STUDY OF THE POSSIBILITY OF USING THE AUTOREGRESSION MODEL FOR THE TEMPERATURE FORECAST

V.I. Akselevich

Russian State Hydrometeorological University, St. Petersburg
Received April 28, 1998

The results of temperature forecast at various altitude levels for different versions of algorithm operation are considered and compared with the inertial and climatic forecasts. Optimum conditions for algorithm operation are established. Different set of initial data required for forecast and possibility of obtaining the information about the vertical profiles of the temperature in various situations are discussed.

The forecast of the vertical temperature distribution is of significant importance, because the information obtained can be used for solving some applied problems. Among them are:

– support for the reliable operation and maintenance of preset power of aviation and rocket engines;

– study of the properties and characteristics of the electromagnetic radiation of various nature;

– determination of the temperature stratification and stability classes of the atmosphere necessary for the successful forecast of weather and environmental pollution;

– climatic and ecological monitoring of the environment;

– interpretation of satellite information;

– forecast of icing, rough air, condensation trails, and positive deviation of the air temperature from its standard value;

– calculation of lidar paths and optimum conditions for reliable communications;

– calculation of flight trajectories for various bodies;

– provision for efficient operation and reliability of balloons.

As a rule, three methodical approaches to the air temperature forecast are distinguished. The first one is to reconstruct the temperature from the data on geopotential obtained by a numerical forecast.[1] The second one is to integrate the heat transfer equation.[2] The third one is to use the methods of statistical interpolation from other prognostic fields.[3,4]

On the basis of these approaches and their various combinations for the forecast of the vertical temperature profiles, several methods are used that differ in fields of their application and amount of necessary initial information.

The method of trajectories is recommended in Ref. 5 for the temperature forecast. It is most suitable for the available standard set of initial data, officially approved by the Russian State Committee on Meteorology and sufficiently reliable. At the same time, there exist a lot of problems for which this traditional approach is incorrect. For instance, if there are omissions in observation data and they must be interpolated or extrapolated, the problem of making up for a deficiency in information often arises. The problem of reconstruction and forecast of the vertical temperature profiles in the middle and upper stratosphere by the physical-statistical methods is acute. The cases of losses or impossibility to obtain meteorological information due to various reasons also should be taken into account. Therefore, together with the traditional approaches to the forecast of the vertical temperature profiles, alternative approaches are developed based on application of the methods of analogy, clustering arguments, standardization of vertical profiles, and autoregression schemes and equations.

Let us consider in detail the autoregression model for temperature forecast at a given reference altitude level. Physically, such an approach is justified by the cyclic change of atmospheric states, presence of wave motion, and the law of energy conservation.

It is expedient to consider the temperature variation as a stationary process and describe it by the well-known autoregression model

$$E\{y[t]\} = \sum_{\tau=1}^{S(0)} QE\{y[t-\tau]\},$$

where $y[\mathrm{t}]$ are the characteristics of the examined process at discrete times $t = 1, 2, \ldots, n$; $y[t] = E\{y[t]\} + \varepsilon[t]$ are the random measurement errors with zero mathematical expectation and finite variance; Q are the parameters of the model; $S(0)$ is the complexity or the order of the model.

Estimation quality must be connected with that of forecasting. Since the forecast error is directly proportional to the errors of parameter estimation, we should to estimate Q as close to its real values as possible. It is expedient to estimate the autoregression parameters by minimization of their deviations from the

0235-6880/98/08 703-03 $02.00

autoregression hyperplane simultaneously along all the directions specified by $y[t]$ and $y[t - \tau]$. Together with estimation of the parameters m, it is necessary to solve the problem of selection of the optimum structure of the autoregression model of the order $S(*)$. It can be reduced to estimation of the parameters $Q(S)$ for different S and selection of $S(*)$ for which the forecast error is minimum.

In practice, the model should be adapted to real conditions and its sensitivity to variation of each parameter should be studied.

The autoregression model permits variation of the length of the initial observation series, the number of forecast steps, the maximum autoregression order, the estimation criterion (the rms forecast error (1) for the check sample, the final forecast error (2), and the rms forecast error (3) from the cross sample), the number of saved models, the running number of the model used for forecast, and the length of the check sample in the case of choosing the first estimation criterion.

To evaluate the efficiency of algorithm operation, all these parameters were varied in different combinations to obtain the best results of forecasting for periods from 1 to 6 terms of future observations. The number of initial elements was varied from 10 to 79; the maximum autoregression order was varied from 1 to 19; the number of saved best models varied from 1 to 29. Because the first estimation criterion requires a given length of the check sample, the variation of this parameter was studied. It also has an effect on the result of forecast. In addition, the experiment was conducted to choose the running model number, which varied from 1 to 29. The temperature values themselves and their deviations from the norm were used as initial data. By the norm we mean the temperature at the observation site averaged over a given period (it is known from climatic reference books). From the calculations, the following conclusions were drawn:

1. The result does not depend on the number of saved models. The minimum error of numerical experiment was obtained for the fourth model. Therefore, it is necessary to save at least four models.

2. The maximum possible number of models depends on the autoregression order. If the order is equal to 1, 2, 3, and 4, this number equals to 1, 3, 7, and 15, respectively, and so on.

