REAL TIME STREAMFLOW FORECASTING USING A SNOWMELT INDEX MODEL AND TIME SERIES ANALYSIS TECHNIQUES

J. C. LaChance*, S. Lee*, P. J. Gabrielsen*, N. J. Gruber*, P. H. Kirshen*

* Gruber, Kirshen and Associates, P.O. Box 710, Brewerton, New York, 13029

ABSTRACT

A snowmelt index model was developed to generate an input time series suitable for forecasting snowmelt and streamflow in the Salmon River Watershed in central New York using a time series model. The snowmelt index model is a modified temperature index model that produces a continuous daily time series. The snowmelt index is generated based on air temperature, a melt factor (which is partially a function of snow density) and a lag coefficient (which is estimated using observed streamflow data). The output of the snowmelt index model is not an estimation of the quantity of melt generated by the snowpack, but rather an index of the magnitude of melt. The snowmelt index is then stochastically related to streamflow using a method of time series modeling known as transfer function noise modeling.

To support this approach a hydro-meteorological monitoring network was setup and operated in a real time fashion using the GOES satellite system.

Time series modeling offers several advantages over traditional conceptual model approaches; these include, simple data requirements, short calibration period and suitability for real time forecasting. When applied to forecasting inflow to a reservoir that is used for hydropower generation, a transfer function noise model, using a snowmelt index as an input series, can forecast short-term streamflow satisfactorily, and increase the efficiency of hydropower generation.

1.0 Introduction

The purpose of this paper is to present the use of a snowmelt index model as a component of a Real Time Streamflow Forecasting/Hydro Scheduling project for Niagara Mohawk Power Corporation (NMPC), Syracuse, New York. This Research and Development project is currently taking place in the Salmon River Basin, Oswego County, New York.

At present, Niagara Mohawk's System Power Control schedules their entire hydro system (ten river basins and over 80 hydro stations) manually. The scheduling system is labor intensive and relatively inflexible. Little extra time is available to test schedule changes, respond to unpredicted streamflow conditions and changes in system requirements. The Real Time Hydro Scheduling Project's goal is to automate in real time the current hydro scheduling procedure that NMPC uses. Given real time knowledge of the system state, a suitable forecast model and a scheduling optimization model, the scheduler can make the most efficient use of the available meteorologic and hydrologic data from the basin.

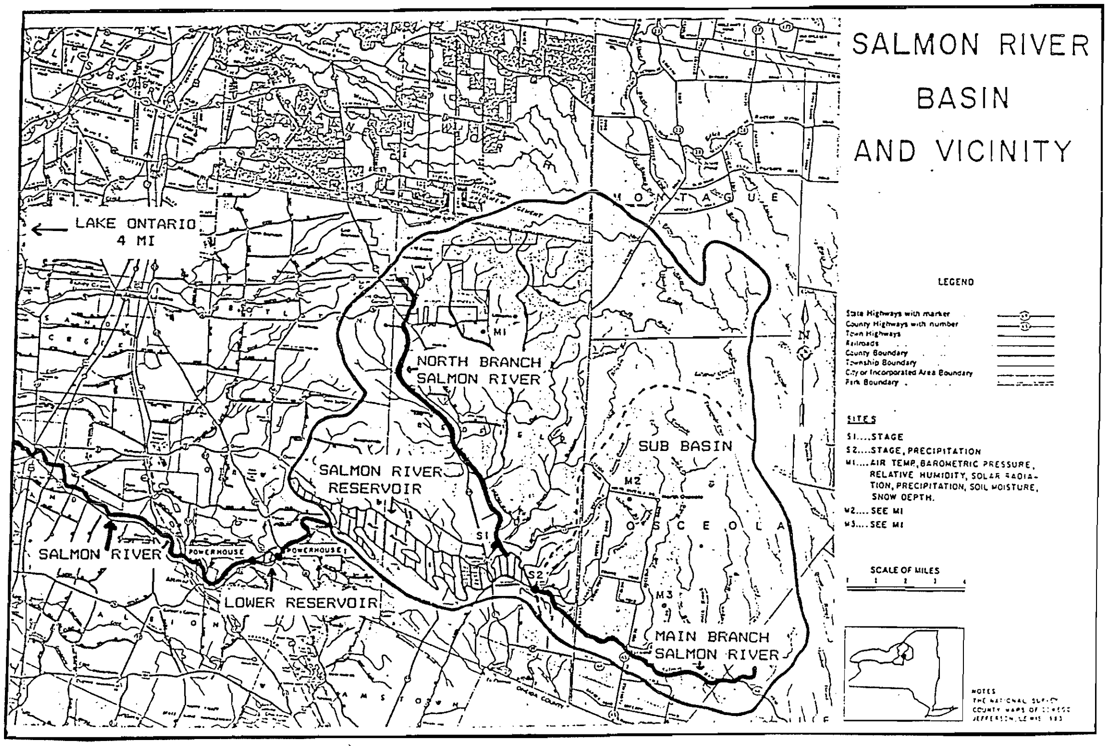

FIGURE 1

1.1 Location and Characteristics of the Study Area

The Salmon River originates in the Wards Hill Plateau of New York State, which is located just west of the foot hills of the Adirondack Mountains in the Tug Hill Region (see fig. 1). This region receives an average of 60 inches (152 cm) of precipitation per year, the majority usually as snow. From a starting elevation of 1750 feet (533.4 m)(AMSL) the river travels in a westerly direction until it empties into Lake Ontario.

