

# STOCHASTIC MODELS OF LOW FLOW FOR THE RISK EVALUATION

Mikhail BOLGOV

*Water Problem Institute Russian Academy of Sciences, Moscow, Russia*

119333, Gubkin st.,3, IWP RAS, Moscow, RUSSIA

## **Abstract.**

The evaluation of the risk, related with water, can be possible with help of stochastic models for river runoff fluctuations for different averaging periods. A simple Markov chain specified by nonlinear regression equation with marginal gamma-distribution, describes adequately the distribution of series of low-flow years. Seasonal (intra-annual) low flow periods are reproduced properly by a non stationary periodic PARMA model. It is shown that the distribution of variables such as sojourn time for a river runoff (time of staying below some specified threshold level) is characterized by a multimodal function caused by genetic non homogeneity of runoff formation conditions in different seasons of a year.

## **INTRODUCTION**

The study of the statistical structure of runoff series in the low-flow range is a key problem of engineering hydrology. The basis for the majority of findings obtained due to this Project, is the threshold level approach proposed by Eyjevich (1967) and involved investigation of the behavior of runoff processes in reference to some specified quantity (a threshold). In essence, this approach means the investigation of the distribution for excursions of the process below some specified runoff level.

The estimates for such values as mean sojourn time (the time of staying below a specified level), obtained from small samples, exhibit an appreciable sample variance and bias (especially in the range of small threshold values). Having at his disposal an adequate stochastic model for the runoff fluctuation process, one is able to obtain the characteristics of interest either analytically or numerically (by means of simulation experiments). The latter approach is preferable since the emerging analytical difficulties are considerable while, on the other hand, there is a wealth of experience in hydrology of constructing stochastic models. Of course some additional technical, mathematical and other problems can arise because reliable methods of constructing stochastic models are still not sufficiently advanced especially if time intervals between observations are shorter than a year (seasons, months, and shorter) when the stationarity hypothesis is not valid.

The procedure of stochastic modeling simulation consists of four stages: identification of

mathematical model; statistical estimation of unknown parameters of the model; numerical simulation of artificial time series; verification of the results.

## ANNUAL RUNOFF TIME SEMES

### *Mathematical model and simulation procedure.*

When describing statistical regularities of long-term annual runoff fluctuations, it is reasonable to restrict oneself to the class of Markov processes (or more precisely, simple Markov chains since the time parameter  $t$  is integer). The characteristic feature of such processes is that a joint probability distribution function for two consecutive time instants defines completely all the statistical characteristics. As a rule, the runoff probability distribution is stationary but non-Gaussian, therefore we consider here a general (not necessarily Gaussian) case when the joint probability distribution of two variables separated by a time interval  $t$  is determined by the probability density  $p(t, x, y)$  satisfying the Markov equation

$$p(t_1 + t_2, x, y) = \int_a^b \frac{p(t_1, x, z)p(t_2, z, y)}{p(z)} dz \quad (1)$$

which specifies a stationary Markov process. Let, in addition,  $p(x)$  be the corresponding marginal probability density. Consider the following integral equation:

$$\int_a^b p(t, x, y)\varphi(y)dy = r(t)p(x)\varphi(x) \quad (2)$$

If some additional conditions for  $p(t, x, y)$  are satisfied (for our applied purpose we may assume this to be true), then this equation has a denumerable set of eigenfunctions  $\varphi_k(x)$  corresponding to numbers  $r_k(t)$ . The following theorem (O.V.Sarmanov, 1961) provides a basis for constructing the probability density for such a process:

**Theorem.** A necessary and sufficient condition for the probability density to satisfy the Markov equation (1) is that the following conditions be valid:

- 1) eigenfunctions  $\varphi_k(x)$ ,  $k=1, 2, \dots$  do not depend on the parameter  $t$ ;
- 2) eigen correlation junctions  $r_k(t)$  are exponential functions:

$$r_k(t) = \exp(-\lambda_k t), t \geq 0, k = 1, 2, \dots \quad (3)$$

where  $\lambda_k$  are positive numbers and

$$0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_k \leq \dots \quad (4)$$

in other words,  $p(t, x, y)$  must be a sum of the form:

$$p(t, x, y) = p(x)p(y) \left[ 1 + \sum_{k=1}^{\infty} e^{-\lambda_k t} \varphi_k(x) \varphi_k(y) \right] \quad (5)$$

To construct a two-dimensional density  $p(t, x, y)$ , one needs to specify a square domain ( $a < x, y < b$ ) (a and b, may be infinite), a nonnegative weight function  $p(x)$  (a marginal probability density), and a set of functions  $\varphi_k(x)$ ,  $k=1,2,\dots$ , which are orthonormal on the interval (a, b) with the weight function  $p(x)$ . If there exist values  $\lambda_k > 0$  such that the series (5) converges to a nonnegative, for all  $t > 0$ , function  $p(t, x, y)$ , then the sum of the series is the probability density of a Markov process.

