivariate TFN models with correlated multiple inputs are critically reviewed. A simple approach to developing multiple input TFN models with correlated inputs is described. This approach is successfully applied to developing a forecasting model for average daily flow of the Mattagami River at Little Long Generation Station in Northern Ontario, Canada. System inputs are upstream and tributary flows. Only three years of daily data for the period April 1st to October 31st were required to calibrate the model. Two further years were used to verify the model. Forecasts at lead times of one and two days were good for both calibration and verification periods. The average standard errors were 8% of average inflows (1-day lead) and 18% (2-day lead). The system produces significantly better forecasts than a univariate time series model.

## Introduction

Hydropower is a major source of electricity in many countries and increasing emphasis is being directed to efficient utilization of this power source. Effective scheduling of hydropower production requires good long and short term forecasts of flows. Long term forecasts can be made off-line but short term flow forecasts can be made only with some type of real-time flow forecasting system. The type and

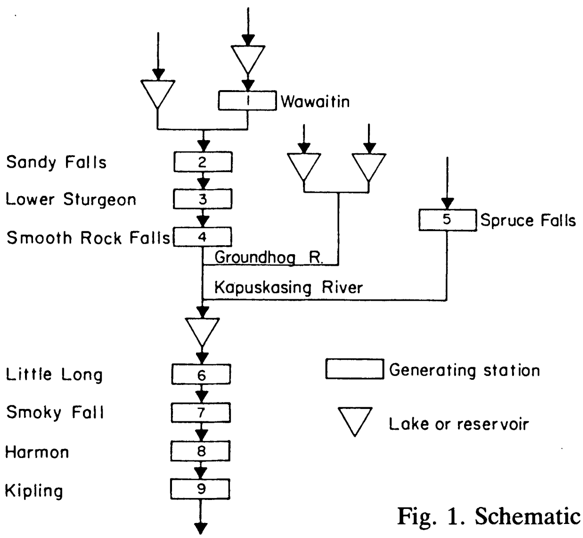
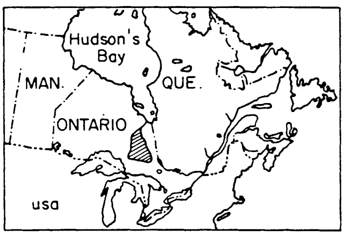



Fig. 1. Schematic illustration of the Mattagami River hydroelectric generating system.

utility of a forecasting system vary because of differences in hydrosystems, hydroelectric generating elements and the constraints imposed on the operation. Despite the differences some general features often occur, e.g., hydropower plants are usually in series (hydraulically coupled) along the main river and its tributaries, the discharge of any upstream plant and possible tributaries making up most of the inflows to the succeeding downstream plant. As an example, Fig. 1 shows a schematic illustration of the Mattagami River hydroelectric generating system in Northern Ontario, Canada. For cases like this, a real-time forecasting system in which upstream and tributary discharges are routed and combined may provide accurate inflow forecasts with sufficient lead times to be of use in short term hydropower scheduling. Traditionally, routing on rivers with tributaries has been done using estimated travel times or multiple correlation methods (WMO 1975). Difficulties in applying the traditional methods arise because discharge time series are generally highly autocorrelated. Furthermore, the tributary discharges are often cross correlated, due to similarities in the runoff generating mechanism. This correlation makes the identification of the systems dynamics very difficult.

Multivariate time series model in their transfer function-noise form are very applicable to multiple routing problems and lend themselves readily to real-time forecasting (Watt and Nozdryn-Plotnicki 1981; Chow *et al.* 1984). The main reasons are that the models involve a small number of variables, a short calibration period is needed, minimal computing power is required, and updating is computationally simple. The development of multivariate TFN models is somewhat complicated and the purpose of this paper is to present a simple method to develop a multivariate TFN model with correlated inputs. The method is used to develop a



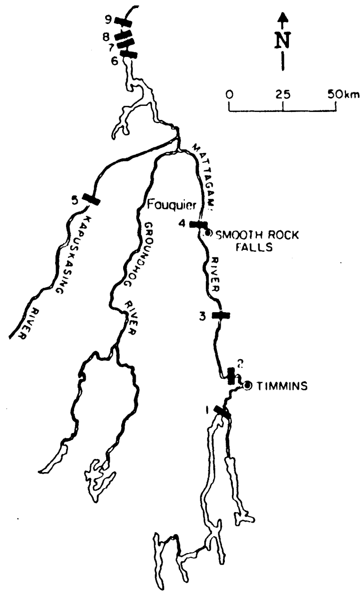




Fig. 2. The Mattagami River watershed.

forecasting system for the Little Long Generating Station on the Mattagami River (Figs. 1 and 2 ). Accurate short term forecasts are particularly important in the scheduling of run-of-river hydropower plants, such as the Little Long GS (Fjeld and Aam 1980; Thompstone *et al.* 1982; Ontario Hydro 1983). The small amount of live storage available can lead to spilling in the case of an unexpected increase in inflows, or loss of head in the case of a decrease in inflows. Significant spill losses occur frequently at Little Long. During a month of high inflows in 1982, spill losses were about \$ 750,000 (Ontario Hydro 1983).

