Temporal error estimate for statistical downscaling regional meteorological models

Egor V. Dmitriev[1], Ilva V. Nogotkov[1], Vladimir S. Rogutov[1], Gueorgui Khomenko[2] and Anatoly I. Chavro[1]

[1] Institute of Numerical Mathematics, Russian Academy of Sciences, 8, ul. Gubkina, 119991 GSP-1, Moscow, Russia, yegor@inm.ras.ru

[2] Laboratoire d'Oceanographie Cotiere du Littoral, ELICO, MREN, 32, Avenue Foch, 62930 Wimereux, France, khomenko@univ-littoral.fr

(Received 27 February 2007; received in revised form 27 April 2007; accepted 30 April 2007)

ABSTRACT:
Statistical downscaling models, which are applied for retrieval of small-scale geophysical fields from large-scale fields, allow obtaining *a priori* estimate of variance of the solution error and some other statistical characteristics of error. However, at given instants or even time periods the solution error of the considered problem can be much higher than its estimate that may be very important when using the results of downscaling. This paper is dedicated to testing the stochastic parameter known as "model reliability" as an indicator of temporal changes of the solution accuracy. For this purpose we considered several basic geophysical applications of downscaling. We show below that the probabilistic parameter "reliability of model" can be used for forecasting time points when the small-scale field is retrieved with high errors.
Key words: statistical downscaling, regional climate, temporal error estimate, inverse problems.

Estimación de errores temporales en modelos meteorológicos regionales de dowscaling estadístico

RESUMEN:
Los modelos de *downscaling* estadístico, que se aplican para la obtención de campos geofísicos de pequeña escala a partir de campos a gran escala, permiten obtener a priori estimaciones de varianza de los errores de las soluciones y algunas otras características estadísticas de los errores. Sin embargo, en determinados instantes o incluso periodos de tiempo los errores de las soluciones de los problemas considerados pueden ser mucho mayores de lo estimado, lo cual puede ser muy importante al usar los resultados del *downscaling.* Este trabajo se centra en comprobar el parámetro estocástico conocido como “fiabilidad del modelo” como un indicador de los cambios temporales de la precisión de la solución. Para este propósito, se han considerado varias aplicaciones geofísicas básicas del *downscaling*. Se muestra que el parámetro probabilista “fiabilidad del modelo” se puede usar para predecir puntos temporales donde el campo de pequeña escala se obtiene con grandes errores.
Palabras clave: *Downscaling* estadístico, clima regional, estimación de errores temporales, problemas de inversión.

1. INTRODUCTION

Resolution of modern general circulation models is about several degrees. It seems to be enough to study global climate changes. However, if for instance, our purpose is the assessment of environmental changes on regional scales, then such resolution is too coarse. Weather forecast models usually work on grids of about half degree, but it is also not enough for predicting weather on scales of a big city. The problem of retrieval of small-scale variability of meteorological fields from the output of global hydrodynamic models was coined the name "downscaling". Downscaling belongs to the class of ill-posed inverse problems. In order to predict or retrieve various small-scale meteorological fields either high resolution mesoscale hydrodynamic models or simple statistical models are employed. Implementation of regional hydrodynamic models is very expensive and time consuming, thus development of statistical downscaling models that work faster is an important problem.

Statistical downscaling methods are based on empirical relationships between large-scale variables and target local variables. It is assumed that global atmospheric models can provide a realistic description of atmosphere at large-scales. Theoretical analysis of predictability of atmospheric parameters showed (Leits 1998; Lorenz 1987; Grotch and MacCraken 1991) that one can expect a reliable long-term forecast only of large-scale fields and states of meteorological variables averaged over a long time interval. During the last decade a considerable progress of general circulation models has been achieved (Diansky and Volodin 2002), however, at the regional scales (< 2-4 thousand km) the results that were obtained in the model experiments are questioned. Assuming existence of statistical correlation between large-scale and regional fields of meteorological variables and employing data of measurements we can try to retrieve the regional field from large-scale data obtained in model experiments. In such a way (Kim et al 1984) retrieved the detailed structures of mean monthly precipitation and temperature fields in the state of Oregon from the spatially averaged values of these variables. A similar approach was also realized by (Storch et al 1991), where the precipitation field over Spain was retrieved from the large-scale pressure field over the North Atlantic. Works (Biau et al 1999; Zorita and Storch 1999) are devoted to the retrieval of the precipitation field over Iberia from the pressure field at sea level. Dmitriev et al. recently reconstructed daily surface air temperature in Moscow region from the large-scale temperature field obtained from reanalysis data (Dmitriev et al 2003).

Nowadays a number of different statistical downscaling models are developed and adjusted for different regions (Wilby and Wigley 1997). The most part of them allow to estimate *a priori* the variance of error and to calculate some other statistics of a similar type. The variance of error characterizes the solution accuracy "in general". It does not allow us to see if at some time points or even time periods the error of reconstruction is extremely high. The last two words mean for instance that the solution error is a few times higher than its *a priori* estimate. Predicted field of meteorological parameters such as temperature and precipitations may be used in critical decision making, when errors may lead to irreparable damage. So it is very important to develop a method allowing the temporal error estimate in statistical downscaling models.

Prediction of accuracy of each reconstructed sample seems to be impossible. However, one does not really need to do this. It is enough to predict only the most considerable error picks, which strongly increase the estimate of the mean-square solution error. The idea can be formulated as follows. Accuracy of statistical downscaling models based on the transfer function technique depends on homogeneity of data series used for the calibration. So it is important that correlation links are independent on time. In practice it rarely happens. The transition of the atmosphere from one stable state to another can cause significant changes of links between small-scale and large-scale fields. Therefore, the prediction of error peaks would be possible if we could recognize *a priori* this transition using some parameter characterizing the spatial distribution of the large-scale field.

In this paper we introduce the stochastic parameter known as the "model reliability" and check on various examples if it can be considered as an indicator of temporal changes of the solution accuracy. For this purpose we considered several different basic geophysical applications of downscaling. We demonstrate how this parameter works in statistical downscaling models for reconstructing local climatic and meteorological fields. We show below that the "model reliability" can be used for forecasting the strongest peaks of error, i.e. such time points, when the small-scale field is reconstructed with high errors.

2. METHODS

Most methods applied for downscaling are founded on three basic techniques: weather generators, transfer functions and weather typing (IPCC report 2001). All of them have their own advantages and disadvantages in representing the local climate. In this paper we consider the most frequently used methods, which are related to the transfer functions technique. Applying statistical downscaling methods we imply that geophysical parameters can be represented as a random function $X(s,t)$ depending on the space coordinate s and time t. If this function corresponds to the stationary ergodic stochastic process, we can consider global and local-scale geophysical fields as random vectors. Only when the spatial correlation length of this process is different from zero, one can assume existence of correlation links between large-scale and small-scale fields.

2.1. BASIC STATISTICAL DOWNSCALING MODELS

To outline the ideas, let us consider the linear problem

$$\xi = Af + \nu \tag{1}$$

where $\xi \in R^n$ is a known stochastic vector of a large-scale field, $f \in R^m$ is an unknown vector of a small-scale field, A is a linear operator, so-called observation matrix (Talagrand 1997), which acts from R^m to R^n and $\nu \in R^n$ is a random noise with an expectation $E\nu = 0$ equals to zero. In the framework of the considered inverse problem it may represent averaging or linear relations between small-scale and large-scale fields of different nature. In the former case the operator A is usually known and in the latter case, as a rule, it is unknown. In practice one often does not need to reconstruct the whole small-scale field, but only some of its patterns. Therefore we introduce a linear operator U acting from R^m to R^k, where $k \leq m$ and define sought vector $Uf \in R^k$ as a predictand and the known vector ξ as a predictor.

If the vector ξ is unknown, the equation (1) describes "averaging problem", which is the well posed problem in considered case. The equation (1) with unknown ξ is a forward problem, while the same equation (1) with known ξ is the inverse problem, or the problem of downscaling.

