On statistics of the tropospheric bioaerosol concentration in Southwestern Siberia

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

* *Institute of Aerobiology,*
State Research Center of Virology and Biotechnology "Vector," Koltsovo, Novosibirsk Region
** *Institute of Atmospheric Optics,*
Siberian Branch of the Russian Academy of Sciences, Tomsk

Received November 28, 2002

We discuss some results obtained during the three-year cycle of measurements of the concentration of total protein and living microorganisms in the tropospheric aerosol in Southwestern Siberia. Analysis of these results shows that the aerosol concentration has a scatter caused by the statistical nature of the aerosol admixtures dispersal. Based on the measurement data available it is shown that the concentration of living microorganisms in the tropospheric aerosol is governed by the discrete Poisson statistics, while the concentration of total protein follows the laws of continual statistics.

The statistics of admixture concentrations in the atmosphere has many times considered and discussed in the literature. However, this issue is still urgent since it is of great fundamental and applied significance. The State Research Center of Virology and Biotechnology "Vector" and Institute of Atmospheric Optics SB RAS are conducting systematic study of the biogenic component of tropospheric aerosol in Southwestern Siberia. We characterize the biogenic component by two its most important parts: total protein, as a basis of all living substance, and viable microorganisms. The experimental technique and our attempts to summarize the obtained data are described in Refs. 1 to 3. One of the features of the obtained array of experimental data is the wide spread in values of the concentration of biogenic component of atmospheric aerosol in Southwestern Siberia. Since this phenomenon cannot be completely explained by measurement errors, its statistical nature should be accepted.

For the general reasons, we can assume that the tropospheric aerosol fraction containing the total protein is formed by particles having a wide size spectrum and includes a very large number of macromolecules with common specific fragments. In these limits, the mechanism of formation of the protein component of tropospheric aerosol can be considered as a result of consecutive fragmentation of some large pattern particles. According to Kolmogorov theorem,[4] such a system should be described, in the limiting case, by the laws of continual statistics, namely, by the lognormal distribution. The atmospheric aerosol component containing viable microorganisms is represented by an ensemble of indivisible particles. Destruction or death of a microorganism removes it from the list of the living. Thus, we should assume that the fraction of viable microorganisms in the tropospheric aerosol is described by the laws of discrete statistics, while the total protein fraction is described by the laws of continual statistics. In this paper, we undertake an attempt to check these hypotheses using the available experimental data.

The lognormal distribution function has the form

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

where C is the argument of the distribution function; erf(...) is the probability integral. The distribution parameters M and Σ can be expressed through the mathematical expectation of the concentration $\overline{C}$ and its variance σ^2 as follows:

$$\overline{C} = \exp\left(M + \frac{\Sigma^2}{2}\right); \quad \sigma^2 = \exp(2M + \Sigma^2)[\exp(\Sigma^2) - 1]. \qquad (2)$$

Let us consider another continual distribution function of the concentration of atmospheric admixtures[5]:

$$F(C) = 1 + \frac{1}{2}\left[\mathrm{erf}\left(\frac{C - \overline{C}}{\beta}\right) - \mathrm{erf}\left(\frac{C + \overline{C}}{\beta}\right)\right], \qquad (3)$$

where β is the second parameter of the distribution function. This equation is the exact analytical solution of the Fokker–Planck–Kolmogorov equation obtained assuming the random process of variation of the atmospheric admixture concentration at a given point of space to be a Markovian process. It is convenient to determine the parameter β from the following equations:

0235-6880/03/05–06 479-04 $02.00

$$\frac{\sigma^2}{\overline{C}^2} = \mathrm{erf}(\beta_0)\left(1+\frac{1}{2\beta_0^2}\right) - 1 + \frac{1}{\sqrt{\pi}\beta_0}\exp(-\beta_0^2);$$

$$\beta_0 = \frac{\overline{C}}{\beta}. \qquad (4)$$

The discrete statistics of the concentration of atmospheric admixtures is considered in Ref. 6, where the binomial distribution for k particles in a given volume is justified based on some natural assumptions. If the number of particles is large enough (in our case, the number of living organisms in 1 m^3 is about 1000 in the order of magnitude), then the binomial distribution can be approximated by the Poisson distribution

$$p(k) = \exp(-\overline{k})(\overline{k})^k/k!, \qquad (5)$$

where $p(k)$ is the probability of observation of k particles in a unit volume; $\overline{k}$ is the mathematical expectation of the number of particles.

