Estimation and prediction of atmospheric state parameters using the Kalman filtering algorithm.
Part 2. Results of research

V.S. Komarov and Yu.B. Popov

Institute of Atmospheric Optics,
Siberian Branch of the Russian Academy of Sciences, Tomsk

Received September 21, 2000

The results of experimental research into the quality of spatial prediction of the atmosphere state parameters (temperature and wind) by the Kalman filtering method are discussed. The long-term data obtained at a typical mesometeorological polygon were used.

It can be *a priori* thought that spatial prediction of mesometeorological fields based on the Kalman filtering method[1] is a great improvement over the same prediction made by one of alternative extrapolation methods (polynomial approximation or optimal extrapolation), because it simultaneously accounts for both the peculiarities of some or other field at individual atmospheric levels and its time dynamics. However, to justify this assumption, a specialized research is needed.

The aim of this paper is to confirm experimentally the efficiency of the synthesized algorithms[1] and to estimate them qualitatively.

In the first turn, we dwell on characteristics of the initial material and methodic aspects of statistical estimation of the quality of spatial prediction.

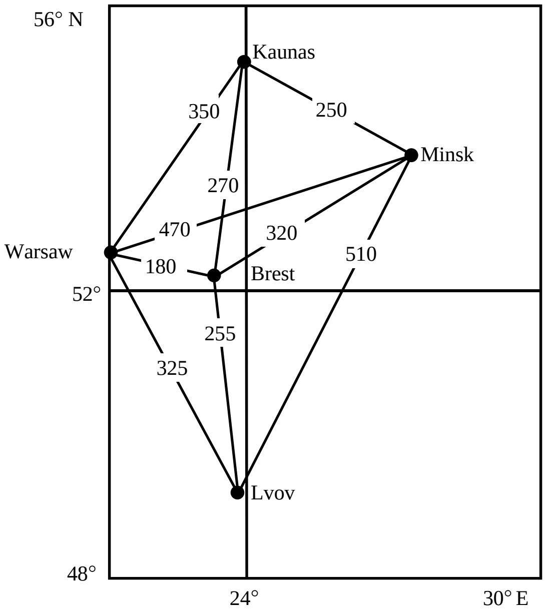

Fig. 1. Mesometeorological polygon (distances are given in km).

For spatial prediction, based on the Kalman filtering method, and estimation of its quality, we used the data of long-term (1961–1975) balloon observations at five upper-air stations: Warsaw (52°11′N, 20°58′E), Kaunas (54°53′N, 23°53′E), Brest (52°07′N, 23°41′E), Minsk (53°11′N, 27°32′E), and Lvov (49°49′N, 23°57′E), which form a typical mesometeorological polygon (see the figure).

For experimental research, we selected synchronous (for all stations) data (at 00:00 and 12:00 G.M.T.). The selected data were reduced (using linear interpolation and taking into account singular points) to a single system of geometric heights including the following levels: 0 (ground layer), 200, 400, 800, 1200, 1600, 2000, 2400, 3000, 4000, 5000, 6000, and 8000 m. This system allows us to describe almost the entire troposphere (including the boundary layer) with high vertical resolution.

The use of the listed levels is caused by requirements of numerical prediction of technogenic pollutants spatial spread in the troposphere (with their maximum transport within the atmospheric boundary layer[2]). The procedure of spatial extrapolation of mesometeorological fields (in our case, fields of temperature and orthogonal components of wind velocity) using the Kalman filtering method was realized just for such a prediction.

It is well known[3] that to calculate in practice the pollution spread (first of all, of technogenic origin), the layer-averaged values of temperature and wind (these meteorological parameters play an important part in transport of pollutants), are commonly used rather than their values at different layers. Therefore, for formation of initial data arrays used in spatial prediction of mesometeorological fields, we apply, as in Ref. 4, the procedure of layer-by-layer averaging of the temperature T, as well as zonal V_x and meridional V_y components of wind velocity. The averaging is performed using the following equations:

$$<T>_{h_0,h} = \frac{1}{h - h_0} \int\limits_{h_0}^{h} T(z)\, \mathrm{d}z \;; \tag{1}$$

$$<V_x>_{h_0,h} = \frac{1}{h - h_0} \int\limits_{h_0}^{h} V_x(z)\, \mathrm{d}z \;, \tag{2}$$

0235-6880/01/04 235-04 $02.00

$$<V_y>_{h_0,h} = \frac{1}{h - h_0} \int\limits_{h_0}^{h} V_y(z)\, dz\ . \qquad (3)$$

In Eqs. (1)–(3), the symbol <•> denotes averaging over the vertical in some atmospheric layer $h - h_0$ (here h_0 and h are bottom and top heights of this layer) with the height h_0 corresponding to the ground layer and, therefore, taken equal to zero.

