Doklady Biological Sciences, Vol. 385, 2002, pp. 285–287. Translated from Doklady Akademii Nauk, Vol. 385, No. 1, 2002, pp. 113–115.

BIOCHEMISTRY, BIOPHYSICS, AND MOLECULAR BIOLOGY

Statistics of the Concentration of Tropospheric Bioaerosol

A. I. Borodulin*, A. S. Safatov*, B. D. Belan, and M. V. Panchenko****

Presented by Academician L.S. Sandakhchiev February 22, 2002

Received February 22, 2002

Although statistical aspects of the distribution of atmospheric impurity concentration have long been a subject of extensive discussion in the literature, this problem is still of considerable theoretical and applied importance. The biogenic component of tropospheric aerosol in the southern regions of Western Siberia is a subject of systematic research at the Research Institute of Aerobiology, Vector State Research Center for Virology and Biotechnology, in collaboration with the Institute of Atmosphere Optics, Siberian Division, Russian Academy of Sciences [1, 2]. The term "biogenic component" is assumed to include only two fractions of atmospheric aerosol: total protein and live microorganisms. The results of the study and preliminary results of their generalization showed that there is a significant scatter of the value of the biogenic component of tropospheric aerosol within the altitude range from 0.5 to 7 km. This effect cannot be attributed to the measuring error alone. Therefore, it should be explained by the statistical nature of the scatter.

Based on general principles, it may be suggested that the protein component of tropospheric aerosol is a result of fragmentation of certain sufficiently large initial particles. According to Kolmogorov's theorem [3], in the limiting case this system should be described by continual statistic equations (a logarithmically normal distribution, in particular). The atmospheric-aerosol component containing live microorganisms is an ensemble of indivisible subunits. Disintegration and death of a microorganism excludes it from "the list of living creatures." Therefore, it may be suggested that the fractions of live microorganisms of tropospheric aerosol and the fraction of total protein should be described by equations of discrete statistics and continual statistics, respectively. This work describes the results of testing this hypothesis.

* *Research Institute of Aerobiology, State Research Center of Virology and Biotechnology Vector, pos. Kol'tsovo, Novosibirsk oblast, 630559 Russia*

** *Institute of Atmosphere Optics, Siberian Division, Russian Academy of Sciences, pr. Akademicheskii 1, Tomsk, 634055 Russia*

The logarithmically normal distribution function can be written as:

$$F(C) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\ln C - M}{\sqrt{2}\Sigma}\right)\right], \tag{1}$$

where C is the distribution-function argument; M and Σ are parameters depending on the mathematical expectation of concentration $\bar{C}$ and its variance σ_2; erf(…) is the probability integral.

Another distribution function is of substantial practical importance [4]:

$$F(C) = 1 + \frac{1}{2}\left[\operatorname{erf}\left(\frac{C - \bar{C}}{\beta}\right) - \operatorname{erf}\left(\frac{C - \bar{C}}{\beta}\right)\right], \tag{2}$$

where β is the second parameter of the distribution function depending on $\bar{C}$ and σ^2. Eq. (2) is an exact analytical solution of the Fokker–Planck–Kolmogorov equation. This solution was obtained assuming that the change in the concentration of the atmospheric impurity at the given point of space is a Markovian process.

It was shown [5] that the distribution of concentrations of some atmospheric impurities is described by the Poisson equation:

$$p(k) = \exp(-\bar{k})\frac{(\bar{k})^k}{k!}, \tag{3}$$

where $p(k)$ is the probability of observation of k particles; $\bar{k}$ is mathematical expectation.

Analysis of experimental data showed that the mean concentrations of total protein and live microorganisms are almost independent on the flight altitude [1, 2]. Based on this finding, the ensemble of experimental data on the concentrations of total protein (C_p) and live microorganisms (C_b) were processed as follows. The mean values of the concentrations of total protein (C_{pm}) and live microorganisms (C_{bm}) averaged over observation altitudes were calculated. Then the values of $\varphi_p = C_p/C_{pm}$ and $\varphi_b = C_b/C_{bm}$, as well as the corresponding variance values $\sigma^2_{\varphi p}$ and $\sigma^2_{\varphi b}$ were calculated. The ensemble of total protein concentrations was represented by $n = 245$ experiments, whereas the ensemble

0012-4966/02/0708-0285$27.00 © 2002 MAIK "Nauka / Interperiodica"

Table 1. Testing the hypothesis on equal values of variance and mathematical expectation

Parameter	Total protein	Microorganisms
Mean values $\sigma^2_{\varphi p}$ and $\sigma^2_{\varphi b}$	$\overline{\sigma^2_{\varphi p}} \pm S_p = 0.29 \pm 0.32$	$\overline{\sigma^2_{\varphi p}} \pm S_b = 2.24 \pm 1.83$
Statistics T_p and T_b	$T_p = -6.28$	$T_b = 1.92$
Quantile $\lvert t_m \rvert_{1-\alpha}$; $\alpha = 0.05$; $m = 8$	2.31	

of the concentrations of live microorganisms was represented by $n = 197$ experiments.

