ON GLOBAL WEATHER MODELING

(Empiric approach)

Romanov L.N.

Federal State Budgetary Institution "Siberian Regional Hydrometeorological Research Institute" (FSBI "SibNIGMI")
Senior Researcher, Novosibirsk

Science of Europe vol. 1, No 4(4) (2016)

Abstract

This paper deals with an approach, which allows constructinglarge-scale weather models as well as the modals of the global size. The construction of such models is being realized on the basis of average risk minimization method, using regular data. The scheme of representation of the initial data, which allows obtaining stable relations, connecting physical parameters in space and time is .described. The comparison of the suggested approach relative to the hydro dynamical approach is considered. Perspectives of the applying of the approach in practice are discussed. The results of the of the one-step prediction of the global temperature are presented.

Key words: average risk, approximation, time series, multidimensional fields, process.

INTRODUCTION

Up-to-date global models of weather prediction in most cases are carried out using hydrodynamic approach. This means that for solving the problem of global atmospheric prediction, as well as for constructing of the global monitoring systems, are used hydrodynamic equations with subsequent applying of the finite-difference methods. However, the very original system of partial differential equations represents a certain idealization of the processes occurring in nature. This

idealization takes place whenever such quantities as derivatives, gradients or density obtained as a result of the limiting process are attributed to the physical meaning. To simplify the description of the physical world around us, this idealization within hydrodynamic approach seems quite natural and in fact inevitable [1].

Moreover, the use of differential equations for modeling purposes implies the boundary conditions to be known and given as continuous functions, which in practice of the weather forecasting are not available. Overcoming these obstacles inevitably lids to the costs, which far and away exerts negative influence on the result. This requires a change over switch to discreet data and to finite-difference representation of derivatives of the function, in addition to the development of methods for solving systems of finite-difference equations

In fact, when model of the hydrodynamic prediction is being created, one may observe double transition in the forward and reverse: first we idealize reality, making thus the differential equations, then, starting from these equations, we pass to the discrete case and consider the finite-difference equations, which as a result lead us to the system of algebraic equations. In circumstances where the original information represent the discrete data, such a transition may not arise much optimism. Moreover, it arises the desire to start the construction immediately from algebraic systems, using data obtained directly at the points of observation and to look for the solution itself in the continuous form. Especially as practical solutions in the form of continuous functions are most preferable. The ability to create such a model would allow us to talk about the general empirical approach for modeling, based entirely on observational data. Such an approach would allow us, if experimental base isavailable, to simulate the processes of global scale and predict physical fields without involvement of the partial differential equations. For realization of such an approach one needs only two conditions: the first - a recovery mechanism, which would allow determining the functional relationships using experimental data, the second - the availability of data that meet certain requirements.

RECOVERY OF THE FUNKTIONS USING EXPERIMENTAL DATA

The problem of the recovery of the unknown functions using experimental data is widely presented in literature. We will not stay to long on the problem, assuming that ifmechanism of solving the problem exists and if the data meeting certain requirements are available, the problem may be effectively solved. In practice, in order to restore the function one should carry out the following steps:

a. tobuild up the matrix X (N, n) and the vector Y (N) (lines – situations, columns - parameters). Y- column representing the known values of the unknown function.

b. to build up a functional representing average risk. In case when the mean square error is minimized, it looks as follows:

$$I(\varphi) = \int \big(y - \varphi(x)\big)^2 P(x,y)dxdy \qquad (1)$$

c. to obtain average risk estimation, which is no longer dependent on an unknown probability density P (x, y)

$$J(\tilde{\varphi}) = \sum_{i=1}^{N} \left(\frac{y_i - \tilde{\varphi}(x_i)}{F(x_i)}\right)^2 \qquad (2)$$

d. to find minimum of the average risk estimation (2) for all the samples of the original situations *(X, Y)*, and for all types of functions belonging to a given class, or group of classes. As a result, we will have a functionφ,

$$J(\hat{\varphi}) = \min\ J(\tilde{\varphi}), \qquad (3)$$

which approximates in some of the best ways the unknown relationship.

The up-to-date methods of recovery of the continuous functions formally fit

into the above three steps. Differences may consist in using estimate (2) only, which can be obtained both theoretically [2], as well as by means of experimental data [3.4]. Hereinafter, to confirm the effectiveness of the proposed approach, while minimizing (2), we use the approach based on the data, and the minimization of this estimation (2) is carried out by using exhaustive search algorithm [5], which allows to reach the point of extreme of the (2) within a reasonable time. Limitations in this case can be associated with the size of aggregatesonly, which should includenot more than 12-13 parameters. However, dealing with practical problems, such limitation does not present very strong restriction, while the most of the physical laws are associated, with a small number of variables. Note that exhaustive search involves not only the measured parameters, but also the values of the functions of these parameters. The number of such values, in order to achieve a deeper extremum must exceed many times the number of input parameters.