In what follows, we consider mainly two forms of the decomposition (5); employing orthonormal bases of either the Laguerre or the Legendre polynomials. Let first  $p(x)$  be the gamma-distribution marginal density:

$$p(x) = \frac{\gamma^\gamma}{\Gamma(\gamma)} x^{\gamma-1} e^{-\gamma x} \quad (6)$$

commonly used in hydrological applications. If we take a set of normalized generalized Laguerre polynomials as functions  $\varphi_k(x)$ , and some positive number  $R < 1$ , then it is easy to obtain from (5) the following expression for the joint probability density of the stationary Markov Chain  $x(t)$ , ( $t = 1, 0, 1, 2, \dots$ )

$$f(x, y) = p(x)p(y) \left[ 1 + \sum_{k=1}^{\infty} R^k L_k^{\gamma-1}(\gamma x) L_k^{\gamma-1}(\gamma y) \right], \quad (7)$$

where  $x=x(t)$ ,  $y=x(t+1)$ , and  $R$  equals to the correlation coefficient between two consecutive values of the Markov chain.

Secondly, let the functions  $\varphi_k(x)$ , be Legendre polynomials  $P_k(x)$  orthonormalized on the interval (0,1) with the constant weight  $p(x)=1/2$ . Then the formula for  $p(x,y)$  can be written as follows (O.V.Sarmanov, 1961):

$$p(t, x, y) = \frac{1}{4\pi} \int_0^\pi \frac{(1 - e^{-2\lambda t}) d\omega}{\left[ 1 - e^{-\lambda t} xy + \sqrt{(1-x^2)(1-y^2)} \cos \omega + e^{-2\lambda t} \right]^{3/2}} \quad (8)$$

We emphasize that both the formulas (7) and (8) imply (O. V. Sannanov, 1961) that, for these two examples, the general regression equation for two variables  $x(t)$  and  $x(t+1)$  should be linear which is not necessarily true for runoff series. That's why the model (7) is rarely used for modeling such series. On the other hand, D.Ya. Ratkovich (1976), who had analyzed long-term (40 years length) runoff records throughout the world, came to a conclusion that the transformed variables,  $F(x)$ , where

$$F(x) = \int_x^{\infty} p(y)dy \quad (9)$$

obey the linear regression equation better. As for the variable  $F(x)$ , it is known to be uniformly distributed on an interval (0,1), therefore the Markov chain  $F(x(t))$  may have the probability density (8) and can be modeled using the linear regression equation. Finally, the appropriate transformation of these variables should be made to transfer to gamma-distributed variables with the probability density (I. O. Sarmanov, 1973):

$$f(x, y) = p(x)p(y) \left\{ 1 + \sum_{k=1}^{\infty} \lambda^k P_k[F(x)]P_k[F(y)] \right\} \quad (10)$$

where  $x=x(t)$ ,  $y=x(t+1)$ ,  $\lambda$ - is the correlation coefficient between  $F(x(t))$  and  $F(x(t+1))$ . As the result, we have developed the procedure for modeling a stationary Markov chain with gamma-distribution marginal densities and nonlinear regression equation.

Fig.1 demonstrates the distinction between the linear (4) and nonlinear (6) models. Here are shown level curves and regression lines for the probability distributions in question.

#### ***Application to the problem of permanent series of low-flow years.***

The stochastic nature of river flow fluctuations may lead to the occurrence of long series of low-flow (drought) years. To define the probability distribution of low-flow series length, we should consider the Markov Chain with two states, corresponding to two alternative events: "runoff value exceeds the threshold T" and "runoff value is below T" (i.e., the year is low-flow) with unconditional probabilities  $(1-Q)$  and  $Q$ , respectively. Our purpose is to calculate the probability  $P(m)$  that the series of low-flow years has exactly the length m which is:

$$P(m) = (1-Q)RuS^{m-1}(1-S) \quad (11)$$

where Ru is (conditional) probability of low-flow year under the condition of wet year, and S is (conditional) probability of low-flow year under the condition of low-flow year. All these probabilities can be calculated from the probability density  $p(x,y)$

