STOCHASTIC MODELS FOR FORECASTING SNOWMELT RUNOFF

By

J. P. Haltiner[1] and J. D. Salas[1]

INTRODUCTION

The use of conceptual hydrologic models for short-term snowmelt/rainfall runoff forecasting (several hours to several days in advance) has a number of drawbacks including large computer storage requirements, difficulty in including real-time measured data in the forecast, and overparameterization (more model parameters than can be justified given the available data). A simpler conceptual model, the "Snowmelt Runoff Model" (SRM) developed by Martinec and Rango (Martinec, 1960; Rango and Martinec, 1979) has been shown to give good results on a number of basins of varying size (Rango, 1983). Difficulties associated with the use of the SRM include the following: parameter estimation, determination of the model form (number of previous runoff and snowmelt/ppt terms to include), use of real-time information to update the model, and difficulty in obtaining satellite imagery for snow-covered areas.

In this paper, a class of stochastic, time-series models referred to as ARMAX (Auto Regressive-Moving Average with Exogenous Inputs) or transfer function models (Box and Jenkins, 1970) are presented. It is shown that the Snowmelt Runoff Model can be viewed as a particular case within this general class of linear stochastic models. This offers a number of advantages. (1) Analytical techniques are available for adapting the model format to basins of varying characteristics and to varying time frames (daily, hourly, etc.). (2) A number of efficient parameter estimation techniques are available. (3) Confidence limits describing the accuracy of the forecast can be included with the forecast. (4) Diagnostic checks are available to determine if the model is performing properly. (5) Previous forecast errors can be included in the model to improve future forecasts. (6) The model can be cast in "systems format" and the Kalman filter can be used to update the parameters or status of the model in real time.

MODEL FORMULATION

The general univariate ARMAX model for river flow (adapted from Box and Jenkins, 1970) can be expressed as:

$$Q_t = \sum_{i=1}^{p} \phi_i Q_{t-i} + \sum_{j=0}^{q} \omega_j I_{t-j} + \sum_{k=0}^{r} \theta_k \varepsilon_{t-k} \tag{1}$$

where Q = streamflow

I = "moisture input" (combined snowmelt/rainfall) at time $t-j$

ε = white noise process

p,q,r = number of previous river flow, input and noise terms included in the model

ϕ,ω,θ = model parameters to be estimated

t = time sequence

Next, consider the Snowmelt Runoff Model equation in its commonly used form:

$$Q_t = Q_{t-1}K_t + \sum_{i=1}^{3} C_t\left[\{a_t^i(T_t^i - T_b)S_t^i + P_t^i\}A^i\right](1-K_t) \tag{2}$$

Presented at the Western Snow Conference 1985

[1]Department of Civil Engineering, Colorado State University, Fort Collins, Colorado 80523

where:

Q = streamflow

C_t = runoff coefficient

a_t = degree day factor

T_t = air temperature

T_b = base temperature

S_t = snow-covered area

P_t = precipitation

A = watershed area

K_t = recession coefficient

i = elevation zone

It can be seen that the forecast of flow at the next time step is a function of the flow at the current time step and effective snowmelt/rainfall during the next time step. If we denote this combined snowmelt/rainfall input from all elevation zones as I_t, let ω_0 equal the product of C_t and $(1-K_t)$ and call the recession coefficient ϕ_1, eqn. (2) becomes:

$$Q_t = \phi_1 Q_{t-1} + \omega_0 I_t \tag{3}$$

where

$$I_t = \sum_{i=1}^{3} (a_t^i (T_t^i - T_b) S_t^i + P_t^i) A^i \tag{4}$$

Equation (3) can be seen to represent a special case of eqn. (1), without the inclusion of a noise term. In adapting the SRM to a relatively large basin in Colorado (3419 km^2), Shafer et al. (1982) also included the snowmelt/rainfall input from the previous day, resulting in a model of the form

$$Q_t = \phi_1 Q_{t-1} + \omega_0 I_t + \omega_1 I_{t-1} \tag{5}$$

Again, this can be seen to be a special case of eqn. (1) with $p = 1$, $q = 1$, and no noise term.

