Stellar Flare Detection with Isolation Forests

Alexander Faassen (1005187174)

April 15, 2025

Abstract

This project investigates the use of Isolation Forests for detecting stellar flares in time-series brightness data from the TESS mission. The flux data is detrended using Seasonal-Trend decomposition with Loess (STL), and anomaly detection is applied to the residuals. Model performance is evaluated through repeated injection-recovery experiments using simulated flares of varying intensities, with comparisons made against a baseline $3 - 3\sigma$-clipping method. Results show that while both models perform well on large flares, the Isolation Forest outperforms the baseline in detecting small and varied flares. These findings highlight the robustness and flexibility of Isolation Forests as an unsupervised tool for flare detection in astrophysical time series.

1 Introduction

Stellar flares are intense bursts of radiation caused by magnetic reconnection near starspots. These flares, while brief, can have significant impacts on planetary atmospheres and are important for understanding stellar behavior. Detecting them automatically is essential for large-scale surveys, like the NASA TESS Mission, which produces massive volumes of time-series brightness (flux) data.

Isolation Forests are random forest-based anomaly detection models. Their efficiency, scalability, and ability to detect local outliers, such as sudden flux spikes, without requiring labelled training data make them well-suited for flare detection. Alongside the IF algorithm, classical time series techniques such as ARIMA modelling and Seasonal-Trend decomposition using Loess (STL) can preprocess the data to isolate flare signatures. The target dataset comprises a brightness time series from an M dwarf star observed by the TESS Mission, TIC 031381302, chosen for its periods of both seemingly quiescent and flaring behavior.

The key research question is:

Can an Isolation Forest be used to detect stellar flares by the brightness of a star over time?

In addition to these technical goals, this project also placed an emphasis on developing personal skills. In particular, working with data in python (e.g. pandas library) and time series analysis which inspired some of the directions explored.

2 Data

The dataset comes from the TESS Mission, a survey designed to detect exoplanets, planets orbiting nearby stars. The NASA space telescope observes different sectors of the sky for approximately 27 days at a time. The time series of interest is the Pre-search Data Conditioning SAP (PDCSAP) flux, which represents Simple Aperture Photometry (SAP) flux with long-term instrumental trends and errors removed (e.g., scattered light). SAP flux quantifies the aggregate brightness (in electrons/second) of each pixel associated with a given stellar object. These measurements are recorded at a 2-minute cadence.

Importantly, flare events are not labelled in this dataset, necessitating an unsupervised learning approach. Moreover, the data contain irregular gaps due to satellite orbital constraints, technical issues, and data quality filtering.

For this project, we analyze 17,719 consecutive observations, roughly 7 days, of PDCSAP flux from the M dwarf star TIC 031381302. This target was selected for its mix of seemingly quiescent and flaring periods, providing a diverse testing ground for anomaly detection. Of these, 686 observations are missing flux values, while an additional 1,137 are missing entirely, corresponding to a large continuous gap, likely from a technical interruption. Altogether, the series contains approximately 1,823 missing observations, prompting an investigation into imputation strategies using forecasting models.

