CHAPTER 12

STOCHASTIC HYDROLOGY

12-1. Introduction.

a. A stochastic process is one in which there is a chance component in each successive event and ordinarily some degree of correlation between successive events. Modeling of a stochastic process involves the use of the "Monte Carlo" method of adding a random (chance) component to a correlated component in order to construct each new event. The correlated component can be related, not only to preceding events of the same series, but also to concurrent and preceding events of series of related phenomena.

b. Work in stochastic hydrology has related primarily to annual and monthly streamflows, but the results often apply to other hydrologic quantities such as precipitation and temperatures. Some work on daily streamflow simulation has been done.

12-2. Applications.

a. Hydrologic records are usually shorter than 100 years in length, and most of them are shorter than 25 years. Even in the case of the longest records, the most extreme drought or flood event can be far different from the next most extreme event. There is often serious question as to whether the extreme event is representative of the period of record. The severity of a long drought can be changed drastically by adding or subtracting 1 year of its duration. In order that some estimate of the likelihood of more severe sequences can be made, the stochastic process can be simulated, and long sequences of events can be generated. If the generation is done correctly, the hypothetical sequence would have as equal likelihood of occurrence in the future as did the observed record.

b. The design of water resource projects is commonly based on assumed recurrence of past hydrologic events. By generating a number of hydrologic sequences, each of a specified desired length, it is possible to create a much broader base for hydrologic design. While it is not possible to create information that is not already in the record, it is possible to use the information more systematically and more effectively. In selecting the number and length of hydrologic sequences to be generated, it is usually considered that 10 to 20 sequences would be adequate and that their length should correspond to the period of project amortization.

c. It must be recognized that the more hydrologic events that are generated, the more chance there is that an extreme event or combination of events will be exceeded. Consequently, it is not logical that a design be based on the most extreme generated event, but rather on some consideration of the total consequences that would prevail for a given design if all generated events should occur. The more events that are generated, the less proportional weight each event is given. If a design is tested on 10 sequences of hydrologic events, for example, the benefits and costs associated with each sequence would be divided by 10 and added in order to obtain the "expected" net benefits.

12-3. Basic Procedure. Successful simulation of stochastic processes in hydrology has been based generally on the concept of multiple linear regression, where the regression equation determines the correlated component, and the standard error of estimate determines the random component. Figure 12-1 illustrates the general nature of the process. In this case, a low degree of correlation is illustrated, in order to emphasize important aspects of the process. It can be seen that, if every estimate of the dependent variable is determined by the regression line (Figure 12-1a), the estimated points would be perfectly correlated with the independent variable and would have a much smaller range of magnitude than the actual observed values of the dependent variable. In order to avoid such unreasonable results, it is necessary to add a random component to each estimate (Figure 12-1b), and this random component should conform to the scatter of the observed data about the regression line.

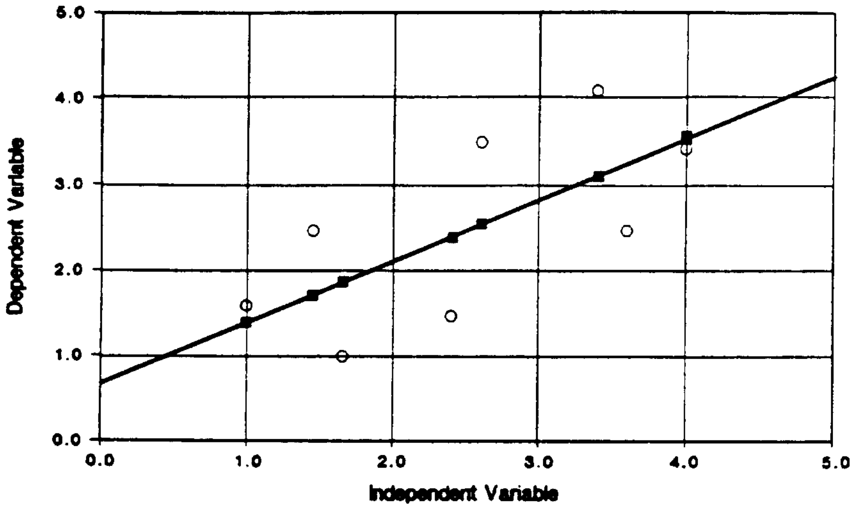

Figure 12-1a. Data Estimation from Regression Line.

