"A COMPARISON OF GRAPHICAL AND ANALYTICAL CORRELATION METHODS AS APPLIED TO SPRING-RUN-OFF STUDIES."

By

George S. Cavadias

1. GENERAL

1.1 The volume of the Spring run-off depends on many interrelated factors and therefore it is necessary to use special mathematical methods for determining a reliable forecasting relationship. For this purpose, hydrologists use the so called "correlation" methods which deal with Statistical associations between observations.

1.2 An examination of current hydrological publications shows that there are two types of correlation methods used for forecasting purposes:

a) Analytical or least squares correlation methods which are described in detail in all standard Statistical texts.

b) Graphical correlation methods which are described in more specialized texts on Statistical Association (Ref. 1 and 2) and in the well-known book "Applied Hydrology" by Linsley, Paulus and Kohler, (Ref. 3)

This graphical approach can be readily understood by a hydrologist without a statistical background, and therefore it has been used extensively for determining forecasting relations.

1.3 The purpose of this paper is to compare the analytical and the graphical correlation methods from the points of view of

a) Underlying assumptions
b) Applicability to the forecasting problem
c) Convenience of application

Before attempting a more detailed comparison of the two methods, the basic concepts of the relationship between observed values of physical variables will be briefly reviewed.

2. TYPES OF RELATIONSHIPS BETWEEN PHYSICAL VARIABLES

Two variables are functionally related if a unique relationship exists between them. This is the case when the variables are related by some physical law and the variation of the observations about the relationship is due entirely to experimental errors.

Exhibits 1 and 2 give examples of functional relationships common in engineering work.

Exhibit 3 shows the relationship between the water equivalent of snow and the run-off, as given by Dr. Church. (Ref. 4).

Even if the run-off during a certain period were a mathematical function of the total precipitation on the watershed during the same or another period, the scatter diagram would not exhibit a functional relationship. The reason is that, the "water equivalent of snow" is a weighted mean of the observations at a number of snow courses, and therefore a more or less biased sample of the amount of water retained in the watershed in the form of snow. Therefore, the scatter of the points in the diagram is due not only to experimental errors but mainly to the sampling variation of the observations. In this case, the water equivalent of snow, as measured, and the total run-off during a certain period, are associated statistically" or "correlated". This means precisely that, to a given value of the water equivalent of snow, there may correspond many values of the run-off and not just one, as in the case of a functional relationship.

The relationships used in the prediction of the Spring run-off from the water equivalent of snow and other physical factors generally belong to the class of Statistical Associations.

3. STUDY OF STATISTICAL ASSOCIATIONS

If two variables x_1, x_2 are statistically associated, there are many possible values of x_1 for a given value of x_2.

The purpose of studying a statistical association is to derive an equation of the form $x_1' = f(x_2)$ where x_1' is a measure of central tendency (usually the mean) of all values of x_1 corresponding to a given value of x_2. This is called a "regression equation."

A method for obtaining a regression equation must satisfy the following conditions in order to be useful for forecasting purposes:

1) To determine the measure of central tendency which is useful for the particular application.

2) To establish a "confidence interval" for the estimated value x_1'.

3) To require an amount of computations consistent with the accuracy of the final result.

4) To be based on <u>explicit</u> assumptions.

   This condition is necessary for determining the applicability of the method to a particular case.

These four requirements will be used as a basis for the comparison of analytical and graphical correlation methods as applied to Spring Run-off Studies.

4. ANALYTICAL CORRELATION METHODS

Analytical methods for determining regression lines are based on the method of least squares. According to this method the regression line is determined from the condition that the sum of the squares of the deviations of all points from it should be a minimum.

The use of this method can be justified as follows:

1) From the point of view of prediction we are interested in the mean $\bar{x}_1$ corresponding to a given x_2 or $\bar{x}_1 = f(x_2)$. By definition, this is a regression line.

2) The mathematical form of the regression curve can be determined if the joint frequency function of x_1 and x_2 or the conditional frequency function $f(x_1/x_2)$ are known, which is not generally the case.

3) The following theorem is proven in texts on Mathematical Statistics, (Ref. 5).
"Among all possible functions $g(x_2)$, the regression curve $\bar{x}_1 = f(x_2)$ has the property of minimizing the sum $\sum \left[x_1 - g(x_2)\right]^2$ of the squares of the deviations from it."

