MINISTRY OF EDUCATION AND SCIENCE
OF THE RUSSIAN FEDERATION

NATIONAL RESEARCH UNIVERSITY –
NOVOSIBIRSK STATE UNIVERSITY

Department of Mechanics and Mathematics

V. A. Ogorodnikov

M.2-V13 "Stochastic models of meteorological processes"

STOCHASTIC MODELS OF METEOROLOGICAL PROCESSES

Teaching aid

Master Educational Programme for teaching foreign students
in English 'Numerical Statistical Modelling and Simulation.
Monte Carlo Methods'

Discipline M.2-V13

Novosibirsk
2013

Ogorodnikov V. A. Stochastic models of meteorological processes: Teaching aid / National Research University – Novosibirsk State University, 2013. 92 p.

This Teaching aid is published in the framework of master educational programme NRU NSU "Stochastic models of meteorological processes" for teaching foreign students in English. It devoted to development numerical stochastic models of meteorological processes and fields and investigation with help of them different meteorological phenomena.

INTRODUCTION

Methods of statistical simulation of time series and spatial fields are applied nowadays to solve both methodological and practical problems, for example, in meteorology, oceanology, hydrology, etc. In meteorology, statistical simulation is used, in particular, in the statistical analysis of time series and spatial fields of meteorological parameters. The corresponding probabilistic models, as a rule, are used for the interpretation of aerology observation data, probability interpolation of spatial meteorology fields and for probability and statistical forecasting. Among these problems investigation of the probability characteristics of extreme properties of time series takes up an important place, and are used, for example, in the analysis of extreme weather conditions.

The simulated time series and spatial fields also find their direct application in calculations of the dynamic action of physical, in particular, meteorological processes on different systems, structures, objects, etc. Here, of primary interest is the response of a system to a random process, and, moreover, both the response and the action can be of a vector form.

Another field of application of the statistical simulation methods is the forecast of scalar and vector time series. Especially important these methods are in the cases, where it is necessary to obtain conditional probability distributions of the quantity to be predicted.

The methods of statistical simulation of random processes and fields are used in filtration problems, in image processing, in applied problems of the turbulence theory, etc. The problems connected with synthesis of dynamic and probabilistic models are of considerable promise. In these models dynamic and probabilistic features mutually complement each other. The considered list of applications may be essentially extended. In this Chapter, only some approaches to constructing probabilistic models of real processes and fields and their use for solving applied problems are considered.

1. HYDROMETEOROLOGICAL APPLICATIONS OF STATISTICAL SIMULATION METHODS

1.1. Influence of uncertainty in initial data on forecasting accuracy

A limited number of meteorological obserbation stations and the presence of measurement errors, approximate character of boundary conditions, forecasting equations and their solution methods result in the fact that the accuracy of hydrodynamics forecasting appears impossible. For making optimal decisions in these conditions, the apparatus of the random process theory may be used, which is aimed at the construction of mathemetical models for synthesis of hydrodynamics and probability methods of forecasting [25].

One of approaches to realization of this synthesis is to take into account the uncertainty in initial data. The initial data are measured only at observation stations, but the necessary initial information for forecasting must be concentrated on the finite difference grid. In the cases, when observation stations are rather far from each other, the methods of optimal interpolation yield errors, which may substantially affect the forecasting results. In this Section, on an example of the barotropic medium-range forecast, the effect of uncertainty in the initial data on some statistical characteristics of forecasting fields is investigated.

If only the vector $\vec{x}^0 = (x_1^0, x_2^0, ..., x_k^0)^T$ of weather element values at stations is given, then at the initial point of time any value of a random vector $\vec{\xi}^0 = (\xi_1^0, \xi_2^0, ..., \xi_N^0)^T$ from the ensemble X^0 of its all possible values distributed with a conditional probability density $W_0(\vec{X}^0|\vec{x}^0)$ may be realized at N points of a finite difference grid. Two-dimensional fields, given at N points of the regular grid are interpreted as N-dimensional vectors; or two-index enumeration of grid points is substituted for one-index enumeration. By solving a certain forecasting system of equations in the interval (t_0, t) for each point of this ensemble, we will obtain the corresponding ensemble X^t of the forecasting vectors $\vec{\xi} = (\xi_1, \xi_2, ..., \xi_N)^T$, distributed with the conditional probability density $W_t(\vec{X}^t|\vec{x}^0)$. The most general formulation of the forecast is just a specification of the probability density $W_t(\vec{X}^t|\vec{x}^0)$. It is clear that calculation of this density for a non-linear system of forecasting equations

is a fairly difficult problem. However, even if the function $W_t(\vec{X}^t|\vec{x}^0)$ is calculated, nevertheless, for many applications of interest would be a concrete forecast, i.e., only one value of the vector $\vec{\xi}$ from the ensemble X^t. The choice of such a vector, for example, as it was suggested by B.I. Tatarsky [25], is done on the basis of the requirement for the probability density $W_t(\vec{X}^t|\vec{x}^0)$ be maximal. The more general approach consists in choosing a point $\vec{z} \in \vec{X}^t$ from the conditions of minimum of the mean square difference

$$S = \int (\vec{X}^t - \vec{z})^T (\vec{x}^t - \vec{z}) W_t(\vec{X}^t|\vec{x}^0) d\vec{X}^t.$$

The required minimum is reached if $\vec{z}$ coincides with the conditional expectation $\vec{\mu}$ of the vector $\vec{\xi}$ provided that $x_1^0, x_2^0, ..., x_k^0$ are fixed. By using the method of E.S. Epstain [4] in this situation one may show that the system of equations for conditional expectations μ_i and conditional covariances σ_{ij} as applied to the forecasting equation of the type

$$\dot{\xi}_i = a_{ik}\xi_k + b_{ikm}\xi_k\xi_m, \quad \xi_{i|t=t_0} = \xi^0, \quad i = 1, ..., N \tag{1}$$

will have the form

$$\begin{array}{l} \dot{\mu}_i = a_{ik}\mu_k + b_{ikm}(\mu_k\mu_m + \sigma_{km}), \\ \dot{\sigma}_{ij} = a_{ik}\sigma_{jk} + b_{ikm}(\sigma_{jk}\mu_m + \sigma_{jm}\mu_k) + \\ + a_{jk}\sigma_{ik} + b_{jkm}(\sigma_{ik}\mu_m + \sigma_{im}\mu_k), \\ i, j, = 1, ..., N, \quad \mu_{i|t=t_0} = \mu_i^0, \quad \sigma_{ij|t=t_0} = \sigma_{ij}^0. \end{array} \tag{2}$$

In the case when equation (1) is the vorticity equation [12], ξ_i denotes a geopotential at the i-th point of a regular grid, a_{ij} and b_{ijk} are the known coefficients, and points over the variables in the left part of equations (1) and (2) denote the differentiation with respect to time. The summation from 1 to N is made for repeated indices.

System of equations (2) may be solved, in principle, on high-performance computers. In its computer costs this problem is equivalent to solving $N^2/2$ equations of the kind of (1). However, even without allowance for considerable computer costs of this problem, one should remember about the artifical manner of its closure. This circumstance can essentially affect the result of forecasting.

Another method of calculation of conditional expectations and covariances characterizing the optimal forecast is possible. The essence of this method consists in the following [18]:

1. simulation of an ensemble of initial random vectors $\tilde{X} = \{(\xi_1^0, \xi_2^0, \ldots, \xi_N^0)^T\}$ provided that the data at observation stations are fixed;

2. solution of the forecasting problem for each element of the ensemble;

3. statistical processing of the ensemble of forecasting vectors.

Averaging over a sufficient number of independent forecasting vectors gives stable, in the statistical sense, conditional expectations and covariances, so that one may obtain practically acceptable results with greately lesser computer costs than in the numerical solution of equation system (1), (2). Moreover, the realization of this method is not connected with a closure problem for the system of equations for the moments of distribution.

Let us consider a case, when $W_t(\vec{X}^t|\vec{x}^0)$ is the Gaussian probability density, which is determined by two multivariate parameters, namely, the vector of conditional expectations $\vec{\mu}^0$, linearly dependent on $\vec{x}^0 = (x_1^0, x_2^0, ..., x_k^0)^T$, and the conditional covariance matrix

$$\Sigma_{00|1} = \Sigma_{00} - \Sigma_{01}\Sigma_{11}^{-1}\Sigma_{10}, \tag{3}$$

where Σ_{00} is a unconditional covariance matrix of the vector $\vec{\xi}^0$, $\Sigma_{01} = \Sigma_{10}^T$ is a cross-covariance matrix between $\vec{\xi}^0$ and $\vec{x}^0$, Σ_{11} is the covariance matrix of the vector $\vec{x}^0$. Thus, the problem consists in simulation of the Gaussian vector $\vec{\xi} = (\xi_1, \xi_2, ..., \xi_n)^T$ with the expectation

$$\vec{\mu}^0 = \Sigma_{01}\Sigma_{11}^{-1}\vec{x}^0,$$

and the covariance matrix $\Sigma_{00|1}$ of type (3).

Let us consider an approximate method for solving this task with the help of algorithms, described in Chapters 1 and 2.

As applied to equations (1) we assume that $\vec{\xi}^0$ and $\vec{x}^0$ are formed from values of the geopotential and are subvectors of the vector $\vec{\xi} = (\vec{\xi}^{0T}, \vec{x}^{0T})^T$, where $\vec{\xi}^0 = (\xi_1, \xi_2, ..., \xi_N)^T$, $\vec{x}^0 = (x_1^0, x_2^0, ..., x_k^0)^T$. The subvector $\vec{\xi}^0$ is interpreted as a discrete geopotential field on a two-dimensional requlat grid (this grid is determined in some rectangular region D) with the number of points m and n along the corresponding coordinate axes x and y, $N = m \times n$ is the general number of grid points. The subvector $\vec{x}^0$ is interpreted as values of the field $\vec{\xi}$ at observation stations with the coordinates (x_p, y_p), $p = 1, ..., k$. We will also consider that components of the vector $\vec{\xi}$ have the joint Gaussian distribution with zero expectation (i.e., we consider deviations of the geopotential from the corresponding norms), and the joint covariance

matrix

$$\Sigma = \left\| \begin{array}{cc} \Sigma_{00} & \Sigma_{01} \\ \\ \Sigma_{10} & \Sigma_{11} \end{array} \right\|. \tag{4}$$

And, finally, we assume that the field of the geopotential is homogeneous and isotropic with the spatial correlations in the following forms, [6]

$$\rho_{x_1,y_1;x_2,y_2} = \exp\{-\alpha[(x_1 - x_2)^2 + (y_1 - y_2)^2]\}. \tag{5}$$

If (x_1, y_1) and (x_2, y_2) are the coordinates of grid points, then correlations (5) corresponds to the matrix Σ_{00}. If they are the coordinates of observation stations, then correlations (5) correspond to the matrix Σ_{11}. Finally, if (x_1, y_1) are the coordinates of grid points and (x_2, y_2) are the coordinates of the observation stations, then correlations (5) correspond to the matrix Σ_{01}. In calculations, the parameter α was equal to $0.55 \cdot 10^{-3}$, and the step of the regular grid was equal to 300 km.

The algorithm for simulation of conditional fields is reduced to the following operations:

(1a) A realization $\vec{\xi}^0_* = (\vec{\xi}^{0T}_*, \vec{x}^{0T}_*)^T$ of the Gaussian field $\vec{\xi} = (\vec{\xi}^{0T}, \vec{x}^{0T})^T$ with zero expectation and covariance matrix Σ of the form (4) is sampled.

(2a) The field $\vec{\xi}^0_{**}$ at grid points, provided that the vector $\vec{x}^0_*$ of values of field $\vec{\xi}^0_*$ at observation stations is fixed, is built by the method of optimal objective analysis using correlations of the form of (5).

(3a) The realization of field in the form $\vec{\Delta}^0_* = \vec{\xi}^0_* - \vec{\xi}^0_{**}$ is built. The random field $\vec{\Delta}^0$, built in this manner, has the Gaussian distribution with zero expectation and covariance matrix, approximately equal to $\Sigma_{00|1}$.

The initial data for the barotropic forecast are obtained by addition of the background fields $\vec{\mu}^0 = \Sigma_{01}\Sigma_{11}^{-1}\vec{x}^0$, calculated at the grid points by the method of optimal interpolation of the simulated field $\vec{\Delta}^0$.

As a forecasting system of equations, the finite difference analog of the vorticity equation was used with the Jacobian presented by Arakowa's scheme [7]. The two-day forecast at 500 mb surface with the time step $\Delta t = 1.5$ hours was calculated in the bounded region consisting of 22×26 grid points with the spatial step $\Delta h = 300$ km. The background field of the geopotential was built with the help of the linear optimal interpolation method by the measured data at 235 observation stations. Two different synoptical situations were considered. The first situation, corresponding to December, 1, 1970, was characterized by especially pronounced meridional

transfer and powerful cyclonic forms, and the second situation corresponding to March, 27, 1965, was characterized by the pronounced west transfer and non-clear baric forms. An ensemble of initial fields consisted of a hundred independent realizations of the vector $\vec{\xi}^0$. The forecast of conditional expectations $\mu_i(t)$ and mean-square deviations $\sigma_i(t)$ from $\mu_i(t)$ were calculated at each grid points for the moments of time 0(12)48 hours with the climatic standard deviation of the geopotential $\vec{\sigma}(0) = 5$ dkm.

Dependence of the mean-square deviations $\vec{\sigma}(t)$, averaged over all the grid points, on the term of a forecast for the two above-discussed situations are represented in Fig. 1. Note that the growth of mean-square deviations with time for the situation with nonclear baric forms is conspicually less than for the situation with powerful cyclonic forms.

The question about stability of conditional expectations and covariances depending on the number of realizations of the initial vector $\vec{\xi}^0$ was studied, too. Experiments showed that these characteristics averaged over all the grid points remain unchanged for the number of realizations more than 50. A considerably larger number of realizations are required for obtaining stable statistical characteristics at every grid point.

In a similar way we can study the influence of uncertainty in initial data on the forecasting accuracy as applied to more perfect forecasting models. It is necessary to simulate complexes of three-dimensional meteorological fields for this case (for example, field complexes of air temperature, vector speed of winter, etc.) with allowance for their non-homogeneity and non-isotropy.

1.2. On accuracy of temperature vertical profiles expansion into a series by eigenvectors of sampling covariance matrix

For solving some meteorological problems, particularly, the problems connected with hydrodynamic forecasting, the expansion of vertical profiles of meteoelements into a series by eigenvectors of a covariance matrix is used [12,16]. Let $\vec{\psi}_1, \vec{\psi}_2, \ldots, \vec{\psi}_m$ be a set of orthonormal m-vectors forming the basis in m-dimensional space. Then any random m-vector $\vec{\xi}$ may be represented in the form

$$\vec{\xi} = \sum_{i=1}^{m} c_i \vec{\psi}_i, \quad c_i = \vec{\xi}^T \vec{\psi}_i.$$

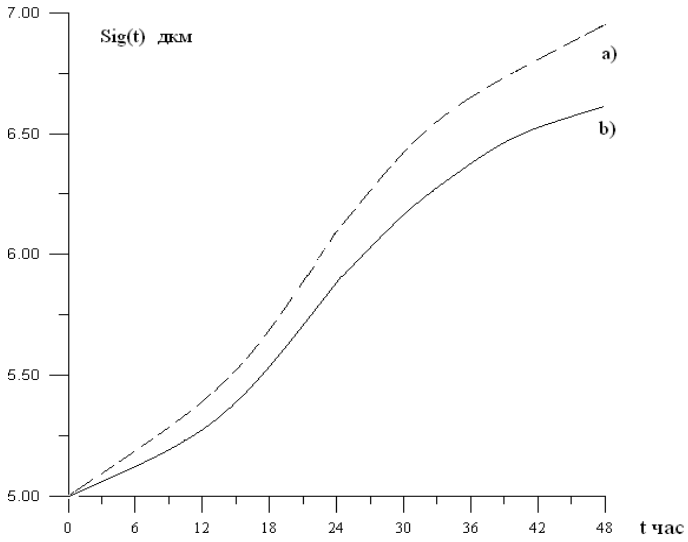

Figure 1: The dependence of the mean-square deviations $\vec{\sigma}(t)$ for two synoptical situations on the term of forecast; $1 - 1.12.1970$; $2 - 27.03.1965$; $\vec{\sigma}(0) = 5$ dkm.

If in this expansion one restricts himself only to the first several terms, then the vector

$$\vec{\xi}_* = \sum_{i=1}^{k} c_i \vec{\psi}_i, \quad k = 1, 2, \ldots$$

will differ from its accurate value in quantity, whose mean square is equal to

$$\sigma^2[k] = M|\vec{\xi}_* - \vec{\xi}|^2 = \sum_{i=k+1}^{m} \vec{\psi}_i^T \Sigma \vec{\psi}_i,$$

where Σ is a covariance matrix of the vector $\vec{\xi}$. It is known [16] that the orthonormal basis is the best in this sense for the expansion of a random vector, if this basis is formed of the eigenvectors $\vec{\alpha}, \vec{\alpha}_2, \ldots, \vec{\alpha}_m$ of the covariance matrix Σ of this vector, respectively, numbered in order of decreasing the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$.

Then, obviously,

$$\sigma^2[k] = \lambda_{k+1} + \ldots + \lambda_m.$$

If a vector $\vec{\xi}$ is formed of values of a certain weather element (for example, geopotential, air temperature, etc.) at different atmospheric levels over an observation station, then, as known [16], several first eigenvalues are predominating in the spectrum $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$, therefore the relative error of a shortened expansion

$$\delta^2[k] = \frac{\sigma^2[k]}{\lambda_1 + \lambda_2 + \ldots + \lambda_m} \tag{1}$$

is small, even if one uses a small number of expansion terms.

In practice, a true covariance matrix Σ of the vector $\vec{\xi}$ is unknown. Therefore the orthonormal basis $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n$ formed of eigenvectors of the sampling covariance matrix

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (\vec{\xi}_i - \vec{\mu})(\vec{\xi}_i - \vec{\mu})^T, \quad \vec{\mu} = \frac{1}{n} \sum_{i=1}^{n} \vec{\xi}_i$$

is used for expansions of such a type, where $\vec{\xi}_1, \vec{\xi}_2, \ldots, \vec{\xi}_n$ is a sample of observations of $\vec{\xi}$. Note that this sample consists, as a rule, of correlated observations, since they are separated from each other by small time intervals, for example, the interval of 6 hours at four-term aerological observations. It seems interesting to elucidate how the sampling nature of the basis

$\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_m$ affects the accuracy of representation of the vector $\vec{\xi}$ in the form of a shortened series

$$\vec{\xi}[k] = \sum_{i=1}^{k} f_i \vec{a}_i, \quad f_i = \vec{\xi}^T \vec{a}_i. \tag{2}$$

The solution of this problem by analytical techniques is of serious difficulty. Even in the case, when $\vec{\xi}$ is a normal vector and observations of it are independent, simple theoretical results for the covariance matrices Σ, differing from the identity matrix, are absent. The problem can be solved by methods of statistical modelling of an "observation" sample of the vector $\vec{\xi}$ provided that the covariance matrix Σ is given.

The mean square of the difference of $\vec{\xi}[k]$ and $\vec{\xi}$ is equal to

$$s^2[k] = M|\vec{\xi}[k] - \vec{\xi}|^2 = \sum_{i=k+1}^{m} \vec{a}_i^T \Sigma \vec{a}_i. \tag{3}$$

Here the operator M denotes averaging over all possible values of a random vector $\vec{\xi}$ with its fixed observation sample, so that $s^2[k]$ is a random variable with probability distribution on a set of all possible samples of the volume n. It is convenient, nevertheless, instead of the quantity $s^2[k]$ to use a relative error

$$d^2[k] = \frac{s^2[k]}{\lambda_1 + \lambda_2 + \ldots + \lambda_m}. \tag{4}$$

It is clear that $d^2[k]$ is a random variable with values from the interval $[0, 1]$. Along with (4) we will also consider the following random variable:

$$\theta[k] = \left| \frac{d[k] - \delta[k]}{\delta[k]} \right|, \quad k = 1, 2, \ldots, m-1. \tag{5}$$

Probability distributions of both variables depend not only on k, but also on the dimension m of the vector $\vec{\xi}$ and on the volume n of a random sample.

Our main task is to find such a minimal volume n of a random sample (provided that the values of k and m are fixed), when $\theta[k]$ with a given, close to unit probability p, will fall within the interval $(0, \varepsilon)$, where ε is a given quantity, for example $\varepsilon = 0.05,\ 0.1$, etc. (this quantity denotes the expansion accuracy). The task was solved as applied to vertical profiles of the air temperature with the values at levels 0.5, 1, 1.5, 2,... 16 km. It was assumed that these profiles were the Gaussian vectors. For simplicity,

but without loss of generality we will consider expectation of the vector $\vec{\xi}$ to be equal to zero. In other words, we will consider a deviation of the air temperature from its normal at the corresponding levels.

As a true covariance matrix, the sampling matrix of covariances of the air temperature at 18 levels for the station Tashkent was used. Now the task reduces to simulation of different volume samples from the normal distribution with the given covariance matrix. The covariance matrix we used was proposed in paper [23]. We will simulate samples formed both of independent and dependent simulated "observations"; we do this in order to investigate the connection affect of a sample for the sampling characteristic accuracy.

Let us consider the case of independent simulated "observations" $\vec{\xi}_1, \vec{\xi}_2, \ldots, \vec{\xi}_n$. Since Σ is a non-stationary (non-Toeplitz) matrix, then each element of the sample is simulated as $\vec{\xi}_i = A\vec{\varphi}$, where $\vec{\varphi}_i$ is a Gaussian vector with zero expectation and unit covariance matrix, A is the lower triangular matrix such that $AA^T = \Sigma$. The realization of this procedure is not difficult, because the dimension of the vector $\vec{\xi}$ is rather small. In our calculations three values $m = 6(6)18$ were assigned to the dimension of this vector. Calculations were made independently for each vector dimension.A thousand of independent from each other samples were used for the calculation of probabilities $P(\theta[k] < \varepsilon)$. It means that the matrix S was calculated for each sample $\vec{\xi}_1, \vec{\xi}_2, \ldots, \vec{\xi}_n$, all eigenvectors and corresponding values of the variable $\theta[k]$ were found; on the basis of one thousand realizations of the variable $\theta[k]$, the probability $P(\theta[k] < \varepsilon)$ was determined.

Some results of calculations are presented in Table 1. The formulated problem is solved with its help. For m and k given we must find in this table such a minimal value of n, beginning with which the probability $P(\theta[k] < \varepsilon)$ becomes greater than the given probability p. For example, if $m = 8$, $k = 5$, it is sufficient to use approximately 60 independent "observations" of the vector $\vec{\xi}$ for the variable $\theta[k]$ with the probability $p = 0.95$ be less than $\varepsilon = 0.1$. As it was expected the probability $P(\theta[k] < \varepsilon)$ provided that n and m are fixed decreases with the growth of k, and provided that n and k are fixed, it increases with the growth of m.

In practice, sampling covariance matrices are calculated by the dependent observation data of the vector $\vec{\xi}$, therefore it is interesting to elucidate how the connectedness of a sample $\vec{\xi}_1, \vec{\xi}_2, \ldots, \vec{\xi}_n$ affects the above-mentioned characteristics. One should expect that the accuracy of the expansion in eigenvectors of a sampling covariance matrix for a connected sample will be worse than for a non-connected sample of the same volume.

Table 1: Probability $P(\theta[k] < \varepsilon)$ for non-correlated samples, $\varepsilon = 0.1$

$m = 18$

k $\backslash n$	20	60	90	120	180
1	0.99	1.00	1.00	1.00	1.00
2	0.95	0.99	1.00	1.00	1.00
3	0.83	0.99	1.00	1.00	1.00
4	0.62	0.99	1.00	1.00	1.00
5	0.33	0.95	0.99	1.00	1.00
6	0.13	0.90	0.99	1.00	1.00
7	0.02	0.79	0.99	1.00	1.00
8		0.62	0.96	0.99	1.00
9		0.42	0.93	0.99	1.00
10		0.21	0.86	0.99	1.00
11		0.08	0.70	0.96	0.99
12		0.02	0.47	0.87	0.99

$m = 12$

k $\backslash n$	20	60	90	120	180
1	0.99	1.00	1.00	1.00	1.00
2	0.97	0.99	1.00	1.00	1.00
3	0.83	0.99	1.00	1.00	1.00
4	0.57	0.96	0.99	1.00	1.00
5	0.33	0.91	0.99	1.00	1.00
6	0.13	0.88	0.99	1.00	1.00
7	0.04	0.75	0.97	0.99	1.00
8	0.03	0.58	0.93	0.99	1.00

$m = 16$

k $\backslash n$	20	60	90	120	180
1	0.99	1.00	1.00	1.00	1.00
2	0.89	0.98	0.99	1.00	1.00
3	0.81	0.99	1.00	1.00	1.00
4	0.57	0.97	0.99	1.00	1.00

Connectedness of the sample $\vec{\xi}_1, \vec{\xi}_2, \dots \vec{\xi}_n$, formed of vertical profiles of the temperature ξ_i with the interval between adjacent measurements equal to 6 hours, is characterized by time-autocovariance and cross-covariance functions $\rho_{k,k;|i-j|} = \text{cov}\ (\xi_i^k, \xi_j^k)$ and, correspondingly, $\rho_{k,l;|i-j|} = \text{cov}\ (\xi_i^k, \xi_j^l)$, where ξ_i^k is a component of the vector $\vec{\xi}_i$. It is assumed that $\vec{\xi}_1, \vec{\xi}_2, \dots, \vec{\xi}_n$ is a stationary connected set of vectors, therefore covariances $\rho_{k,l;|i-j|},\ k, l = 1, \dots, m,\ i, j, = 1, \dots, n$ are formed of the block-Toeplitz matrix $R_{(n)}$ of the form [13, 20] . Concrete values of the discrete functions $\rho_{k,l;|i-j|}$ are presented in paper [23].