2.1 Exploratory Data Analysis

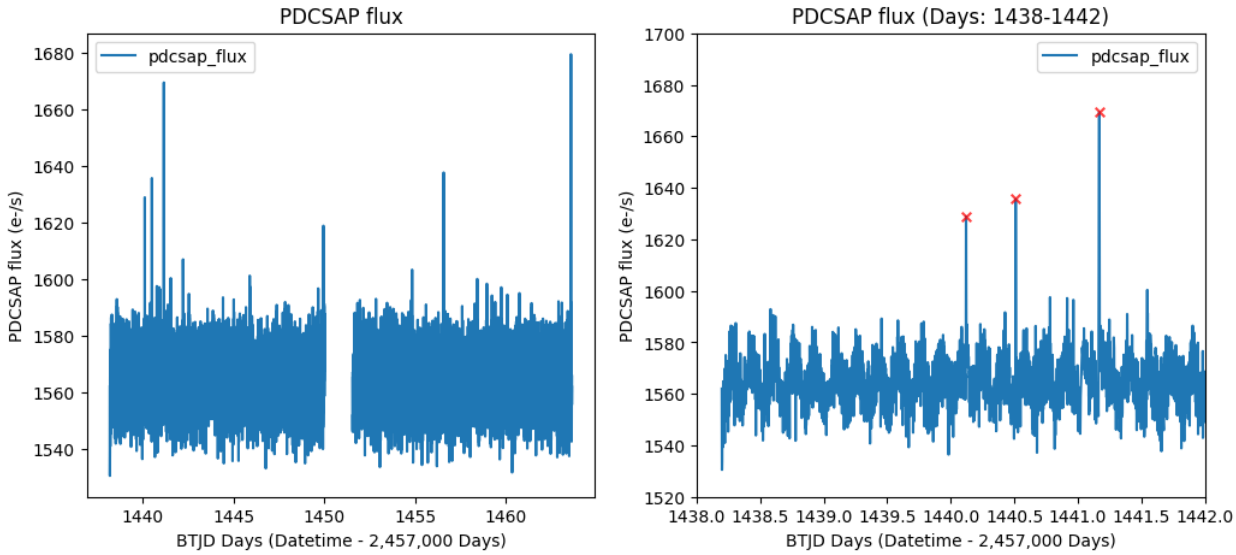


Figure 1: Left: PDCSAP flux time series of interest. Right: Same time series filtered to Days 1438-1442 to better observe seasonal patterns and identify outliers in red.

Figure 1 (left) provides a holistic view of the full PDCSAP flux time series, where several outliers are clearly visible. To better examine periodic patterns and potential flare activity, we zoom in on Days 1438–1442 (Figure 1, right). This closer view reveals a repeating oscillatory pattern, occurring approximately three times per day, as well as several sharp flux spikes marked in red, likely candidates for flare events.

To assess whether the time series is stationary, we apply the **Augmented Dickey-Fuller (ADF)** unit root test. The test yields a statistic of –30.35 and a p-value near zero, providing strong evidence against the null hypothesis of a unit root. Thus, the time series appears stationary, having consistent behaviour over time.

A brief examination of the **autocorrelation function (ACF)** revealed further support for this conclusion, displaying a clear periodic structure. This suggests regular fluctuations, likely arising from stellar rotation or instrumental effects. Meanwhile, the **partial autocorrelation function (PACF)** exhibited strong correlations at lower lags before diminishing, indicating that an Auto-Regressive (AR) model may be appropriate.

The left panel of Figure 2 shows a **periodogram** that reveals sharp peaks at low frequencies. This suggests the presence of long-term or low-frequency variations in the signal. Combined with the results of the ADF and autocorrelation functions, this points to persistent cyclic behaviour, although the absence of sharp harmonic peaks implies that the oscillations are not purely sinusoidal.

On the right of Figure 2, we display the output of an STL decomposition using a period of 240 observations (approximately 8 hours). The trend component reveals slow-moving fluctuations, consistent with the low-frequency peaks seen in the periodogram. The seasonal component exhibits smooth, periodic variations, reinforcing the impression of regular cycles. The residuals show clear spikes and irregularities, likely corresponding to flare events.

