IOP Conf. Series: Earth and Environmental Science **96** (2017) 012010 doi:10.1088/1755-1315/96/1/012010

Statistical analysis of experimental data for mathematical modeling of physical processes in the atmosphere

P A Karpushin[1,3], Yu B Popov[2], A I Popova[2], K Yu Popova[1], N P Krasnenko[1,3] and A V Lavrinenko[4]

[1]Tomsk State University of Control Systems and Radioelectronics, Russia, Tomsk,
[2]Surgut State University, Russia, Surgut,
[3]Institute of Monitoring of Climatic and Ecological Systems SB RAS, Russia, Tomsk,
[4]V.E. Zuev Institute of Atmospheric Optics SB RAS, Russia, Tomsk,

E-mail: pavel.a.karpushin@tusur.ru, kseniia.i.popova@tusur.ru, krasnenko@mces.ru

Abstract. In this paper, the probabilities of faultless operation of aerologic stations are analyzed, the hypothesis of normality of the empirical data required for using the Kalman filter algorithms is tested, and the spatial correlation functions of distributions of meteorological parameters are determined. The results of a statistical analysis of two-term (0, 12 GMT) radiosonde observations of the temperature and wind velocity components at some preset altitude ranges in the troposphere in 2001–2016 are presented. These data can be used in mathematical modeling of physical processes in the atmosphere.

In the last few years, a dynamic stochastic approach based on the theory of Kalman filtering has been widely used in mathematical modeling of physical processes in the atmosphere and for the development and study of algorithms of reconstruction of mesometeorological fields. The use of the Kalman filter for solving these problems requires testing for Gaussian distribution of the sought-for meteorological parameters [1]. Therefore, we performed corresponding investigations, in which long-term aerologic observations in the troposphere, given in standard bulletins and telegrams of the World Meteorological Organization (WMO) for 2001–2016 in the KN-04 code, were used as empirical data. In the present work, the hypothesis of normality of the empirical data distributions is tested, the spatial correlation functions of the meteorological parameters are calculated, and the probabilities of faultless operation of the observation channels are analyzed.

The data of aerologic stations were grouped according to their geographical coordinates into typical mesometeorological experimental sites. In the first stage of the initial aerologic data processing, the code combinations of telegrams were analyzed and the coding errors were corrected. Using a linear interpolation, the vertical profiles of the aerologic data of each of the chosen stations of the Russian Federation were reduced to a system of standard isobaric altitude levels and geometric altitudes [2, 3].

With the data array obtained, the aprioristic probabilities of faultless operation of the measurement channels were analyzed, and the allowable threshold values were determined to decide on the presence or absence of 1) useful information in the measurement channels, 2) anomalous data overshoots in the measurements, 3) an atmospheric front within the mesoscale experimental site being considered.

To estimate the probability of faultless operation q and the probability of measurement channel failure $(1 - q)$, data arrays were compiled for each station from the entire aerologic data archive for a fixed period of time. The data arrays were formed for two crucial months: January and July. A preliminary analysis of the meteorological data on the presence of data omissions demonstrated that the aprioristic probability of faultless operation was 0.89 for both winter and summer.

In our analysis, the presence in the observation data of anomalous errors arising in the course of compiling, re-writing, and deciphering of the standard meteorological telegrams was taken into account. The presence of these inhomogeneities may lead to an increase in the errors and to failures in the operation of forecast algorithms. The analysis of the calculated statistical characteristics allowed us to reveal the anomalous errors. The difference between the verified value and the arithmetic average was used for this purpose. According to the theory of hypothesis testing, the modulus of this difference was compared with $N\sigma$, where σ is the standard deviation of the chosen meteorological parameter [4] and $N = 3$.

An important point in the analysis of the aerologic data is the detection of the passage of an atmospheric front. For this purpose, a comparison of the differences between the increments of pressure, temperature, and wind direction at two subsequent *k*th and (k+1)*th* time moments were used. A sharp change in the wind direction that breaks the monotonic behavior of the meteorological parameters is an obvious indication of the passage of an atmospheric front. The aprioristic probabilities of front detection and the threshold values in decision-making on the absence or presence of a front were calculated for the meteorological measurements at each of the stations.

