Basically, regression analysis is used to evaluate random sampling, where parameters or fiducial quantities can be pre-determined in terms of the information to be determined from the analysis of the random samples. Sampling long established snow courses is not random sampling. Attempting to analyze the results of Grid Sampling by combining regression and correlation analysis, over-complicates the analysis and leaves variables that have to be fitted into the calculation as assumed factors or coefficients. Correlation analysis in this area has provided an accurate forecast with the minimum of expense. Standard deviation and probable errors were calculated for each course only as a check against the correlation analysis.

---

## PRELIMINARY REPORT ON EVALUATING THE UTILITY OF WATER SUPPLY FORECASTS

by

Max A. Kohler 1/

### INTRODUCTION

At last year's meeting of the Western Snow Conference in Bozeman, Mr. R. A. Work presented a paper[1] entitled "Basic Data Characteristics in Relation to Runoff Forecast Accuracy." In essence, that paper constitutes a comparative verification of the Soil Conservation Service and Weather Bureau water supply forecasts for the period 1944-57. The results of the verification were no surprise to us, but we feel Mr. Work's interpretation of the results was not entirely realistic.

The Weather Bureau also verifies the forecasts of both agencies, year by year, as the streamflow data become available. Verifying all 1947-57 common forecasts, we find that the overall average errors of the Soil Conservation Service and Weather Bureau April 1st forecasts are 27 and 29 percent, respectively, when computed in the manner described by Mr. Work. These figures, 27 and 29, are to be compared with the values 24 and 27 cited in the Work-Beaumont paper. Our analysis covers a period beginning 3 years later; it contains more 1956 and 1957 data; and we have excluded those cases where our forecast period terminated prior to the end of the water year. We do not believe, however, that these or Mr. Work's results substantiate the implied or stated conclusions in the referenced paper to the effect that:

1. Snow surveys constitute "... a more precise method of sampling the greatest factor in streamflow production..."
2. The Soil Conservation Service forecast service is of more value to the water user.
3. Forecasts of river flows should be based on "... data secured as nearly as possible at the water sources..."

It seems obvious that the index value of precipitation observations would increase as the location of the gages approaches the source areas, and snow surveys certainly must be made at elevations where snow accumulates. But even these logical contentions cannot be proven by unrestricted verification of published forecasts.

### VERIFICATION METHODS

The method of verification selected by Work and Beaumont utilizes errors of individual forecasts expressed as percent of the observed flow. As pointed out by Work and Beaumont in their paper, this becomes an unrealistic measure of forecast value in extremely dry years. To illustrate this point, consider the Weather Bureau forecast of April 1, 1950, for Embudo Creek at Dixon, New Mexico. The error in our water-year forecast was 960 percent of the April-September flow (incorrectly given as 1033 percent in the Work-Beaumont paper). There are about 6,000 acres of irrigated land above the station and largely because of the less-than-record minimum summer precipitation the April-September flow was only 3,050 acre-feet. A little arithmetic will show that a decrease of six inches in the irrigation diversions would have decreased our error to 430 percent. More important, however, is the fact that an increased diversion of just over six inches would have resulted in sheer calamity — with zero observed flow, the over-all average error of all Weather Bureau and Soil Conservation Service forecasts for all time to come would have been infinite.

---

1/ Chief Research Hydrologist, U. S. Weather Bureau, Washington, D. C.

Methods of deriving Weather Bureau seasonal forecasts are discussed later, but it is apparent that subtracting the flow-to-date from the water-year forecast magnifies the error by telescoping it into a short period. In low years this makes our forecasts particularly vulnerable to large errors (in percent of observed flow), and results in inordinate weighting of individual forecasts. Thus, one bad forecast (1033 percent error) out of 1225 increased our average error by almost one percent, whereas it would take more than 30 perfect forecasts to effect a comparable reduction. Of the 1211 cases in our verification, each agency had 22 errors exceeding 150 percent. If these cases are eliminated from the computations, the average errors of the Soil Conservation Service and Weather Bureau April 1st forecasts becomes 22.6 percent and 22.9 percent, respectively — a difference of only 0.3 percent.

