APPLYING THE MODIFIED METHOD OF CLUSTERING OF ARGUMENTS (MMCA) TO FORECASTING OF PARAMETERS OF MEAN WIND

V.S. Komarov, V.I. Akselevich, and A.V. Kreminskii

Institute of Atmospheric Optics, Siberian Branch of the Russian Academy of Sciences, Tomsk
Russian State Hydrometeorological Institute, St. Petersburg
Received November 15, 1993

Perspectives are discussed of applying the modified version of the method of clustering of arguments to the problem in forecasting (retrieving) of the zonal and meridional components of the vector of mean wind in both the boundary layer and free atmosphere (up to 8 km*) for operative estimates of transport of atmospheric pollution. Actual examples are used to show high efficiency of such a technique and its applicability to the problem in developing an automated system for ecological monitoring of the atmosphere over limited areas.*

Among the problems faced now in ecological monitoring of the atmosphere over limited areas the problem in forecasting (retrieving) of the vertically averaged wind (further called mean wind for brevity reasons) is an important one. This is so, because the spatial transport of substances of technogenic origin is first of all controlled by the field of wind velocities.[1,2] Indeed, analysis of the data on pollution transport in the atmosphere indicates that with the steady–state rate of sedimentation, every particle will horizontally move as affected by wind at various heights. Therefore, following Ref. 1, the vector of horizontal displacement **s** of a particle from its source of emission to the point of sedimentation to the ground is proportional to the integral of the vector of wind velocity over the vertical, i.e.,

$$\mathbf{s} \sim \frac{1}{h}\int\limits_0^h \mathbf{V}(z)\,\mathrm{d}z\,, \tag{1}$$

where h is the level of pollution above ground.

In practice, to calculate the propagation of a cloud of pollutants the vector is usually introduced of wind averaged over separate vertical layers, $h - h_0$, otherwise referred to as the vector of mean wind,

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

The characteristics of such wind are given by its zonal v_x and meridional v_y components, which mean the following:

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

$$<v_y>_{h_0,\,h} = \frac{1}{h - h_0}\int\limits_{h_0}^h v_y(z)\,\mathrm{d}z\,; \tag{3}$$

The symbol $<\cdot>_{h_0,\,h}$ introduced into expressions (2) and (3) denotes the procedure of averaging over the vertical in the $h - h_0$ layer. In practice these are most often read starting from the earth surface, which is $h_0 = 0$.

We took into account all the above–discussed in designing an algorithm for calculation of the characteristics of mean wind, which was then applied to the task of 12–hours forecasting of the zonal and meridional components of the vector of mean wind by the modified method of clustering of arguments (MMCA).[3]

One should immediately note that this technique, coming from the group of nontraditional physical statistical techniques, is quite simple and does not demand any large number of initial experimental data or heavy computer time expenditures. This technique does not require any preliminary statistical averaging of long–term series of meteorological observations and makes it possible to synthesize a prognostic model using the *a priori* information, while our knowledge of the structure of the modeled process and of the properties of the noise, which presents in the initial data, remains almost indefinite.

We demonstrated all the above–discussed also with respect to statistical estimating and forecasting of the characteristics of free atmosphere (including the vertical temperature profiles, the zonal and meridional components of wind speed), and their respective results may be found in Ref. 4. However, in this work the initial information was based on data from radiosounding at the basic constant level surfaces only, that was the site level, 850 (~ 1.5 km), 700 (~ 3 km), 500 (~ 5.5 km), 400 (~ 7 km), and 300 GPa (~ 9 km), which prevented estimating the efficiency of MMCA when applied to the problem in forecasting of the characteristics of mean wind at a fine vertical resolution (particularly in the boundary layer) that is needed to calculate the spatial transport of pollutants from their emission source.

This problem is solved here in using, by way of example, the long–term (1966–1970) series from three typical aerological stations: Keflavic (63°57′N, 22°37°W), Rome (41°48′N, 12°38′E), and Miami (25°49′N, 80°17′W), which represent different geographical regions of the northern hemisphere. The initial data were preliminarily formed into vertical profiles v_x and v_y, using the known[1,5] procedure for interpolation of data presented in isobaric system of coordinates onto the grid of standard heights that is more convenient for calculating the propagation of

0235-6880/94/02 125-04 $02.00

atmospheric pollution. For such heights we selected the levels of site: h_0 = 0, 0.1, 0.2, 0.4, 0.8, 1.2, 1.6, 2.0, 2.4, 3.0, 4.0, 5.0, 6.0, and 8.0 km. As for the characteristics of the mean wind vector, they were retrieved either when forming the initial samples (that is, when $\langle v_x \rangle$ and $\langle v_y \rangle$ were directly retrieved by the MMCA algorithm) or from the results of forecasting of the zonal and meridional components of wind speed specified at separate levels.

