JOURNAL OF THE
BOSTON SOCIETY OF CIVIL ENGINEERS

Volume 59 JANUARY 1972 Number 1

HYDROLOGIC MODELING

by

Ven Te Chow*

(The Seventh John R. Freeman Memorial Lecture, presented before the Boston Society of Civil Engineers, 17 February 1972.)

Introduction

During the early years in hydraulic engineering practice, say prior to sixty years ago, a spillway designed to pass flood 50 to 100 percent larger than the largest which had occurred in a period of record as long as 25 years, was considered adequate. This design criterion is essentially a rule of thumb which involves an arbitrary factor of safety. The concept of factor of safety may be applicable to other engineering designs, but it was soon found out that this practice in hydraulic design is after all entirely inadequate. As one example of absurdities to this criterion, the Republican River in Nebraska was noted to experience a flood in 1935 which was over ten times as large as had ever occurred on that river during 40 prior years of record.

Practicing hydrologists and hydraulic engineers then began to search for better methods of hydrologic design. In 1887, for example, Professor Arthur Talbot of the University of Illinois derived the well-known Talbot formula to determine waterway areas.[1] In 1889, the city engineer of Rochester, New York, Mr. Emil Kuichling, devised the popular rational formula for the computation of peak discharges from urban watersheds.[2] In 1921, Allen Hazen, a consulting engineer, suggested the logarithmic probability analysis of flood data.[3] In 1930, a committee of the Boston Society of Civil Engineers[4] reported that flood hydrographs afford the best basis for the study of drainage areas and concluded that "the base of the flood hydrograph appears to be approximately constant for different floods" and "peak flows tend to vary directly with the total volume of flood runoff." Two years later, another consulting engineer Leroy Sherman proposed the theory of unit hydrographs[5], which is practically based on the conclusions of the BSCE committee.

In recent years, a host of mathematically sophisticated methods of hydrologic analysis have been developed. All these methods and those proposed in the earlier years for hydrologic design are essentially techniques of hydrologic

*Professor of Hydraulic Engineering, University of Illinois, Urbana, Illinois.

modeling. Most natural hydrologic phenomena are so complex that they are beyond human comprehension, or that exact laws governing such phenomena have not yet been fully discovered. Before such laws can ever be found, complicated hydrologic phenomena (the prototypes) can only be approximated by modeling.

Type of Hydrologic Models

It has been said that modeling is the process of approximating the prototype for the purpose of evaluating the performance of the prototype. Since the prototypes under consideration are more complex than models can ever be, certain simplifying modeling assumptions must be made in order to provide the model with a presentable or workable form. According to such assumptions, hydrologic models may be classified into various types for the sake of understanding.

Hydrologic models can be divided into two basic categories: models that possess certain physical properties of their prototypes; and models that have only an abstract form. The former category, or the physical models, can be divided into scale models (iconic models), analog models, and simulation models.

A scale model that looks like the prototype is the simplest type. It is exemplified by ordinary hydraulic models of rivers and structures that are investigated in many hydraulic laboratories, and whose scales are based on geometric and force considerations.[6]

An analog model replaces prototype properties with quantities that bear the same relations to each other as do those of the prototype, but they are easier to measure or visualize. For example, the Hele-Shaw model shows the movement of a viscous liquid between two closely spaced parallel plates is analogous to seepage flow in a two-dimensional cross-section of an aquifer.[7] Many electronic analog models for surface and groundwater flows are built on the principle of analogy between the flow of water and the flow of electrical current.[8]

A simulation model retains the essence of the prototype without actually attaining reality itself. It reproduces the behavior of a hydrologic phenomenon in every important detail but does not reproduce the phenomenon itself. In a broad sense, it is commonly used to include the scale and analog models, but the definition adopted here refers specifically to the simulation on digital computers. In hydrology, the Stanford Watershed Model[9] may be therefore described as a simulation model. This model simulates the land phase of the hydrologic cycle in a watershed on a digital computer.

Abstract models, or the second basic category, are generally referred to as theoretical, or mathematical, models since they attempt to represent the prototype theoretically in a mathematical form. These models neither resemble nor imitate prototypes physically but replace the relevant features of the system by a set of mathematical relationships. According to certainty, or uncertainty, of such relationships, on *a priori* basis, the models can be further divided into deterministic and indeterministic types. A differentiation between deterministic and indeterministic models can be assisted by relating them to the concepts of certainty and uncertainty. Certainty implies that no matter how many times a hydrologic phenomenon is processed under a given set of invariant conditions the same outcome is assumed to result always. On the other hand, uncertainty implies that every time a phenomenon is produced it may be different. Theoretically, certainties may be forecasted while the risk aspect of uncertainties can be predicted with an element of probability. In this sense, therefore, deterministic models make forecastings, while indeterministic models make predications.

Abstract models are the product of modern age, since these quantitative models must depend on adequate mathematical tools which have now become available for practical applications. Such models to be useful must inevitably be complex, yet at the same time be workable. These requirements could not be compromised without the availability of high-speed computers for the solution of the models. Since numerous abstract models have been developed lately and they constitute a significant advance in modern hydrology, this paper will emphasize discussions on these models. Major types of these models as well as the physical models are classified diagrammatically in Fig. 1.

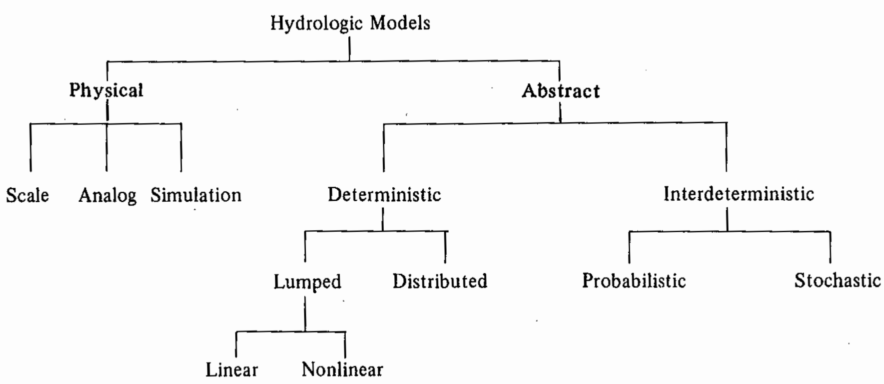

Fig. 1. Classification of Hydrologic Models

In abstract modeling, hydrologic phenomena are treated as systems. By this so-called systems concept. the hydrologic system is considered to consist of an input, an output, and some working medium known as throughput such as the water passing through the system. For example, a watershed can be analyzed as a system. For this system, the input is the rainfall and groundwater inflow; the output is the evapotranspiration, infiltration, and runoff; and the throughput is the water moving over the watershed. By systems concept, a hydrologic phenomenon can be readily interpreted by modern system-analysis techniques, and then modeled mathematically and solved on computers.

Mathematically, the input and output relationship of a hydrologic system may be represented by

$$Q = \phi I \tag{1}$$

in which Q is the output, I is the input, and ϕ is a transfer function which represents the operation performed by the system on the input to transform it into output. For example, the unit hydrograph is a transfer function of the watershed system. It should be noted that input I and output Q are time functions and can be also expressed by I(t) and Q(t), respectively, with t denoting time. The objective of modeling is essentially to derive a mathematical formulation for the transfer function of the system.

