SEQUENTIAL GENERATION
OF RAINFALL AND RUNOFF DATA

by

VEN TE CHOW
and
S. RAMASESHAN

DEPARTMENT OF CIVIL ENGINEERING
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS

October 1965
SEQUENTIAL GENERATION OF RAINFALL AND RUNOFF DATA\textsuperscript{a}

By Ven Te Chow,\textsuperscript{1} M. ASCE, and S. Ramaseshan,\textsuperscript{2} A.M. ASCE

INTRODUCTION

In water resources engineering designs, it is generally more important to assess the information concerning the time distribution rather than the momentary magnitude of a hydrologic event. For example, flood damage is estimated conventionally as a function of the peak discharge, whereas the full extent of damage does not depend entirely on the peak but rather on the time distribution of floods.\textsuperscript{3} In order to evaluate the full-scale damage, it is necessary to perform a stochastic frequency analysis of the time distribution. Unfortunately, such analyses cannot always be made of the available historical data because the latter usually have a limited length of record. However, representative data of large quantity can be produced by a modern hydrologic

\textsuperscript{a} Presented at the ASCE Water Resources Engrg. Conf. at Mobile, Alabama, March 8-12, 1965, where it was available as Conference Paper No. 140.

\textsuperscript{1} Prof. of Hydr. Engrg., Univ. of Illinois, Urbana, Ill.

\textsuperscript{2} Asst. Prof. of Civ. Engrg., A. C. Coll. of Eng. and Tech., Karaikudi, India.

technique known as "sequential generation" and the information so obtained is suitable for the stochastic analysis of river basin hydrologic systems.

The main purpose herein is to demonstrate the technique of sequential generation as applied to a stochastic analysis of the time distribution of hydrologic data. For illustrative purposes, annual storms and the concurrent annual floods in the French Broad River Basin at Bent Creek, North Carolina, are used as the available historical hydrologic data.

The annual storms are those which produced the maximum peak discharge in a water year. They vary with time at a random fashion and thus constitute a stochastic process which is time-dependent and random. Part of the storm rainfall is abstracted by the hydrologic system, and the effective rainfall is transformed by the system to direct runoff. As the historical data are limited in length, they are extended by sequential generation. The rainfall data are sequentially generated by Monte Carlo methods. The generated data are then routed through a hydrologic system, simulating the complex basin system, to yield sequential data of runoff. The latter are compared with the historical data and used in the stochastic analysis.

The hydrologic process under consideration, as it occurs in nature, is extremely complicated for many reasons, such as the apparent nonstationary nature of the rainfall process, the nonlinear response to the basin system, and the uncertain knowledge on abstractions and baseflow separations. In order to demonstrate the proposed method and make it readily amenable to mathematical formulation and analysis, reasonable assumptions or simplifications are necessary. It is believed that for practical purposes they will not affect

---

the results significantly in view of the usual accuracy and reliability of the original hydrologic data.

*NotaHon.*—The letter symbols adopted for use in this paper are defined where they first appear and listed alphabetically in the Appendix.

**ANALYSIS OF RAINFALL DATA**

*Storm-Shifting.*—Rainfall data of the annual storms for a period of 28 yr from 1935 to 1962 are available at a number of rain-gage stations in the French Broad River Basin (Fig. 1). For each annual storm, the average areal hourly rainfall over the basin was calculated by the Thiessen method. When concurrent hourly rainfalls at some stations are missing, they were estimated by regression analysis with the known concurrent rainfalls at other stations.

As the annual storms have different time distribution patterns, it is desirable, if possible, to develop the best orientation of the storms so that the mean, the standard deviation, and the trend and random components of their hourly rainfalls become regular and consistent. This is done by “storm-shifting.”

Consider *N* storms. These storms are assumed to have the same duration of *m* hours, which is taken approximately equal to the longest duration of the storms under consideration. Thus, at some parts of this duration there may be no rainfalls for many storms. That is, all the storms have *m* hourly rainfall values, and some of the hourly values are zero because the actual duration of a particular storm may be less than the selected common duration of *m* hours. Now, all storms are arranged to begin with the first hourly rainfall and end at *m*th hourly rainfall. Actually, it is difficult to provide a criterion for fixing the time of beginning of a storm because in many cases there may be a drizzle before the main part of the storm or there may be one or several breaks during the *m* hours. In this analysis, the first hourly rainfall is taken whenever the rainfall record shows such a value. In order to develop the best orientation for the time distributions, it is necessary to shift the beginning times of all storms so that the cross-correlation coefficients of their hourly rainfalls are maximum.

