Detection and Characterization of Stellar Flares

Todd Eichel, Andrew Hong, Fen-Ly Lo, Guillaume Pignol
Advisors: Peter Freeman, Brian Junker, Daniel Manrique

Abstract

Stellar flares are massive explosions that occur in a star's atmosphere that emit electromagnetic radiation across the entire electromagnetic spectrum. Because of this radiation's potential impact on human activities on Earth, we are interested in detecting stellar flares and being able to describe their properties. The aim of this project was to design a statistical algorithm that automatically detects flares from a given set of data and estimates their significance as well as their starting and ending times. We used two avenues of analysis, exponential fitting and change point analysis, and merged them together to produce an algorithm which was effective on our dataset. This algorithm was also tested by simulating stellar flares and shown to be sensitive to different flare magnitudes and able to estimate the starting and ending times in a biased, yet compact and precise manner.

Introduction

Motivation

We studying stellar flares in order to better understand the flaring behavior of our own star: the sun. Flares on the sun strongly influence our activities here on Earth (disrupting electronics and radio communications), so we have incentive to investigate flaring behavior.

Data Source

For this project, we looked at the flare star AD Leo. AD Leo is a Red Dwarf star of class M, meaning that its surface temperature is relatively cool, and that it is relatively small (less than half the size of our sun). These characteristics make AD Leo an excellent choice for the observation of flares, since we can observe luminosity changes more closely.

Data

Collection

To observe AD Leo, we employ satellites with instruments capable of detecting single photons as they strike a detector. Our datasets come from NASA's Extreme Ultraviolet Explorer satellite in the form of lists of photon arrival times.

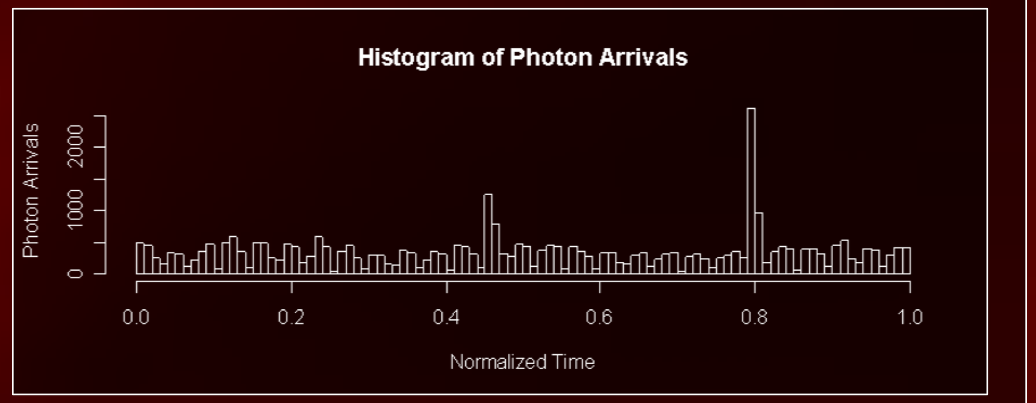
Our particular dataset consists of 34,444 separate photon arrival times, covering a span in real time of about five days.

