At the present time, but not including California, where the snow surveys are an activity of the State's Department of Water Resources, there are 377 Soil Conservation Districts in the West served by the Federal-State Cooperative snow survey network. There are over 10 million acres in Soil Conservation Districts directly served or benefited by snow survey data, and including California something over 20 million acres affected directly or indirectly by the snow survey information.

Farmers and ranchers have to be practical men to keep their operations on a paying basis. We therefore believe that measurements of the snow pack near the very tops of these mountains where the water is produced is a proven basis for the most accurate and practical evaluation that can be made of each season's water supply.

The information, as carried in releases to Soil Conservation Districts, gives the farm and ranch operators the advantage of the basic data of the snow pack itself for their own interpretation as well as carrying the interpretation of the technician forecasting the season in terms of acre-feet for a given seasonal period.

The highly cooperative nature of Soil Conservation District operations, in my opinion, provides an ideal unit in the West for using water supply forecasts for the maximum conservation of soil and water.

AN APPROACH TO FORECASTING THE SPRING RUN-OFF IN QUEBEC

By

George S. Cavadias[1]

1. The purpose of this paper is to describe an approach to the problem of forecasting the Spring run-off. Particular reference will be made to conditions in Quebec and the application of the method will be shown for the case of a large storage reservoir. This is the Gouin reservoir, with a capacity of 6.5 million acre-feet, that regulates the flow for the seven plants of the St. Maurice River hydro system with a total installation exceeding 2,000,000 horsepower (Ex. 1).

Many previous studies of the Spring run-off in Quebec have shown, that the run-off is not well correlated with the water equivalent of snow, as measured during the snow surveys. It appears, on the contrary, that one of the main factors determining the volume of the Spring run-off, is the precipitation during the freshet period, a factor, unknown at the time of the preparation of the forecast.

It was attempted, therefore, to investigate thoroughly all pertinent variables, with the purpose of extracting the maximum amount of information from factors that are known at the date of the forecast.

2. The success of a correlation study depends greatly on the form in which the variables are introduced. I will therefore, describe briefly the preliminary work that led to the selection of the variables.

The first variable to be examined is the Spring run-off to Gouin. In a previous study of the same problem, (Ref. 4) the duration of the flood period was assumed to be known and the variable to be estimated was the total run-off to the Gouin Reservoir during the whole period. Subsequent preliminary studies have shown that a subdivision of the flood period would increase the accuracy of the prediction.

---

[1] George S. Cavadias, The Shawinigan Water and Power Company, and now with the Water Resources Branch, Department of Northern Affairs and National Resources, Montreal, P.Q., 1410 Stanley St., Room 1007. Paper presented on April 17, 1958 at the Joint meeting of the Western Snow Conference and the Upper Missouri River Water Forecast Committee, Montana State College, Bozeman, Montana.

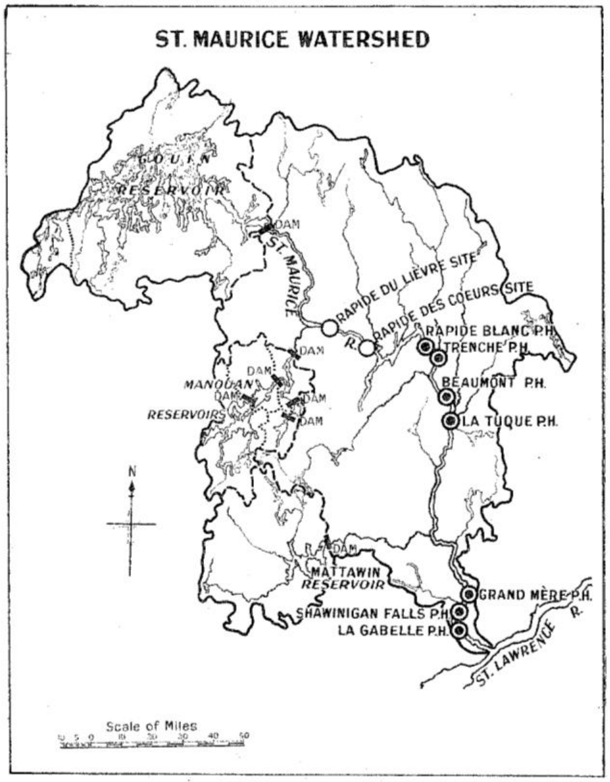


Exhibit 1 - Map of the St. Maurice Watershed.

