EVALUATION OF STREAMFLOW FORECASTING METHODS 1/ 654-79

by

D. P. Lettenmaier and D. C. Garen 2/

The value of streamflow forecasting to operation of water resource systems has long been recognized. In watersheds not affected by snow, runoff is largely effected by precipitation excess and channel precipitation resulting in runoff during and immediately following precipitation events, and by subsequent base flow. Since precipitation forecasts are rarely useful for forecast periods beyond several days, forecast accuracy in such watersheds is largely dependent on the magnitude of the base flow contribution relative to total runoff. Burges and Johnson (1973) have described forecast methods, applicable to such situations, which make use of seasonal runoff correlations. In areas where substantial snow accumulations occur, the opportunity exists for forecasting runoff for a period of several months on the basis of winter snowpack accumulation. In such cases, a variety of regression-based and physical models have been used to forecast runoff volumes during spring and summer snowmelt seasons. In this paper, only models in the former classification are considered, although implications of the results for application and development of physical models are discussed.

Tangborn (1977) has made use of a basin storage accounting model, in which basin storage is defined to be winter (pre-forecast) season accumulated precipitation falling on the basin less runoff and evaporation, where forecasted summer runoff is winter storage less the difference of summer evapotranspiration, net groundwater transport into the basin which occurs as streamflow, and summer precipitation contributing to runoff. In this model, accumulated precipitation is estimated as a weighted sum of observed precipitation at key index stations. An alternate formulation, discussed by Lettenmaier (1978) utilizes weighted snow course snow water equivalent measurements to estimate storage directly, with precipitation measurements used only to update storage from the date of the most recent snow course measurement to the forecast date. In this paper, results obtained using the two approaches are compared with those obtained using streamflow measurements only for eight river basins of widely differing climatic and topographic characteristics, including three each in Washington and Arizona, one in Wyoming and Montana, and one in Colorado.

Model Structure

When forecasts of seasonal runoff volume are made on the basis of previous seasonal flow volumes only, and a lag one Markov model is applicable, Burges and Johnson (1973) have shown that the following model results:

$$\hat{X}_1 \mid X_0 = \mu_1 + \rho_{01} \frac{\sigma_1}{\sigma_0} (X_0 - \mu_0) \tag{1}$$

where $\hat{X}_1$ is the forecasted runoff in period 1, X_0 is the observed runoff during the base period (always taken here as the beginning of the water year to the forecast date), ρ_{01} is the correlation coefficient between runoff in period 0 and 1, μ_0 and μ_1 are the mean flows in the base and forecast periods, respectively, and σ_0 and σ_1 are the corresponding standard deviations. The variance of the forecasted runoff is

$$\mathrm{Var}(\hat{X}_1 | X_0) = (1 - \rho_{01}^2)\sigma_1^2 \tag{2}$$

The square root of the coefficient of σ_1^2 in equation 2 represents the reduction in scale of the distribution of forecasted runoff due to the forecast, and ranges from 0 to 1 as $|\rho_{01}|$ ranges from 1 to 0. In the case where $\rho_{01} = 0$, equation 1 states that the best (in terms of minimum mean square error) forecast is just the mean forecast period runoff, which has

1/ Presented at the Western Snow Conference, Sparks, Nevada, April 18-20, 1979.

2/ Department of Civil Engineering, University of Washington, Seattle, Washington 98195.

variance σ_1^2. In practice, the means (μ_0, μ_1), variances (σ_0^2, σ_1^2) and the correlation coefficient (ρ_{01}) are not known and must be estimated; of interest here is only the root mean square forecast error reduction, which is estimated as $\sqrt{1 - \hat{\rho}_{01}^2}$, with $\hat{\rho}_{01}$ the estimator of ρ_{01}. It should be noted that in practice one would not select a lag one Markov model apriori, but would identify the model which appeared to be most likely to have generated the data. Usually, however, when confronted with record lengths which are commonly available it is difficult to justify a model more complex than that given by equations 1 and 2. This observation is applicable in particular to the test basins examined below, all but one of which have streamflow records of less than forty years duration.