Two rivers, the Main Branch and the North Branch of the Salmon, feed the Salmon River Reservoir. Located at an elevation of 935.0 feet (285.0 m) (AMSL), this reservoir has a drainage area of 191 square miles (494.7 km^2). The reservoir has a storage capacity of 50,000 ac-ft (6.17*10^7 m^3). and is used as a seasonal storage reservoir. Downstream there are two hydropower plants. The reservoir's primary function is to store water for hydro-electric generation, although it also serves as a means of flood control and recreation.

1.2 Real Time Hydro-Met Data Collection

The hydro-met parameters are monitored in the field by meteorological sensors. Parameter information is collected by Data Collection Platforms (DCPs) and transmitted in real time via the GOES satellite system to a direct read out ground station that is owned and operated by NMPC. The direct read out ground station processes the DCP transmissions and stores them in a hydro-met database.

The data collection platforms are a significant component of the hydro-met monitoring network. The DCPs are programmed to scan the sensors every 30 minutes. At four hour intervals the DCPs transmit the stored information to the direct read out ground station at NMPC via the GOES satellite.

1.3 Hydro-Met Network Location and Description

The hydro-met network is located within the watershed boundries of the Salmon River Reservoir. Included in the hydro-met network are two stream gauging and four meteorological monitoring sites. The stage monitoring sites are located on the two major tributaries to the Salmon River Reservoir. Stream stage is converted to discharge through the use of a rating curve.

The locations of the meteorological sites were choosen such that variations in the weather over the watershed would be recorded. The parameters monitored at these sites include: precipitation, insolation, air temperature, relative humidity, barometric pressure, snow depth, and soil moisture.

2.0 Approaches to Modeling Snow Accumulation-Ablation and Forecasting Streamflow

The annual weather pattern of the Salmon River basin has a unique effect on the seasonal hydrologic character of the watershed. During the winter and spring months, in areas that support thick, well developed snowpacks, such as the Salmon River Watershed, the rate and amount of snowmelt is the dominant factor controlling streamflow. Baseflow, during the winter, in the Salmon River Watershed is primarily a function of melting at the base of the snowpack.

A major portion of the yearly discharge observed in the Salmon River occurs during the spring. Significant releases from the snowpack may also occur during periods of winter thaw. Accordingly, a streamflow forecasting model for the Salmon River must account for snowmelt.

Before detailing the snowmelt index - time series analysis approach to forecasting streamflow, it is necessary to first examine the traditional approaches to understand why they may not be the most suitable for the objectives of this project.

2.1 Traditional Approaches

Traditional conceptual models of snow accumulation-ablation and streamflow involve the representation of the responsible physical processes with mathematical expressions. The processes generally estimated are snow accumulation, interception, heat exchange, liquid water retention and transmission within the snowpack, areal variability of melt and the associated land and channel phases of flow (Archer, 1983). Depending on the climatic, vegetative and topographic variability of the area being modeled, various combinations of these components are used to construct a satisfactory model. "Satisfactory," in this sense, refers to the ability to estimate, within acceptable error limits, the amount of melt or discharge from a basin over the snow season.

A watershed (or subsections of a watershed) with fairly uniform topographic features and vegetative cover can be reasonably modeled utilizing only a few of the more important components. Snow accumulation is fairly uniform over areas with similar elevation and vegetative characteristics and snowmelt in such a situation is primarily the result of heat exchanges at the surface and within the snowpack.

2.1.1 Energy-Mass Balance Models

Energy-mass balance models account for changes in mass of the snowpack based on the physics involved with energy-mass exchanges. The production of melt water is calculated by keeping track of fluxes in energy and mass between and within the various layers of the snowpack. Water is stored, transmitted or released depending on the energy balance of the snowpack. If enough energy enters the snowpack melt water will be produced.

This approach to modeling a snow cover would seem to offer the best results. With each factor of the accumulation-ablation process accounted for, generating highly accurate forecasts over a broad spectrum of conditions would appear to be possible. All that would be necessary to model a snow cover in this manner are the appropriate mathematical expressions representing the controlling physics and the measurement of the required input parameters. But this is where energy-mass balance models falter.

One of the drawbacks to using an energy-mass balance model is setting up the required data collection network. A fairly large number of parameters must be accurately measured at a relatively high density level (Colbeck et al., 1979). For example, the data inputs required by E. A. Anderson's point energy-mass balance model of a snow cover include air temperature, vapor pressure, wind speed, incoming solar radiation, reflected solar radiation, incoming longwave radiation, amount and temperature of rainfall, plus the soil temperature and the state of the snow cover (Anderson, 1976).

The expense to set up and maintain such a network is high. Thus, compromises in network design are usually made. Instead of collecting the full realm of necessary parameters at the appropriate density level, some are approximated from the parameters that are sampled and some from empirically derived relationships. A coarser network inherently makes more assumptions about the areal variability of the data gathered.

Another weakness in this approach is that energy-mass balance models usually simulate melt production in terms of a single point. To date, a satisfactory method has not been derived to extrapolate point determinations of snow cover conditions and melt production to basin wide representative values. This poses a problem when attempting to apply these models to operational situations. Also, energy-mass balance models only provide an estimate of the melt water produced; some other method must be employed to transform the melt into streamflow.

