Astronomy & Astrophysics manuscript no. ms_Ryanetal2016_language_corrected
June 14, 2016

Effects of flare definitions on the statistics of derived flare distributions

D. F. Ryan[1,2], M. Dominique[1], D. Seaton[1,3,4], K. Stegen[1], and A. White[5]

[1] Solar-Terrestrial Center of Excellence, SIDC, Royal Observatory of Belgium, Brussels, Belgium
[2] NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
[3] Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA
[4] NOAA National Centers for Environmental Information, Boulder, Colorado, USA
[5] School of Computer Science and Statistics, Trinity College Dublin, O'Reilly Institute, Dublin 2, Ireland

June 14, 2016

ABSTRACT

The statistical examination of solar flares is crucial to revealing their global characteristics and behaviour. Such examinations can tackle large-scale science questions or give context to detailed single-event studies. However, they are often performed using standard but basic flare detection algorithms relying on arbitrary thresholds. This arbitrariness may lead to important scientific conclusions being drawn from results caused by subjective choices in algorithms rather than the true nature of the Sun. In this paper, we explore the effect of the arbitrary thresholds used in the GOES (Geostationary Operational Environmental Satellite) event list and LYRA (Large Yield RAdiometer) Flare Finder algorithms. We find that there is a small but significant relationship between the power law exponent of the GOES flare peak flux frequency distribution and the flare start thresholds of the algorithms. We also find that the power law exponents of these distributions are not stable, but appear to steepen with increasing peak flux. This implies that the observed flare size distribution may not be a power law at all. We show that depending on the true value of the exponent of the flare size distribution, this deviation from a power law may be due to flares missed by the flare detection algorithms. However, it is not possible determine the true exponent from GOES/XRS observations. Additionally we find that the PROBA2/LYRA flare size distributions are artificially steep and clearly non-power law. We show that this is consistent with an insufficient degradation correction. This means that PROBA2/LYRA should not be used for flare statistics or energetics unless degradation is adequately accounted for. However, it can be used to study variations over shorter timescales and for space weather monitoring.

Key words. Sun: corona, Sun: flares, Sun: X-rays, gamma rays, Methods: statistical, Methods: data analysis

1. Introduction

Solar flares are amongst the most energetic events in the solar system, releasing up to 10^{32} erg in minutes or hours. They are believed to arise from a rapid re-ordering of highly stressed magnetic fields in the corona (Shibata & Magara 2011), but this process is still not fully understood. One way of better understanding the flaring process and the role it plays in the corona is to examine flares in a statistical way.

For example, the flaring process may hold the key to the coronal heating problem: the observation that the lower corona is 2–3 orders of magnitude hotter than the photosphere (∼2 MK as opposed to ∼6,000 K). It is thought that ubiquitous small-scale unresolved flaring events, known as nanoflares, may liberate enough energy from the stressed coronal magnetic field to maintain the observed high coronal temperatures. Although this is not the only proposed solution – others involve the dissipation of magnetohydrodynamic (MHD) waves – much work has been done in both modelling the required nanoflare distribution and analysing the observed flare distribution (see reviews by Klimchuk 2006; Reale 2010; Parnell & De Moortel 2012). One method has been to assume that the flare energy size distribution follows a power law which holds over all scales. The power law exponent can then be determined from larger observed flares and extrapolated to the nanoflare regime. Hudson (1991) calculated that the critical value of the power law exponent is 2. If the exponent is greater than this then there is enough energy in the nanoflare ensemble to sustain the corona's temperature. If the exponent is less than this value, then some other mechanism is required.

Another reason to explore the statistical flare distribution is to better understand the flaring process itself. One theory used to try to explain how flares occur is self-organised criticality (SOC). The first SOC models of solar flares were devised by Lu & Hamilton (1991) and Lu et al. (1993) and draw from the earlier theoretical work of Bak et al. (1987, 1988). Since then numerous variations have been devised (see review by Charbonneau et al. 2001). These models hold that solar flares are in fact an avalanche of small-scale magnetic energy release events. This avalanche can be triggered once the magnetic fields in the corona – which slowly grow in complexity owing to the jostling motions of the photospheric plasma in which they are rooted – surpass some critical instability threshold. One of the key predictions of these models is that the flare size or frequency distribution is scale invariant and characterised by a power law (Aschwanden et al. 2016). Although SOC is not the only process that can lead to power law size distri-

Article published by EDP Sciences, to be cited as http://dx.doi.org/10.1051/0004-6361/201628130

butions – others include self-organisation, inverse-cascades, stochastic relaxation, turbulence – a power law size distribution nonetheless remains a fundamental aspect of SOC models of solar flares. Therefore, fully understanding the functional form of the flare size distribution is crucial to testing these models and better understanding the flaring process.

There have been many previous studies which have examined the flare size distribution in soft X-rays (e.g. Hudson et al. 1969; Drake 1971; Shimizu 1995; Lee et al. 1995; Feldman et al. 1997; Shimojo & Shibata 1999; Aschwanden et al. 2000; Veronig et al. 2002b,a; Yashiro et al. 2006; Aschwanden & Freeland 2012). Many previous studies have examined the flare peak size distribution rather than the total energy size distribution. There are a number of reasons for this. First, no single instrument can adequately determine the energy of the entire flare. Second, the duration of some flares are not fully observed because of the instrument duty cycle or because temporally overlapping flares are very difficult to separate in spatially integrated observations. However, if it is true that the flare peak flux scales with total energy, then the peak can be used as a proxy. Under this assumption, the studies mentioned above have found power law slopes in the range 1.64–2.16 with most studies finding a slope less than 2, but not necessarily beyond statistical uncertainty. Other studies (e.g. Berghmans et al. 1998; Parnell & Jupp 2000; Aschwanden et al. 2000; Benz & Krucker 2002) have attempted to more directly observe the small-scale brightenings in the quiet Sun corona that may be responsible for coronal heating. These studies used extreme ultraviolet (EUV) imaging observations from the Transition Region And Coronal Explorer (TRACE) and/or the Extreme ultraviolet Imaging Telescope on board the Solar and Heliospheric Observatory (SOHO/EIT). The power law indices of the energy size distributions of the observed events vary considerably depending on the study. Berghmans et al. (1998) and Aschwanden et al. (2000) found power law slopes or 1.35 and 1.8, implying that coronal heating by nanoflares is highly unlikely. However, Parnell & Jupp (2000) and Benz & Krucker (2002) found power law slopes of 2.52 and 2.31, which strongly support the coronal heating by nanoflares picture. The cause of these discrepancies remains contested. Benz & Krucker (2002) attributed the shallower slopes of other studies to biases in their energy calculation and event selection criteria. In contrast, Aschwanden & Parnell (2002) claimed that the steeper slopes were due to temperature biases of the image passbands used. Thus, the importance of the flare distribution for coronal heating has not yet been conclusively answered.

To date, one aspect of these studies has rarely been discussed in detail: how does the way flares are defined affect the statistics of the derived size distribution? Solar flare event lists are typically generated by simple detection algorithms relying on arbitrary thresholds. The choice of these thresholds could affect not only the power law exponent of the derived flare size distribution, but also whether it is a power law in the first place. Moreover many studies do not provide statistical evidence that their data are truly represented by a power law. Although some studies do statistically test their data (e.g. Wheatland 2010), others only provide a visual observation of a straight-lined log-log histogram. However, it has been shown that first justifying and then fitting a power law model to data via graphical methods can be misleading (Clauset et al. 2009; D'Huys et al. 2016). In light of the importance of the power law nature of the flare distribution to flare theories (e.g. those based on self-organised criticality) and the value of the power law exponent (e.g. for coronal heating), it is crucial to properly understand the effect of flare detection thresholds and the nature of the derived flare size distribution before drawing any physical conclusions.

An example of a study which has examined the effect of event definition on the statistics of derived distributions is Buchlin et al. (2005). They constructed a number of possible types of event definitions inspired by methods used in the literature. They applied these different definitions to time series of energy dissipation from an MHD turbulence simulation. They then examined the size distributions of various properties of the detected events (total energy, peak energy, duration, etc.) and found that their derived distributions changed significantly depending on how the events were defined.

In this paper, we ask a similar question, but apply it to real flare observations and real detection algorithms. We examine the effect of the arbitrary thresholds used in the GOES event list and LYRA Flare Finder (LYRAFF) flare detection algorithms when applied to time-series of GOES/XRS (Geostationary Operational Environmental Satellite/X-Ray Sensor) and PROBA2/LYRA (PRoject for OnBoard Autonomy 2/Large Yield RAdiometer) observations. The algorithms are run on the same data sets using different threshold values. We then perform rigorous statistical analyses on the the resulting flare peak size distributions in order to determine how sensitive they are to the choice of algorithm thresholds. The results lead us to investigate the power law nature of such distributions and the suitability of both GOES/XRS and PROBA2/LYRA for statistical flare studies. Finally, possible reasons for deviations from power law distributions are discussed. In Section 2 we discuss the GOES/XRS and PROBA2/LYRA instruments and their flare detection algorithms. In Section 3 we describe our methodology, including the statistical tools used in analysing the data. In Section 4 we present the results and discuss their causes and implications. Finally, in Section 5 we present our conclusions and outline what is required to determine the nature of the solar flare size distribution.