$$Q = \int_{-\infty}^T p(x)dx \quad (12)$$

$$S = \frac{1}{Q} \int_{-\infty}^T \int_{-\infty}^T p(x, y)dx dy \quad (13)$$

$$Ru = \frac{1}{1-Q} \int_{-\infty}^T \int_T^{\infty} p(x, y)dx dy \quad (14)$$

## SEASONAL STOCHASTIC MODELS

Characteristic features of monthly averaged runoff series and the requirements on the class of models. Analysis of autocorrelation function of monthly runoff series (ACF) shows that now the variable  $x(t)$  is generally determined by two or more preceding variables:  $x(t-1)$ ,  $x(t-2)$ , ... rather than by  $x(t)$  only. Therefore, we cannot restrict ourselves to the class of simple Markov chains and choose the wider class. For simplicity, we consider linear schemes only:

$$x(t) = \phi_1 x(t-1) + \phi_2 x(t-2) + \dots + \phi_p x(t-p) + \dots + a(t) \quad (15)$$

where  $a(t)$  is a white noise.

As a consequence, we must impose the restriction on the modeled variables to be Gaussian because it is the only class known to admit linear transformations like (15). So, we have obtained the well known class of autoregression models of order  $p$ ,  $AR(p)$  with the parameters  $\phi_1, \dots, \phi_p$  to be estimated along with the variance of the white noise  $a(t)$  and the autoregression order  $p$ .

As soon as the Gaussian time series  $z(t)$  (with zero mean and unit variance) has been modeled, the final step should be made consisting of the transformation (of  $z(t)$ ) into the series  $x(t)$  imitating the runoff series with the variables  $x$  having the three-parametric gamma-distribution as their marginal distribution (and, by the reasons explained above, linked with each other by the nonlinear regression equation (Bolgov, 1996).

The shape of the correlation functions calculated from the normalized data  $z(t)$ :

$$r(t, \tau) = \langle z(t)z(t-\tau) \rangle \quad (16)$$

evidences that the correlation structure differs noticeably for different reference months (i.e. for different  $t$ ), especially for different reference seasons. This is the case for any region without regard to the value of its characteristic correlation  $R(1)$  (see section above), and is due to differences in the genesis of runoff formation during different seasons of a year. Indeed, spring-flood runoff is weakly correlated to preceding water discharge, The ACF for flood reference months decay rapidly.

For rivers which are predominantly rain-fed in the warm season, the rapid decay of ACF is typical for all flood months. After cessation of flood period these ACF become flat.

On the contrary, the stable-runoff periods are characterized by significant correlation coefficients between adjacent months and a slowly decreasing correlation function (in the whole) which can be naturally clearly explained by the prevalence of the underground feeding component in this period. Drought months have minimum value of autocorrelation preliminary flood months. Similar minimum correlation can be traced also with the same month season but a year earlier.

Regionally averaged estimates of monthly runoff ACP differ distinctly for different regions for large

(a year and larger) time shifts depending on the correlation  $R(1)$  between adjacent years. The following regularity can be observed: if the correlation  $R(1)$  is negligible, then all the ACF for monthly runoff, even for stable feeding seasons, are close to zero for time shifts exceeding a year. On the contrary, if  $R(1)$  is not small, then the greater it is, the slower ACF decreases for large shifts  $\tau$ .

Our examination and modeling results have shown that, for stable feeding seasons, the runoff process exhibits a long "memory" (reflecting the genetic features of its formation), which must be taken into consideration to reproduce correctly the behavior of processes in question when developing the models to be employed in applied problems.

The analysis made above implies that the monthly runoff time series can be treated as a periodically stationary random process (with the period of 12 months). So, the parameters of the proposed above autoregression scheme (including the order  $p$ ) should periodically depend on reference time  $t$ , and the model is called a periodic autoregression (PAR) model. Moreover, the transformation from normal to gamma-distributed variables also varies periodically with  $t$  which fact is taken into account in (Bolgov, 1996).

## **APPLICATION**

Having at our disposal the stochastic model of seasonal runoff fluctuations, we are able to proceed to study such characteristics as sojourn time (the time of staying below a specified level, in other words, the duration of the deficit period). The corresponding analytical problem being difficult to solve because the model is rather complex, we use the imitation method. So we model an artificial series of 10000 years length (using the method recommended above) for a small catchment in the Central European Russia. On the basis of the modeled series, the probability distribution for sojourn time has been obtained. Fig 2 gives probability densities for duration for several specified levels. These curves have complicated multimodal shapes caused by genetic non homogeneity of the runoff process in different seasons.