Since we apply here the same MMCA algorithm that was used before,[4] we omit any lengthy expositions. We may restrict ourselves to mentioning that the functions used as base to form the set of forecasting models are the mixed differential dynamic–stochastic models of the form:

$$Y_{h,\,N+1} = \sum_{s=1}^{N^*} A_{h,\,s}\, Y_{h,\,N+1-s} + \sum_{j=0}^{h-1} B_{h,\,j}\, Y_{j,\,N+1} + \varepsilon_{h,\,N+1}\,, \quad (4)$$

where $Y_{h,\,t}$ are the data from the spatiotemporal observations (here $h = 0, 1, 2, \ldots$; h^* is height, and $t = 1, 2, \ldots, N$ is time of observations); N^* is the order of the time lag ($N^* < [N - h - 1]/2$); $A_{h,\,1}, \ldots, A_{h,\,N}$ and $B_{h,\,0}, \ldots, B_{h,\,h-1}$ are unknown parameters of the model; and, $\varepsilon_{h,\,N+1}$ are model discrepancies.

Now to determine the best model (4) and to have a successful forecasting on its basis, we use all the initial data, dividing them in advance into set A (it contains observations up to time $t = N - 1$) and set B, containing observations at time $t = N$ alone. Besides, two special techniques are used, in particular: the technique of directional group selection, designed to optimize the model structure, and the technique of minimax estimation, to obtain estimates of the model parameters, which would guarantee high quality of the respective forecast.

We should finally note that the accuracy of fore–calculations of the characteristics of average wind ($\langle v_x \rangle$ and $\langle v_y \rangle$) was estimated using the relative standard errors δ/σ in persents (here δ is the absolute standard forecast error, and σ is the rms deviation, characterizing the natural variability of a meteorological value).

Now consider some preliminary results from numerical experiments in estimating the quality of fore–calculations of the characteristics of average wind (at a time lag of 12 hrs). They are shown in Table I, which contains the relative standard errors in deviations of the retrieved values of $\langle v_x \rangle$ and $\langle v_y \rangle$ from their respective values of the parameters obtained based on the actual radiosonde observations. Note, that due to cumbersome nature of data assembled in Table I, we only cite accuracy estimates for the single station of Rome.

Numerical experiments in estimating the quality of forecasting characteristics of the mean wind have thereby indicated the following:

1. The MMCA is efficient in forecasting the zonal $\langle v_x \rangle$ and the meridional $\langle v_y \rangle$ components of the vector of mean wind, in the case when surface data are available for the moment of observations corresponding to the chosen time lag (i.e., 12 hrs). Moreover, the accuracy of the forecast $\langle v_x \rangle$ and $\langle v_y \rangle$ is much higher in case the values of v_x and v_y are used directly, instead of the above, and averaging over retrieved values is done throughout the $0 - h$ layers.

2. Forecasting by means of the MMCA is typically best (at a relative error of 60 %) for the atmospheric layers $0 - h$ generally 5–6 km of height, and even if the surface observational data are available only, one can reliably estimate the characteristics of average wind up to 2.4–4.0 km.

3. The most successful retrieval of both the characteristics of average wind and the vertical profiles of v_x and $v v_y$ (Ref. 4), occurs when 10 structures are specified which define the (qualitatively) best structure of the prognostic model, while the statistical sets used include 14–16 profiles.

TABLE I. Relative standard errors (θ, %) of deviations of the retrieved components of the mean wind vector from the respective values based on radiosonde data obtained at Rome site for winter and summer separately.

Retrieval layer, km	Informative layers (levels), km												
	0 – 6.0	0 –5.0	0 – 4.0	0 – 3.0	0 – 2.4	0 – 2.0	0 – 1.6	0 – 1.2	0 – 0.8	0 – 0.4	0 – 0.2	0 – 0.1	0
1	2	3	4	5	6	7	8	9	10	11	12	13	14
	1. Zonal component of the mean wind vector, winter												
0 – 8.0	7	14	50	50	48	50	56	63	67	74	92	98	102
0 – 6.0		2	25	28	27	28	35	41	44	52	68	73	77
0 – 5.0			9	15	16	16	26	30	34	42	56	61	66
0 – 4.0				5	5	7	18	21	23	31	45	50	54
0 – 3.0					0	3	9	10	11	18	33	38	41
0 – 2.4						1	3	4	6	12	26	30	33
0 – 2.0							1	1	3	8	20	23	25
0 – 1.6								0	2	5	13	15	17
0 – 1.2									1	2	6	8	8
0 – 0.8										1	2	3	3
0 – 0.4											0	0	1
0 – 0.2												0	0
0 – 0.1													0
	2. Meridional component of the mean wind vector, winter												
0 – 8.0	7	11	12	43	44	47	71	73	76	76	82	97	110
0 – 6.0		1	1	23	23	23	48	50	52	53	61	73	87
0 – 5.0			0	13	13	13	36	37	40	40	50	61	76
0 – 4.0				4	4	5	24	25	26	27	38	49	64

