
WATER RESOURCES AND THE REGIME OF WATER BODIES

Statistical Methods for River Runoff Prediction

V. F. Pisarenko*, A. A. Lyubushin**, M. V. Bolgov***, T. A. Rukavishnikova*,
S. Kanyu*, M. F. Kanevskii*, E. A. Savel'eva*,
V. V. Dem'yanov*, and I. V. Zalyapin*

* *International Institute of the Theory of Earthquake Prediction and Mathematical Geophysics, Russian Academy of Sciences, Varshavskoe sh. 79, korp. 2, Moscow, 111355 Russia*

** *Institute of Earth Physics, Russian Academy of Sciences, ul. Bol'shaya Gruzinskaya 10, Moscow, 112399 Russia*

*** *Water Problems Institute, Russian Academy of Sciences, ul. Gubkina 3, GSP-1, Moscow, 119991 Russia*

Received November 26, 2003

Abstract—Methods used to analyze one type of nonstationary stochastic processes—the periodically correlated process—are considered. Two methods of one-step-forward prediction of periodically correlated time series are examined. One-step-forward predictions made in accordance with an autoregression model and a model of an artificial neural network with one latent neuron layer and with an adaptation mechanism of network parameters in a moving time window were compared in terms of efficiency. The comparison showed that, in the case of prediction for one time step for time series of mean monthly water discharge, the simpler autoregression model is more efficient.

INTRODUCTION

River runoff prediction is a difficult hydrological problem because the physical processes that control it are too complex to allow adequate description by a system of appropriate equations. This is why stochastic methods are widely used for predicting river runoff [1, 8, 14–16]. Several modifications of statistical forecasts proposed in [13] are based on the autoregression analysis of the series in combination with simulation of trends in harmonics. However, all these methods have a significant drawback, since the stochastic component is simulated using the stationary autoregression, while it demonstrates pronounced seasonal variations.

Reliable runoff forecasts are indispensable in various branches of water use (water reserves, power, navigation, flood control, etc.). Clearly, the reliability of an efficient runoff forecast depends on the monitoring systems supplying the required data; therefore, such systems need permanent modernization.

The objective of this study is the development of forecasting methods that are based only on data on the past runoff values. The hydrological experience shows that long-term forecasts are hardly possible in such formulation. However, forecasts with a lead time of one time step (one month) are possible. This allows the power potential of water bodies to be used more efficiently. The theoretical results of this study include conclusions regarding the type and complexity of a stochastic runoff model with a monthly sampling interval (in other words, a model with a seasonal trend). Such models are of particular interest in the simulation-based studies of complex water management systems. Examples of seasonal-trend models are known in hydrology ([2] and others); however, their properties and the qual-

ity of forecasts made with their help have not been adequately studied.

A relatively new statistical method—the theory of periodically correlated processes (PCP) [3]—is used in this case. This model is sometimes referred to as a cyclic stationary process [9]. The formulas for predicting runoff values with a monthly lead based on the previous monthly runoff values were obtained with the use of the standard least-squares method. The PCP approach takes into account both deterministic and stochastic components of seasonal variations in runoff data and allows the correction of the prognostic coefficients for a specific prediction month. The prediction coefficients vary from month to month. Runoff prediction with the use of an artificial neuron network (ANN) [6, 10] was used as a benchmark test of prediction efficiency. This approach was used for river runoff prediction, for example, in [5, 7, 11, 12]. Monthly runoff prediction was made using a perceptron neuron network, containing one latent layer, with training on data from the previous period. Comparison of the two methods shows that the forecasts made by the PCP method are never worse than those made by the ANN method. The former method is as a rule 10–30% more efficient than the latter; however, for some rivers, the results obtained by using these methods are almost identical.

Data on the runoff of nine large rivers in Europe and Asia (Loire, Elbe, Danube, Glomma, Vistula, Oka, Northern Dvina, Irtysh, and Amur) were used in this study. The length of the observational time series for these rivers varied from 84 to 134 years.

Monthly runoff values $X(t)$, $t = 1, \dots, N$, were used to demonstrate the application of the proposed method. The general form of a runoff time series is shown in

Fig. 1. These time series can appreciably differ. A distinct regularity with annual maximums caused by snow melting and spring floods was observed for some rivers, such as the Glomma, Amur, Irtysh, Oka, and Northern Dvina. These time series are typical examples of PCPs. The Vistula, Elbe, and Loire feature less pronounced regularities. The Danube demonstrates the most chaotic runoff variations. This can be explained by the small effect of spring flood on the runoff of these rivers.

Mean monthly runoff values (annual forms) $m(\beta)$, $\beta = 1, \dots, 12$ (hereafter, β will be used only as a number of a month), are given in Fig. 2 along with the respective root-mean-square deviations, which also are different. Some rivers (the Loire, Danube, and Elbe) feature regular quasi-harmonic behavior, which can be attributed to a low spring flood. The Glomma, Vistula, Oka, Northern Dvina, and Irtysh each have a single distinct peak corresponding to the spring flood. Two peaks are recorded in the annual dynamics of the Amur; the lower peak corresponds to the spring flood, and the higher peak, to autumn floods caused by cyclonic precipitation. Root-mean-square deviations (Fig. 2) demonstrate a feature typical of all the rivers: these deviations are proportional to the monthly runoff values. The relative root-mean-square deviations for months with higher runoff values are also relatively large.

Power spectra for runoff time series are shown in Fig. 3. A specific feature of all these spectra is the presence of sharp peaks separated by equal intervals. These peaks are caused by periodic deterministic components with a period of one year and its overtones. The peak magnitudes are closely related to the respective annual curves, since the power spectra are smoothed by the Fourier coefficients of the annual variation curves. Thus, the annual runoff variation curves for the Oka and Northern Dvina have six spectral peaks with virtually equal magnitudes, whereas the spectra of annual runoff variations of the Loire, Danube, and Elbe each have one distinct peak.

PERIODICALLY CORRELATED STOCHASTIC PROCESSES

Let us consider the principal definitions of PCP [3, 9]. A stochastic process $X(t)$ with discrete time t is called a stochastic PCP with a period T if its mean $m(t) = E\{X(t)\}$ and covariation function $C_{xx}(t, s) = E\{(X(t) - m(t))(X(s) - m(s))\}$ are periodic functions with a period of T :

$$m(t + T) \equiv m(t); \quad C_{xx}(t + T, s + T) \equiv C_{xx}(t, s).$$

In this case, $T = 12$, since the runoff time series have a distinct annual seasonal periodicity, which should be used in forecasting. Thus, the annual trend curves in Fig. 2, which in fact are the mean functions $m(t)$, present a simple deterministic forecast of monthly runoff. It is more important, however, to predict the stochastic component of the runoff time series. Thus, let us isolate a centered stochastic component $Y(t) = X(t) -$

$m(t)$ and solve the forecast problem for this purely stochastic part of the runoff time series. In order to satisfy the principle of causal relationship, we are to construct an estimate $m(t)$, which is used in the forecast made at time moment t and is based only on the previous values of $X(s)$, $s < t$.