The following atmospheric layers were used in the averaging: 0–200, 0–400, 0–800, 0–1200, 0–1600, 0–2000, 0–2400, 0–3000, 0–4000, 0–5000, 0–6000, and 0–8000 m.

The quality of spatial prediction of layer-average (or simply average) values of temperature $<T>_{h_0,h}$ and zonal $<V_x>_{h_0,h}$ and meridional $<V_y>_{h_0,h}$ wind with the Kalman filtering method was estimated using a standard (root-mean-square) error of this prediction δ_ξ, as well as the probability of prediction errors $P = P(\Delta\xi < X)$ determined as

$$\delta_\xi = \left[\frac{1}{n} \sum_{i=1}^{n} (\hat{\xi}_i - \xi_i)^2 \right]^{1/2}, \qquad (4)$$

where $\hat{\xi}_i$ and ξ_i are the predicted and measured values of a meteorological parameter, respectively, and n is the number of realizations. The prediction error here is $\Delta\xi_i = |\hat{\xi}_i - \xi|$, and X is some preset threshold.

To estimate the quality of a chosen method of spatial prediction, we used one of the most widespread methods. According to Ref. 5, it is based on comparison of the value of a meteorological parameter predicted (extrapolated) for some control station from the data of neighboring stations with actual values measured at the control station at the corresponding time. Note that as control stations we used the stations Warsaw and Lvov situated 180 and 255 km far from the nearest station Brest having actual data.

To estimate the quality of spatial prediction by the Kalman filtering method, we also compared its results with the results given by other methods. For this purpose, we invoked the estimates of accuracy of two alternative methods: the modified method of clustering of arguments (MMCA) and the widespread method of optimal extrapolation. These estimates were borrowed from our paper[4] (for the station Warsaw) and newly obtained (for the station Lvov).

Now we will discuss the results of studying the quality of spatial prediction based on the Kalman filtering method with the use of mesoscale fields of temperature and orthogonal components of wind velocity. For this purpose, we consider Tables 1 and 2 that give the values of the standard errors δ_ξ of spatial prediction for mean temperature $<T>_{h_0,h}$ and zonal $<V_x>_{h_0,h}$ and meridional $<V_y>_{h_0,h}$ wind, as well as the probability of prediction errors P.

Note that Tables 1 and 2 present, as an example, the results of statistical estimation only for the summer season, when the spatial correlation in midlatitudes of the Northern Hemisphere is much weaker than in winter.[6]

The analysis of Tables 1 and 2 shows that

– first, the algorithm of spatial prediction of mesometeorological fields based on the Kalman filtering method and applied to numerical prediction of layer-average values of temperature and zonal and meridional wind velocity components gives rather acceptable results, especially, in extrapolation to a distance of 180 km. Actually, at this distance, regardless of the chosen atmospheric layer, the probability of errors P (for example, for $\Delta T \le 2$°C and $\Delta V \le 2$ m/s) is 0.71–0.93 (for the layer-average values of temperature) and 0.66–0.71 (for components of the layer-average wind velocity);

– second, this algorithm gives the best prediction results (to a distance of 180 km) for the layer-average temperature, for which the standard errors vary within 1.0–1.5°C regardless of the atmospheric layer, and at the height above 3 km they are at the level of the allowable error of 1°C established for the troposphere by the World Meteorological Organization[7];

– third, the quality of spatial prediction of the parameters $<T>_{h_0,h}$, $<V_x>_{h_0,h}$, and $<V_y>_{h_0,h}$, as expected, decreases markedly, as the distance increases from 180 to 255 km. The results of prediction of layer-average temperature worsen mostly (1.6–2.0 times) with the increase of the distance. This may be attributed to poor selection of the dynamic model playing an important part in the Kalman filtering algorithm.

Consider the results of qualitative estimation of the Kalman filtering method in comparison with two others, namely, the traditional method of optimal extrapolation that is widely used in objective analysis of meteorological fields[5,8–10] and the nontraditional method based on the use of the MMCA algorithm.[11–14] For this purpose, we refer to Table 3 giving the standard errors δ of spatial prediction of layer-average temperature and zonal and meridional wind to a distance of 180 km in the summer season. The prediction was made by three alternative methods: method of optimal extrapolation, method based on the MMCA algorithm, and Kalman filtering method. The data on statistical estimation of the first two methods are borrowed from Ref. 4.

From analysis of Table 3 it follows that

– the method of the optimal Kalman filtering gives the best results, except for the cases of prediction of mean temperature in the atmospheric layers of 0–200, 0–400, and 0–800 m, in which the best results are given by the method based on MMCA. Actually, the Kalman filtering method is characterized by the smallest standard errors of prediction within the entire considered tropospheric layer (up to the height of 8 km). These errors are equal to 1.0–1.5° (for layer-average temperature) and 1.5–2.4 m/s (for layer-average velocities of the zonal and meridional wind);

– the Kalman filtering method gives the most gain in accuracy in the spatial prediction of layer-average temperature at the height $h \ge 4$ km. In these cases, the rms error δ is two and more times smaller than the same error given by the MMCA or optimal extrapolation methods, and the latter gives the worst results.

