

УДК 519.233.5

Н.Н. Щелканов

НОВЫЙ МЕТОД РАЗДЕЛЕНИЯ ФИЗИЧЕСКОЙ ВЕЛИЧИНЫ НА ДВЕ КОМПОНЕНТЫ

Предложен новый метод нахождения коэффициентов линейной регрессии уравнения $Z = K_0 + K_1X + K_2Y$ с учетом случайных погрешностей физических величин Z , X и Y . Метод позволяет получать устойчивые и физически корректные коэффициенты регрессии K_0 , K_1 и K_2 вне зависимости от используемого массива данных.

Ключевые слова: итерационный метод, линейная регрессия, случайные погрешности.

Обычно для разделения физической величины Z на две компоненты (X и Y) используется статистический метод, основанный на линейном множественном регрессионном анализе массива экспериментальных данных [1]. В этом случае уравнение регрессии для величины Z записывается в виде

$$Z = K_0 + K_1X + K_2Y, \quad (1)$$

где K_0 , K_1 , K_2 – эмпирические коэффициенты, определяемые по статистическим характеристикам массивов Z , X , и Y .

Однако статистический метод, основанный на линейном множественном регрессионном анализе, имеет существенные недостатки. Во-первых, даже при отсутствии случайных погрешностей во входных величинах X и Y коэффициенты регрессии зависят от их диапазона изменчивости. Во-вторых, наличие случайных погрешностей во входных величинах X и Y приводит к занижению коэффициентов регрессии K_1 , K_2 и завышению K_0 , причем чем больше эти величины случайных погрешностей, тем меньше K_1 , K_2 и больше K_0 .

В [2–4] представлена обобщенная формула, позволяющая находить коэффициенты регрессии линейного уравнения $Y = K_0 + K_1X$ для общего случая, когда разброс точек в корреляционной связи величин X и Y обусловлен как их случайными погрешностями измерений, так и неконтролируемыми физическими факторами. Формула имеет вид

$$K_1 = \frac{\sigma_Y}{\sigma_X} \frac{B}{A} \frac{1}{2\rho_{XY}} \left\{ \left(\frac{A}{B} - \frac{B}{A} \right) + \sqrt{\left(\frac{A}{B} - \frac{B}{A} \right)^2 + 4\rho_{XY}^2} \right\}, \quad (2)$$

$$\text{где } A = \sqrt{1 - |\rho_{XY}|} \sqrt{\frac{1 - \delta_X^2 / \sigma_X^2}{1 - \delta_Y^2 / \sigma_Y^2}}, \quad B = \sqrt{1 - |\rho_{XY}|} \sqrt{\frac{1 - \delta_Y^2 / \sigma_Y^2}{1 - \delta_X^2 / \sigma_X^2}}, \quad (3)$$

σ_X и σ_Y – среднеквадратические отклонения величин X и Y ; ρ_{XY} – коэффициент корреляции между X и Y ; δ_X и δ_Y – случайные среднеквадратические погрешности измерения X и Y для рассматриваемого массива данных.

1. Постановка задачи

Решается задача нахождения коэффициентов K_0 , K_1 , K_2 уравнения (1), в котором $Z = Z_0 + z$, $X = X_0 + x$, $Y = Y_0 + y$, где Z_0 , X_0 и Y_0 – случайные величины не подверженные случайным погрешностям, z , y и x – случайные погрешности. Разброс точек в корреляционной связи величин Z , X и Y обусловлен как случайными погрешностями, так и неконтролируемыми параметрами. Случайные погрешности z , y и x имеют СКО равные δ_z , δ_y и δ_x , которые представляют собой случайные среднеквадратические погрешности величин Z , Y и X . Предполагается, что величины Z , Y и X не коррелируют с погрешностями z , y и x ; погрешности z , y и x не коррелируют между собой.

2. Новый метод разделения на компоненты

В предлагаемом методе разделения на компоненты используется обобщенная формула линейной регрессии (2). Для нахождения коэффициентов регрессии уравнения (1) применяется численный итерационный метод [5–7]. Процедура вычисления коэффициентов регрессии заключается в следующем.