Centered time series $Y(t)$ are given in Fig. 4. Obviously, the subtraction of sample means $Y(t)$ reduces the variance as compared with the sample variances δ_x^2 of the original time series $X(t)$. The ratio $\gamma_m = \delta_x^2 / \delta_y^2$ can be regarded as an index of deterministic predictability, when the prediction is based only on the cyclic mean $m(t)$. It can be seen that the subtraction of the cyclic mean $m(t)$ appreciably reduces the sample variance (by a factor of 1.09 to 5.21). However, the decrease is not the same in different months. The PCP model enables this heterogeneity to be utilized.

Thus, the most predictable were found to be the runoff variations in rivers with high seasonal floods (the Irtysh, Amur, Northern Dvina, and Oka). The least predictable are rivers with low spring floods (the Danube, Elbe, Loire, and Vistula). However, this classification will change if we take into account the predictability of the stochastic component $Y(t)$.

Correlation matrices for Oka and Irtysh runoff values are shown in Table 1. It can be seen that the Oka features relatively low correlation coefficients for consecutive months; whereas, these coefficients for Irtysh are large and can reach 0.92 (February–March). The PCP model takes into account variations in the correlation coefficients from month to month.

RUNOFF PREDICTION

Suppose that the prognostic value $\hat{X}(t)$ of a process has the form $\hat{X}(t) = \hat{m}(t) + \hat{Y}(t)$, where $\hat{m}(t)$ is an estimate of the cyclic mean $m(t)$ based on the previous values $X(s)$, $1 \leq s < t$. The prediction $\hat{Y}(t)$ for the stochastic component $Y(t)$ can be found in the form of a linear combination of p previous observations:

$$\hat{Y}(t) = a_1(t)Y(t-1) + \dots + a_p(t)Y(t-p), \quad (1)$$

where $[t-p, t-1]$ is the time interval used for prediction. It should be mentioned that the prognostic coefficients $a_j(t)$ really depend on the time moment t of prediction.

Two methods for the evaluation of coefficients $a_j(t)$ in (1) are tested. The first approach, referred to as the standard autoregression (AR) estimate [4, 12] is based on the following procedure. Let L be the number of values within the time window used for the current time moment t of the forecast $Y(t)$. Now, the standard AR-estimate corresponds to the coefficients of linear prediction that minimize the sum of prediction errors squared:

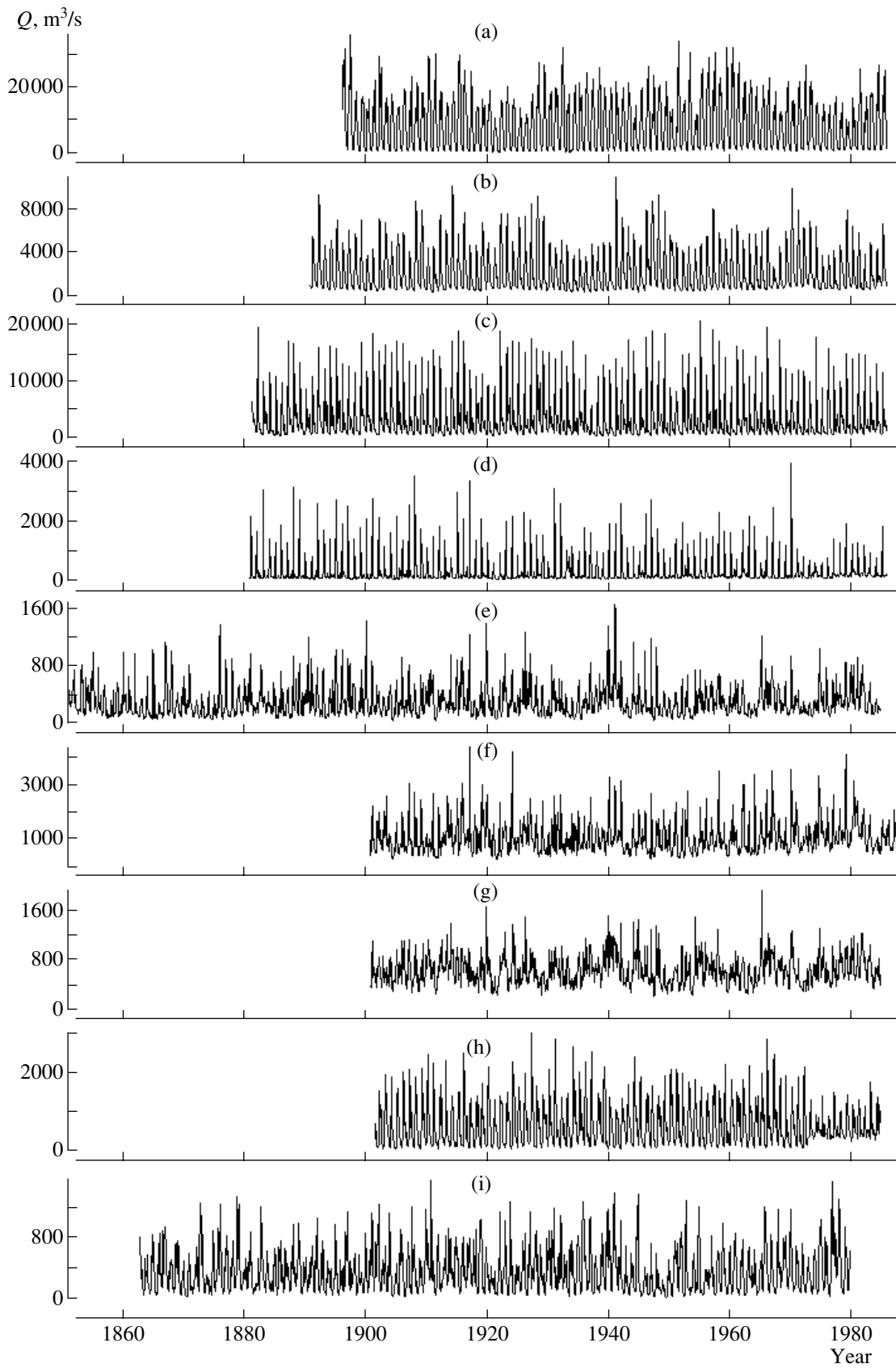


Fig. 1. Mean monthly discharges for nine rivers in Russia and Europe (here and in Figs. 2–4: a–i are Loire, Elbe, Danube, Glomma, Vistula, Oka, Northern Dvina, Irtysh, and Amur, respectively).