The sample $\vec{\xi}_1, \vec{\xi}_2, \dots \vec{\xi}_n$ is simulated by algorithm (5.3), (5.9) considered in the Chapter 1.

For obtaining the final results (Table 2), as before, a thousand independent samples were simulated for the fixed n, m, k. The data of Table 2 show that the connectedness of a sample substantially affects the accuracy of sampling of the covariance matrix and, consequently, the accuracy of representation of $\vec{\xi}$ in the form of (2). Thus, for $m = 18$ and $k = 5$ to fulfil the equality $P(\theta[k] < \varepsilon) = 0.95$ it is necessary to use a sample of the volume $n = 420$, but not $n = 60$ as before. In a similar way we may study other statistical properties of this expansion, for example, expectations, mean-square deviations and probability distribution of the quantity $d[k]$. In addition, there are presented results, connected with the investigation of probabilistic properties of eigenvalues of the sampling covariance matrix.

1.3. Investigation of some features of excursions of air temperature time series

For many applications probabilistic features of strong and long-duration decreases of the air temperature in the cold part of the year is of great interest. It is practically impossible to obtain reliable estimates of these features immediately from observation data, because at most of meteorological stations observations cover a period of a few decades. If, for example, it is required to estimate a probability distribution of the absolute year minimum of the air temperature, then a sufficiently reliable result may be obtained only by using observations during a few centuries. Such long-term observations did not take place even at the "oldest" stations. In these cases, approximate estimations may be got by the statistical simulation of time series, whose

Table 2: Probability $P(\theta[k] < \varepsilon)$ for the correlated samples, $\varepsilon = 0.1$

$m = 18$

k $\backslash n$	30	60	90	120	180	240	360	450	540
1	0.66	0.86	0.94	0.96	0.98	1.00	1.00	1.00	1.00
2	0.39	0.62	0.76	0.83	0.93	0.96	0.99	1.00	1.00
3	0.07	0.22	0.40	0.56	0.78	0.91	0.98	0.99	1.00
4	0.02	0.05	0.16	0.28	0.56	0.76	0.94	0.98	0.99
5		0.01	0.05	0.12	0.36	0.58	0.87	0.97	0.99
6			0.01	0.04	0.16	0.35	0.75	0.95	0.99
7				0.01	0.08	0.22	0.62	0.89	0.97
8					0.02	0.10	0.44	0.84	0.93
9					0.01	0.04	0.28	0.74	0.91
10						0.02	0.18	0.53	0.77
11						0.01	0.09	0.37	0.62
12						0.01	0.05	0.22	0.42

$m = 12$

k $\backslash n$	30	60	90	120	180	240	360	450	540
1	0.68	0.84	0.92	0.94	0.96	0.99	1.00	1.00	1.00
2	0.23	0.54	0.67	0.81	0.95	0.95	0.98	1.00	1.00
3	0.07	0.22	0.41	0.56	0.70	0.82	0.92	0.98	0.99
4	0.01	0.05	0.15	0.29	0.48	0.66	0.84	0.96	0.98
5		0.01	0.06	0.14	0.32	0.52	0.77	0.92	0.96
6			0.02	0.06	0.19	0.37	0.67	0.91	0.96
7				0.01	0.07	0.24	0.52	0.81	0.91
8				0.01	0.07	0.19	0.46	0.73	0.93

$m = 18$

k $\backslash n$	30	60	90	120	180	240	360	450	540
1	0.67	0.85	0.91	0.93	0.97	0.98	0.99	1.00	1.00
2	0.25	0.47	0.61	0.69	0.75	0.83	0.92	0.98	0.98
3	0.10	0.23	0.36	0.48	0.67	0.76	0.90	0.99	1.00
4	0.07	0.13	0.22	0.29	0.48	0.61	0.80	0.94	0.97

properties are close to the corresponding properties of real series. Here we mean only simple properties which one can estimate from available short series of observations, for example, one-dimensional probability distributions and correlation functions. Then the characteristics, we are interested in, are found by the statistical processing of modelling of time series. Note, that to receive some simple characteristics of excursions of meteorological processes, one can use also analytical methods [26]. However it is difficult to use these methods for obtaining intricate characteristics of excursions. One-dimensional probability distributions of most meteorological elements, for example, of the air temperature, as a rule, essentially differ from the normal ones. Moreover, time series of these weather elements are non-stationary, since the number of parameters of one-dimensional distributions try out daily and yearly motions, and time correlations depend not only on the time interval between the correlated variables, but, also, on a specific part of a day or a year, to which this interval belongs. To these series we may refer time series of the air temperature. Only in the winter season it is possible to neglect the non-stationarity of time correlations. As far as one-dimensional probability distributions are concerned, as was shown in [9], they are satisfactorily approximated by a mixture of two normal distributions. Thus, the simulation of temperature time series for a cold part of the year is finally reduced to the simulation of stationary sequences and may be implemented with the help of algorithms from the Chapter 1.

We will simulate the January sequences of the air temperature values $\xi_1, \xi_2, ..., \xi_q$ in a month interval corresponding to six-hour observations at the standard times, i.e., at $01, 07, 13, 19$ o'clock, thus, $q = 31 \times 4 = 124$. In this example the direct simulation is used to estimate such threshold values of the air temperature $C_r(l)$, below which it falls, on the average, once during r years, provided that below this treshold this temperature holds there no less than the $l\Delta t$ hours ($\Delta t = 6$ hours). Thus, we mean the extreme cooling off with the length of a cold wave less than the given one. It is these characteristics of air temperature time series which are interesting for the long-term planning of different economic measures. In this case, the lower bound l of the cold period may be different for the different practical tasks.

If one simulates sequences of monthly length independently, then for some of them the downcrossing under the given level C, or upcrossing from under it, will be beyond these sequences, so that excursions with indefinite beginning and end will have to be excluded. In order to avoid a partial loss of information and at the some time to simplify an algorithm of statistical

processing of simulated sequences, the sequence $\xi_1, \xi_2, ..., \xi_n$ with $n = Nq$, $N = 5000$, was simulated. This is an equivalent to the observation period during 5000 years, but with conservation of the correlation dependence between the values ξ at the boundary of two adjacent years.

Along with the calculation of special values which are necessary for determining the threshold values of the temperature $C_r(l)$, standard deviations of some other characteristics of the time series such as:

1) the number of down crossings below the levels $-10°$C, $-15°$C, etc., on the monthly period of time;

2) duration of the presence of this process under these levels and its probability distribution;

3) the absolute minima of the process within each excursion under these levels

were calculated.

Let us introduce the designations. Assume $\xi_1, \xi_2, ..., \xi_n$ are a simulated sequence of air temperature values, where $n = Nq$. Let us call a random event $E_C^i = (\xi_{i-1} > C, \xi_i \leq C)$ the downcrossing under the level C, and refer it to the element of this sequence with the index i. Let us partition the sequence $\xi_1, \xi_2, ..., \xi_n$ into N subsequences with the number of elements inside each one equal to q. Designate the number of downcrossings of the j-th subsequence, $j = 1, ..., N$, by $\nu_j(E_C; q)$. Then the following quantities will be the mean values and variations of the number of downcrossings on the monthly interval:

$$\begin{aligned} &\bar{\nu}(E_C; q) = \tfrac{1}{N} \sum_{f=1}^{N} \nu_j(E_C; q), \\ &s^2[\nu(E_C; q)] = \tfrac{1}{N} \sum_{f=1}^{N} [(\nu_j(E_C; q) - \bar{\nu}(E_C; q)]^2. \end{aligned} \quad . \tag{1}$$

Let the following random event correspond to each downcrossing under the level C

$$L = L_C^{(i)} = (\xi_{i-1} > C,\ \xi_i \leq C, \ldots, \xi_{i+L-1} \leq C,\ \xi_{i+L} > C). \tag{2}$$

It is clear that $L_C^{(i)}$ is the period (in terms of the time series step) of continuous "residence" of sequence elements, beginning with the element ξ_i, under or at the level C. Let $i_1 < i_2 < ... < i_H$ be numbers of elements of the sequence $\xi_1, \xi_2, ..., \xi_n$, for which the event E_C^i is executed (i.e., a downcrossing under the level C takes place), and $L_C^{(i_1)}, L_C^{(i_2)}, ...L_C^{(i_H)}$ are the

corresponding values of the downcrossing period. In these designations the mean value and variation of a period of the process "residence" under the level C have the following form:

$$\begin{aligned} \bar{L}_C &= \tfrac{1}{H} \sum_{f=1}^{H} L_C^{i_f}, \\ s^2[L_C] &= \tfrac{1}{H} \sum_{f=1}^{H} (L_C^{(i_f)} - \bar{L}_C)^2. \end{aligned} \quad . \tag{3}$$

By the population of quantities $L_C^{(i_1)}, ..., L_C^{(i_H)}$ one can estimate a histogram of probability distribution with the help of conventional methods.

Let us designate the minimal value of the random variable ξ within an excursion under the level C, beginning with $i = i_f$ by $\xi_{i_f}(C)$. Then the mean value and variation of these minimal values are calculated by formulas

$$\begin{aligned} \bar{\xi}(C) &= \tfrac{1}{H} \sum_{f=1}^{H} \xi_{i_f}(C), \\ s^2[\xi(C)] &= \tfrac{1}{H} \sum_{f=1}^{H} [\xi_{i_f}(C) - \bar{\xi}(C)]^2. \end{aligned} \quad . \tag{1}$$

Finally, let us formulate the main task. We partition the sequence $\xi_1, \xi_2, \ldots, \xi_n$ into $N_r = N/r$ subsequences with the number of elements equal to qr in each of them. We may interprete these subsequences as a population of air temperature observations (for example, the January temperature) during r years. Designate the number of process downcrossings under the level C at the j-th subsequence by $\nu_j(E_C, N_r|L > l)$, provided that the period of the downcrossing is not less than the given values l. Then the mean value and variation of the number of downcrossings of this type are calculated by the formulas:

$$\begin{aligned} \bar{\nu}(E_C, N_r|L > l) &= \tfrac{1}{N_r} \sum_{j=1}^{N_r} \nu_j(E_C, N_r|L > l), \\ s^2[\nu(E_C, N_r|L > l)] &= \tfrac{1}{N_r} \sum_{j=1}^{N_r} [\nu_j(E_C, N_r|L > l) - \bar{\nu}(E_C, N_r|L > l)]^2. \end{aligned} \quad . \tag{2}$$

It is clear that with decreasing the level C (in the direction of the low negative temperatures) the mean number of downcrossings $\bar{\nu}(E_C, N_r|L > l)$ must monotonously decrease. Hence, the sought for temperature threshold $C_r(l)$ will be the solution of equation

$$\bar{\nu}(E_C, N_r|L > l) = 1 \tag{6}$$

with respect to C. The corresponding value of the variation is obtained by substitution of the determined value $C_r(l)$ instead of C into the second formula of (5).

1.4. Probabilistic model of air temperature time series.

Here we consider a probabilistic model of the air temperature time series described in paper [9]. According to this model, values of temperature are presented in the form

$$\begin{array}{l} \xi_i = \omega_i(\mu_{1i} + \sigma_{1i}\eta_i) + (1-\omega_i)(\mu_{2i} + \sigma_{2i}\eta_i), \\ i = 1, \ldots, n. \end{array} . \quad (7)$$

Here the sequences $\omega_1, \omega_2, ..., \omega_n$ and $\eta_1, \eta_2, ..., \eta_n$ are independent. The first of them is a stationary Markov sequence of indicator variables with numerical parameters

$$\begin{array}{l} P(\omega_i = 1) = \theta, \quad P(\omega_i = 0) = 1 - \theta = \theta', \\ \beta_h = \text{corr}\ (\omega_i, \omega_{i+h}) = \exp\ (-\beta h), \quad h = 0, 1, \ldots . \end{array} . \quad (8)$$

Hence, the matrix of one-step transition probabilities consists of the following elements:

$$\begin{array}{l} P(\omega_i = 1|\omega_{i-1} = 1) = \theta + \theta'\beta, \quad P(\omega_i = 0|\omega_{i-1} = 1) = \theta'(1-\beta_1), \\ P(\omega_i = 1|\omega_{i-1} = 0) = \theta(1-\beta_1), \quad P(\omega_i = 0|\omega_{i-1} = 0) = \theta' + \theta\beta_1. \end{array} .(9)$$

The stationary sequence $\eta_1, \eta_2, ..., \eta_n$ was formed from the Gaussian random variables with zero expectations and unit variations, connected with each other by correlations

$$\alpha_h = \text{corr}\ (\eta_i, \eta_{i+h}) = \exp\ (-ah^\alpha), \quad h = 0, 1, \ldots . (10)$$

The parameters $\mu_{1i}, \mu_{2i} (\mu_{2i} \geq \mu_{1i})$, $\sigma_{1i} > 0$, $\sigma_{2i} > 0$ in the general case have a daily motion and are calculated by the special method, stated in paper [89]. According to (7), the probability distribution of the air temperature for each j is determined by the mixture

$$P_{\xi_i}(x) = \theta N(x|\mu_{1i}, \sigma_{1i}^2) + \theta' N(x|\mu_{2i}, \sigma_{2i}^2) (11)$$

of two normal distributions with the expectations μ_{1i}, μ_{2i} and variations $\sigma_{1i}^2, \sigma_{2i}^2$, respectively.

Let us take into account the fact, that in view of a daily motion of distribution parameters in (11), the sequence $\xi_1, \xi_2, ..., \xi_n$ will be non-stationary: the correlation coefficient between ξ_i and ξ_j is equal to

$$\begin{array}{l} \rho_{ij} = \mathrm{corr}\,(\xi_i, \xi_j) = (f_i f_j + p_i p_j \beta_{|i-j|})\alpha_{|i-j|} + g_i g_j \beta_{|i-j|}, \\ f_i = \frac{\theta\sigma_{1i} + \theta'\sigma_{2i}}{\sigma_i^2}, \quad p_i = \frac{\sigma_{2i} - \sigma_{1i}}{\sigma_i}\sqrt{\theta\theta'}, \quad g_i = \frac{\mu_{2i} - \mu_{1i}}{\sigma_i}\sqrt{\theta\theta'}, \quad .(12) \\ f_i^2 + p_i^2 + q_i^2 = 1, \quad \sigma_i^2 = D\xi_i. \end{array}$$

The calculations were carried out as applied to the meteorological station "Kiev", where, according to [9], one-dimensional probability distributions of the air temperature are satisfactorily described by the mixture of two normal distributions. The input parameters for the model are presented in Table 3.

Table 3: The input parameters for the model of air temperature time series, Kiev, January

P $\backslash t$	01	07	13	19
μ_{1i}, °C	-10.3	-10.8	-8.2	-9.5
μ_{2i}, °C	-1.0	-1.4	-0.1	-0.7
σ_{1i}, °C	6.2	6.5	5.8	5.9
σ_{2i}, °C	2.2	2.3	2.3	2.1

$$\theta = 0.54, \quad \beta = 0.015, \quad a = 0.056, \quad \alpha = 1.4.$$

Thus, only the daily motion of distribution parameters in (11) is taken into consideration, because no changes take place at the expense of the yearly motion in January.

A satisfactory agreement of simulated and observed time series is verified by the data from Tables 4 and 5.

In Table 4, the mean number of downcrossings $\bar{\nu}(E_C; q)$ under the levels $C = -10, -15, 20$ (°C) on the monthly interval, the standard deviations $s[\nu(E_C, q)]$ of the number of downcrossings per month, the mean "residence" period $\bar{L}_c$ of temperature under these levels and the standard deviations $s[L_C]$, the mean value $\bar{\xi}(C)$ of the temperature minimum within excursions and the corresponding standard deviations $s[\xi(C)]$ are presented.

During the 27-years period, the air temperature in Kiev was falling lower $-25°$C very rarely, therefore statistical characteristics appeared relatively reliable for the values C not lower $-20°$C.

Table 4: Statistical characteristics of air temperature excursions, computed by the observation data, Kiev, January, period 1936 – 1962

C_j °C	$\bar{\nu}(E_C;q)$	t$s[\nu(E_C;q)]$	L_C	$s[L_C]$	$\xi(C)$	$s[\xi(C)]$
-10	4.1	t2.9	6.3	9.1	-14.4	4.5
-15	3.4	t4.0	3.5	3.8	-18.0	3.2
-20	1.1	t1.9	2.5	3.4	-22.8	2.6

Table 5 contains similar characteristics, computed on the basis of 5000 simulated sequences of monthly length ($q = 124$) using the input parameter from Table 3. One can see that the real and the simulated characteristics are satisfactorily consistent with each other. Note that the difference between values of the quantity $s[\nu(E_c;q)]$ at $C = -10°$C takes place because in order to calculate these quantities by observation data, a sample, consisting of only 27 elements was utilized.

Table 5: Statistical characteristics of air temperature excursions, computed by simulated samples, Kiev, January

C_j °C	$\bar{\nu}(E_C;q)$	$s[\nu(E_C;q)]$	L_C	$s[L_C]$	$\xi(C)$	$s[\xi(C)]$
-10	4.7	5.1	6.8	8.2	-14.8	4.7
-15	3.4	3.4	3.9	4.3	-18.2	3.3
-20	1.3	1.7	2.6	2.5	-22.3	2.5
-25	0.3	0.7	1.9	1.4	-26.7	1.7
-30	0.02	0.2	1.4	0.7	-31.2	1.0
-35	0.0	0.0	1.0	1.0	-35.9	0.4

1.4.1.Results of simulation.

We see from (7) that the simulation of the sequence $\xi_1, \xi_2, \ldots, \xi_n$ reduces to the simulation of two independent sequences, i.e., the stationary Markov

chain $\omega_1, \omega_2, ..., \omega_n$ with states 0 and 1 and the stationary Gaussian sequence $\eta_1, \eta_2, \ldots, \eta_n$. The Markov chain is simulated by the well-known algorithms presented. As for the second one of these sequences, the method of its simulation was considered in Chapter 1. In the simulation procedure we took $\beta_{k+1}[k+1] = 0$ for $k \geq m = 40$ with the following use of the corresponding autoregressive procedure, because the correlation function α_h (10) for $h > m \neq 40$ is practically equal to zero.

Experiments show that the quantity $s[\nu(E_C, N_r|L > l)]$ is close to $\bar{\nu}(E_C, N_r|\ L > c)^{1/2}$ for all c, r and l, and, in addition, the order of this closeness is increasing with the growth of all three parameters. One should expect it, because here we deal with rare events, and according to the asymptotic theory [26], the distribution of the number of downscrossings under sufficiently low levels C must approach to the Poisson distributions.

1.5. Probabilistic models of dry and rainy days time series

According to the observation data, time series of daily sums of the rainy precipitation look like an alternation of dry and rainy periods, with their summary number R_t fixed on the days, when the precipitation is falling. Let us agree that the days with traces of precipitation relate to dry ones. There are observed, in principle, two random processes: the first one describes the alternation of dry and rainy days, the second one is the amount of fallen precipitation. At first let us consider the former one.

Let χ_t^*, $t = 1, 2, \ldots$ be a random indicator time series with the values $\chi_t^* = 1$ (rainy days), and $\chi_t^* = 0$ (dry days), where realizations marked with the asterisk are given by observations. During one calendar month this time series is stationary, although its characteristics may change from month to month.

Here we consider three probabilistic models of the process χ_t^* [93]. The three-connected Markov chain χ_t, $t = 1, 2, \ldots, T$ will be the first model for χ_t^*, where T is as large as one likes. As the second one a model of the type $\chi_t = \omega_t \zeta_t$ will be considered, where ω_t and ζ_t are the mutually independent stationary indicator time series of a special type. And, at last, the third model will be based on the threshold transformation of a specially selected stationary Gaussian process ξ_t. For the verification of the models, the distributions of a continuous duration of rainy and dry periods are used. So, a model is acceptable, if it describes these distributions with

sufficient accuracy. In this Section, the formulas are obtained, connecting the transition probabilities and some values of correlation functions with the characteristics of these distributions. This allows one to fit the model to observed data without difficulty.

For illustration, we use the data of 15-year (1969 – 1983) observations in the warm part of a year (May – October) at the 47 gauging stations in the west plane part of the Novosibirsk region. In this way the concrete inferences about the structure and properties of the models are related to this region. In other climatic zones such models, possibly, will be different, but one can construct them in a similar way. Although the south part of the considered territory is slightly drier, nevertheless, it was considered reasonable to combine the observation data to a single sample and thus to obtain relatively stable statistical characteristics, typical of any point of this nearly homogeneous region. Let us present the observation data for each calendar month in the form of the array $\{\chi_i^*[s]\}$, where $s = 1, 2, \ldots, S$, $(S = 47)$ is the number of gauging stations and $i = 1, 2, \ldots, I$ is the order number of a day within a month (I is equal to 30 or to 31), $j = 1, 2, \ldots, J$, $(J = 15)$ is the number of an observation year. It is clear that $\chi_i^{*j}[s] = 1$ for rainy days and $\chi_i^{*j}[s] = 0$ for dry days.

On the basis of realizations of a monthly length it is impossible to precisely define a true duration of rainy or dry periods, as some periods begin or finish outside the calendar month. We have traced periods beginning before the first date of a calendar month (except for May), but finishing within this month and, also, periods beginning within a month, but finishing after its end date (except for October) by the observation data related to the adjacent month. In May and October, periods with an unknown precise duration were excluded from the consideration. Thus, in the above presented array the indices s and j conserve the previous sense, but the index i becomes the order number of a day within of the "extended month", where $i = 1, 2, \ldots, I'^j[s] \geq I$. For obaining the coordinated estimations of the probability $p_1^* = P(\chi_t^* = 1)$ and the mean durations $M_1^* = ML_1^*$, $M_0^* = ML_0^*$ of rainy and dry periods (there must be $p_1^* = M_1^*/(M_0^* + M_1^*)$, see below), in some cases we had to additionally increase "extended months", so that they begin or finish with periods of different types. It was done by addition of one more period of the opposite type to a series beginning or finishing with periods of the same type to that side of a month, where the bound of the new "extended month" is closer to the calendar bound. In May and October these additions were made at the expense of observation data after May and before October. In order

to prevent a loss of long series in the beginning of May and at the end of October, they were cut off.

Finally, the index i has the values $i = 1, 2, \ldots, I^j[s] \geq I'^j[s] \geq I$. All empirical characteristics presented in Table 6, on the whose basis the above-listed models will be built, were obtained with the help of such extended samples.

As in practice sometimes it is impossible to observe a series outside the calendar month, one must make estimates within this month. It is interesting to compare of the empirical characteristics, obtained by these two techniques. As one would expect, the artificial closure of a series results in decreasing the mean values and variations of the duration of dry and rainy periods, and this closure is particularly perceptible for the dry periods. Experiments show that a relative error for the case of dry series is equal to 10% for mean values of the duration, and 30% – for the corresponding variances.

As an estimate for $p_1^* = P$ $(\chi_t^* = 1)$ the relative frequence $\bar{p}_1^* = N_1/N$ was taken, where $N = \Sigma\Sigma I^j[s]$ is the size of the whole indicator sample, and N_1 is the number of units (or rainy days) in it. An estimate for $p_0^* = P(\chi_t^* = 0) = 1 - p_1^*$ can be $\bar{p}_0^* = 1 - \bar{p}_1^*$. Let L_1^* be a continuous duration of a rainy period, and $\{L_1^{*\nu}\}$, $\nu = 1, 2, \ldots, n_1$ be a population of values of this quantity corresponding to the whole indicator sample, i.e., the sample based on observations of all the stations. Then

$$\begin{array}{l} \bar{M}_1^* = \frac{1}{n_1} \sum\limits_{\nu=1}^{n_1} L_1^{*\nu}, \quad \bar{D}_1^* = \frac{1}{n_1} \sum\limits_{\nu=1}^{n_1} (L^{*\nu} - \bar{M}_1^*)^2, \\ \bar{P}(L_1^* = k) = \frac{n(L_1^* = k)}{n_1}, \quad k = 1, 2, \ldots, \end{array} .$$

where $n(L_1^* = k)$ is the number of cases with the event $L_1^* = k$. The estimates $\bar{M}_0^*$, $\bar{D}_0^*$, $\bar{P}(L_0^* = k)$ are similarly calculated by the population $\{L_0^{*\nu}\}$, $\nu = 1, 2, \ldots, n$ of the duration of dry periods. In our sample, due to its special construction, first $n_1 = n_0$, and second $\bar{P}_1^* \equiv \bar{M}_1^*/(\bar{M}_1^* + \bar{M}_0^*)$.

1.5.1. Three-connected Markov model

The model of an indicator time series is, in principle, a mathematical model of the joint probability distribution $P(\chi_1 = \alpha_1, \chi_2 = \alpha_2, \ldots)$, where α_1, α_2 independently take values 1 and 0. We will assume that the simulated series is stationary in a narrow sense. It means that the probabilities $P(\chi_t = \alpha_t, \chi_{t+1} = \alpha_{t+1}, \ldots, \chi_{t+h} = \alpha_{t+h})$ do not depend on t for any $h \geq 0$. The conventional formula

$$\begin{array}{l} P(\chi_1 = \alpha_1, \chi_2 = \alpha_2, \ldots) = P(\chi_1 = \alpha_1) P(\chi_2 = \alpha_2 | \chi_1 = \alpha_1) \ldots \\ \ldots P(\chi_{t+h} = \alpha_{t+h} | \chi_1 = \alpha_1, \chi_2 = \alpha, \ldots, \chi_{t+h-1} = \alpha_{t+h-1}) \ldots \end{array}$$

Table 6: Empirical characteristics of dry and rainy days time series

Month	$\bar{p}_1^*$	M_1^*, days	D_1^*, days	M_0^*, days	D_0^*, days
May	0.31	1.82	1.70	4.04	15.08
June	0.27	1.87	1.88	4.99	27.49
July	0.32	2.07	2.60	4.30	19.87
August	0.32	2.01	2.34	4.17	17.89
September	0.30	2.05	2.33	4.85	24.22
October	0.40	2.40	3.85	3.66	16.25

Table 7: Probability distributions of duration of rainy and dry periods, %

$\bar{P}(L_0 = k)$

Month \k	1	2	3	4	5	6	7	8	9	10
May	56.8	24.1	9.0	5.3	2.0	1.4	0.8	0.4	0.0	0.1
June	55.4	22.8	12.2	4.3	2.5	1.2	0.8	0.3	0.2	0.1
July	52.0	22.8	11.1	5.7	3.5	2.1	1.4	0.7	0.4	0.1
August	52.7	23.2	11.1	5.7	3.3	1.8	1.0	0.7	0.3	0.1
September	50.9	21.5	14.6	6.0	3.6	1.8	0.7	0.4	0.1	0.2
October	45.3	19.7	14.6	10.6	4.2	1.7	1.5	1.5	0.3	0.6

$\bar{P}(L_0 = k)$

Month \k	1	2	3	4	5	6	7	8	9	10
May	28.6	18.7	13.9	8.4	6.0	6.1	3.6	2.7	3.6	1.7
June	24.4	22.3	10.5	7.5	4.7	3.8	5.1	3.9	3.1	2.2
July	28.1	18.7	11.6	8.9	7.4	6.0	4.5	3.0	2.5	1.9
August	29.6	18.1	11.1	9.3	7.7	5.8	3.6	2.8	2.3	1.6
September	29.4	14.8	11.4	9.6	5.4	3.9	3.6	2.7	2.6	3.1
October	36.3	17.9	11.6	9.7	5.4	4.7	3.3	2.4	1.1	1.0

allows us to simulate time series realizations, if the concrete values of probabilities on its right-hand side are given.

