

УДК 004.62

Сравнение методов выявления аномальных выбросов в исходных данных и их применение при обработке данных искусственной вентиляции легких

С. Г. Климанов¹, А. А. Котляров¹, А. В. Крянев^{1,2}, В. Х. Тимербаев³, В. А. Трикозова¹, Д. Д. Царева¹, Н. В. Воронков¹, Э. М. Мухамедзянов¹, А. А. Умаров¹, Ф. С. Утин¹, Е. А. Чилека¹

¹ Национальный исследовательский ядерный университет «МИФИ», Москва, 115409, Россия

² Объединенный институт ядерных исследований, Дубна, 141980, Россия

³ Сеть медицинских центров «MedSwiss», Москва, Россия

В статье рассматривается проблема выявления аномальных выбросов и нивелирования их отрицательного влияния на оценки выделяемых характеристик рассчитываемых показателей. Для решения поставленной задачи в статье рассматриваются различные применяемые на практике робастные методы и основанные на них вычислительные схемы выделения аномальных выбросов в значениях исследуемых показателей. Проведено сравнение эффективности выделения аномальных выбросов различными методами для выборок случайных величин с нормальным законом распределения для различных вариантов числа и расположения аномальных выбросов по отношению к неаномальным значениям с помощью стандартных датчиков случайных величин. Исследуемые в статье робастные методы и схемы используются для выявления аномальных выбросов в показаниях пациентов и параметров вентиляционного потока при искусственной вентиляции легких (ИВЛ). Проведенные численные эксперименты, результаты которых приведены в настоящей статье, показали, что наиболее эффективным методом выявления аномальных выбросов при неизвестной дисперсии основной части неаномальных данных является разработанный в НИЯУ МИФИ модифицированный метод Хьюбера. Этот метод позволяет из сформированной базы данных клинического опыта лечения пациентов на аппаратах ИВЛ эффективно выделять аномальные выбросы, что дает возможность использовать этот метод для создания устойчивой схемы выбора оптимальных значений показателей вентиляционного потока в зависимости от значений показателей текущего состояния пациента.

Ключевые слова: аномальные выбросы, выявление аномальных выбросов, робастные методы, искусственная вентиляция легких, показания пациентов.

Введение

В исходных данных многих задач имеется, как правило, небольшая часть общего количества этих данных, численные значения которых не соответствуют общей тенденции рассматриваемого процесса, и которые принято называть аномальными выбросами. Причины появления аномальных выбросов могут быть разными, но в большинстве случаев ими являются сбои различной природы при фиксации этих значений. В то же время, в силу значительных показателей больших выбросов, их влияние на результаты обработки исходных данных без выделения аномальных выбросов будет, как правило, огромным и приведет к большим погрешностям в значениях показателей результатов обработки. Для выявления аномальных выбросов и нивелирования их вредного влияния на конечные результаты обработки исходных данных различными авторами в течение последних десятилетий были разработаны несколько методов и основанных на них вычислительных схем выявления аномальных выбросов. Целью настоящей работы

✉ А. В. Крянев: AVKryanev@mephi.ru

Поступила в редакцию: 14.12.2024

После доработки: 28.12.2024

Принята к публикации: 14.01.2024

EDN LBYWBJ

является сравнение эффективности выявления аномальных выбросов различными существующими методами, определение среди них наиболее эффективных при обработке данных ИВЛ и применение наиболее эффективного для выявления аномальных выбросов в значениях текущих показаний пациента при ИВЛ.

Робастные методы выявления аномальных выбросов

Ниже приведены математические методы и основанные на них численные алгоритмы выявления и фильтрации аномальных выбросов и сравнение их эффективности [1–7]. Математически эта задача сводится к выделению из совокупности зашумленных значений рассматриваемой случайной величины аномальных значений.

Метод Андрюса

Широкий спектр робастных оценок порожден ММП-схемой, согласно которой минимизируется выражение: $-\ln(p(X_i; u)) = \rho\left(\frac{X_i - u}{\sigma}\right)$. Тогда ММП-оценка $\hat{u}$ -параметра сдвига u минимизирует сумму:

$$\hat{u} = \text{Arg} \min \sum_{i=0}^n \rho\left(\frac{X_i - u}{\sigma}\right), \quad (1)$$

где минимум берется по u , где $p(X_i, u)$ – робастная плотность, $\rho\left(\frac{X_i - u}{\sigma}\right) = -\ln(p(X_i; u))$, а функции $\rho(s)$ – четная выпуклая положительная функция и $\rho(0) = 0$.

Чтобы M -оценка была робастной по отношению к большим выбросам, если исходный закон распределения нормальный, необходим более медленный рост функции $\rho(s)$ по сравнению с квадратичной функцией, по крайней мере, для больших по модулю s , и близкая к квадратичной зависимость для малых по модулю s .

Робастная оценка параметра положения u , порожденная функцией

$$\rho(s) = \begin{cases} a \left(1 - \cos \frac{s}{a}\right), & |s| \leq \pi a, \\ 2a, & |s| > \pi a, \end{cases} \quad (2)$$

называется оценкой Андрюса. При уменьшении a «степень робастности» оценки Андрюса увеличивается.

Оценка Андрюса является нелинейной. Для нахождения этой оценки применяется несколько итеративных методов, чаще всего итеративный метод наименьших квадратов. Запишем минимизируемый функционал в виде:

$$\sum_{i=0}^n (X_i - u)^2 W_i, \text{ где } W_i = \frac{1}{(X_i - u)^2} \sum_{i=0}^n \rho\left(\frac{X_i - u}{\sigma}\right)$$

Если веса W_i фиксированы, то $\hat{u} = \frac{\sum_{i=0}^n X_i W_i}{\left(\sum_{i=0}^n W_i\right)^{-1}}$, σ – стандартное отклонение.