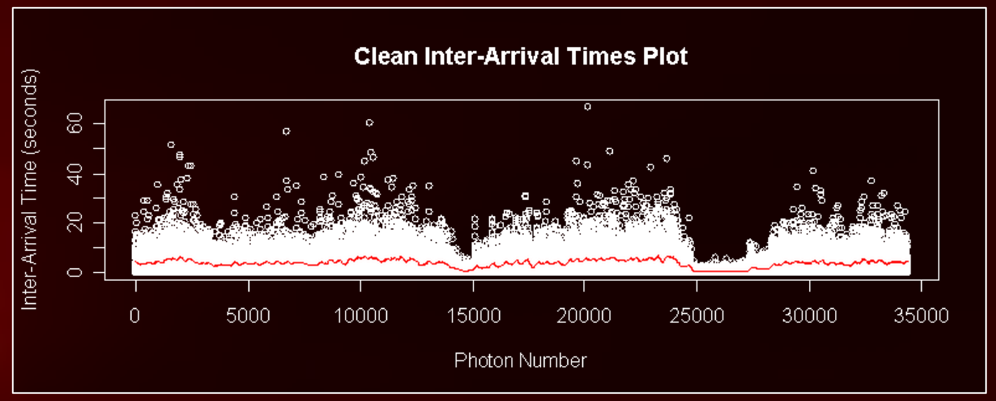
Visualizations

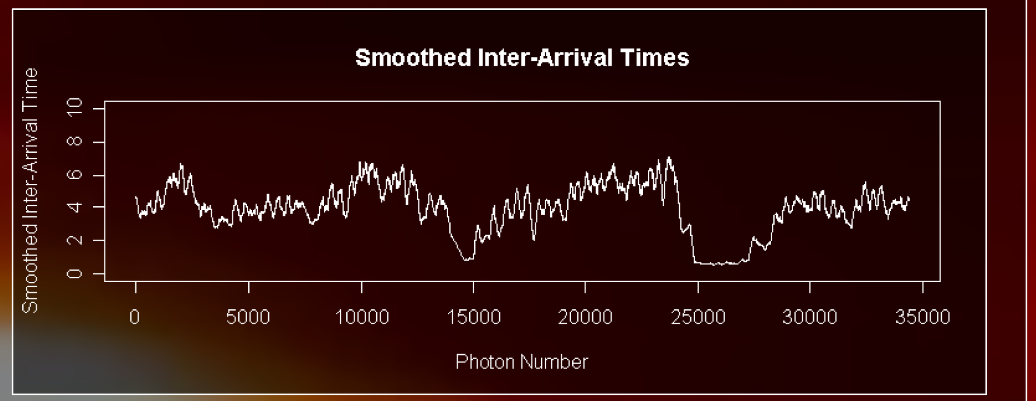
In a histogram of these photon arrivals, we can see two distinct spikes.

A more useful dataset for our project is the time between photon arrivals: inter-arrival times. In a plot of these, the spikes show up as depressions.

We then smooth these data to see the trends more clearly.



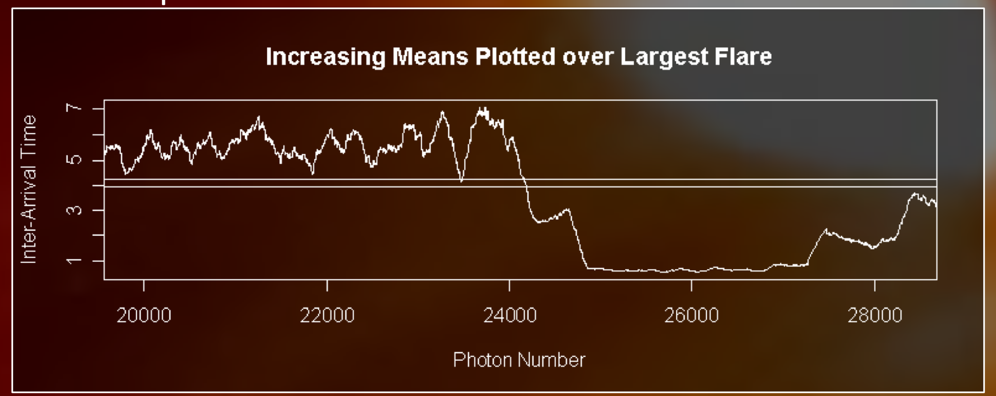




Method: Exponential

Basically, the algorithm compares the photon inter-arrivals of the flare events to the non-flare areas in the data. In order to make such a comparison, the non-flare arrival rates of photons must first be characterized.

In order to estimate the exponential parameter of the distribution of quiescent photon inter-arrival times, the influence of outliers (the flare photon inter-arrivals) must be removed. First, the parameter of interest is calculated and based on it, the 20% outliers are omitted and then the parameter is re-calculated. The process is done reiteratively until the estimator converges within a specified level of precision.