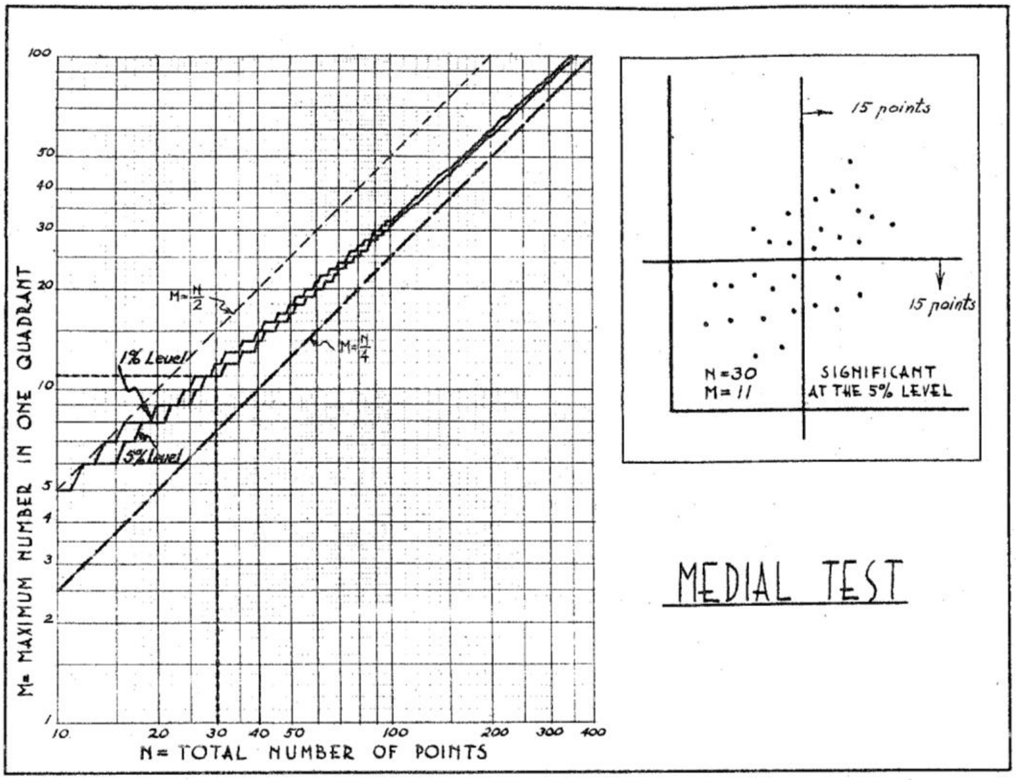


Exhibit 2 - Medial Test.

It was tried therefore, at first, to subdivide the interval between March 15 and July 31 into 15-day intervals. Several trial correlation studies based on this subdivision, resulted finally in the following set of nine periods which was used in the final regression computations:

1) March 16 to March 31
2) April 1 to April 15
3) April 16 to April 30
4) April 1 to April 30
5) May 1 to May 31
6) June 1 to June 15
7) June 16 to June 30
8) June 1 to June 30
9) July 1 to July 31

The variable to be estimated from the regression equations for all periods shown above, is the run-off to the Gouin reservoir during each period. (X_1).

The selection of the independent variables to be introduced in the correlation study is a difficult problem. If we include all the variables that could possibly have an influence on the volume of the spring flood, the degrees of freedom left for the correlation study will be unduly reduced. On the other hand, analytical regressions, which would permit the elimination of some variables, are time-consuming for preliminary work. In our study, a graphical method called the Medial Test (Ref. 1) was used for the preliminary screening of the variables (Ex. 2). According to this method, the degree of correlation between two variables is determined by counting the maximum number of points of the scatter diagram, in anyone quadrant formed by the two medial lines.

3. We now come to a more detailed examination of the variables introduced in the correlation. The water equivalent of snow (X_2) introduced in the study was based on measurements made at two snow courses in the watershed. The total drainage area of the reservoir is about 3600 square miles, and therefore, the coverage of the watershed by snow courses is obviously insufficient. The snow survey is conducted annually around March 15.

The next factor to be examined is the influence of the ground water conditions. Although a number of ground water wells have been installed in the watershed, their record is not satisfactory and therefore, an indirect measure of ground-water water conditions was used in the study: The February run-off of the Batiscan River, (X_3). This is a relatively small, uncontrolled river with topographical and hydrological conditions similar to the St. Maurice River. The variable X_3 is the same for all periods.

The values of the two variables, X_2 and X_3, already discussed, are known at the time of the preparation of the first forecast (March 16). They are the same for all regression equations. The remaining independent variables change from period to period and enable us to revise the forecast in keeping with conditions that occur after the initial forecast is made.

The next variable, (X_4), is the run-off during the period previous to the one considered. The medial test applied to the relationship between the run-offs of two consecutive periods showed that X_4 should be included in all regression equations.

The next variable, (X_5), is defined as the cumulative degree days above 32 degrees F. during the period under consideration. For the calculation of X_5, a Nomogram was prepared (Ex. 3) that was found more convenient than Snyder's chart, given on page 20 of "Applied Hydrology", by Linsley, Paulus and Kohler.

The next variable, (X_6), is the precipitation at Gouin during the period under consideration. The value of this variable is unknown at the time of the preparation of the forecast. In the application of this study made in 1957, X_6 was computed on the basis of Long Range Weather Forecasts. It is well known that the presently attainable accuracy of Long Range Weather Forecasts is not satisfactory. The uncertainty in the forecast introduced by this factor is partly offset, by the other variables considered in the study.

The independent variables X_2, X_3, X_4, X_5, X_6 already described, were found to be satisfactory for the prediction of the run-off during April and June.

Unfortunately, the multiple correlation coefficient of the run-off in May (which constitutes more than half of the total Spring run-off), based on these variables, is not significant. It was attempted, therefore, to introduce composite variables in the correlation study, based on the water and energy budgets.

The water budget variable (X_7) is defined as the water equivalent of snow at Gouin on March 15, plus the precipitation at Gouin from March first to the beginning of the period minus the run-off from March 16 to the beginning of the period. A large number of medial tests was used for the determination of the form in which this variable was introduced in the correlation. Variable X_7 has improved substantially, the correlation coefficient for the period May 1 to 31.

The last variable (X_8) is the cumulative degree days above 32 degrees F. at Gouin, from March first to the beginning of the period. This variable, obviously related to the amount of melting that has taken place before the period, has improved the multiple correlation coefficients for April and May.

All the variables described before were not significant in determinating the run-off of all periods. The decision which variables to introduce in every multiple regression equation was based on several medial tests, and hydrological considerations.

