Nordic Hydrology, 7, 1976, 265-280

Multivariate Distributions in Hydrology and River Regulation

N.A. Kartvelishvili and L.T. Gottschalk
M.A.D.I., Moscow, U.S.S.R., and
S.M.H.I, Norrköping, Sweden

It is assumed that the river runoff process can be approximated by a Markov process. The process is thus described by M distribution functions:

$$F_n(q_t, t; q_{t-1}; t-1; \ldots; q_{t-n}, t-n), \ t \equiv 1, 2, \ldots, M$$

where M is the number of time intervals within the year, n - the order of the Markov process and q_t, in general, is a vector representing runoff at several sites in a river or neighbouring rivers.

Fundamental hypothesis of relations between multivariate distributions and corresponding marginal distributions is given.

A finite difference scheme for multisite and multilag generation of river runoff is derived. The derivation is based on the multivariate normal distribution.

Different methods for determination of the order of the finite difference scheme are discussed as well as the influence of model order and method of parameter estimation on properties of the model.

Introduction

Proper treatment of river regulation calculations assumes a full characterization of the random nature of river runoff. Certain problems allow an analytical solution directly based on the multivariate distribution functions describing the natural river runoff process (Kartvelishvili, 1967, Kartvelishvili and Korganova, 1973). In other cases we use Monte-Carlo methods. For this purpose a scheme for synthetic generation of river runoff series is needed. It is, of course, an advantage if the same basic model can be used for both types of calculations. In this article the theoretical basis for the construction of multivariate distributions to describe the natural river runoff process is given and a finite difference scheme for generation purposes is derived from these multivariate distributions.

Properties of the Runnoff Process.

The instantaneous runoff $Q = Q(t)$ at some river point as a function of time t is a continuous random process. The argument t of such process can take values in the interval$\{-\infty, \infty\}$and the ordinate of the process Q can take any positive value. Let t, ..., t_n be n given moments of time and x_1, ..., x_n-n given numbers. The probability of the event

$$Q(t_1) < x_1, \ldots, Q(t_n) < x_n$$

denoted by

$$P\{Q(t_1) < x_1, \ldots, Q(t_n) < x_n\} = F_n(t_1, x_1; \ldots; t_n, x_n) \qquad (1)$$

is a function of t_1, ..., t_n and x_1, ..., x_n. This probability is called the n-th variate distribution of the process $Q(t)$. If the first distribution function

$$F_1(t, x) = P\{Q(t) < x\}$$

is given as well as a rule for the transition from F_n to F_{n+1}, then the process $Q(t)$ is completely determined. In practice runoff is given as averages over some fixed time period θ.

$$Q^x(t) = \frac{1}{\Theta} \int_{t-\Theta/2}^{t-\Theta/2} Q(z)\, dz$$

for $t = k \times \theta$, where k is an integer, $Q^x(t)$ is a process with discrete time and continuous ordinate. The time period θ equals (dependent on the character of the problem and other conditions) days, pentades, weeks, decades, months or years (hydrological observations are, however, very seldom enough representative to construct the distribution function of daily runoff). If climatic and landscape conditions do not change in time then the distribution function (1) of $Q(t)$ or $Q^x(t)$ is a periodic function with the period of one tropical year T i.e. we are considering a harmonic process. Setting $t_1 = t$, $t_2 = t + \tau_1$, ..., $tn = t+\tau_{n-1}$ we thus have

$$F_n(t_1, x_1; t_2, x_2; \ldots; t_n, x_n) \equiv$$

$$= F_n(t, x_1; t+\tau_1, x_2; \ldots; t+\tau_{n-1}, x_n) \equiv$$

$$= F_n(t+T, x_1; t+T+\tau_1, x_2; \ldots; t+T+\tau_{n-1}, x_n)$$

Harmonic processes (which also implies stationary processes) are the only type of random processes that possesses the very important property of ergodicity. This is very significant for hydrology, dealing with processes, like river runoff, where only one realization is at hand.

Let us turn to another very important property of the runoff process - that it is markovian. If for some n the distribution function F_{n+1} is given, then it is also given for all indices less or equal to n, namely

$$F_n(t_1, x_1 ; \ldots ; t_{n-1}, x_{n-1} ; t_n, x_n) \equiv$$
$$= F_{n+1}(t_1, x_1 ; \ldots ; t_{n-1}, x_{n-1} ; t_n, x_n ; t_{n+1}, \infty)$$