Оценка Андрюса определяется согласно итерационному процессу, который сходится за конечное число итераций:

$$\hat{u}_{j+1} = \frac{\sum_{i=0}^n X_i W_{ij}}{\left(\sum_{i=0}^n W_{ij}\right)^{-1}}, \text{ где } W_{ij} = \frac{1}{(X_i - u_j)^2} \sum_{i=0}^n \rho\left(\frac{X_i - u_j}{\sigma}\right). \quad (3)$$

Данные для обработки были сгенерированы с помощью датчика нормально распределенной случайной величины, в них добавлялись вручную аномальные выбросы. Эта схема использовалась для получения исходных данных, представленных на рис. 1, 3, 5, 7, 9, 11. Были взяты выборки размером 100, 500 1000 элементов. Размер выбросов – 1, 5 и 10 % для каждого размера выборки. Метод показывает себя чуть хуже на данных, в которых выбросы находятся только сверху или только снизу относительно математического ожидания, однако оценка все равно является намного лучше обычного среднего, которое в данном случае будет сильно смещено в сторону выбросов. Можно сделать вывод об эффективности метода с проецированием по сравнению с методом без проецирования.

На рис. 1 приведен один из примеров исходных данных реализации значений нормально распределенной случайной величины с четырьмя аномальными выбросами (АВ), причем несимметрично относительно математического ожидания (три – с большими значениями, а один – с меньшим значением). Все четыре выброса были выявлены с помощью представленного выше итерационного процесса (3), – и на рис. 2 представлен итог фильтрации. На оси абсцисс рис. 1 – 6 отложен номер члена выборки, а на оси ординат значения реализаций нормально распределенных случайных величин, полученных с помощью датчика случайных величин, с последующим наложением аномальных выбросов до и после фильтрации.

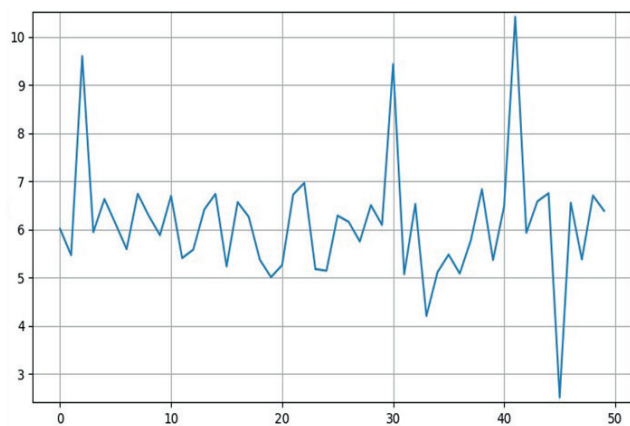


Рис. 1. Выборка с четырьмя аномальными выбросами

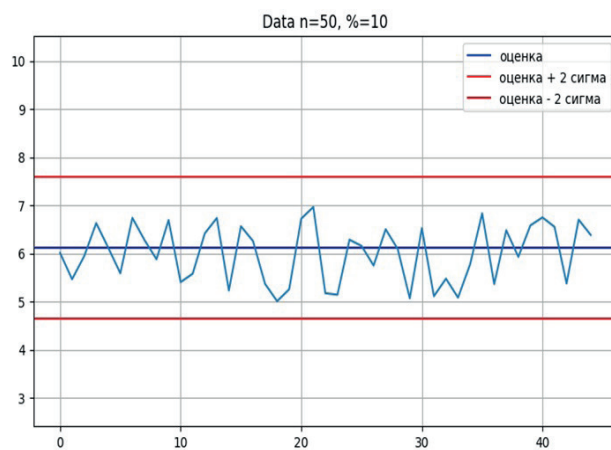


Рис. 2. Отфильтрованная выборка

По совокупности полученных результатов можно сделать вывод об эффективности метода Андруса для фильтрации аномальных выбросов. Метод показывает себя хуже на данных, в которых большая часть выбросов находится только сверху или только снизу относительно математического ожидания. Кроме того, схема метода Андруса с проецированием аномальных выбросов на границы $u - K\sigma_p$, $Y_i + K\sigma_i$ по сравнению с методом без проецирования является более эффективной. Эти два замечания остаются в силе и для других методов выявления аномальных выбросов, представленных в статье ниже.

Метод Рамсея

В методе Рамсея робастная ММП-оценка порождена функцией

$$\rho(s) = \frac{1}{\lambda^2} [1 - (1 + \lambda |s|) e^{-\lambda |s|}]. \quad (4)$$

Параметр λ обычно выбирается так, чтобы точка перегиба $\rho(s)$ функции равнялась 3. При увеличении λ «степень робастности» оценки Рамсея увеличивается.

Оценка Рамсея $\hat{u}$ является нелинейной. Для нахождения оценки в таком случае применяется несколько итеративных методов. Чаще всего используется итеративный метод наименьших квадратов.

Запишем минимизируемый функционал (4) в виде

$$\sum_{i=0}^n (X_i - u)^2 W_i, \quad (5)$$

где

$$W_i = \frac{1}{(X_i - u)^2} \sum_{i=0}^n \rho\left(\frac{X_i - u}{\sigma}\right). \quad (6)$$

Если веса W_i фиксированы, то

$$\hat{u} = \sum_{i=0}^n X_i W_i \left(\sum_{i=0}^n W_i \right)^{-1}. \quad (7)$$

Определим итерационный процесс:

$$\hat{u}_{j+1} = X_i W_{ij} \left(\sum_{i=0}^n W_{ij} \right)^{-1}, \text{ где } W_{ij} = \frac{1}{(X_i - \hat{u}_j)^2} \sum_{i=0}^n \rho\left(\frac{X_i - \hat{u}_j}{\sigma}\right). \quad (8)$$

Данные для обработки были сгенерированы с помощью датчика нормально распределенной случайной величины. Были взяты выборки размером 100, 500 1000 элементов. Размер выбросов – 1, 5 и 10 % для каждого размера выборки. Выбросы генерировались отдельно.

На рис. 3 приведен один из примеров исходных данных реализации значений нормально распределенной случайной величины с двумя аномальными выбросами, причем несимметрично относительно математического ожидания (оба с меньшим значением). Все два выброса были выявлены, и на рис. 4 представлен итог фильтрации.

