

u^b

b
**UNIVERSITÄT
BERN**

Finding Superflares on Stars using Kepler Data

Bachelor Thesis
Faculty of Science
University of Bern

David Jakob Schwarz
2024

Supervised by
Prof. Dr. Lucia Kleint, Pranjali Sharma, Janis Witmer
Astronomical Institute

Declaration of Originality

according to Art.30 RSL Phil.-nat.18

Title of the thesis

Finding superflares on stars using Kepler data

Author

David Jakob Schwarz, 19-936-947

Study program

Physics 120 ECTS, Astronomy 30 ECTS, Mathematics 30 ECTS

Supervisor

Prof. Dr. Lucia Kleint

I hereby declare that I have written this thesis independently and have not used any sources other than those stated. I have marked all passages that have been taken verbatim or in spirit from other sources. I am aware that otherwise the Senate pursuant to Article 36, Paragraph 1, Letter r of the Law of September 5, 1996 on the University is authorised to withdraw the title awarded on the basis of this thesis.

For the purpose of reviewing and verifying compliance with the declaration of independence or the regulations concerning plagiarism, I grant the University of Bern the right, to process the personal data required for this purpose and to carry out acts of utilisation, in particular to reproduce the written work and to store it permanently in a database and to use it for checking the work of third parties or to make it available for this purpose.

Bern, 08.09.2024

Signature: 

Abstract

Context: In a catalogue published by Okamoto et al. (2021) 2341 superflares on 265 solar-type stars have been identified. A recent study by Probst (2023) found additional flares in the same dataset that have not been reported yet.

Aims: The objective is to comprehend how flares can be detected using automated flare detection algorithms and improve Okamoto's catalogue. We also want to study flare properties and compare our findings to actual studies.

Methods: We used and modified two flare finding algorithms; FLATW'RM by Vida and Roettenbacher (2018) and AFD by Althukair and Tsiklauri (2023). For further analysis, multiple scripts examined flare duration, energy, and frequency of flare events for different wait times and stellar rotational periods.

Results: Up to 7336 new flares have been found, with FLATW'RM excelling at finding short low-energy flares and AFD having almost zero false positives. Difficulties were found in determining the flare duration and the flare energy calculation. An exponential decay in flare frequency has been found for increasing stellar rotational periods, whereas the flare energy was found to be independent of the wait time between consecutive flares.

Contents

1	Introduction	1
1.1	Flares on solar-type Stars	2
1.2	Kepler Mission	2
1.3	Identifying Superflares	3
1.4	Goals of this Thesis	3
2	Methods	4
2.1	Dataset	4
2.2	Flare-finding Algorithms	5
2.2.1	FLATW'RM	6
2.2.2	Modifications to FLATW'RM	9
2.2.3	AFD	13
2.2.4	Modifications to AFD	16
2.3	Flare Energy	16
3	Results	17
3.1	Flares found with Algorithms	17
3.2	Quality of the Flares found with different Algorithms	18
3.3	Properties of Flares	19
3.4	Wait Time, Flare Duration and Energy	21
3.5	Stellar rotational Period	25
3.5.1	Rotational Period and Wait Time	25
3.5.2	Rotational Period and Energy	28
4	Discussion	34
4.1	FLATW'RM	34
4.1.1	Implementation and Usage	34
4.1.2	Detected Flares	34
4.1.3	Flare Energy, Wait Time and Duration	35
4.2	AFD	37
4.2.1	Implementation and Usage	37
4.2.2	Detected Flares	37
4.2.3	Flare Energy, Wait Time and Duration	37
4.3	Dependence on rotational Period	38
5	Conclusion and Outlook	40
References		42
A	Flare Energy Calculation	46
A.1	Kepler Respond Function	46
A.2	Temperature and Flare Area Dependence	46

A.3	Flare Energy	47
A.4	Flare Temperature	48
A.5	Changes to Althukair's formulae	49
B	Note on Comparing Flares found with different surveys	51
C	Additional Plots	51
C.1	Wait Time	51
C.2	Flare Duration	57

1 Introduction

The solar atmosphere is highly dynamic, manifesting in the emergence of magnetic flux, sunspots and solar flares. The most prominent feature are sunspots, which Chinese astronomers observed as far back as 314 BC (Rogers, 2011). They form where concentrated magnetic field lines rise from the interior. The more powerful magnetic field causes a higher magnetic pressure in the area where the sunspot forms. This leads to a lower temperature compared to their surroundings, making them appear as dark spots. The formation of sunspots is directly related to the formation of active regions (Solanki, 2003), as can be seen in figure 1, where an eruption occurs over a group of sunspots.

Solar and stellar flares are powerful bursts of energy caused by magnetic field line reconnection near sunspots or starspots (Shibayama et al., 2013). They release radiation in a wide range of the electromagnetic spectrum, from X-rays to ultraviolet light (Davenport, 2016). The impact of solar flares can lead to disruptions in satellite communications, navigation and electric power grid failures. Especially bigger events, like the Carrington event in 1859, would have a global impact, but there is still no model to predict these events (National Research Council, 2008).

Collecting more information on the properties and distribution of flares on solar-like stars might help researchers develop such models, which would not only allow more people to predict auroras but also could prevent billions of dollars in infrastructure damage.

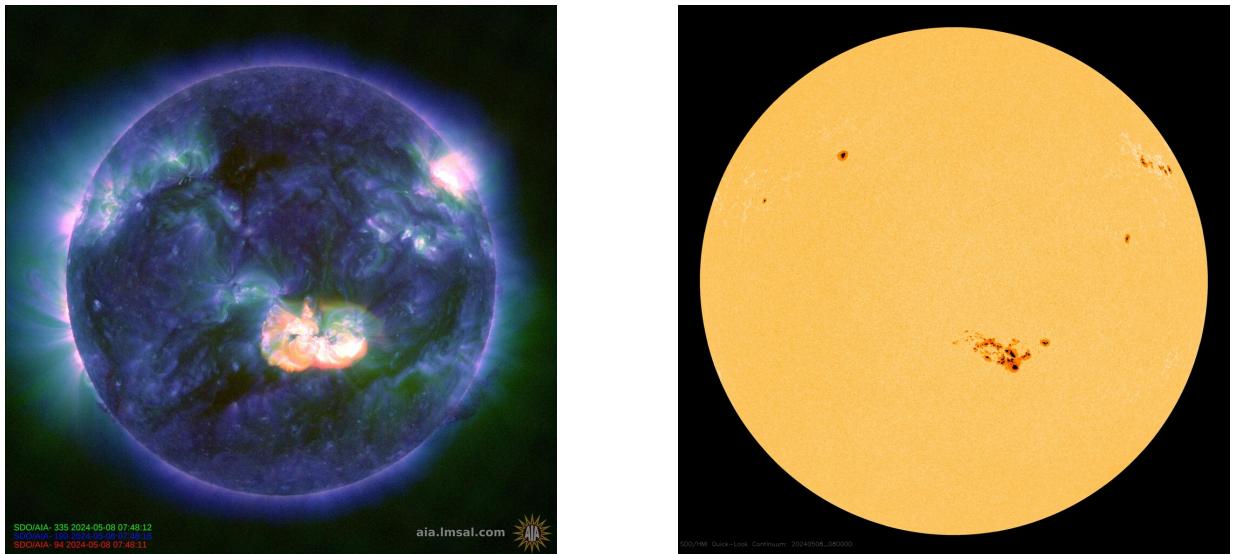


Figure 1: Image of the Sun captured by NASA’s Solar Dynamics Observatory space telescope on 08.05.2024. There is a composite image with wavelengths of 94, 193 and 335 Å (left). The right image shows sunspots. The eruption in the middle of the left image can be correlated with a strong space weather event on Earth, resulting in the aurora borealis being visible in the night sky over Switzerland on May 11. (Images by Courtesy of NASA/SDO and the AIA and HMI science teams, SDO, 2024.)

1.1 Flares on solar-type Stars

Solar flares can last a few minutes to several hours. During these events, the Sun’s local atmosphere can heat up to several million degrees Kelvin, significantly higher than its usual effective temperature of approximately 5'778 K.

Flares can occur on many different stellar types but are generally more common in low-mass stars (Davenport, 2016; Walkowicz et al., 2011). They can produce “Superflares” with energies between 10^{33} and 10^{38} ergs, whereas solar flares typically have energies between 10^{29} and 10^{32} ergs.

The rotation period of a star plays a crucial role in stellar flare activity. The Sun has a rotational period of about 25 days at its equator. In the broader context of solar-like stars ($T_{\text{eff}} = 5100 - 6000$ K), there are many stars whose rotation period is shorter (as short as a few days), where rapid rotation generates stronger magnetic fields through the dynamo effect (Noyes et al., 1984; Skumanich, 1972), which can lead to more frequent and intense flares (Okamoto et al., 2021).

One of the goals of this thesis is to investigate the relationship between rotational period and flare activity in more detail.

1.2 Kepler Mission

The Kepler telescope was launched on March 7 2009 and began operation on May 13. The primary mission lasted until August 15 2013. The mission’s primary goal was to determine how common Earth-like planets are within the Milky Way galaxy. For this, the spacecraft was equipped with a 420-900 nm photometer (J. Van Cleve, 2009), which continuously monitored the brightness of over 150,000 stars in a fixed field of view. The telescope had to be rotated every 90 days to keep the solar panels oriented to the sun, which divided the Kepler dataset in quarters of 90 days (Koch et al., 2010; J. E. Van Cleve and Caldwell, 2016).

Even though the main focus was surveying the brightness of stars to detect dimming caused by transits of exoplanets, the collected light curves also allow us to find flares by looking for a sudden increase in brightness.

Kepler recorded data at two different rates:

1. **Short Cadence Data (1-minute intervals):** This high-frequency sampling was crucial for studying stars with rapid variability and detecting smaller, Earth-sized planets that produce brief transits.
2. **Long Cadence Data (30-minute intervals):** This lower-frequency sampling was used for most stars in Kepler’s field of view. It provided sufficient resolution to detect the periodic dimming caused by transiting planets while enabling the monitoring of a vast number of stars over extended periods.

After a failure on the second of the spacecraft’s four reaction wheels in August 2013, the necessary alignment precision could no longer be guaranteed, and the primary mission was concluded. In May 2014, data collection continued with the K2 mission (Second Light),

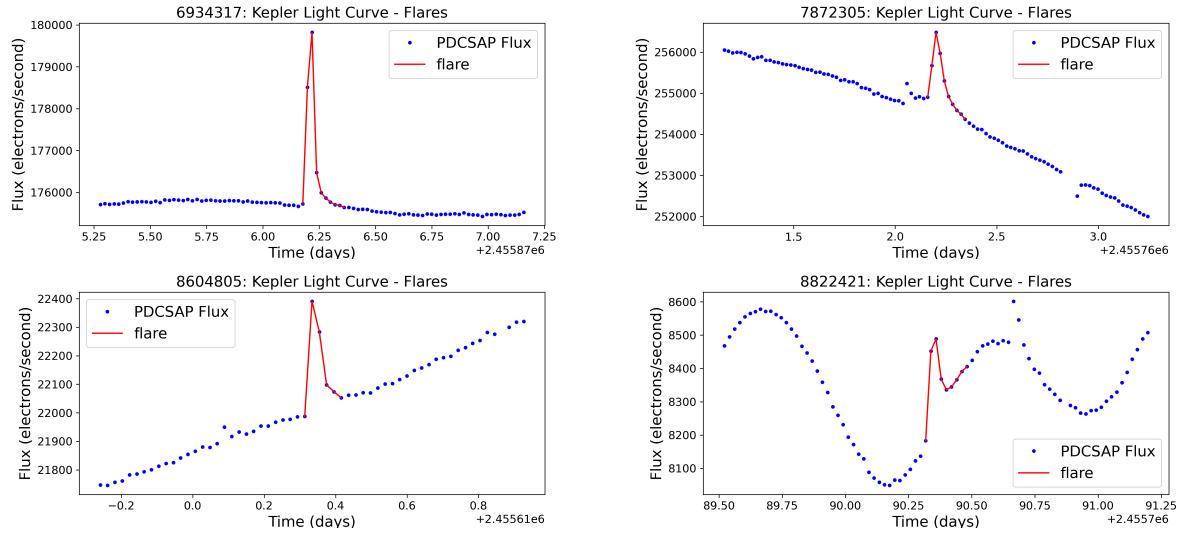


Figure 2: Four light curves with very distinct superflares from Okamoto’s catalogue (Okamoto et al., 2021). The vertical axis shows the flux measured by the telescope in units of electrons per second. The horizontal axis shows the time in the Julian date format (JD). The plot shows long cadence data with roughly 30 minutes between each point.

using the two remaining reaction wheels to investigate smaller and dimmer red dwarf stars (NASA, 2018).

1.3 Identifying Superflares

To distinguish between superflares and data noise or equipment malfunctions, the light curve must be examined. A sudden rise in the flux followed by an exponential decline over several data points indicates a superflare. A more gradual increase or a broad peak is most likely caused by a different process (Maehara et al., 2012). A selection of some very distinguishable superflares can be seen in figure 2.

However, to make this distinction, enough flare points are needed. For long cadence data, this makes it difficult to detect short flares, as the initial increase in brightness is too short to be measured and only the decreasing phase can be seen. Depending on the flare’s energy and the presence of noise, these flares might be impossible to detect. If the flux undergoes a sudden increase in just one or two data points, it could also be that other factors are causing it and thus cannot be categorised as a superflare (Rivera et al., 2023).

But even with more points, some flares might look different. They can have strong fluctuations after the flare peak or even multiple peaks (Okamoto et al., 2021).

1.4 Goals of this Thesis

This thesis focused on two flare-finding algorithms - namely FLATW’RM by Vida and Roettenbacher (2018) and AFD by Althukair and Tsiklauri (2023) - and looked at how

good they are in detecting flares on the same dataset. Both algorithms are described in section 2.2. Interesting questions were:

- Do the algorithms detect flares consistently?
- Do the algorithms detect the same flares? And if not, what are the differences?
- Can the algorithms be improved?

Several scientific questions could be investigated once the algorithms had created new flare catalogues:

- What are the wait times between two consecutive flares? What is the distribution? Are the wait times different for certain stars?
- Is the length of the flares systematic or unusual?
- How can the flare energy be estimated? What is the distribution? Is the energy different for certain stars?

For the dataset, it was chosen to look at stars from Okamoto et al. (2021) because they are solar-like stars known to be flaring. It has been suggested by Probst (2023) that there are more flares in this dataset than those found by Okamoto. Could the flare-finding algorithms detect them?

Okamoto et al. (2021) not only created a catalogue but also studied the properties of those flares. They observed that the upper limit of the flare energy decreased as the rotation period increased. Also, the frequency of superflares decreased as the stellar rotation period increased, which has already been described by Candelaresi et al. (2014).

Balona (2015) found similar results when looking at a range of K–M dwarfs to A–F stars, writing that flaring stars rotate significantly faster than non-flare stars. He also found a strong correlation between flare energies, stellar luminosity, and radius.

The distribution of wait times for Kepler short-cadence M dwarf observations has been studied by Hawley et al. (2014). They have found that an exponential distribution fits the data best for wait times well above 30 minutes. Flares at smaller wait times were underrepresented, likely because many were counted as part of complex events.

2 Methods

2.1 Dataset

A catalogue of superflares made by Okamoto et al. (2021) was used to compare the flares found with different flare-finding algorithms. They selected solar-type stars based on different criteria: They only looked at Kepler 30-minute long cadence data taken from May 2009 to May 2013 that had effective temperatures T_{eff} and stellar radius R_{star} included in the catalogue of Berger et al. (2018). The stars had to be G-type main sequence stars ($T_{\text{eff}} = 5100\text{--}6000$ K), and they had to have their rotational period P_{rot} and amplitude values

reported in McQuillan et al. (2014). They also found one binary candidate, which was included in their list but excluded from this analysis.

In total Okamoto et al. (2021) found 2341 flares on 265 solar-type stars. To detect flare they first calculated the brightness variation between all consecutive data points. The resulting distribution could be used to detect flares by setting a flare threshold of three times the value, where the area is equal to 1% of the distribution. To be classified as a flare two consecutive data points must exceed the threshold and the declining phase must be longer than the increasing phase.

In the selected time frame, there were 4334 long-cadence light curve files belonging to those stars, averaging 16.35 light curve files per star. All light curve files were downloaded with the tools provided by the Mikulski Archive for Space Telescopes (MAST) (Brown et al., 2011; Bunnell, 2018). Kepler light curves contain two different flux types, the Simple Aperture Photometry flux (SAP) and the Pre-searched Conditioning SAP flux (PDCSAP), which has long-term trends removed. For this analysis, only the PDCSAP flux was used. All the codes used for flare finding and further analysis can be found on GitHub.

2.2 Flare-finding Algorithms

Flare-finding algorithms usually search for outliers in light curve data. There are two main problems to overcome:

First, they must find a model for the light curve. One common characteristic of almost all light curves is a quasi-periodic variation in brightness. This is caused by the rotation of a star with starspots as the brightness changes depending on the number and size of the starspots visible on the side facing the telescope. But also other factors could cause these variations, such as orbital motion in binary systems or eclipses in accompanying stars (Maehara et al., 2012).

To illustrate the diversity of light curves, four different light curves are shown in figure 3. The model must be able to adapt to various periods and shapes, while it should be flexible enough to adjust to short-period changes in amplitude and shape. However, high noise levels and gaps can make it difficult to get clean fits.

The second problem is choosing a suitable noise threshold, especially for the identification of weaker flares because of the presence of noise. The noise level is defined by the standard deviation σ of the fluctuations about the detrended light curve. It is common to limit flare identification by selecting a threshold value that significantly exceeds the noise level ($k\sigma$). However, the choice of the k -value is normally made based on an empirical rule (typically $k = 3$), which can lead to a biased threshold level (Rivera et al., 2023).

In other words, a high noise threshold leads low number of false positives, but this is at the cost of missing all of the weaker flares. On the other hand, a low noise threshold increases the number of detected flares but also the number of false positives. Choosing a noise threshold that maximises the information content is essential but also very difficult.

After the data is split into inlier and outlier points, flare events must be distinguished from

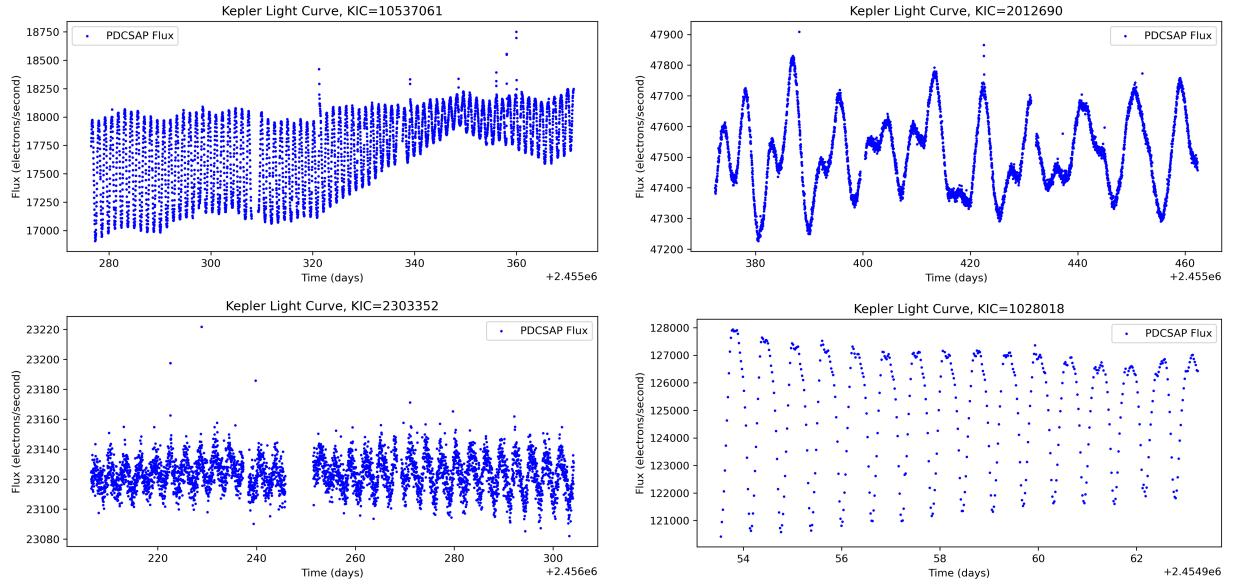


Figure 3: Four light curves for different stars with quasi-periodic variations. The shape of the light curve can be very different depending on the star’s rotational period and the number of starspots present. The vertical axis shows the flux measured by the telescope in units of electrons per second. The horizontal axis shows the time in the Julian date format (JD). The plot shows long cadence data with roughly 30 minutes between each point.

the general noise. This can be done by removing one-point outliers (which can be caused by transmission or instrument errors) and then comparing them to a typical flare shape as, for example, the single-peak model defined by Davenport et al. (2014). But in general, there seem to be different ways to do this. This is why it is interesting to check whether different flare-finding algorithms find the same flares.

For this thesis, the FLAre deTection With Ransac Method (FLATW’RM) algorithm developed by Vida and Roettenbacher (2018) and the Python script for Automated Flare Detection and flare analysis (AFD) developed by Althukair and Tsiklauri (2023) were tested. Both algorithms had to be modified to evaluate multiple light curves sequentially or without reconfiguration. The algorithms and their modifications are described in the following sections.

2.2.1 FLATW’RM

The idea behind FLATW’RM was to reduce the need for visual inspection necessary for traditional fitting and analysis methods by using machine-learning techniques to model light curves. For this, they use the RANSAC (RANdom SAmple Consensus) method, as it is capable of creating smooth fits even for data with a large number of gross outliers (Fischler and Bolles, 1981).

The FLATW’RM algorithm uses the following steps to find a flare inside a light curve:

1. Determine the rotation period P of the light curve.

This is done using `LombScargleFast`, which is an implementation of the Lomb-Scargle Periodogram in `gatspy` (J. T. VanderPlas and Ivezić, 2015; J. VanderPlas, 2015). The period found with this method should be the same as the star's rotational period.

2. For a window size of $1.5 \times P$, the light curve can be described as a low-degree polynomial.

To avoid overfitting, the order of the polynomial is determined with a k-fold cross-validation method. For this, the data is split into k smaller sets (called folds), and a machine learning model is trained on $k - 1$ folds of the data while the remaining part is used for validation. For most machine-learning estimators, both axes have to be normalised. However, since the scaling of the flux variation is no more than a multiplication of the polynomial coefficients, only the time axis needs to be normalised. This is done by subtracting the mean and scaling the light curve by its standard deviation.

3. To model the light curve, a RANSAC (RANdom SAmple Consensus) algorithm is used to find a robust model for the light curve in each window. The original data points that fit the model (consensus set) are considered inlier points.

