

А.И. Бородулин*, А.С. Сафатов*, Б.Д. Белан**, М.В. Панченко**

О статистике концентрации тропосферного биоаэрозоля юга Западной Сибири

*НИИ азробиологии ГНЦ ВБ «Вектор», Кольцово Новосибирской обл.

**Институт оптики атмосферы СО РАН, г. Томск

Поступила в редакцию 28.11.2002 г.

Рассматриваются результаты трехлетнего цикла изучения концентрации суммарного белка и живых микроорганизмов в тропосферном аэрозоле юга Западной Сибири. Анализ результатов измерений показывает, что концентрация таких аэрозолей имеет разброс, обусловленный статистической природой рассеяния атмосферных примесей. На основании имеющегося ансамбля данных измерений показано, что концентрация живых микроорганизмов в тропосферном аэрозоле подчиняется дискретной статистике Пуассона, а концентрация суммарного белка подчиняется законам непрерывной статистики.

Вопросы, касающиеся статистики распределения концентрации атмосферных примесей, неоднократно рассматривались и обсуждались в научных публикациях. Несмотря на это, данная проблема до сих пор остается актуальной, поскольку имеет большое фундаментальное и прикладное значение. ГНЦ ВБ «Вектор» и ИОА СО РАН проводят систематическое изучение биогенной компоненты тропосферного аэрозоля юга Западной Сибири. «Биогенная компонента» характеризуется нами двумя наиболее важными составляющими: суммарным белком, как основой всего живого, и жизнеспособными микроорганизмами. Описание методов проведения экспериментов, результаты проведенных исследований и попытки обобщения полученных данных представлены в [1–3]. Одной из особенностей полученного ансамбля экспериментальных данных является то, что значения концентрации биогенной компоненты атмосферного аэрозоля юга Западной Сибири имеют значительный разброс. Поскольку этот эффект не может быть полностью объяснен погрешностями измерений, следует сделать вывод о статистической природе данного эффекта.

Исходя из общих соображений, можно предположить, что фракция тропосферного аэрозоля, содержащая суммарный белок, формируется из частиц широкого спектра размеров и содержит очень большое количество макромолекул с общими для них специфичными фрагментами. В пределе механизм формирования белковой компоненты тропосферного аэрозоля можно считать результатом последовательного дробления некоторых достаточно крупных «исходных» частиц. Согласно теореме А.Н. Колмогорова [4], такая система должна в пределе описываться законами непрерывной статистики, а именно логарифмически нормальным распределением. Компонента атмосферного аэрозоля, содержащая жизнеспособные микроорганизмы, представлена ансамблем неделимых субъединиц. Разрушение, гибель микроорганизма вычеркивают

его из «списка живых». Таким образом, следует предполагать, что фракция жизнеспособных микроорганизмов тропосферного аэрозоля должна описываться законами дискретной статистики, а фракция суммарного белка — законами непрерывной статистики. В данной статье предпринимается попытка проверки этих гипотез на имеющемся массиве экспериментальных данных.

Логарифмически нормальная функция распределения имеет вид

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

где C — аргумент функции распределения; $\operatorname{erf}(\dots)$ — интеграл вероятности. Параметры закона распределения M и Σ выражаются через математическое ожидание концентрации \bar{C} и ее дисперсию σ^2 следующим образом:

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

Ниже рассмотрим еще одну непрерывную функцию распределения концентрации атмосферных примесей [5]:

$$F(C) = 1 + \frac{1}{2} \left[\operatorname{erf} \left(\frac{C - \bar{C}}{\beta} \right) - \operatorname{erf} \left(\frac{C + \bar{C}}{\beta} \right) \right], \quad (3)$$

где β — второй параметр функции распределения. Данное выражение является точным аналитическим решением уравнения Фоккера–Планка–Колмогорова и получено в предположении, что случайный процесс изменения концентрации атмосферной примеси в данной точке пространства является марковским. Определение параметра β удобно проводить по соотношению

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

$$\beta_0 = \frac{\bar{C}}{\bar{\rho}}. \quad (4)$$

Дискретная статистика концентрации атмосферных примесей рассмотрена в [6], где на основании естественных предположений обосновано биномиальное распределение для k частиц, находящихся в заданном объеме. Если число частиц достаточно велико, а в нашем случае число живых микроорганизмов в кубическом метре по порядку величины составляет тысячи, то биномиальное распределение аппроксимируется распределением Пуассона

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

где $p(k)$ – вероятность наблюдения k частиц в единице объема; \bar{k} – математическое ожидание числа частиц.