For the above reasons, energy-mass balance models are usually limited to research oriented applications. They are also used in situations where the benefits of the ability to accurately model snowmelt outweigh the costs of designing and installing such a system.

2.1.2 Temperature Index or Degree-Day Approach

Because the expense of installing and maintaining data collection networks for energy-mass balance models may be cost prohibitive, even on relatively small watersheds, many snowmelt modelers utilize what are known as temperature index or degree-day models. This group of models, under certain limited conditions, may perform as well as energy-mass balance models.

Temperature index models forecast melt based on the assumption that there is a strong correlation between air temperature (energy available to the snowpack for melting) and streamflow. The diurnal air temperature range is used as a direct index of snow cover outflow.

The correlation between air temperature and snowpack melting is the highest under fairly uniform meteorological conditions (Anderson, 1979). This is typically the case in forested watersheds. Within a forest, temperature regimes, wind speed, snow depth and other meteorological factors have moderate or negligible areal variation.

A simple formulation of the temperature index relationship to snowmelt can be stated as follows:

$$M = a(T-T_c)$$

where "M" is melt; "a" is an empirical coefficient - the melt factor (units cm/°C); "T" is the measured temperature in degrees celsius and "T_c" is the critical temperature (usually 0°C) above which melt occurs (Gray and Male, 1981). The expression ($T-T_c$) is the temperature index, which represents the amount of energy available for producing melt on any given day. This quantity is expressed in average number of degrees above a critical non-melt temperature (T_c) per day. Negative temperature indices are treated as non-melt conditions. The melt factor (a) relates the temperature index to the amount of water (in units of depth) produced by melting on that day (Hawley et al., 1980). By multiplying the depth of melt produced on a given day by the amount of area covered with snow on that day, an approximation of the volume of melt produced can be attained. From this melt estimation and the empirically derived runoff coefficient for the basin, an estimation of streamflow can be a made.

There has been a considerable amount of work done on extrapolating temperature and precipitation data from point measurements to basin wide representative values. The techniques used for performing these extrapolations have been extensively evaluated. For the purpose of estimating melt using temperature index models, these methods have proven to yield satisfactory results.

For the purposes of estimating melt and correspondingly forecasting discharge on the Salmon River Watershed, choosing a temperature index model would initially appear to be a satisfactory choice. The watershed has a fairly uniform vegetation cover and can be subsectioned into areas of similar topographic relief. But there are difficulties configuring this type of model to operate within the constraints of a real time forecasting system. Questions arise as to choosing the proper updating procedure, appropriate computational time interval and importantly, how to route the melt water to the forecast point. However, as will be discussed later, the basic underlying theories of temperature index or degree-day approaches, provide a basis for the development of a modeling method that appears to combine the appropriateness of temperature index models with a structure that supports a real time forecasting operation.

2.1.3 Pseudo Energy Balance Models

Pseudo energy balance models combine the atractiveness of temperature index models, i.e., the use of readily available temperature and precipitation data, with the versatility and modeling accuracy of energy balance models. They are (like true energy-mass balance models) conceptual in nature; they use mathematical expressions to represent the processes responsible for the initiation, development and ablation of a snowpack.

A typical example of a psuedo energy balance model is the snow accumulation-ablation component of the National Weather Service's River Forecasting System (NWSRFS) (Anderson, 1973). It is not a true energy balance model because it uses air temperature alone to approximate the energy balance of the snowpack. It also differs from temperature index models in that information about the diurnal temperature range is used as an index to the energy exchange at the snow-air interface instead of the calculation of snow cover outflow in the basin. Information about the energy exchange and melt is then coupled with the various conceptual components of an energy balance model to account for water retention, transmission and refreezing in the snowpack. Psuedo energy balance models attempt to account for and simulate the various processes that affect the formation of a snowpack and ultimately the generation of snow cover outflow using primarily air temperature and precipitation data.

The major limitation with using such a modeling approach is that an extensive historical database is necessary for calibration purposes. The snow model used in the NWSRFS utilises 6 major parameters and 6 minor parameters. The minor parameters are determined by examining the typical climatic and snow cover conditions, while the major parameters are estimated using the historical data record. Lengthy records of precipitation, temperature and discharge from the basin are required to assemble hydrographs and estimate the model's parameters.

For instance, the National Weather Service recommends using a record of at least 8 to 10 years worth of data for calibrating their river forecasting model (Anderson, 1973). It is clear that not all basins are going to have this long of a database to work with. This is a problem when attempting to use such a model on a watershed like Salmon River.

2.2 Application of Time Series Analysis to the Modeling of a Snow Cover

Because of the requirements of real time hydropower operation and the limitations of the above traditional approaches to snow accumulation-ablation modeling, research into alternative methods of estimating snowmelt and forecasting streamflow was carried out. As discussed previously, snowmelt over a forested watershed can be modeled using only temperature and precipitation data. This relationship holds as long as the watershed has fairly uniform vegetation type and cover and is subdivided to account for topographic variation. Since this is the case for the Salmon River, a modeling approach was sought that would incorporate such an estimation of melt within a suitable model structure for forecasting streamflow.