Let *x*_t and *y*_t, with *t* = 1, 2, ..., *m*, be the sequences of hourly rainfalls of two annual storms, *x* and *y*, respectively. The cross-correlation coefficient for a shift of storm *y* with storm *x* equal to *v* hours is computed by

\[
R_{xy}(v) = \frac{\sum_{t=1}^{m-v} x_t y_{t+v} - \frac{1}{m-v} \sum_{t=1}^{m-v} x_t \sum_{t=1}^{m-v} y_{t+v}}{(m-v-1) s_{x,t} s_{y,t+v}} \quad \ldots \quad (1)
\]

in which the standard deviations *s*_{x,t} and *s*_{y,t+v} are computed from

\[
s_{x,t}^2 = \frac{1}{m-v-1} \left[ \sum_{i=1}^{m-v-1} x_i^2 - \frac{1}{m-v} \left( \sum_{t=1}^{m-v} x_t \right)^2 \right] \quad \ldots \quad (2)
\]

and *s*_{y,t+v}^2, which is expressed by the same equation except that *x* and *t* are replaced by *y* and *t+v*, respectively.
It is generally seen that $R_{xy}(v)$ reaches a maximum for a specific value of $v$. When this shift is applied, the time distributions of the two storms are most agreeable or best oriented to each other at this relative position. When all storms are thus shifted to their best positions with respect to each of the other storms, their relative orientation is considered the best.

In applying the storm-shifting analysis to the hourly rainfalls of the annual storms at the French Broad River Basin, the common total duration $m$ is
taken as 36 hr. Because of the laboriousness of the analysis, the computation was performed on an electronic digital computer. The full-line plots in Fig. 2 represent the variations of the mean and the coefficient of variation of the hourly rainfalls after the annual storms were shifted to best positions for maximum cross-correlation. The dashed-line plots represent the same variations for unshifted data. It can be seen that shifting has adjusted the positions of the annual storms so that their high mean annual rainfalls tend to
cluster over the middle of the common storm duration. The shifted data appear to have less, but more regular, variation than the unshifted data.

Storm-shifting is not an absolutely necessary step in the proposed analysis of rainfall data, but it generally improves the statistical model of the data as an input to the hydrologic system.

Formulation of Stochastic Model.—Let \( x_t \) be the hourly rainfall of any one of \( N \) annual storms at time \( t \)-hour with \( t = 1, 2, \ldots, m \). To develop a suitable model to represent the stochastic process of the hourly rainfalls, several specific linear and nonlinear models having the following general relationship were tried. Thus,

\[
x_t = f_{t,1}(x_{t-1}) + f_{t,2}(x_{t-2}) + \ldots + f_{t,t-1}(x_1) + \epsilon_t \quad \ldots \ldots \ldots (3)
\]

in which \( \epsilon_t \) is the random component of \( x_t \) and the \( f \)'s denote trend components of various functions. It was found that the following nonstationary Markov-chain model is consistently satisfactory and most practical for the purpose. Therefore,

\[
x_1 = \epsilon_1 \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (4a)
\]

and

\[
x_t = r x_{t-1} + \epsilon_t \quad \text{with} \ t = 2, 3, \ldots, m \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (4b)
\]

in which \( r \) is the Markov-chain (or correlation) coefficient to be determined by least squares linear regression from the given rainfall data. Knowing the value of \( r \) for each \( t \) with \( t = 1, 2, \ldots, m \), the trend components \( r x_{t-1} \) can be calculated and deducted from \( x_t \) to give the sample values of \( \epsilon_t \).

For the rainfall data in the French Broad River Basin, \( m = 36 \) hr and \( N = 28 \) yr. The values of \( \epsilon_t \) were computed for both shifted and unshifted data. The mean and the standard deviation of the random components were studied; they have shown that shifting had improved significantly the stability, regularity, and consistency of the proposed model, and it produced relatively regular trends in the results. The shifted data were therefore used in the subsequent analysis.

The random components \( \epsilon_t \) are then analyzed for their probability distribution. Because \( \epsilon_t \) can show both positive and negative values, an arbitrary corrective value \( k \) may be added to it so that the corrected random components

\[
\epsilon'_t = k + \epsilon_t \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (5)
\]

are always positive. For the given rainfall data it was found that the corrected random components \( \epsilon'_t \) fit most satisfactorily a lognormal probability distribution. The parameters of the distribution and the Markov-chain coefficient were determined by the method of least squares and are shown in Table 1.