The solution of the inverse problem (1) consists in constructing such an operator R, which would allow us to obtain an optimal estimate $\hat{U}f$ of vector Uf from the known vector ξ. If $Ef = Ef\nu^T = 0$, then under the "optimal estimate" we will understand the estimate with the minimum variance of error or, more exactly with the minimum norm of cross-covariance matrix of errors. So in the case when the operator A is known, the operator R can be found by minimizing the functional

$$\Phi_1 (R) = tr\, E\, [(RA - U) f + R\nu]\, [(RA - U) f + R\nu]^T \tag{2}$$

Pytiev (Pytiev 1989) found that the solution of the minimization problem (2) is specified by the operator

$$R_1 = UC_f A^T (AC_f A^T + C_\nu)^- \tag{3}$$

where $C_f = Eff^T$ and $C\nu = E\nu\nu^T$ are the covariance matrices of f and ν, and the superscript $(\ldots)^-$ signify pseudo-inversion (see appendix). Thus the optimal solution of the inverse problem (1) and corresponded covariance matrix of errors $\varepsilon \in R^k$ take the form

$$\hat{U}f = R_1 \xi, \quad C_\varepsilon = U\, (C_f - R_1\, AC_f)\, U^T \tag{4}$$

In this case we assume that the model $[A, U, C_f, C_\nu]$ is defined.

If the operator A is unknown, we minimize the functional

$$\Phi_2 (R) = tr\, E\, (R\xi - Uf)\, (R\xi - Uf)^T \tag{5}$$

Equation (5) is equivalent to (2), but does not contain the operator A. The solution of the minimization problem (5) (Chavro and Dymnikov 2000) is specified by the operator

$$R_2 = C_{Uf,\xi}\, C_\xi^-, \tag{6}$$

where $C_{Uf,\xi} = EUf\xi^T$ is a cross-covariance matrix of random vectors Uf and ξ. The solution of the considered problem and the covariance matrix of error have the form

$$\hat{U}f = R_2\xi, \quad C_\varepsilon = C_{Uf} - R_2 C_{\xi,Uf}. \tag{7}$$

When the operator A is unknown, the covariance matrix of noise C_ν also can not be specified. Therefore, in this case we consider the model $[C_{Uf}, C_\xi, C_{Uf,\xi}]$.

The operators R_1 and R_2 are called "reduction operators" and the methods described above are known as the “reduction methods” (Pytiev 1989). It is worth to note that if the covariance matrix C_ξ is non-singular then the solution obtained by reduction method completely coincides with the solution obtained by multiple linear regression (Storch and Zwiers 1999), thus the reduction method can be interpreted as the generalized multiple linear regression.

2.2. FILTERING OF THE PREDICTOR VECTOR

In practice, one does not know exact values of statistical moments for a high-dimensional predictor vector but only their estimates obtained from a finite data series of measurements. In this case, solving equation (1) by the method of reduction may lead to the well-known overfitting problem. This problem arises because estimates $\hat{R}_1$ and $\hat{R}_2$ of the linear operators R_1 and R_2, respectively, are sensitive to small variations of estimate $\hat{C}_\xi$ of the predictor covariance matrix C_ξ. It means that the regression estimate is not robust. To overcome this difficulty, filtering of noninformative components of high-dimensional predictor is usually applied before statistical downscaling. Usually statistical downscaling models based on simple multiple regression are more sensitive to filtering of high-dimensional predictor than reduction methods. One of the best-known techniques of such filtering is founded on the EOF analysis of a predictor (Obukhov 1960; Storch and Zwiers 1999). The predictor vector can be decomposed into eigenvectors $\{\varphi_i\}_{i=1}^n$ of the matrix C_ξ and represented as an expansion in finite series

$$\xi = \sum_{i=1}^{n} (\varphi_i, \xi)\, \varphi_i \equiv \sum_{i=1}^{n} a_i\, \varphi_i$$

that provides maximum convergence rate of given series. Variances of random variables a_i are equal to the corresponding eigenvalues λ_i. Regions of higher variability of geophysical parameters of interest stronger affect their average values that have to be retrieved.

In the framework of considered problem, it appears that the most informative predictor components that contribute to the regression estimate are EOF harmonics with largest variability. Having found the threshold λ_{min}, which determines the noise level, one can assume that more noisy harmonics with a small variability $\lambda < \lambda_{min}$ have to be removed from the predictor. Let $\tilde{n}$ be the number of EOF harmonics retained as a result of predictor reduction or, in other words, the dimension of the predictor after reduction of its original dimension. Then, the parameter

$$\alpha\,(\tilde{n}) = \left(\sum_{i=1}^{\tilde{n}} \lambda_i\right) \left(\sum_{i=1}^{n} \lambda_i\right)^{-1}$$

introduced, in (Chavro and Dymnikov 2000), determines the contribution of the first $\tilde{n}$ harmonics to the variability of the predictor before its filtering. If the number $\tilde{n}$ yields a minimum in $|\alpha(\tilde{n}) - 0.95|$ for all $\tilde{n} \in [\overline{1,n}]$, the EOF filtering of the predictor is said to be done at the 95% level (usually this is denoted as EOF 95%).

In this paper we suggest also an alternative method for estimating the optimal number $\tilde{n}$ of EOF harmonics. Considering only the data remained for the calibration of the statistical model, we start from $\tilde{n} = 1$ and increase it until the norm of the solution error covariance matrix C_ε continues decreasing. It is important to exclude from consideration samples used for estimating the matrix C_ε into the calibration set. Thus samples of the solution error vector ε must be determined, for example, by cross validation method (Storch and Zwiers 1999). This allows us to avoid the overfitting problem, when the matrix C_ε is estimated.

Another type of filtering consists in rejecting spurious components of the predictor vector. For this purpose, the recurrent reduction (RR) method is used (Pytiev and Chulichkov 1989). This method is a recurrent procedure with the maximum number of steps equals the dimension of the predictor vector ξ. Let us consider the model $[A,U,C_f,C_\nu]$ and suppose that the noise vector components $\nu_1,\nu_2,\ldots,\nu_n$ are not correlated, operator $AC_fA^T + C_\nu$ is not degenerated and $\|U\|<\infty$, then the solution of the problem (1) and the estimate of the covariance matrix of errors (4) can be derived from recurrent relations

$$\begin{aligned} f^{(k)} &= f^{(k-1)} + C_f^{(k-1)}\, a_k \frac{\xi_k - (a_k, f^{(k-1)})}{(a_k, C_f^{(k-1)}\, a_k) + \sigma_k^2}, \\ C_f^{(k)} &= C_f^{(k-1)} - \frac{(C_f^{(k-1)}\, a_k)\,(C_f^{(k-1)}\, a_k)^T}{(a_k, C_f^{(k-1)}\, a_k) + \sigma_k^2} \end{aligned} \tag{8}$$

in n steps with initial conditions $f^{(0)} = 0$, $C_f^{(0)} = C_f$. Therefore $\hat{U}f = Uf^{(n)}$, $C_\varepsilon = UC_f^{(n)}\, U^T$, vectors a_k are the transposed lines of the matrix A and the condition $(a_k, C_f^{(k-1)}\, a_k) + \sigma_k^2 > 0$ is satisfied for all steps $k = 1,2,\ldots,n$.