One approach to modeling discharge using an estimate of melt is time series analysis (TSA) or more specifically, transfer function noise (TFN) models. The structure of TFN models is readily adaptable to real time hydrologic applications (Watt and Nozdryn-Plotnicki, 1982). Updating a TFN model is comparatively straightforward and the choice of a computational time step is relatively flexible.

More importantly, these models can be developed and calibrated using relatively short data sets. For instance, W. E. Watt has developed a discharge forecasting model incorporating a TFN component for the Grand River near Marsville, Ontario, Canada (Watt and Nozdryn-Plotnicki, 1982). This model predicted spring discharge fairly accurately after being calibrated with only one spring's worth of data. The results of this model were compared with the NWSRFS model, which was calibrated with 5 years worth of data. The performances of the two models were comparable.

Before examining what a melt index model is and how TSA-TFN models are used to forecast streamflow, it is necessary to first digress a bit and examine the basic theories of TSA.

2.2.1 Time Series Analysis

A time series is a sequence of observations of a random variable. The sequence of occurrence is preserved in the time series and plays an important role in time series analysis. A sequential record of hourly, daily, monthly, or yearly discharge values constitutes a time series. A time series made up of streamflows is statistically dependent; the currently observed value of discharge is dependent on the previous runoff value, i.e., high flows tend to follow high flows and low flows tend to follow low flows.

In much of traditional statistical methodology, the dependence between observations is regarded as a nuisance and efforts are made to make sure that the observations are independent. In time series analysis (Box and Jenkins, 1976) the nature of this dependence is of interest in itself and the dependence pattern is analyzed in order to build a time series model; in essence, the data is allowed to "speak for itself."

Another important feature of time series analysis is that time series models incorporate a stochastic error term in addition to a deterministic function. Thus, when operating such a streamflow forecasting model in real time, the forecast error can be computed as soon as the observed flow is transmitted. This discrepancy between forecasted and telemetered flow (the forecast error) is then incorporated in the time series model to more accurately predict the next-step-ahead forecast value.

Depending on the number of time series involved, time series models may be divided into two categories; univariate time series models (UTS) and transfer function noise models (TFN). Univariate time series modeling is concerned with only one time series. The basic premise behind UTS is that the pattern of flow dependence will repeat in the future. Accordingly, the autocorrelation structure of the time series during the calibration period is analyzed to build the univariate autoregressive moving average (ARMA) model. In streamflow forecasting, ARMA models are rarely used because the driving force of streamflow is not accounted for. Runoff responds to such causal factors as rainfall, snowmelt, or subsequent soil moisture increase. In ARMA modeling of streamflow, these causal factors are not taken into account and the result is either a consistent under or over-estimation of the hydrograph.

TFN models have found practical applications in hydrology and water resources because TFN models represent the runoff (output) time series as a function of one or more causally-related input time series. TFN models have two components; a deterministic transfer function and a stochastic noise model. The transfer function represents the dynamic input/output relationships. Its form is derived by analyzing the cross correlation structure of the causally related time series. The noise model lumps together all the other causal factors which may influence streamflow but which may not have been included in the transfer function.

The most important part of TFN modeling is the identification of the relevant input time series. Precipitation measurements alone are not satisfactory because they constitute a discrete time series (there is the possibility of having values of zero precipitation).

Since only three months of data were available in the Salmon River Watershed, the calibration of the coefficients in a traditional snowmelt model or an API model was considered unmanageable. However, it was felt that the lack of a sufficient historical data record for calibration could be compensated for by using available hydrologic and meteorologic data and time series models. To this end, a snowmelt index model was formulated for defining an input time series for TFN model development. In the following sections the application of the snowmelt index-time series analysis approach to forecasting streamflow on the Salmon River will be addressed.

3.0 Snowmelt Index Model

3.1 Definition

When a snow cover is present, fluctuations in streamflow are a result of changes in the melting rate of the snow cover and direct input from rain. As mentioned earlier, melt occurs when the energy entering the snowpack is significant enough to offset losses and the temperature of the snowpack rises above 0°C. Air temperature and the amount of precipitation are good indicators of the energy balance of the snowpack and of the consequent rate and magnitude of melting and corresponding streamflow (Quick and Pipes, 1976).

In order to forecast streamflow during periods of snow cover it is necessary to consider the amount of actual and potential melting (including precipitation) as the "driving force." If a direct measure of melt could be obtained and effectively extrapolated to the entire watershed, the forecasting of streamflow would be greatly simplified. Since in many cases this is not a practical option, snowmelt models have been developed that attempt to estimate the amount of melt from various hydro-meteorological variables. In all cases there is some error in the estimate of melt and the resulting estimate of stream discharge. Even sophisticated energy-mass balance models can not account for all of the variables affecting melt and streamflow (Colbeck et al., 1979).

From an operational standpoint, for real time forecasting of streamflow (via transfer function noise (TFN) models) during winter and spring months, some method is necessary to combine the controlling factors of melt (and consequently, streamflow) into a suitable "driving force" time series. The use of air temperature data alone would not account for rain on snow events. The use of precipitation alone is not practical because, obviously, without temperature, the form of the precipitation is unknown. Thus, an operationally useful method was sought that would combine the controlling factors of melt into a continuous time series for transfer function noise modeling.

