APPLICATION OF PRINCIPAL COMPONENTS REGRESSION TO STREAMFLOW FORECASTING

by Stephanie Smith[1] and Eric Weiss[2]

ABSTRACT

B.C. Hydro adapted the techniques of principal component regression, cross-validation and a systematic search for optimal combinations of variables (Garen, 1993), to try to improve to the statistical accuracy of streamflow volume forecast equations for the glaciated Bridge River basin in British Columbia. The paper will discuss the trade-offs that B.C. Hydro made in balancing the improvements in forecast accuracy using these techniques against other factors, such as preserving continuity in forecast equation structure from month to month, preferring short-term stations located within the basin to longer term stations outside the basin, and using the thirty-year normals period (1961-90) versus the historical period of record. Results for the Bridge River system demonstrate the balance that can be achieved between optimizing forecast accuracy and selecting variables and stations which make the forecast equations more physically rational.

INTRODUCTION

Regression based and Extended Streamflow Prediction (ESP) methods are used by BC Hydro to forecast seasonal water supply to hydro electric reservoirs. There are advantages and disadvantages to each method of forecasting which have been described by the Columbia River Water Management Group Forecast Committee in their report to the Northwest Power Pool (CRWMGFC, 1993). This paper focuses on statistically based forecast procedures and does not discuss water supply forecasts produced by conceptual watershed models.

BC Hydro issues seasonal water supply forecasts once each month beginning January 1 and ending August 1. The forecasts assist operations and planning personnel in making decisions regarding the medium- to long-range operation of the reservoirs. Mid-month forecasts are not normally required as monthly forecasts satisfy most planning and operation requirements.

Seasonal volume runoff forecasts for the period February through September are generally issued. By issuing the forecast for a common period, variations in the forecasts from month to month can be easily identified. The February through September forecast is computed by adding the residual volume runoff from the forecast date to the end of September to the observed inflow data from February to the forecast date.

This paper discusses the rationale for changes BC Hydro has made from past year in developing statistically based water supply forecast procedures. The Bridge River hydro projects located in the Fraser River basin are used to demonstrate the current development procedures. A comparison of statistics for the equations derived using the new and old procedures is presented.

Types of modeling errors

Three types of errors can occur when applying statistically-based forecast model (CRWMGFC, 1993) as shown in Table 1. The most significant contribution to Type 1 errors for most basins is precipitation that occurs during the forecast period. Another possible contribution is glacier melt for glaciated basins. Any model is only an approximation of the real world, and will be subject to greater or lesser degrees to Type 3 errors. These errors generally result from sampling error and inappropriate selection of model variables.

TABLE 1 Types of modeling errors

Type 1	Due to significant contributors to the seasonal runoff are not known at the time of the forecast
Type 2	The stations used to compute a particular variable are not representative of the variable over the entire basin
Type 3	The model selected is in error

Statistically based equations - past practice at BC Hydro

Following suit with many organizations over the years, BC Hydro used multiple regression analysis to develop water supply forecast equations for its hydro electric reservoirs in the past.

1 Hydrometeorologist, Power Supply Operations, BC Hydro

2 Senior Hydrology Engineer, Power Supply Operations, BC Hydro

Seasonal runoff in much of British Columbia is dependent on snow melt. Therefore, snowpack accumulation tends to be a significant variable in these regression based models. The ability of regression models to accurately forecast seasonal water supply is greatest for basins where the snowmelt freshet is a large proportion of the total seasonal runoff. The statistical significance of regression based equations is reduced for basins that receive much of its runoff from summer rains or glacier melt.

Because snowmelt makes up a significant proportion of much of the runoff in many basins in British Columbia, it is not surprising that indices of snowpack accumulation during the winter tend to be highly correlated to streamflow. In fact, the accuracy of the forecasts improve during the winter as the snowpack builds to its maximum for the year, generally in April or May. Other indices that account for evaporation losses, soil moisture prior to freeze up, and glacier melt during the spring and summer months generally tend to be less significant than the index used to measure winter snowpack.

BC Hydro standardized its forecast development procedure over the years to ensure that the equations being developed adhered to some physical rationale.

It developed regression equations that could use up to 5 pre-defined indices.

X1 - Accumulated snowpack
Accumulated winter precipitation at valley bottom climate stations and snow water equivalent data at high elevation snow courses were combined into one variable. Precipitation was accumulated from November to the forecast date, or November to March inclusive for forecasts made on or after April 1. Accumulated winter precipitation and snow water equivalent were combined because they have a tendency to be intercorrelated. Regression analysis assumes that all independent variables used in a model are independent of one another.

X2 - Spring and summer precipitation
Accumulated spring and summer precipitation using known data only were used in late season forecasts as an index of rainfall runoff. This index is computed by accumulated precipitation from April to the forecast date. This index tends to be more important for those basins that have rainfall runoff comprising a high proportion of the total runoff over the forecast season.

X3 - Antecedent soil moisture
Antecedent soil moisture prior to freeze up was represented by an index of accumulated precipitation for the previous September and October. The index is positively correlated with seasonal runoff, meaning that for a given snowpack, more runoff can be expected following a wet fall than a dry fall.

X4 - Evapotranspiration
Losses due to evapotranspiration can occur from the snow cover or bare ground during the forecast period. The mean of maximum daily temperatures for the period February through the forecast date were used as an indication of evapotranspiration. Higher temperatures will have the effect of lowering the expected seasonal volume runoff.

X5 - Glacier melt
The mean of maximum daily temperature for the period April through the forecast date were used as an index of glacier melt. Higher temperatures will have the effect of raising the expected seasonal volume runoff. The use of this index is restricted to glaciated basins.