### **Amu-Darya example.**

#### *Autocorrelation functions.*

A hydrological problem, like the one considered, puts demands on the model used for reproducing the characteristic features of the observed series in the wide range of periods including time scales from one month up to several (at least two) years. The fact that in the majority of cases the significant value of the inter –annual correlation,  $r(1)$  must be taken into account, along with a complicated (non-monotonous) behaviour of the correlation function  $r(t)$  for a period less than a year, can lead to rather high values of the order  $p$  of the  $PAR(p)$  model. To prevent this undesirable situation, we propose to abandon the  $PAR$  class by adding a proper white noise to the  $PAR$  –process

$z(t)$  thereby entering a wider class of *PARMA* models:

$$r(t) = (a - 1)\delta(\tau) + a \exp(-\alpha\tau) \quad (17)$$

This approach seems to be preferable to the alternative ones, which are rather artificial. The model parameters were determined by the least square method. Their values are presented in Table 1 for some regions. It should be mentioned that the use of individual sample correlation functions is unacceptable for two main reasons: computational instability of the Yule-Walker equations in our case and significant errors in sample estimates. Having this in mind, we used the regionally averaged autocorrelation functions (ACF).

The experience of constructing a stochastic nonstationary periodically correlated model of runoff, using the so-called PAR model, showed that satisfactory results might be obtained for snow-fed rivers. The tests applied showed a satisfactory agreement of the parameters and properties of the observed and simulated series.

Table 1. The parameters of the *PARMA* model for Darganata gauge station.

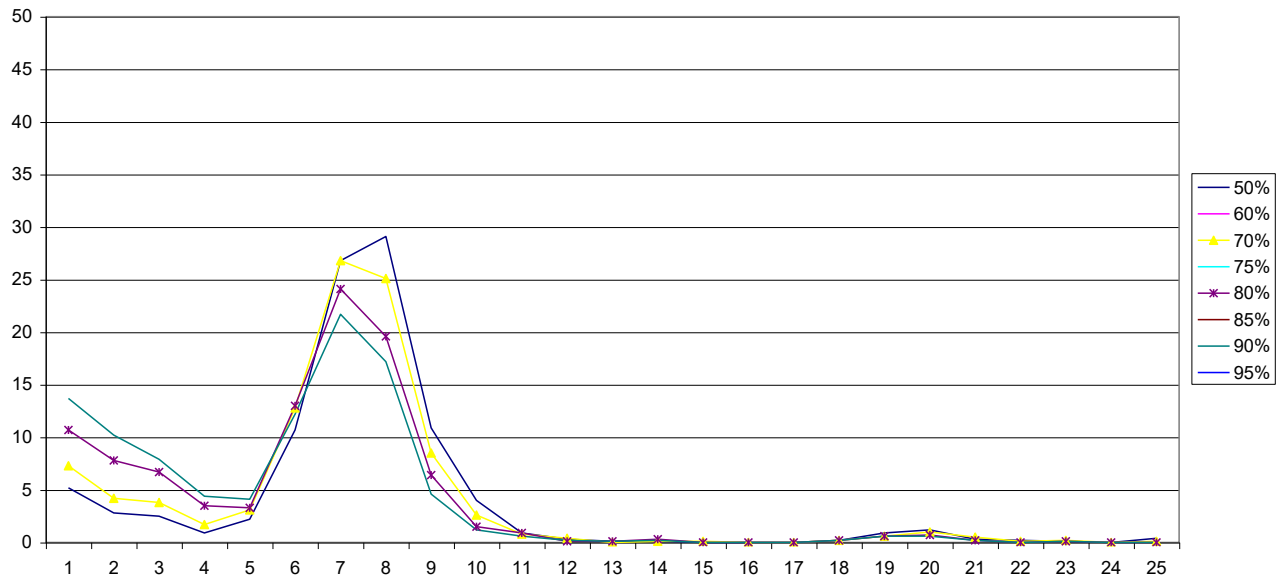
Month	Average	Coef. of variation	Cs/Cv asymmetry	Mean square deviation	Parameters of (17)		r(1)
					a	$\alpha$	
1	622	0.31	1.0	193	0.74	0.10	0.67
2	505	0.39	1.0	196	0.68	0.21	0.55
3	504	0.43	3.0	217	0.80	0.47	0.50
4	758	0.58	2.0	437	0.60	0.10	0.55
5	1490	0.46	0.5	681	1.00	0.43	0.65
6	1920	0.41	0.5	786	0.98	0.29	0.73
7	2400	0.43	0.5	1035	0.89	0.10	0.81
8	1760	0.35	0.5	611	0.92	0.10	0.83
9	1014	0.42	0.5	422	0.80	0.10	0.72
10	618	0.32	0.5	197	0.79	0.16	0.67
11	568	0.28	0.5	160	0.60	0.10	0.55
12	708	0.23	0.5	162	0.60	0.10	0.55