Another deficiency of verification based on errors in percent of observed flow is that variability of flow is not taken into account. If one had forecast 44,186,000 acre-feet (i.e., mean flow) as the April-September flow of the Columbia River at Birchbank for each year 1947-57, the average error would have been 8.9 percent. Yet, by any reasonable standards, such a forecast has no skill and little if any value. Average errors of Weather Bureau and Soil Conservation Service forecasts for this station are 7.6 and 8.3 percent, respectively. In contrast, forecasts of mean 1947-57 flow each year on the Rio Grande near Del Norte, Colorado, would result in an average error of 53.9 percent of observed runoff (compared to Weather Bureau and Soil Conservation Service errors of 18.6 and 22.0 percent, respectively). Although forecast errors for Del Norte are more than twice those for Birchbank, any realistic verification procedure would show that forecast skill is much higher for Del Norte.

There are many verification systems, none of which is entirely satisfactory for testing both forecast skill and value of the forecast to the user. The skill of a forecast method can only be demonstrated by taking into account the variability of the factor being predicted. Figure 1 shows a method used for judging the adequacy of hydrologic forecasts in the USSR[2]. The upper curve shows the error distribution of a climatological forecast (mean flow), while the lower curves show the error distributions for the forecasts issued. In this illustration, the errors are expressed in percent of range, but they could equally well have been given in terms of normal or standard deviation. The ratio of the areas under these two curves is an excellent measure of forecast skill. Depending upon the user's needs, however, it may or may not accurately reflect the value of the forecasts.

Still another aspect of verification which requires further consideration is the method of deriving the April-September forecast from the water-year forecast. Most of the Weather Bureau forecast procedures were derived so as to minimize the errors in predicted water-year streamflow. Without entering into a discussion of the relative merits of water-year, residual, and seasonal forecasts, the user requiring a forecast of April-September flow can interpret the water-year forecast in several different ways. As of April 1st, and knowing the flow to date, he can make a direct subtraction. In this case, the method of verification selected by Work and Beaumont would be applicable. As of any forecast date, however, he can elect to assume that the flow in percent of normal for any part of the year (such as April-September) will be the same as for the water year.

The relative reliability of the two methods depends upon many factors, but it is clear that:

1. Subtracting flow-to-date, thus telescoping the entire forecast error into the remaining period, yields answers of increasingly lower reliability as the end of the water year is approached. Some of the April-September forecasts derived in this manner are even beyond the realm of possibility.
2. With perfect water-year forecasts, the subtraction method yields perfect residual forecasts, whereas the constant percent method will be in error unless the relative distribution of flow between the two parts of the year is normal.
3. The percent method yields the highest relative verification on basins for which the ratio of residual to annual flow is stable, and for which the water-year forecast errors are large but not biased.

We have not verified all Weather Bureau forecasts by the percent method, but analysis for a limited number of points indicates that this method might measurably reduce the over-all average error. The residual forecasts now being published by the Weather Bureau in increasing numbers are usually made with knowledge of the flow to date. The water-year forecast and the residual forecast (derived by subtraction) are compared to the mean relation between the flow of the two periods, and an adjustment is made to both forecasts when this appears appropriate. This procedure would have substantially reduced many of the large errors in past years.

## INTERPRETATION OF VERIFICATION RESULTS

As stated previously, our analysis, using the method of verification proposed by Work and Beaumont, also indicates that the Soil Conservation Service April 1st forecasts have been slightly more accurate than those of the Weather Bureau (average errors 27 and 29 percent). Without attempting to refute the significance of this difference in forecast accuracy, it will be seen from figure 2 that most of the difference came about in only two years (1950 and 1951). These two years were extremely dry in the Rio Grande. We had six forecasts averaging 500 percent in error when telescoped into the dry summer period. If these six forecasts are omitted from our west-wide verification for the two years, the over-all average error would drop about 14 percent. Figure 2 is based on the data presented by Work and Beaumont (in table III), except that additional comparative data are now available for 1956 and 1957. As cited by Work and Beaumont, there has been considerable bias in the Weather Bureau forecasts — particularly in the early years, our forecasts were too low on the average. Steps have been taken to overcome this deficiency in our procedures which should result in improved forecasts.

