Stochastic Hydrol. Hydraul. 1 (1987) 17–36

Stochastic Hydrology and Hydraulics

Stochastic models in hydrology

V. Yevjevich

Research Professor, School of Engineering and Applied Sciences, George Washington University, Washington, DC 20052, USA

Abstract: A stochastic approach to the analysis of hydrologic processes is defined along with a discussion of causes of tendency, periodicity and stochasticity in hydrologic series. Sources of temporal non-stationarity are described along with objectives and methods of analysis of processes and, in general, of information extraction from data. Transferred information as measured by correlation coefficients is compared with the transferable information as measured by entropy coefficients. Various multivariate approaches to hydrologic stochastic modeling are classified in light of complexities of spatial/temporal hydrologic processes. Alternatives of time series structural decomposition and modeling are compared. A special approach to modeling of space properties further contributes to approximate simulations of spatial/temporal processes over large areas. Several aspects of stochastic models in hydrology are concisely reviewed.

Key words: Stochastic hydrology, multivariate analysis, information extraction, information transfer, structure of time series, time series analysis, spatial characteristics, simulation of processes.

1 Determinism and stochasticity

Determinism and stochasticity constitute the two basic approaches to investigation of nature. Axioms of determinism are based on cause-effect relationships. Usually they are described by mathematical equations. Axioms of stochasticity lead to standpoints that relationships often cannot be expressed in simple or complex cause-effect mathematical forms. Instead, the "effect" variables are observed and their properties investigated by using methods of stochastic processes and mathematical statistics (Yevjevich 1974).

Figure 1 presents schematically three cases: (1) a pure deterministic relationship (left graph) as one extreme, (2) a pure stochastic case of cause-effect relationship (center graph) as the other extreme; and (3) transitions (right graph) between the two extremes. Ordinates of these graphs are partial effects on the resulting total "effect" variable by individual causal factors which are given on the abscissa. The left graph of Fig. 1 has a limited number of causal factors which jointly produce the full "effect" variable. It represents the classical case of a deterministic relationship. Errors in measurements act as additional factors, often as random noise. The center graph of Fig. 1 represents the case of effect being dependent on an infinite number of causes, each of them with an infinitesimally small partial effect. Here, no mathematical expression is feasible for a description of the cause-effect relationship. The effect is then conceived and investigated as a random variable.

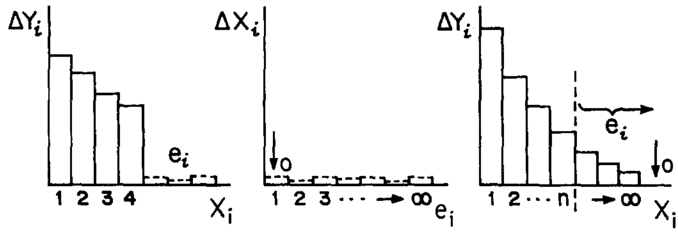

Figure 1. Cause-effect relationships: left graph, classical deterministic case; center graph, classical case of random variables; and right graph, transitional deterministic-stochastic case

The most current case of cause-effect relationship in the geophysical sciences is represented by the right graph of Fig. 1. The partial effects of a large number of causal factors are unequal, and assumed sorted in a descending order. A correlative association function takes into account the partial effects of a limited number of causal variables only. Partial effects of the large number of remaining causal factors are replaced by a random variable. It is composed of all neglected causal factors (which have not been identified, are not identifiable or are not economically observable), plus errors in measurement of variables included in the relationship.

A correlative equation is always composed of two parts: the mathematical regression equation between the $n+1$ variables and the random term

$$Y = f(X_1, X_2, \dots, X_n) + e \tag{1}$$

It may be conceived as a deterministic-stochastic analysis of cause-effect relationship. This case and the pure stochastic case (Fig. 1, center) are assumed here to represent stochastic processes and models in hydrology.

2 Causal factors which produce hydrologic space-time processes

Causal factors which produce continuous or intermittent hydrologic space-time processes as effects are essentially provided by three large sources. They are: (1) astronomic motions of bodies in the solar system acting through variation in input of solar energy to places on the Earth or through tides; (2) thermal processes and movements of fluids in the atmosphere, oceans and surface and subsurface continential environments; and (3) anthropogenic influences.

The first source of causal factors (Fig. 2, left side) basically produce: the short-range periodic solar energy inputs to places on the Earth's surface (with day and year as cycles), the lunar/solar almost-periodic tides, and the long-range almost-periodic astronomic movements (the three Milankovich cycles, the Earth/Moon orbital eccentricity, the precession of equinoxes and the obliquity of Earth's axis). The additional astronomic causal factors may come from the periodic-stochastic process of sunspot activity, which has an average cycle of 11.3 years but is subject to some random time-dependent fluctuation.

Causal factors of the Earth's environments, which affect hydrologic space-time processes (Fig. 2, center), include many thermal and fluid motion random processes on the Earth. The atmosphere is most important in creating randomness because it is a light non-conservative fluid. It not only generates basic randomness, but also acts in transmitting and modifying periodicities and almost-periodicities of solar energy inputs and other causes of periodicity to other environments. In turn these environments further contribute stochasticity and modify randomness and periodicities produced by the atmosphere.

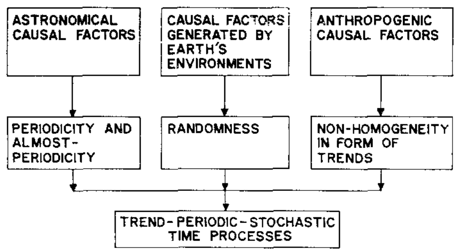

Figure 2. Three basic sources of causal factors of hydrologic relationships (upper line), with main impacts (center line) and resulting structure of time series (lower line)

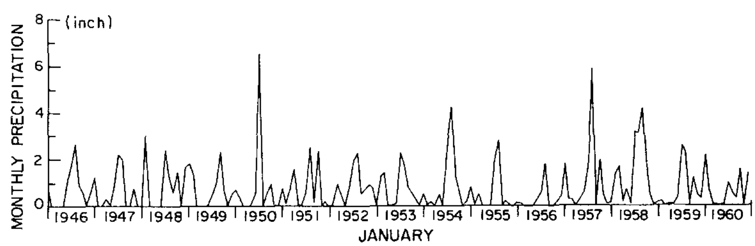

Figure 3. Sequence of monthly precipitation at Hicita, New Mexico, 1946 through 1960, which is composed of periodic mean, periodic standard deviation and an independent stochastic component

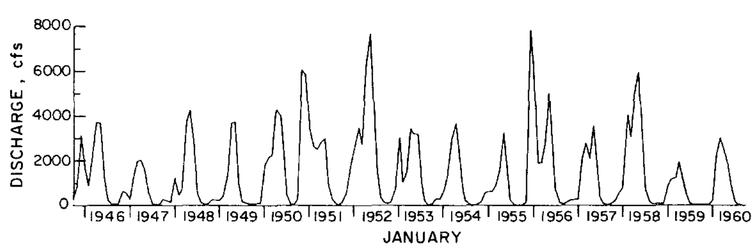

Figure 4. Sequence of monthly river flows for the Middle Fork of the American River near Auburn, California, 1946 through 1960, which is composed of periodic mean, periodic standard deviation and a dependent stochastic component

The resulting continuous or intermittent natural hydrologic time series are combined periodic-stochastic processes. They incorporate partial effects of many causal factors from these two distinct sources.