Analysis of data obtained has shown that the concentrations of total protein and living organisms on the average are almost independent of the flight height. The possible causes of this effect were discussed in Refs. 1 to 3. Therefore, the array of experimental data on the concentrations of total protein C_p and viable microorganisms C_b was processed in the following way. First, for every experiment we determined the concentrations of total protein C_{pm} and living microorganisms C_{bm} averaged over the observation height h. Then we calculated the values of $\varphi_p = C_p/C_{pm}$ and $\varphi_b = C_b/C_{bm}$ and determined the standard deviations of the normalized concentrations of the total protein $\sigma_{\varphi p}$ and viable microorganisms $\sigma_{\varphi b}$ averaged over the whole array of experimental data. The array of data on the total protein concentration was represented by $n = 245$ experiments, while the data on concentration of viable microorganisms included $n = 197$ experiments.

A characteristic feature of the Poisson distribution is that its variance is equal to mathematical expectation $\sigma_k^2 = \overline{k}$. This equality can be used as a hypothesis that the concentration of total protein and viable microorganisms obey the Poisson statistics. In accordance with the method used for processing of experimental data, the values of φ_p and φ_b are proportional to the number of particles in a sample, and their mathematical expectations are equal to unity. Therefore, we have to check the hypotheses $\overline{\sigma_{\varphi p}^2} = 1$ and $\overline{\sigma_{\varphi b}^2} = 1$, where the overbarred quantity denotes averaging over the array of experimental data.

These values, as well as the values of the statistics

$$T_p = \frac{\overline{\sigma_{\varphi p}^2} - 1}{S_p/\sqrt{m}} \text{ and } T_b = \frac{\overline{\sigma_{\varphi b}^2} - 1}{S_b/\sqrt{m}},$$

which should have the Student t-distribution with the number of degrees of freedom $m = 8$ corresponding to eight heights of concentration measurements, are given in Table 1 along with the standard sample deviations S_p and S_b of $\overline{\sigma_{\varphi p}^2}$ and $\overline{\sigma_{\varphi b}^2}$ and the quantiles of the Student t-distribution $|t_m|_{1-\alpha}$ for $\alpha = 0.05$ and $m = 8$. The absolute value of the statistics T_p is greater than the quantile, while that of the statistics T_b is smaller. We can see that the hypothesis put forward for the distribution of the total protein concentration is rejected, while for the distribution of the concentration of viable microorganisms it can be accepted. The check of this hypothesis is the necessary, but not sufficient condition for the conclusion whether the concentration of viable microorganisms obeys the Poisson statistics or not. Therefore, it is of particular interest to check directly whether the array of data corresponds to the Poisson statistics. In the general case, the value of φ_b is not integer. However, the corresponding estimation can be still carried out. Assuming $k = 0$, $\overline{k}$, $2\overline{k}$, ... and since $C_b = k/V$, where V is the sample volume, we obtain a series of integer values $\varphi_b = 0, 1, 2, \ldots$. It corresponds to the number of microorganisms in 1 m^3 multiple to their mean number. The distribution of φ_b should also obey the Poisson statistics. Figure 1 depicts the probabilities $p(\varphi_b)$ calculated by Eq. (5) and sample frequencies of occurrence of the given φ_b values determined for the ensemble of experiments.

Because of measurement errors, in the available sample of φ_b values it is not always feasible to find strictly zero concentrations. Therefore, when estimating the probability of zero values, we used the range $0 \le \varphi_b \le 0.2$. Table 2 presents the results of checking this hypothesis. The first column gives the ranges of φ_b values. The sample frequencies of occurrence of this event h_i are presented in the second column. The third column presents the theoretical values of the probability of this event, and the fourth column gives the statistics terms χ_i^2. As can be seen, the value of the statistics χ^2 is smaller than the distribution quantile $\chi_{m,1-\alpha}^2$ for $\alpha = 0.05$ and $m = 3$. This means that the hypothesis that the distribution corresponds to the Poisson statistics can be accepted with the confidence probability $1 - \alpha = 0.95$.

Table 1. Check of the hypothesis on equal variance and mathematical expectation of the number of particles

Ensemble averaged mathematical expectations of $\sigma_{\varphi p}^2$ and $\sigma_{\varphi b}^2$	$\overline{\sigma_{\varphi p}^2} \pm S_p = 0.29 \pm 0.32$	$\overline{\sigma_{\varphi p}^2} \pm S_b = 2.24 \pm 1.83$
Statistics of T_p and T_b:	$T_p = -6.28$	$T_b = 1.92$
Quantile of the Student t-distribution $\lvert t_m \rvert_{1-\alpha}$, $\alpha = 0.05$, $m = 8$:	2.31	

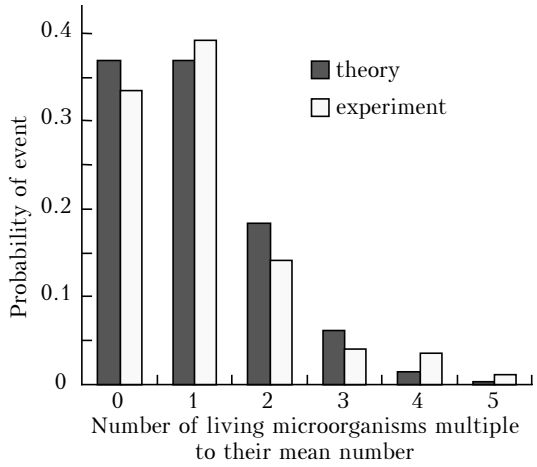

Fig. 1. Theoretical and experimental probability that the number of viable microorganisms in a sample is multiple of their mean number.