In this Section, the method of approximation of the indicator time series χ_t^* by three-connected Markov chains will be considered. The idea of approximation of real sequences, consisting of dry and rainy time series by the Markov chain, is, of course, not new. The method of selecting the connectedness depth and transition probabilities of such a chain is of practical importance. Some relations connecting in the stationary case the quantities $P(\chi_t = 1)$ and transition probabilities of the first three orders $P(\chi_{t+1} = 1|\chi_t = \alpha_t)$, $P(\chi_{t+2} = 1|\chi_t = \alpha_t,\ \chi_{t+1} = \alpha_{t+1})$, $P(\chi_{t+3} = 1|\chi_t = \alpha_t,\ \chi_{t+1} = \alpha_{t+1},\ \chi_{t+2} = \alpha_{t+2})$ with simple distribution characteristics of the duration of dry and rainy periods will be presented below. These distributions are determined by the formulas:

$$\begin{array}{l} P(L_1 = k) = P(\chi_t = 0, \chi_{t+1} = 1, \ldots, \chi_{t+k} = 1, \chi_{t+k+1} = 0)/ \\ \qquad\qquad P(\chi_t = 0, \chi_{t+1} = 1), \\ P(L_0 = k) = P(\chi_t = 1, \chi_{t+1} = 0, \ldots, \chi_{t+k} = 0, \chi_{t+k+1} = 1)/ \\ \qquad\qquad P(\chi_t = 1, \chi_{t+1} = 0), \quad k = 1, 2, \ldots \end{array} .$$

Setting

$$\begin{array}{l} p_{h+1} = P(\chi_t = 1, \chi_{t+1} = 1, \ldots, \chi_{t+h} = 1), \\ q_{h+1} = P(\chi_t = 0, \chi_{t+1} = 0, \ldots, \chi_{t+h} = 0), \quad h = 0, 1, 2, \ldots \\ p_1 = P(\chi_t = 1), \quad q_1 = P(\chi_t = 0) = p_0 \end{array} \quad (1)$$

we have

$$\begin{array}{l} P(L_1 = k) = \frac{p_k - 2p_{k+1} + p_{k+2}}{p_1 - p_2}, \quad P(L_0 = k) = \frac{q_k - 2q_{k+1} + q_{k+2}}{q_1 - q_2}, \\ P(L_1 \geq k) = \frac{p_k - p_{k+1}}{p_1 - p_2}, \quad P(L_0 \geq k) = \frac{q_k - q_{k+1}}{q_1 - q_2}, \quad k = 1, 2, \ldots . \end{array} \quad (2)$$

Hence $p_1 - p_2$,
$M_0 = \sum\limits_{k=1}^{\infty} kP(L_0 = k) = \frac{q_1}{q_1 - q_2}$.

$It follows from the latter that$

$\mathrm{p}_2 = p_1 \left(1 - \frac{1}{M_1}\right), \quad q_2 = q_2 \left(1 - \frac{1}{M_0}\right)$.

(3)(3)Here and in what follows M_1 and M_0 are taken in the non-dimensional form. It is necessary to multiply M_1 and M_0 by the time step Δt for taking the dimensional values. In our case $\Delta t = 1$ day.

It is not difficult to see that $p_1 - p_2 = q_1 - q_2$, hence in accordance with (3)

$$p_1 = \frac{M_1}{M_1+M_0}, \quad q_1 = p_0 = \frac{M_0}{M_1+M_0}. \ (4)$$

We see that for any stationary indicator series the true (i.e., in the general population) values of the probabilities p_1, p_0 and means M_1, M_0 are connected by relation (4). It is natural to require that the sampling estimates for these quantities be connected with each other by analogy with (4). This was obtained by a special expansion of monthly samples, as it was mentioned above. One may show that for any stationary indicator series (not necessarily Markov's one) the following relations hold:

$$P(\chi_{t+1} = 1|\chi_t = 1) = 1 - \frac{1}{M_1}, \quad P(\chi_{t+1} = 1|\chi_t = 0) = \frac{1}{M_0}; \ .(5)$$

$$\begin{array}{l} P(\chi_{t+2} = 1|\chi_t = 1,\ \chi_{t+1} = 1) = 1 - \frac{1-P(L_1=1)}{M_1-1} \\ P(\chi_{t+2} = 1|\chi_t = 0,\ \chi_{t+1} = 1) = 1 - P(L_1 = 1) \\ P(\chi_{t+2} = 1|\chi_t = 1,\ \chi_{t+1} = 0) = P(L_0 = 1) \\ P(\chi_{t+2} = 1|\chi_t = 0,\ \chi_{t+1} = 0) = \frac{1-P(L_0=1)}{M_0-1}; \end{array} ., (6)$$

$$\begin{array}{l} P(\chi_{t+3} = 1|\chi_t = 1, \chi_{t+1} = 1, \chi_{t+2} = 1) = \\ = 1 - \frac{P(L_1\geq 3)}{M_1+P(L_1=1)-2} = p \\ P(\chi_{t+3} = 1|\chi_t = 0, \chi_{t+1} = 1, \chi_{t+2} = 1) = 1 - \frac{P(L_1=2)}{1-P(L_1=1)} \\ P(\chi_{t+3} = 1|\chi_t = 1, \chi_{t+1} = 0, \chi_{t+2} = 1) = \\ = 1 - \frac{\pi_0-(1-P(L_1=1)-P(L_0=1))}{P(L_0=1)} \\ P(\chi_{t+3} = 1|\chi_t = 1, \chi_{t+1} = 1, \chi_{t+2} = 0) = 1 - \frac{\pi_1}{1-P(L_1=1)} \\ P(\chi_{t+3} = 1|\chi_t = 0, \chi_{t+1} = 0, \chi_{t+2} = 1) = \frac{\pi_0}{1-P(L_0=1)} \\ P(\chi_{t+3} = 1|\chi_t = 1, \chi_{t+1} = 0, \chi_{t+2} = 0) = \frac{P(L_0=2)}{1-P(L_0=1)} \\ P(\chi_{t+3} = 1|\chi_t = 0, \chi_{t+1} = 1, \chi_{t+2} = 0) = \\ = \frac{\pi_1-(1-P(L_1=1)-P(L_0=1))}{P(L_1=1)} \\ P(\chi_{t+3} = 1|\chi_t = 0, \chi_{t+1} = 0, \chi_{t+2} = 0) = \\ = \frac{P(L_0\geq 3)}{M_0+P(L_0=1)-2} = 1 - q; \end{array} .(7)$$

$$\pi_1 = \frac{P(\chi_t = 1,\ \chi_{t+1} = 1,\ \chi_{t+2} = 0,\ \chi_{t+3} = 0)}{p_1 - p_2},$$

$$\pi_0 = \frac{P(\chi_t = 0,\ \chi_{t+1} = 0,\ \chi_{t+2} = 1,\ \chi_{t+3} = 1)}{q_1 - q_2},$$

Here only conditional probabilities of the type of $P(\chi_t = 1|A)$ are written. For the sake of completeness one would write out the probabilities $P(\chi_t = 0|A)$ as well, but it is obvious that $P(\chi_t = 0|A) = 1 - P(\chi_t = 1|A)$.

In equalities (7) the probabilities π_1 and π_0 are presented, which in the general case are not expressed by $P(L_1 = k)$ and $P(L_0 = k)$.

It is possible that in some climatic zones one- or two-connected Markov's chain appears a suitable model for an indicator series. The first of them is determined only by mean M_1 and M_0, and the second one requires also the use of the probabilities $P(L_1 = 1)$ and $P(L_0 = 1)$. The detailed analysis has shown that for the region under study one- and two-connected Markov's models are not good enough: they do not describe the sampling distributions $\bar{P}(L_1^* = k)$ and $\bar{P}(L_0^* = k)$ accurately enough. Let us consider a three-connected Markov chain. To determine this chain in accordance with (7) in addition to M_1, M_0, $P(L_1 = 1)$, $P(L_0 = 1)$ it is, also, necessary to set the values of probabilities $P(L_1 = 2)$, $P(L_0 = 2)$ and π_1 and π_0. Let us require that the following conditions be fulfilled:

$$\begin{array}{ll} M_1 = \bar{M}_1^*, & M_0 = \bar{M}_0^*, \\ P(L_1 = 1) = \bar{P}(L_1^* = 1), & P(L_0 = 1) = \bar{P}(L_0^* = 1), \\ P(L_1 = 2) = \bar{P}(L_1^* = 2), & P(L_0 = 2) = \bar{P}(L_0^* = 1), \end{array} \quad .(8)$$

where in the right parts of these equalities, the real characteristics from Table 7 are presented. Then, according to (4), we have $p_1 = \bar{p}_1^*$, i.e., rainy days in the model appear with the same probability, as they do according to the observation data. Moreover, as

$$P(L_1 \geq 3) = 1 - P(L_1 = 1) - P(L_1 = 2),$$

then, according to (8) $P(L_1 \geq 3) = \bar{P}(L_1^* \geq 3)$ and, similarly, $P(L_0 \geq 3) = \bar{P}(L_0^* \geq 3)$. All these relations are fulfilled independent of these concrete values prescribed to the probabilities π_1 and π_0. It is convenient to choose them so as if a three-connected Markov's chain would behave as a two-connected one. This condition gives the following equality

$$\pi_1 = \pi_0 = (1 - P(L_1 = 1))(1 - P(L_0 = 1)). \quad .(9)$$

As a result the corresponding expressions in (7) are essentially simplified. (It is this simplification that serves a justification of selecting the parameters π_1 and π_0.)

One may show, that requirements (8) are feasible. If the sampling characteristics are calculated by the above-presented method, then after their

substitution into (5) – (7) the corresponding quantities actually have the probability sense: all of them are in the interval $[0, 1]$.

In the case of the three-connected Markov chain $p_{h+1} = p_3 p^{h-2}$, $q_{h+1} = q_3 q^{h-2}$, $h \geq 2$, where p and q are determined by the first and the eighth formulas in relation (7). Taking into account relations (8), the final expressions for the model distributions $P(L_1 = k)$ and $P(L_0 = k)$, $k = 1, 2, \ldots$ take the following form:

$$\begin{array}{l} P(L_1 = 1) = \bar{P}(L_1^* = 1), \quad P(L_0 = 1) = \bar{P}(L_0^* = 1), \\ P(L_1 = 2) = \bar{P}(L_1^* = 2), \quad P(L_0 = 2) = \bar{P}(L_0^* = 2), \\ P(L_1 = k) = \bar{P}(L_1^* \geq 3)(1-p)p^{k-3}, \quad k \geq 3, \\ P(L_0 = k) = \bar{P}(L_0^* \geq 3)(1-q)q^{k-3}, \quad k \geq 3, \\ p = 1 - \frac{\bar{P}(L_1^* \geq 3)}{\bar{M}_1^* + \bar{P}(L_1^* = 1) - 2}, \quad q = 1 - \frac{\bar{P}(L_0^* \geq 3)}{\bar{M}_0^* + \bar{P}(L_0^* = 1) - 2}. \end{array} \quad .(10)$$

Values of the parameters p and $q = 1 - p$ are calculated by the above–described samples, and the calculated probabilities $P(L_1 = k)$, $P(L_0 = k)$ are presented in Table 8. The comparison of these probabilities with the corresponding data from Table 7 shows a satisfactory conformity between these characteristics.

Table 8: Three-connected Markov model. Calculated variances of duration of rainy and dry periods D_1 and D_0.

variance	May	June	July	August	September	October
D_1, day^2	1.72	1.84	2.69	2.23	2.13	3.41
D_0, day^2	14.1	27.4	17.6	16.4	23.5	14.3

Let us also compare the variances for the duration of rainy and dry periods for simulated and real time series. According only to (2), i.e., without any restrictions to a model series, except for its stationarity, we may obtain

$$\begin{array}{l} D_1 = \sum\limits_{k=1}^{\infty} k^2 P(L_1 = k) - M_1^2 = \frac{2u_1 - p_1}{p_1 - p_2} - M_1^2, \quad u_1 = \sum\limits_{k=1}^{\infty} p_k, \\ D_0 = \sum\limits_{k=1}^{\infty} k^2 P(L_0 = k) - M_0^2 = \frac{2u_0 - q_1}{q_1 - q_2}, \quad u_0 = \sum\limits_{k=1}^{\infty} q_k. \end{array}$$

Table 9: Three-connected Markov model.Calculated probabilities of duration of rainy and dry periods, %//

$\bar{P}(L_0 = k)$

Month \k	1	2	3	4	5	6	7	8	9	10
May	56.8	24.1	9.4	4.8	2.4	1.2	0.6	0.3	0.2	0.1
June	55.4	22.8	11.1	5.4	2.7	1.3	0.6	0.3	0.2	0.1
July	52.0	22.8	10.9	6.2	3.5	2.0	1.1	0.6	0.4	0.2
August	52.7	23.2	10.9	6.0	3.3	1.8	1.0	0.5	0.3	0.2
September	50.9	21.5	13.8	6.9	3.5	1.7	0.9	0.4	0.2	0.1
October	45.3	19.7	14.5	8.5	5.0	2.9	1.7	1.0	0.6	0.3

$\bar{P}(L_0 = k)$

Month \k	1	2	3	4	5	6	7	8	9	10
May	28.6	18.7	11.9	9.2	7.1	5.5	4.3	3.3	2.6	2.0
June	24.4	22.3	8.8	7.3	6.1	5.1	4.3	3.6	3.0	2.5
July	28.1	18.7	11.0	8.7	6.9	5.5	4.4	3.5	2.7	2.2
August	29.6	18.1	11.1	8.7	6.9	5.4	4.3	3.4	2.7	2.1
September	29.4	14.8	9.9	8.2	6.7	5.5	4.5	3.7	3.1	2.5
October	36.3	17.9	10.3	8.0	6.2	4.8	3.7	2.9	2.2	1.7

For the three-connected Markov chain we have

$$u_1 = p_1 + p_2 + \frac{p_3}{1-p}, \quad u_0 = q_1 + q_2 + \frac{q_3}{1-q},$$

and, also,

$$p_3(1-p) = (p_1 - p_2)P(L_1 \geq 3),$$
$$q_3(1-q) = (q_1 - q_2)P(L_0 \geq 3).$$

If one expresses p_2 and q_2 by M_1 and M_0 then it follows from (3) that

$$D_1 = 2\,\frac{P(L_1 \geq 3)}{(1-p)^2} - (M_1 - 1)(M_1 - 2),$$
$$D_0 = 2\,\frac{P(L_0 \geq 3)}{(1-q)^2} - (M_0 - 1)(M_0 - 2).$$

The results of calculations, presented in Table 10 confirm the closeness between the model and real distributions from the point of view of these characteristics.

1.5.2. Multiplicative model

Here and in the next Section two special models for the processes χ_t^*, determined by the probabilities p_1, p_0 and first several values of β_h, $h = 1, \ldots, k$ of the correlation function

$$\mathrm{corr}\,(\chi_t, \chi_{t+h}) = \beta_h = \frac{P(\chi_t=1,\, \chi_{t+h}=1) - p_1^2}{p_0 p_1}, \quad (11)$$
$$h = 1, 2 \ldots$$

will be considered. The choice of parameters of these models is based on the requirement

$$p_1 = \bar{p}_1^*, \quad \beta_1 = \bar{\beta}_1^*, \ldots, \beta_k = \bar{\beta}_k^*.$$

As an estimation for β_k, the sampling values, which are calculated by standard formulas for estimation of the correlation function of a stationary sequence, may be used. In [3], the detailed analysis of correlation structure for time series of dry and rainy periods for the considered region is presented. Note that all estimations, presented in this Section are obtained by processing of the same archive data, as in the previous Section, but within the limits of a calendar month. Taking into account a special construction of a sample, let us consider another way of estimating the values of the correlation function, based on some general relations, intrinsic of any stationary indicator time series (not necessarily the Makrov's one). This method allows us to connect the first several correlations $\beta_1, \beta_2, \ldots, \beta_k$ with some simple

probabilistic characteristics of the duration of dry and rainy periods. Below we will confine ourselves to consideration of only two first correlations. Using relations (2), (3), and (11) it is not difficult to show that

$$\beta_1 = 1 - \frac{1}{2}\,(1/(p_0 M_1) + 1/(p_1 M_0) = 1 - \frac{M_1 + M_0}{M_1 M_0}$$

or

$$\beta_1 = 1 - 1/(p_0 M_1), \quad \beta_1 = 1 - 1/(p_1 M_0).$$

The correlation coefficient β_2 is connected with $P(L_1 = 1)$, $P(L_0 = 1)$ and β_1 by the relation

$$P(L_1 = 1) + P(L_0 = 1) = 2 - \frac{1 - \beta_2}{1 - \beta_1},$$

whence follows

$$\beta_2 = (1 - \beta_1)[2 - P(L_1 = 1) - P(L_0 = 1)].$$

To the extent, we can rely upon the estimations $\bar{P}(L_1^* = 1)$, $\bar{P}(L_0^* = 1)$, $\bar{M}_1^*$ and $\bar{M}_0^*$, we will consider

$$\bar{\beta}_1^* = 1 - \frac{\bar{M}_1^* + \bar{M}_0^*}{\bar{M}_1^* \bar{M}_0^*}, \quad (12)$$

$$\bar{\beta}_2^* = (1 - \bar{\beta}_1^*)[2 - \bar{P}(L_1^* = 1) - \bar{P}(L_0^* = 1)] \quad (13)$$

as suitable estimations for β_1 and β_2. The quantities 1, $\bar{\beta}_1^*$, $\bar{\beta}_2^*$ are really the initial values of some correlation function, because the conditions $|\bar{\beta}_1^*| < 1$, $1 - 2\bar{\beta}_1^{*^2} + \bar{\beta}_2^* > 0$, denoting positive definiteness of the corresponding correlation matrix, are satisfied here:

$$1 - 2\bar{\beta}_1^{*^2} + \bar{\beta}_2^* = (1 - \bar{\beta}_1^*)[\bar{P}(L_1^* = 1) + P(L_0^* = 1) + 2\bar{\beta}_1^*] > 0.$$

Results of the analysis of the correlation structure of a sequence of dry and rainy series, presented in [3], show that the correlation functions of this process are well described by the power function

$$\beta_0 = 1, \quad \beta_h = \beta^0 \beta^h, \quad h \geq 1. \quad (14)$$

Constants β^0 and β may be determined from the condition $\beta^0 \beta = \bar{\beta}_1^*$, $\beta^0 \beta^2 = \bar{\beta}_2^*$, hence

$$\beta^0 = \frac{\bar{\beta}_1^*}{\bar{\beta}_2^*}, \quad \beta = \frac{\bar{\beta}_2^*}{\bar{\beta}_1^*}.$$

It is known that the correlation function of the form of (14) is intrinsic of the multiplicative indicator succession $\chi_t = \omega_t\zeta_t$, where ω_t and ζ_t are mutually-independent indicator successions, with indicators $\omega_1, \omega_2, \ldots$ such that

$$\theta_0 = P(\omega_t = 1) = p_1 + p_0\beta^0$$

are independent of each other, and ζ_t is one-connected indicator Markov succession with the parameters:

$$\theta_1 = P(\zeta_t = 1) = \frac{p_1}{p_1 + p_0\beta^0}, \quad \text{corr}\,(\zeta_t, \zeta_{t+h}) = \beta^h.$$

Here numerical parameters are chosen so that

$$P(\omega_t\zeta_t = 1) = p_1, \quad P(\omega_t\zeta_t = 0) = p_0.$$

Thus, if

$$\chi_t = \omega_t\zeta_t (15)$$

is a model of a sequence of dry and rainy days with the above-considered numerical parameters, then the simulated sequence will have the same one-dimensional probability (we mean the state probabilities p_0 and p_1) as the real one, and for $h = 1, 2$ it will also have the same correlation function. Methods of the numerical simulation of sequences (15) are well known. The model probabilities $P(L_1 = k)$ and $P(L_0 = k)$ and, also, the model mean values and variances of the duration of dry and rainy series are presented in Table 10.

Table 10: Model $\chi_t = \omega_t\zeta_t$. Calculated mean values and variances of duration of rainy and dry periods M_1, M_0, D_1 and D_0

variance	May	June	July	August	September	October
M_1, day	1.82	1.88	2.06	1.99	2.05	2.38
M_0, day	4.03	4.96	4.29	4.19	4.83	3.63
D_1, day^2	1.49	1.65	2.19	1.97	2.14	3.26
D_0, day^2	14.4	23.7	18.5	18.1	26.1	15.5

1.5.3. Threshold transformation of the Gaussian sequence

In addition to the above-presented models, some other models for describing the process χ_t are possible. For example, a numerical model, based

Table 11: Model $\chi_t = \omega_t \zeta_t$.Calculated probabilities of duration of rainy and dry periods, %//

$\bar{P}(L_1 = k)$

Month \k	1	2	3	4	5	6	7	8	9	10
May	54.9	24.6	11.2	5.3	2.2	1.0	0.4	0.2	0.1	0.0
June	53.3	25.0	11.6	5.3	2.6	1.2	0.6	0.2	0.1	0.1
July	48.8	24.7	12.5	6.9	3.6	1.6	0.9	0.4	0.3	0.1
August	49.9	25.4	12.4	6.2	3.1	1.5	0.8	0.4	0.2	0.1
September	48.7	24.9	12.8	6.7	3.3	1.8	0.9	0.4	0.2	0.1
October	42.0	24.6	14.0	8.2	4.6	2.9	1.6	0.9	0.5	0.3

$\bar{P}(L_0 = k)$

Month \k	1	2	3	4	5	6	7	8	9	10
May	30.2	17.3	12.0	9.2	7.0	5.2	4.1	3.2	2.5	2.2
June	26.3	14.7	10.9	8.9	7.2	5.8	4.8	4.1	3.1	2.7
July	31.9	15.6	10.8	8.3	6.6	5.4	4.2	3.5	2.7	2.2
August	32.5	16.1	10.9	8.2	6.7	5.1	4.2	3.4	2.4	2.1
September	31.3	14.7	9.8	7.5	6.2	5.2	4.3	3.6	3.0	2.4
October	40.7	16.4	9.8	6.7	5.5	4.3	3.3	2.8	2.1	1.7

on a threshold transformation of the Gaussian process ζ_t with the characteristics $M\zeta_t = 0$, $D\zeta_t = 1$, corr $(\zeta_t, \zeta_{t+h}) = \gamma_h$ is a matter of interest. In this case we set

$$\chi_t = \{1, \quad \zeta_t < c, 0, \quad \zeta_t \geq c. \ (16)$$

Here c is chosen from the condition

$$p_1 = P(\chi_t = 1) = P(\zeta_t \geq c) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{c} e^{-u^2/2} du, (17)$$

where p_1 is the given probability, as in the previous models, $p_1 = \bar{p}_1^*$. The correlation function of the process χ_t, given by (16), is connected with a correlation function of the Gaussian process by the relation

$$\beta_h = \text{corr } (\chi_t, \chi_{t+h}) = 1 - \frac{2}{p_1 p_0} T(c, a_h), (18)$$

where

$$T(c, a_h) = \frac{1}{2\pi} \int\limits_{0}^{a_n} e^{-c^2(1+u^2)/2} \frac{du}{1+u^2}, \quad a_h = \sqrt{\frac{1-\gamma_h}{1+\gamma_h}}$$

is the Owen function, presented in [24]. The use of model (16) requires a preliminary calculation of the function γ_h provided that β_h coincides with the real function $\bar{\beta}_h^*$ at a required number of points.

