STREAMFLOW SIMULATION TECHNIQUES

Richard N. DeVries

Department of Civil Engineering, Oklahoma State University, Stillwater, Oklahoma

From application of a streamflow simulation method to the Elkhorn River, it is concluded that long sequences of monthly flows can be simulated for a stream when basic streamflow records are available. Frequency statistics representing characteristics of the natural monthly flows can be preserved in the synthetic flows. Synthesized flows generated for the Elkhorn River at Waterloo, Nebraska meet reasonable statistical criteria and can be used in future hydrologic studies for this basin. Mathematical models for generating monthly runoff are usable only when sufficient records are available to establish characteristics of the streamflow. The model cannot improve faulty or biased records and no mathematical method can replace a reliable long term record.

Simulation as a technique for analyzing the behavior of physical systems has a long history, but only in recent years has its use become widespread as a tool for problem-solving. Its development is directly associated with advances in electronic computers, and the development of operations research and systems analysis.

Simulation is a process which duplicates the essence of a system or activity without actually attaining reality itself. This definition encompasses analog and digital simulations, physical devices, training machines, defense system models, water resources, transportation systems and economic studies.

The design of water resources projects is based primarily on hydrologic and economic data. Streamflow records are a major type of hydrologic data. Surface streamflow data have two major uses. The first is to provide general regional information. This data represents "natural" conditions and may be used in combination with similar data at other sites to gain a regional description of the streamflow of an area. The second major use is for project operation and design purposes.

Inadequate streamflow records nearly always impair the precision of design in river-basin development. The available hydrologic record may lack a critical sequence of years of low or high runoff, and the most severe drought or flood in a short record may not be representative of the statistical population. Design for reservoir projects that are based on a sequence of flows which are not representative of the true potential of a given drainage basin can only result in inefficiencies. A reservoir will be over-designed if the water supply is less than the short term records shows, and the developer, either private or public, will pay for storage capacity that will not be used. Conversely, a reservoir will be designed with inadequate capacity to contain the available flow if the water supply is greater than the short term record shows; consequently, some of the potential benefits from the reservoir may not be realized.

SIMULATION TECHNIQUES

Early streamflow simulation techniques usually employed a series of cards. Annual runoff volumes were recorded on cards, one value per card, and then the cards were shuffled like a deck of playing cards. Cards were drawn from the deck one at a time until all had been drawn. The value of each card was recorded as it was drawn and this array of values was assumed to be a new series of annual volumes.

Next a table of random numbers was used to synthesize a long record of annual flows having the same mean and standard deviation as the original record and assuming a normal distribution. Although an improvement over the method using cards, it neglects any serial correlation that may exist between annual flows.

Thomas and Fiering (1) were the first to develop a mathematical model for sequential generation of nonhistoric streamflows. Their model treats the flow in any period as a linear function of the flow in the preceding period. The inclusion of the serial correlation makes the method applicable to annual flows, monthly flows, or flows of even shorter duration that are serially correlated. The synthetic series is generated to preserve all the characteristics of the original series.

The Army Corps of Engineers Hydrologic Engineering Center uses a second order Markov chain for synthesizing monthly streamflows. A first order Markov chain exists when the probability law that governs the future development of a process depends only on the current state and not on the prior evolution of the processes. When the probability law depends on the current state and the immediately preceding state, it is called a second order Markov chain. In the Corps' method, monthly flow logarithms are generated as normal standard deviates. The deviates are then converted to the logarithm of a specific monthly flow.

Additional work has been done in an attempt to synthesize rainfall data. Higher order Markov chains have been used for deriving hourly rainfalls, but rainfall synthesis is considered an unnecessary intermediate step because in most cases streamflow can be generated directly from streamgaging records.

STREAMFLOW SIMULATION MODELS

A mathematical model is the basis for the derivation of all modern simulation techniques. It is necessary to understand the basic variables of streamflow before a model can be developed for a given drainage basin. Studies of existing runoff records show that flow sequences have two components: a persistence component and a random component.