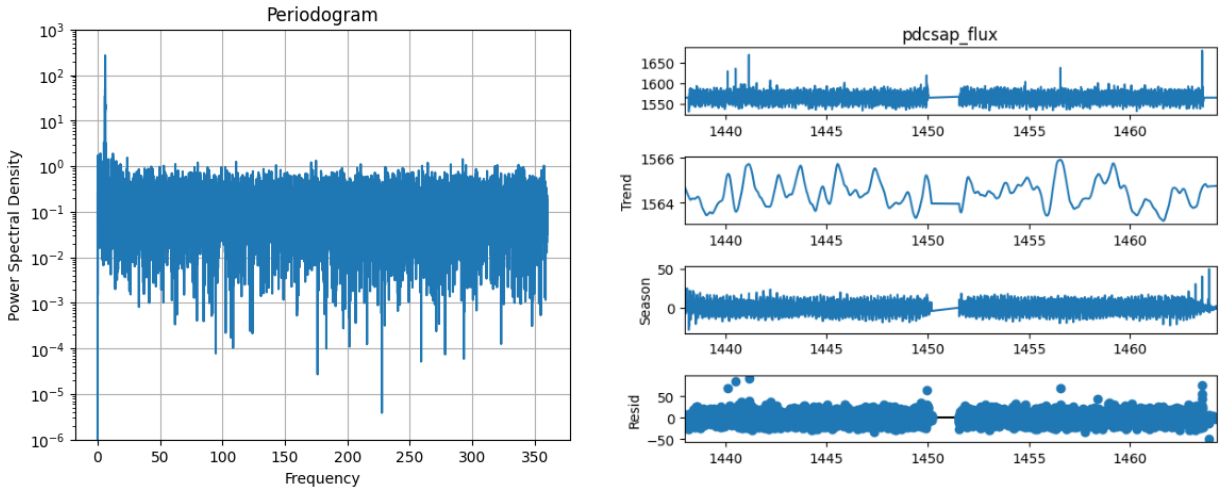


Figure 2: Left: Periodogram. Right: Diagnostic plot of STL decomposition run with a period of 240 corresponding to roughly 8-hour periods.

3 Methods

This section outlines the methodological steps taken throughout the project. Initially, missing value imputation was explored to address the 1,823 missing observations identified during preliminary data analysis. Next, detrending was applied using STL decomposition to separate the signal into trend, seasonal, and residual components, allowing flare signatures to be more clearly isolated from periodic variations.

The final pipeline adopted a two-step approach, applying the anomaly detection model to the residual component produced by the detrending process. Specifically, an Isolation Forest (IF) was used to flag flux anomalies potentially corresponding to stellar flares. Lastly, flare simulation techniques were utilized to assess model performance in the absence of labelled data.

3.1 Missing value imputation

As highlighted earlier, this dataset contains a substantial number of missing values. While imputation can risk introducing bias, it may also provide a more stable input for anomaly detection models. To explore its utility, and as part of a broader learning goals in Python and time series analysis, several ARIMA-based forecasting strategies were attempted. Below is a brief summary of these efforts, including motivations, outcomes, and limitations.

Attempts:

1. **Reindexing:** Missing indices (timestamps) were reintroduced with null PDCSAP flux values at the 2-minute cadence to prepare for imputation.
2. **STL Decomposition + ARIMA on Residuals:** Required an auxiliary imputation method, opting for interpolation, since neither STL nor ARIMA handle missing values. Used in-sample predictions to impute missing values.
3. **SARIMA (Seasonal ARIMA):** SARIMA models do not work well for large periods, such as 240. Requires auxiliary imputation.
4. **ARIMA with Fourier Terms (Exogenous Seasonal Dummy Variables):** Attempted to incorporate seasonal structure using Fourier-transformed regressors. Although promising in theory, results were inconsistent and still required auxiliary imputation.

Despite multiple attempts, none of these methods produced consistently reliable results. Ultimately, imputation was excluded from the final modelling pipeline to avoid introducing artificial structure with bias. Instead, **missing values were removed entirely**. Nonetheless, these exploratory efforts were valuable for skill development in time series modelling and data analysis in Python.

3.2 Detrending

To isolate flare signatures from recurring patterns in the flux data, the time series is first detrended using **Seasonal-Trend decomposition with Loess (STL)**, which separates the series into trend, seasonal, and residual components. Once decomposed, models can either focus solely on the residuals or incorporate information from the trend and seasonal components.

Earlier exploratory analysis indicated that a period of 240 observations, roughly 8 hours, provided strong STL diagnostics and effectively captured the underlying periodic behaviour. However, applying detrending prior to anomaly detection presents a trade-off: while it helps remove regular structure, it may also absorb meaningful variation, including flares, reducing their detectability.

This approach assumes that a well-fitted trend and seasonal decomposition will leave significant anomalies, such as stellar flares, in the residuals. Conversely, failing to detrend risks conflating these regular patterns with genuine anomalies. For this project, we prioritize STL decomposition to enhance flare detection sensitivity by filtering out smooth, periodic fluctuations.