Deterministic Modeling

In deterministic modeling, a hydrologic system is often treated either as lumped or as distributed, although this treatment is equally applicable to indeterministic modeling. A lumped-system model is a gross representation of the hydrologic system as determined from the input and output data pertaining to the system, thus the system is regarded as a single point in space without dimensions. In contrast to this is the distributed-system model which considers the hydrologic processes that are taking place within various distributed points or areas within the internal space of the system. If the internal space is divided into a number of small unit spaces and each unit space is modeled as a lumped system, then the distributed-system model becomes simply a conglomeration of lumped-system models. For demonstrative purposes, each of the two system models may be exemplified as follows:

A. A Lumped-System Model

This example describes a general deterministic model.[10] By applying the basic concept of system continuity to a hydrologic system, the following continuity equation may be written:

$$I - Q = S \tag{2}$$

in which S is the storage of water in the system and thus S is the first derivative of S with respect to time, or $\dot{S} = dS/dt$. The system storage is affected by the throughput and reflects the characteristics of the system. For given physiographic features of a hydrologic system, the storage is likely a function of the input and output and their changes. Therefore, the system storage may be expressed as a mathematical function of the input I and output Q and their derivatives with respect to time, or

$$S = f(I, \dot{I}, \ddot{I}, \ldots, \overset{m}{I}, Q, \dot{Q}, \ddot{Q}, \ldots, \overset{n}{Q}) = f \tag{3}$$

in which $\dot{I}, \ddot{I}, \ldots, \overset{m}{I}$ are the first, second, . . . , n-th derivatives of I; and $\dot{Q}, \ddot{Q}, \ldots, \overset{n}{Q}$ are the first, second, . . . , n-th derivatives of Q. Because I, and Q are time functions, S is also a function of time.

If the input I continues steadily, the output, Q, will increase and finally approach a steady state at which $I = I_*$, $Q \to Q_*$ and all derivatives of I and Q with respect to time will approach zero. The subscript $_*$ indicates the steady state. At the steady state, the storage may be assumed essentially proportional directly to I and Q; thus, Eq. 3 becomes

$$S_* = f(I_*, Q_*) = aI_* + bQ_* \tag{4}$$

in which a and b are coefficients.

Expanding Eq. 4 in Taylor's series about the steady state[11] yields

$$S = \sum_{m=0}^{m} a_m(I,Q) \overset{m}{I} + \sum_{n=0}^{n} b_n(I, Q) \overset{n}{Q} \tag{5}$$

where $a_m(I,Q) = \overset{m}{\partial f/\partial I}$ and $b_m(I,Q) = \overset{n}{\partial f/\partial Q}$. Because these coefficients a_m and b_m are functions of I or Q, or both, the above differential equation is nonlinear. If the coefficients are constant or independent of I and Q, then the equation becomes linear and its solution will be greatly simplified, leading to a linear model. For practical purposes, the coefficients may be assumed as functions of certain characteristic values of I and Q, such as the average input $\bar{I}$ and average output $\bar{Q}$, of the peak input I_p and peak output Q_p. Thus, for a particular hydrologic event, the coefficients are constants and the equation is linear and can be readily solved for the case of that event. However, the characteristic values of I and Q obviously vary from event to event. After the equation is solved linearly for a particular event, the system storage can be then considered as nonlinear by expressing the coefficients as functions of the characteristic values of I and Q.

Substituting Eq. 5 for S in Eq. 2 and solving for Q, a system equation containing a transfer function in the form of Eq. 1 may be written as

$$Q = \left(- \frac{a_m D^{m+1} + a_{m-1} D^m + \ldots + a_o D - 1}{b_n D^{n+1} + b_{n-1} D^n + \ldots + b_o D + 1} \right) I \qquad (6)$$

in which $D^m = d^m/dt^m; D^n = d^n/dt^n$; etc. It can be seen that the transfer function is in the form of a fraction in which the numerator and the denominator are polynomials that may be denoted by M(D) and N(D), respectively. Thus, Eq. 6 is written as

$$Q = \frac{M(D)}{N(D)} I \qquad (7)$$

Eq. 6 or 7 is the general lumped-system hydrologic model. It can be shown that many mathematical models that have been proposed are special cases of this general model.

For example, let $M(D) = - a_o D + 1$ and $N(D) = b_o D + 1$. Then simplifying Eq. 7 and using Eq. 2, it can be shown that

$$S = a_o I + b_o Q \qquad (8)$$

which is the well-known Muskingum equation or the linear model used in the Muskingum flood routing method.[12]

Now, let $N(D) = 1$ and $a_{j-1} = (-1)^{j-1} C^j / j!$ with $j = 1, 2, \ldots, m + 1$, where C is a constant. Then, using Taylor's expansion of I(t-C) about t, it can be shown that

$$M(D)\ I(t) = I(t\text{-}C) \qquad (9)$$

Thus, Eq. 7 becomes

$$Q(t) = I(t-C) \tag{10}$$

which is the mathematical expression for the well-known linear channel model.[13]

Now, let M(D) = 1 and N(D) has a root of 1/K or N(D) = 1 - KD. Then, Eqs. 7 and 2 yield

$$S = KQ \tag{11}$$

which is the mathematical expression for the well-known linear reservoir model.[13] In a similar manner, it can be shown further that if M(D) = 1 and N(D) has n real roots of $1/K_1, 1/K_2, \ldots, 1/K_n$; Eq. 7 will result in the Nash model representing a series of linear reservoirs.[14]

For the application of Eq. 6 or 7 to practical problems it is necessary to know the values of m and n; that is, the highest orders, respectively, of the derivatives of I and Q. In general, the values depend on the problem under consideration. In order to determine the most suitable values of m and n for flood studies, various values m and n have been assumed and hydrologic data were then fitted to Eq. 5. By comparing the storage values computed by different fitted forms of Eq. 5 for various values of m and n, with the corresponding actual storage values computed directly from the hydrologic data, it was found that the most suitable values of m and n are, respectively, 1 and 2. Derivatives of I and Q of higher order were found to be insignificant and they can be dropped without causing appreciable errors in fitting the model. For practical application, therefore, the following simplified form of Eqs. 5 and 6 may be used, assuming constant coefficients:

$$S = a_0 I + a_1 \dot{I} + b_0 Q + b_1 \dot{Q} + b_2 Q \tag{12}$$

and

$$Q = \left(\frac{1 - a_0 D - a_1 D^2}{1 + b_0 D + b_1 D^2 + b_2 D^3}\right) I \tag{13}$$

Using the above lumped-system model to fit the storage, effective rainfall and direct runoff for the storm of April 4-9, 1941 on Wills Creek watershed, near Cumberland, Maryland, the results are shown in Fig. 2.