4. After the determination of the independent variables for each interval, the next step is the study of frequency functions. Strictly speaking, every forecast of a meteorological or hydrological variable should be expressed in terms of probabilities. A forecast based on a regression equation is not an exception to this rule and, therefore, it is not sufficient to compute the most probable value of the Spring run-off from the simple or multiple regression equations. The forecast is complete only if it is accompanied by a "Confidence interval", giving the probability that the actual volume of the run-off will differ from the forecast by a given amount. These confidence intervals are usually determined from the assumption that the deviations from the regression line are normally distributed with the same variability. If this is not the case, it was found that equal variability in each part of the regression line might also be achieved by use of a transformation to normality. In order to find suitable transformations, a study of the frequency functions of all variables was undertaken. This was achieved by postulating certain frequency functions and testing the goodness of fit by a statistical test. In this study three frequency functions were fitted to the observed data:-

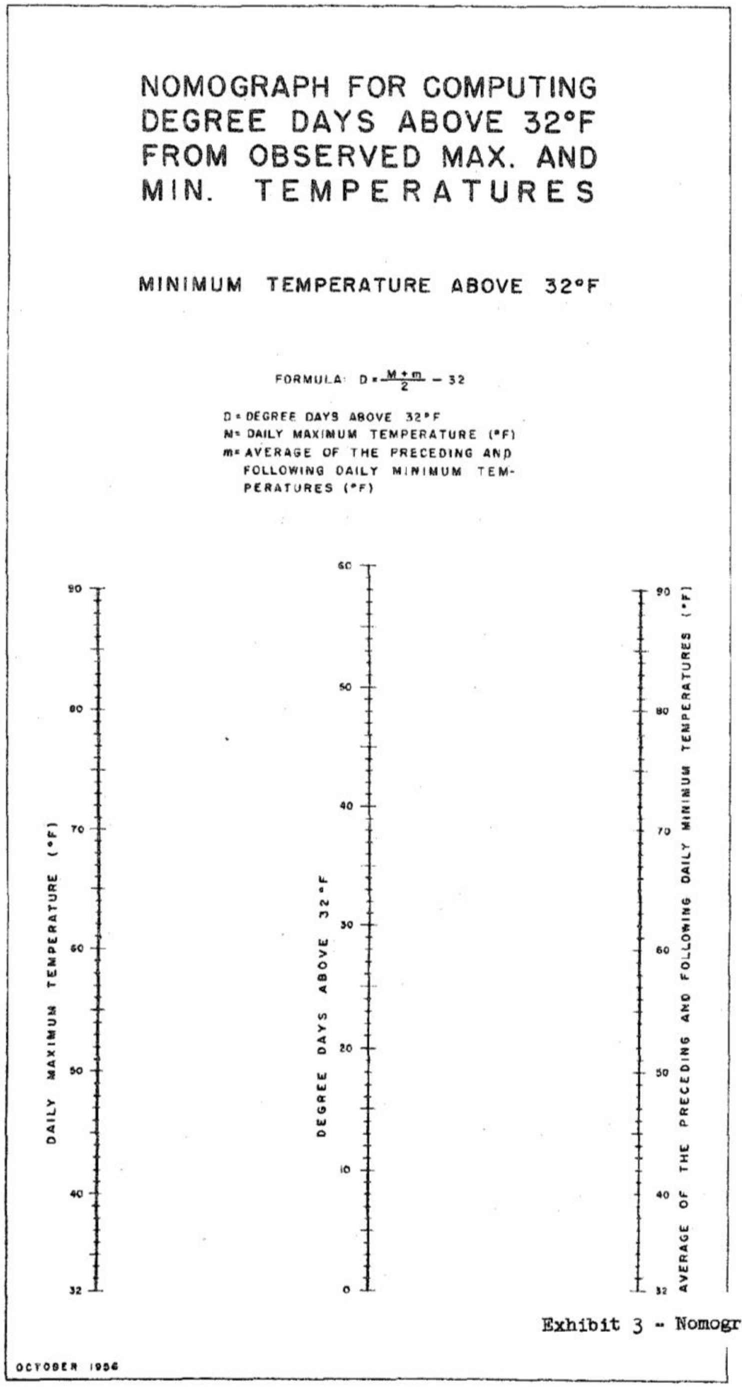
1) Normal frequency function.
2) Log-normal frequency function.
3) Incomplete gamma frequency function.

The reason for selecting these three functions is that the normal function is mathematically simple, the log-normal function is known to fit well many empirical distributions of hydrological variables and the incomplete gamma function was found to provide a good fit to precipitation series (Ref. 2).

The study of the frequency function of each variable is carried out as follows:-

1) First, a histogram of the variable is plotted (Ex. 4). In the case of the variable shown on the chart, (precipitation at Gouin from April 16 to 30) the histogram indicates a skewness in the frequency function.

2) The observed values of the variable are plotted on normal probability paper. The parameters of the frequency function are then estimated and the corresponding straight line drawn on the same probability paper. The goodness of fit of the theoretical function can be determined by means of the Kolmogorov-Smirnov test. (Ref. 3). This test is based on the maximum deviation of the empirical from the theoretical distribution, expressed in terms of probability. In this case, for example, the maximum deviation is 0.16.