Анализ полученных данных показал, что концентрация суммарного белка и жизнеспособных микроорганизмов в среднем практически не зависит от высоты отбора проб. Возможные причины этого эффекта обсуждались в [1–3]. Поэтому ансамбль экспериментальных данных о концентрации суммарного белка C_p и концентрации жизнеспособных микроорганизмов C_b обрабатывался следующим образом. Сначала для каждого эксперимента находились средние по высотам наблюдений h концентрации суммарного белка C_{pm} и живых микроорганизмов C_{bm} . Затем вычислялись значения $\varphi_p = C_p / C_{pm}$, $\varphi_b = C_b / C_{bm}$ и определялись стандартные отклонения нормированных концентраций суммарного белка $\sigma_{\varphi p}$ и жизнеспособных микроорганизмов $\sigma_{\varphi b}$, усредненные по всему ансамблю проведенных экспериментов. Ансамбль концентраций суммарного белка был представлен $n = 245$ экспериментами, а ансамбль концентраций жизнеспособных микроорганизмов $n = 197$ экспериментами.

Распределение Пуассона характерно тем, что его дисперсия равна математическому ожиданию $\sigma_k^2 = \bar{k}$. Это соотношение можно выдвинуть в качестве гипотезы о соответствии функций распределения концентрации суммарного белка и жизнеспособных микроорганизмов статистике Пуассона. В соответствии с использованным методом обработки экспериментальных данных значения φ_p и φ_b про-

порциональны количеству частиц в пробе, а их математические ожидания равны единице. Поэтому в нашем случае проверке подлежат гипотезы: $\overline{\sigma_{\varphi p}^2} = 1$ и $\overline{\sigma_{\varphi b}^2} = 1$, где черта означает усреднение по ансамблю экспериментальных данных. Эти величины, а также значения статистик

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

которые должны иметь t -распределение Стьюдента с числом степеней свободы $m = 8$, соответствующим восьми высотам измерения концентрации, приведены в табл. 1. Там же даны стандартные выборочные отклонения S_p , S_b величин $\overline{\sigma_{\varphi p}^2}$, $\overline{\sigma_{\varphi b}^2}$ и квантили t -распределения Стьюдента $|t_m|_{1-\alpha}$ для $\alpha = 0,05$ и $m = 8$. Модуль значения статистики T_p больше указанного квантиля, а статистики T_b – меньше. Мы видим, что выдвинутая гипотеза для распределения концентрации суммарного белка отвергается, а для распределения концентрации жизнеспособных микроорганизмов – принимается. Проверка данной гипотезы является необходимым, но не достаточным условием для вывода о принадлежности закона распределения концентрации жизнеспособных микроорганизмов статистике Пуассона. Поэтому прямая проверка соответствия ансамбля данных статистике Пуассона (5) представляет особый интерес. В общем случае рассматриваемая нами величина φ_b не является целочисленной. Однако соответствующую оценку все-таки можно произвести. Полагая $k = 0, \bar{k}, 2\bar{k}, \dots$, и поскольку $C_b = k/V$, где V – рассматриваемый объем пробы, получим ряд целочисленных значений $\varphi_b = 0, 1, 2, \dots$. Он соответствует количеству микроорганизмов в 1 м^3 , кратному их среднему числу. Распределение величины φ_b также должно подчиняться статистике Пуассона. На рис. 1 приведены вероятности $p(\varphi_b)$, рассчитанные по (5), и выборочные частоты появления заданных значений φ_b , найденные по ансамблю проведенных экспериментов.

Ввиду наличия погрешностей измерений в имеющейся выборке φ_b не всегда можно выделить строго нулевые значения концентрации. Поэтому при оценке вероятности появления нулевых значений нами был задан интервал $0 \leq \varphi_b \leq 0,2$. Табл. 2 содержит результаты проверки данной гипотезы.

