16558 HYDROLOGIC MODEL FOR DROUGHT MANAGEMENT

KEY WORDS: Distribution functions; **Droughts; Forecasting; Hydrologic data; Probability theory;** Risk; **Rivers;** Stochastic models; **Streamflow forecasting;** Time factors; Water flow

ABSTRACT: A method of producing long-range water-supply forecasts is developed and applied. The method is capable of producing conditional cumulative distribution functions (CDF) for monthly streamflow volumes and producing a set of equiprobable future streamflow traces. Using the latter output, it is possible to evaluate the conditional CDF of reservoir storage for any particular future month, or probabilities of storage falling below some level during the forecast period. The model used in an ARMA (1,1) cyclic model. It differs from an ARMA (1,1) stationary model in that the correlation structure varies with time of year. The model is fit in this analysis to the 81-year long stream-flow record of the Potomac River at Point of Rocks, Maryland.

REFERENCE: Hirsch, Robert M. (Hydro., Systems Analysis Group, Water Resources Division, U.S. Geological Survey, National Center, Reston, Va. 22092), "Stochastic Hydrologic Model for Drought Management," *Journal of the Water Resources Planning and Management Division,* ASCE, Vol. 107, No. WR2, **Proc. Paper 16558,** October, 1981, pp. 303-313

STOCHASTIC HYDROLOGIC MODEL FOR DROUGHT MANAGEMENT

By Robert M. Hirsch[1]

INTRODUCTION

Long-range water-supply forecasting is an integral part of drought management and of water supply management itself. The problems of long-range water-supply forecasting are quite different from those of short-range flow forecasting which is typically used for flood warning and hydropower scheduling. In short-range forecasting (where time horizons are a few hours to a few days), the meteorological events which are responsible for the flow of interest have already occurred or will occur within the time span for which meteorologists have substantial forecast skill. Consequently the problems are those of rapid data acquisition, and computation. The forecast output is appropriately a single trace of future flows (a hydrograph) with some error bands about it which are the result of input error, model error and the meteorological forecast uncertainty.

In contrast, the problems of long-range forecasts are those of understanding the current state of the watershed (streamflow and moisture storage in snow, soil, and aquifers) and incorporating that information with the very broad range of meteorological conditions (for which there is almost no forecasting skill) that may occur in the future. The output of a long-range forecasting model must take one of two possible forms: (1) Probability distributions for streamflow volumes during some forecast periods; or (2) a set of plausible streamflow traces covering the forecast period, each trace having an assigned probability. The presentation of a single "most likely" hydrograph is of no value in long-range forecasting. What is needed is not information on what is most likely but rather information on the likelihood (probabilities) of certain specified outcomes.

There are three broad classes of forecasting methods which are applicable to the long-range water-supply forecasting problem. The three classes are listed here in order of increasing difficulty of application. It should be noted here that little, if anything, is known about the relative appropriateness or forecasting skill of these three classes in comparison to each other.

[1]Hydro., Systems Analysis Group, Water Resources Division, United States Geological Survey, National Center, Reston, Va. 22092.

Note.—Discussion open until March 1, 1982. To extend the closing date one month, a written request must be filed with the Manager of Technical and Professional Publications ASCE. Manuscript was submitted for review for possible publication on July 16, 1980. This paper is part of the Journal of the Water Resources Planning and Management Division, Proceedings of the American Society of Civil Engineers, ©ASCE, Vol. 107, No. WR2, October, 1981, ISSN 0145-0743/81/0002-0303/$01.00.

1. Use historic streamflow traces for the appropriate time of year as a set of possible future streamflow traces, assigning probabilities to each. These probabilities must sum to 1.0 and they may be based on information about antecedent rainfall, soil or ground-water storage associated with each historical trace. The simplest approach is to make all historic traces equiprobable. These methods were applied by Hirsch (4).

2. Develop a stochastic model of streamflow based on the historic record. Initialize the model's state variables to present conditions and use a random number generator to produce multiple (equally likely) traces from this one initial condition. The simplest model is the autoprojective model, where past streamflow alone is then used to forecast future streamflow. This is the method that will be developed in the present paper. A more complicated form is to use past streamflow as well as current or past values of other related state variables (snowpack, soil moisture, ground-water levels). Certainly where snowmelt is a major contributor to streamflow, this multivariate approach would be preferred to univariate autoprojection.