Agreeing, for the time being, that the average error of the published Soil Conservation Service April 1st forecasts is as much as 2 percent less, how can this result be interpreted? For example, does it demonstrate:

1. That snow surveys are a better index of streamflow than is winter precipitation? And that snow surveys are, in general, more useful than precipitation data?
2. That a water-supply forecast service based on snow surveys is more valuable?
3. That one set of forecast procedures is better than the other?

While it is deemed important that we find answers to these questions, none are forthcoming from the analysis presented by Work and Beaumont.

Snow surveys vs winter precipitation

Although snowmelt and drifting can be important modifying factors, it is customary to assume that the April 1st snow survey constitutes an independent measure of winter precipitation. To the extent that this view is realistic, the two types of data are complementary and the use of both should provide the best forecasts[3, 4]. Which constitutes the better index can be determined by statistically weighting the two, basin by basin. It should be no surprise that such analysis shows neither type of data is consistently better than the other. The results for a particular basin depend on numerous factors, such as:

1. number of snow courses and precipitation stations in and adjacent to the basin;
2. location of the courses and stations;
3. reliability and consistency of the observed data; and
4. climate of the area.

We can be certain that some number of precipitation stations, properly located, will provide a better index of streamflow than a single snow course, and vice versa. In this respect, forecasts made from precipitation data only are at a decided disadvantage over much of the West. In fact, there are numerous headwater basins where snow courses outnumber precipitation stations three or four to one.

To be of optimum value, either a snow course or a precipitation station must be in an area of heavy streamflow production, but only the former need necessarily be located above the normal snow-line. The relative value of snow surveys diminishes rapidly as we proceed to the south and to low-elevation basins which may have little or no snow as of April 1st. Both types of data are subject to errors and inconsistencies due to changes in site exposure, observational procedures, etc. These effects alone can entirely mask any superiority of one type of data over the other in a particular basin.

The very nature of the problem thus renders meaningless any bland statement that snow surveys constitute "a more precise method of sampling the greatest factor in streamflow production." Any comparative analysis must be carefully planned in accordance with clearly stated objectives if meaningful results are to be obtained. With this in mind, we might consider the purpose and objectives of a comparative study — snow surveys vs winter precipitation.

