Numerical retrieval of temperature and wind profiles in the boundary atmospheric layer based on the Kalman filter algorithm and 2D dynamical-stochastic model. Part 1. Methodology

E.V. Gorev, V.S. Komarov, A.V. Lavrinenko, and V.V. Budaev

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

Received December 10, 2007

Methodological basis for solution of the problem of numerical retrieval of vertical profiles of temperature and wind in the boundary atmospheric layer is considered based on ground data and individual altitude observations with the Kalman algorithm and 2D dynamical-stochastic model.

Among many problems of the present-day meteorology, connected with information support of applied tasks, of importance is the filling of the missing data for individual atmospheric layers, including its boundary layer (ABL). At the same time, the fields of application of the data filling procedure are quite various. For example, atmospheric optics, related with lidar remote sensing of the environment parameters, is among them. In spite of noticeable advantages of lidar systems over the systems of radio sensing (very high spatial-temporal resolution of lidar measurements at the accuracy quite acceptable for practice), their functioning significantly depends on weather conditions.

For example, optical sensing of the atmosphere cannot be carried out at low continuous cloudiness, fog, and intensive precipitation. So, under such weather conditions (for providing all-weather lidar sensing) it is necessary to use one of the possible algorithms for numerical extrapolation of meteorological fields over height, using the data of ground-based or lower atmospheric layers.

The second example of application of the procedure can be ecology, because estimation and forecast of the processes of pollution of atmospheric air at the local level (especially in large cities) are impossible without preliminary obtained information on the temperature stratification and wind vertical profiles in ABL, where principal transfer of pollutions is observed.[1] Since in this case it is necessary to have data on thermal and wind regime in ABL at many points of the considered mesoscale region, there arises a problem of numerical retrieval of the vertical profiles of temperature and wind at each of such points.

It is quite difficult to realize such procedure in practice, because only one aerological station is usually situated in the vicinity of a large city or an industrial center. Therefore, one of the possible ways out can be the procedure for numerical retrieval of the temperature and wind profiles at a given point, carried out with any method of vertical extrapolation by the data of observations at the available aerological station and the nearest meteorological station.

It is known that in practice for a long time the problem of filling the missed height meteorological information was solved by statistical extrapolation methods based on the use of the multiple regression equations.[2–4] However, these methods give satisfactory results only for atmospheric layers of small thickness and in the absence of essential bends in the vertical profiles of meteorological parameters.

In this connection, another approach to numerical retrieval of missed data was proposed in 90th of the XX century at the Institute of Atmospheric Optics SB RAS. It was based on the modified method of clustering the arguments (MMCA). The detailed description of this method is presented in Ref. 5.

Despite quite satisfactory results obtained at retrieval of the vertical profiles of meteorological parameters (in particular, temperature and wind). using the MMCA algorithm, it requires obligatory fulfillment of the condition that the sample size of height-time observations is of order $N = k + 1$ (where k is the number of atmospheric levels) and no less than 7 profiles in hand.[5] Fulfillment of this requirement leads to impossibility of the real-time use of the MMCA algorithm at $N = 2–3$, i.e., from the very moment of income of the first aerological measurement data.

Therefore, the necessity appeared of development of a more effective method for numerical retrieval of the vertical profiles of meteorological parameters, free of the above restrictions and enable to fill the height missed data (especially in the atmospheric boundary layer, where dramatic breaks on the temperature stratification and vertical distribution of wind are often observed) from the data of single meteorological measurements.

0235-6880/08/04 281-04 $02.00 © 2008 Institute of Atmospheric Optics

In this paper we consider the basis of a technique for solving this problem, which for the first time is based on application of the Kalman filtering algorithm and two-dimensional dynamical-stochastic model, describing variations of a meteorological parameter in height and time.

Such an approach makes it possible on the basis of two preceding altitudinal observations and the ground-based data obtained at the time of retrieval, to fill the missed data at all levels of ABL not only at the site of location of the aerological station (or the system of lidar sensing), but at all arbitrary points of some mesoscale region, where the ground-based meteorological measurements were carried out.