Table 12: Threshold transformation of the Gaussian sequence. Calculated mean values and variances of duration of rainy and dry periods M_1, M_0, D_1 and D_0

variance	May	June	July	August	September	October
M_1, day	1.82	1.88	2.07	2.02	2.04	2.36
M_0, day	4.00	4.94	4.26	4.12	4.84	3.64
D_1, day^2	1.67	1.81	2.56	2.40	2.58	4.22
D_0, day^2	13.5	22.4	16.5	15.4	22.8	12.3

Table 13: Threshold transformation of the Gaussian sequence.Calculated probabilities of duration of rainy and dry periods, %//

$\bar{P}(L_1 = k)$

Month \k	1	2	3	4	5	6	7	8	9	10
May	57.1	22.7	10.5	5.1	2.4	1.2	0.6	0.3	0.1	0.1
June	55.5	22.9	10.8	5.6	2.6	1.4	0.6	0.3	0.2	0.1
July	51.3	22.7	11.6	6.4	3.8	2.0	1.0	0.6	0.3	0.2
August	53.1	21.8	11.7	6.1	3.2	1.9	1.0	0.6	0.3	0.2
September	53.4	21.3	11.2	6.1	3.7	1.9	1.2	0.6	0.3	0.2
October	48.3	21.0	11.5	7.0	4.2	2.8	1.7	1.2	0.8	0.5

$\bar{P}(L_0 = k)$

Month \k	1	2	3	4	5	6	7	8	9	10
May	30.2	17.3	12.0	9.2	7.0	5.2	4.1	3.2	2.5	2.2
June	26.3	14.7	10.9	8.9	7.2	5.8	4.8	4.1	3.1	2.7
July	31.9	15.6	10.8	8.3	6.6	5.4	4.2	3.5	2.7	2.2
August	32.5	16.1	10.9	8.2	6.7	5.1	4.2	3.4	2.4	2.1
September	31.3	14.7	9.8	7.5	6.2	5.2	4.3	3.6	3.0	2.4
October	40.7	16.4	9.8	6.7	5.5	4.3	3.3	2.8	2.1	1.7

In the general case, a function, found by this technique, may be a non-correlation one, hence in each case one must verify it for positive definiteness. One of convenient methods for such a verification is presented in Chapter 1. For the numerical simulation of the sequence χ_t it is necessary to simulate the Gaussian sequence ζ_t with the correlation function γ_n provided that the former exists, and to make use of relation (16), where c is the solution of equation (17) for the given $p_1 = \bar{p}_1^*$. Algorithms for simulation of the Gaussian sequences with an arbitrary given correlation function at the required number of points are presented in Chapter 1. As in the previous model, in calculations we restricted ourselves to the requirement $p_1 = \bar{p}_1^*$, $\beta_1 = \bar{\beta}_1^*$, $\beta_1 = \bar{\beta}_2^*$. It appeared that the corresponding matrices, consisting of the values $1, \gamma_1, \gamma_2$ for each month, satisfy the condition of positive definiteness. In Table 10. as in the previous cases, the probabilities $P(L_1 = k)$, $P(L_0 = k)$, and, also, M_1, M_0, D_1, D_0 are presented. The comparison of data, presented in Tables 9., 10., 11. with the real data from Table 8. shows that each of these models describes the considered process quite satisfactorily in terms of the presented characteristics, however, the three-connected Markov's model is more preferable. In consideration of other characteristics and, also, in investigation of probability properties of sequences of dry and rainy days for different climatic zones, any of the considered models may be used.

1.6. Probability properties of precipitation amount

In the considered region during all months of the warm period of a year both at individual observation stations and on the territory as the whole, the probability distribution of the precipitation amount has an interesting feature, which is in that the probability density $f(x) = -P'(x)$ has the maximum in the vicinity of the point 0.5 mm and to the left of this point (in the narrow interval) tends to zero [32]. Here $P(x)$ denotes the integral probability distribution

$$P(x) = P(R_t \geq x)|A_R^t) = 1 - P(R_t < x|A_R^t) = 1 - F(x),$$

R_t is the daily sum of precipitation, and A_R^t is a random event denoting the appearance of rainy days. To the right of the maximum the density decrease is much slower, and the density has a visible difference from zero even for $x > 10$ mm. Against the general background of strongly stretched

distribution, the growth of $f(x)$ from zero to maximum appears very abrupt. It is absolutely impossible to mathematically describe such a specific form of the probability density by classical distributions. If an empirical distribution is calculated by a large sample, for example, by summing over the whole region, and it does not contain visible irregularities evoked by random deficit or excess of some variable values, then instead of the search for suitable mathematical expressions for $P(x)$ and $f(x)$ it is preferable to use the cubic spline approximation for an empirical function. In the next Section it will be shown that by such a way one may really obtain rather precise representations for these functions.

As we see from Table 12., the correlation connection between R_t and R_{t+h} is very weak, though a probability dependence between them exists. We may see this from the behavior of the joint probabilities

$$P_k(x) = P(R_t \geq x, R_{t+1} \geq x, \ldots, P_{t+k-1} \geq x | A_R^t \cdot A_R^{t+1} \cdots A_R^{t+k-1})$$

with k increasing. In the absence of dependence we would have

$$P_k(x) = P^k(x),$$

where $P(x) = P_1(x)$ is one-dimensional integral distribution, so that actual values of the $\ln(1/P_k(x))$ would group near the line $k \ln(1/P(x))$, going out from zero, because $P_0(x) \equiv 1$. Numerical experiments show that they are really almost lines (almost, because these curves are usually convex upwards), but we see that the values of these curves are not equal to zero for $k = 0$. We may more precisely describe the observed dependence by the formula

$$P_k(x) = P(x)Q^{k-1}(x|x), \quad k = 1, 2, \ldots, (1)$$

where the function $Q(x|x)$ has the sense of the conditional integral probability

$$Q(x|x) = P\left(R_{t+1} \geq x | R_t \geq x\right).$$

We can obtain a probable presentation of this function if we require that for $k = 3$ the probability $P_3(x)$, calculated by formula (1), be equal to the actual one. We take $k = 3$, but not $k = 2$, for the purpose of at least an approximate account of the convextion upward of the curve $\ln(1/P_k(x))$. Thus, we put $Q(x|x) = \left(P_3(x)/P(x)^{1/2}\right)$. In the first approximation

$$Q(x|x) \approx P(x) + \alpha P(x)\left(1 - P(x)\right)^2, (2)$$

where α is a positive parameter, changing from month to month. It is clear that at none concrete values of z it is impossible to provide coincidence of the left and the right parts in (2) for all x. It seems reasonable to consider this parameter as a free one, and to choose its values when fitting a model for solving practical tasks. All above-described allows us to assume that the events $R_t \geq x_1$, $R_{t+1} \geq x_2 \ldots$ form a homogeneous and one-connected Markov chain. Hence,

$$P(R_t \geq x_1, R_{t+1} \geq x_2, R_{t+2} \geq x_3, \ldots) = P(x_1)Q(x_2|x_1)Q(x_3|x_3)\ldots \quad .(3)$$

In addition

$$Q(y|x) = P(y)[1 + \alpha(1 - P(x))(1 - P(y))], \quad (4)$$

so that the joint distribution of quantities R_t and R_{t+1} is given by the function

$$\begin{array}{l} P(R_t \geq x, R_{t+1} \geq y) = P(R_t \geq x)Q(y|x) = \\ = P(x)P(y)[1 + \alpha(1 - P(x))(1 - P(y))] \end{array} \quad (5)$$

which is a modification of the Humbel-Morqenstern two-dimensional distribution, written here in terms of $P(x)$, but not of $F(x) = 1 - P(x)$.

Note, that the parameter α in its absolute value must not be more than unit, otherwise expressions (4) and (5) will loose the probability sense.

Since

$$Q(0|x) = 1, \quad Q(0|0) = 1, \quad Q(y|0) = P(y),$$

random variables R_t and R_{t+h} for $H \geq 2$ are independent. Actually,

$$\begin{array}{l} P(R_t \geq x, R_{t+h} \geq y) = P(R_t \geq x, R_{t+1} \geq 0, \ldots, R_{t+h-1} \geq 0, R_{t+h} \geq y) = \\ = P(x)Q(0|x)Q(0|0)\ldots Q(0|0)Q(y|0) = P(x)P(y) \end{array}$$

(for simplicity we write $R \geq 0$ instead $R \geq 0.1$). Only the correlation coefficient between the adjacent quantities R_t and R_{t+1} may be different from zero in this scheme. Non-difficult transformations give the following result

$$\alpha_1 = \text{corr}\,(R_t, R_{t+1}) = \alpha \, \frac{\left(\int\limits_0^\infty P(x)(1 - P(x))dx \right)^2}{\sigma^2}.$$

Here σ^2 is the variance of the distribution $P(x)$, i.e., the variance of daily precipitation sums.

For an approximate estimate of expected values of α_1 we take $P(x) = \exp(-\alpha x^b)$ with characteristic of the daily precipitation sums parameter $b = 0.8$. Then

$$\left(\int_0^\infty P(x)(1 - P(x))dx \right)^2 = \mu^2 \left(1 - \frac{1}{2^{1/b}} \right)^2 \approx \frac{1}{3}\,\mu^2,$$

where μ is the mean daily sum. So,

$$\alpha_1 \approx \alpha\,\frac{\mu^2}{3\sigma^2} = \frac{\alpha}{3V^2}.$$

Typical for the region in question is the variation coefficient $V = 1.2$. This gives $\alpha_1 \approx 0.2\alpha$. With allowance for the fact that the parameter α is less than unit, we arrive at a conclusion that the correlation coefficient between R_t and R_{t+1}, is also very small according to the observation data.

Thus, we may expect that the joint distribution (3) will be suitable for time series of the daily precipitation sums. Note, that one-dimensional distribution and only one numerical parameter for the account of probability dependence between elements of the sequence R_t are used.

1.7. Approximation of empirical probability distribution of daily rainy precipitation sums

Let a random variable R_t denote a daily layer of rainy precipitation, $F(x) = 1 - P(x) = P(R_t < x)$ denote a probability distribution function and $f(x) = F'(x)$ denotes a probability density. We will consider that R_t is changing in the interval (x_0, ∞), where $x_0 = 0.1$ mm, so that $F(x_0) = 0$. The real density $f(x)$ is strongly extended to the right, but near the reference point it has the maximum at the point $M_0 \approx 0.5$ mm. To the left of this point, the density tends to zero. Sometimes, as a probability density one uses the gamma-density

$$f(x) = (x - x_0)^{\nu-1} \exp\left(-\frac{x - x_0}{\lambda} \right) \Big/ \lambda\Gamma(\nu)$$

with parameters, estimated either by the maximal likelihood method or the moment method [173]; in the latter case $\nu = 1/V^2$, $\lambda = \mu V^2$, where μ and V are the mean value and the variation coefficient for R_t. In the considered

region $\nu < 1$, hence the gamma-density with $x \to x_0$ tends not to zero, as we see from observations, but to infinity. This function may describe only the left part of the empirical distribution, but not the distribution as a whole. The Veibul distribution is also appropriate as applied to the daily rainy precipitation sums. As for the internal parts of the curve, they may be described by polynomials of the third degree. Here this method is illustrated on the example of the total rainy precipitation for the plane part of the Novosibirsk region. By virtue of the physic-geographical homogeneity of this relatively small region, distributions averaged over all observation stations can be considered characteristic of any point of the territory.

In papers [11] the algorithmic part of this problem was considered. Let an empirical distribution function be given at the points $x_0 \le x_1 < \ldots < x_{m+1} < \infty$, and at these points it has the values $0 = F_0 \le F_1 < \ldots < F_{m+1} < 1$. An approximating function is taken in the form

$$F(x) = \begin{cases} U(x), & x_0 \le x < x_1, \\ F_1(x), & x_i \le x \le x_{i+1}, \quad i = 1, \ldots, m, \\ V(x), & x_{m+1} \le x < \infty. \end{cases} , (1)$$

Here $U(x), V(x)$ are preassigned non-negative and non-decreasing functions (distribution "tails") with the properties

$$U(x_0) = 0, \quad U(x_1) = F_1, \quad V(x_{m+1}) = F_{m+1}, \quad V(\infty) = 1,$$

and $F_i(x)$ are polynomials of the third degree relative to x:

$$\begin{array}{l} F_i(x) = F_i + f_i h_i y_i \left[1 + a_i(1 - y_i) + b_i(1 - y_i^2)\right], \\ h_i = x_{i+1} - x_i, \quad h_i y_i = x - x_i, \quad h_i f_i = F_{i+1} - F_i. \end{array} \quad (2)$$

The approximating function is arranged so that for any values of the constants a_i and b_i it coincides with the empirical distribution function at grid points. Some restrictions are imposed on these constants to provide non-negativeness of the density $f(x)$ for all x. The exact restrictions are presented in paper [94], here we consider simpler sufficient conditions. We should have:

$$0 \le u_i, \quad v_i \le 3, \quad i = 1, \ldots, m (3)$$

and, moreover, u_i and v_i are connected with a_i and b_i by the equalities

$$u_i = -3 + 2u_i + v_i, \quad b_i = 2 - u_i - v_i . (4)$$

The linear relations

$$\begin{array}{l} U'(x_1) = f_1 u_1, \\ f_i v_i = f_{i+1} u_{i+1}, \quad i = 1, \ldots, m-1, \\ f_m v_m = V'(x_{m+1}) \end{array} \quad (5)$$

follow from the requirement that the density $f(x)$ is continuous, and, in addition, the first and the last ones from relations (5) determine u_1 and v_m. Note, that for $x_1 = x_0$ the necessity in setting the left "tail" $U(x)$ disappears, and u_1 is determined from the condition to which the density $f(x)$ should satisfy at the reference point. Require, also, that the total curvature of the approximate curve, given by the integral

$$J = \int\limits_{x_1}^{x_{m+1}} \left[F''(x)\right]^2 dx = \frac{1}{4} \sum_{i=1}^{m} \frac{f_i^2}{h_i} \left(3 + u_i^2 + v_i^2 + u_i v_i - 3u_i - 3v_i\right),$$

be the minimal one (note, that it is the usual condition, which is used for spline interpolation). Bearing in mind that $v_1, v_2, \ldots, v_{n-1}$ are unknown, we come to the problem of minimization of the quadratic functional

$$\begin{array}{l} 4J = \frac{f_1^2}{h_1} (u_1 - 3) u_1 + \frac{f_m^2}{h_m} (v_m - 3) v_m + 3 \sum\limits_{i=1}^{m} \frac{f_i^2}{h_i} + \\ + \frac{1}{2} \sum\limits_{i=1}^{m-1} \sum\limits_{j=1}^{m-1} a_{ij} v_i v_j - \sum\limits_{j=1}^{m-1} c_j v_j \end{array} \quad (6)$$

with restriction (3), which with allowance for (5) takes the form

$$0 \le v_i \le \begin{cases} 3, & f_i < f_{i+1}, \\ 3 f_{i+1}/f_i, & f_i \ge f_{i+1} \end{cases}$$

In (6), a_{ij} and c_j denote

$$a_{ii} = 2 f_i^2 \left(\frac{1}{h_i} + \frac{1}{h_{i+1}}\right), \quad a_{ii+1} = a_{i+1i} = \frac{f_i f_{i+1}}{h_{i+1}},$$

$$a_{ij} = 0 \quad \text{for} \quad |i - j| \ge 2,$$

$$c_1 = \frac{f_1^2}{h_1} (3 - u_1) + 3 \frac{f_1 f_2}{h_2},$$

$$c_j = 3 \left(\frac{f_j^2}{h_j} + \frac{f_j f_{j+1}}{h_{j+1}}\right), \quad j = 2, \ldots, m-2, \quad m \ge 3,$$

$$c_{m-1} = \frac{f_m^2}{h_m}\,(3 - v_m) + 3\,\frac{f_{m-1} f_m}{h_{m-1}}.$$

In the case $m = 3$, interior expressions for c_j disappear, and only exterior expressions for c_1 and c_2 are retained.

The search for values $v_1, v_2, \ldots, v_{m-1}$, minimizing (6) with restrictions (3) is numerically done with the help of linear programming methods. The case, when the minimum is achieved at interior points of a set of permissible values for variables v_j, is important and frequently occurs. Then, first, values $v_1, v_2, \ldots, v_{m-1}$ satisfy the system of linear equations

$$\sum_{j=1}^{m-1} a_{ij} v_j = c_j, \quad i = 1, \ldots, m-1 \quad (7)$$

and, second, the approximating density $f(x)$ is not only continuous, but is continuously differentiable for all $x_i < x < x_{m+1}$. In other words, $F_i(x)$, $i = 1, \ldots, m$ in this case are usual cubic splines. It is reasonable to begin solving this problem with equation (7) and with the check of conditions (3) or (2), and only, when these conditions are broken, it is necessary to consider functional (6) directly.

One may judge about the quality of approximation by the closeness of the calculated moments

$$\mu_a = \int_{x_0}^{\infty} x dF(x), \quad \sigma_a^2 = \int_{x_0}^{\infty} (x - \mu_a)^2 dF(x),$$

$$A_a = \sigma_a^{-3} \int_{x_0}^{\infty} (x - \mu_a)^3 dF(x), \quad E_a = \sigma_a^{-4} \int_{x_0}^{\infty} (x - \mu_a)^4 dF(x) - 3$$

to the corresponding empirical values μ, σ^2, A, E, obtained by observation data. The integration is carried out numerically.

The choice of the distribution "tails" is nonformal in this procedure. Due to its strong stretchness to the right, it is especially important to correctly describe a behavior of $V(x)$ for large x. The accurate coincidence of $V(x)$ with several border values of the empirical distribution function is not worth while, because of their non-sufficient statistical significance, although one may get by such a way an acceptable conformity between all four moments of distribution. It is preferable to move the point x_{m+1} away to the left as far as possible and to select the such function $V(x)$ with a monotonic

derivative $V'(x)$, which would satisfy the condition $V(x_{m+1}) = F_{m+1}$, and for $x > x_{m+1}$, even if $V(x_{m+1})$ is not equal to the empirical F_i, but, in a sense, is close to them. For the daily amount of rainy precipitation it turned out to be possible to take $x_{m+1} = 2$, and as the right "tail" to set the function

$$V(x) = (1 - W(x))^{\alpha}, \quad W(x) = e^{-ax^b} \tag{8}$$

depending on the three parameters $a > 0$, $b > 0$, $\alpha > 0$ (for $\alpha = 1$ this function corresponds to the Veibul distribution). The introduction of an additional parameter α allows one to take into consideration a weak convexness upwards, typical of empirical dependencies $\ln\ln(1-F)^{-1}$ on $\ln x$, on the Veibul rectifying lattice. Let us denote

$$\ln x = y(x), \quad \ln\ln(1 - V^{1/\alpha}(x))^{-1} = z_V(x;\alpha)$$

and rewrite formula (8) in the form

$$z_V(x;\alpha) = b\,y(x) + \ln\ a.$$

Parameters α, a and b are found from the conditions

$$\ln\ a + b\,y(2) = z_F(2;\alpha),$$

$$L = \sum_{x\geq 2} [b\,y(x) + \ln a - z_F(x;\alpha)]^2 = \min,$$

with $z_F(x;\alpha)$ denoting $\ln\ln(1-F^{1/\alpha}(x))^{-1}$, where for each $x \geq 2$ the value $F(x)$ is taken from Table 13.

Because of $\ln a = z_F(2;\alpha) - b\,y(2)$, the problem reduces to minimization with respect to b and a of the expression:

$$L = b^2 G - 2bR(\alpha) + S(\alpha),$$

where

$$G = \sum_{x>2} [y(x) - y(2)]^2,$$

$$R(\alpha) = \sum_{x>2} [y(x) - y(2)]\ [z_F(x;\alpha) - z_F(2;\alpha)],$$

$$S(\alpha) = \sum_{x>2} [z_F(x;\alpha) - z_F(2;\alpha)]^2.$$

Table 15: Comparison of empirical (E) and model (M) values of probability functions $F(x)$ on the right tail (%)

x,	May		June		July		Aug		Sent		Octob	
mm	E	M	E	M	E	M	E	M	E	M	E	M
2	50.4	50.4	40.0	40.0	36.7	36.7	42.5	42.5	46.0	46.0	54.4	54.4
3	63.5	63.1	51.9	51.6	47.6	47.4	53.6	53.6	59.7	59.7	69.1	69.3
6	82.9	83.1	72.2	72.2	67.9	67.6	72.9	73.3	81.0	81.5	88.5	88.5
10	93.3	93.0	85.2	84.9	81.4	81.3	85.5	85.7	92.5	92.3	96.5	96.7
15	97.4	97.4	92.0	92.1	89.4	89.8	92.8	92.8	97.1	97.0	99.2	99.2
20	98.7	98.9	95.4	95.6	94.1	94.1	96.4	96.1	98.7	98.7	99.8	99.8
30	99.8	99.8	98.4	98.4	97.8	97.8	98.6	98.7	99.7	99.7	–	–

For any α, the minimum with respect to b is reached for $b = R(\alpha)G^{-1}$, and in this case

$$\min_b \ L = l(\alpha) = S(\alpha) - R^2(\alpha) \cdot G^{-1}.$$

It remains to numerically construct a dependence $l(\alpha)$ on α and to find such $\alpha = \alpha_0$, for which $l(\alpha_0) = \min\ l(\alpha)$. The calculated probabilities, obtained by formula (7), are compared to the empirical probabilities in Table 6.11. One can see that they are distinguished only in the third decimal place. Using the parameters α, a, b we obtain $V'(x_{m+1})$ at the point $x_{m+1} = 2$ and the bounding value $v_m = V'(2)f_m^{-1}$ for the subsequent polynomial approximation.

Let us consider the left part of the distribution. The version, when $x_0 = x_1 = 0,1$, $x_2 = 0,3$, $x_3 = 0,5$, $x_5 = x_{m+1} = 2$ ($m = 4$) was considered and the polynomial of the third degree was built for each interval between these points. The left tail $U(x)$ is absent and the boundary value u_1 is taken equal to zero, which means $f(x_0) = 0$. In the case under study it turned out to be that it is impossible to obtain a solution with the help of equations (7) (except October), because v_1 takes an inadmissibly large values, surpassing 3: for these values v_1, the density $f(x)$ within the interval $[0.1, 0.3]$ at some points is less than zero. In this case one must minimize directly functional (6) with the restrictions of (3). As was to be expected, for May - September it turned out to be that $v_1 = 3$ (for October $v_1 < 3$), which for $u_1 = 0$ gives $a_1 = 0$, $b_1 = -1$ and, hence within the interval $[0.1, 0.3]$ the following

relations are valid:

$$F(x) = F_1(x) = f_1 h_1 y_1^3, \quad f(x) = F'(x) = 3 f_1 y_1^2,$$

$$y_1 = \frac{x - 0.1}{h_1}, \quad h_2 = 0.2, \quad f_1 h_1 = F_2 \,.$$

In addition, since V_1 is at the boundary of the permissible region, then the density $f(x)$ has no continuous derivative at the point $x_2 = 0.3$.

The moments of approximated and empirical distributions (the latter distributions were calculated directly by the original population of 15-year observations) are compared in Table 14.

Table 16: Comparison of empirical (E) and model (M) moments

	May		June		July		Aug		Sent		Octob	
	E	M	E	M	E	M	E	M	E	M	E	M
μ	3.4	3.5	5.4	5.4	6.1	6.2	5.0	5.1	3.7	3.7	2.7	2.8
σ	4.1	4.3	7.4	7.4	7.9	8.2	6.6	6.9	4.4	4.5	2.9	3.0
A	2.7	3.2	3.1	4.0	2.8	3.5	2.9	3.5	2.8	3.3	2.3	2.5
E	9	18	14	29	11	21	12	21	12	20	7	10

The degree of conformity between the standard deviations is rather high. The coefficients of asymmetry conform with each other somewhat worse, and moreover, it is typical that their calculated values for all months are somewhat higher, than empirical values. As for the coefficients of excess, they are approximately two times as large as empirical ones. The main reason of the difference between empirical and calculated higher moments consists, obviously, in that the right tail of the approximating distribution is formed not exactly enough. In this Section an example of distribution of rather a difficult structure is presented. As noted above, the density of this distribution from the right decreases rather slowly, therefore a contribution of the right tail in the values of coefficients of asymmetry and excess is essential. For the approximation of other weather elements distributions, different functions may be used as tails. For the air temperature rather good results with respect to all four moments give, for example, the left and the right tails in the form of the Gaussian curves. In detail, the accuracy of approximation of one-dimensional distributions by polynomials of the third degree for different weather elements was investigated in the papers

[11]. In the next Section this method is used for approximation of one-dimensional distributions of the mean daily air temperature, the daily total rain precipitation and the daily total solar radiation.

2. PROBABILISTIC MODELS OF JOINT TIME SERIES AND SPATIAL FIELDS OF DIFFERENT WEATHER PARAMETERS

In this Section we present a method of the statistical simulation of non-stationary vector processes on an example of non-Gaussian joint time series of the mean daily air temperature, the daily total rain precipitation and the daily total solar radiation. The non-stationary feature here is undestood in the sense that one-dimensional distributions of each of weather elements and auto-correlations and cross-correlations are some functions of time.

The best known techniques for the simulation of a non-stationary process reduce to the use of spectral parametric models, in which parameters are some functions of time. Covariance functions of such processes depend on two variables (for spatial fields these variables are vectorial), and, in addition, one of them is a time increment. If analytical expressions for such functions are known, then the corresponding spectral representation is selected, and this representation determines the algorithm of simulation. Another way for the simulation of non-stationary processes, for example of non-stationary discrete sequences, is simulation of autoregressive sequences with the coefficients depending on time. In some cases one may use the methods, based on the simulation of periodically correlated non-stationary processes.

For many applications a specific feature of the simulation is that observation data of the investigated process are the original information for a model. The correlation structure estimate is made at fixed time points, so for the use of spectral parametric models, it is necessary to approximate the corresponding correlation matrix by some positive definite analytical functions of the proper number of variables. In practice, the structure of correlation matrices may be rather complicated, so it appears difficult to find a suitable approximate function.

In this Section a simple approximate approach, which allows one to avoid these difficulties and to take into account the change in time in the model not only of correlation properties, but also one-dimensional probability distributions, is considered [19]. This approach is based on a stochastic (randomized) joint of random processes, given on some sequence of disjoint intervals. The process inside each time interval has one-dimensional distribution and correlation structure intrinsic only of this interval.

A special technique of simulation of non-Gaussian joint sequences, one of which is the sequence of quantities of the form of $\chi_t \xi_t$ is considered as

well, where χ_t is an element of a non-Gaussian sequence with the given correlation functions.

2.1. Stochastic joint of independent random vector sequences

For simplicity, we will restrict ourselves to the joint of two random vector sequences of finite length. The joint of several sequences is done by the similar way.