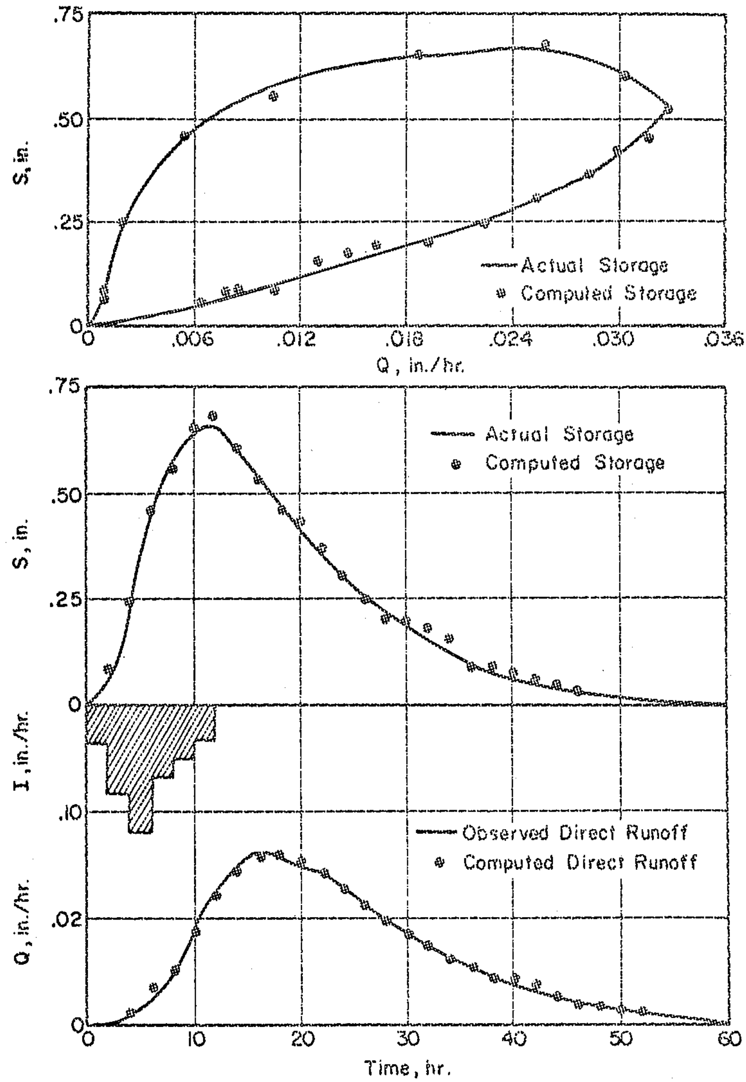

Fig. 2 Storage, effective rainfall and direct runoff graphs for storm of April 4-5, 1941 on Wills Creek watershed

B. A Distributed-System Model

As the distributed-system model considers the internal space of the system, its mathematical formulation must contain space dimensions or coordinates. The simplest distributed-system model to describe the flow of surface water over a watershed is the application of the kinematic-wave theory[15] resulting in a one-space-dimensional model[16]. In fact, such a model has been applied to Mekong Delta flood studies.[17] The example to be given here is a two-space-dimensional model for the surface flow over a watershed due to rainfall.[18] The physical configuration of the watershed model is impervious and rectangular. It consists of two equal-sized rectangular plane surfaces for accommodating overland flows, which intersect to form a trough in the middle of the watershed for serving the main channel flow. The overland flows are assumed to travel in a direction normal to the line of intersection of the two side planes toward the channel flow. The watershed flow thus consists of two flow components, namely the overland flows and the channel flow. Since the watershed is assumed impervious, the watershed flow involves only the surface flow. The influence of the raindrop impact is assumed as an overpressure head added to the hydrostatic pressure head. The surface roughness is represented by the Darcy-Weisbach resistance factor. The distributed-system model of the watershed flow is a mathematical and hydrodynamic formulation considering the curvilinear nature of the streamlines of two dimensions on the watershed surface; while the third dimension, or the depth of flow, is taken as the dependent variable. The velocity and pressure are averaged over a vertical depth of flow and across the flow cross-sectional area. For this model, the unsteady spatially varied equations of continuity and momentum are normalized in order to reduce the equations to dimensionless forms and to save computation effort. The watershed length and the critical values of depth and velocity for the equilibrium discharge at the outlet are used as the terms of normalization. Thus, in a vector form, the normalized hydrodynamic equations to represent the watershed model are as follows:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} + \frac{\partial G}{\partial z} = H \tag{14}$$

where
$$U = \{y;\ q;\ p\}$$
$$F = \{q;\ q^2/y + P_1 y^2;\ qp/y\}$$
$$G = \{p;\ qp/y;\ p^2 y + P_1 y^2\}$$
$$H = \{I;\ P_2 y - P_4(f + f^*)q(q^2 + p^2)^{0.5} y^{-2} + VI;$$
$$P_3 y - P_4(f + f^*)p(q^2 + p^2)^{0.5} y^{-2} + P_5 VI\}$$

in which y = normalized depth of flow; q = normalized discharge in the main channel direction; p = normalized discharge in the lateral direction; I = normalized rainfall intensity; V = normalized vertical raindrop velocity; f = Darcy-Weisbach resistance coefficient; f^* = apparent resistance coefficient due to raindrop impact; $P_1 = (1 - \sin^2\theta - \sin^2\Psi)^{0.5}/2\cos\theta$; $P_2 = I\tan\theta/y_c$; $P_3 = P_2 P_5$; $P_4 = L/8y_c$; and $P_5 = \sin\Psi/\sin\theta$. In addition, θ = watershed slope angle in the main channel direction; Ψ = watershed slope angle in the lateral direction; L = watershed length in the main channel direction; and y_c = critical depth for the equilibrium discharge at watershed outlet.

To examine the nature of Eq. 14, the following two matrices are defined:

$$A = \frac{dF}{dU} = \begin{vmatrix} 0 & 1 & 0 \\ a_2^{\,2} - 2a_1 a_2 & 2a_1 & 0 \\ -a_1\beta_1 & \beta_1 & a_1 \end{vmatrix}$$

$$B = \frac{dF}{dU} = \begin{vmatrix} 0 & 0 & 1 \\ -a_1\beta_1 & \beta_1 & a_1 \\ \beta_2^{\,2} - 2\beta_1\beta_2 & 0 & 2\beta_1 \end{vmatrix}$$

where $\alpha_1 = g/y, \alpha_2 = q/y + (2P_1y)^{0.5}$ and $\alpha_3 = q/y - (2P_1y)^{0.5}$, all being the eigenvalues of A; and $\beta_1 = p/y$, $\beta_2 = p/y + (2P_1y)^{0.5}$ and $\beta_3 = p/y - (2P_1y)^{1/2}$, all being the eigenvalues of B. All these eigenvalues are real and distinct, but they are neither positive definite, nor negative definite. If all three eigenvalues of either A or B were positive, or negative, definite, Eq. 14 would be of elliptic type. If they were definite but had different signs, Eq. 14 would be of hyperbolic type.[19] In the proposed watershed model, the eigenvalues of A and B do not fit these cases; thus Eq. 14 is neither elliptic nor hyperbolic although it is quadratic.

Eq. 14 can be solved on computer by a four-step scheme of numerical integration combining the concepts of Lax-Wendroff,[20] Burstein[21] and Lapidus.[22] This scheme is to express Eq. 14 in a second-order accurate difference form, which includes an artificial viscosity term that sharpens discontinuities but has no significant effect on the continuity of streamlines and the accuracy of the results.

In applying the proposed scheme to the solution of the watershed flow model, the following major assumptions are made: The falling raindrops are sufficiently dense so that their mass and momentum flux can be considered continuous around any grid point of the computer solution. The effect of raindrop impact on the surface flow is taken as an addition to the Darcy-Weisbach resistance factor. The watershed flow is not affected by surface tension and has a momentum coefficient equal to one. Its initial depth and velocity profiles of the flow are given, and there is no flow across any boundary except the outlet. Fig. 3 shows the profiles of depth and velocity vector on half of a square watershed, a typical computer output of the distributed-system model.

