

УДК: 520.88 PACS: 95.10

МЕТОД РЕКУРСИВНОГО РАЗБИЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ В ТРАЕКТОРНЫХ СПУТНИКОВЫХ ДАННЫХ

© 2025 г. П. Р. Запевалин

Астрокосмический центр Физического института имени П.Н. Лебедева РАН, Москва, Россия

e-mail: pav9981@yandex.ru

Поступила в редакцию 24.09.2024 г.

После доработки 15.01.2025 г.

Принята к публикации 31.01.2025 г.

В настоящей работе представлен метод обнаружения аномальных измерений в траекторных данных космических аппаратов, основанный на рекурсивном разбиении временной последовательности наблюдений. Данный метод анализирует среднеквадратическое отклонение данных, эффективно выявляя аномальные измерения, характеризующиеся повышенным шумом. Его преимущество заключается в отсутствии необходимости знания начального приближения орбиты и предварительного обучения. Метод был протестируирован на модельных данных с искусственно введенными аномалиями и на реальных данных космического аппарата “Спектр-Р”. В сравнении с другими традиционными методами обнаружения аномалий на непосредственных наблюдениях данный подход продемонстрировал наименьший процент ложно-отбракованных измерений. Алгоритм этого метода подходит для различных типов орбит и масштабов наблюдений. Код алгоритма доступен для свободного использования.

DOI: 10.31857/S0023420625040079, EDN: QNNFLJ

1. ВВЕДЕНИЕ

Применение траекторных измерений оптического и радио—диапазонов для определения орбиты космического аппарата (КА) — одна из важнейших процедур баллистико-навигационного обеспечения космических проектов. Обычно используются измерения угловых координат КА, дальномерные радио— и лазерные измерения, доплеровские измерения радиальной скорости КА. Каждый из этих типов наблюдений имеет свою математическую модель, на основе которой вычисляются невязки с реальными наблюдениями. Эта информация является ключевой при обработке траекторных измерений и служит для уточнения реальной орбиты КА. Несмотря на то что модель траекторных наблюдений хорошо известна [1–3], в процессе уточнения орбиты возникает сложность в виде наличия аномальных измерений. Эти измерения характеризуются распределением значений, которое не соответствует предположениям модели измерений. Несоответствие ожидаемым значениям делает их неинформативными в контексте уточнения орбиты. Они отличаются заметно большим шумом и не несут в себе полезной информации.

Такие аномалии могут возникать из-за ошибок бортовой и наземной аппаратуры, ошибок приема сигнала при его захвате и потере, а также из-за различных сбоев при прохождении сигнала через слои атмосферы и ионосфера [4]. Например, обработка траекторных наблюдений КА “Спектр-Р” [5] и КА “Спектр-РГ” [6] с отечественных станций слежения в Медвежьих Озерах и Уссурийске выявила большое число аномальных измерений [4].

Присутствие аномальных измерений может значительно повлиять на точность определения орбиты КА. Обнаружить аномальные измерения в автоматическом режиме бывает непросто, т.к. сложно подобрать единый критерий отличия аномального измерения от нормального зашумленного измерения. Наиболее популярным, простым и эффективным методом уточнения орбиты является метод наименьших квадратов [7–10]. Он основан на технике обработки всех имеющихся наблюдений одновременно. Для его корректной работы необходима предварительная отбраковка измерений. Помимо этого, существует группа методов, основанная на последовательной обработке данных. К ним

можно отнести метод фильтра Калмана [11–13]. Но и в этом случае аномальные измерения могут значительно замедлить сходимость процесса уточнения орбиты и требуют предварительной отбраковки.

Для отбраковки измерений можно использовать статистические методы, основанные на различных критериях. К таким критериям можно отнести Z-оценку [14], межквартильный размах [15], медианные абсолютные отклонения [16], критерий Граббса [17], оценку Ходжеса – Лемана [18] и т.д. Обычно для этих методов выбирается некоторое пороговое значение, при превышении которого наблюдение можно считать аномальным. Перед использованием критерия оценки аномальных измерений часто применяют различные модели регрессии и методы сглаживания данных. Если есть информация о начальном приближении орбиты, то ее используют, чтобы анализировать не сами измерения, а их рассогласованные значения с приближенной орбитой — невязки. В этом случае оказывается достаточным использование линейной регрессии для устранения тренда временной последовательности.

В случае если априорная орбита недоступна, и приходится работать с самими на—бледениями, возникают дополнительные сложности. Для того чтобы выбрать единый критерий оценки аномальных измерений, необходимо устраниć временной тренд наблюдений, что не всегда бывает тривиальной задачей. Помимо этого, траекторные наблюдения могут иметь разные масштабы, например для перигея и апогея высокоэллиптической орбиты, от тысяч до сотен тысяч километров. Кроме того, временные последовательности наблюдений могут содержать нелинейные зависимости, к которым сложно подобрать единый критерий оценки аномальности и выбрать нужное пороговое значение.

Помимо классических статистических методов для анализа космических данных используются методики машинного обучения [19, 20]. Для обнаружения аномалий могут использоваться методы кластеризации [21–23], трансформации [24, 25] и предсказания [26] спутниковых данных. Эти методы имеют более сложную форму, чем традиционные статистические методы, но могут предсказывать аномалии для более сложных временных последовательностей. В то же время их основными недостатками являются:

- потребность в большом количестве данных;
- настройка гиперпараметров;
- время обучения;

- проблемы переобучения;
- сложность интерпретации процесса получения результатов;
- вычислительная сложность.

В настоящей работе предложен метод определения аномальных измерений путем рекурсивного разбиения временной последовательности наблюдений. Он не требует начального приближения орбиты или предварительного обучения, подходит к различным типам орбит и масштабам наблюдений и работает быстрее, чем методы машинного обучения. Также проведено сравнение данного метода с другими статистическими подходами на непосредственных наблюдениях и с использованием невязок наблюдений.

2. МОДЕЛИРОВАНИЕ ТРАЕКТОРНЫХ ИЗМЕРЕНИЙ

Для тестирования нового метода необходимо было разработать базу данных траекторных наблюдений с отмеченными в ней аномалиями. Аномальные измерения в реальных данных можно находить вручную, если построить график разностной производной по соседним значениям измерений. В этом случае аномалии будут видны в виде скачков и пиков на графике. При этом, чем меньше будет шаг по времени между измерениями, тем лучше будут они видны. Для больших временных шагов можно использовать вторую и последующие разностные производные измерений по времени, что подразумевает вычитание соседних значений и деление на шаг времени. Однако такой ручной процесс разметки аномалий в реальных наблюдениях довольно неудобен и требует больших временных затрат. В связи с этим было принято решение смоделировать траекторные измерения с заранее известными позициями аномалий, то есть создать модельные размеченные данные.

Также были созданы невязки модельных измерений, содержащих аномалии. Они были получены на основе расчетных значений измерений с использованием той же модельной орбиты. Поэтому невязки не содержат монотонного расхождения измерений и расчетных значений, которое бы наблюдалось в реальности из-за неточности модели движения. А следовательно, они не имеют тренда.

Моделирование траекторных измерений осуществлялось в Геоцентрической небесной системе координат (GCRF). Начало отсчета в системе координат GCRF находится в центре масс Земли, а ее оси закреплены относительно

удаленных радиоисточников (квазаров). Геопотенциал Земли рассчитывался в Международной земной системе координат (ITRF). Это геоцентрическая система координат с началом отсчета в центре масс Земли. Ось Z системы ITRF в пределах 30 мс дуги совпадает с условным международным началом (CIO), что является средним положением земного полюса по измерениям на интервале с 1900 по 1905 г [27]. Ось X лежит в плоскости Гринвичского меридиана и экватора, ось Y лежит в плоскости меридиана с долготой 90° и дополняет систему осей до правой тройки векторов. Эфемериды модельных орбит и модельные измерения привязывались к Всемирному времени UTC (Coordinated Universal Time). Это атомная равномерная шкала времени, отличающаяся от Международного атомного времени (TAI) на целое число секунд, называемых високосными секундами.