Таблица 1

Проверка гипотезы о равенстве дисперсии и математического ожидания количества частиц

Усредненные по ансамблю наблюдений математические ожидания величин $\sigma_{\varphi p}^2$ и $\sigma_{\varphi b}^2$	$\overline{\sigma_{\varphi p}^2} \pm S_p = 0,29 \pm 0,32$	$\overline{\sigma_{\varphi b}^2} \pm S_b = 2,24 \pm 1,83$
Статистики T_p и T_b	$T_p = -6,28$	$T_b = 1,92$
Квантиль t -распределения Стьюдента $ t_m _{1-\alpha}$; $\alpha = 0,05$, $m = 8$	2,31	

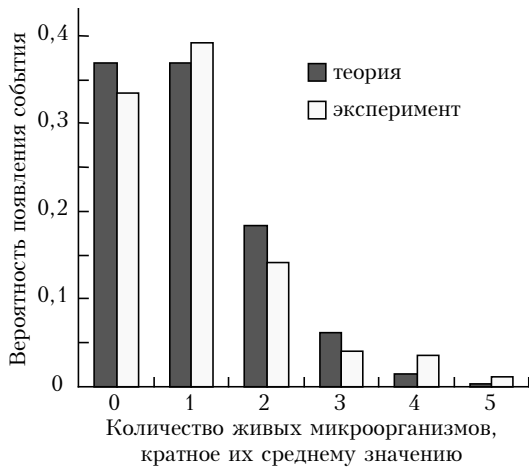


Рис. 1. Теоретические и экспериментально найденные вероятности того, что количество жизнеспособных микроорганизмов в пробе кратно их среднему количеству

Таблица 2

Проверка гипотезы о распределении количества жизнеспособных микроорганизмов, кратного их среднему значению, по закону Пуассона (5)

ϕ_{bi}	Частота появления событий h_i	$p(\phi_{bi})$	$np(\phi_{bi})$	$\chi_i^2 = \frac{[h_i - np(\phi_{bi})]^2}{np(\phi_{bi})}$
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
Статистика				$\chi^2 = \sum_i \chi_i^2 = 6,46$
Квантиль распределения $\chi_{m,1-\alpha}^2$; $\alpha = 0,05, m = 3$				7,81

Первая колонка представляет интервалы значений ϕ_b . Во второй приведены выборочные частоты появления данного события h_i . Третья представляет теоретические значения вероятности появления данного события. В четвертой представлены слагаемые статистики χ_i^2 . Как видно, значение статистики χ^2 меньше квантиля распределения $\chi_{m,1-\alpha}^2$ для $\alpha = 0,05$ и $m = 3$. Это означает, что гипотеза о соответствии данного распределения статистике Пуассона принимается с доверительной вероятностью $1 - \alpha = 0,95$.

Предположение о непрерывности закона распределения концентрации атмосферного белка является альтернативной гипотезой к рассмотренной выше статистике Пуассона. На рис. 2 даны эмпирические гистограммы функций распределения концентрации суммарного белка $F(\phi_p)$ и концентрации жизнеспособных микроорганизмов $F(\phi_b)$, построенные на основании ансамбля проведенных экспериментов. Там же приведены функции распределения (1) и (3).

Согласно использованной методике обработки данных $\bar{\phi}_p = \bar{\phi}_b = 1$. Дисперсии, усредненные по ансамблю данных $\sigma_{\phi_p}^2$ и $\sigma_{\phi_b}^2$, представлены в табл. 1.

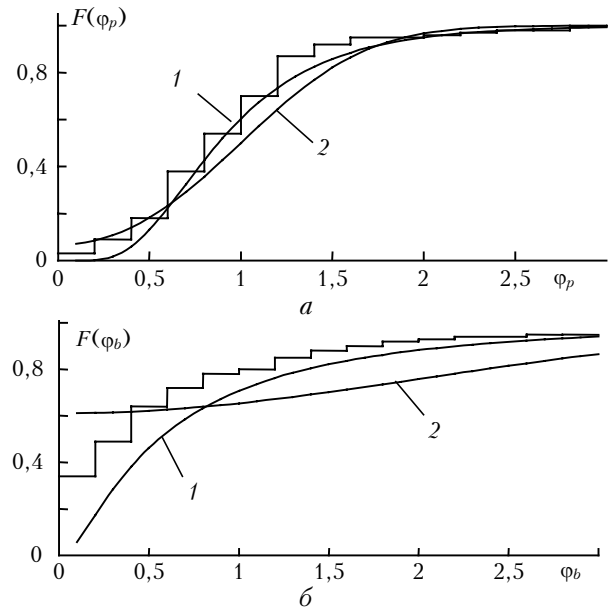


Рис. 2. Эмпирические функции распределения концентрации суммарного белка и концентрации жизнеспособных микроорганизмов (гистограммы) в сравнении с функциями распределения (1) и (3), кривые 1 и 2 соответственно