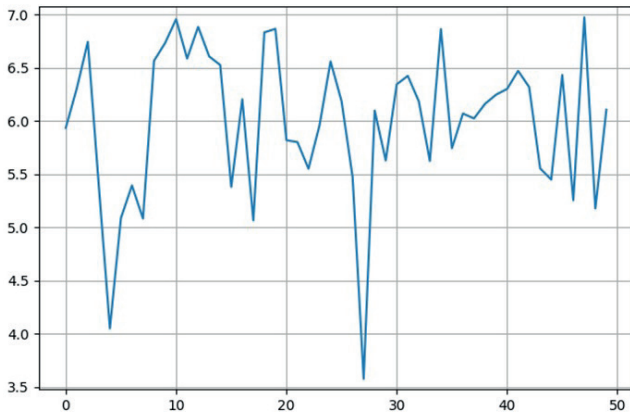


Рис. 3. Выборка с двумя аномальными выбросами

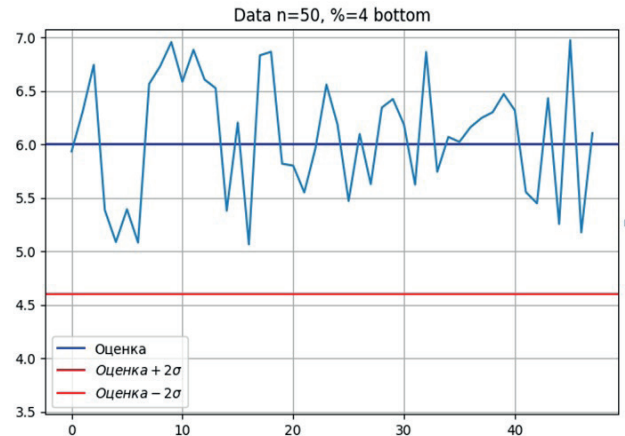


Рис. 4. Отфильтрованная выборка

По совокупности полученных результатов можно сделать вывод об эффективности метода Рамсея для очистки данных от выбросов. Метод иногда показывает результаты хуже на данных, в которых большая часть выбросов находится только сверху или только снизу относительно математического ожидания.

Метод усеченной средней

Одним из самых ранних методов выявления аномальных выбросов и вычисления робастной оценки параметра положения u является метод усеченной средней:

$$\hat{u} = \bar{X}_{yc} = \frac{1}{n - 2m} \sum_{i=m+1}^{n-m} X_i, \quad (9)$$

где $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ — упорядоченные по возрастанию члены выборки $X_1, \dots, X_n$, а $m \leq [n/2]$. Из формулы следует, что усеченная средняя получена отбрасыванием m самых левых и m самых правых членов выборки $X_1, \dots, X_n$ и усреднением $(n - 2m)$ оставшихся.

Этот метод не является достаточно эффективным при выявлении выбросов из выборок данных, где аномальные значения расположены несимметрично по отношению к среднему значению. Поэтому использовалась модификация данного метода: сначала вычисляется количество аномальных значений, которые находятся за пределами интервала $[\bar{X} - k\sigma, \bar{X} + k\sigma]$, где $\bar{X}$ – среднее значение выборки, k – параметр, задающий ширину интервала; затем из упорядоченной по возрастанию выборки удаляются m самых меньших и p самых больших членов выборки.

Имеем итоговую оценку усеченного среднего:

$$\hat{u} = \bar{X}_{yc} = \frac{1}{n - m - p} \sum_{i=m+1}^{n-p} X_i, \quad (10)$$

где $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ – упорядоченные по возрастанию члены выборки $X_1, \dots, X_n$, m – количество элементов выборки, меньших $\bar{X} - k\sigma$; p – количество элементов выборки, превышающих $\bar{X} + k\sigma$. Из формулы следует, что усеченная средняя получена отбрасыванием m самых левых и p самых правых членов выборки $X_1, \dots, X_n$ и усреднением $(n - m - p)$ оставшихся.

Было проведено сравнение метода усеченной средней с методом Хьюбера при 2σ и 3σ . Основное отличие состоит в том, что для выявления всех выбросов метод усеченной средней необходимо применять итеративно несколько раз (два раза для интервала 2σ и больше трех раз для интервала 3σ).

Метод Винзора

Одним из эффективных методов борьбы с выбросами и получения устойчивой оценки по отношению аномальных выбросов является метод Винзора

$$\hat{u} = \hat{X} = \frac{1}{n} \left(\sum_{i=m+1}^{n-m} X_{(i)} + m(X_{(m+1)} + X_{(n-m)}) \right). \quad (11)$$

В отличие от усеченной средней, в оценке Винзора (11) крайние значения выборки не отбрасываются, а заменяются: m самых малых значений заменяются на $(m+1)$ -е, а m самых больших значений – на $(n-m)$ -е. Оценка Винзора похожа на минимаксную оценку Хьюбера, но отличается от нее тем, что просеивание крайних значений выполняется до вычисления оценки и заранее известно при заданном m . Увеличение m повышает степень робастности оценки Винзора. При численном тестировании оценки Винзора было создано 27 выборок данных с использованием генератора случайных чисел на Python: рассматривались три разных объема выборок – 100, 500 и 1000 и три уровня засорения аномальными выбросами: 1, 5 и 10 % с разными вариантами их расположения относительно математического ожидания.

Сравнение метода Винзора с методом Андрюса показало, что метод Андрюса показывает иногда результаты хуже, чем метод Винзора на данных, в которых выбросы находятся только с одной стороны по отношению к среднему значению.

Метод Хьюбера

Робастная оценка параметра положения метода Хьюбера определяется равенством:

$$\hat{u} = \frac{1}{n} \left[\sum_{i \in I_0} X_i + (\hat{n}_+ + \hat{n}_-) \hat{u} + K\sigma(\hat{n}_+ - \hat{n}_-) \right], \quad (12)$$

где $\hat{I}_+$, $\hat{I}_-$, $\hat{I}_0$ определены равенствами

$$\hat{I}_+ = \{i : X_i - \hat{u} > K\sigma\}, \quad \hat{I}_- = \{i : X_i - \hat{u} < -K\sigma\}, \quad (13)$$

$\hat{I}_0 = \{i : |X_i - \hat{u}| \leq K\sigma\}$, $\hat{n}_+$, $\hat{n}_-$ – количество индексов i , принадлежащих множествам $\hat{I}_+$, $\hat{I}_-$ соответственно. Для нахождения оценки Хьюбера использовался итерационный процесс:

$$\hat{u}_{j+1} = \frac{1}{n} \left[\sum_{i \in I_{0j}} X_i + (\hat{n}_{+j} + \hat{n}_{-j}) \hat{u}_j + K\sigma(\hat{n}_{+j} - \hat{n}_{-j}) \right], \quad (14)$$

где $\hat{I}_{+j} = \{i : X_i - \hat{u}_j > K\sigma\}$, $\hat{I}_{-j} = \{i : X_i - \hat{u}_j < -K\sigma\}$, $\hat{I}_{0j} = \{i : |X_i - \hat{u}_j| \leq K\sigma\}$, $\hat{n}_{+j}$, $\hat{n}_{-j}$ – количество индексов i , принадлежащих множествам $\hat{I}_{+j}$, $\hat{I}_{-j}$ соответственно.