To test the statistical hypothesis of normality of the empirical meteorological parameters being analyzed, the coefficients of asymmetry (A) and excess (E) were analyzed together with numerical values of the Kolmogorov–Smirnov normality test criterion [4, 5, 6]. To estimate the deviation of the curve of a normal distribution from the curves of empirical distributions, formulas [7]

$$\sigma_A \cong \sqrt{\frac{6}{n+4}}, \qquad \sigma_E \cong \sqrt{\frac{24*(n-5)}{n*(n+7)}} \tag{1}$$

were used, where σ_A is the standard root-mean-square error of the asymmetry coefficient, σ_E is the same error of the excess coefficient, and n is the number of examined realizations. The series was considered to be normally distributed, if the conditions

$$|A| \leq 3\sigma_A, \ |E| \leq 3\sigma_E \tag{2}$$

were satisfied. During our analysis, we investigated some temporal statistical characteristics of the meteorological fields. To calculate the statistical characteristics, the following expressions were used:

– meteorological parameters averaged over an ensemble of realizations or the first initial moments,

– variance or the second-order central moment of a random process,

– excess coefficient (characterizes the position of the maximum in the curve of an empirical distribution relative to the curve of a normal distribution),

– asymmetry coefficient (characterizes the skewness of the distribution curve).

The results of our investigations demonstrated an insignificant deviation of the empirical distribution functions of the observation data from a normal Gaussian distribution law. Inequality (2) holds true for each station of the experimental site. The insignificant deviations from the normal distribution law are explained by the presence of systematic and random errors in the data of observations of meteorological fields in the atmosphere.

The statistical analysis demonstrated that the empirical distributions of the examined meteorological parameters can be described with sufficient accuracy by ae normal distribution, which is a major condition for the use of the Kalman filter in the algorithms for the reconstruction of mesometeorological fields. This distribution is a good mathematical model for the observable random variables.

The correlation functions for the stations were calculated by the standard formulas [8]:

$$\mu_\xi(\rho) = \frac{1}{n-1}\sum_{i=1}^{n}\left[\xi(r_i) - m_\xi(r_i)\right]\left[\xi(r_i+\rho) - m_\xi(r_i+\rho)\right], \quad (3)$$

where $\xi(r_j)$ are the values of the meteorological parameter at points r_j, respectively, $m_\xi(\rho)$ is the mathematical expectation of the meteorological parameter, and ρ is the distance. Figures 1 and 2 show the spatial correlation functions for the chosen experimental site, indicated altitudes, and January and July periods.

Temperature **Zonal wind velocity component** **Meridional wind velocity component**

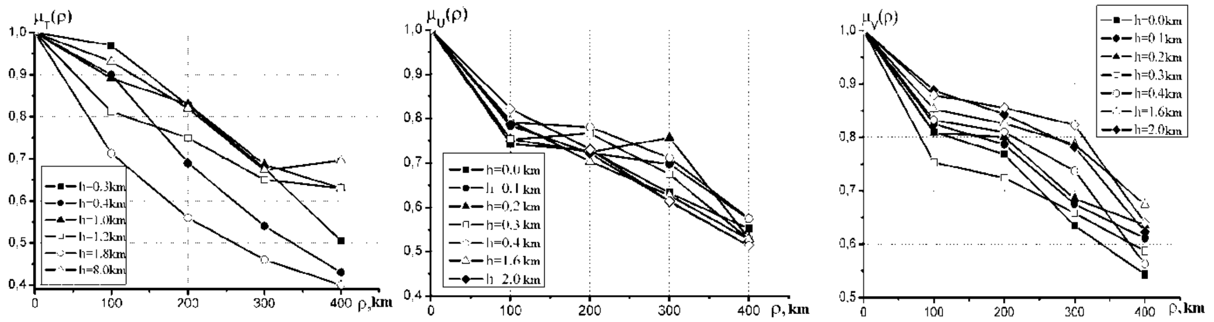

Figure 1. Spatial empirical correlation functions of the meteorological parameters, January.

The spatial correlations of the temperature and wind velocity components at all examined altitudes (Figure 1) decrease with increasing distance ρ.