In practice, however, there is no need to use the wide variety of classes of functions for approximation purposes, as soon as the polynomial with integer powers approximates any continuous function with any desired accuracy. The main requirement is that the power of parameters and coefficients by the monomials provide minimum for the average risk estimation (2). Thus, substituting into the functional (2) a polynomial of the highest possible power

$$\varphi^{(n)} = \sum_{i=1}^{m} \alpha_i \left(x_1^{k_1}, \dots, x_p^{k_p} \right),$$

after optimization we will have some other polynomial, substantially more simple, in which coefficients, $\propto_1, \dots, \propto_m$ represent rational numbers, which can also take zero values. Thus, each factor $x_i^{k_i}$may be part of some monomial $x_i^{k_i}, \dots, x_p^{k_p}$with any integer power$k_i$, satisfying the inequality

$$\left(k_1 + k_2 + \cdots + k_p \leq n\right),$$

where n – is maximal power of the polynomial. The last two points are essential as soon as they express the difference between the considered polynomial approximation, and the ordinary polynomial regression. In fact, presence of any of the original parameter in the monomial with zero power means its absence there and zero values by someof the polynomials means that corresponding monomials are spurious. The optimization process with the help of a polynomial may be regarded as expansion of the function in series of monomials, where order, composition and the very number of monomials are being defined by average risk criteria.

GLOBAL MODELING AND DATA REPRESENTATION

Let us discuss the data problem in greater detail. In which form the data should be presented in order that the restored function, adequately described the process. Firstly, if we represent this data in the form of matrix***A (N, n + 1),*** where the lines - situations and columns - parameters, it should be elongated in the vertical direction. In other words, the number of lines must be at least several times larger than the number of columns. The total size of the matrix must match the complexity of the restored function. If the function to be restored is presumably complicated, the matrix sizes must be increased. However, the main requirement to representation of datais that all the lines of the matrix ***A*** should be obtained randomly and independently according to some unknown, but fixed density distribution ***P (x, y).*** Obviously, when setting targets of atmospheric modeling, the latter condition can be satisfied only with strained interpretation, as soon as the weather situations are coming neither by accident nor independently, but in chronological order. Taking into account that the atmosphere is permanently exposed to exterior influences, the assumption of statistical independence of the initial situations can be done only conditionally.

In general, the problem can be informally presented as follows: we have a global information, measuredat a certain time interval (time series), you need to restore the space-time dependence, which would approximate the processes with allowable accuracy, not only within this time period, but also beyond the bounds of the interval.

But how to summarize all the available global information? How to take into account all the experience of nature accumulated in time series, which are available for modeling at present? Within the initial time series we consider two types of situations: the first type - large situations, consisting of parameters measured at a certain time-point, the second type, - small situations, consisting of a set of measurements in a single geographic point. Obviously, the situation of the first type consist of the situations of the second type, the number of which can be very large. Direct use of the first type situations for the restoration of the unknown function is unacceptable because of the enormous size of the situation, in view of relatively small number of such situations in time series. At the same time, the number of variables and the number of large situations may exceed hundreds of thousands or even more. Matrix A in this case is stretched both in length and in width. Such a representation of initial data must necessarily lead us not only to the necessity of operating with singular matrices, but also to a serious loss of information due to its inefficient use.

Nevertheless, we will try to use all available information to the extent that can be useful for modeling. For this purpose, as the lines of the original matrix $\boldsymbol{A}$, we consider the small situation, representing a set of measurements at the point, and the equation for determining the approximation function

$$T_{t_0+\Delta t} = \psi(Q(V,t), \qquad (t \le t_0)$$

we will write out for all of these points. That instantly solves the problem of the relation of the parameters and situations in the source data. However if we take all situations representing synchronous layers in succession (in chronological order), then the distribution of situations, each of which represents a set of measurements in a single geographic location, may occur to be very complicated, and that will undoubtedly entail the complication of the restored function itself. However, the complexity of the functions obtained in the conditions of measurement errors, which always take place in the experimental data, cannot be regarded as a positive factor.