In detail: A polynomial of order N , which is the model used in this case, requires a minimum of $n = N + 1$ data points. A random subset $S1$ with n data points is selected from the whole dataset P . The number of points in P must be greater or equal to n ($\#[P] \geq n$). The model $M1$ created with $S1$ is used to determine the subset $S1^*$ of points in P , which are within some error threshold of $M1$. The set $S1^*$ is called the consensus set of $S1$.

If $\#[S1^*]$ exceeds a threshold t , use $S1^*$ to compute a new model $M1^*$. If not randomly select a new subset $S2$ and repeat. If no set passes the threshold, the process is stopped after some predetermined number of iterations and the largest consensus set is used. (Fischler and Bolles, 1981)

4. Using the inlier points, the standard deviation σ of the light curve can be calculated.
5. All points above a given detection threshold (3σ by default) are declared first-order flare candidates and given a vote.

The outliers found by the RANSAC method were not reliable enough as they sometimes included the beginning of the light curve and points below it. Using the standard deviation guaranteed that only points of interest (potential flare candidates) were marked as outliers.

6. Shifting the search window through the light curve in given steps (by default, one-fourth of the light-curve window), every point is analysed multiple times.
7. If a light curve section receives enough votes (≥ 3 by default), it is considered a flare.

A light curve section consists of a number of consecutive points (by default, at least two points). Shifting through multiple windows ensures that every point (except close to edge

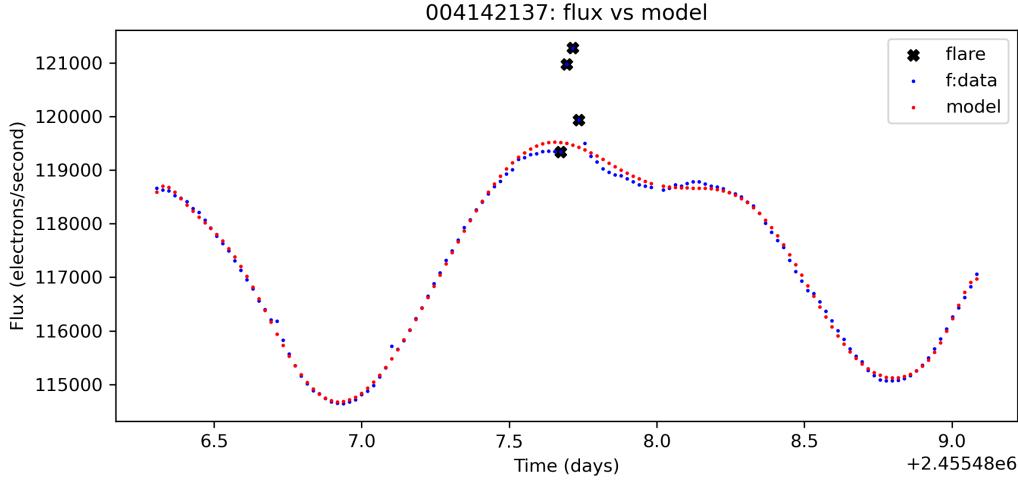


Figure 4: Example of a flare found with the FLATW'RM algorithm for KIC 4142137. The blue dots ($f:\text{data}$) are the PDCSAP flux, and the red dots (model) come from the model of the light curve created by the RANSAC method. The black crosses are the points the algorithm contributes to a flare event. The vertical axis shows the flux measured by the telescope in units of electrons per second. The horizontal axis shows the time in the Julian date format (JD). The plot shows long cadence data with roughly 30 minutes between each point.

points) is evaluated multiple times. As there might be some random outlier points in each window, the voting system ensures that only the points observed in multiple windows are considered flare candidates, making the results more robust.

8. An extra point is added in front of the first flare point for better energy estimation (but only if it is not too far away from the first point).

An example of a flare found by the FLATW'RM algorithm can be seen in figure 4, where it can be seen that an extra point was added in front of the first flare point (points below the model curve can not otherwise be declared as flare points).

There is also a further script in the algorithm that first removes the periodic changes by subtracting the model from the data points. Using the single-peak flare model defined by Davenport et al. (2014) it then tries to fit a model curve to the flare. This fit (see blue curve in figure 5) theoretically allows for a better approximation of the flare duration, as due to the exponential nature of the flare curve and detection threshold, late flare points are not detectable by the FLATW'RM algorithm. Also, it allows for a simple energy calculation as the flare's energy is proportional to the area under the curve. However, in practical terms, it is challenging to get a good fit as it depends on the initial values of the flare amplitude and the flare peak time.

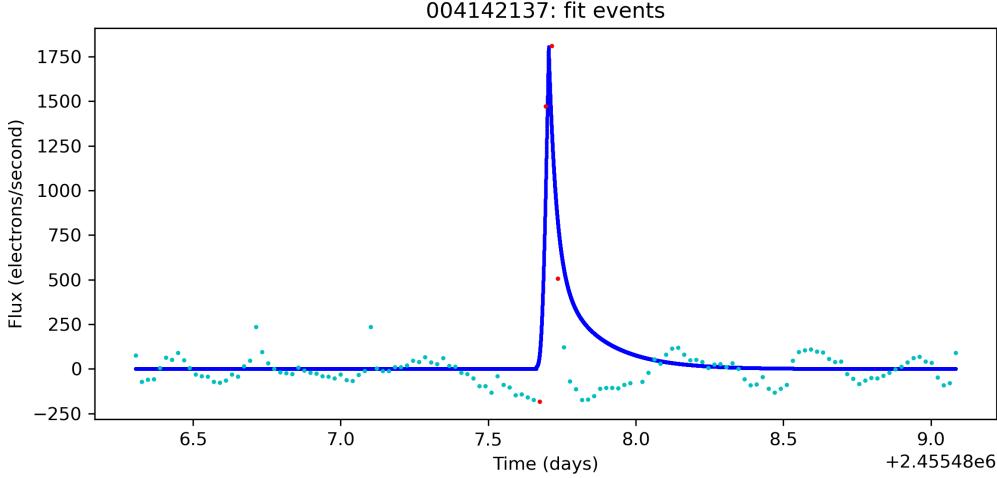


Figure 5: Fitted flare curve using a curve fit of the single-peak model defined by Davenport et al. (2014) on the data-model plot for KIC 4142137. The flare points are marked with red dots, and the single-peak model curve is blue. The vertical axis shows the flux measured by the telescope in units of electrons per second. The horizontal axis shows the time in the Julian date format (JD). The plot shows long cadence data with roughly 30 minutes between each point.

2.2.2 Modifications to FLATW'RM

One of the most notable caveats of the FLATW'RM algorithm is that it can not be applied to multiple light curves (especially if they are from different stars) without adjusting the initial parameters, i.e. the minimal flare points (fp) necessary to declare a flare and the detection threshold Σ . This is essential as some light curves have more intrinsic noise than others and, therefore, need higher thresholds.

For example, in all 17 light curve files for KIC 20303352, 200 flares were initially detected with the FLATW'RM algorithm when using $fp = 2$ and $\Sigma = 3\sigma$. However, only about 13 could be clearly identified as flares when checked by eye. The main problem for this object was, as shown in figure 6, that there was a lot of noise. The solution to this problem was to change the default values of fp and Σ according to the degree of the polynomial used for the model. The rules following rules were chosen by comparing the resulting flares by eye for several stars and with different setups:

1. If n is the degree of the polynomial, then the number of flare points fp is

$$fp = \begin{cases} 3 & \text{if } n < 8 \\ 2 & \text{otherwise} \end{cases}$$

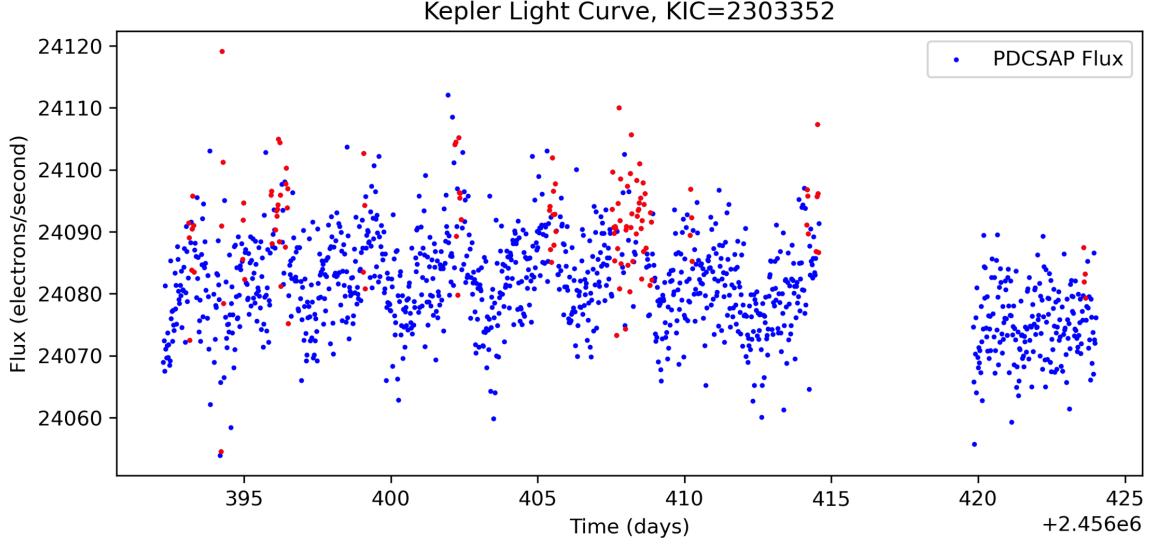


Figure 6: One light curve for KIC 20303352 where all the flares identified by the FLATW'RM algorithm are marked red. However, most of them do not have the exponential rise and slow decay shape characteristic of a flare. The vertical axis shows the flux measured by the telescope in units of electrons per second. The horizontal axis shows the time in the Julian date format (JD). The plot shows long cadence data with roughly 30 minutes between each point.

2. If n is the degree of the polynomial, then the detection threshold Σ is

$$\Sigma = \begin{cases} 5\sigma & \text{if } n < 5 \\ 4\sigma & \text{if } 4 < n < 9 \\ 3\sigma & \text{otherwise} \end{cases}$$

Another problem is the behaviour of the code when there are gaps in the light curve. Either the model failed at the start, end or somewhere in the middle, or some external effects influence the flux. Two examples can be seen in figure 7. Many different strategies were tested to remove these false positives, but in the end, the only thing that proved effective was removing all flare candidates that were too close to the start/end point or to a gap.

For very long events (longer than 0.2 days), yet another problem occurred. Figure 8 shows a selection of such flares. Only two or three of these candidates seem to have the right shape for a flare. The others could be associated with different processes in the stellar atmosphere like stellar pulsation (Maehara et al., 2012) or just bad-fitting model curves. To eliminate this issue, another filter was implemented that looked at the position of the flare peak and only kept the flare candidates with their flare peak near the beginning. However, this might not always be a good indicator if the number of flare points is small.

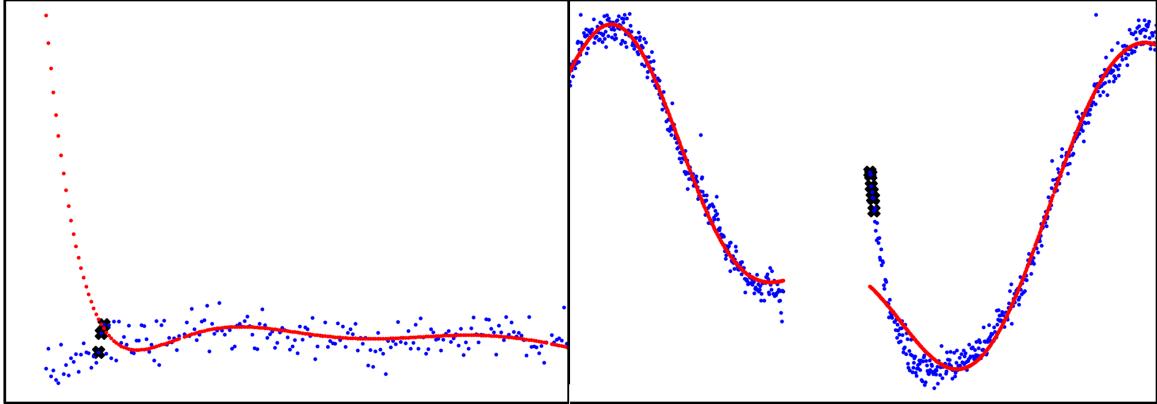


Figure 7: False positive for KIC 4851941 (left, caused by a misbehaving model function) and KIC 3217852 (right, caused by some external influence on the flux) using the FLATW'RM algorithm. The candidate flare points are marked with black crosses, the light curve points with blue dots and the model curve with red dots. The vertical axis shows the flux measured by the telescope in units of electrons per second. The horizontal axis shows the time in the Julian date format (JD). The plot shows long cadence data with roughly 30 minutes between each point.

This has to do with the uncertainty of the position of the peak (long cadence data) and how flare points are detected (the 3σ threshold makes the algorithm miss the last few flare points). Therefore, a minimum of five flare points is required for this rule to take effect.

The last fixed issue concerned the function determining the star's rotational period. As an input, it required a period range besides the time and the flux. When the maximum period was set to the data baseline, the returned period was not always correct (see figure 9). Since the longest rotational period for any star in the Okamoto catalogue was 40.045 ± 0.910 days, it was safe to set the maximum period to 50 days, which fixed the issue. Beware that this is a hard-coded fix, and the input period range may need to be changed if another star catalogue is used.

Several minor bug fixes and other small changes were also needed to get the code to work as intended, but as they did not change the algorithm's method, they will not be discussed here in detail.

Also, there were only estimates of the flare energy based on numerical integration over the flare points using Simpson's rule. Since the flare energies in Okamotos' catalogue were given in ergs, a new energy calculation had to be added. As the AFD algorithm already implemented an energy calculation in ergs, the same function was used to calculate the energies for the FLATW'RM algorithm. Values for the effective temperature T_{eff} and stellar radius R_{star} used in the energy calculation were taken from the Kepler input catalogue.

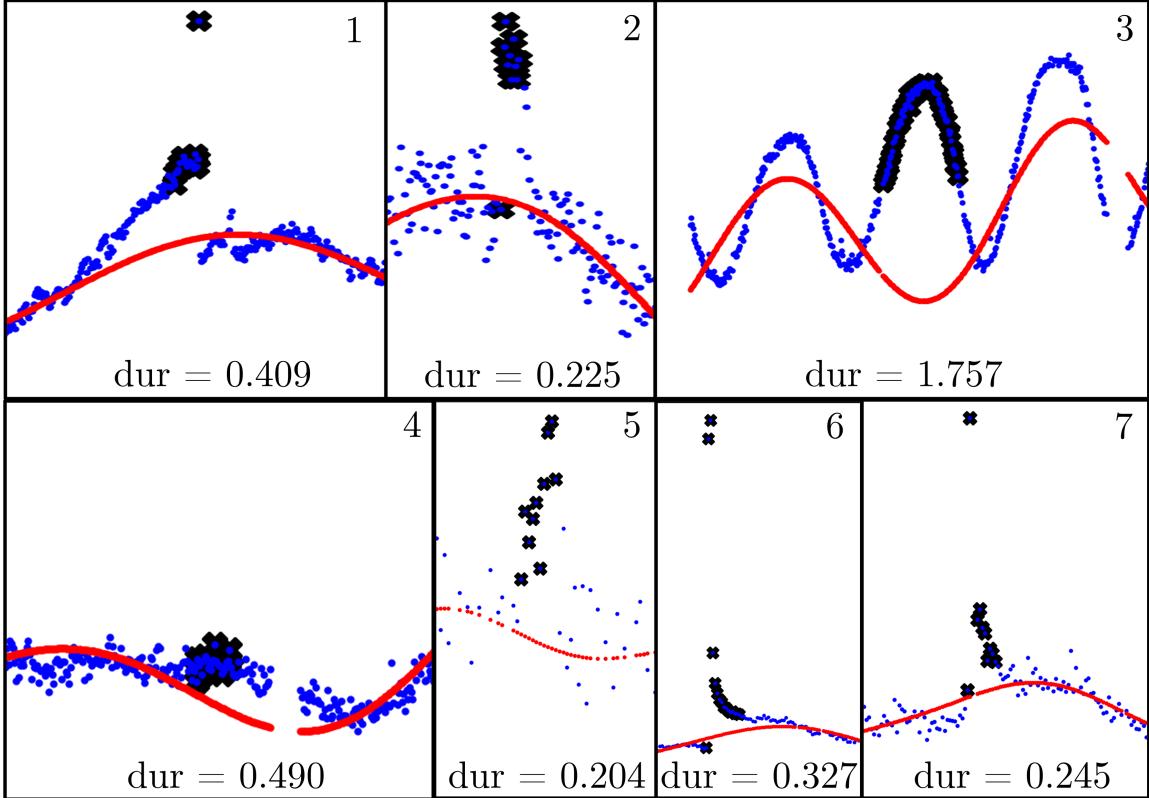


Figure 8: A selection of flares candidates detected by the FLATW'RM algorithm with long flare duration in days. The candidate flare points are marked with black crosses, the light curve points with blue dots and the model curve with red dots. Only numbers 2,6, and 7 have the right shape to be a flare. The other ones do not look like flares and, therefore, are false positives. No axes are displayed in this plot, as the individual images are snippets from individual plots with different scales but identical units on each axis. The vertical axis shows the flux measured by the telescope in units of electrons per second. The horizontal axis shows the time in the Julian date format (JD). The plot shows long cadence data with roughly 30 minutes between each point.

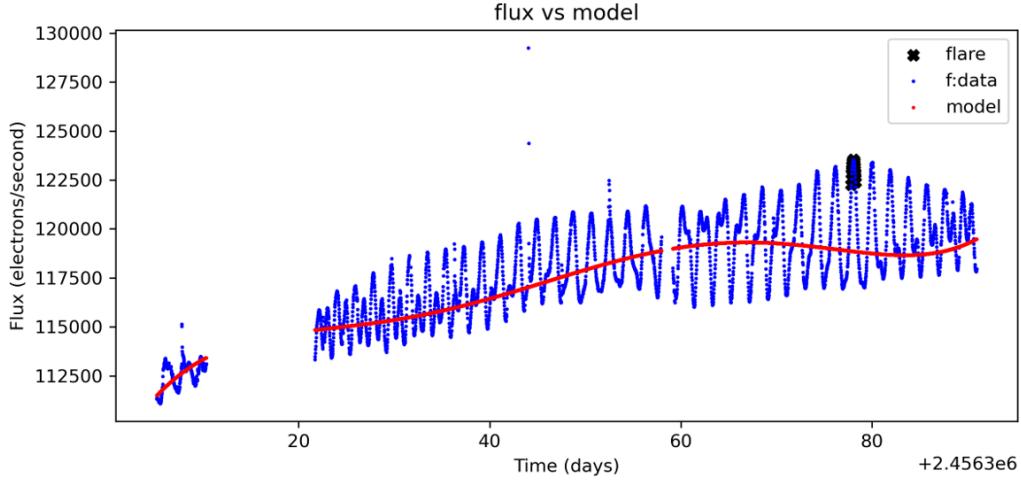


Figure 9: One light curve of KIC 44142137 where the predicted rotational period is incorrect, resulting in a large window size and a bad-fitting model curve. In this case, there is also a false positive flare candidate. The vertical axis is the flux measured by the telescope in units of electrons per second. The horizontal axis shows the time in the Julian date format (JD). The plot shows long cadence data with roughly 30 minutes between each point.

2.2.3 AFD

The AFD algorithm was based on techniques used by Maehara et al. (2012) and Shibayama et al. (2013). The steps are as follows:

1. Normalise flux and time

As with FLATW'RM, this is due to restrictions when using machine-learning estimators. The normalisation is done by dividing by the flux by dividing all values with the first flux value.

2. Calculate the flux difference in adjacent points
3. Find the value of flux difference such that the area under the flux difference distribution is equal to 1% of the entire region. The flare threshold is three times that value.

The same threshold value has been chosen by Shibayama et al. (2013) “as a result of a test run”. Figure 10c and 10d show two examples of the brightness distribution between two adjacent points. The point where the area under the distribution is equal to 1% of the entire region is marked with a green line, and the flare threshold is indicated with a red line. The value of the detection threshold depends on the star’s rotational period (Shibayama et al., 2013).

4. The start time of a flare is when the flux difference of two consecutive points exceeds the threshold for the first time.
5. A B-spline curve is fitted through three points around a flare. The first point is just before the flare, and the second and third are five and eight hours after the flare maximum. Each of these points is an average of five points.

A B-spline curve is a collection of piecewise polynomial functions. The points where the pieces meet, and the derivatives of these points may be continuous (Chaudhuri, 2021). Examples of such curves can be seen in figure 10e and 10f.

6. The end time of the flare is defined as the first point where the subtraction of the B-spline curve from the original relative flux is less than 3σ of the distribution.
7. The amplitude A of the Flare can be calculated from the normalised flux at the flare peak F_{\max} and the average normalised flux of five points before and after the flare $F_{\text{avg},1}$ and $F_{\text{avg},2}$.

$$A = \frac{F_{\max} - F_{\text{avg}}}{F_{\text{avg}}}, \quad \text{where } F_{\text{avg}} = \frac{F_{\text{avg},1} + F_{\text{avg},2}}{2}. \quad (1)$$

All the points that fulfil these steps are considered flare candidates. Conditions are applied to all flare candidates as follows:

- a) There are at least two data points between the peak and the end time of the flare
- b) If two consecutive data points qualify as the start time of the flare for the same Kepler ID, we choose the first point only as the beginning of the flare.
- c) The duration of the flare must be longer than 0.05 days.
- d) The declining phase is longer than the increasing phase
- e) The flare amplitude must be bigger than 0.0007 (normalised flux).

Flare candidates that pass all conditions above are included in the final flares list.

The AFD algorithm creates an output file containing the KIC, start, end, and peak time, as well as flare amplitude, energy, and duration values.

There is also a folder with stellar parameters from some A-, F-, G-, K- and M-type stars downloaded from the Q1-Q17 (DR25) stellar and planet catalogue (Thompson et al., 2017, 2018). The algorithm tries to find the star in this list and, if so, adds the values given for the effective Temperature T_{eff} , the decimal logarithm of the surface gravity $\log g$, the radius R and the rotational period P_{rot} of the star. If it can not find the star in the catalogue, it can get all those values from the Kepler input file, except for the rotational period, which is left empty.

There are three final output files: a list with the final flares, a list with all the flare candidates, and a list of all the errors that occurred when running the script.