В качестве нулевого приближения можно брать среднее арифметическое $\hat{u}_0 = \frac{1}{n} \sum_{i=1}^n X_i$.

Метод Хьюбера позволяет за конечное число шагов удалить все выбросы при их равномерном распределении с интервалом $\pm 2\sigma$. Однако при этом метод Хьюбера иногда дает худшие результаты, когда большинство выбросов находятся с одной стороны от среднего значения, а процент аномальных выбросов больше 5 %.

Модифицированный метод Хьюбера

Пусть рассматриваемый показатель состояния пациента u принимает значения X_i для $i=1, \dots, n$.

$$\hat{u}_1 = \frac{1}{n} \sum_{i=1}^n X_i, \quad \sigma_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{u}_1)^2$$

Определим вектор $\vec{X}^{(1)} = (X_1, \dots, X_n)$ равенством

$$X_i^{(1)} = \begin{cases} X_i, & \text{если } i \in I_0^{(1)}, \\ \hat{u}_1 - K\sigma_1, & \text{если } i \in I_+^{(1)}, \\ \hat{u}_1 + K\sigma_1, & \text{если } i \in I_-^{(1)}, \end{cases} \quad (15)$$

где

$$\begin{aligned} I_0^{(1)} &= \{i : |X_i - \hat{u}_1| \leq K\sigma_1\}, \\ I_+^{(1)} &= \{i : X_i - \hat{u}_1 > K\sigma_1\}, \\ I_-^{(1)} &= \{i : X_i - \hat{u}_1 < -K\sigma_1\}, \end{aligned} \quad (16)$$

K – так называемый параметр Хьюбера, значение которого зависит от доли больших выбросов, и в данной работе взято $K=1.8$.

Реализуем итерационный процесс:

$$\hat{u}_{l+1} = \frac{1}{n} \sum_{i=1}^n X_i^{(l+1)}, \quad (17)$$

$$\sigma_{l+1}^2 = \frac{1}{n} \sum_{i=1}^n (X_i^{(l+1)} - \hat{u}_{l+1})^2, \quad (18)$$

$$X_i^{(l+1)} = \begin{cases} X_i^{(l)}, & \text{если } i \in I_0^{(l+1)} \\ X_i^{(l)} - K\sigma_l, & \text{если } i \in I_+^{(l+1)} \\ X_i^{(l)} + K\sigma_l, & \text{если } i \in I_-^{(l+1)} \end{cases} \quad (19)$$

где

$$\begin{aligned} I_0^{(l+1)} &= \{i : |X_i^{(l)} - \hat{X}_i^{(l)}| \leq K\sigma_l\}, \\ I_+^{(l+1)} &= \{i : X_i^{(l)} - \hat{X}_i^{(l)} > K\sigma_l\}, \\ I_-^{(l+1)} &= \{i : X_i^{(l)} - \hat{X}_i^{(l)} < -K\sigma_l\}, \end{aligned} \quad (20)$$

Введем норму в пространстве E^n :

$$\|\vec{X}\| = \max_{i=1, \dots, n} |X_i|. \quad (21)$$

Тогда, задав приемлемый уровень допустимой погрешности $\varepsilon > 0$, продолжаем описанный итерационный процесс, пока не будет выполнено условие

$$\|\vec{X}^{(l+1)} - \vec{X}^{(l)}\| < \varepsilon. \quad (22)$$

Аналогичные робастные схемы используются для выделения с помощью робастных методов временных трендов при наличии аномальных выбросов в значениях показателей [4–7].

На рис. 5 приведен один из примеров исходных данных реализации значений нормально распределенной случайной величины с тремя аномальными выбросами, причем несимметрично относительно математического ожидания (все 3 с меньшим значением). Все три выброса были выявлены с помощью всех вышеперечисленных методов, и на рис. 6 представлен итог фильтрации.

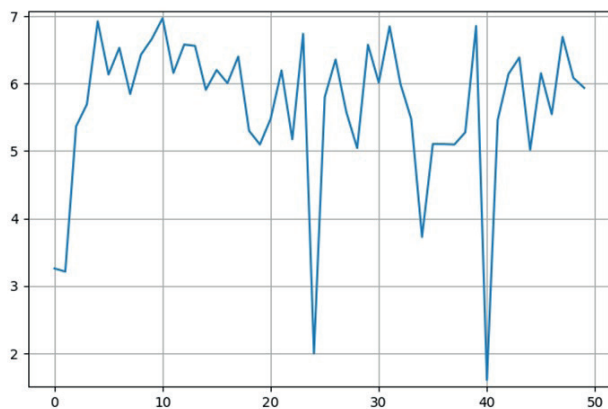


Рис. 5. Выборка с тремя аномальными выбросами

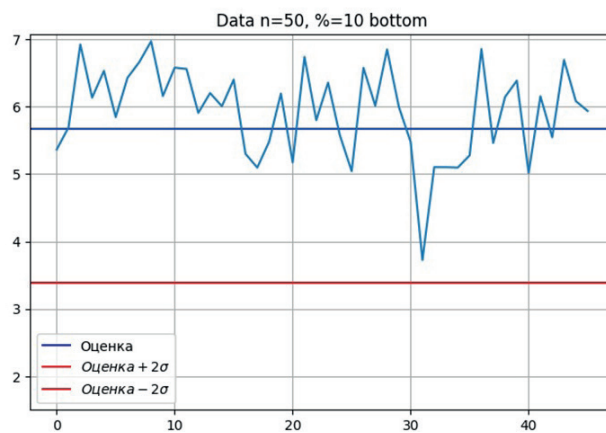


Рис. 6. Отфильтрованная выборка

Было проведено сравнение метода усеченной средней с методом Хьюбера при 2σ и 3σ . Основное отличие состоит в том, что для выявления всех выбросов метод усеченной средней необходимо применять итеративно несколько раз (два раза для интервала 2σ и больше трех раз для интервала 3σ).