In general, the function F_{n+1} does not determine distribution functions with higher indices i.e. F_{n+2}, F_{n+3}, ... In other words, it does not contain all information about the process. In a special case, when the function F_{n+1} has all information, i.e. determine not only distribution functions with lower indices but also these with the higher ones, the process is called a Markov process of order n. For this kind of process we for any $m>n+1$ have:

$$F_m(t_1, x_1 ; \ldots ; t_m, x_m) \equiv F_{n+1}(t_1, x_1 ; \ldots ; t_{n+1}, x_{n+1}) \cdot$$
$$\cdot \prod_{k=n+2}^{m} \frac{F_{n+1}(t_{k-n}, x_{k-n} ; \ldots ; t_k, x_k)}{F_n(t_{k-n}, x_{k-n} ; \ldots ; t_{k-1}, x_{k-1})}$$

We shall assume that river runoff is a Markov process. We should, however, note that actually we do not have any theoretical basis for this assumption. The hypothesis that the runoff process is markovian for some adequate number n and function F_{n+1} can give necessary accuracy. Transition to a more complicated model for the runoff process, than Markov, is not reasonable, as hydrologic information is always not complete enough for its construction. Such transition to a more difficult description of the runoff process does not make it more probable, but only complicates hydrological and water mangement calculations.

The runoff process is not only a continuous one but also differentiable, and consequently, derivative dQ/dt, i.e. the rate of changes in runoff is a continuous process. River runoff cannot be a first order Markov process as such a process is not differentiable.

If we consider runoff to be a continuous process, then the order n should always be greater or equal to two for instantaneous or moving average runoff. For runoff as a discrete process, i.e., considering average runoff at fixed time intervals, the process can be well described by a first order Markov process (n=1) or even n=0, i.e. runoff is not a Markov process but independent for different time intervals. The order of the Markov process with which the runoff process can be best approximated is dependent on the period of averaging. For instance, for the majority of rivers at the foot of the Carpathian mountains already for intervals of one week runoff can be considered to be an independent process. For rivers in the plain of the European part of the USSR monthly average runoff can be considered to be of the third order or even second order and the volume of annual average runoff - a first order Markov process or independent process. The main group of Swedish rivers are well appoximated by a second order model for monthly runoff and also decades, except for the outflows from the big lakes, where higher order models are needed. Annual flow can usually be regarded as independent or a first order Markov process. Criterias for determination of the order of Markov process are discussed below. However it should be noted already now that the order of Markov process is not a simple value but is connected with the criteria used. The examples given above are general conclusions.

Hypothesis about Multivariate Distributions

In water management calculations it is necessary to consider runoff at several sites on a river or neighbouring rivers. The random variable $\xi(t)$ in such tasks is, thus, a vector, and its components: runoff at the different sites - are highly dependent on each other. In the number of the components of the vector $\xi(t)$ we can also include hydrometeorological variables that are connected to runoff. The multivariate distribution function F_n in case of a vector process will be

$$F_n(t_1, x_1^{(1)}, \ldots, x_1^{(m)}; \ldots; t_n, x_n^{(1)}, \ldots, x_n^{(m)}) =$$
$$= P\{\xi^{(1)}(t_1) < x_1^{(1)}; \ldots; \xi^{(m)}(t_1) < x_1^{(m)}, \ldots$$
$$\ldots, \xi^{(1)}(t_n) < x_n^{(1)}, \ldots, \xi^{(m)}(t_n) < x_n^{(m)}\} \tag{2}$$

where m is the number of components of the vector $\xi(t)$. If the vector process is markovian then it is completely determined by the function F_{n+1} for some n, like in the case of the scalar process.

Let us now consider the problem of constructiong the distribution function F_{n+1}. Dimensions of this distribution is $m(n+1)$ and can be very large. Hydrological observations, also the longest ones obviously are not enough even for construction of a two-dimensional distribution function on a pure empirical base (to get the same accuracy as that of a one-dimensional distribution, based on 100 years of observations, one would require about $100^2 = 10000$ years). The distribution function, thus, must be based on some hypothesis, which we check by experiments.

Let us turn to the hypothesis for hydrometeorological processes, given by Kartvelishvili (1956). We shall assume that all random variables $\xi^{(j)}(t_i)$ (in future we shall for brevity write $\xi_i^{(j)}$), are continuously distributed. It is of no importance whether we consider instant runoff or average runoff at fixed time intervals.

Let $F(x) = P(\xi_i^{(j)} < x)$ be the continuous distribution function of the random variable $\xi_i^{(j)}$. Further on we shall use the function $u(x)$, determined by the equation

$$F(x) = \Phi(u(x)) \tag{3}$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{z^2}{2}\right) dz$$

i.e. the unit normal distribution function. Denoting with Φ^{-1} the inverse to the normal distribution we can write

$$u(x) = \Phi^{-1}\{F(x)\} \tag{4}$$

The graphical illustration of equation (3) is given in Fig. 1. It is clear that the random variable $u(j) = u(\xi_i^{(j)})$ is unit normally distributed. The hypothesis consists in that the joint distribution of the variables.

$$u_k^{(1)}, \ldots, u_k^{(m)}; \ldots, u_{k-n}^{(1)}, \ldots, u_{k+n}^{(m)} \tag{5}$$

is a multivariate normal distribution. Mathematically this hypothesis has no background but only a physical base: if marginal distributions of the variables (5) are normal then it is natural to expect that the joint distribution is also normal, because the opposite physically is rather an exception than a regularity.

