APPLICATION OF THE ELECTRONIC COMPUTER TO SEASONAL STREAMFLOW FORECASTING

By

Ashton R. Codd and Phillip E. Fames [1]

Application of the electronic computer to streamflow forecasting was immediately evident upon introduction to the newly installed IBM 650 Electronic Digital Computer at Montana State College in the Fall of 1958.

During the snow survey season, assistance was necessary to check forecast formula computations. Mr. Lynn F. Johnson of the Montana Agricultural Experiment Station, through the generosity of Professor O. W. Monson, came to the rescue. Seeing the series of forecasting equations, Mr. Johnson stated he would like to use the formulas as a programming exercise in a course he was taking on the IBM 650 computer. His work was very successful and his program was used during the 1959 snow survey season.

This pioneer adventure showed many advantages of the robot computer. A cooperative project was started between the Montana Agricultural Experiment Station and the Soil Conservation Service Snow Survey Activity of Montana, the purpose of this project being to develop seasonal streamflow forecasting equations for as many snow-fed streams as possible.

Through Montana State College, a multiple regression program was secured from the IBM Library in New York City. Work was started during the Summer of 1959. To date, approximately 75 stream gaging stations have been investigated and some 310 forecasting equations developed to forecast seasonal runoff at 66 of these stations.

<u>I. Selection of Variables.</u> The Program presented great possibilities for streamflow analysis with a capacity of thirty-three (33) variables. The selection of dependent variables was governed entirely by the purpose of the forecast. Several months during which the snowmelt runoff is predominant was studied. The total runoff in terms of acre feet from April first through September thirtieth was selected as the base period. Shorter periods, April-June and April-July were analyzed for irrigation and reservoir regulation forecasts. May first forecasts were developed using May-September, May- July and May-June periods as dependent variables. In river systems where junior water rights are shut off to supply senior water rights, the number of days the streamflow would exceed a specified flow, and the number of days for the flow to recede from this specified flow to a cut-off flow, were included among the dependent variables.

<u>II. Independent Variables.</u> Snow water equivalent (water content) for several dates of measurement (March, April, May) was used for all courses in and adjacent to the river basin being studied. Fall precipitation (October, November), low flow of the previous fall, spring temperature (March, April, May) and spring precipitation (April, May, June) were selected to complete the remainder of the allowable 33 variables.

In Montana snow survey course records are relatively short. Only those courses with eleven (11) or more years of record parallel with streamflow were considered in the analysis. This small number of samples made it necessary to <u>limit</u> the number of independent variables in the final equation to three or four, and in some cases, possibly five, in order to be assured that the resulting formula would withstand a significance test.

The size of simple correlation coefficients derived from Phase II, was used as a parameter to discard 1pw correlating independent variables. Where independent variables showed almost the same degree of correlation, only one was selected on the basis of accessibility and economy of neasurement. Be- cause of the extremely low correlation coefficients and the decision to use only data that would be available at the time of forecasting, spring precipitation, after several attempts, was eliminated.

Although the simple correlation is a fair parameter, further refinement was made by using the partial correlation coefficients obtained from Phase III of the Program. This parameter indicates the variation accounted for by each independent variable. The size of the multiple correlation coefficient was used as a further guide to determine the total variation accounted for by all variables used in the final regression equation.

[1] Hydraulic Engineers, U. S. Department of Agriculture, Soil Conservation Service, Bozeman, Montana.

During the selection of variables, different snow courses or dates of measurement were found more desirable for an April-July forecast than those used for the April-September forecast.

III. Increase in Accuracy. Comparison between forecasts issued by the Soil Conservation Service *for* Montana in 1959 from the old formulas and those based on computer developed equations show considerable increase in accuracy. The 1959 data were not used to obtain the new equations. Provisional streamflow data have been received for only 26 gaging stations in Montana, which *limits* a complete investigation of forecast accuracy. However, using 66 comparable forecasts at these 26 stations, the average error was decreased from 11.3 percent to 6.6 percent.

IV. Updating. As additional data *become* available, the computer will be used to update forecasting equations. It will be unnecessary to run *a* complete analysis of all variables, since 3 to 5 years additional data will not greatly change a relationship between dependent and independent variables. Only the 6 to 10 variables with the highest correlations will be analyzed, using the second run of Phase II and two runs of Phase IIIa and Phase IIIb with the added data. At present, it is contemplated that all equations will be updated at least every 3 years, and that a complete analysis of all variables will be run every 9 to 12 years. The time lapse for a complete analyses will depend largely upon the number of new variables that have acquired sufficient length of record.

V. Value of Residual Answers. The over-all relationship between snow course data, precipitation data, temperature data, base flow and the dependent variables (streamflow) has never before been available. Remarkable relationships are now evident and are being used to great advantage. The relation between one or several snow courses to another course can be used to estimate data net available at forecast time. A reliable formula can be quickly derived from the residual answers produced by Phase II of the Program. Estimating April first snow data from the March first data is accomplished by this type of equation.

An examination of the many answers derived from these computations has corrected our conception of snow survey data *to* runoff relationship. Time and qualified man-hours have not been available to produce the relationships we now possess through the use of the electronic computer.

VI. Economics. The cost of using an electronic computer appears to be quite large; however, when comparison is made with the number and quality of answers obtained and the speed with which work is accomplished, the electronic computer is very economical. It would take years to complete these computations by desk calculator, even though qualified personnel were used.

During the off-school season in Montana, a competent statistician was employed from 1950 through 1956. This employee developed multiple regression formulas to forecast streamflow of the major streams of the Missouri and Columbia Rivers in Montana. During these seven (7) years, 53 principal formulas, at a labor cost of $6,300, were developed. In 1956, 10 Columbia Basin formulas were updated, bringing the total to 63 formulas at an average cost of $100 per formula.

During the Fall of 1959, using an IBM 650 electronic computer, the same project was started. In the early stages of the project it was found to be as economical to run a complete analyses as it would be to update the original 53 formulas. A complete analyses would also facilitate obtaining a separate equation for each forecasting period. Water right information could also be included. Within three weeks, 310 formulas using 4 to 6 streamflow periods at 66 gaging stations were developed. Machine time, material and labor amounted to $2,600, or approximately $8.50 per equation. Since all machine operation was accomplished by Montana Experiment Station and Soil Conservation Service personnel, time required for tabulating data is also included.

In order to meet publication deadlines, time to compute forecasts on the first of March, April and May necessitates increased man hours. To compute and check the original 53 formulas by the desk calculator method requires 8 to 12 hours. In the early winter months of 1959 an IBM program was written for these 53 equations. This program requires three (3) minutes machine time. The 310 new formulas presented an even greater problem of time-consuming computations on a desk calculator. Since obtaining more experience, a new program that will compute the 310 new formulas in two and one-half (2i) minutes has been written, thereby saving many man hours at this critical time.

Conclusion

The electronic computer has provided a rapid and economical means *of* developing, updating, and computing streamflow forecasts in Montana. It has made possible the investigation of many variables that otherwise would not have been attempted. As the analysis of as many variables as possible is desirable, conclusions drawn from the analysis of Montana streams are:

Each drainage basin has different runoff characteristics and must be analyzed independently.

Factors that correlate well in one basin might not be significant in an adjacent basin. In

all cases, snow water content correlated well with seasonal runoff.

In some basins, March first water content was the best indicator of the runoff; in other basins, April first or May first water content correlated best.

In some drainages, October and/or November precipitation and low streamflow correlated well; in other basins this variable showed no usable relationship.