Because TFN modeling accounts for both the deterministic, linear relationships between some input series and an output series (the transfer function) and the uncertainties, or stochastic randomness of the process (the noise component), an exact measure of melt is not necessary for forecasting flow. The important aspect is the nature or structure of the input time series. It must reflect the impact on melting that changes in the significant controlling factors would have. In other words, the input time series should express peaks that are of the same relative magnitude as the output series and preferably, for forecasting, occur prior to those of the output series (see fig. 2).

Any of the above snow models could have been modified and used as the basis for formulating a model to generate daily estimates of snowpack melting. Martinec's formulation of a degree-day snowmelt model (1975) was chosen because of its wide use, fairly good results, simple structure, modest data requirements, adaptability to real time forecasting and its inclusion of the important dynamic variables controlling streamflow under uniform conditions.

In this case the output of the model is refered to as a melt index rather than an actual estimate of melt because of the uncertainties inherent in an empirical model and in extrapolating point measurements of hydro-meteorological variables to basin wide estimates. Because TFN modeling methods examine the correlation between an input series and an output series, it is not necessary to know the exact amount of melt, at a point or for the whole basin, in order to forecast streamflow. As long as there is good correlation between the two time series, the exact nature of the input series is not a significant consideration.

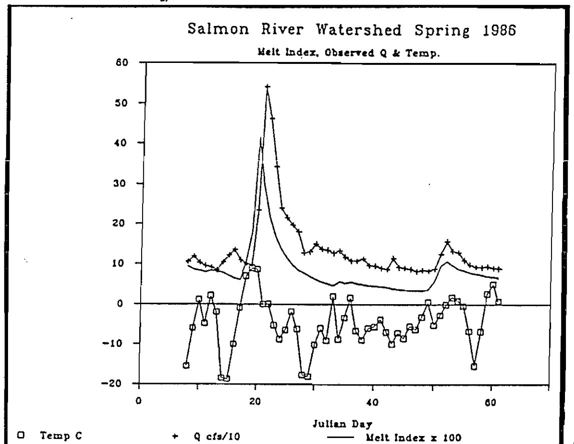

Figure 2

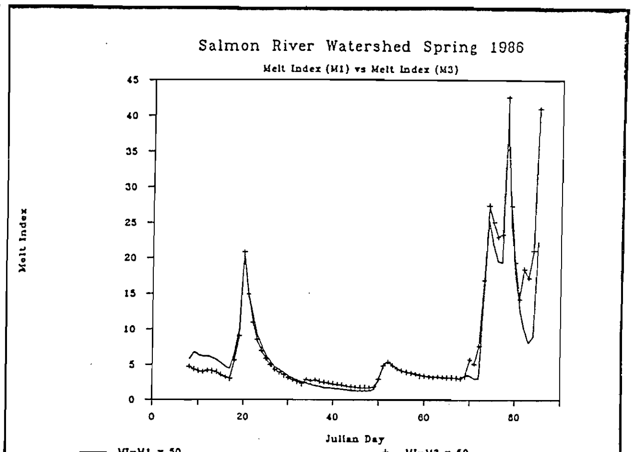

Figure 3

3.2 Snowmelt Index Model Structure and Coefficient Calibration

The formulation of the snowmelt index model (SMIM) is similar to that of Martinec's degree-day melt model (1975):

$$MI_t = ((A * T_t) + PPT_t) * (1 - K_t) + (MI_{t-1} * K_t)$$

where, "MI" is the melt index, "A" is the melt factor (in/C° per day), "T" is the number of degrees above 0 C° per day (values less than 0 are treated as 0) PPT is the amount of precipitation (inches) that falls as rain (T>0 °C) and "K" is a lag coefficient.

The melt factor A is calculated based on work down by Martinec (1975) as .433 * D, where "D" is the average density of the basin's snowpack measured twice a week during the spring season.

The lag factor K is also calculated in the same manner as Martinec's lag factor using a regression equation of the form

$$K_t = a*Q_t^b$$

where Q_t is the current discharge and a and b are regression constants derived from analyses of recession flow events. The lag coefficient behaves such that large Q's result in K values much less than unity and low values of Q result in K values close to unity. Thus, in the above calculation of the melt index, large flows result in a lagging of the melt index and low flows result in the stabilizing of the melt index around some non zero base value. The lag factor, K, is included to account for the lagging and routing of the watershed's snowpack and drainage system.

3.3 Data Requirements

The data required to calculate the melt index include air temperature, precipitation, snow density and streamflow. A complete data set was available for approximately the first three months of 1986 (Julian days 8-85). Air temperature and precipitation from the hydro-met site M3, together with discharge from stream gaging site S2, constituted this database (see fig. 1 for locations of sites). The 4 hour values of temperature and precipitaion are transformed into 24hr averages and totals respectively. Measurements of snow depth, density and water equivalent taken at the hydro-met sites complete the necessary data. From tests of the melt index model using the same temperature record and significantly different precipitation records, it was noted that the dominant factor controlling the magnitude of the melt index was air temperature (see fig. 2). It was also noted that there was little difference in melt index estimates using temperature records from a different site, M1 (see fig. 3). This indicated that using the raw temperature and precipitation records from site M3 without computing basin averages would not have a significant affect on the melt index time series.