FIG 1 EFFECTIVENESS (OR SKILL) OF FORECASTS, RIO GRANDE NEAR DEL NORTE, COLORADO

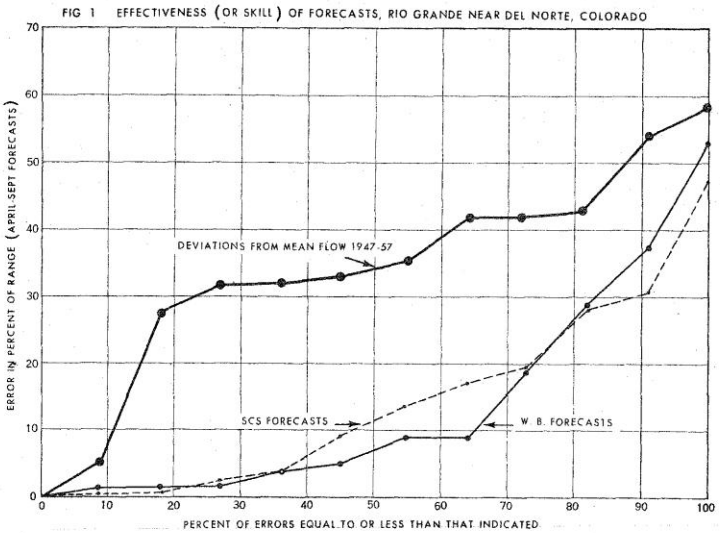
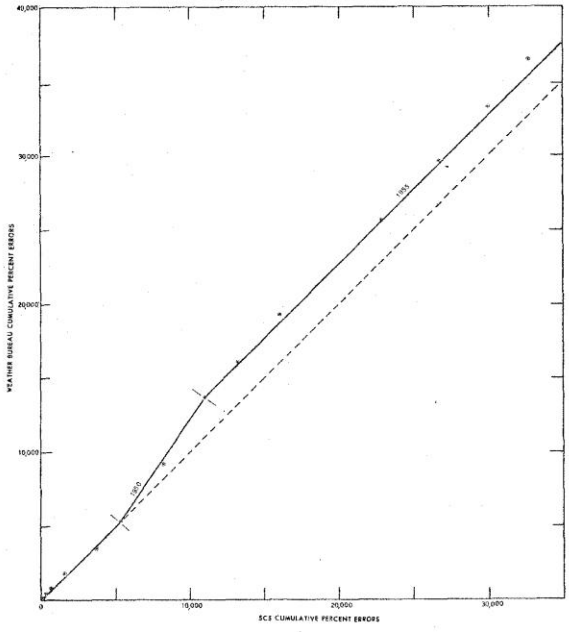



FIG 2 - DOUBLE - MASS CURVE OF WEATHER BUREAU AND SOIL CONSERVATION SERVICE APRIL 1st FORECAST ERRORS (1944-57)



By developing paired forecasting procedures — one on data from a snow course, one on data from an adjacent precipitation station, and otherwise identical — it should be feasible to delineate those basins where one or the other type of data is superior as of any selected forecast date. Even so, the results would depend upon an arbitrary beginning date for "winter precipitation." From a scientific point of view, such an analysis would be of interest and value.

Were the preparation of water-supply forecasts the only purpose to be served, interest would lie in determining the cost-to-benefit ratio of each type of data and of the two in combination. However, any realistic economic analysis of the problem must be based on the value of the data for all purposes, including water-supply forecasting. I believe all those interested in the water resources of the West agree that the existing precipitation network is inadequate, particularly at the higher elevations, and that moderate expansion of the network is economically justified over and above any value the data would have in water-supply forecasting. In the absence of precipitation data, snow surveys are of great climatological value, but precipitation data are to be preferred. Snow surveys made at the onset of heavy melting are very helpful in the preparation of flood forecasts, but scheduled monthly surveys are not particularly suited to this purpose.

These points are raised, not to discredit snow surveys but, rather, to demonstrate the complex nature of the problem. To further complicate the picture, we have appearing on the horizon other types of data still to be evaluated, such as areal observation of snow stakes, radio-isotope water-equivalent observations and radar integrations. It now appears that weather radar will eventually provide a means of evaluating areal precipitation in mountain regions far transcending present techniques. Used in conjunction with isotope and storage gages geared to radar beacons, we will know as of any moment and for any basin the amount of precipitation which has occurred and how much of it remains in the snow pack. With a high-speed communications system embodying modern electronic computers, up-to-the-minute forecasts could be made available for all users. Certainly, it would be unwise to discontinue one snow course without due consideration of the broader aspects of the problem, and the same can be said of any proposed expansion[5].

Value of a forecast service

Returning to the type of verification used by Work and Beaumont, figure 3 displays our results cited earlier. The two lower curves show the error distribution of forecasts published by the two agencies, while the uppermost curve is based on a "hypothetical forecast" of 15-year average flow for each year and basin. Areas under the three curves yield skill scores of 46 and 49 percent for the Weather Bureau and Soil Conservation Service forecasts, respectively.