When runoff forecasts are made on the basis of recorded precipitation, the model is

$$R_t^* = C_w P_w - R_w + B_t \qquad (3)$$

$$R_s^* = C_{w+t} P_{w+t} - R_{w+t} + B_s \qquad (4)$$

where the subscripts w, t, and s denote winter, test, and summer respectively and the seasons are contiguous, i.e., the test season is a period (usually ranging from one to thirty days) during which an initial forecast is made. The subscript w+t denotes the combined winter and test season; C_w, C_{w+t}, B_t and B_s are constants estimated by regression. Once the test season forecast has been made, the summer forecast is updated as

$$R_s^{**} = R_s^* + K(R_t^* - R_t) \qquad (5)$$

where R_t is the observed runoff during the test season and K is a constant determined by regressing forecast period error $R_s^* - R_s$ on test period error $R_t^* - R_t$. The constants B_t and B_s represent the effects of evaporation, subsurface interbasin flow contributing to runoff, and precipitation during the test or forecast season. The precipitation record may represent either a single or weighted sum of measured data. It has been noted (Tangborn, 1977) that low elevation precipitation stations often provide more accurate forecasts than do high elevation stations, even when the low elevation stations are relatively remote from the basin.

When snow course measurements are included, the forecast equations are rewritten as $R_s' = S_w + B_s$, where S_w is the winter basin storage, estimated as $S_w = C_1 Y + C_2 P'$ where Y is the most recent snow course measurement prior to the forecast date, P' is the precipitation from the snow course measurement date to the forecast date, and C_1 and C_2 are constants determined by regression. Although use of a test season is theoretically possible, experience has shown that it does not improve forecast accuracy when snow course measurements are used, so it is not included here. As with the precipitation records, the snow course record Y may consist of either the record at a single site or a composite of several sites.

In determining which precipitation and snow course stations to incorporate in the forecast model, a preliminary screening model was used, where cumulative runoff from the beginning of the water year to the forecast end date was regressed on all the cumulative precipitation or snow course records. It was recognized that the significance levels have little meaning in an absolute sense, since the records are highly correlated, however they do provide the basis for ranking station importance. When more than one record existed with comparable (determined subjectively) significance levels, a composite record was formed as

$$Z = \underline{U}A$$

where Z = n x 1 vector of composite precipitation or snow course data
$\underline{U}$ = n x m matrix of raw station records (precipitation or snow course) ranked from left to right in order of significance
A = m x 1 vector of weighting coefficients, with $A_j = W_j \overline{Y}^* / \overline{Y}_j$

$$W_j = \frac{m-j+1}{\sum_{j=1}^{m} j}$$

$$\bar{Y}^* = \sum_{j=1}^{m} W_j \bar{Y}_j$$

and $\bar{Y}_j$ is the mean of the j th raw record. This weighting function has the advantage that the weighting parameters are not estimated directly from the data, which would burden the limited data sets with further loss of degrees of freedom, whereas the ordering of station importance in forecasting is preserved.

Although the two forms of the storage accounting model are simple, and incorporate only a gross conceptual knowledge of the runoff process (e.g., continuity on a basinwide scale) it is nevertheless difficult to compute a theoretical RMSE reduction as was done for the Markov model. Alternately, the RMSE reduction is estimated here empirically, by comparing forecast with actual runoff using a split sample method, in which model parameters are estimated using an initial calibration period, and forecasts made for several subsequent years. Periodically, parameters are updated using the additional accumulated record, but the forecasts are always made without knowledge of the actual flow until after the forecast has been made. In this work, parameters are updated every third year, although in practice yearly updating would be preferred and might result in small accuracy improvements. After forecasts have been made for the entire period of record (less the initial calibration period) an estimate of the forecast error RMSE was made from the sequence of forecast errors. With minor exceptions, all forecasts were made for the period 1949-75 with the initial nine years used for calibration only. It is emphasized that the RMSE estimate calculated in this manner reflects the actual errors which could be expected, as opposed to residual errors often cited in model assessments, which are generally smaller but do not reflect the operational accuracy of a forecast model (Lettenmaier and Waddle, 1978).