Модифицированный метод Хьюбера позволяет за конечное число шагов удалить все выбросы, когда процент аномальных выбросов не превышает 20 %. При этом модифицированный метод Хьюбера иногда дает худшие результаты, когда все выбросы находятся с одной стороны от среднего значения, а процент аномальных выбросов больше 20 %, что при обработке данных прикладных задач встречается редко.

Выявление аномальных выбросов в значениях показателей текущего состояния пациента, подключенного к аппарату ИВЛ

Ниже для иллюстрации эффективности методов выявления аномальных выбросов представлены результаты их фильтрации для некоторых значимых показателей состояния пациента, подключенного к аппарату ИВЛ.

На рис. 7 представлены исходные фиксируемые значения во времени показателя «Минутная вентиляция легких» (MV). На оси абсцисс рисунков 7–18 отложено время фиксации показателя в минутах, а на оси ординат этих рисунков – значения показателей до и после фильтрации аномальных выбросов.

На рис. 8 – отфильтрованные с помощью метода Андрюса значения во времени показателя «Минутная вентиляция легких» (MV), представленные на рис. 7.

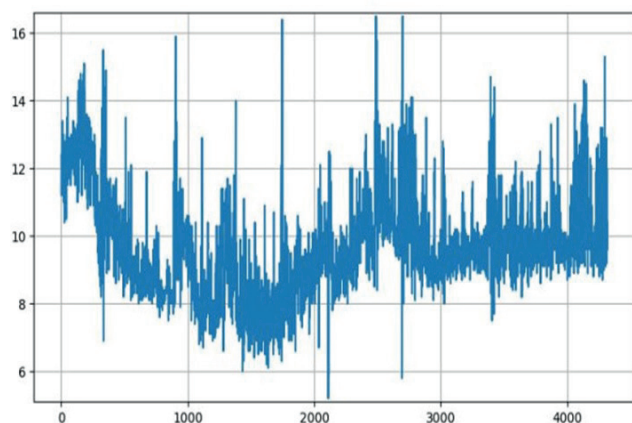


Рис. 7. Показатель MV

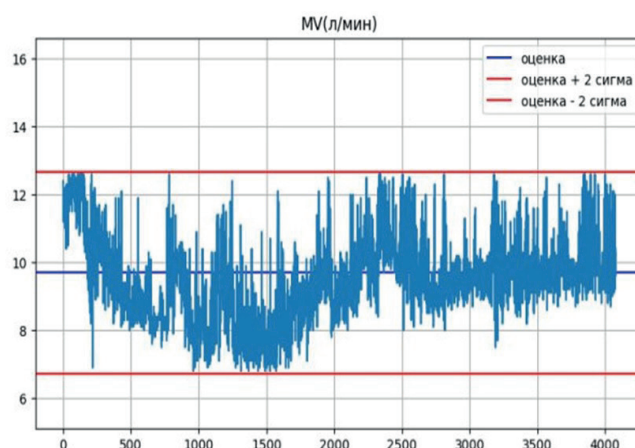


Рис. 8. Отфильтрованный показатель MV

На рис. 9 представлены исходные фиксируемые значения во времени показателя « $f_{\text{общ}}$ – общее количество дыхательных циклов».

На рис. 10 – отфильтрованные с помощью метода Рамсея значения во времени показателя « $f_{\text{общ}}$ – общее количество дыхательных циклов», представленные на рис. 9.

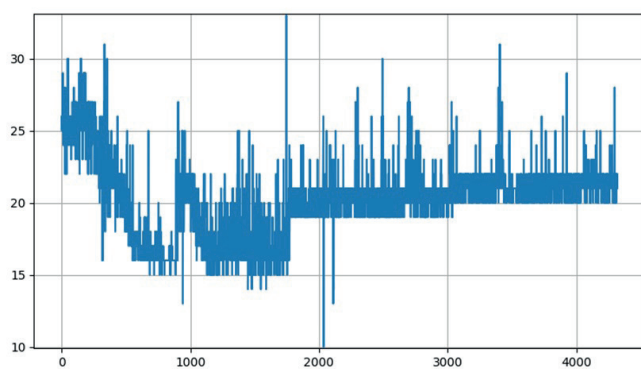


Рис. 9. Показатель $f_{\text{общ}}$

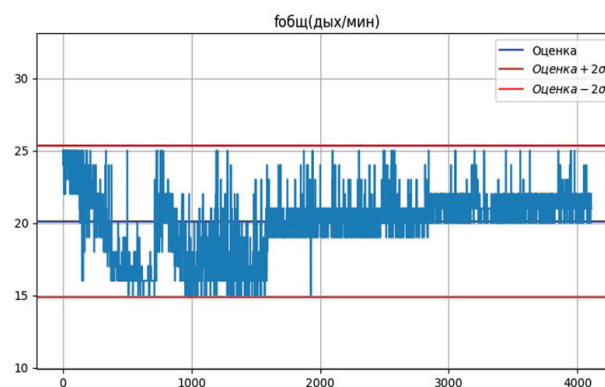


Рис. 10. Отфильтрованный показатель $f_{\text{общ}}$

На рис. 11 представлены исходные фиксируемые значения во времени показателя TVi (мл) – «Объем вдоха».

На рис. 12 – отфильтрованные с помощью метода «Усеченной средней» значения во времени показателя TVi (мл) – «Объем вдоха», представленные на рис. 11.

На рис. 13 представлены исходные фиксируемые значения во времени показателя WOB (Дж/мин) – «Работа дыхания».