At each step k one can determine an influence of each coordinate of the vector ξ on the solution of the inverse problem (1) using the recurrent reduction method. For this purpose we introduce the statistical parameter

$$\beta_k\,(i) = \frac{tr\,[(UC_f^{(k-1)}\, a_i)\,(UC_f^{(k-1)}\, a_i)^T]}{(a_i, C_f^{(k-1)}\, a_i) + \sigma_k^2} \tag{9}$$

called the informativeness. Using this parameter we can construct the following algorithm of filtering. At the first step of the recurrent reduction, i.e. when $k=1$, we find the number $i \in [\overline{1,n}]$ of the predictor vector component having maximum infomativeness. The chosen component is removed from the current predictor vector and it will be the first component of the new predictor. Thus the current predictor has dimensionality n-1. At the next step $(k=2)$ we also find corresponding to the component ξ_i with maximum of infomativeness, but now we already search in the subset of the remained n-1 components. The newly found component ξ_i is also removed from the current predictor vector and it will be the second component of the new predictor. And so on. This process provides the largest decrease of the norm of error covariance matrix and, therefore, the maximum convergence rate at the each step k of recurrent reduction. When on some step k the informativeness of all residuary ξ_i becomes statistically insignificant, the process stops. Using this algorithm we can avoid the overfitting problem and obtain the relatively stable solution of the inverse problem (1). It should be noted that the recurrent reduction algorithm does not use a matrix inversion, which makes it more preferable computationally.

2.3. MODEL RELIABILITY

In order to verify agreement of a statistical downscaling model with its input data we applied the theory of testing of statistical hypotheses. Distribution function of vector ξ, which represents the output data of large-scale model, is determined by properties of the model and the distribution function of vector f is determined by real physical processes and methods of measurements. We consider below the case when the operator A and covariance matrices C_f and C_ν are known and thus the statistical model $[A,U,C_f,C_\nu]$ is specified. We would like to note that matrices C_f and C_ν can be evaluated from various samples and for each sample may be different. Then we obtain different models: $\mu_1 = [A,U,C_f^{(1)},C_\nu^{(1)}]$, $\mu_2 = [A,U,C_f^{(2)},C_\nu^{(2)}]$, … and different reduction operators respectively. In other words, we obtain a certain class of models $\mathbf{M}$.

In order to choose appropriate solution of the inverse problem we introduce the concept of "model reliability", which was suggested in (Pytiev 1989). If one assumes that the vector ν has a normal distribution $N(0, C_\nu)$ in the model $[A,U,C_f,C_\nu]$ defined for the problem (1), then calculation of model reliability is reduced to testing the null hypothesis H_0: $\xi \sim N(0,C_\xi)$ against the alternative hypothesis H_1: $\{\xi \sim N(a, \tilde{C}_\xi), ||a|| \neq 0, ||\tilde{C}_\xi|| > ||C_\xi||\}$. To test the hypothesis H_0 we construct a stochastic value

$$t_n = ||\Omega^{-1/2}\, \xi||^2 \tag{10}$$

t_n must have the χ^2-distribution if the hypothesis H_0 is accepted. For the considered model the matrix Ω takes the form $\Omega = (AC_fA^T + C_\nu)$, but if the model is

$[C_{Uf}, C_\xi, C_{Uf,\xi}]$ then Ω is reduced to C_ξ. In general, the statistics t_n may take any non-negative value, however for some intervals the hit probability is very small. These intervals form so-called critical region. The theoretical probability $P(t_n > t_n^\alpha)$ that t_n lies in the critical region is given by

$$P\,(t_n > t_n^\alpha) = \int_{t_n^\alpha}^{\infty} p_{\chi^2}^{(n)}(x)\,dx, \tag{11}$$

where $\alpha \in [0,1]$ is the significance level specified *a priori*, t_n^α is a boundary of the critical region and $p_{\chi^2}^{(n)}$ is the probability density of the χ^2-distribution with n degrees of freedom. The equation (11) allows computing the parameter t_n^α if the significance level α is specified. The equation (10) allows to evaluate t_n in terms of known ξ. The hypothesis H_0 should be rejected if $t_n > t_n^\alpha$ (t_n belongs to the critical region), however, if $t_n \le t_n^\alpha$ then the hypothesis does not contradict to downscaled realization ξ of the large-scale model.

In contrast with the significance level α, the model reliability τ is determined as a function of known vector ξ

$$\tau_\mu\,(\xi) = \int_{t_n}^{\infty} p_{\chi^2}^{(n)}(x)\,dx. \tag{12}$$

One can see from the equation (12) that the model reliability can be interpreted as a minimum probability of the erroneous rejection of the hypothesis H_0. The model reliability is a probability measure, so its values belong to the interval [0,1]. Models with low values of reliability are usually should be rejected.

Solving the inverse problem we can use the model reliability parameter for selection of a best model μ from the model class **M**. The selection algorithm based on the principle of maximum model reliability (Pytiev 1989), can be formulated as following optimization problem

$$\mu \quad \leftrightarrow \quad \tau_\mu\,(\xi) = max\,\{\tau_{\mu i}\,|\mu_i \in \boldsymbol{M}\} \tag{13}$$

We would like to note that for it's realization we use only observations of vector ξ. Also we can set a minimum reliability threshold τ_{min} and select for the interpretation only those events for which $\tau > \tau_{min}$.

The statistics $t_n(\xi)$ also allows us to construct an indicator $\beta(\xi)$ for optimal choice of the model for downscaling. In this section we show that the parameter

$$\beta\,(\xi) = \frac{\sum_{i=1}^{n} sign\,(\xi_i)}{n}$$

can be used to distinguish extreme fluctuations of atmospheric parameters. Using values $t_n(\xi)$ and $\beta(\xi)$ we divide the dataset of vectors f and ξ, reserved for calibration of the statistical downscaling model, into four subsets

$$S_0 = \{f, \xi: t_n(\xi) < Et_n(\xi)\}$$

$$S_1 = \{f, \xi: \beta(\xi) = 1 \text{ and } t_n(\xi) > Et_n(\xi)\}$$

$$S_2 = \{f, \xi: \beta(\xi) = -1 \text{ and } t_n(\xi) > Et_n(\xi)\}$$

$$S_3 = \{f, \xi: -1 < \beta(\xi) < 1 \text{ and } t_n(\xi) > Et_n(\xi)\}$$

Subsets S_1, S_2 and S_3 contain realizations of vector ξ having unusual spatial distribution and we associate them with extreme situations. Constructing individual statistical models for each of the subsets S_i we can improve solution of the inverse problem (1) if the correlation links between vectors f and ξ significantly vary from one subset to the other.

3. NUMERICAL EXPERIMENTS

In this section we illustrate the methods described above on a few basic examples. We considered statistical downscaling of the fields of a temperature, pressure and precipitations. These fields have different spatial correlation length. The problem is solved at several temporal scales ranging from a day to a season. In each example we evaluated the model reliability and verified if it links with the accuracy of the solution.

3.1. RECONSTRUCTION OF MONTHLY MEAN AIR TEMPERATURE AT THE METEOROLOGICAL STATIONS OF CIS

As a first example we considered reconstruction of the mean monthly surface air temperature field at meteorological stations of Commonwealth of Independent States (CIS) from averaged values of this field at scales of 2000-4000 km. We analyzed data provided by the All-Union Scientific Research Institute of Hydrometeorological Information-World Data Center on mean monthly air temperature measured at 98 weather stations of the CIS for the period of about 35 years from January 1957 to April 1993. The root-mean-square error of measurements is about 0.13°C. Observation data for the first 32 years (384 samples) were used for calibration of the statistical model and the remaining 40 samples were reserved for independent validation of the proposed method. The territory of CIS was subdivided into 15 regions of similar sizes and thus we simulated the global field at the scales > 2000 km. We calculated the annual variation of mean monthly temperature and then subtracted it from observations, so that the inverse problem (1) was solved for temperature anomalies. We considered the model $[A,U,C_f,C_\nu]$ since in this case we know exactly the averaging operator A. We considered only the northern part of Western Siberia and in order to separate components of the predictand vector *f* corresponding to this region, we constructed the operator *U*. The natural variability of mean monthly temperature in this region rised up to 3.27°C. The *a priori* estimate of the solution error equals 1.03°C, so we are able to retrieve approximately 70% of temperature variations in this region.

