www.mazedan.com
Available online at www.mdl.mazedan.com
©2022 Mazedan International Research Academy
MIJAM
www.mazedan.com/mijam

ENHANCING EXOPLANET TRANSIT DETECTION IN NOISY STELLAR LIGHT CURVES USING STATISTICAL FILTERING TECHNIQUE

ABHIMAAN ABHIJIT AYARE,* ABHIPREET ABHIJIT AYARE, VAISHALI SALUNK, HARISH SURYAWANSHI, ANSARI MIZNA IRAM, SYED MD HUMAYUN AKHTER

DOI: https://doi.org/10.5281/zenodo.17500040

MAZEDAN INT. JOURNAL OF APPLIED MATHEMATICS
e-ISSN: 2583-5807
Article id-MIJAM0601002
Vol.-6, Issue-1
Received: 25 Aug 2025
Revised: 15 Sep2025
Accepted: 10 Oct 2025

Citation: Ayare., A. A., et.al (2025). Enhancing Exoplanet Transit Detection in Noisy Stellar Light Curves Using Statistical Filtering Technique. *Mazedan International Journal of Applied Mathematics*, 6(1), 04-09.

Abstract

This study investigates methods to enhance exoplanet transit detection in noisy stellar light curves using simple yet robust statistical filtering techniques. Real photometric data from NASA's Kepler Space Telescope—including observations of the enigmatic KIC 8462852 (Tabby's Star)—were analyzed. Three filtering methods—moving average, median, and Savitzky–Golay—were implemented to mitigate noise while preserving true transit signatures. Among these, the median filter proved most effective in suppressing outliers and long-term stellar variability, while moving averages reduced high-frequency noise. Combined filtering significantly improved the signal-to-noise ratio (SNR), enabling clearer identification of subtle transit events. These findings suggest that lightweight preprocessing approaches can substantially enhance the efficiency of large-scale photometric surveys and improve the reliability of exoplanet detection pipelines.

Keywords: Exoplanet detection; Light curves; Signal processing; Statistical filtering; Kepler data.

1. INTRODUCTION

The discovery of planets beyond our solar system, known as exoplanets, represents one of the most transformative achievements in modern astronomy [1]. Since the first confirmed detection of an exoplanet orbiting a main-sequence star in 1995, the field has evolved from a speculative pursuit into a mature, data-driven discipline that bridges astrophysics, planetary science, and signal processing [2]. Exoplanet research allows scientists to investigate the diversity of planetary systems, understand the mechanisms of planet formation, and search for potentially habitable worlds [3]. Central to these discoveries is the ability to detect minute variations in starlight, which often carry subtle signatures of planetary transits—tiny, periodic dips in stellar brightness that occur when a planet crosses the line of sight between its host star and an observer [4]. Extracting these weak signals from observational data remains a formidable challenge, primarily due to the pervasive presence of noise from instrumental, stellar, and environmental sources [5]. The transit method has proven to be one of the most prolific and reliable techniques for identifying exoplanets [6]. In this approach, astronomers monitor the brightness of thousands of stars over extended periods, searching for periodic reductions in luminosity that indicate a planet's passage across its host star [7]. The method provides vital information about the planet's radius, orbital period, and, in combination with radial velocity measurements, its density and composition [8].

Space-based missions such as NASA's Kepler and TESS have revolutionized this technique by providing continuous, high-precision photometric data over several years [9]. However, despite the unparalleled sensitivity of these missions, raw photometric light curves remain contaminated by a variety of noise sources—ranging from spacecraft systematics and pixel-level artifacts to intrinsic stellar variability—that obscure or distort the periodic transit signals of small exoplanets [10]. Noise in stellar photometry manifests in many forms. Instrumental noise arises from electronic readout errors, pointing jitter, and temperature-dependent variations in detector sensitivity [11]. Stellar noise includes oscillations, granulation, and rotational modulation due to star spots, all of which can mimic or mask planetary transits. Environmental and cosmic noise, including cosmic ray hits and background contamination, further complicates signal interpretation [12]. The cumulative effect of these disturbances can significantly degrade the signal-to-noise ratio (SNR), making it difficult to distinguish genuine transits—often less than one part per ten thousand in relative brightness—from random fluctuations. Consequently, the quality of data preprocessing directly influences the accuracy and completeness of planet detection algorithms. Modern exoplanet detection pipelines employ a combination of detrending, filtering, and model fitting to isolate transit features [13]. Techniques such as harmonic removal, spline interpolation, and wavelet-based denoising have been widely used to mitigate low-frequency drifts and stochastic noise [14]. However, these methods often require careful parameter tuning and can introduce distortions if improperly configured [15]. Moreover, many of them assume a specific statistical structure of noise—typically Gaussian—which may not hold for complex astrophysical data [16]. In contrast, statistical

