3
GENERATION OF SYNTHETIC FLOW SEQUENCES

N. C. Matalas

U.S. Geological Survey
Washington, D.C.

J. R. Wallis

IBM Research Center
Yorktown Heights, NY

3-1 INTRODUCTION

The design of a water-resource system is dependent, in part, upon the sequences of streamflow that are assumed to be realized over the system's economic life. Because these sequences are unknown, the design must in some way be based upon historical sequences, which could be taken to represent the sequences that will be realized in the future. In doing so, there is the implied assumption that the past will be repeated—an unrealistic assumption. The use of historical sequences to represent future sequences subjects the design to risk and provides no basis for assessing the risk or evaluating the losses associated with a system that is under- or overdesigned to an unknown extent.

More realistically, streamflow is assumed to be a stationary stochastic process. Under this assumption, the past is repeated in a statistical sense—the statistical characteristics of the process are independent of absolute time (stationarity). Given the generating mechanism of the process, an ensemble of future sequences can be generated. While no one member of the ensemble can be identified as being the sequence that will be realized in the future, the members collectively represent a set of sequences, each of which is equally likely to be realized in the future.

Given the ensemble of future sequences, alternative designs of a water-resource system can be evaluated in terms of each member of the ensemble. This affords the planner a quantitative measure of both the anticipated performance and the risk associated with each design. The use of such sequences in water-resource system planning has been discussed extensively by Hufschmidt and Fiering,[1] Fiering,[2] and more recently by Fiering and Jackson.[3]

Unfortunately, the generating mechanism of streamflow is unknown, and therefore the ensemble of future flow sequences cannot be generated. The mechanism, however, can be approximated, and with the added assumption that streamflow is an ergodic process, an ensemble of "future" flow sequences, referred to as *synthetic flow sequences*, can be generated. The assumption of ergodicity allows time averages to be used for corresponding ensemble averages.

Over the past decade, an extensive body of literature has developed covering techniques for generating synthetic flow sequences. For the most part, the techniques have been based on the use of short-memory processes to approximate those of streamflow. More recently, long-memory processes have been introduced to approximate long-term persistence that is evidenced by many historical flow sequences. Fiering[2] and Fiering and Jackson[3] have provided excellent summaries on the use of short-memory processes, particularly markovian-type processes, for generating synthetic flow sequences. Mandelbrot and Wallis,[4] Wallis and Matalas,[5] and Matalas and Wallis[6] have considered long-term hydrologic persistence and the use of fractional noise processes to account for this persistence in generating synthetic flow sequences.

In the following paragraphs, some of this literature is briefly reviewed. For the most part, attention is paid to some of the operational problems involved in generating synthetic flow sequences on a multisite, multiseason basis, including those which derive from the "paucity" of streamflow data, as well as those that derive from the operational constraints of model building.

3-2 GENERATION OF SYNTHETIC FLOWS

A historical flow sequence $X(1), \ldots, X(n)$ may be characterized by a set of statistics $\Theta = \{\theta_1, \ldots, \theta_m\}$. For example, θ_1 may denote the mean and θ_2, the standard deviation. From the historical flow sequence, estimates of $\theta_i \ \forall \ i$ may be obtained. To generate synthetic flow sequences, a model must be chosen to approximate the flow's underlying generating mechanism. This model serves to transform a sequence of random numbers, $\eta(1), \ldots, \eta(\tilde{n})$ into a sequence of synthetic flows, $Y(1), \ldots, Y(\tilde{n})$. By using several sequences of random numbers, all drawn from a common population, several synthetic flow sequences may be generated to form the ensemble of "future" flow sequences.

Via simulation of a proposed water-resource system, the ensemble of syn-

thetic flow sequences may be used to derive a set of outputs which measure the system's performance. It should be noted that, for each synthetic sequence, several measures of performance may be considered relative to the purposes to be served and objectives to be met by the system. For each alternative design, a set of system outputs can be obtained. The variations in system performance provide a basis for assessing the risks inherent in each design.

For a realistic assessment of the risks to be taken—risks upon which the selection of the "best" design is derived—the synthetic flow sequences must themselves be realistic realizations of the future flow sequence. Under the assumption that streamflow is an ergodic stochastic process, the statistical characterization of the future flow sequence will not differ from that of the historical sequence. Thus the synthetic flows must be generated in a manner that assures this statistical resemblance. Given the set of statistics Θ, a model and a population from which random numbers are to be drawn must be chosen so as to assure statistical resemblance between the historical and the synthetic flow sequences. Before discussing these choices, a definition of statistical resemblance is in order.

One definition of statistical resemblance is as follows. From a synthetic sequence of length $\tilde{n}$, the set of statistics $\tilde{\Theta}$ corresponding to the historical set Θ is obtained. The synthetic sequence is said to resemble the historical sequence if $\tilde{\theta}_i \to \theta_i \; \forall \; i$ as $\tilde{n} \to \infty$. But infinite synthetic sequences are of little use. What matters is the ensemble of synthetic sequences of length $\tilde{n}$. For each synthetic sequence of length $\tilde{n}$, the set of statistics $\tilde{\theta}$ are obtained whereby the set θ^* may be formed, where θ_i^* is the average of the ensemble of values of $\tilde{\theta}_i$. If $\theta_i^* \to \theta_i$ as $m \to \infty$, where m denotes the number of synthetic sequences forming the ensemble, then each synthetic flow sequence is said to resemble the historical flow sequence.

θ_i is an estimate obtained from a historical sequence of length n. In generating synthetic flows, θ_i assumed the role of a population value and $\tilde{\theta}_i$, a sample value. If indeed $\tilde{\theta}_i$ is an unbiased estimate of θ_i, then the two definitions of statistical resemblance are equivalent. However, estimates for many statistics are biased. If the bias depends upon only $\tilde{n}$ and tends to zero as $\tilde{n} \to \infty$, then an infinite synthetic sequence will statistically resemble the historical sequence, but each finite synthetic sequence will fail to do so.

Later the matter of biased estimates will be discussed, but for now, the assumption is made that $\tilde{\theta}_i$ is an unbiased estimate of $\theta_i \; \forall \; i$. What model should be chosen to approximate the generating mechanism of streamflow? There is no simple answer. More than likely, one will be compelled to choose a prescriptive rather than a descriptive model.[3] By a *prescriptive model* is meant one that will assure statistical resemblance relative to the set Θ. This model may or may not be descriptive of the real world. One of several models may suffice, in which case the choice may be made on the basis of which model places minimum demands on data or computation. As long as the distribution function of the flows is not of concern, an operational choice may be made for the population from which random numbers are to be drawn. The population

need only have finite statistics corresponding to the statistics forming the set Θ. If the distribution of the flows is of concern, problems, some of which are discussed below, arise in choosing the population.

3-3 STREAMFLOW STATISTICS

The above discussions have dealt with the generation of an ensemble of synthetic flow sequences at a single site. In general, the design of a water-resource system will entail the generation of an ensemble of synthetic flow sequences of a multisite, multiseason basis. The definitions of statistical resemblance discussed above apply to the multisite, multiseason case.

The flow for the tth year at the jth location may be denoted as $X(t|j)$, where $t = 1, \ldots, n$ and $j = 1, \ldots, L$. The term *location* is used to indicate a site or a season or a combination of site and season. If N and Z denote the number of sites and seasons, respectively, then $X(t|j)$ may represent one of the following cases:

1. Single site, single season: $L = 1$
2. Multisite, single season: $L = N$
3. Single site, multiseason: $L = Z$
4. Multisite, multiseason: $L = NZ$

A random variable may be described by a set of ensemble averages. Of particular importance are the expected value $E[X(t|j)]$, the second moment or variance $E[x^2(t|j)]$, and the third moment $E[x^3(t|j)]$, where $x(t|j) = X(t|j) - E[X(t|j)]$, which provide some meaningful description of the probability distribution of $X(t|j)$. Under the assumption of stationarity, each ensemble average holds $\forall t$. Unfortunately, for each location, only one realization of the stochastic streamflow process is available, namely, the n years of flow forming the historical sequence. With an infinite historical sequence, the time averages $\mu(X|j)$, $\sigma^2(X|j)$, and $\lambda(X|j)$, corresponding to the ensemble averages $E[X(t|j)]$, $E[x^2(t|j)]$, and $E[x^3(t|j)]$ theoretically could be computed. If the stochastic process is ergodic, then the ensemble averages are equal to their corresponding time averages.

Under the assumption of ergodicity, estimates of the time averages derived from the finite historical flow sequences may be taken to be estimates of the corresponding ensemble averages. These estimates are defined as follows:

$$\hat{\mu}(X|j) = \frac{\sum_{t=1}^{n} X(t|j)}{n} \tag{3-1}$$

$$\hat{\sigma}^2(X|j) = \frac{\sum_{t=1}^{n} x^2(t|j)}{n} \tag{3-2}$$

$$\hat{\lambda}(X|j) = \frac{\sum_{t=1}^{n} x^3(t|j)}{n} \tag{3-3}$$

where $x(t|j) = X(t|j) - \hat{\mu}(X|j)$. Quite often, the coefficient of skewness, defined as

$$\hat{\gamma}(X|j) = \frac{\hat{\lambda}(X|j)}{\sigma^3(X|j)} \tag{3-4}$$

rather than $\hat{\lambda}(X|j)$, is used as a statistical descriptor of streamflow.

To describe the multivariate structure of the historical flow sequences, various statistics may be used. For now, only two statistics are considered: (1) the covariances between the flow at different time points at a given location, and (2) the covariances between the flows at different time points at two given locations.

The estimates of the covariances at a given location are as follows:

$$\hat{C}(X|k,j) = \frac{\sum_{t=1}^{n-k} x(t|j)\, x(t+k|j)}{n-k} \tag{3-5}$$

Note: $x(t|j) = X(t|j) - \hat{\mu}(X|k,j)$ and $x(t+k|j) = X(t+k|j) - \hat{\mu}(X|k',j)$, where

$$\hat{\mu}(X|k,j) = \frac{\sum_{t=1}^{n-k} X(t|j)}{n-k} \tag{3-6}$$

$$\hat{\mu}(X|k',j) = \frac{\sum_{t=1}^{n-k} X(t+k|j)}{n-k} \tag{3-7}$$

The corresponding correlation coefficient is estimated by

$$\hat{\rho}(X|k,j) = \frac{\hat{C}(X|k,j)}{\sigma(X|k,j)\,\sigma(X|k',j)} \tag{3-8}$$

where

$$\sigma(X|k,j) = \left[\frac{\sum_{t=1}^{n-k} x^2(t|j)}{n-k}\right]^{1/2} \tag{3-9}$$

$$\sigma(X|k',j) = \left[\frac{\sum_{t=1}^{n-k} x^2(t+k|j)}{n-k}\right]^{1/2} \tag{3-10}$$

The estimates of the covariances pertaining to two given locations u and

v are as follows:

$$\hat{C}(X|k,u,v) = \frac{\sum_{t=1}^{n-k} x(t|u)x(t+k|v)}{n-k} \tag{3-11}$$

and the corresponding correlation coefficients are estimated by

$$\hat{\rho}(X|k,u,v) = \frac{\hat{C}(X|k,u,v)}{\sigma(X|k,u)\,\sigma(X|k',v)} \tag{3-12}$$

Still other statistics could, and perhaps should, be considered in statistically describing streamflow sequences. One in particular is the Hurst coefficient. To define this statistic, consider once again a historical flow sequence at a single site $X(1), \ldots, X(n)$. The variable

$$Y(i) = \sum_{t=1}^{i} x(t) \tag{3-13}$$

denotes the cumulative departures from the observed sequence mean, $\hat{\mu}(X)$. If $Y(i')$ and $Y(i'')$ denote the minimum and maximum values of $Y(i)$, then

$$\hat{R}(X) = Y(i'') - Y(i') \tag{3-14}$$

the range of cumulative departures, is the minimum storage needed to meet a draft equal to $\hat{\mu}(X)$ over the period of n years

In his studies of long-term storage requirements, Hurst[7,8] noted that for many long sequences of natural phenomena,

$$\frac{R(X)}{\hat{\sigma}(X)} \sim n^h \tag{3-15}$$

Hurst[7] and, independently, Feller,[9] showed that for a purely random normal process, $h \to \frac{1}{2}$ as $n \to \infty$. From a large number of sequences of natural phenomena, Hurst found estimates of h to have a mean and standard deviation of about 0.73 and 0.08, respectively. The tendency for estimates of h to be greater than $\frac{1}{2}$ is referred to as the *Hurst phenomenon*.

The statistic h is a measure of long-term hydrologic persistence—the tendency for high flows to follow high flows and for low flows to follow low flows over long periods of time. The statistic $\rho(X|k)$ is a measure of short-term persistence. If the generating process of streamflow belongs to the Brownian domain of attraction, then $h = \frac{1}{2}$ and $\rho(X|k)$ is a measure of the existing short-term persistence. Outside this domain, there exist generating processes for which $h \neq \frac{1}{2}$ (Mandelbrot and Wallis[4]). If indeed $h > \frac{1}{2}$, then the use of a generating process characterized by $h = \frac{1}{2}$ will yield synthetic sequences such that $R(X)$, the storage requirement, is underestimated. $R(X)$ will be overestimated if the synthetic sequences are generated by a process for which $h > \frac{1}{2}$ when indeed streamflow is generated by a process for which $h = \frac{1}{2}$.

In practice, other factors beside the inflows affect the design capacity of a reservoir. While $R(X)$ may not be the design capacity, it is related to it, such that under- or overestimation of $R(X)$ is associated with under- or overdesign of a water-resource system, either case involving economic losses. Some assessment of these losses may be warranted in deciding whether or not synthetic flow sequences should be generated by a model that assures statistical resemblance in terms of h.

Hurst[7] took as an estimate of h,

$$K = \frac{\log [\hat{R}(X)/\hat{\sigma}(X)]}{\log (n/2)} \tag{3-16}$$

Another estimate of h is defined by Mandelbrot and Wallis[4] as follows. The sequence $X(1), \ldots, X(n)$ is divided into sets of subsequences. Let $\nu(i)$ denote the length of the subsequences for the ith set, where $i = 1, \ldots, \eta$ and $5 \leq \nu(i) \leq n$. For the rth subsequence of the i set, the values $\hat{R}(X|r,i)$ and $\hat{\sigma}(X|r,i)$ are determined. If $\hat{R}(X|i)$ and $\hat{\sigma}(X|i)$ denote the mean values of $R(X|r,i)$ and $\sigma(X|r,i)$ for the ith set, then the slope of the relation between log $[R(X|i)/\sigma(X|i)]$ and log $[\nu(i)]$, denoted by H, is an estimate of h.

Wallis and Matalas[10] utilized Monte Carlo experiments to evaluate some statistical properties of K and H, where H was taken to be a least-square estimate of h. Both K and H were found to be biased and highly variable estimates of h. In general, H is less biased but more variable than K.

While the statistics discussed above are by no means exhaustive, they do provide considerable characterization of the multivariate structure of streamflow. Apart from $\gamma(X)$, only first- and second-order statistics have been considered. Higher order statistics could be considered, but because n is small, generally less than 50, these statistics are subject to large sampling errors. Moreover, to achieve statistical resemblance in terms of higher order statistics, it would be necessary to generate synthetic flow sequences with more complex models. Until such time as it is demonstrated that higher order statistics substantially affect the design of a water-resource system, there is little incentive to generating synthetic flows to achieve statistical resemblance in terms of the higher order statistics.

In the following discussions, statistical resemblance is limited to three sets of statistics:

1 $\Theta_1 = \{\hat{\mu}, \hat{\sigma}, \hat{\gamma}\} \cup \{\hat{\rho}\}$
where $\hat{\mu} = \{\hat{\mu}(X|j) \ \forall \ j\}$
$\hat{\sigma} = \{\hat{\sigma}(X|j) \ \forall \ j\}$
$\hat{\gamma} = \{\hat{\gamma}(X|j) \ \forall \ j\}$
$\hat{\rho} = \{\hat{\rho}(X|k,j) \ \forall \ j \text{ and } k = 1; \ \hat{\rho}(X|k,u,v) \ \forall \ u,v \text{ and } k = 0,1\}$

2 $\Theta_2 = \{\hat{\mu}, \hat{\sigma}, \hat{\gamma}\} \cup \{\rho^*\}$
where $\rho^* = \{\hat{\rho}(X|k,j) \ \forall \ j \text{ and } k = 1; \ \hat{\rho}(X|k,u,v) \ \forall \ u,v \text{ and } k = 0\}$

3 $\Theta_3 = \Theta_2 \cup \{h\}$
where $\hat{h} = \{\hat{h}(X|j) \ \forall \ j\}$; $\hat{h}$ may be defined as either K or H

3-4 LAG-ONE MODEL

If the set of statistics Θ_1 is considered, then the lag-one model, defined as

$$X(t) - \mu = A[X(t-1) - \mu] + B\epsilon(t) \tag{3-17}$$

might be used to generate synthetic flow sequences, where $X(t)$, $X(t-1)$, and $\epsilon(t)$ are matrices of random variables, and μ, A, and B are matrices of coefficients. $X(t)$ and $X(t-1)$ are $(L \times 1)$ matrices whose jth elements $X(t|j)$ and $X(t-1|j)$ are random variables of flows for years t and $t-1$ at location j, where $j = 1, \ldots, L$. $\epsilon(t)$ is an $(L \times 1)$ matrix whose jth element $\epsilon(t|j)$ is a random variable.

To define the elements of the $(L \times 1)$ matrix μ and the $(L \times L)$ matrices A and B, the following assumptions are made:

1 The process is ergodic.
2 $\rho(\epsilon|k,j) = 0 \quad \forall\ k \neq 0, j$
3 $E[\epsilon(t|u)\epsilon(t|v)] = 0 \quad \forall\ t,\ u \neq v$
4 $E[\epsilon(t|u)X(T-1|v)] = 0 \quad \forall\ t, u \neq v$
5 $E[\epsilon(t|j)] = 0 \quad \forall\ t,j$
6 $E[\epsilon^2(t|j)] = 1 \quad \forall\ t,j$

The latter two assumptions, which lead to no loss of generality, are made simply for mathematical convenience.

For the matrix μ, the jth element $\mu(x|j)$ denotes the expected value of the flow for year t at location j, $E[X(t|j)]$. From the first assumption,

$$\mu(X|j) = E[X(t|j)] = E[X(t-1|j)] \quad \forall t \tag{3-18}$$

Matrices M_0 and M_1 are defined as

$$E[x(t)][x(t)]^T = M_0 \tag{3-19}$$

$$E[x(t)][x(t-1)]^T = M_1 \tag{3-20}$$

where $x(t) = X(t) - \mu$ and $x(t-1) = X(t-1) - \mu$. The superscript T denotes the operation of matrix transposition. M_0 and M_1 are $(L \times L)$ matrices whose (u,v)th elements are

$$m_0(X|u,v) = E[x(t|u) \times (t|v)] \quad \forall t \tag{3-21}$$

$$m_1(X|u,v) = E[x(t|u) \times (t-1|v)] \quad \forall t \tag{3-22}$$

Note for $u = v$, $m_0(X|u,u)$ denotes the variance of the flows at location u, and for $u \neq v$, $m_0(X|u,v)$ denotes the covariance of lag $k = 0$ between the flows at locations u and v. For $u = v$, $m_1(X|u,u)$ denotes the covariance between the flows for lag $K = 1$ at location u, and for $u \neq v$, $m_1(X|u,v)$ denotes the cross covariance between the flows at locations u and v for lag $K = 1$.

From the above assumptions,

$$E[\epsilon(t)][\epsilon(t)]^T = I \tag{3-23}$$

$$E[\epsilon(t)][X(t-1)]^T = E[X(t-1)][\epsilon(t)]^T = 0 \tag{3-24}$$

where I is an $(L \times L)$ identity matrix and 0 is an $(L \times L)$ null matrix.

If both sides of Eq. (3-17) are postmultiplied by $[X(t-1)-\mu]^T$, then the expectation of the various matrix products leads to

$$A = M_1 M_0^{-1} \tag{3-25}$$

where M_0^{-1} is the inverse of M_0. Similarly, via postmultiplication of both sides of Eq. (3-17) by $[X(t)-\mu]^T$ (Matalas[11]),

$$BB^T = M_0 - M_1\, M_0^{-1}\, M_1^T \tag{3-26}$$

The elements of B need not have any physical significance. If Δ denotes an $(L \times L)$ matrix, such that $\Delta\Delta^T = I$, then a matrix B^*, defined as $B^* = B\Delta$, where $B^*\, B^{*T} = BB^T$, may be used in place of B in Eq. (3-17). Techniques of principal component analysis (Kendall[12]) may be used to solve Eq. (3-26) for the elements of B. More conveniently, B may be transformed into a lower triangular matrix B^* whereby the elements of B^* may be derived recursively (Young[13]).

To illustrate a procedure for achieving statistical resemblance in terms of skewness, assume the matrix B to be lower triangular. From Eq. (3-17), the flow at time t of the jth location may be expressed as

$$x(t|j) = \sum_{i=1}^{L} a_{ji} \times (t-1|i) + \sum_{i=1}^{j} b_{ji}\epsilon(t|i) \tag{3-27}$$

Cubing both sides of Eq. (3-27) and then taking expectations leads to

$$\begin{aligned}\gamma(\epsilon|j) = [b_{jj}^3]^{-1} \Bigg[& \lambda(X|j) - \sum_{i=1}^{L} a_{ji}^3 \lambda(X|i) - 3 \sum_{i=1}^{L-1} \sum_{r=i+1}^{L} a_{ji}\, a_{jr}^2\, \lambda(X|i,r) \\ & - 3 \sum_{i=1}^{L-1} \sum_{r=i+1}^{L} a_{ji}^2\, a_{jr}\, \lambda'(X|i,r) \\ & - 6 \sum_{i=1}^{L-2} \sum_{r=i+1}^{L-1} \sum_{u=r+1}^{L} a_{ji}\, a_{jr}\, a_{ju}\, \lambda(X|i,r,u) \\ & - \sum_{i=0}^{j-1} b_{ji}^3 \gamma(\epsilon|i) \Bigg]\end{aligned} \tag{3-28}$$

where $\forall\ t$,

$$\lambda(X|j) = E[x^3(t|j)] \tag{3-29a}$$

$$\lambda(X|i) = E[x^3(t|i)] \tag{3-29b}$$

$$\lambda(X|i,r) = E[x^2(t|i) \times (t|r)] \tag{3-29c}$$

$$\lambda'(X|i,r) = E[x^2(t|i) \times (t|r)] \tag{3-29d}$$

$$\lambda(X|i,r,u) = E[x(t|i) \times (t|r) \times (t|u)] \tag{3-29e}$$

$$\gamma(\epsilon|j) = E[\epsilon^3(t|j)] \tag{3-29f}$$

$$\gamma(\epsilon|i) = E[\epsilon^3(t|i)] \tag{3-29g}$$

$$b_{j0} = 0 \qquad \forall\ j \tag{3-29h}$$

Equation (3-28) may be solved recursively to obtain the set of skew coefficients $\gamma(\epsilon|1), \ldots, \gamma(\epsilon|L)$.

As long as the distribution function of $X(t|j)$ is not of concern, $\epsilon(t|j)$ may be assumed to follow any distribution, such that $E[\epsilon(t|j)] = 0$, $E[\epsilon^2(t|j)] = 1$, and $E[\epsilon^3(t|j)] = \gamma(\epsilon|j)$. The use of one distribution, namely, the log-normal distribution, is illustrated. Let

$$\epsilon(t|j) = \frac{Y(t|j) - \mu(Y|j)}{\sigma(Y|j)} \tag{3-30}$$

where $Y(t|j)$ is log-normally distributed with $\mu(Y|j) = E[Y(t|j)]$, $\sigma(Y|j) = \{E[Y(t|j) - \mu(Y|j)]^2\}^{1/2}$, and $\gamma(Y|j) = E[Y(t|j) - \mu(Y|j)]^3/\sigma^3(Y|j) = \gamma(\epsilon|j)$. The random variable $Y(t|j)$ may be expressed as

$$Y(t|j) = \exp[Z(t|j)] \tag{3-31}$$

where $Z(t|j)$ is normally distributed with $\mu(Z|j) = E[Z(t|j)]$ and $\sigma^2(Z|j) = E[Z(t|j) - \mu(Z|j)]^2$. It is convenient to let $\mu(Z|j) = 0$. Then (Aitchison and Brown[14]),

$$\mu(Y|j) = \frac{\exp[\sigma^2 z|j]}{2} \tag{3-32}$$

$$\sigma^2(Y|j) = \exp[2\sigma^2(z|j)] = \exp[\sigma^2(z|j)] \tag{3-33}$$

$$\gamma(Y|j) = \frac{\exp[3\sigma^2(z|j)] - 3\exp[\sigma^2(z|j) + 2}{\{\exp[\sigma^2(z|j)] - 1\}^{3/2}} \tag{3-34}$$

$\gamma(Y|j) = \gamma(\epsilon|j)$, so from Eq. (3-34), $\sigma^2(z|j)$ is determined whereby Eqs. (3-32) and (3-33) yield the values of $\mu(Y|j)$ and $\sigma^2(Y|j)$. Thus, variate values of $\epsilon(t|j)$ may be obtained via the exponental transform of variate values drawn from a normal population having mean zero and variance $\sigma^2(z|j)$.

It should be noted that, to achieve statistical resemblance in terms of the elements of Θ_1, it is necessary to consider third-order terms besides those for defining the skewness of the flows.

3-5 MARKOV MODEL

If Θ_2 instead of Θ_1 is of interest, it may be possible to avoid needless computation by expressing the lag-one model as

$$X(t) - \mu = \tilde{A}[X(t-1) - \mu] + \tilde{B}\epsilon(t) \tag{3-35}$$

where $\tilde{A}$ denotes an $(L \times L)$ diagonal matrix whose (j,j)th element is $\rho(X|k=1,j) \equiv \rho(X|k,j)$, and $\tilde{B}$ is an $(L \times L)$ matrix whose elements are given by the solution of

$$\tilde{B}\tilde{B}^\tau = M_0 - \tilde{A}M_0\tilde{A}^\tau \tag{3-36}$$

The matrix $\tilde{B}$ may be defined as being lower triangular. The flow at time t at the jth location is expressed as

$$x(t|j) = \rho(X|1,j) \times (t-1|j) + \sum_{i=1}^{j} b_{ji}\, \epsilon(t|i) \tag{3-37}$$

which represents a Markov process.

Because for each location the flows are represented by a Markov process, Eq. (3-35) is referred to as a *Markov model*. For the jth element of Eq. (3-35),

$$\rho(X|k,j) = \rho^{|k|}\ (X|i,j) \tag{3-38}$$

which is a basic property of a Markov process. While Eqs. (3-17) and (3-35) are mathematically similar, the Markov property, in general, will not hold for the jth element of Eq. (3-17). If, indeed, the underlying streamflow generating process is markovian, $A = \tilde{A}$ and $B = \tilde{B}$.

The lag-one model could be used to achieve statistical resemblance in terms of the elements of Θ_2. However, it might be possible to reduce the computational load by using the Markov model. In determining the elements of $\tilde{A}$ and $\tilde{B}$, the matrix M_1 need not be considered. Moreover, it is not necessary to determine M_0^{-1}. The skewness of $\epsilon(t|j)$ is defined as

$$\gamma(\epsilon|j) = [b_{jj}^3]^{-1}\left\{ [1-\rho^3(X|1,j)]\lambda(X|j) - \sum_{i=0}^{j-1} b_{ji}^3 \gamma(\epsilon|i) \right\} \tag{3-39}$$

where $b_{j0} = 0\ \forall\ j$. Equation (3-39) may be solved recursively to obtain the set of skew coefficient $\gamma(\epsilon|1), \ldots, \gamma(\epsilon|L)$ whereby the variate values of $\epsilon(t|j)$ may be obtained in the manner described for the lag-one model. It is noted that far fewer computations are required to determine $\gamma(\epsilon|j)$ for the Markov model than for the lag-one model. In particular, with the Markov model, no third-order terms other than those for defining the coefficients of skewness of the flows need be determined.

3-6 FRACTIONAL NOISE MODEL

Lag-one and markovian processes are characterized by values of $h = \frac{1}{2}$, as are all processes which belong to the Brownian domain of attraction. Thus, if the set of statistics Θ_3 is of interest, then a stochastic process not belonging to this domain of attraction must be used to approximate the underlying streamflow generating mechanism. Mandelbrot and Wallis[4] have suggested the use of fractional noise processes which are characterized by infinite memories and values of $h \neq \frac{1}{2}$. For operational purposes, continuous-parameter fractional noise processes must be approximated as discrete parameter processes with large but finite memories.

One approximation suggested by Mandelbrot and Wallis, referred to as

type 2, is defined as

$$Y(t|j) = [h(j) - \tfrac{1}{2}] \sum_{\delta=t-M(j)}^{t-1} (t-\delta)^{h(j)-3/2}\, \eta(\delta|j) \tag{3-40}$$

where $Y(t|j)$ and $\eta(\delta|j)$ are random variables associated with the time points t and δ at location j, $h(j)$ is the Hurst coefficient, and $M(j)$ denotes the memory length of the process. Under the assumption that $\eta(\delta|j)$ and $\eta(\tau|j)$ are linearly independent $\forall \delta \neq \tau$, then

$$\mu(Y|j) = [h(j) - \tfrac{1}{2}] \sum_{i=0}^{M(j)-1} [M(j) - i]^{B(j)}\, \mu(\epsilon|j) \tag{3-41}$$

$$\sigma^2(Y|j) = [h(j) - \tfrac{1}{2}]^2 \sum_{i=0}^{M(j)-1} [M(j) - i]^{2B(j)}\, \sigma^2(\epsilon|j) \tag{3-42}$$

$$\lambda(Y|j) = [h(j) - \tfrac{1}{2}]^3 \sum_{i=0}^{M(j)-1} [M(j) - i]^{3B(j)}\, \lambda(\epsilon|j) \tag{3-43}$$

$$\rho(Y|k,j) = \frac{\sum_{i=0}^{M(j)-1-k} [M(j) - i]^{B(j)}\, [M(j) - i - k]^{B(j)}}{\sum_{i=0}^{M(j)-1} [M(j) - i]^{2B(j)}} \tag{3-44}$$

where $B(j) = h(j) - \frac{3}{2}$.

To reduce the dominance of the low-frequency components which characterize the type-2 process, Matalas and Wallis[6] considered a filtered type-2 process which is defined as

$$X(t|j) = [h(j) - \tfrac{1}{2}] \sum_{\delta=pt-M(j)}^{pt-1} (pt-\delta)^{h(j)-3/2}\, \eta(\delta|j) \tag{3-45}$$

where $X(t|j) = Y(pt|j)$ and $p \geq 1$ is an integer. It can be shown that $\mu(X|j) = \mu(Y|j)$, $\sigma^2(X|j) = \sigma^2(Y|j)$, and $\lambda(X|j) = \lambda(Y|j)\ \forall\ p$. However, $\rho(X|k,j) = \rho(Y|pk,j)$. Matalas and Wallis[6] also point out that for two filtered type-2 processes, where $p(u) = p(v)$, $\rho(X|0,u,v) = \rho(Y|0,u,v)$. If, however, $p(u) \neq p(v)$, then $\rho(X|0,u,v) \neq \rho(Y|0,u,v)$, and moreover, $\rho(X|k,u,v)$ would depend upon $t\ \forall\ k$, $u \neq v$, in which case, the two filtered type-2 processes could not be stationary with respect to their lagged cross correlations.