For independent validation of the statistical model we used 40 samples which were not included in the calibration ensemble. Since inverse problem is usually used in conjunction with general circulation models, it is essential to estimate sensibility of the solution error to the accuracy of input data that usually are output of GCM. This analysis allows us to specify requirements to GCM in order to determine its relevance for downscaling. We performed the test of relevance of input data set of 40 samples perturbing large scale fields by white noise. The result is pseudo-random vector normally distributed with zero expectation and variance with pairwise uncorrelated components. The inverse problem (1) was solved. The RMS error of the large scale field, obtained from GCM, *a priori* estimate of RMS error and the independent estimate of the RMS error were evaluated. The results are presented in the Table 1.

Table 1.- Dependence of inverse problem solution error on input data error.

Root-mean-square error of large-scale fields, °C	**An a priori of root-mean-square error, °C**	**Root-mean-square error obtained from independent validation, °C**
0.000	1.032	1.093
0.130	1.146	1.109
0.173	1.056	1.117
0.548	1.213	1.250
1.000	1.144	1.474
1.225	1.553	1.585
2.500	2.050	2.079

From the Table 1 one can see that the proposed method allows us to retrieve more than 30% of the variations of temperature anomalies if the input data error does not exceed 2.5°C. Natural variability of monthly mean temperature in the considered region rises up to 3°C that is in agreement with estimate 2.5°C. *F*-test showed that the difference between the errors on calibration and validation periods is not significant. It is an additional evidence of the stability of the statistical downscaling model. Typical examples of reconstruction of small-scale monthly mean temperature fields in the north part of Western Siberia are shown in Fig. 1. The error of reconstruction for these samples is close to its root-mean-square value for the whole validation period. We can see that in spite of significant differences in absolute values of real and reconstructed temperature, the structure of the field is reproduced rather well.

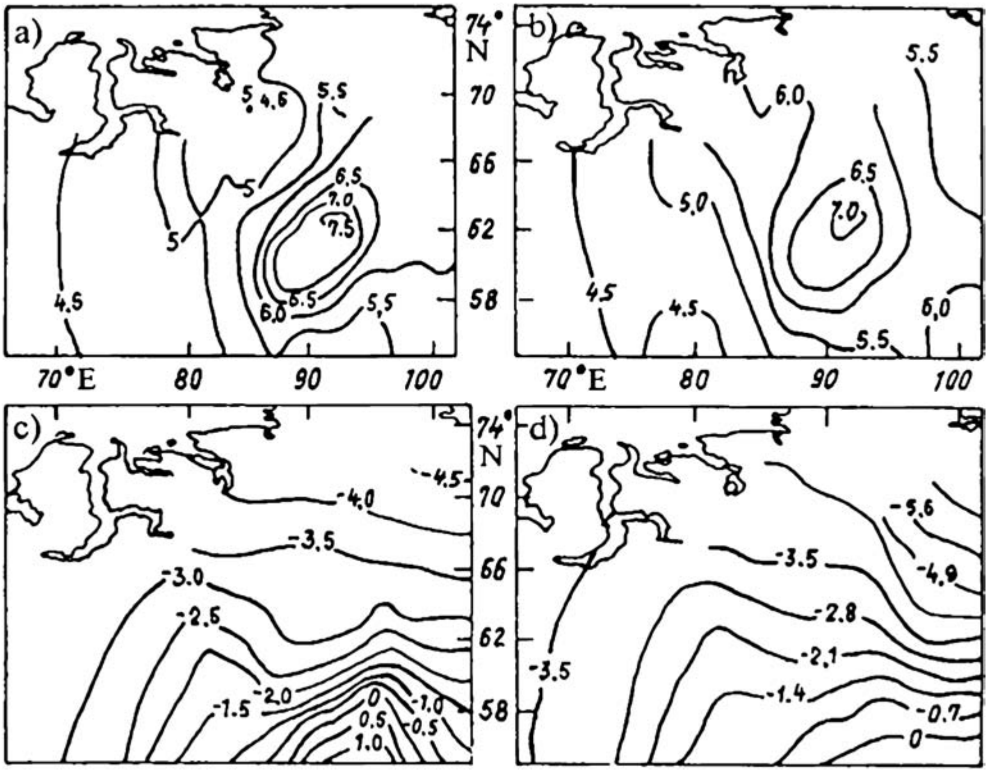

Figure 1.- Reconstruction of monthly mean temperature anomalies in the north part of Western Siberia. On the left - exact fields, on the right - reconstructed fields. a), b) - February 1992; c), d) - November 1992.

We evaluated the model reliability τ for the whole validation period using the equation (12). Let us see now if this parameter can be considered as a measure of agreement between the statistical model and input data. The specific values of solution errors in the north part of Western Siberia and the model reliability are shown in Fig. 2. Obviously there is a negative connection between error and model reliability. The correlation coefficient is relatively low, its value is -0.43. However we can see that the strongest peaks correspond to the almost zero model reliability, without exceptions. It means that the model reliability can be used for filtering samples, which are reconstructed with unusually low accuracy. Of course there is a chance to reject "well-reconstructed" samples applying this approach. However from the practical point of view it is better to reject erroneously good samples than to pass even one wrong sample.

Analysis of this figure shows that we need to introduce a set of models. One can see that the model reproduces the monthly temperature much better in summer than in winter when the model has low accuracy. So we divided the initial *a priori* ensemble into the summer and winter periods, specified the reliability level 0.1 and

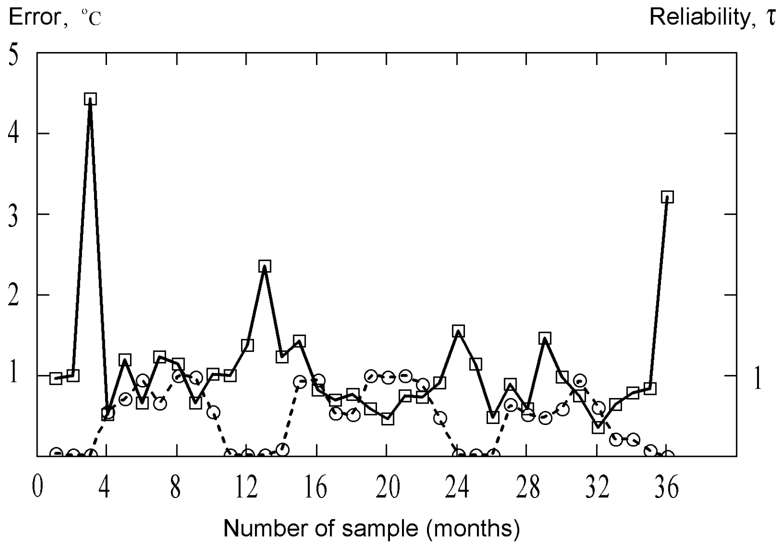

Figure 2.- The error of reconstruction of monthly mean temperature anomalies in the northern part of Western Siberia (solid line) and the model reliability (dashed line).

eliminated events with the model reliability < 0.1. After that the solution error decreases to the value 0.7°C.

3.2. RECONSTRUCTION OF NEAR-SURFACE TEMPERATURE IN THE MOSCOW REGION

Another kind of downscaling is the interpolation of output data of weather forecast models to location points of meteorological stations. The modern short-term forecast models have a resolution of about 0.5 degrees which is not high enough, to describe the weather changes at the scales of a big town so that the downscaling should be applied. Here we consider the test of statistical downscaling of minimum, maximum and daily near-surface temperature in Moscow region. We used the dataset provided by Hydrometeorological center of Russia. It contains six- and three-hourly values of surface air temperature measured at 16 meteorological stations in Moscow region for the period from January 1936 to December 1980. Six meteorological stations are located in Moscow and others are in suburban area. As a predictor for daily temperature we used reanalysis data obtained in cooperation with the U.S. National Meteorological Center (now the National Center for Environmental Prediction, NCEP) and the National Center for Atmospheric Research, NCAR). This project will be referred to as the NCAR/NCEP reanalysis (Batist and Chelliah 1997). In present study we used these data taken on a rather course grid (~ 2.5 x 2.5°), which does not correspond to the resolution of weather forecast models, but it seems to be enough for present example.

