## ГЕОДЕЗИЯ

**УДК 528.112** **DOI 10.52928/2070-1683-2024-39-4-74-79**

# ИСПОЛЬЗОВАНИЕ РАСШИРЕННОЙ МОДЕЛИ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ ДЛЯ ПОИСКА ГРУБЫХ ОШИБОК В РЯДЕ МНОГОКРАТНЫХ ИЗМЕРЕНИЙ

***канд. техн. наук, доц. А.М. ДЕГТЯРЕВ, А.С. ИВАШНЁВА***
***(Полоцкий государственный университет имени Евфросинии Полоцкой)***

*Для поиска грубых ошибок в ряде многократно измеренных значений одной величины использована расширенная модель метода наименьших квадратов, при которой грубые ошибки включены в определяемые величины. Окончательная локализация грубых ошибок проводится на основе статистического тестирования по полученному одномерному алгоритму на основе алгоритма data snooping, разработанного в 50-х годах голландским геодезистом Baarda.*

***Ключевые слова:*** *метод наименьших квадратов, расширенная модель уравнивания, статистическое тестирование, грубая ошибка, многократно измеренная величина.*

**Введение.** Для уравнивания и анализа разного рода геодезических построений разработан очень эффективный математический аппарат на основе метода наименьших квадратов (МНК) и разного рода расширенных моделей. Этот же аппарат можно применить для обработки и анализа одной многократно измеренной равноточной (неравноточной), некоррелированной (коррелированной) величины. Также этот аппарат можно использовать без привлечения дополнительных условий для выявления в ряде измерений грубых ошибок.

Выявления грубых ошибок при обработке одной многократно измеренной величины имеет большое значение для геодезической практики и достаточно давнюю историю [2; 3]. Первые алгоритмы, разработанные классиками математической статистики, относятся к XIX веку [4]. Но алгоритмы носят частный характер и стоят как бы поодаль от общей процедуры получения наиболее эффективной оценки из ряда многократных измерений одной величины. В 50-е годы XX века голландский геодезист Baarda предложил алгоритм уравнивания геодезических сетей с включением в модель обработки грубых ошибок в качестве неизвестных [1]. В настоящее время этот алгоритм под именем data snooping используется в матричной форме, представленный также голландским ученым-геодезистом Teunissen [5].

**Основная часть.** При анализе формул Teunissen [5] было замечено, что если матрицу уравнений поправок $A$ параметрического способа уравнивания с расширенной моделью заменить на единичный вектор-столбец $e$, то получится простой, обоснованный, достаточно информативный и мощный алгоритм локализации грубых ошибок в ряде значений одной многократно измеренной величины. Рассмотрим этот подход подробно.

Пусть некоторая истинная величина $Y$ измерена $n$ раз как $y_i$, с целью определения ее значения и точности. Введем следующие обозначения:

$Y$ – истинные результаты измерения;

$X$ – истинное значение определяемой величины;

$y_i$ – измеренные величины;

$\hat{y}_i$ – уравненное значение (оценка) определяемой величины;

$\hat{x}$ – уравненное значение (оценка) результатов измерений.

Так как истинное значение $X$ определяемой величины должно равняться истинным результатам измерений $Y$ $n$ раз, то в самом общем, матричном виде это можно записать как

$$Y = e \cdot X, \tag{1}$$

где $e$ – вектор-столбец из единиц по числу измерений $e = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$.

Очевидно, что поскольку истинных измерений $Y$ мы не знаем, то их можно заменить на уравненные, вида

$$Y \approx \hat{y}_i = y_i + v_i, \tag{2}$$

откуда следует, что $v = Y - y \approx \Delta$ . Т.е. уклонения $v$ в совокупности должны стремиться к истинным ошибкам $\Delta$, а их оценки

$$\sigma_0 = \sqrt{\frac{\Delta^T \cdot \Delta}{n}} \approx \sqrt{\frac{v^T \cdot v}{n}}$$

достаточно эквивалентны. Если в (1) заменить истинное значение определяемой величины $X$ на ее оценку, $X \approx \hat{x}$, то объединяя с (2) в матричном виде получим

$$y + v = e \cdot \hat{x}. \tag{3}$$

В алгебраическом виде (3) для $n$ измерений имеет вид

$$\begin{cases} y_1 + v_1 = 1 \cdot \hat{x} \\ y_2 + v_2 = 1 \cdot \hat{x} \\ \cdots \\ y_n + v_n = 1 \cdot \hat{x}. \end{cases} \tag{3a}$$