Sequential Generation.—Sequential generation is a statistical process using Monte Carlo methods to produce a random sequence of hydrologic data on the basis of a stochastic model for the hydrologic process. Because of the pos-

---


sibility of obtaining a large quantity of data, the sequence so generated makes possible a relatively comprehensive study of the performance of a simulated hydrologic system as it responds to various possible combinations of hydrologic events, thus helping the development of well-balanced hydrologic designs.

The quantity of the hydrologic data to be generated may be estimated on the basis of required statistical levels of errors and confidence, although the

<table>
<thead>
<tr>
<th>Time t</th>
<th>Random Component $\varepsilon_t$ in inches per hour</th>
<th>Markov-chain coefficient $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>1</td>
<td>0.026024</td>
<td>0.077331</td>
</tr>
<tr>
<td>2</td>
<td>0.008175</td>
<td>0.029459</td>
</tr>
<tr>
<td>3</td>
<td>0.012746</td>
<td>0.050650</td>
</tr>
<tr>
<td>4</td>
<td>0.004524</td>
<td>0.024413</td>
</tr>
<tr>
<td>5</td>
<td>0.001823</td>
<td>0.037586</td>
</tr>
<tr>
<td>6</td>
<td>0.023102</td>
<td>0.059666</td>
</tr>
<tr>
<td>7</td>
<td>0.003615</td>
<td>0.050224</td>
</tr>
<tr>
<td>8</td>
<td>0.035573</td>
<td>0.071749</td>
</tr>
<tr>
<td>9</td>
<td>0.024673</td>
<td>0.040318</td>
</tr>
<tr>
<td>10</td>
<td>0.020016</td>
<td>0.057241</td>
</tr>
<tr>
<td>11</td>
<td>0.063589</td>
<td>0.068550</td>
</tr>
<tr>
<td>12</td>
<td>0.029690</td>
<td>0.057588</td>
</tr>
<tr>
<td>13</td>
<td>0.025850</td>
<td>0.038250</td>
</tr>
<tr>
<td>14</td>
<td>0.039189</td>
<td>0.040759</td>
</tr>
<tr>
<td>15</td>
<td>0.065654</td>
<td>0.068157</td>
</tr>
<tr>
<td>16</td>
<td>0.111417</td>
<td>0.092583</td>
</tr>
<tr>
<td>17</td>
<td>0.136319</td>
<td>0.120192</td>
</tr>
<tr>
<td>18</td>
<td>0.068817</td>
<td>0.123015</td>
</tr>
<tr>
<td>19</td>
<td>0.051496</td>
<td>0.138377</td>
</tr>
<tr>
<td>20</td>
<td>0.101350</td>
<td>0.105346</td>
</tr>
<tr>
<td>21</td>
<td>0.037676</td>
<td>0.121696</td>
</tr>
<tr>
<td>22</td>
<td>0.097949</td>
<td>0.082689</td>
</tr>
<tr>
<td>23</td>
<td>0.032777</td>
<td>0.086571</td>
</tr>
<tr>
<td>24</td>
<td>-0.010003</td>
<td>0.079708</td>
</tr>
<tr>
<td>25</td>
<td>0.00010</td>
<td>0.061605</td>
</tr>
<tr>
<td>26</td>
<td>0.004913</td>
<td>0.038871</td>
</tr>
<tr>
<td>27</td>
<td>0.010790</td>
<td>0.030799</td>
</tr>
<tr>
<td>28</td>
<td>0.017027</td>
<td>0.045676</td>
</tr>
<tr>
<td>29</td>
<td>0.003173</td>
<td>0.025335</td>
</tr>
<tr>
<td>30</td>
<td>0.003515</td>
<td>0.027063</td>
</tr>
<tr>
<td>31</td>
<td>0.0000074</td>
<td>0.013618</td>
</tr>
<tr>
<td>32</td>
<td>0.001185</td>
<td>0.022453</td>
</tr>
<tr>
<td>33</td>
<td>0.008741</td>
<td>0.029584</td>
</tr>
<tr>
<td>34</td>
<td>0.006101</td>
<td>0.052536</td>
</tr>
<tr>
<td>35</td>
<td>0.002127</td>
<td>0.016433</td>
</tr>
<tr>
<td>36</td>
<td>0.004364</td>
<td>0.019318</td>
</tr>
</tbody>
</table>