Задается один из коэффициентов регрессии уравнения (1), например K_1 , и находится среднеквадратическое отклонение разницы $\beta = Z - K_1 X - K_0$

$$\sigma_\beta = \sqrt{\sigma_Z^2 + K_1^2 \sigma_X^2 - 2K_1 \sigma_Z \sigma_X \rho_{ZX}},$$

где σ_Z и σ_X – среднеквадратические отклонения величин Z и X , а ρ_{ZX} – нормированный коэффициент корреляции между Z и X .

Далее находится нормированный коэффициент корреляции β с Y

$$\rho_{\beta Y} = \frac{\rho_{ZY} \sigma_Z - K_1 \rho_{XY} \sigma_X}{\sigma_\beta}, \quad (5)$$

где ρ_{ZY} и ρ_{XY} – нормированные коэффициенты корреляции между ZY и XY .

Затем вычисляется случайная среднеквадратическая погрешность величины β

$$\delta_\beta = \sqrt{\delta_Z^2 + K_1^2 \delta_X^2} \quad (6)$$

и коэффициент регрессии

$$K_2 = \frac{\sigma_\beta}{\sigma_Y} \frac{B_1}{A_1} \frac{1}{2\rho_{\beta Y}} \left\{ \left(\frac{A_1}{B_1} - \frac{B_1}{A_1} \right) + \sqrt{\left(\frac{A_1}{B_1} - \frac{B_1}{A_1} \right)^2 + 4\rho_{\beta Y}^2} \right\}, \quad (7)$$

$$\text{где } A_1 = \sqrt{1 - |\rho_{\beta Y}|} \sqrt{\frac{1 - \delta_Y^2 / \sigma_Y^2}{1 - \delta_\beta^2 / \sigma_\beta^2}}, \quad B_1 = \sqrt{1 - |\rho_{\beta Y}|} \sqrt{\frac{1 - \delta_\beta^2 / \sigma_\beta^2}{1 - \delta_Y^2 / \sigma_Y^2}}, \quad (8)$$

δ_Y – случайная среднеквадратическая погрешность измерения величины Y , σ_Y – среднеквадратическое отклонение величины Y .

Полученное значение коэффициента регрессии K_2 используется для нахождения среднеквадратического отклонения разницы $\mu = Z - K_2 Y - K_0$

$$\sigma_\mu = \sqrt{\sigma_Z^2 + K_2^2 \sigma_Y^2 - 2K_2 \sigma_Z \sigma_Y \rho_{ZY}}; \quad (9)$$

коэффициента корреляции μ с X

$$\rho_{\mu X} = \frac{\rho_{ZX}\sigma_Z - K_2\rho_{YX}\sigma_Y}{\sigma_\mu}; \quad (10)$$

случайной среднеквадратической погрешности величины μ

$$\delta_\mu = \sqrt{\delta_Z^2 + K_2^2\delta_Y^2} \quad (11)$$

и коэффициента регрессии

$$K_1 = \frac{\sigma_\mu}{\sigma_X} \frac{B_2}{A_2} \frac{1}{2\rho_{\mu X}} \left\{ \left(\frac{A_2}{B_2} - \frac{B_2}{A_2} \right) + \sqrt{\left(\frac{A_2}{B_2} - \frac{B_2}{A_2} \right)^2 + 4\rho_{\mu X}^2} \right\}, \quad (12)$$

где

$$A_2 = \sqrt{1 - |\rho_{\mu X}|} \sqrt{\frac{1 - \delta_X^2/\sigma_X^2}{1 - \delta_\mu^2/\sigma_\mu^2}}, \quad B_2 = \sqrt{1 - |\rho_{\mu X}|} \sqrt{\frac{1 - \delta_\mu^2/\sigma_\mu^2}{1 - \delta_X^2/\sigma_X^2}}, \quad (13)$$

δ_X – случайная среднеквадратическая погрешность измерения величины X .

Затем полученное значение коэффициента регрессии K_1 подставляется в (4) и процедура нахождения коэффициентов K_1 и K_2 продолжается до выполнения условий

$$|K_{1,m-1} - K_{1,m}| \leq \theta_1, \quad |K_{2,m-1} - K_{2,m}| \leq \theta_2, \quad (14)$$

где θ_1 и θ_2 – некоторые наперед заданные числа, определяющие погрешность вычисления коэффициентов K_1 и K_2 , m – количество итераций.