Моделирование орбит

Для моделирования траекторных измерений были созданы два набора случайных орбит: низкие околоземные орбиты (НОО) и высокоэллиптические орбиты (ВЭО). Каждый из этих наборов включал 10000 эфемерид, длительностью в один период орбиты. Каждая орбита выбиралась случайным образом из заданного диапазона значений элементов Кеплера. Численное интегрирование уравнений движения проводилось с помощью метода Эверхарта [28]. Модель движения включала в себя расчет геопотенциала согласно модели EGM 2008, а также учет планет Солнечной системы, Луны и Солнца согласно эфемеридам JPL DE433. Для НОО учитывалось разложение геопотенциала до степени и порядка 40, а для ВЭО — до 10. Такой упрощенный вид модели движения достаточен для данной задачи,

поскольку интегрирование орбиты происходит только на один период. В табл. 1 представлены ограничения на значения элементов Кеплера для моделирования случайных НОО и ВЭО.

Модель наблюдений

С помощью модельных орбит были построены модельные траекторные измерения наклонной дальности. Эти измерения должны соответствовать реальным траекторным существующих космических проектов.

Реальные измерения наклонной дальности проекта “Спектр-Р” проводились по запросной схеме, когда сигнал передается на КА с наземной станции слежения, а затем переизлучается на нее обратно. В дальнейшем эти измерения были интерпретированы в однопутные (беззапросные), т.е. наклонную дальность, получаемую при излучении сигнала с борта КА с привязкой к моменту времени приема сигнала на наземной станции. Поскольку в настоящей работе были задействованы именно такие однопутные реальные измерения, было принято решение провести моделирование наблюдений по схожей схеме. В работе использовались две наземные станции слежения: Медвежьи Озера и Уссурийск. Наблюдения моделировались на временном интервале в 15 мин с шагом 10 с. Интервал для наблюдений выбирался случайным образом из интервала модельных орбит с учетом видимости КА с наземной станции. Видимость рассчитывалась без ограничений на минимальный угол места для антенны станции. Каждая последовательность наблюдений записывалась в отдельный файл. Таким образом, получилось 20000 последовательностей измерений, поскольку для каждой орбиты моделировались наблюдения с двух станций.

Таблица 1. Ограничения значений элементов Кеплера для моделирования случайных НОО и ВЭО

Вид орбиты	a , км	e	i	Ω	ω	v
НОО	$R_\oplus + (160, 2000)$	(0, 0.05)	(0, $\pi/2$)	(0, 2π)	(0, 2π)	(0, 2π)
ВЭО	$R_\oplus + (50000, 200000)$ $10000 < rp < 60000$ $ra < 380000$	(0.8, 0.95)	(0, $\pi/2$)	(0, 2π)	(0, 2π)	(0, 2π)

Примечание. R_\oplus — средний радиус Земли; rp — перигей; ra — апогей; a — большая полуось; e — эксцентриситет; i — наклонение; Ω — долгота восходящего узла; ω — аргументperiцентра; v — истинная аномалия.

Моделирование наблюдений включало в себя решение светового уравнения, учет задержки распространения сигнала в тропосфере и ионосфере, а также релятивистскую задержку распространения сигнала. Для упрощения моделирования считалось, что смещения фазового центра наземной и бортовой антенны уже учтены, а ошибки часов КА и наземной станции нулевые. Таким образом, модель наблюдений представляется в виде:

$$P = \varrho(tg, ts) + I + T + c \cdot \delta t_{rel} + \epsilon, \quad (1)$$

где P — модельное измерение; $\varrho(tg, ts)$ — геометрическое расстояние между центрами масс КА в момент передачи сигнала ts и наземной станцией в момент приема tg ; I — влияние ионосферы; T — влияние тропосферы; c — скорость света в вакууме; δt_{rel} — влияние релятивизма; ϵ — шум модельных измерений. Далее приведены формулы расчета каждого из этих эффектов.

Величина ϱ в уравнении (1) представляет собой геометрическое расстояние между центрами масс КА в момент времени передачи сигнала и наземной станции в момент времени приема сигнала. На практике известен только момент времени приема сигнала. Для вычисления момента времени передачи рассчитывается задержка сигнала по формуле:

$$\varrho t_{delay} = R_{start} / c, \quad (2)$$

где R_{start} — расстояние между КА и наземной антенной в момент приема сигнала, c — скорость света в вакууме. Затем определяется вектор состояния КА на момент времени $tg - \delta t_{delay}$ путем численного интегрирования уравнений движения. Находится новое расстояние R между КА в момент времени $tg - \delta t_{delay}$ и наземной станцией в момент времени tg . Это расстояние поступает на вход формулы (2) и вычисляется новая задержка сигнала. Данная процедура повторяется итеративно, пока два последующих расстояния R будут отличаться не больше, чем на некоторое достаточно малое число. После этого задержка и, соответственно, момент времени передачи сигнала считаются найденными. Основная часть ионосферной задержки сигнала I рассчитывается по формуле:

$$I = M(z) \times ZTEC \times \frac{40.309 \cdot 10^{-16}}{f^2}, \quad (3)$$

где $M(z)$ есть картирующая функция, зависящая от зенитного угла z . Величина полного

вертикального содержания электронов ZTEC (Zenith Total Electron Content) рассчитывается согласно ионосферным картам, полученным центром анализа Международной службы Глобальных навигационных спутниковых систем [29]. Данные карты содержат общее содержание электронов в ионосфере для сигнала, испускаемого в зенит из точки на поверхности Земли. Данный формат имеет название IOnosphere map Exchange format (IONEX) [30]. Параметр f — это частота сигнала в герцах, а ионосферная задержка сигнала I по этой формуле вычисляется в метрах.

Задержка распространения сигнала в тропосфере T не зависит от частоты сигнала и в упрощенном виде описывается суммой “сухой” T_d и “влажной” T_w составляющей. Зенитная задержка “сухой” составляющей для сигналов радиодиапазона описывается формулой:

$$T_d = m_d(z) \frac{0.0022786 \cdot P}{1 - 0.00266 \cos(2\phi) - 2.8 \cdot 10^{-7} \cdot H}, \quad (4)$$

где P — атмосферное давление в районе фазового центра антенны; ϕ — широта положения станции; H — высота станции над уровнем моря в метрах. Зенитная задержка “влажной” составляющей T_w полагалась равной 10 % от “сухой” [31]. В качестве картирующей функции $m(z)$ использовалась глобальная функция отображения GMF [32].

Из-за эффекта задержки распространения света при прохождении вблизи массивного тела, известного также как эффект Шапиро [33], возникает релятивистская поправка в модели наблюдений. Она описывается в виде:

$$c\delta t_{rel} = \frac{2\mu}{C^2} \ln\left(\frac{r_s + r_g + \rho}{r_s + r_g - \rho}\right), \quad (5)$$

где μ — гравитационный параметр небесного массивного тела; $\frac{2\mu}{C^2}$ — гравитационный радиус Шварцшильда; r_s — геоцентрическое расстояние до КА; r_g — геоцентрическое расстояние до наземной станции; ρ — расстояние между станцией и КА.

