

УДК 004.91

В. А. Додонов

Институт проблем регистрации информации НАН Украины
ул. Н. Шпака, 2, 03113 Киев, Украина

Применение методов помехоустойчивого оценивания в задачах анализа измерительной информации

Рассмотрена задача оценки измерительной информации с целью выявления и исключения аномальных результатов измерений, определения весовых коэффициентов каждого из оставшихся результатов, согласования отдельных частей измерительной информации между собой для последующего использования в задачах вторичной обработки. Показано, что в ряду различных процедур такого анализа представляется перспективным поочередное использование методов помехоустойчивого (робастного) оценивания.

Ключевые слова: измерительная информация, оценка, робастность, методы, обработка.

Практика натурных испытаний образцов новой техники показывает, что на процесс измерений влияют три рода помех: внешние помехи, сбои измерительной аппаратуры, ошибки персонала [1–3]. Воздействие указанных помех приводит к появлению аномальных результатов измерений, уровень «загрязнения» которых составляет в среднем 5–10 % [4, 5].

Полученная измерительная информация (ИИ) используется для оценивания различных характеристик испытываемых образцов. Широко применяемые в настоящее время классические методы оценивания (ММАВ — метод максимума апостериорной вероятности, ММП — метод максимума правдоподобия, МНК — метод наименьших квадратов), предполагающие априори нормальный закон распределения ошибок измерений, оказываются весьма чувствительными к наличию выбросов, что ведет в большинстве случаев к значительному снижению эффективности процедур оценивания, к снижению точности оценивания характеристик [6–13].

В этих условиях возникает актуальная задача предварительного анализа ИИ с целью выявления и исключения аномальных результатов измерений, определения весовых коэффициентов каждого из оставшихся результатов, согласования отдельных частей ИИ между собой для последующего использования в задачах вторичной обработки. В ряду различных процедур такого анализа представляется пе-

© В. А. Додонов

спективным использование методов помехоустойчивого (робастного) оценивания [7–10]. В статистике под робастностью понимают нечувствительность к малым отклонениям от предположений [7].

При решении задач статистического анализа и, в частности, при вычислении оценок параметров распределений, проблема наличия в выборке аномальных измерений имеет чрезвычайно важное значение. Присутствие единственного аномального наблюдения может приводить к оценкам, которые совершенно не согласуются с выборочными данными. В борьбе с грубыми погрешностями измерений, если они не были обнаружены в процессе измерений, используют два подхода: исключение резко выделяющихся аномальных измерений из дальнейшей обработки; использование робастных методов обработки.

В работе рассмотрена возможность применения методов помехоустойчивого оценивания для задач анализа регистрируемой при испытаниях образцов ИИ и приводится разработанный с применением этих методов алгоритм отбраковки аномальных результатов измерений.

1. Методы помехоустойчивого оценивания

1.1. *Основные используемые в последнее время типы оценок.* Еще в середине 19-го века были предложены первые методы обработки выпадающих измерений. Однако классические методы «жесткой» отбраковки позволяют выявлять лишь крайне грубые ошибки. С их помощью удастся избежать аномальных результатов измерений, однако они ничем не могут помочь в ситуациях «скрытого» засорения [11–20].

МНК и его обобщения широко используются в практике статистических исследований, обладает рядом положительных свойств при независимых (имеющих одинаковую дисперсию) ошибках, распределенных по нормальному закону. Известно, что выбросы, вызванные наличием у распределений, описывающих ошибки измерений, более тяжелых по сравнению с нормальным распределением *хвостов* или просто большими ошибками результатов измерений, оказывают чрезвычайно большое влияние на оценки по МНК. Кривые, подогнанные МНК, сильно смещены в сторону выбросных точек. Окончательная оценка остаточных разностей квадратов может ввести в заблуждение, так как в этом случае выбросы неотличимы от нормальных результатов наблюдений.

Создание помехоустойчивых методов оценивания вызвано стремлением улучшить существующие схемы МНК так, чтобы выбросы оказывали как можно меньшее влияние на конечные результаты. Одним из наиболее удачных является метод помехоустойчивого оценивания, основанный на принципе максимального правдоподобия.

Основные используемые в последнее время типы оценок, выбором структуры которых можно обеспечить более выгодную по сравнению с выборочным средним защищенность от влияния выбросов: линейные комбинации порядковых статистик (*L*-оценки), оценки на основе ранговых тестов (*R*-оценки) и оценки типа оценок максимального правдоподобия (*M*-оценки). В работе для решения задач обработки ИИ предложено использовать *M*-оценки [12].

1.2. *M*-оценки. Пусть X_1, X_2, \dots, X_n — последовательность независимых, одина-

наково распределенных случайных величин, имеющих непрерывную плотность вероятности $f(x - \Theta)$, где Θ — параметр сдвига. Логарифм функции правдоподобия можно записать как

$$\ln L(\Theta) = \sum_{i=1}^n \ln f(x_i - \Theta) = \sum_{i=1}^n \rho(x_i - \Theta),$$

где $\rho(x_i) = -\ln f(x)$. Согласно ММП требуется минимизировать $\ln L(\Theta)$, или минимизировать $K(\Theta) = \sum \rho(x_i - \Theta)$. Предположим, что минимум можно найти путем дифференцирования и решения уравнения $K'(\Theta) = 0$, т.е. поиском соответствующего значения параметра сдвига Θ , которое удовлетворяет условию:

$$\sum_{i=1}^n \Psi(x_i - \Theta) = 0,$$

где $\Psi(x) = \rho'(x) = -f'(x)/f(x)$. Решение этого уравнения, минимизирующее $K(\Theta)$, называется оценкой максимального правдоподобия или M -оценкой параметра Θ и обозначается $\bar{\Theta}$.

В методе M -оценок желательно определить функцию Ψ так, чтобы конечная оценка была защищена от влияния присутствующей в результатах наблюдений небольшой доли (около 10 %) выбросов. Кроме того, необходимо, чтобы метод M -оценок обеспечивал достаточно хорошую эффективность оценок. Например, когда наблюдаемые результаты описываются распределением почти нормального вида, эффективность должна быть порядка 95 %. 5-процентная потеря эффективности — это цена, выплачиваемая за достижение устойчивости оценок в тех случаях, когда распределение ошибок заметно отличается от нормального.