To generate synthetic flows at each of L locations, such that statistical resemblance is achieved with respect to the elements of Θ_3, the flow at the jth location is expressed as

$$x(t|j) = [h(j) - \tfrac{1}{2}] \sum_{\delta=pt-M(j)}^{pt-1} (pt-\delta)^{h(j)-3/2} \sum_{r=1}^{j} b_{ju}\, \epsilon(\delta|r) \tag{3-46}$$

where $x(t|j) = X(t|j) - \mu(X|j)$. The random variable $\epsilon(\delta|u)$ is defined such that the assumption made for the lag-one process holds for the process defined by Eq. (3-46). By considering $E[x(t|u)x(t|v)]\ \forall u,v$, $L(L+1)/2$ equations with $L(L+1)/2$ unknowns, the values of the b's, may be formed. These equations may be solved recursively to obtain the values of the b's in terms of the second-order statistics characterizing the flows at the L locations. It should be noted that $E[x(t|u)x(t|v)]$, where $u \neq v$, involves the summation of $[M(u)-i]^{h(u)-3/2}[M(v)-1]^{h(v)-3/2}$ over $i=0$ to $[M(u)-1,\ M(v)-1]$, where

$$[M(u)-1,\ M(v)-1] = \begin{cases} M(u)-1 & \text{if } M(u) < M(v) \\ M(v)-1 & \text{if } M(v) \leq M(u) \end{cases} \tag{3-47}$$

Cubing both sides of Eq. (3-46) and taking expectations leads to

$$\gamma(\epsilon|j) = \lambda(X|j)\ [b_{jj}]^{-3}\ [h(j)-\tfrac{1}{2}]^{-3} \left\{ \sum_{i=0}^{M(j)-1} [M(j)-i]^{3B(j)} \right\}^{-1} - \sum_{u=0}^{u-1} b_{ju}^{3}\ \gamma(\epsilon|u) \tag{3-48}$$

where $b_{j0} = 0\ \forall j$. Equation (3-48) may be solved recursively to obtain the set of skew coefficients $\gamma(\epsilon|1), \ldots, \gamma(\epsilon|L)$. Given this set of skew coefficients, the variate values for $\epsilon(t|j)$ may be obtained in the manner described for the lag-one process.

Values for $\mu(X|j)$, $\sigma(X|j)$, $\lambda(X|j)$, $\rho(X|k=1,j)$, $\rho(X|k=0,u,v)$, and $h(j) \equiv h(X|j)$ may be estimated from the sequence of historical flows at the jth location. The value for $M(j) \equiv M(x|j)$ may be obtained in the following manner. $\rho(X|k-1,j)$ is substituted for $\rho(Y|k=p,j)$ in Eq. (3-44) whereby

$$\rho(X|k=1,j) = \frac{\sum_{i=0}^{M(j)-1-p} [M(j)-i]^{B(j)}\ [M(j)-i-p]^{B(j)}}{\sum_{i=0}^{M(j)-1} [M(j)-i]^{2B(j)}} \tag{3-49}$$

For $p = p^*\ \forall j$, Eq. (3-49) is solved for $M(j)$. To facilitate this solution, Matalas and Wallis[6] have provided a graphical representation of $\rho(X|k=1,j)$ versus $M(j)$ for various values of p and $h(j)$.

3-7 DISTRIBUTION OF FLOWS

In the above discussions, statistical resemblance was considered with respect to a set of low-order statistics used to characterize streamflow sequences. The probability distribution of the flows, $F[X|t,j]$, was not considered. Under the assumption of ergodicity, $F[X|t,j] = F[X|j]\ \forall t$. While the low-order statistics $\mu(X|j)$, $\sigma(X|j)$, and $\gamma(X|j)$ do characterize a flow sequence, they do not uniquely define $F[X|j]$. As long as this function is not of interest, then the

choice of the distribution of $\epsilon(t|u)$, $F[\epsilon|t,u] = F[\epsilon|u]\ \forall\ t$ is an operational matter. Any one of several distributions of $\epsilon(t|u)$ can be used to achieve statistical resemblance in terms of the low-order statistics. The choice is largely a matter of convenience and ease with which the variate values of $\epsilon(t|u)$ can be generated.

If in addition to the low-order statistics, statistical resemblance is extended to include a specified distribution of $X(t|j)$, then formidable problems may be encountered in generating flow sequences. The problems lie in specifying $F[\epsilon|u]$ given $F[X|j]$. To illustrate the nature of these problems, only the Markov model for the case of a flow sequence at a single site is considered. For convenience, the location index j is omitted. The flow at time t is expressed as

$$x(t) = \rho x(t-1) + b\epsilon(t) \tag{3-50}$$

where $x(t) = X(t) - \mu(X)$, $x(t-1) = X(t-1) - \mu$, $\rho \equiv \rho(X|k=1)$, and $b = [1-\rho^2]^{1/2}\sigma(x)$.

It was noted above that, under the assumption of ergodicity, $F[X|t] = F[X|t-1]\ \forall\ t$. The sum of two normal variables is normal. Thus if $F[X|t-1]$ is assumed to be normal, then $F[\epsilon|t]$ must be assumed to be normal.

Quite often streamflow is assumed to follow either a gamma or a log-normal distribution. The sum of two log-normal variables is not log-normal, and in general the sum of two gamma variables is not gamma. If $F[X]$ is assumed to be log-normal, then the Markov model cannot be used to generate synthetic flows. To accommodate the assumption of log-normally, the following procedure may be used. Let a be the lower bound of a random variable X, where $(X-a)$ is log-normally distributed. $Y = \log\ (X-a)$, where log denotes the logarithm of $(X-a)$ to the base e, is normally distributed. The relations between the low-order statistics of X and Y are given by Eqs. (3-32), (3-33), and (3-34).

If Y is assumed to be generated by a Markov process, then

$$Y(t) = ry(t-1) + (1-r^2)^{1/2}\,\sigma(Y)\epsilon(t) \tag{3-51}$$

where $y(t) = Y(t) - \mu(Y)$, $y(t-1) - \mu(Y)$, and

$$r \equiv \rho(Y|k=1) = \frac{1}{\sigma(Y)} \log\ \{\rho[\exp(\sigma(Y)) - 1] + 1\} \tag{3-52}$$

In terms of X, the generating process is

$$X(t) = a + \{\exp[(1-\rho)\mu(Y)]\}\ [X(t-1) - a]^r\ \delta(t) \tag{3-53}$$

where

$$\delta(t) = \exp\{(1-r^2)^{1/2}\,\sigma(Y)\epsilon(t)\} \tag{3-54}$$

To generate log-normally-distributed synthetic flows with statistical resemblance in terms of $\mu(X)$, $\sigma(X)$, $\gamma(X)$, and $\rho(X|k=1)$, it is first necessary to solve Eqs. (3-32), (3-33), (3-34), and (3-52) for the values of a, $\mu(Y)$, $\sigma(Y)$, and $\rho(Y|k=1)$ given the historical estimates of $\mu(X)$, $\sigma(X)$, $\gamma(X)$, and

$\rho(X|k=1)$. By means of Eq. (3-51), normally distributed synthetic flows in log-space may be generated. Synthetic flows in real space are obtained via adding the constant a to the antilog of each value of $Y(t)$ (Matalas[11]).

From Eq. (3-50), the skewness and kurtosis of $\epsilon(t)$ in terms of the skewness and kurtosis of $X(t)$ are

$$\beta_1(\epsilon) = \gamma^2(\epsilon) = \frac{(1-\rho^3)^2}{(1-\rho^2)^3}\beta_1(X) \tag{3-55}$$

$$\beta_2(\epsilon) = \frac{1+\rho^2}{1-\rho^2}\beta_2(X) - \frac{6\rho^2}{1-\rho^2} \tag{3-56}$$

If $\beta_2(X) = 3$, which is the case for X normally distributed, then $\beta_2(\epsilon) = 3\ \forall\ \rho$. For $\rho^2 > 0$, $\beta_2(\epsilon) > \beta_2(X)$ if $\beta_2(X) > 3$, and $\beta_2(\epsilon) < \beta_2(X)$ if $\beta_2(X) < 3$. For $\rho = 0$, $\beta_2(\epsilon) = \beta_2(X)$. If

$$\beta_2(X) \leq 6\frac{\rho_2}{1+\rho^2} \tag{3-57}$$

then $\beta_2(\epsilon) \leq 0$, which is not admissible. If inequality (3-57) holds, then the Markov process cannot be used to generate synthetic flows (Matalas[15]). In particular, if ϵ is distributed as gamma, then

$$2\beta_2(\epsilon) - 3\beta_1(\epsilon) - 6 = 0 \tag{3-58}$$

$$2\beta_2(X) - 3\beta_1(X) - 6 = \Omega = \frac{4}{1-\rho^2}[\beta_2(X) - 3] - \frac{3\beta_1(x)}{(1-\rho^2)^2}[2(1-\rho^3) - 3\rho^2(1-\rho^2)] \tag{3-59}$$

For X to be distributed as gamma, Ω must equal 0. For $\rho^2 > 0$, $\Omega \neq 0$ so that the Markov process cannot be used to generate synthetic flows that are distributed as gamma if ϵ is assumed to be distributed as gamma.

The above discussions indicate that problems are likely to be encountered in generating correlated synthetic flows that follow a specified distribution. If X is to follow a specified distribution, then it cannot be assumed that ϵ necessarily follows the same distribution. A particular model used to achieve statistical resemblance in terms of certain low-order statistics may not be operational if resemblance is to include a specific distribution for the flows.

3-8 MODEL INCONSISTENCY

Given the mathematical relations between an assumed generating model's parameters and the elements of a set of statistics Θ does not necessarily imply that the model is operational. That is, it may not be possible to solve the equations expressing these relations because the model is inconsistent with the his-

torical flows. The inconsistency relates to the failure of the historical flows to meet structural constraints of the model (Matalas and Wallis[16]).

To illustrate the nature of model inconsistency, the Markov model is considered. For this model to be operational, $\tilde{B}\tilde{B}^T$ must be positive definite, that is, the determinant $|\tilde{B}\tilde{B}^T|$ must be nonsingular. To illustrate that this may not be so, the flows at two locations are considered, where without loss of generality, it is assumed that $\sigma(X|1) = \sigma(X|2) = 1$. To simplify notation, let $\rho \equiv \rho(X|k = 1, 1)$, $r \equiv \rho(X|k = 1, 2)$, and $R \equiv \rho(X|k = 0, 1, 2)$. Thus,

$$\tilde{A} = \begin{bmatrix} \rho & 0 \\ 0 & r \end{bmatrix} \tag{3-60}$$

$$M_0 = \begin{bmatrix} 1 & R \\ R & 1 \end{bmatrix} \tag{3-61}$$

From Eq. (3-36),

$$\tilde{B}\tilde{B}^T = \begin{bmatrix} 1-\rho^2 & R(1-r\rho) \\ R(1-r\rho) & 1-r^2 \end{bmatrix} \tag{3-62}$$

$|\tilde{B}\tilde{B}^T| > 0$, if

$$R^2 \leq \frac{(1-\rho^2)(1-r^2)}{(1-r\rho)^2} \tag{3-63}$$

Inequality (3-63) states that there is an upper bound on the values that R^2 may assume if $\rho = r$, then $R^2 \leq 1$. If, however, $\rho \neq r$, then $R^2 < 1$. The upper bound becomes smaller as the difference between ρ and r becomes larger.

It is possible for the historical flows to yield values of ρ, r, and R which will not satisfy inequality (3-63), in which case, the Markov model cannot be used to generate synthetic flow sequences. It is interesting to note that, if the underlying streamflow process is markovian, the historical flows may yield values of ρ, r, and R which do not satisfy inequality (3-63) (Slack[17]). Thus knowledge of the underlying generating process does not assure the use of that process for generating synthetic flow sequences. Slack[17] points out that if the correlations are very near the upper bound, then the probability that synthetic flows cannot be generated with the underlying generating process increases with the length of the historical flow sequences.

In the case of the fractional noise model,

$$R^2 \leq \frac{\left\{\sum_{i=0}^{[M(1)-1,\,M(2)-1]} [M(1)-i]^{h(1)-3/2}\,[M(2)-i]^{h(2)-3/2}\right\}^2}{\sum_{i=0}^{M(1)-1} [M(1)-i]^{2h(1)-3} \sum_{i=0}^{M(2)-1} [M(2)-i]^{2h(2)-3}} \tag{3-64}$$

where $[M(1)-1, M(2)-1]$ is defined by Eq. (3-47). If $M(1) = M(2)$ and $h(1) = h(2)$, then $R^2 \leq 1$. It is noted that inequality (3-64) does not depend upon the filter parameter p, and if either $M(1) \neq M(2)$ or $h(1) \neq h(2)$, then R^2

cannot assume values equal to unity. For the fractional noise model to be operational, the historical flows must yield a value of R^2 that does not exceed the upper bound given by inequality (3-64).

Model inconsistency may arise in one of two ways. First, in the case where the underlying and assumed generating processes are the same, inconsistency may be attributed to statistical sampling errors. That is, the estimated values of the pertinent statistics may differ sufficiently from their population values to yield inconsistency. Second, the underlying and assumed processes may be different. The historical flows have a multivariate structure which may be inconsistent with that implied by the assumed generating process.

Of course, computational errors or machine round-off errors can lead to model inconsistency. While checks for such errors should be made, the nature of the structural constraints imposed by an assumed generating process should also be investigated if model inconsistency arises.

3-9 STATISTICAL BIAS

In the above discussions, the assumption was made that the estimates of the pertinent statistics forming the set Θ are statistically unbiased. An estimate $\hat{\theta}_i$ of θ_i is said to be statistically unbiased if $E[\hat{\theta}_i] = \theta_i$. With the exception of $\hat{\mu}(X)$, the various estimates given above are not statistically unbiased. Unfortunately, for a particular statistic, the bias does not depend only upon the length of the historical flow sequence from which the estimate of the statistic is derived. To date, few studies have been made to describe the nature of statistical bias in relation to the assumed streamflow generating process.

In general, the bias associated with an estimate of a particular statistic depends upon the underlying streamflow generating process. Because the underlying generating process is unknown, unbiased estimates of the elements of it cannot be derived from the historical flow sequences. Let $\tilde{\theta}_i$ denote the estimate of θ_i derived from a finite synthetic flow sequence. In effect, $\tilde{\theta}_i$ is a biased estimate of $\hat{\theta}_i$, which in turn is a biased estimate of θ_i. While it may not be possible to derive an unbiased estimate of $\hat{\theta}_i$, an unbiased estimate of $\tilde{\theta}_i$ can be derived. Given the assumed generating model, $\hat{\theta}_i$ can be adjusted accordingly to yield an unbiased estimate, $\tilde{\theta}_i$. The following discussions deal with bias corrections for $\tilde{\tilde{\sigma}}(X)$, $\tilde{\rho}(X|k=1)$, and $\tilde{h}(X) \equiv K(X)$, where the assumed generating processes are Markov and filtered fractional noise models (Wallis and Matalas[18]; Wallis and O'Connell[19]). Only results for synthetic flow sequences of length $\tilde{n} = 100$ are given.

The estimate $\tilde{\sigma}(X)$ has expectation $E[\tilde{\sigma}(X)] = \alpha[\hat{\sigma}(X)]$, where $\alpha(1)$ depends upon the assumed generating process. If the process is markovian, $\alpha = f[\tilde{n}, \hat{\rho}(X|k=1)]$, and if the process is filtered fractional noise, $\alpha = f[\tilde{n}, \hat{h}, (X), \hat{M}(X), p]$. Via Monte Carlo experiments, Wallis and Matalas[18] derived values of α for $\tilde{n} = 100$ for markovian and filtered fractional

noise models. The results of these experiments are given in Tables 3-1 and 3-2 for the Markov process and for the filtered fractional noise process where $\hat{M}(X)/p = 1{,}000$.

Table 3-1 α—MARKOV PROCESS: $\hat{n} = 100$

$\hat{\rho}(X\|k=1)$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
α	1	0.99	0.99	0.99	0.98	0.98	0.97	0.95	0.94	0.90

Table 3-2 α—FILTERED FRACTIONAL NOISE PROCESS: $\hat{n} = 100$; $\hat{M}(X)/p = 1{,}000$

p \ $\hat{h}(X)$	0.6	0.7	0.8	0.9
1	0.90	0.88	0.80	0.70
5	0.96	0.94	0.90	0.80
10	0.98	0.96	0.92	0.84
20	0.98	0.96	0.92	0.84

To correct for bias, the estimate of the standard deviation of the historical flows is defined as

$$\sigma^*(X) = \frac{1}{\alpha(1)}\,\hat{\sigma}(X) \tag{3-65}$$

whereby

$$E[\tilde{\sigma}(X)] = \alpha(1)\sigma^*(X) = \hat{\sigma}(X) \tag{3-66}$$

Given $\hat{\rho}(X|k=1)$, values of $E[\tilde{\rho}(X|k=1)]$ for $\tilde{n} = 100$ were obtained by Monte Carlo experiments (Wallis and Matalas[18]) and are given in Tables 3-3 and 3-4, where $\hat{\rho} \equiv \hat{\rho}(X|k=1)$ and $E[\tilde{\rho}] \equiv E[\tilde{\rho}(X|k=1)]$.

Table 3-3 $\hat{\rho}$ VERSUS $E[\hat{\rho}]$—MARKOV PROCESS: $\tilde{n} = 100$

$\hat{\rho}$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$E[\tilde{\rho}]$	0.09	0.18	0.28	0.38	0.48	0.57	0.66	0.76	0.86

Table 3-4 $\hat{\rho}$ VERSUS $E[\hat{\rho}]$—FILTERED FRACTIONAL NOISE PROCESS: $\tilde{n} = 100$; $\hat{M}(X)/p = 1{,}000$

p \ $h(X)$	0.6		0.7		0.8		0.9	
	$\hat{\rho}$	$E[\tilde{\rho}]$	$\hat{\rho}$	$E[\tilde{\rho}]$	$\hat{\rho}$	$E[\tilde{\rho}]$	$\hat{\rho}$	$E[\tilde{\rho}]$
1	0.66	0.58	0.73	0.62	0.80	0.66	0.87	0.71
5	0.35	0.29	0.44	0.36	0.57	0.45	0.68	0.52
10	0.24	0.19	0.33	0.26	0.45	0.34	0.59	0.45
20	0.17	0.13	0.25	0.19	0.36	0.27	0.50	0.37

To generate synthetic flows such that $E[\tilde{\rho}] = \hat{\rho}$, it is necessary to replace the historical value $\hat{\rho}$ by a value $\rho^* > \hat{\rho}$. Given the value ρ^*, the bias in the standard deviation of the synthetic flows is corrected by using $\alpha^* = f[\tilde{n}, \rho^*]$. The following example illustrates the procedure for correcting for bias in both the standard deviation and the lag-one serial correlation coefficient. Assume a Markov process with $\hat{\sigma} = 10$ and $\hat{\rho} = 0.8$. The value of $\hat{\rho}$ such that $E[\tilde{\rho}] = 0.8$ is taken to be ρ^*. From Table 3-3, it is found by linear interpolation that $\rho^* = 0.84$, and from Table 3-1 that $\alpha^* = 0.92$ for $\hat{\rho} \equiv \rho^* = 0.84$. Thus for generating synthetic flows, $\hat{\sigma} = 10$ and $\hat{\rho} = 0.8$ are replaced by $\sigma^* = 10/0.92 \simeq 10.87$ and $\rho^* = 0.84$.

Wallis and O'Connell[19] have investigated various algorithms for estimating $\rho(X|k=1)$ from small samples. Based on an algorithm suggested by Jenkins and Watts[20] and Box and Jenkins,[21] Wallis and O'Connell suggested that $\hat{\rho}(X|k=1)$ be replaced by

$$\rho^*(X|k=1) = \frac{\rho(X|k=1) + (1/n)}{1 - (4/n)} \tag{3-67}$$

where

$$\tilde{\rho}(X|k=1) = \frac{\sum_{t=1}^{n-1} [X(t) - \mu(X)][X(t+1) - \mu(X)]}{n\sigma^2(X)} \tag{3-68}$$

In doing so,

$$E[\tilde{\rho}(X|k=1)] \simeq \rho^*(X|k=1) \tag{3-69}$$

Via Monte Carlo experiments, Wallis and Matalas[5] derived approximate values of $E[K]$, where K is an estimate of the Hurst coefficient h, defined by Eq. (3-16), for both markovian and fractional noise processes. For a Markov process, $h = \frac{1}{2}$, whereas for a fractional noise process h may assume values other than $\frac{1}{2}$. Values of $E[\tilde{K}]$ for $\tilde{n} = 100$ are given in Tables 3-5 and 3-6.

Table 3-5 $E[K]$ **VERSUS** $\hat{\rho}(X|k=1)$**—MARKOV PROCESS:** $\tilde{n} = 100$

$\hat{\rho}(X\|k=1)$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$E[\tilde{K}]$	0.61	0.65	0.66	0.69	0.71	0.73	0.75	0.79	0.82	0.87

Table 3-6 $E[\tilde{K}]$ **VERSUS** $\hat{h}(x)$, p**—FILTERED FRACTIONAL NOISE PROCESS:** $n = 100$, $\hat{M}(x)/p = 1{,}000$

p \ $\hat{h}(X)$	0.6	0.7	0.8	0.9
1	0.81	0.82	0.83	0.85
5	0.74	0.76	0.78	0.81
10	0.71	0.73	0.75	0.79
20	0.68	0.71	0.74	0.78

For the Markov process, the difference between $h = \frac{1}{2}$ and $E[K]$ denotes the bias in estimating the value $h = \frac{1}{2}$. Unlike the estimates of $\sigma(X)$ and $\rho(X|k=1)$, a correction for this bias cannot be made to achieve $E[\tilde{K}] = \frac{1}{2}$. If

for some reason one wishes to generate synthetic sequences such that $E[\tilde{K}] = K$, then $\hat{\rho}(X|k=1)$ must be adjusted. This adjustment involves replacing $\hat{\rho}(X|k=1)$ with a value $\rho^*(X|k=1)$ for which $E[\tilde{K}] = K$. However, it should be noted that, in doing so, the synthetic sequences will yield correlograms quite distinct from the historical one, particularly if $\rho^*(X|k=1)$ is large (Wallis and Matalas[5]).

For filtered fractional noise processes, bias, the difference between $\hat{h}(x) = K$ and $E[\hat{K}]$, can be corrected. This correction is made by selecting a value $h^*(X)$, for which $E[\tilde{K}] = \hat{h}(x) = K$. Note that this correction may also be made by changing the value of p.

In applying the various corrections for bias, it should be noted that correcting one statistic may introduce a constraint in correcting another statistic. Moreover, the corrected statistics may lead to model inconsistency of the type discussed above. While it is possible to derive bias corrections for other statistics, either analytically or by Monte Carlo procedures, there is a need to assess the impact of these corrections on the operational capability of the synthetic flow generating model.

3-10 INCOMPLETE DATA SETS

Throughout these discussions, it has been assumed that, at each location, the historical flow sequences are concurrent and of equal length. In practice, this assumption is unlikely to hold; sequences may be of unequal length, they may span different periods of time, or they may be discontinuous over time. Moreover, for some locations of interest, there may exist no historical flow sequences.

If the historical flow sequences are not concurrent and of equal length, the assumed generating model may frequently prove to be inconsistent. With incomplete data sets, the lag-zero variance-covariance matrix M_0 may be inconsistent. That is, M_0 may not be positive semidefinite. For the lag-one model, Eq. (3-17), the solution for the elements of the matrix A requires the inverse of M_0. If M_0 is inconsistent, its eigenvalues will not all be positive, in which case some of the elements of A will be complex numbers. In such a case, the generated flows would themselves be complex numbers. Beard[22] and Fiering[23] have developed techniques for transforming an inconsistent matrix M_0 into a consistent one.

Crosby and Maddock[24] have pointed out that, for the lag-one model, even if M_0 and M_1 are both consistent, the matrix BB^T, defined by Eq. (3-26), may not be consistent. For a monotone sample, Crosby and Maddock have developed a technique for obtaining a matrix BB^T that is consistent. A monotone sample is defined as a set of sequences, each of which may have originated at different points in time and which are all continuous up to the present. While the technique can be used with the Markov model, it remains to be shown that it can be applied with other flow generators. Corresponding techniques need to

be developed for the case where the historical flow sequences are not continuous.

In attempting to generate synthetic flow sequences at ungaged locations, various problems are likely to be encountered. At each of the ungaged locations, estimates of some of the statistics forming the set Θ may be obtained via regression analysis. For a particular statistic, its estimates at the gaged locations may be regressed on the physiographic and meteorologic factors that characterize the drainage basins associated with each of the locations (Benson and Matalas[25]). Via the regression relation and the physiographic and meteorologic factors characterizing the drainage basins of the ungaged locations, estimates for the particular statistic may be obtained for each of the ungaged locations.

A regression relation for any statistic can be obtained. The question is whether or not the relation should be used to make estimates of the statistic at the ungaged locations. Generally, the answer is provided via classical tests of significance. If, at a preselected level of confidence, the regression relation is significant, then the relation would be used; otherwise, it would not. Perhaps adherence to conventional levels of confidence are unwarranted. Less-binding levels may increase the uncertainty in the estimates derived from "significant" regression relations, but the planning decisions may be insensitive to the increased uncertainty.

Apart from the mean and standard deviation, it may be very difficult to obtain an acceptable regression relation relative to the elements of Θ. If an acceptable relation cannot be developed for a particular statistic, then one may interpret the variations in the estimates of the statistics at the gaged locations as being due entirely to chance whereby the average over the gaged locations may be taken as the estimate at each of the ungaged locations, and perhaps, at each of the gaged locations as well.

However the estimates for the ungaged locations are derived, they do not assure the ability to generate synthetic flow sequences at the ungaged locations. Moreover, if the gaged and ungaged locations are combined, it may not be possible to generate synthetic flow sequences at any of the locations. For example, with the lag-one model, some of the elements of M_0 and M_1 will be the estimates for the ungaged locations. These matrices may not be consistent, nor perhaps, the matrix BB^T. If the matrices prove to be inconsistent, then the estimates for the ungaged locations will need to be modified or the matrices adjusted to achieve consistency.

3-11 ALTERNATIVE MODELS

The lag-one, Markov, and fractional noise models are by no means exhaustive. Three alternative models, which have been suggested in the current literature, are discussed briefly. One of these models is the lag-p model. There are

perhaps two basic reasons for choosing such a model. First, the lag-p covariances are believed to have a significant effect upon the design of a water-resource system, and second, the lag-p covariances may allow synthetic flow sequences to be generated such that statistical resemblance in terms of $h \neq \frac{1}{2}$ is achieved.

At present, there is little evidence to suggest that the covariances for lag $p > 1$ materially affect a system's design. Moreover, these covariances are subject to large sampling errors, and one may question the merit in achieving statistical resemblance in terms of these covariances. Fiering[2] has shown that statistical resemblance in terms of $h \neq \frac{1}{2}$ may be achieved by making p sufficiently large. While the lag-p model provides a mathematically convenient vehicle for achieving this level of statistical resemblance, it does so at considerable computational cost, particularly for the multisite, multiseason case. Because for this model $h = \frac{1}{2}$, one must, for each particular streamflow sequence, find a value for p such that the sample estimates of p derived from synthetic sequences will yield on the average a value for h that is equal to the historical estimate. For this model, h tends to the value $\frac{1}{2}$ as n tends to infinity, where the rate of convergence is inversely related to p. Essentially, increasing p increases the memory of the process, and thus the rate of convergence of h to the value $\frac{1}{2}$ is decreased.

Two processes which seemingly offer greater operational capability than the lag-p model in achieving statistical resemblance in terms of h are the broken-line process introduced by Ditlevsen[26] and being adapted to synthetic flow generation by Mejia[27] and by Garcia, Dawdy, and Mejia,[28] and the ARIMA process (Box and Jenkins[21]). This latter process has been used to describe streamflow sequences by Carlson, MacCormick, and Watts[29] and is being adapted to synthetic flow generation by O'Connell.[30]

A broken-line process consists of a summation of a finite number of simple broken-line processes, which are defined as follows. A simple broken-line process is a sequence of intersecting line segments, where the projections of the line segments on the time axis are of equal length, and where the magnitude of the intersections are randomly distributed. The projection length is allowed to vary in a prescribed manner from one simple broken-line process to another. By selecting appropriate values for the broken-line parameters, which include a basic projection length and the number of simple broken lines to be summed, statistical resemblance in terms of h, as well as other elements of Θ, can be achieved.

Mandelbrot[31] has pointed out that while the broken-line process exhibits a peculiar kind of nonstationarity, the process is a useful approximation to a fractional noise process. This approximation can be made better by increasing the number of simple broken-line processes.

The term ARIMA is used to denote an *a*uto*r*egressive-*i*ntegrated-*m*oving *a*verage process. Such a process is structurally similar to a lag-p autoregressive process, but with two basic differences. First, the random component for the

ARIMA process is a moving average of length q of independent random components. For an autoregressive process, $q = 1$. Second, the random variable at time t denotes a dth order difference of random variables of other time points defined by the $(d-1)$st difference. For $d = 0$, the random variable at time t is the random variable at time t. For an autoregressive process, $d = 0$. Thus one refers to a (p,d,q) ARIMA process. The difference operation leads to a nonstationary process.

O'Connell[30] has noted that ARIMA processes with p, d, and q small exhibit long-memory characteristics and may be used to generate synthetic flows, such that statistical resemblance in terms of h can be achieved over the design life of a water-resource system. While the mathematical tractability of ARIMA processes is attractive, perhaps a more attractive property for hydrologic studies is that ARIMA processes can admit correlograms having negative serial correlation coefficients. Both fractional noise and broken-line processes have positive correlograms. Multivariate ARIMA synthetic flow generators are currently being developed (O'Connell, oral communication, 1972).

3-12 COMMENTS

Following the pioneering work of Thomas and Fiering,[32] interest has grown rapidly in the development of synthetic flow generators and the use of synthetic flow sequences in water-resource system planning. Historically, synthetic flow sequences have been regarded as being of little use by themselves, but coupled with simulation of a water-resource system, they provide a powerful tool for system planning. Apart from system planning, synthetic flow sequences, perhaps with different definitions of statistical resemblance, afford a means for Monte Carlo solutions to a variety of hydrologic problems.

Hydrologists have long recognized the shortcomings of many classical statistical procedures in that the assumptions underlying these procedures are seldom met by hydrologic data. The rather low powers of some tests for serial dependence, particularly when $h > \frac{1}{2}$, has been noted by Wallis and Matalas.[5] Given that streamflow sequences are generated by such models as those discussed above, it may be extremely difficult to develop analytical solutions for a body of statistical procedures such that the procedures are compatible with the assumptions underlying the flow generator. With the availability of large-scale computers, however, Monte Carlo solutions are not out of reach.

At present, a fairly broad spectrum of models exist for generating synthetic flows, and undoubtedly, new models will be developed for achieving still greater degrees of statistical resemblance. From the point of view of the water-resource system planner, it is difficult to justify the need for more complex synthetic flow generators without some assessment of the impact of greater degrees of statistical resemblance on the decision-making process.

Concurrent with future work on model development, there is need for

research in a number of areas, among them being the following. One would be an assessment of the sensitivity of the water-resource planning process to various degrees of statistical resemblance of synthetic flow generators. As the complexities of the generators increase, the demands for historical data are likely to increase, as well as the computational demands. Such assessments could provide inputs to those concerned with the design of hydrologic data-collection systems, as well as guidance in the development of synthetic flow generators. Second, the nature of the bias associated with the statistical estimates derived from synthetic flow sequences should be determined, the effects of the biases on the planning processes assessed, and techniques developed for bias corrections.

The mathematics for a particular synthetic flow generator do not assure its operational capability. The historical data may not satisfy the model constraints or yield consistent matrices of statistical descriptors, particularly for ungaged locations or for gaged locations where there are "missing" data. A closer look at the operational capabilities of synthetic flow generators is warranted.

While numerous problems associated with synthetic flow generators have been discussed, the problems should in no way detract from the use of synthetic flow sequences in water-resource system planning. The use of these sequences in planning affords a means of assessing the performances and the inherent risks associated with alternative system designs. Even under less-than-favorable circumstances such assessments would be perhaps better than none.

REFERENCES

1. HUFSCHMIDT, M. M., and M. B. FIERING, "Simulation Techniques for Design of Water Resource Systems," Harvard University Press, Cambridge, Mass., 1968.
2. FIERING, M.B., "Streamflow Synthesis," Harvard University Press, Cambridge, Mass., 1967.
3. FIERING, M. B., and B. JACKSON, Synthetic Streamflows, *Water Resources Monograph 1*, American Geophysical Union, 1971.
4. MANDELBROT, B. B., and J. R. WALLIS, Computer Experiments with Fractional Gaussian Noises, 1, Averages and Variances, *Water Resour. Res.*, vol. 5, no. 1, pp. 228–241, 1969.
5. WALLIS, J. R., and N. C. MATALAS, Correlogram Analysis Revisited, *Water Resour. Res.*, vol. 7, no. 6, pp. 1448–1459, 1971.
6. MATALAS, N. C., and J. R. WALLIS, Statistical Properties of Multivariate Fractional Noise Processes, *Water Resour. Res.*, vol. 7, no. 6, pp. 1460–1468, 1971.
7. HURST, H. E., Long-Term Storage Capacity of Reservoirs, *Trans. Am. Soc. Civ. Eng.*, vol. 116, pp. 770–799, 1951.
8. HURST, H. E., Methods of Using Long-Term Storage in Reservoirs, *Proc. Inst. Civ. Eng.*, reprint no. 6059, 1956.