Моделирование шума и аномальных измерений

Реальные траекторные измерения также содержат некоторый шум ϵ . В настоящей работе предполагалось, что шум измерений имеет

нормальное распределение. Его моделирование осуществлялось по формуле:

$$\epsilon = AWGN \times N, \quad (6)$$

где $AWGN$ — аддитивный белый гауссовский шум, имеющий нулевое среднее и стандартное отклонение 1; N — коэффициент, характеризующий размер шума. По сути N является среднеквадратическим отклонением (СКО) шума измерений.

Чтобы добавить аномальные измерения в модельные наблюдения, был использован белый гауссовский шум ϵ_A . Для этого выбирался характерный размер аномального шума и минимальное пороговое значение шума, которое считалось бы аномальным.

$$\epsilon_A = AWGN \times S, |\epsilon_A| > E_{\min}, \quad (7)$$

где ϵ_A — аномальная прибавка к модельному наблюдению; S — коэффициент, характеризующий размер аномалий; E_{\min} — минимальное пороговое значение аномального измерения. Также была выбрана частота появления аномалий. В данной работе она задавалась, как процент p аномальных измерений от всех измерений временного интервала. Таким образом, для моделирования наблюдений задавались параметры N , S , E_{\min} и p .

Для тестирования метода были созданы модельные измерения с различным уровнем шума. Параметры моделирования выбирались с учетом реальных данных проекта “Спектр-Р”. Случайные ошибки измерительных систем в данном проекте находились в пределах 10 м [34]. Характерный размер СКО аномалий и частота их появления были выбраны также на основе данных проекта “Спектр-Р”. Параметры моделирования измерений оказались следующие:

$$N=10, S=20N, E_{\min}=5N, p=1\%. \quad (8)$$

Помимо этого, были созданы модельные измерения с другими значениями параметра N для выявления устойчивости метода к различным уровням шума.

3. СУЩЕСТВУЮЩИЕ КЛАССИЧЕСКИЕ МЕТОДЫ

Традиционные методы обнаружения аномалий во временных последовательностях просты в реализации и не требуют значительных вычислительных ресурсов, что делает их популярными в

различных областях анализа временных данных. Одним из наиболее распространенных подходов по обнаружению аномалий является использование статистических характеристик данных, таких как среднее и стандартное отклонение. Например, можно использовать метод Z -оценки [14] для определения аномалий там, где значения превышают определенный порог. Для каждого значения измерения рассчитывается Z -оценка с использованием следующей формулы:

$$z_i = \frac{x_i - \mu}{\sigma}, \quad (9)$$

где x_i — текущее значение измерения, μ — среднее значение данных, а σ — стандартное отклонение. Если Z -оценка превышает заданный порог k_z , где k_z — настраиваемый параметр, значение считается аномальным:

$$|z_i| > k_z, 1 \leq k_z \leq 3. \quad (10)$$

Также существует метод адаптивного порога [35], который динамически регулирует порог обнаружения аномалий на основе статистических характеристик данных. Сначала метод вычисляет среднее значение и стандартное отклонение массива значений измерений, обозначенные как μ и σ соответственно. Порог обновляется по следующей формуле:

$$thr_{at} = \mu + \left(k_{at} \sqrt{\sigma^2 + \sigma_{\text{апр}}^2} \right), 1 \leq k_{at} \leq 3, |x_i| > thr_{at}, \quad (11)$$

где $\sigma_{\text{апр}}^2$ — априорная оценка дисперсии измерений, k_{at} — настраиваемый параметр.

Метод медианных отклонений [16] основывается на применении медианы для оценки центральной тенденции и медианного абсолютного отклонения для оценки вариативности. Сначала вычисляется медиана M значений измерений и медианное абсолютное отклонение MAD , определяемое по формуле:

$$MAD = \text{median}(|x_i - M|, \text{для всех } i), \quad (12)$$

где x_i — текущее значение измерения. Порог для определения аномалий обновляется по следующей формуле:

$$thr_{med} = M + k_{med} \cdot MAD, k_{med} > 0, \quad (13)$$

где k_{med} — настраиваемый коэффициент, который устанавливает уровень чувствительности к

аномалиям. Значения измерений, превышающие порог thr_{med} , считаются аномальными.

Подбор настраиваемых параметров для классических методов обнаружения аномалий представляет собой сложную задачу, поскольку неправильная настройка может привести к высокому уровню ложно-положительных или ложно-отрицательных результатов. Это в свою очередь существенно влияет на качество оценки аномалий и эффективность метода в целом. Кроме того, представленные в этом разделе методы обычно работают с невязками измерений, а для работы с непосредственными наблюдениями необходимо сначала устраниć их нелинейный тренд.

4. УДАЛЕНИЕ ТРЕНДА

Для того чтобы лучше выявлять аномальные измерения, необходимо рассчитать первую разностную производную измерений по времени. В этом случае аномальные измерения будут хорошо видны в виде скачков и пиков на графике разностной производной. Для автоматического обнаружения этих пиков необходимо вычесть тренд графика разностной производной, т.е. привести среднее значение последовательности разностных производных измерений максимально близко к нулю. В работе [4] приведены методы обнаружения аномалий, в которых используется вычитание тренда. В этих методах тренд данных считается линейным, поскольку они работают с невязками измерений. Для самих же измерений тренд может быть линейным только для некоторых участков орбиты и очень коротких временных интервалов. В методах, использующих невязки измерений, вычисляется априорная орбита КА, на основе которой задаются расчетные значения измерений и составляются невязки с реальными наблюдениями. Для этого требуется достаточно точная априорная орбита, которая не всегда бывает доступна. Например, при неправильном учете неконсервативных сил, действующих на движение КА сложной формы, априорная орбита может приносить в невязки измерений значительные нелинейные отклонения. Также влияние на невязки могут оказывать маневры КА, столкновения с космическим мусором и микрометеоритами. Для космических аппаратов, обнаруженных относительно недавно, априорная информация может отсутствовать и вовсе. К примеру, для систем, обрабатывающих информацию в реальном времени требуется работать с непосредственно поступающими наблюдениями. В этом случае априорная орбита

может быть недоступна или неактуальна и может внести дополнительную погрешность. Такие системы могут использовать алгоритм обобщенного фильтра Калмана [36], для которого априорной орбитой служит оценка вектора состояния на предыдущий момент времени, обновляющаяся на каждой итерации алгоритма. В приложениях космической радиоинтерферометрии для уточнения орбиты космического радиотелескопа можно использовать так называемые остаточные задержки и частоты интерференции, полученные в процессе обработки радиоинтерферометрического сеанса наблюдений. В этом случае получение априорных значений измерений задержек невозможно, а аномалии могут возникать в процессе обработки сеанса наблюдений.

Преимущество метода рекурсивного разбиения, представленного в настоящей работе, заключается в отсутствии необходимости в расчетных аналогах измерений. Следовательно, его применение может быть расширено и на другие приложения, в которых в принципе не существует прогнозных значений наблюдений или их построение затруднительно. При работе с непосредственно наблюдениями, а не их невязками, возникает необходимость удаления нелинейного тренда измерений.

В настоящем исследовании для удаления тренда из временных последовательностей был использован метод сглаживания, основанный на полиномиальной регрессии. Этот метод использует фильтр сглаживания данных Савицкого – Голея [37] из библиотеки SciPy для языка программирования Python [38]. Библиотека SciPy имеет открытый исходный код и предназначена для выполнения научных и инженерных расчетов. Фильтр Савицкого – Голея применяет локальную полиномиальную регрессию к выборкам данных. В заданном окне значений вычисляется полином определенного порядка, который затем используется для предсказания центральной точки этого окна. Это позволяет эффективно экстраполировать общую тенденцию в данных, минимизируя искажения, характерные для других методов сглаживания нелинейных данных, таких как метод скользящего среднего. Основное преимущество фильтра заключается в его способности сохранять важные характеристики данных и не размывать аномальные особенности временного ряда. В настоящей работе этот фильтр позволяет выделить четкую линию тренда, которая впоследствии вычитается из временной последовательности измерений. Это дает возможность сосредоточиться на краткосрочных колебаниях и структурных изменениях в данных.