Используя строгое определение помехоустойчивости, Хьюбер [7] нашел соответствующий вид функций для ρ и Ψ :

$$\rho = \begin{cases} \frac{x^2}{2}, & |x| \leq a, \\ a|x| - \frac{a^2}{2}, & |x| > a. \end{cases} \quad \Psi(x) = \begin{cases} -a, & x < -a, \\ x, & |x| \leq a, \\ a, & x > a. \end{cases}$$

Функции ρ и Ψ связаны с распределением в основном нормального вида, однако имеют «утяжеленные хвосты», подчиняющиеся двойному экспоненциальному распределению:

$$\rho(x) = |x| + c, \quad \Psi(x) = \begin{cases} -1, & x < 0, \\ 1, & x > 0, \end{cases}$$

$\sum_{i=1}^n \Psi(x_i - \Theta) = 0$ дает оценку $\bar{\Theta}$, равную медиане выборки.

Соответствующая M -оценка определяется функциями ρ и Ψ , реализующими условие минимакса выражения

$$\min_T \max_F [V(T)],$$

где F представляет собой семейство ε — загрязненных распределений вида $\{(1-\varepsilon)\Phi + \varepsilon H\}$; $V(T)$ — асимптотическая дисперсия; T — класс оценок.

1.3. *Применение метода M -оценок для анализа данных о траекториях движения.* В работе рассматривается решение задачи аппроксимации некоторой траектории движения.

Рассмотрим линейную модель:

$$Y = X\Theta + \delta, \quad (1)$$

где X — матрица условий измерений (конструкционная матрица) размерностью $n \times p$; n — количество измерений; p — количество условий i -го измерения; δ — случайный вектор погрешностей размерностью n , имеющий нулевое среднее значение и ковариационную матрицу $\sigma^2 I$; Θ — параметры функции регрессии, которые нужно оценить.

Отметим, что все описанные далее методы не используют существенным образом линейной структуры модели и могут применяться к анализу нелинейной регрессионной модели.

Требуется найти:

$$\min \sum_1^n \rho(y_i - X_i \Theta). \quad (2)$$

Обозначим через r_i остаточные разности (невязки) оценок Θ :

$$r_i = y_i - X_i \Theta, \quad (3)$$

где X_i — i -я строка конструкционной матрицы X .

Рассмотрим p первых частных производных. Это означает, что необходимо решить систему p уравнений:

$$\sum_1^n \Psi(r/s) x_{ij} = 0, \quad j = \overline{1, p}. \quad (4)$$

Результатом решения системы нормальных уравнений (4) являются оценки Θ_i .

Предложено несколько видов функции Ψ . Для устойчивого оценивания используются наиболее распространенные «одномерные» оценочные функции (Тьюки, Хьюбера, Хампеля, Эндрюса и др.). Эти функции, как и многие другие, обладают одним существенным недостатком. Например, если значения случайной величины X умножить на одну и ту же константу, то начальная оценка не обяза-

тельно изменится на ту же константу. Для того, чтобы M -оценка параметра масштаба обладала свойством инвариантности, выражение (3) нормируется и находится

$$\min \sum_1^n \rho\left(\frac{r_i}{s}\right), \quad (5)$$

где s — помехоустойчивая оценка параметра масштаба.

Для решения системы

$$\sum_1^n \rho\left(\frac{r_i}{s}\right) x_{ij} = 0, \quad j = \overline{1, p} \quad (6)$$

можно использовать итерационные схемы Ньютона, H -метода или средневзвешенного метода наименьших квадратов.

Схема Ньютона:

$$\bar{\Theta}_2 = \bar{\Theta}_1 + \frac{s \sum \Psi[(x_i - \bar{\Theta}_1)/s]}{\sum \Psi'[(x_i - \bar{\Theta}_1)/s]},$$

где в знаменателе второго члена суммирование производится по элементам, удовлетворяющим условию $|x_i - \bar{\Theta}_1|/s \leq a$.

H-метод:

$$\bar{\Theta}_2 = \bar{\Theta}_1 + \frac{s \sum \Psi[(x_i - \bar{\Theta}_1)/s]}{n},$$

где второй член — среднее по псевдоостаткам (винзорированное среднее).

В данном случае МНК применяется к псевдоостаткам.

Средневзвешенный метод наименьших квадратов:

$$\bar{\Theta}_2 = \frac{\sum w_i x_i}{\sum w_i} = \bar{\Theta}_1 + \frac{s \sum \Psi[(x_i - \bar{\Theta}_1)/s]}{\sum w_i},$$

где веса выбираются по формуле:

$$w_i = \frac{\Psi[(x_i - \bar{\Theta}_1)/s]}{(x_i - \bar{\Theta}_1)/s}; \quad i = 1, 2, \dots, n.$$

Следует отметить, что данные итерационные схемы схожи.

Для получения начального приближения итерационного процесса необходи-

мо иметь предварительную оценку вектора $\bar{\Theta}$. Полезно, чтобы такая оценка была помехоустойчивой. Обозначим предварительную оценку, полученную любым способом, через $\bar{\Theta}^1$. В качестве помехоустойчивой меры разброса s применяется медиана (ненулевых) отклонений:

$$s = \frac{\text{med} |y_i - y_i \bar{\Theta}^1|}{0,6745}, \quad (7)$$

где med — оператор вычисления медианы:

$$\text{med } r_i = \begin{cases} r_i, & \text{при } n = 2m - 1; \\ 1/2(r_m + r_{m+1}), & \text{при } n = 2m; \end{cases} \quad (8)$$

где n — объем вариационного ряда; r_m, r_{m+1} — m -й и $(m + 1)$ -й его член.

В этом случае берем медиану только ненулевых отклонений, иначе при больших значениях p многие остаточные разности окажутся равны нулю. Тогда обычная оценка с помощью медианы дает очень малое значение s , поэтому потеря эффективности велика.

В работе [4] предложен другой вариант определения s :

$$s = \frac{y_{0,75} - y_{0,25}}{2 * 0,6745}, \quad (9)$$

где $y_{0,75}, y_{0,25}$ — соответствующие процентилю выборочного распределения.