2. Flare observations

2.1. Instrumentation

GOES/XRS (Hanser & Sellers 1996) and PROBA2/LYRA (Santandrea et al. 2013; Dominique et al. 2013) both take spatially integrated full-disk measurements of the Sun. The GOES/XRS observes in two broadband channels – 1-8 Å (long) and 0.5-4 Å (short) – with a cadence of 3 s (2 s for GOES-15). It is operated by the National Oceanographic and Atmospheric Administration (NOAA) for space weather monitoring. Its near 100% duty cycle and good calibration between satellites have caused it to become a standard candle to which more sophisticated scientific observations are typically compared. Its longevity – the first GOES/XRS was launched in the 1970s – has resulted in a nearly uninterrupted 40-year data set of flare observations and its use in numerous statistical flare studies (e.g. Lee et al. 1995; Feldman et al. 1997; Shimojo & Shibata 1999;

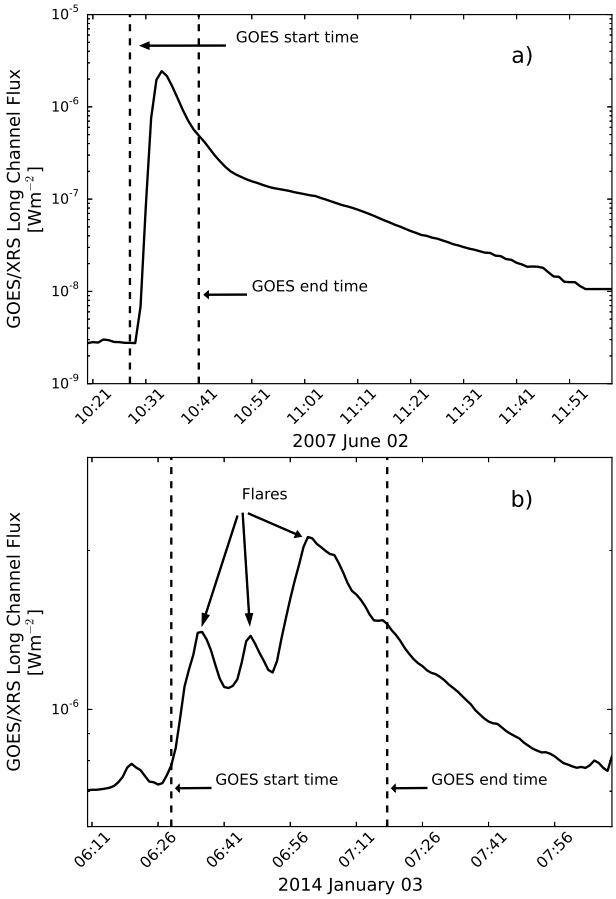

Fig. 1. a) GOES/XRS one-minute-averaged lightcurve of a C2.4-class flare on 2007 June 03. The vertical dashed lines mark the flare's start and end times according to the GOES event list. It can be seen that the flare continues for over an hour after the the official GOES end time. b) Example of three flares incorrectly labelled as one event by the GOES event list owing to crude flare definitions.

Veronig et al. 2002b; Battaglia et al. 2005; Aschwanden & Freeland 2012; Ryan et al. 2012).

PROBA2 is a technology demonstration mission, which was launched in late 2009 and is operated from the Royal Observatory of Belgium (Zender et al. 2013). The LYRA instrument is comprised of three redundant units, each of which takes measurements in four broadband channels – 120-123 nm (Lyman alpha), 190-220 nm (Herzberg), 17-80 nm + < 5 nm (Aluminium), and 6-20 nm + < 2 nm (zirconium) – with a nominal cadence of 20 Hz. However, cadence can be as high as 100 Hz during special campaigns. LYRA's nominal unit (unit 2) typically has a near 100% duty cycle making it another good candidate for monitoring flare activity. However, its duty cycle drops between late September and early March when PROBA2 experiences an eclipse season. Since first light, LYRA's nominal unit has suffered severe degradation, particularly in the Lyman alpha and Herzberg channels, which lost almost all sensitivity within a year. However, the aluminium and zirconium channels are far less degraded and their sensitivity to shorter wavelengths make them well suited to observing flares.

2.2. GOES event list

Because of their suitability for monitoring and recording solar flare activity, automatic flare detection algorithms have been developed for both the GOES/XRS and PROBA2/LYRA. They serve as useful tools for space weather monitoring and valuable resources for scientific investigation.

The GOES event list flare detection algorithm is performed on one-minute-averaged GOES/XRS long channel data. A flare is defined as having started if

1. there are four consecutive minutes of increasing flux;
2. the flux at the end of the fourth minute is at least 40% greater than the flux in the first minute.

The flare peak is defined as the maximum flux observed during the flare, while a flare is defined as having ended once the flux drops to a value halfway between the initial and peak flux values, i.e. the time at which the flux equals $F_{peak} - (F_{peak} - F_{start}) \times 0.5$, where F signifies flux.

These simple criteria make the GOES/XRS flare detection algorithm easy and fast to implement and also fairly reliable for detecting larger flares. This makes it well suited to space weather monitoring. However, the criteria's simplicity can also cause difficulties for scientific analysis. For example, the end time definition causes the flare duration to be systematically underestimated. Therefore analyses of flare properties depending on the duration, e.g. fluence, is unreliable. An example of this can be seen in Figure 1a, which shows a GOES/XRS lightcurve of a flare in the GOES event list with start and end times shown as dashed vertical lines. It can be seen that the flare carries on long after the official end time. An advantage of such an aggressive end time definition is that it enables the algorithm to better distinguish flares that occur temporally close together. Since a new flare cannot be detected until the end of the previous one is found, a more realistic end threshold would cause flares occurring close in time to be counted as a single event, thereby corrupting the flare distribution. In fact, even with this abrupt end time criterion, some events in the GOES event list are still made up of multiple flares. An example can be seen in Figure 1b, which shows three flares counted as one event. In this case the flux of the first flare does not have time to reach the GOES event list end criterion before the second and then third flares dominate the signal.

It is difficult to determine exactly how often events like the one in Figure 1b occur. Ryan et al. (2012) searched the GOES event list for instances during the decay phase of flares which met the GOES event list start criteria. These secondary flares were not included in the GOES event list as the first flare had not yet met its end criterion. They found that just over 5% of listed events between 1991 and 2007 were 'double flares', i.e. two (or more) flares listed as one. This is a lower limit on the true number of double flares. It does not take into account cases where the first flare is smaller than the second. It only includes comparably sized double flares because smaller flares would not satisfy the GOES event list start criteria when placed on the elevated background level of the already decaying earlier flare. It also does not account for flares with overlapping rise phases because they cannot be separated at all by the GOES event list. Therefore, it is expected that double flares make up a small but significant fraction of events in the GOES event list.

2.3. LYRA Flare Finder

The LYRAFF algorithm (LYRA Flare Finder) was developed to enable LYRA to automatically detect flares for space weather monitoring as part of ESA's Space Situational Awareness programme (SSA). The resulting flare list is available online[1] and is updated in near-real time. LYRAFF is very similar to the GOES event list, but has been designed to reduce the drawbacks described in Section 2.2. LYRAFF is applied to one-minute-averaged LYRA zirconium channel data. This channel was chosen because it is sensitive to the shortest wavelengths and therefore exhibits the largest variability due to flares. In addition this channel is the least degraded and so is more consistent over time.

Before flares can be sought, the numerous artefacts often found in LYRA data must be removed (see Table A.1 in Appendix A for a list and explanation of these artefacts). In addition, negative and zero irradiance values and those greater than or equal to 10 Wm^{-2} are deemed unphysical and owing to instrument error. These data points are therefore also removed. Once the data have been cleaned, LYRAFF identifies that a flare has started when

1. four consecutive minutes of increasing irradiance are observed (without data gaps). This is labelled the start threshold period;
2. the irradiance in the fourth minute is 1% greater than the irradiance in the first minute. The first of these four minutes is labelled the reference start time, while the 1% value is labelled the start threshold.

Once the reference start time is found, LYRAFF better estimates the flare's true start time by tracking back to the first minute of continuously increasing flux, whether or not criterion 2) is met. This is defined as the flare start time. The flare peak is simply defined as the time of maximum irradiance. Analogous to the reference start time, the reference end time is defined as the moment the irradiance drops to a level 50% between the flare peak and flare start time irradiance values, just as it is for the GOES event list. The value of 50% is labelled the end threshold. The flare end time is then found by tracking forward to the the time of the latest consecutive irradiance decrease. By finding the earliest/latest times of consecutive increasing/decreasing irradiance, the estimates of the flare start/end times, and hence flare duration, are significantly improved. This can be seen in Figure 2a, which shows a LYRA zirconium channel lightcurve of a flare. The reference start and end times are shown by the dashed vertical lines, while the official start and end times are shown by the solid vertical lines. It is clear that the start and end times better approximate the flare duration than the reference times.