TVi (мл), 2sigma

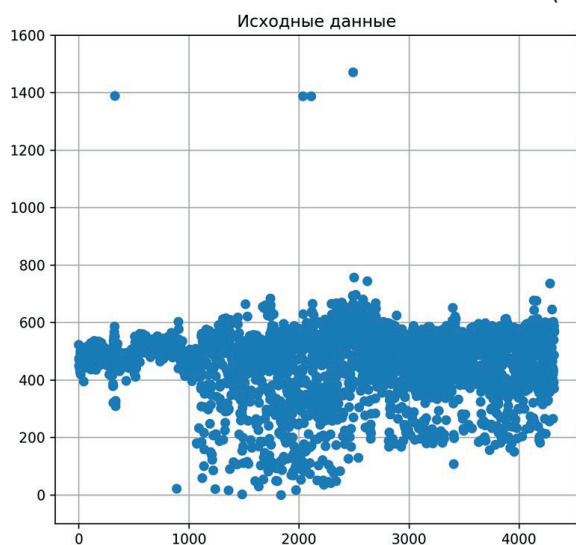


Рис. 11. Показатель TVi

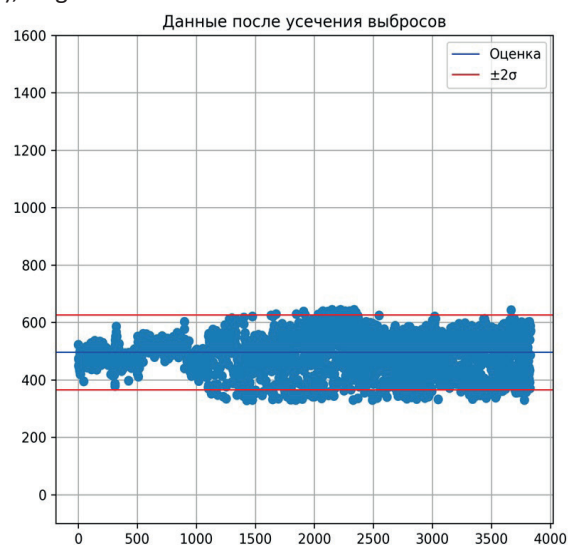


Рис. 12. Отфильтрованный показатель TVi

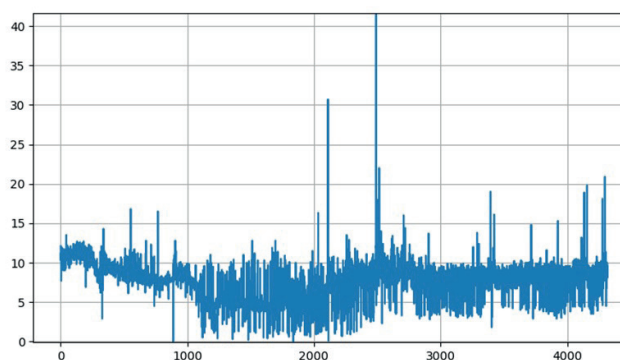


Рис. 13. Показатель WOB

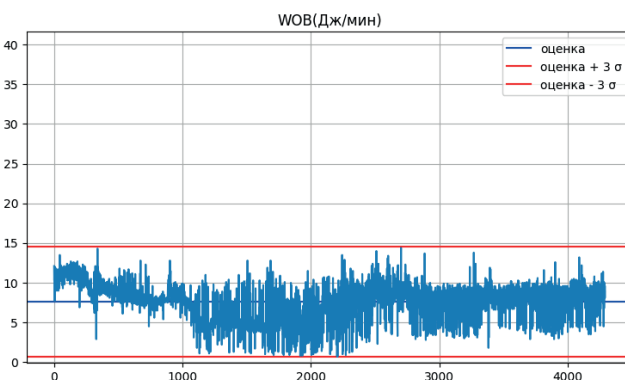


Рис. 14. Отфильтрованный показатель WOB

На рис. 14 – отфильтрованные с помощью метода Винзора значения во времени показателя «WOB (Дж/мин) – «Работа дыхания», представленные на рис. 13.

На рис. 15 представлены исходные фиксируемые значения во времени показателя $P_{\text{пик}}$ (см. H_2O) – максимальное давление в дыхательных путях во время вдоха.

На рис. 16 – отфильтрованные с помощью метода Хьюбера значения во времени показателя $P_{\text{пик}}$, представленные на рис. 15.

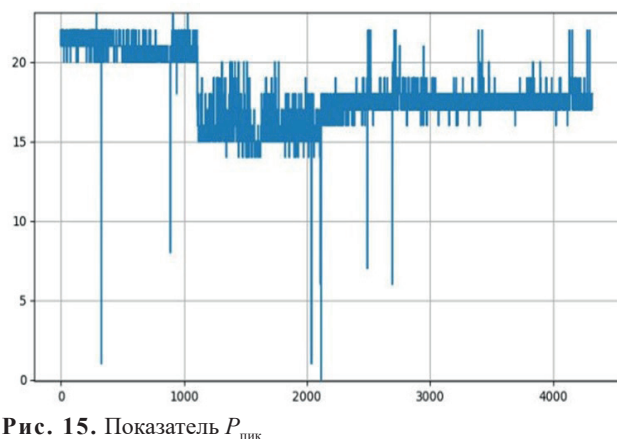


Рис. 15. Показатель $P_{\text{пик}}$

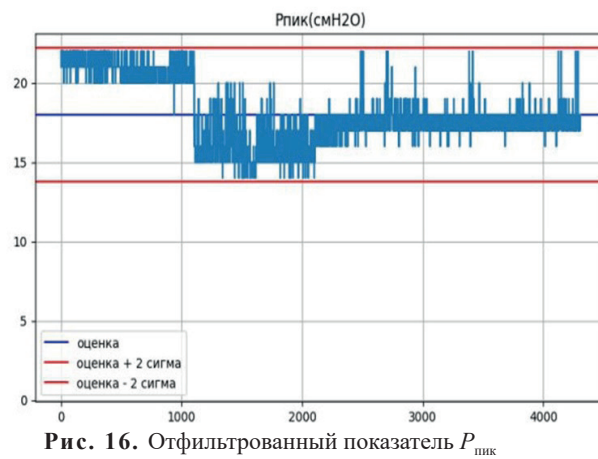


Рис. 16. Отфильтрованный показатель $P_{\text{пик}}$

На рис. 17 представлены исходные фиксируемые значения во времени показателя Re (см. вод ст/л/с) – сопротивление дыхательных путей выдоху.

На рис. 18 – отфильтрованные с помощью модифицированного метода Хьюбера значения во времени показателя Re , представленные на рис. 17.

