

В.И. Цанев, Д.Л. Вындев

МЕТОД ПОСТРОЕНИЯ РЕГРЕССИОННОЙ ЗАВИСИМОСТИ МЕЖДУ АЭРОЗОЛЬНЫМ ОБЪЕМНЫМ КОЭФФИЦИЕНТОМ РАССЕЯНИЯ И МАССОВОЙ КОНЦЕНТРАЦИЕЙ АЭРОЗОЛЯ С УЧЕТОМ ОШИБОК ИЗМЕРЕНИЙ ОБЕИХ ВЕЛИЧИН

Предложен новый метод построения линейной регрессии, учитывающий ошибки измерений массовой концентрации и объемного коэффициента рассеяния атмосферного аэрозоля одночастотным лидаром. Приведен пример использования предложенного подхода в системах контроля за загрязнением воздуха.

Определение массовой концентрации аэрозоля M [$\text{мг} \cdot \text{м}^{-3}$] по измеренному лидаром объемному коэффициенту аэрозольного рассеяния α [км^{-1}] является задачей первостепенной важности для применения одночастотных лидаров в системах контроля за загрязнением воздуха. Возможность успешного решения этой задачи обоснована многочисленными исследованиями, подтверждающими наличие линейной корреляционной зависимости

$$M = C_1\alpha + C_2. \quad (1)$$

Обсуждались разные методы определения в (1) коэффициентов C_1 и C_2 [1, 2]. Первый метод предполагает экспериментальное определение функции распределения аэрозольных частиц по размерам dN/dr , их плотности ρ и комплексного показателя преломления m . Затем проводятся расчеты M и α и строится регрессионная зависимость (1). Второй метод сводится к прямому решению уравнения лазерного зондирования относительно M при задании модельных dN/dr , ρ и m . При третьем подходе M измеряется забором проб воздуха, а α — лидаром. Таким образом, получается выборка экспериментальных данных

$$\{M_i, \alpha_i; i = \overline{1, N}\}, \quad (2)$$

которая используется для построения регрессии (1).

Здесь рассматривается третий подход, который решает поставленную задачу только методами статистики и не требует никаких модельных представлений о микрофизике зондируемого лидаром аэрозоля.

Прежде всего нужно обсудить метод получения выборки (2). Массовая концентрация аэрозоля M обычно определяется отбором проб на мембранные фильтры с последующим взвешиванием [3]. Этой процедуре присущи ошибки, обусловленные природой взаимодействия аэрозоля с мембранными фильтрами, неточностями определения объема воздушной пробы и взвешивания фильтра, необходимостью сохранять фильтры после забора проб. Временная длительность отбора пробы зависит от состояния замутненности атмосферы и от допустимой точности последующего взвешивания. Она варьирует от десятка минут до нескольких часов. Так как аэрозоль переносится ветром, то одно измерение M дает усредненное значение массовой концентрации за время по объему ΔV . Для повышения точности измерения M целесообразно производить одновременный забор проб на несколько мембранных фильтров. Лидар определяет мгновенное распределение α по трассе зондирования. Чтобы получить выборку (2), необходимо усреднить результаты многократных лидарных измерений α , проведенных за время Δt по направлениям, проходящим через объем ΔV . В результате получается выборка

$$\{M_i, \delta M_i, \alpha_i, \delta \alpha_i; i = \overline{1, N}\},$$

где индекс i указывает на то, что измерения проведены в течение Δt_i , а σ означает стандартное отклонение. Можно принять, что измерения M и α независимы и что их ошибки распределены по гауссовым законам $N(M_i, \delta M_i)$ и $N(\alpha_i, \delta \alpha_i)$. Таким образом, получается следующая статистическая модель

$$\{\mathbf{x}_i, \hat{S}_i, i = \overline{1, N}\}, \quad (3)$$

где \mathbf{x}_i — математическое ожидание, $\mathbf{x}_{1i} = \alpha_i$, $\mathbf{x}_{2i} = M_i$; \hat{S}_i — диагональная ковариационная матрица ошибок измерений с элементами $S_{12} = (\delta \alpha_i)^2$ и $S_{2i} = (\delta M_i)^2$. Из способа получения выборки (2) ясно, что нет никаких оснований принять для любого i $S_{1i} = 0$ или $S_{2i} = 0$. Более того, значения элементов

S_i сильно варьируют от одного интервала Δt к другому. Это означает, что традиционные методы построения линейной регрессии неприменимы к модели (3) [4, 5]. Здесь обсуждается метод построения линейной регрессии для модели (3).