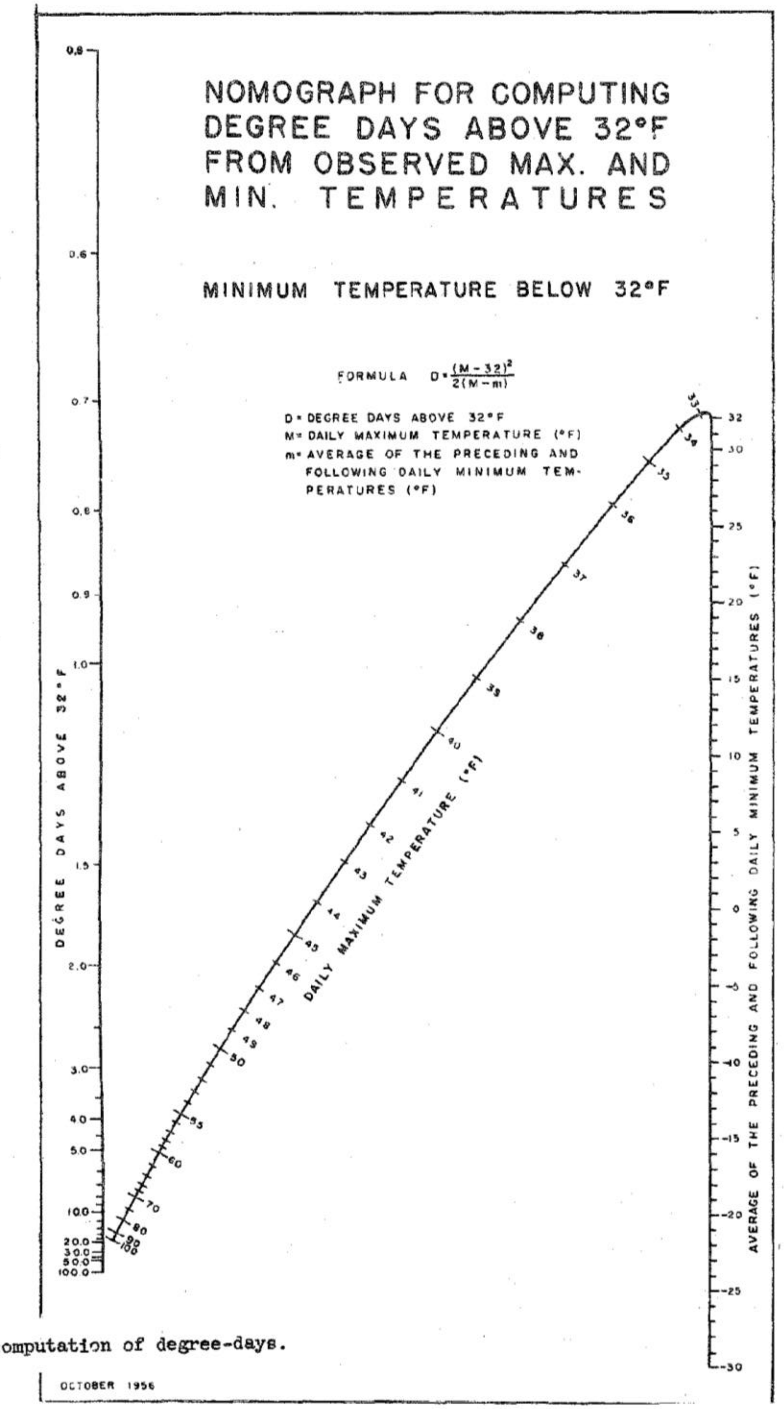


Exhibit 3 - Nomogram for the computation of degree-days.

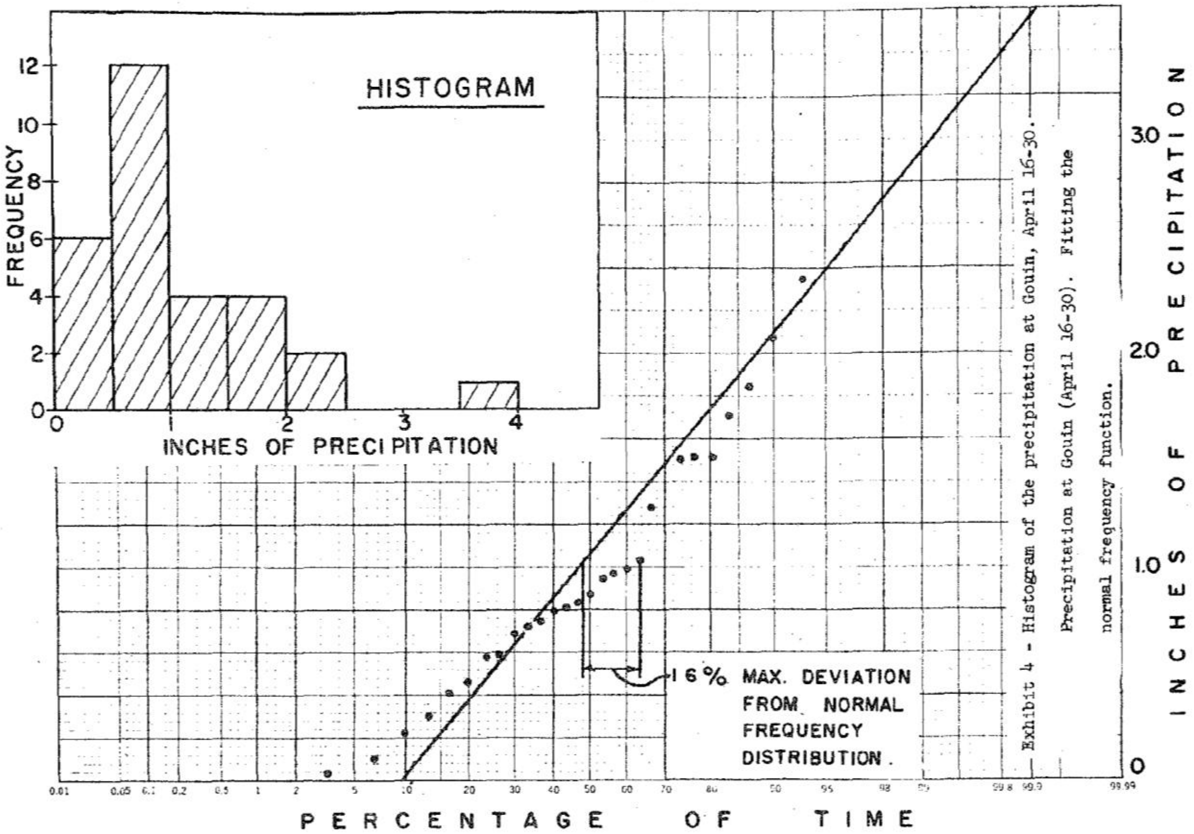


Exhibit 4 - Histogram of the precipitation at Gouin, April 16-30.
Precipitation at Gouin (April 16-30). Fitting the normal frequency function.