Let us consider two independent of each other non-Gaussian sequences

$$\vec{\xi}_{(k+\nu)} = \left(\vec{\xi}_1, \dots, \vec{\xi}_k, \vec{\xi}_{k+1}, \dots, \vec{\xi}_{k+\nu}\right)^T,$$

$$\vec{\zeta}_{(m+k-\nu)} = \left(\vec{\zeta}_{k+\nu+1}, \dots, \vec{\zeta}_{k+m}, \vec{\zeta}_{k+m+1}, \dots, \vec{\zeta}_{m+2k}\right)^T$$

with the block covariance matrices

$$K^{(1)}_{(k+\nu)} = M\left(\vec{\xi}_{(k+\nu)} - \vec{\mu}^{(1)}_{(k+\nu)}\right)\left(\vec{\xi}_{(k+\nu)} - \vec{\mu}^{(1)}_{(k+\nu)}\right)^T = (\sigma^{(1)}_{ij}),$$

$$i, j = 1, \dots, k+\nu,$$

$$K^{(2)}_{(m+k-\nu)} = M\left(\vec{\zeta}_{(m+k-\nu)} - \vec{\mu}^{(2)}_{(m+k-\nu)}\right)\left(\vec{\zeta}_{(m+k-\nu)} - \vec{\mu}^{(2)}_{(m+k-\nu)}\right)^T = (\sigma^{(2)}_{ij}),$$

$$i, j = k+\nu+1, \dots, m+2k,$$

$$\vec{\mu}^{(1)}_{(k+\nu)} = M\vec{\xi}_{(k+\nu)}, \quad \vec{\mu}^{(2)}_{(m+k-\nu)} = M\vec{\zeta}_{(m+k-\nu)}.$$

Here $\vec{\xi}_i$, $i = 1, \dots, k+\nu$ and $\vec{\zeta}_j$, $j = k+\nu+1, \dots, m+2k$ are p–vectors $\vec{\xi}_i = (\xi_{i,1}, \dots, \xi_{i,p})^T$, $\vec{\zeta}_j = (\zeta_{j,1}, \dots, \zeta_{j,p})^T$ with the means $M\vec{\xi}_i = \vec{\mu}^{(1)}$, $M\vec{\zeta}_j = \vec{\mu}^{(2)}$, one-dimensional probability distributions

$$\vec{F}^{(1)}(x) = \left(F^{(1)}_1(x), \dots, F^{(1)}_p(x)\right)^T,$$

$$\vec{F}^{(2)}(x) = \left(F^{(2)}_1(x), \dots, F^{(2)}_p(x)\right)^T$$

and covariance matrices

$$M(\vec{\xi}_i - \vec{\mu}^{(1)})(\vec{\xi}_i - \vec{\mu}^{(1)})^T = \sigma^{(1)}_{ii}, \quad M(\vec{\zeta}_j - \vec{\mu}^{(2)})(\vec{\zeta}_j - \vec{\mu}^{(2)})^T = \sigma^{(2)}_{jj}.$$

Sets of distributions for each component of the vectors $\vec{\xi}_{(k+\nu)}$ and $\vec{\zeta}_{(m+k-\nu)}$ we represent in the form of a vector for simplicity of the record. In the general case, $\sigma_{ij}^{(1)}$ and $\sigma_{ij}^{(2)}$ are square $p \times p$ blocks of the block-covariance matrices $K_{(k+\nu)}^{(1)}$ and $K_{(m+k-\nu)}^{(2)}$, and the matrices $\sigma_{ii}^{(1)}$ and $\sigma_{jj}^{(2)}$ are on the main diagonals of these matrices, correspondingly. The index (l) in the round brackets (for example, $(l) = (k+\nu)$) denotes that the vector sequence consists of l vectors and the block matrix - of $l \times l$ blocks.

The procedure of the stochastic joint of realizations of the sequences $\vec{\xi}_{(k+\nu)}$ and $\vec{\zeta}_{(m+k-\nu)}$ reduces to the following:

$(1a)$ realization of the integer random index ν from the interval $[1, m]$ with probabilities $P(\nu = i) = p_i$, $\sum\limits_{i=1}^{m} p_i = 1$ is sampled,

$(2a)$ from two independent sequences $\vec{\xi}_{(k+\nu)}$ and $\vec{\zeta}_{(m+k-\nu)}$, the sequence

$$\tilde{\vec{\xi}}_{(m+2k)} =$$

$$\left(\tilde{\vec{\xi}}_1, \ldots, \tilde{\vec{\xi}}_{m+2k}\right) =$$

$$\left(\vec{\xi}^T_{(k+\nu)}, \vec{\zeta}^T_{(m+k-\nu)}\right)^T =$$

$$= \left(\vec{\xi}_1, \ldots, \vec{\xi}_k, \vec{\xi}_{k+1}, \ldots, \vec{\xi}_{k+\nu}, \vec{\zeta}_{k+\nu+1}, \ldots, \vec{\zeta}_{k+m}, \vec{\zeta}_{k+m+1}, \ldots, \vec{\zeta}_{m+2k}\right)^T$$

is formed.

We choose the enumeration of indices of the sequences $\vec{\xi}_{(k+\nu)}$ and $\vec{\zeta}_{(m+k-\nu)}$ such that after their joint the enumeration of indices in this combined sequence $\tilde{\vec{\xi}}_{(m+2k)}$ would be through, with the joint being carried out within the interval of values of the indices $\nu = k, \ldots, k+m+1$. Consider the main properties of the constructed sequence. Let

$$\theta_\nu = \begin{cases} 0, & \nu = 1, \ldots, k, \\ \sum\limits_{i=1}^{\nu-k} p_i, & \nu = k+1, \ldots, k+m, \\ 1, & \nu = k+m+1, \ldots, m+2k. \end{cases},$$

In these notations one-dimensional distributions $\vec{F}_i(x)$ for elements of the sequence $\vec{\xi}_{(m+2k)}$ will have the form:

$$\vec{F}_i(x) = (1 - \theta_{i-1})\vec{F}^{(1)} + \theta_{i-1}\vec{F}^{(2)}(x), \quad i = 1, \ldots, m+2k. (1)$$

Note, that two-dimensional probability distributions $F_{ij,pq}(x,y)$ of p-th component of the vector $\tilde{\vec{\xi}}_l$ and q-th component of the vector $\tilde{\vec{\xi}}_j$ for $i,j = k+1,\dots,k+m$ are expressed by θ_i, θ_j and, also, by the given one-dimensional and two-dimensional distributions

$$F^{(1)}_{iq}(x), \quad F^{(2)}_{jp}(x), \quad F^{(1)}_{ij,pq}(x,y), \quad F^{(2)}_{ij,pq}(x,y)$$

of the components of the sequences $\vec{\xi}_{(k+\nu)}$ and $\vec{\xi}_{(k+m-\nu)}$.

In the general case, the elements σ_{ij} of the covariance matrix blocks

$$K_{(m+2k)} = M(\tilde{\vec{\xi}}_{(m+2k)} - \tilde{\vec{\mu}}_{(m+2k)})(\tilde{\vec{\xi}}_{(m+2k)} - \tilde{\mu}_{(m+2k)})^T = (\sigma_{ij}),$$

$$i,j = 1,\dots,m+2k,$$

where $\tilde{\vec{\mu}}_{(m+2k)} = M\tilde{\vec{\xi}}_{(m+2k)}$ satisfy the following relations:

$$\begin{array}{l} \sigma_{i,j} = \sigma_{i,i+h} = (1-\theta_{i+h-1})\sigma^{(1)}_{i,i+h} + \\ +\theta_{i-1}(1-\theta_{i+h-1})(\vec{\mu}^{(2)} - \vec{\mu}^{(1)})(\vec{\mu}^{(2)} - \vec{\mu}^{(1)})^T + \theta_{i-1}\sigma^{(2)}_{i,i+h}, \quad (2) \\ \sigma^T_{ji} = \sigma_{ij}, \quad h = 0,1,\dots,m, \quad i = 1,\dots,2k+m-h. \end{array}$$

The covariances within the interval $\nu = k-h,\dots,k+h$ will be called smoothing functions. The class of these smoothing functions is determined by the corresponding elements of block covariance matrices $K^{(1)}_{(k+\nu)}$ and $K^{(2)}_{(k+m-\nu)}$, and, also, by the probabilities p_i.

In particular, if $p_i = p = 1/(m+1)$, and the matrices $K^{(1)}_{(k+\nu)}$ and $K^{(2)}_{(k+m-\nu)}$ are block-stationary, i.e., $\sigma^{(i)}_{ij} = \sigma^{(1)}_{|i-j|}$ and $\sigma^{(2)}_{ij} = \sigma^{(2)}_{|i-j|}$, then elements of block matrices as functions of i are polynomials of the second degree with respect to i. If $\vec{\mu}^{(1)} = \vec{\mu}^{(2)}$, then expression (2) essentially simplifies, and within the interval $i = k+1,\dots,k+n$ it is a linear function with respect to i. In the block record this expression has the following form:

$$\sigma_{i,i+h} = (1-\theta_{i+h-1})\,\sigma^{(1)}_{i,i+h} + \theta_{i+1}\sigma^{(2)}_{i,i+h}, (3)$$

$$\theta_\nu = \begin{cases} \nu = 1,\dots,k, \\ \frac{\nu-k+1}{m+1}, \ \nu = k+1,\dots,k+m, \\ 1, \qquad\qquad \nu = k+m+1,\dots,m+2k, \end{cases},$$

$$h = 0,1,\dots,m, \quad i = k,\dots,m+k-h.$$

If the process is built by the joint of non-stationary sequences of finite length, then some essential simplifications are possible. Let $\vec{\xi}_{(k+\nu)}$ and $\vec{\xi}_{(m+k-\nu)}$ be two mutually independent sequences with the covariance matrices $K^{(1)}_{(k+\nu)}$ and $K^{(2)}_{(m+k-\nu)}$, $\nu = 0, \ldots, n$. Since ν is an integer random variable, then the maximal block dimension of the matrices $K^{(1)}_{(k+\nu)}$ and $K^{(2)}_{(k+m-\nu)}$ is equal to $k+m$, and the minimal dimension is equal to k. It is natural in some applications to take block elements in each of these matrices, corresponding to the points i and $i+h$ from the interval $k+1, \ldots, k+m$ as equal to each other, i.e., $\sigma^{(1)}_{i,i+h} = \sigma^{(2)}_{i,i+h} = \hat{\sigma}_{i,i+h}$. For example, if non-stationary $K^{(1)}_{(k+m)}$ and $K^{(2)}_{(k+m)}$ are estimated by the data of many - year observations, then for calculation of these matrix elements the variables, related to the same points of the interval $k+1, \ldots, k+m$, are used, and, hence, the corresponding elements of these matrices coincide. In the case $\vec{\mu}^{(1)} = \vec{\mu}^{(2)} = \vec{\mu}$, formula (2) may be written down in the following form

$$\sigma_{i,i+h} = (1 - \theta_{i+h-1} + \theta_{i-1})\, \hat{\sigma}_{i,i+h},$$

$$h = 0, 1, \ldots, m, \quad i = k, \ldots, m+k-h.$$

Thus, the error of smoothing is determined by values of the probabilities $\theta_{i+h-1} - \theta_{i-1}$.

If the Gaussian stationary sequences with zero expectation and with some different correlation functions, provided that $p_i = p$, $i = 1, \ldots, m$, are joint, then the smoothing functions are determined by the linear relation (3). For $p_{i-1} \neq p_i$, these smoothing functions are nonlinear. One-dimensional distributions in the smoothing interval remain standard normal, and two-dimensional distributions are a mixture of the corresponding normal distributions. Note, also, that each realization of the constructed sequence is the joint of independent realizations of limited length sequences, therefore abrupt changes of the simulated variables are possible at the boundary of the joint.

2.2. Probabilistic model of joint time series of air temperature, solar radiation and precipitation

For building non-stationary and non-Gaussian joint time series of meteorological quantities the method, presented in Section 2.1 in the combination with the method, based on normalization of observed time series, is used. The latter method is published in paper [11], and for the scalar case it consists in the following.

Let $\eta_1^*, \eta_2^*, \ldots, \eta_n^*$ be the observed data of a random variable η with the correlation function r_h^*, $h = 1, 2$.

(1*b*) Using $\eta_1^*, \eta_2^*, \ldots, \eta_n^*$, the continuous distribution function $F(x)$ and the correlation function r_n^* are estimated (the piecewise-polynomial approximation of a sampled distribution function, described in Section 6.6, is applied).

(2*b*) The initial series is transformed by solving the equation $\Phi(\xi_t^*) = F(\eta_t^*)$ into the sequence $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$ of standard Gaussian variables with the distribution $\Phi(x)$

$$\Phi(x) = P(\xi_t^* < x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{u^2/2} du. \tag{4}$$

Using this sequence, the correlation function g_n^* is calculated. Note that the joint distributions of this sequence are not Gaussian.

(3*b*) The Gaussian time series $\xi_1, \xi_2, \ldots, \xi_n$ with and the correlation function $g_n = g_n^*$ is simulated.

(4*b*) The concluding model time series $\eta_1, \eta_2, \ldots, \eta_n$ is obtained by solving the equation $F(\eta_t) = \Phi(\xi_t)$.

The one-dimensional distribution of this series coincides with $F(x)$, but its correlation function r_h (which is estimated using the model samples) does not coincide with r_h^* in the general case.

Note the following important property of this method: the time series η_t exists in the probabilistic sense for an arbitrary distribution $F(x)$. The method can be used also to construct non-stationary time series, it is especially simple in the cases where the time dependence appears in one-dimensional distributions but not in correlation functions.

In the general case it is difficult to obtain *a priori* estimate of the difference $r_h - r_h^*$, because it depends not only on $F(x)$, but also on the form

of two-dimensional distribution functions of the time series. The control of the difference $r_h - r_h$ may be empirically done by comparison of the model correlation function with the empirical one.

This method can be easily generalized to a series consisting of vectors of arbitrary dimension. Let $\vec{\eta}_1^*, \ldots, \vec{\eta}_n^*$ be a sequence of p-dimensional observed vectors, where $\vec{\eta}_i^* = \left(\vec{\eta}_{i,1}^*, \ldots, \vec{\eta}_{i,p}^*\right)^T$. Using the series $\eta_{1,k}^*, \ldots, \eta_{n,k}^*$, $k = 1, \ldots, p$, the one-dimensional continuous distribution functions $F^{(k)}(x)$ are estimated. By the transformation $\xi_{i,k}^* = \Phi^{-1}\left(F^{(k)}(\eta_{i,k}^*)\right)$ a vector series

$$\vec{\xi}_1^*, \vec{\xi}_2^*, \ldots, \vec{\xi}_n^*, \quad \vec{\xi}_i^* = \left(\xi_{i,1}^*, \ldots, \xi_{i,p}^*\right)^T$$

is obtained, and the block correlation matrix $G_{(n)}^*$ of this series is estimated. After that we simulate a vector Gaussian series $\vec{\xi}_1, \ldots, \vec{\xi}_n$ with the block-correlation matrix $G_{(n)} = G_{(n)}^*$, which is coordinate-wise transformed into the desired series $\vec{\eta}_1, \ldots, \vec{\eta}_n$, where

$$\vec{\eta}_i = (\eta_{i,1}, \ldots, \eta_{i,p})^T, \quad \eta_{i,k} = \left(F^{(k)}\right)^{-1}\left(\Phi(\xi_{i,k})\right),$$

$i = 1, \ldots, n$, $k = 1, \ldots, p$. If a series $\vec{\eta}_1^*, \ldots, \vec{\eta}_n^*$ is close to a stationary series, elements of the matrix $G_{(n)}^*$ are estimated by standard formulas for the estimation of elements of a correlation matrix of stationary processes, and the simulation of $\vec{\xi}_1, \ldots, \vec{\xi}_n$ is carried out according to scheme presented in [20].

In the cases, when one of components of the vector $\vec{\xi}_t$, forming a vector time series, is an indicator sequence (time series) χ_t, the method of normalization in the form presented in paper [11] is not applicable, because of specific features of the time series χ_t. As an example of such a vector time series let us consider a joint time series of daily amounts of precipitation, the daily mean air temperature and the daily amount of the solar radiation for the warm half of the year (period May-October).

Let $\vec{\eta}_1^*, \vec{\eta}_2^*, \ldots, \vec{\eta}_n^*$, as in the previous case, be the observed vector time series, in which components of the vector $\vec{\eta}_i^*$ are time series $\eta_{i,k}^*$ for $i = 1, \ldots, n$. We will interprete $\eta_{i,k}^*$ for $k = 1, \ldots, p-1$ as time series of different weather elements, for example, the daily amount of the solar radiation, the daily mean temperature, etc., and $\eta_{i,p}^*$ as a sequence of the daily rainy precipitation. Let, also, analytic expressions for one-dimensional distributions of components of the vector $\vec{\eta}_i^*$ for $k = 1, \ldots, p-1$ be known. They can be obtained with the help of the technique described in Section 1.5.

Let us denote one-dimensional distributions of the components $\eta^*_{i,k}$ of the vector $\vec{\eta}^*_i$ as $F_k(x)$, $i = 1, \ldots, n$, $k = 1, \ldots, p-1$. As for $\eta^*_{i,p}$, as was mentioned, these variables form a time series consisting of two sequences: the indicator sequence $\chi^*_{i,p}$ with two states 0 and 1 and, also, the sequence of daily amounts of precipitations $R^*_{i,p}$ with a one-dimensional distribution $F_p(x)$ provided that $\chi^*_{i,p} = 1$, and $R^*_{i,p} = 0$ if $\chi^*_{i,p} = 0$.

According to the technique, presented in the beginning of this Item, the first stage of the simulation is the normalization of observed time series and the estimation of a correlation structure of the normalized time series. For $\eta^*_{i,k}$, $i = 1, \ldots, n$, $k = 1, \ldots, p-1$ the normalization occur in accordance with [11] (see the beginning of this Item), and for the time series $\eta^*_{i,p}$ a special technique is suggested [130]. Let $\xi^*_{i,p}$ be normalized quantities. Consider the following transformation:

$$\begin{array}{l} \xi^*_{i,k} = \Phi^{-1}\left[F_k(\eta^*_{i,k})\right], \\ \xi^*_{i,p} = \Phi^{-1}\left[\chi_{i,p} p^*_1 F_p(R^*_{i,p}) + (1-\chi_{i,p})(p^*_0 x_i + p^*_1)\right], \\ i = 1, \ldots, n, \quad k = 1, \ldots, p-1. \\ (1) \end{array}$$

Here, as noted above, $p^*_1 = P(\chi^*_{i,p} = 1)$, $p^*_0 = 1 - p^*_1$, $\Phi(x)$ is a one-dimensional function of the standard normal distribution (4), x_i is a uniformly distributed variables in the interval $(0, 1)$. As a result of these transformations the vector time series $\vec{\xi}^*_1, \vec{\xi}^*_2, \ldots, \vec{\xi}^*_n$ with one-dimensional normal distribution of the corresponding components is obtained. Then the block correlation function $G^*_{(p)}(\tau)$ of the sequence $\vec{\xi}^*_1, \vec{\xi}^*_2, \ldots, \vec{\xi}^*_n$ is estimated with the help of standard formulas for a stationary vector process, which in combination with one-dimensional distribution will be the input information for building model sequences.

The second stage is the simulation of a stationary vector Gaussian random sequence $\vec{\xi}_1, \vec{\xi}_2, \ldots, \vec{\xi}_n$, $\vec{\xi}_i = (\xi_{i,1}, \ldots, \xi_{i,p})^T$ with the correlation function $G_{(p)}(\tau) = G^*_{(p)}(\tau)$ and its subsequent transformation to the sought for sequence $\vec{\eta}_1, \vec{\eta}_2, \ldots, \vec{\eta}_n$:

$$\begin{array}{l} \eta_{i,k} = F_k^{-1}\left[\Phi(\xi_{i,k})\right], \\ \eta_{i,p} = \left\{ \begin{array}{ll} F_p^{-1}\left[\frac{1}{p^*_1}\Phi(\xi_{i,k})\right], & \text{if} \quad \xi_{i,k} \le c, \\ 0, & \text{if} \quad \xi_{i,k} > c, \end{array} \right. , \quad (2) \\ i = 1, \ldots, n, \quad k = 1, \ldots, p-1. \end{array}$$

where c is determined by the solution of the trancendential equation $p_1 = 1/\sqrt{2\pi} \int\limits_{-\infty}^{c} e^{-u^2/2} du$.

Thus, the non-Gaussian connected vector sequence with the prescribed one-dimensional distributions is obtained. The correlation structure of the sequence $\vec{\eta}_1, \vec{\eta}_2, \ldots, \vec{\eta}_n$ is determined by transformations (1) and (2). In contrast to the models, based on the inverse distribution functions, according to which a certain special Gaussian time series is selected such that its transformation to a non-Gaussian one gives the required correlation structure (provided, of course, that such a Gaussian time series exists), the considered model reproduces the correlation structure only approximately. In some cases one may use a simple technique for increasing the simulation accuracy. Let us consider it on a simple example for a scalar time series χ_{nj}.

In Chapter 3 it was shown that the correlations $g^*_{ip,jp} = \text{corr}\,(\xi^*_{i,p}, \xi^*_{j,p})$ of normalized time series monotonically depend on the corresponding empirical correlations $r^*_{ip,jp} = \text{corr}\,(\xi^*_{i,p}, \xi^*_{j,p})$. The same character of dependence takes place for the correlations of model and empirical time series. Therefore the corresponding correlation functions of these three time series may be essentially different in their values for each i and j, but the character of each of these functions is alike. The accuracy of simulation may be increased, if one appropriately corrects the correlation function of the normalized time series (note, once again, that in the simulation procedure this correlation function is used for the Gaussian time series simulation). Consider a linear combination

$$\tilde{g}_{ip,jp} = \alpha\, r^*_{ip,jp} + (1 - \alpha)\, g^*_{ip,jp}, \quad (3)$$

where α is a real parameter. If $\alpha \in [0, 1]$, then $\tilde{g}_{ip,jp}$, $i, j = 1, 2, \ldots$, is a correlation function, but if $\alpha > 0$, then it is necessary to additionally test the function $\tilde{g}_{ip,jp}$ for its positive definiteness. The experience shows that by the selection of the corresponding α one may essentially increase the simulation accuracy.

Note, that it is reasonable to use the method based on the normalization of the observed time series in those cases, when the method of inverse distribution functions is impossible to apply, i.e., if the functions, obtained as a result of transformations (2.3) from Chapter 3 are not positive definite. Nevertheless, if eigenvalues of the corresponding matrices $G = (g_{ij})$ formed of m first values of these functions are negative, but not large in their absolute magnitude, then one may use another way for the corresponding

correction [106]. This method is in that negative eigenvalues of the matrix G are replaced for small positive ones, and instead of the matrix G the matrix

$$\hat{G} = T\hat{\Lambda}T^{-1} (4)$$

is used, where $\hat{\Lambda} = (\hat{\lambda}_{ii})$ is such that $\hat{\Lambda} = \Lambda + \Lambda'$, T is a matrix of the right normalized eigenvectors of the matrix G, $\Lambda = (\lambda_{ii})$ is the diagonal matrix of eigenvalues of the matrix G, and $\Lambda' = (\lambda'_{ii})$ is a specially chosen diagonal matrix such, that the elements $\hat{\lambda}_{ii}$ of the matrix $\hat{\Lambda}$ are positive ones. If the matrix G is the Toeplitz one (or block-Toeplitz), then transformation (4) disrupts this its property, therefore it is necessary to do some additional transformations, for example, averaging of elements of the matrix $\hat{G}$ along its diagonals, parallel to the main diagonal with a subsequent check for its positive definiteness. If necessary, this procedure may be repeated till the marix $\hat{G}$ be a positive definite one. For estimating the accuracy of the considered regularization, the norm $\max\limits_{i,j} |\hat{g}_{ij} - g_{ij}|$ may be used. Note, that in the Euclidean norm this regularization is close to the optimal one: in the book [22] it was shown that the representation of the correlation matrix G_m in the form of shortened series by its eigenvectors in the Euclidean norm is optimal.

The considered algorithms allow building probabilistic models of the joint weather element time series for different climatical zones with an acceptable degree of approximation. In a number of cases it appears possible to use the method of inverse distribution functions without additional correction of the corresponding matrices. For the station Voeikovo (St-Petersburg) on the basis of many-year observation data, the probabilistic model of the joint non-stationary time series of the daily amount of rainy precipitation, the daily mean air temperature and the daily amount of the solar radiation was built with the help of the inverse distribution function method and the method of a randomized joint of independent sequences of finite random length. It was assumed that inside each of six calendar months time series are stationary ones with specified correlation functions and distribution functions for each month. The observation data processing has shown that empirical correlations and one-dimensional distributions change from month to month. For example, one may judge about the character of the change of distributions by the change of the first four moments corresponding to them. For the daily amount of the solar radiation these moments are presented in Table 6.13. The continuous distribution functions, required for the realization of the inverse distribution function method were

obtained by the cubic spline approximation of discrete empirical distributions. Some special analytical functions were used for the approximation of tails of different weather elements. Thus, the daily mean temterature of the left and the right tails was approximated with the help of the normal distribution function. The same function was used for the approximation of the right tail of the distribution of the daily amount of the solar radiation. Approximation of empirical distributions of the daily amount of rainy precipitation was done in accordance with the technique, presented in Section 6.6. Comparison of distribution moments, calculated by the initial sample with those, calculated on the basis of approximating distribution functions for the daily amount of the solar radiation is presented in Table 15.

In accordance with the method of inverse distribution functions it is required to solve the equation of the form (2.2) from Chapter 3 with respect to unknown values of the Gaussian sequence correlations and to elucidate whether the discrete function g_n, formed of these values of correlation coefficients, would be a positive definite one or not. For the given sample these functions for each month turned out to be positive definite. Thus, obtaining the non-Gaussian series $\eta_{i,k}$ with the given one-dimensional distributions reduces to the simulation of the Gaussian sequence $\xi_{i,k}$ with zero expectation and the correlation function g_n. Then all elements of thise sequence elements are transformed by the formula $\eta_{i,k} = F_k^{-1}(\Phi(\xi_{i,k}))$. Here the index k denotes the number of the corresponding weather elements: $k = 1$ - temperature, $k = 2$ - radiation, $k = 3$ - indicator sequence, $k = 4$ - the daily amount of rainy precipitation. For the indicator sequence, auto-correlations and the corresponding cross-correlations with other weather elements, the relations (2.2) from Chapter 3 are simplified at the expense of specific features of the indicator sequence distribution.