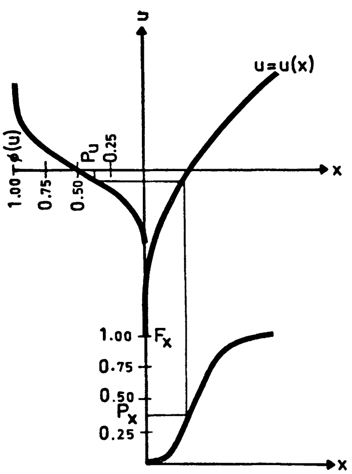

Fig. 1. Graphical illustration of transformation of the unit normal distribution.

According to the accepted hypothesis the $m(n+1)$ - variate density distribution of the random variables (5) is expressed by the formula

$$f(u_k^{(1)}, \ldots, u_k^{(m)}; \ldots; u_{k-n}^{(1)}, \ldots, u_{k-n}^{(m)}) \equiv$$

$$= (2\pi)^{-\frac{(n+1)m}{2}} D^{-\frac{1}{2}} \exp\left\{\frac{1}{2D} D_u\right\} \tag{6}$$

$$D_u = \begin{vmatrix} \Delta_{kk} & \Delta_{k,k-1} & \cdots & \Delta_{k,k-n} & \vec{u}_k^T \\ \Delta_{k-1,k} & \Delta_{k-1,k-1} & \cdots & \Delta_{k-1,k-n} & \vec{u}_{k-1}^T \\ \cdots & \cdots & & \cdots & \cdots \\ \Delta_{k-n,k} & \Delta_{k-n,k-1} & \cdots & \Delta_{k-n,k-n} & \vec{u}_{k-n}^T \\ \vec{u}_k & \vec{u}_k & \cdots & \vec{u}_{k-n} & 0 \end{vmatrix}$$

$$D = \begin{vmatrix} \Delta_{kk} & \Delta_{k,k-1} & \cdots & \Delta_{k,k-n} \\ \Delta_{k-1,k} & \Delta_{k-1,k-1} & \cdots & \Delta_{k-1,k-n} \\ \cdots & \cdots & & \cdots \\ \Delta_{k-n,k} & \Delta_{k-n,k-1} & & \Delta_{k-n,k-n} \end{vmatrix}$$

$$\vec{u}_i = (u_i^1, u_i^2, \ldots, u_i^m)$$

$$\vec{u}_i^T = \begin{Bmatrix} u_i^1 \\ u_i^2 \\ \vdots \\ u_i^m \end{Bmatrix}$$

$$\Delta_{i,j} = \begin{Bmatrix} \rho_{i\,j}^{1\,1} & \rho_{i\,j}^{1\,2} & \cdots & \rho_{i\,j}^{1\,m} \\ \rho_{i\,j}^{2\,1} & \rho_{i\,j}^{2\,2} & & \rho_{i\,j}^{2\,m} \\ \cdots & \cdots & & \cdots \\ \rho_{i\,j}^{m\,1} & \rho_{i\,j}^{m\,2} & & \rho_{i\,j}^{m\,m} \end{Bmatrix} \qquad i,j = k,\ k-1, \ldots k-n$$

and $\rho_{i\,j}^{s\,t}$; $i, j = k, k-1, \ldots, k-n$; $s, t = 1, \ldots, m$ are the correlation coefficient between u_s^i and u_t^j

The density distribution function for the random variables

$$\{ \xi_u^{(1)}, \ldots, \xi_k^{(m)}; \ldots; \xi_{k-n}^{(1)}, \ldots, \xi_{k-n}^{(m)} \}$$

can be given also by equation (6) but with a multiplier:

$$\prod_{\ell=0}^{n} \prod_{j=1}^{m} \frac{du_{k-\ell}^{(j)}}{dx_{k-\ell}^{(j)}}$$

which is always positive.

Analytical Marginal Distribution Functions

Theoretical determination of the distribution function for runoff phenomena is indeed a difficult task, as it involves complete utilization of the probability characteristics of the underlying processes. The a priori information of the form of the distribution curves is thus in most cases restricted to the fact that it should be nonnegative and unbounded upwards.