Results

The eight streams and forecast points considered are given in Table 1. Table 2 reports summary physical and climatic characteristics of the basins. Comparison of the entries in the last column of Table 2 indicates that the normalized annual flow of the basin varies over two orders of magnitude from the marine climate of the Cedar River basin to the near arid climate of the Salt, Verde, and Tonto basins.

Table 1

Streamflow Forecast Points

USGS Gage No.	Location
12-1150	Cedar River near Cedar Falls, Washington
12-4885	American River near Nile, Washington
12-4510	Stehekin River at Stehekin, Washington
8-2200[a]	Rio Grande River near Del Norte, Colorado
6-2075	Clarks Fork Yellowstone River near Belfry, Montana
9-4985	Salt River near Roosevelt, Arizona
9-5085	Verde River below Tangle Creek near Roosevelt, Arizona
9-4990	Tonto Creek above Gun Creek, near Roosevelt, Arizona

[a] adjusted for storage change in Continental, Santa Maria, and Rio Grande rivers.

Table 2

Physical and Climatalogical Characteristics of Test Basins

Basin	GE	DA	MAS	MAP	CS	TME	FA	NAF
Cedar	1560	40.7	440	120	116	3230	77	6.71
American	2700	78.9	350	74	64	4860	91	3.12
Stehekin	1098	344	290	99	137	5130	83	4.15
Rio Grande	7980	1320	NA	27	19	10,200	30	0.69
C.F. Yellowstone	3980	1154	NA	17	76.3	7430	61	0.81
Salt	2177	4306	44	22	23	6190	71	0.20
Verde	2029	5872	32	18	16	5470	67	0.08
Tonto	2523	675	28	24	88	5020	65	0.17

GE = gage elevation, ft
DA = drainage area, $mile^2$
MAS = mean annual snowfall, inches
MAP = Mean annual precipitation, inches
CS = average main channel slope, ft/mile
TME = topographic mean elevation, ft
FA = forested area, per cent
NAF = normalized annual flow, $cfs/mile^2$

The screening model was applied to each of the basins, with the beginning of the winter season fixed at October 1 and the end of the summer season fixed at July 31, with the exception of the Arizona basins where May 31 was used. The snow course and precipitation stations selected by the screening model are given in Table 3. The ratio of station distance from the basin centroid to main stem channel length has been tabulated as an index to the remoteness of the stations which reflects the large differences in size of the basins.

Table 3

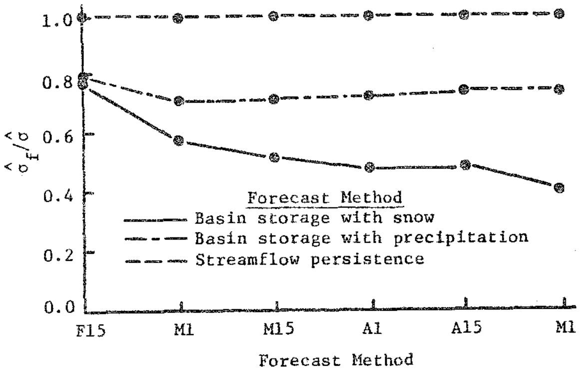
Precipitation and Snow Course Stations Used in Forecasts[a]