To illustrate the periodic-stochastic time series in a simplified way, a series of monthly values of precipitation (Fig. 3) and of runoff (Fig. 4) are presented. Figure 3 shows a time process with evidently periodic mean and periodic standard deviation, and a nearly independent stochastic component. Figure 4 shows evidently periodic mean and periodic standard deviation, but with a highly time dependent stochastic component. No trends seem to exist in these time series.

The third important source of causal factors in hydrology are various human activities (Fig. 2, right side). Hydrology is the geophysical discipline with the largest influence of human activities on its space-time processes. Activities in water resources development, conservation, control and protection are often expressed as water demand time series. Through water deliveries and returns, governed by these demands, the hydrologic processes are often radically changed, usually becoming a combination of natural hydrologic and water-use affected time series. Anthropogenic effects have completely changed entire river basins in the world, including not only the major quantitative water processes but also water quality processes.

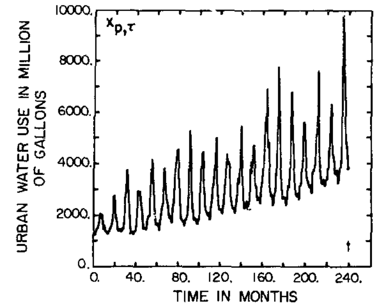

Figure 5. Monthly water use for Dallas, Texas, 1950-1969, with an approximate upward linear trend, periodicities and a stochastic component

These causal factors mainly create trends and slippages (jumps). They also modify existing trends and periodicity in basic parameters of the resulting hydrologic time series, change their space properties and/or modify the inherent stochasticity.

3 Characteristics of resulting processes

The resulting processes are trend-periodic-stochastic time series of hydrology. They are exemplified here by a simple water supply (delivery) series to a large city (Fig. 5). The combined series of natural flows and water-use return series are often very complex processes. Two general characteristics of observed space-time hydrologic processes are particularly important: (1) how the various causal factors and the effects of the three main sources mutually interact and how their effects propagate through the Earth's environments; and (2) the basic structure possessed by the resulting hydrologic time series.

Several questions either do not yet have answers or are only partially answered for the first general characteristic. Examples are: Do astronomical causal factors produce periodicity and almost-periodicity only in the basic parameters of time series, such as mean and standard deviation, or do they also affect series structure beyond the major parameters? Are all parameters periodic or only some of them? Why do the twelve average monthly values of the hydrologic variables often show a nearly sinusoidal pattern (within-the-year temperature and precipitation fluctuations in the Great Plains of the USA), fitted by the 12-month harmonic only, while other variables may need all six harmonics in the Fourier series description of their periodic processes (Fig. 6)? How are periodicities in hydrologic time series related to periodicities of atmospheric and oceanic variables and currents, or how do periodicities propagate through environments? What are the effects of various storage capacities for water and heat of the Earth on frequencies, amplitudes and phases of fitted harmonics to periodicity in parameters of hydrologic time series?

Similarly, questions may be raised regarding various aspects of randomness in hydrologic time series, though processes of heat transfer and transport, and of turbulence and vorticities at various space/time scales on the Earth are well studied and described. Such questions are: What are the major properties of stochastic dependence (persistence) on short, median and long range time scales? How is persistence best defined? Does persistence exist on all time scales? How is stochasticity transformed (attenuated or amplified) and its time dependence changed as water moves through various Earth's environments? How do periodicity and stochasticity interact? How is their significance best measured? What are the major sources and types of non-stationarity in hydrologic time processes?

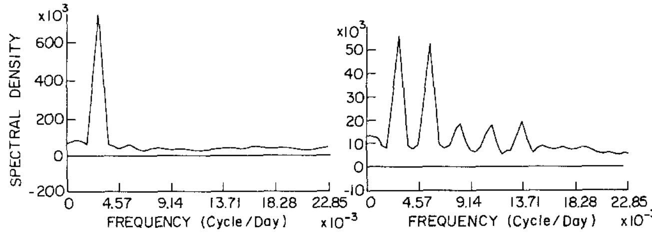

Figure 6. Spectral density graphs of daily river flows: left graph, the Greenbrier River near Alderson, West Virginia (USA); and right graph, the Jump River near Sheldon, Wisconsin (USA), for the period 1921-1960 (40 years) of observation

Versatile aspects of anthropogenic causal factors in hydrology have eluded professionals in the sense that there has not yet been a significant generalization of results and methods of their analysis. The complexity is often overwhelming. The lack of pertinent data is a limiting factor of analysis. Water resource developments have definitely introduced trends and slippages into hydrologic time series. Their effects on periodicity and stochasticity have been much less well investigated and generalized. Extensive studies of water use series in past decades show decisively that they are basically trend-periodic-stochastic time processes. The importance of each of the three properties depends highly on the character of human activity, the type of water use and their direct impacts.

An important factor in analysis of stochastic processes and models in hydrology is the precise definition of random variables. Most hydrologic variables are positive. Their values are greater than or equal to zero (such variables as precipitation. runoff, evaporation, ground water recharge, etc.). Some of them are intermittent if the probability of zero value is not negligible, or are not intermittent if that probability is negligible. The sources of causal factors and definition of random variables lead to the conclusion that the complex, general structure of many hydrologic time processes is composed of four basic properties of tendency, intermittency, periodicity and stochasticity (or the TIPS-structure). Various methods of investigation in stochastic hydrology depend on how these complex processes are conceived, approximated and decomposed in their structural analysis and mathematical modeling. In general and as an example, the three components of tendency, periodicity and stochasticity may be approximated by the following model of the general trend-periodic-stochastic process (Salas and Yevjevich 1972):

$$x_{p,\tau} = \underbrace{Tm_{p,\tau} + Ts_{p,\tau}}_{\text{Trend Component}}\{\underbrace{\mu_\tau + \sigma_\tau}_{\text{Periodic Component}}[\overbrace{\underbrace{\sum_{j=1}^{m} \alpha_{j,\tau-j}\varepsilon_{p,\tau-j} + \Big(1 - \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_{i,\tau-i}\alpha_{j,\tau-j}\rho_{|i-j|,\tau-k}\Big)^{1/2}\xi_{p,\tau}}_{\text{Second-order stationary Independent stochastic component}}}^{\text{dependence structure}}]\} \qquad (2)$$

with $x_{p,\tau}$ = random process under study, p = sequence of years,
τ = sequence of intervals within the year, as fractions of the year,
$Tm_{p,\tau}$ = trend in the mean, $Ts_{p,\tau}$ = trend in the standard deviation,
μ_τ = periodicity in the mean of detrended series,
σ_τ = periodicity in the standard deviation of detrended series,
$\varepsilon_{p,\tau-j}$ = stochastic component having periodic dependence coefficients and having trends and periodicities in the mean and standard deviation removed,
$\xi_{p,\tau}$ = independent, second-order stationary, stochastic component,
$\rho_{|i-j|,\tau-k}$ = periodic autoregressive dependence coefficients of $\varepsilon_{p,\tau}$, and
$\alpha_{|i-j|,\tau-k}$ = periodic autocorrelation coefficients of $\varepsilon_{p,\tau}$.