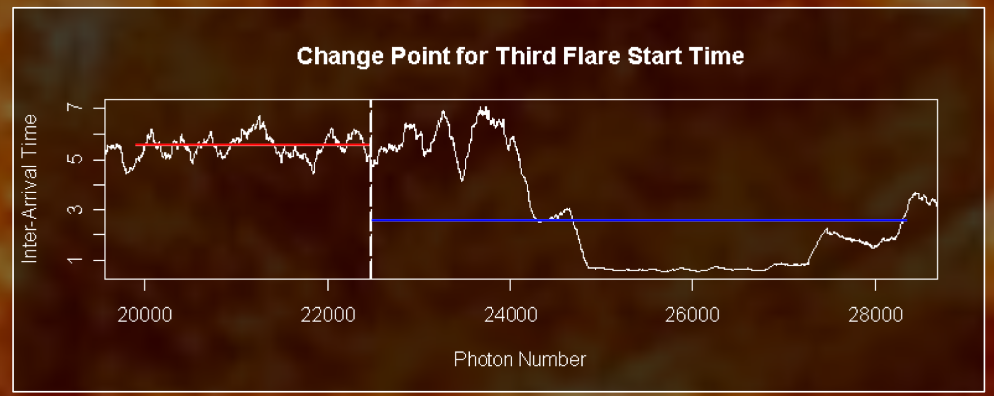


Once a stable estimator is attained, the algorithm uses this mean as a point of reference and obtains the longest continuous intervals (in photon number) where inter-arrival times fall below the mean.

The algorithm calls the change point function on the neighborhood around the first point of the flare partition then another neighborhood the last point of the flare partition. Change point is then run over the neighborhoods to find the exact start and end times of the flare.

Method: Change Point

Change point as a method attempts to detect local moments of change in data distribution. The change point function is distance based and essentially finds the point in the data, where fitting different lines to each side of the division, minimizes the most error. It is appropriate for this particular problem, because in a sense a flaring event is a sharp deviation from the quiescent level preceding it, resulting in one change point and a return following the end of the flare to that quiescent level, resulting in another change.

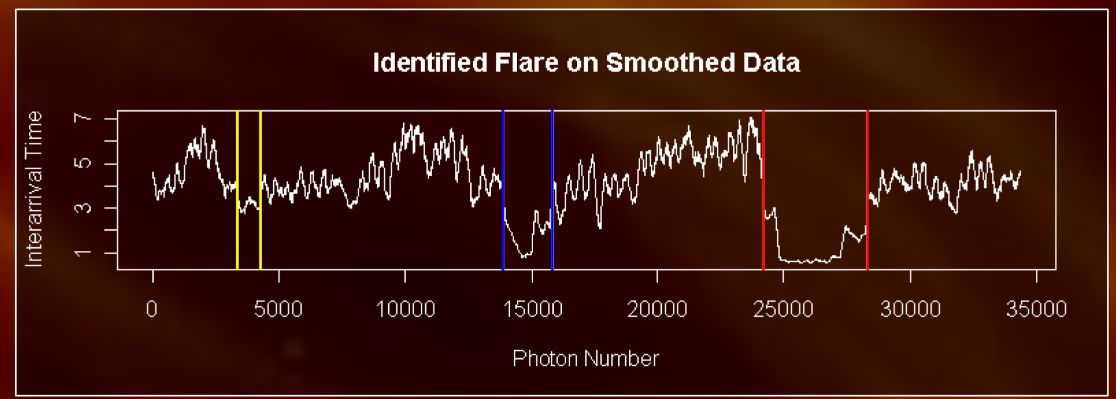


The flare's estimated start and end time mark the boundaries of a particular flare. An exponential model is fit over all the data in the non-flare intervals or observations outside of these newly established boundaries, and the Kolmogorov-Smirnov Goodness of Fit test is used to calculate the probability that the inter-arrival times of a particular flare were generated by the fitted exponential distribution to attribute a p-value to each flare.

Results

Though the method has its limitations, it overcomes a variety of problems in creating an automated method for flare detection. For the data of AD Leo, the identified flares synchronize with the visual expectation of flare start and end times.