To simplify the distribution of the situations, representing measurements in a single geographic point, we will use cyclical nature of the atmospheric changes.

Suppose we have a long-term series of the global data measured every hour, and we are going to predict weather one hour ahead. For this purpose, we obviously need to do one-step forward. Let us fix a certain hour, day and month in the time series, and all the other samples of this series remove from it. As a result, the number of remaining situations will be many times smaller than the original number. Obviously, the multivariate distribution of thus obtained series is much simpler, because it has no daily, monthly or seasonal cycles. However, the number of

situations, each of which represents a geographic location is still very high, since even a temporary one synchronous layer may correspond to tens of thousands of such situations. We can therefore expect that basing on the formed in such way situations; one may obtain stable solution, which would allow predicting the weather element one-step ahead.

From the computational point of view, the desired effect is evident. Namely, we have many situations (many experiments), and relatively few parameters. This guarantees us on the one hand the lack of computational difficulties associated with the handling of singular matrices, and at the same time, the possibility to use all the information of the original series. However, using synchronous situations, measured at some regular points in time, contradicts to the classical formulation of the function recovery problem, which requires the situations to be chosen accidently and independently. Indeed, the situations corresponding to a certain date (large situations) connected with each other not only by time, but by space as well, so that corresponding small situation could not be considered as statistically independent. However if we take into account that the requirement of independence is only associated with estimation of the results of modeling, the obstacle can be largely surmounted by using the group method of sliding control.

SLIDING CONTROL

Sliding control (or jack-knife method) which consists in successive exclusions and inclusions of the situations out of the learning sample and calculating errors on this situations as on the independent material, is known a long time ago (see., E.g. [6]). However, for a long time, this procedure could not find a proper application in modeling because of the large amount of computations required for its realization. In fact, the sliding control could only be used for estimation of the result, but not for minimization of the average risk estimations. The compact formula, which allows estimating the average risk in the conditions of the great arrays of digital information presented in [3]. There was also obtained a formula for realization of the group sliding control

$$\Delta_{\text{rp}} = \frac{1}{N}\sqrt{\sum_{j=1}^{m}\sum_{k=1}^{r}(\lambda_k^{(j)})^2},$$

where

$$\lambda^j = \left(I - K_j^T B^{-1} K_j\right)^{-1}\left(Y_j - K_j^T B^{-1} z_j\right), (1 \le j \le m)$$

which can also be used for the estimation of average risk. Here, the internal summation is carried out according to the number of elements in group, external – according to the number of groups, *B*- covariance matrix obtained for all situations, *K*- matrix of the excluded situations *(n× r).*

Originally, the purpose of obtaining this formula was first of all to reduce the computation

time required for minimizing the average risk. Indeed, according to this formula, the inversion of *B* is carried out only once, and subsequentlymatrix *B*is only corrected. Nowadays computer time is not of such a sacramental significance as previously, but the obtained formula gains in this case, a new, deeper meaning. By using sliding group control, when each group consists of synchronous situation only, we will be able to obtain the results of the constructions not only for each individual situation, but for groups of situations, which can be in first approximation considered as statistically independent. Scattering the groups of situations at a certain distance with respect to time, it is possible to achieve a greater degree of group independence, than in the case,when groups in matrix follow one after another in a chronological order. As a result, we have the initial material for functions recovery in the form of independent groups of situations, and this is the basic requirement for statistical estimation of the results of the constructions.

ANALOGY

Screening in such a way the initial series of global observations we can restore the other functions$\psi_i$for one step prediction of the other elementsT^i. As a result, we have the formula

$$T^i_{t_0+\Delta t}(V) = \psi_i(Q(t, t_0 - \Delta t, V)) , \qquad (4)$$

where$i = 1, \dots, l$, andthetime$t \le t_0$may accept discrete values belonging to the interval of the prehistory of the corresponding processes. Having thus prognostic value of $T^1, \dots, T^l$, we have the opportunity to make a new step forward and thus to obtain a new prognostic values. Thus, after obtaining of the next group of the predicted values, we have to carry out again the screening of a of the initial time series elements, and thus, based on a simplified distribution to find new functions for each of the parameters to be predicted.