Similar or more complex models with the second or higher order stationarity may be designed and implemented.

4 Basic aspects of stochastic hydrology

Stochastic hydrology is conceived as that part of general hydrology which treats the random processes of hydrology and in which the random components should not be neglected. Of all problems encountered in stochastic hydrology, four groups are usually singled out as important in practice: (1) extraction of information from available data; (2) transfer of information from place to place or from variable to variable; (3) condensed description of processes for purposes of their simulation in the form of potential samples; and (4) forecasting of pending occurrence of hydrologic random variables by stochastic and/or deterministic/stochastic forecasting methods.

Stochastic modeling in hydrology is then conceived as the search for the set of mathematical expressions (one or more), with parameters estimated from data, which describe the processes in nature as closely to their true characteristics as the data and other information permit or is economically justified. Equation (2), with the additional equations for Tm, Ts, μ, σ, α and ρ and distributions of ξ, represents a set of equations for the model.

Stochastic modeling in hydrology is as old as the application of mathematics to geophysical disciplines. However, it was not called by that name. Fitting probability distribution functions to frequency curves or correlating hydrologic random variables were likely the first simple stochastic mathematical models used in hydrology. Not until the advent of digital computers were these original simple models exhausted as well as extended to more complex cases of modeling, such as of short-interval discrete time series or of space-time stochastic processes.

An important aspect of stochastic modeling is the problem of non-stationarity in hydrologic time series. Deterministic sources of non-stationarity are periodicities, almost-periodicities and trends and slippages in series parameters. They are relatively easy to treat by series decomposition (Eq. (2)). However, this usually leads only to second-order stationarity, if the variables leading to the stochastic component are not first normalized, leaving higher-order non-stationarity relatively difficult to attain.

An important case of random sources of non-stationarity is the occurrence of rare events conceived as disruptions in nature. Their effects may be sudden, slowly evolving or combined. Such events include earthquakes, large landslides, eruption of volcanos, large basin-wide fires, exceptional avalanches, basin-wide plant diseases, extremely rare floods or droughts, and similar occurrences. In this case the relatively homogeneous random noise in hydrologic processes on short time scales (days, months, years) is superposed by the random noise on long time scales (decades, centuries, millenia) in which the latter act as non-stationarity in the former. It is often nearly impossible to distinguish higher-order deterministic non-stationarity from the eventual effects of rare catastrophic random events.

Stochastic processes in hydrology have been studied from two standpoints: (1) utilitarian, with the purpose of applying results to various water resources and hydrologic problems; and (2) the human urge to understand and describe nature. The first standpoint has prevailed basically by developing models tested on observed data. Recently, the second standpoint has induced hydrologic studies of conceptual and physical nature, particularly through efforts to bridge the cause-effect deterministic type of information and the stochastic analysis, as well as by deriving forms of mathematical models from physical relationships which, in Klemeš' words (1986), is to give the hydrologic content to stochastic modeling. Simulation of samples of known, assumed or fitted stochastic models on computers by the Monte Carlo method, with the purpose of studying properties of complex stochastic processes, gives a further impetus to better understand and model hydrologic stochastic processes.

Stochastic modeling in hydrology is often the product of three distinct inputs: (1) knowledge from probability theory, stochastic processes and mathematical statistics, when axioms and hypotheses underlying the development of their theorems, methods and techniques fit the reality of hydrologic processes; (2) general geophysical knowledge related to hydrology which provides the physical background to stochastic modeling and includes knowledge of atmospheric, oceanic, geologic and other continental backgrounds and effects; and (3) knowledge of various physical processes and relationships of hydrology, which guides stochastic modeling and provides underlying axioms, hypotheses and approaches to modeling.

In general, hydrologic stochastic modeling is a synthesis of inputs rather than an analysis of any specific problem. It stimulates developments through feedbacks in sciences from which knowledge is derived. Contributions by hydrologic research to mathematical statistics and stochastic processes are already significant (theories of range, runs, water storage, etc.). They have stimulated scientists of other disciplines to study problems relevant to hydrology and to contribute to their solutions. Hydrologists in turn have contributed their share to these feedbacks. This symbiosis is evident not only between stochastic hydrology and statistics but also between general hydrology and various environmental and geophysical disciplines.

While scientists from disciplines other than hydrology have contributed to or affected the positive developments in hydrologic stochastic modeling, they have also passed on to hydrology some of their own biases and misconceptions. Simple examples are: (1) for years the search for "hidden periodicities" has claimed many cycles in hydrologic time processes, which could not be substantiated by the rigor of hydrologic tests on the best available data; and (2) often statistical approaches have been uncritically applied to hydrology, with the proper tests proving them not to be appropriate in hydrology.

5 Extraction of hydrologic information

Observed hydrologic data vary in overall reliability. Measuring methods produce random and systematic errors. Significant changes in river basins or in other environments induce non-homogeneities. Observed data are subject to sampling errors due to limited sample sizes. While these latter errors are taken into account by statistical inference techniques, the first three types need special and elaborate techniques for assessment and consideration. These actions are decisive since the accuracy of extracted information is no better than the accuracy of basic data. Furthermore, the study of non-homogeneities helps their extrapolation into the future.

Extraction of information on a random variable X involves several of its properties: (1) estimation of parameters and other characteristic values; (2) estimation of probability distribution functions; (3) modeling of processes in space and time; and (4) determination of properties of random variables derived from the basic processes. Advanced mathematical statistics offers already classical techniques for all the information extraction. The hydrologic reality enters into play basically in two areas of modeling: (1) criteria or approaches for selection of mathematical functions for models, especially probability distribution functions of variables and functions for description of space-time processes; and (2) direct determination of properties of random variables derived from a basic process, such as the model of most intense daily rainfall from the model of hourly precipitation series, or the model of flood peaks from the model of continuous runoff series.

In selecting mathematical forms of models, three approaches seem to have evolved with time:

1. Finding functions which best fit the data from a set of available functions (often from those traditionally used in the past), as the classical statistical approach;
2. Use of model functions which have been already found to best fit a large percentage of cases (say 95 percent or more), instead of making a new selection on each occasion by simple fitting criteria; and
3. Support of the functions selected by physical information, established relationships and analytical derivation based on already used functions for related processes. This third case is often conceived as using both deterministic and stochastic information in the selection, or the use of hydrologic physical knowledge in making the selection of stochastic models.

Simple fitting of functions by statistical techniques using only the available data is inherent in practice. However, many hydrologic cases provide additional information. If one finds that a particular function fits data of a large number of regional or worldwide cases well and fits much better than any other function, this information becomes valuable until changed by new investigations. Furthermore, physical information regarding the variables involved may lead to functions of models under some circumstances. An example is the use of a lognormal probability distribution function for the distribution of the size of sediment grains. This is related to the study of crashing rocks into small sand and gravel particles under some basic assumptions that fit reality (Kolmogorov 1941). Experience with the large number of sediment samples and the theoretical derivation often reinforce each other. The physical processes of transferring precipitation into runoff may often lead to conclusions concerning which models may be most appropriate for runoff variables and their time processes.