2. Решение системы нормальных уравнений для нахождения коэффициентов функции регрессии

2.1. *Схема итерационного метода Гаусса-Ньютона.* На k -м шаге итерации заменим систему p уравнений (6) линеаризацией в окрестности точки $\bar{\Theta}^k$.

$$\sum_i^n x_{ij} \left[\left(\frac{r_i^k}{s} \right) - \frac{\Psi'(r_i^k)}{s} * x_{ij} * (\bar{\Theta}_j^{k+1} - \bar{\Theta}_j^k) \right] \approx 0, \quad j = \overline{1, p}, \quad (10)$$

где

$$r_i^k = y_i - X_i \bar{\Theta}^k; \quad (11)$$

$$\sum_i^n x_{ij} \left[\left(\frac{r_i^k}{s} \right) - \sum_i^n x_{ij} * \frac{\Psi'(r_i^k)}{s} * x_{ij} * (\bar{\Theta}_j^{k+1} - \bar{\Theta}_j^k) \right] \approx 0, \quad j = \overline{1, p},$$

или в матричной форме:

$$\bar{\Theta}^{k+1} = \bar{\Theta}^k + (X^T \Psi'^k X)^{-1} X^T \Psi^k, \quad (12)$$

где Ψ'^k — диагональная матрица с диагональю $diag[\Psi'_1{}^k, \Psi'_2{}^k, \dots, \Psi'_n{}^k]$ и элементами

$$\Psi'_i{}^k = \frac{\Psi'(r_i^k)}{s}; \quad (13)$$

Ψ^k — вектор из n элементов с элементами

$$\Psi_i^k = \left(\frac{r_i^k}{s}\right). \quad (14)$$

Итерационную процедуру продолжают до достижения достаточной степени сходимости процесса.

2.2. *Средневзвешенный МНК.* Заменяем систему p уравнений (6) соответствующими аппроксимациями:

$$\sum_1^n w_i x_{ij} (y_i - X_i \bar{\Theta}) \approx 0, \quad j = \overline{1, p}, \quad (15)$$

где

$$w_i = \frac{\Psi\left[\frac{y_i - X_i \bar{\Theta}^1}{s}\right]}{\frac{y_i - X_i \bar{\Theta}^1}{s}}, \quad i = \overline{1, n}. \quad (16)$$

Решение этой системы уравнений имеет вид:

$$\bar{\Theta}^2 = (X^T W X)^{-1} X^T W X, \quad (17)$$

где W — диагональная матрица с элементами $diag[w_1, w_2, \dots, w_n]$. $\bar{\Theta}^2$ можно использовать для продолжения итерационной процедуры и построения новой весовой матрицы.

Итерационную процедуру продолжают до тех пор, пока не будет достигнута достаточная степень сходимости процесса. Окончательный результат обозначим $\bar{\Theta}$. Окончательный вес, относящийся к данному элементу выборки (результату измерения) указывает, принадлежит ли данный элемент к выборке.

3. Функция пси

Наиболее часто используются следующие виды функции Ψ [4].

Функция Хьюбера:

$$\Psi(x) = \begin{cases} -a, & \text{при } x < -a, \\ x, & \text{при } |x| \leq a, \\ a, & \text{при } x > a. \end{cases} \quad (18)$$

Для функции Хьюбера константа a должна быть порядка 1,5. Установлено, что для достижения асимптотической 95 %-й эффективности в предположении о нормальном распределении помех измерений $a = 1,345$. Для случая загрязнения измерений распределением Коши рекомендовано $a = 1,1$.

Функция Хампеля:

$$\Psi(x) = (\text{sign } x) * \begin{cases} |x|, & 0 \leq |x| < a, \\ a, & a \leq |x| < b, \\ \frac{c-|x|}{c-b} * a, & b \leq |x| < c, \\ 0, & c \leq |x|. \end{cases} \quad (19)$$

Для функция Хампеля рекомендуемые значения констант: $a = 1,7$; $b = 3,4$; $c = 8,5$. Встречаются и значения 3; 6; 9.

Функция Эндрюса:

$$\Psi(x) = \begin{cases} \sin \left| \frac{x}{a} \right|, & \text{при } |x| \leq a\pi, \\ 0, & \text{при } |x| > a\pi. \end{cases} \quad (20)$$

Для функция Эндрюса $a = 2,1$. Фактически, если масштаб известен, величина $a = 1,339$ соответствует 5 %-й потере эффективности. Встречается и значение $a = 1,2$.

Биквадратная функция Тьюки:

$$\Psi(x) = \begin{cases} x \left[1 - \left(\frac{x}{a} \right)^2 \right]^2, & \text{при } |x| \leq a, \\ 0, & \text{при } |x| > a. \end{cases} \quad (21)$$

Для биквадратной функция Тьюки рекомендовано значение $a = 6,0$. Если масштаб известен, тогда $a = 4,685$, что соответствует 5 %-й потере эффективности.

4. Получение начальных приближений

Для решения системы уравнений (6) воспользуемся итерационными методами. В качестве начального приближения $\bar{\Theta}^1$ итерационного процесса необходимо

иметь первоначальную оценку вектора Θ . Полезно, чтобы такая оценка была помехоустойчивой.

Особенно чувствительны к выбору начального приближения оценки на базе немонотонных функций Ψ . При неудачном начальном приближении итерационный процесс может сходиться к локальному минимуму, либо для его завершения потребуется значительное число шагов.

Наиболее естественно за начальное приближение применить оценку, полученную методом наименьших квадратов. Но этот подход может и не обеспечить получение хорошего начального значения $\overline{\Theta}^1$, так как оценка по МНК сильно зависит от наличия в экспериментальных данных резких выбросов. В тех случаях, когда величины y_i и компоненты Θ малы, начальное решение можно задавать в виде $\overline{\Theta}^1 = 0$.

В работах [5–17] описаны три помехоустойчивых метода получения начального приближения: модифицированный метод Тейла; ортогональный метод Брауна–Муда; метод на основе использования коэффициента корреляции Спирмэна. Остановимся на первых двух.