Note that if in the equations (4) to put $t = t_0$, we will have only three values of a discrete-time, and in this case, the analogy with hydrodynamic patterns can be traced most clearly. Indeed, formula (4) in this case corresponds to the prognostic formulas in numerical hydro dynamical schemes, in which the time derivative is represented in central difference form. In the empirical approach, the initial prehistory interval, which in hydrodynamic schemes is usually fixed, one should specifylong enough, giving the possibility for the average risk criterion to choose the best option. Obviously, the initial prehistory interval must be consistent with the scale of the processes under consideration and, accordingly, with the time step. In real models, this step must be chosen according to the actual conditions of the collection and processing of the information, taking into account the scale of the process under consideration. For the elements of weather to be forecasted for the month, season or year, the time step can be selected large enough, and if the

forecast is considered to be up to ten days, a step should be chosen as low as possible, but not less than the interval of measurements.

A distinctive feature of the hydrodynamic approach, compared with the empirical approach, can be illustrated by a simple example. Assume that the system of hydrodynamic equations reduced to one equation with one unknown variable that under certain assumptions is possible. Then, carrying over the partial derivative of the forecasting parameter with respect to time to the left side of the equation, and to the right side - all the other terms of the equation, from which this derivative may depend, wecan write

$$\frac{\partial T}{\partial t} = A(X) \ ,$$

Where X - avector, a A – the differential operator, which may include convective terms and some other combinations of derivatives and parameters. Presentingthis equation in finite-difference form, we will have

$$\frac{T_{t+\Delta t} - T_t}{\Delta t} = \overline{A}(X),$$

where $\overline{A}$ - finite-difference operator corresponding to differential operator. Then, leaving the predicted variable in the left side and transferring all the rest to the right side, we will have the expression $T_{t+\Delta t} = T_t + \Delta t \overline{A}(X)$,

which allows to make a step forward with respect to time. The principal difference is that in the case of a statistical approach the function ψ, with which the time step is being made, is obtained from the experience, and in the case of a hydrodynamic scheme, the corresponding operator is obtained from theoretical considerations. However, any hydrodynamic scheme can be easily synthesized in the above-described statistical scheme. To do this one must the value of the difference operator at various points of measurements to include as a separate parameter in vector Q. Consequent minimization of the average risk would reveal the contribution of the hydrodynamic model to the general synthesized model, and how much this synthesis is advisable.

MODEL

Fig. 1 represents the general scheme of stepwise forecasting of the fields (1≤i≤t), assuming that the approximating functions $\{\varphi_i^j\}$ have already been obtained. After the initial matrix A and matrix of functions, $\{\varphi_i^j\}$ are introduced, the process of forecasting of the fields is carried out. Consequently ***A*** is converted into a matrix ***A (i),*** which is then used as the starting material for computation of the functions on the *i*-th time step. Further, inside the inner cycle matrix ***A (i)*** is converted into matrix ***A (i, j),*** which further is directly used for prediction of the meteorological