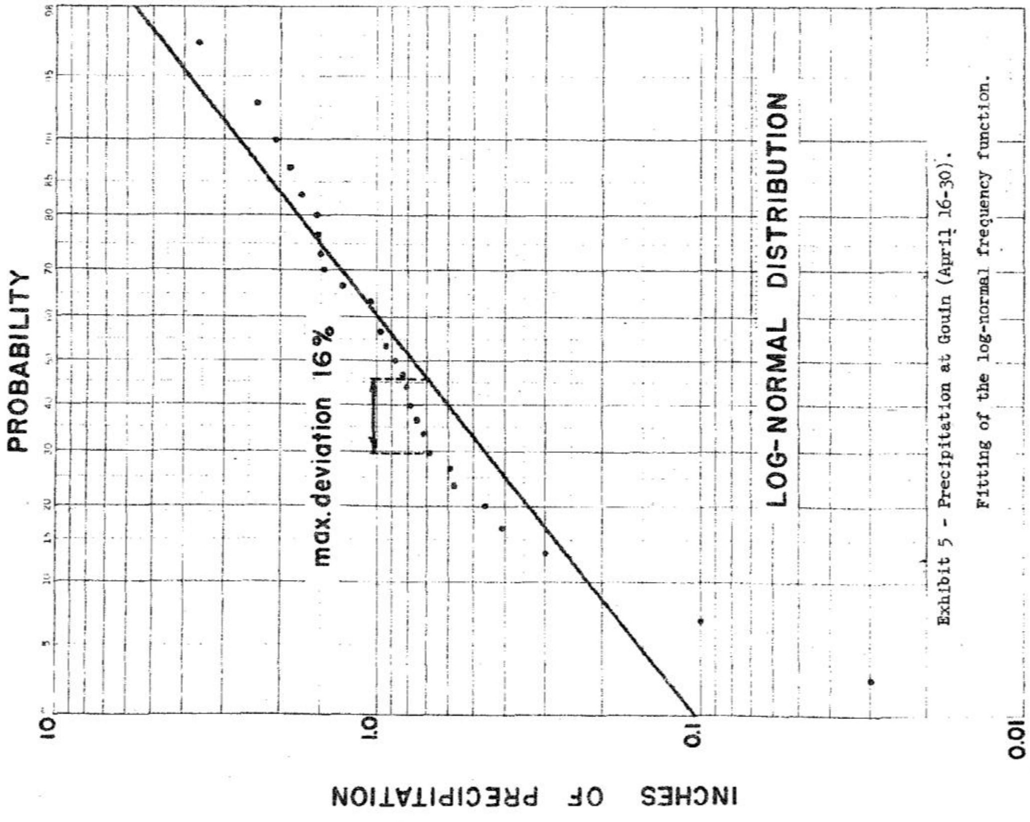
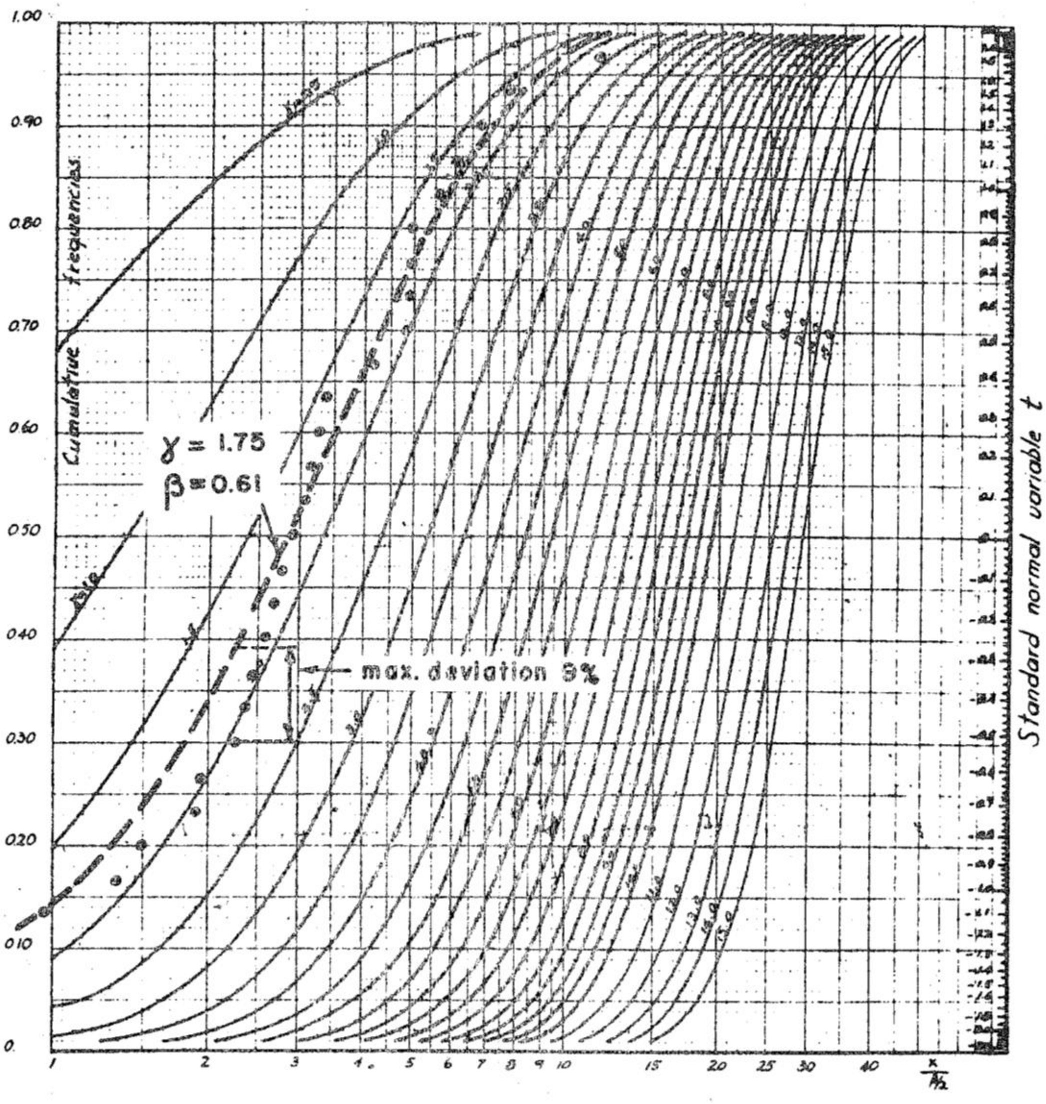
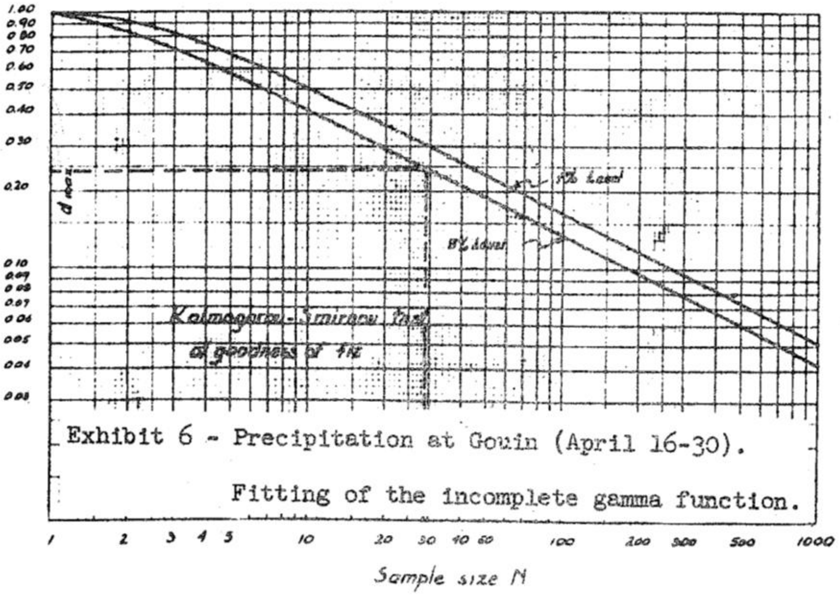


