A new catalogue of solar flare events from soft x-ray GOES signal in the period 1986-2020

Nicola Plutino, Francesco Berrilli*, Dario Del Moro, Luca Giovannelli

Department of Physics, University of Rome Tor Vergata, Via della Ricerca Scientifica, 1, Rome 00133, Italy

Abstract

Solar flares, along with other sun-originated events such as Coronal Mass Ejections, fast solar wind streams, and solar energetic particles are among the most relevant events in Space Weather. Moreover, solar flares are the most energetic processes that occur in our solar system. The in-depth study of their occurrence statistics, both over extended periods or during individual solar cycles, allows us to improve and constrain the basic physical models of their origin. Increasing the number of detected events, especially those of lower intensity, and the number of physical parameters that describe the detected flares is, therefore, a mandatory goal.

In this paper, we present a computationally efficient algorithm for the detection of solar flares in the soft-X solar flux provided by the GOES (NASA/NOAA) satellite constellation. Our code produces a new flare catalogue increasing the number of events with respect to the official GOES list. In addition to increasing the number of identified events, the catalogue contains information such as: an estimate of the total energy released, start and end time of the event, possible overlap with other events, background level of the GOES X-ray emission close to the revealed event. After a detailed description of the detection algorithm, we carry out a preliminary analysis of the flares reported in our catalogue and compare our results with the official list of GOES for the period from 1998 to 2020.

1. Introduction

Space weather events (SWEs), in the form of solar flares, Coronal Mass Ejections (CMEs), fast solar wind streams, solar energetic particles (SEPs), accelerated by magnetic reconnection processes during flares and CME-driven shocks, occur regularly albeit with different frequencies. All these SWEs affect the physical state of the Earth's upper atmosphere, i.e., the magnetosphere, the ionosphere and the thermosphere, and therefore they are important both in defining the physics of Sun-Earth relations and because they have significant operational benefits for satellites in low Earth orbits (LEOs).

The many open questions regarding the physics governing solar activity, the difficulty of predicting instability in solar plasma or identifying the magnetic feature connected to SWEs, including flares (e.g. Park et al., 2020; Cicogna et al., 2021), have led the

---
*Corresponding author:
Email address: francesco.berrilli@roma2.infn.it (Francesco Berrilli)

arXiv:2211.10189v1 [astro-ph.SR] 18 Nov 2022

space agencies to focus their efforts on Space Weather phenomena and precursors (Bobrinsky & Del Monte, 2010; Crown, 2012; Murray et al., 2017; Ishii, 2017; Berrilli et al., 2019; Plainaki et al., 2020).

Solar flares release a huge amount of energy, usually 10^{20} J, however the largest solar flares can reach energies of around 10^{25} J, sometimes up to 10^{26} J, over a timescale of hours (Schwenn, 2006). These impulsive phenomena may trigger spectacular SWEs effects, starting from the historical events that opened the study of the Sun-Earth relationship, such as the Carrington Event of September 1859 (e.g. Smart et al., 2006) or the great aurora of February 1872 (e.g. Berrilli & Giovannelli, 2022).

Solar flares and CMEs do sometimes occur at the same time – and the strongest flares are almost always correlated with CMEs – but the exact relationship between the two events has been long debated. The late idea that CMEs were initiated by large flares (Lin & Hudson, 1976), has been disputed by Zhang et al. (2001) and Zhang & Dere (2006) which showed that CMEs are statistically initiated before the onset of the associated flares. Recent works, making use of space-borne observatories and computational models, effectively investigated the solar surface and atmosphere associated with a flare and CME to analyze the dynamics of the magnetic field before and during the ejection event. Among others, a recent work by Temmer et al. (2017) confirmed that magnetic reconnection processes associated with the CME "were already ongoing before the start of the impulsive flare phase".