Для решения несовместной системы (3) применим МНК с условием качества $\Phi = v^T \cdot P \cdot v \to \min$ . Тогда имеем

$$e^T \cdot P \cdot e \cdot \hat{x} = e^T \cdot P \cdot y \,, \tag{4}$$

а искомая оценка из (4) величины $\hat{x}$ будет

$$\hat{x} = \left(e^T \cdot P \cdot e\right)^{-1} \cdot e^T \cdot P \cdot y \to \frac{e^T \cdot P \cdot y}{e^T \cdot P \cdot e} \,, \tag{5}$$

так как и первый, и второй сомножитель в (5) числа. Здесь $P$ – матрица весов измерений, которая может быть единичной, диагональной или полной. Формула (6) эквивалентна эффективной МНК-оценке при уравнивании геодезической сети параметрическим способом, когда матрица уравнений поправок $A$ равна единичному вектору $e$. Если $P = E$, тогда из (5) имеем среднее арифметическое

$$\hat{x} = \frac{[y]}{n} = \bar{x}_1 \,. \tag{6}$$

Если матрица весов $P$ диагональная, то получаем формулу известного среднего взвешенного значения

$$\hat{x} = \frac{[p \cdot y]}{[p]} = \bar{x}_2. \tag{6b}$$

Если матрица $P$ полная, то для оценки алгебраический вид очень громоздок и необходимо применять формулу (5), которая является наиболее общей и содержит все перечисленные варианты как частные случаи.

Погрешность одного измерения (погрешность единицы веса) после обработки по Бесселю будет

$$\sigma_0 = \sqrt{\frac{v^T \cdot v}{n-1}},$$

погрешность среднего арифметического

$$\sigma_{\bar{x}}^2 = \left(e^{\mathrm{T}} \cdot e\right)^{-1} \cdot e^{\mathrm{T}} \cdot \sigma_0^2 \cdot e \cdot \left(e^{\mathrm{T}} \cdot e\right)^{-1} = \sigma_0^2 \cdot \left(e^{\mathrm{T}} \cdot e\right)^{-1} = \frac{\sigma_0^2}{n}.$$

Уравненные измерения $\hat{y} = y + v$

$$\hat{y} = y + v = y + e \cdot \left(e^{\mathrm{T}} \cdot e\right)^{-1} \cdot e^{\mathrm{T}} \cdot y = \frac{I}{n} \cdot y = \begin{bmatrix} \bar{x} \\ \cdots \\ \bar{x} \end{bmatrix} \cdot e \cdot \bar{x} \,. \tag{7}$$

Поправки в измерения будут

$$v = e \cdot \left(e^{\mathrm{T}} \cdot e\right)^{-1} \cdot e^{\mathrm{T}} \cdot y = \frac{I}{n} \cdot y, \tag{8}$$

где $I$ – матрица из единиц $(n \times n)$;

$E$ – единичная матрица $(n \times n)$.

Погрешность уравненного измерения

$$K_{\hat{y}} = e \cdot \left(e^{T} \cdot e\right)^{-1} \cdot e^{\mathrm{T}} \cdot \sigma_0^2 \cdot e \cdot \left(e^{\mathrm{T}} \cdot e\right)^{-1} \cdot e^{T} = \frac{e \cdot e^{T}}{n} = \frac{\sigma_0^2}{n}. \tag{9}$$

Пусть теперь в модели (3а) имеем в $i$-м измерении грубую ошибку $\nabla l$

$$\begin{gathered} v_1 = 1 \cdot \hat{x} - y_1 \\ \dots \\ v_i = 1 \cdot \hat{x} - \left(y_i + \nabla l\right) \\ \dots \\ v_n = 1 \cdot \hat{x} - y_n. \end{gathered} \tag{10}$$

В матричном виде эту модель как расширенную модель метода наименьших квадратов можно представить следующим образом:

$$v_{\nabla} = e \cdot \hat{x}_{\nabla} - y - c \cdot \nabla l = e \cdot \hat{x}_{\nabla} - c \cdot \nabla l - y, \tag{11}$$

где $c$ – нулевой вектор столбец с единицей на $i$-м месте;

индекс набла ∇ добавляется к поправкам и к оценке, так как они будут искажены дополнительным влиянием – грубой ошибкой $\nabla l$.

