

УДК 524.388:004.62

СРАВНЕНИЕ АЛГОРИТМОВ ОБРАБОТКИ КРИВЫХ БЛЕСКА ДЛЯ ПОИСКА ТРАНЗИТНЫХ ЭКЗОПЛАНЕТ ПО ОТКРЫТЫМ ДАННЫМ КОСМИЧЕСКИХ МИССИЙ

Галанцев С.П.

аспирант физического института им. П.Н. Лебедева РАН (ФИАН)

e-mail: s.p.galantsev@yandex.ru

Аннотация

Рассмотрена методика сравнительного анализа алгоритмов обнаружения транзитных экзопланет в фотометрических временных рядах миссий Kepler и TESS. Сопоставлены метод прямоугольной аппроксимации транзита Box Least Squares (BLS), метод Transit Least Squares (TLS) и влияние способов удаления низкочастотных трендов на полноту поиска. Предложен воспроизводимый дизайн испытания на открытых кривых блеска с добавлением синтетических транзитов, отдельным подбором порогов и контролем ложных срабатываний. Формализованы критерии полноты обнаружения, доли ложных тревог, точности восстановления периода и вычислительных затрат. Рассмотрены опубликованные результаты сравнения BLS и TLS и ограничения их переноса на данные разных миссий. Показано, что улучшение формы шаблона не заменяет контроль инструментальной систематики, звездной активности и неоднородности выборки. Работа имеет обзорно-методический характер; новый массовый вычислительный эксперимент по архивным кривым блеска не заявляется.

Ключевые слова: экзопланеты, транзит, кривая блеска, Kepler, TESS, BLS, TLS, фотометрия, ложное срабатывание, полнота обнаружения.

COMPARISON OF LIGHT-CURVE PROCESSING ALGORITHMS FOR TRANSITING EXOPLANET SEARCHES USING OPEN SPACE-MISSION DATA

Galantsev S.P.

Postgraduate student at the P.N. Lebedev Physical Institute of the Russian Academy of Sciences (FIAN)

e-mail: s.p.galantsev@yandex.ru

Abstract

A reproducible framework is presented for comparing transit-search methods in publicly available Kepler and TESS photometry. Box Least Squares (BLS), Transit Least Squares (TLS), and alternative detrending strategies are examined in terms of detection completeness, false alarm behavior, period recovery, and computational costs. The proposed experiment relies on stratified archival light curves and injection–recovery tests, with thresholds calibrated on independent null cases. Published BLS versus TLS results are distinguished from outcomes not computed in this work. Particular attention is paid to stellar variability, instrumental systematics, heterogeneous observing baselines, and the difference between statistical detections and confirmed planets. This is a methodological and literature-based study rather than a report of a newly executed large-scale archive experiment.

Keywords: exoplanets, transit search, light curve, Kepler, TESS, box least squares, transit least squares, completeness, false alarms.

Введение

Транзитный метод основан на регистрации малого периодического снижения потока звезды при прохождении планеты перед ее диском. Масштаб уменьшения блеска зависит прежде всего от отношения радиусов планеты и звезды; форма события дополнительно определяется параметром прохождения, потемнением к краю диска и временным разрешением наблюдений. Преимущество метода заключается в возможности анализировать длительные ряды фотометрии автоматически. Главная трудность – обнаружить слабый

повторяющийся сигнал среди инструментальных дрейфов, звездной активности, выбросов и пропусков измерений [1, 2].

Данные космических миссий Kepler и TESS доступны через архив MAST и предоставляют основу для сопоставления алгоритмов без собственной наблюдательной аппаратуры. Однако наблюдательные режимы различны: продолжительная база Kepler повышает вероятность регистрации повторных событий с большими периодами, тогда как число секторов и длительность покрытия TESS неодинаковы для разных направлений. Следовательно, единый рейтинг методов по произвольной смешанной выборке отражал бы одновременно свойства алгоритма и свойства наблюдений [6, 7].

Классический BLS аппроксимирует транзит прямоугольным провалом и эффективно перебирает период, фазу и длительность [1]. Алгоритм TLS использует более физичную форму шаблона с учетом входа, выхода и потемнения к краю диска; это потенциально повышает чувствительность к малым планетам, но увеличивает значение выбора параметров и ресурсов [2]. Предобработка фотометрии способна не только уменьшить шум, но и частично удалить сами транзиты, если окно фильтрации соизмеримо с продолжительностью события.