9. FELLER, W., The Asymptotic Distribution of the Range of Sums of Independent Random Variables, *Ann. Math. Stat.*, vol. 22, pp. 427–432, 1951.
10. WALLIS, J. R., and N. C. MATALAS, Small Sample Properties of H and K—Estimators of the Hurst Coefficient, h, *Water Resour. Res.*, vol. 6, no. 6, pp. 1583–1594, 1970.
11. MATALAS, N. C., Mathematical Assessment of Synthetic Hydrology, *Water Resour. Res.*, vol. 3, no. 4, pp. 937–945, 1970.
12. KENDALL, M. G., "A Course in Multivariate Analysis," Charles Griffin & Company, Ltd., London, 1961.
13. YOUNG, G. K., Discussion of Mathematical Assessment of Synthetic Hydrology, *Water Resour. Res.*, vol. 4, no. 3, pp. 681–682, 1968.
14. AITCHISON, J., and J. A. C. BROWN, "The Log-Normal Distribution," Cambridge University Press, London, 1957.
15. MATALAS, N. C., On the Development of an Operational Hydrology, Appendix D, in R. K. Davis, "The Range of Choice in Water Management," pp. 178–185, The Johns Hopkins Press, Baltimore, 1968.
16. MATALAS, N. C., and J. R. WALLIS, Correlation Constraints for Generating Processes, pp. 697–707, Proceedings of the International Symposium on Mathematical Models in Hydrology, Warsaw, Poland, July 26–31, 1971.
17. SLACK, J. R., I Would If I Could, *Water Resour. Res.*, vol. 9, no. 1, pp. 247–249, 1973.
18. WALLIS, J. R., and N. C. MATALAS, Sensitivity of Reservoir Design to the Generating Mechanism of Inflows, *Water Resour. Res.*, vol. 8, no. 3, pp. 634–641, 1972.
19. WALLIS, J. R., and P. E. O'CONNELL, The Small Sample Estimation of ρ_1, *Water Resour. Res.*, vol. 8, no. 3, pp. 707–712, 1972.
20. JENKINS, G. M., and D. G. WATTS, "Spectral Analysis and Its Applications," Holden-Day, Inc., Publisher, San Francisco, 1968.
21. BOX, G. E., and G. M. JENKINS, "Time Series Analysis Forecasting and Control," Holden-Day, Inc., Publisher, San Francisco, 1970.
22. BEARD, L. R., *Monthly Streamflow Simulation*, Computer Program, 723-X6-L2340, Hydrologic Engineering Center, U.S. Corps of Engineers, Sacramento, Calif., 1970.
23. FIERING, M. B., Schemes for Handling Inconsistent Matrices, *Water Resour. Res.*, vol. 4, no. 2, pp. 291–297, 1968.
24. CROSBY, D. S., and T. MADDOCK III, Estimating Coefficients of a Flow Generator for Monotone Samples of Data, *Water Resour. Res.*, vol. 6, no. 4, pp. 1079–1086, 1970.
25. BENSON, M. A., and N. C. MATALAS, Synthetic Hydrology Based on Regional Statistical Parameters, *Water Resour. Res.*, vol. 3, no. 4, pp. 931–945, 1967.
26. DITLEVSEN, O., "Extremes and First Passage Times with Applications in Civil Engineering," doctoral thesis, Technical University of Denmark, Copenhagen, 1970.

27. MEJIA, J. M., "On the Generation of Multivariate Sequences Exhibiting the Hurst Phenomenon and Some Flood Frequency Analysis," doctoral dissertation, Colorado State Univ., Ft. Collins, 1971.

28. GARCIA, L. E., D. R. DAWDY, and J. M. MEJIA, Long-Memory Monthly Streamflow Simulation by a Broken-Line Model, *Water Resour. Res.*, vol. 8, no. 4, pp. 1100–1105, 1972.

29. CARLSON, R. F., A. J. A. MACCORMICK, and D. G. WATTS, Application of Linear Random Mcdels to Four Annual Streamflow Series, *Water Resour. Res.*, vol. 6, no. 4, pp. 1070–1078, 1970.

30. O'CONNELL, P. E., A Simple Stochastic Modeling of Hurst's Law, pp. 327–358, Proceedings of the International Symposium on Mathematical Models in Hydrology, Warsaw, Poland, July 26–31, 1971.

31. MANDELBROT, B. B., On the Broken Line Approximation to Fractional Noise, *Water Resour. Res.*, vol. 8, no. 5, pp. 1354–1356, 1972.

32. THOMAS, H. A., and M. B. FIERING, Mathematical Synthesis of Streamflow Sequences for the Analysis of River Basins by Simulation, "Design of Water Resource Systems," Harvard University Press, Cambridge, Mass., 1962.

4
GROUND-WATER MODELS

Chester C. Kisiel (deceased) and Lucien Duckstein

On joint appointment with
Department of Hydrology and Water Resources
and
Department of Systems and Industrial Engineering
University of Arizona, Tucson

4-1 INTRODUCTION

Ground-water hydrology, until recent years, has been dominated by a focus on well hydraulics and the safe or alternate yield approach to aquifer management. Numerous solutions for idealized pumping cases have been derived for confined, unconfined, and leaky aquifers; evident now is a similar path in the theoretical study of hydrodynamic dispersion of pollutants in aquifers. The gradual accumulation of more field data for a variety of aquifer prototypes has resulted in a rather recent critical evaluation of theoretical solutions and a recognition of the inverse problem.[1,2,3] The coupling of hydraulic and chemical models of aquifers with management models of aquifers is being increasingly aided by digital computer methods.[4,5,6,7,8] With the shift to use of computerized aquifer models for short-term and long-term management of aquifers, important questions arise as to the reliability of forecasts of water levels and chemical patterns and the compatibility of management models with actual socioeconomic processes. This chapter is written with the purpose of identifying the current state of the modeling art and its deficiencies. The focus is on emerging approaches to describing and managing the ground-water system and not on the many theoret-

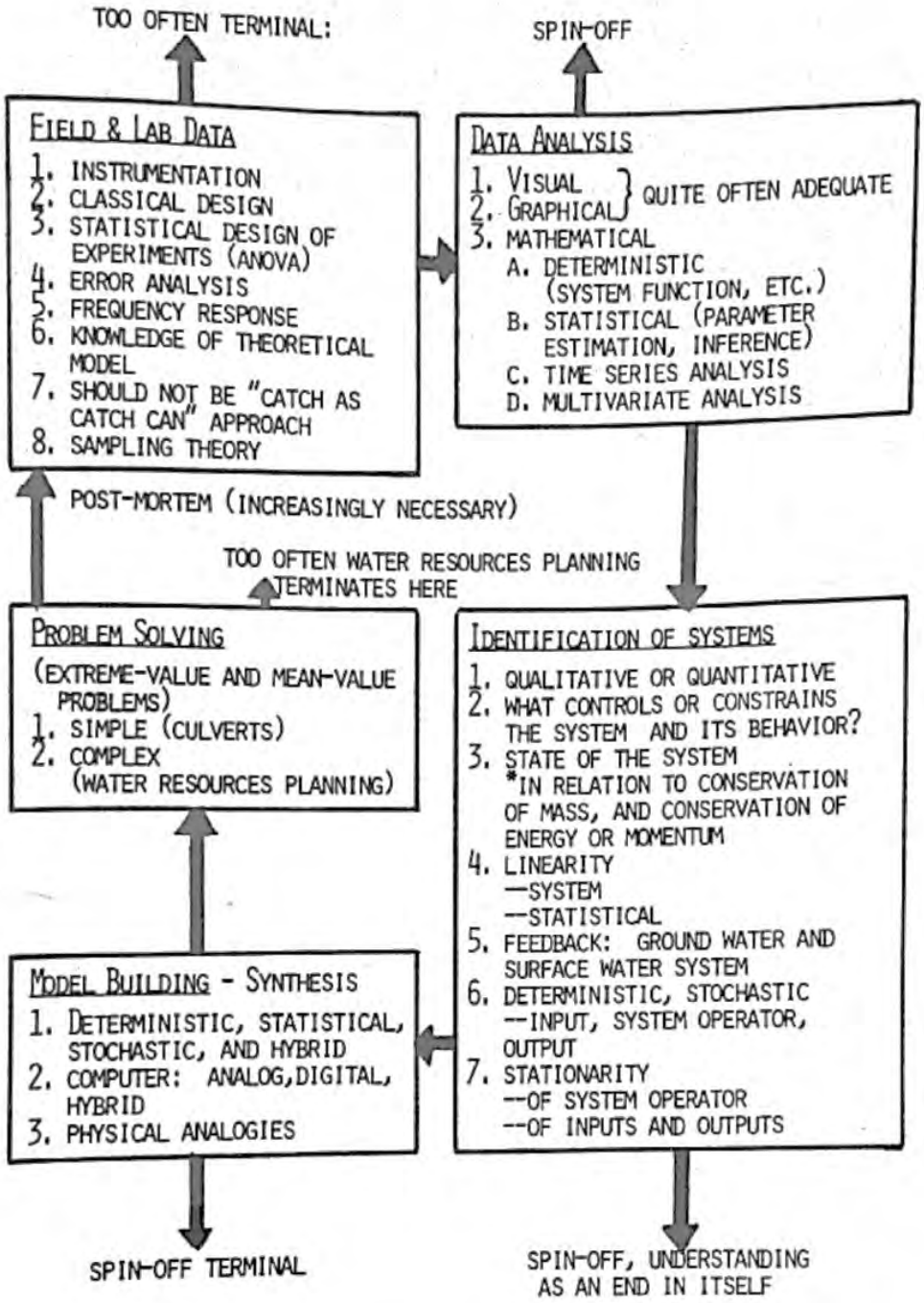


FIGURE 4-1
Cycle of model building.[13]

ical and empirical solutions to regional and specialized ground-water problems.[9,10,11,12] We select some methodologies and show in a few examples how they are used.

An overview of the systems approach may be gained by considering briefly the elements of Fig. 4-1. Figure 4-1 demonstrates that model building does not arise in a vacuum. Aquifer models may be working hypotheses

for further laboratory, field, and computer studies[15] on the road to scientific perfection, or they may be the basis of actual management decisions.[1,4,5,6,7,8] Thus, models can be concurrently conclusion-oriented (in the sense of the scientist) or decision-oriented (in the sense of the water manager). Because of the gaps between reality and model structure, considerable allowances must frequently be made in terms of the use to which the aquifer model is put. It is for this reason that Fig. 4-1 reminds us that the problem-solving activity should not terminate efforts to improve the underlying models used to make decisions and to collect additional data to define more accurately aquifer parameters (hydraulic and chemical dispersivity, and storage coefficient), recharge, pumpage, and water levels. In our view of the systems approach, it is increasingly important to

Identify risks and uncertainties
Evaluate their socioeconomic consequences
Evaluate model forecasts
Evaluate sensitivity of decisions to types of models used for the physical, chemical, and thermal behavior of the aquifer
Evaluate the adequacy of existing ground-water data networks in a more rational manner
Formulate aquifer management as a truly multiobjective problem[16]

This chapter is organized as follows. The problems of ground-water identification and management are first placed in a general systems framework in Secs. 4-2 to 4-6. This orientation allows a formal definition of each type of inverse problem and a clearer identification of nonstochastic and stochastic uncertainties. In Secs. 4-7 to 4-12 methods are presented for coping with inverse problems and identify model choice, calibration, and validation questions. In Secs. 4-13 to 4-18 various methods such as cost effectiveness analysis and sequential multiobjective problem solving as a framework for aquifer management are introduced. Finally, in Secs. 4-19 to 4-22 problems and methods in judging errors and the worth of ground-water data and in designing ground-water data networks are outlined.

4-2 GENERAL SYSTEMS FRAMEWORK

The general systems framework encompasses in a consistent manner the diversity of systems: natural, man-machine, and man-machine-nature. Of prime concern is the change in system states over time. States and time are discrete, continuous, or neither (see Table 4-1). Models may be classified accordingly. States may be discrete by definition; for example, an event has occurred (in state 1) or it has not occurred (in state 0); that is, a pump is on or off, or a well is in a given location or it is not. In digital computer simulation models, states of continuous systems are discretized to permit use of sequential logic (a discrete

operation). This is the basis of a finite-state machine (FSM) to be discussed after introducing the continuous systems framework in the next section.

4-3 GENERAL SYSTEMS THEORY AND GROUND-WATER HYDROLOGY

Rapidly evolving are the concepts of general systems theory—both quantitative[17] and qualitative—in an effort to standardize and unify knowledge in diverse disciplines. The main advantage of this theory is its elegant, logical language and a generality for modeling complex and diverse systems. It permits a clearer specification of the inverse problems to be broached later. The theory is complementary to the basic disciplinary knowledge in ground-water hydrology.

For continuous systems, Wymore [17,18] defines an assemblage as a sixtuple

$$Z = \{S, P, F, M, T, \sigma\}$$

where S = nonempty set of system states

P = nonempty set of input values; includes constraints, controlled inputs, and uncontrolled inputs

F = admissible set of input forcing functions; that is, a set of functions defined on $\{t : t \in T\}$ and with values in P. F is admissible if, for functions $f, g \in F$,

1. F is not empty
2. For $t \in T$, $f(t + \tau) \in F$ for all $\tau \in T$
3. $\{f(t): t > 0\} \cup \{g(t): t > 0\} \in F$; that is, the union of two input functions belongs to the set F

M = set of mappings from the cartesian product $F \times T$ onto σ, that is, the mapping of a set of inputs at a specific time t onto the set of state transition functions σ. For most systems, M is simply the range of σ

T = the set of all real numbers; denotes time

Table 4-1 CLASSIFICATION OF SYSTEM STATES AND TIME AND IDENTIFICATION OF MODELS

		State of systems		
		Discrete	Continuous	Neither
Time	Discrete	Difference equations Digital computer simulation models Dynamic programming	Sampled data systems (measured water levels in aquifer) Dynamic programming	
	Continuous	Electronic analogy models of aquifers	Differential equations	
	Neither	Integer programming Network flow problems	Linear programming Nonlinear programming	Mixed-integer allocation problems

σ = set of all modes of behavior available to the system, that is, σ is the set of state transition functions defined on the mapping of S onto S, including the identity mapping

A system is defined as an assemblage $\mathscr{Z} = \{S, P, F, M, T, \sigma\}$ such that (see Fig. 4-2A):

1 $\sigma(f,t=0)$ = identity mapping of S onto S for all $f \in F$, exists. This ensures that the system is defined in its initial state.

2 $\sigma(f(t+s),t) o \sigma(f(s),t) = \sigma(f(t),s+t)$, where the composition (o) of two functions f and g is defined as $fog = h$, and where f is a mapping from A to B, g is a mapping from B to C, and h is a mapping from A to C; this property means that we can displace the time origin of input function for a fixed system and obtain the same result if we displace the system over time for a fixed input function.

3 $\sigma(f,t) = \sigma(g,t)$ for all $f, g \in F$ and $t \in T$ if $f(t) = g(t)$ for $t \in [o,T)$. Here we ensure that the system is consistent with respect to inputs: if the state transition function remains invariant, it means that the inputs were the same.

Wymore[17] has shown that all differential equations, hence the equations of ground-water flow, satisfy the above three relations. The simpler form of these equations for piezometric head h, chemical concentration ϕ^a for species a, and temperature θ are, respectively, in continuous time and space:[19]

$$S_s \frac{\partial h}{\partial t} = \frac{\partial}{\partial x}\left(K_{xx}\frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_{yy}\frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_{zz}\frac{\partial h}{\partial z}\right) + Q \qquad (4\text{-}1)$$

where Q = source functions (same as the set of input functions F)
S_s = coefficient of specific storage
K_{ii} = the directional permeability in an incompressible medium

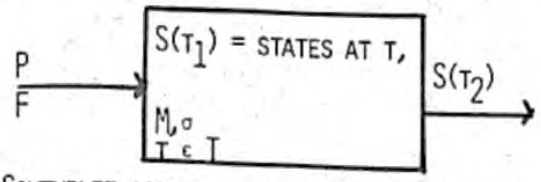
$$\frac{\partial \phi^a}{\partial t} + u\frac{\partial \phi^a}{\partial x} + v\frac{\partial \phi^a}{\partial y} + w\frac{\partial \phi^a}{\partial z} = \frac{\partial J_x^a}{\partial x} + \frac{\partial J_y^a}{\partial y} + \frac{\partial J_z^a}{\partial z} + R_a \qquad (4\text{-}2)$$

where J_i^a = mass flux of chemical a due to diffusion in direction i
R_a = generation or destruction of chemical a (net amount)

$$\rho C\frac{\partial \theta}{\partial t} + \rho_w C_w\left(u\frac{\partial \theta}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}\right) = \left(\frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z}\right) + R_\theta \qquad (4\text{-}3)$$

where ρ = density of rock and fluid complex
ρ_w = density of fluid
H_i = heat flux due to thermal gradient in direction i
C = specific heat of rock and fluid complex
C_w = specific heat of fluid
R_θ = generation or dissipation of heat (net amount)

Use of Darcy's law, $q = -[K]\ \nabla h$, permits evaluation of u, v, and w in Eqs.



(A) SIXTUPLET ASSEMBLAGE ς FOR CONTINUOUS SYSTEMS

SET OF UNCONTROLLED OR PASSIVE INPUTS

- INFLOW FROM ADJACENT NODES
- NATURAL RECHARGE
- EVAPOTRANSPIRATION
- POPULATION GROWTH
- WATER DEMAND
- SUBSIDENCE

SET OF CONTROLLED OR ACTIVE INPUTS

- SCHEDULE OF PUMPAGE
- ARTIFICIAL RECHARGE
- INTEREST RATE
- PUMP TAX
- NEW WELLS
- POLICY FOR CONJUNCTIVE MANAGEMENT OF SURFACE AND GROUNDWATER RESOURCES

STATE OF AQUIFER (T_1)

- HYDRAULIC
- CHEMICAL
- THERMAL
- HYDROGEOLOGIC

STATE AT T_2

- BENEFITS TO COMMUNITY
- COSTS TO COMMUNITY
- OUTFLOW FROM AQUIFER
- SUBSIDENCE

(B) DESCRIPTIVE DEFINITION OF THE AQUIFER MANAGEMENT PROBLEMS (DISCRETE AND CONTINUOUS).

F
RANDOM
DETERMINISTIC

STATE TRANSITION FUNCTIONS

$$\nabla^2 H - \frac{S}{T}\frac{\partial H}{\partial T} = Q$$

$$D^A \nabla^2 \phi^A - \frac{\partial \phi^A}{\partial T} - Q \cdot \nabla \phi^A = R_A$$

$$\nabla^2 \Theta - \frac{\rho C}{K\Theta}\frac{\partial \Theta}{\partial T} - \frac{\rho_W C_W Q}{K} \cdot \nabla \Theta = R_\Theta$$

$H(X,Y,Z,T)$
$\phi^A(X,Y,Z,T)$
$\Theta(X,Y,Z,T)$

(C) STATE TRANSITION FUNCTIONS FOR CONTINUOUS AQUIFER SYSTEMS.

X(T)
RANDOM
DETERMINISTIC

STATE TRANSITION FUNCTIONS S(T)

.HYDRAULIC:	INFLOW - OUTFLOW = TIME RATE OF CHANGE OF STORAGE	$H(X,Y,Z,T)$
.CHEMICAL:	CONSERVATION OF MASS EQUATION AS ABOVE FOR EACH CHEMICAL SUBSTANCE	$\phi^A(X,Y,Z,T)$
.THERMAL:	CONSERVATION OF HEAT EQUATION AS ABOVE	$\Theta(X,Y,Z,T)$

(D) STATE TRANSITION FUNCTIONS FOR DISCRETE AQUIFER SYSTEMS

FIGURE 4-2
System representation of aquifers.

(4-2) and (4-3); whereas, J_i^a and H_i are obtained, respectively, from Fick's law

$$J_i^a = -[D^a]\,\nabla\phi^a \tag{4-4}$$

and from Fourier's law

$$H_i = -[K_\theta]\,\nabla\theta \tag{4-5}$$

where $[D^a] = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}$ = diffusion tensor for species a

$[K_\theta]$ = thermal conductivity tensor

$\nabla\phi^a$, $\nabla\theta$ = gradients of concentration and temperature, respectively

With these constitutive equations and by neglecting heterogeneity and anisotropy, Eqs. (4-1), (4-2), and (4-3) become, respectively,

$$\nabla^2 h + Q = \frac{S}{T}\frac{\partial h}{\partial t} \tag{4-6}$$

$$D^a \nabla^2 \phi^a + R_a = \frac{\partial \phi^a}{\partial t} + \mathbf{q} \cdot \nabla^a \tag{4-7}$$

$$\nabla^2 \theta + R_\theta = \frac{\rho c}{K_\theta}\frac{\partial \theta}{\partial t} + \frac{\rho_w c_w}{K}\,\mathbf{q} \cdot \nabla\theta \tag{4-8}$$

where D^a = diffusion coefficient for species a

K = hydraulic conductivity

$\mathbf{q}$ = vector velocity (u,v,w)

Equations (4-6) to (4-8) are linear partial differential equations (PDEs) for homogeneous isotropic aquifers and are commonly invoked as initial models of hydraulic, chemical, and thermal behavior in aquifers, particularly to obtain analytical solutions for rather special cases.

The correspondence between the assemblage $\mathscr{Z}$ and the above equations is as follows:

$\mathscr{Z}$	Notation in partial differential equations
S	h, ϕ^a, θ
P	Range of values of Q, R_a, R_θ
F	$Q(x,y,z,t)$, $R_a(x,y,z,t)$, $R_\theta(x,y,z,t)$
M	Range of $h(x,y,z,t)$ $\phi^a(x,y,z,t)$, $\theta(x,y,z,t)$, that is, range of S, in response to F
T	Time period over which F is active
σ	Equations (4-6) to (4-8)
$\sigma(f,o)$	Initial conditions

Thus, to model a ground-water system, all elements of the sixtuplet $\mathscr{Z}$ must be specified, but as noted later, this is not necessarily a simple task.

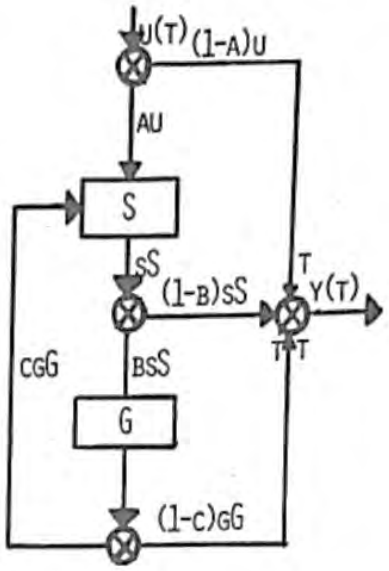


FIGURE 4-3
Flow chart of lake-aquifer system.[21]

4-3.1 Example: Continuous Systems Model of a Lake-Aquifer System

The conditions given previously for the assemblage $\mathscr{Z}$ may be written as a pair of equations which we call the *vector-state equations* for continuous deterministic systems[20]

$$\mathbf{Z}(t_0, t) = \mathbf{n}\,[\mathbf{S}(t_0); \mathbf{X}(t_0, t)] \tag{4-9}$$

$$\mathbf{S}(t) = \mathbf{m}\,[\mathbf{S}(t_0); \mathbf{X}(t_0, t)] \tag{4-10}$$

where both **m** and **n** are singled-valued functions. Equation (4-9) states that the output **Z** over the interval t_0 to t is a single-valued function of the state $\mathbf{S}(t_0)$ at the beginning of the interval and the input $\mathbf{X}(t_0, t)$ over this time interval. The state at the end of the interval is said, in Eq. (4-10), to be a *single-valued function* of the same argument.

If the system can be described by a set of linear ordinary differential equations, the state equations can be written

$$\dot{\mathbf{S}}(t) = \mathbf{A}(t)\mathbf{S}(t) + \mathbf{B}(t)\mathbf{X}(t) \tag{4-11}$$

$$\mathbf{Z}(t) = \mathbf{C}(t)\mathbf{S}(t) + \mathbf{D}(t)\mathbf{X}(t) \tag{4-12}$$

where **A**, **B**, **C**, and **D** are, in general, time-varying matrices and **S**, **X**, and **Z** are vector states, inputs, and outputs (same notation as for the finite-state machine in a later section). Consider the combined lake-aquifer system with storage elements $S(t)$ and $G(t)$, respectively, given in Fig. 4-3; of concern is the operation of this system as a lumped linear storage for water supply purposes $Y(t)$. By mass balance, the system response is, corresponding to the scalar form of Eq. (4-12),

$$Y(t) = (1 - b)sS + (1 - c)gG + (1 - a)U \tag{4-13}$$

in which a, b, c, g, and s are constants that may be subject to control. The input $U(t)$ may be deterministic or stochastic. The response $Y(t)$ may or may not meet municipal requirements.

Similarly, by continuity, the state equations for each of the lumped elements are, respectively,

$$\dot{S} = \frac{dS}{dt} = -sS + cgG + aU \tag{4-14}$$

and

$$\dot{G} = \frac{dG}{dt} = bsS - gG \tag{4-15}$$

In vector matrix form, these coupled differential equations are

$$\begin{bmatrix} \dot{S} \\ \dot{G} \end{bmatrix} = \begin{bmatrix} -s & cg \\ bs & -g \end{bmatrix} \begin{bmatrix} S \\ G \end{bmatrix} + \begin{bmatrix} a \\ o \end{bmatrix} U \tag{4-16}$$

and have the form of (4-11). Duckstein and Kisiel[21] give solutions to Eqs. (4-13) and (4-16) for both the homogeneous and nonhomogeneous cases. They also use these results to solve a control problem wherein the goal is to find a set of control values $a(t)$, $b(t)$, and $c(t)$ such that $Y(t)$ be close to the demand $D(t)$ and the cost of control be minimum. To do so requires specification of a cost function for water deficiencies and its use to solve a calculus of variations problem to obtain the optimal control.

After this introduction to general system modeling, let us examine some specific linear system models of ground water.

4-4 CLASSICAL GROUND-WATER EQUATIONS AND LINEAR SYSTEMS THEORY

Here we introduce linear systems theory, primarily used in surface-water hydrology and its relation to ground-water systems analysis. The primary focus is on unit step and unit impulse response functions and use of the latter in the convolution relation.

The earliest closed-form solution to Eq. (4-6) was obtained by Theis[22] in 1935 for the condition of a fully penetrating well of small diameter in an infinite, homogeneous, isotropic, artesian aquifer. The drawdown $s(r,t)$ for a pumping rate of Q (a step input) is

$$s(r,t) = \frac{Q}{4\pi t} \int_u^\infty \frac{e^{-u}}{u}\, du \tag{4-17}$$

where $(e^{-u}/u)\, du$ is popularly referred to as the "well function" $W(u)$, $u = r^2S/4Tt$, and r = radial distance from well. Equation (4-17) and its graphic analogues (type curves) have found extensive use in ground-water pumping

tests. The type-curve solution to this equation has been modified to include delayed drainage (for the water-table case) and leakage through aquitards (for the confined aquifer case).

Recognizing that both Eq. (4-6) and the boundary conditions are linear, it is apparent that the methodology of linear systems theory would be suitable to solution of aquifer problems governed by Eq. (4-6). Of course, the system is deterministic; impulse response (influence or kernel) functions in one-dimensional form have been derived for a variety of boundary conditions attached to Eq. (4-6).

For a semi-infinite aquifer with boundary conditions, head $h(o,t) = H(o,t)$, and $h(x,t)$ being bounded as $x \to \infty$, and with initial condition $h(x,o) = 0$, Venetis[23] has derived the impulse response function

$$U(x,t) = \frac{\sqrt{S/T}\, \exp\left[(S/T)x^2/4t\right]}{2\sqrt{\pi t^{3/2}}} \qquad \text{for } t > 0 \tag{4-18}$$

The response for zero initial conditions to any input (arbitrary, stochastic, deterministic) $H(o,t)$ is then given by the convolution integral[24]

$$h(x,t) = \int_0^t U(x,t-\tau)\, H(o,\tau)dt \tag{4-19}$$

in which τ is the dummy variable of integration or the time variable that identifies the future effects of past inputs $H(o,\tau)$ in the interval $0 \leq \tau \leq t$.

More appropriate for application to drawdown from a well is the impulse response function as derived by Moench:[25]

$$U(r,t-\tau) = \frac{\exp\left[\dfrac{-u}{t-\tau}\right]}{4\pi T(t-\tau)} \tag{4-20}$$

and

$$h(r,\, t) = \int_0^t Q(\tau)\, U(r,t-\tau)\, d\tau \tag{4-21}$$

If $Q(\tau) = Q$ in Eq. (4-21), we reobtain Eq. (4-17). All the above kernel functions are time-invariant or stationary; i.e., the parameters S and T do not vary over the period of pumping. Also, because they are functions of x or r, their form changes with increasing distance from the pumping well.

The convolution integral, Eq. (4-19) or (4-21), may also be identified as the particular solution, the solution to the nonhomogeneous case, or the forced response. For nonzero initial conditions in a linear system context, we must add the transient response or the solution to the homogeneous case[20]

$$h(x,t) = H(o,o)\, U\,(x,t) + \int_0^t U\,(x,t-\tau)H(o,\tau)d\tau \tag{4-22}$$

At this point, we state that the output $h(x,t)$ of a completely known determinis-

tic system is uniquely determined by knowledge of the initial state of the system and the *future* input function $H(x,t)$ for $t \geq t_0$. These are strong requirements and lead to the definitions of various identification problems in Secs. 4-5 and 4-6.

The importance of the impulse response function lies in the fact that Q in Eq. (4-21) and H in Eq. (4-19) can now be considered both as known and stochastic functions of time. Moench and Kisiel[26] have used the convolution integral approach to estimate the recharge from an ephemeral stream. This approach is more general than the classical methods currently in vogue in ground-water hydrology.

Linear systems theory also includes a study of system response in the time-frequency domain and spatial wave-number domains. Fourier transforms of the kernel functions in Eqs. (4-18) and (4-20) permit a more systematic evaluation of spectral analyses applied to actual time-series data on ground-water systems.[27,28,29] Eriksson[28,29] has applied such methods to ground water, surface water, and meteorologic data in Sweden and developed appropriate stochastic models. However, in general, most ground-water observation networks do not have adequate lengths of sample time series to apply spectral and correlogram analysis. The linear systems framework, however, does permit a check on the assumptions of more traditional deterministic methods of evaluating ground-water data for localized situations.[30]

There are obvious limits to what can be done with linear systems theory and continuous system modeling in terms of differential equations. More general are the discrete models of aquifers. Instead of starting with partial differential equations (PDEs) and then discretizing in order to get numerical solutions, it is more natural and more general to start with discrete representations like the finite-state machine (FSM), as is done in the next section. Too often, the origins of the PDEs are overlooked in subsequent computer solutions to large-scale problems. This is not to denigrate the immense value of closed-form solutions like the Theis equation and others given in this section. Such solutions give the field analyst some base line from which to evaluate data subjectively and to make allowances for the gaps between theory and reality.[10]

4-5 AQUIFERS AS FINITE-STATE MACHINES (DISCRETE SYSTEMS)

A finite number of discrete states observed at discrete times is postulated as being consistent with the manner in which the aquifer system is measured and subsequently modeled.[31,32,33,34] Working simulation models, developed for solution on digital computers, actually function as finite-state machines (FSM). The principal use of the FSM is as a systematic method of simulating *any* system. It distinguishes easily between parameters and decision variables and provides a useful framework in which to conceptualize the interrelation between objective functions and constraints.