Since it is not the purpose of this work to go into the details of the single flare model or of the magnetic instability processes, responsible for the flare triggering of active regions, we refer to specific literature to deepen these aspects (e.g. Priest, 1986; Török et al., 2004; Fletcher, 2005; Pucci & Velli, 2013; Toriumi & Wang, 2019).

Rather, in this work we focus on the need to increase the statistics of flares in order to investigate the nature of physical processes, and the consequent models, necessary to explain their distribution over time and the physical nature that connects them to the solar magnetic cycle (e.g. Vlahos et al., 1995; Boffetta et al., 1999; Priest & Forbes, 2002; Charbonneau et al., 2001; Viticchié et al., 2006; Aschwanden & Freeland, 2012; Sadykov et al., 2019; Lu et al., 2021). In literature, several papers presented the analysis of the distribution of soft X-ray peak fluxes of solar flares suggesting that power laws with slope with a coefficient $\alpha = 1.6 - 2.2$ are a good estimate of this behavior (e.g. Crosby et al., 1993; Boffetta et al., 1999; Veronig et al., 2002; Aschwanden & Freeland, 2012; Li et al., 2016), although other functions, i.e., a lognormal distribution, have recently been proposed (e.g. Verbeeck et al., 2019).

Indeed, as shown in Aschwanden & Freeland (2012), there are still many open and intriguing questions concerning the statistics of solar flares.

The statistical analysis of flares during the different solar cycles, and possibly within the same cycle, requires to increase the detection of events as much as possible and to determine different characteristic parameters of X-ray emissions. These physical parameters are for example: an estimate of the total energy emitted, since the peak flux alone provides a partial information of the energy budget of the flare in the observed spectral range, the multiplicity, which is linked to the release efficiency of the active region where the flare is located, or the start and end times of the flare from which to extract the waiting times, the distribution of which makes it possible to discriminate between different models, i.e., shell model for MHD turbulence (e.g. Giuliani et al., 2000) or fractal-diffusive self-organized criticality (e.g. Aschwanden & Freeland, 2012).

With this purpose in mind, we have developed an algorithm which, based on GOES SXR light curves and on the evaluation of the background level of the GOES X-ray emission close to the revealed event, estimates several of these parameters that are used to populate the catalogue. This result allows us to carry out a preliminary statistical analysis on our data. In particular, an analysis was performed on the 1998-2020 events in order to study the distributions of peak fluxes under different analysis conditions. For a direct comparison of the distributions of the peak fluxes between us and GOES, and especially of the distributions when we consider the evaluation of the background level of the GOES X-ray emission close to the identified event.

To achieve this goal we use as input data of our algorithm the long channel (1-8 Å) in the soft X-ray (SXR) range because of the higher signal-to-noise ratio and better sensitivity to low-intensity events (available at `https://www.ngdc.noaa.gov/stp/satellite/goes/dataaccess.html`).

To implement our algorithm we have relied on the algorithm proposed in Aschwanden & Freeland (2012) by modifying some procedures and parameters (for details about the comparison between our algorithm and that of Aschwanden & Freeland (2012) see appendix A and the discussion in the next sections) but above all by adding new physical parameters. Indeed, compared to Aschwanden & Freeland (2012) work, we have increased the number of physical parameters (attributes) of the revealed events. In fact, as we have discussed earlier, the selection of the most appropriate model to describe the flare distribution requires different parameters from the peak flux alone.

Using GOES SXR light curves in the period 1986-2020 we reveal a total of 334123 events that will form the population of our catalogue. The whole catalogue containing the event attributes estimated by our algorithm for all the identified flares are made available through the web page `https://github.com/nplutino/FlareList` together with Python scripts to retrieve selected data that match user-defined criteria.

2. GOES data set

GOES consists of a series of geostationary satellites (orbiting the Earth at a height of $35,790$ km), which produces an uninterrupted time series of solar X-ray fluxes, together with meteorological observations of the Earth. This continuous stream of data is guaranteed by an overlap in time so that there are always one to three spacecraft recording data.