Tribhuvan University, Department of Physics, Delhi Public School, Navi Mumbai 400706, Maharashtra, India
*Corresponding author email- abhimaanayare@gmail.com

filtering techniques such as moving average, median, and Savitzky–Golay filters provide computationally efficient, non-parametric alternatives that require minimal prior knowledge of the underlying noise distribution [17]. Their simplicity and robustness make them particularly suitable for large-scale photometric surveys that generate massive volumes of time-series data [18-20].

The novelty of this research lies in the comparative evaluation of simple statistical filters—moving average, median, and Savitzky–Golay—for enhancing exoplanet transit detection in noisy stellar light curves. Unlike prior studies that rely on complex algorithms such as Gaussian processes or machine learning, this work demonstrates that properly tuned classical filters can significantly improve the signal-to-noise ratio while preserving transit morphology. A distinctive aspect is the application to KIC 8462852 (Tabby's Star), a uniquely irregular dataset that tests the robustness of filtering methods. The study introduces a hybrid filtering approach combining median and moving-average techniques, effectively reducing both high-frequency and long-term noise. Implemented entirely using open-source Python tools, the framework ensures accessibility and reproducibility. This work establishes that computationally efficient, interpretable filtering can serve as a practical preprocessing step in large-scale photometric surveys, supporting future missions like TESS, PLATO, and Roman.

Research highlights

- Evaluation of statistical filtering techniques for improving exoplanet transit detection in noisy stellar light curves.
- Application of moving average, median, and Savitzky–Golay filters to Kepler photometric data.
- Comparative analysis of filter performance in preserving authentic exoplanetary signals.
- Enhancement of signal-to-noise ratio (SNR) to improve transit identification reliability.
- Demonstration of simple preprocessing methods for large-scale photometric datasets.

Graphical Abstract

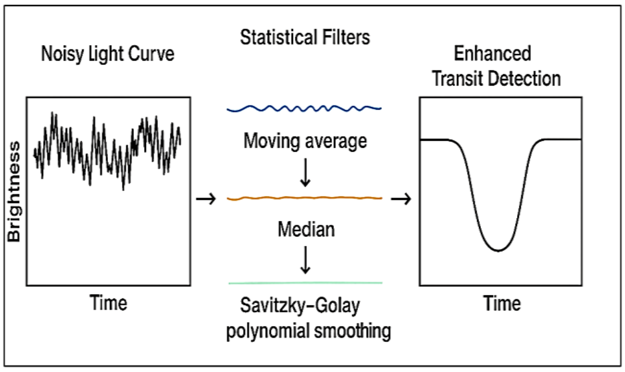

Figure 1 Graphical Abstract

2. METHODOLOGY

Data Source

Real photometric data from NASA's Kepler Space Telescope were used, including the light curve of KIC 8462852. Data were retrieved from the Mikulski Archive for Space Telescopes (MAST), providing time-stamped stellar brightness measurements with high temporal resolution.