Однако следует отметить ограниченную чувствительность данного метода к медленным случайным процессам, связанным с дрейфом измерительной аппаратуры. Для комплексного анализа и выявления подобных систематических ошибок или медленных случайных процессов необходимы дополнительные изыскания. В данной работе основное внимание уделено обнаружению дискретных аномалий и выявлению кратковременных выбросов и скачков в данных.

Алгоритм удаления тренда можно описать следующим образом. Пусть имеется временная последовательность измерений $x(t)$. Для каждого значения $x(t)$ выбирается окно из m соседних точек данных, которое включает выборку измерений $x(t-s), x(t-s+1), \dots, x(t), \dots, x(t+s)$, где $s = \frac{m-1}{2}$. Оценка тренда может быть получена с использованием полинома. Коэффициенты полинома a_k вычисляются с помощью метода наименьших квадратов, осуществляющего минимизацию суммы квадратов отклонений между фактическими значениями и значениями, предсказанными полиномом.

Обозначим вектором y набор измерений в окне:

$$y = \{x(t-s), \dots, x(t+s)\}. \quad (14)$$

Тогда для получения оценки тренда используется следующее выражение:

$$\hat{y}(z) = \sum_{k=0}^p a_k z^k, \quad a = (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \mathbf{y}, \quad (15)$$

где $\hat{y}(z)$ — оценка тренда; z — независимая переменная, отнесенная к центру окна и принимающая значения от $-s$ до s ; p — порядок полинома, который выбирается экспериментальным путем в зависимости от типа данных; $a = \{a_k\}$ — вектор коэффициентов полинома, а m — количество точек данных в окне. Матрица \mathbf{J} представляет собой матрицу Вандермонда следующего вида:

$$\mathbf{J} = \begin{bmatrix} 1 & -s & (-s)^2 & \dots & (-s)^p \\ 1 & -s+1 & (-s+1)^2 & \dots & (-s+1)^p \\ 1 & -s+2 & (-s+2)^2 & \dots & (-s+2)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & s & s^2 & \dots & s^p \end{bmatrix}. \quad (16)$$

После получения оценки тренда вычисляется детрендированная временная последовательность

путем вычитания оценки тренда из исходной временной последовательности:

$$x_{\text{detrend}}(z) = x(z) - \hat{y}(z), \quad (17)$$

где $x(z)$ — исходные значения.

5. МЕТОД РЕКУРСИВНОГО РАЗБИЕНИЯ

В данном разделе приводится описание метода рекурсивного разбиения для обнаружения аномальных измерений. Концепция этого метода основана на разбиении их временного интервала на две равные части и подсчета среднеквадратического отклонения измерений в каждой из них. Предполагается, что в той части, где СКО больше, находится аномальное измерение. Эта половина временного интервала в свою очередь также разбивается на две равные части. Так продолжается до тех пор, пока не будет найдено точное положение предполагаемого аномального измерения. При этом, если СКО предполагаемой аномалии больше некоторого порогового значения, аномалия считается подтвержденной. В этом методе предполагается, что измерения имеют строгий равномерный шаг по времени.

Алгоритм для выявления аномалий начинается с деления последовательности данных, освобожденных от тренда, на две равные части. Предполагается, что часть, имеющая более высокое СКО, содержит аномалии. Эта часть снова делится на две половины. Процесс продолжается рекурсивно, пока не будет обнаружен участок с предполагаемой аномалией. После каждого сравнения СКО двух половинных сегментов, более высокое значение сравнивается с заранее установленным порогом. Это значение задает пользователя в зависимости от шума данных и характерного размера аномалий. Если СКО оказывается меньше порогового значения, то алгоритм останавливается. Это означает, что аномалий в данной последовательности измерений не найдено. Далее представлена краткая схема алгоритма.

Входные данные:

data — массив, последовательность разностных производных траекторных измерений по времени,

thr — пороговое значение.

Определяется начальный и конечный индекс:

$$\begin{aligned} start &= 0, \\ end &= \text{len}(\text{data}) - 1, \end{aligned} \quad (18)$$

где *len* — функция поиска длины массива.

Если:

$$(end - start) \leq 3, \quad (19)$$

то алгоритм завершается и возвращает массив *data*, который содержит аномальный участок. В противном случае переходим к следующему пункту.

Вычисляется средний индекс:

$$mid = \frac{start + end}{2}. \quad (20)$$

Вычисляются стандартные отклонения левой и правой части:

$$\begin{aligned} left_std &= std(data[start:mid]), \\ right_std &= std(data[mid:end]), \end{aligned} \quad (21)$$

где *data* [*start* : *mid*] — означает выборку измерений от индекса *start* до индекса *mid*.

Если:

$$\max(left_std, right_std) < thr, \quad (22)$$

то алгоритм завершается и возвращает пустой массив. Это говорит о том, что аномалии отсутствуют. В противном случае переходим к следующему пункту.

Обновляется массив данных:

$$data = \begin{cases} data[start : mid], & \text{если } left_std \geq right_std, \\ data[mid : end], & \text{иначе} \end{cases}$$

обновляются начальный и конечный индексы:

$$\begin{cases} start = mid, & \text{если } left_std < right_std, \\ end = mid, & \text{если } right_std < left_std. \end{cases}$$

Переходим к пункту 2.

После детектирования эпохи аномального измерения необходимо продолжать поиск других аномалий, поскольку в выбранной последовательности траекторных измерений может присутствовать несколько аномальных значений. Для этого достаточно запустить выше предложенный алгоритм заново. Так будет происходить до тех пор, пока алгоритм не определит, что аномалий на данном участке больше нет. Для того чтобы этот процесс не происходил бесконечно долго в случае выбора малого порогового значения, пользователь задает максимальное количество аномалий, которые могут присутствовать во временной последовательности измерений.

Однако если для поиска других аномалий просто запустить выше предложенный алгоритм заново, то он найдет все ту же аномалию, что и в первый раз. Поэтому возникает необходимость удалить найденную аномалию из временного ряда. При этом в месте удаления может возникнуть скачок данных, который может быть идентифицирован как ложная аномалия. Вместо удаления, в предложенном методе используется замена найденной аномалии на среднее значение последовательности, не содержащей найденных аномалий. Стоит учесть, что если в качестве данных используются первые разностные производные измерений по времени, то аномальное измерение превратится в два аномальных значения разностной производной. В таком случае замене подлежит интервал из этих двух значений, который можно назвать характерным размером аномалии. Поскольку по мере нахождения аномалий среднее значение нормальных данных все время изменяется, необходимо заменять также все найденные участки аномалий новым средним значением. Это позволит избежать возникновения новых скачков и обнаружения ложных аномалий. Алгоритм замены выражается следующим образом.

Допустим имеются следующие входные данные:

- *data* — массив данных;
- *anom* — массив индексов найденных аномалий;
- *size* — характерный размер аномалии.

1. Создаем маску в виде массива значений логического типа:

$$mask[i] = \begin{cases} True, & \text{если } i \notin [a - size, a + size] \\ \text{для } \forall a \in anom, \\ False, & \text{иначе.} \end{cases}$$

2. Вычисляем среднее значение:

$$mean = \frac{1}{M} \sum_j data[j], \quad \text{где } mask[j] = True, \quad (23)$$

где *M* — количество элементов, удовлетворяющих условию.