"Persistence" means that an effect continues after its cause is removed. Here it refers to the tendency for large runoff events to follow large runoff events and for small runoff events to follow small runoff events. The streamflow in any month is dependent on the streamflow in the preceding month in two ways. Precipitation may occur so that the resulting outflow from a watershed occurs over a period of two or more months, and the precipitation in a prior month may have filled the soil moisture reservoir so that the runoff would be higher this month than it would otherwise have been.

There is additional variance in streamflow from one month to the next that is not explained by persistence. This unexplained variance occurs at random and can be referred to as the random component of streamflow. A mathematical model of a drainage basin must include factors that account for both the persistence and random nature of flows.

Thomas and Fiering (1) developed a recursion equation for monthly time intervals in a bivariate model. The equation is:

Eq. 1

$$Q_{i+1} = \bar{Q}_{j+1} + b_j(Q_i - \bar{Q}_j) + t_i S_{j+1} (1 - r_j^2)^{\frac{1}{2}}$$

where

Q_i = the discharge during the i month record from the start of the synthetic sequence.

Q_{i+1} = the discharge during the (i+1) month.

$\bar{Q}_j$ = the mean monthly discharge during the j month with a repetitive cycle of 12 months.

$\bar{Q}_{j+1}$ = the mean monthly discharge during the (j+1) month.

b_j = the regression coefficient for estimating the flow in the (j+1) month from the j month.

t_i = a normal random deviate with zero mean and unit variance.

S_{j+1} = the standard deviation of flows in the (j+1) month.

r_j = the correlation coefficient between the flows of the j and (j+1) month.

This is a linear regression model where flow in any month is a linear function of flow in the preceding month. The sequence of flows obtained from this recursion equation possesses the same general statistical properties as those representing natural streamflows. Variation in sign and magnitude of the random component makes for a continuous, unbounded, and serially correlated sequence of generated data.

The Army Corps of Engineers Hydrologic Engineering Center developed the following recursion equation:

$$K_i = B_i K_{i-2} + B_2 K_{i-1} + K_r (1-R^2)^{\frac{1}{2}} \qquad \text{Eq. 2}$$

where

K = the logarithm of the monthly flow, expressed as a normal standard deviate.

B = beta coefficients.

K_r = a random number from a normal population with zero mean and unit variance.

R = the multiple correlation coefficient.

The subscripts i, i-l, and i-2 represent current and antecedent months.

The logarithms of synthesized monthly flows are expressed as deviates, and then are converted to flows for specific months by the following equation:

$$X = M + K_i S \qquad \text{Eq. 3}$$

where

X = the logarithm of the monthly flow for month i.

M = the mean of the logarithms of the flows in the sample.

K_i = the logarithm of the monthly flow, expressed as a normal standard deviate that is determined from equation 2.

S = the logarithmic standard deviation of the sample.

The computed flow for the given month is obtained by taking the anti-logarithm of X. This transformation alters the correlation characteristics slightly, because of the change from linearity to non-linearity, but this change is considered to be insignificant. The statistical properties of the natural streamflows are preserved in the synthesized data.

A third mathematical model (2) uses a second order Markov chain of logarithmic transforms of data. This model is similar to the Thomas-Fiering model, with two exceptions, viz., two antecedent months are used as independent variables and a logarithmic transform is utilized. The general recursion equation developed is:

$$Q_i = A + B_1 Q_{i-2} + B_2 Q_{i-1} + T_i S_i (1-R^2)^{\frac{1}{2}} \qquad \text{Eq. 4}$$

where

Q = the logarithm of the computed monthly flows

A = a regression constant

B = regression coefficients of the antecedent monthly flows

T = a normal random deviate with zero mean and unit variance

S = the logarithmic standard deviation of the sample

R = the multiple correlation coefficient.