Indeterministic Models

Indeterministic behavior of hydrologic phenomena may be described in many ways. One tangible approach is to hypothesize the risk in uncertainty as definable by an element of probability. In fact, this does not imply that all uncertainties can be measured in terms of probability. On the basis of this understanding, a simple modeling concept can be taken by assuming that hydrologic events are purely random variables. In this way, hydrologic data have been analyzed by many mathematical models of probability distribution. Among the commonly used such probability models are the lognormal distribution,[23] Gumbel's extreme-value distribution,[24] and the log-Pearson Type III distribution recently recommended by the Water Resources Council.[25]

However, the concept of prediction implied in the indeterministic model is more than one of pure randomness since the occurrence of hydrologic event may be affected by its antecedent event or events. In fact, it has been discovered that

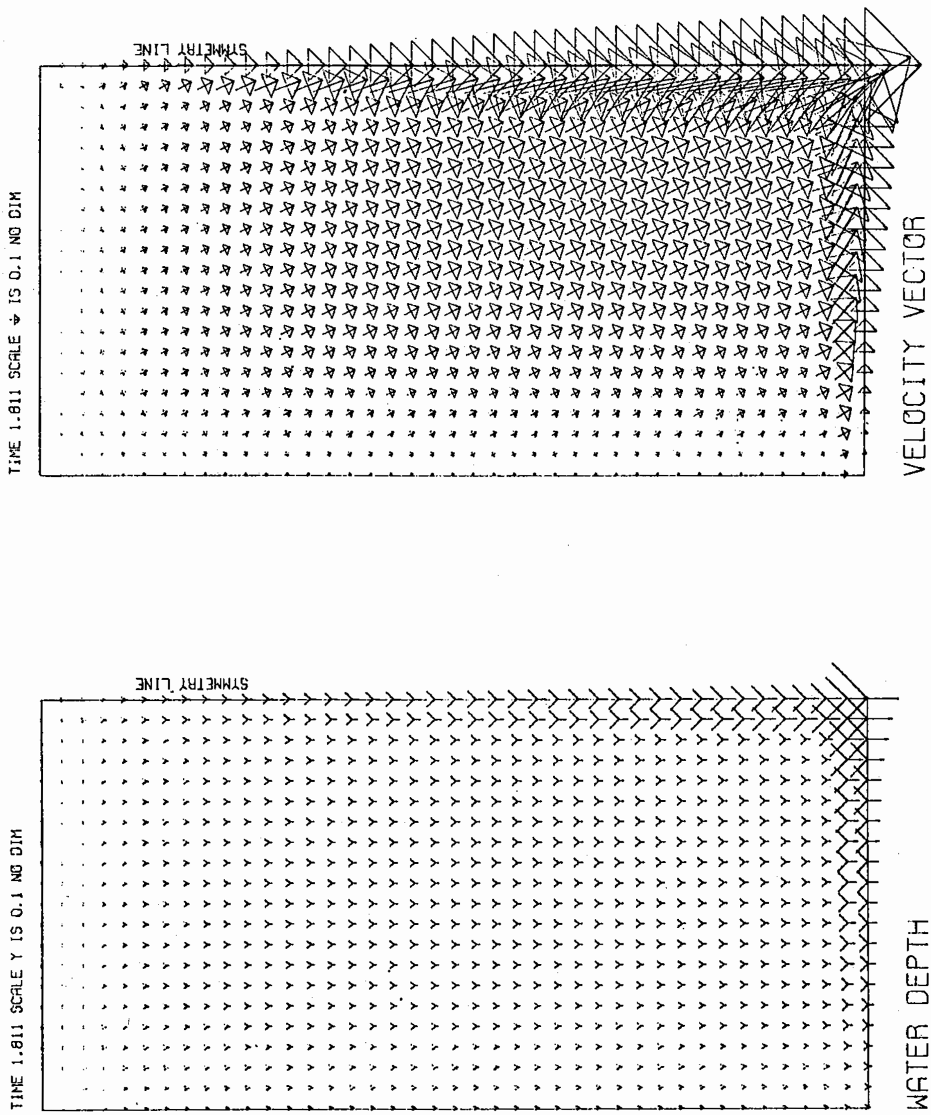

Fig. 3. Computer printout of profiles of water depth and velocity vector on half of a square watershed.

the variability of groups of recorded streamflows in their natural order of occurrence is actually larger than if the same flows occurred in random sequence.[26] This means that hydrologic events do not usually occur in random sequences. By assuming hydrologic events as pure random variables is simply to ignore the effect of sequence since the variables may occur in different sequences but not in a random fashion. In order to cope with this situation, the hydrologic process may be treated as stochastic processes. For example, the record of a hydrologic phenomenon may be analyzed as a time series and thus, available mathematical models of time series can be used as stochastic models to represent the hydrologic process involved. The theoretical meaning of stochastic process may be further elaborated.[27]

A probability distribution is the distribution of a random variable whose specific value cannot be predicted exactly except in terms of chances. Mathematically speaking, a random variable $X(\omega)$, $\omega \epsilon \Omega$ describable by a sample point ω is a function defined on the sample space Ω of an experiment such that for every real number of the variate x there exists a probability $P[\omega; X(\omega) \leq x]$. Once the probability function is formulated for a given problem, it is independent of when or where it occurs except under either a given or an average condition. However, a random variable may have a different probability distribution for each point on the time scale, or for each point on the space coordinates. Because of their dependence on time or space and because there can be, at least in theory, an infinite number of them, these families of random variables constitute the so-called stochastic process. Thus, a stochastic process is a family of random variables $[X(t, s) ; t \epsilon T, s \epsilon S]$ that depend on a parameter or parameters which belong to an indexing set T or S, or indexing sets T and S, of time t or space s, or both.

In a way, it can be seen that the deterministic process and the "purely" probabilistic process are only two special cases of the stochastic process. When the probability or certainty of the random variable is one, the stochastic process simply reduces to deterministic. When the probability is independent of any parameter index, time or space, and the family of random variables belongs to the same population, the stochastic process becomes purely probabilistic, in which no deterministic components exist. On a scale of probability from 0 to 1, the purely probabilistic and the deterministic cases occupy respectively the two extremities, while the stochastic process may occur anywhere between them. Take, for instance, the simple first-order Markov chain model which is a stochastic model. This model consists of two terms; namely, the trend term and the noise term. For the special cases, the noise term may be zero, thus producing a deterministic model; or the trend term may be zero, then resulting in a probabilistic model.

Today, stochastic modeling is at the highest level of hydrologic modeling, although it has not been well developed in view of many practical difficulties yet to be overcome. By stochastic modeling, all components of a hydrologic system can be theoretically described by stochastic processes.[28-32] In the system, the input, the output and the transformation of input to output in the form of throughput passing through the system may be therefore represented mathematically by time series since these component processes, in general, change with time and are functions of time. The transformation of input to output is characterized by the physical features and hydrologic behavior of the system. All the processes are assumed to be governed by mathematically simulated stochastic laws. They may be denoted by $[u_t;\ t \epsilon T]$ where u_t is a stochastic variable at time t which is a parameter running over an index set T or over the time range under consideration. Thus, the input stochastic process is denoted by $[X_t;\ t \epsilon T]$ where X_t is the input stochastic variable; the output stochastic process by $[Y_t;\ t \epsilon T]$ where Y_t is the output stochastic variable; and the throughput stochastic process, representing the transformation of input to output, by $[Z_t;\ t \epsilon T]$ where Z_t is the throughput stochastic variable. These stochastic processes can be simply denoted by $[X_t]$, $[Y_t]$, and $[Z_t]$ respectively. They may not be considered as independent but as a stochastic vector $[X_t, Y_t, Z_t; t \epsilon T]$ or $[X_t, Y_t, Z_t]$.