3. Заменяем аномальные данные средним значением:

$$data[i] = mean, \quad \text{где } mask[i] = False.$$

Данный алгоритм автоматически запускается после алгоритма поиска аномалии. Таким

образом, метод рекурсивного разбиения состоит из следующих шагов:

- 1) вычисление временных разностей измерений;
- 2) удаление тренда последовательности измерений;
- 3) поиск аномального измерения путем рекурсивного разбиения последовательности пополам и сравнения СКО двух половинок;
- 4) замена аномального участка специальным средним значением;
- 5) повторение пунктов 3 и 4 до тех пор, пока все аномалии не будут найдены, либо пока не будет достигнут предел количества найденных аномалий.

При этом пользователь может выбирать следующие параметры метода:

- значение порога аномальных измерений;
- размер окна измерений для удаления тренда;
- порядок полинома для удаления тренда;
- характерный размер аномалии;
- максимальное количество аномальных измерений.

6. РЕЗУЛЬТАТЫ

Предложенный метод был опробован на модельных данных наблюдений для НОО и ВЭО, а также на реальных данных проекта “Спектр-Р”. В качестве измерений выступали наклонные радиодальности. Было проведено сравнение эффективности предложенного метода на непосредственных наблюдениях и их невязках с другими статистическими методами обнаружения аномалий.

Во время тестирования метода анализировались статистические параметры обнаружения аномалий. Разработанный метод содержит встроенную функцию подсчета статистики, что облегчает данную задачу. Данная функция считает следующие величины:

- *True Positives (TP)* — количество корректно обнаруженных аномалий;
- *False Positives (FP)* — количество ложных обнаружений, т.е. когда нормальные измерения ошибочно классифицируются как аномалии;
- *False Negatives (FN)* — количество необнаруженных аномалий, т.е. когда аномальные измерения ошибочно классифицируются как нормальные.

Критерием эффективности метода обнаружения аномалий служит процент отбракованных измерений при нулевом количестве необнаруженных аномалий:

$$q = \frac{TP + FP}{Total} \cdot 100\%, \quad (24)$$

где *Total* — общее исследуемое число измерений. Чем меньше алгоритм находит ложноположительных аномалий при стремлении количества необнаруженных аномалий к нулю, тем меньше *q*, и тем эффективнее работа данного алгоритма.

На рис. 1 показан пример модельных данных. Реальные измерения проекта “Спектр-Р” выглядят аналогично. Из рисунка видно, что аномальные измерения становятся более отчетливо

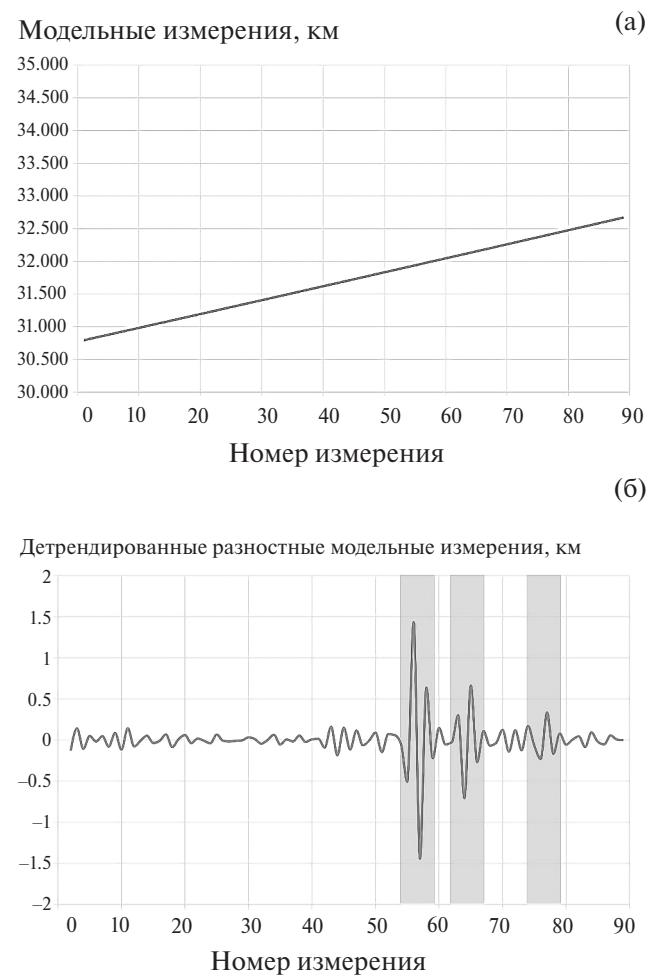


Рис. 1. Пример модельных измерений наклонной дальности для ВЭО (а). Разностная производная по времени модельных измерений для ВЭО с вычтеным трендом (б). Красным цветом отмечены искусственно добавленные аномалии. В данном примере СКО модельных измерений составляет 50 м

видны, если составить временные разности измерений и вычесть тренд. Для этих разностей затем применяется алгоритм метода рекурсивного разбиения, процесс которого показан на рис. 2. После первого прохода метода определяется наибольшая аномалия во всей временной последовательности. Ее координаты заносятся в память компьютера, а сама аномалия заменяется средним значением нормальных данных согласно выражению (23). Затем алгоритм повторяет процедуру поиска других аномалий. В случае, если аномалий больше нет или достигнут предел максимального количества аномалий, алгоритм завершается. В итоге в памяти компьютера остаются координаты всех найденных аномальных измерений, которые затем используются при обработке данных.

На рис. 3 показаны зависимости процента отбракованных измерений от числа необнаруженных аномалий FN в случае использования невязок измерений.

Максимальный процент отбракованных измерений (при $FN = 0$) показан в табл. 2. Метод

абсолютных медианных отклонений дает наименьший процент отбракованных измерений и является наиболее эффективным в этом случае. Методы Z -оценки и адаптивного порога дают q , равный 18.37 % и 12.25 %, соответственно. Эти результаты согласуются с работой [4]. Метод рекурсивного разбиения дает q , равный 4.65 %.

Далее эти же методы были протестированы на непосредственных наблюдениях. Удаление тренда наблюдений во всех методах происходило согласно разделу 4. Размер окна и порядок полинома подбирались экспериментальным путем. Наилучшие результаты были получены с размером окна 35 и порядком полинома 3. На рис. 3б показана зависимость q от FN в случае использования непосредственных измерений. Метод рекурсивного разбиения показал наименьший процент отбракованных измерений по сравнению с другими методами.

Результаты также были получены для других СКО шума с целью определить устойчивость метода и выявить зависимость для порогового значения. Экспериментальным путем было

Таблица 2. Максимальный процент отбракованных измерений q при использовании невязок и непосредственных наблюдений. СКО шума данных 10 м

Процент отбракованных измерений	Метод Z -оценки	Метод адаптивного порога	Метод медианных отклонений	Метод рекурсивного разбиения
При использовании невязок	18.37	12.25	1.24	4.65
При использовании абсолютных измерений	48.44	46.98	70.32	6.21