## Multivariate Time Series Modelling

The univariate ARMA model is a linear stochastic model advocated by Box and Jenkins (1970) which is well suited for modelling discrete time series. A generalization of the univariate ARMA model into a multivariate time series model can be made without difficulties (Granger and Newbold 1976; Tiao and Box 1979, 1981). For $k$ series $(Z_{1t}, \ldots, Z_{kt})$, denote $z_{it} = \nabla^{d_i} Z_{it}$ where the order of differencing $d_i$ is appropriately selected to induce stationarity, then the ARMA model is generally referred to as the ARIMA model and takes the form

$$\Phi_p(B) z_t = \theta_q(B) a_t \tag{1}$$

where $\Phi_p(B) = I - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_p B^p$ and $\Theta_q(B) = \Theta_1 B - \Theta_2 B^2 - \ldots - \Theta_q B^q$ are matrix polynomials in $B$, the $\Phi$'s and $\Theta$'s are $k$ x $k$ matrices. $a_t \equiv \{a_{t1}, a_{t2},$

..., $a_{tk}\}$ is Gaussian white noise with zero mean and covariance matrix $\Sigma$. To ensure stationarity and invertibility, the zeros of the determinantal polynomials $|\Phi_q(B)| = 0$ and $|\Theta_p(B)| \equiv 0$ must lie on, or outside, the unit circle. The ARMA model given by Eq. (1) is a general model that allows for dynamic relations between all the series as well as feedback. However, in practical applications Eq. (1) can usually be simplifed and written in an alternative form. For example, causality and feedback (or lack therof) between hydrological processes can generally be argued on physical bases. If a unidirectional causal relation is assumed between $l$ input variables $X_{1t}$, ..., $X_{lt}$ and $m$ output variables $Y_{1t}$, ..., $Y_{mt}$, and denoting $x_{jt}$ and $y_{it}$ as differenced series, Eq. (1) can be written as

$$y_{it} \equiv \sum_{j=1}^{l} \frac{\omega_{ij}(B)}{\delta_{ij}(B)} B^{d_{ij}} x_{jt} + N_{it} \qquad i = 1, \ldots, m \tag{2}$$

where the 'noise' $N_{it}$ can be represented as a multivariate ARMA model.

$$\Phi_p(B) n_t \equiv \theta_q(B) a_t \tag{3}$$

where $\Phi(B)$ and $\Theta(B)$ are m × m matrix polynomials in B, $n_t = \nabla^{di} N_t$ and $a_t$ is a multivariate white noise sequence with covariance $\Sigma$. The form of ARMA model given by Eq. (2) is also referred to as multiple-input, multiple-output (MIMO) transfer function-noise model. The following discussion is limited to multiple-input, single-output (MISO) cases and so Eq. (2) can written as

$$y_t = \sum_{j=1}^{l} \frac{\omega_j(B)}{\delta_j(B)} B^{d_j} x_{jt} + \frac{\theta(B)}{\Phi(B)} a_t \tag{4}$$

Eq. (4) is often written in an alternative form to show the transfer function and noise components explicitly.

$$y_t \equiv \sum_{j=1}^{l} V_j(B) x_{jt-b_j} + N_t \tag{5}$$

in which $b_j$ is the lag time of the j-th input series and $V_j(B) \equiv V_{oj} + V_{1j}B + \ldots +$ is called the impulse response function of the system. The input series $X_{jt}$ can often supply additional information (leading indicators) not already supplied by past values of the outputs thus resulting in better forecasts.

The TFN model Eq. (5) can be reduced to more simple multivariate models. If there is only one output and the noise $N_t$ is assumed white Gaussian, Eq. (5) reduces to

$$y_t = \sum_{j=1}^{l} V_j(B) x_{jt-b_j} + a_t \tag{6}$$

where $y_t$ is related to current and past values of the $x_{jt}$ series. This model is often referred to as 'distributed lag' regression model. The effect of violation of the assumption of white noise error terms: can be drastic, as demonstrated by Granger and Newbold (1974). Three major consequences of autocorrelated error terms are that

1) estimates of the regression coefficients are inefficient,
2) forecasts based on the regression analysis are suboptimal, and
3) the usual significance tests on the coefficients are invalid.

The basic difference between ARMA models and regression models is that ordinary regression systems are static; a disturbance $a_t$ entering the system at time $t$ affects only $y_t$ but not $y_{t+1}$. The system has no 'memory' or dynamics. However, the ARMA model has memory, i.e., a disturbance affecting the system is 'remembered' and continues to affect the system at subsequent times. It is this memory that gives rise to dependence in the series and is represented by the ARMA model. Therefore, linear regression is often referred to as static regression, and ARMA models as dynamic regression (Pandit and Wu 1983).

## Identification of Systems Dynamics

A major impediment to the identification of systems dynamics is the occurrence of unaccounted for fluctuations, commonly designated as noise. Any adequate model of a dynamic system must include a proper characterization of the noise along with the transfer function. The appropriate model structure should be revealed not assumed *a priori*. Box and Jenkins (1970) suggested a general modelling methodology which consists of three iterative steps:

1) identification of a tentative model form,
2) estimation of parameters in the proposed model, and
3) diagnostic checking of the fitted model.

All of the above steps are repeated until a satisfactory formulation of the relations has been identified and the parameters estimated.

The design of a TFN model involves identifying the number of parameters required in the $V_i(B)$ operator and the $N_t$ terms of Eq. (7). This development can be done either in the frequency or in the time domain. The former approach dates back to the pioneering work of Wiener (1949) and has been successfully applied in hydrology. An example of application to runoff forecasting is given by Huthmann (1981). As for the time domain approach, several different identification methods have been developed and applied. Two widely used identification methods are those suggested by Box and Jenkins (1970) and Haugh and Box (1977). Both methods rely upon cross correlation studies of related series and are designed for single-input, single-output (SISO) cases but can be extended to MISO cases if the input series are uncorrelated. A comprehensive study on the distribution of cross correlation (ccf) estimates and associated covariances was done by Bartlett (1955). The study revealed that if two autocorrelated time series are cross correlated, the autocorrelation inherent in each series can inflate the variance in the cross correlation estimates above what could be expected when correlating two white noise series. Also, the ccf estimates at different lags can be highly correlated (Haugh and Box 1977). This makes identification difficult and, as shown by Box and Newbold (1971), the danger is that some significance can be assigned to apparent patterns in the ccf which are in fact sampling properties of the estimates used. Thus, the identification procedure would be considerably simplified if the input series are white noise series. Time series are usually autocorrelated making the identification of the appropriate structure of distributed lag regression models difficult. Usually several different model forms are fitted and compared on the basis of the coefficient of multiple determination and parameter values. The repeated three steps of identification, estimation and diagnostic checking are often ignored.