In summer, the spatial correlations of the temperature and wind velocity components at all examined altitudes shown in Figure 2 behave similarly to those shown in Figure 1.

The correlation functions shown in Figs. 1 and 2 allow using the analytical expression

$$\mu(\rho) = e^{-\beta\cdot\Delta\rho} \quad (4)$$

as an approximating function. In spite of the fact that the quality of approximation is worse than that in [4, 9], it allows one to simplify significantly the Kalman filter algorithm of spatial forecasting.

Temperature **Zonal wind velocity component** **Meridional wind velocity component**

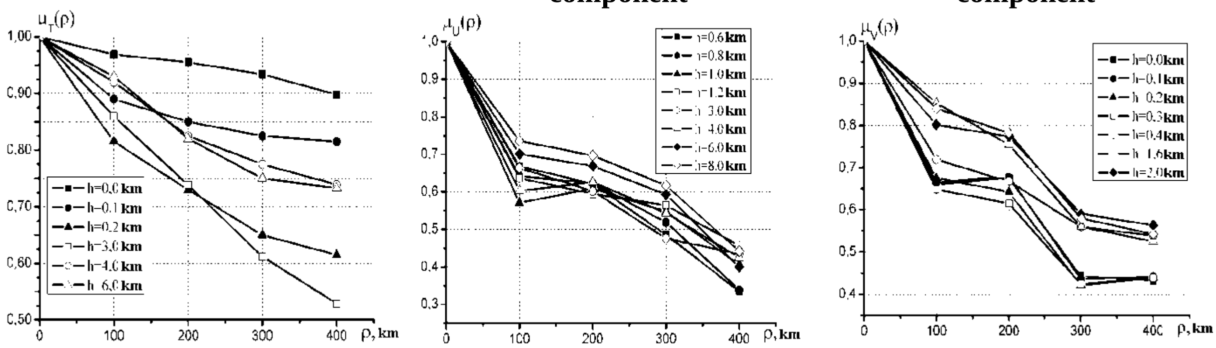

Figure 2. Spatial empirical correlation functions of the meteorological parameters, July.

The results were obtained with financial support of the Ministry of Education and Science of the Russian Federation (project no. 5.3279.2017/PSN) and of the Program of Basic Research of the Siberian Branch of the Russian Academy of Sciences (project no. IX.138.2.5).

References

[1] Popov Yu B 2010 Estimation of the parameters of the atmospheric state using four-dimensional dynamic stochastic model and the linear Kalman filter Part 1 Basic Methodology *TUSUR Rep.* **21** 5-11
[2] Lavrinenko A V 2006 *Multidimensional dynamic stochastic models and their use in applied problems. Candidate's Dissertation* (Tomsk: IAO SB RAS) p 136
[3] Popova A A 2006 *Application of adaptive algorithms in numerical method of spatiotemporal extrapolation of mesometeorological fields. Candidate's Dissertation* (Surgut: IAO SurSU) p 147
[4] Il'in S N 2005 *Dynamic stochastic method of spatial extrapolation of meteorological fields in a mesoscale region Dissertation* (Tomsk: IAO) p 120
[5] Kobysheva N V and Narovlyanskii G Ya 1978 *Climatic Processing of Meteorological Information* (Leningrad: Gidrometeoizdat) p 295
[6] Manita A D 2001 *Probability Theory and Mathematical Statistics: A Textbook.* (Moscow: Publishing Department of Educational and Scientific Center of Pre-University Education) p 120
[7] Bol'shakov V D 1983 *Theory of Observational Errors: A Textbook for High Schools – 2nd Addition Rev. Suppl.* (Moscow: Nedra) p 223.
[8]. Popov Yu B and Popova A A *2008 Optimal Filtration and its Application to Problems of Monitoring of the Parameters of the Atmospheric State within Local Territories* (Khanty-Mansiysk: Poligrafist) p 188
[9] Komarov V S, Popov Yu B, Suvorov S S and Kurakov V A 2004 *Dynamic Stochastic Methods and Their Use in Applied Meteorology* (Tomsk: Publishing House of the IAO SB RAS) p 236