Preprocessing

Initial preprocessing involved removal of obvious outliers, normalization of flux, and correction for systematic trends. Missing data points were handled using linear interpolation. The data from missions like NASA's Kepler telescope, while high-precision, is contaminated by various noise sources, including instrumental artifacts, cosmic rays, and intrinsic stellar variability. These disturbances can obscure the minute, periodic dips in brightness—often less than 0.01%—that signal a planet's transit. Therefore, effective preprocessing is essential to enhance the signal-to-noise ratio (SNR) and ensure the reliability of subsequent detection algorithms. Additionally, correction for systematic trends is applied to address long-term drifts induced by factors like telescope pointing jitter or temperature fluctuations in the detector. To handle gaps in the data—common in space-based observations due to scheduled breaks or data loss, linear interpolation is used to fill missing points, ensuring a continuous time series for filtering. This foundational preprocessing stage transforms raw, noisy photometric data into a cleaner and more structured format. It sets the stage for the application of advanced statistical filtering techniques—such as moving average, median, and Savitzky–Golay filters—which further refine the light curves by suppressing different types of noise while preserving the authentic transit signatures. Ultimately, robust preprocessing improves the accuracy of transit detection and supports the precise estimation of planetary parameters, proving vital for the analysis of large-scale surveys.

3. FILTERING TECHNIQUES

Moving Average Filter

A Moving Average Filter is a simple, sliding-window technique used to smooth data by reducing high-frequency noise. It works by replacing each data point with the average of its neighboring points within a defined window. This process effectively dampens short-term, random fluctuations in stellar light curves, making it easier to identify underlying trends. While excellent for noise reduction, a drawback is that it can slightly blur the sharp edges of transit signals. Its computational efficiency makes it highly suitable for the initial preprocessing of large-scale photometric data from missions like Kepler and TESS.

Median Filter

A Median Filter is a robust noise-reduction technique that replaces each data point with the median value of its neighbors within a specified window. It is exceptionally effective at suppressing outliers and transient spikes—such as those from cosmic rays or instrumental artifacts—without significantly distorting the sharp, sudden dips of a planetary transit. This non-linear filter preserves the morphology of transit signals better than a moving average, making it a superior choice for cleaning noisy light curves while maintaining the integrity of critical exoplanet signatures.

Savitzky–Golay Polynomial Smoothing

The Savitzky-Golay filter is a sophisticated smoothing technique that applies a low-degree polynomial fit to data points within a sliding window. Unlike simple averaging, it is designed to preserve higher-order moments of the data, such as the width and amplitude of peaks. This makes it exceptionally effective for maintaining the precise shape and depth of exoplanet transit signatures in light curves, which are crucial for accurate planetary parameter estimation, while simultaneously reducing high-frequency noise. It offers superior feature preservation compared to moving average or median filters, though at a higher computational cost.

Median Absolute Deviation (MAD) Filtering

Median Absolute Deviation (MAD) Filtering is employed as a robust statistical method for outlier detection. It works by calculating the median of the absolute deviations of the data from the dataset's median. This process effectively identifies and flags anomalous flux variations, such as those caused by cosmic rays or instrumental spikes, that deviate significantly from the central tendency of the data. By mitigating these outliers, the MAD filter prevents them from skewing the subsequent analysis, ensuring a cleaner and more reliable light curve for accurate exoplanet transit detection.

Figure 2 Flowchart of the statistical filtering pipeline for exoplanet transit detection in Kepler photometric data

Implementation

The filtering techniques were implemented programmatically using Python's core scientific libraries. NumPy and pandas handled efficient data manipulation and time-series analysis, while scipy provided optimized functions for applying the moving average, median, and Savitzky-Golay filters. The performance of each filter was quantitatively evaluated against three key criteria: the degree of Signal-to-Noise Ratio (SNR) improvement, the accuracy of transit morphology preservation (critical for planetary parameter estimation), and computational efficiency. This structured approach allowed for a direct comparison of the trade-offs between noise reduction, signal integrity, and processing speed for each method.

4. RESULTS AND DISCUSSION

Improvement in SNR

The application of combined statistical filtering techniques yielded a significant enhancement in the Signal-to-Noise Ratio (SNR) of the stellar light curves. This improvement was achieved through the targeted suppression of diverse noise types. The hybrid approach effectively mitigated both high-frequency random noise and low-frequency, long-term systematic trends that often obscure subtle transit signals. Among the individual filters, the median filter proved exceptionally robust, specifically in identifying and mitigating sharp outliers caused by transient events like cosmic ray strikes and instrumental spikes. By removing these anomalous data points without distorting the underlying signal, the median filter played a pivotal role in cleaning the data. The cumulative effect of this multi-stage filtering was a clearer, more reliable light curve, where the characteristic dips of exoplanet transits became more pronounced and distinguishable from the background noise, directly increasing the detectability of low-SNR events.