	Precipitation Stations		Snow Courses	
Basin	Location	Normalized Distance From Centroid[b]	Location	Normalized Distance From Centroid[b]
Cedar	Snoqualmie Falls (45-7773)	1.5	Olallie Meadows (21B2)	0.4
American	Rainier Ohanapecosh (45-6896)	1.6	Corral Pass (21B13)	0.4
	Yakima (45-9465)	2.9	Bumping Lake (21C8)	0.3
Stehekin	Darrington (45-1992)	1.1	Harts Pass (20A5)	0.8
	Lake Wenatchee (45-4446)	1.4		
	Mazama (45-5133)	0.7		
Rio Grande	Del Norte (05-2184)	0.3	Upper Rio Grande (7M16)	0.3
	Pagosa Springs (05-6258)	0.5	Pool Table Mountain (6M14)	0.2
	Hermit (05-3951)	0.1		
Clarks Fork Yellowstone	Mystic Lake (24-5961)	0.5	Northeast Entrance (10D7)	0.8
			Sylvan Pass (10E5)	0.6
Salt	Chino Valley (02-1654)	0.6	Beaver Head (9S6)	0.4
			Frisco Divide (8S1)	0.5
Verde	Chino Valley (02-1654)	0.3	Milk Ranch (9R1)	0.5
Tonto	Chino Valley (02-1654)	1.7	Milk Ranch (9R1)	1.6
			Chalendar (12P1)	2.1

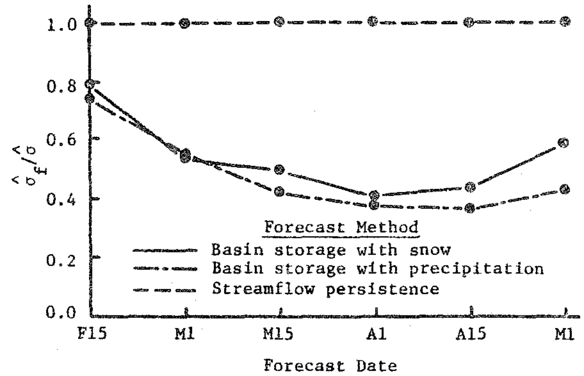
[a] station numbers are U.S. Environmental Data Service and Soil Conservation Service identifiers, respectively.

[b] station distance from basin centroid divided by path length of stream main stem.

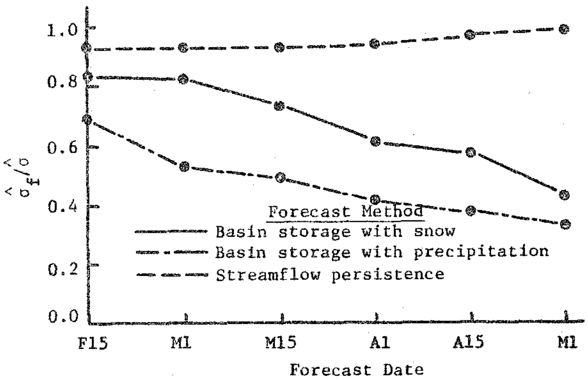
The RMSE reduction factors ($\hat{\sigma}_f/\hat{\sigma}$, with $\hat{\sigma}$ the estimated forecast period standard deviation) obtained in application of the three methods to each of the basins are given in Figures 1a-h. An RMSE reduction factor of 1.0, corresponding to a constant forecast of the mean, was assigned where the estimated factor exceeded 1.0. The earlier forecast dates for the Arizona basins reflect the earlier runoff experienced in these basins. The results show that the preferred forecast method varies between basins. For the Cascade and Rocky Mountain drainages, almost no reduction in RMSE is achievable through forecasting on the basis of streamflow persistence alone. This reflects the facts that most winter precipitation is stored in the snowpack and does not contribute to runoff until late in the year, and that base flow represents a small fraction of total runoff during the snowmelt period. Consequently, the correlation between fall/winter and spring/summer runoff is low for these basins. The three Arizona basins, however, display much larger correlations, reflecting more substantial runoff during the winter (which, in turn, indexes the amount of water