Flare	Start Index	End Index	P-Value
1	3367	4314	3.301×10^{-3}
2	13883	15836	7.838×10^{-10}
3	24198	28314	0



Simulation and Analysis

Method

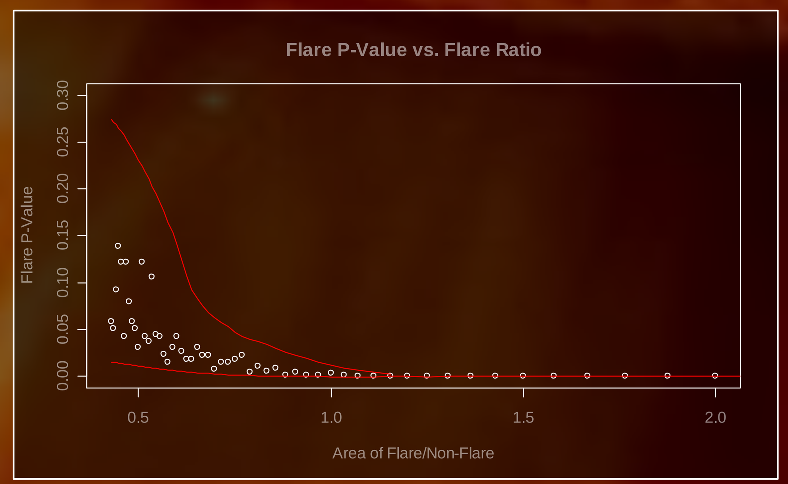
A random number generator is used to build artificial flares by manually defining the shape of the flares. The random number approximate the defined shape and introduce variability in the simulated data. Using the generated datasets, two simulation tests were conducted, one to measure how sensitive the algorithm is by calculating flare significance, and the other by looking at estimated start and end times of flares.

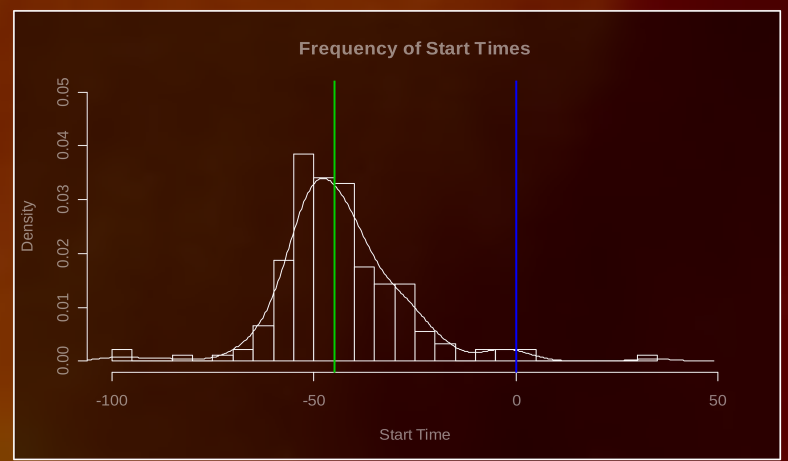
Test 1

For the first test, the ratio of the area of the simulated flare versus the non-flare area under it was changed to range from 0.42 to 2, and for each 55 levels of ratio, 50 flares were created and measured, yielding a total of 2750 flare measurements. The scatterplot shows the median of the 50 p-values for each value of the ratio. The red lines are smoothed representation of the 25% and 75% quartiles of the p-values.

Test 2

The second test involved creating 200 artificial flares and running our algorism to find start and end times. Since these actual values are set in advance when creating flares, we are able to see how much the algorithm's estimate differ from the real values. The histogram and the superposed density curve show the distribution of the difference between the estimated start times and their associated true values.





Analysis

The results show that the 90% sensitivity level of the algorithm is about a flare/non-flare area ratio of approximately 1. The simulation also showed that the estimators for the start and the end of the flares are biased by about 40 photons with a significant flare in a data set of 5000 photons. The distribution of the start and end times is fairly compact with a 95% range of around 40 photons.

Conclusion

By examining inter-arrival times and taking short inter-arrivals as indicative of flaring events, the algorithm runs an exponential-based parametric fitting and non-parametric change point analysis in series to capture the exact boundaries of a flare. The process of partitioning the data into local subsets allowed the algorithm to overcome the sensitivity of the change point function and the overall variability of the data, and the change point function in return complimented the exponential fitting by estimating accurate local bounds on the flare. The flares found in the data were sustained depressions in flare inter-arrival times, and though there were three obvious flares found, the technique begs the question of whether other, less significant flares were present in the data. Given the unstable nature of the star's behavior and the variety of flare morphologies, it is a question unanswerable by statistics. However, using statistics to examine the residue of flare emissions (photons) to give insights into the activities of the actual star is astonishing.