fields

$$y = \varphi\big(A(i,j)\big)$$

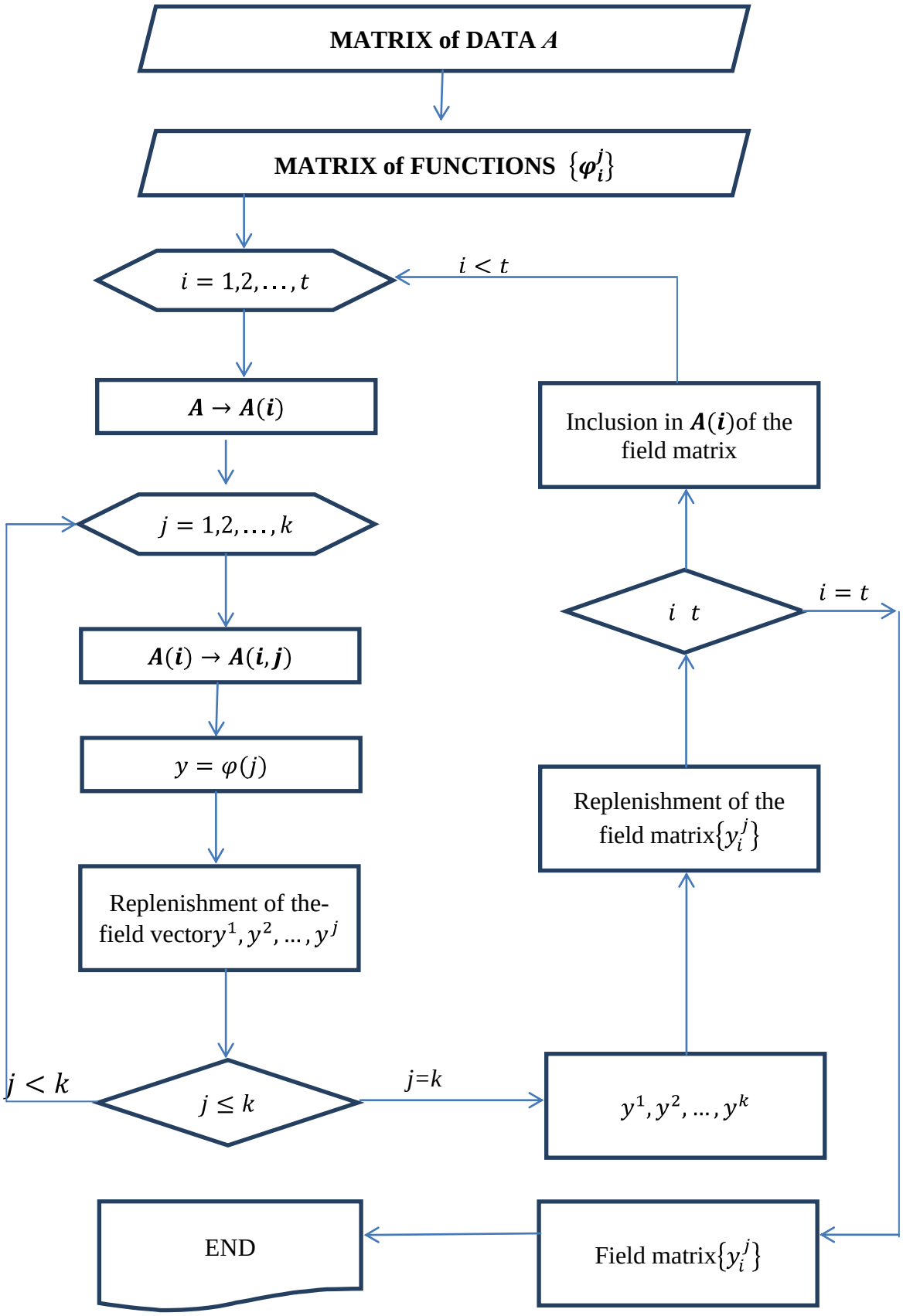

Fig. 1 Prediction of the fields $t$steps ahead

(i-index of the number of steps, j-index of the number of functions)

At the same time the replenishment of the vector fields$y^1, y^2, \ldots, y^j$, arecarried out, and as a result at the final stage of the inside cycle we will have succession$y^1, y^2, \ldots, y^k$, representing the totality of all the fields after the first time step. The cycle of time steps completes the process, and at the last step will have a set of fields corresponding to a given lead-time. At the same time after the next i-th time step, another set of fields is attached to the matrix ***A (i + 1)***to obtain the results on the *(i + 1)* - th step. Thus, in order to make a prediction on the *i*-th time step of the *j*-th element, it is necessary to imagine, conditionally speaking, matrix $A(i, j)$ as an argument of the function φ_i^j.

However, when initial data do not represent direct measurements, but some averaged characteristics of the elements, the multi-step forecasting process is open to question. In this case, it may occur reasonable to make a single time step, covering the lead timeentirely. The same can be said in the case where the original data are not complete and some additional elements for prediction are required. However, the final decision of the question can onlybe obtained through experiments.

Note that the necessity of the detailed time steps may arise mainly by shortrange and very –short range forecasting. This is corroboratednot only by the very design of the hydrodynamic equations, but also by the experience of the hydrodynamic modeling of the atmosphere as a whole. Indeed, the time step in the finite-difference schemes arises from the partial derivatives with respect to time, which implies the rate of change of the function at any given moment. Therefore, any unsound increase of the time step may lead not only to loss of important information, but also toother undesirable effects adversely affecting the result.

EXPERIMENTS

By modeling of the large-scale processes in the atmosphere as the source of information may be used both direct measurements at the stations, and the gridded values of the objective analysis. When modelling is realized using stepwise calculations, both options have its advantages and disadvantages. Of greatest interest is the use of the primary data, as soon as it allows avoiding additional data processing related to objective analysis. However, access to initial information is not simple nowadays, so that using it for modeling purposes is not always reasonable. Unlike the initial data, objective field analysis data are readily available and can be used for the current prediction, and for model construction as well (see., E.g., [7]).