To apply the hypothesis given in part 2 of this paper, we use transformations of the normal distribution. Let us first construct empirical transformation functions from observational data. N observations are given as module values and ordered from the smallest to the biggest:

$$x^1 < x^2 < \ldots < x^i < \ldots < x^n$$

we write equation (3) as

$$p_i = \Phi(u)$$

where p_i is the empirical probability corresponding to x^i. The empirical transformation function is given by sets (x^i, u^i), $i = 1, \ldots, N$, where u^i is the expected value of the transformed ß-variable corresponding to p_i (Blom 1958). A good estimate of u^i for the normal distribution is:

$$u^i = \Phi^{-1}\left(\frac{i = \frac{3}{8}}{N + \frac{1}{4}}\right) \tag{8}$$

Examples of plotted transformation functions of the normal distribution for Swedish conditions are given in Fig. 2. For negative values of u the module x is close to zero and with the increase of u from negative to positive values x also increases. The slope of the transformation curve is described first of all by the coefficient of variation. The transformation usually does not represent a straight line but has a more complicated form. In the lower end the function is convex with an asymptot at $x = 0$ or some positive value. In the upper part the empirical function can be both convex and concave. A convex form can indicate an existence of an upper bound. Plots like those, shown in Fig. 2, give a good complement to the empirical moments for characterization of observation series. For river regulation calculations we can actually do without analytical expressions for the transformation functions but use a visually smoothed empirical function in form of a table or graph. This is due to the

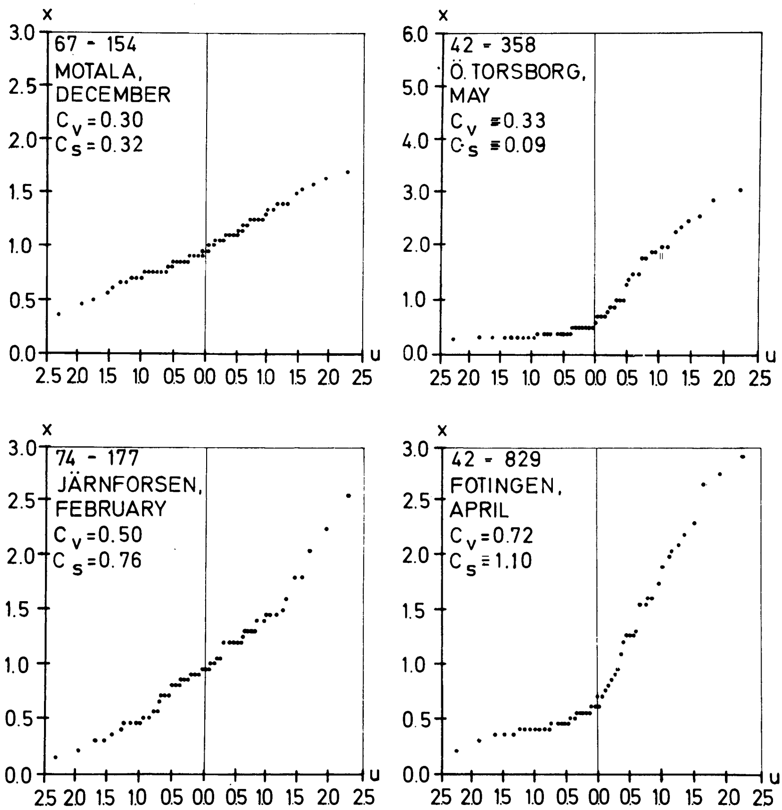

Fig. 2. Empirical transformation functions of the unit normal distribution for monthly river runoff.

fact that the hypothesis about the multivariate distribution is fully indifferent to the kind of univariate distribution. There are no principal advantages in smoothing a curve with analytical expressions instead of visual smoothing. An opinion exists that analytical expressions increase the accuracy of extrapolation outside the area of observations. This opinion strongly overestimates the importance of analytical expressions. Actually, the fact that analytical expression agrees with the empirical distribution function is no guarantee that it will agree with the true distribution function, where no empirical points are at hand. Extrapolation of the distribution function is of importance only for problems concerning calculations of maximum and minimum flow for hydrotechnical installations. As far as river regulation calculations are concerned, the look of the distribution function for very low and high probabilities influences very little the result of these calculations. Analytical expressions are, however, necessary for other reasons. First, they are convenient for calculations on the computer. Second, they are necessary (but not enough accurate) for construction of characteristics of runoff from rivers with few or no observations.

Third, existence of analytical distribution functions makes typization of rivers based on the distribution of runoff within the year easier and more accurate.

To find analytical expressions for a given empirical distribution is, in many cases, a very laborious task but still always possible. It is much more complicated to find such analytical expressions that fit a sufficiently large class of rivers for all months or decades in the year.