The time parameter t in the stochastic processes may be either continuous or discrete. For practical and analytical purposes and for a possible solution of the mathematically simulated model by digital computers, the stochastic processes may be taken as discrete time functions. The index set T represents a length of time long enough to describe the hydrologic phenomenon under consideration. Units of the time parameter t can be chosen in convenient time intervals so that for the integral values of t = 1, 2, . . ., T, the stochastic variables define the respective processes in satisfactory detail. It should be noted that the time interval to be chosen for the discrete time parameter will affect the simulated stochastic laws of the processes. In general, smaller time intervals will make the stochastic laws more complicated as the magnitude and extend of dependence among the stochastic variables based on the historical hydrologic data of a process will be greater and in more detail.

The input-and-output relationship of a stochastic hydrologic system may be represented mathematically by a system equation as Eq. 1:

$$[Y_t] = \phi \{[X_t],\ [Z_t]\} \tag{15}$$

where $\phi \{[X_t], [Z_t]\}$ is the transfer function that represents the operation performed by the system on the input and the throughput in order to transfer them into output.

In most cases, the input, output and throughput of a hydrologic system are amounts of water, although in certain cases they can be taken as energy or other forms of medium. By the basic principle of system continuity, the output is equal to the input minus the throughput which is the amount of flow in the system. Thus, a single transfer function may be written similar to Eq. 2:

$$\phi\{[X_t], [Z_t]\} = [X_t] - [Z_t] \qquad (16)$$

Hence, from Eqs. 15 and 16, the hydrologic system equation becomes

$$[Y_t] = [X_t] - [Z_t] \qquad (17)$$

For the t-th time interval, or the time interval from t to t+1, Eq. 17 may be written as

$$Y_t = X_t - Z_t \qquad (18)$$

where X_t is the input in the t-th time interval, Y_t is the output in the t-th time interval, and Z_t is the change of throughput in the t-th time interval.

In order to demonstrate the use of the above equation, the watershed is taken as a hydrologic system which has the precipitation as its input, the streamflow or runoff as its output, and the change in basin storage, due to storage and depletion as well as to evapotranspiration plus other losses, as its throughput. For the watershed, Eq. 18 becomes

$$Y_t = X_t - E_t - S_{t+1} + S_t \qquad (19)$$

where X_t is the total amount of precipitation input to the watershed during the t-th time interval; Y_t is the total amount of runoff output from the watershed during the t-th time interval; E_t is the total amount of evapotranspiration plus other losses during the t-th time interval; S_t, is the basin storage at time t; and S_{t+1} is the basin storage at time t+1.

To illustrate the application of the above mathematical formulation of stochastic hydrologic systems, a stochastic annual storm-flood model for watershed systems may be discussed. The annual storm-flood is the runoff of a flood produced by an annual storm. The annual storm is a storm which produced the maximum peak discharge of flood flow in a water year. Therefore, for N water years of storm and flood records, there are N annual storms and N annual storm-floods.

The annual storm is treated as the input stochastic process to the watershed system which transforms the annual storm into the annual storm-flood. The stochastic process of the annual storm is denoted by $[X_t;\, t \epsilon T]$ where the time increment for t in general may be conveniently taken as one hour and T is the duration of the storm considered in the analysis. It is evident that the hourly rainfall process in an annual storm is nonstationary as the probability of transition between the hourly rainfalls of a storm changes with time since the storm began and hence depends on the time of transition. The nonstationary discrete-time process can be described by a first-order nonhomogeneous Markov chain to represent the stochastic process of the annual storm under consideration. This Markov chain may be written as

$$X_t = \lambda_t X_{t-1} + \epsilon_t \tag{20}$$

where X_t is the stochastic variable of hourly rainfall in the annual storm, λ_t is the Markov or regression coefficient, and ϵ_t is the random component of X_t. The subscript t implies that the process is nonstationary as the process and its parameters change with time in the process.

The transition probability of the hourly rainfalls in the annual storm process is

$$g_t(x_t \mid x_{t-1}) = P[X_t = x_t \mid X_{t-1} = x_{t-1}] \tag{21}$$

where the subscript t indicates the nonstationarity and x_t is the variate of the stochastic variable X_t. Thus,

$$\begin{aligned} P(X_t = x_t) &= \sum_{x_{t-1}} P[X_t = x_t \mid X_{t-1} = x_{t-1}]\, p(X_{t-1} = x_{t-1}) \\ &= \sum_{x_{t-1}} g_t(x_t \mid x_{t-1})\, p(X_{t-1} = x_{t-1}) \end{aligned} \tag{22}$$

where $p(X_{t-1} = x_t)$ is the probability for $X_{t-1} = x_{t-1}$. Eq. 22 implies that the probability of the outcome x_t depends on the probability of its antecedent x_{t-1} and this dependence is represented by the transition probability of Eq. 21.

The annual storm-flood depends not only on the antecedent annual storm-rainfall but also on the corresponding physical condition of the watershed. It can therefore be assumed that the hourly flood flow Y_t in the t-th time interval depends upon the hourly rainfall X_{t-1} of the annual storm at the (t-1)-th time interval as well as upon the basin storage S_t at time t. Corresponding to the time increment for hourly rainfalls, the increment of t for flow is also taken as one hour. The process $[Y_t; t \epsilon T]$ can then be suitably described by

$$Y_t = \phi_t(X_{t-1}, S_t) + \epsilon_t \tag{23}$$

where $\phi_t(X_{t-1}, S_t)$ is some function of X_{t-1} and S_t to represent the deterministic component of Y_t, and ϵ_t is a random variable uncorrelated with X_{t-1} and S_t but to provide the random component of Y_t. The subscript t implies that the process and its components are all nonstationary. The length of T is the duration of the flood to be considered in the analysis.

When sufficient data are available, the function $\phi_t(X_{t-1}, S_t)$ can be derived by multiple regression. For the case under consideration, a multiple linear regression is found suitable. Thus, Eq. 23 may be written as

$$Y_t = a_t X_{t-1} + b_t S_t + c_t + \epsilon_t \tag{24}$$

where a_t and b_t are nonstationary regression coefficients and c_t is the nonstationary intercept of the linear regression line-of-fit. The random component ϵ_t may be assumed as normally and independently distributed or as distributed according to a probability law suitable to the given data.

In the analysis of annual storms and annual storm-floods, the evapotranspiration and other losses in the hydrologic process may be ignored because they are insignificant in the relatively short durations of the storms and floods under consideration. Thus, Eq. 19 reduces to

$$S_{t+1} = X_t - Y_t + S_t \tag{25}$$

The phenomenon of transforming hourly rainfall to hourly flood flow in the watershed system, as influenced by the basin storage, may be described by a one-step bivariate nonhomogeneous Markov process which is represented by a family of two-dimensional stochastic vectors $[X_t, S_{t+1}]$ as

$$P_{ij}(t) = f_t(x_t, s_{t+1} \mid x_{t-1}, s_t)$$

$$= P[X_t = x_t, S_{t+1} = s_{t+1} \mid X_{t-1} = x_{t-1}, S_t = s_t] \qquad (26)$$

where $P_{ij}(t)$ is the transition probability of the bivariate process from state i at t-1 to state j at t.