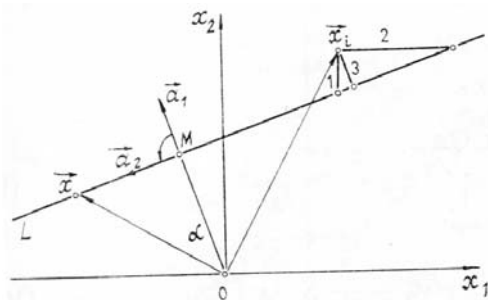


Рис. 1. Расстояния, используемые для построения линейной регрессии. Регрессионная прямая задается нормальным уравнением

На рис. 1 показана гипотетическая прямая регрессии L и одна точка x_i из выборки (3). При построении регрессии « x_2 на x_1 » минимизируется сумма квадратов расстояний от x_i до L . На рисунке этим расстояниям соответствует отрезок 1. Отрезок 2 соответствует расстояниям, используемым для построения регрессии « x_1 на x_2 ». Существует еще одно расстояние от x_i до L — отрезок 3, т. е. расстояние от x_i до L по перпендикуляру. Из дальнейшего рассмотрения будет видно, что отрезок 3 является наилучшим возможным выбором расстояния, если $S_i = E$, т.е. в случаях, когда все ошибки всех измерений либо нули, либо одинаковы.

Прямая регрессии определяется расстоянием $\alpha = OM$ от прямой до начала координат и вектором a_1 . Векторы a_1 и a_2 образуют правую систему координат, причем $a_1^T a_1 = a_2^T a_2 = 1$, $a_1^T a_2 = 0$, где буква „Т” означает транспонирование, т.е. a_1 есть вектор—столбец, а a_1^T — вектор—строка. Естественно считать, что вокруг любой точки x_i расстояние определяется с помощью \hat{S}_i , так как апостериорная вероятность получения x вокруг x_i задается как

$$f(x, x_i) = \frac{1}{\sqrt{2\pi \det(\hat{S}_i)}} \exp\left\{-\frac{1}{2} (x - x_i)^T \hat{S}_i^{-1} (x - x_i)\right\}. \quad (4)$$

Используя метод максимального правдоподобия и неравенство Камера—Рао, можно показать, что среднее \bar{x} является средним взвешенным индивидуальных наблюдений:

$$\bar{x} = \hat{S}^{-1} \sum_{i=1}^N \hat{S}_i^{-1} x_i, \quad (5)$$

$$\hat{S}^{-1} = \left(\sum_{i=1}^N \hat{S}_i^{-1}\right)^{-1}, \quad (6)$$

и представляет собой ковариационную матрицу x , а расстояние p от x_i до L

$$p = a_1^T \hat{S}_i^{-1} x_i - \alpha. \quad (7)$$

Теперь задача о построении линейной регрессии сводится к задаче об условном экстремуме

$$\min \sum_{i=1}^N (a_1^T \hat{S}_i^{-1} x_i - \alpha)^2; \quad a_1^T a_1 = 1, \quad (8)$$

Методом Лагранжа из (8) получается

$$\min \{a_1^T \hat{Q} a_1\}, \quad a_1^T a_1 = 1, \quad (9)$$

где

$$\hat{Q} = \sum_{i=1}^N \left(\hat{S}_i^{-1} \mathbf{x}_i - \frac{1}{N} \hat{S}^{-1} \bar{\mathbf{x}} \right) \left(\hat{S}_i^{-1} \mathbf{x}_i - \frac{1}{N} \hat{S}^{-1} \bar{\mathbf{x}} \right)^T, \quad (10)$$

т.е. прямая L определяется собственным вектором, соответствующим меньшему собственному числу \hat{Q} . Можно показать, что \hat{Q} имеет два различных неотрицательных собственных числа λ_1 и λ_2 . Пусть направление \mathbf{a}_1 выбирается так, как показано на рис. 1. Это означает, что \mathbf{a}_1 определяется в зависимости от знака Q_{11} так, чтобы

$$\alpha = N^{-1} \mathbf{a}_1^T \hat{S}^{-1} \bar{\mathbf{x}} \geq 0. \quad (11)$$

Если $\hat{S}_i = \hat{E}$, то \hat{Q} с точностью до множителя $(N-1)$ равно ковариационной матрице \hat{C}_x , а условие $p = 0$ — нормальное уравнение прямой регрессии L . Расстояние от \mathbf{x}_i к L в этом случае представляет отрезок 3 на рис. 1.