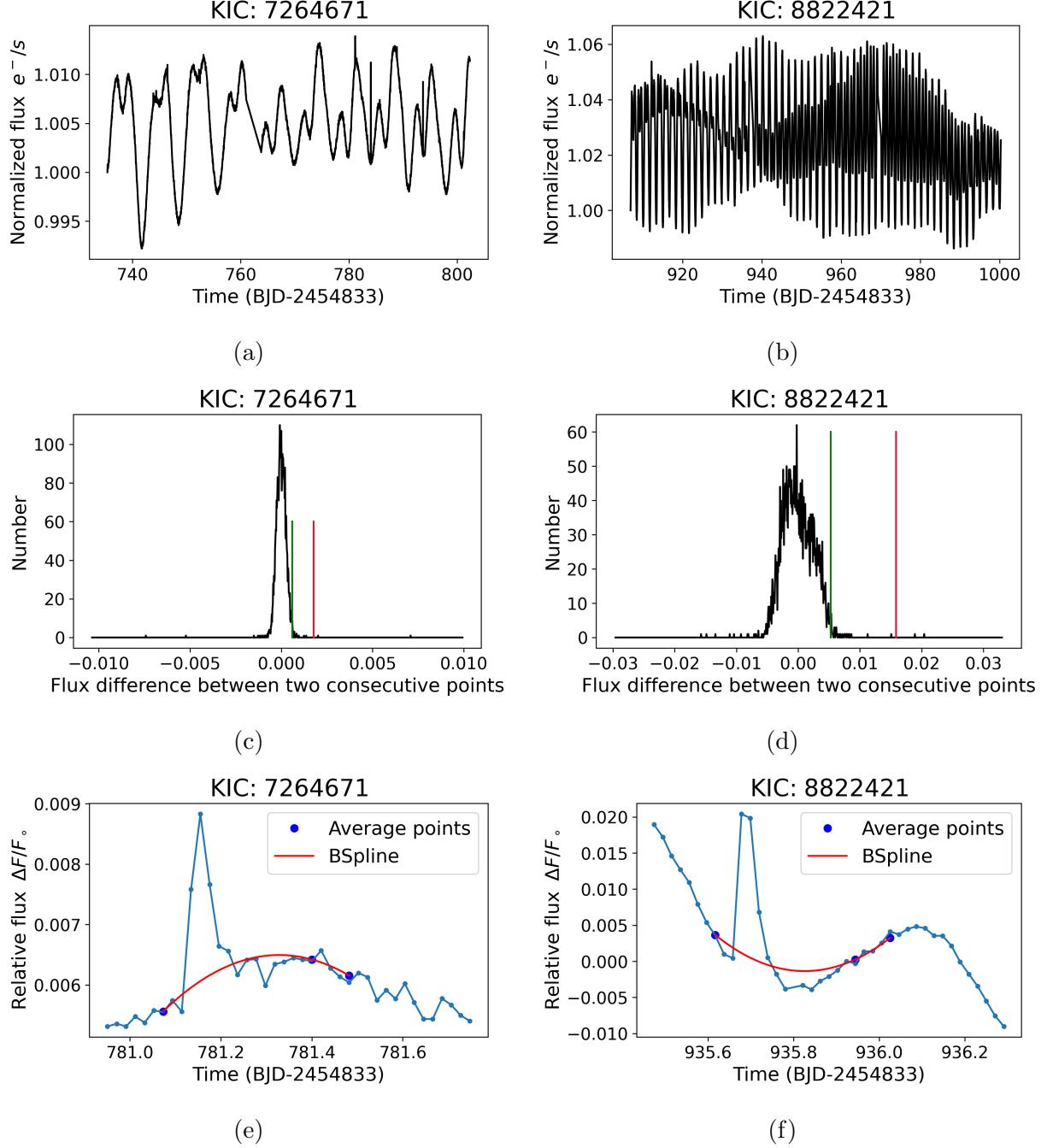


Figure 10: Illustration of the flare detection method used by Althukair and Tsiklauri in one slow one fast rotating star. (a) and (b) show the light curve. The distributions of brightness variation between every two adjacent points in the light curves are shown in (c) and (d). The green line indicates the value of flux difference where the area under the distribution is equal to 1% of the entire area. The red line is the flare threshold at a value three times the value of the green line. (e) and (f) show flares with a BSpline curve.

2.2.4 Modifications to AFD

The only notable change to the flare finding algorithm was applied to the filtering conditions: Additionally to the minimal flare duration of 0.05 days, a maximum flare duration of 0.5 days was added, as all the flares the algorithm detected with longer duration proved to be false positives.

All other changes were due to plotting specifications, data acquisition, and data saving. The energy calculation was not changed, but more on that in the next section.

2.3 Flare Energy

The total energy of a stellar flare can be estimated from the luminosity, flare amplitude and duration. The following calculation is based on Althukair and Tsiklauri (2023), Shibayama et al. (2013), and Yang et al. (2017) as it describes the procedure used for the AFD algorithm. Some changes were made to equation 4 compared to the equivalent equation in Althukairs' paper. The thoughts behind them are described in Appendix A.

The star luminosity L_{star} depends on the star radius R and the effective Temperature T_{eff} on the surface of the star. Assuming a black-body emitting star

$$L_{\text{star}} = \sigma_{\text{SB}} \cdot T_{\text{eff}}^4 \cdot 4\pi R^2, \quad (2)$$

where σ_{SB} is the Stefan-Boltzmann constant and $4\pi R^2$ is the entire surface of the star. Similarly, the luminosity of the flare can be written as

$$L_{\text{flare}}(t) = \sigma_{\text{SB}} \cdot T_{\text{flare}}^4 \cdot A_{\text{flare}}(t), \quad (3)$$

where T_{flare} is the temperature of the flare and A_{flare} the area. The calculations done by Althukair and Tsiklauri (2023) set $T_{\text{flare}} = 9'000$ K. The area of the flare can be estimated as

$$A_{\text{flare}}(t) = A \cdot \pi R^2 \frac{\int R_\lambda \cdot B_\lambda(T_{\text{eff}}) d\lambda}{\int R_\lambda \cdot B_\lambda(T_{\text{flare}}) d\lambda - \int R_\lambda \cdot B_\lambda(T_{\text{eff}}) d\lambda}, \quad (4)$$

where A is the flare amplitude for normalised flux as described in equation 1, R_λ is the Kepler response function (Bryson et al., 2010; Caldwell et al., 2010; J. Van Cleve, 2009) and B_λ is the Plank function at a given wavelength:

$$B_\lambda(T) = \frac{2hc^2/\lambda^5}{e^{hc/\lambda kT} - 1}, \quad (5)$$

with h the Plank constant, c the speed of light, T the Temperature of the black body and k the Boltzmann constant.

The total energy of the flare is the integral of $L_{\text{flare}}(t)$ over the flare duration

$$E_{\text{flare}} = \int_{t_{\text{start}}}^{t_{\text{end}}} L_{\text{flare}}(t) dt. \quad (6)$$

3 RESULTS

3 Results

3.1 Flares found with Algorithms

The following table (1) shows the number of flares found with each algorithm and those found by Okamoto et al. (2021). The number of unfiltered flare candidates given by AFD is also listed in the same table. FLATW'RM only found flares on 263 of the 265 stars in Okamoto's list. AFD found no flares for 48 stars included in Okamoto's list, but in the AFD candidates list, there is at least one flare for each star.

Table 1: Flares found on 265 stars using different flare-finding algorithms.

Catalogue	Number of Flares	Number of flaring stars
Okamoto	2341	265
FLATW'RM	7732	263
AFD	1716	217
AFD candidates	9804	265

The number of common flares between each catalogue is shown in table 2. They are calculated using the timestamps (start, end and peak times) of the flares given by each algorithm. A note on how these numbers were calculated can be found in Appendix B.

Table 2: Comparison of flares found on 265 stars using different flare-finding algorithms.

Compared Catalogues	Number of common Flares
Okamoto & FLATW'RM	1953
Okamoto & AFD	1304
Okamoto & AFD candidates	1771
Okamoto & FLATW'RM & AFD	1163
Okamoto & FLATW'RM & AFD candidates	1544
FLATW'RM & AFD	1469
FLATW'RM & AFD candidates	2221
AFD & AFD candidates	1716

These numbers also allowed us to calculate each set's unique flares. A more detailed illustration of how the different catalogues overlap is shown in the two Venn diagrams in figure 11 (since the AFD flares are a subset of the AFD candidates' flares, they are shown in separate diagrams).

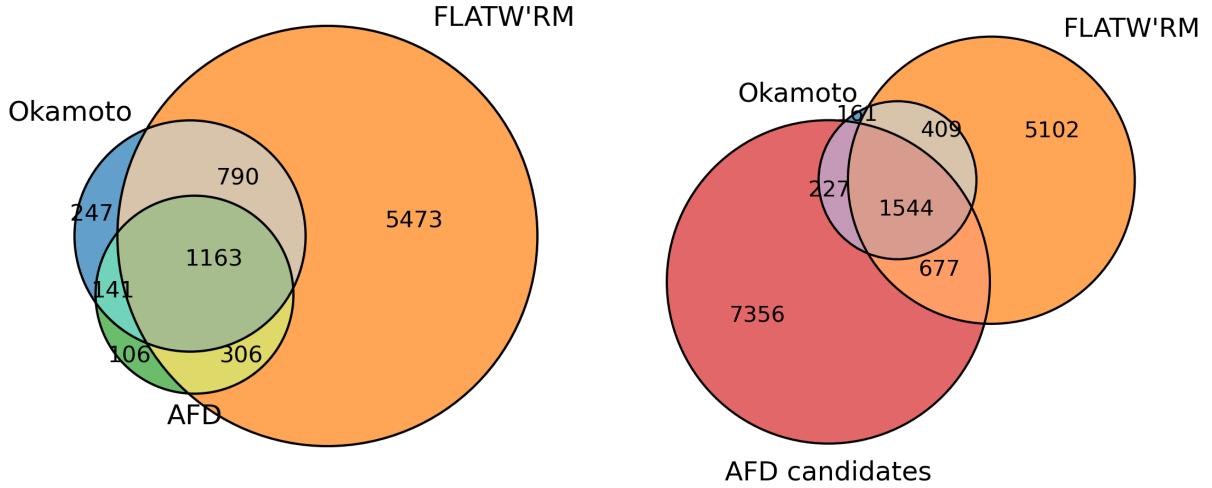


Figure 11: Two Venn diagrams showing the relation between the sets of flares found with each catalogue. Since the AFD flares are a subset of the AFD candidates’ flares, they are shown in separate diagrams. These diagrams show only the events declared as flares without making any assumption about the data quality.

3.2 Quality of the Flares found with different Algorithms

A random flare evaluation was performed to determine whether the flares found with these algorithms were true flares or something else. As there were too many flares to classify them by inspection, a random subset from each catalogue was taken. The flares from those subsets were then evaluated individually, and the number of flare points and shape of the flare were checked by eye. A flare was categorised as ‘Uncertain’ if it had two or more flare points but not quite the shape expected from a flare. For AFD approximately 5% of all flares were evaluated and around 2% for the other lists. This was due to the implementation of the random filtering method and the different sizes of the catalogues. The result from the random flare evaluation can be seen in table 3.

Table 3: Random flare evaluation for different flare-finding algorithms. The percentage of random flares to the total number of flares found in the catalogue is given in the second column. A flare was categorised as ‘Uncertain’ if it had two or more flare points but not quite the shape expected from a flare.

catalogues	Number of random Flares	False Positives	Uncertain	True Positives
Okamoto	51 (2.18%)	0	1	50 (98.0%)
FLATW'RM	151 (1.95%)	30	8	113 (74.8%)
AFD	84 (4.89%)	1	1	82 (97.6%)
AFD candidates	221 (2.24%)	138	4	79 (35.7%)

Using this table, some assumptions about the total number of true flares in the dataset can be made:

From the 2341 flares found by Okamoto et al. (2021) FLATW'RM found 1953 (83.4%) and AFD found 1304 (55.7%). However, there were also many flares that were not included in Okamotos' list: For the calculated precision (percentage of true positives), there were 5784 true positives from the 7732 flares found by FLATW'RM. Therefore, around 3831 new flares. For AFD there were 1675 true positives from the 1716 flares found by AFD. Therefore, around 371 new flares of which up to 306 (82%) were also detected by FLATW'RM (see figure 11, Venn diagram on the left).

For the AFD candidates list, there were even more flares common with both Okamoto and FLATW'RM: From the 2341 flares found by Okamoto, 1771 (75.7%) were also in the AFD candidates list. Accounting for the calculated precision, there were a total of 3'505 true positives from the total of 9804 flares in the AFD candidates list. Therefore, there are around 1'734 new flares, of which up to 677 (39%) were also detected by FLATW'RM (see figure 11, Venn diagram on the right).

3.3 Properties of Flares

Apart from the flare times, each algorithm also returns many other parameters. A summary of the most important ones can be found in table 4. The energy calculation for AFD was added to FLATW'RM as a modification to the original code (see section 2.2.2) and thus marked with brackets.

The effective temperature (T_{eff}), radius (R_{star}) and rotational period (P_{rot}) of each star are not necessarily the same for Okamoto and AFD. AFD took the values from the Q1-Q17 DR 25 stellar catalogue (Mathur et al., 2017a, 2017b) or extracted them from the Kepler input catalogue if the star was not listed in the Q1-Q17 DR 25 stellar catalogue - which was almost always the case. In the latter case P_{rot} was not listed in the final table, as this parameter is not present in the Kepler input catalogue. Okamoto et al. (2021) did their own research on the stellar parameters, which is why they differed from the Kepler input catalogue data. A summary of those differences can be seen in table 5, where also the values for FLATW'RM, which were used to calculate the energy but were not printed in the output file, were compared.

For example the mean difference of the T_{eff} for Okamoto - FLATW'RM is calculated by subtracting $T_{\text{eff},\text{FW}}$ from FLATW'RM from the $T_{\text{eff},\text{Ok}}$ from Okamoto for each star and then calculating the mean of these values. A mean of -4.63 K implies that, on average, the effective temperature of a star in the catalogue used for FLATW'RM is 4.63 K higher than the effective temperature in Okamoto's catalogue.

From these parameters, several plots were created, which are shown in the following sections.

Table 4: Summary of the most important parameters returned by each catalogue. The energy calculation for FLATW'RM was added as a modification to the original code (see section 2.2.2) and thus marked with brackets. Also marked with brackets is the P_{rot} for AFD since this parameter is not included in the final list if the parameters are taken from the Kepler input catalogue. KIC refers to the number given to the star in the Kepler input catalogue. The effective temperature (T_{eff}), radius (R_{star}) and rotational period (P_{rot}) of each star are not necessarily the same for Okamoto and AFD.

catalogue	KIC	t_{start}	t_{end}	t_{peak}	Duration	Energy	T_{eff}	R_{star}	P_{rot}
Okamoto	✓	x	x	✓	✓	✓	✓	✓	✓
FLATW'RM	✓	✓	✓	✓	x	(✓)	x	x	x
AFD	✓	✓	✓	✓	✓	✓	✓	✓	(✓)

Table 5: Table that shows the difference between the different values of effective temperature (T_{eff}) in Kelvin and stellar radius (R_{star}) as a multiple of the solar radius used for each star by the different catalogues. The first row shows how they were compared and the second row shows how many values were not the same from the total number of common stars. The difference occur as each catalogue uses different catalogue values. As AFD only found flares on 217 of the 265 stars from Okamotos' list the number of common stars with other catalogues is also 217.

Catalogue	Okamoto - FLATW'RM		Okamoto - AFD		FLATW'RM - AFD	
	T_{eff}	R_{star}	T_{eff}	R_{star}	T_{eff}	R_{star}
# Different Values	243/265	265/265	195/217	217/217	3/217	2/217
Mean difference	-4.63	0.03	-2.94	0.03	1.29	0.00
Stdev	38.1	0.34	30.33	0.36	19.17	0.00
Max difference	103	0.61	79	0.61	283	0.01
Min difference	-386	-1.34	386	-1.34	-3	0

3.4 Wait Time, Flare Duration and Energy

The wait time between two consecutive flares from the same star is the difference in their flare peak time. Therefore the number of wait times is one less than the number of flares occurring for each star (KIC).

Normalised data was used to display the frequency of different wait times in a histogram to compare the catalogues in one plot, as shown in figure 12. These plots show the normalised frequency, which means that the height for each bar is calculated by dividing the number of measurements in the range of the bar (bin size) by the total number of measurements for each catalogue. The range of the wait times displayed in the plot goes from 0 to 200 days, while the frequency has been calculated for 0 to 1000 days. For each catalogue, the percentage of measurements that lie outside the displayed range of wait times F_1 has been calculated and displayed in the plot legend.

In addition to the histogram, a cumulative distribution is shown in the same figure. In addition to the data for each catalogue an exponential distribution and a quadratic distribution were added to this figure. The formula for both distributions are

$$f_{\text{exp}}(x; m, t) = m \cdot e^{-t \cdot x}, \quad f_{\text{quad}}(x; a, b) = \frac{b}{a \cdot x^2 + 1}. \quad (7)$$

The quadratic distribution fits the data better for wait times over 100 days and the exponential distribution results in a better fit for wait times below 100 days.

Additional plots, where the histogram in different ranges for each catalogue individually, can be found in appendix C.1.

The next figures shows the wait time plotted against the flare energy. The energy can be plotted against the wait time before or after the flare, resulting in slightly different plots. Plots containing all catalogues are shown in figure 13a plotting the flare energy with the wait time to the next flare. Figure 13b plots the flare energy with the wait before the flare. A figure showing the plots for each sensor individually can be found in appendix C.1 figure 34.

Figure 14 shows one plot where the flare duration is plotted against the flare energy for all flares with different coloured catalogues. Individual plots and three plots comparing Okamoto to FLATW'RM, AFD and the AFD candidates can be found in appendix C.2. For the flares from the AFD candidates list, the flare duration range was limited to between zero and 1 day, as there were also flares with negative lengths and extremely long flares with a duration of over 11 days.

Looking at the wait times for FLATW'RM, we note that the number of flares exponentially decreases with longer wait times, which is also what is happening with the flares found by AFD and Okamoto. The only difference occurs for very short wait times, where FLATW'RM has a much higher number of flares (see figure 12).

If we look at the figures for energy and wait times and energy and flare duration of the flares (see figures 13 and 14), we see that the minimum energy for FLATW'RM ($\approx 10^{33}$

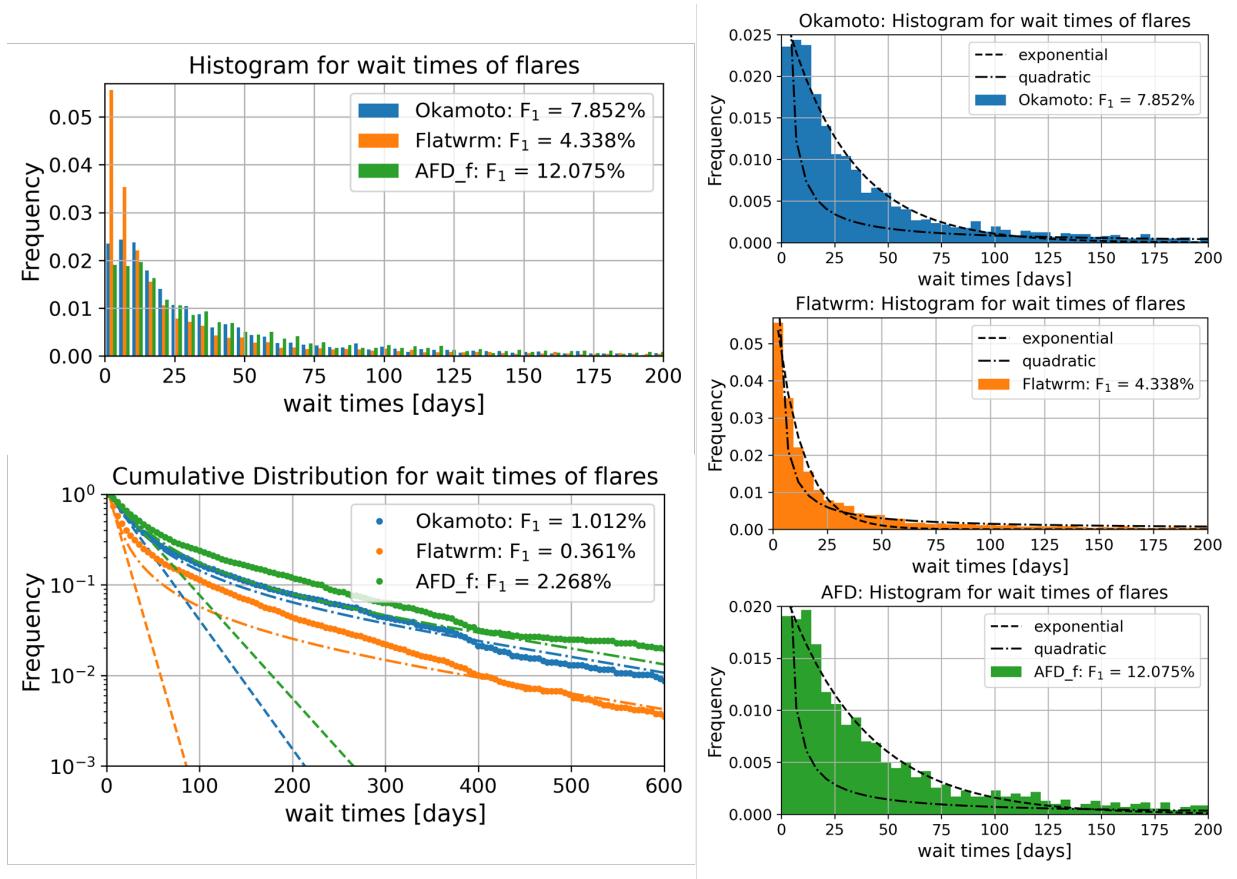


Figure 12: Normalised wait time histogram for the Okamoto, FLATW’RM and AFD flares with wait times up to 200 days (top left) and a cumulative distribution shown on the bottom left with a log scaled y-axis. On the right the histogram for each catalogue is displayed with added exponential and quadratic fit. The horizontal axis displays the wait time in days, and the vertical axis shows the frequency of flares found in the range of each bin. $F_1 \%$ refers to portion of the total received measurements that lie outside the range of values on X-axis of this plot. The dashed lines are exponential fits to the cumulative distribution and the dash-dotted lines are the quadratic fit.