Since Eq. 23 assumes that the flood flow Y_t is a function of X_{t-1} and S_t, the transition probability of the hourly flood flows for the annual storm-flood process is

$$h_t(y_t \mid x_{t-1}, s_t) = P[Y_t = y_t \mid X_{t-1} = x_{t-1}, S_t = s_t] \qquad (27)$$

where x_{t-1}, y_t and s_t are respectively the variates of the variables X_{t-1}, Y_t and S_t. Thus, the probability of the flood flow is

$$P(Y_t = y_t) = \sum_{x_{t-1}, s_t} h_t(y_t \mid x_{t-1}, s_t) P(X_{t-1} = x_{t-1}, S_t = s_t) \qquad (28)$$

From Eqs. 21 and 27, it can be shown[32] that

$$p_{ij}(t) = g_t(x_t \mid x_{t-1}) h_t(y_t \mid x_{t-1}, s_t) \qquad (29)$$

Let the two-dimensional state vector $[X_t, S_{t+1}]$ of the bivariate Markov process assume discrete values $\delta [X_t\ S_{t+1}]_k$ which can be represented by a point in a two-dimensional plane. The coordinates of this point are X_t and S_{t+1}. If X_t can assume discrete states m = 1, 2, . . ., M, where each state represents a convenient range of hourly rainfall amounts, and S_{t+1} can assume discrete states n = 1, 2, . . ., N, where each state represents a convenient range of hourly basin storages, then the state vector $[X_t, S_{t+1}]$ can assume MN discrete states, i.e., K = MN.

Following Eq. 26 the transition probability of the bivariate process from state i at t-1 to state j at t is

$$f_t(x_t, s_{t+1} \mid x_{t-1}, s_t) = p_{ij}(t), i,j = 1, 2, \ldots, K \qquad (30)$$

For each t = 1, 2, . . ., T, there will be a stochastic matrix of size K. Hence, there will be T stochastic matrices, each of size K, for the proposed bivariate process.

Now, assume the initial probability for the bivariate process, $p_j(0)$ for j = 1, 2, . . ., K. Since the initial rainfall is always zero, only the probability of the initial basin storage should be assumed. The absolute probability of the bivariate Markov process can be derived inductively[32] to be

$$p_j(t) = \sum_{i=1}^{K} p_i(t-1)\, p_{ij}(t) \quad \text{for } j = 1, 2, \ldots, K \tag{31}$$

where $p_j(t) = p(X_t = x_t, S_{t+1} = s_{t+1})$ and the state j refers to the state of the state vector $[X_t, S_{t+1}]$.

Summing up the transition probability $p_{ij}(t)$ of Eq. 29 over all the values of x_t,

$$\begin{aligned} \sum_{x_t} p_{ij}(t) &= \sum_{x_t} g_t(x_t \mid x_{t-1})\, h_t(y_t \mid x_{t-1}, s_t) \\ &= h_t(y_t \mid x_{t-1}, s_t) \end{aligned} \tag{32}$$

Substituting this equation in Eq. 28, the probability distribution of the flood flow Y_t is

$$p(Y_t = y_t) = \sum_{x_t, x_{t-1}, s_t} p_i(t-1)\, p_{ij}(t) \tag{33}$$

where $p_i(t\text{-}1) = p(X_{t-1} = x_{t-1}, S_t = s_t)$ is, as shown by Eq. 31, the absolute probability of the bivariate process being in state i at time t-1, and $p_{ij}(t)$ is, as shown by Eq. 30, the transition probability of the bivariate process from state i at t-1 to state j at t.

The basin storage depends on the input rainfall and the storage at the beginning of the rainfall interval. From the joint probability, the marginal probability of the basin storage S_{t+1} can be calculated by

$$p(S_{t+1} = s_{t+1}) = \sum_{x_t} p(X_t = x_t, S_{t+1} = s_{t+1}) = \sum_{x_t} p_j(t) \tag{34}$$

where the summation is over all values of x_t for m = 1, 2, . . ., M.

The stochastic annual storm-flood model is now represented by Eq. 25 in which the components X_t, Y_t and S_t are expressed by Eqs. 20, 24 and 25 itself, respectively, and their probabilities by Eqs. 22, 31 and 32, respectively. As an exercise, this model has been applied to the French Broad River Basin in North Carolina. Twenty-seven annual-storm records (1935-62) and hydrographs of the corresponding floods were used to determine the model.[32] Fig. 4 shows the computed hydrograph by this model, or Eq. 24, as compared with the observed hydrograph for the annual storm-flood of the hydrologic year 1937-38. An exact agreement between the two hydrographs is not expected because the computed hydrograph is generated stochastically and the scattered portion of the observed data may be caused by the random component.

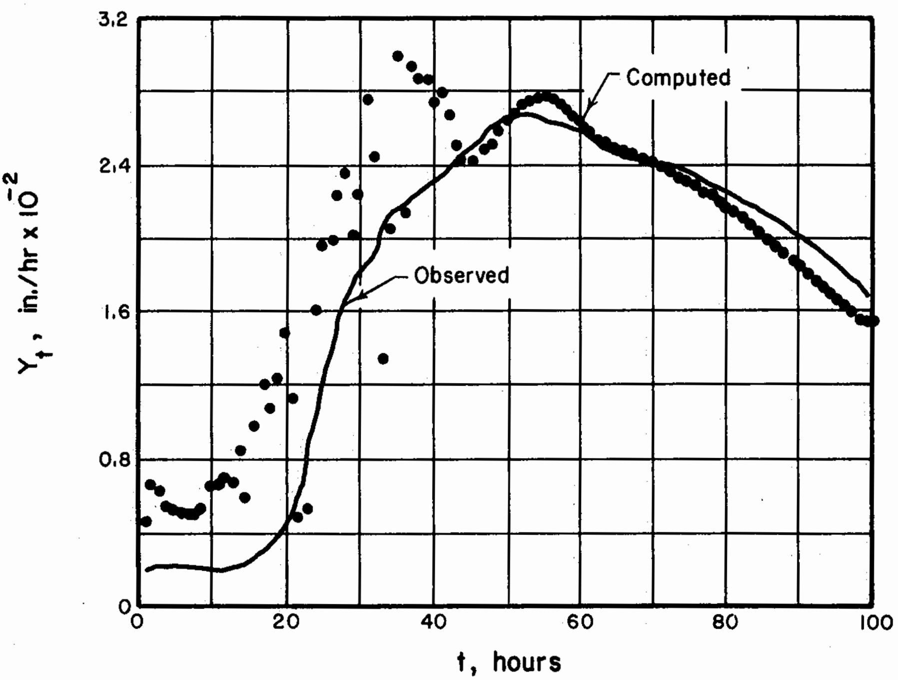

Fig. 4 Computed and Observed Hydrographs for Hydrologic Year 1937-38

The above example is a relatively complex stochastic model as it is the result of a recent research, and it should be further improved for practical use. Such complex models are generally handicapped by the limited amount of existing hydrologic data and the storage capacity of available computers. For the models to be practicable, more data and much larger computer memory would be required. However, for practical purposes, much simpler stochastic models are now available.

Practical Considerations

As mentioned earlier, deterministic models make forecasts, while indeterministic models make predictions. A deterministic model forecasts by transforming one sequence of hydrologic events (the input to a hydrologic system) into another sequence (the output from the system). An indeterministic model predicts by generating sequences of hydrologic events. The generating process may be done by using a mathematical random-number generator to produce random values on a computer and then feed the latter into the random element of the model, such as ϵ_t in Eq. 20 or 24, thus producing various output sequences.[33] The generating process will take into account the order of events in the sequences. Theoretically speaking, if the indeterministic model is correctly built, the statistical characteristics of the historical data which determine the model must be preserved in the generated sequences. This principle provides a criterion to test the validity of an indeterministic model.

From the above discussion, it can be seen that the two types of abstract model are complementary rather than competitive in the practical application. It is perfectly possible to take the generated output from an indeterministic model, a sequence of events that could occur, and transform them using a deterministic model. The deterministic model then produces a forecast of what would happen, given what could happen. By this procedure, Chow and Ramaseshan[34] generated runoff data from historical rainfall data for the French Broad River Basin in North Carolina. In this case, given a predicted rainfall sequence the runoff resulting from the rainfall could be forecast.