It is hypothesized here that integration of information contained in data with the other types of information for selecting hydrologic stochastic model functions is an easier task than derivation of model parameters from physical information. Exceptions are available. Often autocorrelation coefficients may be related to coefficients of equations which describe hydrologic runoff recession curves. The question is then which approach should be used, simple estimation of parameters from data or a combination of that information with information provided by the physics of the processes.

6 Transfer of information in hydrology

Variables observed at fixed points are determined by design of hydrologic networks or by other criteria. Information often is needed: (1) at points others than those of gauging stations, (2) for variables that are related to variables observed, and (3) for ranges of variables outside of their historic observed ranges. These cases then require transfer of information from point to point for a given variable or from variable to variable, extrapolation of information to ungauged ranges of observed variables and combinations of the three cases. Information transfer requires simultaneous observations for a minimum length of time at both or several points for the same variable in case (1) or for two or more variables in case (2) in order to establish the information transfer relationships. For case (3) a relationship is needed (probability function, correlative function, etc.) for purposes of extrapolation. For case (4) complex relationships may be required. Basically, regression techniques are used for cases (1) and (2) with correlation coefficients measuring the degree of association, and fitting procedure of given goodness-of-fit measure for case (3).

In relating variables by using their trend-periodic-stochastic time series, a question seems pertinent, namely how each property of these processes affects the degree of correlative association. The answer is given through an example, Fig. 7 and 8, for the case of correlating 10 years of mean daily flows of the Esencay River in Southwest Turkey (Harmancioglu and Yevjevich 1968) for its two gauging stations, Orenkoy (upstream) and Yapilar (downstream). Water from the upstream station passes the downstream station (defined as "throughflow") together with the intermediate inflow between the two stations. Squares of the correlation coefficient (r^2) for linear correlation (1), and linear correlation of logarithms of mean daily flows (2) are presented in Fig. 7 for five cases of decomposed time series: original data (OR), detrended series in the mean (DT), detrended and deseasonalized series in the mean (DS-M), detrended in the mean and deseasonalized series in the mean and standard deviation (DS-MS), and independent stochastic series with stochastic dependence removed (IND). Regardless of throughflow increasing correlation between station series, the removal from the series of the periodic mean, periodic standard deviation and time dependence in the stochastic component significantly decreased r^2, from 0.88 to 0.60 for the non-linear case and from 0.72 to 0.45 for the linear case. The greatest decrease was achieved by removing the periodic standard deviation in the linear case.

Figure 8 shows the same graphs as in Fig. 7 except for the correlation of the mean daily flows at the upstream station only with the intermediate flow between the two stations (obtained as the downstream station flows minus the concurrent upstream station flows). This figure shows the same pattern of effects of series components on the square of correlation coefficient except that they are smaller in comparison with those of Fig. 7 because of the removed throughflow from downstream station series. Figures 7 and 8 show then three basic properties: (1) correlation with throughflow retained in downstream time series is much greater than without it for all five cases of decomposed series; (2) non-linear correlation (or linear correlation of logarithms of mean daily flows) is stronger than linear correlation; (3) major sources of correlative association result from parallelism in periodic parameters and in stochastic dependence of correlated series of mean daily flows; and (4) correlating only independent stochastic components often does not reach the limit values of r^2 for the transfer of information.

Not only do the concurrent values of the correlated series contribute to the degree of correlation, but the non-zero lag cross-correlation may also be a significant contributor. Figure 9 illustrates this case for the same example as used in Figs. 7 and 8. It gives the square of multiple correlation coefficient versus the largest lag j of the multiple linear lag cross-correlation between the mean daily flows of the two gauging stations (Yapilar, Orenkoy) of the Esencay River. Six cases are: (1) linear correlation of original data (with throughflow), (2) linear correlation of original data with throughflow removed from the downstream station series; (3) linear correlation of independent stochastic component without throughflow; (4) same as in (3) but with throughflow removed in the Yapilar series; (5) same as for (1) but for the multiple non-linear (linear case between logarithms) lag cross-correlation; and (6) same as for (2) but for the multiple non-linear case of lag cross-correlation between the logarithms of values. Figure 9 demonstrates clearly that non-zero lag cross-correlation terms in the multiple regression equation significantly increased the transferred information. Though this figure relates to two stations along the same river, the non-zero lag cross-correlation should be investigated in most cases for an increase in transferable and in transferred information.

The transferable information contained in data, which can be maximally transferred, seems to be best measured by the square of entropy coefficient. Entropy of a discrete random variable with N elementary events of probability $P_n = p(X_n)$, $n = 1,2,\ldots,N$ is defined by (Shannon and Weaver 1949)

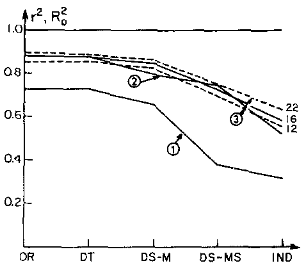

Figure 7

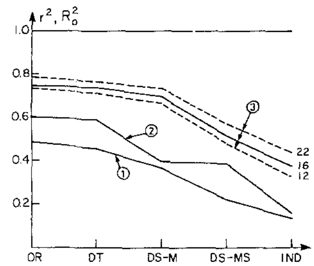

Figure 8

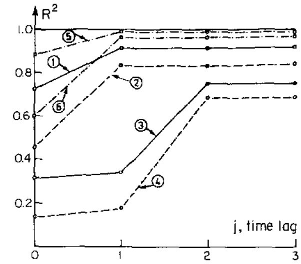

Figure 9

Figure 7. Effects of various decompositions of series on correlation (r^2) for linear, line (1), and non-linear, line (2), correlation cases, with entropy coefficients (R^2) measuring the transferable information between mean daily flows of Yapilar and Orenkoy gauging stations of the Esencay River (Southwest Turkey), for three different numbers of class intervals.

Figure 8. Effects of various decompositions of series on correlation (r^2) for linear, line (1), and non-linear, line (2), correlation cases (without throughflow in downstream station series), with entropy coefficients (R_o^2) measuring the transferable information between mean daily flows of Yapilar and Orenkoy gauging stations of the Esencay River (Southwest Turkey), for three different numbers of class intervals.

Figure 9. Multiple lag cross-correlation coefficients (R^2) between the mean daily flows of Yapilar and Orenkoy gauging stations of the Esencay River (Southwest Turkey): (1) Original data with throughflows; (2) Original data without throughflow in Yapilar series; (3) Independent stochastic component of series with throughflow unremoved in Yapilar series; (4) same as in (3) with throughflow removed in Yapilar series; (5) same as (1) but for non-linear case; and (6) same as in (2) but for non-linear case (or linear case between logarithms of series values)

$$H(X) = K \sum_{n=1}^{N} p(X_n) \log \frac{1}{p(X_n)}; \quad K = 1. \tag{3}$$

The entropy of two independent variables X and Y is $H(X,Y) = H(X) + H(Y)$, and of two dependent variables is

$$H(X,Y) = K \sum_{n=1}^{N} \sum_{m=1}^{N} p(X_n,Y_m) \log \frac{1}{p(X_n,Y_m)}; \quad K = 1, \tag{4}$$

with $T(X,Y) = H(X) + H(Y) - H(X,Y)$ defined as transinformation (information repeated in both X and Y). For X and Y independent, $T(X,Y) = 0$.