We specify a FSM as the assemblage $\{X, S, Z, F, G\}$ wherein the respective symbols are X = input, S = state, Z = output, F = state transition function that maps $\{X,S\}$ into S, and G = output function that maps $\{X,S\}$ into Z. It is important always to also identify the time or iteration interval $\Delta t = 1$ in terms of days, months, 1 year, or 5 years.

4-5.1 Example: FSM of Well Systems

Let us define two subsystems (see Fig. 4-3): wells in a unit square with location P are called WELL (P), and the aquifer around P as the cartesian product

$$GW(P) = \text{EL} \times \text{QUAL} \times \text{HEAT} \times \text{FLOW} \times \text{TRANS} \times \text{STOR} \times \text{IN} \times \text{OUT} \times \cdots \quad (4\text{-}23)$$

where EL = elevation or piezometric head h (continuous range of values between h_1 and h_2)
QUAL = water-quality parameters (a subcartesian product of range of values for several parameters)
HEAT = heat flow in the aquifer
FLOW = water flow in the aquifer
TRANS = transmissivity
STOR = storage coefficient
IN = natural and artificial recharge
OUT = natural and artificial (pumping) discharge

The cartesian product [Eq. (4-23)] emphasizes the dynamic properties of the aquifer around point P, explicitly identifies all state elements associated with the aquifer, and specifies the range of values for each element of $GW(P)$. It is combinatorial in nature, defines the scope of necessary information retrieval, and reflects our subjective confidence in the reliability of the data.

To illustrate the operation of the FSM, consider SCHED as the time schedule of pumped withdrawals. Given SCHED as an active input operator on wells (a human-state element) in a sector of the basin, the output to be achieved may be a certain level of income, firm quantity of water, or certain rate of economic growth. Because economic growth escalates water use, we augment the descriptive FSM model with a feedback element as possibly defined by an econometric relation between growth rate, population income, water use, and level of technology. That is, economic growth operates on SCHED.[6,7,8] To emphasize how the model simultaneously generates information on other effects of SCHED, we note that SCHED operates as well on $GW(P)$; for example, when the aquifer is mined, the model shows that TRANS, STOR, and QUAL may be functions of depth.

A FSM representation of the above well subsystem WELL (P) is developed by letting

$$X = \text{SCHED} = \{Q_1T_1, \ldots, Q_NT_N\}$$

where Q_i is the pumping rate and T_i the duration of the pumping cycle at the ith of N wells

$$S = \{\text{EL, QUAL, HEAT}\} \qquad (4\text{-}24)$$
$$Z = \{\text{EL, QUAL, HEAT, COST, BENEFITS}\}$$

For purposes of simulating the system, one needs the state transition function

$$S(t+1) = F\{S(t), X(t)\} \qquad (4\text{-}25)$$

To obtain $S(t+1)$, Eqs. (4-1) to (4-3) or their simpler forms (4-6) to (4-8) may be discretized in space and time and solved simultaneously according to finite difference[2,35,36] or finite element methods.[37] We also need the goal or objective function for evaluating the simulation as given rather generally by

$$Z(t+1) = G\{S(t), X(t)\} \qquad (4\text{-}26)$$

As noted previously, the output $Z(t)$ may be physical, chemical, biological, or socioeconomic. An economic output requires specification of an objective function, such as cost of treating polluted aquifer waters,

$$Z(\alpha,t) = p_1\phi(\alpha_i t,1) + p_2\phi(\alpha_i t,2) + \cdots = \sum_j p_j\phi(\alpha_i t,j) \qquad (4\text{-}27)$$

where p_j is the cost of removing pollutant α at the jth location (say, a spring or well). Another common objective function is the cost of pumping the aquifer,

$$Z(t) = c,Q(t,1) + c_2Q(t,2) + \cdots = \sum_k c_kQ(t,k) \qquad (4\text{-}18)$$

where c_k is the cost of pumping at the rate Q at the kth location.

4-5.2 Example: Ground Water and a Stochastic Model of Annual Streamflow

A stochastic model of annual streamflows can be shown to include a ground-water storage component.[38] Refer to Fig. 4-4 for definitions of terms. Correlated precipitation inputs are transformed to contributions to ground-water storage $V(t)$, baseflow $\gamma V(t-1)$, evapotranspiration $\beta X(t)$, and streamflow $Z(t)$. During year t, application of the mass balance relations gives

$$Z(t) = (1-\alpha-\beta)X(t) + \gamma V(t-1) \qquad (4\text{-}29)$$

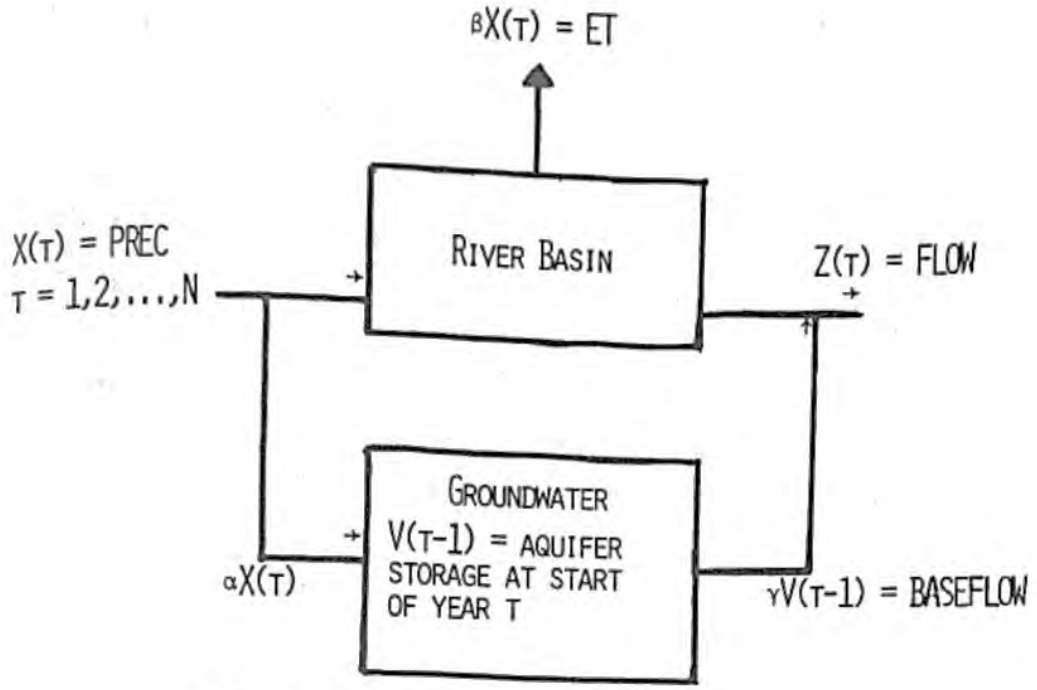
and

$$V(t) = \alpha X(t) + (1-\gamma)V(t-1) \qquad (4\text{-}30)$$

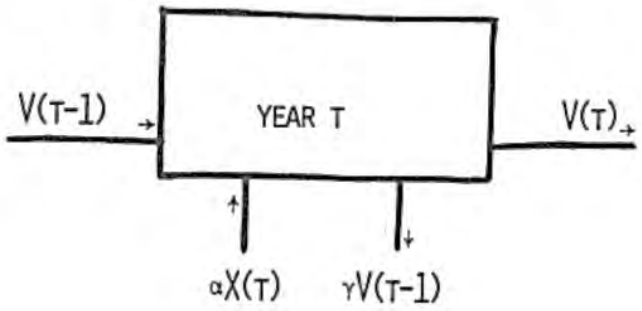
for ground-water storage at the end of year t. These equations are constrained by

$$0 \le \alpha, \beta, \gamma \le 1 \qquad (4\text{-}31)$$
$$0 \le \alpha + \beta \le 1 \qquad (4\text{-}32)$$
$$X(t), V(t) \ge 0 \qquad (4\text{-}33)$$

Values of α, β, and γ reflect the properties of the hydrologic regime (humid,



(A) FLOW CHART FOR ANNUAL WATER BUDGET IN A RIVER BASIN



(B) ANNUAL WATER BUDGET FOR GROUNDWATER STORAGE

FIGURE 4-4
Ground-water storage and stochastic streamflow.

arid, temperate, forested volcanic, glacial, and others) and will vary from year to year. The extent to which the stochastic recharge to an aquifer is important to aquifer management is determined by the ratio of annual recharge volume to aquifer volume and by proximity of important well fields to recharge zones. Development of stochastic space-time models of recharge are desirable[39] for more rational management of aquifers.

On this point, Harshbarger[10] notes that "a part of the inflow from wet years should be held in storage for later use in drought years when the demand is urgent to maintain the economy of the region." A sketch of a stochastic model of recharge along the bed of an ephemeral stream is presented later (Sec. 4-16)

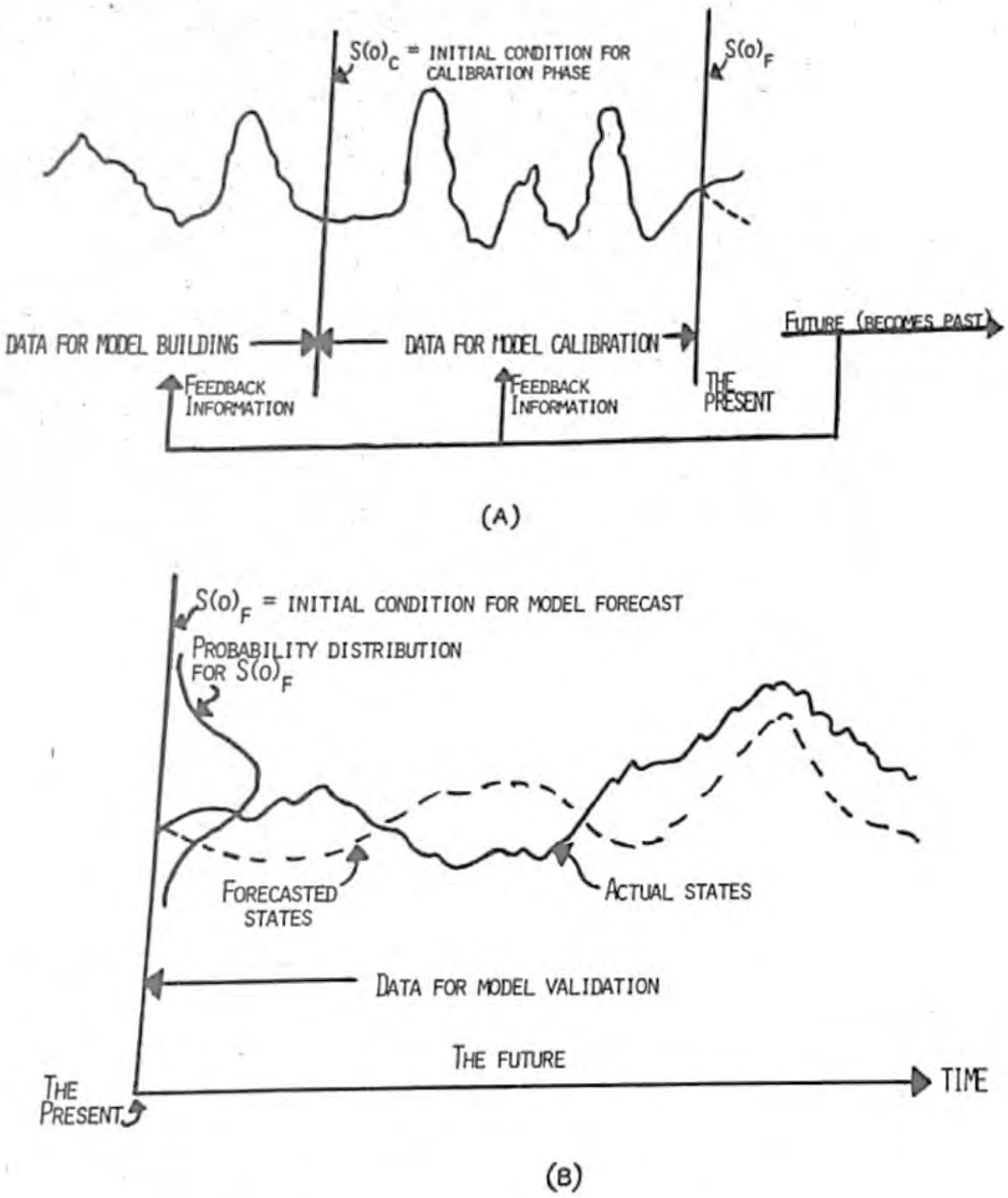


FIGURE 4-5
Data paths for model building, calibration, and validation.

4-5.3 Problems Arising with Aquifer Systems (Using the Finite-State Model)

For the present we focus on the problems of *physical* (chemical and thermal) systems analysis and systems synthesis independent of any managerial uses of the results. By *systems analysis* we mean the decomposition of the aquifer sys-

tem and its interfaces, as in Fig. 4-5, and the evaluation of systems with limited data on inputs, states, parameters, and outputs. By *systems synthesis* we mean the reconstruction of the real aquifer in terms of continuous system models, as in Eqs. (4-1) to (4-3), or in terms of numerical models of the distributed aquifer system and its couplings with the unsaturated zone, land, surface, atmosphere, and river channel.[15,40] There exists a sequential interplay between systems analysis and systems synthesis.

Model building, *calibration*, and *validation* (see Fig. 4-5) encompass both systems analysis and synthesis. But in our judgment there has been a historical tendency to focus on model building (synonymous with system synthesis) and to use the resultant models to "solve" direct problems. Future use of aquifer models in generating information for management decisions depends strongly on a resolution of the calibration and validation questions, that is, the inverse problem. Model calibration involves the matching of historical and simulated aquifer-state variables. Model validation involves truth testing *after* calibration by comparing in a one-to-one manner future aquifer forecasts with actual values of the aquifer state. These difficult requirements are forced on us by the limited data situation that mitigates the complete description of *any* aquifer system.

To use any of the deterministic ground-water models, one needs to identify the state set S, input set X, input functions $X(t)$, parameters of the state transition function F, initial condition $S(0)$, and boundary conditions. The usual assumption is that F is well defined. Even if this were the case, and if $X(t)$ were perfectly known, when $S(0)$ is not defined perfectly at "all" points in the aquifer, subsequent predictions with F constitute sample time series from a *perfectly determined stochastic process*.[41,42] Relevant to identification of $S(0)$ is the realization that Laplace's equation may be concurrently applicable for hydraulic and thermal behavior but not for chemical substances because of the unsteady nature of the aquifer mixing process.

State identification in both steady-state and dynamic systems requires identification of F. Such identification may be based on previously known equations of the process—such as Eqs. (4-1) to (4-3) or (4-6) to (4-8)—or on the assumption of a linear system[26,43,44] but without introducing detailed structure as is embodied in Eqs. (4-1) to (4-3). In both cases, input-output data are required. Table 4-2 gives the identification and other problems for deterministic models in terms of input, initial condition, state transition, and output. The issues surrounding these problems and their resolution are more formally presented in Secs. 4-7 to 4-12. It should be noted that "relative solution" of the identification problem has as its goal the formulation of the direct (prediction) problem and its solution for management of aquifers.

The combined identification problem described in Dooge[43] is very real and deserves serious evaluation. Quite often only output information on quantity and quality of water in a few wells, springs, and baseflow is available. Amount of data in space and time is sparse. Sometimes, stochastic analyses

Table 4-2 INPUT IDENTIFICATION PROBLEMS (MORE COMPLICATED SITUATIONS EXIST BUT ARE NOT SHOWN)

	Given data in		Given prior knowledge about model structure (yes or no)		Required (yes or no)			
Problem	Time	Space	Deterministic	Stochastic	Structure of model	Model parameters	Reliability of estimate for intended use	Appropriateness of model for forecasting
Find hydrograph of natural recharge	Water levels in a well for a single high flow	One well only	Yes; transient function	No	No	Yes	Yes	Yes
Find streamflow hydrograph	Water levels in a well for a single high flow	One well only	No	No	Yes	Yes	Yes	Yes
Regional recharge	Water levels in a well for a single high flow	Very few wells	Yes	No	No	Yes	Yes	Yes
Regional recharge	Water levels in a well over a few years (small sample)	Very few wells	No	Yes; Markov, nongaussian	No No	Yes Yes	Yes Yes	Yes Yes

Regional recharge	Water levels in a well over a few years (small sample)	Very few wells	No	No	Yes	Yes	Yes	Yes
Find streamflow hydrograph	Water levels in a well over a few years (small sample)	Very few wells	No	No	Yes	Yes	Yes	Yes
Find source and path of water parcels at different depths of aquifer	Environmental tracers Controlled or uncontrolled pollution Temperature measurements	Only a few locations	Yes	Perhaps	Yes	Yes	Yes	Yes
Find aquifer parameters	Quality and discharge of springs and baseflow	Only a few locations	Yes	Perhaps	Yes	Yes	Yes	Yes

are invoked to evaluate such situations.[28,29,32,44] Clearly, the indeterminacy can be reduced only by a willingness to invoke prior knowledge on other aquifer systems or to collect more data in space and time. Both strategies engender risks and uncertainties and deserve rather serious formal analysis by groundwater hydrologists. Some aspects of such analysis are given in Secs. 4-13 to 4-22.

Input identification of aquifer systems requires consideration of spatial and temporal issues, of deterministic and stochastic components in the disturbance, and of the past and future (see Table 4-2). This identification problem becomes more important as the need develops for finer tuned management and prediction of chemical quality.

To develop further the *indeterminacy* and *uncertainty* issues, consider a model (machine) with five possible states S_i, $i = 1, \ldots, 5$, and two possible inputs X_j, $j = 1,2$. If the system is completely determined, the model is termed an *information lossless machine* in a finite number of steps. However, most natural systems are indeterminate, and as a consequence there are many possible paths that the FSM could have followed to get from state 2 to 3 in 5 time units. Such possible paths are 25313 or 24523. In fact, the following state transitions are possible:

State at time t	Input sequence	State at time $(t+5)$	Output sequence
2	$\{A\}$	3	$\{\alpha\}$
2	$\{B\}$	3	$\{\beta\}$

Thus, in the absence of output, the input which drove the system from 2 to 3 in five transitions is undetermined. The indeterminacy is in F because, as exemplified above, F is not a one-to-one mapping. Notice that nothing is being said here about stochasticity of initial conditions and measurement and sampling errors in time; these are additional issues superimposed on the problem of system indeterminacy. Another way of noting this deterministic indeterminacy is to state that there are more unknowns than equations. To reduce the indeterminacy would require finding the required additional equations from knowledge on the devious pathways by which the inputs gave rise to the transitions from 2 to 3. From a sampling standpoint this is indeed a difficult task. The presence of an observable output may remove this indeterminacy as shown above. Basically, system indeterminacy is *backward* in character because the analyst does not know how the system got to the state in which it finds itself at any given time.

The problem of sampling to identify X, S, F, and Z for a FSM is more complex than explained thus far. Uncertainty in state transitions may be in both a backward and a forward direction. Not only is it not known how the system arrived at its present state, but also not known is the "direction" in which

the system will go. State transitions might now be modeled probabilistically, like a Markov chain, of which a random walk is a special case. Here we are not referring to microscopic models but to the macroscopic changes in system states, inputs or outputs. Space-time sampling to define uncertain systems would have a goal to estimate the probability of going from one discrete state to another. It remains to be established as to how uncertain is the behavior of pollutants in aquifers.

4-5.4 Advantages of FSM Theory

Increasingly it is desirable to use a consistent language to facilitate dialogue between disciplines. In the face of this need, it is felt that the finite-state machine (FSM) offers the following advantages in thinking about modeling and space-time sampling of properties of aquifers:

1 Classifies various problems relative to each other.

2 Allows transferability of models from one aquifer to another without having to reiterate the whole modeling process.

3 Makes possible a logical conceptual transition from a deterministic model to a stochastic model; more precisely,

a One may have a stochastic input to a system described by a deterministic F.

b One may have a stochastic input with a stochastic F—for example, random recharge in a medium where uncertainty exists in identification of parameters.

For each of the above there is a different sampling requirement in time and space.

4 Distinguishes between:

a Transient model (varying input) and steady-state model (constant input).

b Indeterminacy (backward identification) and uncertainty (forward and backward identification).

5 Enables a simulation with high chance of success at first trial and no error in logical operations on model equations. Finite memory may be incorporated in this machine (model).

6 The following operations are possible; that is, algorithms are available:

a Minimization of system representation (to minimize redundancy or reduce dependent equations) and decomposition of system in logical manner.

b Diagnosing: to find the state $S(t)$ through experiments with input—for example, use of pump test and environmental tracers to assess disturbances induced by sampling scheme. The algorithm provides the minimum number of experiments (that is, minimum input sequence of data from sampling scheme) that are necessary to identify

state of the system; it would give a lower bound to space-sampling requirements.

c Homing: to drive the system toward a desirable state such as a prespecified salt concentration.

d Feedback: to restrict input disturbances and to contribute in monitoring the input as a consequence of predicted or observed $S(t)$ or $Z(t)$.

e Coupling of FSM for thermal, pollutant, and hydraulic behavior of aquifer.

f Study of isomorphism and homomorphism between two machines (models). This can describe the loss in going from one model to another. As an example, a lumped aquifer model (no spatial variability) is homomorphic to a distributed model, but the two are not isomorphic. Only two models of the same kind may be isomorphic such as the FSM of the aquifer in relation to the finite difference or finite element model derived from differential or integral equations. However, in this case, the FSM can be much more general than the numerical models because fewer assumptions about the system are made at the start of the modeling process.

g System representation is independent of goal function (output Z, output function G); that is, a manager, a scientist, or an engineer can use the same model, each with their own objective function. This permits more communication between disciplines.

h Bypasses gross idealizations that lead to the partial differential equations of aquifer behavior, that is, bypasses the continuum concept in favor of the way the system is sampled and modeled in practice.

i Includes reactions between aquifer and input, that is, the way the input acts on the state $S(t)$.

In sharpening the various dimensions of aquifer modeling, FSM theory helps us to identify the goals of modeling, the existence of model choice questions, and the criteria for choosing and evaluating aquifer models. These issues are introduced in the next section.

4-6 THE MODEL CHOICE PROBLEM

In considering the model choice problem in ground-water hydrology, it is critical to differentiate between the meaning of and the acquisition of knowledge on the one hand and the making of decisions on the other hand. When making a decision we must acknowledge that there is a correct decision and a wrong decision and that there are costs involved in gathering data and losses engendered when making erroneous decisions. In the knowledge problem, there is no finality, but rather a sequence of models. At each iteration we seek to

make the model more consistent with the data, our index of reality. Nowhere in this scientific process do we assert that we have the true model. We may reject models but we do not accept models as the final truth (vis à vis our earlier arguments on the backward and forward indeterminacy problems). Nonetheless, when confronted with management decisions, we will accept a model for the present as the best predictor in our possession. Yet there are potential losses when using imperfect models. At this juncture of ground-water modeling, it is quite important for the generator of models to assess the socioeconomic consequences of his creation. Who else is in a better position?

In one form or another, the key question to be answered in ground-water management is: Given constraints on resources (budgetary, technical, human, etc.) and given multiple objectives of satisfying domestic, irrigation, and industrial requirements, how best to obtain the required quantity and quality of water? In order to answer the above question, a knowledge of the following is required:

1 On the regional level:
   - *a* Identify the areas of natural recharge and discharge
   - *b* Quantitative estimates of natural recharge and discharge
   - *c* Direction and velocity of flow
   - *d* Estimate of the total storage
   - *e* Quality of stored water
   - *f* Source and amount of any pollutants

2 On the local level:
   - *a* The geologic structure as to its suitability for drilling
   - *b* Effect of pumping water on the surrounding areas
   - *c* Effects of various shapes of well fields as well as that of the time variation in pumping

None of the ground-water models answers all the above questions raised by the above information requirements, and therefore, the use of any one model gives us only partial knowledge about the system. Even if it were possible to choose a model that is best from physical (chemical and thermal) considerations, it may not be possible to use it either because not enough data are available or because the time and money required for its use are not available. In fact, quite often the data and economic considerations may completely dominate the choice of the model. It is in this context that the model choice problem partly arises.[45]

In choosing one model from a set of alternative models, one must consider:

1 The scale of modeling (space and time)
2 The objective of modeling
3 Budget constraints
4 Technical facilities in terms of personnel and inanimates
5 Amount and kind of data available

6 Sensitivity of the project to wrong (or imprecise) answers or the risk involved

It is no easy matter to decide on criteria or measures of effectiveness for comparing models, because these criteria may be scientific, nonscientific, quantitative, or qualititative. Kisiel and Duckstein[45] identify two classes of criteria: cost and effectiveness. The cost includes expenditure on development, acquisition (if proprietary), development and maintenance of data bases, implementation and maintenance (or updating) of a model. The effectiveness of a model can be judged by the variance of and bias in the forecast, error growth, transferability to another site and use, simplicity, sensitivity to variation in inputs and model parameters, computational stability, etc. Keeping one of the two generalized variables (i.e., cost or effectiveness) constant, the models can be compared with respect to the other.

Various ground-water models have been summarized by Sagar,[3] with some qualitative statements about their properties. To some extent, such a table may help in selecting a suitable model qualitatively, but of course, quantitative criteria need to be developed for a better selection. Lovell[46] has worked on this problem from the point of view of watershed models.

4-7 INVERSE PROBLEMS

A more formal presentation of the inverse problems in ground-water hydrology is now given.

Let ϕ, F be some complete metric spaces, and let $A\phi$ be a function with the domain ϕ and the range of values F. Consider the equation

$$A\phi = f \tag{4-34}$$

If its solution is to be investigated for a given A and f, we have the *direct problem* (DP). Operator A which maps ϕ into F may be an algebraic, differential, or integral equation. A set which contains such possible A's is called a set of models $\{M_i\}$. The problems of choosing one element of this set over all other elements for physical, technological, or economic reasons represents the *model choice problem*. In ground-water hydrology, A is generally the differential operator in Eqs. (4-1) to (4-3) and is usually linear. In numerical form, these operators are then algebraic.

The problem in Eq. (4-34) is properly posed in the sense of Hadamard[48] if and only if (iff) the following conditions are satisfied:

1 The solution of Eq. (4-34) exists for any $f \in F$; that is, the problem is not overdetermined (more equations than parameters to be estimated) and extraneous conditions are not imposed.

2 The solution of (4-34) is unique in ϕ. When A is a differential operator

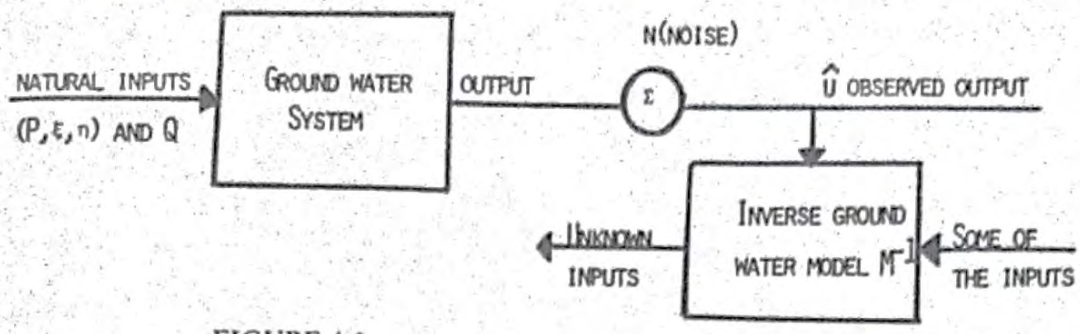


FIGURE 4-6
Schematic representation of the inverse problem.

as in Eqs. (4-1) to (4-3), then the solution is unique if:

a All coefficients in A are perfectly known.

b f is completely known in the sense that an input-output function in f is identified.

c Initial and boundary conditions on ϕ are known.

3 The solution of (4-34) depends continuously on f. In real problems, f is obtained with the aid of measurement instruments and so is known only approximately. Thus, the function Bf can be discontinuous and ϕ cannot be recovered uniquely from Eq. (4-34).

Improperly posed problems in the sense of Hadamard arise if any of the three conditions is not met.

Consider the situation when ϕ, the dependent variable, is completely known, but A and f are only partially identified. Then the problem of completely finding A and f is called the *inverse problem* (IP), or in the vernacular, the *model calibration problem* (also called the system or state identification problem). In effect, all these names refer to making the chosen ground-water model operational. This is the problem of testing the effectiveness of a calibrated model as a predictive tool (see Fig. 4-6). We assume with Sagar[3] that this effectiveness is determined if we can make a probabilistic statement about the uncertainties in the predictions (say, for 1 day or 50 years into the future). We do not say that such a statement has been developed as yet by theoretical and simulation studies.[3]

4-8 CLASSES OF INVERSE PROBLEMS

For ease of exposition, consider Δ as some operator such that the general equation of flow is written as

$$\frac{\partial u}{\partial t} - P\,\Delta u(x,t) + Q = 0 \qquad X \epsilon \Omega \times \Gamma \text{ and } t > 0 \tag{4-35}$$

[e.g., $\Delta = \sum_{i=1}^{n} (\partial^2/\partial x_i^2)(\cdot)$ in the case of Eq. (4-1)], where u is the state or dependent variable [h in the case of Eq. (4-1) and P is the set of parameters (T and S) in Eq. (4-1)]. The parameters may also be considered to be a part of the operator Δ, i.e.,

$$\Delta = \sum_{i=1}^{n} \frac{T}{S} \frac{\partial^2}{\partial x_i^2}(\cdot) \tag{4-36}$$

If an initial condition

$$u(x,0) = \xi(x) \qquad x \in \Omega \tag{4-37}$$

and a boundary condition

$$u(x,t) = \eta(x,t) \qquad x \in \Gamma \text{ and } t > 0 \tag{4-38}$$

are also given, then the solutions of Eqs. (4-35) through (4-38), with suitable hypothesis on ξ and η, uniquely determines $u(x,t)$; $x \in \Omega$, $t > 0$. This determination of the dependent variable is called the *direct problem* in hydrology.

The *inverse problem* arises when $u(x,t)$ has been actually observed (call the observed values of u as $\hat{u}$) for some $X_i \in \Omega$ at $t_i \in (o,T)$ and it is required to solve Eq. (4-34) for:

Inverse problem type I : Parameters $\{P\}$
Inverse problem type II : Initial conditions, i.e., $\{\xi\}$
Inverse problem type III: Boundary conditions, i.e., $\{\eta\}$
Inverse problem type IV: The inputs, i.e., $\{Q\}$
Inverse problem type V : A mixture of the above in the event that decomposition into the above problems is not possible

The name *inverse problem* (IP) seems to be most appropriate because u, the dependent variable in the direct problem, now becomes the independent variable and thus the problem must be solved in reverse. Notice that while Eqs. (4-1), (4-2), and (4-3) are of second order in h, ϕ, and θ, respectively, these same equations are of first order only, when the inverse problem is considered. Both the direct as well as the inverse problems may be of either discrete or continuous type. If we solve the Eqs. (4-35) to (4-38) so that $u(x,t)$ is obtained as a function of x and t such that u and all its existing derivatives can be evaluated at all points $(x,t) \in \Sigma$, then it is a continuous direct problem. The Theis equation (4-17) is an example of this type. This, however, can be done in exceedingly simple cases only. For complicated boundaries and inputs, numerical methods are used which give values of u at a few discrete points $(x_i, t_j) \in \Sigma$ only. This is the discrete form of the DP.

Similarly in the case of the IP, if u is completely known as a continuous function of x and t, then it is a continuous IP. On the other hand if u is known only at few discrete points in Σ, then it is a discrete IP. In almost all field

problems, u would be measured at few discrete points in space and time, and hence the inverse problem would be of the discrete type.

It is obvious that, one way or the other, the inverse problem must be solved, before any model can be successfully employed for its legitimate purpose. One method, and that very often followed, is to use experience, objectivity, and nonnumerical data such as geologic studies to assign values to inverse problem types I through V. A more formal development has been hampered because the inverse problem is improperly posed in the sense of Hadamard.