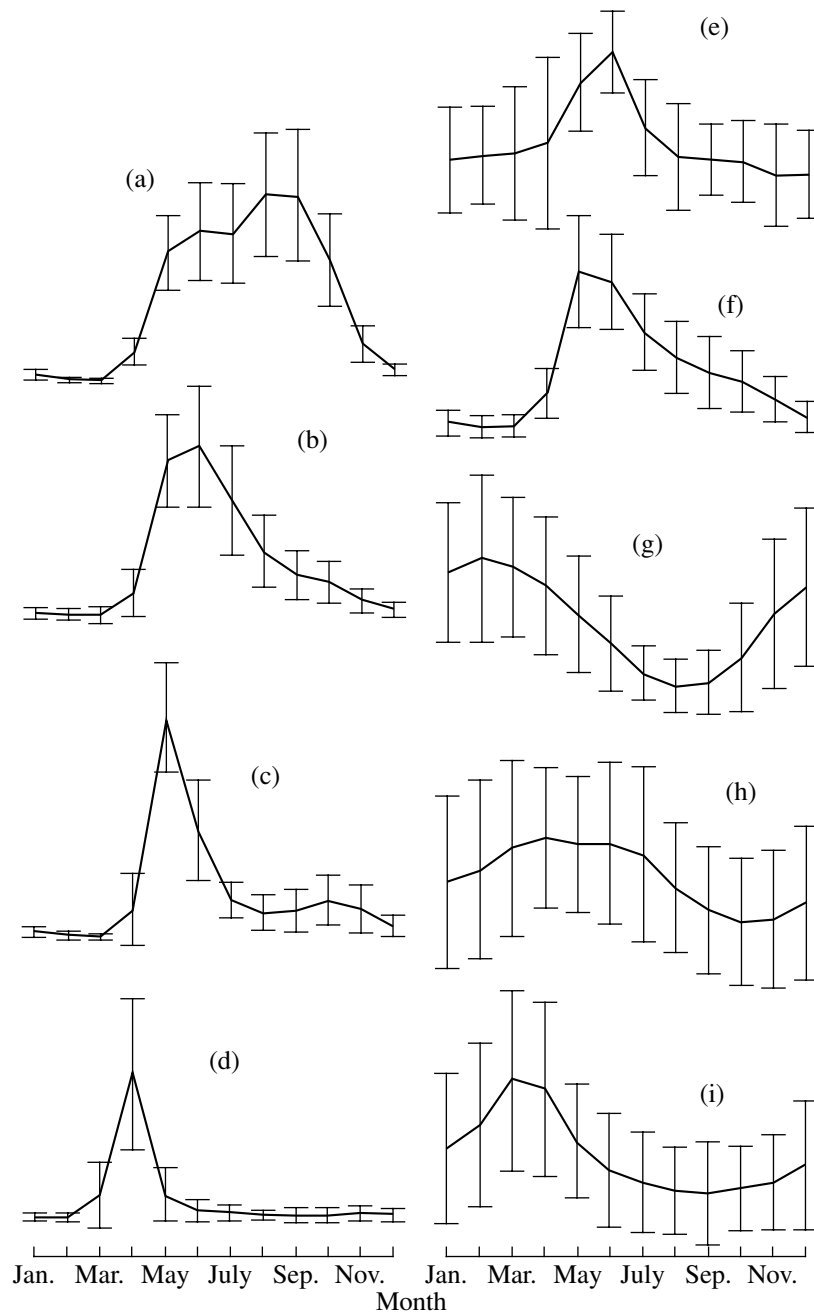


Fig. 2. Mean annual runoff variations $m(\beta)$, $\beta = 1, \dots, 12$ in relative units with standard deviations.

$$\sum_{n=1}^{L-p} \left(Y(t-n) - \sum_{k=1}^p a_k Y(t-n-k) \right)^2 \rightarrow \min_{a_k} \quad (2)$$

Unlike the PCP approach, the prognostic coefficients a_k in (2) are not supposed to depend on the time t .

The solution to the problem (2) must satisfy the system of linear equations

$$\sum_{k=1}^p C_{kj}^{(0)}(t) a_k = C_{0j}^{(0)}(t), \quad j = 1, \dots, p, \quad (3)$$

where

$$C_{kj}^{(0)}(t) = \sum_{n=1}^{L-p} Y(t-n-k) Y(t-n-j). \quad (4)$$

Suppose that $a_k^{(0)}(t|p)$ is the solution to linear system (3) (in this form, a dependence of a_k on t still persists). It should be mentioned that in the case when an infinite set is averaged ($L \rightarrow \infty$), this dependence for a stationary time series $Y(t)$ disappears. However, some dependences of a_k on t can still exist for finite L values.

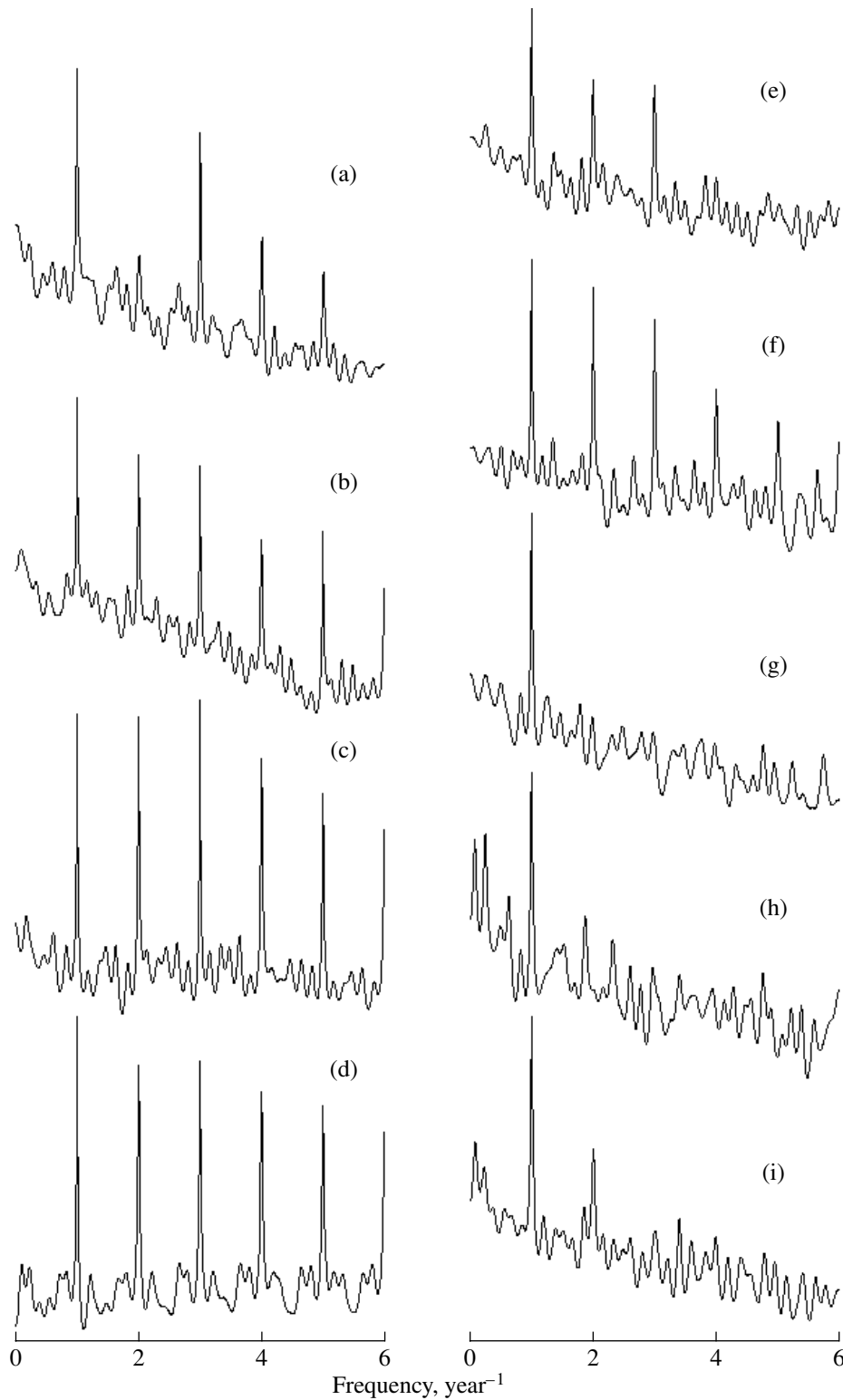


Fig. 3. Estimated power spectra of seasonal runoff variations, relative units.