Систему (5) можно свернуть как

$$v_{\nabla} = \begin{pmatrix} e & -c \end{pmatrix} \cdot \begin{pmatrix} \hat{x}_{\nabla} \\ \nabla l \end{pmatrix} - y. \tag{11a}$$

Применив к системе (5) МНК в виде левой трансформации Гаусса с весами $P$ получим

$$\begin{pmatrix} e^{T} \\ -c^{T} \end{pmatrix} \cdot P \cdot v = \begin{pmatrix} e^{T} \\ -c^{T} \end{pmatrix} \cdot P \cdot \left[ \begin{pmatrix} e & -c \end{pmatrix} \cdot \begin{pmatrix} \hat{x} \\ \nabla l \end{pmatrix} - y \right], \tag{12}$$

или

$$0 = \begin{bmatrix} e^{T} \cdot P \cdot e & -e^{T} \cdot P \cdot c \\ -c^{T} \cdot P \cdot e & c^{T} \cdot P \cdot c \end{bmatrix} \cdot \begin{bmatrix} \hat{x}_{\nabla} \\ \nabla l \end{bmatrix} - \begin{bmatrix} e^{T} \cdot P \cdot y \\ -c^{T} \cdot P \cdot y \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} n & -1 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \hat{x} \\ \nabla l \end{bmatrix} = \begin{bmatrix} [y] \\ -y_i \end{bmatrix}.$$

Обратная матрица системы (5)

$$\begin{bmatrix} 1 & 1 \\ 1 & n \end{bmatrix} \cdot \frac{1}{n-1},$$

а решение

$$\begin{cases} \begin{bmatrix} \hat{x}_{\nabla} \\ \nabla l \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & n \end{bmatrix} \cdot \frac{1}{n-1} \cdot \begin{bmatrix} [y] \\ -y_i \end{bmatrix} \\ \hat{x}_{\nabla} = \frac{[y] - y_i}{n-1} \\ \nabla l = \frac{[y] - y_i \cdot n}{n-1} = \frac{\bar{x} - y_i}{\left(\frac{n-1}{n}\right)} = \frac{n \cdot v_i}{n-1}. \end{cases} \tag{13}$$

Влияние ожидаемой ошибки (дополнительного влияния) на результаты обработки из (6)

$$\nabla\hat{x} = \left(e^T \cdot P \cdot e\right)^{-1} \cdot e^T \cdot P \cdot c \cdot \nabla l = \frac{e^T \cdot P \cdot c \cdot n}{e^T \cdot P \cdot e \cdot (n-1)} \cdot v_i. \quad (14)$$

Ковариационная матрица $Q_{\nabla l}^{-1} = \dfrac{n-1}{n}$, тогда статистика для $\nabla l$ есть [1]

$$T = \frac{\nabla l^2 \cdot Q_{\nabla l}^{-1}}{1 \cdot \sigma_0^2} = \frac{\left(\bar{x} - y_i\right)^2 \cdot \left(\dfrac{n-1}{n}\right)}{\left(\dfrac{n-1}{n}\right)^2 \cdot \sigma_0^2} = \frac{\left(\bar{x} - y_i\right)^2}{\left(\dfrac{n-1}{n}\right) \cdot \sigma_0^2}, \quad \sqrt{T} = \frac{\bar{x} - y_i}{\sigma_0 \cdot \sqrt{\dfrac{n-1}{n}}}. \quad (15)$$

Используя процедуру data snapping [1; 5] для статистики *T*, получаем

$$\sqrt{T} \le z(p),$$

где $z(p)$ – квантиль нормального распределения для двухсторонней доверительной вероятности *p*.

Таким образом, если это условие выполняется для *i*-го измерения, то это измерение не содержит грубую ошибку с доверительной вероятностью *p*.

Сама процедура тестирования достаточно проста. Из (15) видим, что $\bar{x} - y_i = v_i$, а знаменатель – число постоянное. Тогда получаем алгоритм вида

$$\sqrt{T_i} = \frac{v_i}{k} \le z(p). \quad (16)$$

Формула (16) замечательна тем, что не использует никаких дополнительных, часто сомнительных предположений и упрощений, а только строгую математическую модель (11) с включением в обработку грубой ошибки как дополнительного определяемого параметра.

Рассмотрим пример. Пусть значение длины измерено 20 раз в одинаковых условиях с ошибкой σ = 0,005 м. В результате сбоя прибора в 5 измерении появилась ошибка в 0,02 м (изначально неизвестно ни ее значение, ни номер измерения).

$$x = \begin{bmatrix} 436,257 \\ 436,256 \\ 436,248 \\ 436,256 \\ 436,273 \\ 436,252 \\ 436,251 \\ 436,254 \\ 436,247 \\ 436,262 \\ 436,260 \\ 436,257 \\ 436,257 \\ 436,253 \\ 436,262 \\ 436,256 \\ 436,253 \\ 436,264 \\ 436,256 \\ 436,254 \end{bmatrix}.$$

В качестве оценки используем среднее арифметическое $\overline{x}$ = 436,2564 м из ряда, и находим вектор уклонений от среднего $v$

$$v = \begin{bmatrix} 0,001 \\ 0,000 \\ -0,008 \\ 0,000 \\ 0,017 \\ -0,004 \\ -0,005 \\ -0,002 \\ -0,009 \\ 0,006 \\ 0,004 \\ 0,001 \\ 0,001 \\ -0,003 \\ 0,006 \\ 0,000 \\ -0,003 \\ 0,008 \\ 0,000 \\ -0,002 \end{bmatrix}.$$