3.3 Anomaly Detection Model

The Isolation Forest (IF) is an unsupervised, tree-based anomaly detection algorithm designed to identify unusual observations within a dataset. In this case, it operates by splitting points to construct a forest of isolation (decision) trees. The underlying principle is that anomalies, being rare and distinct, are easier to isolate and thus tend to require fewer splits compared to normal observations. Anomaly scores are assigned based on the average path length (tree height) from the root node to each observation across all trees. Observations with shorter paths are flagged as anomalies. This makes IF well-suited for detecting sharp flux spikes associated with stellar flares, as these are expected to deviate significantly from the regular signal. The method is also computationally efficient and naturally scalable to large datasets.

Hyperparameters:

- `n_estimators`: Number of trees in the forest (default: 100).
- `max_samples`: Number of samples used to train each tree (default: 256).
- `contamination`: Expected proportion of anomalies in the data. Requires domain knowledge and directly impacts thresholding.
- `max_features`: Number of features used per split. Not applicable for univariate time series.

For flare detection, `contamination` is the most influential parameter. In Figure 3, the model was applied to the full time series (after imputation via ARIMA with Fourier terms), with `contamination` = 0.001 and all other parameters set to defaults. These results highlight the flexibility of Isolation Forests in handling irregular patterns and inspire opportunities for post-processing, such as mapping flare boundaries with adjacent low anomaly scores. However, without labelled flare data, model evaluation remains challenging. While initial results appear promising, further validation is necessary to confirm flare recovery performance.

3.4 Flare Simulation

Since this is an unsupervised learning setting in which flares are unlabelled, simulation and baseline model comparison are the most straightforward means of model evaluation. Following the methodology of [Esq+24], we inject simulated flares, based on templates from [Dav+14], into the real data, then assess detection performance using sensitivity, PPV, and F1-score classification metrics across repeated trials.

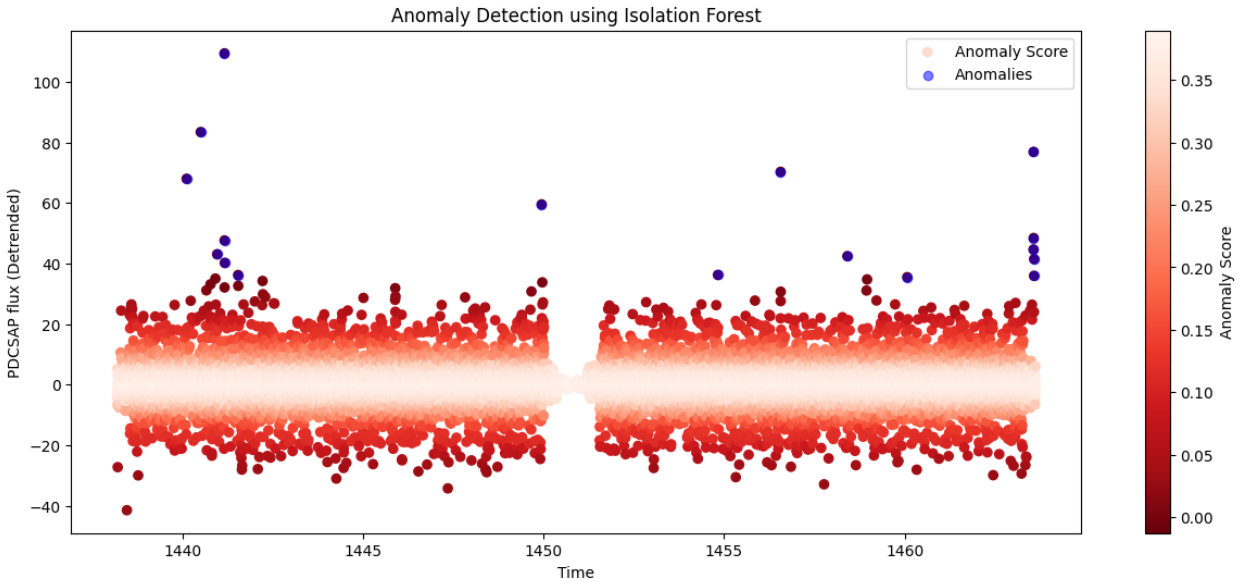


Figure 3: Scatterplot of detrended PDCSAP flux over time, coloured with red gradient by anomaly scores with anomalies highlighted in blue.