Table 17: Comparison of empirical distribution moments (E) and model distribution moments (M) for daily amount of solar radiation

	May		June		July		Aug		Sent		Octob	
	E	M	E	M	E	M	E	M	E	M	E	M
μ	186	184	211	219	199	198	143	141	81.6	85.8	33.8	32.3
σ	67.4	67.2	81.2	71.6	63.5	64.1	58.9	58.1	45.0	43.0	23.1	22.9

The joint of independent sequences for building non-stationary time se-

ries was done at the stage of simulation of the Gaussian sequences on the basis of the scheme, described in Section 2.1. of the present Section for smoothing functions of the form of (3). The obtained joint sequences $\tilde{\xi}_t$ were transformed to non-Gaussian ones on the basis of solving the equation $F(\tilde{\eta}_t) = \phi(\tilde{\xi}_t)$ with respect to $\tilde{\eta}_t$. Note, that in this case the method of inverse distribution functions is realized only approximately, because the joint sequence $\tilde{\xi}_t$ is not Gaussian. As we see from (1), one-dimensional distributions of this sequence near each boundary of any two adjacent months are the mixture of two one-dimensional normal distributions, and two-dimensional distributions are a combination of one-dimensional and two-dimensional distributions of the initial series, etc. In this connection, linear smoothing functions (3) are transformed to non-linear, nevertheless this approach is acceptable for many applications, because it is often difficult to determine a true character of non-stationarity by the real observation data, since the number of observation years in many cases is restricted to several tens.

2.3.Numerical stochastic models of multi-dimensional non-stationary time series of weather elements

In this Section, the algorithm of modeling of two joint non-stationary meteorological daily time series for annual interval has been developed. This algorithm was realized for the time series of daily mean air temperature and indicators of precipitation totals. Numerical stochastic models of these variables are of great importance in assessment of wildfire potential change induced by climate warming.

The synchronous air temperature and precipitation time-series of daily scale in Sverdlovsk (Russia, Ural region) over period 1938-1984 were used as observational basis for modeling. Here n=365 - the number of days in each year.

2.3.1. Stochastic model of a joint time series of temperature and indicators of precipitation In the given work the algorithm for modeling of two joint non-stationary time series of meteorological elements on an annual interval on the basis of the real observation data is considered.

Let $\vec{T} = (T_1, \ldots, T_n)^T$ is daily mean air temperature for days from 1 to n, and $\vec{\Omega} = (\Omega_1, \ldots, \Omega_n)^T$is indicators of precipitation totals. The components

$\Omega_i = 1$, if the precipitation take place and $\Omega_i = 0$, if they are absent. Here $n = 365$ is the number of days in each year (for convenience the data for February 29 were not taken into account).

The elements of a joint time series of temperature and indicators of precipitation are calculated with the help of transformations

$$\begin{array}{ll} T_i = & \xi_{1i}\sigma_i + \mu_i; \\ \Omega_i = & \left\{ \begin{array}{ll} 1, & \xi_{2i} \le c_i \\ 0, & \xi_{2i} > c_i \end{array} \right. , \end{array} \quad (1)$$

where ξ_{1i}, ξ_{2i} are elements of joint Gaussian time series with zero average and block correlation matrix

$$G_{(2)} = \left[\begin{array}{cc} G_{\xi_1\xi_1} & G_{\xi_1\xi_2} \\ G_{\xi_2\xi_1} & G_{\xi_2\xi_2} \end{array} \right].$$

Here

$$\begin{array}{ll} G_{\xi_1\xi_1} = (g_{\xi_1\xi_1}(i,j)), & G_{\xi_1\xi_2} = (g_{\xi_1\xi_2}(i,j)), \\ G_{\xi_2\xi_1} = (g_{\xi_2\xi_1}(i,j)), & G_{\xi_2\xi_2} = (g_{\xi_2\xi_2}(i,j)), \quad i,j = 1,\ldots,n, \end{array}$$

μ_i and σ_i are the components of vectors$\vec{\mu} = (\mu_1,\ldots,\mu_n)^T$ and$\vec{\sigma} = (\sigma_1,\ldots,\sigma_n)^T$, and values $c_i$are determined from the equations

$$p_i = P(\Omega_i = 1) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{c_i} e^{-\frac{1}{2}t^2} dt \quad (2)$$

on the given probabilities p_i, $\vec{p} = (p_1,\ldots,p_n)^T$. The similar transformations for modeling of indicator processes are used in works [2]. In the given work these transformations are considered in the combination with the linear transformation Gaussian time series. The components of vectors $\vec{\mu}$, $\vec{\sigma}$ and $\vec{p}$ determine an annual overpatching of daily average temperature, standard deviation of temperature and annual variations of probabilities of precipitation fall-out.

The correlation matrix of dimension $730 \times 730$of the joint time series on the annual interval also has a block form

$$R_{(2)} = \left[\begin{array}{cc} R_{TT} & R_{T\Omega} \\ R_{\Omega T} & R_{\Omega\Omega} \end{array} \right],$$

where

$$\begin{array}{ll} R_{TT} = (r_{TT}(i,j)), & R_{T\Omega} = (r_{T\Omega}(i,j)), \\ R_{\Omega T} = (r_{\Omega T}(i,j)), & R_{\Omega\Omega} = (r_{\Omega\Omega}(i,j)), \quad i,j = 1,\ldots,n. \end{array}$$

The elements of a correlation matrix of the time series of temperature and precipitation indicators are determined by the ratio

$$r_{TT}(i,j) = g_{\xi_1\xi_1}(i,j), \quad r_{\Omega\Omega}(i,j) = \frac{P(\Omega_i = 1, \Omega_j = 1) - p_i p_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}} \quad , (3)$$

where

$$P(\Omega_i = 1, \Omega_j = 1) = P(\xi_{2i} \le c_i, \xi_{2j} \le c_j) = F(c_i, c_j, g_{\xi_2\xi_2}(i,j)).$$

Using expressions for the function of normal distribution

$$F(c_i, c_j, g_{\xi_2\xi_2}(i,j)) = \frac{1}{2}\Phi(c_i) + \frac{1}{2}\Phi(c_j) - T(c_i, a_1) - T(c_j, a_2) + \frac{1}{2}$$

at $c_i \ge 0,\ c_j \ge 0$ and $c_i < 0,\ c_j < 0$

$$F(c_i, c_j, g_{\xi_2\xi_2}(i,j)) = \frac{1}{2}\Phi(c_i) + \frac{1}{2}\Phi(c_j) - T(c_i, a_1) - T(c_j, a_2)$$

at $c_i < 0,\ c_j \ge 0$ and $c_i \ge 0,\ c_j < 0$,

where

$$\Phi(c_i) = \frac{1}{\sqrt{2\pi}} \int\limits_0^{c_i} e^{-\frac{1}{2}t^2} dt, \ \ T(c,a) = \frac{1}{2\pi} \int\limits_0^{a} e^{-\frac{1}{2}c^2(1+t^2)} \frac{dt}{1+t^2},$$

$$a_1 = \frac{c_j - g_{\xi_2\xi_2}(i,j)c_i}{c_i\sqrt{1 - g^2_{\xi_2\xi_2}(i,j)}}, \ \ a_2 = \frac{c_i - g_{\xi_2\xi_2}(i,j)c_j}{c_j\sqrt{1 - g^2_{\xi_2\xi_2}(i,j)}}.$$

we receive the formula connecting $r_{\Omega\Omega}(i,j)$ with $g_{\xi_2\xi_2}(i,j)$:

at $c_i \ge 0,\ c_j \ge 0 \quad (p_i \ge 0.5, p_j \ge 0.5)$ and $c_i \le 0,\ c_j \le 0 (p_i \le 0.5, p_j \le 0.5)$

$$r_{\Omega\Omega}(i,j) = \frac{0.5p_i + 0.5p_j - p_i p_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}} - \frac{T(c_i, a_1) + T(c_j, a_2)}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}}, (4)$$

at $c_i \ge 0,\ c_j \le 0 \quad (p_i \ge 0.5, p_j \le 0.5)$ and $c_i \le 0,\ c_j \ge 0 (p_i \le 0.5, p_j \ge 0.5)$

$$r_{\Omega\Omega}(i,j) = \frac{0.5p_i + 0.5p_j - 0.5 - p_i p_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}} - \frac{T(c_i, a_1) + T(c_j, a_2)}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}}. \quad (5)$$

Note, that the similar ratio are considered. In the given work these ratio are generalized on a case, when the probabilities p_i accept values more than

0.5 at various combinations p_i and p_j. This situation frequently arise at modeling of non-stationary precipitation time series.

Ratio connecting elements of the mutual correlation matrix $r_{T\Omega}(i,j)$ with the elements of matrix $g_{\xi_1\xi_2}(i,j)$ has the form:

$$r_{T\Omega}(i,j) = \frac{MT_i\Omega_j - MT_iM\Omega_j}{\sqrt{DT_i}\sqrt{D\Omega_j}} = -\frac{g_{\xi_1\xi_2}(i,j)}{\sqrt{2\pi p_j(1-p_j)}}\exp(-\frac{c_j}{2}) \quad (6)$$

Elements of a matrix $r_{\Omega T}(i,j)$ have the similar form.

The input parameters for model of joint time series $(\vec{T}^T, \vec{\Omega}^T)^T$ are the components of vectors $\vec{\mu}$, $\vec{\sigma}$ and $\vec{p}$, and elements of a correlation matrix $R_{(7)}$.

In order to receive estimations $\vec{\mu}^*$, $\vec{\sigma}^*$, $\vec{p}^*$, $R^*_{(2)}$ of these parameters joint time series of daily average air temperature and precipitation indicators are constructed from the real data, and standard formulas for sample statistic are used.

It is necessary to note that the real data are available over period of $N = 47$ on years, therefore rank of a matrix from sampling coefficients of correlation at its dimension 730×730does not exceed $N = 47$, and the statistical error at an estimation of parameters is great enough (approximately it is equal to 13 %). In this situation it is expedient to use procedure of moving averaging of the received characteristics on time.

In the Fig. 2, À) the diagrams of an annual variation of empiric daily mean air temperature averaged by 47-years period (curve 1) and appropriate smoothed curve with a window of bilateral smoothing per 21 day (curve 2) are given.

The curves 5,6 correspond to the similar diagrams for a standard deviation of daily average air temperature, and the curves 3,4 represent functions $\vec{\mu}^*_{sm} \pm \vec{\sigma}_\mu$, where $\vec{\mu}^*_{sm}$ is smoothed estimation daily average air temperature, and $\vec{\sigma}_\mu$ is standard deviation of an estimation $\vec{\mu}^*$, calculated on the modeling data, imitated 47 years temperature time series. In the Fig. 2, Â) the diagrams of an annual chanding averaged by 47 years real probabilities (curve 1) and appropriate smoothed curve (curve 2) are given.

In a Fig. 3. the diagrams of the annual change of correlation coefficient between values of temperature of air with shift per one day (curve 1) and similar correlation coefficient between values of precipitation indicators (curve 3) are given. The estimations were carried out on samples of 47 elements.

The curves 2,4 correspond to the smoothed curve of an annual change of these parameters. The curve 5 corresponds to the smoothed curve of

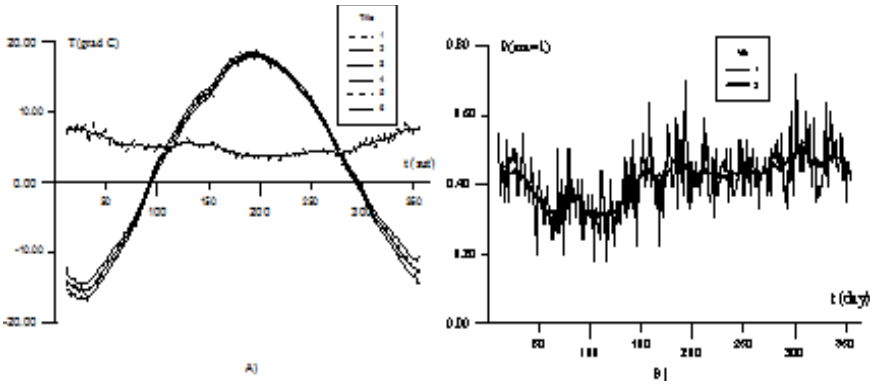

Figure 2: Dependence on time of the statistical characteristics of time series of daily average temperature and precipitation indicators.

an annual change of correlation coefficient with shift per one day between temperature and precipitation indicators.

Table 18: Parameters of the one-dimensional precipitation amount distributions for four seasons.

	Winter	Spring	Summer	Autumn
μ_f	0.602	1.085	2.283	1.086
μ_m	0.580	1.070	2.270	1.068
σ_f^2	0.811	3.414	14.451	3.683
σ_m^2	0.772	3.232	13.710	3.393
A_f	3.222	4.006	4.854	5.409
A_m	3.896	4.395	4.883	6.081
E_f	13.845	21.311	36.084	52.085
E_m	16.839	24.236	34.012	59.864

From the given figures it is visible, that considered joint time series it is essential non-stationary both on the one-dimensional probability distributions, and on correlations.

For construction of a joint non-stationary time series $(\vec{T}^T, \vec{\Omega}^T)^T$ with the given correlation matrix and the given parameters of one-dimensional

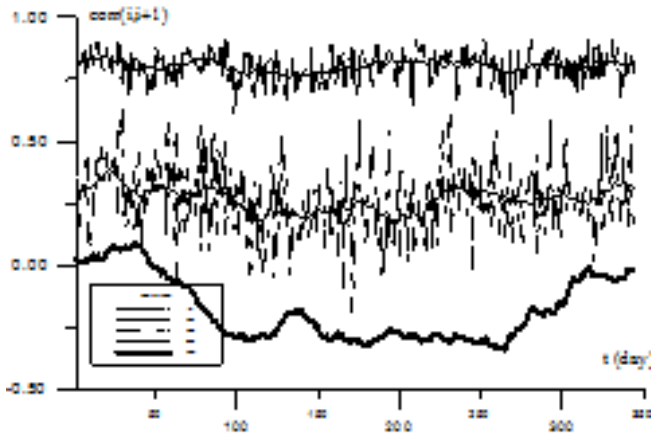

Figure 3: Dependence on time of correlation coefficient of time series of daily average temperature and precipitation indicators.

Table 19: Correlation coefficients r_1 (8) calculated by the real data.

	Winter	Spring	Summer	Autumn
r_1	0.280	0.206	1.165	0.172
ρ_1	0.369	0.340	0.250	0.311

Table 20: Mean values of seasonal precipitation calculated by the real and model data.

	Winter	Spring	Summer	Autumn
R	17.6651	33.1205	92.4875	46.5223
M	18.0264	33.9163	91.4049	45.0622

distributions it is necessary to solve n the nonlinear equations (2), and also $n(2n+1)/2$ appropriate equations from (3) - (5) concerning elements of a matrix $G_{(2)}$ at the given values of elements of a matrix $R_{(2)}$.

As the result we receive a matrix of correlation coefficients, which generally may be out of the class of correlation matrixes, i.e. some its eigenvalues may be negative. In this case the task is solved approximately (e.g. special correction procedure, based on spectral decomposition of the original matrix with replacement of negative eigenvalues for small positive values can be used [7]).

After obtaining of the correlation matrix $G_{(2)}$ for Gaussian time series the generation of joint time series $(\vec{T}^T, \vec{\Omega}^T)^T$with the help of transformation (6) is carried out. The joint Gaussian time series $\vec{\xi} = (\vec{\xi}_1^T, \vec{\xi}_2^T)^T$ with a correlation matrix $G_{(2)}$ using transformation

$$\vec{\xi} = P\Lambda^{1/2}\vec{\varphi}$$

are simulated. Here Λ is diagonal matrix of eigenvalues of matrix $G_{(2)}$, $P = (\vec{P}_1, \ldots, \vec{P}_N)$ is matrix of the appropriate eigenvectors, $\vec{\varphi}$ - vector of independent standard normal values.

The practical calculations have shown that some eigenvalues of the matrix $G_{(2)}$ have appeared to be negative. After updating the matrix $\tilde{G}_{(2)}$ was received, and for this matrix the algorithm (1), (6) is realized. The Figure 3 demonstrates comparative estimates of the correlation coefficients of precipitation indicators using observational data and model series (10000 realizations). Because of large dimension of the correlation matrix, it is possible to present only one fragment of the row, which has maximal (by module) difference between elements of the model and empiric correlation matrixes.

2.3.2. Numerical models of time series of daily precipitation

In this section we consider a more general model of a joint temperature and precipitation series, which takes into account not only the presence of absence of precipitation, but also its amount for the daily interval. Particularly, numerical stochastic models of these variables are of great importance in assessment of wildfire potential change induced by climate warming. Thus, we consider the joint temperature and precipitation series that simulate real observations of these weather elements. According to real data, the precipitation series are the alternation of precipitation periods in which the corresponding precipitation amount is registered for each day with periods without precipitation. Therefore, the precipitation series model is regarded

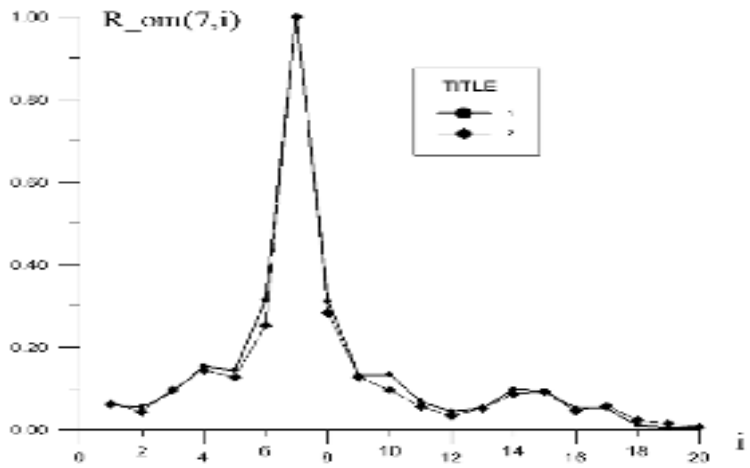

Figure 4: Fragment of lines of real - 1 and modelling – 2 correlation matrixes for precipitation indicators.

as the process of the form

$$\gamma_t = \Omega_t \theta_t(7)$$

Where Ω_t is the time series of precipitation indicators considered in the previous section, which specifies the periods of presence or absence of precipitation, and θ_t is some auxiliary time series imitating the daily precipitation totals for the given interval. These series, as before, are considered for an annual interval. Therefore, the daily precipitation amount in different seasons has a seasonal meaning (snow or rain precipitation). Thus, we first construct the nonstationary auxiliary process of precipitation θ_t which takes into account the one-dimensional distribution of precipitation and season-dependent correlations. Then, using relation (7), we construct the nonstationary precipitation process.

a) One-dimensional distributions of precipitation

According to the observed data precipitation intensity is season-dependent. In the given model, the annual interval is divided into four intervals corresponding to definite seasons. One-dimensional distributions are estimated separately for each season. The process θ_t inside the season is supposed to be a stationary one with respect to one-dimensional distributions and correlations. Figure 5 presents, as an example, the empirical

one-dimensional distribution $F(x_i) = P(\theta^* \leq x_i)$ of precipitation for summer, where θ^* are the real values of precipitation if it takes place. When constructing this distribution, x_i is chosen to more fully show the change in the curve. To construct the θ_t- series model by the inverse distribution function method we must have the continuous distribution function $F(x)$. To construct this function we used the linear interpolation of the empirical function $F(x_i)$. The plot of this function is also presented in Fig. 5.

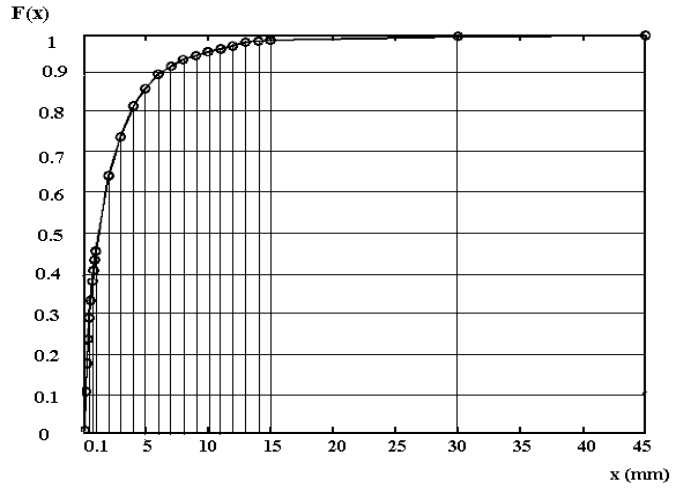

Figure 5: Empirical one-dimensional distribution of precipitation in summer and its linear interpolation.

Calculations showed that the empirical distribution function is substantially season-dependent. This is readily seen from Table 1 that gives the distribution moments (average, variance, asymmetry, and excess) for each season separately, which are calculated by real and model data. The interpolation quality of empirical onedimensional distribution can be judged from the model parameters in Table 18.

b) Correlation functions

As mentioned above, the auxiliary process θ_t inside each season is regarded as a stationary one. Therefore, the correlation functions corresponding to these seasons depend on the time lag only. Since the real precipitation series have a specific character, viz. the precipitation periods alternate with those without precipitation, the standard methods for estimating the corre-

lation function in the stationary series are not applicable to the given case. We estimate the corresponding correlation coefficients in the framework of the precipitation series by the formula

$$r_h \approx \tilde{r}_h = \frac{\tilde{M}\,\theta_i\,\theta_{i+h} - \tilde{M}\,\theta_i\,\tilde{M}\,\theta_{i+h}}{\sqrt{\tilde{D}\,\theta_i\,\tilde{D}\,\theta_{i+h}}} \quad (8)$$

Here

$$\tilde{M}\,\theta_i\,\theta_{i+h} = \frac{1}{N}\sum_{k=1}^{N_h}\sum_{i=1}^{N_{k-h}}\theta_i\,\theta_{i+h}$$

$$N = \sum_{k=1}^{N_h} N_k$$

where N_k is the number of h-shift pairs inside the k-th series whose length is larger than h, N_h is the number of series whose length is larger than h

$$\tilde{M}\,\theta_i = \frac{1}{N}\sum_{k=1}^{N_h}\sum_{i=1}^{N_{k-h}}\theta_i\,, \quad \tilde{D}\,\theta_i = \frac{1}{N}\sum_{k=1}^{N_h}\sum_{i=1}^{N_{k-h}}\theta_i^2 - (\tilde{M}\,\theta_i)^2$$

$$\tilde{M}\,\theta_{i+h} = \frac{1}{N}\sum_{k=1}^{N_h}\sum_{i=h}^{N_k}\theta_i\,, \quad \tilde{D}\,\theta_{i+h} = \frac{1}{N}\sum_{k=1}^{N_h}\sum_{i=h}^{N_k}\theta_i^2 - (\tilde{M}\,\theta_{i+h})^2$$

It is these correlations that are used to model the auxiliary series θ_t. We should emphasize that the Toeplitz matrix composed of these correlation coefficients does not always turn out to be positive definite. However, in our case for $h = 0, \ldots, 10$, this matrix turned out to be a correlation one. Since the values of the correlation function for $h > 1$ are small, hereafter when constructing the model we considered only the correlation coefficient r_1 unique to each season. These correlation coefficients are given in the first row of Table 19.

c) Modelling the series θ_t

The series θ_t is constructed by the inverse distribution function method using the Gaussian autoregression model. Let χ_t be a Gaussian series with correlation function ρ_h. We consider the transformation

$$\theta_t = F^{-1}(L(\chi_t)). \quad (9)$$

where $F(x)$ is the given one-dimensional distribution and $L(x)$ is the standard normal distribution function. The correlations r_h between the

elements of series (9) are connected with the corresponding correlations ρ_h of the series χ_t by the known relation $r_h = f(\rho_h)$, its concrete form is given, for example, in [Ermak].

To model the series θ_t we must construct the Gaussian series χ_t with correlation function ρ_h the elements of which are calculated by solving the equations $r_h = f(\rho_h)$ for ρ_h, given the correlations r_h. As a result, we obtain a sequence of correlation coefficients for the process χ_t which, in the general case, may not be a correlation function. However, since for each season we take into account only one correlation coefficient in modelling (the corresponding correlation coefficients ρ_1 are given in the second row of Table 19.), our matrix is a correlation one.

In order to take into account the season dependence of the model parameters, we construct the series χ_t by the autoregression model with variable coefficients

$$\chi_t = \rho_t \chi_{t-1} + \sqrt{1 - \rho_1^2 \varphi_t} (10) where($$

φ_t are independent standard normal values. The coefficient ρ_1 in formula (10) changes from season to season and the series χ_t is constructed continuously within the annual interval. Further, we construct the series θ_t by the transformation (9). In this transformation, we use the distribution function $F(x)$ corresponding to the current season. Finally, we construct the daily precipitation amount series γ_t by the constructed series θ_t and relation (8).

To verify the model accuracy we consider seasonal precipitation totals. These characteristics are not input data for the given model and can be used as test data to verify its accuracy. Table 20 gives the mean values of the seasonal precipitation amount calculated by the real and model data (1000 annual realizations were used for the corresponding estimates by the model data).

As seen from Table 20, from the viewpoint of the seasonal precipitation amount, the constructed model simulates the real data rather exactly. The values of the precipitation amount in the model given in Table 3 are affected by the length of series consisting of units in the indicator sequence, as well as the distribution and correlations of the series γ_t. However, the connection between the series θ_t and the series T_t was not taken into account in the given model. A further improvement of the model can be based on considering these connections, as well as the unsteady state of the series θ_t in more detail.