4.1. *Модифицированный метод Тейла.* В этом методе в линейной модели выделяется среди других член Θ_0 . Далее, для получения остальных независимых переменных применяется процедура ортогонализации Грамма–Шмидта; значения ортогональных переменных x'_{ij} :

$$x'_{i1} = x_{i1}, \quad x'_{ij} = x_{ij} - \sum_k^{j-1} r'_{jk} x'_{ik},$$

$$r_{jk} = \frac{\sum_1^n x_{ij} * x'_{ik}}{\sum_1^n (x'_{ik})^2}, \quad i = \overline{1, n}, \quad j = \overline{2, p}. \quad (22)$$

Линейная модель (1) в терминах ортогональных независимых переменных записывается так:

$$Y = \Theta_0 + X' \Theta' + \delta'. \quad (23)$$

Оценки коэффициентов регрессии получают по схеме:

$$1. \quad dm(i, j) = (y_j - y_i) / (x'_{jm} - x'_{im}), \quad j > i, \quad i = \overline{1, n-1}; \quad (24)$$

$$2. \quad \delta \Theta'_m = med \, dm(i, j); \quad (25)$$

$$3. \quad \Theta'_m := \Theta'_m + \delta \Theta'_m; \quad (26)$$

$$4. \quad y_i := y_i - \delta \Theta'_m * X'_{im}, \quad i = \overline{1, n}. \quad (27)$$

5. Повторять пп. 1–4, пока не будет достигнут нужный уровень сходимости.

$$6. \quad \overline{\Theta}_0 = med_i y_i. \quad (28)$$

Чтобы вернуться к исходному базису, применим преобразование Грама-Шмидта к Θ^1 :

$$\begin{aligned} \Theta_p &= \Theta'_p, \\ \Theta_{p-i} &= \Theta'_{p-i} - \sum_{j=0}^{i-1} r_{p-j} \overline{p-i}^* \Theta_{p-j}, i = 1, \overline{p-1}. \end{aligned} \quad (29)$$

Число значений $dm(i, j)$, равное

$$p^* C_{N-1}^2 = p^* \frac{(n-1)(n-2)}{2}, \quad (30)$$

слишком велико даже для не очень большого n .

Одна из возможностей уменьшения объема вычислений $dm(i, j)$ заключается в следующем. Для каждого m строится упорядоченная выборка величин $X_{im}, i = 1, n$, в порядке возрастания их значений, и пусть

$$N^* = \left[\frac{n+1}{2} \right].$$

Тогда (при нечетном n) $X_{N^*, m}$ является медианой случайных величин X'_{im} . Затем строятся величины

$$dm(i) = \frac{y_{N^*+i} - y_i}{x'_{N^*+i, m} - x'_{im}}, \quad (31)$$

где $i = \begin{cases} 1, N^*, & \text{при четном } n, \\ 1, N^*-1, & \text{при нечетном } n. \end{cases}$

Полученные значения $dm(i)$ используются в п. 2 схемы (24)–(27).

4.2. *Ортогональный метод Брауна–Муда.* В этом методе, также с помощью схемы Грама–Шмидта (22), осуществляется переход к ортогональным независимым переменным. Метод реализуется по следующей схеме, начиная с $\overline{\Theta}^0 = 0$:

$$1. x_m^* = \text{med}_i x'_{im}; \quad (32)$$

$$2. x_{m+} = \text{med}_{i \in I_U} x'_{im}; \quad \overline{x}_m = \text{med}_{i \in I_L} x'_{im}, \quad (33)$$

где $I_U = \{i : x'_{im} > x_m^*\}, \quad I_L = \{i : x'_{im} \leq x_m^*\};$

$$3. y_{i+} = \text{med}_{i \in I_U} y_i, \quad \overline{y}_i = \text{med}_{i \in I_L} y_i; \quad (34)$$

$$4. \delta\Theta'_m = \frac{y_i^+ - y_i^-}{x_m^+ - x_m^-}; \quad (35)$$

$$5. \Theta'_m := \Theta'_m + \delta\Theta'_m; \quad (36)$$

$$6. y_i := y_i - \delta\Theta'_m * x'_{im}, \quad i = \overline{1, n}. \quad (37)$$

7. Повторять пп. 3–6 до достижения сходимости (выполнить пп. 1–7 для каждого из $m = 1, p$).

$$8. \quad \overline{\Theta}_0 = \underset{i}{med} y_i. \quad (38)$$

Затем происходит возвращение к исходному базису с помощью преобразований (25).

5. Описание программы

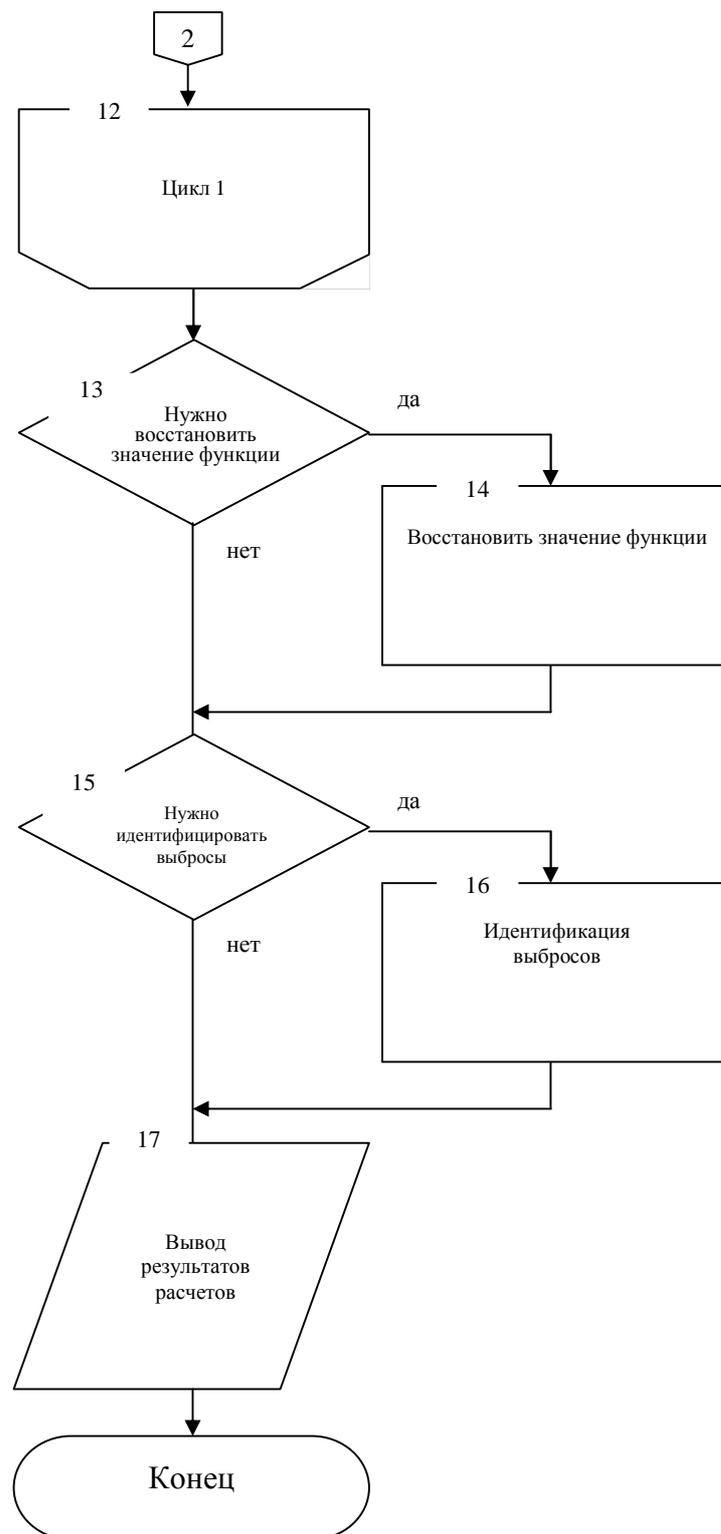
Разработана программа *robust.exe*, предназначенная для обработки данных о траекториях движения. Она позволяет аппроксимировать некоторую функцию от времени $y(t)$, представленную массивом наблюдений $y_i(t_i)$, многочленом заданной степени p :