optimal size may be determined more realistically by comparing the cost of the increased sample size with the benefits of the corresponding increase in accuracy, provided the benefit and cost data are available. For practical purposes, a univariate population may be assumed. The proportion $p_n$ of a
sample size $n$ taken from the population can be estimated within an error level of $\alpha$ per cent of its true value at a confidence level of $\beta$ per cent. The variance of the estimate $\hat{p}_n$ of $p_n$ is

$$v(\hat{p}_n) = \frac{1}{n} \left[ \hat{p}_n (1 - \hat{p}_n) \right]$$

When $n$ is large, the sample value of $p_n$ in $\beta$ per cent of the cases will fall in the range

$$\hat{p}_n = p_n \pm t_{\beta} \left[ p_n (1 - p_n) \right]^{0.5}$$

in which $t_{\beta}$ is the standard normal deviate corresponding to the confidence level $\beta$. To keep $p_n$ within an error level $\alpha$ per cent of its true value at $\beta$ per cent confidence,

$$\left| \frac{\hat{p}_n - p_n}{p_n} \right| \leq \alpha$$

Combining Eqs. 7 and 8, the sample size is

$$n \geq \left[ \frac{t_{\beta}}{\alpha} \right]^2 \left( \frac{1 - p_n}{\hat{p}_n} \right)$$

In the present analysis it is assumed that $\beta = 80\%$, $\alpha = 10\%$, and $p_n = 15\%$. Thus, Eq. 9 gives $n = 930$ or approximately 1,000. That is, the sample size required to have a confidence level of 80% that any of the 15% proportions of the sample is not different from the actual value by more than 10% is approximately 1,000. Therefore, 1,000 storms were to be generated in the analysis.

For use in the sequential generation, there are many methods to generate random numbers. In this study, random numbers were generated by two subroutine computer programs that are available at the University of Illinois, Urbana, Ill., to represent the following relationships:

$$x_n = 5^{15} x_{n-1} \pmod{2^{36}}$$

and

$$x_n = \left[ 2^9 + 1 \right] x_{n-1} + 0.788 \times 2^{35} \pmod{2^{35}}$$

in which $x_n$ is the $n$th sample generated. These programs generate pseudo-random numbers$^4$ for a uniform distribution; i.e., a rectangular distribution over the interval $(0, 1)$.

The generated random numbers of a uniform distribution for computing any hourly rainfall should be tested for their representativeness by applying the chi-squared and $t$ tests of significance to the means. If appreciable differences between the population means and the means of the generated data are indicated by these tests, either the generating function should be changed or, if the differences are not too large, the generated data should be so transformed or corrected that the differences become small, thereby eliminating the bias.

Because the probability distribution of the rainfall data was found to be different from a uniform distribution, it is necessary to convert the generated
random numbers of uniform distribution to those of the desired distribution by means of inverse probability integral transformation.

The cumulative probability function \( y = P(x) \) of any distribution of \( x \) has a range of \( 0\% \) to \( 100\% \) and therefore can be considered to possess a uniform distribution over the interval \((0, 1)\). By taking a uniform \((0, 1)\) sample of \( y \) and then taking the inverse \( x = P^{-1}(y) \) of the probability integral transformation \( y = P(x) \), a sample \( x \) is produced. By repeating this for 1,000 samples, a sequence of random numbers having the desired distribution can be developed from a sequence of uniform \((0, 1)\) random numbers.

In practice, the probability integral transformation can be performed in various ways. In this study, a relatively simple method was used; this was to interpolate \( x \) values for corresponding \( y \) values from a table of values for \( y = P(x) \), applying a polynomial of suitable degree for the interpolation.

The sample rainfall data are generated for each successive hour on the basis of the rainfall in the previous hour according to the Markov-chain model formulated before. Knowing the Markov-chain coefficients derived from the historical data and the generated random components, a random sample of 1,000 storms can be developed. According to Eq. 4a, the first hourly rainfall is equal to the random component. Thus, 1,000 first hourly rainfalls or random components were generated by the procedure previously described. The generated first hourly rainfalls were used to compute 1,000 second hourly rainfalls by means of Eq. 4b to which the coefficient \( r \) was derived from the historical data as given in Table 1 and the random components \( E_2 \) were generated beginning with an initial random number continuous with the random number used in the generation of the last value of the first hourly rainfalls.