2.4. Special stochastic models of joint meteorological time series

Numerical stochastic simulation of meteorological elements complexes was incorporated into practice of hydrometeorology several decades ago, but became widely-used with the beginning of global impact studies in climate change investigations. Within the given problem, the approach under study is used when interpreting results of modeling of the global climate and development of the regional scenarios of climate changes, for estimation of productivity of agricultural cultures, in the models of water balance, etc. The obtained results, however, are sensitive to the quality of reproduction of a probabilistic structure of simulated time series. Significant simplifications used when constructing stochastic models, can result in serious restrictions in their application. For obtaining more reliable results and expansion of opportunities of using numerical stochastic simulation in climatic applications, an increase in accuracy of the model description of real variability of meteorological time series is necessary.

In the given Section the development of a stochastic model based on real data for joint daily meteorological time series is considered. We consider such meteorological elements as surface air temperature, components of wind speed, total cloudiness.

Approach is based on representation of one-dimensional distributions as a mixture of two normal distributions. The corresponding algorithms are generalized to a multi-dimensional case and are used for modeling of the joint time series of the surface air temperature and the wind speed vector.

2.4.1. Stochastic models of joint time series with one-dimensional distributions as a mixture of two normal distributions

Let us present n time series with one-dimensional distributions

$$P_{\xi_{it}}(x) = \frac{\theta_{it}}{\sqrt{2\pi}\sigma_{it}^{(1)}} e^{-\frac{(x-\mu_{it}^{(1)})^2}{2(\sigma_i^{(1)})^2}} + \frac{1-\theta_i}{\sqrt{2\pi}\sigma_{it}^{(2)}} e^{-\frac{(x-\mu_{it}^{(2)})^2}{2(\sigma_{it}^{(2)})^2}}, \qquad 0 \le \theta_{it} \le 1 (1)$$

in the form of the following transformations:

$$\xi_{1t} = \omega_{1t}(\mu_1^{(1)} + \sigma_1^{(1)}\eta_{1t}) + (1-\omega_{1t})(\mu_1^{(2)} + \sigma_1^{(2)}\eta_{1t}),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots (2)$$

$$\xi_{nt} = \omega_{nt}(\mu_n^{(1)} + \sigma_n^{(1)}\eta_{nt}) + (1-\omega_{nt})(\mu_n^{(2)} + \sigma_n^{(2)}\eta_{nt}),$$

$$t = 1, 2, \dots \ .$$

Here $\mu_i^{(1)}$, $\sigma_i^{(1)}$, $\mu_i^{(2)}$, $\sigma_i^{(2)}$, $(i = 1, \dots, n)$ are parameters of mixtures, η_{it} are Gaussian time series with joint matrix correlation function $G_{t,t+\tau} = (g_{i,j;t,t+\tau})$, $i, j = 1, \dots n$, $t, \tau = 0, 1, \dots$ and ω_{it} are binary time series, which, in turn, are constructed with the help of the transformation

$$\omega_{it} = \begin{cases} 1, & \psi_{it} \le c_{it} \\ 0, & \psi_{it} > c_{it} \end{cases} , (3)$$

where ψ_{it} are independent of η_{jt}, $(i, j = 1, \dots, n)$ Gaussian time series with a joint matrix correlation function $Q_{t,t+\tau} = (q_{i,j;t,t+\tau})$. The quantities c_{it} are connected to the probabilities $P(\omega_{it} = 1) = \theta_{it}$ by the equation

$$P(\omega_{it} = 1) = \theta_{it} = \Phi(c_{it}) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{c_{it}} e^{-\frac{1}{2}x^2} dx \, . (4)$$

The matrix correlation function

$$R_{t,t+\tau} = (r_{i,j;t,t+\tau}), \ i, j = 1, \dots n, \ t, \tau = 0, 1, \dots \ (5)$$

of the time series ξ_{it} is determined by the functions $G_{t,t+\tau}$, $Q_{t,t+\tau}$ and transformations (2), (3). The relations connecting the correlation functions $G_{t,t+\tau}$ and $Q_{t,t+\tau}$ with the correlation function $R_{t,t+\tau}$ for the case of one time series are described in Section 1.3. They are easily generalized to the case of modeling n joint time series. When modeling random processes with a given correlation structure we consider the correlation function $R_{t,t+\tau}$ to be given, and the functions $G_{t,t+\tau}$, $Q_{t,t+\tau}$ are found from appropriate nonlinear equations connecting these three functions. In general case, the solution is not unique. For simplification it is possible to set $G_{t,t+\tau} = Q_{t,t+\tau} = K_{t,t+\tau} = (k_{i,j;t,t+\tau})$, $i, j = 1, \dots, n$, $\quad t, \tau = 0, 1, \dots \ .$ In this case, the equations connecting entries of the matrices $R_{t,t+\tau}$ and $K_{t,t+\tau}$ take the form

$$C_{1;ij} = C_{1;ij}(\theta_{it}, \theta_{j,t+\tau}, \mu_{it}^{(1)}, \mu_{it}^{(2)}, \mu_{j,t+\tau}^{(1)}, \mu_{j,t+\tau}^{(2)}, \sigma_{it}^{(1)}, \sigma_{it}^{(2)}, \sigma_{j,t+\tau}^{(1)}, \sigma_{j,t+\tau}^{(2)}) =$$

$$= \frac{(\theta_{it}\sigma_{it}^{(1)} + (1-\theta_{it})\sigma_{it}^{(2)})(\theta_{j,t+\tau}\sigma_{j,t+\tau}^{(1)} + (1-\theta_{j,t+\tau})\sigma_{j,t+\tau}^{(2)})}{2\sqrt{\theta_i(1-\theta_i)\theta_{j,t+\tau}(1-\theta_{j,t+\tau})}(\sigma_{it}^{(2)} - \sigma_{it}^{(1)})(\sigma_{j,t+\tau}^{(2)} - \sigma_{j,t+\tau}^{(1)})} +$$

$$+ \frac{(\mu_{it}^{(2)} - \mu_{it}^{(1)})(\mu_{j,t+\tau}^{(2)} - \mu_{j,t+\tau}^{(1)})}{2(\sigma_{it}^{(2)} - \sigma_{it}^{(1)})(\sigma_{j,t+\tau}^{(2)} - \sigma_{j,t+\tau}^{(1)})},$$

$$C_{2;ij} = C_{2;ij}(\theta_{it}, \theta_{j,t+\tau}, \sigma_{it}^{(1)}, \sigma_{it}^{(2)}, \sigma_{j,t+\tau}^{(1)}, \sigma_{j,t+\tau}^{(2)}) = (6)$$

$$= \frac{\sigma_{it}\sigma_{j,t+\tau}}{2\sqrt{\theta_{it}(1-\theta_{it})\theta_{j,t+\tau}(1-\theta_{j,t+\tau})}(\sigma_{it}^{(2)} - \sigma_{it}^{(1)})(\sigma_{j,t+\tau}^{(2)} - \sigma_{j,t+\tau}^{(1)})},$$

$$k_{i,j;t,t+\tau} = C_{1;ij} + \sqrt{C_{1;ij}^2 + 2C_{2;ij}r_{i,j;t,t+\tau}},$$

$$k, l = 1, \ldots, n,$$

$$-\frac{C_{1;ij}^2}{2C_{2;ij}} \le r_{i,j;t,t+\tau} \le 1.$$

In some cases the matrix function $K_{t,t+\tau}$ can is not a correlation function and it is necessary to correct it in an appropriate way [**?**].

2.4.2. Models of joint meteorological time series with given discrete distributions and distributions as mixtures of two normal distributions

Let us consider the matrix correlation functions of the form (5) for the case of two joint time series ξ_{kt} and ξ_{lt}, $(t = 1, 2, \ldots)$ of various meteorological parameters, where elements of this function are auto and mutually correlation functions of two various time series: the surface air temperature, wind speed components, precipitation indicators, total cloudiness, etc. For example, in problems associated with estimation of probabilities of forest fire occurrence one can use time series of the air temperature and precipitation. Below we consider models of various pairs of meteorological time series with one-dimensional distributions of two types: in the form of a mixture of two normal distributions and distributions of a discrete type. Let us consider a discrete random time series ξ_{qt}, with integer values $k = 1, \ldots, N$ (for total cloudiness $N = 10$), which is constructed as threshold transformations of Gaussian time series η_{qt}

$$\xi_{qt} = \begin{cases} 0, & \eta_{qt} \le c_{qt}^{(0)} \\ 1, & c_{qt}^{(0)} < \eta_{qt} \le c_{qt}^{(1)} \\ & \cdots \\ N-1, & c_{qt}^{(N-2)} < \eta_{qt} \le c_{qt}^{(N-1)} \\ N, & c_{qt}^{(N-1)} < \eta_{qt} \end{cases}, \quad (7)$$
$$t = 1, 2, \ldots ,$$

where $c_{qt}^{(0)}, \ldots c_{qt}^{(N-1)}$ are some given numbers (levels), dividing an interval $(-\infty, \infty)$ into intervals $A_0 = (-\infty, c_{qt}^{(0)})$, $A_1 = (c_{qt}^{(0)}, c_{qt}^{(1)}), \ldots, A_N =$

$(c_{qt}^{(N-1)}, \infty)$. Probabilities $P(\xi_{qt} = k) = P(\eta_{qt} \in A_k)$ are determined by the relations

$$P(\xi_{qt} = k) = p_{qt}^{(k)} = \frac{1}{\sqrt{2\pi}} \int\limits_{A_k} e^{-\frac{1}{2}x^2} dx\,, \;\; k = 0, \ldots, N. \quad (8)$$

Let us introduce the designations: $H(\mu, \nu; g) = T(\mu;\ a_1(\mu, \nu; g)) + T(\nu;\ a_2(\mu, \nu; g))$, $g = g_{qq;t,t+\tau}$, where

$$T(c, a) = \frac{1}{2\pi} \int\limits_0^a e^{-\frac{1}{2}c^2(1+u^2)} \frac{du}{1+u^2} \quad (9)$$

is the Owen function, in which parameters a_1 and a_2 have the form

$$a_1(\mu, \nu; g) = \frac{\nu - g\mu}{\mu\sqrt{1-g^2}}, \qquad a_2(\mu, \nu; g) = \frac{\mu - g\nu}{\nu\sqrt{1-g^2}}.$$

In particular, for $p_{qt}^{(N-1)} \le 0.5,\ \ p_{q,t+\tau}^{(N-1)} \le 0.5$ or $p_{qt}^{(N-1)} \ge 0.5,\ \ p_{q,t+\tau}^{(N-1)} \ge 0.5$ with allowance for (**??**) the relations, connecting elements of the correlation function $r_{qq;t,t+\tau}$ of time series ξ_{qt} with the function $g_{qq;t,t+\tau}$ takes the form

$$r_{qq;t,t+\tau} =$$

$$= \frac{\sum\limits_{k,l=1}^{N-1} lk \left\{ H(c_{qt}^{(l)}, c_{q,t+\tau}^{(k-1)}) + H(c_{qt}^{(l-1)}, c_{q,t+\tau}^{(k)}) - H(c_{qt}^{(l)}, c_{q,t+\tau}^{(k)}) - H(c_{qt}^{(l-1)}, c_{q,t+\tau}^{(k-1)}) \right\}}{S_{qt} S_{qt+\tau}} +$$

$$+ \frac{N \sum\limits_{l=1}^{N-1} l \left[0.5 p_{qt}^{(l)} - 0.5 p_{qt}^{(l-1)} + H(c_{qt}^{(l)}, c_{q,t+\tau}^{(N-1)}) - H(c_{qt}^{(l-1)}, c_{q,t+\tau}^{(N-1)}) \right]}{S_{qt} S_{qt+\tau}} + \quad (10)$$

$$+ \frac{N \sum\limits_{k=1}^{N-1} k \left[0.5 p_{q,t+\tau}^{(k)} - 0.5 p_{q,t+\tau}^{(k-1)} + H(c_{qt}^{(N-1)}, c_{q,t+\tau}^{(k)}) - H(c_{qt}^{(N-1)}, c_{q,t+\tau}^{(k-1)}) \right]}{S_{qt} S_{qt+\tau}} +$$

$$+ \frac{N^2 \left[1 - 0.5 p_{qt}^{(N-1)} - 0.5 p_{q,t+\tau}^{(N-1)} - H(c_{qt}^{(N-1)}, c_{q,t+\tau}^{(N-1)}) \right] - \left(\sum\limits_{l=1}^{N} l p_{qt}^{(l)} \right) \left(\sum\limits_{l=1}^{N} l p_{q,t+\tau}^{(l)} \right)}{S_{qt} S_{qt+\tau}},$$

$$S_{qt} = \sqrt{\sum\nolimits_{l=1}^{N} l^2 p_{qt}^{(l)} - \left(\sum\nolimits_{l=1}^{N} l p_{qt}^{(l)} \right)^2},$$

$$S_{q,t+\tau} = \sqrt{\sum_{l=1}^{N} l^2 p_{q,t+\tau}^{(l)} - \left(\sum_{l=1}^{N} l p_{q,t+\tau}^{(l)}\right)^2}.$$

The relations connecting elements of the mutually correlation function $r_{iq;t,t+\tau}$ of the time series ξ_{it} and ξ_{qt} with the function $g_{iq;t,t+\tau}$ have the form

$$r_{iq;t,t+\tau} = -\frac{[\theta_{it}\sigma_{it}^{(1)} + (1-\theta_{it})\sigma_{it}^{(2)}]g_{iq;t,t+\tau}}{\sqrt{2\pi}S_{it}S_{q,t+\tau}} \sum_{l=1}^{N-1} \exp\left(-\frac{\left(c_{qt}^{(l)}\right)^2}{2}\right),$$

$$S_{it} = \sqrt{\theta_{it}[(\sigma_{it}^{(1)})^2 + (\mu_{it}^{(1)} - \mu_{it})^2] + (1-\theta_{it})[(\sigma_{kt}^{(2)})^2 + (\mu_{it}^{(2)} - \mu_{it})^2]}, \quad (11)$$

$$S_{q,t+\tau} = \sqrt{\sum_{l=1}^{N} l^2 p_{q,t+\tau}^{(l)} - \left(\sum_{l=1}^{N} l p_{q,t+\tau}^{(l)}\right)^2}, \; k = 1, \ldots, n.$$

2.4.3. Numerical experiments

From the standpoint of climatic applications the use of numerical models of joint meteorological time series to estimate probabilities of rare and dangerous combinations of values for the variables in question is of a great interest. For example, currently when studying of fire risk conditions much attention is given to long periods with a high air temperature and deficiency of precipitation.

Table 21:

$T \leq -10$ T	P_{real}	P_{model}
$\rho \geq 3\,m/s$	0,2344	0,2693
$\rho \geq 4\,m/s$	0,1494	0,1676
$\rho \geq 5\,m/s$	0,0836	0,1012
$\rho \geq 6\,m/s$	0,0494	0,0611
$\rho \geq 7\,m/s$	0,0224	0,0359
$\rho \geq 8\,m/s$	0,0132	0,0287
$\rho \geq 9\,m/s$	0,0053	0,0132

About quality of the model reproduction of the joint distribution of air temperature and wind speed one can judge from Table 21., where the

comparative estimations of probabilities for various combinations of the parameters calculated using the model and real samples are given. In the calculations 10000 simulated vectors and 1519 real observations of temperature and wind speed were used.

2.5. Numerical stochastic model of spatial fields of daily sums of rainy precipitation

In this section a numerical stochastic model of spatial fields of daily sums of liquid precipitation constructed with the use of real observation data is proposed. Based on this model, simulation algorithms for conditionally distributed fields are constructed and certain examples of the use of this model in studying extreme properties of liquid precipitation fields are considered.

In order to construct precipitation fields on a spatial grid, we use the method of inverse distribution functions [15]. To do this, one has to be able to simulate Gaussian fields on grids with an arbitrary correlation structure, and the size of such fields may be sufficiently large. In this paper the simulation of a Gaussian field is performed by the method of conditional distributions [20]. In this case the use of block Toeplitz matrices for simulation of uniform fields allows us to increase essentially the admissible sizes of these fields [13, 20]. In this paper we use the data of fifteen years of observations of the daily sums of liquid precipitations registered at 47 stations in the Western part of the Novosibirsk region.

The specifics of spatial fields of daily precipitation sums $\eta(x, y),\ x, y \in S$, is that domains S_0 without precipitation alternate with domains S_R with precipitation, $S_0 \cup S_R = S$. Thus, a random precipitation field takes zero values in some random domains S_0 of a given domain S, and takes nonnegative values in domains S_R.

According to observation data, the minimal amount of daily precipitation sums registered by precipitation gauges is equal to 0.1 mm; a precipitation amount less than 0.1 mm, but greater than zero is registered as a 'precipitations trace', and zeros correspond to the absence of precipitation. In this paper we consider a trace precipitation as zero precipitation. We also assume that the field of daily precipitation sums is uniform and hence its one-dimensional distribution does not depend on spatial coordinates and the correlation function does not depend on a parallel translation of the point coordinates of the field where the correlations are specified. Such approximation is justified for the considered domain by the fact that geo-

graphic conditions pertaining to this domain are sufficiently uniform and many field characteristics rather weakly depend on the coordinates of the observation posts [3]. In order to construct a random field of daily liquid precipitation sums, we use an approximate modification of the method of inverse distribution functions where the absence of precipitation is identified with zero amount. By $P(\eta(x,y)=0)=q$ we denote the probability that the amount of precipitation equals zero at the point $x,y \in S$ and by $P(\eta(x,y)=R(x,y))=p=1-q$ we denote the probability that the amount of precipitation equals $R(x,y) \quad > 0.1$ mm at that point. For the sake of convenience of using the method of inverse distribution functions, the distribution function for the daily precipitation sums $F(x)$ is specified as a certain function of a continuous argument in the interval $[0,\infty)$ and in the interval $[0,0.1)$ the function $F(x)$ satisfies the condition $F(0.1)-F(0)=q$. All preliminary work in implementation of the inverse function method is performed with this distribution function.

In this paper we use the algorithms of approximation of the empirical distribution function considered in Section 1.5. and the correlation function on the stations by special continuous functions. Another topic relates to the algorithms of simulation of Gaussian vectors and fields on a regular grid and their use in construction of fields of daily liquid precipitation sums with a given one-dimensional distribution function and a correlation function. The algorithms of simulation of conditionally distributed Gaussian vectors and fields are considered on regular and irregular grids for given values at fixed points. Using these models, we construct an approximate model of a conditionally distributed non-Gaussian field of daily liquid precipitation sums. Application examples of this model are presented for estimation of some characteristics of extreme precipitations.

2.5.1 One-dimensional distribution of uniform spatial fields of daily liquid precipitation sums

In this paper the field $\eta(x,y)$ is constructed on a regular grid with $m \times n$ nodes. Hereafter we denote it by $\{\eta_{ik}\}, \quad i=1,\ldots,n, \quad k=1,\ldots,m$. In the construction of the numerical stochastic model of this field by the method of inverse distribution functions

$$\eta_{ik} = F^{-1}(\xi_{ik}) \tag{6}$$

the one-dimensional distribution function $F(x)$ of the elements of the non-Gaussian field η_{ik} and its correlation matrix $R=(r_{ik,hl}), \quad h,i = 1,\ldots,m, \quad l,k=1,\ldots,n$, are assumed to be given (under the condition of their mutual compatibility [21]), where ξ_{ik} is an appropriate Gaussian

field $\{\xi_{ik}\}$ with the correlation matrix $G = \{g_{ik,hl}\}$. The elements of this matrix are determined from the given elements of the matrix R with the use of the well-known relation [21]:

$$r_{ik,hl} = f(g_{ik,hl})$$

The monotonicity of this function and its other properties were indicated in [10].

The empirical function of one-dimensional distribution of the non-Gaussian precipitation field $F^*(x_i) = P^*(\eta \leq x_i),\ i = \overline{1,k}$ is calculated from a data array where, as was indicated above, the absence of precipitation is identified with a zero precipitation amount. In order to approximate the empirical distribution function $F^*(x)$ by a continuous function $F(x)$, we use a special modification of the method considered in Section 1.5. for the case when the absence of precipitation is considered as zero precipitation sum. The empirical function of the one-dimensional distribution of daily precipitation probabilities is determined at the points

$$0 = x_0 \leq x_1 < \ldots < x_{m+1} < \ldots < x_{n+1} < \infty$$

(in this case $x_1 = 0.1$ mm) and takes the values with corresponding to formula (1.5.1) The approximating function $F(x)$ is constructed so that for any values of the constants $a_i,\ b_i$ it coincides with the empirical distribution function at the nodal points. Certain restrictions are imposed on those constants in order to ensure the nonnegativity of the density $f(x) = F'(x)$ for all $x$and the minimal curvature of the function $F(x)$ in the interval $x_1 < x < x_{m+1}$. For the function $U(x)$ one can take any monotone increasing function taking the values $F_0^* = 0$ and F_1^* at the points x_0 and x_1, respectively. In the simplest case this can be the linear function

$$U(x) = cx, \quad 0 \leq x < x_1$$

where the parameter $c$is chosen from the condition

$$\int_0^{0.1} cx dx = q$$

where q is the probability $P(\eta_{ik} < 0.1)$, i.e., the probability of the absence of precipitation.

The quality of approximation essentially depends on the choice of the function $V(x)$ and the nodal point x_{m+1}. The function $V(x)$ is generally used in the form

$$V(x) = (1 - e^{-ax^b})^\alpha \tag{7}$$

where $a > 0,\ b > 0,\ \alpha > 0$. For fixed x_{m+1} and under the condition that F^*_{m+1} and $V(x_{m+1})$ coincide at this point, the value a is connected with b and α by the relation

$$a = -\ln(1 - F^{*\frac{1}{\alpha}}_{m+1})/x^b_{m+1}$$

and the values b and α are obtained from the condition

$$b, \alpha = \arg\min_{b,\alpha} \sum_{x_i \geq x_{m+1}} (F^*_i - V(x_i))^2.$$

In this paper we take

$$x_0 = 0,\ x_1 = 0.1,\ x_2 = 0.3,\ x_3 = 0.5,$$

$$x_4 = 1, x_5 = x_{m+1} = 2,\ (m = 4)$$

and cubic polynomials were constructed for each interval between these nodes. For the nodes $x_{m+1}, ..., x_n$ used in the approximation of the right 'tail' of the empirical distribution function $F_{*(x)}$ by the function $V(x)$ we took the two points $x_{m+1} = 2, x_{m+2} = 3$ $(m + 2 = n)$, where $m + 2 = n$. The accuracy of approximation was checked by comparing the mean value, variance, asymmetry, and the excess of the actual and approximating distribution functions.

2.5.2. Correlation structure of spatial fields of liquid precipitation daily sums

Since the aim of the paper was the construction of a field on a regular grid and we had data describing precipitation amount only at the stations, for the construction of the model we need an approximation of the empirical correlation function by a certain special function describing the field structure at arbitrary points of the considered domain. To do that, we used the function

$$\begin{aligned} r(x_s, y_s; x_h, y_h) = r(x_s - x_h, y_s - y_h) = r(x, y) = \\ = \exp(-[ax^2 + bxy + cy^2]^\theta) \end{aligned} \tag{8}$$

This function is correlative, which follows from the transformation of a random uniform and isotropic field with the correlation function $r(\rho) =$

$\exp(-\alpha\rho^{\beta})$, $\rho = \sqrt{x^2+y^2}$, $\alpha, \beta > 0$, in the form of the corresponding contraction, expansion, and rotation of the coordinates. Functions of such type are often used in description of the correlation structure of spatial fields of liquid precipitation [6]. The choice of the parameters a, b, c, and θ is based on the minimality condition for the mean square difference between the actual and approximating functions. The values of these parameters are presented in the first row of Table 22.

Table 22: Parameters of the correlation function of a Gaussian field.

a	b	c	θ
0.0043	−0.0016	0.0057	0.3125
0.00081	−0.00031	0.00111	0.3225

In the construction of spatial fields of daily sums of liquid precipitations with the distribution function $F(x)$ based of the method of inverse distribution functions, we have to calculate the corresponding correlation function $g(x, y)$ of the Gaussian field. In this case we assume that the correlation function of the Gaussian field as well has form (8), and the problem is reduced to the calculation of the parameters of this function. The values of the parameters of function (8) for a Gaussian field are presented in the second row of Table 22.

2.5.3. Numerical simulation of spatial fields of daily sums of liquid precipitations

In this paper we have used the method of conditional distributions [20] to simulate Gaussian fields in a given rectangular domain with the grid $\{x_{ik}\}$ of $m \times n$ nodes, which was necessary for the implementation of the method of conditional distributions [20]. The field $\{\xi_{ik}\}$ can be represented as a stationary sequence of Gaussian vectors $\xi_{(n)} = (\xi_1^T, \ldots, \xi_n^T)^T$, where $\xi_k = (\xi_{k1}, \ldots, \xi_{km})^T$ are the vectors of dimension $m$and the block Toeplitz matrix

$$G_{(n)} = \{G_{|i-l|}\} = \{G_h\}, \;\; i, l = 1, \ldots, n$$

with the blocks

$$G_h = \{g_{ik,jk+h}\}, \qquad i, j = 1, \ldots, n, \qquad h = 0, \ldots, n-1.$$

The elements $g_{ik,jk+h}$ of this matrix $G_{(n)}$ are determined by function (8). The method of conditional distributions for the vector sequence $\xi_{(n)}$ is re-

duced to realization of the transformations

$$\begin{aligned}
&\vec{\xi}_1 = C_0\vec{\varphi}_1,\\
&\vec{\xi}_2 = \vec{B}^T[1]J_{(1)}\vec{\xi}_{(1)} + C_1\vec{\varphi}_2,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&\vec{\xi}_n = \vec{B}^T[n-1]J_{(n-1)}\vec{\xi}_{(n-1)} + C_{n-1}\vec{\varphi}_n,
\end{aligned}$$

Where $\vec{\varphi}_k,\ k = 1, \ldots, n$, are independent Gaussian vectors of the dimension m:

$$J_{(k)}G^{-1}_{(k)}J_{(k)}\vec{G}_k = \vec{B}[k], \vec{G}_k = (G_1^T, \ldots, G_k^T)^T, \quad C_kC_k^T = Q_k,\ Q_k = R_0 - \vec{B}^T[k]\tilde{R}_{(k)}\vec{B}[k],$$

$$J_{(k)} = \left\| \begin{array}{ccc} 0 & \cdots & I \\ \cdots & \cdots & \cdots \\ I & \cdots & 0 \end{array} \right\| -$$

is a block permutation matrix, where 0 and I are the zero and identity matrices of the dimension $m \times m$, respectively.