3. Develop and calibrate a deterministic hydrologic model of the river basin where the model inputs are the meteorological inputs to the basin (rain, snow, temperature, humidity, wind), and the output is streamflow. Such models have a number of state variables representing storages in the basin (snowpack, aquifers, soil zones). By simulating the antecedent hydrologic conditions (for several months), a set of starting values for the state variables are determined and then a set of historic meteorological traces (with assigned probabilities) are used as inputs to the model in order to produce a set of streamflow output traces (with the same assigned probabilities). Such a procedure has been used by the National Weather Service (6) and is known as ESP (extended streamflow prediction).

LT-ARMA (1,1) Cyclic Flow Model

The remainder of this paper deals with a stochastic flow model which was first described by Hirsch (5). It is a model of monthly streamflow volumes and the application of it described here and in the previous paper is to the Potomac River at Point of Rocks, Maryland. The probability distribution of monthly volumes is modeled as a two-parameter lognormal distribution with a different mean and variance in each month. The serial dependence is modeled in a form closely related to the autoregressive moving average (ARMA) process described by Box and Jenkins (2). The model of streamflow volumes consists of Eqs. 1 and 2:

$$Z_i = \phi Z_{i-1} + e_i - \theta_j e_{i-1} \quad \ldots\ldots (1)$$

$$X_i = \exp\,[m_j + Z_i S_j] \quad \ldots\ldots (2)$$

The streamflow volume in month, i, is denoted X_i in which i may take on integer values (..., −2, −1, 0, 1, 2, ...); Z_i = a transformation of X_i; m_j and S_j^2 = the mean and variance respectively, of the natural log of streamflow volume in month, j, in which j may take on integer values (1, 2, ..., 11, 12) and j is related to i by $i = 12 \cdot g + j$ in which g = an integer. The variables, e_i, are independent and normally distributed with mean, 0, and variance β_j^2.

The Z_i are normally distributed with mean 0, variance 1, but are serially correlated. The total number of independent parameters is 37 (ϕ, 12 $\theta_j's$, 12 $m_j's$, and 12 $S_j's$, the 12 $\beta_j's$ are determined by ϕ and the 12 $\theta_j's$). The method of parameter estimation used for this model is given by Hirsch (5) under the heading of the "LT-ARMA" model. The important point to be noted about the estimates of the parameters of Eq. 1 is that the lag one serial correlation between the log transformed flow values for all 12 months are exactly preserved in the long run, and that serial correlation of log flow values at lags 2–12 are preserved in a least-squares sense. The serial correlation structure of Z_i may be derived from (1). Let $\rho_j(l)$ be the correlation between Z_i and Z_{i-l}, for $i = 12 \cdot g + j$:

$$\rho_j(l) = E\,[Z_i Z_{i-l}] \quad \ldots\ldots (3)$$

then $\rho_j(l) = \phi^{l-1}\,[\phi - \theta_{j+1-l}\,\beta_{j-l}^2]$ $l = 1, 2, \ldots$; and $j = 1, 2, \ldots, 11, 12$. It should be noted that where the index of any model parameter θ, β, m or S is outside the range (1, 2, ..., 11, 12), it is assumed to be recomputed by adding or subtracting integer multiples of 12 to bring it into the range (1, 2, ..., 11, 12). For example, $\rho_2(3) = \phi^2\,[\phi - \theta_{12}\beta_{11}^2]$.

Using related notation, the stationary ARMA (1,1) model is expressed

$$Z_i = \phi Z_{i-1} + e_i - \theta e_{i-1}$$

and $\rho(l) = \phi^{l-1}\,[\phi - \theta\beta^2]$ in which $\rho(l)$ = the lag l correlation coefficient; and β^2 = the variance of the e_i values.

Use of ARMA (1,1) Cyclic Model in Forecasting

The derivation of the forecasting method will begin with the problem of forecasting Z values and then proceed to the forecasting of X (streamflow volume). Given the model for Z (Eq. 1), assume that Z_{i-1} and e_{i-1} are known, then

$$E\,[Z_i | Z_{i-1}, e_{i-1}] = \phi Z_{i-1} - \theta_j e_{i-1} \quad \ldots\ldots (4a)$$

$$\text{Var}\,[Z_i | Z_{i-1}, e_{i-1}] = \beta_j^2 \quad \ldots\ldots (4b)$$

in which $E\,[Z_i|Z_{i-1}, e_{i-1}]$ represents the expectation of Z_i conditional on Z_{i-1} and e_{i-1} and $\text{Var}\,[Z_i|Z_{i-1}, e_{i-1}]$ is the conditional variance of Z_i given Z_{i-1} and e_{i-1}.