Elaborate trial-and-error procedures were used to weight both station values and monthly values used to construct the indices. The ultimate weightings used were those that produced the highest correlation coefficients for the regression equation. The equations were also updated once every few years. The updating procedure was generally not as rigorous as the initial development. Rather, the independent variables and the stations used to comprise those variables were not changed. The coefficients of the variables however were re-computed using additional data that had been collected since the last update.

Current review of BC Hydro's forecast procedures

There are a number of reasons why BC Hydro has undertaken a review of it existing water supply forecast equations. First, BC Hydro began operating a more dense and elaborate automated station network in the early 1980s. More recently, a good number of climate stations have been decommissioned due to cutbacks in the federal government's budgets. In some basins like the Bridge River, the data collection network has been sparse and many of the stations used in existing procedures lie outside of the basin. BC Hydro wanted to include more of its own stations and reduce the dependency on data collected by the federal government. We hoped that by including the new stations located within the basin, Type 2 errors would be reduced.

Recent studies by Garen (1992) and Koch and Buller (1993) heralded the use of principal component analysis (PCA) as a better tool than multiple regression analysis to use in the development of water supply forecasting procedures. The statistics for equations based on PCA showed significant improvements in forecasting accuracy over equations developed using multiple regression analysis.

Finally, the operation of reservoirs for the purpose of power production has been increasingly constrained by environmental considerations. This is particularly true with respect to fisheries issues, and flooding issues along the shores of reservoirs and downstream of dams. BC Hydro is committed to environmental stewardship, and would be better able to meet both power and environment objectives if forecasts of future water supply were more accurate and used the denser climate network now available.

With the objective of considering a greater proportion of BC Hydro stations in the forecast procedures and improving the accuracy of the water supply forecasting procedures, BC Hydro began developing new forecast equations for the Bridge River projects. The procedure used to develop these new equations was a hybrid of procedures BC Hydro used in the past, and procedures recently finding favor among practitioners.

PHILOSOPHY BEHIND A REVISED METHODOLOGY

The CRWMGFC report describes a host of procedures now being practiced by various agencies in the Western United States and Canada to develop statistically-based water supply forecasting procedures. The following sections describe some of the decisions made and philosophy adopted by BC Hydro in revising our methodology of developing statistically-based seasonal water supply forecasts..

Data preparation

The first task was to collect all available data that could be potentially used in the new forecast procedures. Both BC Hydro and Environment Canada climate data were used in the analysis. We decided to use BC Hydro climate stations in preference to Environment Canada stations provided significant degradation in the forecast equation statistics didn't result.

Previous files maintained by BC Hydro contained both observed plus estimated data. Unfortunately, estimates had not been flagged. These estimates were generally made using simple or multiple inter-station regression analysis. Hirsch (1982) notes that while estimates made using regression preserve the mean of a data set, the variance will be reduced. If a significant number of data are estimated, the data set used to develop the equations will underestimate the variance of the real time data, which could produce real time forecast residuals inconsistent with the those of the developed equations.

Before estimates of missing data were made however, data record consistency was checked using double mass curve analysis (Smith and Weiss; 1996). Historical records were adjusted as required to be consistent with catch characteristics of the station currently being operated. To ensure that both the mean and the variance of the estimated data were preserved, MOVE 1 and MOVE 2 organic regressions were used to estimate missing monthly precipitation and snow water equivalent data (Smith and Weiss; 1996).

Additional data types

Koch and Buller (1993) shows some success in using the Southern Oscillation Index (SOI) in providing forecasts of seasonal water supply. We attempted to see if SOI would improve our ability to forecast seasonal water supply, particularly for early season forecasts.

Brugman and Pietroniro (1995) showed that glacier melt in the Illecillewaet River basin in the Columbia was highly correlated with short-wave radiation data collected on site. Until now, BC Hydro has used mean monthly maximum temperatures as a proxy for short-wave radiation data. We attempted to use hours of sunshine during the month for some stations as alternative proxy for short-wave radiation.

Principal component analysis

Based on the success of Garen and Koch, principal component analysis was selected in preference to multiple regression analysis for developing the water supply forecasting equations. A copy of Garen's optimal search program was obtained and used to identify the mostly statistically optimal results. However, we chose to conduct our analysis more along the traditional methods used in the past than along the procedures used by Garen and Koch.

We were uncomfortable with Garen and Koch's approach of using individual station months individually as independent variables in the PCA. It is possible for equations developed to use precipitation for a given station in December and February, but not January or March. It is true that statistically, these equations could be "optimal" based on the sample statistics. But the optimality is dependent on the sample of data used. We think it unlikely that the same stations for the same months would be selected when additional years of data are collected and analyzed. In our opinion, the statistically "optimal" equations developed are in part dependent on the sample used to develop the equation.

One main argument for not predefining accumulated variables but keeping individual station months of precipitation is that the period selected over which to accumulate precipitation is arbitrary. PCA should be allowed to select the optimal combination of winter precipitation. However, we would argue that even using monthly data is arbitrary. It so happens that precipitation is neatly packaged and made available in monthly format. But why not two week accumulations, or even daily?

It would be a particularly tough sell to someone who is directly responsible for the operation of the reservoir if you were to tell them that precipitation in January is not statistically significant in the equation you chose. What do you do with a forecast that calls for average water supply because December and February precipitation was near normal, but January saw a record event. We suspect that the potential for introducing Type 3 errors into the forecasting procedures are increased by allowing individual station months stand on their own in the development of the equations. We believe that an optimal equation is one that uses input variables that follow rational physical principals that are understood and accepted by the user of the forecast. Once these physically rational variables are designed, only then do you try to optimize the equations. Accumulating precipitation over an extended winter period follows the principal that the maximum accumulated winter snowpack is physically linked with future runoff.