The dataset with observations at meteorological stations in Moscow region contains a lot of gaps, but after calculating daily mean temperatures their number decreases. Therefore when we reconstructed the small-scale daily temperature we calibrated the statistical model within the first 35 years (1936-1970) and the remaining 10 years (1971-1980) were used for verification. For the reconstruction of extreme temperatures we considered only the period from 1965 to 1974 year. Indeed, for estimating minimum and maximum daily temperature, we need to have available whole measurements during a day, because in this case the interpolation cannot be applied successfully. We have defined the calibration period from 1965 to 1969 and the validation period from 1970 to 1974.

Our goal is to reconstruct minimum, maximum and daily near-surface temperatures at 6 meteorological stations in Moscow. As a predictor we used observations in Moscow region and for reconstruction of daily temperature we used also the NCAR/NCEP reanalysis. The predictors have the similar spatial scales, but observations in Moscow region reproduce the observations in Moscow much better than the reanalysis. To make our test more realistic we have perturbed the data in Moscow region with a white noise 50%. It means that the noise has the same variance as clear signal, thus the contribution of noise raises to one half of the total signal. This level of noise corresponds to the level of error of 3-day forecast of the temperature. To reduce the influence of within-year variability on the estimate of statistical parameters we calculated annual variations of the considered fields with regard to the leap years. The inverse problem (1) was solved for temperature anomalies. The results are presented in the Table 2.

Table 2.- Results of near-surface air temperature reconstruction in Moscow.

Predictor	***A priori* of absolute error estimate, °C**	***A priori* of relative error estimate, %**	**Absolute error estimated within the validation period, °C**	**Relative error estimated within the validation period, %**	**Correlation between the model reliability and the error**
Minimum daily temperature					
noise 0%	1.27	23.8	1.22	22.8	-0.55
noise 50%	2.06	38.6	1.98	37.2	-0.20
Maximum daily temperature					
noise 0%	1.16	22.5	1.02	19.8	-0.47
noise 50%	1.96	38.0	1.85	35.9	-0.31
Daily mean temperature					
noise 0%	0.73	15.1	0.79	16.5	-0.41
NCAR/NCEP	1.84	38.3	2.09	43.4	-0.38

Natural variability of the daily temperature in Moscow is approximately 4.8°C and for extreme temperature it is a little more than 5.2°C. Our experiments show that in real downscaling of the weather forecast output we can hope to retrieve more than 60% of variations of small-scale temperature field. We can see that the reconstruction of daily temperatures is more accurate than the reconstruction of extreme temperatures. The daily temperature field is smoother and has larger spatial correlation length then the extreme values of temperature that explains more accurate retrieval of daily temperature. Root-mean-square errors estimated using validation data set, in general, correspond to *a priori* estimates. But in some experiments the difference is significant. We revealed that this difference can not be explained by fluctuations of statistical moments only, but rather related to gaps in the data sets. Indeed, the mean-square solution error differs from one station to the other by 0.7°C. It means that if, for instance, there is a long series of missing data at one or more stations in the validation period and in the calibration period we have another gap pattern, then the general mean-square error will be shifted to the value at one of stations.

The model reliability τ was calculated for all reconstructed temperature fields. In Figs. 3 and 4 the parameter τ is shown only partially, but the correlation coefficients between the model reliability and the solution error were calculated in the whole

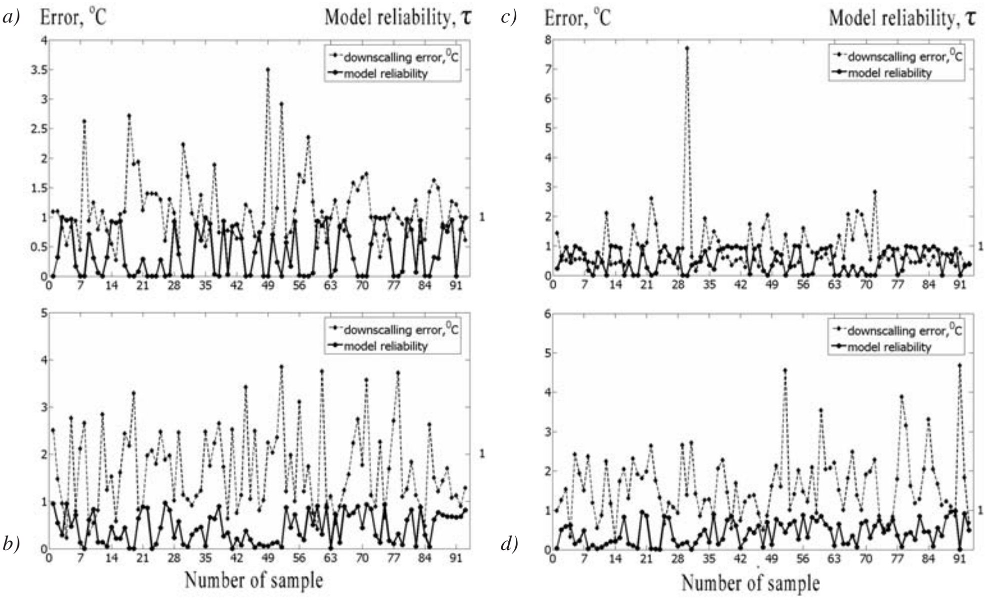

Figure 3.- The error of reconstruction of extreme near-surface air temperature at the meteorological station in Moscow (dashed line) and the model reliability (solid line). a)- reconstruction of the minimum daily temperature at 0% noise level; b) - the same, but at 50% noise level; c) - reconstruction of the maximum daily temperature at 0% noise level; d) - the same, but at 50% noise level.

validation period. The results are presented in table 2. In this case we also can see the negative association between error and model reliability. Correlation coefficients are similar to the previous example. However, the correlations strongly decrease when the predictor is perturbed by white noise. From the Fig. 3 we can conclude that in this case the model reliability can not be successfully used even for predicting error peaks.

The correlation coefficient between the model reliability and the error remains almost the same in both cases, when the NCAR/NCEP reanalysis data are used as the predictor and for reconstruction of daily temperature in Moscow from the observations in Moscow region. Furthermore, at Fig. 4 we can see that all strongest peaks of the solution error are reproduced by the model reliability. Therefore the question is: why the behavior of the model reliability is so different in these two cases? As a matter of fact this happens, because the NCAR/NCEP reanalysis data are based on the physical rules in contrast to the white noise. Of course the reanalysis may reproduce the real data with large errors, but at the same time it preserves characteristics responsible for the spatial distribution, for instance, such as the correlation range. The white noise has the spatial correlation length equals to zero, which is never realized for real temperature fields. Adding white noise to the predictor we affect the spatial distribution of large-scale temperature field and therefore break prerequisites, which were used in the definition of the model reliability. So in this sense the reconstruction small-scale temperature field in Moscow from the NCAR/NCEP reanalysis data is more accurate as a test, than the reconstruction from noisy observations in the Moscow region.