For continuous density functions, $p(X)$ is approximated by $f(X)\Delta X$, for small ΔX, with the entropy defined by

$$H(X;\Delta X) = \int_{-\infty}^{+\infty} f(X) \log \frac{1}{f(X)}\, dX + \log \frac{1}{\Delta X} \tag{5}$$

and the joint entropy of X and Y by

$$H(X,Y;\Delta X,\Delta Y) = \int_{-\infty}^{+\infty} f(X,Y) \log \frac{1}{f(X,Y)}\, dxdy + \log \frac{1}{\Delta X \Delta Y}, \tag{6}$$

with both entropies dependent on the selected class interval ΔX, or ΔX and ΔY, respectively.

The entropy information coefficient R_o is defined by (Linfoot 1957)

$$R_o = (1 - e^{-2T_o})^{1/2}, \tag{7}$$

where R_o is a dimensionless measure of dependence between variables, and T_o is the upper limit of transferable information (mutual information between variables). This does not assume variables to be normal (Rajski 1961) as classical correlation does.

The transferred information T_1 by regression is related to the square of correlation coefficient as

$$T_1 = [\ln(1 - r^2)]/2. \tag{8}$$

By comparing T_o and T_1, a new dimensionless measure

$$t_i = (T_o - T_1)/T_o, \tag{9}$$

indicates how much information a regression equation leaves untransferred out of the total transferable information.

In the bivariate case, r^2 shows how much of information is transferred out of the total transferable information, measured by R_o^2 (or by T_0, or t_i, whichever is appropriate). In the multivariate regression case (including the case of non-zero lag cross multiple regression), the multiple correlation coefficient, R, is compared with R_o (Harmancioglu et al. 1986). A drawback is that R_o depends on the size of the class interval and therefore on the selection of the number of class intervals in the case of continuous density functions (Amorocho and Espildora 1973). This difficulty is best resolved by using several numbers of class intervals for a given frequency distribution, showing a relatively narrow band of variation of R_o in comparison with the difference $R_o - r$, or $R_o - R$, respectively for bivariate and multivariate case.

Figures 7 and 8 show also examples of comparison of transferred and transferable information for the case of the Esencay River. Three numbers of class intervals are used in these figures for X and Y variables: 12, 16, and 22, with R_o values being in a relatively narrow band. In cases of non-linear correlation of mean daily flows between gauging stations (Fig. 7), the transferred information (2) nearly coincides with the transferable information (3) of an average of the three curves. With or without throughflow in the downstream station series and for linear cases the transferred information in Figs. 7 and 8 is much smaller than the transferable information.

The above presentation shows that tools are available at present to measure, at least approximately, whether the method used in information transfer extracts none, some or most of the transferable information. This can then guide investigators to search for those regression functions and/or for variables to be correlated which minimize the parameter t_i of Eq. (9). Using entropy coefficients as the measure of transferable information and correlation coefficients as the measure of transferred information definitely deserves further analysis and development.

7 General properties of hydrologic stochastic modeling

Physical, spatial/temporal, hydrologic natural processes are continuous in space and time (each point in space has either a finite value or a zero value, at each time instant). Their mathematical descriptions are generally approximated by three types of multivariate processes: (1) as a set $X_1, X_2,...,X_k$ of k well-defined random variables (such set as precipitation, evaporation, infiltration, runoff and other variables) for a point in finite space and of finite time length, basically for studying their relationships; (2) as a set $Y_1, Y_2, \ldots, Y_m$ of each variable X, for m discrete in space, which approximate continuous spatial process of one or more X variables, basically for studying space properties as a multipoint spatial multivariate process; and (3) as a set $Z_1, Z_2, \ldots, Z_n$ for each X variable for n discrete points or intervals in time, either along a cycle or along the entire sample of data, which approximates the continuous temporal process (including zero values), basically for studying time properties as a multipoint temporal multivariate process. In this third case an extension is made so that the multivariate process implies that the size n (either cycle or sample) can be repeatedly observed in form of many sets of n values, with each datum of the cycle or sample, $i = 1,2,\ldots,n$, having outcomes of the marginal random variable.

As the right graph of Fig. 1 demonstrates, the larger k is in studying the stochastic processes of hydrology, the closer the mathematical description is to a true cause-effect relationship. Similarly, the larger the selected numbers m and n of discrete points of space and time, for limited space and given time, the more accurate is the description of a process. The parameter $p = kmn$, with $kmn > 0$, determines the accuracy of approximation of nature. It also measures the cost of providing and processing the necessary data in modeling hydrologic stochastic processes. For $k = 1, m = 1$, and $p = n$, this is the simple temporal point discrete process. This case is then studied as a time series. For $k = 1, n = 1$, and $p = m$, this is the simple spatial or lines, or area, process for give time point or interval. The case is studied as synoptic maps of times $t = 1,2,\ldots$ for selected variables. For $m = 1, n = 1$, and $p = k$, this is the case of space/time point values of a set of k distinct variables. By combining various numbers of k, m and n, all practical study cases of hydrologic stochastic processes are covered.

The question of why the mathematical modeling is needed should be often re-examined. One reason is utilitarian, namely to better solve practical problems. Another reason is that mathematics is the stenography of description of natural and human-affected processes, whenever their complexities permit such an approach. In the general human urge to know and describe, whether or not there is any utilitarian effect from it, mathematical modeling is the standard approach used in description. For larger values of k, m and n, one is usually searching for that knowledge. For the cases where some of these three determinants are small, utilitarian objectives most often prevail.

In general, modeling of stochastic processes of hydrology requires estimation of the model of multipoint spatial/temporal multivariate process. For the three "marginal multivariates",

$$F_x(X_1, X_2, \ldots, X_k) = 0 \tag{10}$$

represents the search for hydrologic relationships and marginal, conditional or joint probability distributions;

$$F_y(Y_1, Y_2, \ldots, Y_m) = 0 \tag{11}$$

for any X_i enables the study of spatial dependence and spatial variation with coordinates x, y, z, of parameters of X; and

$$F_z(Z_1, Z_2, \ldots, Z_n) = 0 \tag{12}$$

for any X_i at a space point (x_o, y_o, z_o) enables the study of temporal structure and dependence, as well as the type and degree of non-stationarities.

The study of non-zero space-lag and non-zero time-lag relationships among variables in hydrology is important. A variable may be related to upwind or upstream non-zero spatial and temporal lag values apart from being related to values of zero lag at the time point and the same time. Furthermore, one should account for the TIPS-structure of complex time series of hydrology (tendency-intermittency-periodicity-stochasticity). The three cases of Eqs. (10)-(12), two cases of non-zero lag relationship, and four cases of structural time properties, help to describe the complexities of hydrologic stochastic processes.

History of contemporaneous stochastic modeling in hydrology is likely tied either to the individual analysis and description of the above $3 \times 2 \times 4 = 24$ cases of single modeling problems, or to the joint analysis and description of two or more of these 24 cases. Most hydrologic variables are mutually related, however, individual variables are also spatially and temporally dependent. This three-pronged dependence is likely the main reason why complex multipoint spatial/temporal multivariate processes should be investigated in depth, but also it produces complexities which make this joint investigative approach extremely difficult. Therefore, modeling of stochastic processes in hydrology requires the symbiosis of analysis of well-defined individual problems and synthesis of their properties and results into the general space/time multivariate models.