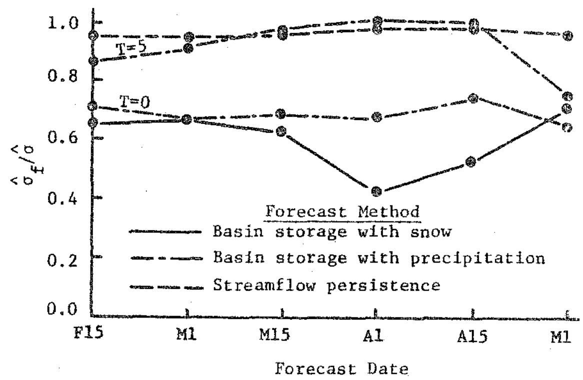
a. Cedar River



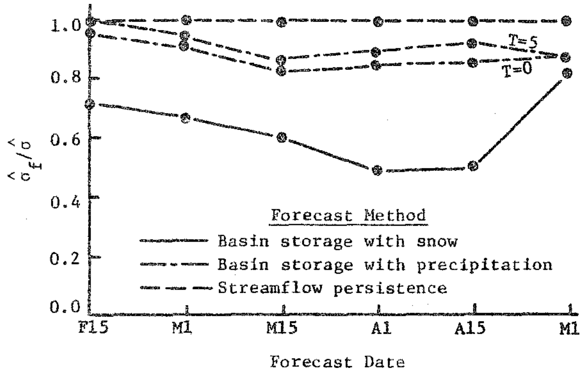
b. American River



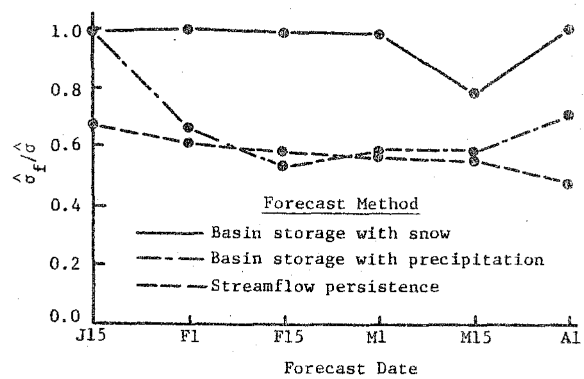
c. Stehekin River



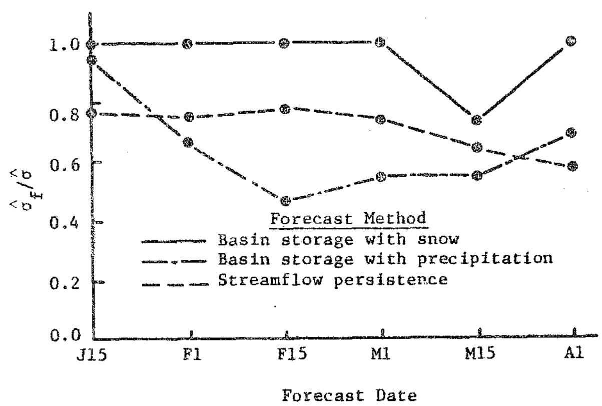
d. Rio Grande River



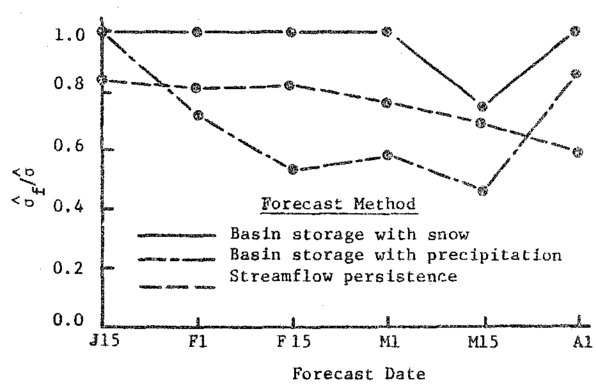
e. Clark, Fork Yellowstone River



f. Salt River



g. Verde River



h. Tonto Creek

Figures 1a-h. Estimated Forecast Root Mean Square Error Reduction Factors for Forecasts Through July 31 (Figures 1a-e) or May 30 (Figures 1f-h); T = test season length (Figures 1d-e).