Table 1. Standard errors δ_T and probabilities *P* of errors of spatial prediction of layer-average temperature by the Kalman filtering method to a distance of 180 (1st column) and 255 km (2nd column)

Layer, m	Probability, $P \times 10^2$										δT, °C	
	$\Delta T \le 1$°C		$\Delta T \le 2$°C		$\Delta T \le 3$°C		$\Delta T \le 4$°C		$\Delta T > 4$°C			
	1	2	1	2	1	2	1	2	1	2	1	2
0–200	38	24	71	50	87	65	94	80	06	20	1.5	2.6
0–400	38	24	71	50	88	65	94	80	06	20	1.5	2.6
0–800	39	24	71	49	88	65	94	79	06	21	1.5	2.6
0–1200	40	24	72	49	90	65	97	79	03	21	1.4	2.6
0–1600	41	24	73	49	92	66	98	79	02	21	1.4	2.6
0–2000	42	25	74	49	93	67	99	79	01	21	1.4	2.6
0–2400	44	26	76	50	94	68	99	80	01	20	1.3	2.5
0–3000	51	27	81	51	96	69	99	81	01	19	1.2	2.4
0–4000	56	28	87	52	98	70	99	83	01	17	1.1	2.3
0–5000	63	29	90	56	98	71	99	87	01	13	1.1	2.2
0–6000	64	30	92	58	98	73	99	89	01	11	1.0	2.1
0–8000	68	32	93	60	99	76	99	91	01	09	1.0	2.0

Table 2. Standard errors δ_V and probabilities *P* of errors of spatial prediction of layer-average zonal and meridional wind by the Kalman filtering method to a distance of 180 (1st column) and 255 km (2nd column)

Layer, m	Probability, $P \times 10^2$										δV, m/s	
	$\Delta V \le 1$ m/s		$\Delta V \le 2$ m/s		$\Delta V \le 3$ m/s		$\Delta V \le 4$ m/s		$\Delta V > 4$ m/s			
	1	2	1	2	1	2	1	2	1	2	1	2
	Zonal wind V_x											
0–200	44	41	68	67	87	86	94	93	06	07	1.8	2.0
0–400	43	40	62	61	86	83	93	92	07	08	1.9	2.1
0–800	39	36	60	59	84	78	93	90	07	10	2.0	2.3
0–1200	36	33	61	58	83	78	92	88	08	12	2.1	2.4
0–1600	34	31	63	57	82	76	92	86	08	14	2.1	2.5
0–2000	36	29	63	57	82	74	92	85	08	15	2.1	2.5
0–2400	37	28	63	57	82	73	92	85	08	15	2.1	2.5
0–3000	38	26	65	53	84	72	93	85	07	15	1.9	2.5
0–4000	39	23	71	50	87	71	94	83	06	17	1.6	2.6
0–5000	42	22	71	45	88	70	95	82	05	18	1.5	2.7
0–6000	40	22	69	42	86	68	95	81	05	19	1.5	2.8
0–8000	38	22	66	41	85	63	94	79	06	21	1.6	3.0
	Meridional wind V_y											
0–200	45	40	69	65	87	85	94	93	06	07	1.6	1.9
0–400	46	39	69	64	81	80	91	90	09	10	1.7	2.1
0–800	42	39	68	64	81	79	91	89	09	11	1.7	2.3
0–1200	39	37	68	63	81	79	91	88	09	12	1.7	2.5
0–1600	39	37	68	61	82	77	91	87	09	13	1.7	2.6
0–2000	39	35	68	61	83	76	91	86	09	14	1.8	2.7
0–2400	41	34	68	60	84	75	92	86	09	14	1.8	2.8
0–3000	40	32	68	60	83	74	91	86	09	14	1.9	2.8
0–4000	39	26	68	58	83	73	91	86	09	14	2.0	2.9
0–5000	38	25	68	52	82	69	91	84	09	16	2.1	3.1
0–6000	38	23	67	43	81	66	90	81	10	19	2.3	3.5
0–8000	34	21	66	40	78	55	89	78	11	22	2.4	3.9

Table 3. Standard errors δ of spatial prediction of layer-average temperature and zonal and meridional wind by the method of optimal extrapolation (1st column), method based on MMCA (2nd column), and Kalman filtering method (3rd column) to a distance of 180 km