GOES X-ray Sensor (XRS) data have been recorded by NOAA satellites since 1974 and are archived at the NOAA National Center for Environmental Information (NCEI). Each GOES satellite is equipped with two X-ray sensors which provide solar fluxes for the wavelength bands of 0.5-4 Å (short channel) and 1-8 Å (long channel).

NOAA Space Weather Prediction Center (SWPC) has used this data to produce the 1-minute and 5-minute averaged X-ray data sets. Moreover, it is responsible for producing an official X-, M-, and C-class solar flare catalog, observed by the 1–8Å channel of the GOES XRS, which we will be referred to as the GOES catalog from now on. This catalog, derived from an automatic procedure applied to XRS lightcurves, is the gold standard for studies on the statistical properties of flares. Given its relevance, we will describe in an extremely synthetic way, the standard NOAA algorithm applied to the XRS signal and which define a GOES flare event.

SWPC algorithm populates GOES catalogs with the following flare detection criteria:

1. a *GOES X-ray flare event* starts when 4 consecutive 1-minute X-ray flux are above the B1 threshold, i.e. above a flux threshold of 10^{-7} W/m^2;

2. all 4 flux values monotonically increase;

3. the fourth flux value is greater than 1.4 times the first value in the sequence of 4 data points.

Solar flares are classified (A-B-C-M-X scale) taking in consideration the flux in the soft-X band at the peak of the flare emission. The event ends when the flux, during the decaying phase, drops below half of the peak flux. The data set used in this work covers the period from 1986 to 2020 (three solar cycles), with data from GOES-6 to GOES-16. Data from GOES-6 to GOES-15 have been downloaded in their February 2021 version (`https://www.ngdc.noaa.gov/stp/satellite/goes/dataaccess.html`), while GOES-16 data have been downloaded in their March 2021 version (`https://www.ngdc.noaa.gov/stp/satellite/goes-r.`

html). However, as described in the GOES XRS Operational Data document, (`https://www.ngdc.noaa.gov/stp/satellite/goes/doc/GOES_XRS_readme.pdf`), to get true fluxes for data up to GOES-15, users must remove the SWPC scaling factors. Dividing the long band flux by 0.7, such corrected fluxes and corresponding flare indices will agree with those obtained by GOES-16. We applied this correction to the GOES light curves, analyzed by our procedure, in order to obtain true fluxes for the entire list of events. At the same time, since data is recorded from different satellites, data points in the whole series do not show the same temporal resolutions: until 2010 data have a temporal resolution of 3s, while more recent data have a 2s resolution. In addition, data present gaps and instrumental spikes which must be handled before the event detection itself. For this reason, data manipulation starts with a smoothing process.

3. Flare detection algorithm

In this section we describe our mining procedure applied to GOES soft X-ray light curves to identify X-ray events (i.e., flares). The methodology includes a data pre-processing step, preparatory to the identification of events, and subsequent steps in which the events are identified and different physical quantities associated with the events are calculated. For the sake of clarity we divide this description in subsections, starting with the data pre-processing, and the localization of x-ray emission local minima and maxima, which is critical for our definition of flare event. The last two subsection are devoted to the handling of overlapping flare events and to the description of the outputs of the flare search algorithm. In subsection 3.3 we discuss how to deal with the pre-flare background. The estimation of this flux, which includes not only the contribution of the active region hosting the event, but the integrated flux over the entire solar disk (sun-as-a-star) and beyond the limb, is necessary to maximize the efficiency in detecting flare of a minor class and in correctly estimating the energy associated with the identified event.