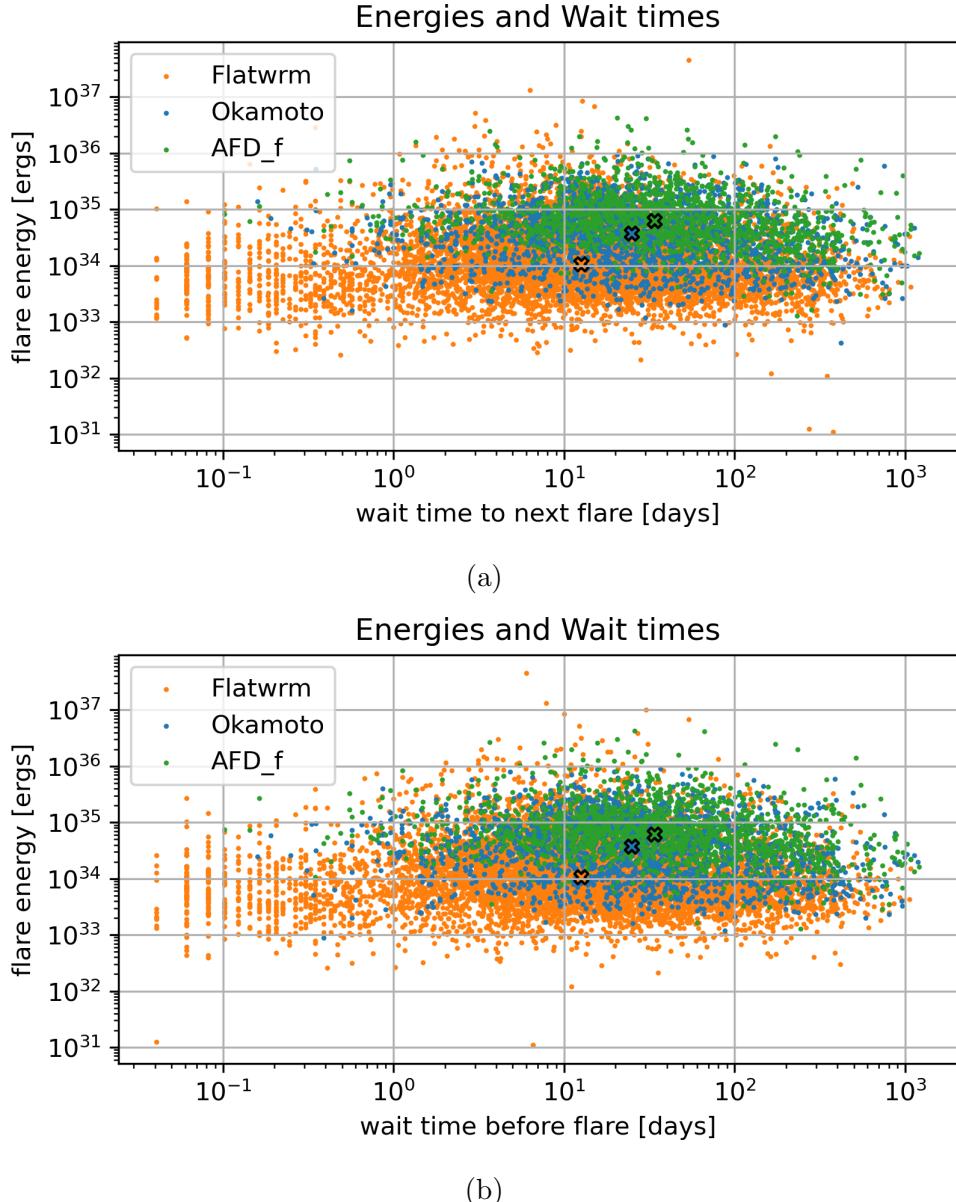


Figure 13: Plot with the wait time *to the next* (top) or *before* (bottom) the flare on the horizontal axis and the flare energy on the vertical axis for Okamoto (blue), FLATW'RM (orange) and AFD (green). The median in both axes is marked with an 'x' for each catalogue. Both plots show the distribution of wait times for different energies.

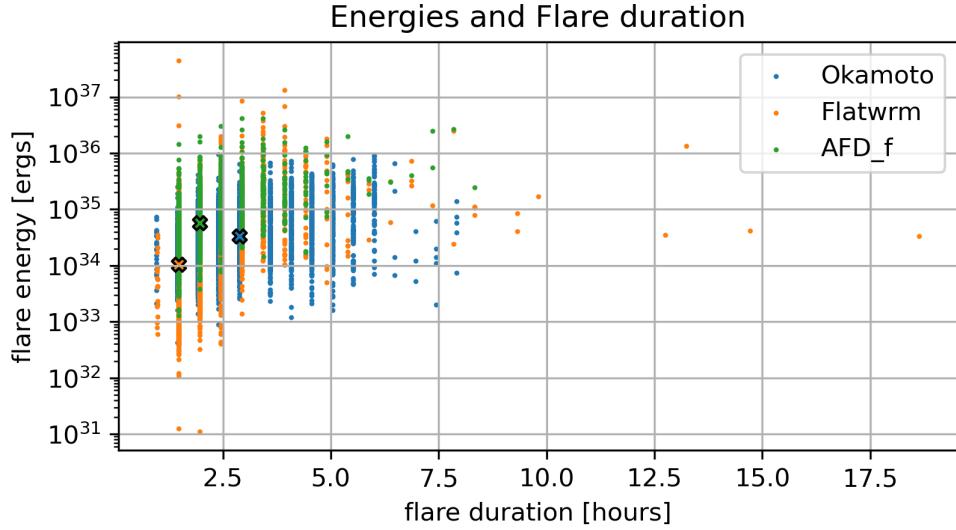


Figure 14: Plot showing the flare duration on the horizontal axis and the flare energy on the vertical axis for Okamoto (blue), FLATW'RM (orange), AFD (green). The crosses mark the point of median flare duration and median energy.

ergs) is lower than the minimum energy for Okamoto ($\approx 5 \cdot 10^{33}$ ergs), which again is lower than the minimum energy for AFD ($\approx 10^{34}$ ergs). The difference in the minimum energies is bigger for short wait times and nearly identical for long wait times. The maximum flare energy is similar for all three catalogues.

FLATWR'M has a unique set of flares with energies below $\approx 10^{33}$ ergs and wait times shorter than 3.5 hours, and for wait times shorter than 1 day, where AFD and Okamoto do not show many flare events.

For each flare the duration is given, either directly in the output file or by calculating the difference between the start and end time of the flare. Since the flux measurements occur in discrete time steps of 30 minutes (Kepler data sampling rate for long cadence data, see 1.2) the flare lengths also appear in discrete steps. For Okamoto and FLATW'RM, the minimum number of flare points required is two; thus, the shortest flares are approximately one hour long. For AFD the minimum number of flare points is three and hence, the shortest flares are 1.5 hours long.

Table 6: Number of stars and flares for different rotational periods for Okamoto (a), FLATW'RM (b) and AFD (c).

(a) Okamoto						
Rotational Period [days]	0-2	2-5	5-15	15-25	25-35	>35
Number of Stars	23	40	108	65	26	3
Number of Flares	685	854	611	145	43	3
Flares per Star	29.78	21.35	5.66	2.23	1.65	1.00
Error of mean	1.01	0.61	0.10	0.04	0.05	0.00
Standard deviation	23.33	24.54	10.95	2.54	1.27	0.00
(b) FLATW'RM						
Rotational Period days	0-2	2-5	5-15	15-25	25-35	>35
Number of Stars	22	40	107	65	26	3
Number of Flares	1468	2378	2067	1208	561	50
Flares per Star	66.73	59.45	19.32	18.58	21.58	16.67
Error of mean	1.96	1.41	0.25	0.17	0.48	4.37
Standard deviation	43.2	56.45	26.45	11.25	12.41	13.12
(c) AFD						
Rotational Period days	0-2	2-5	5-15	15-25	25-35	>35
Number of Stars	22	35	88	47	22	3
Number of Flares	458	576	495	138	46	3
Flares per Star	20.82	16.46	5.62	2.94	2.09	1
Error of mean	0.77	0.53	0.11	0.07	0.07	0.00
Standard deviation	16.85	18.39	9.83	3.09	1.50	0.00

3.5 Stellar rotational Period

To differentiate between stars with different stellar rotation periods, the stars were grouped into different ranges. For 265 stars and 2341 flares in the Okamoto dataset, table 6a shows a separation into different rotational periods. The same separation is shown for the 263 stars and 7732 flares found by FLATW'RM in table 6b and for the 217 stars and 1716 flares found by AFD table 6c. This separation was used for all the following subsections.

3.5.1 Rotational Period and Wait Time

Similarly to the section 3.4 figure 15 shows histograms where the wait time between two consecutive flares is shown on the horizontal axis and the frequency inside a bin is shown on the vertical axis as well as the cumulative distribution. On the left are stacked histograms for each catalogue showing the number of flares in binned ranges for the wait time. On the right, the cumulative distribution for each P_{rot} range.

For each P_{rot} range, the percentage of measurements that lie outside the range of wait times has been calculated and displayed in the plot legend. Also, the plot legend notes the

Table 7: Median, standard deviation (Stdev) and median for wait time (WT) grouped depending on their rotational period for Okamoto (a), FLATW'RM (b) and AFD (c). The rotational period, mean, standard deviation and median are in days. # Gaps refers to the number of gaps or wait times present in each range.

(a) Okamoto				
Rotational Period	# Gaps	Median WT	Mean WT	Stdev
0-2	662	21.6	42.4 ± 2.7	68.6
2-5	814	19.8	48.1 ± 3.2	91.6
5-15	503	33.1	89.7 ± 6.3	141.0
15-25	80	151	211 ± 24	217
25-35	17	320	329 ± 57	234
>35	0	-	-	-
(b) FLATW'RM				
Rotational Period	# Gaps	Median WT	Mean WT	Stdev
0-2	1446	10.84	20.64 ± 0.88	33.32
2-5	2338	8.62	22.62 ± 0.94	45.66
5-15	1960	20.5	57.4 ± 2.1	94.3
15-25	1143	23.3	68.3 ± 3.5	119.7
25-35	535	14.4	57.5 ± 4.8	111.7
>35	47	14	80 ± 21	145
(c) AFD				
Rotational Period	# Gaps	Median WT	Mean WT	Stdev
0-2	436	27.7	59.5 ± 4.5	93.4
2-5	541	25.1	61.5 ± 4.9	114.1
5-15	407	40.9	108.9 ± 8.4	168.9
15-25	91	164	214 ± 23	223
25-35	24	283	329 ± 56	274
>35	0	-	-	-

total wait times for each range of rotational period as 'gaps' as the wait time represents the gap between two consecutive flares.

For the different P_{rot} ranges the mean, standard deviation and median of the wait times could be calculated. These values are shown in table 7a for Okamoto, table 7b for FLATW'RM and table 7c for AFD.

3 RESULTS

3.5 Stellar rotational Period

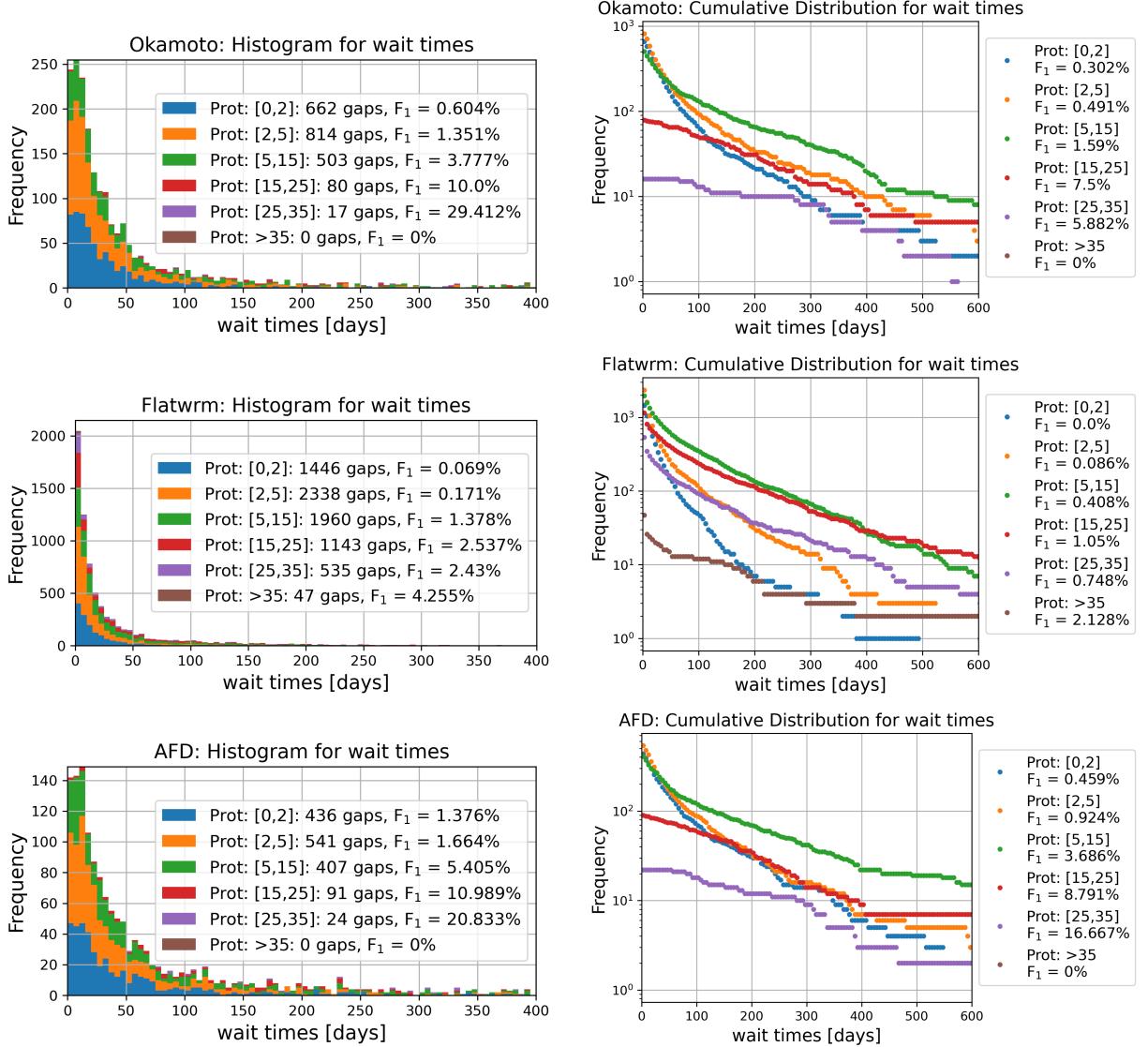


Figure 15: Stacked wait time histogram for the Okamoto, FLATW'RM and AFD flares with wait times up to 400 days (left) and a cumulative distribution of up to 600 days shown on the right with a log-scaled vertical axis. Wait time is defined as the time between two consecutive flares from the same star. The horizontal axis displays the wait time in days, and the vertical axis shows the number of flares (labelled frequency) found in the range of each bin. F₁ % refers to the portion of the total received measurements that lie outside the range of values on X-axis of this plot. The flares in each plot have been grouped by their stellar rotational period.

3.5.2 Rotational Period and Energy

The next plots in figure 16 show the wait time plotted against the flare energy. As described in section 3.4, the energy can be plotted against the wait time before or after the flare, which results in slightly different plots. The number of wait times (gaps) was again denoted in the plot legend.

Also, the flare duration could be plotted against the energy. This is shown in figure 17. As each flare has a distinct energy and duration, there is no need to count gaps; the number of events in each range is shown in the plot legend.

For the different P_{rot} ranges, the mean, standard deviation and median of the flare duration and flare energy could be calculated. These values are shown in table 8a and 9a for Okamoto, 8b and 9b for FLATW'RM, and 8c and 9c for AFD.

The rotational period was plotted against the energy of the flare, which is shown in figure 18. A polynomial of degree one and three were fitted to the data points. The value for the adjusted R^2 indicates how well the polynomial fits the data while adjusting for the number of terms in the model. If useless variables are added to a model, the adjusted R^2 will decrease. If useful variables are added, the adjusted R^2 will increase. For all catalogues, the polynomial of order three results in a better fit.

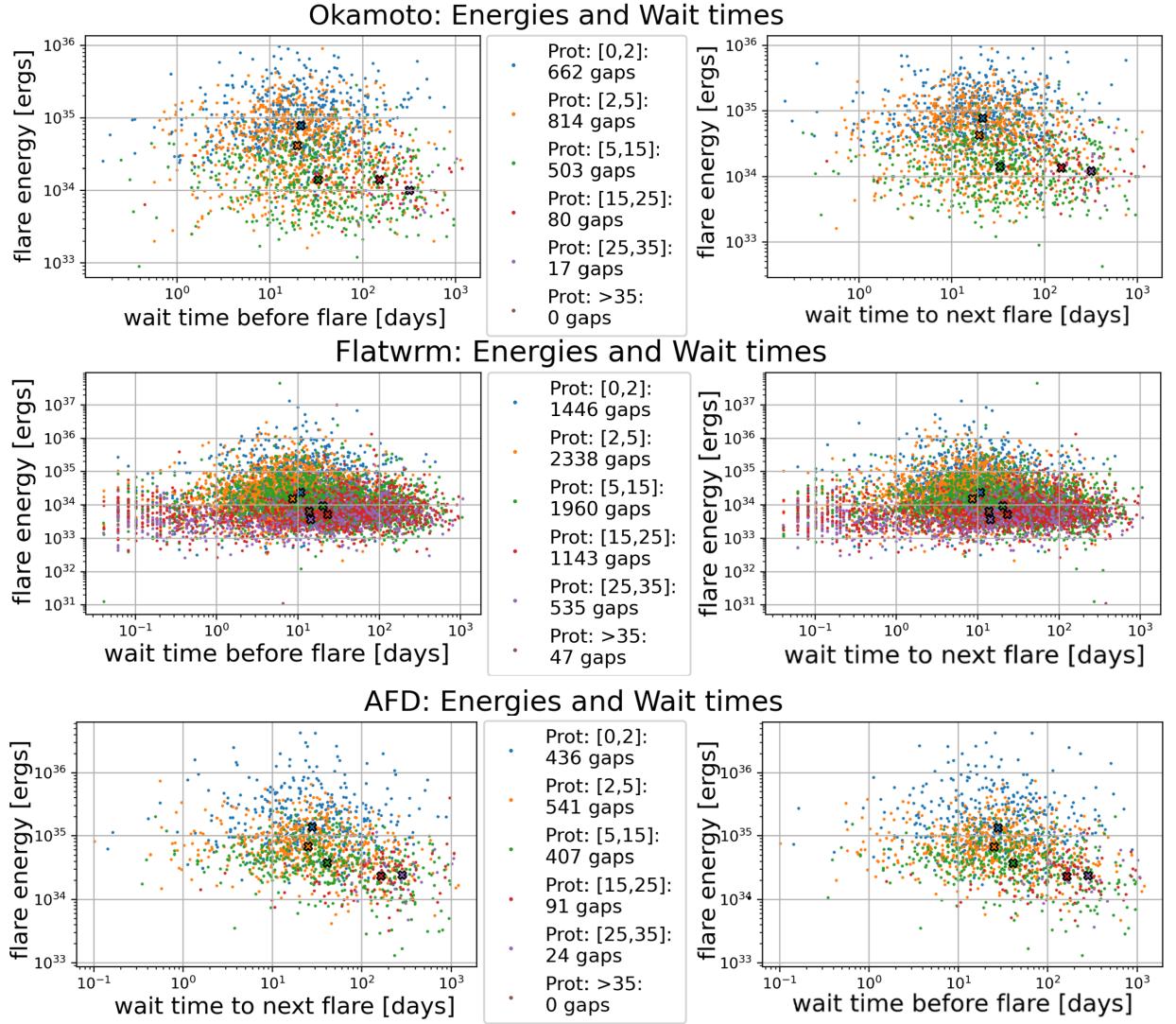


Figure 16: Plot with the wait time *before* (right) or *after* (left) the flare on the horizontal axis and the flare energy on the vertical axis splitted into different ranges of rotational periods. The plots for Okamoto are on the top, for FLATW'RM in the middle and for AFD on the bottom. Wait time is defined as the time between two consecutive flares from the same star. The median in both axes is marked with a cross for each range. The number of wait times (gaps) for each range is denoted in the plot legend.

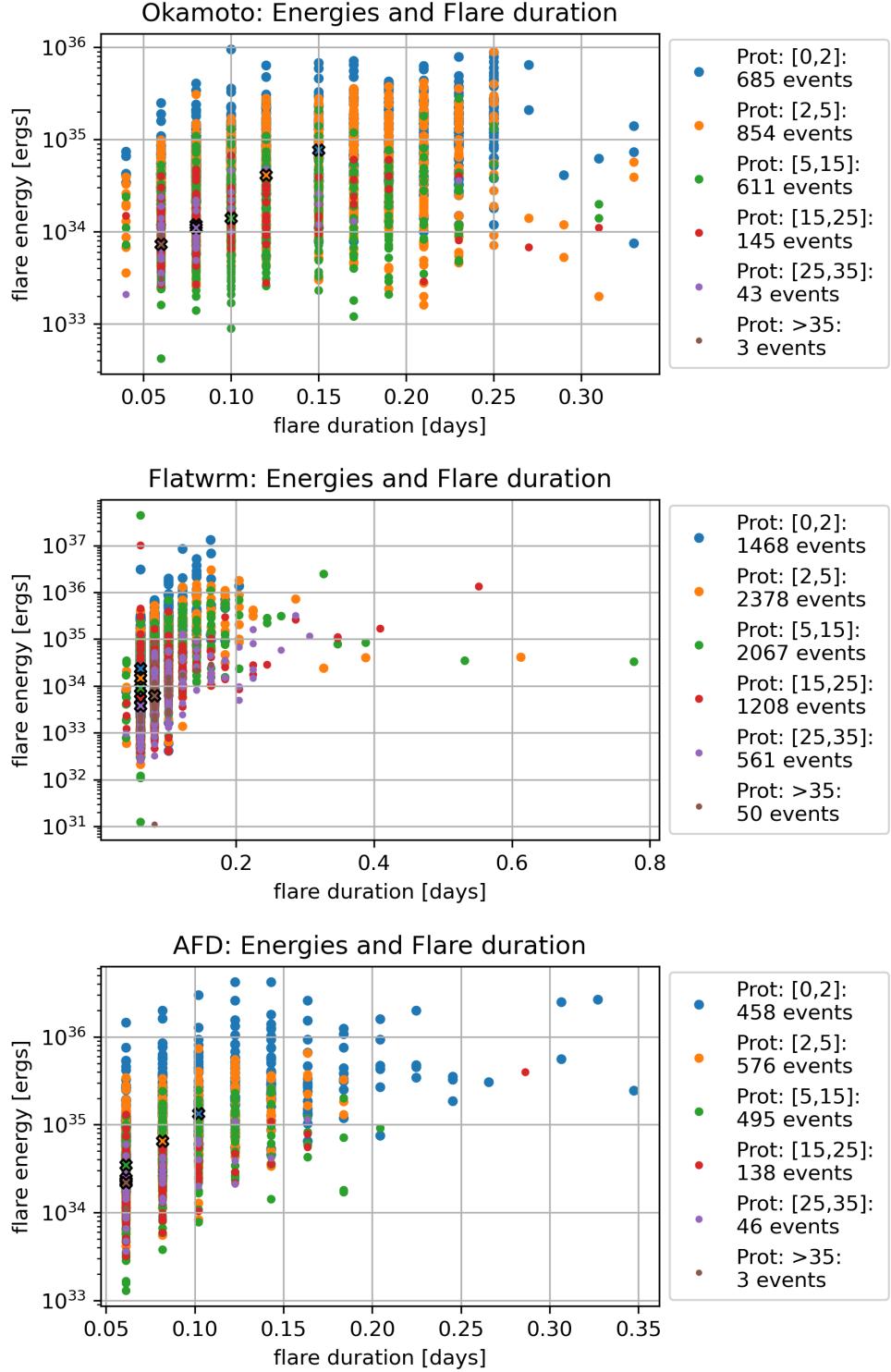


Figure 17: Plot showing the flare duration on the horizontal axis and the flare energy on the vertical axis for Okamoto (top), FLATW'RM (middle) and AFD (bottom) split into different ranges of rotational periods. The crosses mark the point of median flare duration and median energy.