Although it is apparent that in any practical field situation, in the physical sense, the problem should always be properly posed, yet, unfortunately, in the usual mathematical formulation (given in this section), the inverse problem is inherently ill-posed. This is mainly because of insufficient and inaccurate data, as well as because of the fact that different combinations of P, ϵ, η, and Q may produce the same result thus giving rise to nonuniqueness. Insufficient information makes the system physically underdetermined. Because the data are inexact and the solution ordinarily is nonunique, we speak of optimal estimation which may improve as more and more data are gathered. Thus the history of the system becomes of utmost importance.

Since the solution of the inverse problem greatly depends on the data, the design of the data-gathering network and development of solution procedures are complementary. Hence, an attack on one may indicate a solution for the other.

Equal attention has not been paid to all the five kinds of inverse problems. Type I (i.e., to find the parameters) has been the most widely investigated. A short review of the known methods of solving the inverse problem is provided below. Type I will be discussed first.

4-9 SOLUTIONS TO INVERSE PROBLEMS—A CRITIQUE

Extensive literature is available on the general problem of system identification and this will not be reviewed here (for such a review, see Astrom and Eykhoff[49]) except when pertinent to the ground-water flow problem. The usual method of solving the inverse problem is shown in Fig. 4-7. This method consists of:

1 Assuming certain values for the unknown parameters.

2 Solving the direct problem with this assumed value.

3 Comparing the results obtained in step 2 with the actual observations, and if the two do not correspond within a certain limit, then

4 With the help of a suitable algorithm (adjustment algorithm), changing the values of parameters in step 1

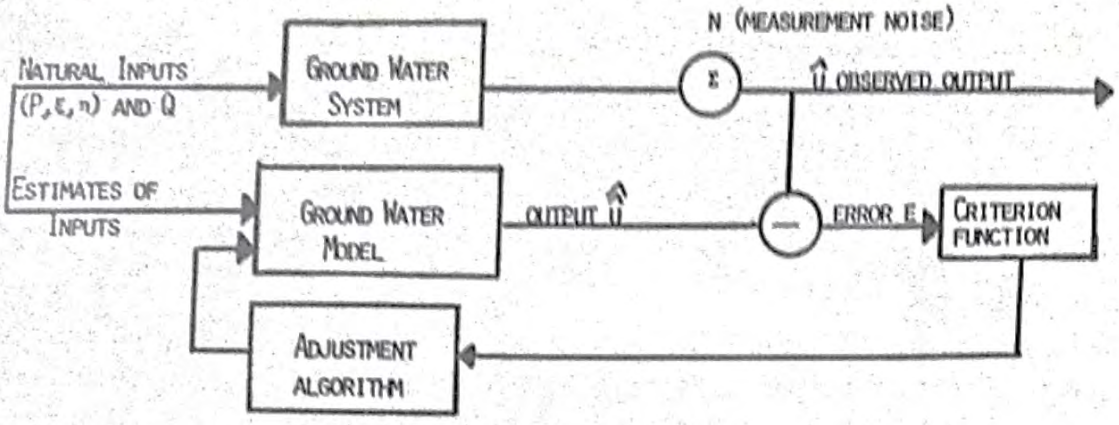


FIGURE 4-7
Usual method of solving the inverse problem.

5 Repeating steps 2 to 4 until satisfactory similarity (using a criterion function) between the observed and the computed values is obtained.

Different adjustment algorithms and criterion functions differentiate one investigator's method from another.

4-10 CRITERION FUNCTIONS

As discussed by Eykhoff,[50] different criteria can be used for optimality; i.e., if b is the true parameter vector and β the estimated one, then our interest may primarily lie in:

1 The minimization of some function of $(\beta - b)$. Since b is unknown, we can only minimize the expectation of this difference if sufficient a priori knowledge is available.

2 The minimization of some function or functional of $e = \hat{u} - \hat{\hat{u}}$ (Fig. 4-8), i.e., the difference between the measured output (including noise) and the model output. This error may be used because e can be made measurable.

3 The minimization of some functional containing the measurable process output and the estimates of the state vector and the parameter vector.

Often the second criterion is used because the correspondence of input-output relations is considered more important than parameter correspondence. This is specifically true if the main aim of modeling is prediction in the future and not the basic understanding of the process itself. Using a least-square criterion the problem then reduces to the variational problem of finding β so that the functional $J(\beta)$ is extremized:

$$J(\beta) = ||\hat{u}(b) - \hat{\hat{u}}(\beta)|| \tag{4-39}$$

Phillipson[51] proves that, if n, the noise, is considered random (i.e., gaussian and additive in nature), then for linear systems, the "maximum likelihood a posteriori estimate" of b is the same obtained for the least mean-square-error criterion in Eq. (4-39).

4-11 ADJUSTMENT ALGORITHMS

A general survey of the search techniques using Eq. (4-39) as the optimality condition is provided by Bekey,[52] with the method of gradients being the most popular in ground-water hydrology. It is based on the algorithm that lets the $(i+1)$st parameter estimate be

$$b^{(i+1)} = b^{(i)} - K \nabla_b J(b^{(i)}) \tag{4-40}$$

where $\nabla_b J$ is the gradient vector. Usually it is desired that the adjustment algorithm converge fast. If possible it should also guarantee that the parameter values to be obtained in any iteration be definitely better than those obtained in the previous iteration. $J(\beta)$ in Eq. (4-39) should be a continually decreasing functional.

In this methodology it is important to note that we are required to solve the PDE of interest repeatedly in step 2. This causes a number of difficulties, such as:

1 Complicated cases may be difficult to solve (such as when anisotropy is considered).
2 When more than one parameter is to be determined, no unique answer is guaranteed.
3 When solving for parameters, initial and boundary conditions and inputs will have to be exactly known, and vice-versa.

The various algorithms available in the literature are discussed below.

4-11.1 Linear Programming

Kleinecke[53] uses linear programming to obtain optimum values of the aquifer parameters. He imposes a rectangular grid on the aquifer and discretizes Eq. (4-6) at any node i to

$$\sum_i K_{ij} D_{ij} W_{ij} \frac{h^k_j - h^k_i}{L_{ij}} + Q^k_i = S_i A_i \frac{d}{dt} h^k_i \qquad \begin{cases} i,j = 1,2,\ldots,m \\ k = 1,2,\ldots,n \end{cases} \tag{4-41}$$

or

$$\sum_i T_{ij}(h^k_j - h^k_i) + Q^k_i = C_i \frac{d}{dt} h^k_i \qquad \text{for all } i, j, \text{ and } k \tag{4-42}$$

where K_{ij} = permeability between nodes i and j
$T_{ij} = P_{ij} D_{ij} W_{ij}$ = transmissivity between nodes i and j

D_{ij} = average depth of flow between nodes i and j
W_{ij} = average width of flow between nodes i and j
L_{ij} = average length of flow between nodes i and j
h_i^k = hydraulic head at node j and at time k
Q_i = pumping or recharge at node i
S_i = storage coefficient at node i
A_i = area of flow associated with node i
$C_i = S_i \cdot A_i$ = storage capacity of node i
m = number of nodes
n = number of time periods for which observations are available

Each node is assumed to be surrounded by a rectangular area. The criterion function preferred by Kleinecke is to minimize Z, where

$$Z = \sum_i \max_k |x_i^k| \tag{4-43}$$

where x_i^k = error at node i for time k. If we call the $\max_k |x_i^k|$ as x_i, then the constraints are

$$\sum_j (h_j^k - h_i^k)\, T_{ij} - \frac{d}{dt} h_i^k C_i + U_i^k - X_i = Q_i^k \tag{4-44}$$

$$\sum_j (h_i^k - h_i^k)\, T_{ij} - \frac{d}{dt} h_i^k C_i - V_k^k \times X_i = -Q_i^k \tag{4-45}$$

where U_i^k and V_i^k are nonnegative slack variables. The linear programming problem then is to find T_{ij}, C_i in Eqs. (4-44) and (4-45) so that Z in Eq. (4-43) is minimized. The total number of constraints would obviously be equal to $2mn$.

The main drawback of this approach is that many of the T_{ij} and C_i do not appear in the optimal solution, which means that these parameters will have to be zero for Z to be minimum. There is no satisfactory explanation for this, although Kleinecke speculates that, with data base spread over longer time, this may not happen. Whatever the length of data, however, there is nothing in the linear programming algorithm that would ensure the presence of all the parameters in the optimal solution.

4-11.2 Method of Multilevel Optimization

Haimes et al.[54] formulate the problem of estimating the aquifer parameters with a nonlinear criterion function and solve it with multilevel optimization. They consider a rather specialized case of a cluster of production wells in an infinite aquifer. The aquifer is divided into N wedge-shaped homogeneous regions each enclosing a single well. A well is supposed to draw water only from the area enclosed in the wedge, and thus the lines delineating the regions act as impervious boundaries for the well. If the total number of wells is N, then val-

ues of transmissivity T_i, storage coefficient S_i, and location of the boundary α_i, $i = 1,2,\ldots,\eta$ are to be found such that Z is minimized:

$$Z = \sum_{i=1}^{N} \sum_{k=1}^{M} (\hat{\hat{h}}_{i,k} - \hat{h}_{i,k})^2 \tag{4-46}$$

where M is the number of pressure observations in a well, and $\hat{\hat{h}}_{i,k}$ is the calculated and $\hat{h}_{i,k}$ the observed pressure heads at well i at time k. Since the boundaries in this case are simple (straight lines), $\hat{\hat{h}}_{i,k}$ is obtained in closed form by the method of images (see Walton[11]) as

$$\hat{\hat{h}}_{i,k} = h_o - \frac{Q_i}{4\pi T_i} W(u_{i,1}) - \frac{Q_i}{4\pi T_i} \sum_{l=2}^{m_i} W(u_{i,2}) \tag{4-47}$$

where $w(u_{i,1})$ and $w(u_{i,2})$ are well functions as defined in Eq. (4-17)

$u_{i,1} = r_i^2 S_i / 4T_i t_i$, where r_i is the radius of ith well

$u_{i,2} = a_{i,l}^2 S_i / 4T_i t_i$, where a_i is the distance of the lth image from the production well

m_i = number of images considered of the ith well

t = time

h_o = the initial head

Equation (4-47) can be written for $i = 1,2,\ldots,N$ and $k = 1,2,\ldots,M$, and hence there are $M \times N$ such equations. Besides these, the other constraints are, for $i = 1,2,\ldots,N$:

$$\begin{aligned} T_i &> 0 \\ S_i &> 0 \\ \theta_i < \alpha_i &< \theta_{i+1} \qquad \text{(see Fig. 4-10)} \\ \alpha_i - \alpha_{i-1} + 2\pi\delta_{i,j} &= \frac{2\pi}{m_i} \end{aligned} \tag{4-48}$$

where $\delta_{i,j}$ is the Kronecker delta. Pseudovariables σ_i, $\sigma_i = \alpha_{i-1}$, $i = 1,2,\ldots,N$, were introduced so that by considering σ_i as a parameter in the ith subsystem, the overall system was effectively decomposed into N independent subsystems. Optimal values for T_i, S_i, and α_i were now determined for each subsystem for a fixed σ_i (first-level optimization) by a direct search technique and the optimal values of σ_i were then calculated (second-level optimization) by a Gauss-Seidel-type algorithm.

It is obvious that the problem considered by Haimes et al. is somewhat artificial. It will be hard to find a cluster of fully penetrating production wells in an infinite aquifer all pumping at the same rate. Also, the division of the aquifer into wedge-shaped subregions is arbitrary and does not correspond to real-world situations. Any one wedge may be heterogeneous in itself, and areas contributing water to wells may be overlapping so that the existence of a dividing line has no basis.

4-11.3 Hybrid Computations

Vemuri and Karplus[55] look upon the problem of parameter estimation of an unconfined aquifer as a control problem in distributed parameter systems and use a hybrid computer to solve it. For a heterogeneous isotropic unconfined aquifer, the flow equation in two space dimensions is

$$\frac{\partial}{\partial x}\left[T(x,y,h)\frac{\partial h}{\partial x}\right]+\left[\frac{\partial}{\partial y}T(x,y,h)\frac{\partial h}{\partial y}\right] + Q(x,y,t)=S(x,y,h)\frac{\partial h}{\partial t} \qquad (4\text{-}49)$$

The problem then is to compute in a region R, the values of T, S, and the boundary of R (say, ∂R) such that the functional J is minimized:

$$J=\int_{t_i}^{t_{i+1}}\int_R [\hat{\hat{h}}(x,y,t)-\hat{h}(x,y,t)]^2\,dR\,dt \qquad (4\text{-}50)$$

in which $\hat{\hat{h}}$ and $\hat{h}$ are defined before, and the time interval of interest $[o,T]$ has been divided into intervals t_i, $i=1,2,\ldots,N$, each interval small enough so that $T(x,y,h)$ in Eq. (4-49) may be considered independent of h.

The identification was achieved by first assuming a nominal shape of ∂R and a nominal set of values to S. The only unknown parameter in Eq. (4-49) is then T, which is obtained by minimizing J in Eq. (4-50). The minimization of J in Eq. (4-50), however, is equivalent to extremizing (by the so-called maximum principle in control theory: see Sage[56]) a suitably defined scalar-valued hamiltonian in Hilbert space. Vermuri et al.[57] derived this hamiltonian, along with a set of necessary conditions in the form of a pair of PDEs, to minimize (or maximize) it, which in turn determined optimum estimates of transmissivity. This pair of PDEs was solved on a hybrid computer and a steepest gradient method was employed to search for the extremum of the hamiltonian. Changes in S are accomplished by changing a suitable number in the computer memory via the console. If a change in the boundary geometry is warranted, it was accomplished by changing a few wires on the analog patch board of the hybrid computer.

This method effectively computes only one parameter, i.e., T, the other two, i.e., S and ∂R, being assigned values subjectively by the investigator. There is no indication of the sensitivity of the computed values of T to changes in S and ∂R, but if the sensitivity is high, this may create problems. The authors advocate the use of a hybrid computer because the solution of a differential equation (two in their method) on a digital computer is expensive; but the construction of an analog model of an aquifer required to implement their method may not be easy either.

Vemuri et al.[57] also suggest the use of "sensitivity analysis" as a method for parameter identification. This procedure, however, has not yet been applied to aquifers.

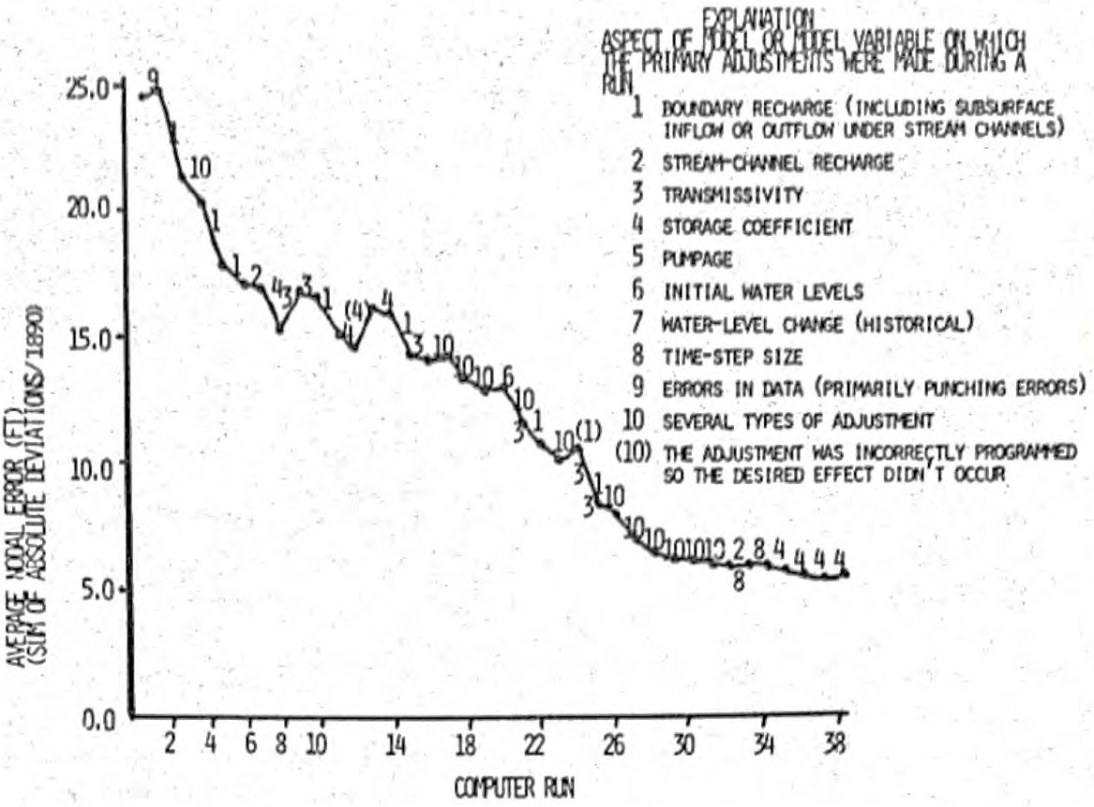


FIGURE 4-8
Average nodal error of successive computer runs during calibration of the Tucson basin model.[2]

4-11.4 Method of Subjective Calibration

As used by Gates[2] to calibrate the Tucson basin digital model, this method is effectively a trial-and-error procedure. A suitable numerical scheme is set up to solve Eq. (4-6). The unknown parameters and the recharge and discharge are now assigned values subjectively, and Eq. (4-6) is solved to obtain $\hat{h}$ at grid points. This calculated $\hat{h}$ is now compared to the observed $\hat{h}$ at the same points. Any of the unknown parameters are now changed subjectively (without any automated algorithm) to obtain a better fit between $\hat{h}$ and $\hat{h}$ at as many grid points as possible

This method is perhaps the simplest and therefore is preferred by field hydrologists who have a considerable knowledge of the aquifer with which they are working. Notice, however, that the fit between h and $\hat{h}$ after any iteration may not always be better than the previous iteration. There is no fixed criterion function, and therefore even the same investigator will not be sure of obtaining the same answers twice in an aquifer. The computer time involved will also depend upon the skill and the knowledge of the individual performing the calibration.

Illustrative of the results obtained by subjective calibration is that given in Table 4-3 and Fig. 4-8. These results, for the unconfined Tucson aquifer

on an annual basis, are in terms of a 1,890-node model with ¼-square-mile nodal areas and a 509-node model with 1-square-mile nodal areas. Average nodal errors have been reduced to about 5.3 ft (either too high or too low in relation to observed water levels) after about 38 iterations on the computer. Note the substantial reduction in maximum error and in number of nodes with large errors. The computer can print out a map of the errors and their signs so as to give a quick view of those parts of the aquifer where the calibration fit is poorest.

4-11.5 Method of Collective Adjustment by Subjective Programming

A definite improvement over the subjective calibration described above, this method was developed by Lovell et al.[58] and tested on a digital model of a part of the Tucson aquifer. A considerable amount of subjective information about the aquifer is still required in this method, but this is now quantified and used in a logical way.

A regular square grid is imposed on the aquifer, and Eq. (4-6) is discretized on this grid. The coherence of the data is now tested by evaluating two indices: imbalance and trouble index at each node. "Imbalance" was defined as the amount of water in feet per year at any node, not accounted for by the continuity equation, and "trouble index" was defined as a number assigned to each node in order to give an indication of the amount of adjustment needed at each node to remove the imbalance. Such information was also used to evaluate the quality of the data entering into the modeling process. In addition, subjective "confidence bounds" were placed on data at each node and "sensitivity" of water balance at a node to change in each parameter value evaluated. The trouble index, the confidence bounds, and sensitivity then guided the determination of a parameter perturbation coefficient. By an iterative procedure, the parameters are adjusted until the desired improvement in model performance is obtained.

TABLE 4-3 COMPARISON OF INITIAL AND FINAL ERRORS IN THE LARGE-SCALE AND SMALL-SCALE MODELS OF THE TUCSON AQUIFER. (*From Gates and Kisiel*[2])

	Average nodal error, ft	% nodes with errors < 10 ft	% nodes with errors < 20 ft	% nodes with errors < 30 ft	% nodes with errors < 40 ft	Maximum error, ft
Large-scale model (1,890 nodes)						
Initial run	24.4	26	48	65	82	190.6
Final run	5.3	87	99.5	100	···	28.6
Small-scale model (509 nodes)						
Initial run	6.2	80	89	99.5	100	34.4
Final run	5.7	82	99.5	100	···	24.7

In Lovell's study of the 512-node model of the Tucson aquifer, imputed value tests were used to make a node-by-node analysis of data consistency. Subjected to these were the data on pumpage-recharge, water levels, transmissivity, and storage coefficients. One of these four was computed (or imputed) by assuming all others to be correct. The discretized form of the heat diffusion equation, commonly used to describe the behavior of isotropic unconfined aquifers, was used to compute the imputed values.

For the 19-year data set (1947–1966) all results reflect serious inconsistencies within the 512-node model:

1 *Storage coefficient:* 357 nodes had values that were clearly infeasible (either negative or in excess of 1.0). Only 89 were within the likely range of 0.0 to 0.25.

2 *Transmissivity:* at least 180 had physically impossible negative values. Only 89 had values falling between 50 and 200 percent of the originally assigned values.

3 *Net pumping/recharge (QR):* over 200 of the nodes showed sign changes between net outflow and net inflow. Fewer than 60 of the nodes had imputed values of QR in the range of 50 to 200 percent of the originally assigned values. These results reflect in part a major deficiency in knowledge of mountain-front recharge.

The earlier mentioned subjective-objective procedure provides a basis for getting weighting factors for the parameter estimation step. Brute force optimization techniques are discouraged and a modified optimization (using subjective knowledge) is used to ameliorate the inconsistencies. It should be noted that it is difficult to discern whether the errors are to be assigned to the data, to the model, or to the imputed value test. Research is needed on these issues.

4-11.6 Automatic Solution

Emsellem and de Marsily[59] developed a method which they called "An Automatic Solution for the Inverse Problem" and applied it to the determination of T for *steady-state* flow in heterogeneous isotropic aquifers. T is determined as an average value over an area; the area defining the scale of T is determined in the process. They start out by assuming the entire area as homogeneous and find a value of T such that the norm of the error (or the selected criterion function) in computed flow is a minimum. Subsequently the area is subdivided into smaller portions and, with the same criterion function, the best T value is calculated for each subarea. This process is terminated when no further improvement in the criterion function takes place. The major point of their analysis is their choice of criterion function, which, they claim, makes the answer to the inverse problem unique. Nonetheless, their method was developed only for the steady-state case.

4-11.7 Method of Energy Dissipation

Nelson[60] developed an energy dissipation method to evaluate K for a heterogeneous, isotropic, and now compressible medium (so that $S = 0$ for this medium) for which the PDE is

$$K\nabla^2 h + \frac{\partial K}{\partial x}\frac{\partial h}{\partial x} + \frac{\partial K}{\partial y}\frac{\partial h}{\partial y} + \frac{\partial K}{\partial z}\frac{\partial h}{\partial z} = 0 \tag{4-51}$$

This is the equation of flow in the eulerian framework. The lagrangian equation for the same case in terms of stream function Ψ is

$$\frac{D\Psi}{Dt} = \frac{\partial \Psi}{\partial x}\frac{dx}{dt} + \frac{\partial \Psi}{\partial y}\frac{dy}{dt} + \frac{\partial \Psi}{\partial z} + \frac{\partial \Psi}{\partial t} = 0 \tag{4-52}$$

These two equations are then used to devise a PDE in terms of the unknown K:

$$\frac{d(\ln K)}{dt} = \frac{-\nabla^2 h}{|\nabla h|^2}\left(\frac{Dh}{dt} - \frac{\partial h}{\partial t}\right) \tag{4-53}$$

where Dh/dt represents the total derivative of h, that is,

$$\frac{Dh}{Dt} = \frac{\partial h}{\partial x}\frac{dx}{dt} + \frac{\partial h}{\partial y}\frac{dy}{dt} + \frac{\partial h}{\partial t}\frac{dz}{dt} + \frac{\partial h}{\partial t} \tag{4-54}$$

Equation (4-53) being in the lagrangian frame, t in this equation is a parameter that divides the location in the stationary eulerian x, y, t coordinates. Equation (4-54) is a first-order PDE in K and can be solved for K, provided boundary values on K are supplied; i.e., once the value of K is known at one point on a streamline, its value at all other points can be computed. Nelson tested his method on both the steady and the unsteady flow. The main drawback of the method of course is the assumption of zero storage coefficient.

4-11.8 Related Work

Chen and Seinfeld[61] discuss the estimation of spatially varying parameters in partial differential equations. Depending upon whether the unknown parameter vector is considered as a control variable or a state variable, they formulate the problem in the context of control theory or state estimation (or filtering) theory, respectively. As the boundary conditions become more complex and the number of unknown parameters increases, both formulations become very difficult computationally.

Cannon[62] solves a different kind of parameter identification problem. He considers heat flow in a homogeneous isotropic rod of unknown diffusivity. He assumes that no measurements on temperature were taken but that an additional (besides the usual initial and boundary conditions) condition on heat flow such as heat input at one boundary at a given time was known. Under these conditions, Cannon proves that diffusivity can be obtained uniquely. Under

the same conditions Jones[63] also gives a method to solve for diffusivity when it is a function of time. The methods depend upon a closed-form solution of the flow equation and therefore cannot be applied to a regional scale which is of intent here.

Yeh and Tauxe[64] apply the quasi-linearization method as developed by Bellman and Kalaba[65] to a rather idealized and local ground-water pumping problem. This method may be an alternative to the graphical-type-curve method but is not applicable to the regional flow problem.

Table 4-4 summarizes the various methods discussed so far.

4-12 INVERSE PROBLEM TYPES II TO V

Comparatively very little has been done on the remaining types of inverse problems. Specifically, no work has been done on them in the particular area of ground-water hydrology. The general work in this area that has come to the attention of these authors is as follows:

4-12.1 Work on Inverse Problem Type II

This problem is discussed by Lattes and Lions.[66] In context of Eq. (4-35) defining Δ as in Eq. (4-36), the problem is to determine the initial condition $u(x,o) = \xi(x)$; $x \in \Omega$, given $u(x,T) = \chi(x,T)$ at $T > 0$ and appropriate boundary conditions.

It is important to notice in this problem that Eq. (4-35) is irreversible in time, i.e., cannot ordinarily be solved backward in time. Lattes and Lions solve the problem by a method which they called the method of quasi-reversibility. Basically the method consists of choosing a new operator O_ϵ "near" the original operator $[\partial/\partial t - \Delta]$ in Eq. (4-35) for which Eq. (4-35) can be integrated backward in time, i.e., for which the problem becomes reversible (hence the name of the method). O_ϵ would be considered "near" the original operator, if the functional $J(\xi) \leq \epsilon$, that is,

$$J(\xi) = \int_\Omega |u(x,\mathrm{T};\xi) - \chi(x,T)|^2 \, dx \leq \epsilon \qquad \text{with } \inf_\xi J(\xi) = 0 \qquad (4\text{-}55)$$

In Eq. (4-55) $u(x,T;\xi)$ is the solution of Eq. (4-35) at time T for an initial condition ξ, and $\chi(x,T)$ is the given solution at the same time.

Lattes and Lions show that many possible O_ϵ's are available and that the solution to this inverse problem is unique only with respect to the O_ϵ selected. Usually that O_ϵ is selected which leads to the simplest calculations. For Eq. (4-35), Lattes and Lions for example show that with O_ϵ taken as

$$O_\epsilon = \frac{\partial u_\epsilon}{\partial t} - \Delta u_\epsilon - \epsilon \Delta^2 u_\epsilon \qquad \epsilon > 0 \qquad (4\text{-}56)$$

Table 4-4 TABULATION OF KNOWN METHODS FOR SOLVING THE INVERSE PROBLEM

Author(s) (see references)	Name of method	General assumptions made	Results and comments
Kleinecke[53]	Linear programming	1. Linear objective function (sum of max. error at each node) 2. Isotropic medium with two-dimensional flow, T and S independent of time	1. No guidelines on subdivision of the basin into a grid 2. No way to force all the unknowns into the optimum solution
Yeh and Tauxe[64]	Quasi-linearization	1. Infinite, homogeneous, and isotropic aquifer 2. Drawdown data on a constant discharge well is required 3. A good initial estimate of the parameters for the method to converge	1. Proposed as an alternative to the type curve method but not applicable to leaky aquifers 2. Not applicable to a regional identification of the parameter 3. If initial estimates are bad, the method may not converge
Vemuri and Karplus[55]	Hybrid computations	1. Unconfined aquifer 2. Initial estimate of the values of the parameters required 3. Sufficient subjective knowledge of S and boundaries needed	1. No guarantee of a global minimum of the objective function 2. Two of the three parameters are arbitrarily adjusted to obtain a best fit in the T parameter. Thus unicity not guaranteed 3. Hybrid computer required

Vemuri et al.[57]	Sensitivity analysis (maximum gradient)	Scalar problem, i.e., only lumped estimates are to be obtained	1. Not applied to the partial differential equation of ground-water flow 2. Lots of computational difficulties
Emsellem and de Marsily[59]	Automatic solution of inverse problem	1. Two-dimensional isotropic medium 2. T and S are continuous functions of space 3. T and S are independent of time	1. Scale of determination of S and T also determined 2. No initial estimate required
Haimes et al.[54]	Decomposition and multilevel optimization	The total region can be divided into wedge-shaped homogeneous isotropic portions, each enclosing a single production well	Initial estimates of S and T are required
Gates[2]	Subjective trial and error	1. Enough subjective knowledge is available 2. Valid for confined and unconfined aquifers	Solely based on the investigators experienced and subjective knowledge; there is no specific algorithm to adjust the parameter values
Lovell et al.[58]	Subjective optimization	Subjective knowledge can be quantified by defining certain notions	1. Not posed as an optimization problem 2. Certain parameters quantifying subjective knowledge have to be defined
Nelson[60]	Method of energy dissipation	1. Two-dimensional isotropic medium 2. Storage coefficient is zero 3. Permeability value is available at every streamline	1. It will be hard to treat anisotropic aquifers 2. Assuming storage coefficient to be zero makes it inapplicable to the usual ground-water flow models

the inverse problem type II is properly posed and, for $\epsilon \to 0$, the results obtained are quite close to the desired ones [i.e., $J(\xi)$ in Eq. (4-55) becomes small].

4-12.2 Work on Inverse Problem Type III

This inverse problem has been discussed by Phillipson[51] in a special context. He assumes that inexact measurements on initial and boundary conditions are given and considers the problem of finding the "time" estimates of the initial and boundary conditions which are optimal in some sense. All three kinds, i.e., Dirichlet, Neumann, and mixed-boundary conditions, were considered. With reference to Eq. (4-35) with Δ as in Eq. (4-36) and with Dirichlet boundary conditions, the problem is defined as follows:

$$\begin{aligned} &\frac{\partial u}{\partial t} - \Delta u = Q \qquad \text{in } \Omega \times [0,T) \\ &\text{Boundary condition:} \quad u(\Sigma) = v_1(\Sigma) \qquad \text{on } \Sigma \\ &\text{Initial condition:} \quad u(x,0) = v_2(x) \qquad \text{in } \Omega \end{aligned} \tag{4-57}$$

Obtain the best estimates of $v_1(\Sigma)$ and $v_2(x)$ when inexact measurements on u are given in $\Omega \times [0,T)$, on Σ, and in Ω as $\hat{u}$, $\hat{u}_1$, and $\hat{u}_2$, respectively. Phillipson[51] formulated the problem in the least-square sense; i.e., if $u(x,t;V)$ is a solution of Eq. (4-57), then obtain $v_1(\Sigma)$ and $v_2(x)$ so as to minimize $J(V)$:

$$\begin{aligned} J(V) = &\int [u(x,t;V) - \hat{u}(x,t)]^2 dx\, dt \\ &+ \int_{\Omega \times [0,T]} [v_1(\Sigma) - \hat{u}_1(\Sigma)]^2 d\Sigma \\ &+ \int_{\Omega} [v_2(x) - \hat{u}_2(x)]^2 dx \end{aligned} \tag{4-58}$$

Phillipson proved that a unique solution to this problem exists and devised an iterative method to obtain the solution.

4-12.3 Work on Inverse Problem Type IV

Moench and Kisiel[26] invert the convolution relation in discrete form to obtain an input recharge function from a flood flow in an ephemeral channel (see Sec. 4-4). A check on the estimated function is not possible, because there are no available discharge hydrographs in the area.

Changes in ground-water levels can lead to estimates of recharge, but their credibility still requires corroboration, in terms of error analysis.