Both the Box and Jenkins and Haugh and Box methods are based on the assumption that the series can be adequately modelled by an ARMA model. Thus the series can be pre-whitened by fitting appropriate ARMA models, yielding white-noise residual series. The Box and Jenkins approach is to determine an appropriate ARMA model to fit the input series, following the three stages of model construction. The ARMA filter from the univariate modelling of the input series is then used to transform the output series. Box and Jenkins (1970) showed that the estimate of the ccf between the pre-whitened input and transformed output is directly proportional to the impulse response function. The noise model can then be identified from the residual series of the TFN model, using standard Box and Jenkins procedures.

In the Haugh and Box approach both the input $X_t$ and the output $Y_t$ series are prewhitened by fitting an appropriate ARMA model to each series. The linear nature of the ARMA filters ensures that if a meaningful relation exists between the original series $Y_t$ and $X_t$, the innovation series, $a_{xt}$ and and $a_{yt}$ driving each filter can be expected to show the same relation independently of the ARMA filters fitted. The first step is to develop a TFN relating the two innovation series; the form and initial estimates of the impulse response function can be identified from the estimated ccf between the innovation series. The final TFN model relating $Y_t$ to $X_t$ is constructed by recombining the two univariate models for the $Y_t$ and $X_t$ series with the TFN model connecting $a_{yt}$ and $a_{xt}$. Standard Box and Jenkins procedures can be used to fit the model and to estimate the noise term.

The identification and fitting procedures given above can be extended to MISO systems as long as the inputs are mutually uncorrelated. The procedure is to develop a transfer function between each input series and the output series using standard procedures. A combined model consisting of the transfer functions plus a single noise term is constructed by combining the models. This has been referred to as the 'superposition' approach by Wright and Bacon (1974), who showed that, in the case of correlated inputs, the superposition approach can lead to an acceptable form of the combined model. They indicated that parameter estimates can be obtained by simultaneously reestimating all the parameters, but that for models with many parameters, this procedure may result in serious convergence problems.

Several procedures that require transformation of the inputs series have been developed and applied to correlated inputs. All input series except one are decomposed into an explained part and orthogonal residual part. The transformation of the original inputs into a set of uncorrelated inputs and the building of the transfer function is done in a stepwise manner. In any single step the additional contribution of each input series to explaining the output series is added to the model. The transformation becomes complicated in the case of several correlated inputs; it involves many parameters and is difficult to apply. For example, in the case of three correlated inputs say $X_1$, $X_2$ and $X_3$, a model relating the output variable Y to one of the inputs say, $X_1$, is first developed

$$Y = V_1(X_1) + e_1 \tag{7}$$

where $e_1$ is the portion of Y unaccounted for by $X_1$. Another TFN model relating the second input $X_2$ to the first is then developed

$$X_2 = V_2(X_1) + e_2 \tag{8}$$

where $e_2$ is the portion af $X_2$ unaccounted for by $X_1$, that is the portion of $X_2$ that is uncorrelated with $X_1$. A third TFN model relating the third input to the first and the portion of the second that is uncorrelated with $X_1$ ($e_2$) is developed

$$X_3 = V_3(X_1) + V_4(e_2) + e_3 \tag{9}$$

where $e_3$ is the portion of $X_3$ unaccounted for by $X_1$ and $X_2$. A fourth model relating $e_1$ to $e_2$ and $e_3$ is developed

$$e_1 = V_5(e_2) + V_6(e_3) + e_4 \tag{10}$$

or by substituting for $e_2$ and $e_3$

$$e_1 = V_5(X_2 - V_2(X_1)) + V_6(X_3 - V_3(X_1) - V_4(e_2)) + e_4 \tag{11}$$

The final model is obtained by combining Eqs. (7) and (11)

$$Y = V_1(X_1) + V_5(X_2 - V_2(X_1)) + V_6(X_3 - V_3(X_1) - V_4(e_2)) + e_4 \tag{12}$$

A procedure based on this principle has been successfully applied to runoff modelling (two correlated inputs) by Snorrason *et al.* (1984). Another approach to modelling correlated inputs was developed by Akaike (1976) using a canonical correlation analysis. The approach is based on a Markovian representation of stationary time series and the identification is based on minimizing Akaike's information criterion for different models that have been fitted using maximum likelihood method. Few applications of this approach to modelling stochastic processes have been reported in the literature. It may be very difficult to apply due to the complexities of the calculations required.

## Overview

The Mattagami River, located in northern Ontario (see Fig. 2.), flows from south to north into James Bay; the total drainage area above Little Long GS is 36,470 $km^2$. Two major tributaries are the Groundhog and the Kapuskasing Rivers.

Streamflow records were available for the three years from 1980 to 1982 for the period April 1st to October 31st each year. These records consist of average daily discharge for the Mattagami River measured at Little Long GS and at Lower Sturgeon GS (8,300 $km^2$); the Groundhog River at Fauquier (11,900 $km^2$) and the Kapuskasing River at Spruce Falls GS (6,760 $km^2$). The existing real-time forecasting system is a simple graphical extrapolation; that is, the forecaster plots the inflows to LLGS for the previous few days and projects graphically.