To generate the third hourly rainfall, the trend components in Eq. 4b were computed from the generated second hourly rainfalls and the corresponding Markov-chain coefficient, and the random components were generated beginning with the random number continuous with the one used in the generation of the last value of the second hourly rainfalls. This procedure was then repeated for other successive hourly rainfalls until 1,000 hourly rainfalls for all the 36 hr were generated. The sequentially generated data are compared with the historical data in Figs. 3 to 5.

**ANALYSIS OF BASIN SYSTEM**

*Separation of Abstractions and Baseflow.*—Part of storm rainfalls over the drainage basin is abstracted by infiltration, evapotranspiration, depression storage, etc. For practical purposes and simplicity, the rate of abstractions is assumed to be uniform when it is less than the rate of rainfall, and otherwise it is equal to the rate of rainfall. This uniform rate is referred to as the "abstraction index."

Because the baseflow cannot be precisely understood at the present stage of knowledge, its estimation is generally empirical and conventional. The method used here is to represent the baseflow by an exponential curve plus a straight line of slope \( S_0 \) which follows the curve. The exponential curve corresponds to the composite groundwater recession curve derived from the runoff hydrographs. This curve is then fitted between the time of rise and the time of peak discharge. The initial baseflow at the time of rise is assumed equal to the actual discharge. After the time of peak discharge, the baseflow
is fitted with the straight line until the end point of the direct runoff is reached. The end point on a runoff hydrograph is estimated from comparison of the hydrograph with the composite recession curve.

According to the foregoing procedures, the abstraction index, the initial baseflow, and the slope \( S_0 \) of the linear part of the assumed baseflow were found to be quite random (Figs. 6 and 7). Because of the small percentage of baseflow in cases of high runoff, such randomness may be ignored. However, the variation of the abstraction index with the total storm rainfall, as shown in Fig. 8, indicates that the random component is significant. A model for the abstraction index was therefore assumed so that the standard deviation of the random component varies linearly from zero at 1 in. of total rainfall to 0.076 in. per hour at 3 in. of total rainfall and again to zero at 10 in. of total rainfall.

![FIG. 8.—VARIATION OF ABSTRACTION INDEX WITH TOTAL STORM RAINFALL](image)

Routing Model.—The routing model of a basin system is used to simulate the system for transforming the effective rainfall to direct runoff. Many conceptual routing models of the type of an instantaneous unit hydrograph have been developed in hydrology.\cite{11} For practical purposes and simplicity, a series of \( n \) equal linear reservoirs of storage coefficient \( K \) is used in this analysis. The instantaneous unit hydrograph for this model is in the form of a gamma distribution.
in which \( u(t) \) is the ordinate of the instantaneous unit hydrograph at time \( t \). The product \( nK \) represents the lag time of the centroid of the instantaneous unit hydrograph. The rainfall and runoff relationship is represented by the convolution integral

\[
Q(t) = \int_0^{t'} u(t-\tau) I(\tau) d\tau \tag{13}
\]

in which \( Q(t) \) is the ordinate of the direct runoff at time \( t \), \( u(t-\tau) \) is the ordinate of the instantaneous unit hydrograph given by Eq. 12 for time \( t-\tau \), \( I(\tau) \) is the ordinate of the effective rainfall hyetograph at time \( \tau \), \( t_0 \) is the time at the end of the effective rainfall duration, and \( t' \) is the upper limit of the integral.

From the values of effective rainfall and direct runoff derived from the historical data, the values of \( n \) and \( K \) were evaluated by the method of least squares for fitting the data to the assumed model. Each direct runoff hydrograph was shifted in time until the sum of squared errors between the routed and the actual hydrographs became minimum and the corresponding values of \( n \) and \( K \) were determined. The relationships between the total effective rainfall and the values of \( nK \), \( K \), and the actual direct runoff peak discharge are shown in Fig. 9 by curves which are fitted to the plotted data.

ANALYSIS OF RUNOFF DATA

Sequential Generation.—The runoff data can be generated by routing the sequentially generated rainfall data through the assumed routing model for the basin system.