Коэффициенты регрессии выражаются через α и \mathbf{a}_1 :

$$c_1 = -a_{11}/a_{12}; \quad c_2 = \alpha/a_{12} \quad (12)$$

Математическое ожидание \bar{c} и ковариационную матрицу \hat{C}_c вектора $\mathbf{c} = (c_1, c_2)$ можно получить методом статистических испытаний, решая многократно задачу для ряда имитированных выборок

$$\{\mathbf{x}'_i, \hat{S}'_i, i = 1, N\}, \quad \mathbf{x}'_i = \bar{\mathbf{x}} + \sqrt{\lambda_1} a_{11} \xi_{1i} + \sqrt{\lambda_2} a_{21} \xi_{2i}, \quad (13)$$

где $\xi = (\xi_{1i}, \xi_{2i})$ — двумерная гауссова случайная величина $N(0, \hat{E})$. Можно показать, что $\{\mathbf{x}_i\}$ и $\{\mathbf{x}'_i\}$ имеют одинаковые математические ожидания и ковариационные матрицы.

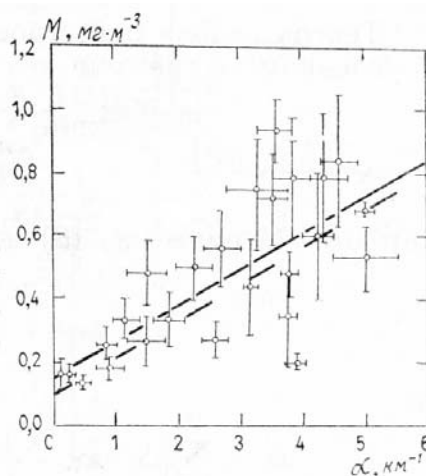


Рис. 2. Результаты измерений M и α : \circ — средние значения M и α в течение временного интервала Δt ; I — стандартное отклонение M за время Δt ; $-$ — стандартное отклонение α за время Δt . Прямая регрессии с коэффициентами (15) показана сплошной линией, а прямая регрессии с коэффициентами (16) — штриховой

Рассмотрим пример применения предложенного метода построения прямой регрессии. На рис. 2 показаны результаты измерений $\{\alpha_i, \delta\alpha_i, M_i, \delta M_i; i = 1, 25\}$. Если игнорировать ошибки измерений, то стандартным методом [3] для регрессий M на α и α на M получается

$$\begin{aligned} c'_1 &= (0,125 \pm 0,021) \text{ мг} \cdot \text{м}^{-3} \cdot \text{км}^{-1}; \\ c'_2 &= (0,136 \pm 0,063) \text{ мг} \cdot \text{м}^{-3}; \\ c''_1 &= (4,777 \pm 0,818) \text{ мг} \cdot \text{м}^{-3} \cdot \text{км}^{-1}; \\ c''_2 &= (0,367 \pm 0,422) \text{ мг} \cdot \text{м}^{-3}. \end{aligned} \quad (14)$$

Для рассматриваемого набора экспериментальных данных из (10) находим

$$\hat{Q} = \begin{pmatrix} 56,937 & 6,498 \\ 6,498 & 1,396 \end{pmatrix}.$$

Меньшему собственному числу $\lambda_1 = 0,6457$ соответствует собственный вектор

$$\begin{aligned} a_1 &= (-0,1147 + 0,9934); \\ c_1 &= 0,149 \text{ мг} \cdot \text{м}^{-3} \cdot \text{км}^{-1}; \\ c_2 &= 0,115 \text{ мг} \cdot \text{м}^{-3}. \end{aligned} \tag{15}$$

В результате 500 статистических испытаний имитированной выборки $\{x'_{ik}; i = 1,25; k = 1,500\}$ определены следующие статистические характеристики вектора:

$$\begin{aligned} \bar{c} &= (0,116; 0,099); \\ \hat{C}_c &= 10^{-4} \begin{pmatrix} 3,913 & -8,927 \\ -8,927 & 28,031 \end{pmatrix}, \end{aligned} \tag{16}$$

На рис. 2 регрессия с коэффициентами (15) показана сплошной линией, а результаты, полученные с помощью статистических испытаний, — штриховой.