Testing the hypotheses on the consistency of the ensemble of data with the Poisson statistic. As the first hypothesis, we tested the properties of the Poisson distribution, according to which $\sigma^2_k = \bar{k}$, where σ^2_k is the variance of the number of particles. In the case considered in this work, this means that $\overline{\sigma^2_{\varphi p}} = 1$ and $\overline{\sigma^2_{\varphi b}} = 1$, where the bar indicates averaging over the ensemble of data. The results of testing this hypothesis are given in Table 1. Statistics $T_p = \frac{\overline{\sigma^2_{\varphi p}} - 1}{S_p/\sqrt{m}}$ and $T_b = \frac{\overline{\sigma^2_{\varphi b}} - 1}{S_b/\sqrt{m}}$ should follow Student's t distribution with the number of degrees of freedom $m = 8$ corresponding to eight altitudes of concentration measurement. Table 1 also contains standard sample deviations S_p and S_b of $\overline{\sigma^2_{\varphi p}}$ and $\overline{\sigma^2_{\varphi b}}$ and Student's t distribution quantile $\lvert t_m \rvert_{1-\alpha}$ at a confidence level of $\alpha = 0.05$. It follows from the results given in Table 1 that the suggested hypothesis fails to describe the distribution of total protein concentration, but it is acceptable for describing distribution of live microorganisms. Testing this hypothesis is a necessary but not a sufficient condition for making decision on the consistency of the distribution of live microorganisms with the Poisson statistic. Therefore, a direct testing of the hypothesis on the consistency of the ensemble of experimental data with the Poisson statistic (Eq. (3)) is of particular interest. In the general case, the value of φ_b considered above is not integer. However, the corresponding estimates can be obtained. Because $C_b = k/V$, where V is the sample volume, assuming that $k = 0, \bar{k}, 2\bar{k}, \ldots$ we obtain a series of integer values $\varphi_b = 0, 1, 2, \ldots$ corresponding to the number of live microorganisms per m^3 divisible by their mean count. The next hypothesis suggests that the distribution of φ_b divisible by the mean number of particles should also be consistent with the Poisson equation. The empirical and theoretically calculated probabilities $p(\varphi_b)$ and the frequencies of appearance of φ_b values are shown in Fig. 1. The results of testing this hypothesis with the use of the chi-square test are given in Table 2 using the generally accepted notation. It follows from Table 2 that this hypothesis is acceptable at a confidence level of $\alpha = 0.05$.

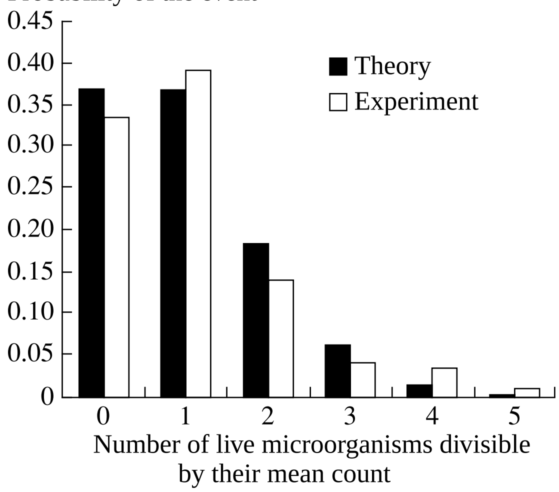

Fig. 1. Theoretically calculated and experimentally measured probabilities that the number of live microorganisms is divisible by their mean count.

Testing the hypothesis on the consistency of the empirical function of the total protein concentration

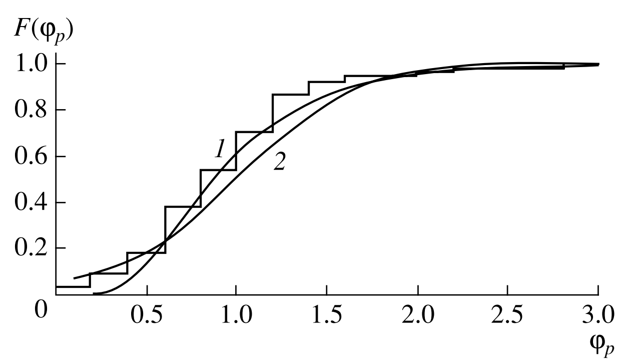

Fig. 2. The empirical function of the distribution of the total protein concentration (histogram) and its comparison with distribution functions (1) and (2) (curves *1* and *2*, respectively).