Define $\hat{Z}_i \equiv E\,[Z_i|Z_{i-1}, e_{i-1}]$, then $\hat{Z}_i = \phi Z_{i-1} - \theta_j e_{i-1}$ . . . . . (5)

The one-step ahead forecast error is $Z_i - \hat{Z}_i$. Subtracting Eq. 5 from Eq. 1 we have

$$Z_i - \hat{Z}_i = e_i \quad \ldots\ldots (6)$$

So e_i = the one-step ahead forecast error. Therefore, given the model, Z_{i-1} and e_{i-1}, are the necessary and sufficient information for making the forecast, $\hat{Z}_i$, and this forecast and the realization, Z_i, are necessary and sufficient information for determining e_i. It would appear that there is some difficulty in making the first forecast because only Z_1 would be known. In practice, an arbitrary choice of e_1 values can be made for the initial month of a historical time series and the computation of successive forecasts, $\hat{Z}_i$, and forecast errors,

e_i, can be carried out up to the present month. For any reasonable set of parameter values the choice of e_1 will make an insignificant contribution to the e_i value several months later.

Thus, by making predictions and computing errors in an interative fashion through the entire history of the time series $Z_1, \ldots Z_{i-2}, Z_{i-1}$ up to time $i - 1$, it is possible to "know" the values of e_{i-1} even though it is not a variable which is subject to direct measurement.

The conditional expectations and variances for forecasts more than one month ahead are

$$E[Z_{i+k}|Z_{i-1}, e_{i-1}] = \phi^k[\phi Z_{i-1} - \theta_j e_{i-1}] \quad \ldots\ldots\ldots\ldots\ldots\ldots (7)$$

$$\text{Var}[Z_{i+k}|Z_{i-1}, e_{i-1}] = \beta_{j+k}^2 + \sum_{h=1}^{k} \beta_{j+k-h}^2(\phi^h - \phi^{h-1}\theta_{j+k-h+1})^2 \quad \ldots (8)$$

for $k = 1, 2, 3, \ldots$.

Recalling that the purpose of the forecasting exercise is not to produce a most likely streamflow trace but rather to produce a risk assessment, then these forecasts must be translated into cumulative distribution functions (CDF). It can be shown that the conditional distribution of Z_i (given Z_{i-1}, e_{i-1}) is approximately normal with mean $\hat{Z}_i$ and variance β_j^2. Thus, knowing the relationship between X_i and Z_i (Eq. 2), one can specify completely the CDF for X_i. Let the unconditional CDF of X_i be denoted $F_{X_i}(x) = \text{Prob}(X_i \le x)$, and the conditional CDF of X_i be denoted $F_{X_i}(x|Z_{i-1}, e_{i-1}) = \text{Prob}(X_i \le x|Z_{i-1}, e_{i-1})$. Let the standard normal CDF be denoted $F_z(t)$. $F_z(t) = \int_{-\infty}^{t} 1/\sqrt{2\pi}\ e(-u^2/2)du$.

$$\text{Then} \quad F_{X_i}(x) = F_z\left(\frac{\ln x - m_j}{S_j}\right) \quad \ldots\ldots\ldots\ldots\ldots (9)$$

$$\text{and} \quad F_{X_i}(x|Z_{i-1}, e_{i-1}) = F_z\left(\frac{\dfrac{\ln x - m_j}{S_j} - \hat{Z}_i}{\beta_j}\right) \quad \ldots\ldots\ldots\ldots (10)$$

and more generally the conditional CDF for any number of time steps ahead is

$$F_{X_{i+k}}(x|Z_{i-1}, e_{i-1}) = F_z\left(\frac{\dfrac{\ln x - m_{j+k}}{S_{j+k}} - E[Z_{i+k}|Z_{i-1}, e_{i-1}]}{(\text{Var}[Z_{i+k}|Z_{i-1}, e_{i-1}])^{1/2}}\right) \quad \ldots\ldots (11)$$

