REGRESSION MODELS IN WATER SUPPLY FORECASTING[1/]

715-82

By

A. Leon Huber and David C. Robertson[2/]

INTRODUCTION

Water supply forecast models may be categorized as either index (regression) type models or as conceptual simulation models. Conceptual simulation models of the snowmelt/runoff process have their greatest utility in assessing alternative management or operational scenarios, i.e., providing answers to "what if" type questions and will not be discussed here. This paper reviews some of the most prominent index methods and outlines a strategy for applying regression methodology to the problem of real-time water supply forecasting. The proposed method could be used conjunctively with the Soil Conservation Service (SCS) SNOTEL data acquisition system to enable water supply forecast updates on a daily or real-time basis.

REVIEW OF INDEX MODELS FOR WATER SUPPLY FORECASTING

Zuzel and Cox (1978) reported on work which compared five of the most commonly used index methods for forecasting seasonal water supply. The index models tested were:

1. A precipitation index model of the form

$$Q_s = a + bP_w$$

2. A water-balance model of the form

$$Q_s = a + b(P_w - Q_w)$$

3. The Tangborn-Rasmussen model of the form

$$Q_s = a + bP_w - Q_w$$

4. The general multiple regression model of the form

$$Q_s = a + \sum_{i=1}^{k} b_i X_i \quad \text{and}$$

5. The pattern search model which is functionally identical to the multiple regression model, except that the coefficients multiplying the index variables are constrained to be non-negative.

In the above equations, Q_s is the seasonal runoff to be forecast, a and b's are regression coefficients, P_w is the cumulative winter precipitation from October 1 through the forecast date, and Q_w is the cumulative winter runoff corresponding to P_w.

It may be noted that all five of the index methods are actually functional subsets of the general multiple regression model. The only real difference is in how the coefficients are obtained and in the index variables used. A comparison of the forecasting ability of the five methods applied to a stream gaging station in the Boise River drainage basin is given in Table 1.

Poster Paper presented at the Western Snow Conference, April 19-23, 1982, Reno, Nevada.

1/ Contribution from the Northwest Watershed Research Center, USDA-ARS; and Soil Conservation Service, USDA.

2/ Hydrologist and Hydrologic Technician, respectively; Northwest Watershed Research Center, USDA-ARS, 1175 South Orchard, Suite 116, Boise, Idaho 83705.

Table 1

Comparison of five index type water supply forecast models with respect to their correlation coefficients (r), standard errors of estimate as a percent of mean flow (% S.E.), and the indexing variables for forecasting seasonal runoff for the Boise River near Twin Springs, Idaho, for six forecast dates (from Zuzel and Cox, 1978).

Forecast Date		Forecast Model: Precipitation Index	Water Balance	Tangborn-Rasmussen	Multiple Regression	Pattern Search
JAN. 1	r	0.828	0.787	0.812	0.828	0.924
	%S.E.	18	20	19	18	13
	Variables[1/]	1	1,2	1,2	1	1,4,6
FEB. 1	r	0.917	0.880	0.893	0.917	0.969
	%S.E.	13	15	14	13	8
	Variables	1	1,2	1,2	1	1,3
MAR. 1	r	0.974	0.921	0.953	0.988	0.969
	%S.E.	7	13	10	6	8
	Variables	1	1,2	1,2	1,3,5	1,3
APR. 1	r	0.931	0.833	0.833	0.985	0.975
	%S.E.	12	18	15	7	8
	Variables	1	1,2	1,2	1,3,5	1,3
MAY 1	r	0.928	0.835	0.915	0.977	0.978
	%S.E.	13	19	14	9	8
	Variables	1	1,2	1,2	1,4,5,6	1,3,4,6
JUN. 1	r	0.899	0.360	0.719	*[2/]	*
	%S.E.	17	37	27	*	*
	Variables	1	1,2	1,2	*	*

1/ Variables:
1. Arrowrock Precipitation
2. Twin Springs Runoff
3. Atlanta Summit SWE[3/]
4. Bad Bear SWE
5. Mores Creek SWE
6. Trinity SWE

2/ * Insufficient June 1 Data.

3/ Snow Water Equivalent.

From Table 1 it may be observed that the standard error of estimate (S.E.) varied from 6 to 37 percent of the mean seasonal runoff. The general multiple regression model exhibited a standard error of 6 to 18 percent, and is typical of the forecast accuracy attainable by the regression approach.