Below are given analytical expressions for transformation of the normal distribution of the module of flow $x = Q/\bar{Q}$ that in our opinion are flexible enough for application to monthly and decade runoff.

$$u(x) = \frac{\ell n\, x - a}{\sigma} \tag{9}$$

$$u(x) = \frac{x - 1/x - a}{\sigma} \tag{10}$$

$$u(x) = a\, \ell n\, \ell n\, (1+x) + \beta\, \ell n\, (x) + \gamma\sqrt{x} \tag{11}$$

The first expression gives the two parameter log normal distribution. The second one - a distribution with three parameters $(\bar{Q}, a, \sigma)$ and the third one - a distribution with four parameters $(\bar{Q}, a, \beta\ \gamma)$. We should note, that according to what have been discussed above, these analytical distributions cannot be regarded as any final recommendations, but only as examples. However, they have proved to be useful and applicable in many cases.

Parameter Estimation

In many applications of multivariate distributions in hydrology the method, used for parameter estimation, can be of vital importance for the result of the calculations. One should be especially careful when using transformations like the ones described in the previous part. In hydrology the method of moments and the method of maximum likelihood are most frequently used methods.

Let us first turn to the marginal distributions. A general advice is not to use the method of moments if the coefficients of variation and skew coefficients have high values. For two parameter distributions, for instance the two parameter log-normal distribution, differences between parameters, estimated by two different methods, can be large. For log-normal distribution the two first moments of the logarithms of the sample, which also are the maximum likelihood estimates, can differ considerably from the same parameters, estimated from the two first moments of the original sample. The deviations are the largest in cases, when the skewness of the sample is not in agreement with the theoretical one. Using the m.l. method, we always get a variability in the process larger than the one, estimated by the method of moments. This can be of great importance for water management calculations. Differences in the variability of the process are smaller if we use three parameter distributions, though the individual parameter values can be significantly different.

Problems that have been discussed above, examplify some of the difficulties, when using transformations. In many situations it is advisable to use the method of moments, in spite of its drawbacks.

Theoretical and numerical calculations for determination of parameters of the multivariate distribution Eq.(6) will be much more complicated, than in the univariate case, when performing a joint derivation of all parameters of respective distributions. Not only the specific parameter that describes the dependence between the variables at different times and places, the time and space correlation coefficients, should be calculated but also the parameters of the marginal distributions.

If we use the maximum likelihood method to estimate the parameters of the bivariate normal distribution, it comes out that the marginal samples contain all information about parameters of the marginal distributions. When using other than linear transformations of the normal distribution, this is not the case. Formulas for estimation of parameters of the marginal distribution for same certain time interval and place involve, thus, observations from adjacent time intervals and places. It is difficult to foretell theoretically the amount of extra information, that can be obtained from them. It is possible to guess that the marginal estimates will be close to the estimates from the joint distribution, as one can assume, that the main part of the information about the marginal parameters should be contained in the marginal sample.

In agreement with the discussion above we can estimate the parameters of the transformations Eqs. (9) - (10) from the marginal samples. The correlation coefficients between the transformed variables in the multivariate distribution Eq. (6) can now be calculated by the method of moments as the correlation coefficients between the transformed marginal samples.

We can use the empirical transformation described above for calculation of this correlation coefficient. The variance σ_u of the transformed sample $u^i(x_k^i)$, $i = 1, ..., N$ of the vector component k, is

$$\sigma_u = \frac{1}{N-1} \sum_{i=1}^{N} \{ \Phi^{-1} (\frac{i = \frac{3}{8}}{N + \frac{1}{4}}) \}^2 \tag{12}$$

σ_u is only dependent on N and takes always a value that is a little less than one. Let m_i and n_i be the rank numbers of the ordered samples of vector components j and k. The correlation coefficient between these two components then is

$$\rho_u = \frac{1}{N-1} \sum_{i=1}^{N} \Phi^{-1} (\frac{m_i = \frac{3}{8}}{N + \frac{1}{4}}) \; \Phi^{-1} (\frac{n_i = \frac{3}{8}}{N + \frac{1}{4}}) \tag{13}$$

The correlation coefficients calculated from the transformed samples differ from the one calculated from the original sample. As a rule, they are somewhat larger.