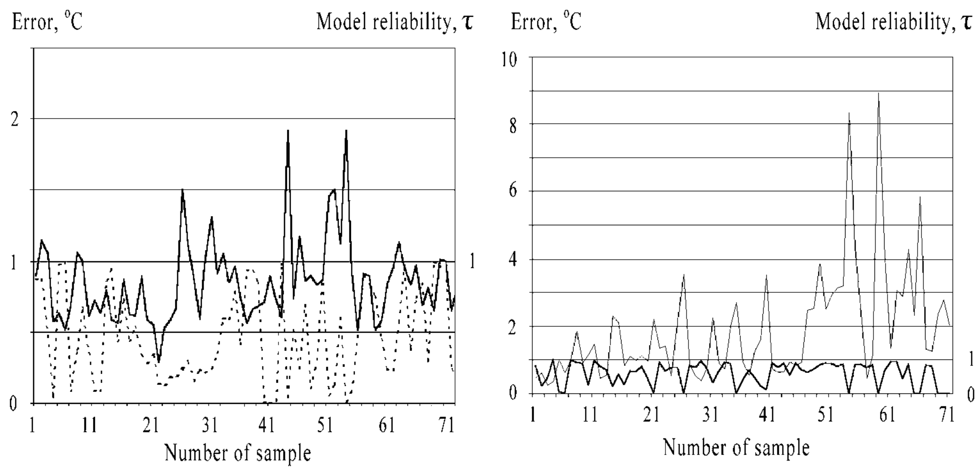

Figure 4.- The left plot represents the reconstruction error of daily near-surface air temperature at weather stations in Moscow (solid line) obtained from averaged observation data in Moscow region and the model reliability (dashed line). The right plot represents the reconstruction error of daily near-surface air temperature at weather stations in Moscow (solid thin line) obtained from NCAR/NCEP reanalysis and the model reliability (thick line).

3.3. RECONSTRUCTION OF DAILY AIR PRESSURE AT STATION LEVEL IN THE CENTRAL REGION OF RUSSIA

A downscaling of the near-surface pressure is not the subject of wide discussion. However we would like to consider this field because it is smooth enough and we will see how the behavior of the model reliability changes under different spatial correlations of the reconstructed field. We used the data measured of air pressure at stations level provided by All-Russian Research Institute of Hydrometeorological Information World Data Centre (RIHMI-WDC). The measurements were made 4 and 8 times per day at 233 meteorological stations in the territory of CIS in the period from 1936 to 1986 year. We disregarded the data from 1984 to 1986 year, because of the huge number of gaps in that period. After the calculation of daily pressure its accuracy amounted to 0.1 mb.

The large-scale field was simulated by means of interpolation from stations to the grid 10° x 10° using the near-neighbour method. For the validation of statistical downscaling model we chose the central region of Russia. There are 49 meteorological stations in this region, which are almost uniformly distributed with a small thickening in areas with complex orography, and it seems to be enough for the reliable validation of our methods. The first 35 years (the period from 1936 to 1970) were used for the calibration of the statistical model and the next 13 years were used for the independent validation. An a priori estimate of the solution error amounted to 2.1 mb. The natural variability of the near-surface daily pressure is about 8.7 mb in this region and it means that we can reproduce approximately 75% of variations of this field. The difference between the mean-square error estimated at the validation period, which equals to 2.3 mb, and the *a priori* error is small but significant, since both calibration and validation periods contain a huge number of samples. This is caused by strong within-year variability of the covariance between small-scale and large-scale fields of pressure.

We applied two different methods for the downscaling of daily near-surface pressure anomalies in the central region of Russia. The first one consists in the simple bilinear interpolation. Since there is no large-scale data on the west part of the border of this region, we applied the Cressman analysis (Cressman 1959) for the extrapolation. The other method is based on the recurrent reduction, which is described above. A typical reconstruction is shown in the Fig. 5. This example shows that the simple linear interpolation cannot be successfully applied even for downscaling of such smooth value as near-surface pressure, because the whole small-scale structure is lost (see Fig. 5). The result of statistical downscaling appears much better than the bilinear interpolation in spite of the significant error of statistical downscaling, which is comparable in magnitude with the error of bilinear interpolation. The spatial variations of the pressure field are slightly smoothed, but the small-scale structure is reproduced well. We also tried to correct the solution employing the pressure field from the previous day, but the improvement was not significant.

The validation period contains so many samples that it would be hard to compare visually the model reliability and the solution error. However, since the covariance

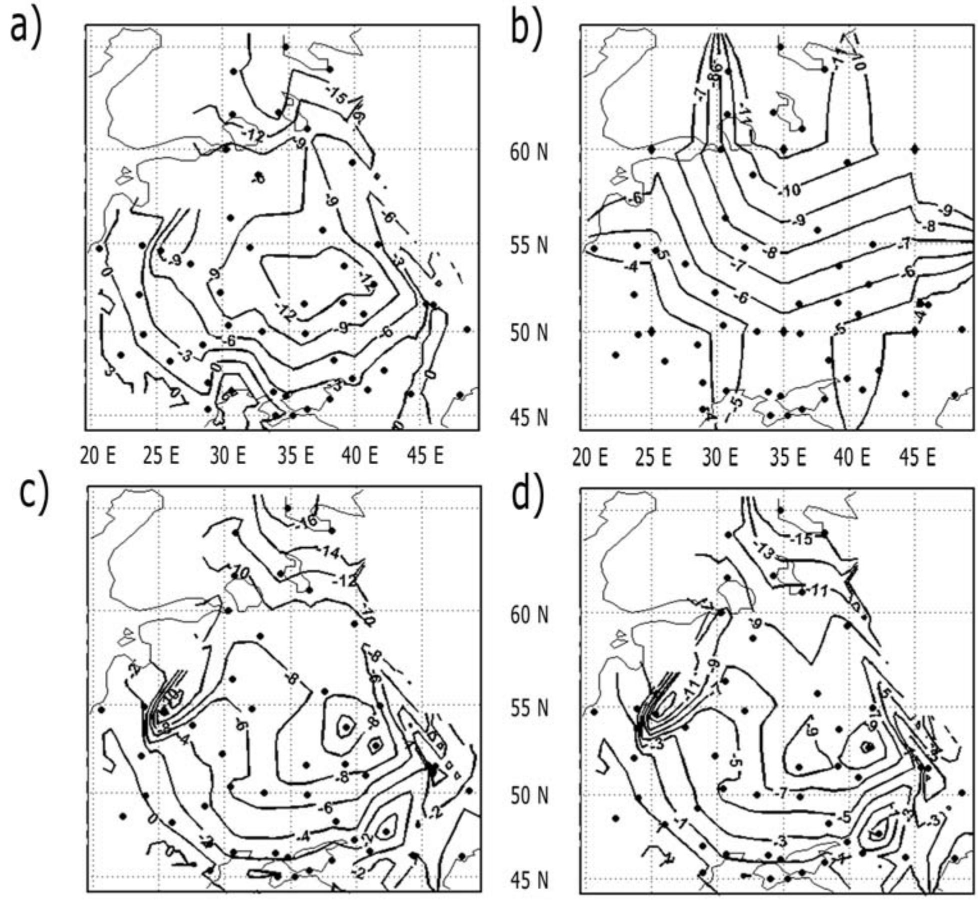

Figure 5.- Fields of daily near-surface pressure anomalies in the central region of Russia. a) - exact; b) - reconstructed by combined linear interpolation (bilinear interpolation and Cressman analysis); c) and d) - reconstructed by recurrent reduction method without and with the use of measurements from previous day, respectively.

between small-scale and large-scale fields of pressure undergoes strong within-year changes, then it is natural to expect strong within-year changes of the solution error and the model reliability. These characteristics were smoothed by Gaussian filter mean and represented on the Fig. 6. We can see that there is a strong negative link between error and model reliability. If one rejected non-homogeneity between 1000 and 1500 samples, the correlation coefficient equals to -0.85. The state and variability of the atmosphere in the middle latitudes are considerably different in summer and winter periods. Thus the correlation links between small-scale and large-scale near-surface pressure fields undergo significant seasonal variations. This is the cause of seasonal fluctuations of the solution error, which are successfully reproduced by the model reliability.