stored which will occur as subsequent runoff) and the larger contribution of base flow. For these basins, the forecasts using snow course data are of very low accuracy, primarily because the snow accumulations at the snow courses tend to be quite small and are zero for a number of years of the record. Consequently, the linear relationship assumed between basin storage and snow depth is inappropriate, and the precipitation method performs better, although its accuracy is also undoubtably limited by the assumption of linearity. For both the Rocky Mountain drainages the highest accuracy was achieved using the storage accounting model with the snow course index. For these basins, forecast accuracy using the precipitation method is sensitive to test season length. Substantial differences were observed when forecast accuracy obtained using the best test season length for the calibration period was compared with the best aposteriori test season length. In practice, the test season length could be adjusted periodically, as are the other model parameters; this might result in RMSE reduction intermediate between those plotted. In either case, however, better performance was achieved using snow course data. The results for the Washington streams are similar to those reported by Lettenmaier (1978); the snow course method was preferred for the Cedar, whereas the precipitation method provided the best results for the Stehekin and American Rivers.

Figure 2 summarizes the results for each basin, where the best (smallest) RMSE reduction factor for each forecast date is plotted. Despite the large differences in basin climatological characteristics and the type of forecast used, the RMSE reduction factors are quite similar between the basins. The primary difference appears to be a shift corresponding to the length and timing of the snowmelt season. The significance of the similarities is emphasized by Figure 3 in which the best forecast RMSE is normalized by the mean, rather than the standard deviation of the forecast period runoff. When summarized in this manner, differences between the basin become apparent, insofar as the RMSE reductions are relatively similar, Figure 3 largely reflects differences in streamflow variability (coefficient of variation) between the basins. Consequently, it is seen that the fractional forecast error ranges from well under 10% of the mean flow for the Stehekin to almost 100% for the Arizona drainages, depending on the forecast date. The implications of these variations are discussed below.

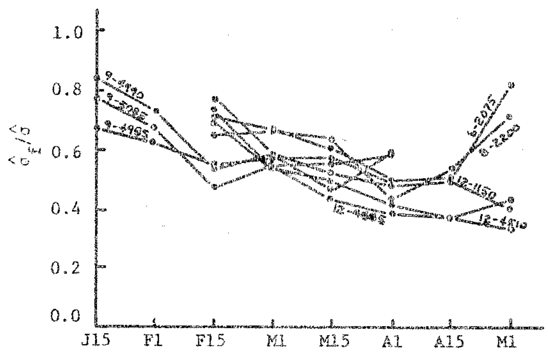


Figure 2. RMSE Reduction Factors for Best Forecasts from Figures 1a-h.

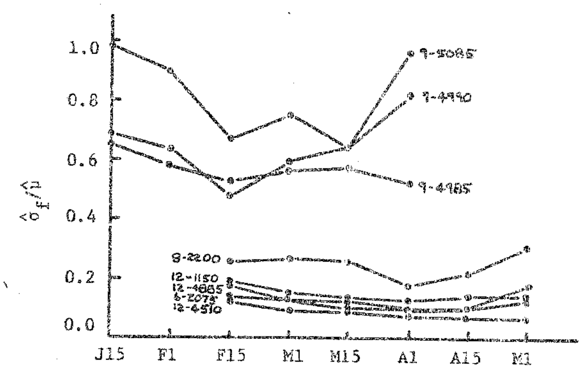


Figure 3. Forecast RMSE as a Fraction of Mean Flow for Best Forecasts from Figures 1a-h.

Discussion

At least three types of error contribute to inaccuracy in forecast of seasonal runoff volumes: model error, parameter error, and input data error. Model error in the storage accounting model arises in part from the assumption of linearity, i.e., forecasted runoff increases in direct proportion to rainfall or snow accumulation, regardless of soil moisture, groundwater level, etc. Parameter error arises because model parameters are estimated from a finite record length. The variance of parameter errors, when parameters are estimated by regression, is inversely proportional to record length; the record lengths used here are short in a statistical sense, (each year is, in essence, a single sample) so parameter error is potentially significant. Input data error results from variable catch efficiency of precipitation gages, spatial variation in snow course depths, and from the error inherent in characterizing storage by a small number of point measurements, for instance, substantially different basin storage could be associated with the same snow water equivalent at a point.