3.1. Data pre-processing: smoothing procedure

An initial smoothing procedure is applied to the XRS signal as a one-dimensional data noise reduction. Data smoothing techniques for GOES soft x-ray light curves have been described extensively (e.g. Aschwanden & Freeland, 2012). We replicate some of these procedures using Python libraries and analysing 24 hours of data at a time. The use of 24 hour batches was decided to facilitate the execution of the procedure, but introduces some possible problems if a flare occurs near the end of the batch and extends after the end of the day. At the moment we neglect this fact for two substantial reasons: first statistically no more than 4% of the events occur in the last hour of the day, second, since the duration of the event depends on its intensity (e.g. Kashapova et al., 2021) only the flare more energetic that occur around 24:00 UTC extend significantly after the end of the day leading to a significant underestimation of descriptors such as total energy. We checked all the identified class X events and we report that of the 528 flares recorded for the entire 1986-2020 period only 17 have a peak after 23:00 UTC (i.e., 3.2%) and only two events have a peak near the midnight. A procedure that considers these few cases, taking into account the initial hours of the following day, has been implemented and will allow to correct the values of total energy and end time of the flare. In any case, since in the present work we will carry out the analysis of peak fluxes only, this problem has no effect.

To deal with the varying temporal resolution of the data we perform a temporal sub-sampling by replacing a set of data points with their average flux. Since the typical solar flare duration is in the range of minutes (e.g. Reep & Knizhnik, 2019), this operation can be performed without significant loss of information. This procedure not only produces a uniform time-spaced data set, but also reduces the number of data points to be processed by the algorithm, improving efficiency. The sub-sampling operation is performed using the function *resample* of the module *pandas* for Python (McKinney, 2011). This function correctly re-samples our data every 12s and is adept at excluding data gaps under typical conditions.
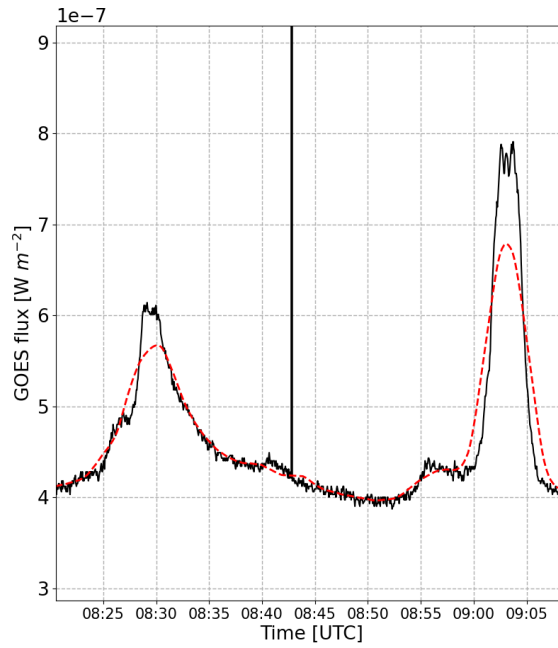
Fig. 1. Data smoothing. The black continuous line represents the raw data. A major data spike can be observed at about 8:42 UTC. The red dashed line represents the signal obtained after the smoothing operations. Data refer to 2nd November 1987.

The main purpose of the smoothing procedure is to mitigate or eliminate the effect of instrumental spikes on the light curve to be analyzed. Indeed, sudden flux increases of instrumental origin, i.e., non-solar origin, can produce a false flare detection. This problem has already been pointed out and solved by introducing a parameter that describes how fast the slope of the light curve changes (Aschwanden & Freeland, 2012). We use a moving time window, with a time window size of 60s, which associates each data point with a value

$$Q = F_{max}/F_{min} \tag{1}$$

where F_{max} and F_{min} are the maximum and minimum flux value in the time window. Therefore, the Q-value for each data point contains information on the neighbour of that data point and on how fast the flux changes in this time interval. Subsequently, the algorithm performs a classification procedure. Any point with a value of $Q > Q_{crit} = 10$ is identified as a spike and removed from the data set. We tested this method with data from the year 1987, one of the years greatly affected by the spike problem. The spike detection resulted in a total of 4968 events, less than the number of data points in a single day. Therefore, it should be emphasized that if on the one hand the elimination of these instrumental spikes must be carried out to eliminate false identifications, on the other hand the effect on the total statistics is in practice negligible.