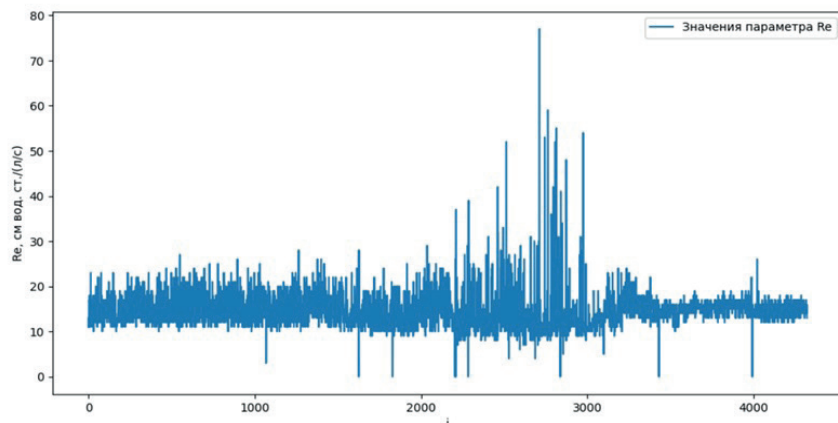


Рис. 17. Показатель Re

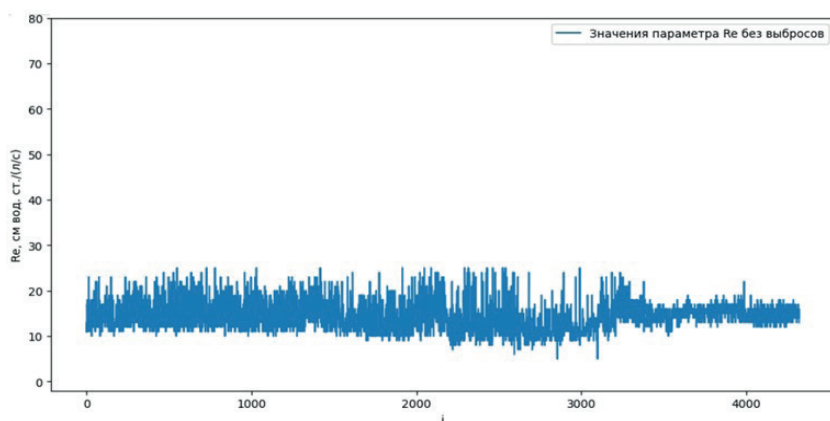


Рис. 18. Отфильтрованный показатель Re

Анализ полученных результатов и выводы

Проведенные исследования по сравнению существующих робастных методов выявления аномальных выбросов, результаты которых приведены в настоящей статье, показали, что наиболее эффективным из них при неизвестной дисперсии основной части данных, за исключением аномальных выбросов, является разработанный в НИЯУ МИФИ модифицированный метод Хьюбера, который позволяет на основе сформированной базы данных клинического опыта лечения пациентов на аппаратах ИВЛ и использования методов искусственных нейронных сетей или методов метрического анализа [8–13], устанавливать устойчивую зависимость между совокупностью значений показателей текущего состояния пациента и оптимальными значениями регулируемых параметров вентиляционного потока аппарата ИВЛ и прогнозировать будущее состояние пациента.

Финансирование

Работа частично выполнена при поддержке Министерства науки и высшего образования Российской Федерации (проект государственного задания № FSWU-2023-0031).

Вклад авторов

Климанов С.Г. – участие в проведении исследования и обсуждении результатов, анализ экспериментальных данных;

Котляров А.А. – участие в формулировке идеи и целей исследования, участие в проведении исследования и обсуждении результатов.

Крянев А.В. – участие в формулировке идеи и целей исследования и обсуждении результатов, постановка задачи, выбор методов исследования, анализ экспериментальных данных, выявление закономерностей, подбор литературных источников, редактирование текста статьи.

Тимербаев В.Х. – участие в формулировке идеи и целей исследования, предоставление клинических данных показаний состояний пациентов, участие в проведении исследования и обсуждении результатов.

Трикозова В.А. – выбор методов исследования, анализ экспериментальных данных, участие в составлении части программного кода и проведение с его помощью серии численных расчетов, включая анализ их результатов.

Царева Д.Д. – выбор методов исследования, анализ экспериментальных данных, участие в составлении части программного кода и проведение с его помощью серии численных расчетов, включая анализ их результатов.

Воронков Н.В. – участие в составлении части программного кода и проведение с его помощью серии численных расчетов, включая анализ их результатов.

Мухамедзянов Э.М. – участие в составлении части программного кода и проведение с его помощью серии численных расчетов, включая анализ их результатов.

Умаров А.А. – участие в составлении части программного кода и проведение с его помощью серии численных расчетов, включая анализ их результатов.

Утин Ф.С. – участие в составлении части программного кода и проведение с его помощью серии численных расчетов, включая анализ их результатов.

Чилека Е.А. – участие в составлении части программного кода и проведение с его помощью серии численных расчетов, включая анализ их результатов.

Конфликт интересов

Конфликт интересов отсутствует.

Список литературы

1. Смоляк С.А., Титаренко Б.П. Устойчивые методы оценивания. М.: Статистика, 1980. 254 с.
2. Хьюбер П. Робастность в статистике. М.: Мир, 1984. 303 с.
3. Арсенин В.Я., Крянев А.В., Цупко-Ситников М.В. Применение робастных методов для решения некорректных задач // ЖВММФ, 1989, Т. 29. № 5. С. 653–661.
4. Хампель М., Рончетти Э., Рауссеу П., Штаэль В. Робастность в статистике. Подход на основе функций влияния. М.: Мир, 1989. 512 с.
5. Цыпкин Я.З. Информационная теория идентификации. М.: Наука, 1995. 336 с.
6. Шурыгин А.М. Прикладная статистика: робастность, оценивание, прогноз. М.: Финансы и статистика, 2000. 223 с.
7. Крянев А.В., Лукин Г.В., Удумян Д.К. Метрический анализ и обработка данных. М.: Физматлит, 2012. 308 с.
8. Ананьев Е.П., Полупан А.А., Мацковский И.В. [и др.]. Использование режима IntelliVent-ASV для поддержания целевого диапазона EtCO₂ у пациентов с тяжелой ЧМТ // Журнал «Вопросы нейрохирургии» имени Н.Н. Бурденко, 2017. Т. 81 (5). С. 63–68.
9. Artificial intelligence and machine learning show promise in cancer diagnosis and treatment // Medical Xpress. [Электронный ресурс]. URL: <https://medicalxpress.com/news/2022-03-artificial-intelligence-machine-cancer-diagnosis.html> [дата обращения 20.11.2024].
10. Kryanev A.V., Udumyan D.K. Metric analysis, properties and applications as a tool for interpolation // International Journal of Mathematical Analysis, 2014. V. 8. № 45. P. 2221–2228.
11. Kryanev A.V., Udumyan D. K. Metric Analysis, Properties and Applications as a Tool for Forecasting // International Journal of Mathematical Analysis, 2014. V. 8. №. 60. P. 2971–2978.
12. Ivanov V.V., Kryanev A.V., Udumyan D.K., Lukin G.V. Metric Analysis Approach for Interpolation and Forecasting of Time Processes // Applied Mathematical Sciences, 2014. V. 8. № 22. P. 1053–1060.
13. Климанов С.Г., Крянев А.В., Трикозова В.А., Царева Д.Д. Вычислительные схемы выявления аномальных выбросов в значениях текущих показаний пациента при искусственной вентиляции легких // Вестник НИЯУ МИФИ, 2024. Т. 13. № 2. С. 76–82.