The two types of model can also be used independently. For example, an indeterministic model built on historical streamflow could predict future streamflows. Such predicted streamflows can be used to design a water resource system over the period of planning projection. Fig. 5 shows one historical sequence and 98 stochastically generated sequences of average monthly flows which are used for the design study of the Trans-Texas Water System.[35] Selected sequences from such water supply and demand sequences were then used to help find the optimal implementation plan, or the plan to be developed at various stages of the planning horizon having the least total expected cost. For such a study, Fig. 6 shows the present value cost response surface where the least expected cost of 4.34 billion dollars occurs when the change in canal capacity is $\Delta C = -1000$ cfs, i.e., so much less than a given design capacity, at every interval of time $\Delta T = 10$ years.

On the other hand, for short period data, the actual input to a deterministic model may already be available. For example, deterministic flood-routing models are widely used for flood forecasting in river control works. Such a model is an operational one and would, for this case, forecast the

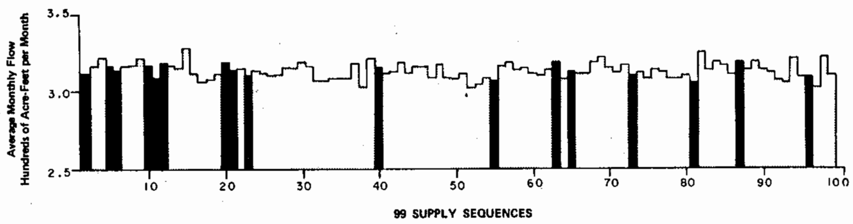

Fig. 5. 99 Average Monthly Flow Sequences Used in the Trans-Texas Water System Study[35].

downstream flood that would occur from the observed upstream flood. Used independently as so described, the two types of model have separate purposes: indeterministic models are tools for planning, while deterministic models are tools for operation.

The procedure of hydrologic modeling may be either analytic or synthetic, although the distinction cannot be too exact. The analytic procedure is to assume certain physical principles that can be theoretically described. Given these principles, a means of analysis is developed, and the product is an analytic model. For example, Eq. 3 for the development of a general deterministic model is a principle which is described by a mathematical expression and soluble by mathematical analysis. The procedure is therefore analytic. As for the synthetic procedure, it is to combine the elements, or subsystems, of a system into an entity which will represent the system; and the product is a synthetic model. For example, The Stanford Watershed Model[9] is built by this procedure, and the entity to be represented is the land phase of the hydrological cycle of a watershed. If the way of combining the subsystems is not well defined or generally agreed upon, the synthetic model may vary between model builders. The subsystems are building blocks, and they may be individual analytic models. Thus, an analytic model usually describes a narrow and restricted system in hydrology so that the problem is made manageable, whereas a synthetic model can be complicated and cover as broad a scope of hydrologic phenomena as one wishes.

Whatever form the model takes it inevitably contains a number of unknowns, i.e., parameters, that serve to characterize a given hydrologic phenomenon. To be justifiable in a model's structure and to make it practically useful, it is desirable that the parameters be physically meaningful and capable of estimation. This point is of vital importance if the model is used to assess the effects of natural or man-made changes in the hydrologic system, or if a model developed from a known hydrologic region is to be transposed to an unknown

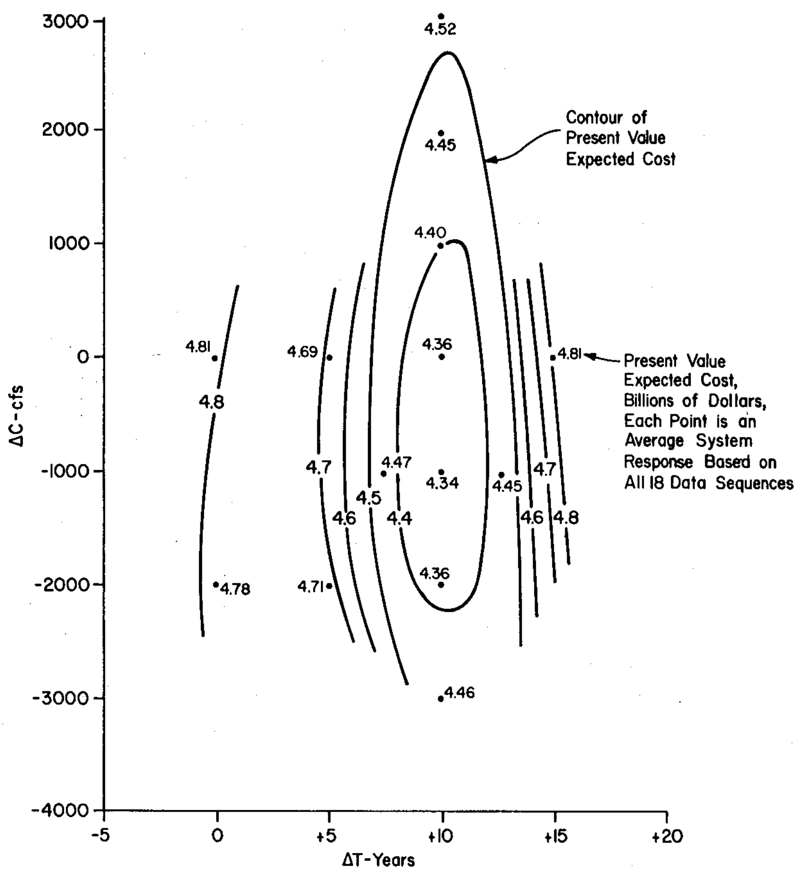

Fig. 6. Present value cost response surface [35]

hydrologic region. At present, hydrologic modeling technique has not advanced sufficiently to allow reliable estimates of the absolute values of the parameters to be determined. Furthermore, the parameters are not independent of one another; it is their relative magnitudes that count. This explains why certain models can satisfactorily reconstruct events for the hydrologic system to which they were fitted but are useless when applied to situations where a knowledge of the absolute values of certain parameters is required. When too many parameters are involved, usually in synthetic models, the interdependence of parameters is frequently aggravated. In a synthetic model, a wide variety of data can be used in the determination of the parameters. The model structure is not predetermined and it can be built to use whatever data are available. This flexibility is both a virtue and an evil, because each model can be tailored to fit a given hydrologic phenomenon but this has led to a vast proliferation of models whose relative merits are uncertain.

During the process of modeling or selecting of a model, one is invariably faced with a decision concerning simplicity versus completeness of the model. A simpler model is easier to understand and apply, and probably cheaper to use. This preference has resulted in an oversimplified and unrealistic model, such as the rational formula. On the other hand, a good model should involve minimum assumptions and approximations. This choice has led to certain components being more sophisticated than others and thus has biased the overall model structure towards special objectives. Both the analytic and synthetic modelings face the same problem, but the decisions are not easy to make as the effects of assumptions and approximations are not so obvious. To be effective, the model builder should be at least capable to distinguish the relative significance of the assumptions and approximations and between the components of the model. A desire for completeness in a model may result in one which is unwieldly to manage and may contain various components that are of incomparable significance. It has been suggested that if a simple model will do, no more complexity is necessary. This rule fails in practice unless "do" is better defined.