Цель исследования – разработать воспроизводимый протокол сравнения BLS и TLS на архивной фотометрии, явно разделяя качество восстановления сигналов и частоту ложных обнаружений. Задачи состоят в выборе сопоставимых кривых блеска, регламентации их предобработки, формализации метрик, анализе известных опубликованных сравнений и описании статистического плана независимого вычислительного эксперимента.

Методы и методики

Исходными данными служат ряды времени, нормированного потока, неопределенности и служебных битовых флагов качества из файлов FITS. Для Kepler предпочтительно фиксировать одну разновидность потока и процедуру объединения кварталов; для TESS – один тип продуктов и явное указание секторов. Вопрос о выборе SAP или систематически скорректированного PDCSAP нельзя решать автоматически: более гладкая фотометрия не всегда означает лучшее сохранение астрофизических событий [6, 7].

Для базовой выборки целесообразно сформировать три независимые группы: звезды с хорошо установленными транзитными кандидатами и известным периодом; кривые без установленных периодических транзитов, используемые в тестах ложной тревоги; звезды с выраженной переменностью, на которых оценивают устойчивость метода. Статус «нет известной планеты» не эквивалентен доказанному отсутствию транзитов, поэтому такую группу нельзя называть абсолютно отрицательным эталоном.

Критерии включения должны задаваться до поиска: доступность исходного файла и метаданных, достаточная продолжительность и доля валидных точек, допустимый диапазон звездных величин, известная частота отсчетов и отсутствие критических разрывов в исследуемом фрагменте. Не следует исключать сложные кривые лишь по факту низкой успешности обнаружения: это искусственно повышает оценку чувствительности. Все исключения вместе с идентификаторами архивных продуктов фиксируются в воспроизводимом списке. Табл. 1 представляет источники и характеристики будущего набора данных.

Таблица 1 – Источники данных

Источник	Наблюдения	Методическое ограничение
Kepler / MAST	Кварталы, длительные ряды	Разрывы и межквартальное нормирование
TESS / MAST	Секторы, неодинаковое покрытие	Различие баз наблюдений и каденса
Каталоги кандидатов	Периоды и параметры объектов	Неравнозначность статусов кандидатов

В перспективном эксперименте для каждой кривой сохраняются идентификатор наблюдения, миссия, сектор или квартал, длительность ряда, каденс, величина

фотометрического шума, доля пропусков, версия обработки и набор использованных флагов. Данные разделяются на набор настройки и отдельный набор итоговой проверки по звездным объектам, а не по фрагментам одной кривой: иначе свойства отдельной звезды попадут одновременно в обучающую и контрольную части.

Результаты

Базовая последовательность обработки включает удаление недопустимых значений, маскирование точек по флагам качества, ограниченную коррекцию выбросов, нормирование потока и устранение низкочастотного тренда. Важно проводить маскирование резких событий с осторожностью: агрессивная сигма-отсечка способна удалить глубокие транзиты вместе с инструментальными выбросами. Для сравнения желательно фиксировать общий входной ряд для обоих методов поиска.

Простой вариант устранения тренда – скользящая медиана или локальный полином. Их преимущество состоит в малом числе параметров и высокой скорости, но при коротком окне они уменьшают глубину длительных транзитов. Более гибкий подход использует регрессию гауссовских процессов, моделирующую временную переменность или систематические зависимости; в опубликованной системе K2SC такие компоненты учитываются совместно [3]. При этом чрезмерно гибкая модель способна подгонять аномалии, совпадающие с реальным транзитом.

Выбор временного масштаба фильтрации должен согласовываться с ожидаемой максимальной длительностью транзита. Для предотвращения подгонки известных событий можно применять итерационный поиск: предварительное обнаружение, маскирование транзитных интервалов, повторную оценку тренда и новый поиск. Такой подход не гарантирует полноты в системе нескольких планет, поэтому число итераций и условие прекращения необходимо задавать заранее.