Table 2. Check of hypothesis on Poisson distribution of the number of viable microorganisms multiple of their mean number

φ_{bi}	Frequency of occurrence of events h_i	$p(\varphi_{bi})$	$np(\varphi_{bi})$	$\chi_i^2 = \frac{[h_i - np(\varphi_{bi})]^2}{np(\varphi_{bi})}$
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
			Statistics:	$\chi^2 = \sum_i \chi_i^2 = 6.46$
Distribution quantile $\chi^2_{m,1-\alpha}$; $\alpha = 0.05$, $m = 3$:				7.81

The assumption of the continual law of distribution of the atmospheric protein concentration is an alternative hypothesis to the Poisson statistics considered above. Figure 2 depicts empirical distribution histograms of the total protein concentration $F(\varphi_p)$ and the concentration of viable microorganisms $F(\varphi_b)$ drawn based on the ensemble of the experiments, along with distribution functions (1) and (3).

According to the data processing technique used, $\overline{\varphi}_p = \overline{\varphi}_b = 1$. The variances averaged over the array of data $\overline{\sigma^2_{\varphi p}}$ and $\overline{\sigma^2_{\varphi b}}$ are given in Table 1. The parameter β_0 was determined from these values by solving Eq. (4). For the sample of the total protein concentrations, β_0 is equal to 1.3, and for the sample of the concentration of viable microorganisms it is 0.36. The parameters of lognormal distribution (1) calculated by Eq. (2) for the array of data on the total protein concentrations are $M = -0.13$ and $\Sigma = 0.50$. For the concentration of viable microorganisms they are $M = -0.59$ and $\Sigma = 1.08$. It can be seen that the empirical distribution function of the total protein concentration qualitatively corresponds to continual distribution functions (1) and (3). Grouping of the data allows the χ^2-criterion for lognormal distribution (1) to be met at the confidence level $\alpha = 0.01$. These results are presented in Table 3.

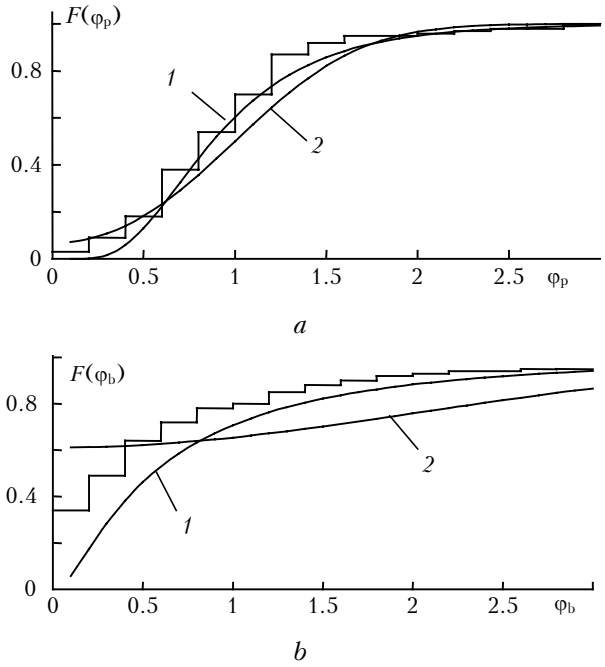

Fig. 2. Empirical distribution functions of the total protein concentration and the concentration of viable microorganisms (histograms) compared with distribution functions (1) and (3) (curves *1* and *2*, respectively).

Table 3. Check of hypothesis on the lognormal distribution of the total protein concentration

φ_{pi}	Frequency of occurrence of events h_i	$p(\varphi_{pi})$	$np(\varphi_{pi})$	$\chi_i^2 = \frac{[h_i - np(\varphi_{pi})]^2}{np(\varphi_{pi})}$
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
			Statistics:	$\chi^2 = \sum_i \chi_i^2 = 10.32$
Distribution quantile $\chi^2_{m,1-\alpha}$; $\alpha = 0.01$, $m = 3$:				11.34

The hypothesis put forward at the confidence level $\alpha = 0.005$ and with grouping of the data allows the criterion χ^2 to be met for distribution law (3) as well. These results are given in Table 4.