4.0 Application of Snowmelt Index and TFN Models to Forecasting Streamflow

4.1 Development of TFN Model

A statistical package prepared by Automatic Forecasting Systems, Inc. was used to develop a TFN model relating the melt index time series to daily discharge. The software is designed to be run on a personal computer (PC) and offers a complete array of Box-Jenkins modeling techniques. The Box-Jenkins approach to time series analysis allows the data to "speak for itself" rather than relying on pre-defined assumptions. This results in the development of the most appropriate model for describing and forecasting the time series.

Output from the melt index model for the spring of 1986 was divided into two segments. The first, Julian days 8 thru 58, was used to calibrate the TFN model and the second, Julian days 59 thru 85, was used to verify the TFN forecast equation. Calibration involved using the TSA software package to analyze the autocorrelations of the input and output series, identify and apply the appropriate pre-whitening filters, identify the TFN model form, estimate the model's parameters and check the final TFN model for statistical adequacy.

The final TFN model for the calibration period is as follows:

$$Q_t = 5.0 + 0.742Q_{t-1} + 1161.2MI_{t-1} - 615.7MI_{t-2} - 182.5MI_{t-3} + a_t$$

The current discharge, Q_t (CFS) is a function of portions of past values of discharge and melt index. Fortunately, for forecasting purposes, there is a time lag between input and output (see fig. 2). However, the form of the above equation limits the forecast timestep to 1 (24hrs):

$$Q_t(1) = 5.0 + 0.742Q_t + 1161.2MI_t - 615.7MI_t - 182.5MI_{t-1}$$

Forecast time steps greater than 1 would require forecasts of melt index. For example, the two-step-ahead forecast would have the following form:

$$Q_t(2) = 5.0 + 0.742Q_{t+1} + 1161.2MI_{t+1} - 615.7MI_{t-1} - 182.5MI_{t-2}$$

From this equation it is also evident that forecasts of discharge would be required. This is not an untenable condition, as forecasts of discharge (Q_{t+1}) can be generated from the one-step-ahead forecast equation. Obviously, though, inclusion of a forecasted variable in a forecast equation would result in a widening of the confidence limits of the final forecast. The extent of uncertainty would depend on the effectiveness of the input time series in capturing the driving forces of the output series and on the forecasting accuracy of the TFN model.

However, in order to provide forecasts of the melt index an additional univariate or TFN model must be developed. The confidence limits of the forecasted discharge would again be widened. The amount of widening would depend, in part, on the accuracy of the melt index forecast. As we have seen, univariate forecasts are usually unsuitable for modeling hydrologic time series because they fail to account for the driving force. Thus, including forecasts of melt index in the above two-step-ahead (or greater) discharge forecast equation would result in an undesireable widening of the forecast confidence limits.

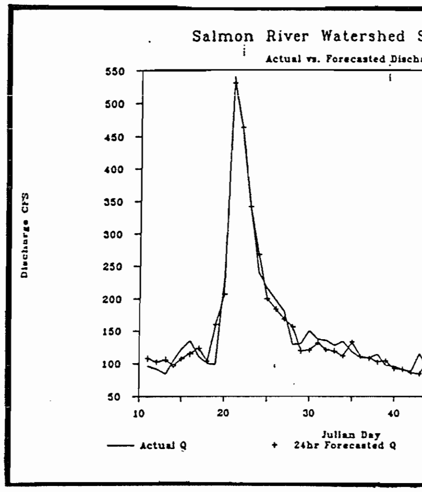

Figure 4

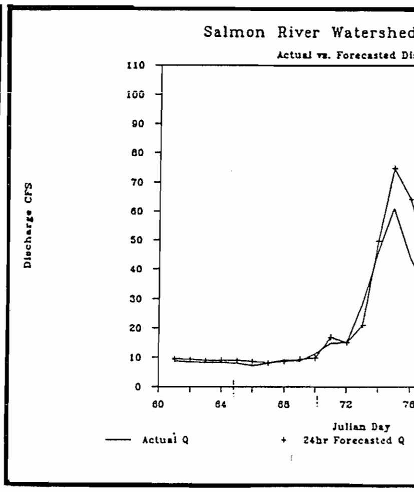

Figure 5

From figure 4 we can see the final fit of the calibrated model. In general, there is good agreement between the actual and forecasted discharge, particularly with regards to the model's forecasting of the hydrograph peak. The accuarcy of the model is reflected in the root mean square error for the calibration period, which was 17.48 or 12.3% of the output series' mean.

4.2 Performance of TFN Model

At first glance the performance of the calibrated TFN model during the forecast period (Julian days 59-85) appears to be unsatisfactory (see fig. 5). However, given the limited calibration period (50 days with only one significant melt event) the TFN model does a fairly good job of approximating the complex nature of the hydrograph using the melt index as an input driving force time series. It is important to note the excellent correspondence between the timing of the observed and forecasted discharge hydrograph peaks. With a more extensive calibration period, one including more complex hydrographs, the performance of the TFN model should improve.

4.3 Discussion

The results were compared with two other techniques which could have been used to forecast streamflow. The first comparison was between a univariate model using only past values of streamflow and a TFN model. Figure 6 shows the comparison. As expected, the univariate model, which was identified as an AR(2) model, forecasts the hydrograph well in magnitude but at the expense of a consistent timing error. The forecast is repeatedly one-day off. This is attributed to the omission of the driving force in the model development. The second comparison was between a lagged regression model and the TFN model. To find out the independent variables in the regression equation, a stepwise regression was used. In the first starting model, Q_{t-1}, Q_{t-2}, Q_{t-3}, MI_{t-1},

MI_{t-2} and MI_{t-3} were included. The stepwise regression selected only Q_{t-1}, MI_{t-1}, MI_{t-2} and MI_{t-3} as important variables in explaining the variance of Q_t. The final regression coefficients were close to the TFN parameters.