Purposes of modeling vary widely. In general, if there is a pool of data, the corresponding model should be an extraction of all or most information contained in it. In other words, if samples are generated by the model (and the Monte Carlo method), the new samples should be as close to the true but unknown population as the inferred population from the original data permits. The implication is that a reliable model can replace data in applications. In practice, this is not yet the case for two reasons: (1) models are often considered only as an idealized approximation, not accepted as fully equivalent to the information contained in data; and (2) there is an unfortunate but prevailing attitude that samples generated by the model should closely reproduce nearly all the parameters inferred from the original data, even those parameters which have very large sampling variations; this decreases confidence in the results of modeling and simulation. Modelers may be partially responsible for these two attitudes. Often, processes are modeled by fits before their structure and physical interactions and relationships are well-understood.

In general, models are useful in helping solutions of hydrologic and water resource problems, for which planners or designers must look at the long time horizons (say, large overyear storage capacities). In this case, historic samples give only one or a small number of solutions for a design random variable, therefore they may be unreliable for decision making and lead to overdesign or underdesign. Simulation of stochastic processes on computers is in some way equivalent to physical experimental laboratories. The simulated samples provide data points like a string of experimental runs in physical laboratories. Thus, one is able to study properties of processes with assumed stochastic models, and to test various related hypotheses.

8 Time series analysis

In the TIPS-structural analysis (Yevjevich 1984) and modeling of hydrologic time series, various approaches may be used:

1. Tendency is basically assumed in parameters, usually in the mean T_m and standard deviation T_s of the variable X, removed simply by (using Eq. (2)):

$$U_{p,\tau} = (X_{p,\tau} - Tm_{p,\tau})/Ts_{p,\tau}, \quad (13)$$

with $U_{p,\tau}$ considered as trendless series (regardless that trends may exist in other parameters).

2. Intermittency is treated in one of the three ways: (i) as point processes, YES-NO, and then modeled within the YES-process with the variable distribution and time structure as the periodic-stochastic process (Chin 1977; Roldan and Woolhiser 1982; Waymire and Gupta 1981; etc.); (ii) as the spell (storm) processess, with each event (cell, spell, storm, ...) having random characteristics of time, volume, shape, etc. (Todorovic and Yevjevich, 1969); and (iii) as inferred non-truncated positive-negative process, with its model estimated from data on the positive but truncated process at $X = 0$ (Kelman 1977; Richardson 1977).

3. Periodicity of hydrologic series is treated in two ways: (i) for each cycle (say day, year) there is a given number w of intervals or points in a discrete series, with values of each interval considered as a marginal random variable out of w temporal multivariates of periodicity (say $w = 12$ for monthly, $w = 52$ for weekly and $w = 365$ for daily series, in case of the annual cycle), with periodic parameters estimated from data by a non-parametric approach (not by the fit of periodic functions); and (ii) basic periodic parameters are modeled in mathematical forms (Fourier series fits), then removed by leaving the stochastic process with constant (or even some periodic) parameters, with such removal leading to

$$\varepsilon_{p,\tau} = (U_{p,\tau} - \mu_\tau)/\sigma_\tau, \quad (14)$$

with $U_{p,\tau}$ that of Eq. (13) and μ_τ and σ_τ the mathematical Fourier series models of periodic mean and periodic standard deviation of $U_{p,\tau}$. When feasible, Equation (14) is adjusted to produce a stochastic process and is studied as such.

4. Stochasticity is studied also in two ways: (i) similar to the first approach of treating periodiocity, namely stochasticity is studied as marginal random variables of $\varepsilon_{p,\tau}$ and $\xi_{p,\tau}$ of Eq. (2); and (ii) similar to the second approach of treating periodicity, with the unique time stochastic (preferable stationary) process (with or without periodic parameters) studied as the $\varepsilon_{p,\tau}$ - process of Eq. (2).

Each of the four structural TIPS-characteristics requires a corresponding number of parameters of tendency, intermittency, periodicity and stochasticity, with the total number:

$$a = a_t + a_i + a_p + a_s \quad (15)$$

While methods are available to infer the most parsimonious number of model parameters for stationary stochastic processes (Akaike 1974), there is no method available for optimizing the sum of Eq. (15). Inclination tends to the use of too many parameters. For a weekly hydrologic series the sum $a_t + a_i + a_p$ may be very large. For periodicity only, the mean, standard deviation, skewness coefficient and two autocorrelation coefficients of each week of the year account for 260 parameters. For daily series it becomes 1825 parameters. Therefore, the second method of analysis, fitting periodic functions with a small but necessary number of harmonics to periodic parameters becomes necessary in order to avoid the avalanche of parameters required using the non-functional approach to description of periodicity.

In the first approach using the w marginal variables, the resulting stochastic variables $\varepsilon_{p,\tau} = (U_{p,\tau} - m_\tau)/s_\tau$, with m_τ and s_τ the estimates of the mean and

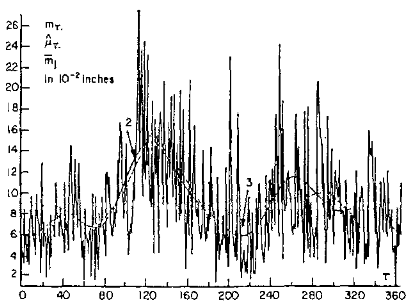

Figure 10

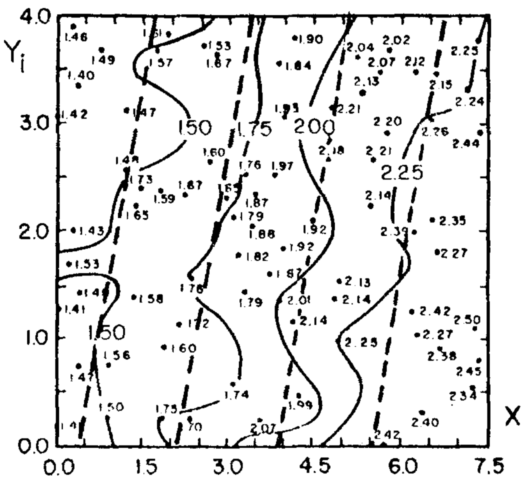

Figure 11

Figure 10. Periodicity of daily precipitation means of the Austin (Texas) Station: (1) 365 estimates of means, m_τ; (2) Fitted periodic function μ_τ by the Fourier series to the 365 values of m_τ, with five significant harmonics; and (3) Daily means, m_j, averaged over the 28-day intervals (13 average values)

Figure 11. Isolines (full lines) and fitted plane (dashed lines) for the general monthly mean $\bar{x}$ for 40 years of data in the Northern Great Planes of the United States

standard derivation of the $U_{p,\tau}$ marginal variables, are all standardized ($\bar{\varepsilon} = 0$, var $\bar{\varepsilon} = 1$), and all the independent components $\xi_{p,\tau}$ of Eq. (2) will have consecutive pairwise zero autocorrelation coefficients. In the second approach, with the Fourier series modeling of periodic parameters, $\varepsilon_{p,\tau}$ of Eq. (14), the resulting $\xi_{p,\tau}$ independent components of Eq. (2) are standardized as a series, with zero autocorrelation coefficient for the entire sample only, and not for each consecutive pair of marginal variables. In this approach periodicity as a deterministic property is separated from the remaining stochastic stationary or non-stationary process.