Публикация Vanderburg и Johnson показала, что уменьшение систематических ошибок фотометрии K2 дает существенное улучшение точности по сравнению с необработанными рядами [4]. Однако приведенные авторами величины относятся к конкретным данным и методу коррекции, а не к универсальной прибавке полноты BLS или TLS. При переносе результатов следует различать фотометрическую точность и непосредственно вероятность успешной регистрации планетного транзита (табл. 2).

Таблица 2 – Методы анализа

Подход	Преимущество	Риск для транзитов
Скользящая медиана	Простота и скорость	Подавление длительных событий
Локальный полином	Гибкость при плавном тренде	Переобучение около разрывов
Гауссовский процесс	Учет коррелированной изменчивости	Удаление части слабого сигнала

Ни один подход к детрендингу не является безусловно предпочтительным. Его влияние следует проверять путем внесения одинаковых синтетических транзитов до и после выбранного этапа предварительной обработки.

Обсуждение

В методе BLS предполагается, что внутри транзитного окна поток отличается от внезатменного приблизительно на постоянную глубину. Для сетки периодов вычисляется критерий улучшения согласия по сравнению с моделью постоянного блеска. Упрощение формы события обеспечивает низкую стоимость поиска; однако реальные переходы на входе и выходе транзита образуют потери соответствия шаблона, особенно для слабых сигналов и высокого временного разрешения [1].

TLS использует шаблон с закругленным дном и конечной продолжительностью переходов, что позволяет применять согласованное сопоставление с ожидаемой формой транзита. Дополнительные астрономические предпосылки, например, о плотности звезды, ограничивают пространство физически возможных длительностей. Для корректного

сравнения необходимо согласовать диапазон периодов, число проверяемых циклов, требование к количеству событий и процедуру измерения статистики обнаружения [2].

Сам алгоритм поиска следует отделять от процесса подтверждения планеты. BLS и TLS возвращают периодические кандидаты, но затменно-двойные звезды, переменные звезды и загрязнение апертуры соседними источниками могут создавать похожие сигналы. После поиска необходимы независимые проверки: сравнение нечетных и четных транзитов, анализ вторичных минимумов, диагностика центроидов изображения и согласование с каталогами объектов [5].

При работе с несколькими возможными планетами сигнал наиболее сильного кандидата маскируют либо вычитают модель и запускают поиск повторно. Итеративная процедура может порождать ложные пики после неточного вычитания транзита, особенно при наличии звездных вспышек, поэтому полезно повторно исследовать исходный ряд и требовать воспроизведения сигнала в различных участках наблюдений.

$$\delta \approx \left(\frac{R_p}{R} \right)^2, \quad (1)$$

В приближении равномерно светящегося звездного диска глубина δ определяется отношением радиусов планеты R_p и звезды R . Эффект потемнения диска и геометрия траектории уточняют форму реального транзита.

$$\chi^2(P, t_0, \tau, \delta) = \sum_i \left[\frac{f_i - m_i(P, t_0, \tau, \delta)}{\sigma_i} \right]^2, \quad (2)$$

где P – пробный период;

t_0 – эпоха транзита;

τ – длительность;

δ – глубина;

f_i и σ_i – наблюдаемый поток и его ошибка;

m_i – модель BLS или TLS.

Ключевой показатель чувствительности – доля правильно восстановленных инжектированных транзитов среди всех внесенных сигналов при заданном диапазоне глубин, периодов и длительностей. Совпадение обнаруженного периода с исходным нужно проверять с учетом гармоник $P/2$ и $2P$; эти случаи следует учитывать отдельно как частично верное обнаружение периодичности, а не автоматически считать точным восстановлением физических параметров.

Частоту ложных срабатываний оценивают на независимых кривых без внесенного сигнала либо с использованием временно перемешанных данных при условии сохранения требуемых свойств шума. Перемешивание отдельных точек разрушает корреляционный шум и может сильно занижить ложную тревогу; предпочтительны блочная перестановка и тщательно подобранные нулевые выборки. Пороги статистики BLS и TLS калибруют раздельно при одинаковом целевом уровне ложных срабатываний.