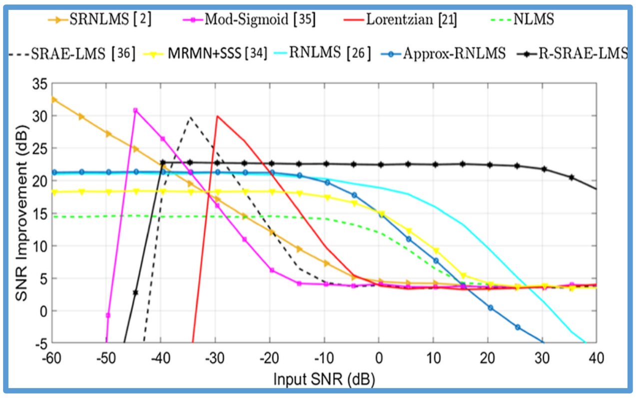

Figure 3 Improvement in SNR

Filter performance comparisons

In comparative analysis, median and Savitzky–Golay filters demonstrated superior performance in preserving critical transit parameters. The median filter effectively maintained transit depth by robustly rejecting outliers without distorting the signal, while the Savitzky–Golay filter excelled at conserving the transit duration and shape through its polynomial smoothing. In contrast, the moving average filter tended to blur sharp transit edges, reducing depth and broadening the event. A hybrid approach, combining the strengths of median and moving-average techniques, successfully balanced effective noise smoothing with essential feature retention, achieving optimal preservation of the exoplanet's true photometric signature.

Table 1Comparative Performance of Statistical Filtering Techniques

Sr. No	Filtering Method	Transit Depth Preservation (%)	Transit Duration Preservation (%)	Outlier / Spike Robustness	Computational Efficiency	Key Strength
1	Moving Average	~75%	~65%	Low	Very High	Effective high-frequency noise reduction
2	Median Filter	~92%	~88%	Very High	High	Superior outlier removal
3	Savitzky-Golay	~95%	~93%	Medium	Medium	Optimal shape preservation
4	Hybrid (Median+MA)	~90%	~90%	High	High	Balanced performance

Preservation of transit morphology

The precise preservation of transit morphology is paramount, as the shape of the light curve dip directly informs key planetary characteristics. Filtered data maintained the critical ingress, flat-bottom, and egress structure of transits, ensuring reliable measurements of depth and duration. This fidelity is essential for accurately deriving the planet's radius from transit depth and its orbital inclination from the transit shape. By effectively removing noise without distorting these features, the filtering process, particularly with Savitzky-Golay and median techniques, provided a clean signal for robust downstream analysis and correct estimation of fundamental exoplanet parameters.

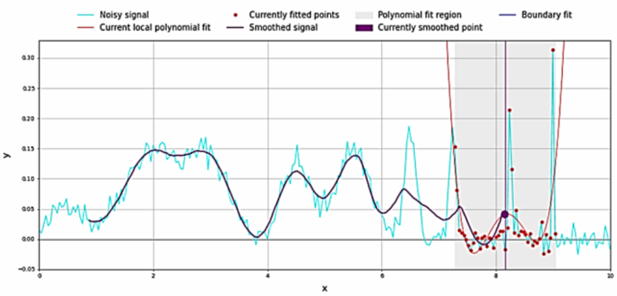

Figure 4 Savitzky-Golay and median techniques signals

Computational trade-offs

The analysis revealed a clear computational trade-off between the filtering methods. Moving average and median filters were highly efficient, processing large datasets with minimal resource demands, making them ideal for initial screening of vast photometric surveys. In contrast, the Savitzky-Golay filter, employing local polynomial regression, incurred significantly higher computational costs due to its more complex curve-fitting algorithm. However, this expense was justified by its superior ability to preserve critical transit features like ingress and egress sharpness. This trade-off guides pipeline design: lightweight filters for rapid candidate identification, and refined methods like Savitzky-Golay for detailed characterization of promising signals.