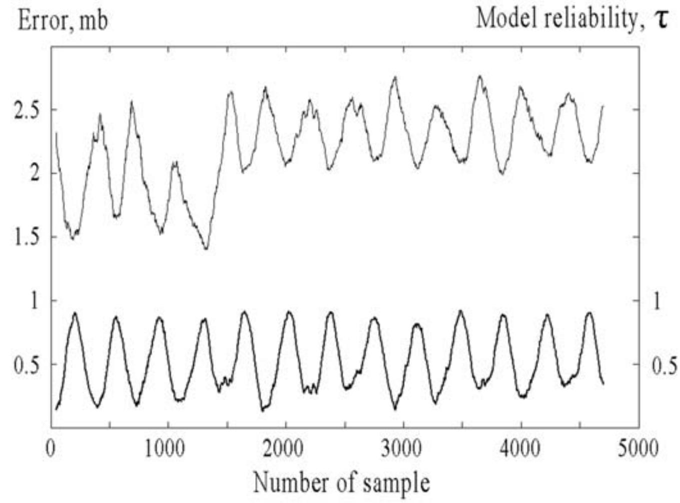

Figure 6.- The error of reconstruction of daily near-surface pressure in the central region of Russia (upper curve) and the model reliability (lower curve). Both parameters were smoothed by Gaussian filter.

3.4. RECONSTRUCTION OF WINTER PRECIPITATION IN EUROPE FROM PRINCIPAL EOFS OF THE SEA-LEVEL PRESSURE IN THE NORTH ATLANTIC

One of the most difficult problems in geophysics is the modeling of precipitation. It is especially difficult to reconstruct precipitation observed at meteorological stations. We used the dataset maintained by the Global Historical Climatology Network (GHCN), which can be obtained from the website of the National Climatic Data Center. It contains monthly precipitation measured at ~20000 meteorological stations for the period from 1697 to 2000. In Europe there is a great number of meteorological stations, but at many of them the precipitations were not observed regularly. So we have chosen 271 stations that provided series of monthly precipitation data with no more than 5 gaps for the period from 1950 to 2000.

As a predictor for winter precipitation in Europe we used NCAR/NCEP reanalysis data of monthly sea-level pressure obtained in cooperation with the National Center for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR). The data set covers the period from 1950 to 2000, so we have 51 samples of mean winter sea-level pressure. Instead of the global pressure field we considered only that part, which related to the North Atlantic and to Europe, and therefore the predictor is specified at 348 grid points. The spatial dimension of the predictor is much more than the number of samples and thus it is necessary to apply filtering. We used the EOF decomposition to reduce the dimension of the predictor. The number of principal components was calculated individually for each station by the method based on the principles of cross validation, which is described above. Since the number of samples is relatively small, we verified our statistical model also by the cross validation method.

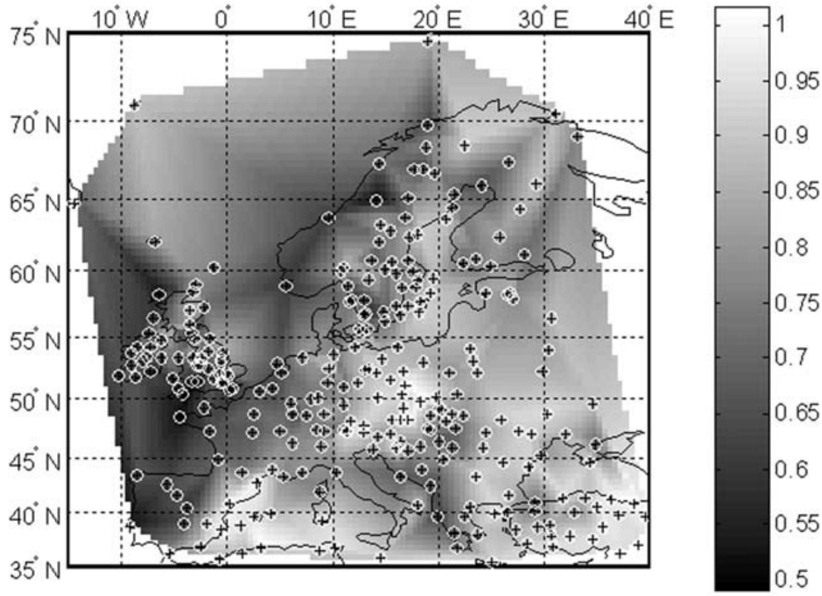

Figure 7.- Relative errors of the reconstruction of winter precipitation in Europe.

Mean-square errors of the reconstruction of winter precipitation in Europe normalized on the station values of natural variability are presented in Fig. 7.

The numerical experiments have shown that the solution accuracy is very low and we can not reconstruct more then 55% of the natural variability of precipitation. The best results can be obtained for the west coast of Europe, especially for Spain, the North of England and the western side of the Scandinavian Peninsula. In these areas it was possible to use the most number of EOFs. In Eastern Europe the reconstruction accuracy is very low.

The model reliability must be calculated individually for each station, because the dimension of the predictor varies from one station to the other. The correlation coefficients between the model reliability and the solution error are shown in Fig. 8. They vary in the interval from -0.4 to 0.3, but in general the correlations are either positive or so low that we can not prove their significance.

We can not see the results obtained in the previous examples even for the stations at which the correlation takes considerable negative values. For instance, the solution error and the model reliability for the station Bodo, presented in the Fig. 9 (left) do not reveal considerable negative association. Many error peaks correspond to high values of the model reliability. At the same time the correlation coefficient between them is about -0.31. In contrast with this example, the comparison of the model reliability and the error averaged for all stations in Spain reveals much stronger negative association (see Fig. 9 right). In this case the correlation coefficient amounts to -0.51 and peaks of error correspond to minimums of the model reliability. However, for predicting high solution errors we need to define a critical value and to rejects realizations, which correspond to the model reliability

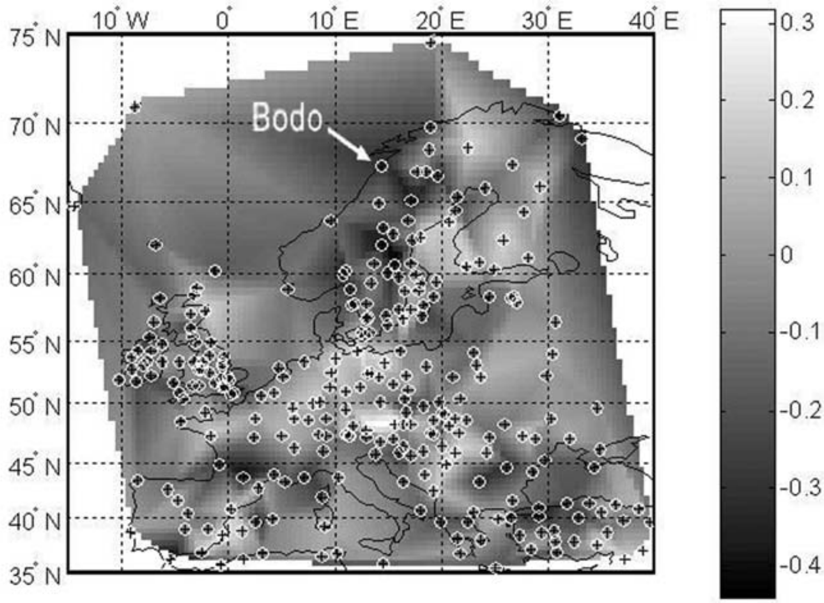

Figure 8.- Correlation coefficients between the solution error and the model reliability.

lower than this value. Looking at the Fig. 9, it is difficult to decide how to choose correctly the critical value, because the model reliability corresponding to the strongest peaks of errors varies in a wide range. So we can conclude that, for the time being, the temporal error control for downscaling of precipitation is questionable and we need to construct another characteristic for such kind of problems. Also the considered example shows that the model reliability can not be used to predict the solution error peaks at individual stations, but only in a region.