Isotope hydrology[67] has the potential of helping in the resolution of this problem, but in our judgment the methodology must be developed in a more formal space-time sampling framework.

Inverse problem type V is not yet found in the literature and hence is not discussed here.

4-13 APPROACHES TO MANAGEMENT OF THE SYSTEM: INTRODUCTION

In Secs. 4-13 to 4-18 we examine successively the following classes of problems:

1 Extraction of water from an aquifer
2 Natural and artificial recharge into an aquifer
3 Chemicals and pollutants in an aquifer
4 Conjunctive operations of surface- and ground-water resources
5 A combination of the above problems, which often constitutes the real situation

For each problem we attempt to identify management goals. In the cases in which the goals can be collapsed in a single objective, say (benefit − cost), a standard operations research model can be used. For example, we can position a certain number of wells in a well field so as to minimize total capital and operation cost, subject to a given demand to be satisfied and to the equations of motion of water both under and above ground, i.e., in the aquifer and in the pipelines.

Hotelling[68] has done pioneering work on optimum exploration of depletable resources such as metal ore, using calculus of variations. Clearly, if the goal is to exploit a ground-water resource until exhaustion, as considered by Kelso[69] and Cummings,[70] then such a calculus method can be used. Dupnick[71] has considered adverse effect, or externalities, which may be caused by ground-water resource depletion; he distinguishes two kinds of externalities:

1 Horizontal externalities, of a spatial nature (pumping decreases water supply downstream)
2 Vertical externalities, of a temporal nature (rights of people in the future are neglected)

In order to control these externalities and perhaps compensate losses, Dupnick proposes taxes on each kind of externality, which is a form of pricing policy.

Seldom is such a single objective or criterion realistic. A water utility company, for example, seeks to provide reliable service first, political considerations may come next, and then cost. A single criterion to represent service is difficult to find; thus when de Neufville[72] attempted to define a proxy variable (pressure at a given point of the network) to represent quality of water service in New York City, he and his coworkers were criticized by city engineers for their arbitrary choice.[73]

We are thus led to multiobjective management models, preferably where the multiple objectives are not immediately converted into a single-dimensional function by a priori weighting. We are thus excluding national cost-benefit analysis and recommend the use of a cost-effectiveness approach,[74] where no weighting is necessary, or the use of an interactive multiobjective programming approach,[16] where weighting is introduced in a dynamic but subjective manner.

The basic premise of a multiobjective approach is that maximizing or minimizing is not always a valid goal in real-world problems. Several years ago, Dorfman[75] emphasized that economic decision makers satisfice rather than optimize systems. In other words, a compromise is sought between a perfectly reliable (in terms of quantity and quality) water supply network at an enormous cost and a poor supply system at a low cost. The multiple objectives may be, in this example:

1 Average pressure in this system
2 Minimum pressure in this system
3 Probability of service interruption
4 Quality of water (a vector itself!)
5 Emergency capabilities (fire protection)

Note that the number of goals in a decision situation should usually be less than seven[76] because of human limitations. The number of decision variables may be considerably larger; it appears, however, that the human mind cannot grasp more than seven—at most 10—objectives for a given problem.

The multiobjective nature of decision-making in water resources is gradually being recognized by the leading scientific groups both within and outside of the United States. However, so far, only a few methods are available;[77] furthermore, to our knowledge, none of these methods has been applied to a ground-water management example. For the sequel, we shall identify problem areas warranting such a multiobjective approach and illustrate the methodology by means of examples.

Another striking feature of available literature on management of ground-water systems is the almost universal use of deterministic models, as we had remarked earlier in the sections on system description (Secs. 4-2 to 4-6). Since one of the purposes of this chapter is to point out future research needs, we shall carefully point out the various types of uncertainties that are encountered in ground-water management models that suggest methods to account for those uncertainties.

Each class of problems is subdivided into the following elements:

1 Management system description
2 Multiobjective features
3 Uncertainty aspects

The first class of problems examined corresponds to the most common use of an aquifer, namely, water supply.

4-14 EXTRACTION OF WATER FROM AN AQUIFER

4-14.1 Management System Description

Water is pumped out of an aquifer in a given region or basin in order to satisfy, fully or in part, the demand by various categories of users: domestic, commercial, industrial, agricultural, or public (recreation, navigation, fishing).

If recharge and demand are assumed to be known, the problem usually consists of one or more of the following elements:

1 To design a well field such that water can be extracted at a sufficiently high rate to satisfy a design peak demand, at an acceptable cost

2 To trade off, in this design, well size versus storage and transport facilities to satisfy demand

3 To design a minimum cost schedule that still ensures a certain quality of service or reliability

4 To minimize interaction between wells, land subsidence, salt-water intrusion, and other water-quality problems

5 To compute a withdrawal schedule that ensures optimum economic growth,[69,78] i.e., calculating of an optimum mining policy of the water resource

6 To ensure safe yield

7 To ensure alternative yield concept[79]

The constraints in each of these problems are the equations of motion of water, salts, and heat through process media. These constraints may appear under the form of a simulation model of the aquifer based on the partial differential equations of motion, [3,9] a finite element model,[37] a finite-state-machine model,[32,33] or a completely lumped model.[9,78]

Note that, as said before, the finite-state-machine model may be used to represent any simulation model whether distributed, lumped or in between. In every case the goal of the system is represented by the output, namely, a function defined on the state of the system. As we will show in the next section, this output may well be a vector.

4-14.2 Multiobjective Features

It is sufficient to choose two or more of the elements described in Sec. 4-14.1 to obtain a multicriterion decision problem. For example, a real-world problem may seek the maximizing of net benefits stemming from the allocation of ground water while minimizing quality problems and land subsidence. One way that many authors have avoided this problem is by entering all objectives but one into the calculation as a constraint.[80,81,82] Here we would set a quality standard and an acceptable land subsidence as constraints, and maximize net benefits. Still other authors are avoiding the problem by considering one use for the ground water, especially irrigation.[6,7] Finally, even economists who are

well aware of the difference between maximizing net income and ensuring efficient income distribution end up lumping both objectives into one benefit function.[69,83]

The above remarks are not meant to negate the usefulness of classical single-criterion operations research models for the allocation of ground water; an excellent review of these models is given by Domenico,[9] and so we shall not repeat that material here. Our point is that applied research needs in ground-water management in the next few years will most probably include the development and implementation of multiobjective decision models: under this hypothesis, is it worthwhile to invest considerable effort in a model that can look at only one aspect of the problem? Furthermore, it should be kept in mind that the evaluation of a computer simulation model involves much more than considering central processor time.[2,39] The choice of an aquifer model for management study must itself be considered in a broader framework than computer time: data requirement, programming time, and level of analyst competence necessary to view the model are a few of the additional considerations that should enter into the choice.[45,84] See Sagar[3] on these issues as they enter the model-choice problem.

4-14.3 Uncertainty Aspects

As noticed before, most ground-water models are deterministic; the question is then: can they handle uncertainty at all? If an aquifer is used for water supply purposes, how can we account for the variability of storage and transmissivity coefficients and for the errors in the initial or boundary conditions in the allocation scheme? Another element here is the stochastic nature of recharge, which will be discussed separately.

Next, it is quite common to assume that water demand is given. Clearly, it is an oversimplification of reality and a dangerous one. Let us examine why for various categories of uses:

1 *Agricultural uses*: irrigation demand Q depends on the weather [precipitation $\mathbf{R}(t)$ and temperature $\mathbf{T}(t)$ during a given time period].[85,86] In a manner akin to bayesian decision theory,[87] we can define a profit function $Z(Q,\mathbf{R},\mathbf{T})$. The boldface $\mathbf{R}$ and $\mathbf{T}$ are random functions. Two typical profit functions are shown in Fig. 4-9 for two pairs of values $(\mathbf{R}_1,\mathbf{T}_1)$ and $(\mathbf{R}_2,\mathbf{T}_2)$. Note that $\mathbf{R}=\mathbf{R}(t)$ and $\mathbf{T}=\mathbf{T}(t)$ are functions of time; however, one may choose proxy variables to represent these time functions—for example the value of $\mathbf{R}$ and $\mathbf{T}$ during a critical growing phase of the plants. A deterministic model assumes that the irrigation demand is Q_E^*, which corresponds to $Z[Q_E^*|E(\mathbf{R}),E(\mathbf{T})]=$ maximum. But many other values of $\mathbf{R}$ and $\mathbf{T}$ are possible besides the expected values $E(\mathbf{R})$ and $E(\mathbf{T})$ and the cost of overirrigating is generally not the same as that of underirrigating (depending on the type of crops and the type of soils). If the joint distribution of rainfall and temperature $f_{R,T}(r,t)$ is known, then the optimum de-

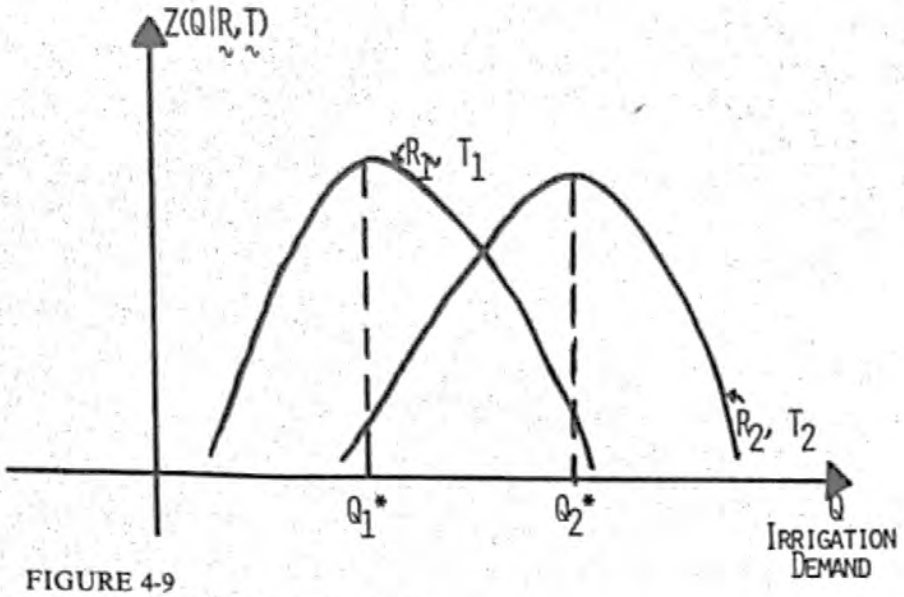


FIGURE 4-9
Profit functions for an irrigation system.

mand Q^* should be the one that maximizes the expected profit:

$$Z(Q^*) = \max_{Q} \int \int Z(Q,r,t) f_{R,T}(r,t)\, dr\, dt \qquad (4\text{-}59)$$

We can also envision a maximin strategy, where we seek to *max*imize the *min*imum profit, as a large land corporation would try to do.

2 *Industrial and domestic uses*: the quantity of concern here is the yearly peak demand, which conditions the design of water supply facilities. Short-term forecasts (few years) are usually adequate. However, long-term demand depends on population and technology, the forecast of which is quite difficult. Also, demand varies as a function of pricing of policy:[88,89,90] it is quite important to consider pricing or taxing as a control variable in this demand for water. Maybe setting the price of water or the value of a pump tax should be one of the objectives of ground-water management.

Then, the response to a given pricing structure is stochastic: such an uncertainty could be incorporated into the loss or profit function as we have done for irrigation demand.

4-15 NATURAL AND ARTIFICIAL RECHARGE INTO AN AQUIFER

4-15.1 Management Problem Description

Recharge into a large aquifer is not easy to measure. Yet safe or alternate yield values must be decided in face of this uncertainty.

Assuming that the distribution of the amount of natural replenishment of

an aquifer is known, what can be done to increase it? Conversely, how much recharge will be lost if a riverbed is lined with concrete? The U.S. Army Corps of Engineers has had to evaluate such losses when planning flood-control projects.

In semiarid countries, for intermittent flows, natural recharge is an increasing function of flow duration, especially if recharge occurs only along the riverbed;[91] small retention dams, which retard the flow, may provide such a benefit.

Another method to increase recharge may be to drill a series of wells along the river shore, or bank, so as to power the water table at the place when recharge may occur. Such a problem has been described in terms of costs of Type I and Type II errors,[92] upon which we shall dwell just below.

Disposal of toxic water in deep wells, especially of radioactive wastes is another problem that has received attention in the past few years. Artificial recharge may be obtained through spreading or injection—which method should be chosen to attain the set goals?

4-15.2 Multiobjective Features

Whenever an action is considered that may change natural recharge or, more generally, the course of nature, at least two objectives should be considered:

1 The consequences of a Type 1 error, which corresponds to rejection of a null hypothesis when it is true

2 The consequence of a Type II error, which occurs when one accepts the null hypothesis which turns out to be false

An example of a null hypothesis is that our action (wells along the river, retention dams) increases recharge by 50 percent. If all the consequences of our errors can be measured in monetary terms, then we have reduced our problem to a single-criterion-type problem as in ref. 92. Thus, if we fail to take an action that would indeed increase recharge by 50 percent, this may either cause a faster decrease of the ground-water level or require the purchase of water elsewhere; the cost of the second solution may serve as a monetary measure of the consequence of the Type I error made.

However, such a monetary evaluation is not always possible. If retention dams and drilled wells are installed to no avail (from the goal viewpoint), the damage may be more than financial: the environment or ecology may have suffered irreversibly. It is very often the case that the consequences of a Type II error, which corresponds to consumer wish (the public), contain nonquantifiable elements. This is perhaps why standard statistical tests are mainly based on the value of α—the probability of Type I error, while the probability of Type II error β is computed afterward, as power of the test $(1 - \beta)$.

Other objectives in aquifer recharge problems may be

1 Minimizing pollution from agricultural, industrial, or municipal effluents
2 Stopping salt-water intrusion
3 Avoiding land subsidence
4 Eliminating lawsuits because of well interference

4-15.3 Uncertainty Aspects

Uncertainty has already been introduced in Sec. 4-14.2 under the form of probabilities α, β of Type I and Type II errors, respectively. This uncertainty originates in our prior knowledge of infiltration parameters, not the process model. But before more data on the infiltration mechanism are taken, the question should be asked whether or not such data would be worth the effort and expense. More data sharpens the PDF (probability density function) of water amount recharged and hence reduces α and β: but how much is a reduction worth in terms of the measure of effectiveness (money, environmental impact, conservation of a scarce resource)? In recent literature[93] an error of Type III has been defined, which corresponds to solving the wrong decision problem.

Returning to natural recharge caused by intermittent flows, there is no point in attempting to estimate a mean annual recharge with the data record length in hand: the confidence band would just be meaninglessly wide. This is a characteristic of data in semiarid lands.[91] Building a stochastic model is suggested as an excellent method to obtain the PDF of recharge with limited data: it is a case of replacing data with hypothesis and a model, which the writers have been advocating for a certain time.[31]

Under intermittent flow conditions, recharge along the riverbed is best modeled as an event-based stochastic process. Clearly the opportunity for recharge exists only when there is flow, and the amount of recharge may be assumed proportional to the flow duration. Thus, for the jth flow event,

$$\begin{aligned} \mathbf{V}(j) &= a(\mathbf{D}(j) - b) \qquad \text{if } \mathbf{D}(j) > b \\ &= 0 \qquad\qquad\quad \text{if } \mathbf{D}(j) \le b \end{aligned} \tag{4-60}$$

(i.e., recharge occurs only if flows last longer than b time unit)

where $\mathbf{V}(j)$ = volume of recharge
$\mathbf{D}(j)$ = duration of flow
a,b = constants depending upon the physical and hydrologic characteristics of the riverbed, which may be estimated from elements given by Harshbarger[10]

Assume that, for a given season, the $\mathbf{D}(j)$ are independent identically distributed random variables and that there are N events per season, $j = 1,2,\ldots,N$.

In order to calculate the seasonal recharge volume **W**, we need to sum a random number of random variables:

$$\mathbf{W} = \mathbf{V}(1) + \mathbf{V}(2) + \cdots + \mathbf{V}(\mathbf{N}) \tag{4-61}$$

If the mean and variance of D and N are known or estimated from data, the mean and variance of the seasonal recharge volume can be computed:[94]

$$\begin{aligned} E(\mathbf{W}) &= E(\mathbf{V})\, E(\mathbf{N}) \\ &= a(E(\mathbf{D}) - b) \cdot E(\mathbf{N}) \end{aligned} \tag{4-62}$$

$$\begin{aligned} \mathrm{VAR}(\mathbf{W}) &= E(\mathbf{N})\mathrm{VAR}(\mathbf{V}) + [E(V)]^2\mathrm{VAR}(\mathbf{N}) \\ &= a^2 E(\mathbf{N})\mathrm{VAR}(\mathbf{D}) + a^2[E(\mathbf{D}) - b]^2\mathrm{VAR}(\mathbf{N}) \end{aligned} \tag{4-63}$$

Furthermore, the distribution of W can be obtained, if that of **D** and **N** are known. The probability density function of **D** may be obtained from empirical data or from a watershed model or from simulation.[95] For example, Kisiel et al.[91] found that, for a small river in Arizona, the Tucson Arroyo, the PDF of **D** could be taken as negative binomial, with $p = 0.56$, $r = 2$ in summer, and $p = 0.65$, $r = 2$ in winter, where p is the probability of a runoff event and r is the other parameter of the negative binomial distribution.

The negative binomial distribution is written[96]

$$f_{\mathbf{D}}(k) = \binom{r + k - 1}{k} p^k (1 - p)^r \tag{4-64}$$

and the corresponding generating function is[96]

$$F_{\mathbf{D}}(s) = \left(\frac{1 - p}{1 - ps}\right)^r \tag{4-65}$$

By a proper choice of units of time and volume, Eq. (4-60) is transformed into

$$\begin{aligned} \mathbf{V}(j) &= \mathbf{D}(j) - 1 \qquad \text{if } \mathbf{D}(j) > 1 \\ &= 0 \qquad\qquad\ \ \text{if } \mathbf{D}(j) < 1 \end{aligned} \tag{4-66}$$

The generating function of **V** is[96]

$$F_{\mathbf{V}}(s) = \frac{1}{\mathrm{s}}\left[\left(\frac{1 - p}{1 - ps}\right)^r - (1 - p)^r\right] \tag{4-67}$$

Flow events occur with a PDF $f_{\mathbf{N}}(k)$. In summer, when flows are caused by thunderstorms, **N** is likely to be Poisson-distributed:[91,97]

$$f_{\mathbf{N}}(k) = e^{-m}\frac{m^k}{k!} \tag{4-68}$$

with generating function

$$F_{\mathbf{N}}(s) = e^{-m+ms} \tag{4-69}$$

In winter, due to persistence in the weather patterns, **N** does not appear to

be Poisson-distributed. Kisiel et al.[91] found that winter flows in the Tucson Arroyo might follow a geometric distribution, with parameter $p = 0.6$:

$$f_N(h) = (1 - p)p^k \tag{4-70}$$

with generating function

$$F_N(k) = \frac{1 - p}{1 - ps} \tag{4-71}$$

The generating function of the sum W of a random number N of independent variables V is given by

$$F_W(s) = F_N[F_V(s)]$$

which, using, for example, Eqs. (4-67) and (4-69), yields

$$F_W(s) = \exp\left\{-m + \frac{m}{s}\left[\left(\frac{1 - p}{1 - ps}\right)^r - (1 - p)^r\right]\right\} \tag{4-72}$$

By definition, if Eq. (4-72) is expanded as a power series of s, the coefficient of s^j is $f_W(j)$, namely, the probability that total summer recharge volume is equal to j units of water. In a similar manner, the PDF of winter recharge may be computed and convoluted with summer recharge to obtain the PDF of yearly recharge. Note that the convolution of PDFs is enacted by multiplying generating functions. The use of discrete densities is not convenient. This methodology is illustrated for rainfall-runoff models by Fogel et al.[98] and Duckstein et al.[99]

Once the PDF of yearly recharge is obtained, various economic calculations may be made:

1 What risks are incurred if a certain yield policy is applied? Wells may dry out on certain years of low recharge, or recharge opportunity may be lost in wet years because of shallow water tables.

2 A minimum drought risk policy may be found by considering the minimal distribution of W over n years. Let

$$\Phi(a,n) = P(W(1) < a, W(2) < a, \ldots, W(n) < a)$$

A yield policy based on $\Phi(a,n)$ will minimize the maximum risk due to a dry year over a horizon of n years.

Furthermore, to apply this methodology to perennial flows, assume that recharge occurs on the sides of a channel only if stage height H is superior to a given value h_0. The duration of the event $\mathbf{H}(j) > h_0$ (a crossing kind of model) replaces duration $\mathbf{D}(j) > b$ [Eq. (4-60)], and the whole derivation above can be applied.

In a more refined manner, the infiltration, hence recharge, is a function of flow volume. The random process of recharge can thus, in principle, be derived from the random process of flow. However, no closed-form solution is obtainable unless the transformation of random variables involved is relatively

simple; but numerical solutions are sufficient to evaluate the distribution of the random functions of recharge and their potential effect on aquifer management.

4-16 CHEMICALS AND POLLUTANTS IN AN AQUIFER

4-16.1 Management System Description

Chemical species are naturally present in ground water. Species have been known since antiquity. Further, compounds and pollutants are introduced by man's activities. Whitsell[100] gives a very clear summary of the chemical engineering of ground water as it relates to water supply:

1 Pumping producing chemical changes by disturbing the equilibrium in an aquifer

2 Fluid-waste injection wells which may one day cause very serious problems because of insufficient knowledge about motion and diffusion processes

3 Filtration and chemical complexes in the "great ground-water factory"

"Undesirable" water may come from agriculture (fertilizer, pesticides), industry (mine smelters, tailing ponds, brine from desalinization plant), municipal sector (sewage effluent, street runoff), private sector (outhouses)—not to mention natural causes, such as salt-water intrusion.

Management problems consist essentially in preventing or controlling the pollution of ground water, while faced with one or more of the above pollutant sources. Clearly, enormous amounts of data are necessary to ascertain the origin of pollutants by means of comprehensive models, such as the Upper Santa Ana River Basin Water Quality Model.[101] In this respect, it is quite surprising that so few time-series measurements of water quality and quantity are available throughout the United States when so many wells have been drilled; as Whitsell[100] writes it:

Many Holes: Little Information

> When one considers the millions of water wells drilled in the United States just in the last decade,[102] he cannot but be dismayed to note that such enormous volume of activity has yielded so little useful information....

Farther along in his paper, he sounds an alarm bell:

> Uncontrolled drilling, construction, pumping and waste injection practices threaten eventual destruction of this vast natural resource. And unlike so many other resources, its destruction may very well be irreversible.

Management implications are clear: one should not act unless the uncertainty about possible negative consequences is reduced to a very small quantity. Worth of data and model choice problems enter under this category: until when should hydrogeologic data be taken before injecting waste or letting effluent water infiltrate with a set very small risk of polluting water supply?

Which model should be chosen to describe the motion of chemicals underground? The advantages of using a finite-state-machine model here have been described earlier; let us just point out that the management problem examines the output sequence $Z_1, Z_2, \ldots, Z_n$ corresponding to a given input sequence $X_1, X_2, \ldots, X_n$. The crucial part here is the output function G which maps the product X by S into S:

$$Z(t+1) = G(X(t), S(t)) \tag{4-74}$$

The output sequence may be a cost function depending on the cumulative concentration of phosphates and nitrates in pumped water (QUAL); the input sequence may be agricultural fertilizing practices, weather-dependent variables, and municipal waste water released in the river channel over the past months (or years). Note that we are not "optimizing," but only examining the consequence of a policy on our goals: this is the essence of simulation models as pointed out by Simpson et al.[32] and in Sec. 4-5.

4-16.2 Multiobjective Features

We have seen at least two conflicting objectives: to continue taking data and developing a model (before injecting brine, for example), or to take an action (inject the brine). One may also visualize the conflict between economic development of industry or agriculture and pollution of ground-water supply a few years later. The story of a Long Island county forbidding the sale of detergents is an example where the public has ranked the alternatives—probably on emotional rather than scientific grounds.

One possible approach to examine marginally different alternatives of reusing waste water, which radically suppresses the aquifer pollution problem, is collective utility.[103] An example has been given for the case of reuse of secondary-treated sewage effluent by Ko and Duckstein.[104] The same example has been expanded into a cost-effectiveness framework, where every measure of effectiveness corresponds to a different goal:[105] ground-water pollution, lawsuit possibilities stemming therefrom, and cost of tertiary treatment are a few of these measures of effectiveness.

4-16.3 Uncertainty Aspects

Here again, rather than represent concentration C of a given pollutant by a deterministic value, we seek a PDF $f_C(x)$ of **C**.

For example, to measure the effectiveness of artificial recharge near the coastline to prevent salt-water intrusion, we may look at the change of $f_C(x)$ as we increase the injected quantity of fresh water: we are interested not only in the mean value of C but also in its variance, as well as the extreme value properties of $f_C(x)$.

In this respect, it is quite unfortunate that most reports on modeling water quality contain average-value curves only: no confidence bands or brackets are given. This often represents loss of precious information.

4-17 CONJUNCTIVE OPERATION OF SURFACE- AND GROUND-WATER RESOURCES

4-17.1 Management System Description

In any region where pumpage of ground water affects the flow of surface water, it is logical and necessary to manage ground and surface water conjunctively. Harshbarger[10] writes that one of the advantages of proper management is the possibility of capturing waters that would otherwise evaporate: in such a case, we seek to maximize the ground-water recharge by proper pumping in the vicinity of the recharge region. This problem was mentioned in the previous section.

Another problem symmetric to the previous one is to find the pumping policy of an aquifer which does not cause undue diversion of surface water that was planned for other uses further downstream: the aquifer simulation model of Young and Bredehoeft[6,7] was developed to examine a problem of this kind in the South Platte Valley of eastern Colorado; however, the only use considered in that study was irrigation, and no uncertainties were considered.

Operations research models for the conjunctive use of ground and surface water were first proposed by Buras and Hall[106,107] and by Burt.[108] Buras[109] gives a very good review of sophisticated models for maximizing net discounted return from such a conjunctive operation. Included in Buras' presentation are most aspects of storage (surface and ground), sequential models which lead to the use of dynamic programming, uncertainty of input, and quality questions (natural salinity and salt-water intrusion). The uncertainty of demand, of time preference (discount rate), and of cost functions are considered only indirectly through a marginal or sensitivity analysis.

Lesourne's concept of collective utility,[103] which was exposed by Dupnick[71] and then by Duckstein and Kisiel,[90] could be used effectively to compare two conjunctive use policies. The problem could be presented as follows:

Problem Compare a policy (S,G) with a policy (S',G'), where S and G are the amounts of surface and ground water stored during a given time period, in the initial state, (S',G') are the corresponding amounts in the transformed state, and the quantities S and S' on the one hand and G and G' on the other hand do not differ too much (say, by 20 percent).

Region of interest A well-defined basin.

Goods Surface water; ground water; dissolved salt, respectively; kilowatt-hours (for pumping); crops produced; recreation produced.

Firms h The various production sectors (water authorities); the various consumption sectors (industrial, municipal, commercial, agricultural, recreational).

Individuals k Consumers of water; consumers of recreation.

Type of economy Open if water import or export is our desire; closed otherwise.

A simpler but more precise example of use of collective utility follows.

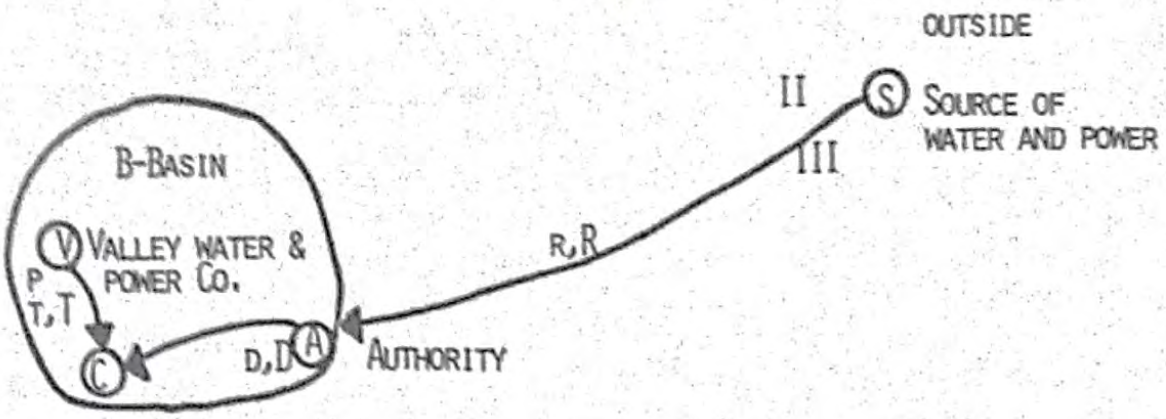


FIGURE 4-10
Definition sketch of regional economy.

A regional authority A (Fig. 4-10) controls the production of water and power (by autonomous companies) which comes from several sources within and outside of the basin B: ground water, reservoirs (salvaged and imported). Uses of the water include hydroelectric dams and steam and nuclear power plants.

The water and power production of a valley V controlled by the authority is gradually being decreased; a *recession plan* is thus applied, by depletion of sources free of contamination or pollution, obsolescence of equipment, and shifting population.

As the water and power needs of the region increase, the following question is raised: should the recession plan be applied as is or should it be slowed down? If the recession plan were applied, water and power would have to be imported from a source S.

The individuals may be separated into two groups: those in V working for the authority A, and the others.

The firms under consideration include:

The water and power producers in V
The companies in charge of transporting water and power from V to other parts of B, such as city C
The local companies adjacent to V buying power and/or water
The firms which would transport the goods under consideration from S to the purchasing agency A in the basin B
The firms which distribute the imported goods from A to the demand location in B

The goods are divided into power ($i = 1$), water ($i = 2$), and other goods ($i = 3$).

Assuming that $dU = \sum_i p(i)\, dQ(i)$, the most general model would be an *open economy* (with exports) with *incomplete utilization of resources* (layoff due to recession, facilities abandoned). We shall simplify the presentation by

assuming that

$$\sum_i p(i)\, dQ(i,E) = 0 \qquad \text{(closed economy)}$$

which is approximately true if the balance of payments of the authority A is not affected by the transformation; i.e., the transformation does not affect the budget of A's considered from the collectivity standpoint.

We shall also assume that we have full utilization of resources: this would be the case if, for example, the individuals laid off by the recession are automatically reemployed. We can thus write

$$dU = \sum_{i,h} p(i)\, dQ(i,h) \tag{4-75}$$

The initial state is the recession state; the alternative is the slowed-down recession state.

Let us consider only goods $i = 1$ and $i = 2$, that is, respectively, electricity and water. Let:

$p(i)$ = unit price of i at production site in V

$c(i)$ = unit cost of slowing down recession from regular to slowed down plan

$t(i)$ = unit transportation rate from V to C

$T(i)$ = marginal transport productivity of good i from V to C:

$$\mathrm{T}(i) = \frac{\partial t(i)}{\partial Q(i)}$$

$\pi(i)$ = unit price of i at production site in S

$r(i)$ = unit transportation rate from S to A; $R(i)$ = marginal productivity:

$$R(i) = \frac{\partial r(i)}{\partial Q(i)}$$

$d(i)$ = unit transportation rate from A to C

$D(i)$ = marginal transport productivity from A to C:

$$D(i) = \frac{\partial d(i)}{\partial Q(i)}$$

$dQ(i)$ = change in quantity of good i coming from V, a negative quantity

The change in collective utility is found to be the sum of the following terms:

1 Change of income of power and water company in V:

$$(p(1) - c(1))\, dQ(1) + (P(2) - c(2))\, dQ(2)$$

2 Change of income of the V to C transportation company:

$$\sum_i (t(i) - T(i))\, dQ(i)$$

3 Change of income of the S to A transportation company, if owned by a firm belonging to our basin B—otherwise, zero:

$$-\sum_i (r(i) - R(i))dQ(\mathrm{i}) \qquad [dQ(i) > 0]$$

4 Change of income of A to C transportation company:

$$-\sum_i (d(i) - D(i))\, dQ(i)$$

There again, the negative sign stems from our definition of $dQ(i) > 0$.