The second approach referred to as AR-PCP-model represents a modification of the least-squares problem (2) that takes into account the periodic correlation property. Suppose that q is the integer part of the relationship $(L-p)/T$: $q = [(L-p)/T]$. The integer $q = q(t)$, gen-

erally speaking, depends on t . Now we have to solve the following problem:

$$\sum_{n=1}^{q(t)} \left(Y(t-nT) - \sum_{k=1}^p a_k Y(t-nT-k) \right)^2 \rightarrow \min_{a_k} \quad (5)$$

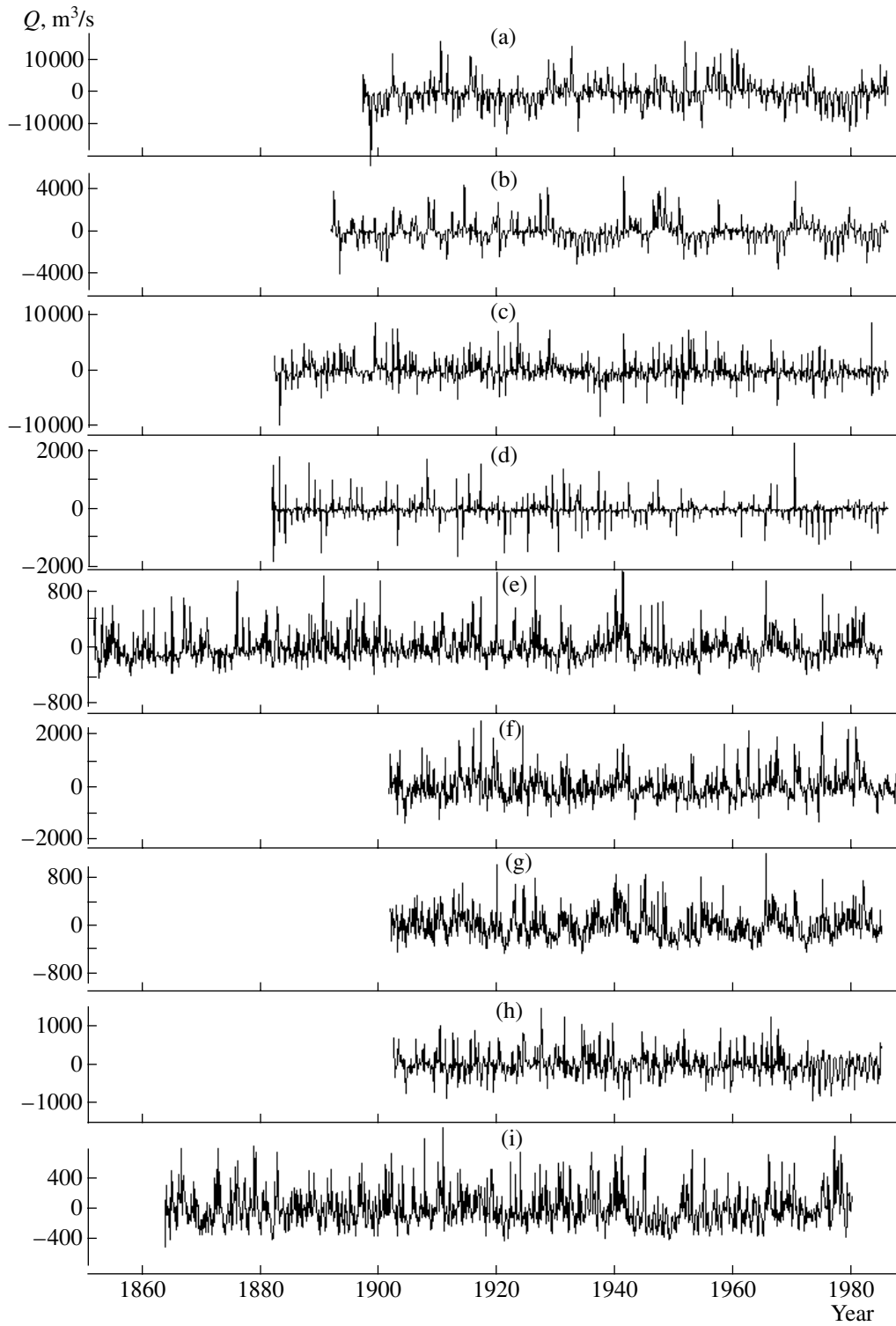


Fig. 4. Deviations from current estimates of cyclic means (signals $Y(t)$) for mean monthly water discharges.

From (5), we obtain the following system

$$\sum_{k=1}^p C_{kj}^{(1)}(t) a_k = C_{0j}^{(1)}(t), \quad j = 1, \dots, p, \quad (6)$$

$$C_{kj}^{(1)}(t) = \sum_{n=1}^{q(t)} Y(t-nT-k)Y(t-nT-j). \quad (7)$$

Let us denote the solution to system (6) by $a_k^{(1)}(t|p)$. Thus, we have two different predictors of the p th order:

Table 1. Matrices of correlation coefficients between different months for deviations from the cyclic mean (signal $Y(t)$) for the Oka (top number) and Irtysh (bottom number)