Table 8: Median, standard deviation (Stdev) and median for the flare duration grouped depending on their rotational period for Okamoto (a), FLATW'RM (b) and AFD (c). The rotational period are in days, the mean, standard deviation and median are in hours.

(a) Okamoto				
Rotational Period	# Events	Median Duration [h]	Mean Duration [h]	Stdev [h]
0-2	1468	3.60	3.58 ± 0.06	1.49
2-5	854	2.88	3.16 ± 0.04	1.30
5-15	611	2.40	2.61 ± 0.05	1.13
15-25	145	1.92	2.22 ± 0.09	1.06
25-35	43	1.92	2.05 ± 0.13	0.87
>35	3	1.44	1.44 ± 0.00	0
(b) FLATW'RM				
Rotational Period	# Events	Median Duration [h]	Mean Duration [h]	Stdev [h]
0-2	1468	1.47	1.75 ± 0.01	0.48
2-5	2378	1.47	1.76 ± 0.01	0.63
5-15	2067	1.47	1.74 ± 0.02	0.72
15-25	1208	1.47	1.75 ± 0.02	0.70
25-35	561	1.47	1.90 ± 0.03	0.75
>35	50	1.96	2.06 ± 0.07	0.46
(c) AFD				
Rotational Period	# Events	Median Duration [h]	Mean Duration [h]	Stdev [h]
0-2	458	2.45	2.46 ± 0.05	1.04
2-5	576	1.96	1.99 ± 0.02	0.58
5-15	495	1.47	1.90 ± 0.03	0.60
15-25	138	1.47	1.79 ± 0.06	0.65
25-35	46	1.47	1.89 ± 0.09	0.60
>35	3	1.47	1.47 ± 0.00	0

Table 9: Median, standard deviation (Stdev) and median for the flare energy grouped depending on their rotational period for Okamoto (a), FLATW'RM (b) and AFD (c). The rotational period are in days, the mean, standard deviation and median are in ergs.

(a) Okamoto				
Rotational Period	# Events	Median Energy [ergs]	Mean Energy [ergs]	Stdev [ergs]
0-2	685	$7.7 \cdot 10^{34}$	$(1.199 \pm 0.049) \cdot 10^{35}$	$1.3 \cdot 10^{35}$
2-5	854	$4.1 \cdot 10^{34}$	$(6.34 \pm 0.25) \cdot 10^{34}$	$7.2 \cdot 10^{34}$
5-15	611	$1.4 \cdot 10^{34}$	$(2.14 \pm 0.10) \cdot 10^{34}$	$2.5 \cdot 10^{34}$
15-25	145	$1.2 \cdot 10^{34}$	$(1.61 \pm 0.10) \cdot 10^{34}$	$1.2 \cdot 10^{34}$
25-35	43	$1.1 \cdot 10^{34}$	$(1.43 \pm 0.16) \cdot 10^{34}$	$1.0 \cdot 10^{34}$
>35	3	$7.3 \cdot 10^{33}$	$(6.5 \pm 1.5) \cdot 10^{33}$	$2.5 \cdot 10^{33}$
(b) FLATW'RM				
Rotational Period	# Events	Median Energy [ergs]	Mean Energy [ergs]	Stdev [ergs]
0-2	1446	$2.4 \cdot 10^{34}$	$(1.12 \pm 0.14) \cdot 10^{35}$	$5.2 \cdot 10^{35}$
2-5	2338	$1.5 \cdot 10^{34}$	$(4.32 \pm 0.25) \cdot 10^{34}$	$1.2 \cdot 10^{35}$
5-15	2067	$9.1 \cdot 10^{33}$	$(4.5 \pm 2.2) \cdot 10^{34}$	$9.9 \cdot 10^{35}$
15-25	1208	$5.2 \cdot 10^{33}$	$(2.02 \pm 0.85) \cdot 10^{34}$	$2.9 \cdot 10^{35}$
25-35	561	$3.8 \cdot 10^{33}$	$(8.48 \pm 0.83) \cdot 10^{33}$	$2.0 \cdot 10^{34}$
>35	50	$6.3 \cdot 10^{33}$	$(9.2 \pm 1.7) \cdot 10^{33}$	$1.2 \cdot 10^{34}$
(c) AFD				
Rotational Period	# Events	Median Energy [ergs]	Mean Energy [ergs]	Stdev [ergs]
0-2	458	$1.3 \cdot 10^{35}$	$(2.88 \pm 0.22) \cdot 10^{35}$	$5.2 \cdot 10^{35}$
2-5	576	$6.6 \cdot 10^{34}$	$(9.07 \pm 0.36) \cdot 10^{34}$	$2.0 \cdot 10^{34}$
5-15	495	$3.5 \cdot 10^{34}$	$(4.52 \pm 0.18) \cdot 10^{34}$	$2.0 \cdot 10^{34}$
15-25	138	$2.3 \cdot 10^{34}$	$(3.04 \pm 0.32) \cdot 10^{34}$	$2.0 \cdot 10^{34}$
25-35	46	$2.5 \cdot 10^{34}$	$(3.06 \pm 0.33) \cdot 10^{34}$	$2.0 \cdot 10^{34}$
>35	3	$2.2 \cdot 10^{34}$	$(1.89 \pm 0.32) \cdot 10^{34}$	$1.2 \cdot 10^{33}$

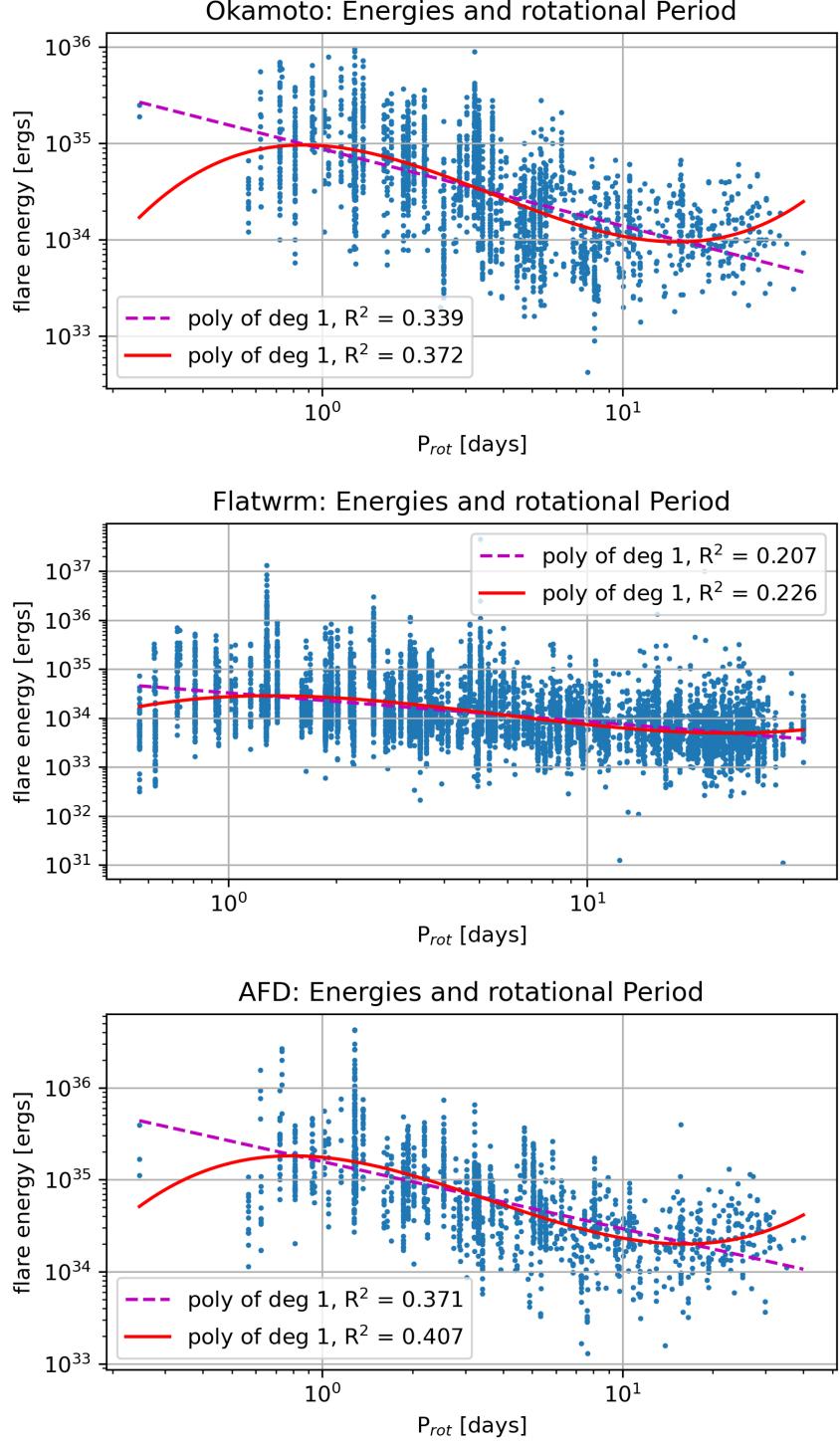


Figure 18: Plot showing the flares' rotational period P_{rot} on the horizontal axis and the flare energy on the vertical axis for Okamoto (top), FLATW'RM (middle) and AFD (bottom). To show possible trends a polynomial of degree one and three were fitted to the data points. R^2 (adjusted r-squared) measures how well the polynomial fits the data. A higher R^2 is indicating a better fit.

4 Discussion

4.1 FLATW'RM

4.1.1 Implementation and Usage

As mentioned in Vida and Roettenbacher (2018), applying the algorithm to a large dataset with different light curves is not possible without changing the searching parameters. Therefore, significant changes were made to the code for this study. The final setup for the input values described in 2.2.2 was based on visual inspection of the resulting flares and could be biased by the size and stars of the test sample and the human factor. A better approach would have been to choose a variation of random light curves from the whole dataset and have multiple human inspectors.

The final setup was a compromise between maximising the number of true flares and minimising the number of false positives, as a higher precision would have resulted in fewer true flares being found and vice versa.

Ultimately, a point was reached where trying to modify the algorithm more would have resulted in creating a new algorithm. Since that was not the goal of this thesis, it was left at the point presented here.

4.1.2 Detected Flares

According to the flares inspected in the random flare evaluation, the false positives found with FLATW'RM are short, low-energy flares in noisy light curves (left plot in figure 19). These outliers could be removed if the minimal number of flare points were increased to four or higher, but this would also negate the main strength of FLATW'RM, which is finding short and low-energy flares. Other false positives occurred when the model fitted to the light curve was not good enough (right plot in figure 19). However, there were also some false positives for longer flares. For those the position of the peak was by chance at the right point to pass through the filtering, but apart from that they had not the right shape or smoothness expected from a flare.

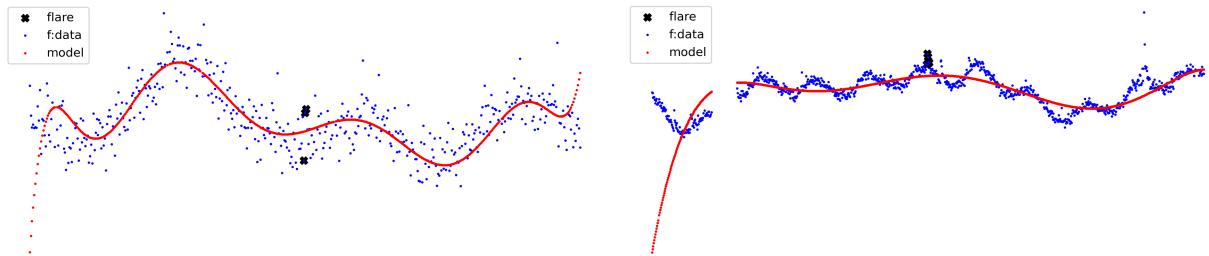


Figure 19: Two flares found by the FLATW'RM algorithm. The one on the left is a false positive occurring because of noisy data and some chance alignment of two consecutive points. The one on the right could be a flare, but since the model curve (red) does not fit the data well, it would only be by accident.

4.1.3 Flare Energy, Wait Time and Duration

The high number of flares found by FLATW'RM could be due to the algorithm's ability to detect short and low-energy flares. However, some flares have strong fluctuations right after the flare peak or even multiple peaks (Okamoto et al., 2021), which FLATW'RM might interpret as another flare, but Okamoto does not.

Also, there is no correlation between flare energy and the wait time. With longer wait times, more energy could accumulate in the magnetic field, which should theoretically result in higher energies. This is not the case, which is an interesting result. Maybe there is a correlation between wait times for subsequent flares as suggested by Snelling et al. (2020), but for this a different analysis method would be needed.

The difference in flare duration is primarily caused by different methods of determining the flare endpoint. When they are not the same, FLATW'RM's flare endpoint is usually first (see figure 20), which results in an offset between these two catalogues. This offset is especially notable in the top plot of figure 35.

But even with this offset, short-duration flares have a different distribution. For FLATW'RM, the minimum and maximum flare energy increases with the flare duration until a duration of at least four hours (see figure 17 or 36). For Okamoto, the maximum flare energy no longer increases at 3.6 h, and the minimum flare energy has no clear trend but stays more or less the same for all flare duration. This means that FLATW'RM can detect more low-energy flares with a duration shorter than four hours than Okamoto.

FLATW'RM also seems to find flares with higher energies ($> 10^{36}$ ergs) than Okamoto. This might be counter-intuitive at first, as one might assume that a shorter average flare duration would also imply a lower average energy. But since those extra flare points that Okamoto includes are very close or sometimes even below the model curve (see bottom plot on figure 20), they do not have any notable impact on the flare energy calculation. Therefore, the difference must be in how the flare energy was calculated.

To investigate this hypothesis, the flares with the ten highest energies from both the FLATW'RM and Okamoto were compared, and three flares were found, where the energy calculated with FLATW'RM is about an order of magnitude higher than the energy calculated by Okamoto (see table 10).

The different energy calculations are due to multiple effects: Not only could the method be slightly different (as explained in appendix A), but more importantly, the stellar parameters used for the energy calculations are different, as described in section 3.3. This makes it very difficult to compare more than the distribution of the energies involved.

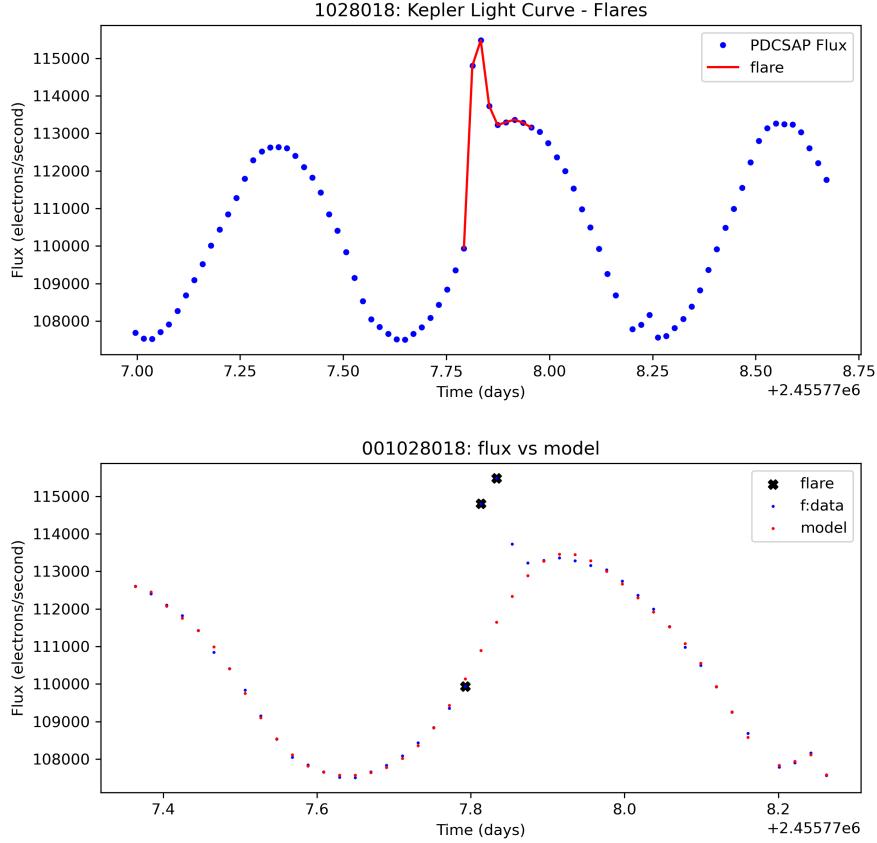


Figure 20: The duration of the flare for KIC 1028018 found by Okamoto et al. (2021) (top) is visibly longer than the flare found with FLATW'RM (bottom). Besides the measured flux (blue dots in both plots), the flare points are connected with a red line for Okamoto and marked with black crosses for FLATW'RM. The FLATW'RM plot also shows a model curve used to determine the flare points.

Table 10: When looking at the top 10 highest flare energies for both the FLATW'RM and the Okamoto catalogue, these three flares are in both lists. Notable, the flare energy for FLATW'RM is about an order of magnitude higher than the energy calculated by Okamoto, even though the flare duration is always higher for Okamoto.

KIC	10537061	10537061	10537061
Peak Time [JD]	2455621.2	2455759.8	2456171.8
FLAWRM energy [erg]	$1.32 \cdot 10^{37}$	$3.14 \cdot 10^{36}$	$5.23 \cdot 10^{36}$
Okamoto energy [erg]	$8.90 \cdot 10^{35}$	$9.60 \cdot 10^{35}$	$7.90 \cdot 10^{35}$
FLAWRM duration [h]	3.9	1.5	3.4
Okamoto duration [h]	6	2.4	5.5

4.2 AFD

4.2.1 Implementation and Usage

As with FLATW'RM, the algorithm was not designed to handle multiple light curves at once, but contrary to the latter, it did not need any input parameters which needed to change depending on the light curves.

It is also interesting, that it provides a list of flare candidates, but as the precision (estimated percentage of true positives) was only 35% it is not very useful for further analysis.

4.2.2 Detected Flares

AFD detects flares with high precision, meaning there are almost no false positives in the final list of flares. This is done by a strict filtering process but is at the cost of losing more than 50% of the true positives from the AFD candidates list.

Therefore it would be interesting to change the filtering criteria to get a higher recall. For example, the restriction on the normalised flare amplitude and the flare duration could be modified. Also, as has been seen for FLATW'RM, the restriction that the declining phase must be longer than the increasing phase limits the number of short flares found by the algorithm, as the number of flare points in the declining phase is generally underestimated by automated flare-finding algorithms.

4.2.3 Flare Energy, Wait Time and Duration

There seem to be more flares with lower energy for Okamoto than for AFD. This is again caused by different values for the effective temperature and stellar radius and also by different methods to calculate the energy (more about that in Appendix A).

Nonetheless, the possibility that this difference is caused by the different flares found in each catalogue can also not be excluded.

Flares in Okamoto's list sometimes have longer durations than the same flares found by AFD. This is again due to the more manual method used by Okamoto compared to the automated methods used by AFD and FLATW'RM. Even though the detecting methods for AFD and FLATW'RM were very different, the flare lengths seem to be similar on average. For example, the flare in figure 14 had the same length of 0.063 days for AFD and FLATW'RM, whereas the length was 0.17 days for Okamoto. That does not mean that the flare lengths given by AFD and FLATW'RM are better than those given by Okamoto: Since they miss the last few points, they always underestimate the actual flare duration. As there is no way to fix this issue without assuming anything about the shape of the flare, we must accept this as a property of automated flare-finding algorithms.

For the wait times, there seems to be a small trend, that suggests lower energies for longer wait times. However, as there is no such trend in the data from Okamoto and FLATW'RM, this is more likely to be a characteristic of the algorithm and not caused by

a physical process.

The number of flares first decays exponentially with longer wait times (figure 12), which has also been detected by Hawley et al. (2014) when they looked at short cadence light curves and wait times up to 10 hours. However, the number of flares decays slower than predicted by the exponential model and for wait times between 50 and 100 days the exponential fit does no longer correlate well with the data. For longer wait times a quadratic distribution fits better.

4.3 Dependence on rotational Period

Many new dependencies are revealed when the rotational period is included as an additional parameter.

For the mean wait time, the data from AFD and Okamoto are nearly identical and imply an exponential increase in wait time with longer rotational periods (see figure 21b), meaning that there are fewer flares for stars like our sun compared to faster rotation stars. This different to the result by Candelaresi et al. (2014) who found a quadratic increase, which could be due to the addition of the convective turnover time as a parameter, different binning or different stellar properties as they also included stars with lower effective temperatures. Balona (2015) did also mention a decrease flaring stars for lower rotational periods but did not make any statement about the decreasing rate.

Since the number of flaring stars also decreases with longer rotational periods, this means that there should be fewer flares per star. This trend can be seen for AFD and Okamoto in figure 21a. For FLATW'RM, the trend is a bit different, which can be interpreted in two ways:

1. The wait time/flare per star ratio stabilises for $P_{\text{rot}} > 5$ days.
2. There are some flares (or false positives) in the range between $25 > P_{\text{rot}} > 5$ days that are uncommon to the flares of AFD and Okamoto who disturb this trend. If those flares were removed, the same trends would be seen with all catalogues.