*Preliminary results.*





Density of the distribution of the deficit duration  
(Amu-Darya - Darganat, simulated data, 2000 years)



## CONCLUSION

Study of low-flow fluctuations by stochastic modeling leads to the following results.

1. The model of simple Markov chain with marginal gamma-distribution and nonlinear regression equation connecting two adjacent variables, can be used to describe the distribution for the series of low-flow years.
2. On the basis of the nonstationary periodic model consisting of the set of 12 equations corresponding to 12 months, the sequence of low-flow periods can be adequately reproduced. In general, the model is the mixture of white noise and simple Markov chain with a slowly decaying correlation function.
3. The probability distribution of the variables like the sojourn time, is characterized by a complicated multimodal probability density caused by the genetic nonhomogeneity of the runoff formation conditions during different seasons and by the periodic correlations.

## Acknowledgment

The results of researches, described in this publication, was made possible in part by Grant Jayhun and Grant RFBR No 08-05-00807-a

## REFERENCES

1. Bolgov M. 1996, Estimate of the parameters of stochastic flow models. International conference on Water Resources & environment research: towards the 21th century. Vol.2,

Kyoto, Japan

2. Gustard A., Bullock A. and Dixon J.M 1992, Low flow estimation in the United Kingdom. Report 108, Inst. of Hydrology, Wallingford, UK.
3. Ratkovich.D.Ya., 1976. Mnogoletnie Kolebaniya Rechnogo Stoka (Long-Term Variations in River Runoff), Leningrad: Gidrometeoizdat.
4. Sarmanov, O.V., 1961, Issledovanie stationamih Markovskih proeessov metodom razlozheniya po sobstvennim funkciyam, TrudyMIAN, Vol.60, pp.238-261.(In Russian).
5. Sarmanov, I.O., 1973, Problemy Regulirovaniya i ispol'ziovaniya Vodnykh Resursov (Problems of Management and Use of Water Resources), Moscow: Nauka, p.87 (In Russian)
6. Yevjevich, V. 1967, An objective approach to definition and investigations of continental hydrologic drought. Hydrology papers, 23, Colorado State University, Fort Collins.

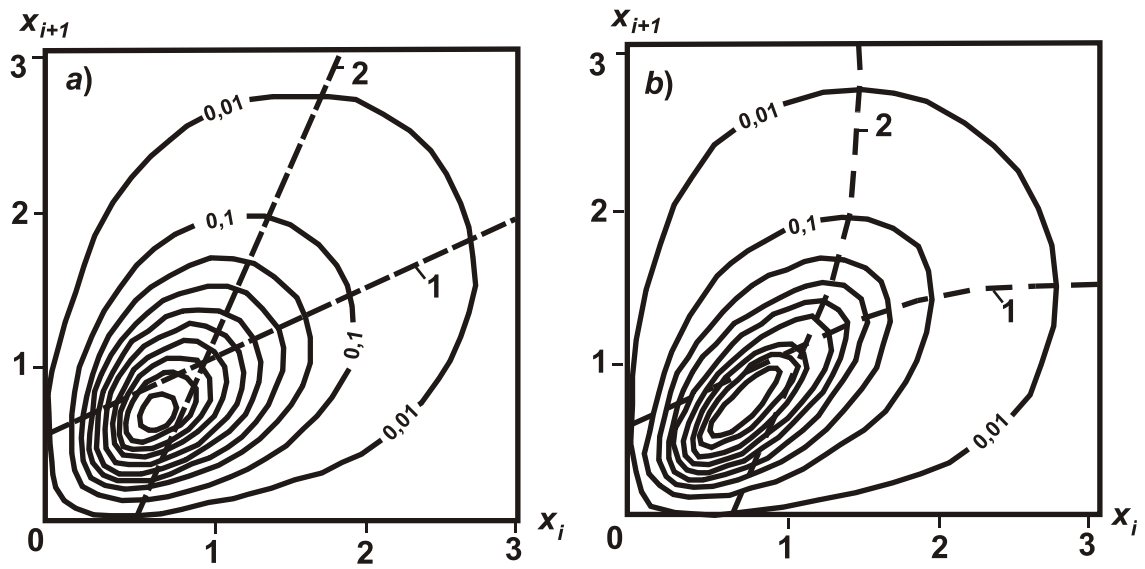


Figure I. Level curves of equal probability density for the two-dimensional gamma-distribution when both the correlation coefficient and the coefficient of variation are 0.5 for the linear (a) and nonlinear (b) Markov model.