Для статистического сравнения одной и той же кривой между алгоритмами необходим парный дизайн. Доли полноты по бинам периода и отношения радиусов сопровождают биномиальными доверительными интервалами, а разность полноты при совпадающих инъекциях – оценкой неопределенности с повторной выборкой по звездам. Усреднение десятков инъекций в одной звезде как полностью независимых наблюдений переоценивает точность результата.

Время выполнения измеряется на одинаковом аппаратном обеспечении и одинаковой сетке требований к периоду. При этом процесс удаления тренда учитывают отдельно от времени собственно поискового алгоритма. Сравнение одного TLS-прохода с BLS-пайплайном, включающим многократную очистку и проверку гармоник, не показывает истинное соотношение вычислительных затрат [2].

$$C = \frac{N_r}{N_i}; \quad FPR = \frac{N_f}{N_0}, \quad (3)$$

где C – полнота;

N_r – число восстановленных внесенных транзитов;

N_i – число всех внесенных;
 FPR – частота ложных тревог,
 N_f – число отрицательных тестов с сигналом-кандидатом;
 N_0 – общее число отрицательных тестов.

$$\varepsilon_p = \frac{|\hat{P} - P|}{P}, \quad (4)$$

Относительная ошибка периода ε_p рассчитывается с отдельной регистрацией гармонических совпадений; для них рекомендуется фиксировать дополнительный признак кратности периода. В табл. 2 выделены критерии, которые должны регистрироваться одновременно; отдельный высокий показатель полноты без контроля ложных срабатываний не может рассматриваться как достаточное свидетельство преимуществ метода.

Таблица 2 – Критерии регистрации

Показатель	Основание расчета	Условие корректности
Полнота С	Инжектированные транзиты	Фиксированные интервалы глубины и периода
Частота FPR	Нулевая независимая выборка	Общий заданный уровень ложных тревог
Ошибка периода	Сопоставление $\hat{P}$ и P	Раздельный учет гармоник
Время работы	Повторные запуски	Одинаковое оборудование и диапазоны

В работе Hippke и Heller проведен injection-recovery эксперимент с моделируемыми транзитами планет земного размера на фоне белого шума 110 ppm при 30-минутной частоте отсчетов. При настройке порогов по SDE на частоту ложных срабатываний 1% авторы сообщают полноту порядка 93% для TLS и 76% для BLS [2]. Разница относится именно к условиям указанного синтетического теста; она не доказывает такое же превосходство на каждой реальной кривой TESS или Kepler.

При низком отношении сигнал/шум преимущества физически обоснованного шаблона особенно существенны, если длительность и форма события оценены удовлетворительно. В задачах с большими транзитами, ограниченными вычислительными ресурсами либо недостаточно известными параметрами звезды BLS сохраняет практическую ценность. Нельзя сводить сравнение к одной доле обнаружений: цена более чувствительного поиска может проявиться в дополнительной чувствительности к систематическим повторяющимся структурам.

Публикации по обработке K2 показывают, что подавление инструментальной систематики заметно меняет характеристики входных рядов [3, 4]. Поэтому сравнение методов поиска имеет смысл только при документированном выборе фотометрического продукта, одинаковой маске качества и общей стратегии детрендинга. Замена способа предобработки может оказаться по влиянию на результат сопоставима с переходом от прямоугольного к физическому шаблону.

Наконец, открытые архивы содержат не только подтвержденные планеты, но и кандидатов различной надежности. Оценка ложных обнаружений по простому несовпадению с каталогом является методически слабой: реальный еще не внесенный в каталог транзит может быть ошибочно объявлен ложным. Более надежная стратегия комбинирует контролируемые инъекции, отдельно подготовленную нулевую выборку и независимую астрофизическую проверку найденных событий.

Таблица 3 – Оценка Hippke и Heller

Условие теста из [2]	TLS	BLS
Шум 110 ppm / 30 мин	Одинаковая симуляция	Одинаковая симуляция
FPR, целевой, %	1	1
Доля восстановленных, %	93	76

Числа в табл. 3 воспроизводят опубликованную оценку Hippke и Heller для конкретного синтетического опыта, а не результат повторного анализа автора этой статьи. Их статистическая и физическая переносимость ограничена исходными условиями моделирования.