In the present paper for verification of the empirical approach were used the results of objec-

tive analysis, obtained from the website [7]. These data present the values of meteorological elements in grid points NCEP / NCAR with space and time steps making up 2,5 grad and 6 hours, correspondingly. The experiments were carried out using data for July and January to obtain the prediction of the global temperature field one step ahead. Fig. 2 shows curves characterizing the behavior of average risk estimation, according to the dimension of the parameter vector, constructed for one-step temperature forecast in July. In this case, the upper curve is constructed using geographical coordinates of the grid, the lower curve - without their participation. As it seen from the figures, the average risk estimation by predicting the temperature without using coordinates is significantly higher than the corresponding estimate obtained with latitude and longitude involved. That is despite the fact that latitude and longitude, taken as separate parameters are sufficiently informative. This result, initially seeming paradoxical, can be explained quite simply. Latitude and longitude in the points of measurement are involved in the selection process as the two parameters, which must necessarily belong to the final set. However, the structure of the data used for construction allows us, not to use these variables at all, since coordinates of the stations, presenting finite number of points corresponding to the number of grids, is indirectly present in the database, and this order must remain the sameall-time steps. In fact, in this case, the coordinates of the station (or cell location) are replaced by some sequence of number (or symbols) which are not used when the functions are recovered, but which allow to identify any current small situation by the routine forecasting.Therefore, all further experiments to restore the temperature fields were carried out without the participation of latitude and longitude.

Note that in this figure, as in all the following, not the squared error is plotted along the y-axis, but the magnitude, which is proportional to the squared error. However, the efficiency of the approximation, defines, above all, not the verysquared error but its decrease velocity. As additional reference point on the plot may serve an error on the first parameter, which corresponds to the average climate forecasts.

Fig. 3 shows the behavior of the curve by forecasting temperature 6 hours ahead, calculated for January 2, starting from 00 hours. The maximal prehistory interval of the process makes up 18 hours. Thus, by the recovery of functional dependence, were used 8 parameters, of which 4 were surface height at 500 mb, and the rest - temperature at the same level.

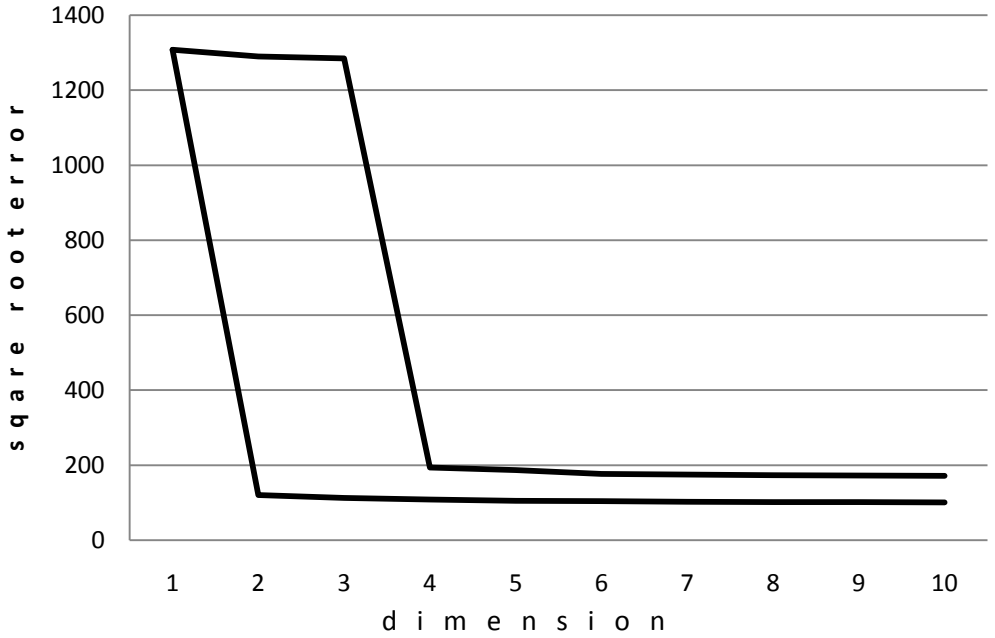

Fig. 2. Curves of average risk estimations obtained using the same material. The upper curve - with coordinates, the lower curve - without.