Далее находится коэффициент регрессии K_0 по известной формуле

$$K_0 = \bar{Z} - K_1\bar{X} - K_2\bar{Y}, \quad (15)$$

где $\bar{Z}$, $\bar{X}$, $\bar{Y}$ – средние значения физических величин Z , X и Y .

Заключение

Предложен новый метод разделения физической величины на компоненты с учетом случайных погрешностей измеряемых величин, который позволяет получать устойчивые и физически корректные коэффициенты регрессии независимо от используемого массива данных. Предложенный метод может быть использован в работах [8–12].

ЛИТЕРАТУРА

1. Крамер Г. Математические методы статистики. М.: Мир, 1975. 648 с.
2. Щелканов Н.Н. Построение регрессионной зависимости между аэрозольными оптическими толщами атмосферы с учетом их случайных погрешностей // II заседание Рабочей группы проекта «Аэрозоли Сибири». Тез. докл. Томск: ИОА СО РАН, 1995. С. 16.
3. Щелканов Н.Н. Обобщенный метод построения линейной регрессии и его применение для построения однопараметрических моделей аэрозольного ослабления // Оптика атмосферы и океана. 2005. Т. 18. № 1–2. С. 86–90.
4. Щелканов Н.Н. Новый метод нахождения коэффициентов линейной регрессии между двумя физическими величинами // Вестник ТГУ. Управление, вычислительная техника и информатика. 2010. № 4 (13). С. 91–96.
5. Щелканов Н.Н. Итерационный метод разделения коэффициентов ослабления на две компоненты // Аэрозоли Сибири. XIII Рабочая группа: тез. докл. Томск: Изд-во ИОА СО РАН, 2006. С. 58.

6. *Shchelkanov N.N.* Iterative method of separation of the extinction coefficients into two components // Proc. of SPIE – The International Society for Optical Engineering. 2006. V. 6522. 65221O. 5p.
7. *Shchelkanov N.N.* Comparison of two methods of separation of the extinction coefficients into components // XIV Int. Symp. «Atmospheric and Ocean Optics. Atmospheric Physics»: Abstracts. Tomsk: IAO SB RAS, 2007. P. 159, 160.
8. *Рублев А.Н., Григорьев Г.Ю., Удалова Т.А., Журавлева Т.Б.* Регрессионные модели для оценки углеродного обмена в бореальных лесах // Оптика атмосферы и океана. 2010. Т. 23. № 01. С. 21–26.
9. *Лысенко С.А., Кугейко М.М.* Методика определения концентрации респираторной фракции атмосферного аэрозоля по данным трехчастотного лидарного зондирования // Оптика атмосферы и океана. 2010. Т. 23. № 02. С. 149–155.
10. *Лысенко С.А., Кугейко М.М.* Восстановление оптических и микрофизических характеристик поствулканического стратосферного аэрозоля из результатов трехчастотного лидарного зондирования // Оптика атмосферы и океана. 2011. Т. 24. № 04. С. 308–318.
11. *Афонин С.В.* Апробация способа восстановления АОТ над сушей по спутниковым измерениям MODIS в ИК-диапазоне спектра // Оптика атмосферы и океана. 2011. Т. 24. № 08. С. 703–705.
12. *Абдуллаев С.Ф., Назаров Б.И., Салихов Т.Х., Маслов В.А.* Корреляции температуры приземной атмосферы и оптической толщи аридного аэрозоля по данным AERONET // Оптика атмосферы и океана. 2012. Т. 25. № 05. С. 428–433.

Щелканов Николай Николаевич

Институт оптики атмосферы им. В.Е. Зуева СО РАН, Томск

E-mail: snn@iao.ru

Поступила в редакцию 25 мая 2012 г.

Shchelkanov Nikolai N. (Zuev Institute of Atmospheric Optics of the SB RAS). A new method for dividing a physical parameter into components.

Keywords: iterative method, linear regression, random errors.

A new method is proposed for determining the linear regression coefficients of the equation $Z = K_0 + K_1 X + K_2 Y$ taking into account random errors of the physical parameters Z , X and Y . The method allows obtaining the stable and physically correct regression coefficients K_0 , K_1 and K_2 independently on the considered data array.