$$y_i = \overline{\Theta}_0 + \sum_{j=1}^p \overline{\Theta}_j (t_i - t)^j.$$

Оценивание коэффициентов $\overline{\Theta}_j$, $j = \overline{0, p}$ производится помехоустойчивыми методами, что позволяет исключить или существенно уменьшить влияние аномальных результатов измерений на вычисляемые оценки. Начальное приближение может быть либо задано, либо вычислено обычным МНК, либо помехоустойчивым методом. Дальнейшее уточнение оценок $\overline{\Theta}_j$ может производиться в несколько этапов одним из приведенных выше методов, либо их чередованием. Этап заключается в выполнении заданного числа итераций или в достижении заданной величины модуля относительного приращения.

Кроме того, программа позволяет восстанавливать результаты наблюдений внутри интервала между двумя наблюдениями, а также идентифицировать неполные результаты измерений. Все исходные данные для работы программы находятся в файле *input.dat*, который должен находиться в том же каталоге, что и исполняемая программа. Файл исходных данных имеет структуру, которая описана в таблице. Результатом работы программы является файл *output.res*, который создается в процессе работы программы, и где представлены все результаты ее работы в удобном для восприятия виде.

Программа написана на языке программирования C++ с использованием компилятора *Borland C++* версии 2.0 и имеющейся в этой версии стандартной библиотеки классов. Программа включает в себя главную функцию *main()* и 15 функций, описанных далее. Также созданы новые классы *vector(double)*, *vector(int)* и *matrix(double)*, *matrix(int)*. Их описание и реализация приведена в тексте программы. Алгоритм работы главной программы представлен на рисунке.



Алгоритм работы главной программы

5.1. Типы данных

В программе определены следующие типы данных:

Enum Bool { *false*, *true* };

enum FirstCalcMethod { *MNK*, *BrownMood*, *fullTheil*, *shortTheil*};

enum IterMethod (*IMNK*, *GaussNewton*);

enum funcPSItype { *Andzews*, *Hampel*, *Huber*, *Tukey*};

enum StopType { *Steps*, *Preciz*};

Структура файла исходных данных *input.dat* приведена в таблице.

Структура файла исходных данных *input.dat*

| №№ поля | Характеристика | Наименование переменной программы | Тип | Значение по умолчанию |
|---|---|-----------------------------------|-------------------------|-----------------------|
| 1 | Число измерений | N | unsigned int | нет |
| 2 | Степень полинома | P | unsigned int | 4 |
| 3 | Признак наличия начального приближения | first_exist | Bool | 0 |
| 4 | Начальные значения $\Theta, j = \overline{0, P}$ (задаются, если поле 3 равно 1) | Tetta_0, Tetta_P | double, vector (double) | 0, {0} |
| 5 | Число этапов решения | EtapNumb | unsigned int | 1 |
| Далее для каждого этапа решения задаются поля | | | | |
| 6 | Метод получения начального приближения (имеет смысл только для первого этапа) | MN | FirstCalc-Method | нет |
| 7 | Итерационный метод | MI | Iter-Method | нет |
| 8 | Признак используемой функции пси | FuncPSI | funcPSI-type | нет |
| 9 | Критерий прекращения итераций | StopAt | StopType | нет |
| 10 | Значение критерия прекращения итераций | StepNumb (если StopAt=Steps) | unsigned int | 10 |
| | | Precision (если StopAt=Preciz) | double | 0,01 |
| 11 | Признак задания изменяемых начальных значений массива весов W | setw | Bool | 0 |
| 12 | Значения вектора-аналога диагональной матрицы весов | W(N) | vector (double) | {1} |
| | Последующие поля исходных данных задаются один раз для всего файла исходных данных: | | | |
| 13 | Признак восстановления аппроксимруемой функции | Values | Bool | нет |
| 14 | Шаг восстановления значений функции (имеет смысл, если в поле 13 значение равно 1) | dT | double | нет |
| 15 | Признак идентификации выбросов | IW | Bool | нет |

5.2. Описание функций программы

Главная программа *main()* : 1. Объявление данных. 2. Открытие исходного файла данных. 3. Ввод размерностей. 4. Объявление массивов. 5. Ввод остальной информации. 6. Вычисление матрицы условий наблюдения X . 7. Начало цикла расчетов; если достигнуто заданное количество этапов расчета, переход на п. 15. 8. Вычисление начального приближения (если оно не задано). 9. Начало цикла итераций. 10. Выполнить один шаг итерационного метода: если $MI==MНК$, то расчет методом наименьших квадратов, *iterMНК()*; если $MI==GaussNewton$, то расчет методом Гаусса–Ньютона, *itezGN()*. 11. Вычисление критерия остановки: — если $StopAt==Steps$, то сравнение с заданным количеством шагов; если $StopAt==Preciz$, то определение модуля относительного приращения, *verify_good()*. 12. Если выполнен критерий остановки, конец цикла итераций, переход к п. 13, иначе к п. 10. 13. Вычисление Θ_0 , *TetNull()*. 14. Переход на п. 7. 15. Восстановить значения аппроксимируемой функции, если $Values==true$. 16. Идентификация выбросов, если $IW==true$. 17. Вывод полученной в ходе расчетов информации.