β	1	2	3	4	5	6	7	8	9	10	11	12
1	$\frac{1.00}{1.00}$											
2	$\frac{0.57}{0.70}$	$\frac{1.00}{1.00}$										
3	$\frac{0.08}{0.41}$	$\frac{0.19}{0.92}$	$\frac{1.00}{1.00}$									
4	$\frac{-0.07}{0.36}$	$\frac{-0.13}{0.55}$	$\frac{-0.51}{0.53}$	$\frac{1.00}{1.00}$								
5	$\frac{-0.00}{0.52}$	$\frac{-0.07}{0.40}$	$\frac{-0.31}{0.25}$	$\frac{0.27}{0.44}$	$\frac{1.00}{1.00}$							
6	$\frac{0.01}{0.44}$	$\frac{0.02}{0.26}$	$\frac{-0.12}{0.10}$	$\frac{0.06}{-0.17}$	$\frac{0.39}{0.66}$	$\frac{1.00}{1.00}$						
7	$\frac{0.10}{0.36}$	$\frac{0.06}{0.19}$	$\frac{0.11}{0.05}$	$\frac{0.02}{-0.17}$	$\frac{0.31}{0.42}$	$\frac{0.68}{0.87}$	$\frac{1.00}{1.00}$					
8	$\frac{0.10}{0.30}$	$\frac{0.02}{0.19}$	$\frac{0.02}{0.07}$	$\frac{-0.11}{-0.06}$	$\frac{0.16}{0.27}$	$\frac{0.44}{0.63}$	$\frac{0.63}{0.85}$	$\frac{1.00}{1.00}$				
9	$\frac{0.13}{0.27}$	$\frac{0.03}{0.20}$	$\frac{-0.01}{0.08}$	$\frac{-0.07}{-0.02}$	$\frac{0.10}{0.31}$	$\frac{0.48}{0.51}$	$\frac{0.48}{0.65}$	$\frac{0.71}{0.84}$	$\frac{1.00}{1.00}$			
10	$\frac{0.16}{0.14}$	$\frac{0.02}{0.07}$	$\frac{-0.10}{0.01}$	$\frac{0.01}{-0.04}$	$\frac{-0.02}{0.28}$	$\frac{0.17}{0.37}$	$\frac{0.23}{0.47}$	$\frac{0.40}{0.61}$	$\frac{0.57}{0.83}$	$\frac{1.00}{1.00}$		
11	$\frac{0.12}{0.15}$	$\frac{0.01}{0.12}$	$\frac{-0.02}{0.06}$	$\frac{-0.01}{0.00}$	$\frac{0.05}{0.30}$	$\frac{0.22}{0.33}$	$\frac{0.28}{0.39}$	$\frac{0.45}{0.52}$	$\frac{0.59}{0.69}$	$\frac{0.71}{0.85}$	$\frac{1.00}{1.00}$	
12	$\frac{0.19}{0.26}$	$\frac{0.09}{0.24}$	$\frac{0.03}{0.18}$	$\frac{-0.12}{0.08}$	$\frac{-0.08}{0.33}$	$\frac{0.00}{0.34}$	$\frac{0.11}{0.35}$	$\frac{0.21}{0.45}$	$\frac{0.20}{0.59}$	$\frac{0.37}{0.66}$	$\frac{0.51}{0.78}$	$\frac{1.00}{1.00}$

$$\widehat{Y}^{(\alpha)}(t|p) = \sum_{k=1}^p a_k^{(\alpha)}(t|p)Y(t-k), \quad \alpha = 0, 1. \quad (8)$$

$$e^{(\alpha)}(t|p) = Y(t) - \widehat{Y}^{(\alpha)}(t|p), \quad \alpha = 0, 1, \quad t = L_0 + 1, \dots, N. \quad (9)$$

If we compare (2) with (5), it can be easily seen that the number of terms in the sum (2) is greater than that in the sum (5) by a factor of $T = 12$, and the sum (2) ensures more reliable averaging. However, if the time series examined is periodically stationary rather than stationary, such averaging can worsen the forecast. On the other hand, averaging (7), which takes into account the principal property of periodicity of PCP, can enable a more efficient forecast. Thus, the advantage of the PCP forecast depends on the extent to which the time series in question is periodically correlated, that is, its correlation varies periodically. Let us consider both these methods and their possible combinations used to reduce the prediction error.

A time series of prediction errors can be constructed for both methods

Here, $L_0 + 1$ is the first time moment for prediction, N is the number of values in the time series considered. The first L_0 values are used to initialize the algorithm. In the case of AR-PCP-estimate, the calculation of all time series is more convenient to start from January. Since some time series do not meet this requirement (Fig. 1), the first values (≤ 11) were not used in order to make all the series to start from January. Again, the length of initialization L_0 or the length of the moving time window L was taken equal to an integer number of years: L_0 or $L = Tn_0 = 12n_0$, where n_0 is an integer. In all cases, $n_0 = 30$ years, $L = L_0 = 360$ values. Thus, the time series of prediction errors (9) also started from January. Let N_Y be the integral part of $N/12$, that is, the number of complete years of observations. Note that the time series $Y(t)$ can be determined only starting from the second year of observations.

The time series $Y(t)$ can be divided into 12 partial series corresponding to months:

$$Y(\tau|\beta) = Y(12(\tau - 1) + \beta), \quad (10)$$

$$\beta = 1, \dots, 12; \quad \tau = 2, \dots, N_Y,$$

where β denotes months, and τ denotes successive observational years. Thus, $Y(\tau|\beta)$, $\tau = 2, \dots, N_Y$ is the time series of deviations from the cyclic mean value that corresponds to the month with the number β . Similarly to formula (9), we can divide the time series of prediction errors (8):

$$e_\beta^{(\alpha)}(\tau|p) = e^{(\alpha)}\{12[\tau - (n_0 + 1)] + \beta|p\}, \quad (11)$$

$$\tau = n_0 + 1, \dots, N_Y.$$

We can introduce an index to characterize the efficiency of prediction for month β :

$$\mu^{(\alpha)}(\beta|p) = \sum_{\tau = n_0 + 1}^{N_Y} Y^2(\tau|\beta) / \sum_{\tau = n_0 + 1}^{N_Y} (e_\beta^{(\alpha)}(\tau|p))^2. \quad (12)$$

In addition to that, we can introduce two indices to characterize the overall efficiency of the forecast averaged over all months:

$$\gamma_0 = \sum_{t = L_0 + 1}^N X^2(t) / \sum_{t = L_0 + 1}^N e^2(t), \quad (13)$$

$$\gamma_1 = \sum_{t = L_0 + 1}^N Y^2(t) / \sum_{t = L_0 + 1}^N e^2(t).$$

Both moving and increasing time windows were used for linear AR-predictors. The increasing time series always yield the best result. This is why the results presented below were obtained only with the use of increasing time windows.

Table 2 gives the results of experiments with AR-prediction for all nine time series. The prediction algorithm based on AR-model with $p = 1-12$ was tested for each series. An increasing time window beginning from the 361st value (January that follows the first 30 years used for the initialization) was used. The results of prediction with a standard AR-estimate and with an estimate using AR-PCP ($\alpha = 0$) were compared. For each month $\beta = 1, \dots, 12$ and for each method $\alpha = 0, 1$, the efficiency of the monthly prediction $\mu^{(\alpha)}(\beta|p)$ was evaluated. Suppose that $\alpha^*(\beta)$ and $p^*(\beta)$ is a solution of the simple maximization problem:

$$\alpha^*(\beta), p^*(\beta) : \mu^{(\alpha)}(\beta|p) \longrightarrow \max_{\alpha, p}. \quad (14)$$

We will refer to the series of pairs $(\alpha^*(\beta), p^*(\beta))$ as a prediction scenario, because it determines the choice law of the type α of AR-estimate for any AR-order p and month β . Prediction scenarios for each river are given in Table 2. The lines marked by a symbol μ contain the prediction efficiency for each month and the

overall efficiency γ_0 and γ_1 obtained as a result of the use of the given prediction scenario. The last column of Table 2 contains the values of γ_m for comparison with γ_0/γ_1 . In this case, $\gamma_0 \sim \gamma_m \gamma_1$, since the ratio γ_m was calculated for the time interval $[\tau + 1, N]$, while equation (13) was written for time interval $[L_0 + 1, N]$. The value γ_1 shows the prediction efficiency for the stochastic component.