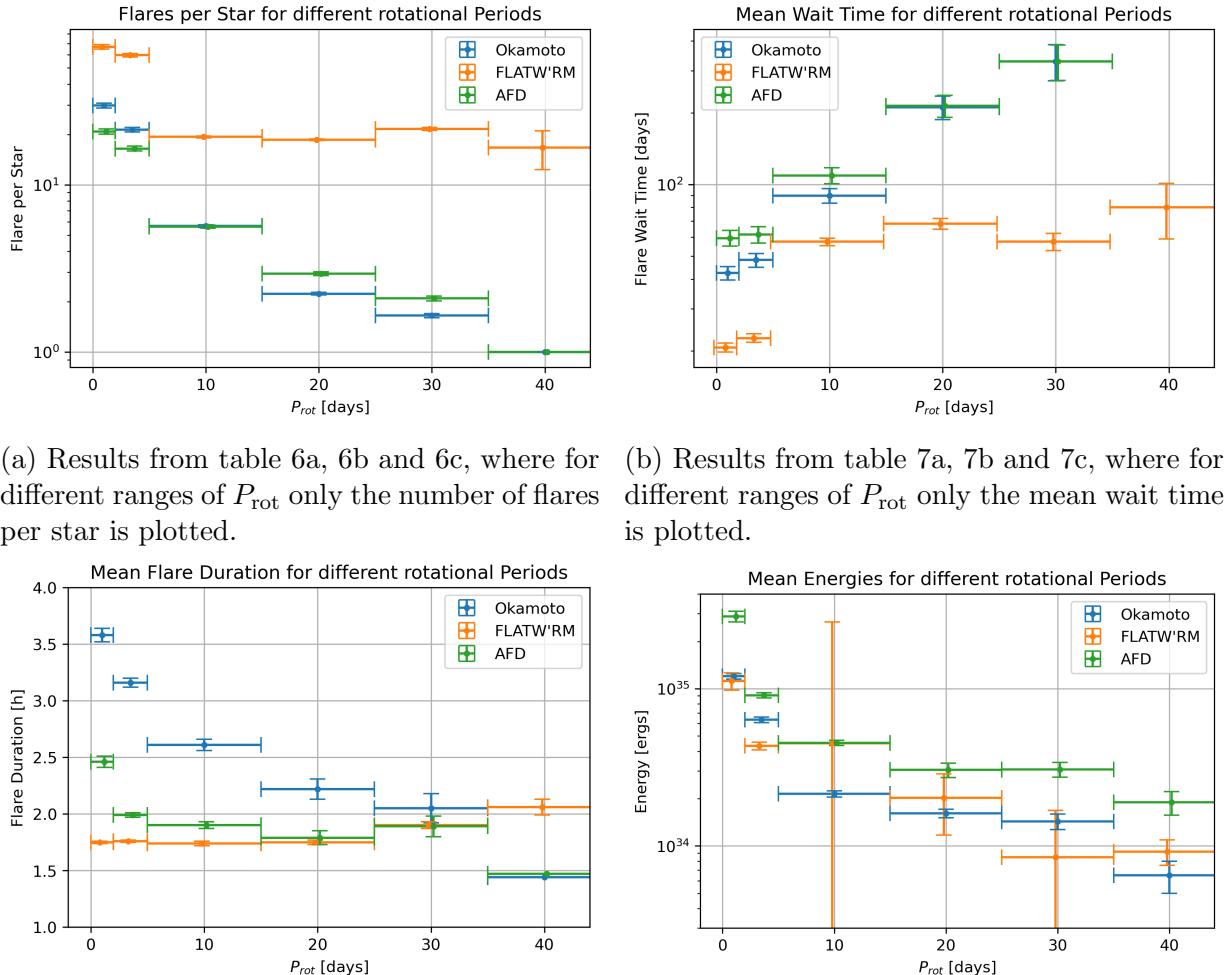
Similarly, the trends for mean flare energy for AFD and Okamoto in figure 21d look almost identical apart from a constant offset. The offset is probably due to the different catalogue values for the effective temperature and stellar radius, as described in the previous section. The energy values for FLATW'RM seem to follow the same trend, but the high uncertainties make it difficult to suggest anything specific.

But even with the high uncertainty in FLATW'RM's values, the data from all three catalogues predict an exponential decrease in flare energy and flare frequency for longer stellar rotational periods. This has already been observed by Okamoto et al. (2021) for slightly different bins.

The flare energy stagnates for $35 > P_{\text{rot}} > 15$ days. As this trend was not observed by Okamoto et al. (2021), this might only be caused by the different bin sizes. But if we look at plots of the energy and rotational period without bins (see figure 18) there seems to be an increase in flare energy for $P_{\text{rot}} > 10$ days as indicated by the third order polynomial

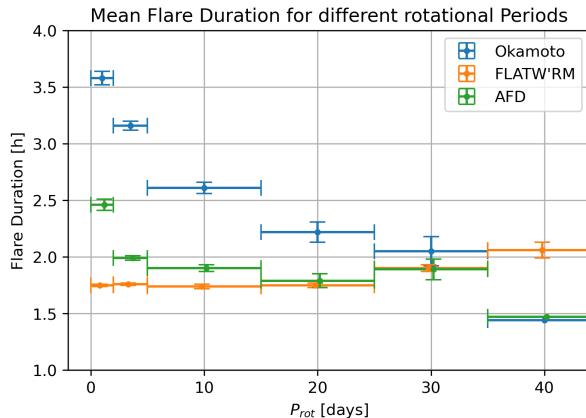
fit. But as before, this trend can only be seen in the data from AFD and Okamoto and not in the data from FLATW'RM.

When looking at the mean flare durations, there are some discrepancies between the different catalogues (figure 21c). For Okamoto, there is a negative correlation between the flare duration and the rotational period visible, which is - apart from one point - also visible for AFD. For FLATW'RM, it is the other way around.



(a) Results from table 6a, 6b and 6c, where for different ranges of P_{rot} only the number of flares per star is plotted.

(b) Results from table 7a, 7b and 7c, where for different ranges of P_{rot} only the mean wait time is plotted.



(c) Results from table 8a, 8b and 8c, where for different ranges of P_{rot} only the mean flare duration is plotted.

(d) Results from table 9a, 9b and 9c, where for different ranges of P_{rot} only the mean energy is plotted.

Figure 21: Plots showing the change in mean flares per star, wait time, duration and flare energy depending on different ranges in the rotational period. The error bar in the horizontal axis shows the range of values used to calculate the mean and error for each bin. For better visibility the centre points have been slightly offset from each other in the horizontal axis.

Summarising the trends observed in figure 21: Short rotational periods correlate with shorter wait times and higher flare energies (see figure 21b). Likewise, longer rotational periods result in higher wait times and lower flare energies (see figure 21d). However, the data also has very high variations, which means that these trends are only visible when looking at a large dataset. For individual light curves, this might not be true.

Since the results from FLATW'RM differ from those of the other algorithms, it would be interesting to find a reason for this. It could be caused by the false positives in the dataset. Therefore, as a first step, all the false positives must be removed manually before any other cause can be guessed. But this would involve going through over seven thousand light curves, which is very tedious and therefore unlikely to be done anytime soon, if at all.

5 Conclusion and Outlook

If we trust the precision calculated in section 3.2 the total number of new flares found in the dataset by the two algorithms could be as high as 7'336 flares (3'831 new FLATW'RM flares and 3'505 new AFD candidates flares, ignoring overlaps), resulting in a total number of 9'677 flares present in the dataset.

This is certainly an upper boundary to the actual number of true flares. If we only look at the number of flares that AFD declared as flares in the final list, AFD only finds 371 new flares, reducing the maximum number of flares found in the dataset to 6'543.

A more conservative approach would be to ignore the calculated precision and only look at the number of flares in at least two surveys (not counting the AFD candidates list). This would bring the total number of flares to 2'400, which is about the number of flares found by Okamoto.

The strength of the FLATW'RM algorithm is detecting low energy and short flares, as it finds much more of them than AFD and Okamoto. The main disadvantage is its low precision of approximately 75% - compared to AFD and Okamoto, who both have an precision of almost 100%. This means that the data from FLATW'RM still needs to be post-processed to do further analysis.

As this was not done here, there were some inconsistencies between FLATW'RM and the other two surveys, especially when looking at the flare properties in different ranges of the rotational period (see figure 21). However, it could also be that FLATW'RM found flares with different properties due to the difference in the flare-finding method.

Something that seems to be very difficult for flare-finding algorithms is getting the correct flare duration. When looking at individual flares detected in all three surveys, the flare durations of AFD and FLATW'RM were very similar, if not the same, and the duration for the flares found by Okamoto could be more than twice as long.

One reason for this is the noise threshold: As the luminosity of the flare decreases exponentially, the latter flare points are quickly below this threshold. As the flare sometimes goes even below the model curve of the light curve for a quiet star (see figure 20), it might be better to use the second point where the measured points cross the model points as the

5 CONCLUSION AND OUTLOOK

flare endpoint. Another option would be to lower the noise threshold after the flare peak, but this might be problematic for data with high noise levels.

The detection threshold for both algorithms should be challenged as they use the 3σ threshold first suggested by Shibayama et al. (2013) without questioning whether it is also applicable to their data. According to a paper by Rivera et al. (2023), the information content could be increased if the threshold was lowered to 1.56σ . Still, as their study is only based on the light curves from one star, this value is certainly not applicable to all stars.

A general solution would be to change the detection threshold according to the properties of the light curve. This has been done to some extent for the FLATW'RM algorithm as described in section 2.2.2, but the setup found there is far from perfect as the algorithm's precision is still very low. Rivera et al. (2023) suggest that mutual information could be used on the stellar wait time distribution to find a detection threshold that maximises the information content of the data.

One of the most surprising properties found while analysing the data from all three surveys was that there is no correlation between the wait time between flares and the flare energy. It would be interesting to study if there is a relation between subsequent flares, as suggested by Snelling et al. (2020), but for this, a more detailed analysis would be necessary.

The median wait time and energy, however, depend on the rotational period of the star. Short rotational periods correlate with shorter wait times and higher flare energies (see figure 21b). Likewise, longer rotational periods result in higher wait times and lower flare energies (see figure 21d). For the flare duration, no clear trends are visible.

The inverse dependency of rotational period and flare energy has already been published in multiple papers (Notsu et al., 2019; Okamoto et al., 2021), but the independence of the energy from the wait time between flares and that the wait time increases with longer rotational periods seem to be new findings.

As the need for automated algorithms will not decrease in the near future it would be interesting to compare even more algorithms. For example, FLATW'RM2 by Vida et al. (2021), which was initially planned as a third algorithm for this thesis but had to be cut for time management reasons, or AltaiPony by Davenport (2016) and Ilin, Ekaterina et al. (2021). The goal of further studies could be to use the insights gained by comparing these algorithms to create a new algorithm or to use different algorithms to create an extended flare catalogue.

References

- Althukair, A. K., & Tsiklauri, D. (2023). Main Sequence Star Super-flare Frequency based on Entire Kepler Data. *Research in Astronomy and Astrophysics*, 23(8), 085017. <https://doi.org/10.1088/1674-4527/acdc09>
- Balona, L. A. (2015). Flare stars across the H-R diagram. *Monthly Notices of the Royal Astronomical Society*, 447(3), 2714–2725. <https://doi.org/10.1093/mnras/stu2651>
- Berger, T. A., Huber, D., Gaidos, E., & van Saders, J. L. (2018). Revised Radii of Kepler Stars and Planets Using Gaia Data Release 2. *The Astrophysical Journal*, 866(2), 99. <https://doi.org/10.3847/1538-4357/aada83>
- Brown, T. M., Latham, D. W., Everett, M. E., & Esquerdo, G. A. (2011). Kepler Input Catalog: Photometric Calibration and Stellar Classification. *The Astronomical Journal*, 142(4), Article 112, 112. <https://doi.org/10.1088/0004-6256/142/4/112>
- Bryson, S. T., Tenenbaum, P., Jenkins, J. M., Chandrasekaran, H., Klaus, T., Caldwell, D. A., Gilliland, R. L., Haas, M. R., Dotson, J. L., Koch, D. G., & Borucki, W. J. (2010). THE KEPLER PIXEL RESPONSE FUNCTION. *The Astrophysical Journal Letters*, 713(2), L97. <https://doi.org/10.1088/2041-8205/713/2/L97>
- Bunnell, J. (2018, October). *Using Kepler Data to Plot a Light Curve*. Retrieved June 25, 2024, from https://github.com/spacetelescope/notebooks/tree/master/notebooks/MAST/Kepler/Kepler_Lightcurve
- Caldwell, D. A., van Cleve, J. E., Jenkins, J. M., Argabright, V. S., Kolodziejczak, J. J., Dunham, E. W., Geary, J. C., Tenenbaum, P., Chandrasekaran, H., Li, J., Wu, H., & von Wilpert, J. (2010, July). Kepler instrument performance: an in-flight update. In J. Oschmann Jacobus M., M. C. Clampin, & H. A. MacEwen (Eds.), *Space telescopes and instrumentation 2010: Optical, infrared, and millimeter wave* (p. 773117, Vol. 7731). <https://doi.org/10.1117/12.856638>
- Candelaresi, S., Hillier, A., Maehara, H., Brandenburg, A., & Shibata, K. (2014). Superflare Occurrence and Energies on G-, K-, and M-type Dwarfs. *The Astrophysical Journal*, 792(1), Article 67, 67. <https://doi.org/10.1088/0004-637X/792/1/67>
- Carrington, R. C. (1859). Description of a Singular Appearance seen in the Sun on September 1, 1859. *Monthly Notices of the Royal Astronomical Society*, 20(1), 13–15. <https://doi.org/10.1093/mnras/20.1.13>
- Chaudhuri, A. (2021). B-Splines. <https://arxiv.org/abs/2108.06617>
- Davenport, J. R. A. (2016). THE KEPLER CATALOG OF STELLAR FLARES. *The Astrophysical Journal*, 829(1), 23. <https://doi.org/10.3847/0004-637X/829/1/23>
- Davenport, J. R. A., Hawley, S. L., Hebb, L., Wisniewski, J. P., Kowalski, A. F., Johnson, E. C., Malatesta, M., Peraza, J., Keil, M., Silverberg, S. M., Jansen, T. C., Scheffler, M. S., Berdis, J. R., Larsen, D. M., & Hilton, E. J. (2014). KEPLERFLARES. II. THE TEMPORAL MORPHOLOGY OF WHITE-LIGHT FLARES ON GJ 1243. *The Astrophysical Journal*, 797(2), 122. <https://doi.org/10.1088/0004-637x/797/2/122>

- Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6), 381–395. <https://doi.org/10.1145/358669.358692>
- Hawley, S. L., Davenport, J. R. A., Kowalski, A. F., Wisniewski, J. P., Hebb, L., Deitrick, R., & Hilton, E. J. (2014). KEPLER FLARES. I. ACTIVE AND INACTIVE M DWARFS. *The Astrophysical Journal*, 797(2), 121. <https://doi.org/10.1088/0004-637X/797/2/121>
- Ilin, Ekaterina, Schmidt, Sarah J., Poppenhäger, Katja, Davenport, James R. A., Kristiansen, Martti H., & Omohundro, Mark. (2021). Flares in open clusters with K2 - II. Pleiades, Hyades, Praesepe, Ruprecht 147, and M 67. *Astronomy & Astrophysics*, 645, A42. <https://doi.org/10.1051/0004-6361/202039198>
- Koch, D. G., Borucki, W. J., Basri, G., Batalha, N. M., Brown, T. M., Caldwell, D., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Gautier, T. N., Geary, J. C., Gilliland, R. L., Gould, A., Jenkins, J., Kondo, Y., Latham, D. W., Lissauer, J. J., Marcy, G., ... Wu, H. (2010). KEPLER MISSION DESIGN, REALIZED PHOTOMETRIC PERFORMANCE, AND EARLY SCIENCE. *The Astrophysical Journal Letters*, 713(2), L79. <https://doi.org/10.1088/2041-8205/713/2/L79>
- Maehara, H., Shibayama, T., Notsu, S., Notsu, Y., Nagao, T., Kusaba, S., Honda, S., Nogami, D., & Shibata, K. (2012). Superflares on solar-type stars. *Nature*, 485(7399), 478–481. <https://doi.org/10.1038/nature11063>
- Mathur, S., Huber, D., Batalha, N. M., Ciardi, D. R., Bastien, F. A., Bieryla, A., Buchhave, L. A., Cochran, W. D., Endl, M., Esquerdo, G. A., Furlan, E., Howard, A., Howell, S. B., Isaacson, H., Latham, D. W., MacQueen, P. J., & Silva, D. R. (2017a). *Kepler Object of Interest (KOI) Catalog - Q1-Q17 DR 25*. Retrieved August 14, 2024, from <https://exoplanetarchive.ipac.caltech.edu/docs/Q1Q17-DR25-KOIcompanion.html>
- Mathur, S., Huber, D., Batalha, N. M., Ciardi, D. R., Bastien, F. A., Bieryla, A., Buchhave, L. A., Cochran, W. D., Endl, M., Esquerdo, G. A., Furlan, E., Howard, A., Howell, S. B., Isaacson, H., Latham, D. W., MacQueen, P. J., & Silva, D. R. (2017b). Revised Stellar Properties of Kepler Targets for the Q1-17 (DR25) Transit Detection Run. *The Astrophysical Journal Supplement Series*, 229(2), Article 30, 30. <https://doi.org/10.3847/1538-4365/229/2/30>
- McQuillan, A., Mazeh, T., & Aigrain, S. (2014). ROTATION PERIODS OF 34,030 KEPLER MAIN-SEQUENCE STARS: THE FULL AUTOCORRELATION SAMPLE. *The Astrophysical Journal Supplement Series*, 211(2), 24. <https://doi.org/10.1088/0067-0049/211/2/24>
- NASA. (2018). *Kepler / K2*. Retrieved May 24, 2024, from <https://science.nasa.gov/mission/kepler/>
- National Research Council. (2008). *Severe space weather events: Understanding societal and economic impacts: A workshop report*. The National Academies Press. <https://doi.org/10.17226/12507>

- Notsu, Y., Maehara, H., Honda, S., Hawley, S. L., Davenport, J. R. A., Namekata, K., Notsu, S., Ikuta, K., Nogami, D., & Shibata, K. (2019). Do Kepler Superflare Stars Really Include Slowly Rotating Sun-like Stars?—Results Using APO 3.5 m Telescope Spectroscopic Observations and Gaia-DR2 Data. *The Astrophysical Journal*, 876(1), Article 58, 58. <https://doi.org/10.3847/1538-4357/ab14e6>
- Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K., & Vaughan, A. H. (1984). Rotation, convection, and magnetic activity in lower main-sequence stars. *The Astrophysical Journal*, 279, 763–777. <https://doi.org/10.1086/161945>
- Okamoto, S., Notsu, Y., Maehara, H., Namekata, K., Honda, S., Ikuta, K., Nogami, D., & Shibata, K. (2021). Statistical Properties of Superflares on Solar-type Stars: Results Using All of the Kepler Primary Mission Data. *The Astrophysical Journal*, 906(2), 72. <https://doi.org/10.3847/1538-4357/abc8f5>
- Probst, S. (2023). *Investigating Pre-Flare Stellar Flux Fluctuations and Starspot Activity on Solar-Type Stars* [Bachelor's Thesis]. Astronomical Institute, University of Bern.
- Rivera, E. C., Johnson, J. R., Homan, J., & Wing, S. (2023). How noise thresholds affect the information content of stellar flare sequences. *Astronomy & Astrophysics*, 670, A143. <https://doi.org/10.1051/0004-6361/202245309>
- Rogers, L. (2011). *Early Astronomy and the Beginnings of a Mathematical Science*. Retrieved June 27, 2024, from <https://nrich.maths.org/6843>
- SDO. (2024, May). Retrieved June 27, 2024, from <https://sdo.gsfc.nasa.gov/data/>
- Shibayama, T., Maehara, H., Notsu, S., Notsu, Y., Nagao, T., Honda, S., Ishii, T. T., Nogami, D., & Shibata, K. (2013). SUPERFLARES ON SOLAR-TYPE STARS OBSERVED WITH KEPLER. I. STATISTICAL PROPERTIES OF SUPERFLARES. *The Astrophysical Journal Supplement Series*, 209(1), 5. <https://doi.org/10.1088/0067-0049/209/1/5>
- Skumanich, A. (1972). Time Scales for Ca II Emission Decay, Rotational Braking, and Lithium Depletion. *The Astrophysical Journal*, 171, 565. <https://doi.org/10.1086/151310>
- Snelling, J. M., Johnson, J. R., Willard, J., Nurhan, Y., Homan, J., & Wing, S. (2020). Information Theoretical Approach to Understanding Flare Waiting Times. *The Astrophysical Journal*, 899(2), Article 148, 148. <https://doi.org/10.3847/1538-4357/aba7b9>
- Solanki, S. K. (2003). Sunspots: An overview. *The Astronomy and Astrophysics Review*, 11(2-3), 153–286. <https://doi.org/10.1007/s00159-003-0018-4>
- Thompson, S. E., Coughlin, J. L., Hoffman, K., Mullally, F., Christiansen, J. L., Burke, C. J., Bryson, S., Batalha, N., Haas, M. R., Catanzarite, J., Rowe, J. F., Barentsen, G., Caldwell, D. A., Clarke, B. D., Jenkins, J. M., Li, J., Latham, D. W., Lissauer, J. J., Mathur, S., ... Borucki, W. J. (2017). *Kepler Object of Interest (KOI) Catalog - Q1–Q17 DR 25*. Retrieved June 27, 2024, from <https://exoplanetarchive.ipac.caltech.edu/docs/Q1Q17-DR25-KOIcompanion.html>
- Thompson, S. E., Coughlin, J. L., Hoffman, K., Mullally, F., Christiansen, J. L., Burke, C. J., Bryson, S., Batalha, N., Haas, M. R., Catanzarite, J., Rowe, J. F., Barentsen, G., Caldwell, D. A., Clarke, B. D., Jenkins, J. M., Li, J., Latham, D. W., Lissauer,

- J. J., Mathur, S., ... Borucki, W. J. (2018). Planetary Candidates Observed by Kepler. VIII. A Fully Automated Catalog with Measured Completeness and Reliability Based on Data Release 25. *The Astrophysical Journal Supplement Series*, 235(2), Article 38, 38. <https://doi.org/10.3847/1538-4365/aab4f9>
- Van Cleve, J. (2009). *Kepler Instrument Response Function (high resolution)*. Retrieved June 28, 2024, from <https://nexsci.caltech.edu/workshop/2012/keplergo/CalibrationResponse.shtml>
- Van Cleve, J. E., & Caldwell, D. A. (2016, April). Kepler Instrument Handbook.
- VanderPlas, J. T., & Ivezić, Ž. (2015). Periodograms for Multiband Astronomical Time Series. *The Astrophysical Journal*, 812(1), Article 18, 18. <https://doi.org/10.1088/0004-637X/812/1/18>
- VanderPlas, J. (2015). *gatspy: General tools for Astronomical Time Series in Python*. Retrieved August 16, 2024, from <http://www.astroml.org/gatspy/>
- Vida, K., Bódi, A., Szklenár, T., & Seli, B. (2021). Finding flares in Kepler and TESS data with recurrent deep neural networks. *Astronomy & Astrophysics*, 652, A107. <https://doi.org/10.1051/0004-6361/202141068>
- Vida, K., & Roettenbacher, R. M. (2018). Finding flares in Kepler data using machine-learning tools. *Astronomy & Astrophysics*, 616, A163. <https://doi.org/10.1051/0004-6361/201833194>
- Walkowicz, L. M., Basri, G., Batalha, N., Gilliland, R. L., Jenkins, J., Borucki, W. J., Koch, D., Caldwell, D., Dupree, A. K., Latham, D. W., Meibom, S., Howell, S., Brown, T. M., & Bryson, S. (2011). WHITE-LIGHT FLARES ON COOL STARS IN THE KEPLER QUARTER 1 DATA. *The Astronomical Journal*, 141(2), 50. <https://doi.org/10.1088/0004-6256/141/2/50>
- Yang, H., Liu, J., Gao, Q., Fang, X., Guo, J., Zhang, Y., Hou, Y., Wang, Y., & Cao, Z. (2017). The Flaring Activity of M Dwarfs in the Kepler Field. *The Astrophysical Journal*, 849(1), Article 36, 36. <https://doi.org/10.3847/1538-4357/aa8ea2>

A Flare Energy Calculation

The following text, calculations and plots are based on private notes from Prof. Lucia Kleint.