The problem of numerical retrieval of the vertical profiles of some meteorological parameter at a given point of space (x_0, y_0) lies in estimation of its value at the fixed time moment t_0 from the measurements at lower layers using some mathematical model. An important condition for such retrieval of the vertical profiles is the presence of measurements at lower levels at the moment t_0 and, besides, a sequence of measurements of the vertical profiles obtained at preceding time moments $t_0 - j$, where $j = 1, 2, \ldots, K$, at the same point of space (x_0, y_0) or at a neighboring point (x_1, y_1). One can use measurements of the vertical profile at the point (x_1, y_1) for the moment t_0, however, in this case the correlation structure of the vertical distribution of the meteorological parameter to be retrieved should not essentially change between these two points.

The two-dimensional dynamical-stochastic model served as a mathematical model used for solution of the stated problem. It describes the dependence of variation of the field ξ in height and time:

$$\xi_h(k) = \sum_{m=h}^{h+i}\sum_{j=0}^{K} d_{j,m}\xi_m(k-j) + \varepsilon_h(k), \qquad (1)$$

where $\xi_h(k)$ is the value of the field of meteorological parameter at the fixed height h at the moment k; m is the current number of the height level, at which the retrieval is performed; m changes from h to $h + i$ (here $i = 1, 2, \ldots, n$ is the maximum number of the height data levels; j is the current value of the discrete time changing from 0 to K and determining the size of predictor of the Kalman filter algorithm; $d_{j,m}$ are the unknown coefficients to be estimated, which determine the correlation between the field $\xi_h(k)$ and its values at preceding moments at a given height and above-resting height levels, i.e., $\xi_m(k-j)$; $\varepsilon_h(k)$ is the discrepancy of the model determined by stochastics of the considered atmospheric processes.

Thus, according to Eq. (1), the value of the field $\xi_h(k)$ at the fixed h and at a given moment $k = t_0$ at the point of retrieval (x_0, y_0) is a linear combination of the vector of unknown parameters $d_{j,m} = \mathbf{D}$ and the values of this field at the considered and preceding time moments to the depth K, at the same level and at i above-resting levels at the point of aerological measurements (x_1, y_1). Note that a case can appear, when measurements and retrieval are carried out at the same spatial point, i.e., at $x_0=x_1$ and $y_0=y_1$.

Then, when considering the algorithm for retrieval of the vertical profiles, we take $i = 2$, i.e., only the fixed and two higher layers are used in calculations, and $j = 3$, that corresponded to application of three preceding measurements.

For balloon measurements, the centered values of the meteorological field are taken in Eq. (1) as the initial data, and the procedure for retrieval of the vertical profile of the meteorological parameter $\xi_h(k)$ is performed using the two-channel scheme.[6] According to this scheme, the resulting estimate of the value of the field ξ at the point of retrieval (x_0, y_0) is the sum of the estimate of the regular component of the field $\bar{\xi}$ and the estimate of the fluctuating component ξ', i.e.,

$$\xi=\xi'+\bar{\xi}. \qquad (2)$$

The prime is missed in Eq. (1). Further this sign also will be omitted.

Regular component of the field $\bar{\xi}$ was calculated as the smoothed average value over several previous observations at a fixed height:

$$\bar{\xi}_h(k) = \frac{1}{p}\sum_{j=1}^{p}\xi_h(k-j), \qquad (3)$$

where p is the depth of the temporal window, used for estimation of the regular component of the field. Note that in general case $p \neq K$.

Consider now the technique for retrieval of the vertical profile based on the Kalman filter and the model (1).

It follows from Eq. (1) that the value of $\xi_h(k)$ parametrically depends on values of $\xi_h(k-j)$ obtained at earlier moments, therefore, the problem of retrieval of the vertical profile at some point of observation is solved in two stages. At the first stage, the parameters of the model $d_{j,m}$ are estimated based on the Kalman filter algorithm from measurements of ξ at the point (x_1, y_1) taken at the moment t_0 and earlier moments to the depth K for the fixed level h and i above-resting levels, as well as from its ground-based measurements at the point (x_0, y_0).

At the second stage, we obtain the estimate of the field at a given point (x_0, y_0) at the moment t_0 at the height $h + 1$ based on the assumption on a weak variability of the vector of the estimated coefficients $\mathbf{D}$ within the limits of the considered atmospheric layer, using the equation for retrieving the vertical profile.