In order to reduce cases of double flares (multiple events being incorrectly labelled as a single event), LYRAFF simultaneously searches for a new reference start time once the irradiance has dropped 20% from the peak towards the flare start irradiance. This is labelled the first fall threshold. If a new reference start time is found before the reference end time, the flare is defined as ending here and is added to the flare list and is flagged as incomplete. This technique improves the identification of flares which occur temporally close together. Figure 2b shows an example of two

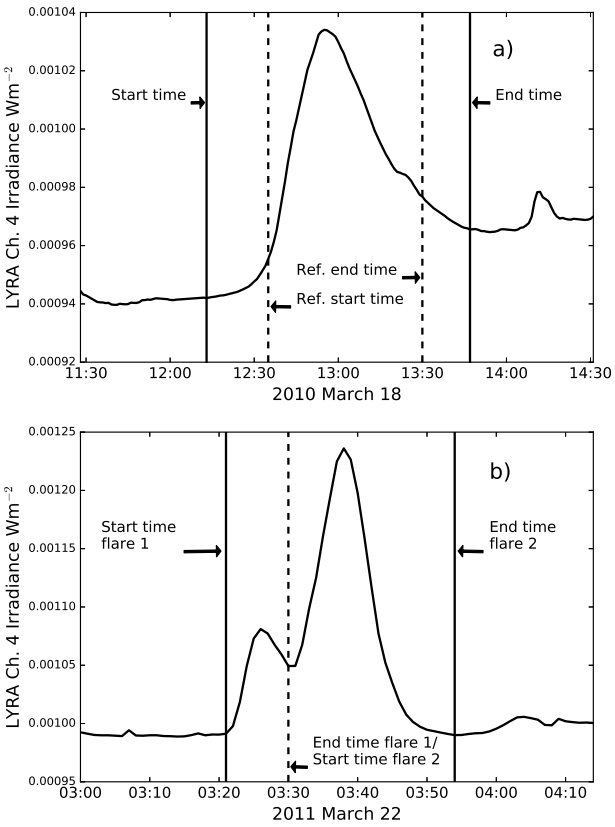

Fig. 2. a) LYRA zirconium one-minute-averaged lightcurve of a flare which occurred 2010 March 18. The start and end times identified by the LYRAFF algorithm are marked by the solid vertical lines while the reference start and end times are marked by the dashed lines. b) Example of two flares from 2011 March 22 which occurred temporally close together being successfully separated by LYRAFF.

temporally close flares being successfully separated by the LYRAFF algorithm. However, it is clear that in this case LYRAFF does not do a good job at approximating the flare duration. Therefore, when examining properties related to duration, e.g. fluence, such events must be treated with caution. Moreover, LYRAFF does not eliminate all cases of multiple flares being counted as one event. If two flares happen simultaneously or temporally close enough together that the irradiance does not drop to the 20% threshold, LYRAFF still cannot distinguish multiple events.

As a test of how effective LYRAFF is at detecting double flares, it was applied to the one-minute-averaged GOES/XRS time series for 1991 to 2007 (inclusive) by altering the start threshold to 40% (the GOES event list start threshold). This period was chosen because it was the same that Ryan et al. (2012) used to search for double flares in the GOES event list (Section 2.2). It was found that LYRAFF correctly identified and separated 799 double flares. This was 2.6% of all the flares listed in the GOES event list for this period and so is about half the percentage of double flares identified by Ryan et al. (2012). Thus, LYRAFF significantly reduces the misidentification of double flares.

[1] http://proba2.oma.be/data/LYRA/LYRAFF

3. Methodology

In order to explore how the choice of arbitrary thresholds in flare detection algorithms can affect the distributions they produce, we examined the variation in the flare peak flux frequency distribution. As we discussed in Section 1, such distributions are typically described as a power law. We therefore determined the relationships between the power law slope of the background-subtracted peak flux size distribution and the algorithm threshold used to derive it. If the power law slope of the flare peak distributions is independent of the choice of these thresholds, then statistical flare analyses based on these algorithms can be trusted. However, if this is not the case, then the effect of threshold choice must be taken into account before drawing any conclusions about the statistical nature of flares. In this section we outline the methodology followed in this study in three broad steps: producing flare lists using different algorithms and thresholds; performing a suitable background subtraction to the flare peak flux data; and fitting the flare peak flux frequency distributions with a power law.

3.1. Producing Flare Lists

First we focussed on the GOES event list and GOES/XRS data. We used the one-minute-averaged long channel GOES/XRS data produced by NOAA[2] from 1986 to 2014, inclusive. Data points marked as bad in the data files were removed, as were negative and unphysically high flux values. We then ran the GOES event list algorithm several times for values of the start threshold in the range 5%–90%[3] while keeping the end threshold fixed at its standard value of 50%. We then repeated the process for different values of end threshold in the range 5%–90% while keeping the start threshold fixed at its standard value of 40%. Thus two sets of GOES flare lists were produced reflecting the effect of the start and end thresholds.

Next we turned to LYRAFF and PROBA2/LYRA. We constructed a one-minute-averaged LYRA zirconium channel time series from January 2010 to May 2015 inclusive. The time series was cleaned as outlined in Section 2.3. Eclipse season observations (mid-September to mid-March) were also removed as attenuation by the atmosphere causes the irradiance signal to smoothly but dramatically rise and fall over the course of each orbit, making it impossible for LYRAFF to reliably detect flares. LYRAFF was then run several times on this time series using different values of start threshold in the range 0.1%–100% while keeping the first fall and end thresholds fixed at their standard values (20% and 50%, respectively). The process was repeated for different values of the first fall threshold in the range 5%–90%, while keeping the start and end thresholds fixed at their standard values (1% and 50%). The process was repeated once more for different values of end threshold in the range 5%–90%, while keeping the start and first fall thresholds fixed at their standard values (1% and 20%). This resulted in three sets of LYRA flare lists corresponding to varying start, first fall, and end thresholds, respectively.

[2] http://satdat.ngdc.noaa.gov/sem/goes/data/new_avg/

[3] For clarity, a 5% end threshold means the flare ends when the flux has drops to a level equal to $F_{peak} - (F_{peak} - F_{start}) \times 0.05$ where F is flux. Thus, a 5% end threshold results in a short flare duration, while a 90% end threshold results in a long flare duration.

Finally, we also applied LYRAFF to the GOES/XRS time series. We produced three sets of flare lists where one LYRAFF threshold was varied in the range 5%–90%, while the other two were held fixed at their standard values (40%, 20%, and 50% for the start, first fall, and end thresholds, respectively). Thus we produced eight sets of flare lists in total, each allowing us to explore the effect of a certain threshold.

3.2. Background subtraction

Before the flare peak flux size distributions could be reliably analysed, the background flux had to be subtracted. Otherwise, the derived distributions would be unphysically steep. Since the rise phase of many flares is of the order of minutes, the background typically has little chance to evolve between the start and peak of the flare. Therefore in many cases the flux at the flare start adequately approximates the background at the peak time. For some flares though, such as those occurring on the tails of other flares, this will not be the case. However, as the GOES event list does not reliably determine the flare end time, more sophisticated methods of determining the background – such as fitting a polynomial to periods of pre-flare and post-flare flux – are also likely to be unrepresentative of the true background. For flare distributions derived using LYRAFF, which does a better job of determining the flare duration, we compared the results of subtracting the initial flux and a value at the peak time found from linear interpolation between the start and end of the flare. However, we found no significant difference in the subsequent results. We therefore simply subtracted the initial flare flux for all cases.

3.3. Fitting flare distributions

Figure 3 shows a histogram of the background-subtracted flare peak flux size distribution corresponding to the standard GOES event list (start threshold=40%, end threshold=50%). It can be seen that above $\sim 10^{-6}$ Wm^{-2} the distribution appears be a power law. The roll-over below this level is caused by undersampling of events owing to the variation in the solar background, which typically reaches low C-class at solar maximum. All GOES flare peak flux frequency distributions derived in this study exhibited this same qualitative behaviour.

In order to quantitatively determine the effect of the different flare algorithm thresholds, we fit each distribution with a power law and examined the variation of the exponent. To do this we used maximum likelihood estimation (MLE), which is a standard statistical technique for fitting parametric models to empirical data. It is far more robust and easy to use than graphical methods such as fitting a linear model to a log-log histogram. For further discussion of the advantages of MLE see Clauset et al. (2009) and specifically in the context of solar physics, see D'Huys et al. (2016).

Given a parametric model (e.g. a power law) described by a parameter, α, and a set of observations of a property, x, it is possible to determine the likelihood function, $L(\alpha, x)$, which describes the probability that the data is drawn from a distribution with a given α value. According to MLE, the maximum of this likelihood function gives the most likely estimate of α (denoted $\hat{\alpha}$), i.e. the best fit. This

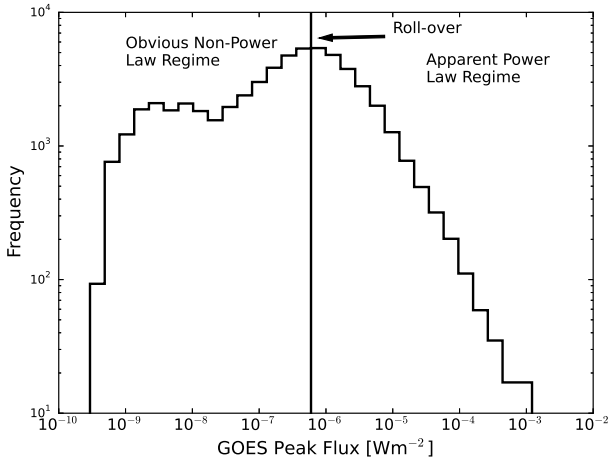

Fig. 3. Frequency distribution of the background-subtracted GOES long channel peak flux for flares detected using the standard GOES event list algorithm (start threshold=40%, end threshold=50%). Above the roll-over the distribution appears to follow a power law, while below the roll-over it deviates owing to undersampling of the true flare distribution.

can simply be determined by differentiating the likelihood function, setting it to zero, and solving for α. Therefore, given a power law of the form

$$f(x) \propto x^{-\alpha} \tag{1}$$

the most likely estimate of α is given by

$$\hat{\alpha} = 1 + n_{tail} \left(\sum_{i=1}^{n_{tail}} ln \frac{x_i}{x_{min}} \right)^{-1}, \tag{2}$$

where x_{min} is the lower limit above which the data obey a power law (e.g. the roll-over in Figure 3); $x_{i=1,...,n_{tail}}$ is the subsample of data, where $x_i \geq x_{min}$; and n_{tail} is the number of data points in the subsample $x_{i=1,...,n_{tail}}$. To a first-order approximation, the statistical error on $\hat{\alpha}$ is then given by

$$\sigma = \frac{\hat{\alpha} - 1}{\sqrt{n_{tail}}}. \tag{3}$$

For a derivation of Equations 2 and 3, see Appendix B of Clauset et al. (2009).