For this reason, we have chosen to accumulate precipitation beginning arbitrarily on November 1 each year, to represent an index of the snowpack that we are expecting to melt during the forecast period. We have also abandoned our own past practice of weighting individual months differently to try to minimize the standard error of the forecast. There may be some physical rationale for doing so, that being there may be a greater chance for snow that fell in November to sublimate or melt prior to the forecast season than snow that fell later in the season. However, the weightings applied in the past have not always been such that early season precipitation is weighted less than later season precipitation. Again, random weightings of months will likely result because of sampling error, rather than truly being random.

Cross validation standard errors

BC Hydro has used the standard error of estimate of the forecast equations in the past to express uncertainty in the forecast. We have observed that this standard error tends to underestimate the error of the forecasts made in real time. Based on Garen's work and others, we have decided therefore to adopt the cross validation, or jack knife, standard error (CVSE) to more realistically describe the uncertainty in the forecast.

Using only known variables at the time of forecast

BC Hydro's past practice, and that endorsed by Garen and Koch is to include only known input at the time of the forecast. We agree that including future variables, such as summer precipitation in an April 1 forecast, doesn't reduce the standard error of the April 1 forecast. Therefore, our operational forecasts do not include future variables.

We do, however, endorse the practice of adding a future precipitation variable in the development of equations to determine what types of errors the equations are subject to. For example, if an April 1 forecast equation has a relatively low correlation coefficient, which improves substantially with the inclusion of a future variable, that gives the forecaster an indication that Type 1 and not Type 2 errors exist. The forecaster should probably not invest substantially in the establishment of additional stations in the basin in the hope of getting more representative samples. On the other hand, if an April 1 forecast equation has a relatively low correlation coefficient, which does not improve substantially with the inclusion of a future variable, the forecaster might suspect a Type 2 error, and establish new sites in an attempt to get more representative data to base forecasts on.

Month to month consistency in forecast

BC Hydro in the past has generally optimized the forecast equations for a given month based on the standard error of estimate for that month. If several equations gave similar statistics, equations for different forecast dates that looked similar to one another were chosen. However, reducing the standard error of estimate was the prime consideration in selecting operational equations.

From the standpoint of credibility with the reservoir operator, a change in a forecasts from month to the next should reflect changes in the meteorological conditions observed in the intervening month. It should not reflect the fact that different independent variables are being uses.

Planning and operation of reservoirs rely quite heavily on the results of the April 1 forecast. The standard error of the April 1 forecast is generally less than that of previous forecasts because of the reduction in Type 1 errors. We therefore have decided that we would develop equations for April 1 first. Then we would develop equations for January 1, February 1 and March 1 using where possible the same variable used in the April 1 forecast. For example, in the extremely simple case of a seasonal runoff forecast procedure relying solely on one snow course on April 1, the equation for March 1 would use the March 1 observation for that same snow course. In the case of accumulated winter precipitation variables, the same stations used to compute the April 1 variable would be used in the March 1 forecast. The only difference in the variable would be that for the April 1 forecast, the accumulation period would be from November to March inclusive, whereas, the accumulation period for the March 1 equations would be from November to February inclusive. Keeping the same stations to define the input variables for the forecast equations during the snowpack accumulation phase of the winter assures that a change in forecast from one month to the next results only from deviations from normal in weather during the intervening month.

Basing equations on the normals period

Past practice at BC Hydro was to use as much historical data for a given watershed as possible to develop equations, the logic being that increasing the number of sample points reduces the confidence interval about a forecast. However, differences in percentage forecasts between neighbouring watersheds were sometimes observed to be only the result of different periods of record used to develop the equations for each basin.

Almost all forecasts disseminated by agencies in the Pacific Northwest express forecasts in terms of percent of the 30-year normal. Currently, the normals period is defined by the WMO as 1961 to 1990. If the mean seasonal runoff for the historical record used to develop the equation is much different than that of the normal period, some problems can occur in presenting percent of normal forecasts. For example, if you input normal values for all value of the independent variables, a forecast other than normal inflow can result.

To ensure that seasonal runoff forecast will be 100% of normal when normal independent variables are input to the equation, all equations are developed based on the 1961 through 1990 historical record. This goes against the philosophy of using as many years as possible to reduce the width of the confidence interval. But the overall width of the confidence interval for one-tailed, alpha=95% will reduce by less than 2% when a 60-year is used to develop an equation, versus a 30-year record. (The figures from the Student's t table are 1.671 and 1.697 respectively). Therefore, it seems a small statistical price to pay to have equations that complement the presentation of forecast results in terms of percent of normal.

Forecasts after April 1

As suggested by the CRWMGFC report, alternate procedures were tested for forecasts after April 1 to account for changes that may occur to forecasting after the spring freshet begins or after it starts to recede. Starting with the May 1 forecast, additional variables were added to the principal components search, such as summer accumulated precipitation (from April 1), summer mean monthly maximum temperature, and monthly total sunshine hours. Methods other than principal components regression, such as residual and recession curve forecasts, were also tried for later in the season. The residual forecast takes the forecast from the previous month and subtracts the flow that has been observed for that month. Recession forecasts assume that the peak runoff has occurred and use a standard equation to distribute the remaining flow through the rest of the forecasting season. The equation we used to describe the recession limb of the seasonal volume runoff hydrograph was:

$$Q_2 = Q_1 k^{-\Delta t}, \qquad (1)$$

where Q_1 and Q_2 are the monthly volume at months t_1 and t_2 respectively, k is the recession constant, and Δt is the change in time from t_1 to t_2 (Gray, 1970). Use of these alternate procedures are meant to distribute the remaining expected flow over the rest of the season. As noted previously, we do not consider including any future variables, even though it tends to be future events (precipitation and glacier melt) that dominates the late season runoff.