The equation for retrieval of the vertical profile for one height level upward is as follows:

$$\hat{\xi}_{h+1}(k) = \sum_{m=h}^{h+i}\sum_{j=0}^{K}\hat{d}_{j,m}\xi_{m+1}(k-j). \qquad (4)$$

Here $\hat{\xi}_{h+1}(k)$ is the estimate of the meteorological field at the moment k at the height $h + 1$ at the point (x_0, y_0); $\hat{d}_{j,m}$ are estimates of the unknown parameters of the model made for the h-th height level; $\xi_{m+1}(k-j)$ are the measured values of the meteorological field at the point (x_1, y_1) at the height levels starting from $h+1$ up to $h+i+1$ at the moments from k to $k-K$.

Then the estimate $\hat{\xi}_{h+1}(k)$, obtained for the level $h+1$ was used in the left part of Eq. (1) as the observation at fixed level. The right part of Eq. (1) is filled with observations for the fixed atmospheric layer shifted to one level upward. Thus, recurrently, repeating the two-stage procedure of estimation of the vertical profile, we successively obtain estimates of ξ at each height level of ABL up to $h_{\max}-i$ (here $h_{\max}$ is the maximum height corresponding to the top of the atmospheric boundary layer).

To estimate the unknown parameters of the model (1), i.e., $d_{j,m}$, a system of difference equations in the matrix form is set[7]:

$$\mathbf{x}_{k+1}^t = \mathbf{\Psi}_k \mathbf{x}_k^t + \boldsymbol{\omega}_k^t, \tag{5}$$

where $\mathbf{x}_k^t$ is the vector-column of the dimension $n=(i+1)K$, including the unknown parameters of the state of the dynamical system to be estimated (the "true" vector of state) for the current discrete time k; $\mathbf{\Psi}_k$ is the transition matrix for the discrete system of the dimension

$$(n\times n)=[(i+1)K\times(i+1)K];$$

$\boldsymbol{\omega}_k^t = \left\|\omega_1^t, \omega_2^t, \ldots, \omega_n^t\right\|^{\mathrm{T}}$ – is the vector-column of the random disturbances of the system (vector of the state noises) of the dimension $n=(i+1)K$,

$$\left\langle \boldsymbol{\omega}_k^t \right\rangle = 0, \quad \left\langle (\boldsymbol{\omega}_k^t)(\boldsymbol{\omega}_l^t)^{\mathrm{T}} \right\rangle = \mathbf{Q}_k \delta_{kl},$$

where δ_{kl} is the Kronecker sign; $\mathbf{Q}_k$ is the matrix of covariance of the state noises; T is the transposition operator.

Provided parameters $\mathbf{x}_k^t$ are unchanged on the average in time on the considered time interval, then the transition matrix $\mathbf{\Psi}_k$ corresponds to the unit matrix of the dimension $(n\times n)$.

In this case, equation (5) is written in the form

$$\mathbf{x}_{k+1}^t = \mathbf{x}_k^t + \boldsymbol{\omega}_k^t. \tag{6}$$

The mathematical model of measurements, on the basis of which the state of the system is estimated by the Kalman filter algorithm, in general case is described by the additive mixture of the useful information and the measurement error:

$$\mathbf{y}_k^0 = \xi_k^0 = \mathbf{H}_k \mathbf{x}_k^t + \boldsymbol{\varepsilon}_k^0, \tag{7}$$

where $\mathbf{y}_k^0$ is the vector of actual measurements. In our case $\mathbf{y}_k^0$ is the scalar-number, representing the measurement at the fixed height h at the moment k; $\mathbf{H}_k$ is the vector of observations of the dimension $n=(i+1)K$ determining the functional relation between true values of variables of state and actual measurements, i.e., the vector-row, the elements of which are successive predictors, $\boldsymbol{\varepsilon}_k^0$ is the error in measuring at the moment k (noise of measurements); it also is the scalar value

$$\left\langle \boldsymbol{\varepsilon}_k^0 \right\rangle = 0, \quad \left\langle (\boldsymbol{\varepsilon}_k^0)(\boldsymbol{\varepsilon}_l^0)^{\mathrm{T}} \right\rangle = \mathbf{R_k}\delta_{kl}, \ \left\langle (\boldsymbol{\varepsilon}_k^0)(\boldsymbol{\omega}_l^t)^T \right\rangle = 0,$$

where $\mathbf{R}_k$ is the covariance matrix of the measurement noises.