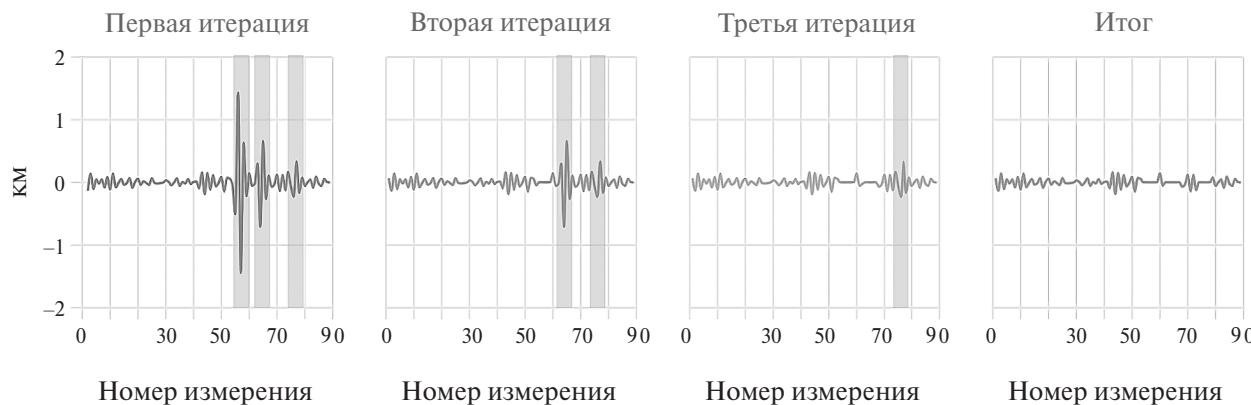


Рис. 2. Пример работы метода рекурсивного разбиения для обнаружения аномальных измерений. См. пояснения в тексте

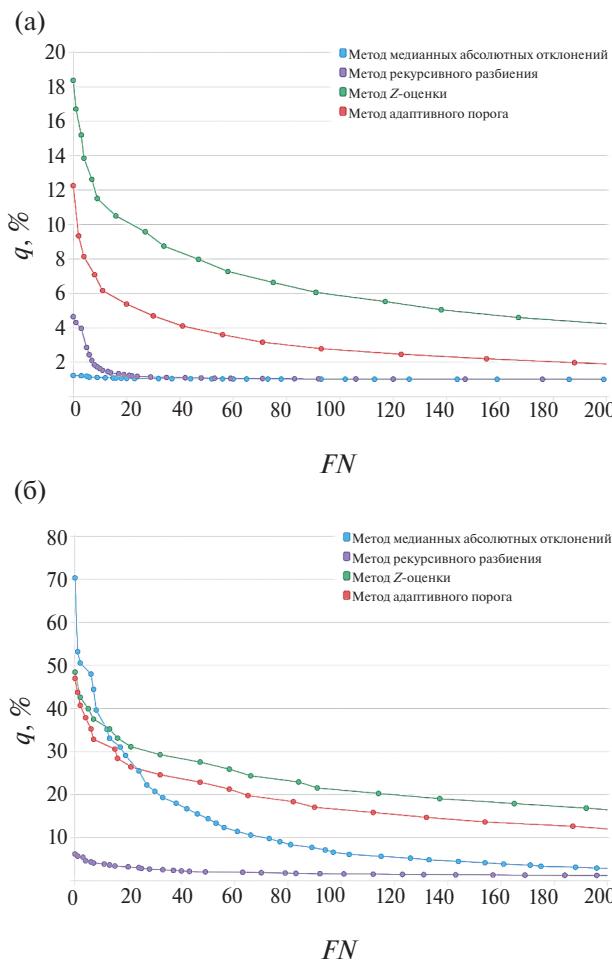


Рис. 3. Зависимость процента отбракованных измерений от числа необнаруженных аномалий FN в случае использования невязок измерений (а) и в случае использования абсолютных измерений (б). СКО модельных измерений составляет 10 м

установлено, что использование априорной оценки шума в качестве порогового значения является наиболее оптимальной стратегией выбора порога. При таком значении порога достигаются нулевые значения FN . Также было установлено, что метод рекурсивного разбиения работает в различных условиях шума измерений. На рис. 4 отображена зависимость количества ложноположительных аномалий при $FN = 0$ от параметра N в моделировании измерений, отвечающего за СКО шума измерений. Из рисунка видно, что количество ложноположительных аномалий остается примерно на одном уровне для всех случаев СКО шума измерений.

Метод рекурсивного разбиения был опробован на реальных данных проекта “Спектр-Р”. Результаты обнаружения приведены на рис. 5.

Так как аномалии в этих данных не размечены, то проверка корректности обнаружения осуществлялась визуально. Случайным образом было выбрано 100 различных интервалов наблюдений наклонной дальности. Визуально было установлено, что во всех выбранных временных последовательностях аномалии были обнаружены корректно.

7. ОБСУЖДЕНИЕ

Настоящая работа была инициирована в связи с наличием аномалий в реальных данных траекторных измерений, полученных в рамках миссий “Спектр-Р” и “Спектр-РГ”. Целью исследования стала разработка метода, способного автоматически обнаруживать эти аномалии, поскольку они оказывают значительное влияние на процесс уточнения орбиты. Для решения этой задачи было проведено моделирование данных с тщательным подбором параметров моделирования экспериментальным путем, для того чтобы обеспечить близкое соответствие между модельными данными и наблюдениями из реальных космических проектов. Для выбора параметров использовались случайные интервалы измерений из различных эпох проекта “Спектр-Р”.

Было изучено несколько различных методов для обнаружения аномалий во временных последовательностях, в том числе методы машинного обучения. Несмотря на то что подходы машинного обучения демонстрируют хорошую эффективность в обработке данных сложной структуры, они часто требуют значительного времени на обучение и наличие размеченных тренировочных данных. Для обнаружения аномалий в траекторных измерениях спутников гораздо удобнее оказываются классические методы. Обычно они применяются к невязкам измерений, т.е. используются расчетные значения измерений согласно модели движения и модели наблюдений КА. В настоящей работе был предложен альтернативный метод обнаружения аномалий — метод рекурсивного разбиения, который может применяться к непосредственным наблюдениям без использования расчетных измерений. Данный метод был создан на основе документации программного обеспечения для определения параметров орбит Bernese [39]. Он был доработан и усовершенствован так, как описано в настоящей работе, что и представляет собой ключевой результат данного исследования.

Результаты работы показали, что при использовании невязок измерений метод медианных абсолютных отклонений оказался более

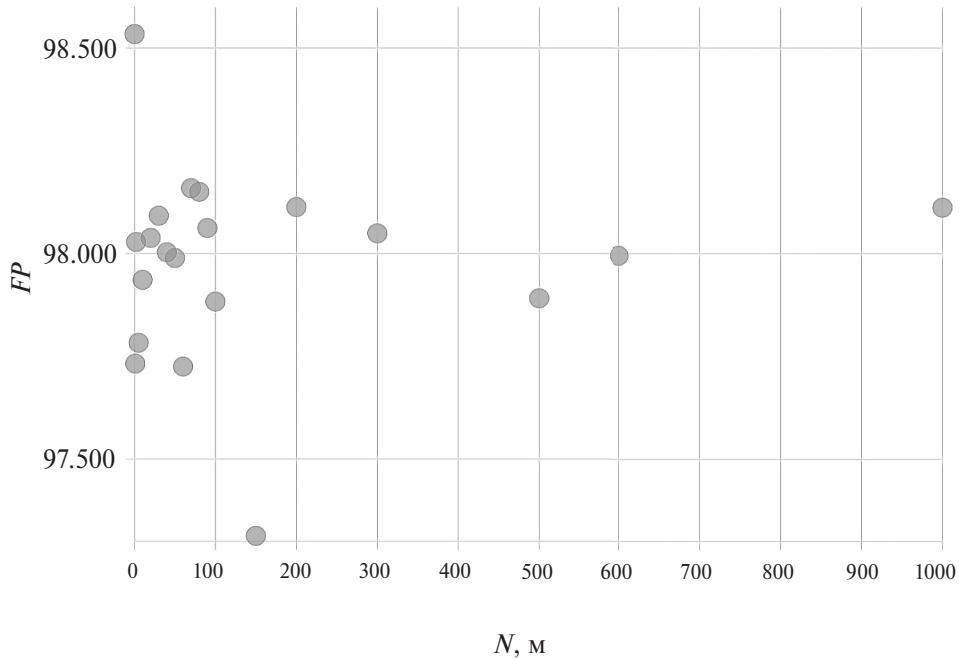


Рис. 4. Зависимость количества ложноположительных аномалий при $FN = 0$ для различных уровней СКО шума измерений N . Общее число измерений эксперимента 1800000