A.1 Kepler Respond Function

Since the Kepler telescope is only sensitive to specific wavelengths given by the Kepler Response Function (figure 22) the observed energies will differ from those expected from a black body (figure 23). However, this can be neglected as only the ratio between the non-flaring and the flaring star is needed for the energy calculation.

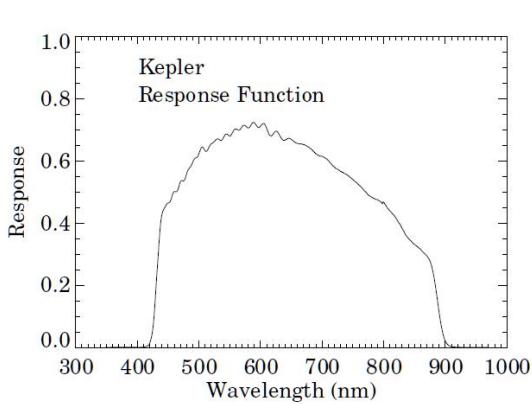


Figure 22: Kepler Response Function with courtesy to the Kepler Instrument team (J. Van Cleve, 2009).

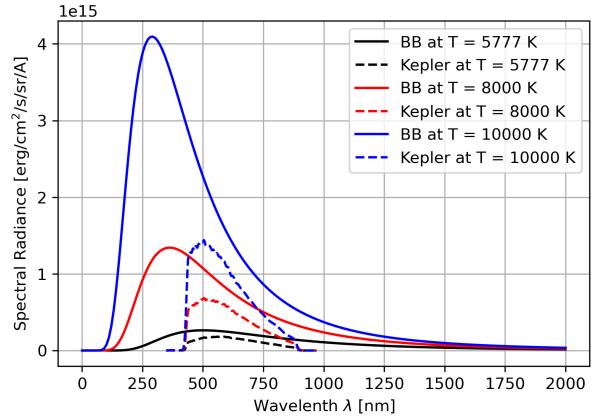


Figure 23: Difference between what Kepler sees (dashed) and the spectral radiance of a body with the same wavelength and temperature for a black body (solid).

A.2 Temperature and Flare Area Dependence

The luminosity L of a quiet star and the flare can be calculated using the formula:

$$L_{\text{star}} = A_{\text{star}} \cdot \int R_{\lambda} \cdot B_{\lambda}(T_{\text{star}}) d\lambda, \quad (8)$$

$$L_{\text{flare}} = A_{\text{flare}} \cdot \int R_{\lambda} \cdot B_{\lambda}(T_{\text{flare}}) d\lambda, \quad (9)$$

where A_{star} and A_{flare} are the stars resp. the flaring area (see figure 24), R_{λ} is the Kepler response function (Bryson et al., 2010; Caldwell et al., 2010; J. Van Cleve, 2009) and B_{λ} is the Plank function at a given wavelength.

During a flare, the area of the quiet star is A_{QS} and the area of the flare is A_{flare} . In percentages of the total area A_{star} we write $p_{\text{QS}} = \frac{A_{\text{QS}}}{A_{\text{star}}} = \frac{A_{\text{star}} - A_{\text{flare}}}{A_{\text{star}}}$ and $p_{\text{flare}} = \frac{A_{\text{flare}}}{A_{\text{star}}}$. The

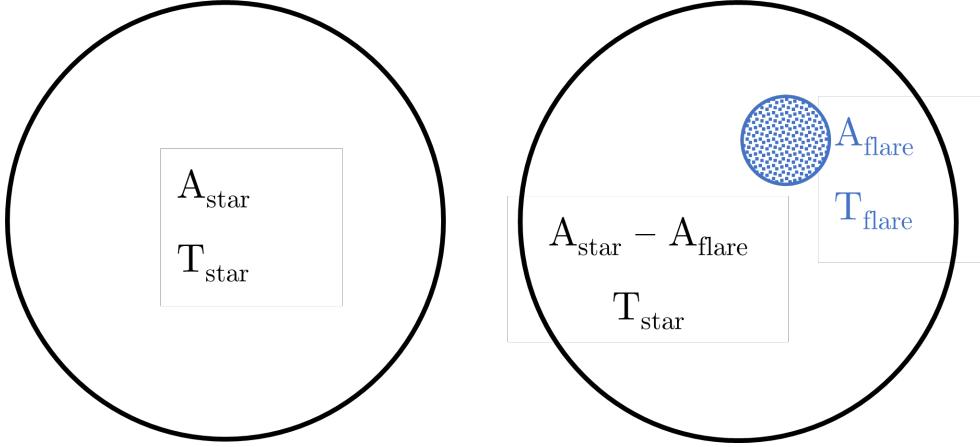


Figure 24: Star without (left) and with flare (right). The area of the quiet star changes with the size of the flare.

luminosity in the Kepler transmission band for different temperatures and areas can be calculated for a flaring star as

$$\begin{aligned} L_{\text{FS}} &= p_{\text{QS}} \cdot A_{\text{star}} \cdot \int R_\lambda \cdot B_\lambda(T_{\text{star}}) d\lambda + p_{\text{flare}} \cdot A_{\text{star}} \cdot \int R_\lambda \cdot B_\lambda(T_{\text{flare}}) d\lambda \\ &= (A_{\text{star}} - A_{\text{flare}}) \cdot \int R_\lambda \cdot B_\lambda(T_{\text{star}}) d\lambda + A_{\text{flare}} \cdot \int R_\lambda \cdot B_\lambda(T_{\text{flare}}) d\lambda. \end{aligned} \quad (10)$$

The difference in luminosity - or flux F , since $L = F \cdot A$ can now be written as

$$\begin{aligned} \frac{\Delta F}{F_{\text{star}}} &= \frac{F_{\text{FS}} - F_{\text{star}}}{F_{\text{star}}} = \frac{L_{\text{FS}} - L_{\text{star}}}{L_{\text{star}}} \\ &= \frac{A_{\text{flare}} \cdot \int R_\lambda \cdot B_\lambda(T_{\text{flare}}) d\lambda - A_{\text{flare}} \cdot \int R_\lambda \cdot B_\lambda(T_{\text{star}}) d\lambda}{A_{\text{star}} \cdot \int R_\lambda \cdot B_\lambda(T_{\text{star}}) d\lambda} \\ &= \frac{A_{\text{flare}}}{A_{\text{star}}} \cdot \frac{\int R_\lambda \cdot B_\lambda(T_{\text{flare}}) d\lambda - \int R_\lambda \cdot B_\lambda(T_{\text{star}}) d\lambda}{\int R_\lambda \cdot B_\lambda(T_{\text{star}}) d\lambda}. \end{aligned} \quad (11)$$

A.3 Flare Energy

The total energy of the flare is the integral of $L_{\text{flare}}(t)$ over the flare duration

$$E_{\text{flare}} = \int_{t_{\text{start}}}^{t_{\text{end}}} L_{\text{flare}}(t) dt, \quad (12)$$

where the luminosity of the flare L_{flare} depends on the flare temperature T_{flare} and the flaring area A_{flare} . Assuming a black-body emitting star

$$L_{\text{flare}}(t) = \sigma_{\text{SB}} \cdot T_{\text{flare}}^4 \cdot A_{\text{flare}}(t), \quad (13)$$

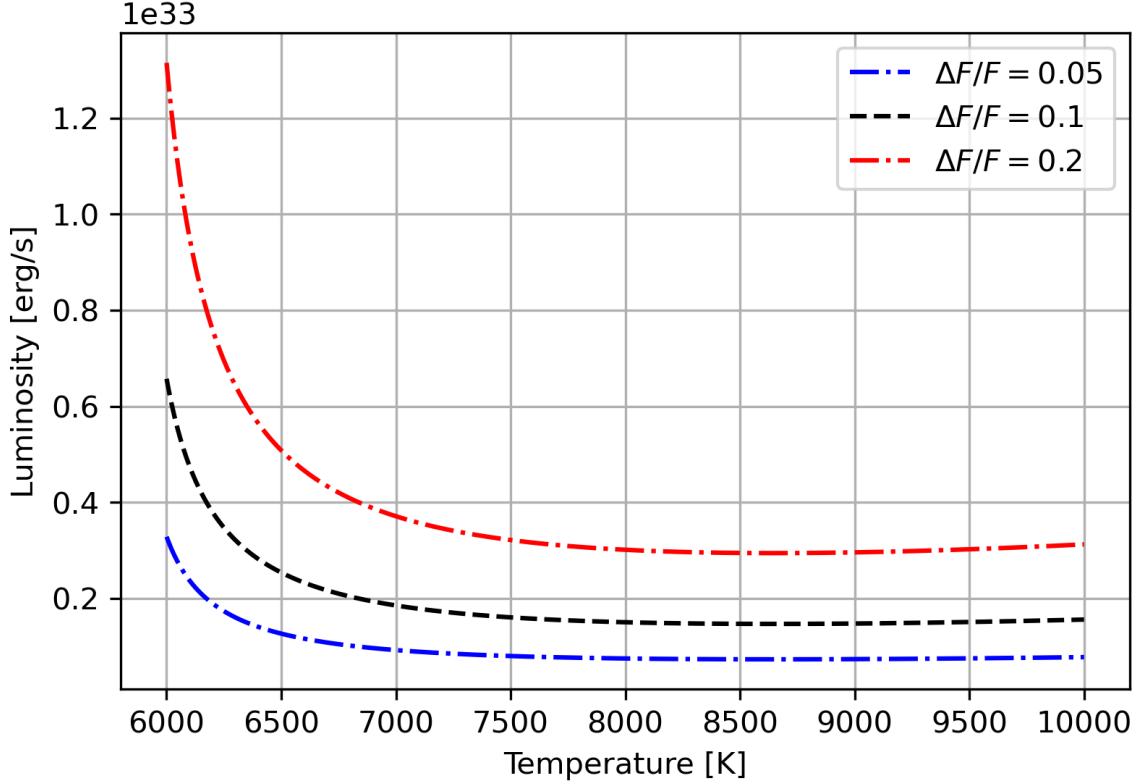


Figure 25: Luminosity for a solar flare for different flare temperatures and fixed amplitude.

where σ_{SB} is the Stefan-Boltzmann constant. From equation 11 we can write the flaring area as

$$A_{\text{flare,new}}(t) = \frac{\Delta F}{F_{\text{star}}} \cdot A_{\text{star}} \cdot \frac{\int R_{\lambda} \cdot B_{\lambda}(T_{\text{star}}) d\lambda}{\int R_{\lambda} \cdot B_{\lambda}(T_{\text{flare}}) d\lambda - \int R_{\lambda} \cdot B_{\lambda}(T_{\text{star}}) d\lambda}. \quad (14)$$

With $\frac{\Delta F}{F_{\text{star}}}$ being the amplitude of the normalised flux of the flare (see equation 1), the only remaining unknown is the temperature of the flare.

A.4 Flare Temperature

For a given amplitude of the normalised flux $\frac{\Delta F}{F_{\text{star}}}$ the luminosity of the flare can be calculated for different temperatures. Figure 25 shows the calculated luminosity for a solar flare ($T_{\text{star}} = 5777$ K and $R_{\text{star}} = 6.96342 \cdot 10^{10}$ cm) for different different flare temperatures and fixed amplitude. As can be seen, there is not much difference in luminosity for a flare temperature between 8'000 and 10'000 K. Therefore, the difference between calculations of the flare energy by Okamoto using 10'000 K for the flare temperature and Althukair using 9'000 K should be minimal.

A.5 Changes to Althukair's formulae

Althukair and Tsiklauri (2023) calculated the area of the flare with the following equation (equation (5) from their paper, rewritten with the notation used for this thesis):

$$A_{\text{flare,old}}(t) = \frac{\Delta F}{F_{\text{star}}} \cdot A_{\text{star}} \cdot \frac{\int R_\lambda \cdot B_\lambda(T_{\text{star}}) d\lambda}{\int R_\lambda \cdot B_\lambda(T_{\text{flare}}) d\lambda}. \quad (15)$$

The difference between this equation and equation 14 occurred since the change in the stars quiet area was not considered. To calculate the formula used by Althukair and Tsiklauri (2023) ($A_{\text{star}} - A_{\text{flare}}$) needs to be changed to (A_{star}) in equation 10.

The difference between the luminosity calculated with each formula is

$$\frac{L_{\text{flare,old}}}{L_{\text{flare,new}}} = \frac{A_{\text{flare,old}}}{A_{\text{flare,new}}} = \frac{\int R_\lambda \cdot B_\lambda(T_{\text{flare}}) d\lambda - \int R_\lambda \cdot B_\lambda(T_{\text{star}}) d\lambda}{\int R_\lambda \cdot B_\lambda(T_{\text{flare}}) d\lambda}, \quad (16)$$

which is also shown in figure 26. For the flare temperature used by Althukair (9'000 K) the factor is about 0.8 and for the flare temperature used by Okamoto (10'000 K) the factor is about 0.85.

Since Okamoto et al. (2021) seem to use the same method for calculating the flare energy as Althukair there might be a difference in energy of approximately 6% compared to the energies in Okamotos' list. If Okamoto uses a different method this change might be even higher.

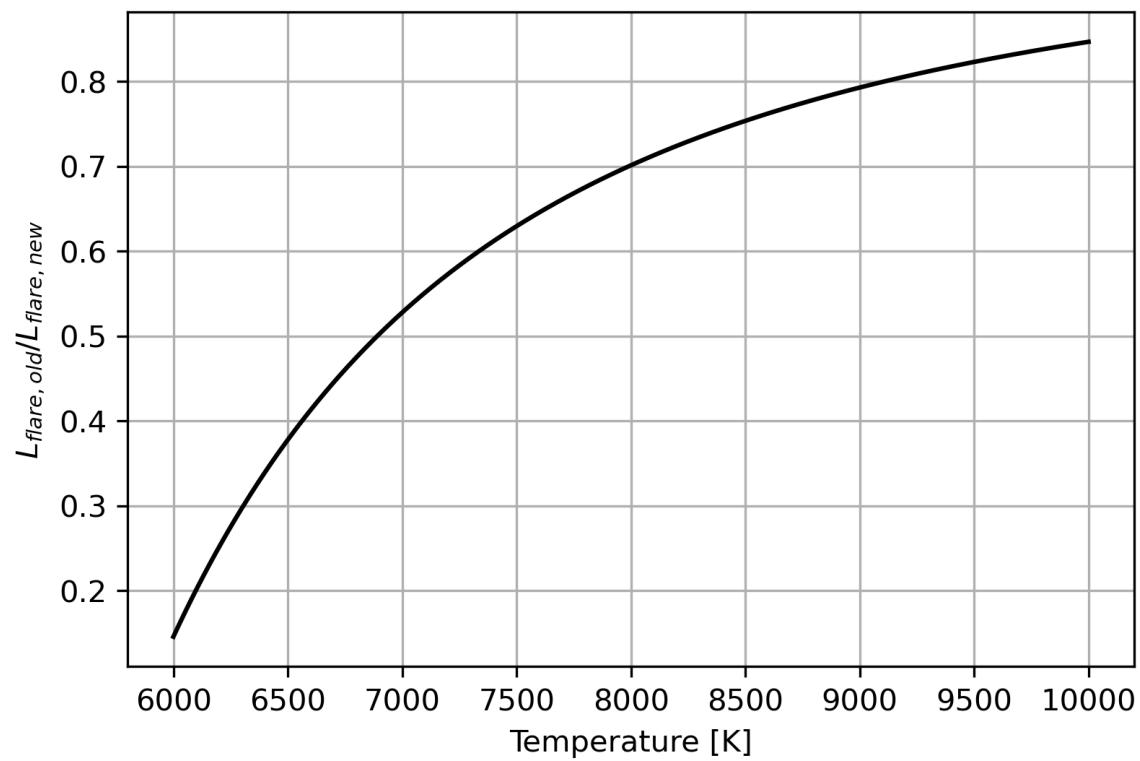


Figure 26: Difference between the flares' luminosity calculated by Althukair and Tsiklauri (2023) and with the formula calculated in this thesis.

B Note on Comparing Flares found with different surveys

Each survey returns slightly different numbers regarding the time when flares occurred. Okamoto states the flare peak time and duration, whereas both FLATW'RM and AFD return the flare's start, end and peak time (see figure 27).

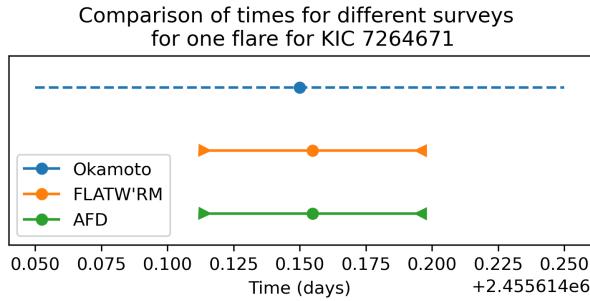


Figure 27: Plot of the different results for the flare time given by Okamoto, FLARW'RM and AFD for a flare of KIC 7264671. The circle indicates the peak time of the flare, and the triangles indicate the start and end times of the flare. For Okamoto, no such times were given, and the uncertainty given by the flare duration is indicated with a dashed line.

Thus a flare of the same KIC is common in

1. Okamoto and FLATW'RM, if the Okamoto peak time is between FLATW'RM's start and end time.
2. Okamoto and AFD, if the Okamoto peak time is between AFD's start and end time.
3. AFD and FLATW'RM, if the AFD peak time is between FLATW'RM's start and end time.
4. Okamoto, AFD and FLATW'RM, if 1 and 2 are true.

C Additional Plots

Here are some plots which are referred to in section 3.3 but would have overcrowded the results.

C.1 Wait Time

The following three figures show wait time histograms in different ranges for Okamoto, FLATW'RM and AFD.

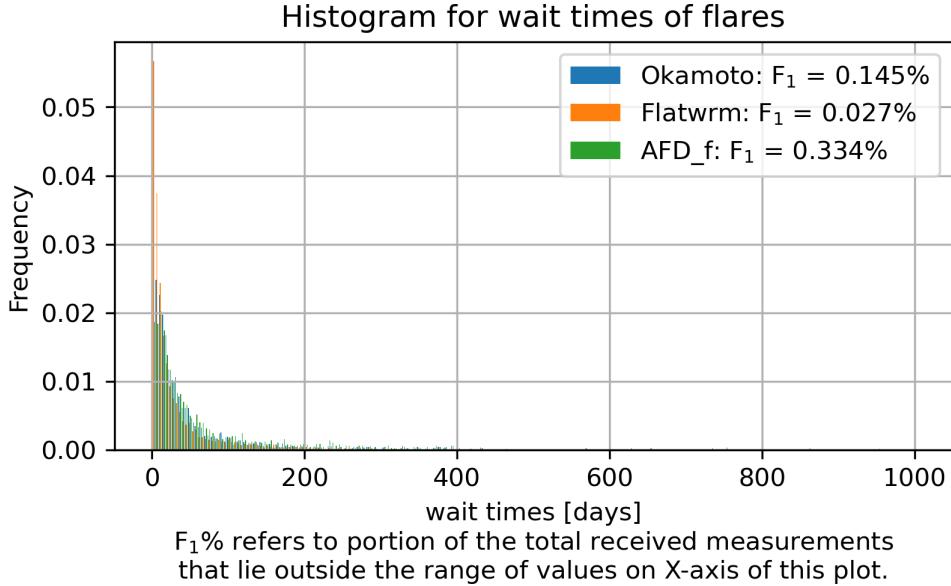


Figure 28: Normalised wait time histogram for the Okamoto, FLATW'RM and AFD flares with wait times up to 1000 days. The horizontal axis displays the wait time in days, and the vertical axis shows the number of flares found in the range of each bin. The bin size depends on the range of the timeline.

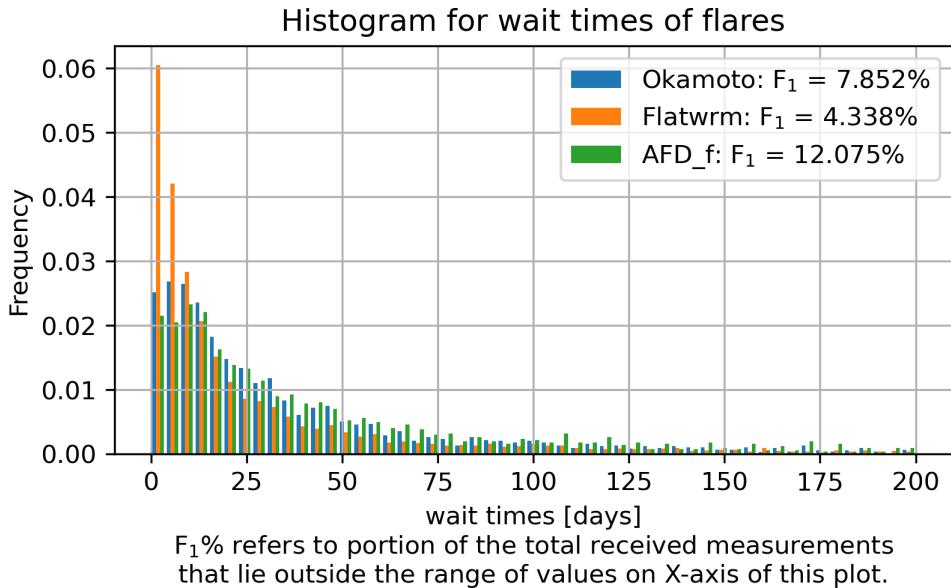


Figure 29: Normalised wait time histogram for the Okamoto, FLATW'RM and AFD flares with wait times up to 200 days. The horizontal axis displays the wait time in days, and the vertical axis shows the number of flares found in the range of each bin. The bin size depends on the range of the timeline.