The flare model from [Dav+14] separates each flare into a rapid rise followed by a slower exponential decay. The flare duration is controlled via the parameter `t_half`, which defines the time it takes for the flare to decay to half of its peak flux. Like [Esq+24], we model flare amplitudes using a Pareto distribution.

Key simulation parameters include:

- `t_half` = 5e-5 days $\approx$ 4.32 seconds (converted to 2-minute indices).
- `xm` = mean flux $\times$ 0.01 $\equiv$ scale paraemeter affecting the likely minimum amplitude,
- `alpha` = 0.65 $\equiv$ shape parameter controlling skew,
- `offset` = 0 $\equiv$ shift,
- `upper` = mean flux $\times$ 0.15 $\equiv$ relative maximum amplitude.

These settings were chosen to capture a range of flare intensities. The left panel of Figure 4 demonstrates a 5-flare injection with these parameters, while the right panel zooms in on the second flare to show its shape.

3.5 Evaluation

To assess the effectiveness of the two-step approach, detrending via STL followed by IF, we conduct repeated injection-recovery experiments across varying flare sizes. Evaluation begins with a **hyperparameter tuning** phase for IF via grid search, followed by 100 simulations to compare our model against a baseline $3-3\sigma$-clipping method using standard classification metrics.

Exploring model performance across flare magnitudes provides insight into the detection limits of the model. To simulate different flare sizes, we vary flare amplitudes in the ranges of 1-5%, 5-25%, and 1-25% of quiescent flux (using mean flux) to simulate "small", "large", and "varied" flares respectively. While there do not seem to be standardized guidelines, some related studies ([Kow+10], [YLG+17], [MKS+20]) utilize amplitude relative to quiescent flux to roughly categorize flare sizes.

As this pipeline targets individual flare detection, recovery is measured at the **flare event level**. If any point within a flare's span is flagged as anomalous, it is counted as a successful detection. This approach does not distinguish between single and composite (overlapping) flares, which may bias the model toward

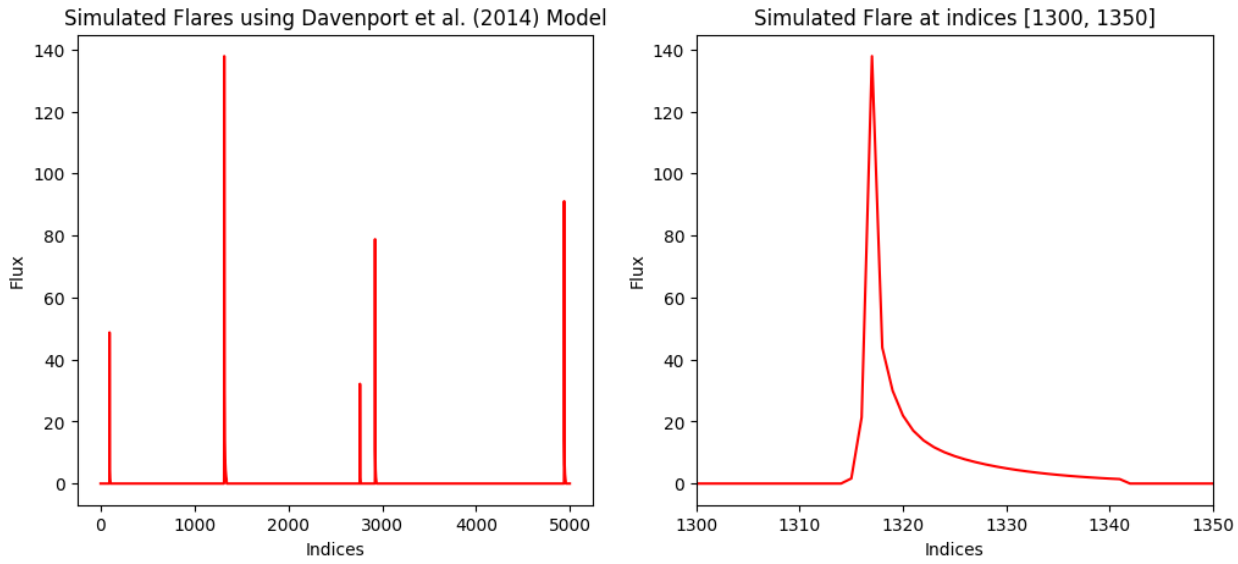


Figure 4: Left: Simulation with 5 flares of varying sizes. Right: Granular view of the second flare.