Preset now the matrix of observations $\mathbf{H}_k$. It becomes clear from comparison of the formulas for the model (1) and the mathematical model of measurements (7) that the matrix $\mathbf{H}_k$ elements are the measurements of $\varepsilon_h(k)$ at the atmospheric levels within the limits of the fixed height window at the considered and preceding moments to the depth K. Thus, the matrix $\mathbf{H}_k$ can be written in the form

$$\mathbf{H}_k = \left\| y_0(k-1),\ y_0(k-2),.,\ y_0(k-K), y_1(k-1),.,y_{i+1}(k-K) \right\|. \tag{8}$$

After determining all elements included into Eqs. (6) and (7), the problem of estimation of $d_{j,m}$ in model (1) is solved using the linear Kalman filter,[8] which provides for estimation of the elements of the vector of state with minimal root-mean-square errors.

In this case, the equation for optimal estimation of the vector of state has the form

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k^0 - \mathbf{H}_k \cdot \mathbf{x}_k^f), \tag{9}$$

where $\mathbf{x}_k^a$ is the vector of analyzed values (estimate of the vector of state) at the moment k; $\mathbf{x}_k^f$ is the vector of predicted values at the moment k; $\mathbf{K}_k$ is the matrix of the weight coefficients of the dimension $(n\times n)$.

The weight coefficients in the linear Kalman filter are calculated by the recurrent matrix equations.[8]

The value of the field $\xi_{h+1}(k)$ is calculated at the second stage based on the obtained estimates of the vector of state $\mathbf{x}_k^a$. The algorithm of retrieval of the vertical profile by Eq. (4) can be presented in the matrix form:

$$\xi_{h+1}(\mathbf{k}) = \mathbf{y}_k^0 = \mathbf{H}_k^* \cdot \mathbf{x}_k^a, \tag{10}$$

where $\mathbf{H}_k^*$ is the transition vector of the dimension $n=(i+1)K$ used for calculation of the field value at the height level $h+1$ at the moment k; $\mathbf{x}_k^a$ is the obtained estimate of the vector of state at the moment k.

The formulas for estimation of the vector of state (9) and the retrieved vertical profile (10) completely determine the procedure for the use of the two-dimension dynamical-stochastic model and the Kalman filtering formalism for the problems of retrieval of the vertical profiles of the prescribed meteorological parameter ξ at an individual point.

Note for the conclusion that the efficiency of the proposed approach to solving the problem of numerical retrieval of the vertical profiles of meteorological parameters can be judged from the results of field experiments. This is the subject of the second part of this paper.

References

1. A.M. Vladimirov, Yu.I. Lyakhin, L.T. Matveyev, and V.G. Orlov, *Preservation of Environment* (Gidrometeoizdat, Leningrad, 1991), 423 pp.
2. A.S. Marchenko, L.A. Minakova, and A.G. Semochkin, in: *Application of Statistical Methods in Meteorology* (CC SB AS USSR, Novosibirsk, 1991), pp. 82–113.
3. I.L. Byzova, A.A. Lunina, and O.M. Khachaturova, Trudy IEM, Issue 41 (126), 25–50 (1987).
4. V.S. Komarov and Yu.B. Popov, Proc. SPIE **3983**, 591–599 (1999).
5. V.S. Komarov, *Statistics in Application to the Problems of Applied Meteorology* (SB RAS, Tomsk, 1997), 255 pp.
6. V.S. Komarov, A.V. Lavrinenko, N.Ya. Lomakina, Yu.B. Popov, A.I. Popova, and S.N. Il'in, Atmos. Oceanic Opt. **17**, No. 8, 584–589 (2004).
7. E.P. Sage and J.L. Melsa, *Theory of Estimation and its Application to Communications and Control* (Svyaz', Moscow, 1976), 496 pp.
8. K. Brammer and G. Ziffling, *Kalman–Bucy Filter* (Nauka, Moscow, 1982), 200 pp.