## Model Development

Correlation analyses of the time series for each of the three realizations available revealed that the series are non-stationary and differencing of order one was identified as sufficient to induce stationarity. Furthermore, the inputs (Lower Sturgeon, Kapuskasing and Fauquier) are highly correlated. In order to identify the relations between the output and the inputs it is necessary to pre-whiten the series. Univariate time series models were developed using the three stages of model building. The selected models are listed in Table 1 and parameter estimates in Tables 2 to 5.

The series were pre-whitened using the models selected and parameter estimates obtained for each year. The ccf estimates between the pre-whitened inputs and the pre-whitened output are shown in Fig. 3. The cross correlation structure is very similar; there is significant correlation at lags 0-2. The ccf estimates between each of the pre-whitened input flows series shown in Fig. 4 confirm the relations between the input series, not that there is any causal relations among the tributary flows, but that the innovation series 'driving' each system are related.

Because the routing dynamics from each of the of the gaging sites appear to be similar and the flow series are correlated, the separate identification of the individual effects of each input series on the output series is difficult and the superposition approach is not applicable. The development of a MISO TFN model by transforming the three correlated inputs into a set of orthogonal inputs is compli-

Table 1 – Selected time series model for each of the four sites.

| | |
|---|---|
| Little Long GS ($Y_t$) | ARIMA (1,1,0) |
| Lower Sturgeon ($X_{1t}$) | ARIMA (0,1,2) |
| Kapuskasing ($X_{2t}$) | ARIMA (2,1,0) |
| Fauquier ($X_{3t}$) | ARIMA (1,1,1) |

Table 2 – Parameter estimates and standard diviations for LLGS ARIMA (1,1,0) model

| parameter | 1980 | | 1981 | | 1982 | |
|---|---|---|---|---|---|---|
| | estim. | st.dev. | estim. | st.dev. | estim. | st.dev. |
| $\phi_1$ | 0.65 | 0.07 | 0.51 | 0.06 | 0.64 | 0.07 |
| residual st. error (m$^3$/s) | 66.1 | | 69.2 | | 83.8 | |

Table 3 – Parameter estimates and standard deviations for Fauquier ARIMA (1,1,1) model.

| parameter | 1980 | | 1981 | | 1982 | |
|---|---|---|---|---|---|---|
| | estim. | st.dev. | estim. | st.dev. | estim. | st.dev. |
| $\phi_1$ | 0.53 | 0.07 | 0.25 | 0.08 | 0.62 | 0.06 |
| $\theta_1$ | −0.46 | 0.07 | −0.71 | 0.06 | −0.49 | 0.07 |
| residual st. error (m$^3$/s) | 20.2 | | 18.2 | | 19.7 | |

Table 4 – Parameter estimates and standard deviations for Lower Sturgeon ARIMA (0,1,2) model.

| parameter | 1980 | | 1981 | | 1982 | |
|---|---|---|---|---|---|---|
| | estim. | st.dev. | estim. | st.dev. | estim. | st.dev. |
| $\phi_1$ | 0.50 | 0.07 | −0.15 | 0.06 | −0.15 | 0.07 |
| $\theta_2$ | 0.08 | 0.08 | −0.09 | 0.07 | −0.36 | 0.06 |
| residual st. error (m$^3$/s) | 13.0 | | 17.9 | | 26.9 | |

Table 5 – Parameter estimates and standard deviations for Kapuskasing ARIMA (2,1,0) model.

| parameter | 1980 | | 1981 | | 1982 | |
|---|---|---|---|---|---|---|
| | estim. | st.dev. | estim. | st.dev. | estim. | st.dev. |
| $\phi_1$ | 0.50 | 0.06 | 0.56 | 0.08 | 0.46 | 0.07 |
| $\phi_2$ | 0.26 | 0.06 | −0.13 | 0.06 | 0.26 | 0.07 |
| residual st. error (m$^3$/s) | 12.2 | | 14.9 | | 12.9 | |

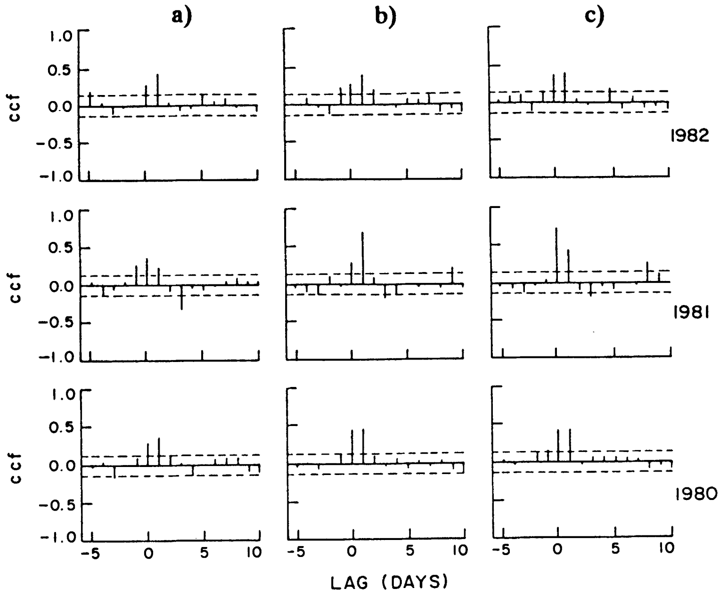



Fig. 3. Ccf estimates between pre-whitened inputs and pre-whitened output, with 95% C.L.
a) Lower Sturgeon, LLGS b) Kapuskasing, LLGS c) Fauquier, LLGS