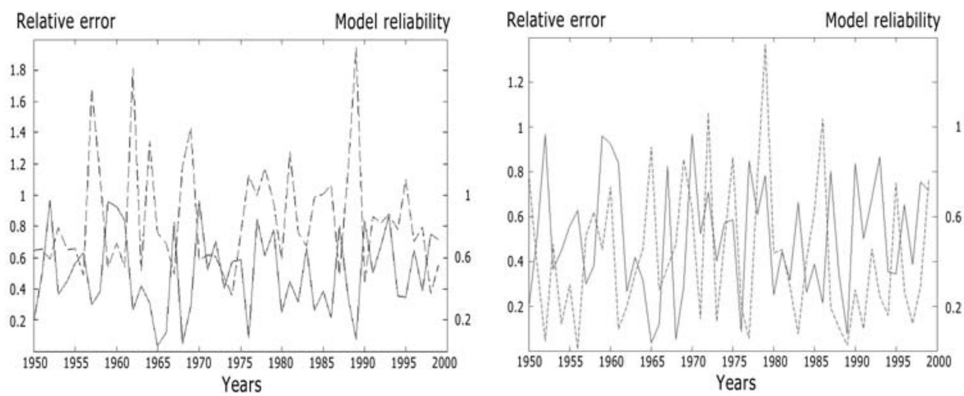

Figure 9.- The model reliability (solid line, the both plots) and the solution error (dashed line) calculated for the station Bodo (the left plot, see Fig. 8) and the mean error for Spain (the right plot).

4. CONCLUSIONS

In this paper we considered the problem of temporal error estimate for statistical downscaling models. We have introduced the probabilistic characteristic named "the model reliability" and studied the possibility to use it for predicting temporal changes of the solution accuracy. For this purpose we have considered several typical problems of statistical downscaling at different temporal and spatial scales. The numerical experiments have shown that there is an obvious negative association (or mutual correspondence) between the model reliability and the solution error. It seems to be possible to recognize *a priori* such cases when the statistical downscaling model reconstructs the small-scale field with a high error. This considerably improves an estimate obtained from the calibration period. At the same time we see that the correlation coefficient between the model reliability and the error has relatively low values. This means that we can not estimate all variations of the error. Otherwise, we could use the model reliability as an alternative predictor. At present we can dependably predict only the strongest peaks of the solution errors.

We considered here several typical problems of downscaling to illustrate the idea that variations of spatial distribution of large-scale fields may be a sign of changes of correlation links between geophysical fields at different scales. The given examples show that the use of the model reliability in statistical downscaling is justified when a reconstructed field has considerably large spatial correlation length, but we have seen that the practical utility of the model reliability was questionable when we considered the reconstruction of winter precipitation in Europe from near-surface pressure patterns that have much shorter spatial correlation range. We are sure that the usage of model reliability is a promising approach in the statistical downscaling of temperature fields obtained both from general circulation models and from weather forecast models. We hope that the concept proposed in this paper will be deepened and extended to other meteorological fields.

5. ACKNOWLEDGEMENTS

The work was supported by the Russian Foundation for Basic Research (projects 07-05-00328-a and 06-05-64916-a) and a grant from President of the Russian Federation MK-661.2006.5.

6. APPENDIX. PSEUDO-INVERSION, DEFINITION AND GENERAL PROPERTIES OF THE MORE-PENROSE MATRIX

In the framework of this article it is enough to consider only the space of real matrices. Let us consider a matrix $A \in M_{m,n}$ and its singular value decomposition $A = V\Lambda V^T$. The matrix $A^- = W\Lambda^- V^T$ is called the More-Penrose matrix or the pseudo-inverted matrix (Horn and Johnson 1990), if Λ^- is obtained from Λ through the transposition and the substitution of its positive elements on inverted values. This matrix has following properties:

1) $(AA^{-})^T = AA^{-}$ and $(A^{-}A)^T = A^{-}A$;
2) $AA^{-}A = A$;
3) $A^{-}AA^{-} = A^{-}$.

These properties uniquely define the matrix A^{-}.

It is easy to show that $A^{-T} = A\,(A^TA)^{-1}$ in the case $m \geq n$. Indeed

$A(A^TA)^{-1} = V\Lambda W^T\,(W\Lambda^T V^T V\Lambda W^T)^{-1} = V\Lambda(\Lambda^T\Lambda)^{-1}\,W^{-1} = V\Lambda^{-T}W^T = (W\Lambda^{-}V^T)^T = A^{-T}$.

This matrix also has another useful property. For the system of linear equations $Ax = b$, the vector $\hat{x} = A^{-}b$ minimizes the Euclidean norm of the discrepancy $||Ax - b||_2$ and $\hat{x}$ is known as the least square solution. For example if matrices F and Ξ rows of which are realizations of stochastic vectors f and ξ then the estimate of the reduction operator reads

$$\hat{R}_2 = F^T\,\Xi\,(\Xi^T\Xi)^{-1} = F^T\Xi^{-T}$$

REFERENCES

BIAU, G., E. ZORITA, H. von STORCH & H. WACKERNAGEL (1999). Estimation of precipitation by Kriging in the EOF space of the sea level pressure field. *J. Climate*, 12, 1070-1085.

BATIST, Ä.N. & M. CHELLIAH (July 1997). Comparison of tropospheric temperatures derived from the NCEP/NCAR reanalysis, NCEP operational analysis, and the microwave sounding unit. *Bull. Amer. Meteorol. Soc*, 78, 7, 1431-1447.

CHAVRO, A.I. & V.P. DYMNIKOV (2000). *Methods of mathematical statistics in physics of the atmosphere.* INM RAS, Moscow, 210 pp. (In Russian)

CRESSMAN, G.P. (1959). An Operational Objective Analysis System. *Monthly Weather Review*, 87, 367-374.

DIANSKY, N.A. & E.M. VOLODIN (2002). Simulation of Present-Day Climate with a Coupled Atmosphere–Ocean General Circulation Model. *Izvestiya, Atmospheric and Oceanic Physics*, 38, 6, 824-840.

DMITRIEV, Y.V., RUBINSHTEIN K.G. & CHAVRO A.I. (2003). Downscaling the large-scale surface temperature for the Moscow region. *Meteorology and Hydrology,* 7, 19-30.

GROTCH, S. & M. MACCRAKEN (1991). The use of general circulation models to predict regional climate change. *J. Climate*, 4, 203-303.

HORN, R.A. & C.R. JOHNSON (1990). *Matrix Analysis*. Cambridge University Press, Cambridge, 575 pp.

KIM, J.-W., J.-T. CHANG, N. L. BAKER & W. L. GATES (1984). The statistical problem of climate inversion: determination of the relationship between local and large-scale climate. *Monthly weather review*, 112, 10, 2069-2077

LEITS, S. (1988). Theoretical and practical predictability. *In:* Hoskins B. & R. Piers (Eds.) *Large-Scale Dynamic Processes in the Atmosphere*, Mir, Moscow. 401-422. (Russian translation)

LORENZ, E. (1987). Some aspects of predictability of almospheric behavior. *In:* Burridge D. & E. Challen (Eds.) *Long- and Medium-Term Weather Forecasts. Problems and Prospects*, Mir, Moscow. 10-32. (Russian translation)

PYTIEV, U.P. (1989). *Mathematical methods for interpretation of the experiment.* Vysshaya Shkola, Moscow, 352 pp. (In Russian)

STORCH, H., E. ZORITA & V. CUBASCH (1991). *Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in winter time.* Max Planck Institut fur Meteorologie, Report N 64, Hamburg, Germany, 36 p.

STORCH, H. & F.W. ZWIERS (1999). *Statistical analysis in climate research*. Cambridge University Press, Cambridge, 484 pp.

TALAGRAND, O. (1997). Assimilation of Observations, an Introduction. *Journal of the Meteorological Society of Japan*, 75, 1B, 81-99.

WILBY, R. L. & T.M.L. WIGLEY (1997). Downscaling general circulation model output: A review of methods and limitations. *Prog. Phys. Geogr.*, 21, 4, 530–548.

ZORITA, E. & H. STORCH (1999). The analog method as a simple statistical downscaling technique; Comparison with more complicated methods. *J. Climate*, 12, 2474-2489.