The advantages of using the ARMAX formulation of eqn. (1) will be addressed in the following sections.

DESCRIPTION OF STUDY AREA AND DATA[1]

The ARMAX model described by eqn. (1) was used to forecast daily streamflows in the headwaters of the Rio Grande River in southern Colorado. The riverflow forecast point was at the Del Norte, Colorado gaging station. Drainage area above this point is 3419 km^2 (1320 mi^2), with elevations ranging from 2432 m (7980 ft) to 4215 m (13,830 ft). Riverflow behavior is dominated by snowpack accumulation and melt, resulting in a low-flow (snowpack accumulation) period between September and April and a high-flow (snowpack melting) period which begins in early April and persists through the summer. This basin was selected because extensive modeling with the SRM has already been conducted on the basin (Shafer et al., 1982), allowing use of identical data for model comparison and evaluation.

Daily precipitation, streamflow, average daily temperature and snow-covered area data for the period 1973 to 1980 were used in the forecasting evaluation. Only the peak flow period (April to September) in each year was examined. The precipitation, temperature and snow-covered area were used to calculate the moisture input, I_t, for each of three elevation zones using the SRM approach shown in eqn. (4). These zonal moisture inputs were then combined to represent a single basin moisture input, I_t, for use in eqn. (1). Although the SRM approach to snowmelt was used in this study to facilitate model comparison, any available snowmelt/precipitation model could be used to calculate the input term in eqn. (1).

[1]A complete basin description and listing of streamflow and climate data are available in the report by Shafer et al. (1982).

MODEL ESTIMATION

1. Model Format

The first step in estimation is to determine the model order (the values of p, q, and r) in eqn. (1). This can be done empirically, analytically or using a combination of both. Empirical identification is based on examining the physical processes involved. For example, Shafer et al. (1982) examined hourly flow records for the Rio Grande at Del Norte and determined that about 35 percent of a particular day's snowmelt/rainfall appears as runoff on that same day and 65 percent on the following day. Using this information, they determined that the SRM model should include two input terms (eqn. (5)). The analytical approach uses the cross-correlation structure between the input and output (flow) series to identify model order. In this approach, the input series is "prewhitened" by fitting a standard time series model to the data, resulting in a white noise process. This same model is then applied to the output series, and the cross-correlation at various time lags is calculated. This cross-correlation is scaled using the standard deviation of the transformed input and output processes to produce an impulse response function (Box and Jenkins, 1970):

$$v_k = \frac{\sigma_{Q'}}{\sigma_{I'}} \rho_{I'Q'}(k) \qquad k = 0,1,2,\ldots \tag{6}$$

where v = impulse response function

$\sigma_{I'}, \sigma_{Q'}$ = standard deviation of the transformed input and flow series

$\rho_{I'Q'}$ = cross-correlation between the transformed input and flow

k = time lag.

The impulse response function for daily input/flow data for the Rio Grande is shown in Fig. 1. When compared with known impulse response curves (Box and Jenkins, 1970), it indicates that the model should include one past flow term and two input terms, the same as that identified in Shafer et al. (1982). The analytical approach is especially useful in complex cases, such as hourly modeling or large basin modeling, where there may be several time periods of pure delay and a large number of previous flow and input terms to be included.