Using the function $g(x, y)$ considered in the previous section, we constructed the block Toeplitz matrix $G_{(n)}$ for the regular grid of 30×25 nodes and simulated the Gaussian field $\{\xi_{ik}\}$ on this grid. Using transformation (6), we constructed the non-Gaussian field $\{\eta_{ik}\}$ with one-dimensional distribution function (7). The field of daily sums of liquid precipitations was obtained after the values below 0.1 mm were replaced by zeros in the field $\{\eta_{ik}\}$.

Figure 6 presents the graphs of correlation function (8) for the spatial field of sums for fixed values of the arguments $y = 0$ and $x = 0$, respectively (curves 1, 3). Curves 2, 4 are obtained by the model. The estimate was based on 10000 implementations with additional averaging over the spatial variables.

2.5.4. Simulation of conditionally distributed Gaussian fields

Consider some problems related to the use of the method of conditional distributions and the corresponding algorithms modelling random sequences for solution of applied problems in hydrometeorology. One of the problems where the application of this method is natural is the stochastic interpolation of daily precipitation fields from an irregular grid of the meteorological stations network to the nodes of a regular grid. Algorithms simulating realizations of conditionally distributed fields with given values at fixed points are well known for Gaussian fields [6, 20, 21]. These algorithms are based on different variants of simulation of Gaussian vectors with a given vector

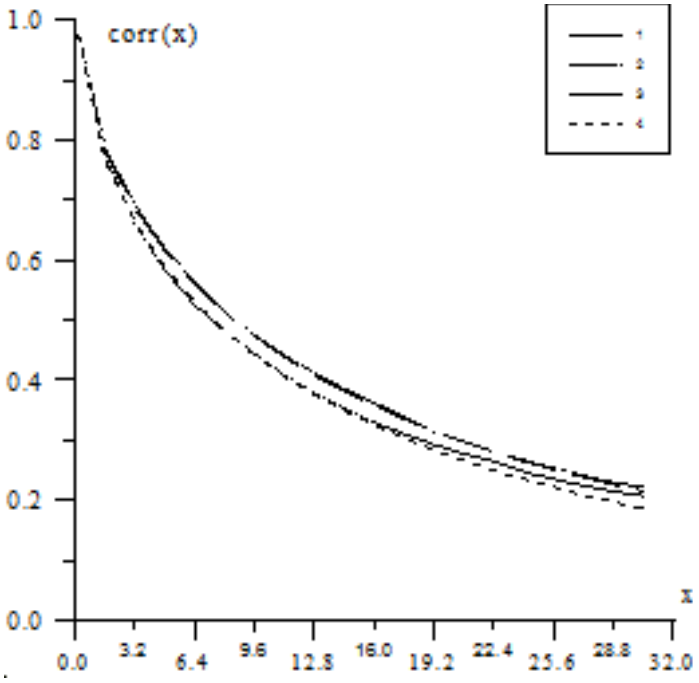

Figure 6: Comparison of given and model correlation functions of the fields of daily liquid precipitations

of the mean values and a given covariation matrix. In the modelling of conditionally distributed Gaussian vectors, the given characteristics are the vector of conditional means dependent on fixed values at given points and the conditional covariation matrix not dependent on these values.

In the simulation of a field at the nodes of a regular grid it is convenient in some cases to use the well-known property of the normal distribution consisting in the fact that if $\vec{\zeta} = (\vec{\zeta}_1^T, \vec{\zeta}_2^T)^T$ is a normal vector with zero mean and the joint covariation matrix

$$R = M(\vec{\zeta} - \vec{\mu})(\vec{\zeta} - \vec{\mu})^T = \left\| \begin{array}{cc} R_{11} & R_{12} \\ R_{21} & R_{22} \end{array} \right\|, \tag{9}$$

then the vector with zero mean

$$\vec{\eta}_1 = \vec{\zeta}_1 - R_{12} R_{22}^{-1} \vec{\zeta}_2, \tag{10}$$

has the covariation matrix

$$R_{11.2} = R_{11} - R_{12} R_{22}^{-1} R_{21} \tag{11}$$

Now let $\vec{\xi} = (\vec{\xi}_1^T, \vec{\xi}_2^T)^T$ be a normal vector with the mean $\vec{\mu} = (\vec{\mu}_1^T, \vec{\mu}_2^T)^T$and the covariation matrix R. Adding the mean value

$$\mu_{1.2} = \vec{\mu}_1 + R_{12} R_{22}^{-1} (\vec{x}_2 - \vec{\mu}_2) \tag{12}$$

to $\vec{\eta}_1$, where vector $\vec{\xi}_2$ is fixed and equal to $\vec{x}_2$, we get the normal vector

$$\vec{\xi}_1 = \mu_{1.2} + \vec{\eta}_1 \tag{13}$$

with mean (12) and covariation matrix (11). If the dimensions of the matrices in (9) are not too large, it is easier to use the triangular decomposition of matrix (11) directly and the standard method of simulation of a Gaussian vector with the given vector of means and the covariation matrix. If the dimension of the matrix R is large, the implementation of both these approaches is sufficiently laborious, however, an approximate modelling algorithm for conditionally distributed Gaussian fields can be constructed rather easily in the realization of the first approach for uniform fields on a regular grid [17]; such algorithm is based on the method of optimal interpolation [6].

The algorithm is realized in four stages.

(1a) Modelling a uniform field $\{\hat{\zeta}_{ik}\}$ on a regular grid $\{\hat{x}_{ik}\}$ with a sufficiently small step in the given domain S, in this case the grid $\{x_{ik}\}$ is embedded into the grid $\{\hat{x}_{ik}\}$.

(2a) A linear interpolation of the values of the Gaussian field $\{\hat{\zeta}_{ik}\}$ from the closest grid nodes $\{\hat{x}_{ik}\}$ to the stations. In the simplest case this is a translation of the values of the field from the closest node to a station. As the result, we have the values of the field $\{\zeta_{ik}\}$ on the grid $\{x_{ik}\}$ and the values of the field on the stations, or the vector $\vec{\zeta} = (\vec{\zeta}_1^T, \vec{\zeta}_2^T)^T$ with covariation matrix (9). It should be noted that the approximate spectral model of a Gaussian field [21] does not require simulation of the field on an auxiliary grid for the construction of this field at grid nodes and at the stations.

(3a) Implementation of transformation (10) for the vector $\vec{\zeta} = (\vec{\zeta}_1^T, \vec{\zeta}_2^T)^T$. If the dimensions of the matrices R_{12} and R_{22} are not very large and the matrix R_{22} can be inverted easily, then transformation (10) can be realized exactly. If this condition does not hold, at this stage we may use the optimal procedure of linear interpolation of the field from the stations nearest to the grid nodes.

(4a) Construction of the field $\{\xi_{ik}\}$, or the vector $\vec{\xi}_1$ with the use of transformation (7). If the matrices R_{12} and R_{22} do not satisfy the conditions from the previous item, for the calculation of $\mu_{1.2}$ we use the optimal linear interpolation of the fixed values $\vec{x}_2$ from the stations nearest to the grid nodes.

2.5.5. Approximate algorithm of numerical simulation of conditionally distributed fields of liquid precipitation daily sums.

For the sake of convenience, we consider the field of precipitation sums as a vector $\vec{\eta} = (\vec{\eta}_1^T, \vec{\eta}_2^T)^T$ whose components all have the same distribution function $F(x)$, where $\vec{\eta}_1$ is the field at the grid nodes and $\vec{\eta}_2$ is the field at the stations. The problem consists in construction of the field $\vec{\eta}_1$ on a grid under the condition that the field values are fixed at the stations, i.e., $\vec{\eta}_2 = \vec{\eta}_2^*$. Actual observation data from the stations are taken as $\vec{\eta}_2^*$. As was indicated above, there are zeros in a particular realization of a real field on the stations. Represent the vector $\vec{\eta}_2^*$ in the form $\vec{\eta}_2^* = \left(\vec{\eta}_{21}^{*T}, \vec{\eta}_{22}^{*T}\right)^T$, where $\vec{\eta}_{21}^* = \vec{0}$ and $\vec{\eta}_{22}^* \neq \vec{0}$. The first modelling stage consists in the 'normalization' of nonzero components of this vector, i.e., in the following transformation of this vector:

$$\vec{\zeta}_{22}^* = \Phi^{-1}\left(F(\vec{\eta}_{22}^*)\right) \quad . \tag{14}$$

Since the components of the vector $\vec{\eta}_{21}^*$ are equal to zero, these zero values

should be associated with some values of the corresponding components of the vector $\vec{\zeta}_{21}$ from the interval $\left(-\infty, \Phi^{-1}\left(F(0.1)\right)\right)$. Here the vectors $\vec{\zeta}_{21}$ and $\vec{\zeta}_{22}$ are subvectors of the vector $\vec{\zeta}_2 = \left(\vec{\zeta}_{21}^T, \vec{\zeta}_{22}^T\right)^T$, and the dimension of this vector and the dimensions of its subvectors coincide with the corresponding dimensions of the vector $\vec{\eta}_2^*$. Consider one approximate method of modelling the vector $\vec{\zeta}_{21}$. It is based on the approximation consisting in the fact that the vector $\vec{\zeta}_2 = \left(\vec{\zeta}_{21}^T, \vec{\zeta}_{22}^T\right)^T$, where the subvector $\vec{\zeta}_{22}$ is normalized according to (14), is a normal vector with zero mean and the covariation matrix

$$R_{22} = \left\| \begin{array}{cc} R_{22}^{11} & R_{22}^{12} \\ R_{22}^{21} & R_{22}^{22} \end{array} \right\|$$

determined by correlation function (8). The algorithm is reduced to the following transformations:

(1b) In accordance with procedure (10), (12), (13) and a fixed value of the vector $\vec{\zeta}_{22}^* = \Phi^{-1}\left(F(\vec{\eta}_2^*)\right)$, we construct the conditionally distributed normal vector $\vec{\zeta}_{21}$. The covariation matrix of this vector equals $R_{22}^{11} - R_{22}^{12}\left(R_{22}^{22}\right)^{-1} R_{22}^{21}$ and the mean is $R_{22}^{12}\left(R_{22}^{22}\right)^{-1}\vec{\eta}_{22}^*$.

(2b) The vector $\vec{\hat{\zeta}}_{21}$ is chosen, all components of this vector lie within the interval $\left(-\infty, \Phi^{-1}\left(F(0.1)\right)\right)$.

(3b) The vector $\vec{\hat{\zeta}}_2 = \left(\vec{\hat{\zeta}}_{21}^T, \vec{\zeta}_{22}^{*T}\right)^T$ is formed from the vectors $\vec{\hat{\zeta}}_{21}$ and $\vec{\zeta}_{22}^*$. In this case the vector $\vec{\zeta}_{22}^*$ is fixed and the vector $\vec{\hat{\zeta}}_{21}$ is random. The next stage in the simulation of a conditionally distributed Gaussian field consists in the implementation of algorithm (1a)–(4a) where for the fixed vector $\vec{x}_2$ we take the vector $\vec{\zeta}_2^* = M\vec{\hat{\zeta}}_2$. As the result, we have the conditionally distributed Gaussian vector $\vec{\xi}_1$ (or the field $\{\xi_{ik}\}$ at the grid nodes $\{x_{ik}\}$) with the fixed vector $\vec{\zeta}_2^*$. The mathematical expectation and covariation matrix of this vector are equal to $\mu_{1.2} = R_{12}R_{22}^{-1}\vec{\zeta}_2^*$ and $R_{11.2} = R_{11} - R_{12}R_{22}^{-1}R_{21}$respectively, where the elements of the matrices R_{11},R_{12} and R_{22} are calculated with the use of the uniform field correlation function of form (8) from the corresponding coordinates of the grid nodes and the stations. It is worth noting that, in contrast to an unconditional field, a conditional field is not uniform.

The one-dimensional conditional distribution function for the values of the field ξ_{ik} has the following form for each grid node x_{ik} under fixed values

of the field at the stations:

$$\Phi(x_{ik} \left| \vec{\zeta}_2 = \vec{\zeta}_2^* \right.) = \frac{1}{\sqrt{2\pi}\sigma_{ik.2}} \int_{-\infty}^{x_{ik}} \exp\left(-\frac{1}{2} \left(\frac{u - \mu_{ik.2}}{\sigma_{ik.2}} \right)^2 \right) du \qquad (15)$$

where the conditional mean $\mu_{ik.2}$ at the node x_{ik} depends on the values at the stations and the correlations determining the conditional mean and variance depend on the distances from the grid node to the stations.

Further, the realization of the conditionally distributed field is constructed with the use of the transformation

$$\eta_{ik} = F^{-1}\left(\Phi(\xi_{ik})\right) \qquad (16)$$

After that, the values η_{ik} less than 0.1 mm are replaced by zeros. It is worth noting that we do not know the joint distribution of the precipitation field.

Therefore, based on transformation (16), we obtain only a certain approximate model of conditionally distributed fields of daily sums of liquid precipitation.

2.5.6. Numerical experiments

Using the numerical models of daily sums of liquid precipitations and the method of direct modelling, we calculate certain characteristics of the spatial field of daily sums of liquid precipitations. Some examples related to the use of conditional field model in the study of extreme precipitation regimes have been considered. Figure 7 shows the probabilities that the total amount of precipitations is not less than a given level at a fixed point of the considered domain. Two cases were considered. In the first case, many stations are positioned around a point (curve 1), in the second case, there are no stations in a sufficiently large neighborhood of the chosen point (curve 2). Curves 3, 4 correspond to the unconditional model of daily fields of liquid precipitation for the same points of the domain. Since the information concerning precipitation is given only on the stations and it is known that considerable precipitations are registered there, the probability of substantial precipitations at points positioned closely to stations is greater than at distant ones where the influence of the stations is small. The graphs show that the considered probability at points positioned far from stations based on the conditional mode is close to the same probability calculated by the unconditional model.

Using the models constructed here and applying direct simulation of daily spatial fields of liquid precipitation, we constructed histograms of probability distribution for the total precipitation amount within a day in the

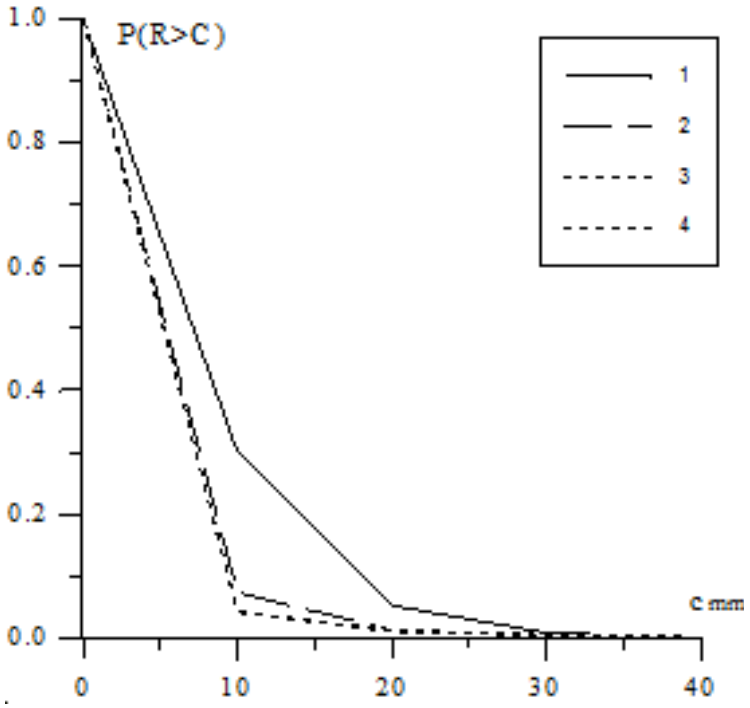

Figure 7: Probabilities that the sums of precipitations are not less than a given level (mm) for two points of the domain according to the conditional and unconditional models.

whole considered area. The corresponding graphs are presented in Fig. 8 where the solid line corresponds to the conditional model and the dotted line corresponds to the unconditional one. The total precipitation values are marked on the abscissa axis in km^3. The same fixed values of precipitation sums at the stations have been used in all experiments.

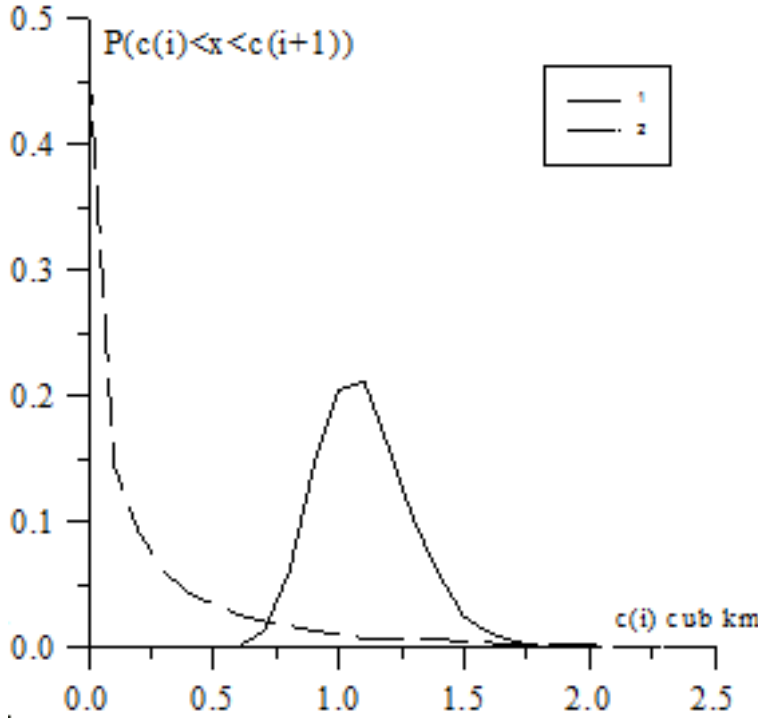

Figure 8: Histograms of probability distribution for the total precipitation amount in a day in the whole considered area according to conditional and unconditional models.

In conclusion we note that the algorithms considered here can be used in simulation of various fields of meteorological elements under given values at the observation posts. The distribution functions of some elements, for example, air temperature, are close to Gaussian ones, therefore, the method of inverse distribution functions is more accurate for this case. In some cases, for example, in the calculation of total moisture resources in a given area for

a long period of time by spatial-time models, the problem becomes rather laborious and the implementation of the computations requires state-of-the art multiprocessor computers.

BIBLIOGRAPHY

1 Anderson, Introduction to Multivariate Statistical Analysis / / M: Gos. Izdat. Physical and Mathematical Literature, 1963. -500 p.

2 A. Arakawa. Computation design for long-term numerical integration of the equation of fluid motion of two-dimensional incompressible flow. Part 1. J. Comput. Phys. (1966) 1, No. 1, pp. 119-143

3 A.D. Drobyshev, A.S. Marchenko, V.A. Ogorodnikov and V.D. Chizikov. Statistical structure of daily amounts of rainy precipitation time series. Proceedings ZapSibNIGMI Goskomgidrometa (1989) 86, pp. 44-74 [in Russian]

4 E.S. Epstein. Stochastic dynamic prediction. Tellus (1969) 21, No. 6, pp. 739-759

5 A. L Evstafieva, E. 1. Khlebnikova, and V. A. Ogorodnikov, Numerical stochastic models for complexes of time series of weather elements. Russ. J. Numer. Anal. Math. Modelling (2005) 20, No. 6, 535-548.

6 L.S. Gandin and R.L. Kagan. Statistical Methods of Interpretation of Meteorological Data. Gidrometeoizdat, Leningrad, 1967 [in Russian]

7 N. I. Gubina and V. A. Ogorodnikov, Stochastic filling of homogeneous isotropic random fields. Russ. J. Numer. Anal. Math. Modelling (2006) 21, No. 6, 514–524.

8 A. S. Marchenko, Approximation of empirical probability distribution for daily sums of liquid precipitation. Proc. West Sib. Research Inst. Goskomgidromet (1989) 86, 66– 74 (in Russian).

9 A.S. Marchenko and L.A. Minakova. A probabilistic model of air temperature time series. J. Meteorology and Hydrology (1980), No. 9, pp. 39-48 [in Russian]

10 A.S. Marchenko and V.A. Ogorodnikov. Probabilistic models of dry and rainy days time series. Preprint No. 993, Comp. Cent. Sib. Division, USSR Acad. Sci., Novosibirsk, 1991 [in Russian]

11 A.S. Marchenko and A.G. Seomochkin. Piece-polynomial approximation of groupped empirical distribution function. In: Metody Statist. Modelirov. Comp. Cent. Sib. Division, USSR Acad. Sci., Novosibirsk, 1986, pp. 16-24 [in Russian]

12 G.I. Marchuk. Numerical Methods in Weather Forecasting. Gidrometeoizdat, Leningrad, 1967 [in Russian]

13 S. L. Marple, Digital Spectral Analysis with Applications. Prentice-Hall, 1987.

14 G.A. Mikhailov. Optimization of Weighted Monte Carlo Methods. Nauka, Moscow, 1987 [in Russian]

15 G.A. Mikhailov and A.V. Voitishek, Numerical Statistical Modelling. Monte Carlo Methods. Akademia, Moscow, 2006 (in Russian).

16 A.M. Obukhov. On statistically orthogonal decompositions of empirical functions. Izv. Akad. Nauk SSSR, Ser. geofiz., (1960), No. 3. pp. 432-439 [in Russian]

17 V. A. Ogorodnikov, Simulation of three-dimensional geopotential fields with given statistical structure. In Monte Carlo Methods in Numerical Mathematics and Mathematical Physics, Novosibirsk, 1979, pp. 73–78 (in Russian).

18 V.A. Ogorodnikov. On dynamic-stochastic forecasting. Izv. Akad. Nauk SSSR, FAO (1975) 11, No. 8, pp. 851-853 [in Russian]

19 V.A. Ogorodnikov and L.A. Minakova. Probabilistic models of nonstationary vectorial sequences. In: Teoria i pril. Stat. Modelirov., Comp. Cent. Sib. Division, USSR Acad. Sci., Novosibirsk, 1989, pp. 17-28 [in Russian]

20 V. A. Ogorodnikov and S.M. Prigarin, Numerical Modelling of Random Processes and Fields: Algorithms and Applications. VSP, Utrecht, The Netherlands, 1996.

21 S. M. Prigarin, Conditional spectral models of Gaussian homogeneous fields. Russ. J. Numer. Anal. Math. Modelling (1998) 13, No. 1, 514–524.10. S. M. Prigarin, Numerical Modelling of Random Processes and Fields, Inst. Comp. Math. Math. Geoph., Novosibirsk, 2005 (in Russian).

22 C.R. Rao. Linear statistical inference and its applications. John Wiley & Sons. Inc., New York [Russian transl., Nauka, Moscow, 1968]

23 T.P. Romanenko. On time statistical structure of air temperature in the troposphere and the lower stratosphere. In: Primenenie statistricheskik metodov v meteor., Comp. Cent. Sib. Division, USSR Acad. Sci., Novosibirsk, 1971, pp. 122-138 [in Russian]

24 N.V. Smirnov, L.N. Bolshev. Tables for calculating the normal distribution functions. Izv. Akad. Nauk SSSR, 1962 [in Russian]

25 B.V. Tatarsky. The use of dynamic equations for a probabilistic forecasting of the pressure fields. Izv. Akad. Nauk SSR, FAO (1969) 5, No. 3, pp. 293-297 [in Russian]

26 V.I. Tikhonov. Excursions of Random Processes. Nauka, Moscow, 1970 [in Russian]

CONTENTS

INTRODUCTION 3
1. HYDROMETEORLOGICAL APPLICATIONS OF STATISTICAL SIMULATION METHODS 4
1.1. In influence of uncertainty in initial data on forecasting accuracy .. 4
1.2. On accuracy of temperature vertical profiles expansion into a series by eigenvectors of sampling covariance matrix 8
1.3. Investigation of some features of excursions of air temperature time series 14
1.4. Probabilistic model of air temperature time series. 19
1.4.1.Results of simulation. 21
1.5. Probabilistic models of dry and rainy days time series 22
1.5.1. Three-connected Markov model 24
1.5.2. Multiplicative model 31
1.5.3. Threshold transformation of the Gaussian sequence 33
1.6. Probability properties of precipitation amount 37
1.7. Approximation of empirical probability distribution of daily rainy precipitation sums 40
2. PROBABILISTIC MODELS OF JOINT TIME SERIES AND SPATIAL FIELDS OF DIFFERENT WEATHER PARAMETERS 48
2.1. Stochastic joint of independent random vector sequences 49
2.2. Probabilistic model of joint time series of air temperature, solar radiation and precipitation 53
2.3.Numerical stochastic models of multidimensional non-stationary time series of weather elements 53
2.3.1. Stochastic model of a joint time series of temperature and indicators of precipitation 53
2.3.2. Numerical models of time series of daily precipitation 65
2.4. Special stochastic models of joint meteorological time series 70
2.4.1. Stochastic models of joint time series with one-dimensional distributions as a mixture of two normal distributions 70
2.4.2. Models of joint meteorological time series with given discrete distributions and distributions as mixtures of two normal distributions 72
2.4.3. Numerical experiments 74
2.5. Numerical stochastic model of spatial fields of daily sums of rainy precipitation 75
2.5.1 One-dimensional distribution of uniform spatial fields of daily liquid precipitation sums 76

2.5.2. Correlation structure of spatial fields of liquid precipitation daily sums 78
2.5.3. Numerical simulation of spatial fields of daily sums of liquid precipitations 79
2.5.4. Simulation of conditionally distributed Gaussian fields 80
2.5.5. Approximate algorithm of numerical simulation of conditionally distributed fields of liquid precipitation daily sums. 83
2.5.6. Numerical experiments 85
BIBLIOGRAPHY 89