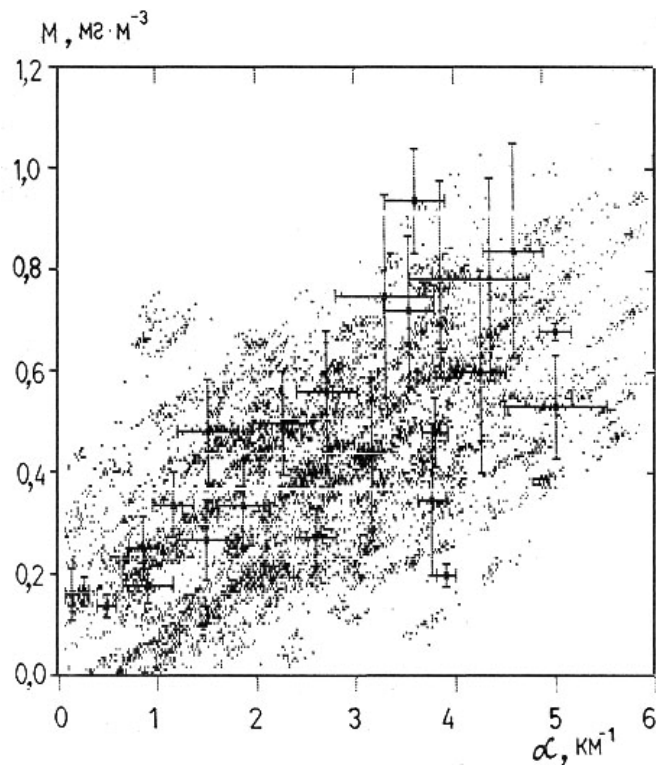


Рис. 3. Результаты измерений M и α и 500 имитированных с помощью (13) выборок $\{x'_k, i = 1,25; k = 1,500\}$

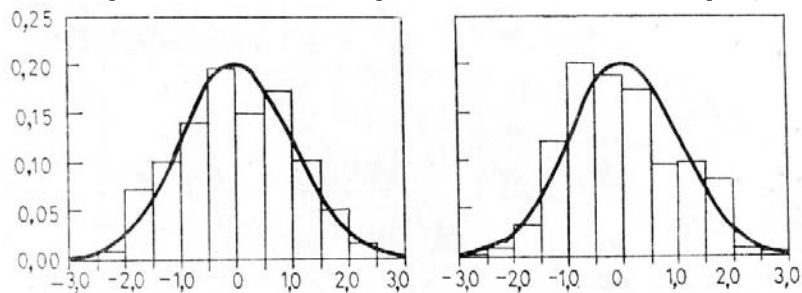


Рис. 4. Гистограммы нормализованных коэффициентов \bar{c}_1 и \bar{c}_2 . Для сравнения показана функция гауссового распределения $N(0, 1)$, пересчитанная с учетом ширины разрядов статистического ряда

На рис. 3 показаны одновременно результаты измерений и 500 имитированных с помощью (13) выборок $\{x'_i; i = \overline{1, 25}; k = \overline{1, 500}\}$. На рис. 4 показаны гистограммы для нормализованных коэффициентов

$$\begin{aligned}\tilde{c}_{1i} &= (c_{1i} - \bar{c}_1) / \sqrt{c_{11}}; \\ \tilde{c}_{2i} &= (c_{2i} - \bar{c}_2) / \sqrt{c_{22}}.\end{aligned}$$

Для сравнения показана функция гауссового распределения $N(0,1)$, пересчитанная с учетом ширины разрядов статистического ряда.

1. Лазерный контроль атмосферы. / Под ред. Э.Д. Хинклн. — М: Мир, 1979. — 416 с.
2. Зуев В.Е., Кауль Б.В., Самохвалов И.В. и др. Лазерное зондирование промышленных аэрозолей. — Новосибирск: Наука, 1986. — 188 с.
3. Временные методические указания по определению содержания примесей в атмосфере. / Под ред. М.Е. Берлянда, Н.Ш. Вольберга. — Л.: Гидрометеоздат, 1971. — 120 с.
4. Бранд З. Статистические методы анализа наблюдений. — М.: Мир, 1985. — 312 с.
5. Kendall M., Stuart A. The Advanced Theory of Statistics. Vol. 2. Inference and Relationship. London, Charles Griffin, 1979. — 748 p.

Институт электроники
Болгарской Академии наук,
София, НРБ
Институт математики
Болгарской Академии наук,
София, НРБ

Поступила в редакцию
5 ноября 1987 г.

V.J. Tsanev, D.L. Vyn dev. **Method for the Construction of the Linear Regression Dependence of the Aerosol Volume Scattering Coefficient on Atmospheric Aerosol Mass Concentration Assuming Measurement Errors.**

A new method for the construction of the linear regression accounting for the experimental errors involved in the atmospheric aerosol mass concentration and volume scattering coefficient measured by a single-frequency lidar system is reported. The application of the proposed approach to the monitoring of air pollution is discussed.