For each of the 1,000 generated annual storms, the total rainfalls and effective rainfalls are known. The corresponding abstraction index, lag time \( nK \), and parameter \( K \) can be estimated from Figs. 8 and 9. The trend line in Fig. 8 indicates, however, only the deterministic component of the abstraction index. For each storm the random component of the abstraction index was developed by sequential generation using the abstraction-index model described previously. This was then added to the deterministic component to give the sample value of abstraction index. The generated rainfall storm is corrected by the generated abstraction-index to produce a generated effective rainfall hyetograph. The instantaneous unit hydrograph is determined in terms of \( n \) and \( K \) by Eq. 12. The generated effective rainfall hyetograph is convoluted with the determined instantaneous unit hydrograph according to Eq. 13 to produce the generated direct runoff hydrograph. The peak of the direct runoff is then noted from the hydrograph, and the corresponding values of lag time and \( K \) are determined from Fig. 9. Using these new values of system parameters, effective rainfall is again routed through the system. This procedure is repeated until the resulting peak direct runoff and the system parameters become reasonably consistent.

The preceding procedure was applied to 1,000 generated annual storms, resulting in 1,000 generated direct runoff hydrographs. The initial baseflow
and the slope $S_0$ were estimated for each storm from Figs. 6 and 7. The hourly baseflow was then determined and added to the generated direct runoff hydrographs, producing 1,000 generated total runoff hydrographs. The peak discharges of these generated hydrographs are plotted with those of the historical data on a lognormal probability paper in Fig. 10. The generated data appear to differ somewhat from the historical data at the lower and the
very high ranges of the discharges. Because the lower ranges are not of considerable importance to this study which deals with floods, the differences at the lower ranges can be ignored although they could be further reduced by improving the assumption on the variation of the standard deviation of the abstraction index. At the high ranges, the generated data are believed to be more representative than the historical data when two higher floods on record were also taken into account. As detailed precipitation data are unavailable for these floods, they were not used in this study. However, their peak discharges are known and their effect should not be ignored. On the basis of the proposed statistical analysis and the assumed basin system simulation, the generated runoff data can be considered a satisfactory representative sample of the annual floods in the French Broad River Basin at Bent Creek, North Carolina.
FIG. 12.—CUMULATIVE PROBABILITY DISTRIBUTION OF THE DURATION $D_i$ FOR WHICH A GIVEN DISCHARGE $q_i$ IS EXCEEDED IN AN ANNUAL FLOOD

FIG. 13.—STOCHASTIC FLOW-DURATION CURVES FOR FLOODS OF GIVEN RECURRENCE INTERVALS
Stochastic Analysis.—The generated annual floods can be used for a numerical analysis of the stochastic characteristics of the floods in the basin. The time distribution of floods may be represented by the statistical distribution of the time interval that a given discharge is exceeded during the annual flood.

Referring to the definition sketch of a flood hydrograph as shown in Fig. 11, a given discharge $q_i$ is exceeded at the interval $D_i$ during the flood. The value of $D_i$ varies from flood to flood, and it is zero for floods with peaks less than $q_i$. Thus, $D_i$ will have a continuous distribution censored at zero. This distribution may be evaluated from the historical data but the results will be very approximate because the historical data usually provide only few sample points at the higher discharges. From the generated data the distribution can be estimated more broadly because the generated data will provide a large number of sample plots for the analysis.

From the generated flood hydrographs, sample values of $D_i$ for which a given $q_i$ is exceeded can be found and their probability distributions of $D_i$ for given $q_i$ can be determined. For this study, the cumulative probability distributions so obtained are plotted on normal probability paper in Fig. 12, in comparison with those for the historical data. All the probability curves appear to show a definite trend and tend to become vertical at lower durations and horizontal at higher durations.

In this analysis, a flood of given frequency may be defined as a hypothetical flood during which the discharge exceeds any rate of flow for the time-duration that corresponds to the given frequency. Obviously, the peak discharge for the flood of a given frequency will be identical with that used in the conventional frequency analysis. From the stochastic analysis of the generated floods, it is then possible to develop the flow-duration curve of the annual flood with a given frequency. For example, consider the annual flood with a recurrence period of 100 yr which corresponds to a probability of 0.01 for being equal or greater than $q_i$. The intercept of the probability curve in Fig. 12 for the given $q_i$ with the probability of 0.99 for being less than $q_i$ will give the duration $D_i$ for which $q_i$ is exceeded with the given frequency. The durations $D_i$ corresponding to several values of $q_i$ to the given frequency can be thus determined and then plotted against $q_i$ to obtain the flow-duration curve for the flood of 100-yr recurrence interval. This procedure is repeated for other recurrence intervals of 50 yr, 200 yr, 500 yr, and 1,000 yr. The flow-duration curves for all these frequencies are shown in Fig. 13. Because the stochastic characteristics of the floods are considered in this analysis, such curves may be called "stochastic flow-duration curves." The curve for average annual floods will apparently be near the conventional flow-duration curve.