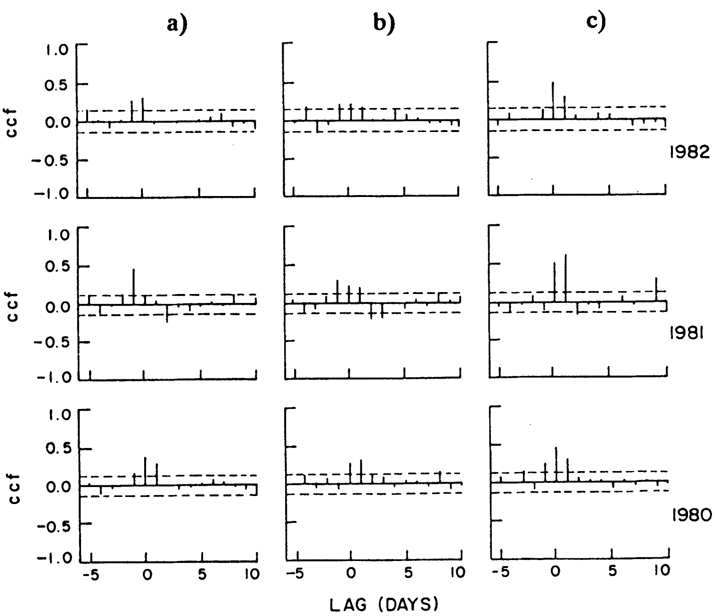



Fig. 4. Ccf estimates between each of the pre-whitened inputs, with 95% C.L.
a) Lower Sturgeon, Kapuskasing b) Lower Sturgeon, Fauquier c) Kapuskasing, Fauquier

cated and involves a large number of parameters as Eqs. (7) to (12) demonstrate clearly. Furthermore, for three correlated inputs a preordering of inputs is necessary. The ordering would be determined in such a way that each successive input variable would be correlated with its predecessors only at positive lags. However, such preordering of the input series is difficult in this particular case. There is no direct causal relationship between the tributary flows, but the correlation is caused by similarities in the runoff generating mechanism, i.e. spring breakup occurs at about the same time and rain on one subbasin is likely to occur at the same time as on the other two. There are however differences between events or years as Fig. 4 demonstrates. The response of the tributaries depends on spatial and temporal variability in rainfall and/or snowmelt inputs. There is no consistency in the correlation between the tributary flows making it difficult to express one input as a function of another plus orthogonal residuals in the form of Eqs. (7) to (12). Therefore the transformed input approach was not pursued.

The primary difficulty in the development of MISO TFN models with correlated inputs is the identification of a parsimonious model capturing the dynamic relationship between the output and the inputs. By carrying the analysis out in a stepwise manner, identifying, fitting and checking a SISO TFN model between the output and each of the inputs pairwise, a tentative identification of the MISO model can be made. However, the parameter estimates obtained from the SISO models are biased since they reflect the effects of the correlated inputs excluded from each analysis. The parameters do however provide good initial estimates to be used in simultaneous reestimation of the parameters in the combined MISO model. With a parsimonious form of the MISO model and good initial estimates of the parameters, convergence problems in the simultaneous reestimation of parameters can generally be avoided.

The structure of the estimated ccfs between the pre-whitened Fauquier $X_{3t}$ and LLGS $Y_t$ and between Kapuskasing $X_{2t}$ and LLGS led to the identification, fitting and checking of a model of the form

$$\nabla Y_t = (\omega_{0i} - \omega_{1i} B)\, \nabla X_{it} + (1-\theta_1 B-\theta_2 B^2)\, a_t \tag{13}$$

Parameter estimates and standard deviations are listed in Tables 6 and 7. The model relating Lower Sturgeon flows $X_{1t}$ to LLGS flows was identified as

$$\nabla Y_t = (\omega_{01} - \omega_{11} B - \omega_{21} B^2)\, \nabla X_{1t} + (1-\theta_1 B-\theta_2 B^2)\, a_t \tag{14}$$

Fitting and diagnostic checking confirmed the identification; parameter estimates along with standard deviations are listed in Table 8.

The final step is to combine the models and reestimate the parameters. The $\omega_{21}$ parameter is not significant in the combined model. The final model, of the form

$$\begin{aligned}\nabla Y_t &= (\omega_{01} - \omega_{11} B)\, \nabla X_{1t} + (\omega_{02}-\omega_{12} B)\, \nabla X_{2t} \\ &+ (\omega_{03} - \omega_{13} B)\, \nabla X_{3t} + (1-\theta_1 B-\theta_2 B^2)\, a_t\end{aligned} \tag{15}$$

Table 6 – Parameter estimates and associated standard deviations for the TFN model relating LLGS inflows to Fauquier flows.

| parameter | 1980 | | 1981 | | 1982 | |
|---|---|---|---|---|---|---|
| | estim. | st.dev. | estim. | st.dev. | estim. | st.dev. |
| $\omega_{03}$ | 1.64 | 0.16 | 2.50 | 0.13 | 1.89 | 0.21 |
| $\omega_{13}$ | –0.83 | 0.16 | –0.50 | 0.16 | –0.85 | 0.21 |
| $\theta_1$ | –0.13 | 0.07 | 0.04 | 0.07 | –0.14 | 0.07 |
| $\theta_2$ | –0.01 | 0.07 | –0.18 | 0.07 | 0.01 | 0.07 |
| residual st. error ($m^3/s$) | 48.4 | | 33.5 | | 63.5 | |

Table 7 – Parameter estimates and associated standard deviations for the TFN model relating LLGS inflows to Kapuskasing flows.