Параметр β_0 определялся по этим значениям решением уравнения (4). Для выборки концентраций суммарного белка β_0 равно 1,3, а для выборки концентраций жизнеспособных микроорганизмов – 0,36. Параметры логарифмически нормального закона распределения (1), вычисленные по (2) для массива данных о концентрации суммарного белка, составляют $M = -0,13$ и $\Sigma = 0,50$. Для концентрации жизнеспособных микроорганизмов $M = -0,59$ и $\Sigma = 1,08$. Видно, что эмпирическая функция распределения концентрации суммарного белка качественно соответствует непрерывным функциям распределения (1) и (3). Группировка данных позволяет на уровне значимости $\alpha = 0,01$ удовлетворить критерию χ^2 для логарифмически нормального распределения (1). Эти результаты приведены в табл. 3.

Таблица 3

Проверка гипотезы о распределении концентрации суммарного белка по логарифмически нормальному закону (1)

ϕ_{pi}	Частота появления событий h_i	$p(\phi_{pi})$	$np(\phi_{pi})$	$\chi_i^2 = \frac{[h_i - np(\phi_{pi})]^2}{np(\phi_{pi})}$
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
Статистика				$\chi^2 = \sum_i \chi_i^2 = 10,32$
Квантиль распределения $\chi_{m,1-\alpha}^2$; $\alpha = 0,01, m = 3$				11,34

Выдвинутая гипотеза на уровне значимости $\alpha = 0,005$ и при группировке данных позволяет удовлетворить критерию χ^2 и для закона распределения (3). Эти результаты приведены в табл. 4.

Таблица 4

Проверка гипотезы о распределении концентрации суммарного белка по закону (3)

Φ_{pi}	Частота появления событий h_i	$p(\Phi_{pi})$	$np(\Phi_{pi})$	$\chi_i^2 = \frac{[h_i - np(\Phi_{pi})]^2}{np(\Phi_{pi})}$
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

$$\text{Статистика} \quad \chi^2 = \sum_i \chi_i^2 = 14,74$$

Квантиль распределения $\chi_{m,1-\alpha}^2$;

$$\alpha = 0,005, m = 4$$

$$14,86$$

Таким образом, гипотезы о том, что функция распределения концентрации суммарного белка соответствует законам (1) или (3), принимаются с доверительной вероятностью $1 - \alpha \approx 0,99 - 0,995$. В противоположность вышесказанному эмпирическая функция распределения концентрации жизнеспособных микроорганизмов, как это видно из рис. 2,б, даже качественно не соответствует непрерывным законам распределения (1) и (3).

В дальнейшем, по мере накопления экспериментальных данных, следует обратить внимание на взаимное сравнение непрерывных законов распределения концентрации (1) и (3), поскольку в них заложен различный физический смысл. Логарифмически нормальный закон распределения строго соответствует нулевой вероятности наблюдения нулевых значений концентрации $F(0) \equiv 0$. В противоположность этому вероятность наблюдения нулевых значений концентрации для закона распределения (3) составляет $F(0) = 1 - \text{erf}(\beta_0)$. В полученном массиве экспериментальных данных мы имеем два строго нулевых значения концентрации $\Phi_p = 0$ и четыре достаточно близких к нулю значения $0 < \Phi_p \leq 0,1$. Таким образом, частота появления «нулевых» значений концентрации в ансамбле соответствует $F(0) = 0,024$. Согласно (3) теоретически вычисленная вероятность появления нулевых значений концентрации $F(0)$ составляет 0,019. Следовательно, имеются достаточно веские основания при анализе данных о концентрации суммарного белка отдать предпочтение непрерывному закону распределения (3).