Lagged Multiple Regression Model:

$$Q_t = 7.38 + 0.725Q_{t-1} + 1156.0MI_{t-1} - 588.6MI_{t-2} - 202.8MI_{t-3} + \dot{e}$$

TFN Model:

$$Q_t = 5.0 + 0.742Q_{t-1} + 1161.2MI_{t-1} - 615.7MI_{t-2} - 182.5MI_{t-3} + a_t$$

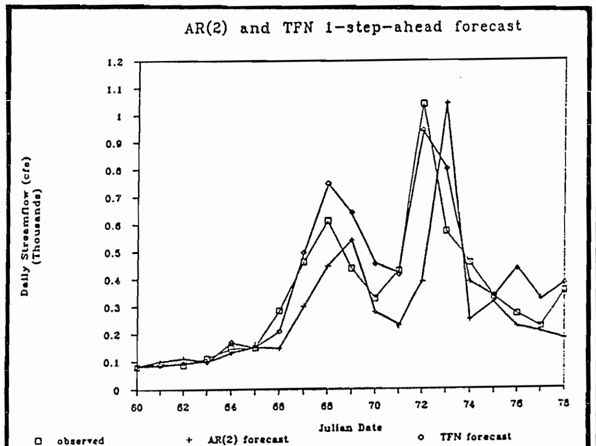

Figure 6

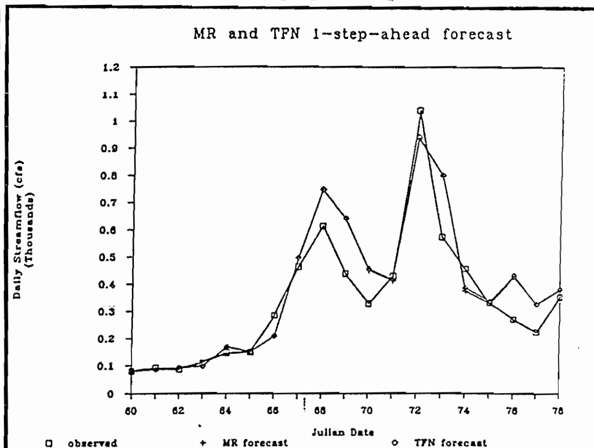

Figure 7

Figure 7 shows the graphical comparison between multiple regression and TFN forecasts. For all practical purposes they are identical. What then are the advantages of using a TFN model for streamflow forecasting instead of a multiple regression model? The arguments favoring TFN modeling over multiple regression modeling are as follows.

First, the assumptions about the error terms should be compared. In multiple regression the error term, e, is supposed to be uncorrelated with previous error. The violation of this assumption will result in the instability of the regression coefficients (Granger and Newbold, 1970). The practical result of this violation leads to the observation that adding or deleting a couple of observations will change the regression coefficients considerably. In TFN modeling, the error term, a_t, is random white noise. Consequently, the parameters of the TFN model are stable.

Second, multiple regression analysis takes more computing time because the modeler does not know how many independent terms should be included in the starting model. Normally, the modeler includes as many terms (lagged terms) as possible. In contrast, TFN modeling examines the dependence structure between input and output time series and the final forecast equation is derived automatically from the structures of the series.

Third, the TFN model can accommodate the forecast error from the previous step and self-correct the forecast in the next-step-ahead forecast. This feature is not found in multiple regression. For example, even if the multiple regression forecast consecutively underestimates the real time flow, there is no convenient way, other than to replace the forecast equation, to modify the forecast model. In TFN modeling, previous errors are usually incorporated in the next-step forecast. In general this promises to provide a more accurate forecast. Unfortunately, this point is not clear in the above two forecast equations because a previous error term $(a_{t-1} \ldots a_{t-n})$ is not included in the TFN one-step-ahead forecast equation. This is due to the nature of the time series being modeled and the resultant structure of the TFN model. The form of the TFN one-step-ahead forecast equation does include the current forecast error (a_t); however, since the a_t series is assumed to be a random white noise process, the best guess of the current and future expected values of the error series would be 0. If the model structure included past errors, then the discrepency between previously forecasted and observed streamflow values (usually not 0) would be included in the current forecast.

5.0 Conclusion

5.1 Advantages of Snowmelt Index Approach

The primary advantage of using a snowmelt index model to generate an input time series for TFN modeling of discharge is the ease with which the necessary parameters can be attained and extrapolated. The spatial and temporal distribution of temperature over a watershed is normally relatively uniform. Although it is not as easy to extrapolate point measurements of precipitation to basin wide values, the gaging density of the Salmon River watershed is high enough to ensure a fairly good estimate of the net basin precipitation.

Another advantage is that since the melt model is providing only an index to the cumulative melting potential of the significant hydro-met parameters, the calibration process is limited to deriving the lag coefficient K, which defines the recession characteristics of the snow covered basin, based on the structure of recession events. In Martinec's snowmelt model (1975) a basin runoff factor must also be calibrated. In more complex snowmelt-streamflow models there is the problem of calibrating the runoff generating component of the model. By using TFN modeling techniques, this factor is accounted for in the formulation of the transfer function. An important consideration and an aspect unique to TFN models is that they readily provide state updating and error correction capabilities. The uncertainties of any estimate of the physical processes governing the production of melt and streamflow are accounted for in the stochastic component of the formulation.