Функция *calc_first()*. Вычисление первого приближения. Параметры: *method* — задает метод получения начального приближения; *tet_0, tetta* — начальное (первое) приближение; *y* — вектор исходных измерений; X — матрица условий наблюдения; W — вектор-аналог диагональной матрицы весов; *preciz* — точность вычислений. Возвращаемое значение: нет.

Алгоритм: 1. Объявляется копия вектора измерений y . 2. Если метод вычислений — МНК, то переход к п. 3, иначе к п. 6. 3. Вычисление Θ_0 — вызов *TetNull()*. 4. Вычисление $\Theta_1 \dots \Theta_N$ — вызов *funcMНК()*. 5. Возврат в функцию *main()*. 6. Объявление ортогональной матрицы X и массива аналогов коэффициентов корреляции при построении ортогонального базиса. 7. Если метод вычислений — метод Брауна–Муда, переход к п. 8, иначе к п. 12. 8. Вычисление ортогональной матрицы X — вызов *to_ortho()*. 9. Вычисление вектора Θ — вызов *funcBM()*. 10. Переход от ортогонального Θ к Θ в исходных координатах — вызов *from_ortho()*. 11. Возврат в функцию *main()*. 12. Если метод вычислений — метод Тейла, переход на п. 13, иначе к п. 18. 13. Вычисление ортогональной матрицы X — вызов *to_ortho()*. 14. Вычисление вектора Θ — вызов *fTheil()*. 15. Возврат к Θ в исходных координатах — вызов *from_ortho()*. 16. Вычисление Θ_0 — вызов *TetNull()*. 17. Возврат в функцию *main()*. 18. Если метод вычислений — сокращенный метод Тейла, то переход к п. 19, иначе к п. 23. 19. Вычисление ортогональной матрицы X — вызов *to_ortho()*. 20. Вычисление вектора Θ — вызов *sTheil()*. 21. Возврат к Θ в исходных координатах — вызов *from_ortho()*. 22. Вычисление Θ_0 — вызов *TetNull()*. 23. Возврат в функцию *main()*.

Функция *from_ortho()*. Переход от вычисленного в ортогональном базисе Θ к Θ в исходных координатах. Параметры: *tetta_ortho* — вектор Θ в ортогональном базисе (типа *vector(double)*); *corr* — массив аналогов коэффициентов корреляции, вычисленный при переходе к ортогональному базису — *vector(double)*. Возвращаемое значение: Значения вектора Θ в исходных координатах (типа *vector(double)*).

Алгоритм: По формулам (25) вычисляются значения элементов Θ (r_i в фор-

муле эквивалентно $corr.elem(i)$.

Функция $fTheil()$. Получение начального приближения методом Тейла. Величины $dm(i, j)$ вычисляются по формуле (24) схемы (24)–(27) по варианту с большим количеством вычислений. Параметры: y — вектор измерений типа $double$ (изменяется в процессе вычислений); $orthoX$ — ортогональная матрица условий наблюдения типа $matrix(double)$; $tetta$ — вектор оцениваемых коэффициентов регрессии (вычисляется) типа $vector(double)$; $preciz$ — требуемая точность вычислений, тип $double$. Возвращаемые значения: нет.

Алгоритм: 1. Объявления промежуточных данных. 2. Начало цикла вычислений. 3. Цикл по $m(m_i = 1)$: если $m \leq P$, то переход на п. 4, иначе переход на п. 8. 4. Цикл вычисления $dm(i, j)$. 5. Вычисление $(delta_tetta - mediana())$. 6. Выполнение итерации для m -го элемента — $iteration()$. 7. $m := m + 1$, переход на п. 3. 8. Проверка сходимости — $verify-good()$. Если результат положительный, то переход на п. 9, иначе переход на п. 2. 9. Возврат в вызывающую процедуру.

Функция $funcBM()$. Получение начального приближения ортогональным методом Брауна–Муда. Параметры: Y — вектор измерений типа $vector(double)$ (изменяется в процессе вычислений); $orthoX$ — ортогональная матрица условий наблюдений типа $matrix(double)$; $tetta$ — массив оцениваемых коэффициентов регрессии (вычисляется) типа $vector(double)$; $preciz$ — требуемая точность вычислений, $double$. Возвращаемые значения: нет.

Алгоритм: 1. Объявления промежуточных данных. 2. Цикл по m : если $m \leq P(m_1 = 1)$ то переход на п. 3, иначе — переход на п. 12. 3. Начало цикла вычислений. 4. Сортировка m -го столбца матрицы $orthoX$ по возрастанию — вызов $variat()$. 5. Вычисление медианы сортированного m -го столбца — вызов $mtdiana()$. 6. Цикл определения множеств $I_y, I_1, X_{E1}, X_{lu}, y_{lu}$: y_{lu} (по формуле (33)). 7. Определение $xm_plus, xm_minus, y_plus, y_minus$. 8. Вычисление $delta-tetta$ по формуле (34). 9. Вычисление итерации — вызов $iteration()$, вычисление относительного приращения Θ_m $r_delta.elem(m)$. 10. Если $r_delta.elem(m) > preciz$, то переход на начало цикла, п. 3, иначе переход на п. 11. 11. $m = m + 1$. Переход на п. 2. 12. Возврат в вызывающую программу.

Функция $funcMNK()$. Выполняет оценку параметров Θ по нормальным уравнениям метода наименьших квадратов (17). Параметры: I — массив данных типа $vector(double)$; X — матрица условий наблюдений типа $matrix(double)$; ww — вектор-аналог диагональной матрицы весов $vector(double)$; $tetta$ — вектор оцениваемых параметров типа $vector(double)$. Возвращаемое значение: нет.