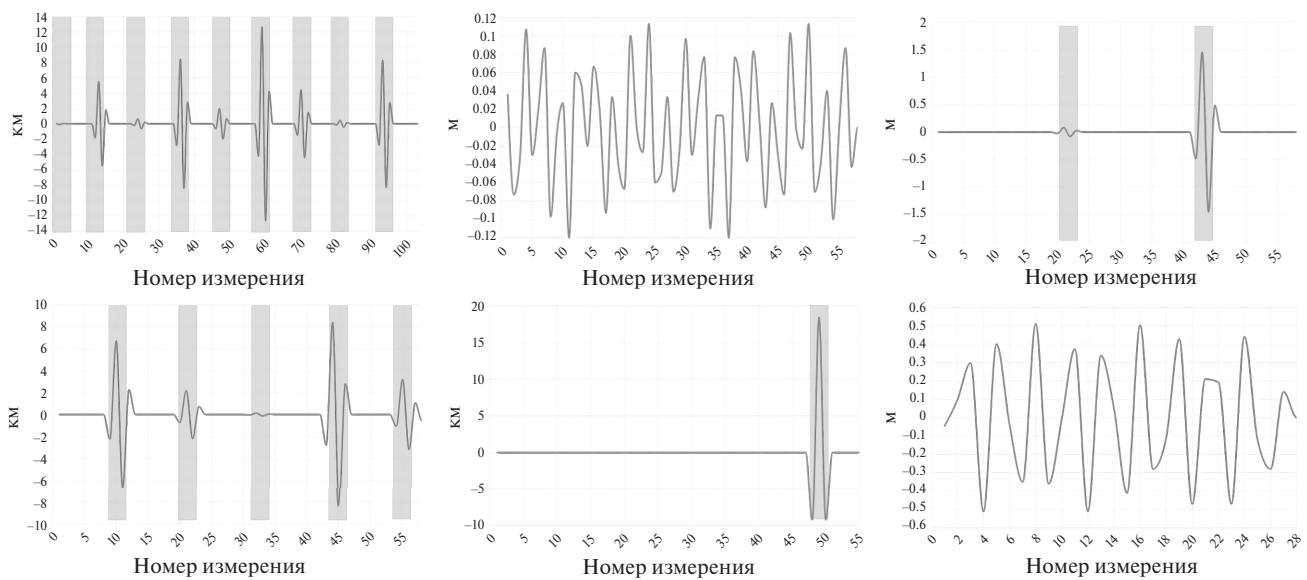


Рис. 5. Пример работы метода рекурсивного разбиения для обнаружения аномальных измерений в реальных данных наклонной дальности проекта “Спектр-Р”. Зеленым показаны разности измерений по времени с вычтеным трендом. Красным отмечены найденные аномалии

эффективным для обнаружения аномалий. Преимущество метода рекурсивного разбиения становится очевидным при поиске аномалий без использования априорной орбиты и расчетных

значений. Максимальный процент отбракованных измерений оказался наименьшим среди рассматриваемых методов. Этому есть объяснение. Метод рекурсивного разбиения определяет

аномалии поочередно, в порядке их величины. Поэтому вероятность правильно обнаружить аномалию в начале обработки временного интервала выше, чем в конце алгоритма. Кроме того, метод имеет параметр, ограничивающий максимальное число аномалий в выборке, что не позволяет заходить алгоритму достаточно далеко, туда, где выше вероятность обнаружить ложноположительную аномалию. В других же статистических методах обработка выборки измерений осуществляется пакетным образом, используется единый пороговый уровень для всех аномалий.

Алгоритм и программный код метода рекурсивного разбиения (<https://github.com/paulzap/RPAD>) доступны любому пользователю, что способствует дальнейшему исследованию и применению метода. Также к нему прилагается краткая инструкция и набор тестовых данных.

Предложенный метод будет использоваться для обработки данных миссий “*Спектр—P*”, “*Спектр-РГ*” и будущей миссии “*Спектр-М*” [40]. Благодаря его проверке на модельных данных, схожих с реальными, метод может считаться надежным инструментом для анализа любых типов данных со схожей структурой. Важно отметить, что он не требует предварительного обучения и зависит только от порогового значения.

Следует также признать, что, несмотря на свою эффективность, метод рекурсивного разбиения имеет свои ограничения. В частности, его время выполнения увеличивается линейно с ростом количества аномалий, что делает его вычислительно затратным для наборов данных с высокой плотностью аномалий. Для решения этих проблем необходимо дальнейшее исследование.

8. ЗАКЛЮЧЕНИЕ

В результате проведенного исследования был представлен метод рекурсивного разбиения для автоматического обнаружения аномальных измерений в траекторных данных космических аппаратов. Этот метод не требует предварительного обучения, прост в использовании и демонстрирует точность, превосходя традиционные статистические подходы по максимальному проценту отбракованных измерений. Результаты тестирования метода на модельных и реальных данных подтверждают его универсальность и эффективность в различных условиях обработки космических данных.

Метод обладает рядом преимуществ. Он не зависит от длины интервала измерений, что делает его удобным для использования в задачах различного объема, а его гибкость позволяет адаптироваться к различным уровням шума в данных. Метод не требует грубой оценки орбиты, что упрощает его применение для уточнения орбиты космических аппаратов. Набор различных параметров настройки дает возможность пользователю адаптировать метод под конкретные задачи. Кроме того, в отличие от других методов, которые обычно ограничиваются вычитанием линейного тренда в спутниковых измерениях, в настоящей работе показано, что можно эффективно работать и с различными видами орбит благодаря использованию полиномиальной регрессии.

Тем не менее метод рекурсивного разбиения имеет свои ограничения: время его выполнения зависит от количества аномалий в данных. Поэтому необходимо продолжать исследование и совершенствование алгоритма с целью его оптимизации и адаптации к различным типам данных. Наконец, стоит отметить, что представленный метод имеет открытый код и содержит подробные инструкции к использованию. Это расширяет возможности его применения в самых разных областях, помимо анализа траекторий космических аппаратов и других спутниковых данных.

СПИСОК ЛИТЕРАТУРЫ

1. Zapevalin P. Multi-GNSS Observations Simulation // Cosmic Research. 2024. V. 62. Iss. 5. P. 424–435.
2. Zakhvatkin M.V., Andrianov A.S., Avdeev V.Yu. et al. RadioAstron orbit determination and evaluation of its results using correlation of space-VLBI observations //Advances in Space Research. 2020. V. 65(2). P. 798–812.
3. Montenbruck O., Gill E. Satellite Tracking and Observation Models // Satellite Orbits: Models, Methods and Applications. Springer-Verlag Berlin Heidelberg, 2000. P. 193–232.
4. Tuchin D.A. Selection of anomalous measurements during primary processing in orbit determination from the trajectory information of the ground station Cobalt-R // Keldysh Institute of Applied Mathematics Preprints. 2022. Iss. 58.
5. Zakhvatkin M.V., Andrianov A.S., Avdeev V.Yu. et al. RadioAstron orbit determination and evaluation of its results using correlation of space-VLBI observations // Advances in Space Research. 2020. V. 65. Iss. 2. P. 798–812.