Disaggregation of forecasts

BC Hydro has traditionally disaggregated the seasonal runoff forecast to monthly values required in other energy studies by prorating the seasonal forecast based on average monthly inflow values. Grygier et al. (1993) described a method for improving on this disaggregation method. By applying a statistical logit model to the seasonal forecast, the resulting monthly flows are guaranteed to be non-zero and less that the total forecast. In their study, Grygier et al. (1993) found the logit function to outperform both exponential and polynomial models. By using a combination of three explanatory variables, (total seasonal forecast, previous month's flow, and future monthly precipitation) they were able to achieve superior results with the logit model than by using a simple linear function to disaggregate the seasonal forecast into monthly flows. We found for Bridge that the inter-monthly correlations of inflow were significantly different than presented by Grygier et al. (1993), and that the disaggragation procedure produced no better results than the simple prorating procedure for the Bridge. However, the method will be tested again for other basins where higher inter-monthly correlations might be observed.

THE BRIDGE RIVER

BC Hydro operates two reservoirs on the Bridge River system that are used for hydro electric power production. Figure 1 shows a map of the project area. The Bridge River basin is located on the leeward side of the southern coastal mountains, approximately 200 km north-east of Vancouver. Predominantly westerly weather systems over the basin topography create a highly-transitional precipitation pattern across the basin, from 1800 mm in the west to 600mm in the east. The flow of the Bridge River has been regulated by the construction and operation of several dams, diversions and hydroelectric generating stations developed between 1927 and 1960.

FIGURE 1. Bridge River Basin

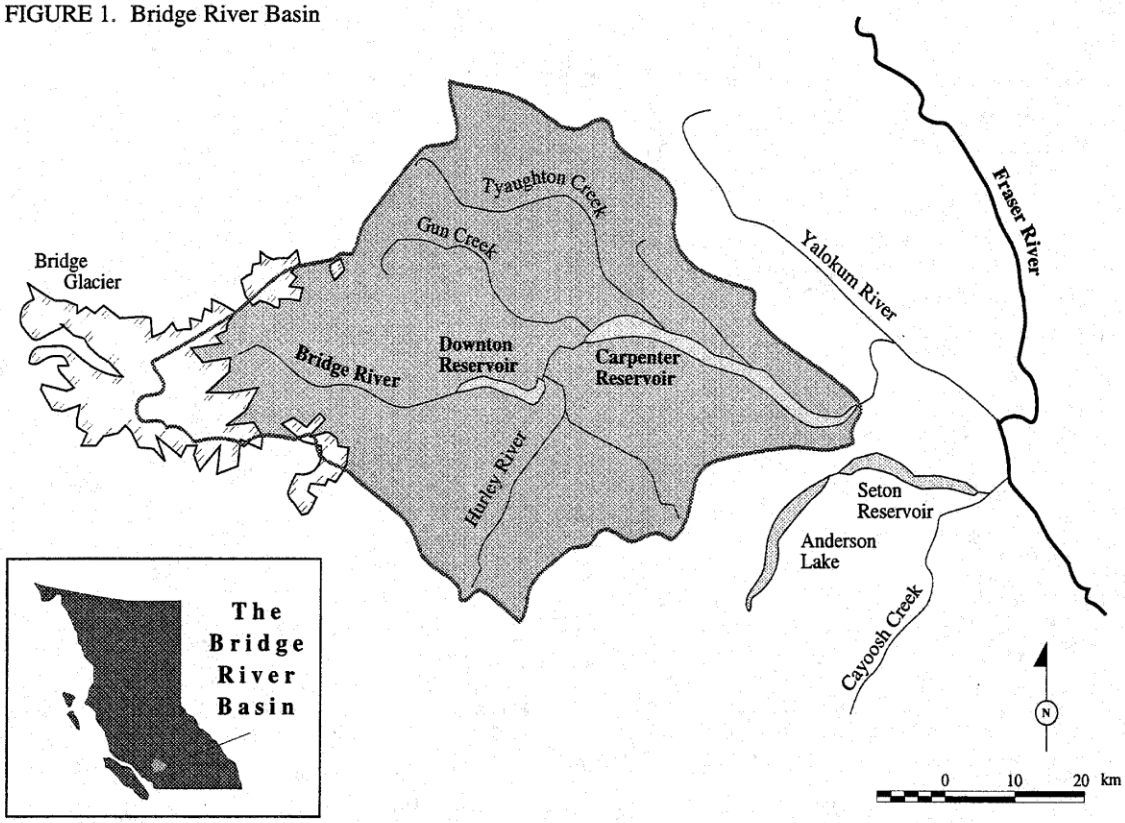


Table 2 outlines some general information about Carpenter and Downton reservoirs and their generating capacity.

TABLE 2

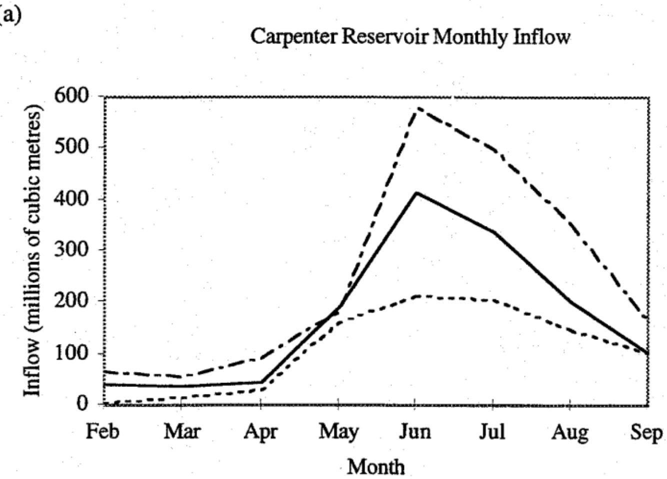
Description	Carpenter	Downton
Basin Area (square kilometres)	2719	984
Mean Runoff (millions of cubic metres)	1366	1145
Storage (millions of cubic metres)	1,011	722
Operating Range (metres)	44.5	49
Nameplate Capacity (MW)	428	22
Average Energy Capability (GWh)	2670	170

The basin hydrology is affected by the presence of Bridge Glacier at the extreme west end of the basin. A significant amount of runoff occurs from the glacier, which tends to have a moderating effect on the variability of annual flows, particularly to Downton Lake. Figure 2 plots the average, high, and low monthly inflow to the two reservoirs. While the glacier tends to moderate the inflows to Downton reservoir, it also tends to make forecasting for the reservoir more difficult, as much of the late season runoff is from by glacier melt rather than from snowmelt.