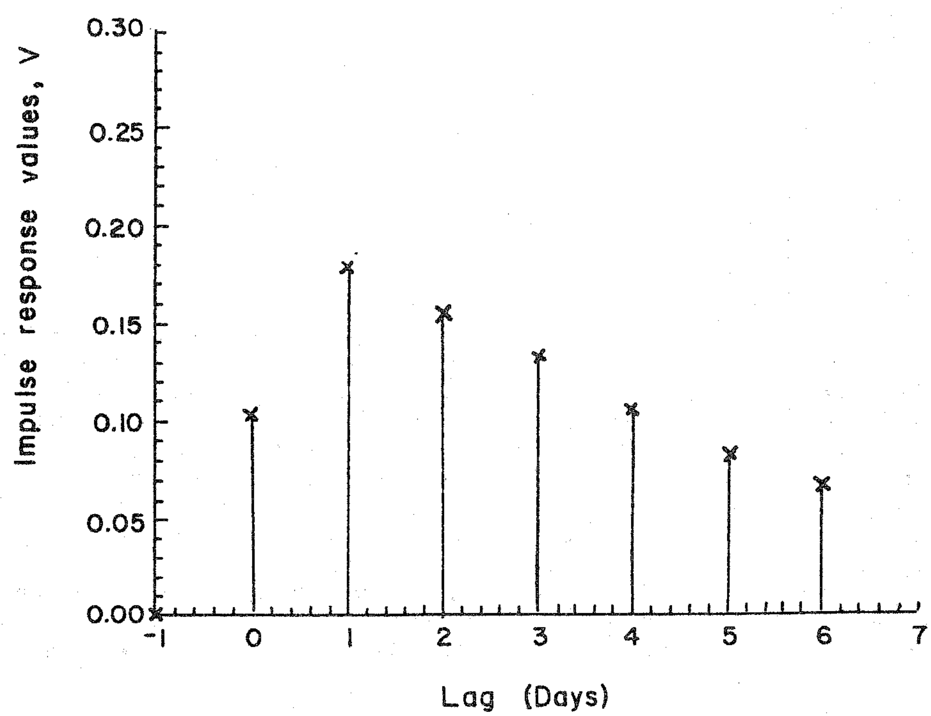

Figure 1. Impulse-response function for the snowmelt/precipitation input (impulse) and streamflow (response) on the Rio Grande River.

The number of noise terms to be included in the model will be determined by analyzing the residual series, ε_t, after preliminary estimation of the ϕ and ω model parameters.

2. Parameter Estimation

The model parameters (ϕ, ω and θ) were estimated using a nonlinear least squares (NLS) algorithm. In the NLS approach, eqn. (1) is solved for ε_t, which represents the difference between the observed and forecast flows. Parameters are selected using a Newton-Raphson iterative technique which minimize $\Sigma\varepsilon_t^2$. Under the conditions of stationarity and normality of the data, nonlinear least squares is an approximate maximum likelihood estimator. Since these conditions are not strictly met, the estimator must be considered suboptimal. Nonetheless, the NLS approach was found to provide consistent, stable parameter estimates.

Several schemes to account for the variation of basin hydrologic conditions with the ARMAX model were investigated. In the seasonal version, parameters were estimated on a monthly and biweekly basis. In a second approach, the level of flow in the river was used as an index of antecedent moisture conditions. Six thresholds of flow were chosen and a set of parameters was calculated for each level. The ϕ and ω parameters for the threshold and seasonal models are shown in Figs. 2 and 3. In Fig. 2 it can be observed that the ϕ_1 parameter is inversely proportional to flow while the ω_0 and ω_1 parameters (accounting for the current and previous days input) are generally proportional to flow. Thus, increasing flows indicate greater antecedent soil moisture and correspondingly, an increased amount of snowmelt/rainfall appearing as streamflow. The seasonal behavior of ω_0 and ω_1 in Fig. 3 suggests that early in snowmelt season, the proximity of the snowpack to the river results in a majority of a current day's snowmelt appearing as runoff on that same day. As the season progresses, the snowpack recedes and most of the snowmelt produced on a given day does not appear as runoff until the following day (represented by the fact that $\omega_1 > \omega_0$).