CONCLUSIONS

This study demonstrates that storm rainfalls may be treated by a finite-discrete-duration, nonstationary, stochastic process that is amenable to mathematical formulation and analysis. For the river basin under consideration, the process for annual storms can be represented by a nonstationary Markov-chain model with lognormally distributed random components. Using suitable model formulations and Monte Carlo methods, a large representative
sample of rainfall and runoff data may be developed by sequential generation and simulation.

The storm-shifting analysis seems to be useful in the analysis of finite duration stochastic processes. In this study, it results in consistent, stable and regular models for the process of hourly rainfall distribution.

Sequentially generated data are shown to be useful in analyzing stochastic characteristics of the complex hydrologic process and also on investigating the relative importance and interdependence of several components in the process. For practical applications, the analysis will produce stochastic flow-duration curves or possibly stochastic hydrographs for use in hydrologic system designs.

An important use of the sequentially generated runoff data is, of course, for the design of optimal water resources systems by simulation. For example, the generated floods may be routed through alternate system designs. By comparing the resulting performances of the various designs, the optimal design may be evolved.

The generated hydrologic data may also be used to study other stochastic characteristics of rainfall and runoff. For example, they may be used to estimate the conditional probability distribution of the peak discharge or stage for given rainfalls or runoffs, or both, at a given time interval after the beginning of the storm. Such a study may help develop sound flood regulating procedures.

ACKNOWLEDGMENTS

For most numerical analyses in this study the IBM 7094-1407 system, partially supported by a grant GP-700 from the National Science Foundation, of the Digital Computer Laboratory and the facilities at the Statistical Services Unit, both of the University of Illinois, Urbana, Ill., were used. For the collection of hydrologic data on the French Broad River Basin, generous help and cooperation were received from the Tennessee Valley Authority and the U. S. Geological Survey. This study is a part of a continuing research project on stochastic hydrology directed by the senior author. It also constitutes the major work performed for a doctoral thesis\(^{15}\) by the junior author under the direction of the senior author.

APPENDIX.—NOTATION

The following symbols have been adopted for use in this paper:

\[
D_i = \text{time interval during a flood that a given discharge is exceeded;}
\]

\[
f_{t,1}(x_t-1) = \text{trend component of hourly rainfall;}
\]

\(^{15}\) Ramaseshan, S., "A Stochastic Analysis of Rainfall and Runoff Characteristics by Sequential Generation and Simulation," thesis presented to the Univ. of Illinois, at Urbana, Ill., in 1964, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Civil Engineering.
I(τ) = ordinate of an effective rainfall hyetograph;
K = storage coefficient of a linear reservoir;
k = a corrective value;
m = order of hourly rainfall;
N = number of annual storms;
n = sample size; also, number of linear reservoirs in series;
P(x) = cumulative probability function of x;
p_n = proportion of a sample size from population;
\hat{P}_n = estimate of p_n;
Q(t) = ordinate of a direct runoff hydrograph;
q_i = a given discharge on a flood hydrograph;
R_{xy}(v) = cross-correlation coefficient for a time shift of \ V \ hours \ of \ storm \ y \ with \ reference \ to \ storm \ x;
r = the Markov-chain coefficient;
S_o = slope of a straight line for baseflow separation;
s_{x,t} = standard deviation of x with respect to t;
t = time, in hours;
t_0 = time at end of the effective rainfall duration;
t_r = time for a hydrograph to rise to a given discharge;
t' = upper limit of the convolution integral;
u(t) = ordinate of an instantaneous unit hydrograph;
v = time shift, in hours;
v(\hat{p}_n) = variance of \hat{p}_n;
x = designation for an annual storm;
x_n = n-th sample to be generated;
x_t = sequence of hourly rainfalls of an annual storm;
y = designation for an annual storm; also, P(x);
y_x = sequence of hourly rainfalls in an annual storm;
α = error level, in per cent;
β = confidence level, in per cent;
ε_t = random component of hourly rainfall;
ε'_t = corrected ε_t; and
τ = time, in hours.