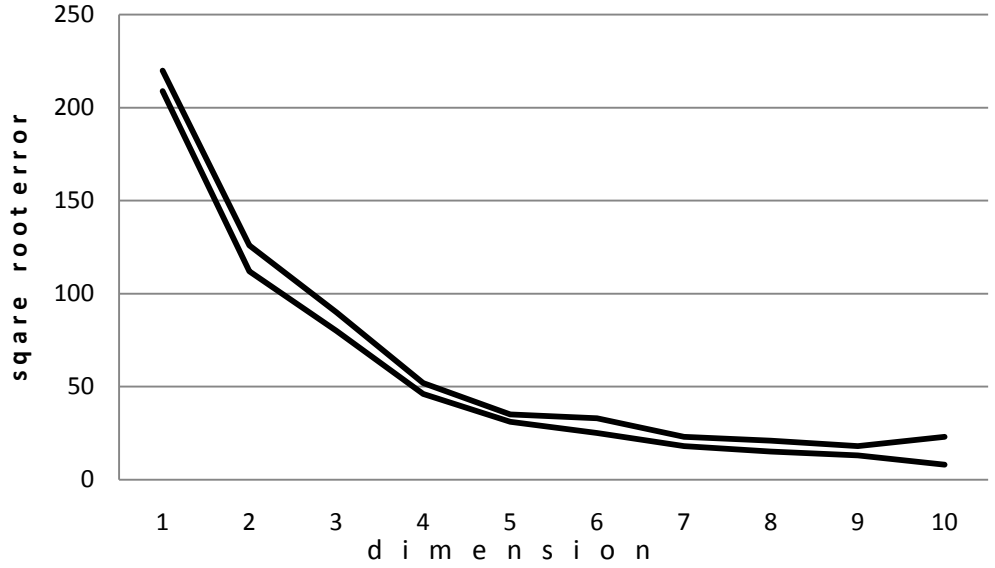

Fig. 3. Curves of error behavior. The lower curve corresponds to mean square error, the upper curve – to the errors of sliding control

Asit is seen from the figure, the curves are located very close to each other, which naturally can be explained by the usage of a large number of situations for constriction under the relatively small number of potential arguments.

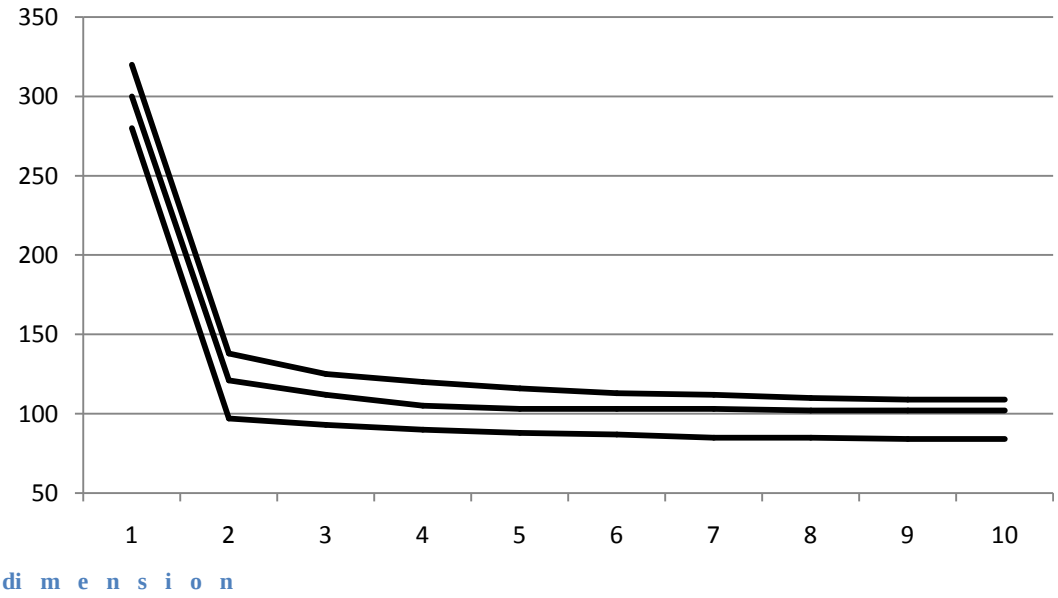

Figure 4. Curves of average risk estimations for sphere (upper curve), hemisphere (middle curve), and southern hemisphere (bottom curve)