In order to obtain a reliable fit via MLE, a good estimation of x_{min} must first be made. There are a number of ways of doing this. It can be approximated by eye from histograms. However, this depends on the arbitrary choice of bin width and the scientist's subjective choice of where the histogram deviates from a power law. Therefore, in order to determine the lower limit more robustly, we performed the MLE power law fit for x_{min} equal to each flux value in the sample. We then selected the x_{min} which produced the lowest Kolmogorov-Smirnov (KS) statistic[4]. Using this method we fit power laws to the peak flux/irradiance frequency distributions of all the flare lists derived in Sections 3.1 and 3.2.

It should be pointed out that MLE is not the only viable fitting technique available. Other studies (e.g. Ireland et al. 2015) have used Markov chain Monte Carlo (MCMC) methods to fit power laws to solar observational data. The advantages of MCMC methods are that they provide a more rigorous determination of the fit uncertainties and tend to be more robust in analysing complex likelihood functions that may include numerous local maxima. However, for well-behaved likelihood functions, MLE is expected to produce similar results to MCMC. What is more, MCMC methods come at an additional computational cost. In their analysis of the Fourier power spectra of the dynamics of different regions of the solar atmosphere, Ireland et al. (2015) ran each MCMC chain for 150,000 samples. In our case, implementing this procedure for over 100 separate data sets would be prohibitive. We therefore chose MLE as our preferred fitting technique for this study. In the next section we discuss the results of our analysis.

4. Results and discussion

4.1. Effect of arbitrary thresholds on GOES-derived flare distributions

Figure 4 shows the relationships between the thresholds used to define GOES flares and the power law exponent of the derived GOES flare peak flux size distribution. The power law exponents of each distribution were obtained using x_{min} in the range $5 \pm 1 \times 10^{-6}$ Wm^{-2}. The mathematical justification for this choice is that an x_{min} in this range gives the lowest KS statistic. The physical justification is that this is the maximum level the solar X-ray background can reach during solar max. This causes the flare detection algorithms to undersample the true flare distribution, thereby making the statistics unreliable below this level.

Figure 4a shows the variation of the power law exponent with the GOES event list start threshold. A shallow but clear decreasing dependence is evident. For a start threshold of 5% the exponent is greater than 2. As the threshold is increased the exponent decreases steadily until it reaches ∼1.92 for start thresholds of 90%. Moreover this trend is clearly beyond the range of statistical uncertainty. We might naively expect this result if we used crude fitting techniques and believed that the change in start threshold simply caused the detection algorithm to undersample the flare population below a given peak flux. This would only cause the lower section of the distribution to deviate from a power law and would hence flatten the average slope. However, by robustly determining x_{min} and excluding events below this value, we would expect the fitted power law slope not to change as the larger flares should be common to all flare distributions. However, this is not what we see in Figure 4a. We see that despite robustly determining x_{min}, the start threshold still affects the power law slope. This occurs because the start threshold defines a flare based on its impulsiveness rather than its peak flux. Therefore, some larger flares above the best found value of x_{min} may occasionally be missed because they are not impulsive enough. This shows that – depending on how the start of a flare is defined – different flare distributions can be obtained.

[4] The KS statistic is a standard measure of how well an empirical distribution is described by a fit. It is given by the maximum difference between the empirical cumulative distribution function (CDF) of the data and the theoretical CDF of the fit. The smaller the KS-statistic, the more closely the fit approximates the data.

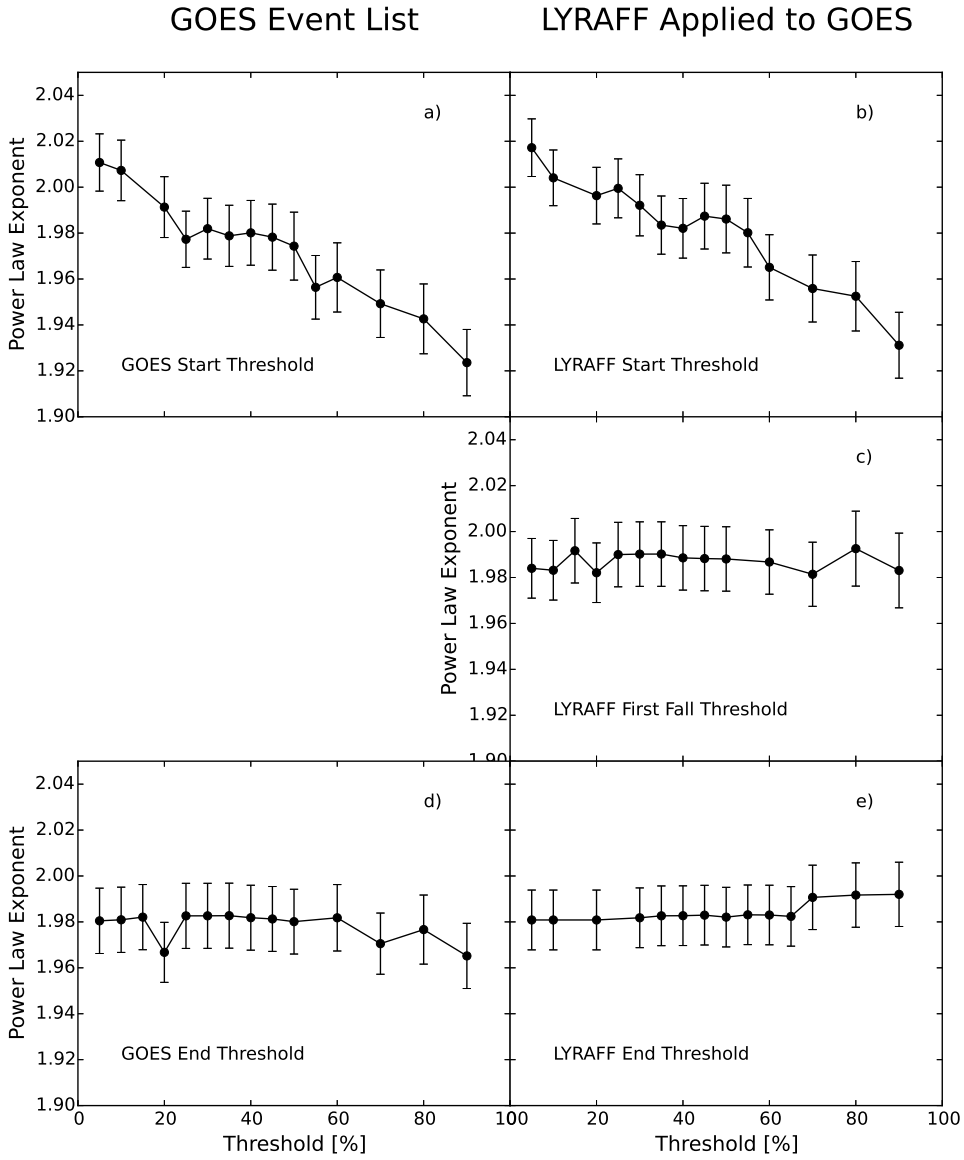

Fig. 4. Power law slope of the background-subtracted GOES long channel peak flux frequency distribution as a function of the GOES event list and LYRAFF thresholds. The left column corresponds to the thresholds of the GOES event list algorithm, while the right column corresponds to the LYRAFF algorithm applied to GOES/XRS observations. Each row then corresponds to the start (top), first fall (middle), and end (bottom) thresholds. We note that there is no plot in the middle left panel as the GOES event list does not use a first fall threshold. The error bars represent the statistical uncertainty of the power law fits given by Equation 3. We note that there is a clear downward trend in both start threshold plots implying that the derived flare distribution depends on the choice of start threshold. In contrast, the first fall and end threshold panels show flat relationships implying that the derived flare distribution is independent of the choice of these thresholds.

To verify this result, we examined the same relationship for the LYRAFF algorithm (Figure 4b). The improved ability of LYRAFF to detect temporally nearby flares helps determine whether the above result is owing to incorrect classifications of multiple events or aggressive start and end time definitions. However, a very similar behaviour is observed. Although the decreasing trends in Figures 4a and 4b are small, they could be significant. For example, when trying to determine whether nanoflaring can explain the coronal heating problem by extrapolating the flare distribution, the key value of the power law slope is 2. If we naively define a flare with a 5% or 50% start threshold, we could come to totally opposite conclusions. Therefore, it is vital to fully understand the effect of the chosen flare definitions before interpreting the results.