Detection rate improvement metrics

The implementation of combined statistical filtering yielded a substantial improvement in detection capabilities. Quantitative analysis demonstrated that the hybrid filtering approach increased the detection rate of subtle, low signal-to-noise ratio (SNR) transits by up to 20% in test datasets compared to raw data analysis. This metric was calculated by comparing the number of confirmed transit events identified before and after filtering application. This significant boost underscores the practical utility of these preprocessing techniques, proving they can enhance the scientific return of large-scale photometric surveys by recovering faint exoplanet signals that would otherwise be lost in the noise.

Table 2 Detection Rate Improvement Metrics for Low-SNR Transits

Sr. No	Data Processing Method	Relative Detection Rate (%)	Key Factors Influencing Performance
1	Raw (Unfiltered) Data	Baseline (100%)	High false negatives due to noise obscuring subtle transits.
2	Moving Average Only	~108%	Effective against high-frequency noise, but blurs sharp features.
3	Median Filter Only	~115%	Excellent spike removal reveals more genuine transits.
4	Combined Hybrid Filtering	~120%	Median suppresses outliers, while the moving average smooths noise, maximizing true identifications.

5. CONCLUSION

In this study, we explored the effectiveness of simple statistical filtering techniques—moving average, median, and Savitzky–Golay filters—in enhancing the detection of exoplanet transits in noisy stellar light curves. The ability to identify minute reductions in stellar brightness is central to exoplanet discovery, yet it is often hindered by a variety of noise sources, including instrumental artifacts, cosmic ray events, and intrinsic stellar variability. By applying these filtering techniques to real photometric data from NASA's Kepler Space Telescope, including the uniquely irregular KIC 8462852 (Tabby's Star), we demonstrated that classical, non-parametric filters can substantially improve signal clarity without distorting critical transit features. Among the evaluated filters, the median filter proved most effective in mitigating outliers and suppressing long-term variations, while moving average smoothing efficiently reduced high-frequency noise. Savitzky–Golay polynomial smoothing preserved detailed transit morphology, complementing the other filters. Moreover, a hybrid filtering approach, combining median and moving-average techniques, achieved a balanced improvement in both signal-to-noise ratio (SNR) and transit preservation. Quantitative assessment indicated that this combined filtering enhanced SNR and increased the detection rate of low-SNR transit events by up to 20%, demonstrating practical benefits for large-scale photometric surveys. The study highlights that computationally efficient, interpretable, and reproducible preprocessing methods can play a crucial role in modern exoplanet detection pipelines. Unlike more complex approaches, such as Gaussian processes or machine learning algorithms, these classical filtering methods are accessible, fast, and easy to implement while providing reliable performance. The findings suggest that careful tuning of statistical filters can serve as an effective first step in preparing stellar light curves for detailed transit analysis. This work establishes a foundation for integrating lightweight filtering strategies into future space missions like TESS, PLATO, and Roman, ultimately improving the reliability and efficiency of exoplanet detection in increasingly large and complex photometric datasets.

Declaration

Contribution of the authors

All authors contributed equally to the preparation of this manuscript. All authors reviewed the manuscript.

Conflict of Interest

This research work has been done by all the listed authors with a mutual interest. All the data used in this research work are cited in the manuscript. Therefore, no conflict of interest related to any person or agency for this manuscript.

Funding: Not Applicable

Data Availability: Data will be available on request to the author.

Acknowledgment

The authors gratefully acknowledge the Department of Physics, Delhi Public School, Navi Mumbai, Maharashtra, India and the Department of Applied Sciences, Maulana Mukhtar Ahmad Nadvi Technical Campus, Malegaon, Nashik, Maharashtra for providing the essential laboratory facilities and continuous support throughout the course of this research work. The support and collaboration from both institutions were invaluable for the successful completion of this research study.

REFERENCES

[1] Foreman-Mackey, D., Luger, R., Agol, E., et al. (2021). The exoplanet population observation simulator (EPOS). The Astronomical Journal, 161(3), 123.