It is worth mentioning that scenarios (14) in Table 2 were obtained by processing the entire observational intervals for each river. Therefore, such scenarios can be of use for prediction in the future (prediction efficiency estimates are available for them). At the same time, it would be of interest to test the forecast algorithm using the standard method: dividing the observational interval into two equal parts and using one of them to find a scenario and the other for its examination. The principal obstacle for such experiments is the short length of the observational intervals. The duration (the number of observations used) for the AR-PCP-estimate is the number of years N_Y . To initialize AR-predictor (i.e., to evaluate AR-coefficients), we will use the initial time interval with a length of $n_0 = 30$ years, which seems not sufficient for a reliable AR-PCP-estimate, especially, in the case of large p values. This is why n_0 cannot be reduced. However, if we divide the interval in half, we will have only $(N_Y/2 - n_0)$ observations for the search for a scenario in the first part and for studying its quality in the second part. Taking into account that $N_Y = 84-134$ years, we obtain 12-37 observations for the latter number rather than 54-104 ($N_Y - n_0$ for the results given in Table 2).

Nevertheless, a validation experiment was made for five rivers (the Irtysh, Northern Dvina, Oka, Elbe, and Loire) with the longest observational intervals. The observational intervals were divided into two approximately equal parts (not strictly equal, because the first and second parts have to begin from January). Table 3 gives the results of validation.

The application of prediction scenarios derived from the first half of the observational interval to the second half yielded analogous results for all months and for all full efficiencies. Comparison of Tables 2 and 3 shows that the scenarios and efficiencies sometimes can radically differ (because of the small number of samples), though the efficiencies, especially full ones, are almost the same. This fact demonstrates certain stability of the prediction results, given in Table 3.

ARTIFICIAL NEURON NETWORK MODEL

Let us apply the method of artificial neuron network to the problem of prediction of time series $Y(t)$ —a stochastic component of river runoff. We will make an ANN-prediction in a moving time window with a length of L and compare it with an ordinary AR-prediction in the same time window (288 observations or 24 years). The need to use moving time window is

Table 2. Prediction scenarios and its efficiency

River	β	1	2	3	4	5	6	7	8	9	10	11	12	γ_m	γ_1	γ_0
Amur	α^*	1	0	1	1	0	1	0	1	0	0	1	1	4.38	1.73	8.23
	p^*	10	2	12	1	12	5	1	1	3	3	2	5			
	μ	1.42	5.22	5.91	1.13	1.11	1.80	1.71	1.60	1.64	3.20	4.08	2.34			
Irtysch	α^*	1	1	0	1	1	0	0	1	1	1	1	1	4.15	2.56	10.59
	p^*	2	1	1	1	9	10	11	4	2	2	1	1			
	μ	5.11	5.74	5.17	1.18	1.64	2.57	4.33	4.60	3.54	2.95	3.42	2.47			
North. Dvina	α^*	0	1	0	0	1	1	1	1	1	1	1	1	5.21	1.27	6.65
	p^*	2	11	4	10	2	2	1	1	1	1	3	3			
	μ	2.52	4.64	3.04	1.01	1.18	1.35	1.03	1.25	1.56	2.07	2.00	2.45			
Oka	α^*	1	1	0	1	1	1	1	1	1	0	1	1	2.95	1.21	3.58
	p^*	1	1	11	1	1	1	1	1	3	4	1	1			
	μ	1.61	1.36	1.02	1.26	1.08	1.15	2.04	1.67	1.48	0.93	1.95	1.48			
Vistula	α^*	0	1	0	1	0	0	0	1	0	0	1	1	1.39	1.40	2.02
	p^*	10	3	6	2	5	4	5	1	10	4	1	4			
	μ	1.59	1.29	1.25	0.96	1.92	1.30	1.33	1.90	1.72	1.40	2.66	3.01			
Elbe	α^*	0	0	0	0	0	0	0	0	1	0	0	1	1.27	1.27	1.61
	p^*	10	1	10	10	1	1	1	9	1	8	5	1			
	μ	1.29	1.17	1.04	1.38	1.28	1.21	1.21	1.27	1.43	1.27	1.56	1.72			
Danube	α^*	0	0	1	0	0	0	0	0	0	0	1	1	1.09	1.41	1.57
	p^*	12	10	5	2	1	5	11	10	9	7	1	5			
	μ	1.31	1.21	1.08	1.46	1.78	1.47	1.14	1.52	2.02	1.45	1.78	2.16			
Loire	α^*	0	0	1	0	0	1	0	0	0	1	1	1	1.47	1.31	2.00
	p^*	12	11	1	3	6	2	1	1	1	1	1	1			
	μ	1.21	1.15	1.07	1.21	1.45	1.93	1.76	1.52	1.41	1.17	1.32	1.81			
Glomma	α^*	1	1	1	0	1	1	0	0	1	1	1	0	3.39	1.21	3.90
	p^*	12	1	2	6	2	1	12	12	1	7	3	5			
	μ	2.62	5.31	2.97	1.06	1.10	0.97	1.33	1.22	1.32	1.26	1.74	1.39			

caused by the computational complexity of the assessment of ANN parameters for an increasing time window (the moving window will require only a “slight modernization” of the current parameters of ANN). We use for comparison only the standard AR-estimate with a constant AR-order $p = 12$. AR-PCP-estimate is not used for this case because the number of observations ($288/12 = 24$) is too small to allow a reliable statistical estimate of AR-coefficients. However, 288 observations enable us to make a standard AR-estimate. Moreover, the maximum AR-order is used ($p = 12$). Such choice does not make AR-prediction worse.

We will try to use a simple ANN architecture with one latent level comprising p neurons to make a forecast of the process for time moment t . For this purpose, we consider q previous terms of the time series, calculate their weighted sum, and use this scalar value as an input for every neuron. Thus, this procedure is based on the formation of p scalar signals:

$$z_j(t) = \sum_{i=1}^q w_{ji}Y(t-i) + c_j, \quad j = 1, \dots, p. \quad (15)$$

Each signal (15) is passed through the nonlinear activation function $f(s)$. The same function

$$\xi_j(t) = f(z_j(t)), \quad f(s) = s/(|s| + 1/2), \quad (16)$$

where $\xi_j(t)$ are neuron output signals, is used for any neuron. We define the forecast value $Y(t)$ as the weighted sum of neuron signals plus the shift parameter:

$$\widehat{Y}^{(n)}(t) = \sum_{j=1}^p \alpha_j \xi_j(t) + \beta, \quad (17)$$

where (n) is ANN predictor. Thus, the complete vector of ANN parameters has the form

$$\theta = (\beta, \alpha_j, c_j, w_{ij}), \quad j = 1, \dots, p; \quad i = 1, \dots, q. \quad (18)$$

Table 3. Verification of prediction scenarios for rivers with longest observational series