Layer, m	Temperature, °C			Zonal wind, m/s			Meridional wind, m/s		
	1	2	3	1	2	3	1	2	3
0–200	1.6	1.2	1.5	2.8	1.8	1.8	3.0	2.2	1.6
0–400	1.6	1.3	1.5	2.8	2.0	1.9	3.1	2.3	1.7
0–800	1.6	1.3	1.5	2.7	2.1	2.0	3.1	2.3	1.7
0–1200	1.7	1.5	1.4	2.7	2.2	2.1	3.1	2.4	1.7
0–1600	1.9	1.6	1.4	2.7	2.2	2.1	3.0	2.4	1.7
0–2000	2.0	1.7	1,4	2.6	2.2	2.1	2.9	2.4	1.8
0–2400	2.2	1.8	1.3	2.6	2.2	2.1	2.9	2.4	1.8
0–3000	2.5	2.0	1.2	2.6	2.2	1.9	2.9	2.4	1.9
0–4000	2.7	2.2	1.1	2.6	2.2	1.6	2.8	2.4	2.0
0–5000	2.8	2.3	1.1	2.6	2.2	1.5	2.8	2.4	2.1
0–6000	3.0	2.5	1.0	2.6	2.2	1.5	2.8	2.4	2.3
0–8000	3.2	2.6	1.0	2.5	2.2	1.6	2.8	2.4	2.4

Thus, the conducted experiments on estimating the quality of spatial prediction of mesometeorological fields (in particular, temperature and wind fields) have shown that the Kalman filtering method is efficient enough and exceeds in the accuracy both the method based on MMCA and, especially, the method of optimal extrapolation.

Consequently, this method can be successfully used in the procedure of objective analysis of such fields. The results of analysis form the basis for solution of the problem of diagnostics and prediction of atmospheric pollution processes on the local and regional scales.

In conclusion, it should be emphasized that the considered method can be significantly improved. For this purpose, it should be revised in the following parts:

1) possible use of adaptive structures of the Kalman filter (KF) allowing:

– parallel inclusion of several KF's initiated with different initial values of unknown intervals of space and time correlation with the following choice of the best filter;

– use of a specialized KF providing the real time determination of correlation intervals with following application of the obtained values to the filters for prediction of meteorological parameters;

2) application of more complex analytical models describing the dynamics of the meteorological parameters behavior in space and time and offering more adequate description of physical processes;

3) use of measuring stations giving higher frequency and higher accuracy of measurements;

4) optimization of the number of measuring stations from the point of view of their information content.

The listed problems call for specialized development and will be the subject of our further research.

References

1. V.S. Komarov and Yu.B. Popov, Atmos. Oceanic Opt. **14**, No. 4, 230–234 (2001).
2. A.M. Vladimirov, Yu.I. Lyakhin, L.T. Matveev, and V.G. Orlov, *Environmental Protection* (Gidrometeoizdat, Leningrad, 1991), 423 pp.
3. F.F. Bryukhan', *Methods of Climatic Processing and Analysis of Aerological Information* (Gidrometeoizdat, Moscow, 1983), 112 pp.
4. V.S. Komarov, A.V. Kreminskii, and Yu.B. Popov, Atmos. Oceanic Opt. **11**, No. 8, 692–702 (1998).
5. L.S. Gandin and R.L. Kagan, *Statistical Methods of Interpretation of Meteorological Data* (Gidrometeoizdat, Leningrad, 1976), 359 pp.
6. L.S. Gandin V.I. Zakhariev, and R. Tselnai, eds., *Statistical Structure of Meteorological Fields* (Budapest, 1976), 365 pp.
7. Agenda of the World Meteorological Organization (Geneva, 1959), Appendix No. 2 (Geneva, 1963).
8. P.N. Belov, *Practical Methods of Numerical Weather Forecast* (Gidrometeoizdat, Leningrad, 1967), 335 pp.
9. V.A. Gordin, *Mathematical Problems of Hydrodynamic Weather Forecast* (Gidrometeoizdat, Leningrad, 1987), 264 pp.
10. V.F. Kramar and T.V. Menzhulina, Meteorol. Gidrol., No. 12, 20–28 (1991).
11. V.S. Komarov and A.V. Kreminskii, Atmos. Oceanic Opt. **8**, No. 7, 488–500 (1995).
12. V.C. Komarov, *Statistics in Application to Problems of Applied Meteorology* (Spektr, Tomsk, 1997), 256 pp.
13. V.S. Komarov, A.V. Kreminskii, and Yu.B. Popov, Meteorol. Gidrol., No. 8, 37–45 (1999).
14. V.S. Komarov, A.V. Kreminskii, N.Ya. Lomakina, Yu.B. Popov, and K.Ya. Sineva, in: *Proceedings of Russian Scientific Conference on Climate, Environmental Monitoring, Hydrometeorological Forecast and Service* (UNIPRESS, Kazan, 2000), pp. 214–216.