Despite the similarity between Figures 4a and 4b, the results are not exactly the same. With the exception of the 10% start threshold, the LYRAFF exponents are all greater than their GOES event list counterparts for the same threshold by between 0.1% and 1.2%. This is a very small difference and within the range of uncertainty. Nonetheless, a slightly higher exponent is expected when employing LYRAFF. As discussed in Section 2.3, LYRAFF successfully identified and separated 2.5% of the flares in the standard GOES event list as double flares, using the standard GOES thresholds (start threshold=40%, first fall thresh-

old=20%, end threshold=50%). These additional flares are the main difference between the flare lists derived by applying the GOES event list and LYRAFF algorithms to GOES/XRS time series. The peak flux frequency distribution of these additional flares has a best fit power law exponent of 2.59±0.15 above the x_{min} value of 5×10^{-6} Wm^{-2} commonly used in this study. This is significantly steeper than the corresponding LYRAFF exponent in Figure 4b as the distribution of the additional flares is highly skewed to smaller events. This occurs because it is the smaller flare that is masked from the peak flux size distribution when a double flare occurs. Therefore, LYRAFF's ability to better separate double flares is what leads to marginally steeper flare distributions. However, as also highlighted in Section 2.3, LYRAFF cannot identify all temporally overlapping flares. If more of these flares could be identified, it would only serve to further steepen the overall flare size distribution. This shows the small but important impact that correctly identifying temporally overlapping flares can have.

Next, we examined the effect of the first fall and end thresholds of the GOES event list and LYRAFF algorithms, seen in Figure panels 4c–4e. More promisingly, flat relationships around 1.98–1.99 were found. This demonstrates that the derived flare distribution is not influenced by the choice of first fall or end thresholds in either algorithm. However, it must be noted that we are only looking at peak flux distributions which are independent of flare duration. If we were to examine fluence, the choice of first fall and end thresholds and the effectiveness of our algorithm to accurately determine start and end times would play a very important role.

4.2. Stability of the power law nature of flare distributions

4.2.1. Stability plots

In order to further investigate the reliability of the power law fits discussed in Section 4.1, we examined the stability of the MLE-derived power law exponent as a function of x_{min}. Figure 5a shows a cartoon of a stability plot, i.e. the dependence of the derived fit on the choice of x_{min}, for a hypothetical stable power law fit (for similar plots, see Clauset et al. 2009). There are three regimes in a stability plot for a reliable MLE-derived fit. At low values of x_{min} there is an increasing relationship because x_{min} is lower than the roll-over (Figure 3) caused by the undersampling of events. As x_{min} is decreased below the roll-over, the fit is made artificially shallower. At high x_{min} there is another steeper increasing relationship because the choice of x_{min} is now so high that we only have a few events in our sample which is insufficient for fitting a power law. At intermediate values of x_{min} the relationship is flat. This means that within this range the derived power law exponent is independent of the choice of x_{min} since the distribution is now adequately sampled. However, as the choice of x_{min} is increased, the number of events in the sample is decreased, causing the statistical uncertainty of the fit to increase. Therefore, it is important to choose the lowest value of x_{min} for which the fit is stable. If the dependence of the derived power law exponent on x_{min} exhibits this flat regime, we say that the power law fit is stable.

Figure 5b shows the stability plot for the GOES peak long channel flux size distribution derived using the standard GOES event list definitions (start threshold=40%, end threshold=50%). It can be seen that the intermediate flat regime is absent. The stability plots of all the derived flare peak flux size distributions in this study show the same qualitative behaviour. This implies that these distributions are not well fit by a single power law. This is in contrast to numerous studies over the past decades which have treated the distribution of peak GOES long channel flux as such.

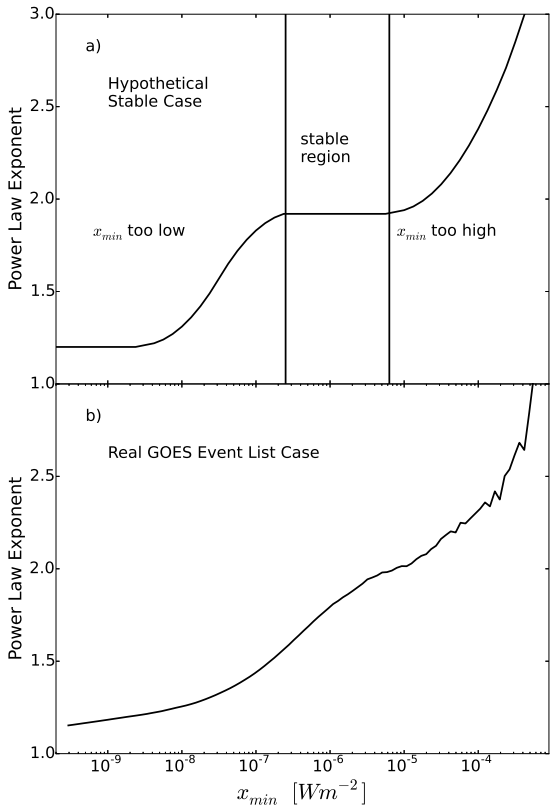

Fig. 5. a) Hypothetical relationship between the best fit power law slope and the lower limit (x_{min}) for a stable power law fit. The flat region in the middle of the plot shows the range over which the fit is independent of the choice of x_{min}, i.e. stable. b) The same relationship as a) for the real GOES long channel peak flux frequency distribution derived with the standard GOES event list (start threshold=40%, end threshold=50%). It can be seen that there is no flat region where the power law slope is independent of choice of x_{min}. This suggests that the distribution may not be best approximated by a single power law.

4.2.2. Assessing goodness-of-fit

To better understand this result, we assessed the goodness-of-fit for all the flare data sets derived in this study to their respective power law fits. The methods described below were implemented with the **poweRlaw** package (Gillespie 2015) using the statistical software language R. We initially operated under the assumption that the observed data greater than $x_{\min}$ follows a power law distribution (our null hypothesis). We then use p-values to assess the plausibility of this assumption in each case. The p-value is the probability that a finite sample drawn from a parent distribution will deviate from that parent distribution by at least a given amount due to random chance. In our case, the

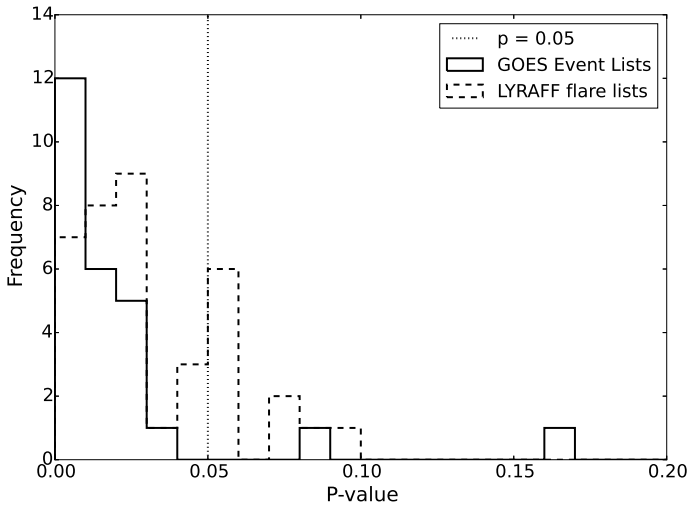

Fig. 6. Histogram of p-values for power law fits to all GOES/XRS flare peak irradiance frequency distributions determined with the GOES event list (solid) and LYRAFF (dashed). The dotted vertical line denotes the 5% level below which we reject the null hypothesis, i.e. that the flares are drawn from a power law distribution.

finite sample is the observed peak flux frequency distribution, while the believed parent distribution is a power law. If p is large, then it is impossible to rule out random noise as the cause of the deviation of the data from a power law. If p is small, then there is a very low probability of this being the case and our assumption that the null hypothesis is true becomes untenable. (For a description of how p-values were calculated in this study, see Appendix B.)

Figure 6 shows a histogram of the p-values derived for all the GOES/XRS flare lists derived with different thresholds in the GOES event list (solid) and LYRAFF (dashed) algorithms. The vertical dotted line shows $p = 0.05$, which is a threshold typically used to define whether a p-value is significant. It can clearly be seen that the majority of flare lists have p-values less than 0.05, while almost all have a p-value less than 0.1. This means that almost all flare distributions derived in this study have less than a 1 in 20 chance that their deviation from a power law is due to random noise. Moreover, almost all have less than a 1 in 10 chance. Although this does not categorically prove that the distributions are not power laws, it does cast significant doubt over our above assumption that flares detected with GOES/XRS are appropriately modelled by a power law parent distribution.

4.2.3. Possible Causes Of Non-Power Law Behaviour

There are two different explanations for why the observed flare peak flux size distributions are not power law-like. Either the true flare distribution is power law-like and the observational and/or analytical techniques are masking this or the true flare distribution is in fact not a power law.