5 Change of income of firms or companies buying power and water in C:

$$\text{(cost of water from S)} - \text{(cost of water from V)}$$

or

$$\sum_i [\mathrm{II}(i) + r(i) + d(i) - p(i) - t(i)]\, dQ(i)$$

Adding up, we obtain

$$dU = \sum_i [\Pi(i) + D(i) + R(i) - T(i) - c(i)]\, dQ(i)$$

If the transportation from S to A were done by a company outside B, the expression for dU would be (see item 3 above)

$$dU = \sum_i [\Pi(i) + r(i) + D(i) - T(i) - c(i)]\, dQ(i)$$

These simple expressions for dU have been derived under precisely stated assumptions. It is a matter of routine to introduce into dU factors that have not been considered above.

4-17.2 Multiobjective Features

By definition, when we write "conjunctive use" we are implying two coupled objectives: the management of both surface water and ground water. That both objectives could be expressible in a common monetary unit might or might not be a valid assumption. How do we compare the various advantages of ground and surface storage, for example? There is no evaporation in surface storage, but, also there is no heat-sink capability, In fact, because of the thermal gradient of the earth, water stored underground may be heated. Surface storage may cause water to contain sediments, while ground storage filters the water.

Using only the economic efficiency concept is the sole criterion fallacy denounced by Kazanowski.[74] Furthermore, spillover effects are difficult to evaluate in terms of one single unit—money. Economists often lump those under the name externalities,[70] but we think that externalities should be spelled out and become part of the goals.[71]

An example of how an externality may be accounted for in a post-mortem analysis is given in a slightly different context by El Ramly,[110] who defines the "uncertainty cost" as a deterministic quantity explaining the fear of decision makers for untested excellent but new systems: this "uncertainty cost" is certainly an externality which causes systems to be chosen on other grounds than pure economics and may be applied to the choice of a conjunctive ground- and surface-water management scheme.

4-17.3 Uncertainty Aspects

The uncertainties already described for pure ground-water systems are now compounded with uncertainties due to surface water. Not only is the inflow stochastic, but so is the evaporation. Even the effect of wind waves may have to be taken into account for controlling the reservoir release,[111] which in turn may constrain the flexibility of a control scheme.

Physical uncertainties are sometimes considered in comprehensive conjunctive-use schemes;[109] at other times they are not.[7,107,108,112] However, economic uncertainties concerning demand and pricing have rarely, if ever, been considered. For example, Buras[109] describes the case of Lake Kinneret in Israel, where water must satisfy quantity and quality requirements. What if those requirements change? Would not an estimation of the PDF of demand be a useful input into a management model? This will be discussed in greater detail in Secs. 4-19 to 4-22.

The concept of collective utility developed in Sec. 4-17.1 may be extended to uncertainties in the future states of the world. For example, state 1 may be a 15 percent increase of water consumption during the time horizon, with a probability 0.3, state 2 a 10 percent increase with probability 0.2, etc. . . . In another example, the states correspond to values of demand elasticity X, or else the states represent the cost of power to transfer water from the ground to the surface.

4-18 THE REAL WORLD: A COMBINATION OF THE ABOVE

In this section we blend the three parts—description of the system, multiobjective features, and uncertainty aspects—into one general framework: that of cost effectiveness, as defined by the standardized approach of Kazanowski,[74] updated recently by the same author.[113] The problem on hand is to manage a ground-water basin, coupled with surface-water resources and subject to polluting agents, in the "best possible manner for the benefit of all." How can we attack such a vaguely stated problem? The following describes a revised version[114] of the 10-step procedure proposed by Kazanowski and gives examples along the way.

1 Define the system goals or objectives: this may be the vague state-

ment just made. However, let us emphasize after Kazanowski[113] that the definition or identification of appropriate goals is one of the sources of uncertainties in this kind of analysis.

2 Identify the system specifications, which is a translation of the goals ("benefit of all") into marketable criteria (average income, income distribution, dissolved salts in water) or nonmarketable criteria (taste of water, quality of life factors, such as risk or land subsidence). Constraints such as standards and legal requirements are included in this step.

3 Establish criteria or measures of effectiveness (MOE) that relate the performance of each alternative system to the specifications. This is another delicate step in the approach which is discussed by Kazanowski[113] and David and Duckstein.[114] The multiobjective nature of ground-water resource management (with interfaces) is appearing again at this point. For example, if the standard to be reached is concentration x of nitrates and y of phosphates, how much better is $(0.9x, 0.95y)$—or how does one compare violations of the standard: $(0.9x, 1.1y)$ versus $(1.05x, 0.8y)$? How do we judge that demand is satisfied? Spatial and temporal allocation problems may be distinguished here; however, the real world mixes the two!

Note that a system with maximum effectiveness (a nonmarketable measure) and minimum cost (a monetary measure) does not exist. This point is discussed further in Chaemsaithong et al.[115]

4 Select a fixed-cost or a fixed-effectiveness approach. Because of cost uncertainties, this is the third difficult point of the approach mentioned by Kazanowski. He states that "skepticism of the validity or merit of cost effectiveness evaluations frequently originates from a recognition of the apparent inaccuracies of cost estimates of implemented systems." This is a point that we would like to see considered more often in the literature on ground-water system management. It can be made even more sharply about operations research and optimization models.

5 Develop distinct alternative systems to reach the goals; the word "distinct" must be emphasized. A well field with 30 wells coupled to 2 reservoirs is distinct from 8 wells coupled with 5 reservoirs. But the difference between 30 and 32 wells is not worthy of a cost-effectiveness analysis. Examples of distinct waste-water reuse systems being compared in a cost-effectiveness framework are found in Ko and Duckstein:[105] one scheme exchanges the treated effluent with ground water used by agriculture; another one consists in tertiary treatment followed by recharge into the basin aquifer.

6 Determine the performance of alternative systems in terms of MOE defined in item 4: this is when simulation models may be used and when general system models may be useful. Suboptimization of each alternative system may be in order here so that an "honest" comparison may be made.

7 Generate an array of alternative systems versus MOE (monetary and nonmarketable ones). An example of such a table can be found in Kazanowski,[74] in Ko and Duckstein,[105] and in Chaemsaithong et al.[115] The last example deals with alternative dam systems on the Mekong River, which has implication for ground-water use in the basin.

8 Analyze the merits of alternative systems by ranking MOE, not by weighting them, as done by Maass et al.,[116] Major,[117] and Howe.[118] For example, income distribution may be the most important MOE, then cost, then risk of land subsidence (low, average, high), etc. . . .

9 Perform a sensitivity analysis:

a Perturb goals, MOEs, alternatives, approach (fixed cost or fixed effectiveness), hypotheses (demand, pricing policy, discount factors, state of technology such as desalinization)

b Account for uncertainty in the design, the models, the data used throughout the analysis.

10 Document the above nine steps, stating explicitly the hypotheses and rationale used. These steps include the often omitted but fundamental implementation and feedback steps.

4-19 UNCERTAINTIES, ERROR GROWTH, WORTH OF DATA, AND ECONOMIC LOSSES

One dimension of uncertainty is concerned with the amount of information available on a specific problem; another dimension is concerned with sources or causes of the uncertainty of outcomes. *Strategic uncertainty* and *technological uncertainty* are suggested as two major types.

Strategic uncertainty arises from our inability to forecast future states of the world—for example, cultural attitude toward pollution or water wastage, technological change, economic uncertainty in future water demands and waste loads on aquifers, and political and institutional uncertainty. Included is the risk of using input-output econometric models as forecasts of future economic conditions and, in turn, as disturbances on aquifers.[119] The importance of the uncertainties has already been evaluated in Sec. 4-13.

Technological uncertainty arises in the uncertain nature of hydrologic, engineering, geologic, and scientific calculations, in the random inputs of nature, and in our ignorance about nature's modes of behavior. Such uncertainty occurs because of errors in instrumental measurements, space-time sampling, and modeling. These give rise to the inverse problems that are identified in Sec. 4-7.

In what follows, we present results on error propagation (giving it an economic interpretation) and on the relative importance of data, aquifer parameters, pumpage, recharge, and water levels.

4-20 ERROR GROWTH (OR THE LIMITS OF DETERMINISTIC PREDICTABILITY)

If we accept that Eqs. (4-1) to (4-3) are sufficient to represent the dynamics of the system, then we must also concede that the future states of the system are determined and absolutely predictable, provided we could specify exactly the initial and boundary conditions and also the parameters. These quantities, however, are estimated from few discrete observations in space and time and hence are not completely known. We should also notice that we cannot counter this problem by having a dense sampling system, since it would alter the nature of the system being measured. Assuming that the data contained no instrument, reading, roundoff, or transmission errors, the estimation of the parameters in initial and boundary conditions would still not be the same as the true state. At various points the estimated state will differ from the true state by an amount that depends on distance from the point in question to the nearest observation station. The *analysis error*, defined by Thompson[120] as the difference between the true and the reconstructed state, is zero at or near the observation points. The error would increase as the distance between stations increases. Since the observation stations are usually fixed, and the disturbances that they are supposed to measure are randomly distributed relative to them, the analysis error at any point is also random. As a matter of fact, the best that can be done in practice is to define an ensemble of possible starting points. The hydrodynamical equations of flow [e.g., Eqs. (4-1) to (4-3)] are equally applicable to any point in this ensemble. If the error in the initial state is large compared to the true state, the predictions may be completely unreliable.

In a purely deterministic framework, one of the various points in the ensemble is considered as the "best" to the exclusion of all others, and the prediction is based on this best value. This best state is usually obtained by averaging out (e.g., average T and S in space) or by some other method. The problem, however, is that the errors in predicted value are not static, but are functions of time and may grow (or decrease) with time. This idea has been explored much more extensively in meteorology than in hydrology. This may be mainly due to the fact that the atmosphere is (compared to a ground-water system) a very unstable system so that high-frequency disturbances often alter the atmosphere's character. For example, even with a small analysis error to start with, in the prediction of atmosphere, with perfect hydrodynamic equations, the error may grow so fast that no reliable prediction of the atmosphere can be made, say, after 24 hours. On the other hand, the ground-water system is usually a highly dampened system, and on a large scale, high-frequency disturbances are dissipated quickly.

In meteorology, Thompson,[120] working with winds, trying to find the limit of predictability, related the root-mean-square error of the predicted value at any time to the initial root-mean-square error, period of forecasting, and various other meteorological factors. The error in this case was assumed to

have certain statistical properties, and similar properties of the predicted value at various times were computed theoretically. Gleeson,[121] working along similar lines, examined the spread of the ensemble of the initial states with time and pointed out that the variance of any parameter increases or decreases in time according to whether the correlation between the current value of the parameter and its time derivatives is positive or negative. Epstein,[122] in addition to the spread of points, also examined the behavior of the mean of the ensemble and came to the conclusion that, in general, the best forecast will not be obtained by applying deterministic equations to the average initial conditions.

In ground-water flow, the possibility of making an error in the estimation of initial conditions and parameters is admitted by many hydrologists, but under ordinary circumstances no effort is made to ascertain the manner in which the errors propagate through the forecast procedure. By a frequency analysis of field data on hydraulic conductivity, McMillan[123] found that, in a single geologic formation, it has a log-normal distribution (see also Freeze[124]). Assuming this distribution and steady-state flow conditions (Laplace's equation) he came up with a relation:

$$\frac{S_d}{\Delta H} = F_d \frac{S_k}{\bar{k}} \tag{4-76}$$

where S_d = standard deviation of differences in potential values for homogeneous and heterogeneous solutions of Laplace's equation
ΔH = mean difference in potential between adjacent nodes in the principal direction of flow
S_k = standard deviation of conductivity
$\bar{k}$ = mean or the homogeneous value of conductivity
F_d = an empirical factor which was found to range between 0.05 to 1.95, with a value near 0.5 for typical basin conditions

The propagation of error with time, however, was not considered by McMillan, because he did not extend his analysis to unsteady flow. Bibby[125] analyzed an unsteady system represented by Eq. (4-6) under the assumption of measurement errors in T, S and initial conditions. Assuming the errors to have a normal distribution, to be independent of the observed values, and to be mutually uncorrelated, he concluded that the error in the initial head is the only significant error in predicting error on final head. He also found that, while the errors were linearly related for short periods of time, the error in the final head approached a constant value after a certain time.

No investigation of this nature with respect to Eqs. (4-7) and (4-8) has yet come to the notice of these authors. One reason why such investigations have not yet been carried out may be the mathematical complexities involved. The treatment of differential equations with stochastic parameters or initial and boundary conditions, called stochastic differential equations in mathematics, is a comparatively new field (see Saaty[126] and Beran[127]). This is especially true of stochastic partial differential equations, treatment of which is definitely in the early stages. The simple example (somewhat artificially constructed) in this

section, therefore, does not pretend to deal with stochastic partial differential equations, but is aimed to focus attention on the need for this kind of study.

4-21 ERROR PROPAGATION IN THE THEIS EQUATION

We assume a nonleaky, artesian aquifer of semi-infinite extent with a small-sized, fully penetrating well discharging at constant rate. For the moment we also consider that the values of the parameters T and S have been measured at a few points in the aquifer by pumping tests and averaged to get the homogeneous value for the aquifer. For these conditions we use the Theis equation (4-22). The integral in Eq. (4-17) is evaluated from the series

$$\int_u^\infty \frac{e^{-u}}{u}\,du = -0.5772 - \log_e u + u + \frac{u^2}{2.21} + \frac{u^3}{3.31} + \cdots$$

If $u \leq 0.01$, all terms beyond $\log_e u$ can be neglected (Cooper and Jacob;[12] see Walton[11]) and Eq. (4-17) can be written as

$$s = \frac{Q}{4\pi T} \log_e \frac{2.25Tt}{r^2S}$$

Changing the units to the gallon-day-foot system:

$$s = \frac{114.60}{T} \log_e \frac{.3\,Tt}{r^2S} \tag{4-77}$$

where s = drawdown, feet
Q = discharge, gallons per minute (gpm)
T = coefficient of transmissivity, gallons per day per foot (gpd/ft)
t = time after pumping started, days
S = coefficient of storage, fraction
r = distance from pumped well to observation point, feet

and $h(r,t)$ can be written as

$$h(r,t) = h_0 - \frac{114.6Q}{T} \log_e \frac{.3\,Tt}{r^2S} \tag{4-78}$$

where $h_0 = H(r,o)$ represents the initial condition. Normally, Eq. (4-78) would be used with average values of h_0, S, and T, but we now assume that the range of values for them is known and assign them a uniform distribution; i.e., the probability density function of any parameter λ is

$$f(\lambda) = \frac{1}{\beta - \alpha} \qquad \alpha \leq \lambda \leq \beta \tag{4-79}$$

Notice that, in this case,

$$\begin{aligned} \text{Mean} &= \bar{\lambda} = E(\lambda) = \frac{\alpha + \beta}{2} \\ \text{Variance} &= \alpha_\lambda^2 = E(\lambda - \bar{\lambda})^2 = \frac{(\beta - \alpha)^2}{12} \end{aligned} \tag{4-80}$$

so that $\alpha = \bar{\lambda} - \sqrt{3}\alpha_\lambda$ and $\beta = \bar{\lambda} + \sqrt{3}\alpha_\lambda$

Thus we assume that the probability of the parameter value lying anywhere between α and β is constant. We also assume S, T, and h to be statistically independent and consider the standard deviation of λ as the measure of error in λ.

Taking expectation of both sides of Eq. (4-78), we have

$$\begin{aligned} E(h) &= \bar{h} \\ &= E(h_0) - 114.6Q\left[E\left(\frac{1}{T}\right)(\log .3 + \log t - 2\log r)\right. \\ &\qquad \left. + E\left(\frac{\log T}{T}\right) - E\left(\frac{\log S}{T}\right)\right] \end{aligned} \tag{4-81}$$

and variance of $h = \sigma_h^2 = E(h^2) - (E(h))^2$. This variance has been evaluated by Sagar and Kisiel.[47] Because it has a rather involved form, it is not given here.

Thus, given the means and variance of parameters, the mean and variance of the final head can be evaluated. Four cases were considered:

1 Error in initial head h_0 only
2 Error in transmissivity T only
3 Error in storage S only
4 Equal errors in all the three parameters λ

For the numerical example, the following data was assumed:

$Q = 1{,}000$ gpm
Effective radius of well $= 0.5$ ft
$E[h_0] = +100.00$ ft
$E[T] = 100{,}000$ gpd/ft
$E[S] = 0.2$

Errors in final head at the pumping well (with $r = 0.5$ ft) are calculated and plotted in Figs. 4-11 to 4-14. Notice that the ordinate in these figures is the standard deviation of the final head expressed as a percentage of the expected value of the final head. The results are as follows:

1 The Theis equation is most sensitive to errors in initial head and least sensitive to errors in coefficient of storage. The error in the final head at any time is greater in magnitude than the error in the initial head to start with. On the other hand, the error in the final head is considerably less than the initial errors in S and T.
2 The plots of errors in final head versus time are approximately straight lines on a semilog paper, indicating an exponential relationship between error in final head and time.
3 While the errors in final head increase with time, the time rate of increase decreases. This is to be expected due to the damped nature of the ground-water system.

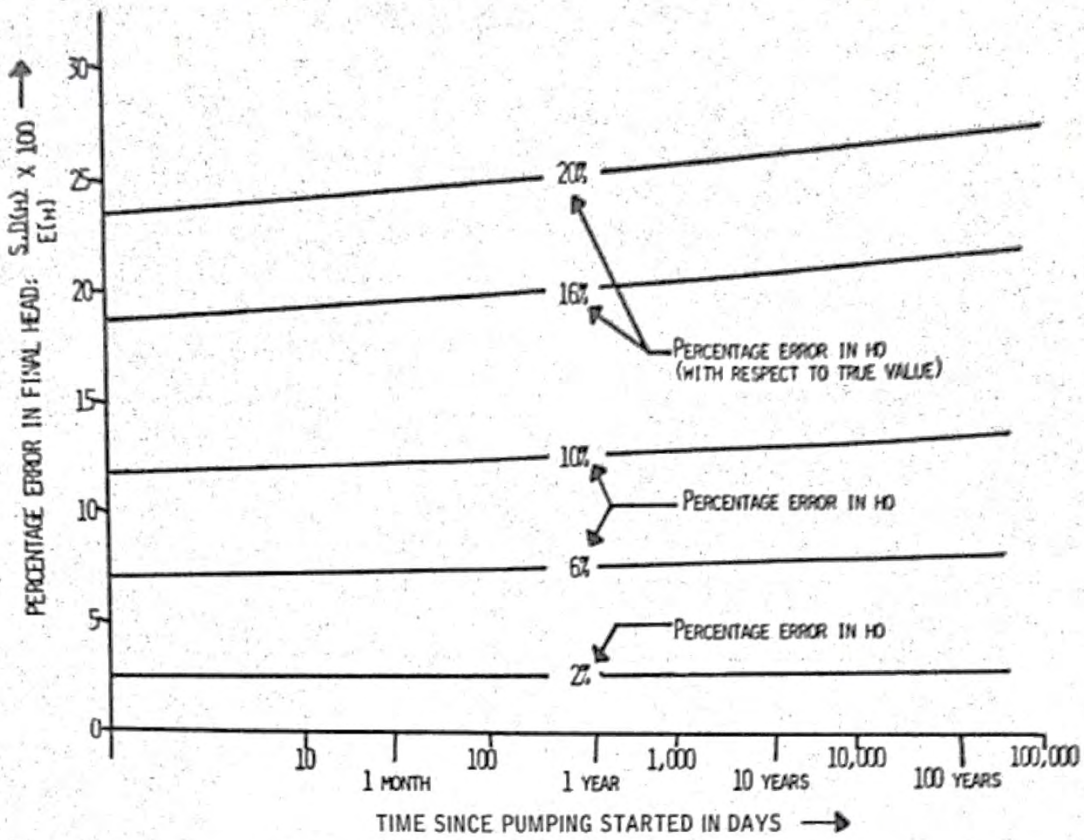


FIGURE 4-11
Error in final head versus time for different amount of error in initial head h_0.

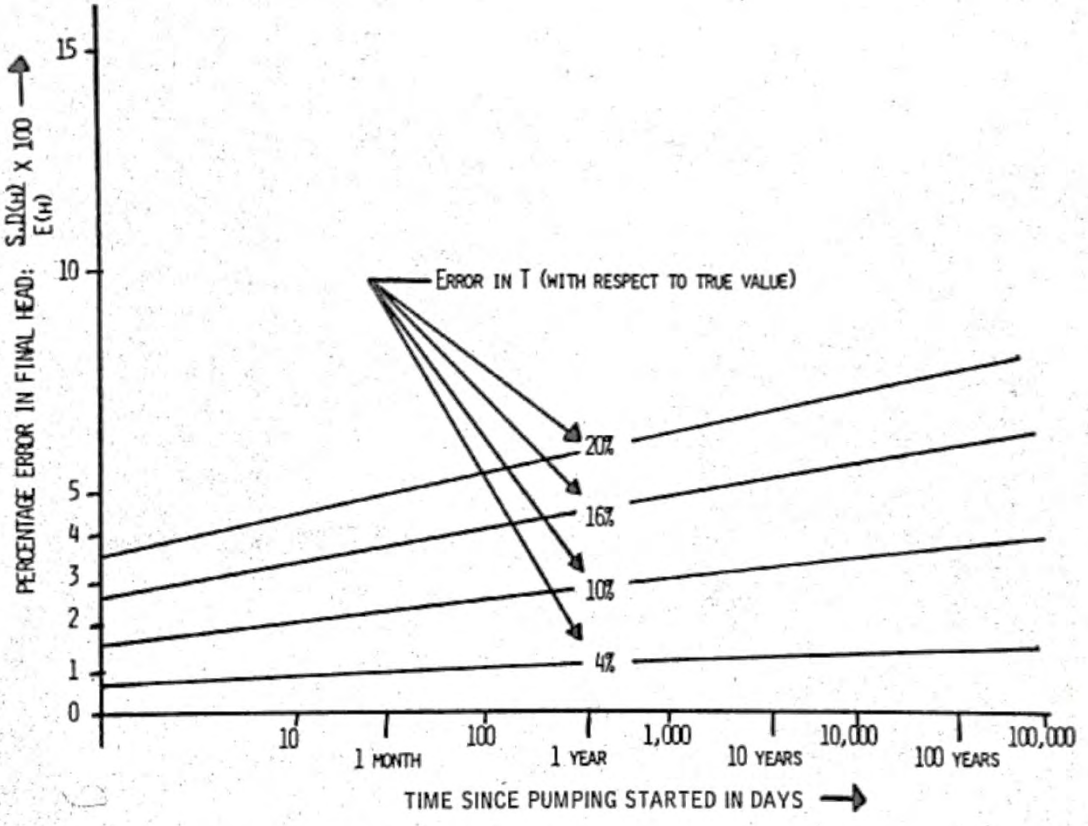


FIGURE 4-12
Error in final head versus time for different amount of error in transmissivity T.

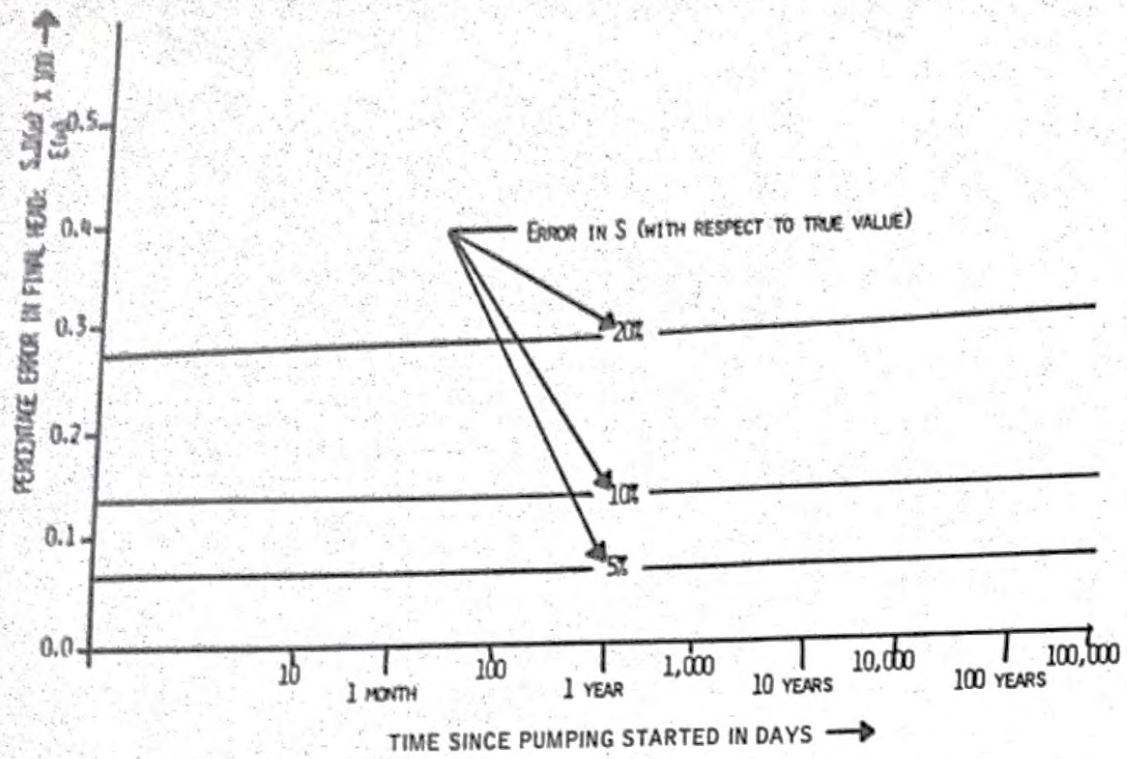


FIGURE 4-13
Error in final head versus time for different amount of error in storage coefficient S.

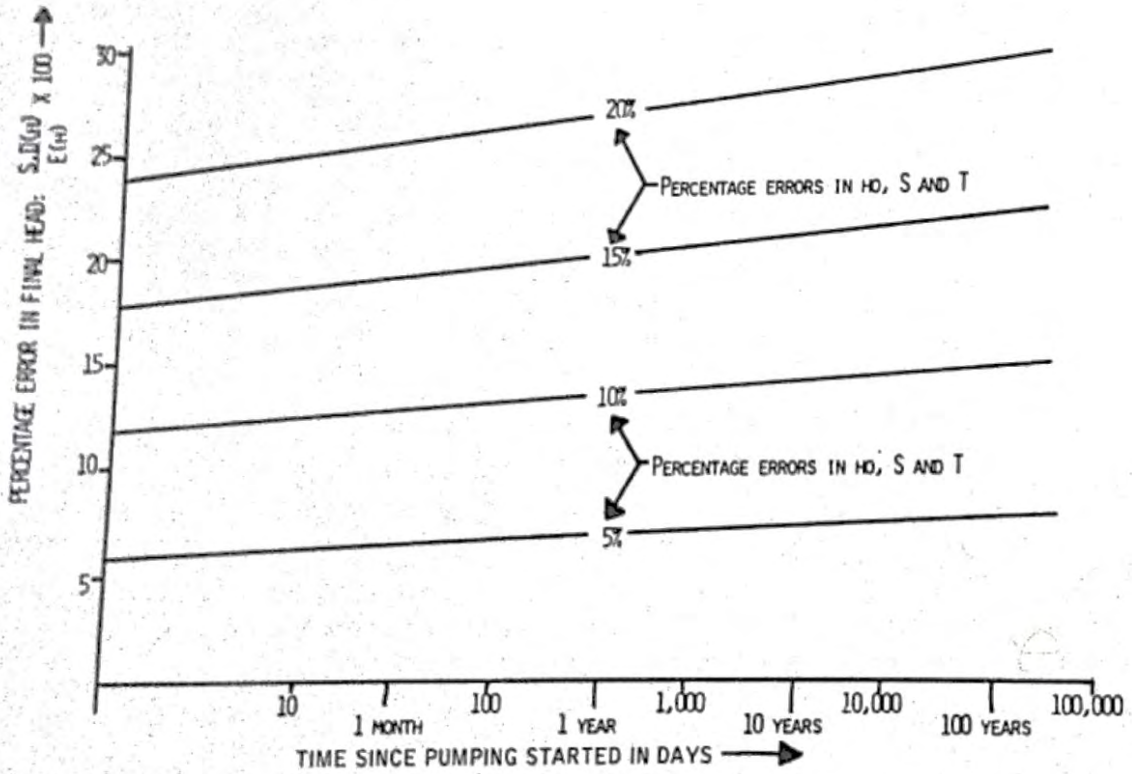


FIGURE 4-14
Error in final head versus time for equal errors in h_0, S, and T.

The mean value of head as calculated from Eq. (4-78) was found to be the same as calculated directly from Eq. (4-81), using the mean values of the parameters. This is generally true only for linear systems such as are defined by Eqs. (4-1) to (4-3). Thus if the average values of the parameters are considered to be their true values, then Eq. (4-17) produces correct results; otherwise, for a given deviation in the parameter values, the corresponding deviation from the calculated mean head increases with time.

It is a common practice in designing a ground-water utilization system to consider the aquifer as homogeneous, i.e., a lumped system as far as parameters are concerned. The lumped value of a parameter is usually obtained by simply averaging the measured values at a few discrete points. A standard deviation of 50 percent of the mean is not uncommon.[2] These uncertainties in the parameter values cause corresponding uncertainties in the predicted head. What may not be obvious, however, is that the uncertainty in the final head increases with time.

If the true value of the head is not what we predict, obviously our design (for example, of a well) will not be correct. Thus the actual area dewatered, the quantity of water that can be economically extracted, and the lift for the pumps will all be in error. This is true irrespective of the fact that the system may be underdesigned or overdesigned.

Assigning a monetary value to the opportunity loss incurred due to these errors, one may write for loss:

$$L = \sum_{\text{area}} \sum_{\text{time}} c(i,j) \cdot \text{error } (i,j) \tag{4-82}$$

where c is the cost of a unit error in model area i and at time j. For example, the electric energy required to pump 1 acre-foot of water is, according to Nelson and Busch,[129]

$$\text{KWH} = \frac{1.024 \text{ lift}}{E_e} \tag{4-83}$$

where E_e is the overall efficiency of the pumping unit. Therefore, for each foot of error in the lift we over- or underassign $1.024/E_e$ KWH of electricity.

For intelligent design, therefore, it is important to understand the propagation of errors in the deterministic equations and also to understand that the reliability of predictions decreases with time.

We conclude from the above study that an initial uncertainty in the estimation of parameters or initial conditions causes uncertainty in the result which grows with time. Thus, there is a limit to the length of time up to which the predictions would be valid.

This analysis, however, does not give the full probability distribution of the error. Higher order moments can be calculated, although much greater computational effort would be required. The sign of the error (i.e., whether above or below the true value) is also not predicted by this method. This

problem can be circumvented by considering the expected economic or opportunity loss (EOL) over the range of possible errors;[2] the EOL is based on a weighing of costs of each error with the probability of occurrence of that error.

The appropriateness of the results of this paper depends entirely on the suitability of the uniform distribution as a basis for weighting of errors. In the bayesian viewpoint this is a "know-nothing" prior distribution and indicates that the geologist or hydrologist has no preference for any particular parameter value and that he is willing only to specify the bounds (α and β) on the parameters. There is need to extend the analysis, if judged to be of value, to "richer" or unimodel prior distributions that reflect greater confidence in certain parameter values than in others.

A far more general and better method of investigating error growth would be to consider the parameters as stochastic processes and the differential operator in the linear differential equation as a linear random operator and then to evaluate the properties of the resulting stochastic process.[130] This approach, however, is too complex mathematically and is not yet fully developed.

The theoretical approach to error analysis in the use of partial differential equations of hydrologic systems has merit because of the weakness of classical sensitivity analysis and the expense and provincial character of computer experiments on errors in ground-water modeling.

4-22 WORTH OF DATA

To date, worth of data studies on the Tucson aquifer has been confined to hydraulic data, even though earlier we alluded to the value of thermal data for hydraulic and chemical modeling of the aquifer. Gates and Kisiel[2] constructed two digital computer hydraulic models of the Tucson aquifer to evaluate the worth of additional basic hydraulic data in modeling. The two models differ primarily in degree of detail: the large-scale model consisted of 1,890 nodes at a ½-mile (0.80-km) spacing, and the small-scale model consisted of 509 nodes at a 1-mile (1.60-km) spacing. The large-scale model was the forerunner of the current updated digital model used in the work for the City of Tucson Department of Water and Sewers.