Two additional model parameterizations were evaluated. In the first, the snowmelt and rainfall terms are included in the model separately. The model for the Rio Grande River then becomes:

$$Q_t = \phi_1 Q_{t-1} + \omega_0 M_t + \omega_1 M_{t-1} + \omega_0 P_t + \omega'_1 P_{t-1} + \sum_{k=0}^{r} \theta_k \varepsilon_{t-k} \tag{7}$$

where $M_t = \sum_{i=1}^{3} (T_t^i - T_b)(S_t^i) \frac{A_i}{A^T}$

A^T = total area

$P_t = \sum_{i=1}^{3} P_t^i \frac{A_i}{A^T}$

In eqn. (7), M_t represents an areal average value of the product of degree days and snow-covered area and P_t represents an areal average value of precipitation. The degree day factor is now included in the ω_0 and ω_1 parameters. In addition to eliminating the need for estimating the degree day factor, the model form of eqn. (7) allows separate consideration of the effects of snowmelt and rainfall on river flow. Perhaps the primary difficulty in using the SRM approach is obtaining values of the snow-covered area (see eqn. (2)). Since aerial photography is too expensive to be used on a regular basis, satellite imagery has received primary emphasis. Problems associated with satellite data include cloud cover, estimation between time of satellite passes, and difficulty in obtaining imagery quickly for use in real-time forecasting. Alternatives to satellite imagery include snow-cover depletion curves based on accumulated degree days, or possibly relating areal snow cover to daily snow water equivalent (SWE) as measured at SNOTEL sites. An approach investigated here was to eliminate the snow-covered area term from the model. The melt term in eqn. (7) then becomes:

$$M_t = \sum_{i=1}^{3} (T_t^i - T_b) \frac{A_i}{A^T} \tag{8}$$

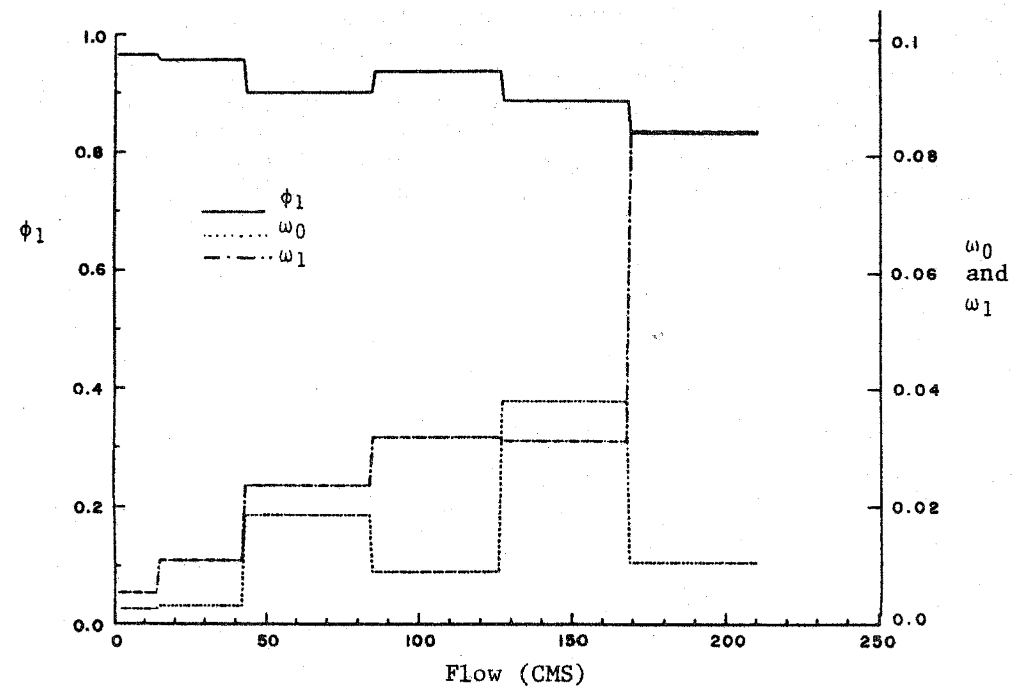

Figure 2. ARMAX model parameter estimates based on flow thresholds.