large-amplitude tuning. To limit composite flare occurrence, we inject only five flares into a relatively quiet 7-day segment (Days 1442–1449). Figure 5 shows the results of applying STL (period = 240) and IF with default parameters (as in Figure 3) to the 5-flare simulation from Figure 4. Three of the five injected flares are recovered with no false positives, resulting in a precision of 1.00, recall of 0.60, and F1-score of 0.75. These preliminary results suggest that the method is highly conservative, successfully avoiding false positives but missing lower-intensity events. Future tuning efforts may improve recall while maintaining strong precision.

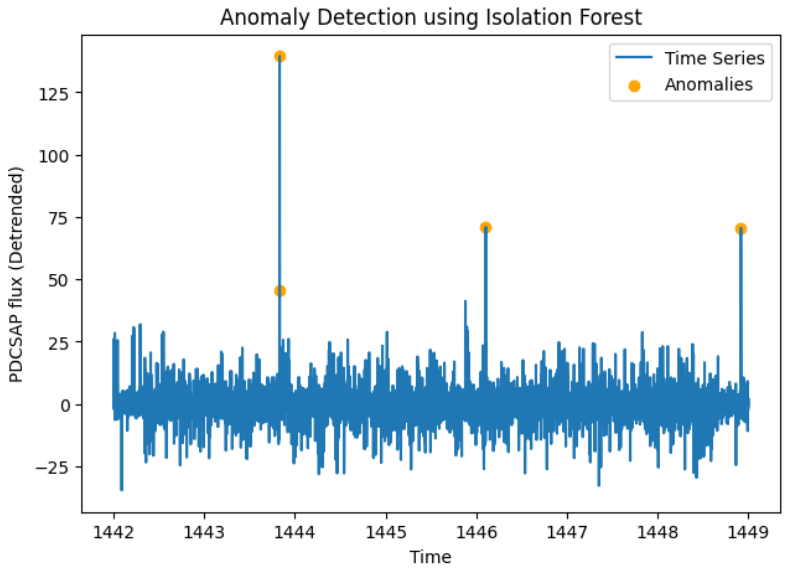
4 Results

To evaluate model performance, synthetic flares were injected into a real PDCSAP light curve using a parameterized Kepler flare model. The light curve was first detrended using STL before applying both the Isolation Forest and the standard $3-3\sigma$-clipping method to enable a fair comparison. By varying the intensity of simulated flares, we assessed the models' respective strengths and limitations. Performance was measured using Precision, Recall, and F1 Score classification metrics, calculated at the flare event level.

Metric	**Large**		**Varied**		**Small**	
	Isolation Forest	$3-3\sigma$	Isolation Forest	$3-3\sigma$	Isolation Forest	$3-3\sigma$
Precision	1.000	0.998	0.755	0.850	0.424	0.150
Recall	0.838	0.890	0.792	0.288	0.376	0.030
F1 Score	0.893	0.934	0.718	0.413	0.394	0.050

Table 1: Simulation results in terms of Precision, Recall, and F1 Score for Isolation Forest and 3–3σ-clipping across flare sizes.

As shown in Table 1, the Isolation Forest outperforms the $3-3\sigma$ clipping method in most settings, particularly for small and varied flares. For large flares, both methods achieve near-perfect detection, with F1 Scores exceeding 0.89. However, in the varied flare setting, Isolation Forest demonstrates greater balance between precision (0.755) and recall (0.792), while the $3-3\sigma$ method suffers from poor recall (0.288), resulting in an F1 Score of 0.413. The contrast is most striking for small flares, where the IF maintains moderate sensitivity (F1 = 0.394), while $3-3\sigma$ fails to recover most events (F1 = 0.050). These results highlight the robustness of Isolation Forests in detecting subtle and heterogeneous flares relative to traditional thresholding methods when searching for low-amplitude events.