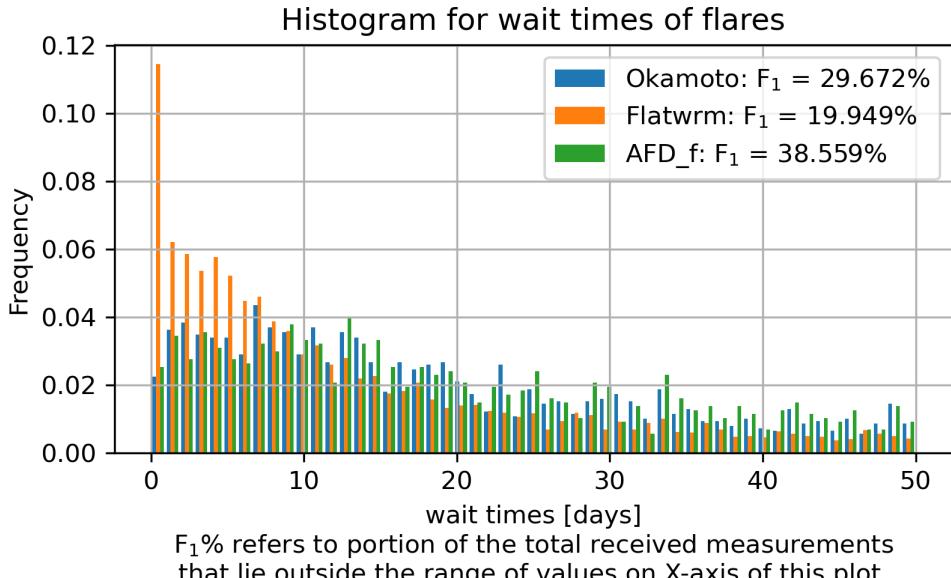


Figure 30: Normalised wait time histogram for the Okamoto, FLATW'RM and AFD flares with wait times up to 50 days. The horizontal axis displays the wait time in days, and the vertical axis shows the number of flares found in the range of each bin. The bin size depends on the range of the timeline.

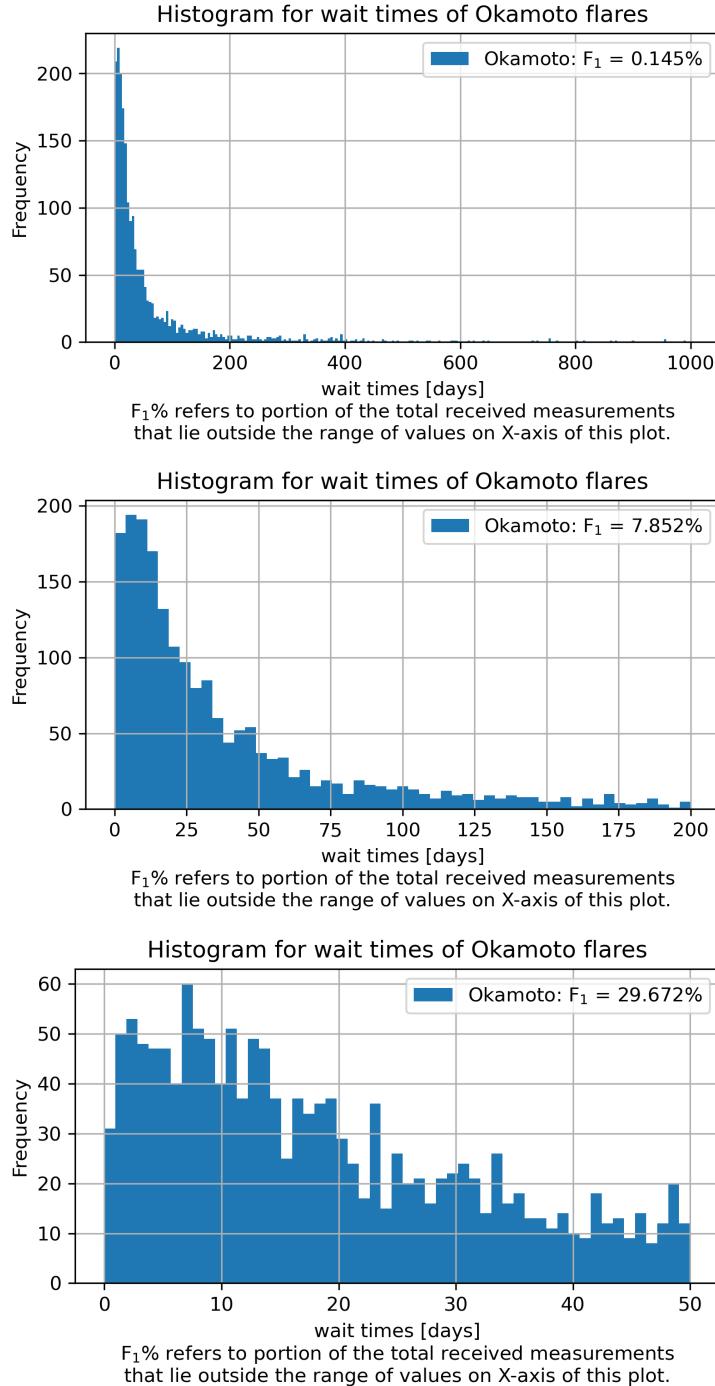


Figure 31: Wait time histograms for the Okamoto flares. The horizontal axis displays the wait time in days, and the vertical axis shows the number of flares found in the range of each bin. The bin size depends on the range of the timeline.

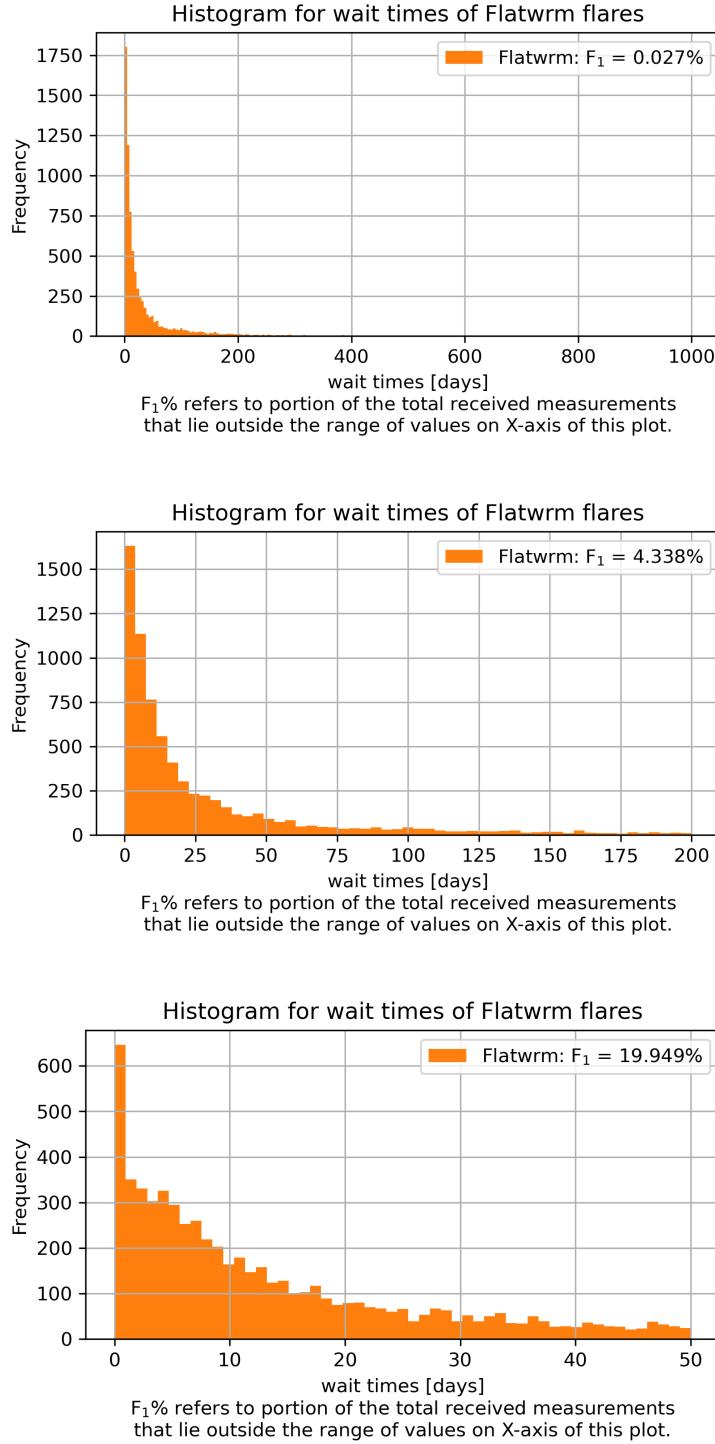


Figure 32: Wait time histograms for the FLATW'RM flares. The horizontal axis displays the wait time in days, and the vertical axis shows the number of flares found in the range of each bin. The bin size depends on the range of the timeline.

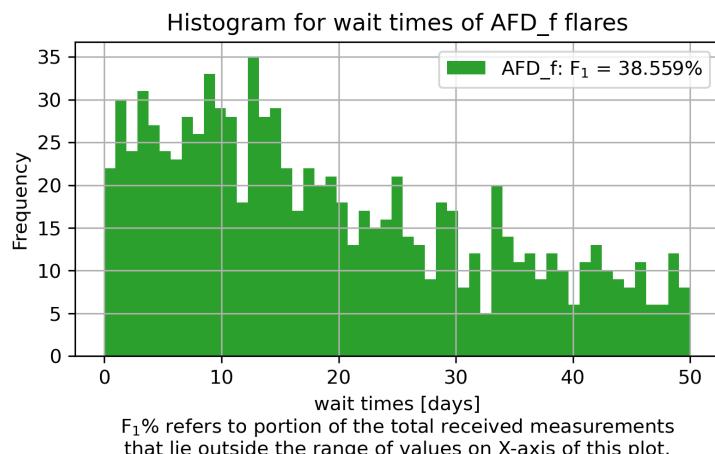
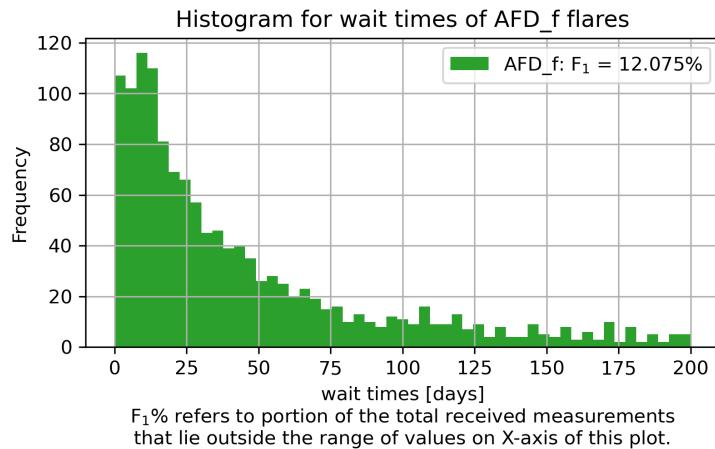
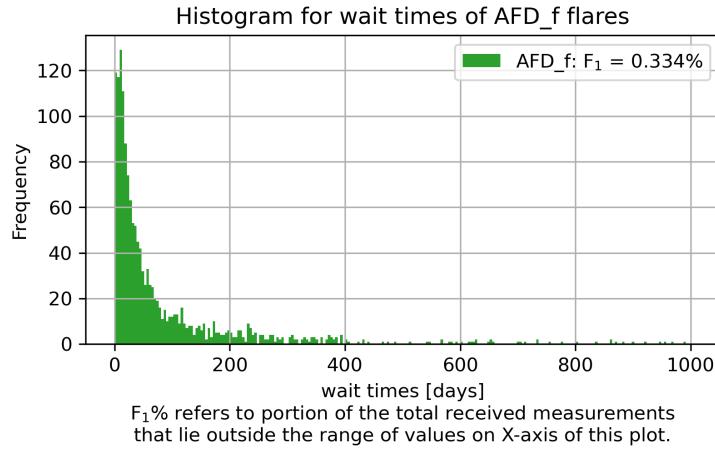


Figure 33: Wait time histograms for the AFD flares. The horizontal axis displays the wait time in days, and the vertical axis shows the number of flares found in the range of each bin. The bin size depends on the range of the timeline.

C.2 Flare Duration

The following figures shows flare duration plotted against the flare's energy for each survey individually.

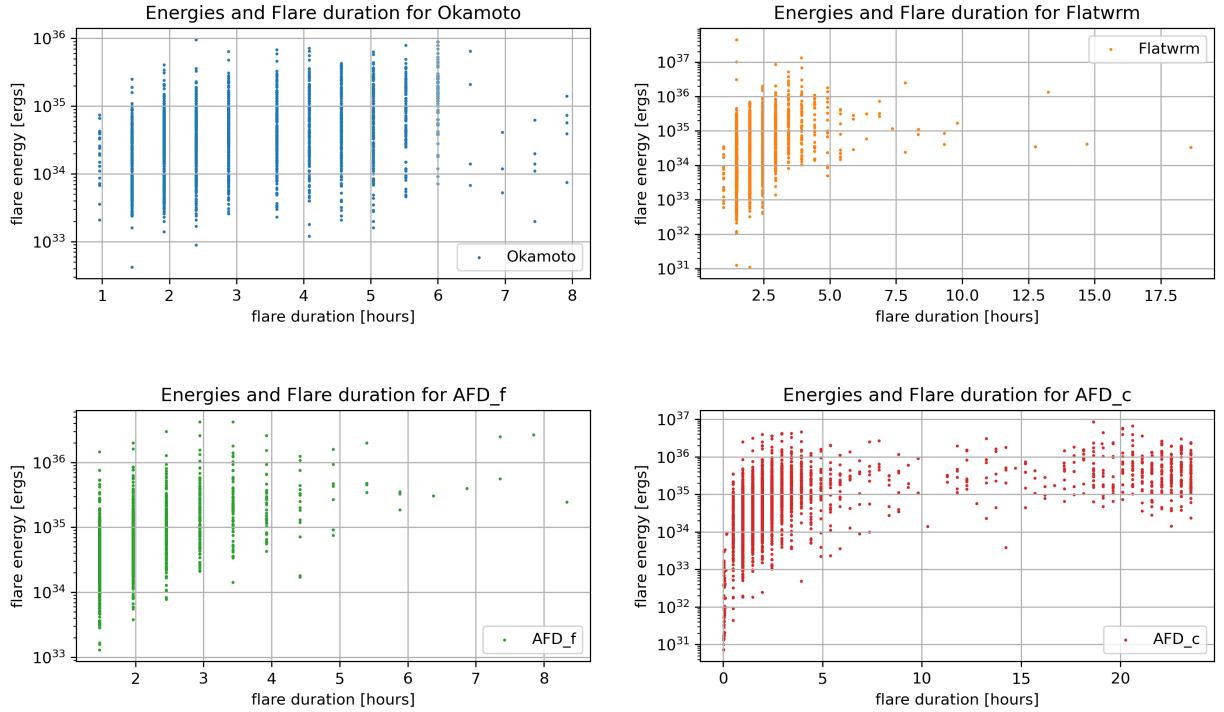


Figure 36: Plots with the flare duration on the horizontal axis and the flare energy on the vertical axis for Okamoto (blue) to FLATW'RM (orange), AFD (green) and AFD candidates (red). For the flares from the AFD candidates list the flare duration range was limited to be between zero and 1 day as there were also flares with negative lengths and extremely long flares with a duration of over 11 days.

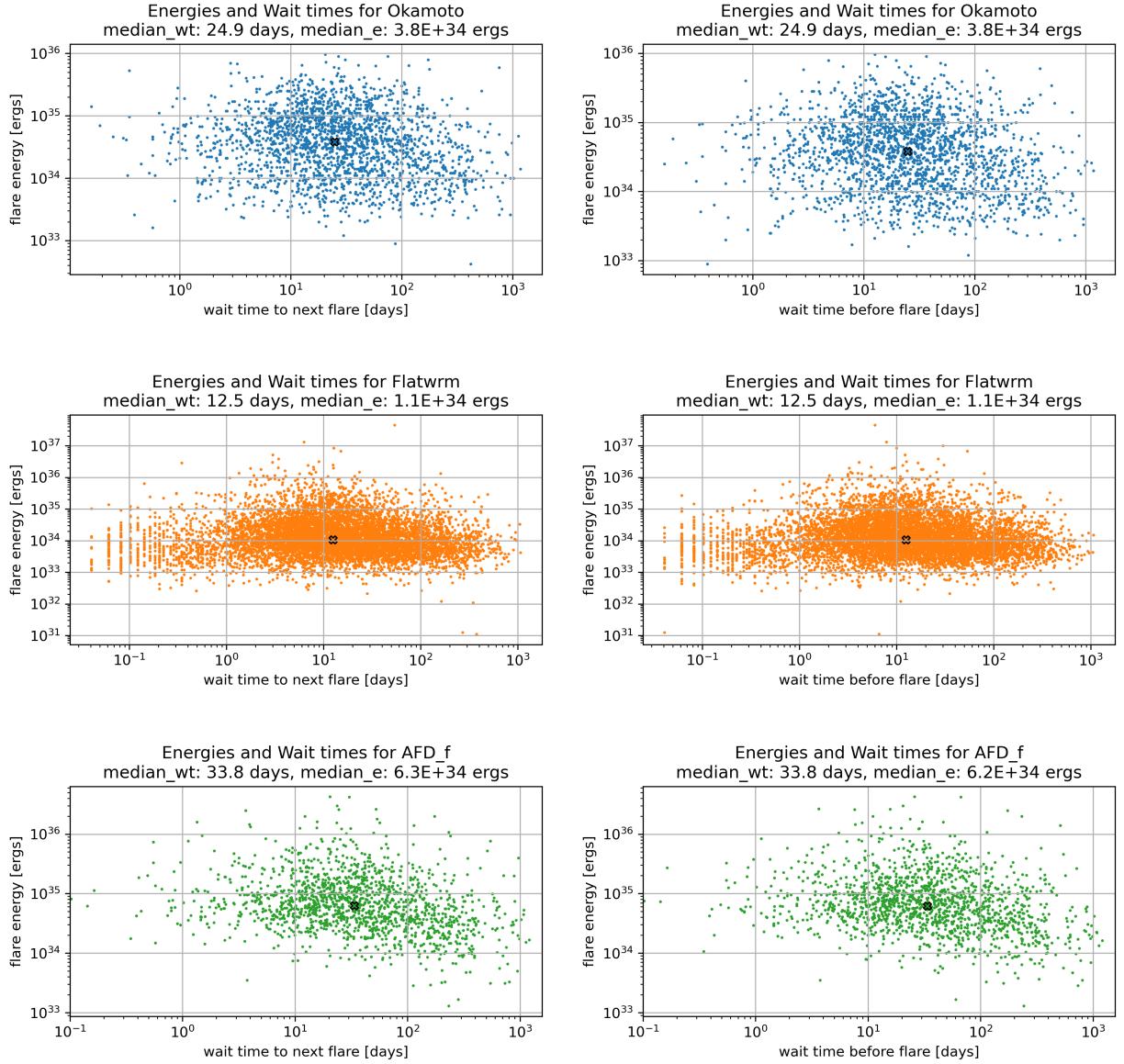


Figure 34: Plots with the wait time on the horizontal axis and the flare energy on the vertical axis for Okamoto (blue), FLATW'RM (orange) and AFD (green). On the left, the wait time to the next flare is shown, and on the right, the wait time before the flare is shown. The median in both axes is marked with an 'x'.

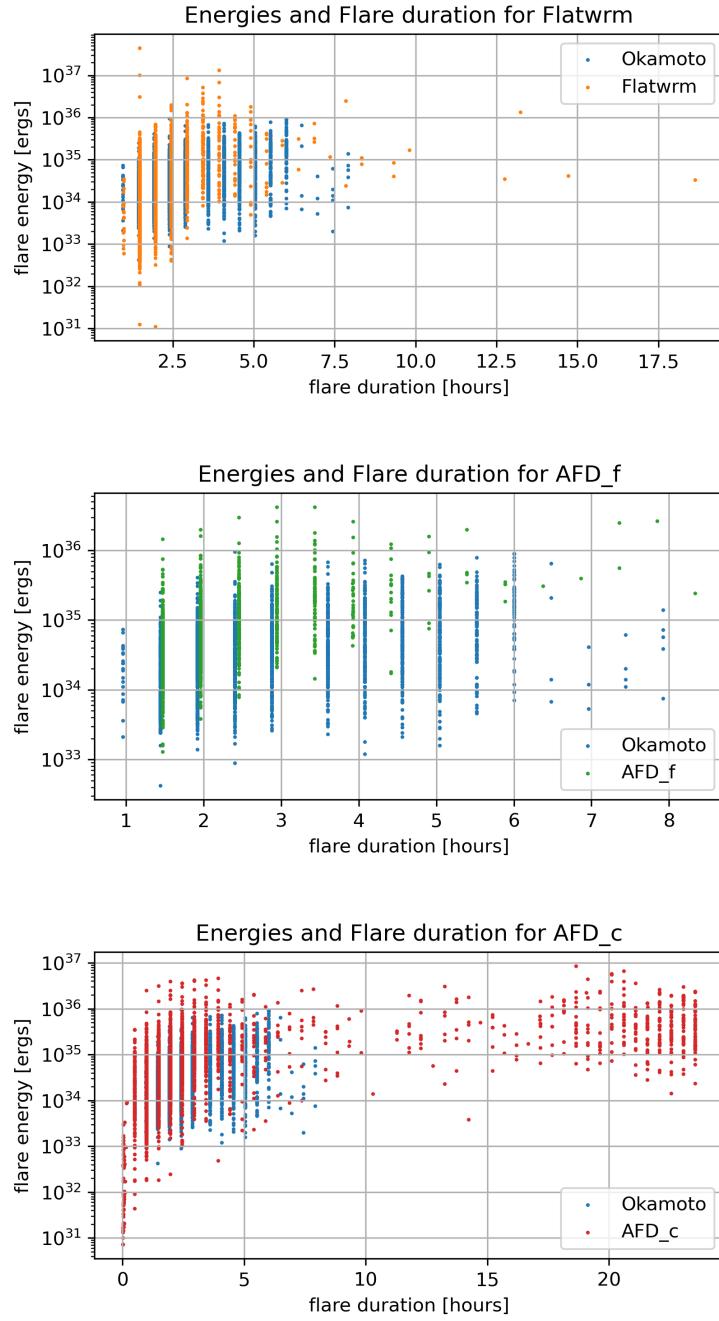


Figure 35: Plots with the flare duration on the horizontal axis and the flare energy on the vertical axis comparing Okamoto (blue) to FLATW'RM (orange), AFD (green) and AFD candidates (red). For the flares from the AFD candidates list the flare duration range was limited to be between zero and 1 day as there were also flares with negative lengths and extremely long flares with a duration of over 11 days.