Finite difference equation

The multivariate distribution Eq. (6) can be transfered to a finite difference form (Kartvelishvili 1975, Gottschalk 1975). In matrix notations it looks like:

$$\vec{u}_k = H_1 \vec{u}_{k-1} + H_2 u_{k-2} + \ldots + H_n \vec{u}_{k-n} + \Lambda \vec{\varepsilon} \qquad (14)$$

here $\vec{u}_k$ is the m-dimensional vector of normalized runoff at m different places for the k-th time interval, H_s, $s = 1, \ldots, n$ and Λ are (mxm) parameter matrices and $\vec{\varepsilon}$ is a m-dimensional vector which components are independent unit normally distributed random numbes. It should be noted, that the matrices H_s, $s = 1, \ldots, n$ and Λ, in general case, are dependent on the time interval k. The components H_{ij} of the parameter matrix

$$H_s = \begin{Bmatrix} H_{1,sm+1} & \ldots & H_{1,(s+1)m} \\ \ldots & & \ldots \\ H_{m,sm+1} & \ldots & H_{m,(s+1)m} \end{Bmatrix}$$

are calculated as

$$H_{ij} = \{(-1)^i E_{1i} A_{1j} + (-1)^{i+1} E_{2i} A_{2j} + \ldots$$
$$\ldots + (-1)^{i+m-1} E_{mi} A_{mj}\} / E$$

where A_{ij}; $i,j = 1, m(n+1)$ are the algebraic complements to the determinant D (Eq. 6),

$$E = \det \begin{Bmatrix} A_{11} & \ldots & A_{1m} \\ \ldots & & \ldots \\ A_{m1} & \ldots & A_{mm} \end{Bmatrix}$$

and E_{ij}; $i,j = 1, m$ are the minors to the determinant E. The parameter matrix Λ is defined by

$$\Lambda = \sqrt{D} \begin{Bmatrix} \sqrt{\lambda_1} & 0 & \ldots & 0 \\ 0 & \sqrt{\lambda_2} & \ldots & 0 \\ \ldots & & & \ldots \\ 0 & 0 & \ldots & \sqrt{\lambda_m} \end{Bmatrix} \cdot (\beta_{ji})$$

where λ_i; $i = 1, \ldots, m$ are the solutions to the secular equation:

$$\det \begin{Bmatrix} A_{11}-\lambda & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22}-\lambda & \cdots & A_{2m} \\ \cdots & \cdots & & \cdots \\ A_{m1} & A_{m2} & \cdots & A_{mm}-\lambda \end{Bmatrix} = 0$$

and (β_{ji}) the corresponding unitar matrix, which has the property: $(\beta_{ij}) \times (\beta_{ji}) = I$.

It is obvious that Eq. (14) is convenient for generation of synthetic series of runoff. Let $\vec{u}_{k-1}, ..., \vec{u}_{k-n}$ be given. With calculated parameter matrices H_s, $s = 1, ..., n$ and Λ we generate with the help of normally distributed random numbers ε_k, $k = 1, ..., m$ and calculate $\vec{u}_k$. When $\vec{u}_k$ is given we can in the same way find $\vec{u}_{k+1}$ and so on. The transition from the realization of the process $\vec{u}_k$ to the river runoff process $\vec{Q}_k$ is obvious: for every k and ℓ, $\ell = 1, ..., m$ we solve the equation $u_k^\ell = u_k^\ell (Q_k^\ell)$ (ℓ denotes the vector component).

Order of the Markov Process

Determination of the order of the Markov process is based on the information we get from the correlation matrix. Any true order does not, of course, exist. We only talk about the best order based on some certain criterion. This criterion can be defined in many different ways. Applying different criteria we can get contradictory results.

As we are dealing with a nonstationary process the order can be different for different time intervals of the year. Within a year different processes in the watershed have predominant influence on the correlation structure of runoff for different seasons (Gottschalk 1976). The nonstationarity is also connected to the length of the time intervals used, as the time scales of these different process show large variations. The order must therefore be determined by the maximum order over all time intervals in the year.

Let us first for simplicity discuss determination of the order of a scalar process i.e. $m = 1$ in Eq. (6). For this case H^n, $s = 1, ..., n$ and Λ^n are simple parameters. These parameters change with the order n. With the growth of n differences between the parameters become smaller and we can always define some $n = n_x$, when these differences can be considered to be negligible. n_x then defines the order. Fig. 3 shows the behaviour of the parameter Λ^n as a function of n for monthly flows in two rivers, Amur at Chabarovsk and Volga at Kuibychev. We see that Λ^n is stable for different n for different months. The values of n for the twelve months are given in Table 1.