The subscripts i, i-l and i-2 represent the current and antecedent months.

Equations of this general form are developed and a solution of monthly flow computed for each month of the year. A computer sub-routine program will generate four digit random numbers whose magnitude lies between zero and one. These random numbers, which come from a rectangular distribution, are converted to normal random deviates by a transformation equation.

AN APPLICATION OF A STREAMFLOW SIMULATION MODEL

The above simulation model was applied to data for the Elkhorn River at Waterloo, Nebraska. This U. S. Geological Survey stream gaging station has complete flow data for the period from October, 1928 through September, 1968.

The Elkhorn River is a major tributary of the Platte River. The drainage area of about 5,000 square miles is upstream from Waterloo; it is approximately 190 miles long and 60 miles wide at its point of maximum width. The topography varies from sand hills to rolling farm land.

Runoff characteristics vary widely throughout the basin. Runoff in the sand hills region is characterized by slow peaking times and relatively low peak discharges. The downstream tributaries are characterized by flash runoff which generally occurs a few hours after a runoff-producing rainfall. Rainfall of sufficient magnitude to cause runoff from the entire basin would be a rare event. About 30% of the peak annual discharges recorded at the Waterloo gaging station occur as a result of snowmelt.

Three regression studies of existing monthly streamflow data were made. The first consisted of computing serial correlations for each month, using the previous monthly flow as an independent variable. The second study consisted of computing serial correlations for each month using the logarithmic transforms of the antecedent monthly flow as the independent variables. The third regression study consisted of a correlation of the logarithmic transforms of the monthly data using two antecedent months as independent variables. The correlation and skew coefficients that were computed for each month and for each condition studied are shown in Table 1.

The coefficients of serial correlation for the Elkhorn data are higher when they are computed from logarithmic transforms of the data than from the natural data. When the second antecedent monthly flow is included as a variable in the equations, the average correlation coefficient increases slightly, from 0.683 to 0.709.

The coefficient of skew computed from a normal distribution is zero. As shown in Table 1, the average monthly skew coefficients computed from the natural and logarithmic transformed data are 2.18 and 0.36, respectively. Thus, the transformed data more nearly conform to a normal distribution than do the natural data.

The constants, regression coefficients, standard deviations, and correlation coefficients which were computed for each month of the year are presented in Table 2.

GENERATION OF DATA

Synthetic monthly streamflows were computed for the Elkhorn River using an RCA 301 electronic digital computer. Equations which were derived for each month were stored on tape. These equations, together with the generated normal random deviates, were used for computing the synthetic flows. The procedure proceeds as follows: the January flow in synthetic year i is generated from the January equation using the synthesized flow for November and December of the i-1 year as independent variables. This procedure was continued sequentially from month to month and year to year until 400 years of record had been generated.

Since it is necessary to use average values of the antecedent flows to start the simulation procedure in year one, the first few generated monthly flows will not necessarily occur at random. The first five years were arbitrarily discarded to insure that the important statistical properties were retained.

TESTING THE SYNTHESIZED DATA

It is necessary to test the generated data to insure that the method has preserved the natural streamflow characteristics. Synthetic years 6 through 385 were divided into ten consecutive segments of record of the same length as the natural streamflow record. The mean, standard deviation, and coefficient of skew were computed for each month of the ten segments of synthetic record. The coefficients of serial correlation and the generated normal random deviates were also examined.

The means computed from both the monthly and annual segments of synthesized data were in close agreement with those computed from the record. Standard deviations computed from the monthly synthesized data were similar to those computed from the record, but some variability in natural annual flows was lost in the simulation computations. Coefficients of skew computed from the natural monthly flow data were inconsistent; however, the average value was slightly positive. Simulated flows have a slight tending toward negative skewness. These values refer to the serial correlation coefficients of the natural and synthetic flows. The computed mean and standard deviation of the random deviates were −0.010 and 0.997, respectively, which proved that these deviates came from a normal distribution.