| parameter | 1980 | | 1981 | | 1982 | |
|---|---|---|---|---|---|---|
| | estim. | st.dev. | estim. | st.dev. | estim. | st.dev. |
| $\omega_{02}$ | 2.72 | 0.23 | 1.26 | 0.17 | 2.71 | 0.31 |
| $\omega_{12}$ | –2.22 | 0.23 | –3.33 | 0.16 | –2.95 | 0.31 |
| $\theta_1$ | 0.10 | 0.07 | 0.20 | 0.07 | 0.06 | 0.07 |
| $\theta_2$ | –0.03 | 0.07 | 0.07 | 0.07 | 0.23 | 0.07 |
| residual st. error ($m^3/s$) | 41.3 | | 36.1 | | 61.5 | |

Table 8 – Parameter estimates and associated standard deviations for the TFN model relating LLGS inflows to Lower Sturgeon flows.

| parameter | 1980 | | 1981 | | 1982 | |
|---|---|---|---|---|---|---|
| | estim. | st.dev. | estim. | st.dev. | estim. | st.dev. |
| $\omega_{01}$ | 1.55 | 0.31 | 1.57 | 0.25 | 0.91 | 0.19 |
| $\omega_{11}$ | –1.63 | 0.31 | –1.41 | 0.25 | –1.66 | 0.20 |
| $\omega_{21}$ | –1.46 | 0.38 | –0.47 | 0.24 | –0.70 | 0.19 |
| $\theta_1$ | –0.36 | 0.07 | –0.38 | 0.07 | –0.47 | 0.07 |
| $\theta_2$ | –0.21 | 0.07 | –0.15 | 0.07 | –0.24 | 0.07 |
| residual st. error ($m^3/s$) | 58.6 | | 61.1 | | 73.6 | |

provides a good fit. Furthermore, the residual standard error (Table 9) is significantly smaller than that of the time series model (Table 2) or the SISO TFN models (Tables 6, 7 and 8). The residuals are white noise, uncorrelated with any of the input series, and no evidence of model deficiency is apparent. As indicated in Table 9, there is some variation in parameter estimates from year to year. This variability is also evident in the SISO models (Tables 6, 7 and 8) and is not caused

Table 9 – Parameter estimates and associated standard deviations for the model relating LLGS inflows to tributary flows.

| parameter | 1980 | | 1981 | | 1982 | |
|---|---|---|---|---|---|---|
| | estim. | st.dev. | estim. | st.dev. | estim. | st.dev. |
| $\omega_{01}$ | 0.54 | 0.19 | 0.27 | 0.13 | 0.15 | 0.13 |
| $\omega_{11}$ | –0.74 | 0.19 | –0.40 | 0.13 | –0.73 | 0.14 |
| $\omega_{02}$ | 1.42 | 0.26 | 0.67 | 0.17 | 1.85 | 0.34 |
| $\omega_{12}$ | –1.55 | 0.23 | –1.67 | 0.24 | –1.49 | 0.34 |
| $\omega_{03}$ | 0.72 | 0.14 | 1.28 | 0.16 | 0.74 | 0.21 |
| $\omega_{13}$ | –0.10 | 0.14 | –0.06 | 0.12 | –0.21 | 0.20 |
| $\theta_1$ | 0.49 | 0.07 | 0.37 | 0.07 | 0.14 | 0.07 |
| $\theta_2$ | 0.14 | 0.07 | 0.07 | 0.12 | 0.26 | 0.07 |
| residual st. error ($m^3/s$) | 33.4 | | 29.6 | | 51.8 | |

by the fitting technique for the MISO model. A likely cause for this variation in parameter estimates is an actual difference in systems dynamics from year to year (i.e., the system is not time invariant). Recognition of this variation in parameter estimates leads to two points. First, can the variation be attributed to some physically measurable quantity that could be used to improve parameter estimates? Secondly, is the variation significant for forecasting purposes or will constant parameter values lead to acceptable forecasts? Consider first physical causes for the variation. In this regard, a useful function of the model parameters is the gain, g which is defined as the ratio of the steady state output to a steady state input and can be shown to be

$$g = \omega_{01} - \omega_{11} + \omega_{02} - \omega_{12} + \omega_{03} - \omega_{13} \tag{16}$$

Differences in the gain reflect differences in the additional contribution to streamflow from areas lying between the upstream flow gaging sites and Little Long GS. The gain for each of the three calibration years is listed in Table 10. The gain for the year 1981 is significantly different from 1980 and 1982. A comparison of the time series for the three years revealed that the only significant flood in 1981 was a snowmelt/rainfall induced spring flood. Hence, the spring flood had dominating effects on the parameter estimates for 1981. The time series from 1980 and 1982 showed similar behavior with floods in the spring, summer and fall. Ontario Hydro operates six snow courses in the basin and 1981 was characterized by a below average accumulation of snow, particularly in the northern part of the basin. The average April 1st water equivalents of snowpack from the six snow courses are listed in Table 11. The below average snow accumulation in 1981 in the areas downstream of the tributary flow gaging sites explains why the gain for 1981 differs from those for 1980 and 1982.

Table 10 – Steady state gain for each of the calibration years.

| year | 1980 | 1981 | 1982 |
|---|---|---|---|
| gain | 5.07 | 4.35 | 5.17 |

Table 11 – Average April 1st water equivalent of snowpack in the Mattagami River watershed.

| year | 1980 | 1981 | 1982 |
|---|---|---|---|
| water eq. (mm) | 130 | 94 | 150 |