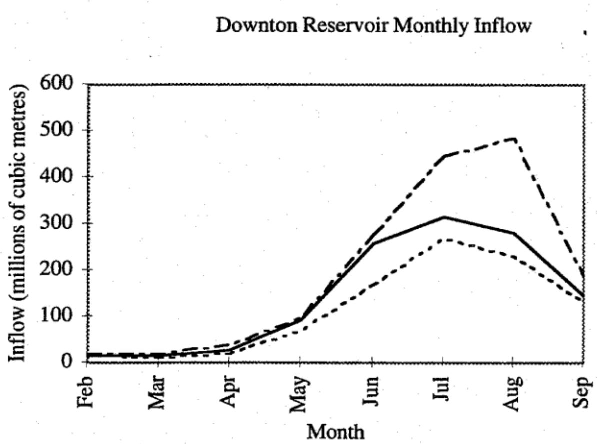
The hydrometeorologic data collection network in the Bridge River basin is fairly sparse. However, Dam Safety has operated a number of climate stations for several years, and there was some data available that could be potentially used in developing water supply forecast equations.

Figure 2. Mean monthly inflow for (a) Carpenter and (b) Downton reservoirs

(a)



(b)



Results for the Bridge River

The update procedures outlined in the preceding sections were applied to the Downton and Carpenter reservoirs in the Bridge River system in an effort to update the water supply forecast equations for those basins. Results from these procedures will be presented and compared to the previous equations created using multiple regression. We will also present the variations possible in the April 1 forecast equation depending on the methodology chosen to select the final equation. We will attempt to demonstrate the trade-offs necessary between statistical accuracy, as measured by cross-validation standard error (CVSE), and station and variable selection.

Comparison of Equation Selection Techniques

Equations were developed starting with the April 1 forecast of April through September volume runoff. It was our focus to balance improvements in statistical accuracy with physically sound station and variable selections. Our technique for developing the equations was very interactive, as we explored the capabilities of the principal components analysis procedure, and experimented with the trade-offs between optimizing CVSE and forcing certain stations and variables into the equation.

To evaluate the contribution of each variable to the forecast, we started developing the April 1 equation by adding in one variable at a time, starting with the April 1 snow course readings. Then winter precipitation was added. Both accumulated and individual monthly data were tried, to ensure that our decision to accumulate precipitation from November would not greatly degrade the forecast accuracy. Other variables that were included in the analysis were accumulated fall precipitation (Sep. - Oct.), mean monthly maximum temperature, winter streamflow, and southern oscillation index. With the exception of fall precipitation, none of these other variables were found to be significant to the water supply forecast for either Carpenter or Downton reservoir.

Table 3 presents the summary statistics for a number of the alternatives tried in developing the April 1 forecast equation for Carpenter Lake. Equation 1 lists the standard error and correlation coefficient for the origi-

TABLE 3. Comparison of April 1 forecast equations for Carpenter Lake

	Run	Standard Error	CVSE	Correlation Coefficient
1	VOLCAST	85.6		0.930
2	uncombined winter precipitation	79.8	86.5	0.944
3	accumulated winter precipitation	86.1	93.5	0.924
4	preliminary 1	90.3	98.5	0.920
5	preliminary 2	93.0	101.6	0.911
6	Final Equation	98.2	112.0	0.920

nal equation used in the past. Based on CVSE alone, Equation 2 would appear to be the "best" forecasting equation, showing an eight percent improvement over the next best equation. This equation was developed using individual monthly precipitation values (not accumulated) plus April 1 snow course readings. Unfortunately, this equation uses three different stations to represent four months of precipitation data from October through January. No precipitation station was selected for February or March. It would be very difficult to try to explain the physical processes driving this equation to the end users of the forecast and the choice of stations is likely dependent on particular period of record used.