3. Efficiency of the forecast depends on the running number of the model rather than on the autoregression order. The fifth number proved to be optimum from the viewpoint of the result.

4. As the number of the initial observations changed, the best results varied as functions of the forecast step. In forecasting for one or two steps, the best results were obtained with 30 elements in the initial sample; in forecasting for 3–4, 5, and 6 steps, the best results were obtained with 25...30, 40, and 20 elements, respectively.

5. The second criterion, i.e., the final error of the forecast, is most convenient for current estimation of forecast quality.

Thus, the model is most sensitive to the choice of the running model number and the autoregression order. The number of initial observations and replacement of absolute values of the parameters by their deviations from norms have a somewhat weaker effects on the efficiency of forecast.

After determining the optimal parameters for periods of 12, 24, 36, 48, 60, and 72 hours (ignoring the diurnal behavior) and 1–6 days (with allowance for the diurnal behavior), the temperature and its deviations from the climatic norms were determined at the surface level and at barometer altitudes of 700, 500, and 300 hPa at the station Keflavik for winter and summer from the 10-year sample. Table I presents the results of the temperature forecast (the mean absolute error of the forecast and the rms forecast error) for the autoregression model for July (from the ten-year sample) for a period of 12 h for the station Keflavik. The forecast results can be estimated from Table II presenting the relative rms errors of temperature forecast that were obtained by comparing the rms forecast error for the autoregression model with the climatic standard deviation[6] and the rms error of the inertial forecast.

TABLE I. Results of temperature forecast at the surface level and barometer altitudes of 700, 500, *and* 300 hPa *(mean absolute error/rms error,* °C*).*

Level	Forecast period, h					
	12	24	36	48	60	72
Surface	1.4/0.9	1.5/1.0	1.9/1.1	1.8/1.0	1.9/1.2	1.9/1.3
700 hPa	1.7/1.3	2.2/1.8	2.4/1.8	2.8/2.4	3.1/2.4	3.4/2.7
500 hPa	3.0/2.0	3.1/2.4	2.8/1.8	2.7/1.9	3.2/2.7	3.5/3.1
300 hPa	2.4/1.9	3.1/1.9	3.4/2.0	3.4/2.0	3.5/2.3	3.6/2.4

TABLE II. Relative rms errors in temperature forecast at the surface level and barometer altitudes of 700, 500, *and* 300 hPa *(the ratio of the model error to the climatic standard deviation/ the ratio of the model error to the rms error of the inertial forecast,* %).

Level	Forecast period, h					
	12	24	36	48	60	72
Surface	39/82	43/111	48/100	43/125	52/120	57/217
700 hPa	36/108	50/120	50/95	67/120	67/126	75/142
500 hPa	50/87	60/92	45/58	48/56	68/75	78/103
300 hPa	59/95	59/86	63/65	63/69	72/121	75/100

Analysis of Table I demonstrates that the least mean values of the absolute forecast error at all the levels, except 500 hPa, correspond to a forecast period of 12 h, whereas the largest ones correspond to 72 h. Moreover, at the surface level and a level of 500 hPa, monotonic increase of the error is not observed. If the values of errors are compared at different levels, the least values correspond to the surface level, and the largest ones, except for a forecast period of 12 h, correspond to a level of 300 hPa. Similar results are also characteristic of the rms forecast error except for a level of 500 hPa.

Analysis of the data presented above makes it possible to conclude that the forecast by the autoregression method is much better than the forecast by the climatic reference book. At the same time, the autoregression model, being compared with the inertial forecast from the same initial data, has an advantage only at levels of 500 and 300 hPa. The best advantage is obtained for forecast periods of 36–48 h. It is evident that the model should be refined. Deviations yield worse results than the temperatures themselves. Main hopes for the improvement of forecast quality for the autoregression model are connected with changes in the form of the initial data. The coefficients of expanding of the parameters into their natural orthogonal components (NOC) proposed in Refs. 7 and 8 are supposed to be used as the initial data.

REFERENCES

1. P.N. Belov, E.P. Borisenkov, and B.D. Panin, *Numerical Methods of Weather Forecast* (Gidrometeoizdat, Leningrad, 1989), 375 pp.
2. L.T. Matveev, *General Meteorology* (Gidrometeoizdat, Leningrad, 1984), 751 pp.
3. L.S. Gandin and R.L. Kagan, *Statistical Methods of Interpolation of Meteorological Data* (Gidrometeoizdat, Leningrad, 1976), 359 pp.
4. V.V. Bykov and G.P. Kurbatkin, Izv. Akad. Nauk SSSR, Geofiz., No. 2, 307–318 (1961).
5. K.G. Abramovich and A.A. Vasil'eva, eds., *Manual of Forecast of Meteorological Conditions for Aviation* (Gidrometeoizdat, Leningrad, 1985), 301 pp.
6. O.B. Mertsalov, ed., *Reference Book on Vertical Correlation Relationships of the Temperature, Geopotential, and Density in the Free Atmosphere over the Northern Hemisphere* (NIIAK, Moscow, 1966), 369 pp.
7. V.S. Komarov and V.A. Remenson, Opt. Atm. **1**, No. 7, 3–16 (1988).
8. D.K. Faddeev and V.N. Faddeeva, *Numerical Methods of Linear Algebra* (Fizmatgiz, Moscow, 1963), 734 pp.