По величинам отклонений пятое измерение наиболее подозрительное. Используя (15) и заданное значение стандартного отклонения, получаем значения статистик для $i$-го измерения

$$T = \begin{bmatrix} 0,21 \\ 0,00 \\ -1,64 \\ 0,00 \\ 3,49 \\ -0,82 \\ -1,03 \\ -0,41 \\ -1,85 \\ 1,23 \\ 0,82 \\ 0,21 \\ 0,21 \\ -0,62 \\ 1,23 \\ 0,00 \\ -0,62 \\ 1,64 \\ 0,00 \\ -0,41 \end{bmatrix}.$$

Выдвигаем гипотезу, что с вероятностью 0,95 любое из измерений ряда является грубым. Для вероятности 0,95, согласно процедуре data snooping (16), получаем квантиль нормального закона распределения z = 1,96 и сравниваем ряд статистик $T$ с квантилем. Неравенство не выполняется только для пятого измерения, которое необходимо с вероятность 95% признать грубым измерением, изъять из ряда и провести повторно обработку и контроль.

Далее, используя формулу (13), для величины дополнительного влияния по модели (11) имеем $\nabla\hat{l} = -0,018,$ что достаточно хорошо определяет величину грубой ошибки в 0,02 м, которая должна быть изъята из измерения, поэтому знак вычисленного значения противоположный.

**Заключение.** Таким образом, можно сделать вывод, что используя расширенную модель (11), когда грубая ошибка включена в определяемые параметры, и используя способ тестирования data snooping без привлечения дополнительных предположений, можно достаточно корректно определить номер грубого измерения и предрассчитать значения грубой ошибки.

ЛИТЕРАТУРА

1. Klassische und robuste Ausgleichungsverfahren / R. Jager, T. Muller, H. Saler et al. – Heidelberg: Herbert Wichmann Verlag, 2005. – 340 p.
2. Коугия В.А., Павлов В.И. Современные проблемы уравнивания инженерно-геодезических сетей. – СПб.: Нац. минер.-сырьевой ун-т «Горный», 2012. – 105 с.
3. Голубев В.В. Теория математической обработки геодезических измерений. – М.; Вологда: Инфра-Инженерия, 2021. – 424 с.
4. Дегтярёв А.М. Теория погрешностей и статистический анализ [Электронный ресурс]: электрон. конспект лекций. – Новополоцк: Полоц. гос. ун-т им. Евфросинии Полоцкой, 2023. – 1 эл. опт. диск (CD-R).
5. Teunissen P.J.G. Testing theory: an introduction. – Delft: VSSD, Delft university press, 2024. – 199 p.

REFERENCES

1. Jager, R., Muller, T., Saler, H. & Schwable, R. (2005). *Klassische und robuste Ausgleichungsverfahren*. Heidelberg: Herbert Wichmann Verlag.
2. Kougiya, V.A. & Pavlov V.I. (2012). *Sovremennye problemy uravnivaniya inzhenerno-geodezicheskikh setei*. Saint Petersburg: Nats. miner.-syr'evoi un-t «Gornyi». (In Russ.).
3. Golubev, V.V. (2021). *Teoriya matematicheskoi obrabotki geodezicheskikh izmerenii*. Moscow; Vologda: Infra-Inzheneriya. (In Russ.).
4. Degtyarev, A.M. (2023). *Teoriya pogreshnostei i statisticheskii analiz* [Elektronnyi resurs]: elektron. konsp. lektsii. Novopolotsk: Polots. gos. un-t im. Evfrosinii Polotskoi. (In Russ.).
5. Teunissen, P.J.G. (2024). *Testing theory: an introduction*. Delft: VSSD, Delft university press.

*Поступила 28.10.2024*

## USING AN EXTENDED LEAST SQUARES MODELTO FIND GROSS ERRORS IN A SERIES OF MULTPLE MEASUREMENTS

***A. DEGTJAREV, A. IVASHNIOVA***
***(Euphrosyne Polotskaya State University of Polotsk)***

*To find gross errors in a series of repeatedly measured values of one quantity, an extended model of the least squares method is used, in which gross errors are included in the determined values. The final localization of gross errors is carried out on the basis of statistical testing according to the obtained one-dimensional algorithm based on the data snooping algorithm developed in the 50s by the Dutch surveyor Baarda.*

***Keywords:*** *least squares method, extended adjustment model, statistical testing, gross error, repeatedly measured value.*