The forecast of flows (x_i as opposed to Z_i) can be expressed in either of two ways. One is to give a median forecast $\hat{x}_i(\text{med})$ in which

$$\hat{x}_i(\text{med}) = \exp(m_j + S_j\hat{Z}_i) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

This forecast is the median of the conditional distribution, that is

$$\text{Prob}[X_i \le \hat{x}_i(\text{med})] = 0.50 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

The other is the expected value or mean forecast $\hat{x}_i(\text{mean})$ in which

$$\hat{x}_i(\text{mean}) = \exp\left[\frac{1}{2}\cdot S_j^2\beta_j^2 + m_j + S_j\hat{Z}_i\right] \quad \ldots\ldots\ldots\ldots\ldots (14)$$

This forecast is the mean of the conditional distribution.

The conditional standard error of these $\hat{x}_i(\text{mean})$ forecasts is

$$\text{SE}(\hat{x}_i) = \exp[m_j + S_j\hat{Z}_i]\sqrt{\lambda^2(\lambda^2 + 1)} \quad \ldots\ldots\ldots\ldots\ldots (15)$$

in which $\lambda^2 = \exp[S_j^2\beta_j^2] - 1$.

Note that λ is the conditional coefficient of variation of x_i and it is not dependent on the streamflow forecast itself. The conditional coefficient of skewness of x_i is γ_x in which $\gamma_x = \lambda^3 + 3\lambda$; thus, the skew does not depend on the flow forecast.

The other mode of application of the model is to simulate multiple equally likely streamflow traces. For simplicity of computation rewrite the generating Eqs. 1 & 2 as

$$Z_{i+k-1} = \phi Z_{i+k-2} + \beta_{j+k-1}\epsilon_{i+k-1} - \theta_{j+k-1}\beta_{j+k-2}\epsilon_{i+k-2} \quad \ldots\ldots\ldots (16)$$

$$X_{i+k-1} = \exp[m_{j+k-1} + Z_{i+k-1}S_{j+k-1}]; \quad k = 1, 2, \ldots, n \quad \ldots\ldots\ldots (17)$$

simulation of

storage

Examples

The model parameters were fit to the 81-yr monthly flow record for the Potomac River at Point of Rocks, Maryland. The full set of parameters and the sample values of $\rho_j(1)$ are given in Table 1. The 12 values of $\rho_j(1)$ are shown to demonstrate their wide month-to-month differences. In Monte Carlo experiments it was determined that such a wide variation is very unlikely to have occurred by chance if all 12 values $\rho_j(1)$ were equal. Noting the variation in β_j^2 values it is clear that the predictive ability of the model for month ahead forecasts is highly dependent on the month. For example, forecast variance for April is 0.91 but for November it is 0.56.

The explanatory power of this model is compared to several others in Table 2. These models all assume the same unconditional flow distribution (lognormal) with the same m_j and S_j parameters given in Table 1. The measure of the explanatory power is in all cases the following estimate of the error variance

$$1/960\sum_{i=13}^{972}\hat{e}_i^2$$

in which $\hat{e}_i$ = the estimated error term for month i. The reason that the first 12 values of $\hat{e}_i$ are not used is because of the arbitrary start-up rule (which was $\hat{e}_1 \equiv Z_1$). More complicated techniques involving hindcasting could have been used to prevent this "waste" of 12 values but given the wealth of data it was felt to be unnecessary. The following six models are defined:

ARMA (1,0) stationary $Z_i = \phi Z_{i-1} + \beta\epsilon_i$
ARMA (1,0) cyclic $Z_i = \phi_j Z_{i-1} + \beta_j \epsilon_i$
ARMA (1,1) stationary $Z_i = \phi Z_{i-1} + \beta\epsilon_i - \theta\beta\epsilon_{i-1}$
ARMA (1,1) cyclic $Z_i = \phi Z_{i-1} + \beta_j\epsilon_i - \theta_j\beta_{j-1}\epsilon_{i-1}$
ARMA (2,1) stationary $Z_i = \phi_1 Z_{i-1} + \phi_2 Z_{i-2} + \beta\epsilon_i - \theta\beta\epsilon_{i-1}$
ARMA (2,2) stationary $Z_i = \phi_1 Z_{i-1} + \phi_2 Z_{i-2} + \beta\epsilon_i - \theta_1\beta\epsilon_{i-1} - \theta_2\beta\epsilon_{i-2}$