Exhibit 5 - Precipitation at Gouin (April 16-30).
Fitting of the log-normal frequency function.





INCOMPLETE
Γ-DISTRIBUTION

Exhibit 6 - Precipitation at Gouin (April 16-30).

Fitting of the incomplete gamma function.

FREQUENCY FUNCTIONS

VARIABLE	NUMBER OF FREQUENCY FUNCTIONS			
	NORMAL	LOG-NOR.	INCOMP Γ	TOTAL
RUN-OFF	3	9	4	16
WATER EQUIV. OF SNOW	1	1	1	1
DEGREE-DAYS	2	4	3	9
PRECIPITATION	1	1	8	10
WATER-BUDGET	2	2	1	4
EVAPORATION	1	2	1	4
TOTAL	9	19	16	44

Exhibit 7 - Results of the investigation of frequency functions.

REGRESSION EQUATIONS

PERIOD	COMPUTED EQUATIONS	EQUATIONS USED IN FORECAST	EQUATIONS BASED ON TEST OF SIGNIFICANCE
MARCH 16-31	$X_1=a_1+b_5X_5+b_6X_6+b_3X_3$ $X_1=a_1+b_5X_5+b_6X_6$	$X_1=a_1+b_5X_5-b_6X_6$	$X_1=a_1+b_5X_5+b_6X_6$
APRIL 1-15	$X_1=a_1+b_4X_4+b_5X_5+b_3X_3+b_6X_6+b_8X_8+b_2X_2$ $X_1=a_1+b_4X_4+b_5X_5+b_3X_3+b_6X_6+b_8X_8$ $X_1=a_1+b_4X_4+b_5X_5+b_3X_3+b_6X_6$ $X_1=a_1+b_4X_4+b_5X_5+b_3X_3$ $X_1=a_1+b_4X_4+b_5X_5$	$X_1=a_1+b_4X_4+b_5X_5+b_3X_3+b_6X_6+b_8X_8+b_2X_2$ The method of successive elimination of variables was applied to all subsequent equations	$X_1=a_1+b_5X_5+b_4X_4+b_8X_8+b_3X_3+b_6X_6$
APRIL 16-30	$X_1=a_1+b_4X_4+b_5X_5+b_6X_6+b_7X_7+b_8X_8$ $X_1=a_1+b_5X_5+b_4X_4+b_6X_6+b_3X_3+b_7X_7$	$X_1=a_1+b_5X_5+b_4X_4$	$X_1=a_1+b_5X_5+b_4X_4$
APRIL 1-30	$X_1=a_1+b_5X_5+b_4X_4+b_6X_6+b_7X_7+b_8X_8$ $X_1=a_1+b_5X_5+b_6X_6+b_4X_4+b_3X_3+b_7X_7$ $X_1=a_1+b_6X_6+b_5X_5+b_4X_4+b_8X_8+b_3X_3$	$X_1=a_1+b_5X_5+b_6X_6+b_4X_4+b_3X_3$	$X_1=a_1+b_5X_5+b_4X_4+b_3X_3$
MAY 1-31	$X_1=a_1+b_6X_6+b_4X_4+b_7X_7+b_2X_2+b_8X_8+b_3X_3$ $X_1=a_1+b_6X_6+b_5X_5+b_4X_4+b_3X_3+b_7X_7+b_8X_8$	$X_1=a_1+b_6X_6+b_5X_5+b_4X_4+b_3X_3+b_7X_7+b_8X_8$	$X_1=a_1+b_3X_3+b_7X_7+b_4X_4+b_6X_6+b_5X_5$
JUNE 1-15	$X_1=a_1+b_4X_4+b_6X_6$	$X_1=a_1+b_4X_4+b_6X_6$	$X_1=a_1-b_4X_4+b_6X_6$
JUNE 16-30	$X_1=a_1+b_4X_4+b_6X_6$	$X_1=a_1+b_4X_4+b_6X_6$	$X_1=a_1+b_4X_4+b_6X_6$
JUNE 1-30	$X_1=a_1+b_4X_4+b_6X_6$	$X_1=a_1+b_4X_4+b_6X_6$	$X_1=a_1+b_4X_4+b_6X_6$

Exhibit 8 - Tabulation of multiple regression equations.