REGRESSION MODEL STRATEGY FOR REAL TIME FORECASTING

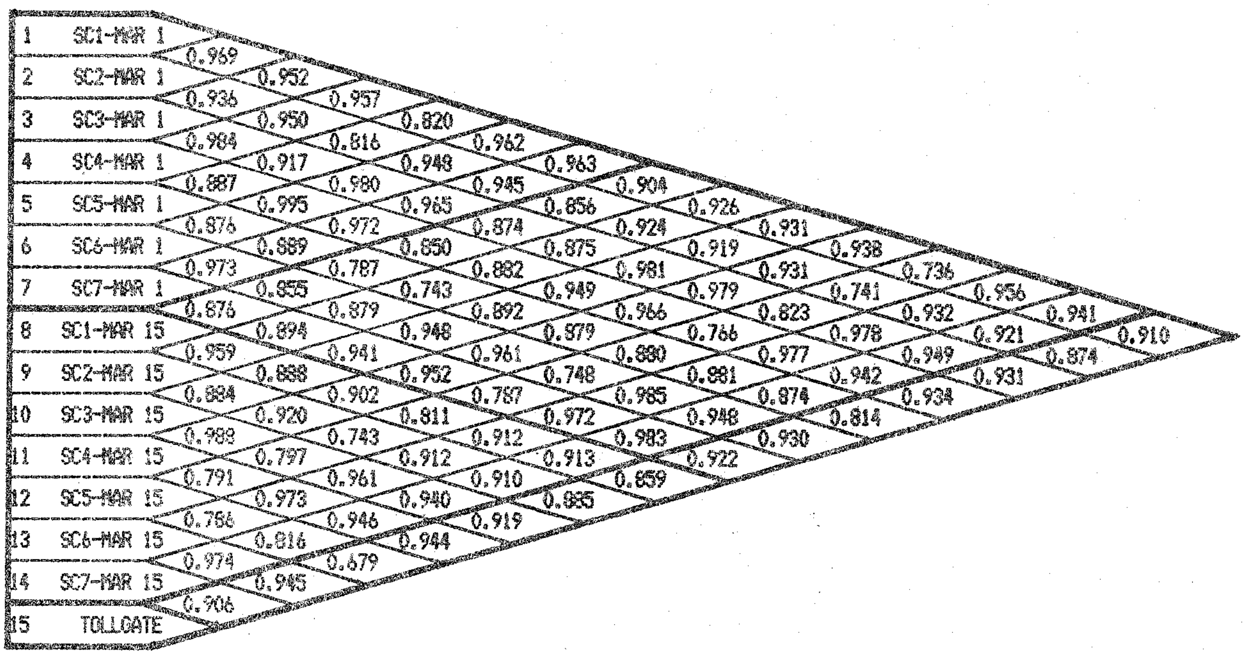
A strategy for using regression analysis techniques for real-time seasonal water supply forecasting is outlined as follows:

1. Correlation Analysis and Variable Selection

The index variables that are available for inclusion in a forecasting model are assembled and a simple correlation map is computed which provides a basis for selecting different sets of the index variables for inclusion in the regression models to be tested. The selection criteria may be somewhat subjective, but should include consideration of the intercorrelations existing among the various candidate variables as well as their correlation with the seasonal runoff to be forecast. An example of a correlation map computed from data for seven Reynolds Creek Experimental Watershed snow courses and the Tollgate Weir is shown in Figure 1. A correlation map is analogous to a highway map mileage table. The product moment correlation coefficients (r) between all two way combinations of variables are printed at the intersections of the diagonal lines projected from the respective variable labels. For example, the correlation between the March 1 SWE at snow course No. 4 (SC4-MAR 1) and the runoff at Tollgate is shown as 0.934 in Figure 1.

Figure 1

Correlation map of seven Reynolds Creek Experimental Watershed snow course March 1 and March 15 snow water equivalents, and Reynolds Creek at Tollgate March through July seasonal runoff. The data included measurements for the years 1966 through 1981.



2. Regression Trials

The second step is to test various regression equations (models) composed of selected variables as candidates for the regression analysis. The best model or models is determined by considering the analysis of variance standard error of estimate, rationality, and stability when tested with data not used in the derivation of the model. Table 2 is a summary of the results of the March regression trials for the variables shown in Figure 1.