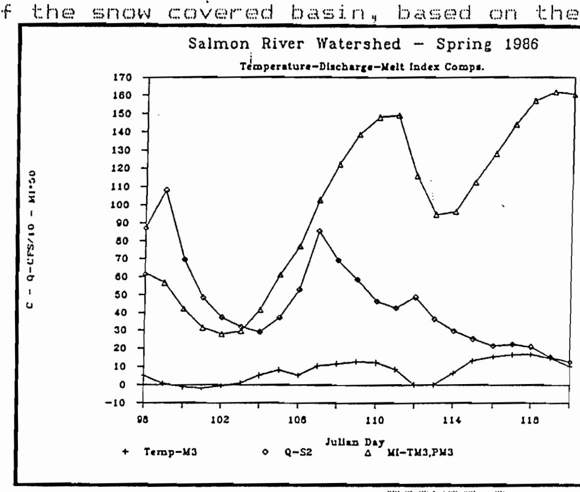

FIGURE 8

5.2 Limitations

The major limitation of using the snowmelt index-transfer function noise approach to forecast streamflow is the lack of a direct method to measure the areal extent of the snow cover. As can be seen from figure 8, the correlation between the melt index time series and streamflow breaks down around the middle of April. The melt index, driven by the rise in temperature, indicates a significant amount of melt potential even though the streamflow hydrograph is actualy in a recession stage. This kind of anomolous response can be attributed to the loss of the snow cover. Thus, deciding when the snow cover is gone and the TFN model relating the melt index to discharge is no longer applicable, is a problem.

This problem could be solved by keeping track of the progressive melting of the snow cover and its eventual elimination. To this end, auto-sensing instrumentation that monitors the depth of the snow cover was installed. Unfortunately, it was not operational for the spring of 1986.

Another limitation is the forecast interval. Presently, using the current snowmelt index model, forecasts of streamflow are limited to 24 hours in the future. One way to handle this kind of limitation would be to make use of quantitative precipitation forecasts. To date these are not available, but the feasability of this approach is being considered. Nevertheless, given the small size of the watershed, 24 hour ahead forecasts based on currently available knowlege of the hydro-met conditions in the basin appear quite satisfactory.

Acknowlegements

Funding for this research was provided by Niagara Mohawk Power Corporation, Syracuse, New York. In particular, we would like to express our appreciation for the support and encouragement provided by John Stojka, Richard Laycock, and Charles H. Saylor.

6.0 References

Anderson, E. A., "National Weather Service River Forecasting System--Snow Accumulation and Ablation Model,"NOAA Technical Memorandum NWHydro-17, U.S. Dept. of Commerce, Silver Spring, Md., 217 pp., 1973.

Anderson, E. A., "A Point Energy and Mass Balance Model of a Snow Cover,"NOAA Technical Report NWS 19, 150 pp., U.S. Dep. of Commer., SilverSpring, Md., 1976.

Anderson, E. A., "Streamflow Simulation Models for Use on Snow Watersheds," in CRREL Proceedings, Modeling of Snow Cover Runoff, U.S. Army CRREL, Hanover, NH, pp. 336-350, 1979.

Archer, D. R., "Computer Modeling of Snowmelt Flood Runoff in North-East England," in Proceedings of the Institution of Civil Engineers, Part 2, 75, pp. 155-173, 1983.

Box, G. E. P. and G. M. Jenkins. Time Series Analysis Forecasting and Control. 2nd. ed. San Francisco: Holden-Day, 1970.

Chow, K. C. A., W. E. Watt and D. G. Watts. "A Stochastic-Dynamic Model for Real Time Flood Forecasting," Water Resources Research. Vol. 19. No. 3, pp. 746-752 1983.

Colbeck, S. C., E. A. Anderson, V. C. Bissel, A. G. Crook, D. H. Male, C. W. Slaughter and D. R. Wiesnet, "Snow Accumulation, Distribution, Melt and Runoff," EOS, 60(21), pp. 464-468, 1979.

Granger, G. W. J. and P. Newbold. "Spurious Regression in Econometrics," Journal of Econometrics Vol. 2, pp. 111-120, 1970.

Hawley, M. E., R. H. McCuen, A. Rango, "Comparison of Models for Forecasting Snowmelt Runoff Volumes," Water Resources Bulletin, 16(5), pp. 914-920, 1980.

Martinec, J., "Snowmelt - Runoff Model for Stream Flow Forecasts," Nordic Hydrology, vol. 6, pp. 145-154, 1975.

Quick, M. C., A. Pipes, "A Combined Snowmelt and Rainfall Runoff Model," Can. J. Civ. Eng., 3(2), pp. 449-460, 1976.

Watt, W. E. and M. J. Nozdryn-Plotnicki. "Real-time Flood Forecasting for Flood Damage Control," In Proceedings of International Symposium on Real-time Operation of Hydrosystems. Vol. II, pp. 741-760. ed. by T. E. Unny and E. A. McBean, University of Waterloo, Ontario, 1981.