TABLE I continued

1	2	3	4	5	6	7	8	9	10	11	12	13	14
0 – 3.0					0	2	11	12	12	12	24	35	48
0 – 2.4						0	4	5	5	5	16	26	37
0 – 2.0							1	2	2	2	12	20	29
0 – 1.6								0	0	0	8	14	21
0 – 1.2									0	0	4	8	12
0 – 0.8										0	2	3	5
0 – 0.4											0	0	1
0 – 0.2												0	0
0 – 0.1													0
3. Zonal component of the mean wind vector, summer													
0 – 8.0	10	27	37	53	53	55	69	73	73	73	77	80	83
0 – 6.0		4	12	25	25	26	44	50	48	46	50	51	54
0 – 5.0			4	13	13	16	33	39	38	38	37	38	40
0 – 4.0				4	3	9	23	28	28	24	25	27	29
0 – 3.0					0	4	12	15	15	16	15	17	20
0 – 2.4						1	4	5	6	7	11	13	15
0 – 2.0							1	2	3	6	8	10	12
0 – 1.6								0	1	4	5	6	8
0 – 1.2									0	2	2	3	4
0 – 0.8										0	1	1	2
0 – 0.4											0	0	0
0 – 0.2												0	0
0 – 0.1													0
4. Meridional component of the mean wind vector, summer													
0 – 8.0	21	53	54	85	89	90	95	102	107	120	133	147	152
0 – 6.0		6	6	58	59	62	70	74	76	86	98	110	116
0 – 5.0			0	33	33	35	48	53	55	61	73	85	92
0 – 4.0				10	10	12	29	35	38	41	51	63	70
0 – 3.0				0	0	3	12	15	18	25	34	43	49
0 – 2.4						1	4	6	8	17	25	32	37
0 – 2.0							1	2	4	11	18	24	27
0 – 1.6								0	2	1	11	15	18
0 – 1.2									0	3	5	1	9
0 – 0.8										1	2	3	4
0 – 0.4											0	0	0
0 – 0.2												0	0
0 – 0.1													0

It is also important that the MMCA algorithm also may yield good results (in particular, the noticeable increasing of the height h) in case the data from wind lidar soundings may be used to estimate the characteristics of mean wind. The wind lidars feature high vertical resolution, although they are limited to low heights of sounding (around 1.0–1.5 km), since their accuracy at higher altitudes is insufficient.[6] This conclusion follows, in particular, from data presented in Table II, which presents the heights of upper boundary h and the relative errors θ in retrieving the components of mean wind by the MMCA at the stations of Keflavik, Rome, and Miami, as retrieved via MMCA from lidar wind observations in the 0–1.2 km layer. That is the basic layer in which lidar wind observations are usually taken. Indeed, it follows from Table II that the limit upper boundary height of forecast h for the parameters $\langle v_x \rangle_{0-h}$ and $\langle v_y \rangle_{0-h}$ are much higher in that case, than when the surface data alone are used, and reaches 5–6 km, while the standard relative error nowhere did not exceed 60%.

Thus, the results of numerical experiments indicate that applying the MMCA algorithms to the problem in forecasting the characteristics of average wind, which controls the distribution of atmospheric pollution, is effective, so that such a technique may be successfully applied to building an automated system of atmospheric ecological monitoring of limited areas. However these results are still preliminary and need further substantiation, using comprehensive experimental data. This is the problem for our further studies.

TABLE II. Values of upper boundary height h (km) and of the relative standard error $\theta = \delta/\sigma$ *(%) for the characteristics of the mean wind forecasted by the MMCA algorithm, as applied to data of radiosounding of the* 0–1.2 km *layer*

Station	Zonal components of the vector of mean wind		Meridional components of the vector of mean wind	
	h	θ	*h*	θ
	Winter			
Keflavic	5	60	5	51
Rome	6	41	6	50
Miami	5	58	6	60
	Summer			
Keflavic	5	60	5	61
Rome	6	50	5	53
Miami	5	58	6	56

REFERENCES

1. F.F. Brjukhan', *Techniques for Climatologic Processing and Analysis of Aerologic Data* (Moscow, Gidrometeoizdat, 1983), 112 pp.
2. A.M. Vladimirov, Yu.I. Lyakhin, L.T. Matveev, and V.G. Orlov, *Environmental Protection* (Leningrad, Gidrometeoizdat, 1991), 423 pp.
3. Yu.L. Kocherga, Automatics, No. 5, 80–87 (1991).
4. V.S. Komarov, V.I. Akselevich, and A.V. Kreminskii, Atmos. Oceanic Opt. **7**, No. 2, 121–124 (1994).
5. V.E. Zuev and V.S. Komarov, *Statistical Models of the Temperature and Gaseous Components of the Atmosphere* (D. Reidel Publishing Company, Dordrtecht–Boston–Lancaster–Tokyo, 1987), 306 pp.
6. V.S. Komarov, V.I. Akselevich, A.V. Kreminskii, and G.G. Matviyenko, Atmos. Oceanic Opt. **7**, No. 2, 96–99 (1994).