There are a few possible reasons why this type of analysis may not reveal a power law flare distribution. First, the simple background subtraction methods used here – subtracting the flux at the flare start time or a linearly extrapolated flux between the flare start and end times – may not be adequate and may be corrupting the distribution. It is difficult to fully rule this out without performing manual background subtraction for all events and comparing results. Second, the range over which we can fit a power law may be insufficient to adequately determine the power law stability. The x_{min} values used in the fits described in Section 3.3 are around $5{\times}10^{-6}\,\mathrm{Wm}^{-2}$ owing to a combination of the upper limit of solar background and the GOES/XRS lack of spatial resolution. Moreover, the largest flares in our sample are of the order of $10^{-3}\,\mathrm{Wm}^{-2}$, while we have only very small number statistics above $10^{-4}\,\mathrm{Wm}^{-2}$. This gives us less than two orders of magnitude over which to fit a power law. If we were able to extend this range, perhaps a more reliable, stable power law could be found. Third, even in the range 10^{-6}–$10^{-4}\,\mathrm{Wm}^{-2}$ we could still be undersampling the true flare distribution owing to large flares masking smaller ones.

In order to investigate whether this third possibility is plausible, we attempted to determine how many flares would have to be missed by GOES/XRS and/or the standard GOES event list algorithm (start threshold=40%, end threshold=50%) to make the observed flare frequency distribution a stable power law. To do this we added synthetic flares to the original distribution above the upper limit of solar background variation ($\sim 5{\times}10^{-6}\,\mathrm{Wm}^{-2}$) until the CDF of the flare distribution closely matched the CDF of the theoretical power law distribution. The difficulty in this method is selecting the power law exponent with which to generate the theoretical CDF. Since the stability plot in Figure 5 is continuously increasing there is very little reason to select one exponent value over another. On the one hand, exponents corresponding to an x_{min} close to the upper limit of the solar background are, according to our hypothesis, artificially shallow because of undersampling. On the other, exponents corresponding to a high x_{min} suffer from small number statistics and are therefore artificially steep. To give us an idea of the importance of this problem, we performed this process for exponents of 1.98, 2.02, 2.2, and 2.3 (which correspond to fits using x_{min} values of $5{\times}10^{-6}$, 10^{-5}, $5{\times}10^{-5}$, and $10^{-4}\,\mathrm{Wm}^{-2}$, respectively; see Figure 5). The number of additional flares required to make the distribution stable around an x_{min} of $\sim 5{\times}10^{-6}\,\mathrm{Wm}^{-2}$ for each of the exponents were found to be as follows: an exponent of 1.98 required ∼300–400 additional flares; an exponent of 2.02 required ∼700–800 additional flares; an exponent of 2.2 required ∼2500–4000 additional flares; and an exponent of 2.3 required ∼4000–6000 additional flares. As percentages of the observed number of flares with peak fluxes above $5{\times}10^{-6}\,\mathrm{Wm}^{-2}$, these numbers translate to 6–8%, 14–16%, 25–80%, and 80–120%. The distributions of additional flares were predominantly high C- and low M-class flares. The percentage corresponding to an exponent of 1.98 is very similar to the 5% of flares in the GOES event list found by Ryan et al. (2012) to be double flares. As discussed in Section 2.2, this is a lower limit for the number 'missing flares' from the GOES event list, so perhaps the true number may be as high as ∼15%. This would agree with the 14–16% of additionally required flares corresponding to a power law exponent of 2.02. Therefore if the true power law exponent of the flare size distribution is around 2, it is not impossible that the deviation from a power law is due to undersampling of high C- and low M-class flares. If the true power law exponent is much above 2, this becomes increasingly unlikely. Unfortunately, however, it is not pos-

sible to objectively determine the true exponent from the GOES/XRS observations alone.

Other possible explanations for the behaviour of the observed flare frequency distribution is that the true distribution is close to but not actually a power law. If the global flare distribution is a convolution of multiple heavy-tailed distributions, e.g. if each active region has a different flaring distribution or if there are different flaring mechanisms which produce different distributions, then the resulting distribution would converge to a power law under the central limit theorem. Evidence for active region flare distributions deviating from a power law was found by Wheatland (2010). He used GOES/XRS to examine AR 11029, which produced 70 B- and C-class flares. Using a Bayesian maximum likelihood estimation technique he found that a power law plus exponential rollover model was 200 times more likely to represent the data than a simple power law. This functional form qualitatively agrees with the unstable single power law fits found in this study. Wheatland (2010) interpreted this power law deviation as being due to the fact that the active region only had a finite amount of free magnetic energy available for flaring. If this behaviour is found for all or many active regions, it is possible that the convolution of all active region flare distributions studied here would also exhibit this behaviour.

Another possibility is related to the fact that traditionally the GOES flare peak flux has been used as a proxy for the flare energy. This is valid if the total flare energy scales reliably with GOES peak flux. However, this may not be the case. It may be possible that some flares have a lower proportion of their thermal emission in the GOES/XRS long passband owing to their temperature or differential emission measure (DEM) distribution. This may cause the flare distribution as observed by GOES/XRS to slowly flatten at lower peak fluxes. Previous studies have examined the effect of using X-ray and EUV flux measurements as a proxy for the total flare energy (Aschwanden & Parnell 2002; Parenti et al. 2006). These studies have found that the power law slope of the flare size distribution as observed with a given emission line/passband can be different from the size distribution of flare thermal energies which ultimately drives the observed emission. This difference was found to depend on the formation temperature/temperature response of the observed line/passband. Parenti et al. (2006) further concluded that high temperature lines/passbands tend to be less different from the thermal energy size distributions. These studies, however, did not focus explicitly on the validity of a power law model for the size distribution, but rather the best fit power law slopes.

The interplay between conductive and radiative cooling is another potentially important factor. This interplay varies depending on the flare temperature. The GOES/XRS temperature response lies in the range 4–40 MK and typically leads it to record flare peak temperatures in the range 10–25 MK. At these temperatures, conduction is more efficient than radiative cooling, with radiation becoming increasingly relatively efficient as the temperature drops (Cargill et al. 1995). The cooling rates change from flare to flare based on physical conditions (Ryan et al. 2013). Therefore, a greater or lesser fraction of a flare's thermal energy may be transported through conductive or radiative processes and hence emitted in the GOES/XRS passbands at the flare peak.

Furthermore, flare energy may not be consistently partitioned between thermal energy, accelerated particles, non-thermal emission, etc. A comprehensive understanding of the energy partition between the different physical processes is a large and difficult undertaking and is beyond the scope of this paper. There have been some previous studies which have made important progress in this endeavour (e.g. Emslie et al. 2012; Aschwanden et al. 2015). However, more work is required to fully understand how the energy partition changes as a function of peak GOES flux across multiple orders of magnitude.

4.3. Using LYRA for flare statistics

We performed a similar analysis as described in Section 4.1 on the LYRA flare lists derived with LYRAFF. Figure 7a shows a log-log histogram of the background-subtracted peak LYRA zirconium channel irradiance of flares detected with the standard LYRAFF thresholds (start threshold=1%, first fall threshold=20%, end threshold=50%). The first thing to note is that the distribution visibly deviates from a straight line even above the roll-over. Similar behaviour was observed in most of the LYRA peak irradiance size distributions derived with different LYRAFF thresholds. This implies more strongly than in the case of GOES that these distributions are not well described by a power law. To confirm this, we determined the p-values of each LYRA peak irradiance frequency distribution using the same process described in Section 4.2. The results are shown in Figure 8a. It can be seen that most of the distributions have a p-value less than 0.05, which suggests that our null hypothesis – that the flares are drawn from a parent power law distribution – is incorrect. This agrees with our visual observation from Figure 7a. However, there is also a minority of distributions with very high p-values. We investigate this in Figure 8b, which shows the p-values as a function of the number of flares used in the fitting process, i.e. the number of flares greater than x_{min}. It can be seen that all the distributions with high p-values have less than 500 events greater than x_{min}. This is because the x_{min} was so high that it excluded the majority of flares from the fitting process. This implies that a power law is not appropriate for describing the distribution.

Although the LYRA flare distributions are clearly not power laws, they are typically much steeper than those found using GOES/XRS. For example, the best fit power law exponent for the distribution in Figure 7a was found to be $\sim 2.3 \pm 0.1$. For comparison, the exponents for almost all the GOES/XRS flare lists derived with both the GOES event list and LYRAFF algorithms were less than 2. In order to determine the cause of this discrepancy, we split the LYRA flares into two subdistributions: those observed by both LYRA and GOES/XRS and those observed by LYRA only. For this exercise we used the LYRA flare list derived with the standard LYRAFF algorithm (start threshold=1%, first fall threshold=20%, end threshold=50%) and the standard GOES event list (start threshold=40%, end threshold=50%). We found that out of the 5253 flares observed by LYRA between 2010 and 2014 inclusive (excluding eclipse seasons), 2276 were also in the GOES event list, while 2977 were not.