Figure 10 demonstrates the difference between the two approaches visually in an extreme case of $w = 365$ for daily precipitation process. It is clear that the use of non-functional methods of treating periodicity with w values of each m_τ and s_τ in removing periodicity in parameters, also removes a large random sampling variation in the means, and similarly in the standard deviations, from the resulting stochastic components. A question to ask is whether and what kind of distortion is being introduced in the generated samples, and if the generated $\varepsilon_{p,\tau}$ values are located around the extreme mean values of Fig. 10? This distortion decreases with a decrease of w.

9 Analysis of spatial characteristics

Spatial distributions of model parameters and other relationships of characteristics are usually used to describe spatial properties of hydrologic processes. A simple example is given here for the monthly precipitation space/time process of the Northern Great Plains of the United States (latitude 43.75°N - to 47.75°N, longitude 92.50°W - 100.00°W), with 77 stations of monthly precipitation series of 40 years of record used (Yevjevich and Karplus 1973).

Basic parameters (mean, standard deviation, amplitude and phase of harmonics of the mean and standard deviation) are areally studied by drawing their isolines. Figure 11 gives the case of the areal variation of the overall monthly mean precipitation. It shows an increase in values from West to East, and a slight increase from South to North. A fit of the plane $\mu_i = a + bX_i + cY_i$, with X_i and Y_i the latitude and longitude, through 77 point values of $\overline{m}_\tau$ gives then the fitted function of the overall mean monthly precipitation μ_i. Because flat areas predominate in landscape, the fit of a plane in Fig. 11 looks good. However, the isolines are less regular in case of mountains, and fits of functions are complex.

Monthly precipitation in the above example appears to be composed only of the 12-month harmonic (pure sine or cosine periodic mean) and an independent stochastic component, as the spectrum of all monthly values of a station demonstrates in Fig. 12. One-harmonic Fourier series functions are fitted to parameters, with their ampitudes and phases also regionalized by isolines and isolines fitted by functions of X_i, Y_i. The resulting random processes then become independent stochastic components. Frequency distribution curves of these time independent components are often indistinguishable among themselves in a homogeneous precipitation region, such as the Great Plains. These components are spatially dependent. Figure 13 shows the relationship of the simple linear bivariate correlation coefficient of pairs of station series to the distance between stations for the region of Fig. 11. There is also a slight effect of the azimuth of station-connected straight lines on the correlation coefficient. A function $r = f(d)$ is fitted in Fig. 13, with its confidence limits at the 95% probability level also drawn. They show that spatial dependence can be mathematically modeled.

With Fourier functions for periodic parameters, regional functions for parameters of fitted Fourier functions, probability distribution functions for independent stochastic components (only one or several, depending on homogeneity of precipitation over the region), and the $r = f(d)$ function (only with distance d, or with both distance and azimuth), the model may be used for simulation of spatial characteristics of the process. This model then enables: (1) generation of spatial/temporal samples at a grid of stations, namely at other positions than those of gauging stations; (2) selection of a desired density of stations (each station assumed to be at the center of a regular network area); (3) application of principal components into the m spatially independent stochastic components; (4) preservation of spatial variation of parameters; (5) preservation of spatial dependence; and (6) preservation of basic periodic characteristics of the temporal process.

10 Forecasting hydrologic variables

A special area of stochastic modeling is the hydrologic forecast. The usual deterministic methods of forecast show that various errors in forecast variables are spatially and temporally dependent. Therefore, they still contain information for improving forecast. The use of conditional random variables and the Kalman filtering technique has provided tools to increase the forecast accuracy in many cases. (For details see Szöllősi-Nagy 1987). It seems that a combination of deterministic and stochastic methods of forecast will most often lead to smaller forecast errors than either a purely deterministic or a purely stochastic approach.

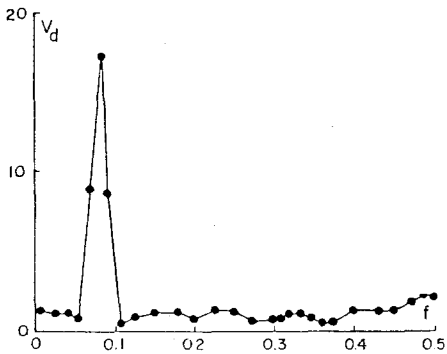

Figure 12

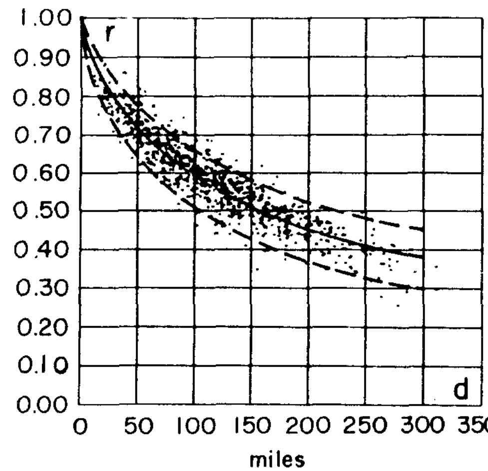

Figure 13

Figure 12. Spectral density graph of monthly precipitation series in Northern Great Plains of the United States (Alexandria Airport Station, latitude 45.867N and longitude 95.383W, elevation 1421 feet)

Figure 13. Correlation coefficient r of monthly precipitation n between the pairs of stations versus their distances, with relationship fitted by the model $r = (1 + AD)^{-n}$ with $A = 0.016$ and $n = 0.534$, for the pairs of station series in Northern Great Plains of the United States (position as in Fig. 11)

11 Various aspects of stochastic modeling

Most hydrologic random variables are non-normally distributed. The majority of techniques for investigation of stochastic processes are based on the assumption of normality. Therefore, hydrologic variables are often normalized by transformations in order to take advantage of these techniques. This approach makes the linear dependence of transformed normal variables nonlinear for the original variables. The question arises whether this nonlinearity is or is not a realistic, physically based condition. If no transformation is made, what nonlinear models can be developed to treat various types of asymmetrically distributed stochastic components (Bernier 1970; Tong 1983)?

Classical statistical techniques often standardize random variables. By doing so, the variables are studied for various other properties (skewness, kurtosis, shape, boundaries, sequential dependence, etc.). However, in water resources the removed mean is a crucial parameter because all or most effects of various water resource projects are either proportional to or in some complex way related to the mean. Similarly, the removed standard deviation is also very important, since the needed storage capacity, with all other factors being the same, is proportional to the standard deviation of the series of storage input minus storage output. By standardization it is feasible to study effects of each major characteristic of an original random process on water resources problems and solutions.