Вероятность наблюдения $k = 0$ микроорганизмов зависит от объема пробы и составляет $p(0) = \exp(-\bar{k})$. В нашем случае типичная концентрация живых микроорганизмов равна $\log_{10}(C_b) \approx 3,7$, что соответствует $\bar{k} \approx 5000$ шт. в 1 м^3 . Объем отобранных проб воздуха составляет примерно $0,75 \text{ м}^3$.

A.I. Borodulin, A.S. Safatov, B.D. Belan, M.V. Panchenko. On the statistics of tropospheric aerosol concentration in Southwestern Siberia.

The results of a three-year series of measuring the concentration of total protein and living microorganisms in the tropospheric aerosol in Southwestern Siberia are considered. The analysis of the results of measurements shows that the aerosol concentration has a spread caused by statistical nature of diffusion of aerosol admixtures. It was shown on the basis of the existing ensemble of measured data that the concentration of living microorganisms in the tropospheric aerosol is governed by the Poisson discrete statistics while the concentration of total protein follows the laws of continual statistics.

Оценки показывают, что вероятность наблюдения нулевого количества жизнеспособных микроорганизмов в наших пробах должна быть пренебрежимо мала. Вместе с тем при высеве каждого образца на питательную среду используется 1/50–1/250 часть пробы. Поэтому вероятность проявления на питательной среде $k = 0$ микроорганизмов составляет порядка $3 \cdot 10^{-7}$. Эта вероятность также очень мала. Однако в наших пробах все-таки наблюдаются случаи нулевого количества жизнеспособных микроорганизмов. Мы считаем, что этот вопрос требует дальнейшего анализа.

Таким образом, предположение о различной физической природе статистики концентрации суммарного белка и концентрации жизнеспособных микроорганизмов находит свое подтверждение. Статистика концентрации жизнеспособных микроорганизмов описывается дискретным законом – распределением Пуассона. Статистика концентрации суммарного белка подчиняется законам непрерывной статистики (1) или (3). Однако для объективного представления вероятностей наблюдения нулевых значений концентрации следует, по видимому, использовать закон распределения (3).

1. Белан Б.Д., Бородулин А.И., Марченко Ю.В., Олькин С.Е., Панченко М.В., Пьянков О.В., Сафатов А.С., Бурак Г.А. Изучение изменчивости белковой компоненты атмосферного аэрозоля над лесными массивами юга Западной Сибири // Докл. РАН. 2000. Т. 374. № 6. С. 827–829.
2. Андреева И.С., Белан Б.Д., Бородулин А.И., Бурак Г.А., Жуков В.А., Панченко М.В., Пененко В.В., Петрищенко В.А., Сафатов А.С. Изучение изменчивости содержания живых микроорганизмов в атмосферном аэрозоле на юге Западной Сибири // Докл. РАН. 2001. Т. 381. № 2. С. 278–282.
3. Андреева И.С., Бородулин А.И., Бурак Г.А., Жуков В.А., Зыков С.В., Марченко В.В., Марченко Ю.В., Олькин С.Е., Петрищенко В.А., Пьянков О.В., Резникова И.К., Решин В.Е., Сафатов А.С., Сергеев А.Н., Рапуца В.Ф., Пененко В.В., Цветова Е.А., Аршинов М.Ю., Белан Б.Д., Панченко М.В., Анкилов А.Н., Бакланов А.М., Власенко А.Л., Куценогий К.П., Макаров В.И., Чуркина Т.В. Биогенная компонента атмосферного аэрозоля на юге Западной Сибири // Химия в интересах устойчивого развития. 2002. Т. 10. № 5. С. 547–561.
4. Фукс Н.А. Механика аэрозолей. М.: Изд-во АН СССР, 1955. 351 с.
5. Бородулин А.И., Майстренко Г.А., Чалдин Б.М. Статистическое описание процесса турбулентной диффузии аэрозолей в атмосфере. Новосибирск: Изд-во Новосиб. ун-та, 1992. 124 с.
6. Бородулин А.И., Десятков Б.М., Котлярова С.С. Дискретная функция распределения числа аэрозольных частиц в единице объема // Оптика атмосфер и океана. 1997. Т. 10. № 10. С. 1241–1246.