River	β	1	2	3	4	5	6	7	8	9	10	11	12	γ_1	γ_0
Irtysh	α^*	1	0	1	0	1	1	0	1	1	1	1	1	2.36	9.31
	p^*	3	1	2	9	3	4	12	5	1	1	2	2		
	μ	4.87	7.73	4.78	1.15	0.98	4.53	7.15	4.74	8.41	1.75	3.21	2.96		
North. Dvina	α^*	0	1	0	0	1	1	1	1	1	1	1	1	1.12	6.80
	p^*	1	11	7	2	2	2	1	1	1	1	3	7		
	μ	1.39	4.86	1.01	1.00	0.99	1.46	1.02	1.64	1.55	1.56	1.00	1.23		
Oka	α^*	1	1	0	1	1	1	1	1	1	1	1	1	1.06	3.02
	p^*	5	2	4	1	2	1	1	1	1	1	2	2		
	μ	1.65	1.65	1.01	1.05	0.91	1.22	1.94	1.77	1.08	1.66	3.77	1.19		
Elbe	α^*	0	1	0	0	1	0	0	0	1	0	0	1	1.36	1.61
	p^*	12	1	10	9	1	1	1	9	2	2	10	1		
	μ	1.68	1.36	1.18	1.44	1.46	1.78	0.87	1.11	1.15	1.34	1.52	1.67		
Loire	α^*	1	0	1	1	1	0	1	0	0	1	1	0	1.23	2.02
	p^*	11	11	1	3	2	2	1	2	10	4	8	1		
	μ	0.92	1.19	1.27	1.22	1.29	2.10	4.26	1.65	1.10	1.08	1.04	1.64		

Vector θ has the dimension $pq + 2p + 1$. Thus, (17) can be written as

$$\widehat{Y}^{(n)}(t) = \widehat{Y}^{(n)}(t|\theta).$$

Vector θ is determined by minimizing the quadratic criterion of the adjustment quality:

$$J(\theta) = \sum_{t=\tau-L+1+q}^{\tau} (Y(t) - \widehat{Y}^{(n)}(t|\theta))^2 \rightarrow \min_{\theta} \quad (19)$$

where τ takes the value L for the first time window of initialization. Hereafter, τ is the time moment corresponding to the right-hand end of the current moving time window, $\tau = L, \dots, N$. Let us suppose that $\widehat{\theta}(\tau)$ is the vector of parameter estimates obtained by solving problem (19).

It should be mentioned that function $J(\theta)$ has, as a rule, a number of local minimums because of the high dimension of θ and a nonlinear character of this function. The problem of searching for the global minimum is difficult to solve. To find the respective minimum (19), let us consider 10^3 random initial values θ from a parallelepiped, containing the zero point, with a side of 2×10^{-3} . For every initial value θ , we implement the gradient method to search for a local minimum with a maximum step of gradient of 10^{-4} , so that when the cost function increases, the step of gradient is reduced by 1/2. The total number of steps of the gradient method is limited either by the number of 10^4 steps or by the moment when the step becomes less than 10^{-8} . The global minimum is chosen from all the local minimums obtained.

Once the estimate $\widehat{\theta}(\tau)$ is obtained for the first time window, this window starts moving from the left to the right with a step of 1. In this case, the previous vector θ

is used as the initial point. Thus, we can determine a “one-step-forward” prediction for each time window: $\widehat{Y}_F^{(n)}(\tau + 1) = \widehat{Y}^{(n)}(n)(\tau + 1|\widehat{\theta}(\tau))$ for $\tau = L, \dots, N - 1$. In this case, the error of the ANN-predictor

$$\delta^{(n)}(\tau) = Y(\tau) - \widehat{Y}_F^{(n)}(\tau), \quad \tau = (L + 1), \dots, N, \quad (20)$$

and the sample estimate of its variance in the moving time window of the same length L depends on the right-hand end of this window:

$$\widehat{\sigma}_F^2(\tau) = \sum_{t=\tau-L+1+q}^{\tau} (\delta^{(n)}(t))^2 / (L - q), \quad \tau = 2L, \dots, N. \quad (21)$$

For the ANN-predictor, we have $L = 288, p = 2, q = 3$; the total number of ANN parameters is 11, that is, approximately the same as that for the AR-predictor.

Now we can compare the four estimates of variance in the moving time window: for the initial time series $X(t)$, for the stochastic component $Y(t)$, for forecast errors of standard AR-estimate with an order of $p = 12$; and for ANN prediction errors (21).

Four plots of prediction variance for the Irtysh, Amur, Danube, and Elbe are shown in Fig. 5. Note that both AR- and ANN-predictors exhibit a decrease in variance, but AR-predictors are the best in all cases. This conclusion holds for other rivers as well. Thus, the statements, regarding the advantage of the ANN-approach to the formation of time series, which can be found in some publications, are not confirmed. Simpler and faster classical methods proved to be more efficient. It appears reasonable to continue discussion of this problem.