USE OF SYNTHETIC FLOWS IN THE DESIGN PROCESS

Flows generated for the Elkhorn River were used to demonstrate how a change in design could result from the application of synthetic flows in project planning. Estimates of storage requirements for a reservoir on the Elkhorn River at Waterloo were made, using the natural streamflow record, and using the generated streamflow record.

Maximum drought periods during the natural and synthetic records occurred from calendar years 1938 through 1940 and from synthetic years 80 through 82, respectively. Cumulative runoff volumes for these periods were plotted by the mass-diagram method.

The storage capacity required to maintain an assumed flow of 400 cfs downstream from an Elkhorn River reservoir would be 101,000 acre feet if the drought from the natural record was used as the design criteria. If the most severe drought from the synthetic record was used as the design criteria, 143,000 acre feet of storage capacity would be required to maintain a flow of 400 cfs downstream.

TABLE 1. *Correlation and skew coefficients computed from monthly streamflow data.*

	Serial Correlation Natural Data		Serial Correlation Logarithmic Transform		Second Order Markov Chain Log Transform
Month	Correlation Coefficient	Skew Coefficient	Correlation Coefficient	Skew Coefficient	Correlation Coefficient
Jan	0.845	1.58	0.833	0.01	0.854
Feb	0.719	2.51	0.694	0.48	0.705
Mar	0.255	2.16	0.387	0.70	0.555
Apr	0.523	3.41	0.699	1.28	0.706
May	0.680	1.28	0.759	0.31	0.762
Jun	0.387	2.16	0.494	0.16	0.502
Jul	0.455	1.43	0.588	−0.03	0.623
Aug	0.704	2.79	0.645	0.42	0.646
Sept	0.635	2.93	0.584	−0.14	0.618
Oct	0.895	2.93	0.799	0.85	0.808
Nov	0.873	2.03	0.920	0.53	0.928
Dec	0.874	0.93	0.793	−0.26	0.801
Average Value	0.654	2.18	0.683	0.36	0.709

TABLE 2. *Constants and coefficients of the monthly stream flow simulation equations (second order Markov chain).*

Month	Regression Constant A	Regression Coefficient B_1	Regression Coefficient B_2	Standard Deviation S	Correlation Coefficient R
Jan	0.854	0.288	0.513	0.152	0.854
Feb	−0.706	−0.358	1.556	0.267	0.705
Mar	0.191	1.073	0.004	0.296	0.555
Apr	1.291	0.113	0.615	0.277	0.706
May	0.897	−0.083	0.909	0.308	0.762
Jun	1.832	0.191	0.468	0.373	0.502
Jul	1.697	0.232	0.379	0.300	0.623
Aug	0.937	0.050	0.713	0.349	0.646
Sept	1.841	0.240	0.328	0.276	0.618
Oct	1.265	0.103	0.609	0.239	0.808
Nov	1.646	−0.121	0.757	0.167	0.928
Dec	0.609	−0.102	0.955	0.172	0.801

REFERENCES

1. H. A. THOMAS, JR. and M. B. FIERING, The Mathematical Synthesis of Streamflow Sequences. Chapt. 12 *in* A. Maass et al., Design of Water Resource Systems, Harvard University Press, Cambridge, Mass., 1962.
2. D. L. SVEUM, An Investigation and Application of Streamflow Simulation Techniques, M.S. Thesis, University of Nebraska, Lincoln, Neb., 1969.
3. K. L. PERRY, The Effects of Historical Record Lengths on Generating Synthetic Data Using a Stochastic Model of the Markov Chain, M.S. Thesis, Oklahoma State University, Stillwater, Okla., 1969.
4. M. M. HUFSCHMIDT and M. B. FIERING, Simulation Techniques for Design of Water Resource Systems, Harvard University Press, Cambridge, Mass., 1966.