Алгоритм: 1. Транспонировать матрицу X . 2. Восстановить матрицу из вектора ww . 3. Произвести вычисления по формуле (17).

Функция $iteration()$. Получение оценок тетты m -е на очередном шаге итерации, когда приращение дельта тетта m -го получено каким-либо помехоустойчивым методом. Возвращает модуль относительного приращения тетты. Параметры: y — вектор измерений (уточняется); $orthoX$ — ортогональная матрица условий наблюдений; $tetta$ — вектор оценок тетта (уточняется); $delta_tetta$ — приращение тетта m -го; m — индекс параметра, оценка которого вычисляется. Возвращаемое значение: модуль относительного приращения Θ_m , типа $double$.

Алгоритм: 1. Если $\Theta_m = 0$, то относительное приращение равно GREAT, иначе расчет по формуле относительного приращения Θ_m . 2. $\Theta_m = \Theta_m + d\Theta$.

3. Вычисление нового значения y на данном шаге итерации. 4. Возврат значения относительного приращения в вызывающую процедуру

Функция *iterGN*(). Итерационный метод Гаусса-Ньютона. Основное уравнение метода:

$$dQ = (X^t * psi_abg * X^{-1}) * X^t * psi_val;$$

Вычисляется новое приближение *tetta*, а также значения остаточных разностей z_ost , относительных приращений тетты *zel_delta*, значение среднеквадратичной ошибки SKO. Параметры: y — вектор измерений; X — матрица условий наблюдения; *tet_0* — тетта нулевое, нулевой коэффициент функции регрессии; *tetta* — коэффициенты функции регрессии; r_ost — остаточные разности; *rel_delta* — относительные приращения тетты; SKO — значение среднеквадратичной погрешности; *psitype* — признак типа функции пси. Возвращаемое значение: нет.

Алгоритм: 1. Вычисление остаточных разностей. 2. Оценка SKO — вызов *variati()*, *mediana()*. 3. Объявление векторов значений Ψ и Ψ' . 4. Вычисление значений Ψ и Ψ' — вызов *psi()*. 5. Вычисление вектора приращений Θ . 6. Вычисление вектора относительных приращений Θ . 7. Вычисление вектора Θ . 8. Возврат в вызывающую функцию *main()*.

Функция *iterMNK*(). Итерационный средневзвешенный метод наименьших квадратов. Вычисляется новое приближение *tetta*, а также значения остаточных разностей z_ost , относительных приращений тетты *zel_delta*, значение среднеквадратичной ошибки SKO. Параметры: y — вектор измерений; X — матрица условий наблюдения; *tet_0* — тетта нулевое, нулевой коэффициент функции регрессии; *tetta* — коэффициенты функции регрессии; r_ost — остаточные разности; *rel_delta* — относительные приращения тетты; SKO — значение среднеквадратичной погрешности; *psitype* — признак типа функции пси. Возвращаемое значение: нет.

Алгоритм: 1. Вычисление остаточных разностей. 2. Оценка SKO — вызов *variati()*, *mediana()*. 3. Определение вектора W . 4. Вызов *funcMNK()*. 5. Вычисление относительного приращения Θ . 6. Возврат в вызывающую функцию *main()*.

Функция *mediana*(). Определение медианы упорядоченной выборки по заданному количеству членов. Параметры: *vec* — упорядоченная выборка; *end_numb* — ограничитель, по первым *end_numb* членам определяется медиана. Возвращаемое значение: медиана выборки типа *double*.

Алгоритм: 1. $mid=end_numb/2$; $okr=[mid]$. 2. Если $okr=mid$, возврат в вызывающую процедуру значения *vec(mid)*, иначе переход к п. 3. 3. Возврат в вызывающую процедуру значения $(vec(okr)+vec(okr+1))/2$.

Функция *psi*(). Вычисление функции Ψ и ее производной Ψ' . Возвращает значение функции или ее производной в зависимости от параметра *psitype*. Параметры: *aggit* — аргумент функции; *psitype* — тип функции Ψ ; *psitype* — указывает на возвращаемое значение: функция пси (*false*) или ее производная (*true*). Возвращаемое значение: Значение Ψ или Ψ' типа *double*.

Алгоритм: 1. Если тип функции — функция Эндрюса — вычисление пси по формуле Эндрюса, возврат в вызывающую функцию. Иначе переход к п. 2. 2. Ес-

ли тип функции — функция Хампеля, вычисление пси по формуле Хампеля, возврат в вызывающую функцию. Иначе переход к п. 3. 3. Если тип функции — функция Хьюбера — вычисление пси по формуле Хьюбера, возврат в вызывающую функцию. Иначе переход к п. 4. 4. Если тип функции — функция Тьюки — вычисление пси по формуле Тьюки, возврат в вызывающую функцию. Иначе переход к п. 5. 5. Возврат значения GREAT.

Функция *sTheil*(). Получение начального приближения модифицированным методом Тейла. Величины $dm(i)$ вычисляются по варианту с меньшим количеством вычислений, т.е. по формуле (30). Параметры: y — вектор измерений типа *double* (изменяется в процессе вычислений) $ogthoX$ — ортогональная матрица условий наблюдения типа *matrix(double)*; $tetta$ — вектор оцениваемых коэффициентов регрессии (вычисляется) типа *vector(double)*; $preciz$ — требуемая точность вычислений, тип *double*. Возвращаемые значения: нет.

Алгоритм: 1. Объявления промежуточных данных. 2. Начало цикла вычислений. 3. Цикл по $m(m=1)$: если $m \leq P$, то переход на п. 4, иначе переход на п. 8. 4. Сортировка m -го столбца матрицы $ogthoX$ — вызов *variati*(). 5. Цикл вычисления $dm(i)$. 6. Вычисление $delta_tetta$ — вызов *mediana*(). 7. Выполнение итераций для m -го элемента — вызов *iteration*(). 8. Проверка сходимости — вызов *verify_good1*(). Если результат положительный, то переход на п. 9, иначе переход на п. 2. 9. Возврат в вызывающую процедуру.

Функция *TetNull*(). Вычисление Θ_0 . Параметры: y — вектор исходных измерений. Возвращаемое значение: Θ_0 типа *double*.