Conclusions

This paper has introduced briefly the status of hydrologic modeling today. Hydrologic models are classified according to various assumptions. Emphasis is given to abstract models and their examples are illustrated. In practice, hydrologic modeling is partly science and partly art. It is an art since ingenuity and judgement enter into the modeling process and its assessment. As a science, it requires an advanced knowledge in physical principles and mathematical methods, but a main problem is to develop models with their practical application in mind. Many practicing engineers hesitate to accept new techniques, because they find difficulties in understanding and applying them. It is therefore apparent that the practical value of hydrologic modeling in the future will depend on a better communication between practicing engineers and theoretical hydrologists.

Acknowledgments

Part of the material used in this paper was developed by funds provided by the U.S. Department of the Interior Office of Water Resources Research as authorized under the Water Resources Research Act of 1964, P.L. 88-379, Agreement No. 14-01-0001-1632, and by the National Science Foundation research grant GK-11292.

References

1. Talbot, A.N., "The determination of water-way for bridges and culverts", Selected Papers of the Civil Engineers' Club, Technograph No. 2, University of Illinois, 1887-8, pp. 14-22.
2. Kuichling, E., "The relation between the rainfall and the discharge of sewers in populous districts", Transactions, American Society of Civil Engineers, Vol. 20, 1889, pp. 1-56.
3. Hazen, A., Discussion of a paper on "Probable variations in yearly runoff" by L.S. Hall, Transactions, American Society of Civil Engineers, Vol. 84, 1921, pp. 214-224.
4. "Report of the Committee on Floods", Journal of Boston Society of Civil Engineers, Vol. 17, No. 7, 1930, pp. 285-464.
5. Sherman, L.K., "Stream flow from rainfall by the unit-graph method", Engineering News-Record, Vol. 108, 1932, pp. 501-505.
6. "Hydraulic Models", Manuals and Reports on Engineering Practice No. 25, American Society of Civil Engineers, 1942; reprinted 1958.
7. Hele-Shaw, H.S., "Experiments on the nature of the surface resistance in pipes and on ships", Transactions, Institution of Naval Architects, Vol. 39, 1897, pp. 145-156.
8. Chow, V.T., "Application of electronic computers in hydrology" in Handbook of Applied Hydrology, McGraw-Hill Book Co., New York, 1964, Sec. 29, pp. *29*-3 to *29*-14.
9. Crawford, N.H., and Linsley, R.K., The synthesis of continuous streamflow hydrographs on a digital computer, Department of Civil Engineering, Stanford University, Technical Report No. 12, July 1962.
10. Chow, V.T., and Kulandaiswamy, V.C., "General hydrologic system model", Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers, Vol. 97, No. HY6, June 1971, pp. 791-804.
11. Cosgriff, R.L., "Nonlinear Control Systems", McGraw-Hill Book Co., New York, 1958.
12. Method of flood routing, "Report on Survey for Flood Control, Connecticut River Valley", Vol. 1, Sec. 1, appendix, U.S. Army Corps of Engineers, Office of the District Engineer, Providence, R.I., 1936.
13. Chow, V.T., "Runoff", Handbook of Applied Hydrology, McGraw-Hill Book Co., New York, 1964, Sec. 14, p. *14*-27.
14. Nash, J.E., "The form of the instantaneous unit hydrograph", Publication No. 45, International Association for Scientific Hydrology, Vol. 3, 1957, pp. 114-121.
15, Lighthill, M.J., and Whitham, G.B., "Kinematic waves", Proceedings, Royal Society of London, Vol. A-229, 1955, pp. 281-316.
16. Wooding, R.A., "A hydraulic model for the catchment-stream problem", Journal of Hydrology, Vol. 3, Nos. 3-4, 1965, pp. 254-282; Vol. 4, No. 1, 1966, pp. 21-37.
17. Zanobette, D., Lorgeré, H., Preissmann, A., and Cunge, J.A., "Mekong Delta mathematical model program construction", Journal of the Waterways and Harbors Division, Proceedings, American Society of Civil Engineers, Vol. 96, No. WW2, May 1970, pp. 181-199.
18. Chow, V.T., "Role of WES in the development of hydrodynamic watershed model", Proceedings, International Symposium on Mathematical Models in Hydrology, Warsaw, Poland, 26-31, July 1971, Vol. 2, Pt. 2, Topic 5, pp. 1-14.
19. Garabedian, P.R., "Partial Differential Equations", John Wiley, New York, 1964.
20. Lax, P.D., and B. Wendroff, "Difference schemes for hyperbolic equations with higher order of accuracy", Communications on Pure and Applied Mathematics, Vol. 17, 1964, pp. 381-393.
21. Burstein, S.Z., "Finite-difference calculations for hydrodynamic flows containing discontinuities", Journal of Computational Physics, Vol. 1, No. 2, 1966, pp. 198-222.
22. Lapidus, A., "A detached shock calculation by second order finite differences", Journal of Computational Physics, Vol. 2, No. 2, 1967, pp. 154-177.

23. Chow. V.T., "The log-probability law and its engineering applications", Proceedings, American Society of Civil Engineers, Vol. 80, Paper No. 536, November 1954, pp. 1-25.
24. Gumbel, E.J., "The return period of flood flows", Annals of Mathematical Statistics, Vol. 12, No. 2, June 1941, pp. 163-190.
25. "A uniform technique for determining flood flow frequencies", Water Resources Council, Washington, D.C., December 1967.
26. Leopold, L.A., "Probability analysis applied to a water supply problem", U.S. Geological Survey, Washington, D.C., Circular 410, 1959.
27. Chow, V.T., "Stochastic hydraulics – a challenging field of study, Stochastic Hydraulics", Proceedings of the First International Symposium on Stochastic Hydraulics, Pittsburgh, Pennsylvania, May 31–June 2, 1971, pp. 3-8.
28. Chow, V.T., "A general report on new ideas and scientific methods in hydrology" (Simulation of the hydrologic behavior of watersheds), Proceedings, International Hydrology Symposium, Fort Collins, Colorado, 6-8 September 1967, pp. 50-65.
29. Chow, V.T., "Hydrologic systems for water resources management", Conference Proceedings of Hydrology in Water Resources Management, Water Resources Research Institute Report No. 4, Clemson University, Clemson, South Carolina, March 1968, pp. 8-22.
30. Chow, V.T., "Stochastic analysis of hydrologic systems", Research Report No. 26, Water Resources Center, University of Illinois, Urbana, Illinois, 1969; or P.B. No. 189791, Clearinghouse of Federal, Scientific and Technical Information, U.S. Department of Commerce, Springfield, Virginia, 1969.
31. Chow, V.T., "Systems approaches in hydrology and water resources", in "The Progress of Hydrology", Proceedings, The First International Seminar for Hydrology Professors, Vol. 1, 1970, pp. 490-509.
32. Chow, V.T., "Stochastic hydrologic systems", "Systems Approach to Hydrology", Proceedings of the First Bilateral U.S.-Japan Seminar in Hydrology, Honolulu, Hawaii, January 11-17, 1971, Water Resources Publications, Fort Collins, Colorado, pp. 1-19
33. Chow, V.T., "Sequential generation of hydrologic information," Handbook of Applied Hydrology, McGraw-Hill Book Co., New York, Sec. 8-I, 1964, pp. *8*-91 to *8*-97.
34. Chow, V.T., and Ramaseshan, S., "Sequential generation of rainfall and runoff data", Journal of Hydraulics Division, Proceedings, American Society of Civil Engineers, Vol. 91, No. HY6, July 1965, pp. 205-223.
35. "Stochastic optimization and simulation techniques for management of regional water resources systems", Report 131, Texas Water Development Board, Austin, Texas, July 1971.