Thus, the hypotheses that the distribution function of the total protein concentration obeys laws (1) or (3) are accepted with the confidence probability $1 - \alpha \approx 0.99 - 0.995$. As can be seen from Fig. 2*b*, the empirical distribution function of the concentration of viable microorganisms, to the contrary, does not correspond even qualitatively to continual distribution laws (1) and (3).

Table 4. Check of the hypothesis that the total protein concentration obeys distribution law (3)

φ_{pi}	Frequency of occurrence of events h_i	$p(\varphi_{pi})$	$np(\varphi_{pi})$	$\chi_i^2 = \frac{[h_i - np(\varphi_{pi})]^2}{np(\varphi_{pi})}$
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
			Statistics:	$\chi^2 = \sum_i \chi_i^2 = 14.74$
Distribution quantile $\chi^2_{m,1-\alpha}$; $\alpha = 0.005$, $m = 4$:				14.86

In the future, with compiling more experimental data, an attention should be paid to comparison of continual distribution laws (1) and (3), since they have different physical meaning. The lognormal distribution law rigorously corresponds to the zero probability of observation of zero concentrations $F(0) \equiv 0$. To the contrary, for distribution law (3) the probability of observation of zero concentrations is $F(0) = 1 - \mathrm{erf}(\beta_0)$. In the obtained array of experimental data, we have two strictly zero values of the concentration $\varphi_p = 0$ and four close-to-zero values $0 < \varphi_p \le 0.1$. Thus, the frequency of occurrence of zero concentrations in the ensemble is $F(0) = 0.024$. According to Eq. (3) the theoretically calculated probability of zero values of the concentration is $F(0) = 0.019$. Consequently, we have a good reason to prefer continual distribution law (3) when analyzing the data on the total protein concentration.

The probability of observation of $k = 0$ microorganisms depends on the sample volume and is $p(0) = \exp(-\overline{k})$. In our case, the typical concentration of living microorganisms is $\log_{10}(C_b) \approx 3.7$, which corresponds to $\overline{k} \approx 5000$ in 1 m^3. The volume of collected air samples is roughly 0.75 m^3. The estimates show that the probability of observation of no viable microorganisms in oursamples is negligibly low. At the same time, only 1/50 to 1/250 fractions of a sample are used for seeding onto a nutrient medium. Therefore, the probability of manifestation of $k = 0$ microorganisms in the nutrient medium is about $3 \cdot 10^{-7}$. This probability is very low as well. However, in our samples we still observed the cases of zero number of viable microorganisms. We believe that this situation calls for further analysis.

Thus, the assumption on different physical nature of the statistics of the total protein concentration and the concentration of living microorganisms finds its confirmation. The statistics of concentration of viable microorganisms is described by the discrete law of the Poisson distribution. The statistics of the total protein concentration obeys continual statistics laws (1) or (3). However, for the objective representation of the probabilities of observation of zero concentrations, distribution law (3) should likely be used.

References

1. B.D. Belan, A.I. Borodulin, Yu.V. Marchenko, S.E. Ol'kin, M.V. Panchenko, O.V. P'yankov, A.S. Safatov, and G.A. Buryak, Dokl. Ros. Akad. Nauk **374**, No. 6, 827–829 (2000).
2. I.S. Andreeva, B.D. Belan, A.I. Borodulin, G.A. Buryak, V.A. Zhukov, M.V. Panchenko, V.V. Penenko, V.A. Petrishchenko, and A.S. Safatov, Dokl. Ros. Akad. Nauk **381**, No. 2, 278–282 (2001).
3. I.S. Andreeva, A.I. Borodulin, G.A. Buryak, V.A. Zhukov, S.V. Zykov, V.V. Marchenko, Yu.V. Marchenko, S.E. Ol'kin, V.A. Petrishchenko, O.V. P'yankov, I.K. Resnikova, V.E. Repin, A.S. Safatov, A.N. Sergeev, V.F. Raputa, V.V. Penenko, E.A. Tsvetova, M.Yu. Arshinov, B.D. Belan, M.V. Panchenko, A.N. Ankilov, A.M. Baklanov, A.L. Vlasenko, K.P. Kutsenogii, V.I. Makarov, and T.V. Churkina, Khimiya v Interesakh Ustoichivogo Razvitiya **10**, No. 5, 547–561 (2002).
4. N.A. Fuks, *Mechanics of Aerosols* (AS USSR Publishing House, Moscow, 1955), 351 pp.
5. A.I. Borodulin, G.A. Maistrenko, and B.M. Chaldin, *Statistical Description of the Process of Turbulent Diffusion of Aerosols in the Atmosphere* (Novosibirsk University Publishing House, Novosibirsk, 1992), 124 pp.
6. A.I. Borodulin, B.M. Desyatkov, and S.S. Kotlyarova, Atmos. Oceanic Opt. **10**, No. 10, 777–780 (1997).