Дополнительные ограничения связаны с неполной наблюдательной базой, временными разрывами, разной звездной активностью и недостаточно известными параметрами звезд. Поиск планет с длительным периодом в короткой фотометрической серии может быть неидентифицируемым при любом алгоритме; алгоритмическая ошибка здесь должна быть отделена от фундаментального отсутствия повторных транзитов. Также требуется исключить систематическое улучшение метрик за счет подбора порогов на итоговой контрольной выборке.

Воспроизводимая публикация полноценного эксперимента должна включать список объектов и файлов MAST, версии Lightcurve/BLS/TLS, правила маскирования и интерполяции, параметры сеток поиска, семена генератора для инъекций, диапазоны физических параметров и программный код оценки метрик. Эти условия важны потому, что даже малые различия в длине ряда и алгоритме коррекции тренда могут изменить ранжирование кандидатов.

Перспективно дополнить методику оценкой устойчивости к звездным вспышкам, транзитным вариациям времени, системам нескольких планет и коррелированным шумам. Для каждого такого фактора желательно строить отдельную карту полноты, а не единую усредненную характеристику. Практически полезным итогом станет выбор не единственного «лучшего» алгоритма, а разных рабочих конфигураций для больших и малых транзитов, коротких и длинных рядов.

Заключение

Обоснована сопоставимая методика анализа BLS и TLS, включающая единую входную фотометрию, контролируемые варианты удаления тренда, инъекцию синтетических транзитов и независимую калибровку порогов по частоте ложных срабатываний. Такой дизайн позволяет исследовать собственно алгоритмическую чувствительность, не смешивая ее с различиями миссий и сложностью отдельных звездных объектов.

По опубликованному контролируемому тесту TLS демонстрировал более высокую полноту обнаружения землеподобных транзитов при сопоставимом уровне ложных тревог, однако результаты не должны переноситься без проверки на архивные ряды иных характеристик. BLS остается вычислительно экономичным и прозрачным базовым методом, особенно для сигналов с выраженными и относительно глубокими провалами блеска.

Показано, что критерии чувствительности, ложных срабатываний, точности периода и вычислительных затрат должны оцениваться совместно. В качестве следующего этапа необходим расчет на зафиксированной выборке Kepler/TESS с открытым перечнем объектов и публикацией программных параметров.

Литература

1. Kovács G., Zucker S., Mazeh T. A box-fitting algorithm in the search for periodic transits // *Astronomy & Astrophysics*. 2002. Vol. 391. PP. 369-377.
2. Hippke M., Heller R. Optimized transit detection algorithm to search for periodic transits of small planets // *Astronomy & Astrophysics*. 2019. Vol. 623. Art. A39.
3. Aigrain S., Parviainen H., Pope B.J.S. K2SC: flexible systematics correction and detrending of K2 light curves using Gaussian process regression // *Monthly Notices of the Royal Astronomical Society*. 2016. Vol. 459, № 3. PP. 2408-2419.
4. Vanderburg A., Johnson J.A. A Technique for Extracting Highly Precise Photometry for the Two-Wheeled Kepler Mission // *Publications of the Astronomical Society of the Pacific*. 2014. Vol. 126, № 944. PP. 948-958.
5. Jenkins J.M. Overview of the Kepler Science Processing Pipeline // *The Astrophysical Journal Letters*. 2010. Vol. 713, № 2. – PP. 87-91.

6. Mikulski Archive for Space Telescopes. Kepler mission data. URL: <https://archive.stsci.edu/missions-and-data/kepler> (дата обращения: 21.10.2026).
7. NASA Goddard Space Flight Center. TESS Data Products Information. URL: <https://heasarc.gsfc.nasa.gov/docs/tess/data-products.html> (дата обращения: 21.10.2026).
8. NASA Exoplanet Archive. Bulk Data Download: Kepler Time Series. URL: https://exoplanetarchive.ipac.caltech.edu/bulk_data_download/ (дата обращения: 21.10.2026).
9. Hounsell R. An introduction into the tools and tutorials available for the analysis of TESS data [Электронный ресурс] // NASA HEASARC. URL: <https://heasarc.gsfc.nasa.gov/docs/tess/TS-C2-Intro.html> (дата обращения: 21.10.2026).