## Forecasting Results

Before the MISO model described above can be used for forecasting, two steps are required. First, the dynamics of this system are such that a forecast one time step ahead requires a one time step ahead forecast of the tributary flows and therefore, for forecasting, the univariate time series models developed for each of the tributaries were used in combination with the MISO TFN model given by Eq. (15). Secondly, values for the parameters must be selected. A logical approach might be to simply average the values from the three years. However, this might not give the best forecasts. To evaluate this approach and to assess the robustness of the model, the following analyses were carried out. One and two day ahead forecasts were generated using four sets of parameter estimates: one set for each of the three calibration years (Tables 3, 4, 5 and 9) and a set defined by the average of the three years. The average standard errors of the forecasts are compared in Table 12. Within the range of the parameters used, the accuracy of the forecast results is not very sensitive to parameter values. This is an indication of the robustness of the model and indicates that even if data for only one year were available, an acceptable forecast model could be developed. Because the parameter set for 1982 yielded the smallest error, it was selected. A schematic illustration of the forecast system is shown in Fig. 5. For purposes of verifying the model, data for the years

Table 12 – Average standard error of LLGS inflow forecasts.

| | | Standard error ($m^3/3$) | | | |
|---|---|---|---|---|---|
| parameter set | | 1980 | 1981 | 1982 | Avg. |
| lead time | 1 | 58 | 60 | 52 | 56 |
| (days) | 2 | 114 | 123 | 110 | 120 |

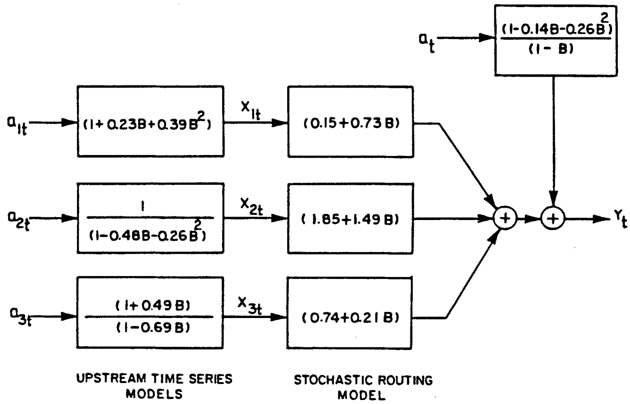



Fig. 5. Schematic illustration of the forecast system.

1983 and 1984 were used. The results given in Table 13 indicate that the forecasting accuracy did not seriously deteriorate using data outside the calibration data set. The one and two day ahead forecasts plotted in Figs. 6 to 9 show close agreement of forecast and observed flows.

Table 13 – Standard error of LLGS inflow forecasts.

| model | lead time (days) | Standard error ($m^3/s$) 1980 | 1981 | 1982 | 1983 | 1984 |
|---|---|---|---|---|---|---|
| | | calibration period | | | verif. | period |
| Time series | 1 | 66 | 70 | 84 | 90 | 79 |
| | 2 | 125 | 133 | 161 | 152 | 155 |
| | 3 | 189 | 191 | 237 | 219 | 223 |
| | 4 | 248 | 234 | 302 | 283 | 287 |
| MISO TFN | 1 | 47 | 44 | 64 | 62 | 60 |
| | 2 | 100 | 103 | 127 | 116 | 137 |
| | 3 | 159 | 165 | 193 | 177 | 205 |
| | 4 | 216 | 213 | 259 | 241 | 266 |

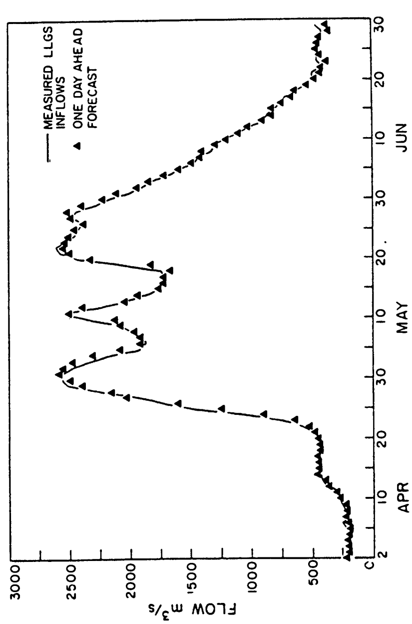



Fig. 6. Time series plot of one day ahead forecast of LLGS inflows for 1983.

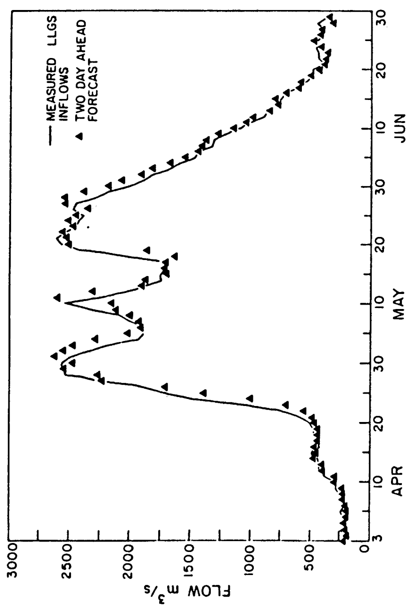



Fig. 7. Time series plot of two day ahead forecast of LLGS inflows for 1983.

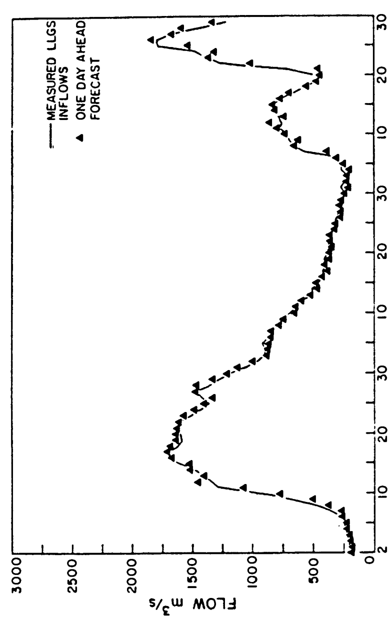



Fig. 8. Time series plot of one day ahead forecast of LLGS inflows for 1984.