Table 2. Testing the hypothesis on the consistency of the distribution of the number of live organisms divisible by their mean count with the Poisson equation

φ_{bi}	h_i	$p(\varphi_{bi})$	$n\,p(\varphi_{bi})$	χ^2_i
0	66	0.37	72.5	0.58
1	77	0.37	72.5	0.28
2	28	0.18	36.3	1.90
>3	8	0.08	15.6	3.70
Statistic χ^2				6.46
Distribution quantile $\chi^2_{m,\,1-\alpha}$; $\alpha = 0.05$; $m = 3$				7.81

Table 3. Testing the hypothesis on the consistency of the distribution of the total protein concentration with a logarithmically normal equation

φ_{pi}	h_i	$p(\varphi_{pi})$	$n\,p(\varphi_{pi})$	χ^2_i
0–0.4	21	0.06	18.48	0.34
0.4–0.8	71	0.37	89.43	3.80
0.8–1.2	80	0.31	74.97	0.34
1.2–1.8	60	0.19	46.55	3.89
>1.8	13	0.08	19.11	1.95
Statistic χ^2				10.32
Distribution quantile $\chi^2_{m,\,1-\alpha}$; $\alpha = 0.01$; $m = 3$				11.34

distribution with distribution equations (1) and (2). The empirical histogram of the distribution functions of total protein concentration $F(\varphi_p)$ and theoretical distribution functions (1) and (2) are shown in Fig. 2. It follows from Fig. 2 that the empirical function of concentration distribution is a qualitatively satisfactory fit of the continuous distribution functions (1) and (2). The results of testing of the hypotheses on the consistency of the empirical function of the total protein concentration distribution with distribution Eqs. (1) and (2) are given in Tables 3 and 4. It follows from Tables 3 and 4 that these hypotheses are acceptable at a confidence level of $\alpha = 0.01$–0.005.

Thus, it may be concluded that the results of this study support the suggestion that the statistics describing the concentrations of protein and microbiological components of atmospheric aerosol are of different physical nature. The distribution of the concentration of live microorganisms is described by the discrete Poisson distribution equation. In contrast, the distribution of the total protein concentration is described by continuous statistic Eqs. (1) and (2). However, Eq. (2) seems to be more adequate, because it also describes zero concentrations observed in samples with a probability of $F(0) \approx 0.02$. In contrast to Eq. (2), the logarithmically normal Eq. (1) always gives a zero probability of the zero concentration, i.e., $F(0) \equiv 0$.

Table 4. Testing the hypothesis on the consistency of the distribution of the total protein concentration with Eq. (2)

φ_{pi}	h_i	$p(\varphi_{pi})$	$n\,p(\varphi_{pi})$	χ^2_i
0–0.4	21	0.08	19.60	0.10
0.4–0.8	71	0.22	53.08	6.05
0.8–1.0	39	0.14	35.01	0.45
1.0–1.2	41	0.14	35.15	0.99
1.2–2.0	61	0.32	79.28	4.21
>2	12	0.03	7.35	2.94
Statistic χ^2				14.74
Distribution quantile $\chi^2_{m,\,1-\alpha}$; $\alpha = 0.005$; $m = 4$				14.86

REFERENCES

1. Belan, B.D., Borodulin, A.I., Marchenko, Yu.V., *et al.*, *Dokl. Akad. Nauk*, 2000, vol. 374, no. 6, pp. 827–829.
2. Andreeva, I.S., Belan, B.D., Borodulin, A.I., *et al.*, *Dokl. Akad. Nauk*, 2001, vol. 381, no. 2, pp. 278–282.
3. Fuchs, N.A., *Mekhanika aerozolei* (Mechanics of Aerosols), Moscow: Akad. Nauk SSSR, 1955.
4. Borodulin, A.I., Maistrenko, G.A., and Chaldin, B.M., *Statisticheskoe opisanie protsessa turbulentnoi diffuzii aerozolei v atmosfere* (Statistical Description of Turbulent Diffusion of Aerosols in the Atmosphere), Novosibirsk: Novosib. Gos. Univ., 1992.
5. Borodulin, A.I., Desyatkov, B.M., and Kotlyarova, S.S., *Opt. Atmosph. Okeana*, 1997, vol. 10, no. 10, pp. 1241–1246.