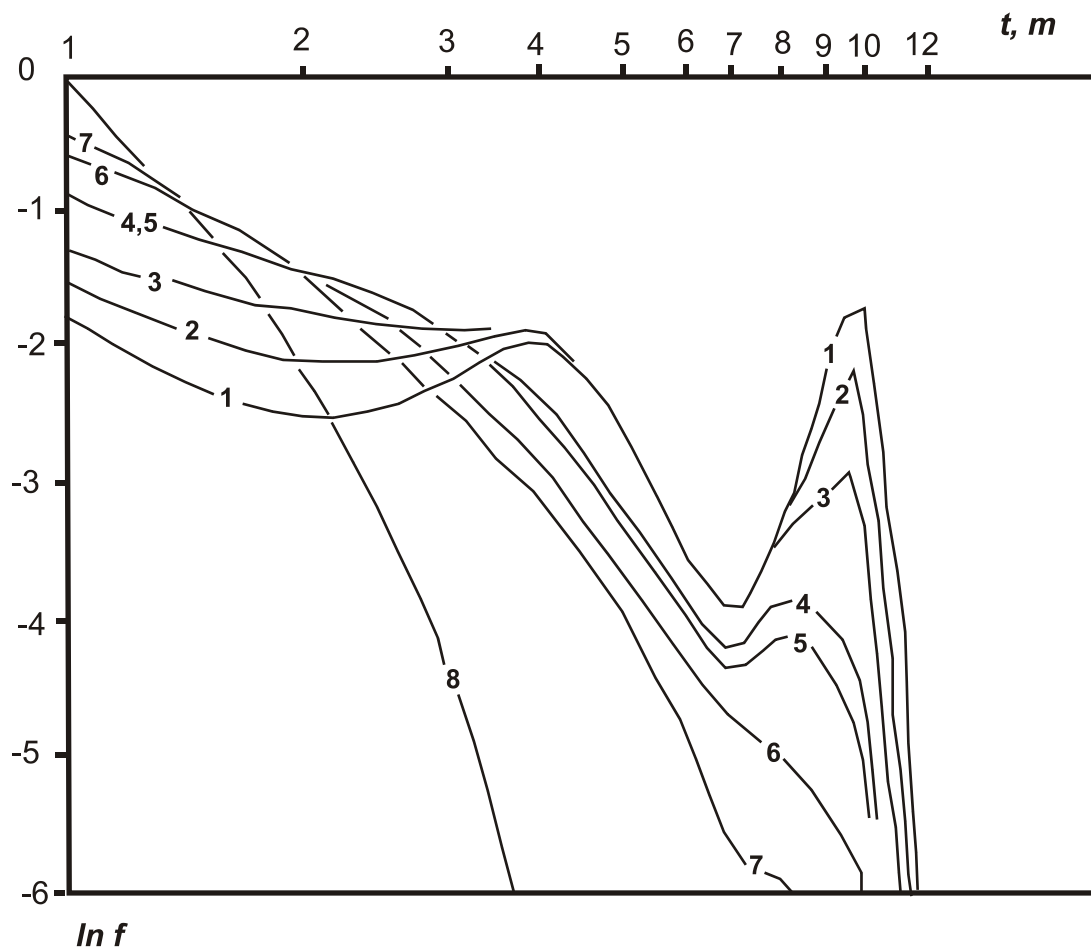


Figure 2. Probability densities of sojourn time (the time of staying below the specified level  $T$ ) for the periodically correlated river runoff process. The level is given in percentage of the runoff norm  $X_0$ :

1 - 100%  $X_0$ , 2 - 75%  $X_0$ , 3 - 50%  $X_0$ , 4 - 25%  $X_0$ , 5 - 20%  $X_0$ , 6 - 10%  $X_0$ , 7 - 5%  $X_0$ , 8 - 1%  $X_0$ .