Comparison of methods for detecting abnormal emissions in the original data and their application in processing artificial ventilatory data

S. G. Klimanov¹, A. A. Kotlyarov¹, A. V. Kryanev^{1,2,✉}, V. Kh. Timerbaev³, V. A. Trikozova¹, D. D. Tsareva¹,
N. V. Voronkov¹, E. M. Mukhamedzyanov¹, A. A. Umarov¹, F. S. Utin¹, E. A. Chileka¹

¹ National Research Nuclear University “MEPhI”, Moscow, 115409, Russia,

² Joint Institute for Nuclear Research, Dubna, 141980, Russia,

³ Network of medical centers MedSwiss, Moscow, Russia

✉ AVKryanev@mephi.ru

Received December 14, 2024; revised December 28, 2024; accepted January 14, 2025

The article considers the problem of detecting abnormal outliers and mitigating their negative impact on the estimates of the allocated characteristics of the calculated indicators. To solve the problem, the article considers various robust methods used in practice and computational schemes for detecting abnormal outliers in the values of the studied indicators based on them. A comparison is made of the efficiency of detecting abnormal outliers by various methods for samples of random variables with a normal distribution law for various options for the number and location of abnormal outliers in relation to non-anomalous values using standard random variable sensors. The robust methods and schemes studied in the article are used to detect abnormal outliers in patient readings and ventilation flow parameters during artificial lung ventilation (ALV). The numerical experiments conducted, the results of which are presented in this article, showed that the most effective method for detecting abnormal outliers with an unknown variance of the main part of non-anomalous data is the modified Huber method developed at MEPhI. This method allows us to effectively identify abnormal emissions from the formed database of clinical experience in treating patients on ventilators, which makes it possible to use this method to create a stable scheme for selecting optimal values of ventilation flow indicators depending on the values of the indicators of the patient's current condition.

Keywords: abnormal outliers, identifying abnormal outliers, robust methods, artificial lung ventilation, patient readings.

References

1. Smolyak S.A., Titarenko B.P. Ustojchivye metody ocenivaniya [Robust estimation methods]. Moscow, Statistika Publ., 1980. 254 p.
2. Huber P. Robastnost' v statistike [Robustness in statistics]. Moscow, Mir Publ., 1984. 303 p.
3. Arsenin V.Ya., Kryanev A.V., Tsupko-Sitnikov M.V. Primenenie robustnykh metodov dlya resheniya nekorrektnykh zadach [Application of robust methods for solving ill-posed problems]. JVMMF. 1989, vol. 29, no. 5, pp. 653–661 (in Russian).
4. Hampel M., Ronchetti E., Rausseau P., Stael W. Robastnost' v statistike. Podhod na osnove funktsij vliyaniya [Robustness in statistics. An approach based on influence functions]. Moscow, Mir Publ., 1989. 512 p.
5. Tsytkin Ya.Z. Informatsionnaya teoriya identifikatsii [Information theory of identification]. Moscow, Nauka Publ., 1995. 336 p.
6. Shurygin A.M. Prikladnaya statistika: robustnost', ocenivanie, prognoz [Applied statistics: robustness, estimation, forecast]. Moscow, Finansy i statistika Publ., 2000. 223 p.
7. Kryanev A.V., Lukin G.V., Udumyan D.K. Metricheskij analiz i obrabotka dannykh [Metric analysis and data processing]. Moscow, Fizmatlit Publ., 2012. 308 p.
8. Ananyev E.P., Polupan A.A., Matskovsky I.V. [et al.]. Ispol'zovanie rezhima IntelliVent-ASV dlya podderzhaniya tsелеvogo diapazona EtCO₂ u pacientov s tyazhelej CHMT. [Using the IntelliVent-ASV mode to maintain the target EtCO₂ range in patients with severe TBI]. Zhurnal «Voprosy neirohirurgii» imeni N.N. Burdenko. 2017, vol. 81 (5), pp. 63–68 (in Russian).

9. Artificial intelligence and machine learning show promise in cancer diagnosis and treatment. Medical Xpress. Available at: <https://medicalxpress.com/news/2022-03-artificial-intelligence-machine-cancer-diagnosis.html> (accessed 20.11.2024).
10. Kryanev A.V., Udumyan D.K. Metric analysis, properties and applications as a tool for interpolation. International Journal of Mathematical Analysis, 2014, vol. 8, no. 45, pp. 2221–2228.
11. Kryanev A.V., Udumyan D.K. Metric Analysis, Properties and Applications as a Tool for Forecasting. International Journal of Mathematical Analysis, 2014, vol. 8, no. 60, pp. 2971–2978.
12. Ivanov V.V., Kryanev A.V., Udumyan D.K., Lukin G.V. Metric Analysis Approach for Interpolation and Forecasting of Time Processes. Applied Mathematical Sciences. 2014, vol. 8, no. 22, pp. 1053–1060.
13. Klimanov S.G., Kryanev A.V., Trikozova V.A., Tsareva D.D. Vychislitel'nye skhemy vyyavleniya anomal'nyh vybrosov v znacheniyah tekushchih pokazanij pacienta pri iskusstvennoj ventilyacii legkih. [Computational Schemes for Detecting Abnormal Outliers in the Values of Current Patient Readings during Artificial Lung Ventilation]. Vestnik NIYaU MIFI, 2024, vol. 13, no. 2, pp. 76–82 (in Russian).