TABLE 1.—Statistics and Model Parameters for ARMA (1,1) Cyclic Model of Potomac River at Point of Rocks, Maryland—Monthly Volumes

Month (1)	Index j (2)	Mean of log volumes m_j (3)	Standard deviation of log volumes S_j (4)	θ_j (5)	β_j (6)	Lag one correlation coefficient $\rho_j(1)$ (7)
October	1	5.50	0.838	0.499	0.905	0.357
November	2	5.63	0.746	0.040	0.745	0.667
December	3	6.15	0.776	0.085	0.756	0.653
January	4	6.57	0.633	0.067	0.749	0.661
February	5	6.74	0.602	0.734	0.886	0.288
March	6	7.16	0.483	0.584	0.940	0.242
April	7	6.95	0.511	0.505	0.954	0.254
May	8	6.64	0.576	0.455	0.949	0.286
June	9	6.20	0.614	0.484	0.954	0.264
July	10	5.67	0.558	0.128	0.811	0.589
August	11	5.54	0.672	0.423	0.884	0.421
September	12	5.29	0.672	0.187	0.829	0.554

TABLE 2.—Results of Fitting Six Alternative Stochastic Models to Potomac River Data. Numbers in Parentheses Indicate Ranks (1—best, 6—worst)

Model (1)	Error variance (2)	Number of ARMA parameters (3)	AIC (4)
ARMA (1,0) stationary	0.810 (6)	1	2,523.94 (6)
ARMA (1,0) cyclic	0.782 (2)	12	2,512.29 (5)
ARMA (1,1) stationary	0.792 (5)	2	2,504.01 (2)
ARMA (1,1) cyclic	0.752 (1)	13	2,476.61 (1)
ARMA (2,1) stationary	0.791 (4)	3	2,505.77 (3)
ARMA (2,2) stationary	0.791 (3)	4	2,507.28 (4)

in which ϵ_i = in all cases white noise. All of the stationary ARMA models were fit using maximum likelihood estimation, the ARMA (1,0) cyclic model was fit to assure long-run preservation of the sample values of $\rho_j(1)$ for all 12 months, and the ARMA (1,1) cyclic model was fit by the procedures previously mentioned.

One may argue that the relative size of the forecast variances are simply a consequence of the number of parameters used and, in fact, for these six models there is a perfect inversion of ranks between forecasts variance and number of ARMA parameters (lower variance corresponding to higher number of parameters). Hipel (3) suggests using the Akaike Information Criterion (AIC) (1) in model selection to incorporate the criterion of parsimony along with that of error minimization. They recommend selecting the model that gives the minimum value of the AIC defined by: AIC = -2 ln (maximum likelihood) + $2K$, in which K = the number of ARMA parameters. Using this criterion, rather than minimum variance alone, the ARMA (1,1) cyclic remains the preferred model, the rather parsimonious ARMA (1,1) stationary model moves from being the fifth best model to the second best. Undoubtedly, with a substantially shorter data set, the ARMA (1,1) stationary model may be preferred. The appropriate degree of model complexity remains a major problem in the construction of forecasting models.

Applications

Three different conditional CDF's (for three different antecedent conditions) are shown in Figs. 1, 2, and 3. All are historical cases in which the forecast is of the July volume based on volumes for the entire record up to and including June. Also shown on the figures is the unconditional July CDF. Table 3 gives a description of the three situations. In the first case there was a sustained period of very high flow through June—so the July forecast is quite high. The

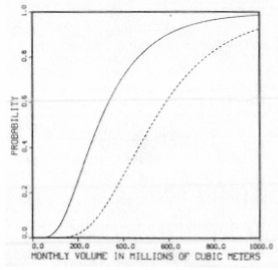

FIG. 1.—Unconditional CDF (Solid Line) and Conditional CDF (Dashed Line) for July Volumes for Potomac River at Point of Rocks, Maryland, when Previous May and June Volumes were 2,042 × 10^6 m^3 and 1,408 × 10^6 m^3, Respectively

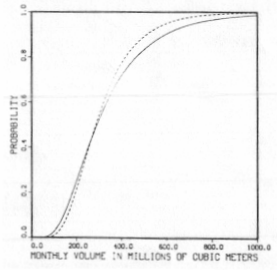

FIG. 2.—Unconditional CDF (Solid Line) and Conditional CDF (Dashed Line) for July Volumes for Potomac River at Point of Rocks, Maryland, when Previous May and June Volumes were 341 × 10^6 m^3 and 512 × 10^6 m^3, Respectively

conditional probability of a volume below 200 m^3 is only 0.017 as compared to the unconditional probability of 0.251.