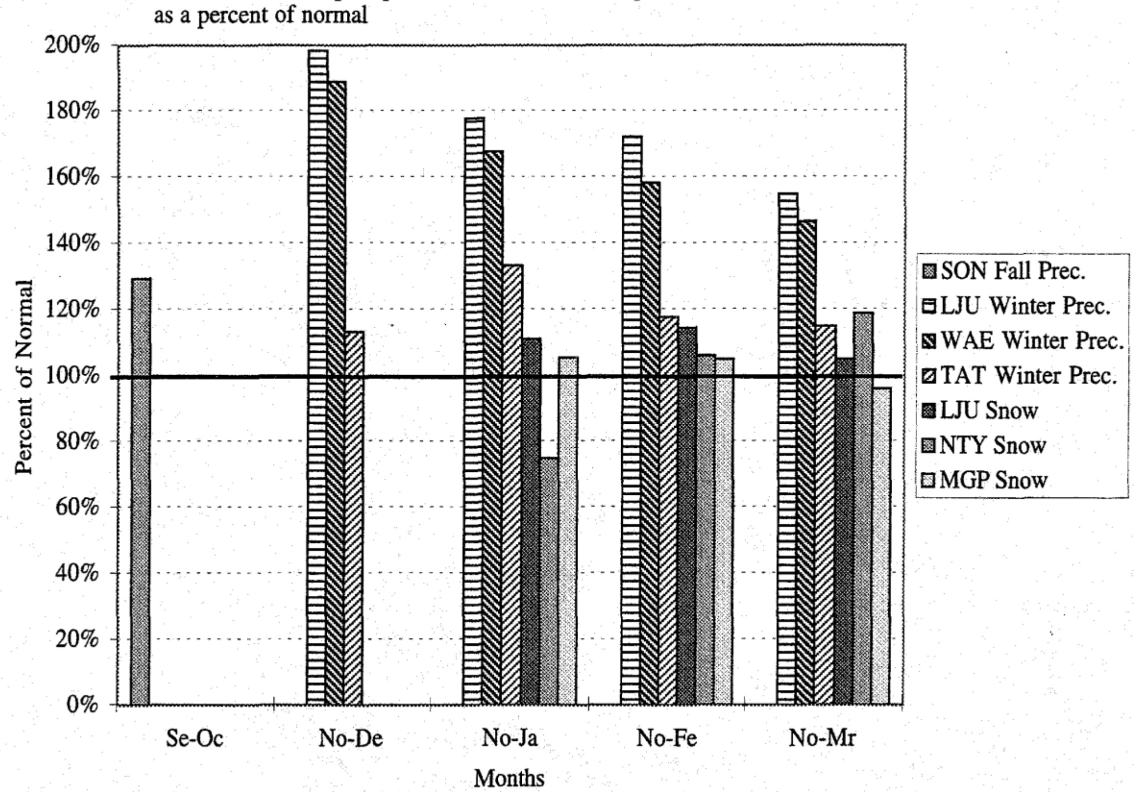
When the winter precipitation term, accumulated beginning in November, is included in the PCA program, there is a slight increase in standard error over using the individual monthly values, but the ability of the forecaster to explain the physical basis of the equation increases considerably. There is another increase in CVSE when the procedure was forced to use particular precipitation stations identified as the key stations in previous runs (Equation 4). Equation 5 was developed (with another increase in CVSE) to force the procedure to use a station that is located within the Bridge River basin, as both the climate stations selected for Equation 4 are located outside the basin and are not operated by B.C. Hydro. The development of the Equation 6, which excludes any winter precipitation term, was not necessary until after the new water supply forecasting equations were implemented beginning in January 1996.

1996 Forecasting

1996 has been a very unusual year hydrologically in much of the Pacific Northwest and elsewhere. The Bridge River basin is no exception. Figure 3 plots climate and snow course data some key stations in the Bridge River area. The fall weather pattern was characterized by heavy precipitation (with some stations reaching record monthly amounts) combined with higher than normal temperatures. These two features working together meant that while the precipitation early in the winter was well above normal, the snowpack accumulation was very close to normal. But when it came time to issue a water supply forecast on February 1 using both snowpack readings and accumulated winter precipitation, the conflicting reports caused a problem.

Generally in British Columbia, one expects the seasonal runoff to be directly related to the amount of snow accumulated on the mountainsides by early spring. Winter precipitation is expected to add a little extra information to the snowpack readings, particularly earlier in the season, while the snowpack is still developing. As seen

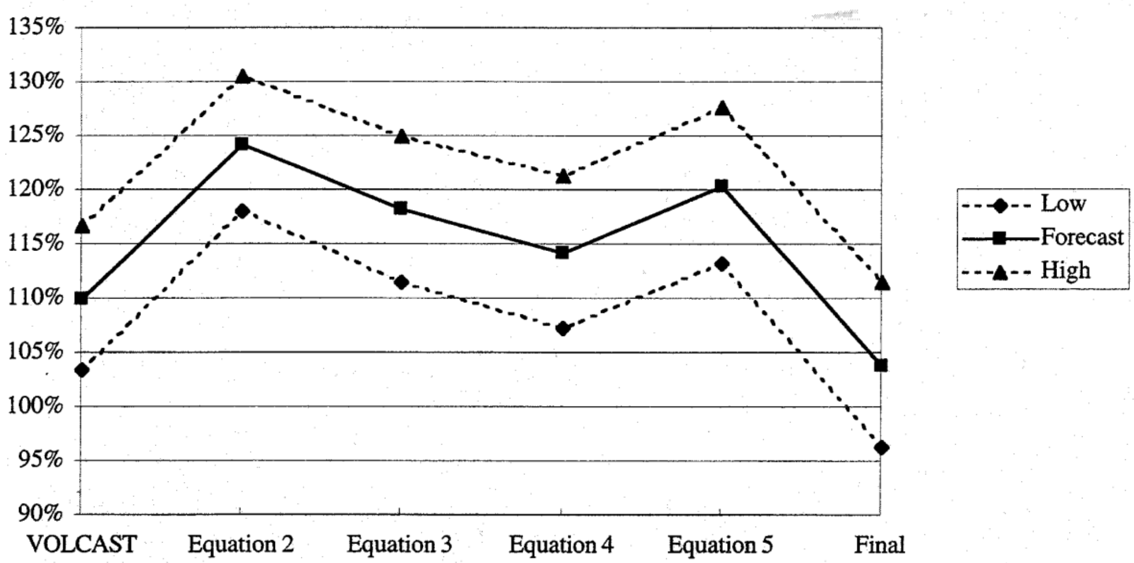
FIGURE 3. 1996 Accumulated precipitation and snow water equivalent as a percent of normal



in 1996, however, this correlation between winter precipitation and snowpack does not always hold up. Winter precipitation was well above normal, while the accumulated snowpack was normal to slightly above normal. As a result of this discrepancy between precipitation and snowpack, we were forced to remove winter precipitation from the equation development, with the explanation that in near normal years, the winter precipitation will be highly correlated to the snowpack accumulation, but in non-normal years, the winter precipitation will be misleading and cloud the forecast. Equation 6 is based on the April 1 snowpack along with a fall precipitation term to account for the antecedent soil moisture. The winter precipitation term was removed from the equation, with a further 10 percent increase in CVSE over Equation 5. While it may seem to be a large jump in CVSE from Equation 2 to the final equation (6), as a proportion of the normal April through September runoff, it is only an increase from 6.7 percent to 8.7 percent.