Without questioning the statistical significance of the difference in accuracy, it remains to be shown that the improved accuracy is worth the additional cost. Also, one might ask if any substantial part of the operators using the forecasts are interested only in an April 1st prediction of the most likely flow.

It is readily demonstrated[6] that the efficient use of forecasts in most operational planning requires information other than (or in addition to) the most likely flow. Hydro-electric operations are often based on the minimum anticipated flow, flood control operations on the maximum anticipated flow, and many operators require a full range of possible flows with corresponding probabilities. Although not entirely successful, perhaps, we have nevertheless attempted to pattern our forecasts to meet these needs. So doing has admittedly led to confusion on the part of some recipients, but more and more favorable comments are being received.

Our techniques are entirely flexible with respect to forecast date and permit the issuance of a series of forecasts throughout the season which are consistent in the light of intervening weather. We find this to be an important aspect of our service since the key forecast for many users is much earlier than April 1st. Indeed, many users can or must revise plans from time to time in a manner entirely compatible with our forecast service. The verification of our early season forecasts is shown in figure 4. While the skill deteriorates as the time range of the forecast increases, even the January 1st forecast is of value.

The Soil Conservation Service has published some early season forecasts and the relative verification of the common cases for the Colorado, Rio Grande, and Great Basins (1947-57) is shown in the following table.

FIG 3 COMPARISON OF APRIL 1st WATER - SUPPLY FORECASTS PUBLISHED BY THE WEATHER BUREAU AND THE SOIL CONSERVATION SERVICE (1947-57)

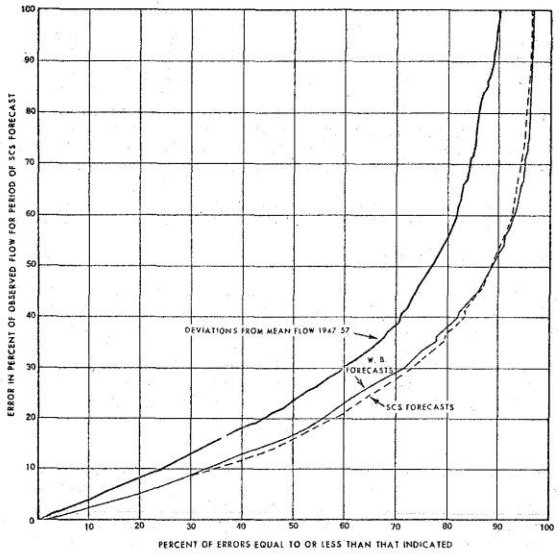
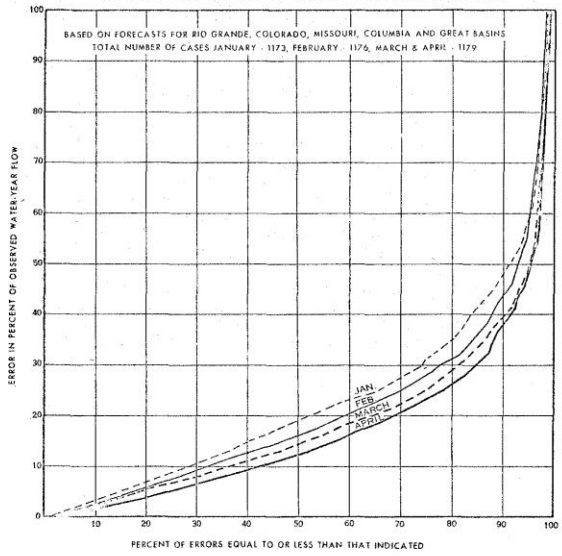



FIG 4 - COMPARISON OF WEATHER BUREAU WATER - SUPPLY FORECASTS ISSUED AS OF JANUARY 1st, FEBRUARY 1st, MARCH 1st & APRIL 1st (1947-57)



| Date of Forecast | Number of Cases | Average Error (% Obs. Flow) SCS | WB |
|---|---|---|---|
| February 1 | 41 | 49.1 | 43.5 |
| March 1 | 173 | 39.9 | 39.3 |
| April 1 | 513 | 34.2 | 36.4 |

As in the case of the west-wide verification, the Soil Conservation Service April 1st forecast errors were about 2 percent less, but Weather Bureau March 1st forecasts were slightly better than those of the Soil Conservation Service, and the February 1st much better. Perhaps more important is the fact that our coverage of the forecast points is complete even on January 1st.