Fig. 4 represents the curves obtained by temperature field prediction for the entire sphere, and for the northern and southern hemisphere separately. As it is seen from the figure, the errors in the global case are almost the same as they are for the northern hemisphere. As for the southern hemisphere, the errors are consistently smaller than the errors for northern hemisphere, as well as for the entire area. The reasons for this may be different.One of the reason - a change of seasons at the time of the forecast. In southern hemisphere, in contrast to the northern hemisphere, takes place summer. Moreover, obtained in such a way the prognostic formula does not differ much according to its form from each other's. At the same time, for the whole sphere, we have the formula

$$T(t+\Delta t) = \ \propto_0 + \alpha_1 T(t-3\Delta t) + \alpha_2 T(t-2\Delta t) +$$
$$+\alpha_3 H(t-3\Delta t) + \alpha_4 H(t-2\Delta t) + \alpha_5 T(t),$$

whereasthe formula for hemisphere

$$T(t+\Delta t) = \propto_0 + \alpha_1 T(t-3\Delta t) + \alpha_2 T(t-2\Delta t) +$$
$$+ \alpha_3 H(t-3\Delta t) + \alpha_4 H(t-2\Delta t) + \alpha_5 T(t) + \alpha_6 T^3(t) H(t-2\Delta t),$$

differs from the preceding formula only by the last term. The formula for the southern hemisphere differs from the first formula only by the last two terms

$$T(t+\Delta t) = \propto_0 + \alpha_1 T(t-3\Delta t) + \alpha_2 T(t-2\Delta t) +$$
$$+ \alpha_3 H(t-3\Delta t) + \alpha_4 H(t-2\Delta t) + \alpha_5 T^3(t) + \alpha_6 T^3(t) H(t-\Delta t),$$

which in reality make a negligible contribution as compared to the previous four. This result shows that the chosen distributions of the selected situations are fairly uniform, and the formulas obtained using these distributions are stable enough. It is noteworthy that in all three formulas, the monomials of the first degree are predominating. However, this does not indicative on the simplicity of the modeling processes. The above experiments correspond to one-level model with two input parameters, and therefore, the description of the simulated processes cannot possibly claim to completeness. If you include in the experiment additional parameters, as well as additional levels at which these parameters are measured, the polynomial representation of the fields will match to more adequate description of real processes, and in this case, the predominance of monomials of higher degree would be much more probable.

SUMMARY

In conclusion, we formulate the basic differences and the main advantages of the new approach compared to the traditional methods of global forecasting. Comparisons with other statistical methods of global forecasting is hardly appropriate in this case, taking into account that the global modeling was associated so far with hydrodynamic methods only. As to the hydrodynamic approach, the major advantage consists in the fact, that there is no need to calculate derivatives, which connected with idealization of the reality and the necessity of solving ill-posed problems. Besides, there is no need to operate with the system of difference equations, which is associated with the solution of finite-difference scheme stability problems. The question of the unique existence of the solution in this case is not as

acute as it is by hydrodynamic modeling, since the existence of solutions, as well as uniqueness, is defined by the empirical approach in the process of modeling with the help of the efficient statistical criteria.

The selection of time-step and space-step, by solving system of differential equations, is realized taking into account the stability of the finite-difference schemes thatis not directly connected to the aim of construction. In the above considered construction, the problem of choice of the steps in time and space is not of a current importunes, because this choice is completely dictated by the structure of the available and the incoming hydro meteorological information. In the case of long term forecasts, when averaged characteristics of the elements are predicted, the time step can be significantly increased, but in all cases it must be a multiple of the interval of measurements

The absents of the additional steps in the constructions, which, as a rule, do not fit into a single controlled computing process, such as adjustment of the fields, or objective analysis, represents the undoubted advantage of this approach compared with hydrodynamic methods. However the main advantages of the approach is that the whole process of the model construction is controlled by a single criterion (average risk criterion), and this criterion is directly connected to the ultimate goal of construction. All this testifies to the prospects of successful development and application of the empirical approach for global modeling.

LITERATURE

1. Курант Р., Гильберт Д., Методы математической физики, т. 1, Гостехиздат, М.-Л., 1951
2. Vapnik.V (1998)Statistical Learning Theory, John Wiley, 1998, NY, p.732.
3. Поляков Г.Г., Романов Л.Н. Скользящий контроль и линейная регрессия. Метеорология и Гидрология, 1988, N 9.
4. Романов Л.Н. Минимизация риска и восстановление пропусков в атмосферных данных. Сиб. журн. вычисл. математики, РАН. Сиб. отд-ние, 2009, Т. 12, № 2.
5. Поляков Г.Г., Романов Л.Н. Об аппроксимации зависимостей с помощью линейных функций. Труды ЗапСибНИГМИ, вып. 83, 1988.
6. Quenouille M.H. Approximate tests of correlation in timeseries. J.R. Statist. Soc., 1949, B 11, p. 68-84
7. www.tsrl.noa.gov/psd/data/composites/hour,