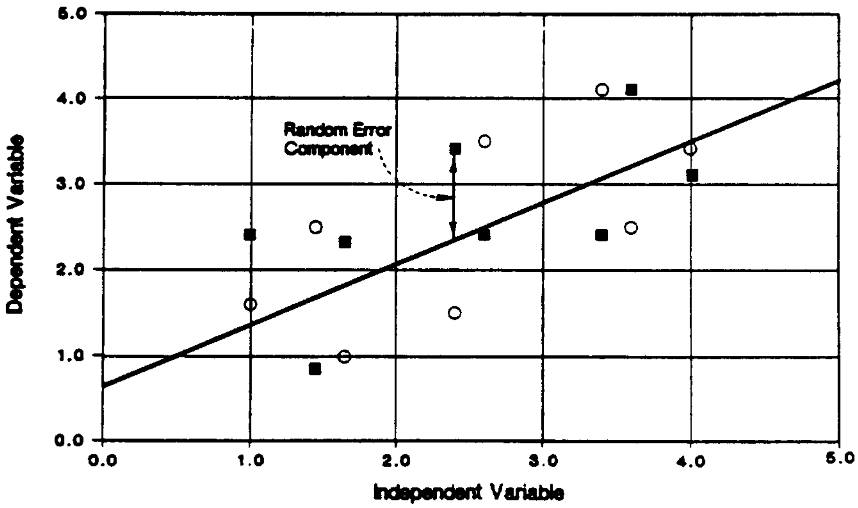

Figure 12-1b. Data Estimation with Addition of Random Errors.

12-4. Monthly Streamflow Model.

a. In accordance with the above basic procedure, a simulation model for generating values of a variable which can be defined only partially by a deterministic relation is:

$$Y = a + b_1X_1 + b_2X_2 + ZS_Y(1-R^2)^{½} \quad (12\text{-}1)$$

where:

Y = dependent variable

a = regression constant

b_1, b_2 = regression coefficients

X_1, X_2 = independent variables

Z = random number from normal standard population with zero mean and unit variance

S_Y = standard deviation of dependent variable

R = multiple correlation coefficient

b. This type of simulation model can be used to generate related monthly streamflow values at one or more stations. Multiple linear regression theory is based on the assumed distribution of all variables in accordance with the Gaussian normal distribution. Therefore, mathematical integrity requires that each variable be transformed to a normal distribution, if it is not already normal. It has been found that the logarithms of streamflows are approximately normally distributed in most cases. For computational efficiency it is convenient to work with deviations from the mean which have been normalized by dividing by the standard deviation. This deviate is sometimes called the Pearson Type III deviate and can be computed as follows:

$$t_i = (X_{i,j} - \overline{X}_i)/S_i \quad (12\text{-}2)$$

where:

t = Pearson Type III deviate

i = month number

j = year number

X = logarithm of flow

$\overline{X}$ = mean of flow logarithms

S = standard deviation of flow logarithms

c. If these deviates exhibit a skewness, they can be further transformed, if necessary, to a distribution very close to normal by use of the following approximate Pearson Type III transform equation:

$$K_i = (6/G_i)\,\{[(G_i t_i/2) + 1]^{1/3} + 1\} + G_i/6 \qquad (12\text{-}3)$$

where:

K = normal standard deviate

i = month number

G = skew coefficient

t = Pearson Type III deviate as defined in Equation 12-2

An equation for generating monthly streamflow is:

$$K'_{i,k} = \beta_1 K'_{i,1} + \beta_2 K'_{i,2} + \dots + \beta_{k-1} K'_{i,k-1} + \beta_k K'_{i-1,k} + \beta_{k+1} K'_{i-1,k+1} + \beta_n K'_{i-1,n} + Z_{i,k}\,(1-R^2_{i,k})^{1/2} \qquad (12\text{-}4)$$

where:

K' = monthly flow logarithm, expressed as a normal standard deviate

β = beta coefficient, defined as $b_{i,m}S_{i,m}/S_{i,k}$ where m is a station not equal to k and b is the regression coefficient.