Thus, we feel that the verification results of one forecast (made at the time when the index value of snow surveys is a maximum) is not particularly indicative of the value of our service.

Relative reliability of forecast procedures

It should be emphasized that verification of forecasts published over the past 10 or 15 years is in no manner indicative of the reliability of present-day forecast procedures because of numerous reasons. For example, the published values are often largely subjective in nature, at least to the extent of embodying adjustments to one or more objective "answers." Also, the method of arriving at a particular forecast has usually changed at least once during the period. Virtually all of the Weather Bureau forecast procedures have undergone a revision of some form since 1947, and it is assumed that this applies to the Soil Conservation Service as well.

## SUMMARY AND CONCLUSIONS

The primary objective of this paper has been to raise and discuss the issues involved in the evaluation of water-supply forecasts — to show that verification of the forecasts made on a selected date by an arbitrary method cannot provide a yardstick for comparing the value of two forecast services. Although some discussion of the relative value of snow survey and precipitation data could not be avoided under the circumstances, there is no intended implication that any of the hydrologic data now being collected is not of sufficient value to justify the cost. Snow survey and precipitation data are complementary, and both can be used to advantage.

The Weather Bureau has used and will continue to use such snow survey and other hydrologic data as are available to us at the time the forecasts are prepared. I am certain, on the other hand, that the forecasters in the Soil Conservation Service feel that winter precipitation at so-called "valley stations" is of some value, since these data are included in their forecast bulletins. When the courses in Arizona are barren of snow, precipitation is a far better index of moisture storage in the basin than a few point soil-moisture observations made by any of the available techniques.

Although the conclusions to be drawn from this paper are largely negative in nature, principally because time was not available to make conclusive studies, constructive suggestions have been made concerning certain types of analyses. Moreover, it is obvious that little is to be gained by comparative studies unless they are planned and executed with inter-agency support. One of the stated objectives of the erstwhile Soil Conservation Service-Weather Bureau cooperative program was the standardization of forecast verification methods. The program was formally discontinued by the Soil Conservation Service before this and other objectives were achieved.

In closing, it should be pointed out that forecasts for California have been excluded in the verification analysis presented here, in conformance with the Work-Beaumont paper, since the Soil Conservation Service does not issue forecasts for that State. Those forecasts are prepared by the State of California Department of Water Resources and the Weather Bureau, and are fully coordinated prior to publication.

## ILLUSTRATIONS



## REFERENCES

(1) WORK, R. A., and R. T. BEAUMONT, "Basic Data Characteristics in Relation to Runoff Forecast Accuracy," Proc. Western Snow Conference, pp. 45-53, 1958.
(2) LIGHT, P., and M. A. KOHLER, "Forecasting Seasonal Runoff by Statistical Methods," Trans. American Geophysical Union, Part II, pp. 719-736, 1943.
(3) POLOS, A. J., "The Use of Precipitation and Snow Survey Data in Water Supply Forecasting," Proc. Western Snow Conference, pp. 30-35, 1953.
(4) HANNAFORD, J. F., C. G. WOLFE and R. W. Miller, "Graphical Method for Determination of Area-Elevation Weighting of Snow Course Data," Proc. Western Snow Conference, pp. 73-82, 1958.
(5) KOHLER, M. A., "Design of Hydrological Networks," Technical Note No. 25, World Meteorological Organization, 1958.
(6) BLANCHARD, F. B., "Operational Economy Through Applied Hydrology," Proc. Western Snow Conference, pp. 35-48, 1955.