Using the April 1 forecast for consistency with the previous section, Figure 4 presents the calculated April through September forecast (as a percent of normal) by each of the 6 equations, plus the expected range in the forecast as given by +/- one standard error. Except in very extreme years, the actual seasonal runoff is expected to fall within this range of the forecast. For 1996, with a snowpack near normal, the seasonal runoff would also be expected to be near normal. The April 1 forecast equations that include winter precipitation (Equations 2 through 5) all forecast the seasonal runoff to be in the neighbourhood of 115 to 125 percent of normal. Even if the low end of the forecast range is chosen, this is still higher than the forecast made excluding winter precipitation (Final equation). In particular, note that the forecast made using the individual monthly precipitation values (Equation 2), having the "optimal" statistical accuracy, was 124 percent of normal with an expected low range forecast of 117 percent. Since believability of the forecast was one of our main goals in updating the water supply forecast, we could not accept any of the equations using winter precipitation and decided accept a higher standard error with an equation that is more physically rational.

FIGURE 4. April 1 forecast +/- 1 standard error for Carpenter Lake



Limiting Input Variables

The decision to remove the winter precipitation term from the Carpenter forecasting equation was preceded by the experimentation with several alternate techniques to somehow reduce the impact of very high precipitation values that were not accumulated as snow. Our first thought was to somehow split the 30 year sample and produce different equations for different climate regimes. This method has two major obstacles. First, the 30 year period was chosen as it is just large enough to be representative of the expected range of events for the basin, and thus any attempts to reduce the sample size would yield unrepresentative equations. Secondly, producing different equations for different climate regimes could cause inconsistencies later in the forecast season if the climate were to shift over the course of the season from one regime to another, forcing a switch in the equations used.

Our next thought was to somehow place a cap or limit on how large any variable could be by setting an upper limit or applying a log or cube root transformation to the data. Unfortunately, placing a limit at one end of a distribution or transforming the data would violate the assumption that each independent variable is normally distributed.

Finally, we attempted to somehow account for the precipitation that fell as rain in the fall and early winter, by generating a forecast of November through September runoff and then subtracting the observed runoff to forecast date. This technique was not successful for the Bridge River basin, but may have greater usefulness for coastal basins which may have a higher frequency of precipitation falling as rain and running off during the winter.

The only alternative remaining was to remove the winter precipitation term and accept the losses to statistical accuracy.

Explaining the Error - using future variables

Table 4 presents the summary statistics for Downton Reservoir generated during the development of its April 1 water supply forecast equations. The obstacles encountered while developing forecasting equations for Downton were very different from those for Carpenter. Most of the obstacles could be related to the fact that Downton Lake is glacier-fed and much of its seasonal water supply is influenced by glacier melt. The correlation coefficient for the April 1 forecast for Downton Reservoir was only 0.653 for the "best" equation selected using the PCA/search technique, which was based on a single snow course for the forecast (Equation 2). There are two possible reasons for these results: either the stations used in the procedure were not representative of the basin hydrology (Type 2 error), or the seasonal volume runoff can not be predicted very well with only the variables known at forecast time (Type 1 error).

TABLE 4. Comparison of April 1 forecast equations for Downton Lake

Equation	Description of variables	Standard Error	CVSE	Correlation Coefficient
1	VOLCAST	90.5		0.693
2	April snow (single snow course)	90.9	97.3	0.653
3	April snow, summer precip. & temperature	78.0	87.7	0.753
4	Fall and winter precipitation, April snow	93.4	99.8	0.629
5	April snow (Final Equation)	95.1	102.1	0.627

To check the latter case, the procedure was repeated with the inclusion of the future summer precipitation and temperature variables in the analysis. As can be seen from Equation 3 in Table 4, by adding the summer variables to the April 1 forecast, the correlation coefficient rose from 0.65 to 0.75 and the CVSE dropped by 14 percent. This would indicate that the relatively poor statistics for Equation 2 are a result of factors unknown at the time of the forecast and that trying to add in more variables or stations would not greatly improve the forecast accuracy. Other variables such as winter streamflow and the previous year's volume runoff were also tested in the analysis with little success.

Equation 5 is the final equation chosen as the April 1 forecast equation for Downton Reservoir. A second snow course was added to the forecast to guard against missing or anomalous data at the original snow course included in Equation 1. Equation 4 also includes fall and winter accumulated precipitation, but neither of these terms added much to improve the accuracy of the forecast, and were eliminated.

Glacier Melt

Because of the importance of the glacier melt component to the seasonal water supply at Downton reservoir, as a part of the review of the forecasting equations, we attempted to better predict the expected glacier melt. As was mentioned earlier, mean monthly maximum temperature was used to forecast glacier melt. It was our hope to find some link earlier in the season to predict the late season melt.

First we tried correlating the date of snow disappearance at the Mission Ridge snow pillow to the late season runoff with the expectation that the sooner the snow cover is gone off the glacier, the higher the glacier melt component will be. Unfortunately, there were only eight years of data available for the snow pillow and the dates were all very close together, so the results were inconclusive.