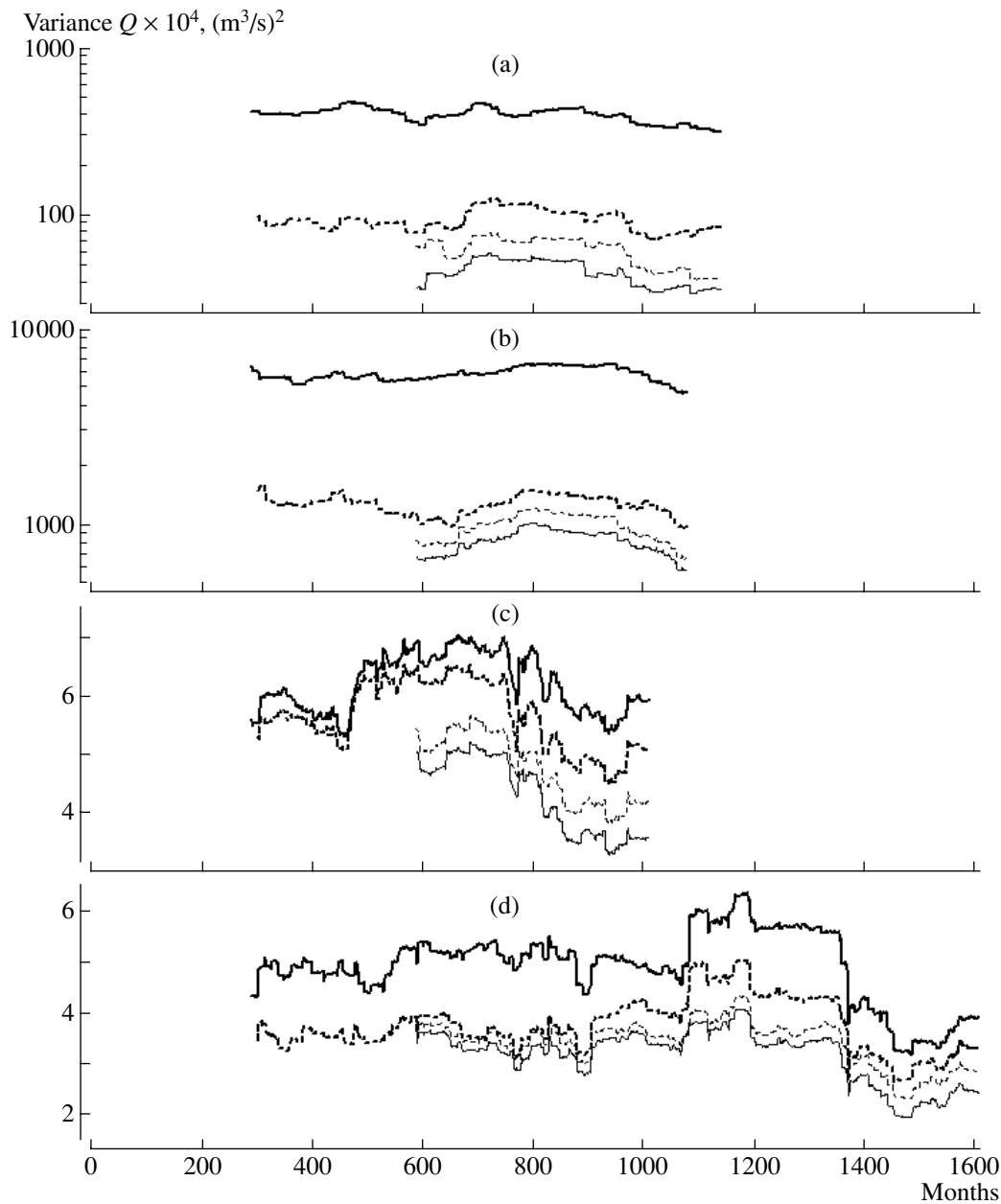


Fig. 5. Estimated prediction variance in moving time window for (a) the Irtysh, (b) Amur, (c) Elbe, and (d) Danube. Full thick lines are for the variance of the initial time series $X(t)$, dashed lines are for variance estimates for the stochastic component $Y(t)$, dash-and-dot lines are for the error of ANN prediction (21); full thin lines are for the errors of standard AR-prediction in a moving time window with an order of $p = 12$. All the curves are given as functions of the right-hand end of the moving time window. Time is given in months from the beginning of observations.

CONCLUSIONS

Statistical analysis allows the assessment of some parameters characterizing the features of the concrete hydrological regime. First, we should mention that the cyclic mean can be used for prediction. The nine rivers can be divided into two groups in accordance with the value of index γ_m . The first group has high seasonal (spring) floods. Such rivers can be called “seasonally controlled.” Contrary to that, the second group consists

of the rivers the regime of which is only slightly dependent on snowmelt floods.

The next characteristic of river regime is determined by the shape of its spectra (Fig. 2). The number of harmonics in the power spectra is different for different rivers. In accordance with this characteristic, the rivers can be divided into three groups. The first group (Loire and Danube) has a single basic harmonic. Maybe, the Elbe can also be included in this group, although, a weak second overtone can be seen in the power spec-

trum. A single convex spectral peak means that the trend resembles the harmonic, which confirms the analysis of Fig. 3. In other cases, the power spectrum can contain the maximum possible number of peaks (in this case, when the cyclic mean for 12 months is expanded in the Fourier series, there are six peaks). Again, all these spectral peaks have almost equal magnitudes. This group of rivers includes the Oka, Northern Dvina, and maybe the Irtysh and Glomma. One peak means that the cyclic mean is somewhat similar to a delta function, that is, one peak out of the 12 monthly runoff values dominates over all others. Finally, there are an "intermediate" number of peaks in the Fourier transformation of the cyclic mean. This is characteristic of the Vistula and Amur. This characteristic of river regime can be called "the spectral complexity of the cyclic mean." The Monthly cyclic mean was simulated in [13] with the help of trigonometric polynomials of the 2–4 orders. Thus, these rivers have an intermediate "complexity." As can be seen from the examples considered above, cases with different degree of "complexity" are also possible.

In general, the predictability of monthly runoffs, estimated with the use of the PCP model and other statistical characteristics gives useful data for both theoretical studies of the hydrological regime of rivers and for practical use.

ACKNOWLEDGMENTS

This study was financially supported by INTAS, project no. 99-00099, National Scientific Foundation EAR, project no. 9804859, and ISTC, project no. 1293-99.

REFERENCES

1. Privalsky, V.E., Panchenko, V.A., and Asarina, E.Yu., *Modeli vremennogo ryada s prilozheniyami k geofizicheskim issledovaniyam* (Time Series Models with Applications of Geophysical Studies), Leningrad: Gidrometeoizdat, 1992.
2. Ratkovich, I.A. and Bolgov, M.V., *Stokhasticheskoe modelirovanie kolebanii komponentov vodnogo balansa rechnogo stoka* (Stochastic Modeling of Variations in Water Balance Components of River Runoff), Moscow: IVP RAN, 1997.
3. Rytov, S.M., *Vvedenie v statisticheskuyu radiofiziku* (Introduction to Statistical Radiophysics), Moscow: Nauka, 1976.
4. Box, G. and Jenkins, G., *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day, 1970.
5. Campolo, M., Soldati, A., and Andreussi, P., Forecasting River Flow Rate During Low-Flow Period Using Neural Networks, *Water Resour. Res.*, 1999, vol. 35, no. 11, pp. 3547–3552.
6. Cheng, B. and Titterton, D.M., Neural Networks: A Review from a Statistical Perspective, *Statistical Sci.*, 1994, vol. 9, no. 1, pp. 2–54.
7. Dawson, C.W. and Wilby, R., Artificial Neural Network Approach to Rainfall-Runoff Modelling, *Hydrol. Sci. J.*, 1998, vol. 43, no. 1, pp. 47–66.
8. Fernandez, B. and Salas, J.D., Periodic Gamma Autoregressive Processes for Operational Hydrology, *Water Resour. Res.*, 1974, vol. 22, pp. 1385–1396.
9. Gardner, W.A. and Franks, L.E., Characterization of Cyclostationary Random Signal Processing, *IEEE Trans. Inf. Theory*, 1975, vol. 21, no. 1, pp. 1–14.
10. Haykin, S., *Neural Networks, a Comprehensive Foundation*, New Jersey: Prentice Hall, 1999.
11. Jayawardena, A.W. and Fernando, A., Use of Artificial Neural Networks in Estimation of Evaporation, *Proc. Int. Conf. on Water Resour. Environ. Res.*, Kyoto, 1996, vol. 1, pp. 141–148.
12. Jayawardena, A.W., Fernando, T.M.K.G., Chan, C.W., and Chan, W.C., Comparison of ANN, Dynamical Systems and Support Vector Approaches for River Discharge Prediction, *19th Chinese Control Conf.*, Hong Kong, vol. 2, pp. 504–508.
13. Kashyap, R.L. and Rao, A.R., *Dynamic Stochastic Models from Empirical Data*, New York: Academic, 1976.
14. Loucks, D.P., Stedinger, J.R., and Haith, D.A., *Water Resource System Planning and Analysis*, Englewood Cliffs: Prentice Hall, 1988.
15. Moss, M.E., Bryson, M.E., Autocorrelation Structure of Monthly Stream Flows, *Water Resour. Res.*, 1974, vol. 10, pp. 734–747.
16. Vecchia, A.V., Periodic ARMA (PARMA) Model with Application to Water Resources, *Water Resour. Bull.*, 1985, vol. 21, pp. 721–730.