For short-interval time series (say daily), the sample sizes are very large (for 40 years of data the size is 14,600 values). It is known in statistics that very large samples are as troublesome as are the small samples. It is quite difficult in such large samples for normalized singled-out independent (and ideally stationary) stochastic components of complex series of the TIPS type to pass tests of normality and independence. Therefore, it is easy to explain why professionals have been either avoiding, or studying in an approximate way, the short-interval large size hydrologic time series.

The common stochastic models in hydrology are ARMA (autoregressive moving average), with variations such as ARIMA and others (Fernandez and Salas 1986; Salas et al. 1984). The physical background of AR and MA parts of ARMA models are relatively easy to explain. An infinite MA model can be transformed into a finite AR model and vice versa the finite MA model into an infinite AR model. So, a long memory MA model (very long tail of hydrograph of groundwater or lake and swamp outflows) can be replaced by an AR model with a small number of terms, while the surface flow response (unit hydrograph) of a limited number of ordinates already represents an MA model of finite number of terms. This simplest physical explanation, however, requires either constant unit response hydrograph throughout the year if the ARMA parameters are constant, or periodic unit hydrographs for the periodic parameters of ARMA models (PARMA models). If in the final analysis the changing unit hydrographs over the year come up to follow the periodic-stochastic processes, with the classical constant unit hydrograph being the average values of ordinates of these processes, the simple ARMA or ARIMA models would need some adjustments as further generalization beyond periodic ARMA and periodic ARIMA models. Furthermore, there are all transitions of tails between finite short tail and nearly infinite tail of response hydrographs. Therefore, the AR/MA mixture is often based on the presence of these two extreme cases of response hydrographs, which may not be always a realistic assumption.

The detailed study of continuous hydrologic space/time processes appears then often to be very complex. The problem is always how far one should probe into these complexities in stochastic modeling. Objectives of study, sensitivity analysis, size and accuracy of data, cost of modeling and testing, and other factors affect this decision. As a consequence one should clearly stay either with the utilitarian objectives of modeling or with the desire to probe and describe nature in detail by attacking its complexities. If theoretical papers for a hydrologic symposium are solicited, chances are very high that predominately utilitarian (applied) contributions would be submitted. This is to be expected since applicability of results is often the leading objective of hydrologic research.

Models in practice are tested in several ways: (1) data of only one or few stations are used; (2) regional data of many stations are jointly analyzed; (3) representative stations over large continental areas are used as the basic modeling material; and (4) global data of stations selected worldwide are collected and used in modeling. The model development from small number of stations is useful for local applications. Claims for models to be general or universal becomes credible only if data of a large number of representative, regional or global stations are properly screened, eventually corrected for man-made non-homogeneities, and used. Computers and data banks facilitate such tasks at present.

12 Conclusions

Developments of stochastic modeling in hydrology have been significant for the last 25 years. Often modeling was not preceeded by detailed structural and physical analysis of processes. Recent efforts were directed at bridging the physical (deterministic) properties with stochastic modeling, or to have as much physical background to estimate models and their parameters as feasible. Complex hydrologic space/time processes require approximations which can be justified by the utilitarian objectives of modeling. They may not be acceptable from the point of view of the basic understanding and description of nature. Future progress will likely be through evolution of current trends in modeling rather than through breakthroughs. This progress may depend in large measure on critical evaluation of existing modeling principles and techniques. Results of reassessment of some prevailing, classical, hydrologic physical concepts may further affect stochastic modeling in hydrology and water resources.

Acknowledgement

This paper is based on research sponsored by the U.S. National Science Foundation, Grant CEE-8541631. This support is gratefully acknowledged.

References

Akaike, H. 1974: A new look at the statistical model identification. IEEE Transactions on Automatic Control AC-19, 716-723

Amorocho, J.; Espildora, B. 1986: Entropy in the assessment of uncertainty in hydrologic systems and models. Water Resour. Res. 9, 1511-1522

Bernier, J. 1970: Inventaire des modèles de processus stochastiques applicable à la description des dêbits journaliers des rivières. Rev. Inst. Int. Stat. 38, 39-61

Chin, E.H. 1957: Modeling daily precipitation occurrence process with Markov chain. Water Resour. Res. 13, 949-956

Fernandex, B.; Salas, J.D. 1986: Periodic Gamma autoregressive processes for operational hydrology. Water Resour. Res. 22, 1385-1396

Harmancioglu, N.B.; Yevjevich, V.; Obeysekera, J.T.B. 1986: Measures of information transfer between variables. Proceedings of Fourth International Hydrology Symposium of July 15-17, 1985. "Multivariate analysis of hydrologic processes", Colorado State University (H.W. Shen), Fort Collins, Colorado, USA, 481-499

Harmancioglu, N.; Yevjevich, V. 1987: Transfer of hydrologic information among River Points. J. Hydrology

Kelman, J. February 1977: Stochastic modeling of hydrologic intermittent daily processes. Colorado State University Hydrology Papers Series 89, 1-56

Klemeš, V. August 1986: Dilettantism in Hydrology: Transition or Destiny? WRR 22, Supplement 9, 1775-1885

Kolmogorov, A.N. 1941: Ueber das logaritmisch normale verteilungsgesetz der dimensionen der teil chen by Zerstueckerung (On the Lognormal Distribution of Particle Dimension in Breakage), C. R. Acad. Sci., USSR, 31, 99

Linfoot, E.H. 1957: An information measure of correlation. Information and control 1, 85-89

Rajski, C. 1961: Entropy and metric spaces. In: Cherry, C. (ed.) Information theory: 6th Landon Symposium, 1960, Butterworths, Washington, D.C., 41-45

Richardson, V.C. 1977: A model of stochastic structure of daily precipitation over an area. Colorado State University Hydrology Papers Series 91

Roldan, J.; Woolhiser, D.A. 1982: Stochastic daily precipitation models, I. A comparison of occurrence processes. Water Resour. Res. 18, 1451-1459

Salas, J.D.; Delleur, J.W.; Yevjevich, V.; Lane, W.L. 1980: Applied modeling of hydrologic time series. Water Resour. Publ., Littleton, Colorado

Salas-La Cruz, J.D.; Yevjevich, V. 1972: Stochastic structure of water use time series. Colorado State University Hydrology Papers Series 52

Shannon, C.E.; Weaver, W. 1949: The mathematical theory of communication. The University of Illinois Press, Urbana

Szöllősi-Nagy, A. 1987: Chapter on hydrologic forecasts. In: Starosolszky, O. (ed.) Applied surface hydrology, Water Resour. Publ., Littleton, Colorado

Torodovich, P.; Yevjevich, V. 1969: Stochastic process of precipitation. Colorado State University Hydrology Papers Series 35

Tong, H 1983: Threshold models in nonlinear time series analysis. Lecture notes in statistics 21, New York: Springer

Waymire, E.; Gupta, V.K. 1981: The mathematical structure of rainfall representations. 2. A review of the theory of point processes. Water Resour. Res. 17, 1273-1286

Yevjevich, V.; Karplus, A.K. 1973: Area-time structure of the monthly precipitation process. Colorado State University Hydrology Papers series 64

Yevjevich, V. 1974: Determinism and stochasticity in hydrology. J. Hydrology 22, 225-258

Yevjevich, V. 1984: Structure of daily hydrologic series. Water Resour. Publ., Littleton, Colorado

Accepted January 2, 1986.