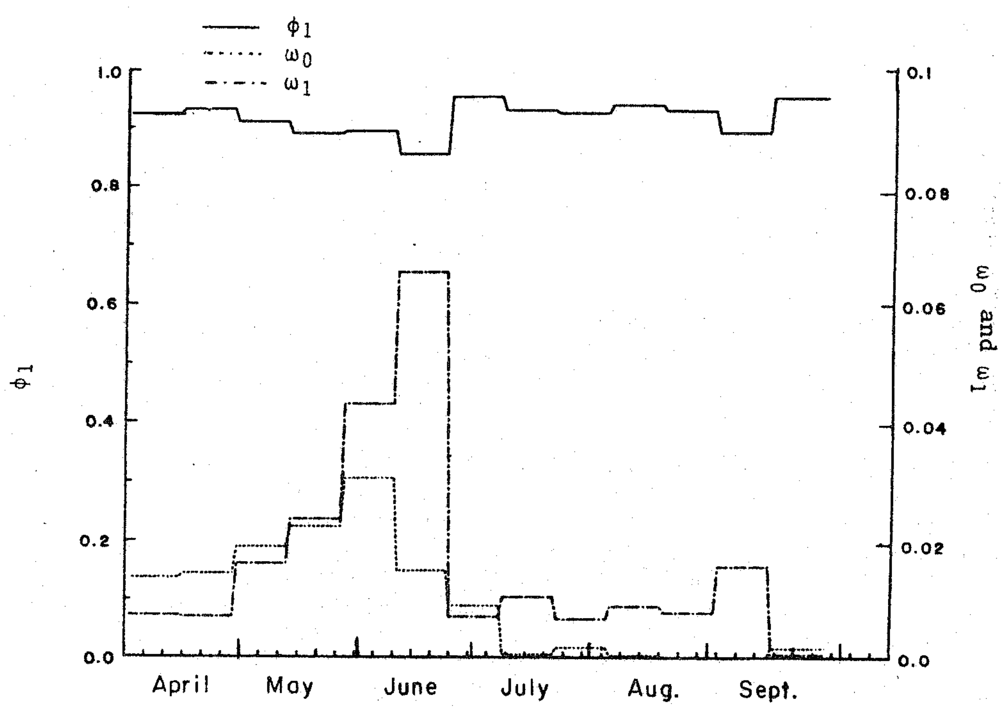

Figure 3. Biweekly estimates of the ARMAX model parameters.

It is obvious that the parameters ω_0 and ω_1 in eqn. (7) must also now account for the areal extent of snow available for melting, as well as the "effectiveness" of degree days in causing melt and the amount of melt to appear as runoff. ω_0 and ω_1 could not be seasonally estimated because of the dramatic differences year to year in the SCA. It was determined that the parameters could be allowed to vary on a daily basis using a Kalman filter to update the parameters based on measured data. This has been used in rainfall-runoff models for a number of years (Todini and Bouillot, 1975) and initial attempts in snowmelt runoff (Burns and McBean, 1985) appear promising. Daily values of ω_0 and ω_1 for the years 1975 and 1976 are shown in Fig. 4. It can be seen that the parameters