4) It is known that if the two variables are jointly normally distributed, the regression curve is a straight line. Therefore, in this case, the application of the method of least squares to a linear function is sufficient for determining the coefficients of the regression equation $x_1 = a + b_2 x_2$.

5) Many frequency functions of hydrological variables show some similarity to the normal frequency function. (They are approximately bell-shaped, unimodal etc). In some cases they approach the normal function after a suitable transformation. On this basis, if the scatter diagram suggests a linear association, we <u>postulate</u> a linear regression equation and determine the coefficients by the method of least squares.

In order to determine the "confidence interval" the additional assumption is necessary, that the deviations from the regression line are normally distributed with the same variance.

If the last condition is not satisfied, a transformation of the variables may be useful (Ref. 6).

In the case of many variables, the regression equation is defined similarly as the locus of the points $\bar{x}_1 = f(x_2, x_3, \ldots x_n)$. This regression surface in n- dimensional space satisfies the theorem mentioned above and it can be determined by the method of least squares. For carrying out the computations for multiple correlation studies a systematic arrangement (Ref. 7) or electronic computing equipment can be used advantageously.

5. GRAPHICAL CORRELATION METHODS

The method most commonly used is the "coaxial" correlation method described by Linsley, Kohler and Paulus (Ref. 3). A careful examination of the description of the method, given in Ref. 3 indicates that it is a search for a functional relationship between the variables.

From the point of view of the criteria outlined in Section 3, the coaxial method is not well suited for the study of statistical associations: In respect to the first criterion, the value of x_1, estimated by the coaxial method is not a precisely defined measure of central tendency.

The second criterion concerning the establishment of confidence intervals can be satisfied by the graphical method only if an additional assumption is made concerning the frequency of the deviations from the estimate (Ref. 8). A more important shortcoming of the graphical approach is that it does not permit the determination of confidence limits for individual forecasts, which depend on the values of the independent variables.

Concerning the third criterion, Ezekiel comments as follows on the speed of convergence of the graphic method (Ref. 1). Mathematical studies have shown that in linear multiple correlation the graphic method gives results which tend to approach the lines secured by a least squares solution, even if the first approximations are pure arbitrary guesses. Further, they have shown that the speed of convergence depends on the intercorrelation among the independent variables. The higher their intercorrelation, the slower tends to be the speed of convergence.

A direct comparison of the least squares and the coaxial methods was made by the writer, using data from a paper by Burns and Strauss. (Ref. 8). The calculations were carried out according to the outline given in Reference 7 and the results included the multiple regression equation, the multiple correlation coefficient, and the confidence intervals for the regression coefficients and for one individual forecast. A graph similar to fig. 1 of Ref. 8 was also prepared from the multiple regression equation. The time required for these computations was two days.

Exhibit 4 shows a comparison of the solutions of this problem by the graphical and the analytical methods. The scatter diagram indicates that there is no significant difference between the two estimates and therefore no particular importance should be attributed to the slight curvature and the non-uniform spread of the lines in fig. 1 of Ref. 8.

The application of the coaxial correlation, primarily a method for finding functional relationships, was made possible by the high value of the multiple correlation coefficient $\overline{R}_{1.2345} = 0.93$.

Finally, it is not possible to formulate precisely the basic assumptions of the coaxial correlation method and, therefore, it is difficult to determine the applicability of the method to a particular problem.