6. Kovalenko I.D., Eismont N.A. Orbit design for the Spectrum-Roentgen-Gamma mission // Acta Astronautica. 2019. V. 160. P. 56–61.
7. Эскобал П.Р. Методы определения орбит. М.: Мир, 1970. 471 с. (Перевод с англ. В.И. Ноздрина и В.М. Рудакова. Под ред. В.Г. Демина).
8. Губанов В.С. Оценивание стохастических параметров обобщенным методом наименьших квадратов. СПб.: Институт прикладной астрономии РАН, 1994. 12 с. (Сообщения ИПА РАН; № 60).
9. Tapley B.D., Schutz B.E., Born, G.H. Statistical Orbit Determination. Elsevier Academic Press, 2004. 563 p.
10. Эльясберг П.Е. Определение движения по результатам измерений. М.: Наука, 1976. 416 с.
11. Kalman R. A new approach to linear filtering and prediction problems transaction of the asme journal of basic // J. Basic Engineering (American Society of Mechanical Engineers). 1960. V. 82. Iss. 1. P. 35–45.
12. Sorenson H. Kalman filtering : theory and application. IEEE Press, 1985. 457 p.
13. D'Souza C. Fundamentals of Kalman Filtering and Estimation in Aerospace Engineering. NASA // Johnson Space Center. Houston, Texas, 2013. 83 p.
14. Coronel D., Guevara C. Anomaly Detection in the Uni-axial Vibration in Wind Turbines Applying Linear Regression and Z-score // Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. Seattle WA, USA. 2024. P. 169–176.
15. Chen D. and Meng D., Wang F. et al. A study of ionospheric anomaly detection before the August 14, 2021 Mw7.2 earthquake in Haiti based on sliding interquartile range method // Acta Geodaetica et Geophysica. 2023. V. 58. P. 539–551.
16. Romo-Chavero M.A., Cantoral-Ceballos J.A., Pérez J.A. et al. Median Absolute Deviation for BGP Anomaly Detection // Future Internet. 2024. V. 16(5). Art. ID. 146.
17. Grubbs F.E. Sample Criteria for Testing Outlying Observations // The Annals of Mathematical Statistics. 1950. 21(1). P. 27–58.
18. Duchnowski R., Wiśniewski Z. Accuracy of the Hodges–Lehmann estimates computed by applying Monte Carlo simulations // Acta Geodaetica et Geophysica. 2016. V. 52. P. 511–525.
19. Baireddy S., Desai S., Mathieson J. et al. Spacecraft Time-Series Anomaly Detection Using Transfer Learning // Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. Nashville, TN, USA. 2021.
20. Zapevalin P.R., Novoselov A., Zharov V.E. Artificial neural network for star tracker centroid computation // Advances in Space Research. 2022. V. 71. Iss. 9. P. 3917–3925.
21. Herrmann L., Bieber M., Verhagen W. et al. Unmasking overestimation: a re-evaluation of deep anomaly detection in spacecraft telemetry // CEAS Space J. 2024. V. 16. P. 225–237.
22. Cuéllar Carrillo S., Santos Penas M., Alonso F. et al. Explainable anomaly detection in spacecraft telemetry // Engineering Applications of Artificial Intelligence. V. 133(4). P. 1–15.
23. Biswas G., Khorasgani H., Stanje G. et al. An Application of Data Driven Anomaly Identification to Spacecraft Telemetry Data // Annual Conference of the PHM Society. 2016. V. 8(1).
24. Liu L., Tian L., Kang Zh. et al. Spacecraft Anomaly Detection with Attention Temporal Convolution Network // Neural Computing and Applications. 2023. V. 35. P. 9753–9761.
25. Ré N.P., Popplewell M., Caudill M. et al. Transformers for Orbit Determination Anomaly Detection and Classification // Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. Seattle, WA, USA. 2024. P. 6819–6827.
26. Hundman K., Constantinou V., Laporte C. et al. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding // Proc. 2018 ACM SIGSAC Conference on Computer and Communications Security. London, United Kingdom. 2018. P. 387–395.
27. Montenbruck O., Gill E. Satellite Orbits. Models, Methods, and Applications. Springer-Verlag Berlin Heidelberg, 2000. 382 p.
28. Kara I., Bazyey O. Using of Everhart's method of 15, 17, 19 and 21st-order for computation of celestial bodies' trajectories in the circumplanetary space // Astronomical School's Report. 2009. V. 6(1–2). P. 155–157.
29. Jin S., Cardellach E., Xie F. Introduction to GNSS // GNSS Remote Sensing. Remote Sensing and Digital Image Processing. V. 19. Springer, 2014. P. 3–16. ISBN: 978-94-007-7481-0.
DOI: 10.1007/978-94-007-7482-7.
30. Gulyaeva T. Investigation of total electron content deduced from TOPEX-JASON and GPS-IONEX maps over land and oceans // Proc. 37th COSPAR Scientific Assembly. Montréal, Canada. 2008. V. 37. P. 31–36.
31. Mendes V., Langley R. Tropospheric zenith delay prediction accuracy for airborne GPS high-precision positioning // Proc. Institute of Navigation 54th Annual Meeting. 1998. P. 337–347.
32. Boehm J., Niell A., Tregoning P. et al. Global Mapping Function (GMF): A new empirical mapping function based on numerical weather model data // Geophysical Research Letters. 2006. V. 33(7).
33. Kutschera M., Zajicek W. Shapiro effect for relativistic particles—testing General Relativity in a new window// Acta Physica Polonica Series B. 2009. V. 41(6).
34. Zakhvatkin M., Ponomarev Yu., Stepanyants V. Navigation support for the RadioAstron mission // Cosmic Research. 2014. V. 52(4). P. 342–352.

35. Тучин Д.А. Определение орбиты на борту космического аппарата // Известия РАН. Теория и системы управления. 2020. № 3. С. 126–147.
36. Lam Q., Junker D., Anhalt D. et al. Analysis of an Extended Kalman Filter Based Orbit Determination System // Proc. AIAA Guidance, Navigation, and Control Conference. Toronto, Ontario, Canada. 2010.
37. Luo J., Ying K., Bai L. Savitzky–Golay smoothing and differentiation filter for even number data // Signal Processing. 2005. V. 85(7). P. 1429–1434.
38. Virtanen P., Gommers R., Oliphant T.E. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python // Nature Methods. 2020. V. 17. (Suppl. 1). P. 1–12.
39. Dach R., Andritsch F., Arnold D. et al. Bernese GNSS Software Version 5.2. Bern: Astronomical Institute, University of Bern, 2015.
40. Rudnitskiy A., Mzhelskiy P.V., Shchurov M. et al. Analysis of orbital configurations for Millimetron space observatory // Acta Astronautica. 2022. V. 196(1). P. 29–41.

A RECURSIVE PARTITIONING FOR ANOMALY DETECTION IN TRACKING SATELLITE DATA

© 2025 P. R. Zapevalin

Astro Space Center of P.N. Lebedev Physics Institute, Moscow, Russia

e-mail: pav9981@yandex.ru

This study presents a method for detecting anomalous measurements in the trajectory data of spacecraft, based on recursive partitioning of the time series of observations. This method analyzes the standard deviation of the data, effectively identifying anomalous measurements characterized by elevated noise levels. A significant advantage of this approach is the lack of requirement for prior knowledge of the initial orbital approximation and the absence of a need for pre-training. It has been tested on synthetic datasets with artificially introduced anomalies, as well as on real data from the “Spektr-R” spacecraft. The results demonstrated an accuracy of 96 % compared to other traditional anomaly detection methods. The algorithm of this method is applicable to various types of orbits and scales of observations. Its code is available for public use.