Efforts have been made to reduce the magnitude of each of these sources of error. Model error has been addressed through development of physically-based models which attempt to describe the distribution of snow and subsurface water storage and the associated accumulation and ablation process throughout the watershed. Twedt, et.al. (1978) describe application of one such model to California and Nevada watersheds. Parameter error in regression models can only be reduced through extension of record lengths. In physical models, parameters are estimated from hourly or daily streamflow records; features of subsets of these records can be used to estimate some of the model parameters (e.g., recession coefficients) and calibration record lengths are usually reduced by comparison with regression models. The difficulty with such models, however, is that while record length dependence is reduced, the number of parameters to be estimated may be quite large and it is sometimes difficult to isolate the effect of each of the variables. Further, although efforts have been made to implement automated calibration schemes, the necessity for user familiarity with the model cannot be avoided altogether, and a certain amount of art is inherent in the calibration procedure. Finally, input data error is present regardless of which model form is used. It is difficult to determine the characteristics which define a good precipitation station or snow course. Certainly, considerations such as design of data collection equipment and small-scale location (e.g., effects of trees, buildings, etc.) are important, however it is difficult, on the basis of these considerations alone, to explain variations in forecast accuracy among potential stations. Station location on a large scale apparently also plays a role; however, until the recent past potential station sites were restricted by logistical limitations and so the issue of station siting has generally not been given much consideration. However, with development of improved remote data collection and transmission facilities, the station location problem appears to warrant more emphasis, in short, data input error may to some substantial extent be amenable to control.

The results summarized in Figures 2 and 3 may prove useful in assessing the potential for forecast accuracy improvements. These results suggest that forecast accuracy is approximately proportional to streamflow variability (i.e., coefficient of variation). There does not appear, however, to be any inherent property of the snowmelt runoff process which would lead to this result, for instance it should not necessarily be more difficult to forecast extreme years than average years so long as the process is properly modeled. Difficulties of forecasting of extremes have, however, been cited as one of the limitations of regression models; it appears that physical models, the primary competitor to regression, may have the most to offer in the more variable regimes. On the other hand, there may be relatively little to be gained in use of physical models where natural variability is low, such as in the Northwest. Although conceptual models are gaining acceptance in practice, there appear to have been few attempts to establish the operational accuracy of such models in seasonal streamflow forecasting; consequently, the existence of relationships such as that discussed above remains hypothetical. We hope, however, that the results presented here may serve as a baseline for future comparisons of alternate forecasting methods.

REFERENCES

Burges, S.J. and A.E. Johnson, "Probabilistic Short-Term River Yield Forecasts", Journal of the Irrigation and Drainage Division, ASCE, Vol. 99, No. IR2, June, 1973, pp. 143-155.

Lettenmaier, D.P., "Worth of Snow Course Data in Forecasting Snowmelt Runoff", Proceedings, Workship/Meeting on Modeling of Snow Cover Runoff, Hanover, N.H., September, 1975.

Lettenmaier, D.P. and T.J. Waddle, "Forecasting Seasonal Snowmelt Runoff: A Summary of Experience with Two Models Applied to Three Cascade Mountain, Washington Drainages", Technical Report No. 59, C.W. Harris Hydraulics Laboratory, Department of Civil Engineering, University of Washington, November, 1978.

Tangborn, W.V., "Application of a New Hydrometeorological Streamflow Prediction Model", Proceedings, Western Snow Conference, Albuquerque, N.M., 1977.

Twedt, T.W., R.J.C. Burnash and R.L. Ferral, "Extended Streamflow Prediction During the California Drought", Proceedings, Western Snow Conference, Otter Crest, Oregon, 1978.