LIST OF REFERENCES

1. M. Ezekiel "Methods of Correlation Analysis" John Wiley & Sons, New York 1941.

2. M. H. Quenouille "Associated Measurements" Butterworths Scientific Publications, London 1952

3. Linsley, Paulus, Kohler "Applied Hydrology" McGraw-Hill Co. 1949.

4. O. Meinzer "Hydrology" Dover Publications.

4. H. Cramer "Mathematical Methods of Statistics" Princeton University Press. 1957.

6. G. S. Cavadias "An Approach to forecasting the Spring Run-off in Quebe c". Proceedings Western Snow Conference 1958.

7. G. S. Cavadias "Reappraisal of Snow Melt as a factor in Quebec Stream Flow" Proceedings Eastern Snow Conference 1955.

8. Joseph I. Burns and Fred A. Strauss "Graphical Forecast Errors" Proceedings Western Snow Conference 1957.

EXHIBIT I

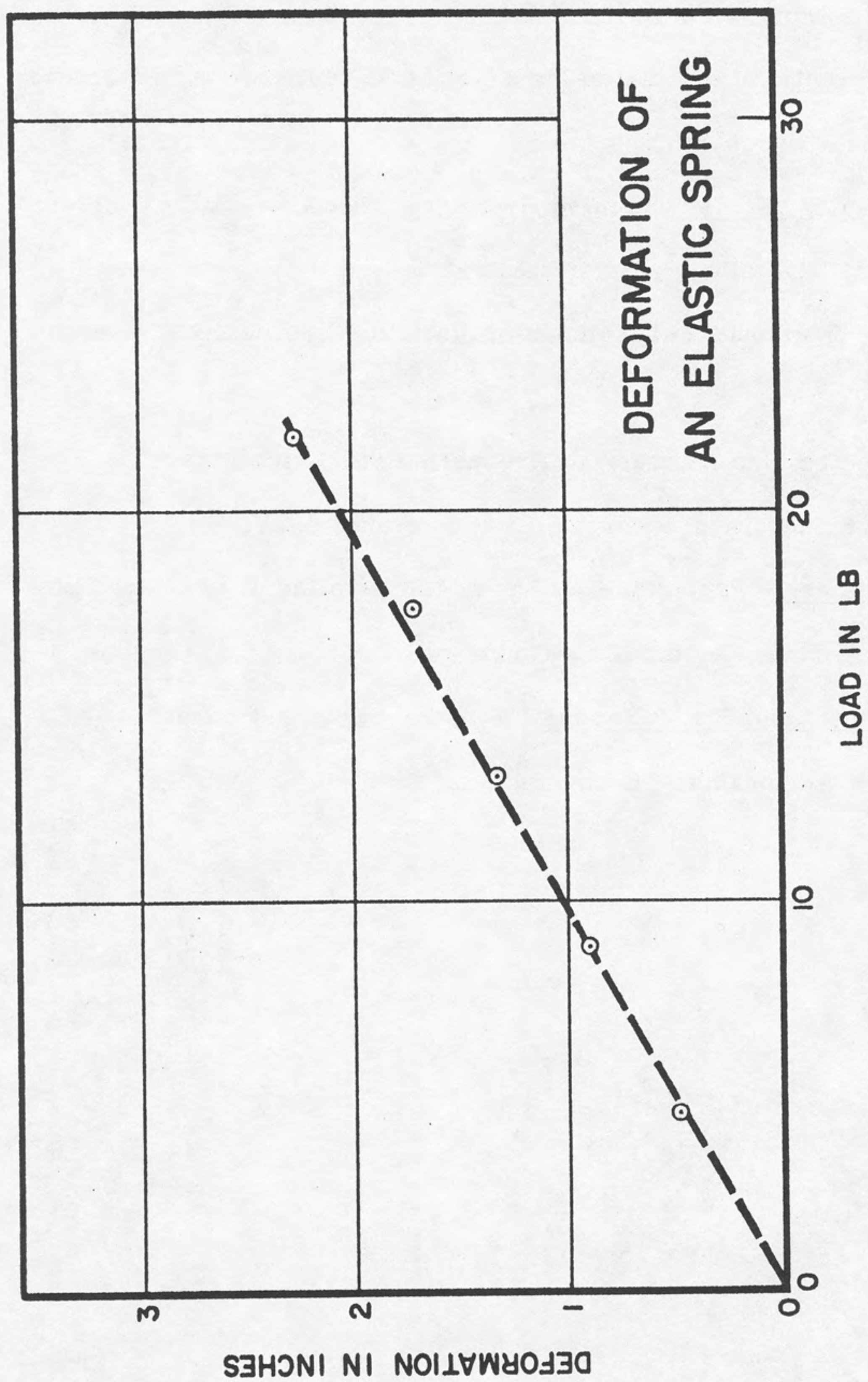
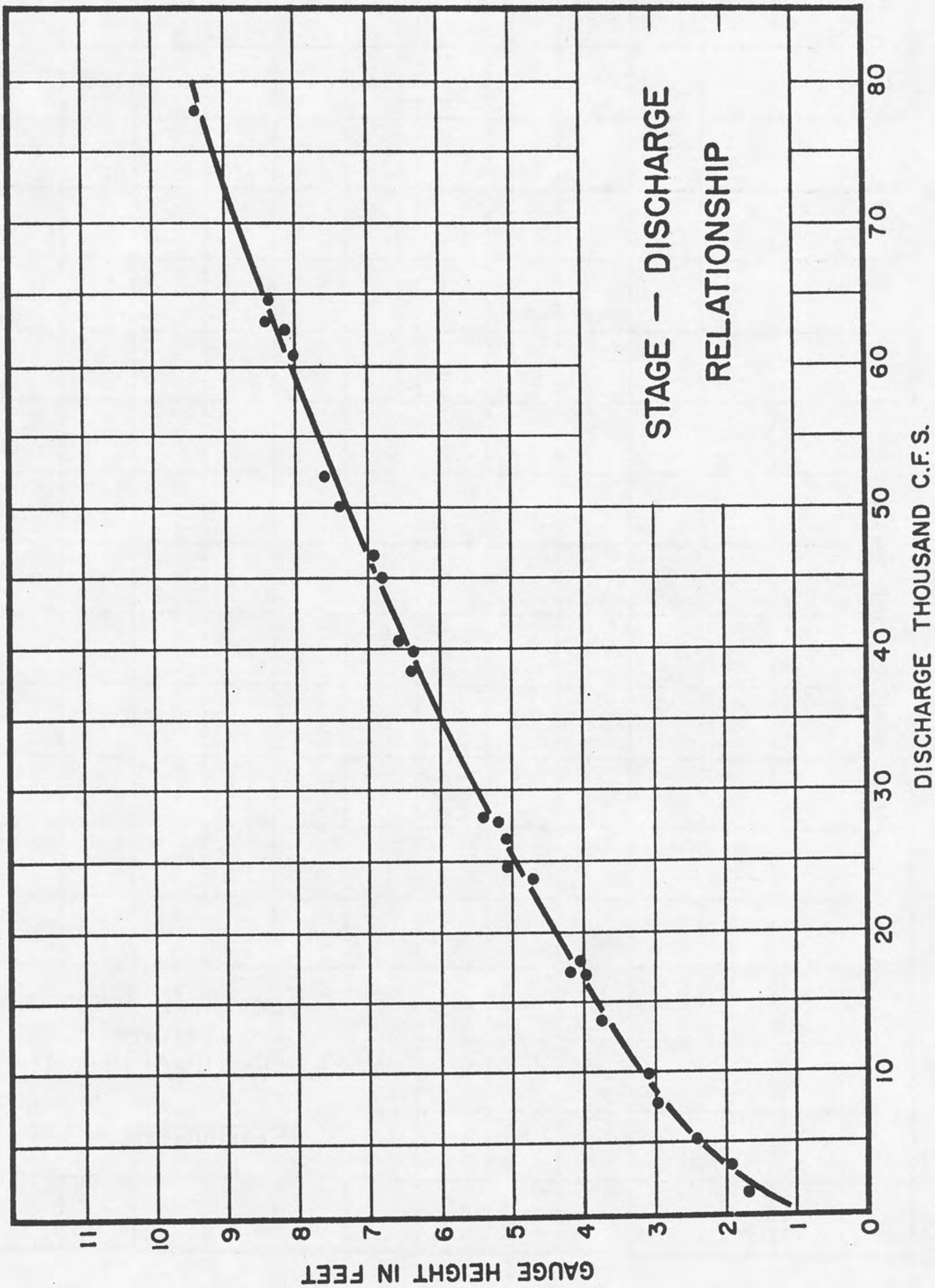