3) The next step is to plot the same observed values of the precipitation at Gouin on log-probability paper (Ex. 5). The parameters of the log-normal distribution fitted to the observed points are then calculated and the maximum deviation determined. (0.16).

4) The last theoretical function investigated is the incomplete gamma function. It is described by H. C. S. Thom in the final report of the Advisory Committee on Weather Control. The details of fitting the incomplete gamma function to a given series of observed values are given by Thom and therefore, only the results of the fitting will be given here. Exhibit 6 shows the fitting of this frequency function to the precipitation at Gouin from April 16 to 30. The maximum deviation is, in this case, reduced to .09 and therefore, we assume that the variable in question follows the incomplete gamma frequency function with $\gamma = 1.75$ and $\beta = 0.61$. The lower part of the chart shows that the observed maximum deviation is significant at the 5 percent level.

This study of the theoretical frequency functions was carried out for all 44 variables introduced in the correlation studies. Exhibit 7 shows the results of this investigation. It is apparent that the variables introduced in the study are predominately non-normal and therefore, a transformation of the original data is indicated. In the case of the log-normal function, the transformation is simple: we introduce in the correlation the logarithms of the original data. In the case of the imcomplete gamma function, the equi-probability transformation described by Thom was used.

Because of the use of a transformation, the multiple regression equations give an estimate of the <u>normalized</u> value of the Spring run-off. This value is then re-transformed into inches.

The next Exhibit (Ex. 8) shows a tabulation of the variables introduced into the regression equations of each period. The program of the digital computer used for the calculation of the regression coefficients included a step-by-step deletion of the variables, as shown on the tabulation. Computer programs now available, delete automatically the variables which prove to be not significant for the correlation.

After the calculation of all multiple regression coefficients, the best equation is selected, on the basis of the corresponding unbiased multiple correlation coefficient. These best equations for each period are shown on the second column of the Exhibit. The third column of this Exhibit shows the sequence of significance of the variables, as determined from the multiple correlation studies. This sequence will be used for any future studies of the Spring run-off to the same reservoir.

The next Exhibit (Ex. 9) shows a comparison of the forecast prepared on April first, 1957, with the actual run-off. A 90% confidence interval was also determined for this forecast. This confidence interval was determined from the simple regression between the normalized actual and estimated X_1.

The final Exhibit (Ex. 10) shows the application of the method to 29 Spring run-offs of the Gouin reservoir. The points are not scattered symmetrically around the 1:1 line, and therefore, the transformation of the original variables is justified.

This graph does not show any striking improvements, over the results achieved in the 1955 study (Ref. 4). The main advantages of the method described in this paper are that:

1) The assumption regarding the total length of the flood period becomes unnecessary.

2) A better understanding of the nature of the variables that influence the volume of the Spring run-off, is achieved.

LIST OF REFERENCES

(1) "Associated Measurements" M. H. Quenouille, Butterworth's Scientific Publications, London, 1952.
(2) "Statistical Methods in Evaluating Cloud-Seeding Operations". H.C.S. Thom. Final Report of the Advisory Committee on Weather Control, Washington, 1957.
(3) "The Kolmogorov - Smirnov test of goodness of fit". Frank Hassey. Journal of the American Statistical Association, March 1951.
(4) "Reappraisal of Snow melt as a factor in Quebec Stream Flow". G. S. Cavadias, Proceedings of the Eastern Snow Conference. Vol. 3.

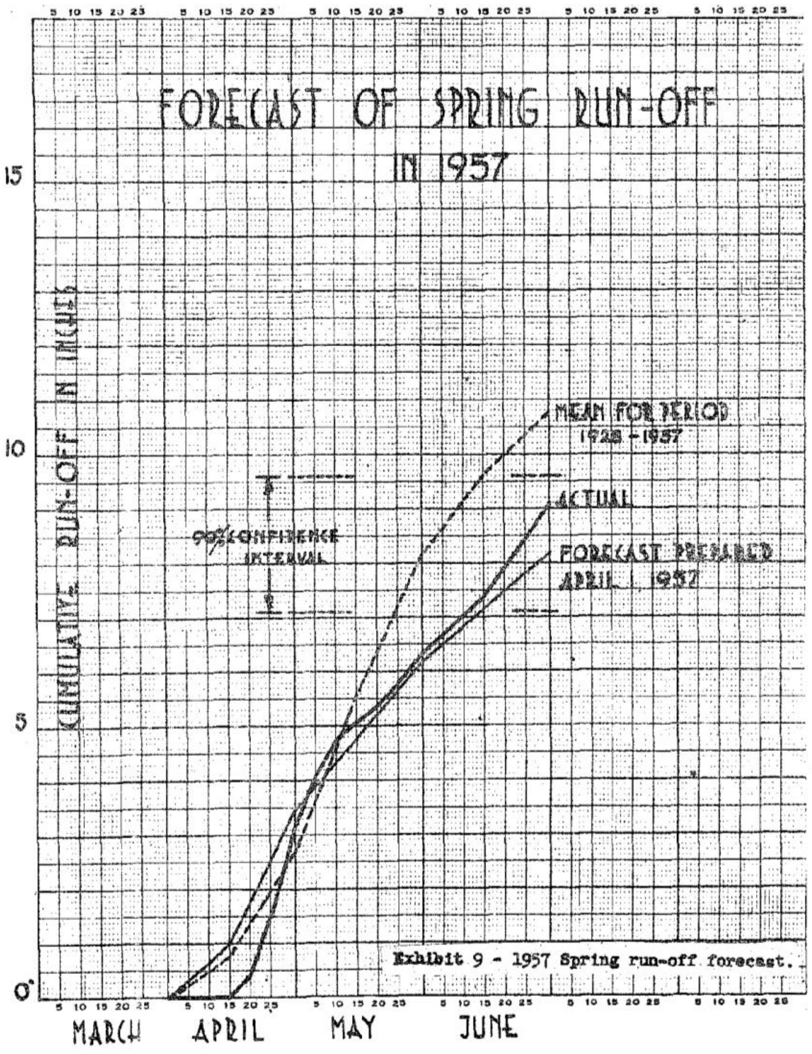


Exhibit 9 - 1957 Spring run-off forecast.