Detection Performance

Metric	Value
Precision	1.00
Recall	0.60
F1 Score	0.75

Figure 5: Left: Anomaly scores over time. Smaller values indicate higher chance of being anomalous. Right: Scatterplot of time series over time, coloured by anomaly scores with anomalies highlighted in blue.

5 Conclusion

This project explored the application of Isolation Forests for detecting stellar flares in brightness time series data from the TESS mission. By detrending the flux data with STL decomposition and evaluating performance via simulated flare injections, we demonstrated that Isolation Forests offer a competitive and flexible alternative to a traditional $3 - 3\sigma$-clipping method that maintains computational efficiency. While both approaches performed well on large flares, Isolation Forests significantly outperformed baseline methods on small and varied flares, highlighting their robustness in detecting subtle and heterogeneous events.

Several limitations emerged during the project. An early effort to impute the substantial number of missing flux values using ARIMA-based forecasting models ultimately proved ineffective, as these methods introduced unreliable structure or required auxiliary interpolation strategies. As a result, missing values were excluded from the final modelling pipeline to preserve data integrity. Additionally, while STL decomposition improves flare isolation by removing smooth periodic variation, it can also suppress low-amplitude flares, making them harder to detect. Lastly, because real flare labels are unavailable, evaluation relied entirely on synthetic flare injections, which may not fully represent the complexity of astrophysical phenomena.

In conclusion, Isolation Forests represent a promising direction for unsupervised flare detection, particularly when paired with appropriate preprocessing. Their simplicity, scalability, and performance across flare types suggest that they could form the basis of more sophisticated detection pipelines.

5.1 Explorations and Future Work

Future work could focus on improving the **detrending process**, as STL may suppress low-amplitude flares. Although not detailed in this report, alternative methods showed potential and merit further exploration. **Feature engineering** may enhance detection by incorporating additional components, such as trends, seasonal signals, rolling statistics, and derivatives, to reduce reliance on residuals alone. Other preprocessing ideas include applying **data transformations** that emphasize positive outliers, such as flares. Finally, the **anomaly scores** generated by the Isolation Forest could be used more flexibly for soft labelling or to guide flare boundary mapping.

5.2 Reflections

This project strengthened my skills in Python and time series analysis, particularly in working with real, incomplete data. Although my initial efforts to impute missing values using ARIMA were ultimately time-consuming and set aside, they offered valuable insight. I also learned the importance of clear planning and evaluation when working in an unsupervised setting. Overall, I now feel more confident tackling applied time series problems and exploring more advanced modelling techniques.

References

[Dav+14] James. R. A. Davenport et al. "KEPLER FLARES. II. THE TEMPORAL MORPHOLOGY OF WHITE-LIGHT FLARES ON GJ 1243". In: *The Astrophysical Journal* 797 (2014). URL: https://api.semanticscholar.org/CorpusID:119298940.

[Esq+24] J. Arturo Esquivel et al. "Detecting Stellar Flares in Photometric Data Using Hidden Markov Models". In: *The Astrophysical Journal* (2024). URL: https://api.semanticscholar.org/CorpusID:269293497.

[Kow+10] A. F. Kowalski et al. "Time-resolved Properties of a Large UV Ceti Flare Observed with the GEB and Kepler". In: *The Astrophysical Journal Letters* 714.1 (2010), pp. L98–L102. DOI: 10.1088/2041-8205/714/1/L98.

[MKS+20] A. Medina, D. Kipping, T. E. Sartori, et al. "Flares in the Context of M-dwarf Planets: The Virtual Planetary Laboratory". In: *The Astrophysical Journal* 905 (2020), p. 107. DOI: 10.3847/1538-4357/abc9bc.

[YLG+17] H. Yang, M. C. Liu, Q. Gao, et al. "Superflares on Late-type Dwarfs from the First Year Observations of TESS". In: *The Astrophysical Journal* 849 (2017), p. 36. DOI: 10.3847/1538-4357/aa8d1f.