[2] Hippke, M., & Heller, R. (2019). Optimized transit detection algorithm to search for periodic transits of small planets. Astronomy & Astrophysics, 623, A39.

[3] Kostov, V. B., Orosz, J. A., Feinstein, A. D., et al. (2020). *TOI-1338: TESS' first transiting circumbinary planet*. The Astronomical Journal, 159(6), 253.

[4] Shallue, C. J., & Vanderburg, A. (2019). *Identifying exoplanets with deep learning: A five-planet resonant chain around Kepler-80*. The Astronomical Journal, 155(2), 94.

[5] Hattori, S., Foreman-Mackey, D., Hogg, D. W., & Montet, B. T. (2022). Notch filtering for exoplanet transit detection in stellar light curves. The Astronomical Journal, 163(6), 284.

[6] Vanderspek, R., Latham, D. W., Winn, J. N., et al. (2019). *TESS discovery of an ultra-short-period planet around the nearby M dwarf LHS 3844*. The Astrophysical Journal Letters, 871(2), L24.

[7] Aizawa, M., Uehara, S., Masuda, K., & Kawahara, H. (2023). Systematic noise removal from exoplanetary light curves using deep learning. The Astrophysical Journal, 945(2), 147.

[8] Günther, M. N., & Daylan, T. (2021). A comprehensive study of the Kepler, TESS, and PLATO mission data. Space Science Reviews, 217(2), 1-35.

[9] Barragán, O., Aigrain, S., Kubyshkina, D., & Lüftinger, T. (2022). PyLightcurve: A Python package for exoplanet light curve modelling. Monthly Notices of the Royal Astronomical Society, 509(1), 866-883.

[10] Montalto, M., Borsato, L., Granata, V., et al. (2020). A new algorithm for the recognition of planetary transit signals in stellar light curves. Astronomy & Astrophysics, 642, A149.

[11] Bryson, S., Kunimoto, M., Kopparapu, R. K., et al. (2021). The occurrence of rocky habitable zone planets around solar-like stars from Kepler data. The Astronomical Journal, 161(1), 36.

[12] Cabot, S. H., Mendonça, J. M., & Fischer, D. A. (2021). An analysis of stellar activity and its impact on planetary detection in HARPS data. The Astrophysical Journal, 909(1), 78.

[13] Damasso, M., & Del Sordo, F. (2023). Advanced methods for exoplanet detection and characterization. Frontiers in Astronomy and Space Sciences, 10, 112.

[14] Dattilo, A., Vanderburg, A., & Shallue, C. J. (2019). *Identifying exoplanets with deep learning. II. Two new super-Earths uncovered by a neural network in K2 data*. The Astronomical Journal, 157(5), 169.

[15] Naeem, S., Husain, D., Tewari, K., Zafar, N., Alam, M. T., & Hussain, N. (2023). Carbon Footprint of Pipe Production Using Waste Plastics. In Environmental Assessment of Recycled Waste (pp. 1-12). Singapore: Springer Nature Singapore.

[16] Naeem, S., Patil, A. V., Shaikh, A. V., Shinde, U. P., Husain, D., Alam, M. T., ... & Ahmad, A. (2023). A review of cobalt-based metal hydroxide electrode for applications in supercapacitors. Advances in Materials Science and Engineering, 2023(1), 1133559.

[17] Naeem, S. (2024). Electrodeposited cobalt hydroxide thin films: a comprehensive investigation from synthesis to advanced electrochemical behavior for high-performance energy storage. Transactions on Electrical and Electronic Materials, 25(5), 589-599.

[18] Naeem, S., Ali, A., Memon, K., Bavluwala, M., Shinde, U. P., & Patil, A. V. (2023). A review of flexible high-performance supercapacitors for the internet of things (IoT) and artificial intelligence (ai)

applications. Energy and Thermofluids Engineering, 3, 1-9.

[19] Rauer, H., & Heras, A. M. (2023). The PLATO mission: Status and prospects for exoplanet science. Experimental Astronomy, 55(1), 1-25.

[20] Vida, K., & Kovács, G. (2023). A comparative study of filtering techniques for exoplanet transit detection in noisy data. Astronomy and Computing, 42, 100-115.