Next, we verified that the flares not in the GOES event list were not bad detections caused by data artefacts or deficiencies in the LYRAFF algorithm. We randomly se-

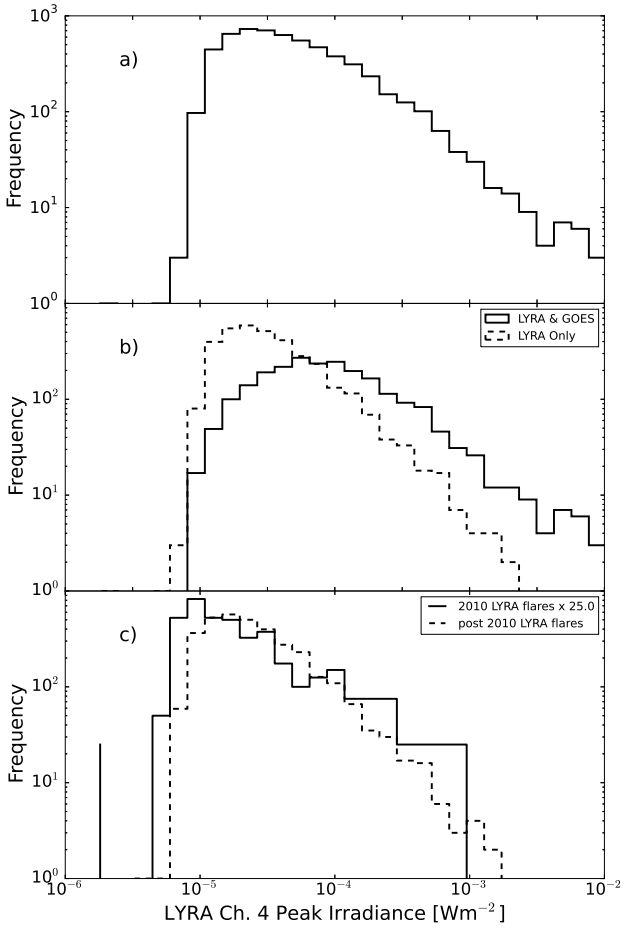

Fig. 7. a) Frequency distribution of background-subtracted LYRA zirconium channel peak irradiance of all non-eclipse-season flares between January 2010 and May 2015 derived with the standard LYRAFF algorithm (start threshold=1%, first fall threshold=20%, end threshold=50%). The distribution above the roll-over does not follow a straight line indicating this distribution is not a power law. b) The same distribution as in a) broken into flares observed by LYRA and GOES/XRS (solid) and those observed by just LYRA (dashed). The flares observed by only LYRA tend to be smaller and have a steeper distribution. c) The same distribution as in a) broken into flares which occurred in 2010 and after 2010. The slope of the 2010 distribution is less steep. We note that the 2010 distribution has been scaled by 25 to make it easier to visually compare its slope with the post 2010 distribution.

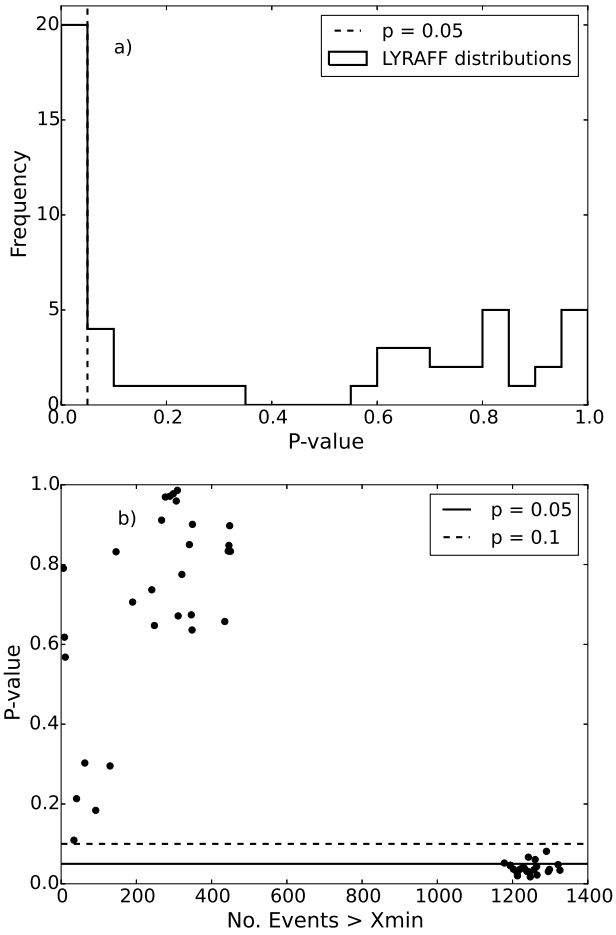

Fig. 8. a) Histogram of p-values for power law fits to all LYRA flare peak irradiance size distributions determined with the LYRAFF. The dashed vertical line denotes the 5% level below which we reject the null hypothesis, i.e. that the flares are drawn from a power law distribution. b) p-value vs. number of flares greater than x_{min} for the same flare distributions as in a). We note that distributions with high p-values have less than 500 events above x_{min}, causing the power law fits to be unreliable.

lected 100 of the LYRA flares not found in the GOES event list and examined their lightcurves. By visual inspection we determined that 91% were good detections, while only 9% appeared not to be a flare. This confirms that the majority of the LYRAFF detections are real and that there is a population of flares visible to LYRA but not detectable by GOES/XRS and the GOES event list.

Figure 7b shows histograms of the two LYRA subdistributions discussed above. Flares detected in LYRA and GOES/XRS observations are represented by the solid line while flares only detected with LYRA are represented by the dashed line. Two things are clear from this plot. Firstly, the LYRA-only flares tend to be small, which may be a clue as to why they are not detected with GOES/XRS. Since less intense flares tend to be cooler (Ryan et al. 2012), the LYRA-only flares may be too cool to exhibit significant thermal emission in the GOES/XRS channels. Secondly, the LYRA-only flares have a much steeper distribution. If this were physical it could imply that the flare distribution actually steepens as flares become smaller and would have interesting consequences for the coronal heating problem. However, another possible explanation is that LYRA's degradation has not been properly corrected and is skewing the results.

Although LYRA's zirconium channel has experienced less degradation than the other channels, its degradation is still significant. It is known that this degradation is more severe in the EUV part of the spectral band (6–20 nm) than the soft X-ray (SXR; <2 nm), which is more sensitive to flare variability. In order to avoid artificially enlarging the

flare variability in the signal – assumed to be dominated by the SXR contribution – an additive degradation correction is included in the publicly available data. However, if the flare variability also has a significant contribution from the EUV part of the passband, then the additive correction would cause flare peak irradiances to be increasingly underestimated as degradation increases. This would cause the flare distribution to become steeper over time. To check this, we plotted histograms in Figure 7c of the LYRA flares not found in the GOES event list from 2010 (solid histogram) when the degradation was not too bad and after 2010 (dashed histogram) when the degradation had become more significant. It can clearly be seen that the 2010 distribution is shallower than that for flares after 2010. This is consistent with the degradation explanation. Additionally, it is expected that the difference in slopes would be smaller if all flares observed by LYRA were included because the flares large enough to also be detected in GOES/XRS would have a higher contribution from SXR and the additive degradation correction would affect them less. This is exactly what we found. Therefore, we conclude that the steeper slope seen in LYRA flare distribution is neither physical nor due to deficiencies in the LYRAFF algorithm. It is most likely due to insufficient degradation correction of LYRA irradiances. This means that although a combination of LYRA and LYRAFF can be used to detect the occurrence of flares, absolute LYRA irradiances are not reliable and should not be used for flare statistics or energetics unless the degradation is adequately corrected.

5. Conclusions

In this paper we examined the effect of arbitrary thresholds in the GOES event list and LYRAFF flare detection algorithms when applied to GOES/XRS and PROBA2/LYRA observations. We found that that the power law index of the GOES/XRS-derived flare peak flux size distribution depends on the choice of start threshold, but not on those of the first fall and end thresholds. This was found to be true when both the traditional GOES event list and GOES-adapted LYRAFF algorithms were used. The power law slope value ranged from ∼1.92–2.02 depending on whether start thresholds of 90%–5% were used. This highlights the importance of understanding the biases introduced by flare definitions before drawing conclusions on the nature of the derived flare distribution. This is particularly true if trying to address coronal heating by nanoflares as changing the flare start threshold can cause the flare distribution to have a power law slope either above or below the critical threshold of 2.

We also found that the LYRA flare peak irradiance size distribution is artificially steep and not well modelled by a power law. This was shown to be consistent with an insufficient degradation correction. Therefore, although LYRA can be used to detect flares, its absolute irradiances should not be used for statistics or energetics unless the degradation is adequately corrected. However, this does not affect LYRA's ability to chart relative irradiance changes over short timescales, for example within flares.

We consider the discussion in the two previous paragraphs to be the main conclusions of this paper. However, in the process of this analysis we came to a number of secondary conclusions. Firstly, we found that that the LYRAFF algorithm does a better job of separating temporarily overlapping flares than the GOES event list. This leads to a slightly higher proportion of smaller flares being detected. This may cause the power law slope of the flare peak flux size distribution to steepen. However, any such steepening observed as part of this study was found to be within the range of statistical uncertainty. Despite LYRAFF's improved ability to detect temporally overlapping flares, it is still limited. If a greater number of such flares could be detected and separated, it is expected that the power law slope of the flare distribution would steepen further.

Secondly, we showed that – independent of the choice of threshold – the GOES flare peak flux size distribution is visually approximate to but is not strictly a power law. However, we have not been able to adequately determine the reason for this. We showed that by assuming the true flare distribution is a power law with an exponent of ∼2, it is plausible that the observed distribution's deviation from a power law may be owing to flares being missed by the detection algorithm. However, if the power law exponent is much steeper than 2, this becomes increasingly unlikely. The issues with flare detection, solar background, energy partition, and instrumental response suggest that conclusively answering this question with an instrument as basic as GOES/XRS may be impossible.