DEFORMATION OF
AN ELASTIC SPRING
LOAD IN LB
DEFORMATION IN INCHES
0
10
20
30
1
2
3

EXHIBIT 2



STAGE — DISCHARGE
RELATIONSHIP
DISCHARGE THOUSAND C.F.S.
0
10
20
30
40
50
60
70
80
GAUGE HEIGHT IN FEET
1
2
3
4
5
6
7
8
9
10
11

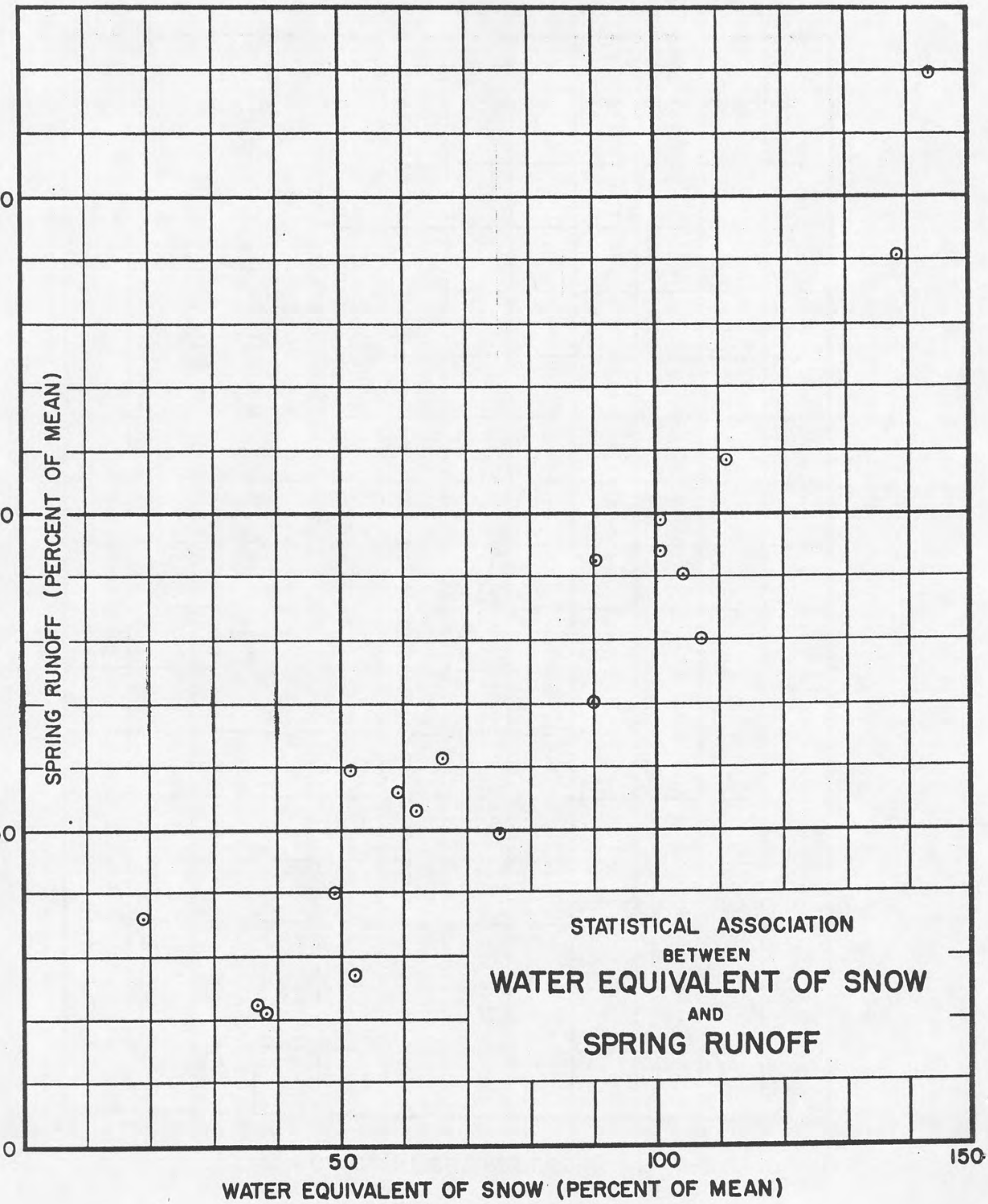
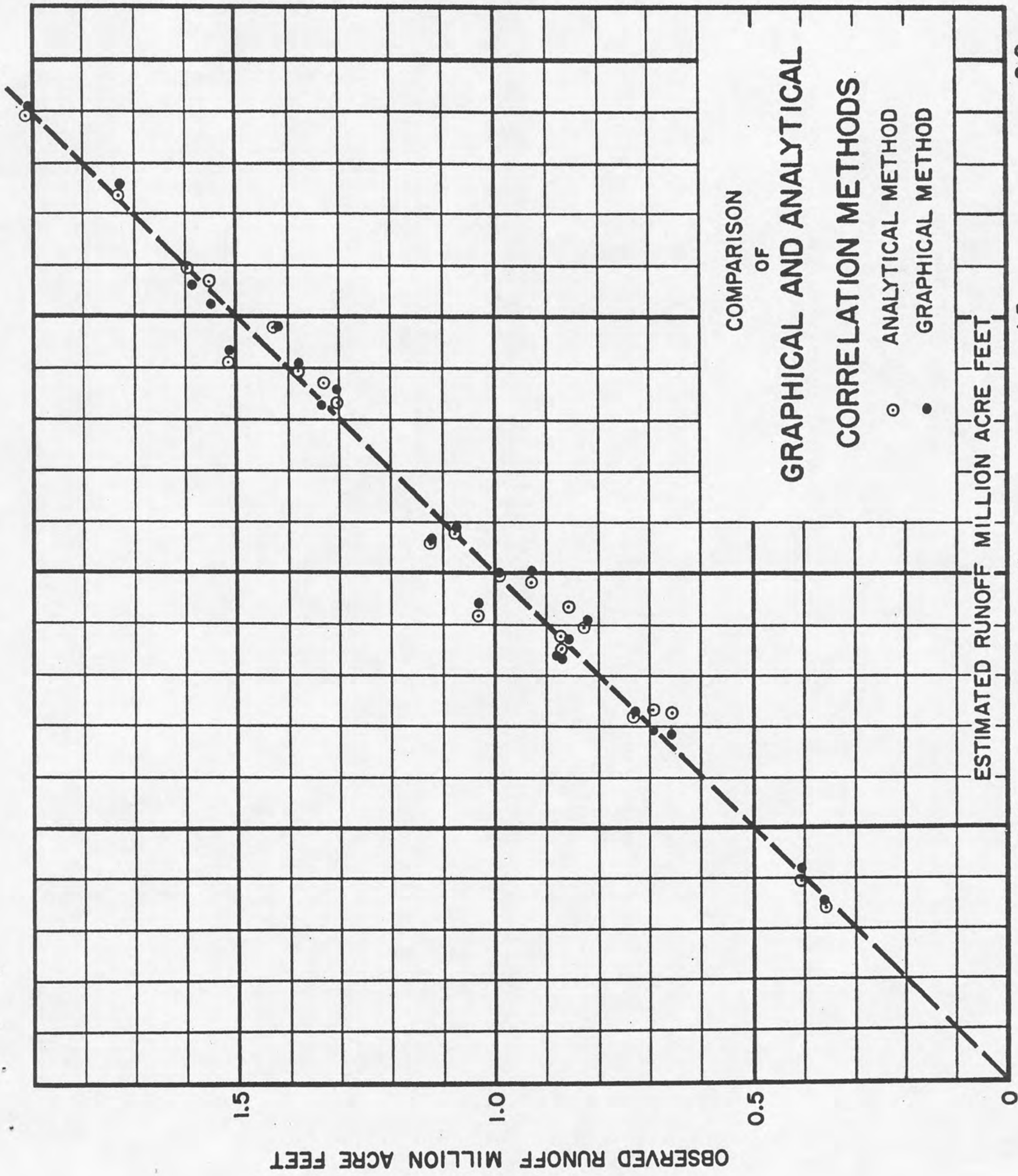

STATISTICAL ASSOCIATION
BETWEEN
WATER EQUIVALENT OF SNOW
AND
SPRING RUNOFF
SPRING RUNOFF (PERCENT OF MEAN)
WATER EQUIVALENT OF SNOW (PERCENT OF MEAN)
0
50
100
150

EXHIBIT 4


COMPARISON
OF
GRAPHICAL AND ANALYTICAL
CORRELATION METHODS
ANALYTICAL METHOD
GRAPHICAL METHOD
ESTIMATED RUNOFF MILLION ACRE FEET
OBSERVED RUNOFF MILLION ACRE FEET
0
0.5
1.0
1.5