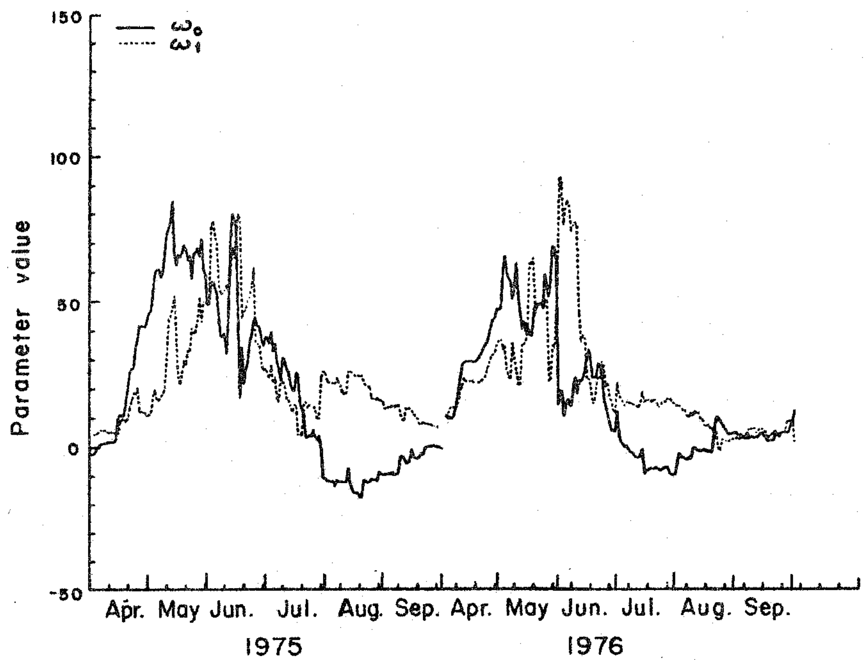

Figure 4. Kalman filter estimates of snowmelt parameters (ω_0 and ω_1) in the ARMAX model of eqn. (7) for the years 1975-76.

increase rapidly in the early snowmelt season, as the degree day effectiveness in causing melt and soil moisture are both increasing. At the peak of the snowmelt season, the parameters are approximately constant, as degree day effectiveness continues to increase, but snow-covered area is decreasing. This latter factor becomes dominant as the season progresses, and the parameters decrease. It can also be seen that in the early season, ω_0 (current day's melt) is dominant. As the season progresses, ω_1 (previous day's melt) becomes more important. The application of Kalman filtering has some inherent problems. The user must determine the noise covariance matrices in the filter, which determine how rapidly the parameters change. Analytical methods of estimating these matrices resulted in overestimation and subsequent filter instability. It appears that the problem results from inaccuracies in the measurement and areal averaging of precipitation and degree days. The fairly large errors inherent in these data may cause the parameter values to fluctuate too rapidly.

Two additional advantages of the ARMAX formulation are the availability of diagnostic checks to determine model adequacy and the calculation of confidence limits for a forecast. The presence of significant autocorrelation in the forecast error sequence is an indication of model inadequacy. This represents the presence of additional "information," in the data which the model is not using. Tests of the SRM model and ARMAX models without moving average terms showed the presence of residual autocorrelation. The inclusion of moving average terms increased the forecasting accuracy of the ARMAX model and eliminated the autocorrelation. Confidence limits for the forecast can be obtained using the variance of the forecast errors. Theory and results are available in Haltiner (1985).

RESULTS

The previously described models were used to make one- and three-day ahead flow forecasts for the period 1973-1980. In addition a simple ARMA(1,1) model which only uses previous values of streamflow (no snowmelt/rainfall inputs) was also included; it represents the forecasting accuracy which can be achieved relatively simply without attempting to model the rainfall or snowmelt. Models were compared using a mean squared error criterium (MSE). The model forecasting results are presented in Table 1. This table also includes the number of parameters which are estimated during the forecasting period and an indication of the presence of significant autocorrelation in the forecast errors.

Table 1. Comparison of one-day-ahead forecasting results for the period 1973-1980.

Model	MSE	No. of Param.	Res. Corr. Test
1. SRM-1 (cont. varying param.)	40.1	672	Failed
2. SRM-2 (seasonal param.)	41.3	84	Failed
3. ARMA(1,1) (threshold param.)	56.3	12	Passed
4. ARMAX-1 (combined input/threshold param.)	38.8	60	Passed
5. ARMAX-2 (combined input/ seasonal param.)	35.1	88	Passed
6. ARMAX-3 (sep. snow/ppt input, seasonal param.)	33.5	78	Passed
7. ARMAX-4 (K.F. est./no SCA data)	42.2	N/A	Passed

Models 1 and 2 represent the Snowmelt Runoff Models described by Shafer et al. (1982), adapted to include measured data. In Model 1, the parameters were re-estimated every 15 days for the eight-year forecasting period. In Model 2, average parameters were calculated for each 15-day period. Model 3 is the simple time series model with six sets of parameters based on flow thresholds. Models 4 and 5 correspond to the ARMAX models described in eqn. (1), using a combined snowmelt/rainfall input. Model 6 is the ARMAX model described in eqn. (7), with separate snowmelt and rainfall inputs. Model 7 uses the Kalman filter algorithm to update the parameters on a daily basis; it does not use the SCA data used in the other models.