The second approach we tried was to use monthly total hours of sunshine as a proxy for solar radiation. Incoming solar radiation has been identified as one of the key forces driving glacier melt (Brugman and Pietroniro, 1995). After identifying a reasonable correlation between hours of sunshine and measured solar radiation, hours of sunshine data from a station just outside the Bridge River basin were added to the late season forecast analysis for Downton Lake. Results showed that the monthly sunshine hours did not correlate well to seasonal runoff.

Late Season Forecasts

Principal component analysis provided satisfactory results in developing water supply forecast equations for the forecast dates from January 1 to June 1. By adding in additional snow course readings to the April 1 snowpack the May and June equations continued to reduce in CVSE. By July 1 however, the peak runoff due to snowpack contributions will have already occurred and runoff becomes a results of summer precipitation and glacier melt. As we are unable to predict these components of the runoff in advance as yet, we decided to try using residual and recession forecasts to distribute the remaining expected flow through the rest of the season. Residual forecasts, that is the June forecast minus the observed flow since June, were used for both Carpenter and Downton for the July 1 and August 1 forecast. The forecast statistics were marginally better than those for the PCA results.

Final Equations and Discussion

The final water supply forecast equations with the variables and their regression coefficients and summary statistics are presented in Table 5 (a) and (b) for the Carpenter and Downton reservoirs.

TABLE 5 (a). Carpenter Lake Volume Forecast Equations

	Forecast Date						
	February	March	April	May	June	July	August
Station Data	Feb-Sep	Mar-Sep	Apr-Sep	May-Sep	Jun-Sep	Jul-Sep	Aug-Sep
Downton UpperSWE							
February	0.395						
March		0.402					
April			0.533	0.248	0.244		
May				0.246			
June					0.24		
Mission Ridge Pillow SWE							
February	0.619						
March		0.533					
April			0.256				
May				0.326			
June					0.311		
Seton							
Sep-Oct precip	1.384	1.376	1.137	0.699			
Intercept	697.6	588.4	513.5	505.3	550.7	residual	residual
Jackknife Standard Error	159.6	124.3	112.0	95.0	80.1	n/a	n/a
Jackknife Corr. Coeff.	0.784	0.865	0.9	0.918	0.933	n/a	n/a
Standard Error	150	116.4	98.2	89.9	75.2	68.6	57.5
Correlation Coeff.	0.811	0.883	0.92	0.927	0.941	0.899	0.713

TABLE 5 (b). Downton Lake Volume Forecast Equations

	Forecast Date						
	February	March	April	May	June	July	August
Station Data	Feb-Sep	Mar-Sep	Apr-Sep	May-Sep	Jun-Sep	Jul-Sep	Aug-Sep
Downton UpperSWE							
February	0.125						
March		0.166					
April			0.148	0.098	0.09		
May							
June							
McGillivray Pass SWE							
February	0.262						
March		0.337					
April			0.3	0.201	0.181		
May				0.175			
June					0.142		
Intercept	954.5	817.7	790.0	765.3	752.6	residual	residual
Jackknife Standard Error	126.0	108.1	102.1	100.3	92.2	n/a	n/a
Jackknife Corr. Coeff.	0.335	0.581	0.627	0.622	0.606	n/a	n/a
Standard Error	117.3	99.3	95.1	93.0	84.8	75.2	65.9
Correlation Coeff.	0.461	0.659	0.686	0.685	0.678	0.476	0.352

We are satisfied that the new equations meet the criteria that we established at the start of this study, that is to:

- use principal components analysis and systematic search to identify optimal combinations of variables;
- use B.C. Hydro- operated stations located within the Bridge River basin where possible;
- preserve month-to-month consistency in the forecast;
- show that slight increases in the standard error can be balanced by improved confidence in station and variable selection, and;
- try new variables and techniques, particularly in the late-season forecasts.

CONCLUSIONS

By employing new techniques in the development of regression-based water supply forecasting equations, we were able to update the forecasts of February through September runoff for the Carpenter and Downton reservoirs in the Bridge River system. While the new techniques, particularly principal component analysis, show the ability to significantly reduce the standard error of forecast, these procedures should also be tempered by interactively ensuring that the choice of stations and variables being input into the process represent the expected physical hydrological relationships present in the basin. Such direct interference may cause some loss in statistical accuracy, but this is balanced by the ability to explain the physical basis of the forecast and the resulting gain in the credibility of the forecast to the end users.

With the updating procedure defined by the update of the Bridge River water supply forecasts, we are currently reviewing the forecasts for the Columbia River projects using a similar procedure.

REFERENCES

Brugman, M.M. and A. Pietroniro. 1995. "Glacier runoff modeling." Proceedings of the Mountain Hydrology Conference. Vancouver, B.C.

Columbia River Water Management Group Forecast Committee (CRWMGFC). 1993. "Review of runoff forecasting on the Columbia River and Pacific Slope basins" Report to Northwest Power Planning Council.

Garen, D.C. (1993). "Improved techniques in regression-based streamflow volume forecasting." J. of Water Resources Planning and Management, 118(6), 654-670.

Gray, D.M. 1970. Handbook on the Principles of Hydrology. Water Information Center, Inc. NY.

Grygier, J. et al. (1993). "Disaggregation models of seasonal streamflow forecasts". Proceedings of the 61st Western / 50th Eastern Snow Conference, Quebec City, PQ.

Hirsch, R.M. (1982). "A comparison of four streamflow record extension techniques." Water Resources Research, 18(4), 1081-1088.

Koch, R.W. and D. Buller. (1993). "Forecasting seasonal streamflow: Columbia River at The Dalles". Report prepared for Bonneville Power Administration.

Smith, S.E. and E. Weiss. (1996) "Record estimation and extension for precipitation data." Proceedings of the 64th Western Snow Conference, Bend OR.