i = month number for value being generated

k = station number for value being generated

n = number of interrelated stations

R = multiple correlation coefficient

Z = random number from normal standard population

For the case of a single station, this resolves to:

$$K_i' = R_{i,i-1} K_{i-1}' + Z_i (1-R_{i,i-1}^2)^{1/2} \qquad (12\text{-}5)$$

d. Note that Equation 12-5 is very similar to Equation 12-1. The differences result from using normal standard deviates. When this is done, the regression constant, a, equals zero, the regression coefficients, b, become beta coefficients, β, and the standard deviation, S, does not appear in the random component since it equals 1. Note also that one of the independent variables is the flow for the preceding month in order to preserve the inherent serial correlation. The flow value in the original units is computed by reversing the transformation process, i.e., from normal standard deviate to Pearson Type III deviate, to logarithm of flow and finally flow value.

e. A step-by-step procedure for generating monthly streamflows for a number of interrelated locations having simultaneous records is as follows:

(1) Compute the logarithm of each streamflow quantity. If a value of zero streamflow is possible, it is necessary to add a small increment, such as 0.1 percent of the mean annual flow, to each monthly quantity before taking the logarithm.

(2) Compute the mean, standard deviation and skew coefficient of the values for each location and each month, using equations given in Chapter 2.

(3) For each month and location, subtract the mean from each event and divide by the standard deviation (Equation 12-2).

(4) Transform these "standardized" quantities to a normal distribution by use of Equation 12-3.

(5) Arrange the locations in any sequence, and compute a regression equation for each location in turn for each month. In each case, the independent variables will consist of concurrent monthly values at preceding stations and preceding monthly values at the current and subsequent stations.

(6) Generate standardized variates for each location in turn for each month, starting with the earliest month of generated data. This is accomplished by computing a regression value and adding a random component. The random component, according to Equation 12-5, is a random selection from a normal distribution with zero mean and unit standard deviation, multiplied by the alienation coefficient which is $(1 - R^2)^{1/2}$.

(7) Transform each generated value by reversing the transform of Equation 12-3 with the appropriate skew coefficient, multiplying by the standard deviation and adding to mean in order to obtain the logarithm of streamflow.

(8) Find the antilogarithm of the value determined in step (7) and subtract the small increment added in step (1). If a negative value results, set it to zero.

f. It is obviously not feasible to accomplish the above computations without the use of an electronic computer. A computer program, HEC-4 Monthly Streamflow Simulation (51) can be used for this purpose.

12-5. Data Fill In. Ordinarily, periods of recorded data at different locations do not cover the same time span, and therefore, it is necessary to estimate missing values in order to obtain a complete set of data for analysis as described above. In estimating the missing values, it is important to preserve all statistical characteristics of the data, including frequency and correlation characteristics. To preserve these characteristics, it is necessary to estimate each individual value on the basis of multiple correlation with the preceding value at that location and with the concurrent or preceding values in all other locations. A random component is also required, as indicated in Equation 12-1.

12-6. Application In Areas of Limited Data. The streamflow generation models discussed so far have assumed that sufficient records were available to derive the appropriate statistics. For instance, the monthly streamflow model requires four frequency and correlation coefficients for each of the 12 months, or 48 values for one station simulation. A model has been developed (51) that combines the coefficients into a few generalized coefficients for the purpose of generating monthly streamflow at ungaged locations. (Procedures for determining generalized statistics for use in generating daily flows have not yet been developed.) The generalized model considers the following:

- season of maximum runoff
- lag to season of minimum runoff
- average runoff
- variation between maximum and minimum runoff
- standard deviation of flows
- interstation and serial correlations of flows

12-7. Daily Streamflow Model.

a. Generation of daily streamflows can be accomplished in a manner very similar to the generation of monthly streamflow quantities. Although a computer program has been prepared for this purpose, it is capable only of generating flows at a single location and does not provide a totally satisfactory hydrograph. Since it is desired in many reservoir operation studies to use a monthly interval most of the time, and to perform daily operation computations for only a few critical periods, the program has been designed to generate daily flows after the monthly total runoff has been generated by another program. Flows for any particular day are correlated with flows for the preceding day and for the second antecedent day.

b. A procedure that will give a reasonable shaped hydrograph, as well as coordinated hydrographs at many locations in a basin, would consist of (1) stochastic generation of precipitation over the basin, and (2) using a precipitation-runoff model to derive the resulting streamflow.

12-8. Reliability. While the simulation of stochastic processes can add reliability in hydrologic design, the techniques have not yet developed to the stage that they are completely dependable. All mathematical models are simplified representations of the physical phenomena. In most applications, simplifying assumptions do not cause serious discrepancies. It is important at this "state of the art," however, to examine carefully the results of hydrologic simulation to assure that they are reasonable in each case.