A comparison of the MSE shows that the ARMA(1,1), which does not include rain or snowmelt, gives the poorest forecasts. The ARMAX Models 4, 5 and 6 give better forecasts with fewer parameters than the SRM models. Since the form of the SRM and ARMAX models is similar, the smaller MSE of the latter result from the analytical parameter estimation techniques and the inclusion of moving average terms. The presence of residual autocorrelation in the SRM models indicates that the model forecasts could be improved by including the previous model errors in the forecasts.

The number of parameters in Models 4 and 5 is somewhat misleading, compared with Model 6, since 4 and 5 each contain 36 estimates of the degree day factor (a_t^i in eqn. (2)), which were not analytically estimated. In general it was found that for models using equal numbers of parameters, seasonal and threshold parameter estimation gave about equal forecasting accuracy. Model 7 results show that relatively good forecast accuracy can be obtained without the snow-covered area data by using the Kalman filter approach. However, problems with parameter stability and estimation of the KF noise matrices suggest that this approach is still a research topic and not an operational tool at this time. Figure 5 shows the one-day-ahead forecast and measured streamflows for the 1973 snowmelt season using Model 5. Three-day-ahead forecasts for 1980 using Model 4 are shown in Fig. 6. (It should be noted that in using the ARMAX and SRM models in this study, the input data (temp. and ppt) on the forecast day are assumed known. In operational use, these data would be forecast, either using NWS forecast or autoregressive type forecasts.)

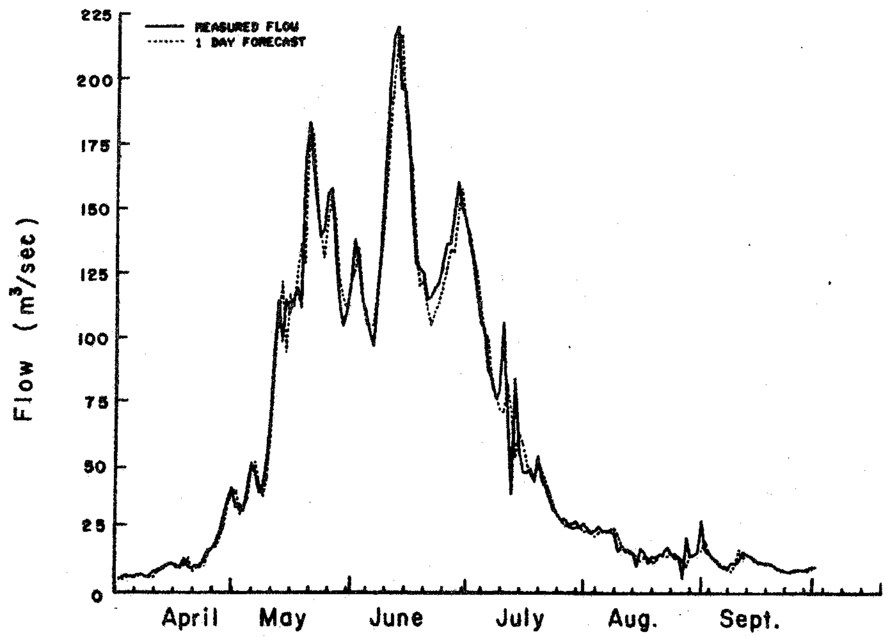

Figure 5. Comparison of measured flows and one-day-ahead forecasts for 1973 using Model 4 of Table 1.