Алгоритм: 1. Построение вариационного ряда по вектору Y (сортировка) — вызов *variati*(). 2. Вычисление Θ_0 как медианы вариационного ряда — вызов *mediana*(). 3. Возврат в вызывающую процедуру.

Функция *to_ogtho*(). Построение ортогонального базиса методом Грама–Шмидта. Параметры: X — матрица условий наблюдений типа *matrix(double)*; $corr$ — массив аналогов коэффициента корреляции при построении ортогонального базиса (вычисляется). Возвращаемое значение: Возвращает ортогональную матрицу условий наблюдений, построенную на основе матрицы X , типа *matrix(double)*.

Алгоритм: 1. Вычисляются значения элементов ортогональной матрицы и массива $corr$ по формулам (22).

Функция *variati*(). Построение вариационного ряда (сортировка) выборки. Параметры: v — вектор выборки; end_numb — ограничительный номер, вариационный ряд строится по первым end_numb членам. Возвращаемое значение: вариационный ряд, типа *vector(double)*.

Алгоритм: 1. Объявление промежуточного вектора $vec=v$; $i = 1$. 2. $Stop=da$. 3. Если $1 < end_numb$, то п. 4, иначе п. 10. 4. Если $vec(i + 1) \geq vec(i)$ — переход к п. 9. 5. $buf=vec(0)$. 6. $vec(0)=vec(i + 1)$. 7. $vec(i + 1)=buf$. 8. $stop=нет$. 9. $i = i + 1$ — переход к п. 3. 10. Если $stop=нет$ — переход к п. 2, иначе к п. 11. 11. Возврат значения вектора vec в вызывающую процедуру.

Функция *verify_good*(). Проверка сходимости оценок тетты. Возвращает истину в случае сходимости. Параметры: $delta_zel$ — вектор модулей относительных приращений тетты; $preciz$ — требуемая точность вычислений. Возвращаемое значение: типа *Bool*.

Алгоритм: Если модуль относительного приращения хотя бы одного элемента вектора Θ превышает допустимую погрешность вычислений — возврат значения *false*, иначе — *true*.

Выводы

Показано, что в ряду различных процедур анализа измерительной информации представляется перспективным использование методов помехоустойчивого (робастного) оценивания. Обоснована возможность применения методов помехоустойчивого оценивания для задач анализа регистрируемой при испытаниях образцов измерительной информации, приведен разработанный с применением этих методов алгоритм отбраковки аномальных результатов измерений. Предлагаемый алгоритм позволяет аппроксимировать измеряемую функцию на заданном интервале полиномом заданной степени. При оценивании коэффициентов полинома используются помехоустойчивые (робастные) методы. К достоинствам алгоритма следует отнести возможность гибкого применения (настройки) используемых вычислительных процедур.

1. Додонов А.Г., Пулятин В.Г., Валетчик В.А. Обработка оптических измерений траектории летательных аппаратов // Реєстрація, зберігання і оброб. даних. — 2004. — Т. 6, № 4. — С. 38–52.
2. Ершов А.А. Стабильные методы оценки параметров. Обзор // Автоматика и телемеханика. — 1978. — № 8. — С. 66–100.
3. Стогов Г.В., Макианов А.В., Мусаев А.А. Устойчивые методы обработки измерений. Обзор // Зарубежная радиоэлектроника. — 1982. — № 9. — С. 3–46.
4. Хогг Р.В. Введение в помехоустойчивое оценивание / В кн.: Устойчивые статистические методы оценки данных. — М.: Машиностроение, 1984. — С. 86–105.
5. Аджи У.С., Тернер Р.Х. Применение методов помехоустойчивого оценивания в анализе данных о траекториях движения / В кн.: Устойчивые статистические методы оценки данных. Пер. с англ. — М.: Машиностроение, 1984. — С. 12–26.
6. Алдонин Г.М. Робастность в природе и технике. — М.: Радио и связь, 2003. — 336 с.
7. Хьюбер П. Робастность в статистике / Пер. с англ. — М.: Мир, 1984. — 304 с.
8. Мактанов А.В., Мусаев А.А. Робастные методы обработки результатов измерений: Учебное пособие: Радиоэлектронные системы и комплексы. — Вып. 3. — М.: МО СССР, 1980. — 144 с.
9. Huber P.J. Robust Regression: Asymptotics, Conjectures and Monte Carlo // Ann. Math. Statist. — 1973. — N 1. — P. 799–821.
10. Huber P.J. Robust Statistics: A Review // Ann. Math. Statist. — 1972. — Vol. 43. — P. 1041–1067.
11. Лемешко Б.Ю. Робастные методы оценивания и отбраковка аномальных измерений // Заводская лаборатория. — 1997. — Т. 63, № 5. — С. 43–49.
12. Лемешко Б.Ю., Чимитова Е.В. Оптимальные L -оценки параметров сдвига и масштаба распределений по выборочным квантилям // Заводская лаборатория. Диагностика материалов. 2004. — Т. 70, № 1. — С. 54–66.
13. Линник Ю.В. Метод наименьших квадратов и основы обработки наблюдений. — М.: Наука, 1962. — 269 с.

14. Поляк Б.Т. Устойчивые методы оценки параметров / В сб.: Структурная адаптация сложных систем управления. — Воронежский политехнический ин-т, 1977. — С. 66–71.
15. Тьюки Дж. Анализ результатов наблюдений. — М.: Мир, 1981. — 692 с.
16. Микешина Н.Г. Выявление и исключение аномальных значений // Заводская лаборатория. — 1966. — № 3. — С. 185–198.
17. Agee W.S., Nurner R.H. Robust Regression: M — Estimated. Tehnical Report. — White Sand and Missile Range, 1978.
18. Andrews D.F. A Robust Method of Multiple Linear Regression. — Technjmetrics, 1979. — Vol. 16, Nov. — P. 523–531.
19. Грановский В.А., Сиряя Т.Н. Методы обработки экспериментальных данных при измерениях. — Л.: Энергоатомиздат, 1990. — 288 с.
20. Фомин А.В., Новоселов О.Н., Плющев А.В. Отбраковка аномальных результатов измерений. — М.: Энергоатомиздат, 1985. — 198 с.

Поступила в редакцию 11.01.2008