In the second case the June volume was almost exactly the historic median value (Z_{t-1} is very near 0.0), however, the previous month's volume was quite low. If the forecasting model were ARMA (1,0) rather than ARMA (1,1), the

TABLE 3.—Examples of Three Forecasts for Potomac July Streamflow Volumes in 10^6 m^3

Z_{t-2} (1)	Z_{t-1} (2)	Z_t (3)	x_{t-2} (4)	x_{t-1} (5)	$\hat{x}_t$ (median) forecast (6)	$\hat{x}_t$ (mean) forecast (7)	Standard error of $\hat{x}_t$ (mean) (8)	Probability ($X_t < 200$) (9)
1.696	1.705	1.047	2,042	1,408	522	578	275	0.017
−1.414	0.060	−0.032	341	512	286	316	151	0.215
−0.614	−1.715	−0.993	540	172	167	185	88	0.654
Unconditional					291	340	205	0.251

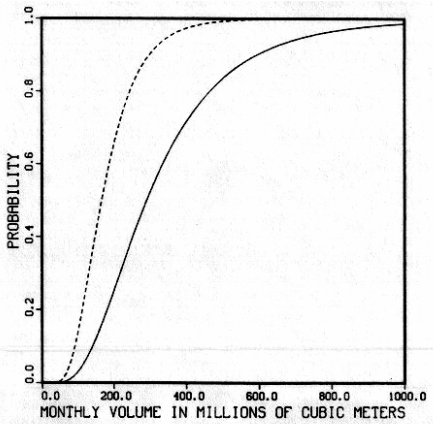

FIG. 3.—Unconditional CDF (Solid Line) and Conditional CDF (Dashed Line) for July Volumes for Potomac River at Point of Rocks, Maryland, when Previous May and June Volumes were 540 × 10^6 and 172 × 10^6 m^3, Respectively

forecast value $\hat{Z}_t$ would have been positive because Z_{t-1} was positive. The ARMA (1,1) formulation contains "memory" of the very low May volume and this more than compensates for the very slightly positive June value resulting in a small negative $\hat{Z}_t$ value. In the third case the two previous months had low volumes so the forecast is for a low volume.

The conditional standard error of the log volumes for July is 0.452 (= $S_J\beta_J$), independent of past flows, and the unconditional standard error for July is 0.558 (= S_J). Thus in log space the knowledge of past flows always reduces forecast error. In contrast, the conditional standard error of the flows themselves depends on the history of flows. When $\hat{Z}_t$ is high the standard error is higher than the unconditional standard error, but when $\hat{Z}_t$ is near zero or negative. the standard error may be a good deal below the unconditional.

If a water-supply system had storage only sufficient for regulating the delivery of water over periods of about a month, then these CDF's would be highly useful for making risk assessments. If the storage were only sufficient for a few days, then one would need a stochastic model with a shorter time step in order to evaluate the risk of shortage. If the storage capacity were sufficient to carry over a few months, then it would be appropriate to use this kind of forecasting model in a synthetic streamflow generating mode.

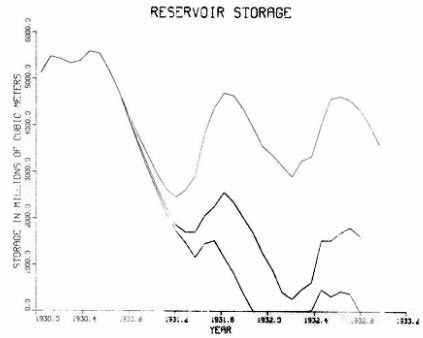

FIG. 4.—Example of Position Analysis for June, 1930, Showing Storage up to June, 1930, and 2% (Lower Curve), 10% (Middle Curve), and 50% (Upper Curve) Risk Storage Levels through September, 1932