Table 1-Values of n_x for different months

River	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Okt	Nov	Dec
Amur	3	2	1	1	1	8	1	2	3	2	1	5
Volga	6	2	1	1	7	9	1	8	9	8	9	7

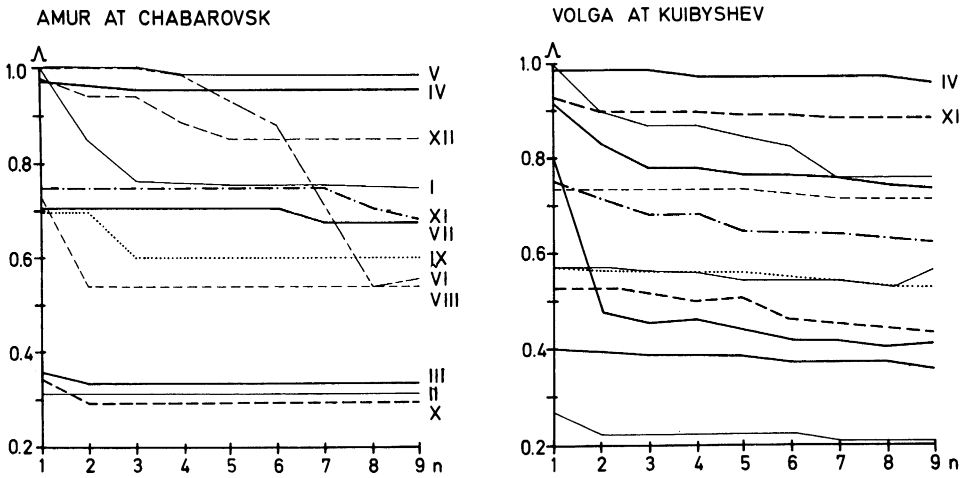

Fig. 3. The dependence of λ^n on n for different months for rivers Amur at Chabarovsk and Volga at Kuibyshev.

We can also use an F-test on the residual variances to see if any significant improvement can be obtained by using order $n + 1$ instead of order n. This way of determining the order, as a rule, gives lower values than the one exemplified above.

Similar critera determine the order based on the parameter H^n can be set up. The behaviour of this parameter is however less stable.

We have assumed that the residual ε is unit normally distributed and independent. Tests for control whether these two assumptions are fullfilled do not actually point a certain order. There exists, however, a tendency in such tests to prefere lower orders. As an example, unit normality was checked by a χ^2-test for 38 Swedish rivers with monthly flow for models of order from one to five. The result is shown in Table 2 which certifies what was said above.

Table 2 - Number of months for which the hypthesis that the residuals are unit normally distributed was rejected. Totally 456 months were tested.

Order of the model	Level of significance 5%	1%	0.1%
1	28(6.5%)	3(0.7%)	2(0.5%)
2	30(6.9%)	9(2.9%)	2(0.5%)
3	27(6.2%)	10(2.3%)	2(0.5%)
4	42(9.7%)	10(2.3%)	3(0.7%)
5	34(7.9%)	16(3.7%)	3(0.7%)

In principle, the same methods as those applied to the scalar case can be used for the vector model. The individual matrix components are less stable compared to the scalar case. Kartvelishvili (1975) suggests to use the entropy

$$\epsilon(\tilde{\Lambda}\varepsilon) = \ln \prod_{j=1}^{n} \Lambda_j \qquad (15)$$

to determine the order. That n for which the product of $\Lambda^1 \ldots \Lambda^n$ stabilizes or has a minimum determines thus the order. Results got above for the scalar Markov process represent a special case for the vector process.

Inconsistensies in Annual Moments

One, of course, cannot accept simulated series totally without reservations and set them alike with historical data. A model is always an approximation. It will preserve properties of the natural runoff process that are directly described by its parameters but can show inconsistensies with respect to other properties, if this is not specially taken care of.

Let us discuss the problem of preserving the moments of the annual runoff process, when our model is based on time intervals like months, decades and weeks. We derive annual runoff as the sum over M time interval within the year

$$Q = \frac{1}{M} \sum_{j=k}^{k+M-1} Q_j$$

The mean of annual runoff is thus:

$$m = E(Q) = \frac{1}{M} \sum_{j=k}^{k+M-1} E(Q_j) = \frac{1}{M} \sum_{j=k}^{k+M-1} m_j \qquad (16)$$

The annual variance is equal to

$$\mathrm{var}(Q) = \frac{1}{M^2} \{ \sum_{j=k}^{k+M-1} \mathrm{var}(Q_j) + \sum_{i=k}^{k+M-2} \sum_{\ell=i+1}^{k+M-1} \mathrm{cov}(Q_i Q_e) \} \qquad (17)$$

From Eqs. (16) and (17) we learn that the annual mean and variance are not only dependent on the mean and variance of different time intervals within the year but also on the correlation structure i.e. both individual values of correlation coefficients and order of Markov process.

We shall also draw attention to the fact that Var (Q) is dependent on the definition of the year i.e. k in Eq. (17). This is illustrated in Fig. 4, where the annual coefficient of variation as a function of the beginning month of the year for some Swedish observation series is given. This small but significant periodicity is explained by the nonstationarity in the runoff for time intervals less than one year.