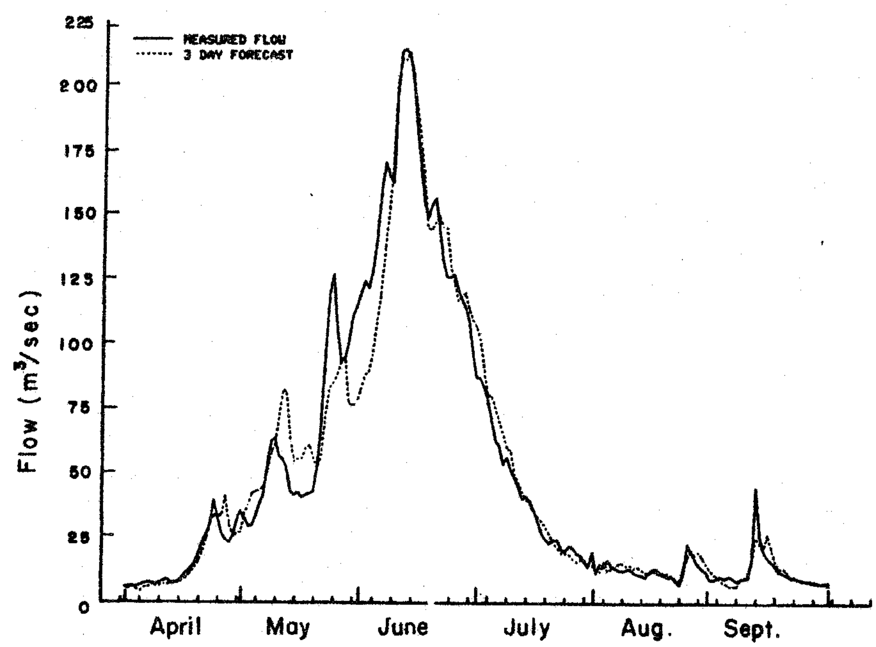

Figure 6. Comparison of measured flows and three-day-ahead forecasts for 1980 using Model 5 of Table 1.

CONCLUSIONS

The ARMAX formulation of the SRM model appears to offer a number of advantages over current procedures. Analytical techniques are available to identify model format and estimate model parameters. Diagnostic checks are available to insure that the model is operating properly, and confidence limits can be included with a forecast. The model can be updated in real time using previous forecast errors; in addition, the use of Kalman filtering to update the model parameters in real time appears promising.

ACKNOWLEDGMENT

This work was supported by the Colorado Experiment Station, Project Number 151141; Dr. H. W. Shen is the principal investigator. We would like to acknowledge Dr. B. Jones for making previously collected data available to us.

REFERENCES

Box, G.E.P. and G. M. Jenkins. 1970. Time Series Analysis, Forecasting and Control. Holden-Day Inc. San Francisco, California.

Burn, D. H. and E. A. McBean. 1985. River flow forecasting model for Sturgeon River. ASCE J. of Hyd. Eng., Vol. III, No. 2, February.

Haltiner, J. P. 1985. Modeling of Snowmelt Runoff. Unpublished Ph.D. dissertation (in progress). Colorado State University, Fort Collins, Colorado.

Martinec, J. 1960. The degree day factor for snowmelt-runoff forecasting. Proceedings: General Assembly of Helsinki, 1960. Commission on Surface Runoff. Publ. IASH No. 51, pp. 468-477.

Rango, A. 1983. Application of a simple snowmelt-runoff model to large river basins. Proceedings of the 51st Western Snow Conference, Vancouver, Washington, pp. 89-99.

Rango, A. and J. Martinec. 1979. Application of a snowmelt-runoff model using Landsat data. Nordic Hydrology, Vol. 10, pp. 225-238.

Shafer, B. A., E. B. Jones and D. M. Frick. 1982. Snowmelt runoff modeling in simulation and forecasting modes with the Martinec-Rango model. NASA Report CR 170452, Goddard Space Flight Center, Greenbelt, Maryland 20771.

Todini, F. and D. Bouillot. 1975. A rainfall-runoff Kalman filter model, System Simulation in Water Resources, ed. G. C. Vansteenkiste, N. Holland, Amsterdam, 1975.