In order to improve our ability to examine the flare distribution, it is necessary to use spatial information. This will allow flares to be identified when they overlap temporally but not spatially. It will also improve our ability to accurately subtract the background flux. Both of these steps will improve our statistics. The flare distributions from multiple individual active regions should also be studied in this way to determine whether the global flare distribution is a single power law or a convolution of numerous heavy-tailed non-power law distributions. Studies that more comprehensively chart the energy partition of flares and eruptions must be undertaken to better determine what biases we introduce by using measurements from a single passband as a proxy for the flare energy. Finally, algorithms that can better identify the event duration will allow us to examine fluences rather than just peak fluxes, which are likely to be a better proxy of the total flare energy.

Acknowledgements. The authors would like to thank Ingolf Dammasch for his helpful discussions. D. Ryan wishes to thank the Solar-Terrestrial Centre of Excellence and the SIDC Data Exploitation project for their financial support. M. Dominique's work has been funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office (IAP P7/08 CHARM). Support for D. Seaton was provided by PRODEX grant No. 4000103240 managed by the European Space Agency in collaboration with the Belgian Federal Science Policy Office (BELSPO) in support of the PROBA2/SWAP mission and by the European Union's Seventh Framework Programme for Research, Technological Development and Demonstration under grant agreement No. 284461 (Project eHeroes, www.eheroes.eu). A. White has been supported by the the STATICA project, funded by the Principal Investigator programme of Science Foundation Ireland, contract number 08/IN.1/I1879. LYRA is a project of the Centre Spatial de Liege, the Physikalisch-Meteorologisches Observatorium Davos and the Royal Observatory of Belgium funded by the Belgian Federal Science Policy Office (BELSPO) and by the Swiss Bundesamt für Bildung und Wissenschaft. This research has made use of SunPy, an open-source and free community-developed solar data analysis package written in Python (SunPy Community et al. 2015).

References

Aschwanden, M. J., Boerner, P., Ryan, D., et al. 2015, ApJ, 802, 53
Aschwanden, M. J., Crosby, N. B., Dimitropoulou, M., et al. 2016, Space Sci. Rev., 198, 47
Aschwanden, M. J. & Freeland, S. L. 2012, ApJ, 754, 112
Aschwanden, M. J. & Parnell, C. E. 2002, ApJ, 572, 1048
Aschwanden, M. J., Tarbell, T. D., Nightingale, R. W., et al. 2000, ApJ, 535, 1047
Bak, P., Tang, C., & Wiesenfeld, K. 1987, Physical Review Letters, 59, 381
Bak, P., Tang, C., & Wiesenfeld, K. 1988, Phys. Rev. A, 38, 364
Battaglia, M., Grigis, P. C., & Benz, A. O. 2005, A&A, 439, 737
Benz, A. O. & Krucker, S. 2002, ApJ, 568, 413
Berghmans, D., Clette, F., & Moses, D. 1998, A&A, 336, 1039
Buchlin, E., Galtier, S., & Velli, M. 2005, A&A, 436, 355
Cargill, P. J., Mariska, J. T., & Antiochos, S. K. 1995, ApJ, 439, 1034
Charbonneau, P., McIntosh, S. W., Liu, H.-L., & Bogdan, T. J. 2001, Sol. Phys., 203, 321
Clauset, A., Shalizi, C. R., & Newman, M. E. J. 2009, SIAM Rev., 51, 661
D'Huys, E., Berghmans, D., Seaton, D. B., & Poedts, S. 2016, Solar Physics, 1
Dominique, M., Hochedez, J.-F., Schmutz, W., et al. 2013, Sol. Phys., 286, 21
Drake, J. F. 1971, Sol. Phys., 16, 152
Emslie, A. G., Dennis, B. R., Shih, A. Y., et al. 2012, ApJ, 759, 71
Feldman, U., Doschek, G. A., & Klimchuk, J. A. 1997, ApJ, 474, 511
Gillespie, C. S. 2015, Journal of Statistical Software, 64, 1
Hanser, F. A. & Sellers, F. B. 1996, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 2812, GOES-8 and Beyond, ed. E. R. Washwell, 344–352
Hudson, H. S. 1991, Sol. Phys., 133, 357
Hudson, H. S., Peterson, L. E., & Schwartz, D. A. 1969, ApJ, 157, 389
Ireland, J., McAteer, R. T. J., & Inglis, A. R. 2015, ApJ, 798, 1
Klimchuk, J. A. 2006, Sol. Phys., 234, 41
Lee, T. T., Petrosian, V., & McTiernan, J. M. 1995, ApJ, 448, 915
Lu, E. T. & Hamilton, R. J. 1991, ApJ, 380, L89
Lu, E. T., Hamilton, R. J., McTiernan, J. M., & Bromund, K. R. 1993, ApJ, 412, 841
Parenti, S., Buchlin, E., Cargill, P. J., Galtier, S., & Vial, J.-C. 2006, ApJ, 651, 1219
Parnell, C. E. & De Moortel, I. 2012, Philosophical Transactions of the Royal Society of London Series A, 370, 3217
Parnell, C. E. & Jupp, P. E. 2000, ApJ, 529, 554
Reale, F. 2010, Living Reviews in Solar Physics, 7, 5
Ryan, D. F., Chamberlin, P. C., Milligan, R. O., & Gallagher, P. T. 2013, ApJ, 778, 68
Ryan, D. F., Milligan, R. O., Gallagher, P. T., et al. 2012, ApJS, 202, 11
Santandrea, S., Gantois, K., Strauch, K., et al. 2013, Sol. Phys., 286, 5
Shibata, K. & Magara, T. 2011, Living Reviews in Solar Physics, 8
Shimizu, T. 1995, PASJ, 47, 251
Shimojo, M. & Shibata, K. 1999, ApJ, 516, 934
SunPy Community, T., Mumford, S. J., Christe, S., et al. 2015, Computational Science and Discovery, 8, 014009
Veronig, A., Temmer, M., & Hanslmeier, A. 2002a, Hvar Observatory Bulletin, 26, 7
Veronig, A., Temmer, M., Hanslmeier, A., Otruba, W., & Messerotti, M. 2002b, A&A, 382, 1070
Wheatland, M. S. 2010, ApJ, 710, 1324
Yashiro, S., Akiyama, S., Gopalswamy, N., & Howard, R. A. 2006, ApJ, 650, L143
Zender, J., Berghmans, D., Bloomfield, D. S., et al. 2013, Sol. Phys., 286, 93

Appendix A: LYRA Artefacts Removed By LYRAFF

Table A.1 shows the different artefacts removed from LYRA observations before applying the LYRAFF algorithm. These artefacts are defined in the PROBA2 time annotation files. For more information see http://proba2.oma.be/data/TARDIS.

Appendix B: Calculating p-values

The method of calculating p-values used in this study is outlined in Clauset, Shalizi, and Newman (2009). First, we used the methods described in Section 3.3 to find the best fit power law to our empirical data set, giving us an exponent $\hat{\alpha}$, a lower limit $x_{\min}$, and a KS statistic KS_{emp}. The total number of flares in the data set is n, while the number greater than $x_{\min}$ is n_{tail}. Next, we created an ensemble of synthetic data sets. Data points above $x_{\min}$ were randomly generated from a power law with exponent $\hat{\alpha}$, while data points below $x_{\min}$ were randomly sampled with replacement from the corresponding subset of the empirical data. Each synthetic data point had a probability of n_{tail}/n of being in the regime above $x_{\min}$ and $1-n_{tail}/n$ of being in the regime below this value. Thus, the synthetic data sets should mimic the empirical data set well across its entire range if the data above the threshold value of $x_{\min}$ follows a power law.

The MLE method outlined in Section 3.3 was then used to fit each synthetic data set with its own power law model, and its corresponding KS statistic of the fit, KS_{synth} was calculated. Finally, the p-value was given by $p = n_{synth>emp}/n_{synth}$, where $n_{synth>emp}$ was the number of synthetic data sets with a $KS_{synth} > KS_{emp}$ and n_{synth} was the total number of synthetic data sets. In this study we created 1000 synthetic data sets for each empirical one so that our p-value would be reliable to two decimal places (Clauset et al. 2009).

Table A.1. Standard artefacts removed from LYRA observations before applying LYRAFF.

Artefact	Description
ASIC reload	Reload of the application apecific integrated circuit
Calibration	Calibration campaign
Glitch	Undefined glitch.
LAR	Large angle rotation. 90^o rotation of satellite lasting a few minutes. Cause spikes and dips in the signal.
Moon in LYRA	Moon in LYRA field of view which can cause non-solar variations in signal.
Offpoint	PROBA2 offpointed from disk centre potentially altering signal artificially.
Operational Anomaly	Unexplained anomaly.
Recovery	Time taken for LYRA detectors to get back to full efficiency after not being exposed to sunlight. Can take many hours.
SAA	Transition of PROBA2 through the South Atlantic anomaly causing increased noise in signal.
UV occultation	Solar irradiance attenuated by Earth's atmosphere. Defined as when PROBA2's tangential altitude is $\leq$350 km.
Visible occultation	Solar irradiance is blocked by the disk of the Earth.