To preserve the annual variance for any definition of the year we should manipulate with the correlation structure.

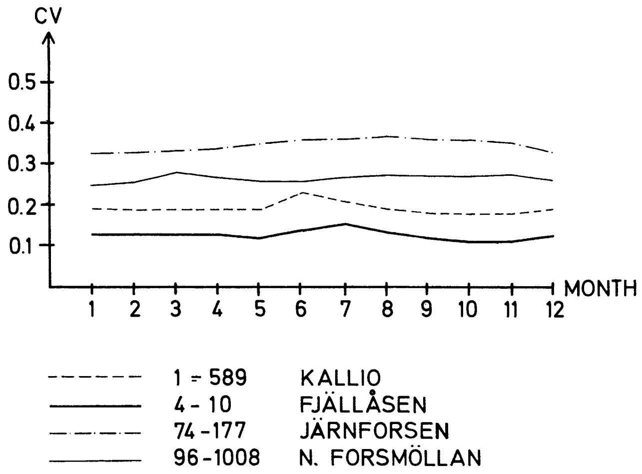

Fig. 4. Annual variation coefficient as a function of the beginning month of the year.

In Fig. 4 is shown the annual lag one correlation coefficient as a function of the beginning month of the year. We observe the same periodic variation as for the annual variance. We can derive an equation for the annual lag one correlation coefficients between the chosen time intervals. In this case it is also necessary to manipulate with the correlation structure to assure that the annual lag one correlation coefficient is preserved.

We can go on like this to higher order moments and correlation coefficients with larger lags.

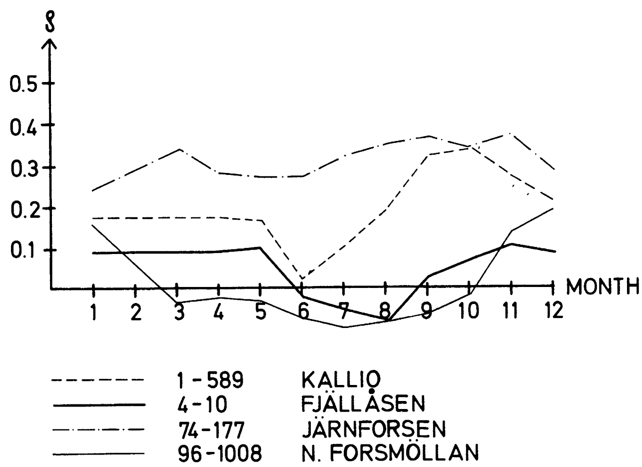

Fig. 5. Annual lag one correlation coefficient as a function of the beginning month of the year.

Conclusions

The presented model, first of all, should be understood as a working tool to be applied in water managements calculations. We must remember, that when we use statistical methods to determine our model, the order as well as parameter values, there is actually no guarantee that we can reflect in all respects the true nature of the runoff process. We only find a model that judging from the observation material is the best one under given assumptions. For most applications this is satisfactory.

The stochastic model gives us a good complement to observed runoff series to carry out water management calculations. The main adventages seems to be:

- Possibility to compare the sensitivity of different alternatives to variations in hydrological data
- Differences between different alternatives can be better defined
- Possibility to evaluate hydrological extremes and plan for such extreme situations in future.

References

Blom, G. (1958) Statistical estimates and transformed ß-variables, Almquist & Wiksell, Stockholm.

Gottschalk, L. (1975) Stochastic modelling of monthly river runoff, Bullentin series A No 45, Department of Water Resources Engineering, Lund Institute of Technology, Lund.

Gottschalk, L. (1976) Correlation structure and time scale of simple hydrologic systems. To be presented at Nordic Hydrological Conference, Reykjavik 1976.

Kartvelishvili, N. A. (1956) O matematicheskom opisasii i metodike raschetov regulirovaniya rechnogo stoka, Izvestiya AN SSSR, OTN No 1.

Kartvelishvili, N. A. (1967) Teoriya veroyatnostnykh protsessov v gidrologii i regulirovanii rechnogo stoka, Gidrometeoizdat, Leningrad.

Kartvelishvili, N. A. and Korganova, N. S. (1973) Integralnye uravneniya regulirovaniya rechnogo stoka, Problemy gidroenergetiki i vodnogo khozyajstva, Vyp 10, Alma Ata.

Kartvelishvili, N. A. (1975) Stokhasticheskaya Gidrologiya, Gidrometeoizdat, Leningrad.

First received: 29 April, 1976
Revised version received: 13 August, 1976

Address:
Lars Gottschalk
The Swedish Meteorological and Hydrological Institute Fack,
S-601 01 Norrköping
Sweden.
N. A. Kartvelishvili,
M.A.D.I.,
Moscow,
U.S.S.R.