Consider the following examples of a use of the ARMA (1,1) cyclic model of the Potomac flows. (The example is in no way related to the actual water supply configuration of the Potomac River.) There is a reservoir of 6 × 10^9 m^3 capacity located at the gage site, and the withdrawal rate for use is 20 × 10^6 m^3 per day throughout the year and evaporation and rainfall on the reservoir surface are ignored. The history of flows was used to simulate the history of reservoir storage from October 1896–June 1930 (1930 was a severe drought year). Knowing the Z_t and e_t values for June 1930, 100 realizations of streamflows through the remainder of water years 1930, 1931, and 1932 were generated using Eqs. 16 and 17. Using the end of June 1930 storage conditions for each realization, the continuity equation was used to develop 100 realizations of reservoir storages through the end of water year 1932. The 100 realizations of storage in any month may be viewed as a sample CDF. From these synthetic traces one may estimate the risk of storage falling below some specified level

in one particular month or over all months. For graphical purposes a few selected risk levels are traced out over the simulation period (Fig. 4). Given 100 realizations the 2% risk curve follows the second smallest storage values through all the months, the 10% follows the tenth smallest, and 50% follows the median. Note that any one risk level curve will not be identical with any one of the 100 traces but, rather, will jump from one to the other.

The results suggest that the probability of storage reaching zero is less then 2% through August of 1931 but it is never as great as 10% during the simulation period. Another simulation was run taking the starting time as September 1930 after three more months of continuing drought conditions (Fig. 5). The time for which the probability of zero storage exceeds 2% remains the same but the new estimate of zero storage probability now exceeds 10% for the period October–December of water year 1932.

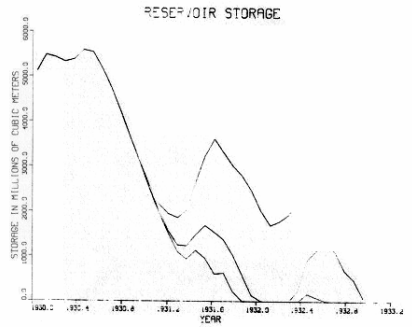

FIG. 5.—Example of Position Analysis for September, 1930 Showing Storage up to September, 1930, and 2% (Lower Curve), 10% (Middle Curve), and 50% (Upper Curve) Risk Storage Levels through September, 1932

Using such an analysis one can obtain, every month, an updated estimate of the conditional probability of storage falling below some level (such as zero) over some specified time horizon. In fact, given a large enough number of synthetic traces and a long enough time horizon, one can obtain results showing the gradual evolution from the conditional CDF of storage to the unconditional CDF of storage as the memory of the initial values of storage and flow die away.

Conclusions

With any of these tools for long-range water supply forecasting the ultimate aim is to provide information to improve the management of a water supply system. The power of the tools comes when they are used to provide information on the outcome of management options. For example, Figs. 4 or 5 could be re-created under a different withdrawal rate assumption. This kind of approach allows for the quantification of the risk side of the risk-cost tradeoff.

There is need for more investigation into the problems of parameter estimation for stochastic forecasting models. Finding the proper balance between preservation of detailed seasonally varying phenomena (flow correlation and CDF's) with the need for parsimony and the desire to account explicitly for the uncertainty in some 36 or more parameters remains a challenging task.

Appendix.—References

1. Akaike, H., "A New Look at the Statistical Model Identification," *Transactions Automat. Cont.*, IEEE, Vol. AC-19, No. 6, 1974, pp. 716–723.
2. Box, G. E. P., and Jenkins, G. M., *Time Series Analysis: Forecasting and Control*, Holden-Day Inc., San Francisco, Calif., 1970.
3. Hipel, K. W., McLeod, A. I., and Lennox, W. C., "Advances in Box-Jenkins Modeling, 1 Model Construction," *Water Resources Research*, Vol. 13, No. 3, June, 1977, pp. 567–575.
4. Hirsch, R. M., "Risk Analyses for a Water-Supply System—Occoquan Reservoir, Fairfax and Prince William Counties, Virginia," *Open-File Report 78-452*, United States Geological Survey, Reston, Va., 1978.
5. Hirsch, R. M., "Synthetic Hydrology and Water Supply Reliability," *Water Resources Research*, Vol. 15, No. 6, Dec., 1979, pp. 1603–1615.
6. Schaake, J., "Risk Analysis During the Occoquan Water Shortage of 1977," *Proceedings of the Thames/Potomac Seminars, ICPRB General Publication 79-2*, Interstate Commission on the Potomac River Basin, July, 1979, pp. 68–69.