Gates and Kisiel[2] classified potential errors in the Tucson basin models as errors associated with computation, errors identified with mathematical assumptions, and errors in basic data on (1) model parameters—transmissivity and coefficient of storage at each node of the digital model—and (2) initial water levels, discharge (pumpage), and recharge. Not all these data are available at every node of the model—a very common situation that creates modeling difficulties and requires the exercise of considerable judgment.

The need for a more objective approach to judging the worth of more data on the above parameters and variables prompted the investigation of the role of statistical decision theory. Worth of data (or value of information) is defined in terms of expected (or weighted averages) values of parameters and variables.

Because their true values are not known, a probability distribution is used to weight each possible value to find an *expected value or worth*. An extension of this approach, not taken as yet in our research on aquifers, is to weight the errors with respect to the *unknown* true value with an economic loss function. To obtain such a function requires specification of use of the data for well design, well operation, location of a well field, and so on. The loss function would assign an economic value to each error of overestimation and underestimation. Determination of such loss functions is an area for collaboration between researchers, planners, and economists.

Statistical decision theory was used to compute expected error in predicted water levels and expected (average) worth of more sample data (expected or average reduction in error with new data) over the 509-node model associated with uncertainty in a model variable at one given node. Discrete probability distributions with largely subjectively determined parameters were used to characterize tested variables. Ninety-one parameters and variables at 61 different locations in the model were tested, using 6 separate error criteria (loss functions). Of the tested parameters and variables, 67 were chosen because their expected errors were likely to be large at some nodes, and, for the purpose of comparison, 24 were chosen because their expected errors were *not* likely to be particularly large at other nodes. Of the imperfectly known parameters and variables, discharge/recharge and transmissivity have the largest expected errors that average 155 and 115 ft (47.24 and 35.05 m), respectively, per 509 nodes for the criterion of absolute value of error. Similarly, discharge/recharge and transmissivity have the largest averages: 29 and 14 ft (8.84 and 4.27 m), respectively, per 509 nodes. In contrast, initial water level and storage coefficient have lesser values in each case. Of the more certain parameters and variables, transmissivity and initial water level generally have the largest expected errors [a maximum of 73 ft (22.25 m) per 509 nodes] and expected sample worths [a maximum of 12 ft (3.65 m) per 509 nodes]; whereas storage coefficient and discharge/recharge have smaller values. These results are likely not typical of those from many ground-water basins, and they may apply only to the Tucson basin.

The largest expected errors are associated with nodes at which values of discharge/recharge are large or at which prior estimates of transmissivity are very uncertain. Large expected sample worths are associated with variables which have large expected errors or which could be sampled with relatively little uncertainty. Results are similar for all six of the error criteria used.

Tests were made of the sensitivity of the method for simplifications and assumptions, such as the type of probability distribution function assumed for a variable, the values of the estimated standard deviations of these distributions, and the number and spacing of the elements of each distribution. The results are sensitive to all the assumptions and, therefore, are likely correct only in order of magnitude. However, the ranking of the types of parameters and variables in terms of magnitude of expected error and expected sample worth is not sensitive to the assumptions. Proceeding from highest to lowest expected error

and expected sample worth, the rankings are

1 For the large-error category (based on 67 parameters and variables): discharge/recharge, transmissivity, initial water level, and coefficient of storage

2 For the small-error category (based on 24 parameters and variables): transmissivity, initial water level, storage coefficient, and discharge/recharge

Thus, the general conclusions on relative effects of errors for different variables are likely valid.

Limited studies of error propagation indicated that errors in predicted water levels associated with extreme erroneous values of a parameter or variable at one node are commonly less than 4 ft (1.22 m) per node at a distance of 1 mile (1.60 km) from the tested node (from which the error is propagated). This suggests that, in many cases, prediction errors associated with errors in basic data are not a major problem in digital modeling.

Importance of error analysis in the context of worth of data studies cannot be overemphasized. Classical sensitivity analysis alone is not enough. Hydrologists putting forth estimates of water budgets, ground-water supply, or hydrologic, chemical, and thermal properties of an aquifer (among many kinds of hydrologic estimates) have an increasing responsibility to identify the reliability of the estimates in terms of upper and lower bounds of the estimate, variance of the estimate, and, preferably and more generally, the probability distribution of the errors within the error bounds. Typically, the hydrologist does not, but should, assess at least the economic consequences of both high and low estimates. A high estimate may lead to overdesign and perhaps optimism on available resources; an opportunity loss or regret ensues, in that the investors later wish that the money could have been used for another socially desirable purpose. A low estimate leads to underdesign and, probably, a real socioeconomic loss. Both kinds of losses require explicit evaluation in the spirit of project uncertainty analysis. Water-resources engineers and planners have a responsibility to know enough of hydrologic and other uncertainties so as to insist on explicit information on these uncertainties and to be able to undertake socioeconomic interpretations of such uncertainties. It is the essence of professionalism to attack these issues in a frontal manner. To say that there is not enough data is to expect the user to believe the implied assumptions of estimates (customarily taken to be true). All parties concerned owe each other a statement of assumptions, because to do otherwise is to impose on the other parties one's own value structure.

4-23 CONCLUSIONS

The issues on model building, calibration, and validation apply equally to models of socioeconomic processes as coupled to physical models of the aquifer. Models falling in this category include operation research models, microecono-

mic and macroeconomic models, and econometric models. These issues have been commonly overlooked in the context of water resources and other natural resource fields. In addition, much work remains to be done on the design of space-time sampling networks in ground-water hydrology[32,131,132] and the related inverse problems.[133,134]

ACKNOWLEDGMENTS

Considerable recognition is in order for the substantial review of inverse problems in ground-water hydrology undertaken by B. Sagar and used in the preparation of Secs. 4-6 to 4-11 of this chapter. Close collaboration with many individuals involved in ground-water hydrology has facilitated our orientation; they include E. S. Simpson, J. W. Harshbarger, J. S. Gates, D. Supkow, J. G. Ferris, and others. Research contracts and grants from the Office of Water Resources Research (U.S. Department of the Interior) and The City of Tucson have provided the necessary resources for our collective efforts.

REFERENCES

1. WITHERSPOON, P. A., and R. A. FREEZE, The Role of Aquitards in Multiple-Aquifer Systems (Results of a Penrose Conference), *Geotimes*, pp. 22–24, April 1972.
2. GATES, C. S., and C. C. KISIEL, Worth of Additional Data to a Digital Computer Model of a Groundwater Basin, *Water Resour. Res.*, vol. 10, no. 5, pp. 1031–1038, 1974.
3. SAGAR, B., Calibration and Validation of Aquifer Models, Tech. Rept. No. 17, Department of Hydrology and Water Resources, University of Arizona, Tucson, 1973.
4. KISIEL, C. C., D. SUPKOW, and W. TETLEY, Role of Digital Computer Models of Aquifers in Water Resources Planning: Case Study in Tucson, Arizona, Proceedings of the International Symposium on Planning of Water Resources, Mexico City, December 4–8, 1972.
5. ENGINEERING-SCIENCE, INC., and MARUM AND MARUM, INC., "A Regional Plan for Water, Sewerage and Solid Waste Management," Urban Systems Engineering Demonstration Program, Pima Association of Governments, Tucson, Arizona, 1972.
6. BREDEHOEFT, J. D., and R. A. YOUNG, The Temporal Allocation of Groundwater—A Simulation Approach, *Water Resour. Res.*, vol. 6, no. 1, pp. 3–21, 1970.
7. YOUNG, R., and J. BREDEHOEFT, Digital Computer Simulation for Solving Management Problems of Conjunctive Groundwater and Surface Water Systems, *Water Resour. Res.*, vol. 8, no. 3, 1972.
8. KNOWLES, T. R., B. G. CLABORN, and D. M. WELLS, A Digital Computer Model for Predicting Regional Aquifer Response, Report 71–7, Water Resources Center, Texas Tech. University, Lubbock, December 1971.
9. DOMENICO, P. A., "Concepts and Models in Groundwater Hydrology," McGraw-Hill Book Company, New York, 1972.

10. HARSHBARGER, J. W., Groundwater Resources Evaluation and Exploration, *Proceedings of the Seminar on Groundwater*, Food and Agricultural Organization of the United Nations, Granada, Spain, October 1971.
11. WALTON, W., "Groundwater Resource Allocation," McGraw-Hill Book Company, New York, 1971.
12. DEJU, R. A., "Regional Hydrology Fundamentals," Gordon and Breach, Science Publishers, New York, 1971.
13. KISIEL, C. C., and L. DUCKSTEIN, Operations Research Study of Water Resources—Part II: Case Study of the Tucson Basin (Arizona), *Water Resour. Bull.*, vol. 6, no. 6, pp. 857–867, December 1970.
14. ACKERMANN, W. C., Hydrology Becomes Water Science, *Trans. Am. Geophys. Union*, vol. 50, no. 3, 1969.
15. FREEZE, R. A., Role of Subsurface Flow in Generating Surface Runoff: 2. Upstream Source Areas, *Water Resour. Res.*, vol. 8, no. 5, pp. 1272–1283, 1972.
16. MONARCHI, D., C. C. KISIEL, and L. DUCKSTEIN, SEMOPS: An Interactive Algorithm for Multiobjective Problem-Solving, *Water Resour. Res.*, vol. 9, no. 3, pp. 837–850, 1973.
17. WYMORE, A. W., "A Mathematical Theory of Systems Engineering: The Elements," John Wiley & Sons, Inc., New York, 1967.
18. WYMORE, A. W., "Notebook on Systems Engineering," Department of Systems & Industrial Engineering, University of Arizona, Tucson, 1972.
19. HIMMELBLAU, D. M., and K. B. BISCHOFF, "Process Analysis and Simulation: Deterministic Systems," John Wiley & Sons, Inc., New York, 1968.
20. DERUSSO, P. M., R. J. ROY, and C. M. CLOSE, "State Variables for Engineers," pp. 325–329, 347, John Wiley & Sons, Inc., New York, 1965.
21. DUCKSTEIN, L., and C. C. KISIEL, Control of Hydrologic Systems for Multiple Uses in a Closed-Loop Framework, *J. Hydrol.*, vol. 15, pp. 69–76, 1972.
22. THEIS, C. V., The Relation between the Lowering of the Piezometric Surface and the Rate and Duration of Discharge of a Well Using Groundwater Storage, *Trans. Am. Geophys. Union*, vol. 16, pp. 519–524, 1935.
23. VENETIS, C., On the Impulse Response of an Aquifer, *Bull. Int. Assoc. Sci. Hydrol.*, vol. 13, no. 8, pp. 136–139, 1968.
24. SCHWARZ, R. J., and B. FRIEDLAND, "Linear Systems," McGraw-Hill Book Company, New York, 1965.
25. MOENCH, A., Ground Water Fluctuations in Response to Arbitrary Pumpage, *Ground Water*, vol. 9, no. 2, pp. 4–8, 1971.
26. MOENCH, A., and C. C. KISIEL, Application of the Convolution Relation to Estimating Recharge from an Ephemeral Stream, *Water Resour. Res.*, vol. 6, no. 4, pp. 1087–1094, 1970.
27. KISIEL, C. C., Time Series Analysis of Hydrologic Data, in Van Te Chow (ed.), "Advances in Hydrosciences," pp. 1–121, Academic Press, Inc., New York, 1969.
28. ERIKSSON, E., Groundwater Time Series, *Nord. Hydrol.*, vol. I, no. 3, pp. 181–205, 1970.

29. ERIKSSON, E., Cross-Spectrum Analysis of Groundwater Levels in an Esker, *Nord. Hydrol.*, vol. I, no. 4, pp. 245–259, 1970.
30. KISIEL, C. C., and L. DUCKSTEIN, Time and Frequency Domain Identification of a Causal Bivariate Stochastic Process, pp. 364–371, Proceedings of the International Symposium on Mathematical Models in Hydrology, Warsaw, Poland, July 1971.
31. DUCKSTEIN, L., and C. C. KISIEL, General Systems Approach to Groundwater Problems, pp. 100–115, Proceedings of the National Symposium on Analysis of Water Resources Systems, American Water Resources Association, Denver, Colo., July 1968.
32. SIMPSON, E. S., C. C. KISIEL, and L. DUCKSTEIN, Space-Time Sampling of Pollutants in Aquifers, pp. 443–469, Proceedings of the Symposium on Ground Water Pollution, 15th General Assembly of International Union of Geodesy and Geophysics, Moscow, USSR, August 1971.
33. GILL, A., "Introduction to the Theory of Finite-State Machines," McGraw-Hill Book Company, New York, 1962.
34. BOOTH, T. L., "Sequential Machines and Automata Theory," John Wiley & Sons, Inc., New York, 1968.
35. BREDEHOEFT, J., and G. PINDER, Digital Analysis of Areal Flow in Multiaquifer Groundwater Systems: A Quasi Three-Dimensional Model, *Water Resour. Res.*, vol. 6, no. 3, pp. 883–888, 1970.
36. REMSON, I., G. M. HORNBERGER, and F. J. MOLZ, "Numerical Methods in Subsurface Hydrology," Interscience Publishers, a division of John Wiley & Sons, Inc., New York, 1971.
37. NALLUSWAMI, M., R. A. LONGENBAUGH, and D. K. SUNADA, Finite Element Method for the Hydrodynamic Dispersion Equation with Mixed Partial Derivatives, *Water Resour. Res.*, vol. 8, no. 5, pp. 1247–1250, 1972.
38. FIERING, M., "Streamflow Synthesis," p. 67, Harvard University Press, Cambridge, Mass., 1967.
39. KISIEL, C. C., and D. SUPKOW, Digital Modeling of the Tucson Basin, Reports to the City of Tucson, Department of Hydrology and Water Resources, University of Arizona, Tucson, 1971–1973.
40. WOOLHISER, D. A., Deterministic Approach to Watershed Modeling, *Nord. Hydrol.*, vol. II, pp. 141–146, 1971.
41. PAPOULIS, A., "Probability, Random Variables, and Stochastic Processes," McGraw-Hill Book Company, New York, 1965.
42. COX, D. R., and H. D. MILLER, "The Theory of Stochastic Processes," pp. 286–289, John Wiley & Sons, Inc., New York, 1965.
43. DOOGE, J. C. I., The Hydrologic System as a Closed System, vol. 2, pp. 98–113, Proceedings of the International Hydrology Symposium, Fort Collins, Colo., 1967.
44. KISIEL, C. C., H. K. QASHU, and J. J. ROGERS, Error Analysis and System Identification of Water Balance and the Rainfall Excess Component in Real and Model Watersheds, Paper Presented at the National Winter Meeting, AGU, San Francisco,

California, December, 1970. Available as "Systems Approach in Hydrology," Water Resources Publishers, Inc., Fort Collins, Colo.

45. KISIEL, C. C., and L. DUCKSTEIN, Economics of Hydrologic Modeling, pp. 319–330. Proceedings of the International Symposium on Mathematical Modeling Techniques in Water Resources Systems, Ottawa, Canada, May 1972.
46. LOVELL, R. E., Hydrologic Model Selection in a Decision Making Context, Tech. Rept. No. 26, Department of Systems and Industrial Engineering, University of Arizona, Tucson, 1975.
47. SAGAR, B., and C. C. KISIEL, Limits of Deterministic Predictability of Saturated Flow Equations, pp. 289–300, Proceedings of the International Symposium on Flow through Porous Media, University of Guelph, Canada, 1972.
48. ISAACSON, E., and H. B. KELLER, "Analysis of Numerical Methods," pp. 21–45, 442–444, John Wiley & Sons, Inc., New York, 1966.
49. ASTROM, K., and P. EYKHOFF, Identification and Process Parameter Estimation, pp. 112–131, Proceedings of the Second Symposium on System Identification, International Federation for Automatic Control, Prague, Czechoslovakia, 1970.
50. EYKHOFF, P., Process Parameter and State Estimation, *Automatica*, vol. 4, pp. 205–233, 1968.
51. PHILLIPSON, G. A., "Identification of Distributed Systems," American Elsevier Publishing Company, Inc., New York, 1971.
52. BEKEY, G. A., System Identification—An Introduction and Survey, *Simulation*, vol. 15, no. 4, pp. 151–166, 1970.
53. KLEINECKE, D., Use of Linear Programming for Estimating Geohydrologic Parameters of Groundwater Basin, *Water Resour. Res.*, vol. 7, no. 2, pp. 367–381, 1971.
54. HAIMES, Y. Y., R. L. PERRINE, and D. A. WISMER, Identification of Aquifer Parameters by Decomposition and Multilevel Optimization, Contribution no. 123, University of California, Water Resources Center, 1968.
55. VEMURI, V., and W. J. KARPLUS, Identification of Nonlinear Parameters of Groundwater Basins by Hybrid Computations, *Water Resour. Res.*, vol. 5, no. 1, pp. 172–185, 1969.
56. SAGE, A. P., "Optimum Systems Control," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1968.
57. VEMURI, V., J. A. DRACUP, and R. C. ERDMANN, Sensitivity Analysis Method of System Identification and Its Potential in Hydrologic Research, *Water Resour. Res.*, vol. 5, no. 2, pp. 341–349, 1969.
58. LOVELL, R., L. DUCKSTEIN, and C. C. KISIEL, Use of Subjective Information in Estimation of Aquifer Parameters, *Water Resour. Res.*, vol. 8, no. 3, pp. 680–690, 1972.
59. EMSELLEM, Y., and G. DE MARSILY, An Automatic Solution for the Inverse Problem, *Water Resour. Res.*, vol. 7, no. 5, pp. 1264–1283, 1971.
60. NELSON, W. R., In-Place Determination of Permeability Distribution for Heterogeneous Porous Media through Analysis of Energy Dissipation, *Soc. Pet. Eng. J.*, vol. 38, pp. 32–42, March 1968.

61. CHEN, W. H., and J. H. SEINFELD, Estimation of Spatially Varying Parameters in Partial Differential Equations, *Int. J. Control*, vol. 15, no. 3, pp. 487–495, 1972.
62. CANNON, J. R., Determination of Certain Parameters in Heat Conduction Problems, *J. Math. Anal. Appl.*, vol. 8, pp. 188–201, 1964.
63. JONES, F. B., JR., Various Methods for Finding Unknown Coefficients in Parabolic Differential Equations, *Commun. Pure Appl. Math.*, vol. XVI, pp. 33–44, 1963.
64. YEH, W., and G. W. TAUXE, Quasilinearization and the Identification of Aquifer Parameters, *Water Resour. Res.*, vol. 7, no. 2, 1971.
65. BELLMAN, R. E., and R. E. KALABA, "Quasilinearization and Non-Linear Boundary Value Problems," Rand Corporation Research Study, Elsevier Publishing Company, Amsterdam, 1965.
66. LATTES, R., and J. L. LIONS, "The Method of Quasi-Reversibility, Application to Partial Differential Equations" (trans. from French by R. Bellman), American Elsevier Publishing Company, New York, 1969.
67. INTERNATIONAL ATOMIC ENERGY AGENCY, Isotopes in Hydrology, Proceedings of a Symposium, Vienna, Austria, 1967.
68. HOTELLING, H., The Economics of Exhaustible Resources, *J. Pol. Econ.*, vol. 39, pp. 137–75, April 1931.
69. KELSO, M. M., The Stock Resource Value of Water, *J. Farm Econ.*, vol. 43, no. 5, pp. 1112–1129, December 1961.
70. CUMMINGS, R. G., Some Extensions of the Economic Theory of Exhaustible Resources, *Western Econ. J.*, vol. VII, no. 3, pp. 201–210, September 1969.
71. DUPNICK, E., Collective Utility in the Management of Natural Resource Use: A Systems Approach, Tech. Rept. No. 5, Department of Hydrology and Water Resources, University of Arizona, Tucson, June 1971.
72. DE NEUFVILLE, R., Cost-Effectiveness Analysis of Civil Engineering Systems: New York City's Primary Water Supply, *Oper. Res.*, vol. 18, no. 5, pp. 785–804, 1970.
73. TERENZIO, V. G., M. HAUPTMANN, and L. ERICSON, Discussion of Systems Analysis of Water Distribution Networks by R. de Neufville, J. Schaake, Jr., and J. H. Stafford, *J. Sanit. Eng. Div., Am. Soc. Civ. Eng.*, vol. 98(SA 5), pp.789–791, 1972.
74. KAZANOWSKI, A. D., A Standardized Approach to Cost-Effectiveness Evaluations, in J. M. English (ed.), "Cost-Effectiveness," pp. 113–150, John Wiley & Sons, Inc., New York, 1968.
75. DORFMAN, R., Operations Research, *Am. Econ. Rev.*, vol. 50, no. 4, pp. 575–623, 1960.
76. JOHNSEN, E., "Studies in Multiobjective Decision Models," Monograph No. 1, Economic Research Centers in Lund, Lund, Sweden, 1968.
77. ROY, B. "Problems and Methods with Multiple Objective Functions." North-Holland Publishing Company, Amsterdam, 1971.
78. BROWN, G., and R. DEACON, Economic Optimization of a Single-Cell Aquifer, *Water Resour. Res.*, vol. 8, no. 3, pp. 557–564, 1972.
80. BURT, O., Temporal Allocation of Groundwater, *Water Resour. Res.*, vol. 3, no. 1, pp. 45–56, 1967.

81. BURT, O., Groundwater Management under Quadratic Criterion Functions, *Water Resour. Res.*, vol. 3, no. 3, pp. 673–682, 1967.

82. CUMMINGS, R., and D. L. WINKLEMAN, Water Resource Management in Arid Environs, *Water Resour. Res.*, vol. 6, no. 6, pp. 1559–1568, 1970.

83. MARTIN, W. E., and R. A. YOUNG, The Need for Additional Water in the Arid Southwest: An Economist's Dissent, Unpublished Paper Presented at the Eighth Annual Meeting of the Western Regional Science Association, Newport Beach, California, February 1969.

84. DUCKSTEIN, L., and C. C. KISIEL, General Report on "Model Choice and Validation," pp. 1282–1308, Proceedings of the International Symposium on Uncertainties in Hydrologic and Water Resource Systems, University of Arizona, Tucson, December 1972.

85. BURAS, N., "Scientific Allocation of Water Resources," American Elsevier Publishing Company, Inc., New York, 1972.

86. FOGEL, M. M., L. DUCKSTEIN, and C. C. KISIEL, Optimum Control of Irrigation Water Applications, *Automatica*, vol. 10, no. 6, pp. 579–586, 1974.

87. DAVIS, D., C. C. KISIEL, and L. DUCKSTEIN, Bayesian Decision Theory Applied to Design in Hydrology, *Water Resour. Res.*, vol. 8, no. 1, pp. 33–41, 1972.

88. HOWE, C. W., and F. P. LINAWEAVER, JR., The Impact of Price on Residential Water Demand and Its Relation to System Design and Price Structure, *Water Resour. Res.*, vol. 3, no. 1, pp. 13–32, First Quarter, 1967.

89. HOWE, C. W., and B. BOWER, Policies for Efficient Regional Water Management, *J. Irrig. Drain. Div., Am. Soc. Civ. Eng.*, vol. 96, IR4, pp. 387–393, December 1970.

90. DUCKSTEIN, L., and C. C. KISIEL, Collective Utility: A Systems Approach to Water Pricing Policy, pp. 881–888, Proceedings of the International Symposium on Mathematical Models in Hydrology, Warsaw, Poland, July 1971.

91. KISIEL, C. C., L. DUCKSTEIN, and M. FOGEL, Stochastic Analysis of Ephemeral Flow in Aridland Channels, *J. Hydraul. Div., Am. Soc. Civ. Eng.*, vol. 97, no. HY 10, pp. 1699–1717, October 1971.

92. DUCKSTEIN, L., and C. C. KISIEL, Efficiency of Hydrologic Data Collection Systems: Role of Type I and II Errors, *Water Resour. Bull.*, vol. 7, pp. 592–604, 1971.

93. MITROFF, I. I., and F. BETZ, Dialectical Decision Theory: A Meta-Theory of Decision Making, *Manage. Sci.*, vol. 19, pp. 11–24, 1972.

94. BENJAMIN, J., and C. A. CORNELL, "Probability, Statistics and Decision for Civil Engineers," McGraw-Hill Book Company, New York, 1970.

95. YAKOWITZ, S., A Statistical Model for Daily Streamflow Records, with Application to the Rillito River, pp. 273–283, Proceedings of the International Symposium on Uncertainties in Hydrologic and Water Resource Systems, University of Arizona, Tucson, December 1972.

96. FELLER, W., "Introduction to Probability Theory and Its Applications," 3d ed., vol. 1, John Wiley & Sons, Inc., New York, 1966.

97. TODOROVIC, P., and E. ZELENHASIC, A Stochastic Model for Flood Analysis, *Water Resour. Res.*, vol. 6, no. 6, pp. 1641–1648, 1970.

98. FOGEL, M. M., C. C. KISIEL, and L. DUCKSTEIN, Space-Time Validation of a Rainfall Model for Summer-Type Precipitation, *Water Resour. Bull.*, vol. 7, no. 2, pp. 309–316, March 1971.

99. DUCKSTEIN, L., M. FOGEL, and C. C. KISIEL, A Stochastic Model of Runoff Producing Rainfall for Summer Type Storms. *Water Resour. Res.*, vol. 8, no. 3, pp. 327–350, 1972.

100. WHITSELL, W.J., Chemical Engineering, Municipal Ground Water Supplies and Public Health, Unpublished Paper Presented at the 71st National Meeting of the American Institute of Chemical Engineers, Dallas, Texas, February 1972.

101. WATER RESOURCES ENGINEERS, INC., "Final Report on Task III-3, Study Flow in Unsaturated Zone," Walnut Creek, Calif., March 1968.

102. TODD, D. K., "The Water Encyclopedia," Water Information Center, Port Washington, N.Y., 1970.

103. LESOURNE, J., "Le Calcul Economique," 2d ed., Dunod, Paris, 1972.

104. KO, S., and L. DUCKSTEIN, Collective Utility of Exchanging Treated Sewage Effluent for Irrigation and Mining Water, Proceedings of the 16th Annual Meeting of the Arizona Academy of Sciences, Prescott, Ariz., April 1972.

105. KO, S., and L. DUCKSTEIN, Cost-Effectiveness Analysis on Waste Water Reuses, *J. Sanit. Eng. Div., Am. Soc. Civ. Eng.*, vol. 98, no. SA6, pp. 869–881, December 1972.

106. BURAS, N., Conjunctive Operation of Dams and Aquifers, *J. Hydraul. Div., Am. Soc. Civ. Eng.*, vol. 89, no. HY6, pp. 111–131, 1963.

107. BURAS, N., and W. A. HALL, An Analysis of Reservoir Capacity Requirements for Conjunctive Use of Surface and Groundwater Storage, Publication No. 57, pp. 556–563, International Association of Scientific Hydrology, Gentbrugge, Belgium, 1961.

108. BURT, O., The Economics of Conjunctive Use of Ground and Surface Water, *Hilgardia*, vol. 36, no. 3, 1964.

109. BURAS, N., "Scientific Allocation of Water Resources," pp. 155–171, American Elsevier Publishing Company, Inc., New York, 1972.

110. EL RAMLY, N., Technological Uncertainity in the Economic Evaluation of New Water Resources, Proceedings of the International Symposium on Uncertainties in Hydrologic and Water Resource Systems, University of Arizona, Tucson, December 1972.

111. BOGARDI, I., L. DUCKSTEIN, and C. C. KISIEL, Distribution of Dynamic Water Level in a Shallow Lake, *Hydrol. Sci. Bull.*, vol. 18, no. 4, pp. 473–484, 1972.

112. DRACUP, J. A. Optimum Use of a Ground-Water and Surface-Water System: A Parametric Linear Programming Approach, Tech. Rept. 6–24, Hydraulics Laboratory, University of California, Berkeley, 1966.

113. KAZANOWSKI, A. D., Treatment of Some of the Uncertainties Encountered in Conduct of Hydrologic Cost-Effectiveness Evaluation, pp. 771–785, Proceedings of the International Symposium on Uncertainties in Hydrologic and Water Resources Systems, University of Arizona, Tucson, December 1972.

114. DAVID, L., and L. DUCKSTEIN, Multicriterion Ranking of Alternative Long-Range Water Resources Systems, *Water Resour. Bull.* (in press).

115. CHAEMSAITHONG, L., L. DUCKSTEIN, and C. KISIEL, Alternative Water Resource Systems in the Lower Mekong, *Hydraul. Div., Am. Soc. Civ. Eng.*, vol. 100, no. HY3, pp. 461–475, March 1974.
116. MAASS, A., et al., "Design of Water Resource Systems," Harvard University Press, Cambridge, Mass., 1962.
117. MAJOR, D. C., Benefit-Cost Ratios for Projects in Multiple Objective Investment Programs, *Water Resour. Res.*, vol. 5, no. 6, 1969.
118. HOWE, C. W., Benefit-Cost Analysis for Water Systems Planning, Water Resources Monograph 2, American Geophysical Union, Publication Progress, Inc., Baltimore, 1971.
119. MARTIN, W. E., T. BURDAK, and R. A. YOUNG, Projecting Hydrologic and Economic Relationships in Groundwater Basin Management, *Am. J. Agr. Econ.*, vol. 51, no. 5, 1969.
120. THOMPSON, P. D., Uncertainty of Initial States as a Factor in the Predictability of Large Scale Atmospheric Flow Patterns, *Tellus*, vol. 9, pp. 275–285, 1957.
121. GLEESON, T. A., Statistical-Dynamic Prediction, *J. Appl. Meteorol.*, vol. 9, no. 3, pp. 333–344, 1970.
122. EPSTEIN, E. S., The Role of Initial Uncertainties in Prediction, *J. Appl. Meteorol.*, vol. 8, pp. 180–198, 1969.
123. MCMILLAN, W. D., Theoretical Analysis of Groundwater Basin Operations, Water Resources Center Contribution No. 114, University of California, 1966.
124. FREEZE, R. A., "Regionalization of Hydrogeologic Parameters for Use in Mathematical Models of Groundwater Flow," Thomas J. Watson Research Center, IBM, Yorktown Heights, N.Y., 1971.
125. BIBBY, R., "Statistical Error Analysis of Groundwater Systems," doctoral dissertation, Colorado State University, Fort Collins, 1971.
126. SAATY, T. L., "Modern Non-Linear Equations," pp. 346–456, McGraw-Hill Book Company, New York, 1967.
127. BERAN, M. J., "Statistical Continuum Theories," Interscience Publishers, a division of John Wiley & Sons, Inc., New York, 1968.
128. COOPER, H. S., and C. E. JACOB, A Generalized Graphical Method for Evaluating Formation Constants and Summarizing Well-Field History, *Trans. Am. Geophys. Union*, vol. 27, no. 4, 1946.
129. NELSON, A. G., and C. D. BUSCH, Cost of Pumping Irrigation Water in Central Arizona, Tech. Bull. 182, Arizona Agricultural Experiment Station, University of Arizona, Tucson, 1967.
130. ADOMIAN, G., Random Operator Equations in Mathematical Physics, *J. Math. Phys.*, vol. II, no. 3, pp. 1069–1084, 1970.
131. CANCEILL, M., Rationalization of Piezometric Measurement Networks Using Multivariate Techniques, pp. 493–507, Proceedings of the International Symposium on Uncertainties in Hydrologic and Water Resources Systems, University of Arizona, Tucson, December 1972.

132. MOODY, D. W., and T. MADDOCK, A Planning Model for Preliminary Network Design, p. 448, Proceedings of the International Symposium on Uncertainties in Hydrologic and Water Resource Systems, University of Arizona, Tucson, December 1972.

133. SAGAR, B., S. YAKOWITZ, and L. DUCKSTEIN, A Direct Method for the Identification of the Parameters of Dynamic Nonhomogeneous Aquifers, *Water Resour. Res.*, vol. 11, no. 4, pp. 563–570, 1975.

134. NEUMAN, S. P., Analysis of Pumping Test Data from Anisotropic Unconfined Aquifers Considering Delayed Gravity Response, *Water Resour. Res.*, vol. 11, no. 2, pp. 329–342, 1975.