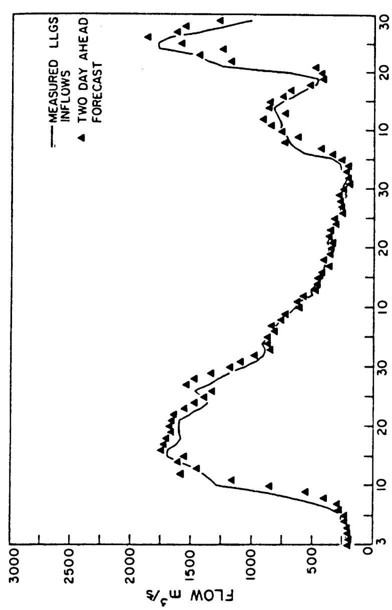



Fig. 9. Time series plot of two day ahead forecast of LLGS inflows for 1984.

## Conclusions

The development of a MISO TFN model with correlated inputs using the superposition of SISO models and simultaneous reestimation of parameters resulted in an acceptable model. This method is simple and is a viable alternative to complex methods based on transformation of the correlated inputs.The MISO TFN model explains more of the variability in inflows into LLGS and gives better forecasting accuracy than a univariate time series model of inflows.

The MISO TFN model produced good forecasts of daily flows at leads one and two days. The average standard errors were 8% of average inflows (1-day lead) and 18% (2-day lead).

Increased accuracy of the forecasts as well as increased lead time can be achieved by the development of more detailed models for the tributaries and possibly by the use of an on-line recursive parameter estimation algorithm. The period under consideration is April to October and a model relating geophysical inputs to runoff would have to account for snowmelt and losses due to evapotranspiration and would have to incorporate a seasonally variable transfer function.

## Acknowledgements

Financial assistance from a research grant under the Water Resources Research Program of the Inland Waters Directorate, Environment Canada is acknowledged.

## References

Akaike, H. (1976) Canonical correlation analysis of time series and the use of an information criterion, *Systems* identification (advances and case studies), ed. Mehra, F. and D. Lainiotis, Academic Press, New York.

Bartlett, M.S. (1955) *Stochastic processes*, University Press, Cambridge.

Box, G.E.P., and Jenkins, G.M. (1970) *Time series analysis, forecasting and control*, Holden-Day, San Francisco.

Box, G.E.P., and Jenkins, G.M. (1976) *Time series analysis, forecasting and control, 2nd edition*, Holden-Day, San Francisco.

Box, G.E.P., and Newbold, P. (1971) Some comments on paper of Coen, Gomme and Kendall, *J. Econometrics, Vol. 2,* 229-240.

Chow, K.C.A., Watt, W.E., and Watts, D.G. (1984) A stochastic-dynamic model for real-time flood forecasting, *Water Res. Res., Vol. 19(3),* 747-752.

Fjeld, M., and Aam, S. (1980) An implemation of estimation techniques to a hydrological model for prediction of runoff to a hydroelectric powerstation, *IEEE Transaction on Automatic Control, Vol 25(2)*, 151-163.

Granger, G.W.J., and Newbold, P. (1974) Spurious regression in econometrics, *J. Econometrics, vol 2,* 111-120.

Granger, G.W.J., and Newbold, P. (1976) *Forecasting econometric time series,* Academic Press, New York.

Haugh, L.D., and Box, G.E.P. (1977) Identification of dynamic regression (distributed lag) models connecting two time series, *J. Am. Statist. Assoc., Vol. 72,* 121-130.

Huthmann, G. (1981) Real-time flow forecasting by multiple input-single output model in the frequency domain, Proc. Int. Symp. on Real-time Operation of Hydrosystems. Waterloo, Ontario, 2, 885-889.

Ontario Hydro (1983) Inflow forecasting model for Little Long GS, Technical report No. 0534H, 28, Civil Works Department, Ontario Hydro, Toronto, Ont.

Pandit, S.M., and Wu, S.M. (1983) *Time series and system analysis with applications*, John Wiley and Sons, New York.

Snorrason, A., Newbold, P., and Maxwell, W.H.C. (1984) Multiple input transfer function-noise modeling of river flow, *Frontiers in Hydrology,* ed. by Maxwell, W.H.C., and Beard, L.R., Water Resources Publication, Fort Collins.

Thompstone, R.M., Hipel, K.W., and McLeod, A.I. (1982) Transfer function noise model for run-of-river operations, Proc. of IFAC Symp. on Theory and Application of Digital Control, New Delhi, India.

Tiao, G.C., and Box, G.E.P. (1979) An introduction to applied multiple time series analysis, Technical report No. 582, University of Wisconsin, Madison.

Tiao, G.C. and Box, G.E.P. (1981) Modelling multiple time series with applications, *J. Am. Stat. Assoc., vol. 78,* 802-811.

Watt, W.E., and Nozdryn-Plotnicki, M.J. (1981) Real-time forecasting for flood damage reduction, Proc. Int. Symp. on Real-time Operation of Hydrosystems. Waterloo, Ontario, 2, 741-760.

Wiener, N. (1949) *The extrapolation, interpolation and smoothing of stationary time series,* MIT Press, Boston.

World Meteorological Organization (1975) Operational hydrology, report No. 6, Hydrological forecasting practices, WMO No. 425, Geneva, Switzerland.

Wright, J.D., and Bacon, D.W. (1974) Statistical identification of dynamic behavior of a dual input heat exchanger network, *Can. J. Chem. Eng., Vol. 52,* 405-414.

First version received: 26 May, 1986
Revised version received: 24 July, 1986

**Address:**
Department of Civil Engineering,
Ellis Hall,
Queen's University,
Kingston,
Canada K7L 3N6.