3. Develop Set of Forecast Equations Corresponding to the Historical Data

The regression trials described under step 2 are now repeated for all historical data and the best equations selected for each forecast date. Traditionally these have been the first of each month from January through April. In some cases, historical data are available to develop an additional midmonth forecast equation. The resulting set of equations becomes the basis for the real-time forecast procedure.

4. Develop Interpolating Procedure for Using Real-Time Data

The forecast for a date other than the one for which the forecast equation was derived is a function of the two estimates obtained by calculating forecast estimates using the real-time data in the two forecast equations that bracket the desired date. For example, suppose that a forecast update is desired March 10th, but only March 1 and March 15 equations have been developed from the historical data. Using 1982 data from Reynolds Creek, a March 10 forecast is made by computing Q_s from both equations using March 10 SWE data at snow course No. 4 and linearly interpolating with time between them. With a March 10, 1982 SWE of 711 mm (28.0 inches) at SC4, the March 1 and March 15 equations forecast 262 mm (10.35 inches) and 246 mm (9.68 inches) of runoff, respectively. Linearly interpolating for the time difference between March 10 and the 2 equation dates of March 1 and March 15 yields a March 10 forecast estimate of 252 mm (9.92 inches) of seasonal runoff at the Tollgate Weir. Other interpolating procedures, such as a multipoint Lagrange interpolating polynomial, could also be tested and used if proven better.

Table 2

Summary of regression results for forecasting seasonal March-July runoff from Reynolds Creek measured at the Tollgate Weir. All variables in the regression were snow water equivalent (SWE).

EQUATION NUMBER	VARIABLES IN REGRESSION	DEGREES OF FREEDOM	F	R	PERCENT S.E.	PERCENT ERROR FOR EXTREME YEARS 1972 HIGH	1977 LOW	1982 FORECAST (INCHES)
1	7 SNOW COURSE SWE FOR MARCH 1 AND MARCH 15	(14,1)	4.34	0.992	24.8	-3.1	-12.3	7.58
2	SC1-MAR1, SC1-MAR15, SC6-MAR15, SC7-MAR15	(4,11)	24.55	0.948	18.7	4.6	129.1	10.17
3	7 SNOW COURSE SWE FOR MARCH 15 (1966-1980 DATA SET)	(7, 7)	13.18	0.964	18.2	1.5	86.4	10.15
4	7 SNOW COURSE SWE FOR MARCH 15 WITH 1981 DATA ADDED	(7, 8)	14.37	0.962	18.7	2.6	126.9	10.82
		MARCH 1 EQUATIONS						
5	SC4 AND SC5	(2,13)	44.85	0.936	19.2	6.0	12.1	9.69
6	SC5 AND SC6	(2,13)	41.68	0.930	19.8	10.6	64.8	9.52
7	SC4	(1,14)	95.78	0.934	18.6	9.2	-37.0	9.56
8	SC6	(1,14)	89.78	0.930	19.1	10.5	64.1	9.51
		MARCH 15 EQUATIONS						
9	SC4 AND SC5	(2,13)	61.16	0.951	16.8	-2.5	79.6	10.73
10	SC5 AND SC6	(2,13)	61.70	0.951	16.7	9.2	120.6	10.55
11	SC4	(1,14)	115.23	0.944	17.1	-5.2	57.0	10.40
12	SC6	(1,14)	117.85	0.945	17.0	5.3	93.9	10.27

5. Produce Upper and Lower Forecast Estimates Using Variances from Analysis of Historical Data

A confidence interval about the forecast derived in the previous step may be made by applying traditional statistical techniques. This, of course, requires variance estimates of the error in the forecast equations which would have to be obtained from the regression analysis of variance produced in deriving the forecast equations.

CONCLUSION

A procedure has been outlined which enables the traditional regression based water supply forecast methods to be adapted to the problem of forecasting in real or near real-time. It offers the following advantages to the water supply forecaster:

1. Conceptual simplicity and widespread acceptance.
2. Confidence intervals about the estimate may be easily calculated.
3. Requires a minimum of computational facilities.
4. Forecasts are available instantaneously.
5. Forecast errors are acceptably small.

The regression approach is a viable tool of the water supply forecaster and will probably remain the primary method for forecasting the seasonal water supply prior to the snowmelt runoff period.

REFERENCES

Zuzel, J. F., and L. M. Cox, 1978: A review of operational water supply forecasting techniques in areas of seasonal snowcover, in 46th Annual Western Snow Conf. Proc., Otter Crest, Oregon, April 18-20, p. 69-77.