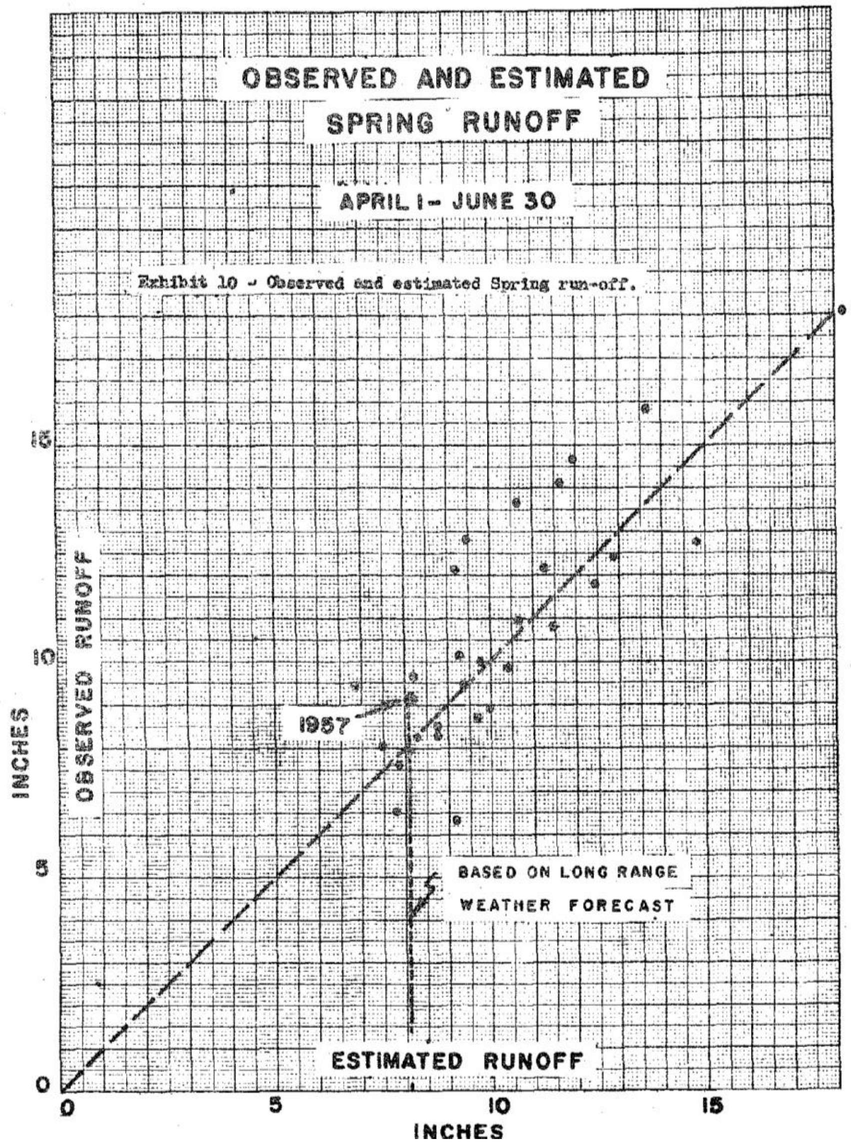


Exhibit 10 - Observed and estimated Spring run-off.

LIST OF EXHIBITS

BASIC DATA CHARACTERISTICS IN RELATION TO RUNOFF FORECAST ACCURACY[1]

By

R. A. Work and R. T. Beaumont[2]

Abstract

The analyses presented in this paper stress the importance of basing forecasts of river flows upon data secured as nearly as possible at the water sources. Data of the most simple and direct character are most efficient. The basic data, as gathered by snow surveys from the heart of the water-producing areas, generally result in the most accurate forecasts because it is a more precise method of sampling the greatest factor in streamflow production in mountainous western areas.

Introduction

Numerous agencies conduct snow surveys in order to forecast the seasonal runoff of western rivers. Prominent among such agencies are the California Department of Water Resources, the British Columbia Water Rights Branch, a considerable number of private or public utilities, numerous irrigation and soil conservation districts. The U. S. Soil Conservation Service has since 1935 coordinated most of the western snow survey activities outside of California and Canada. The Service has been strongly supported in the activity by various State Engineers and Agricultural Experiment Stations of western states, and by federal agencies specifically concerned with water and natural resource problems, including the U. S. Bureau of Reclamation, Forest Service, Corps of Engineers, Geological Survey, Bonneville Power Administration, National Park Service, Indian Service, Fish and Wildlife Service, and others. The results and interpretations of the snow surveys have been made available for the past 23 years through the published "Federal-State Private Cooperative Snow Survey and Water Supply Forecast" reports.

More recently, beginning in 1944, forecasts of the annual runoff of western rivers have also been developed. These forecasts, based largely upon precipitation measurements, are issued by the U. S. Weather Bureau in its publication entitled "Water Supply Forecasts for the Western United States." The methods used have been detailed in numerous references. The merits of the water year forecast have been described by Kohler (1).

---

[1] Paper presented at Western Snow Conference, Bozeman, Montana, April 17, 1958.

[2] The authors are respectively, Head, Water Supply Forecast Section, and Head, Analysis Unit, Water Supply Forecast Section, both of Soil Conservation Service, Portland, Oregon. Mrs. Helen Woodbury, Statistical Clerk, developed and assisted in interpretation of the tabular material herein. The authors are deeply indebted to Mrs. Woodbury for her interest and assistance.