The last step of the smoothing procedure is a boxcar average operation, which is crucial to improve algorithm efficiency. Indeed, our algorithm computation time strongly depends on the number of maximum and minimum points to be analysed. With a smoother curve, we are automatically excluding maximum and minimum points coming from signal noise. Given the boxcar size (the number of data points $2n + 1$ to be used in the average procedure), fluxes from x_1 to x_{2n+1} are summed up and the total flux is then divided by $2n + 1$. The boxcar size is 21 data points, corresponding to approximately 4 minutes. The flux obtained this way is attributed to the middle point x_n. Then the average operation moves on points from x_2 to x_{2n+2}. We can describe the whole boxcar averaging
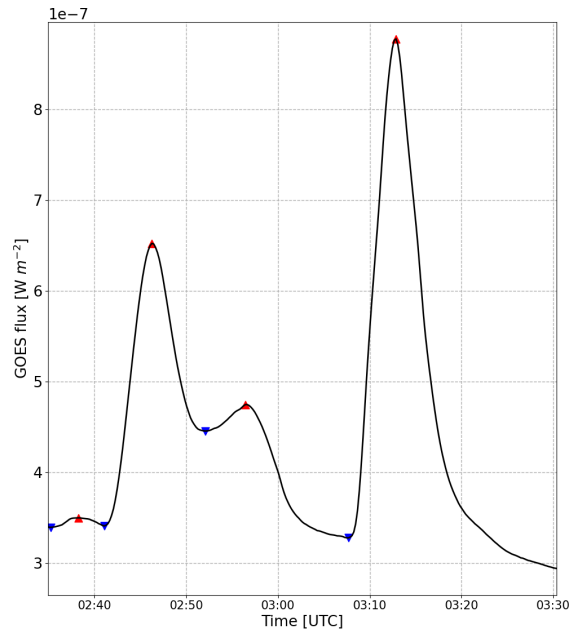
Fig. 2. Maximum and minimum detection procedure. Red upward triangles mark maximum values, while blue downward triangles mark minimum values. Data refer to 2nd November 1987 GOES XRS signal.

operation with the recursive formula:

$$x_i = \frac{1}{2n+1} \sum_{j=i-n}^{i+n} x_j \qquad (2)$$

All the steps described so far produce the smoothed data sets. In Figure 1 an example of the smoothing procedure on real data is shown.

Once this pre-processing of the data has been completed, it is possible to move on to identifying the events. The flare detection algorithm will be described in the next sections.

3.2. Flare identification: detection of local maximum and minimum

The identification of a flare event starts with the detection of local maxima and minima. More in detail, local minima will be taken as possible event start points, while local maxima will be used as candidates for flare peak flux values. Indeed, GOES light curves present small variations even after the smoothing procedure has been applied. For this reason, the search for the maximum (or minimum) is carried out in a moving sub-sample of 5 points. A data point is considered maximum (minimum) only if its value is higher (lower) than the values of the two points on its left and of the two points on its right. This operation considerably reduces the number of extremes that the algorithm has to analyze to identify events. An example of this procedure is shown in Figure 2.

3.3. Flare identification: definition of an event

Once all minimum and maximum values are listed, a new step of the flare identification procedure searches for candidate flare events. However, as reported in the previous sections, the soft X-ray flux level background must be taken into account in order to identify event candidates. As already reported in the paper by Wagner (1988), the flux integrated soft X-ray produced by all the active regions on the whole solar disk (sun-as-a-star) and beyond the limb changes continuously and is also modulated by the solar activity cycle. This flux level can be greater than the expected flux for flare of a minor classes (A or B), and during the maximum of activity also of C-class flares. Particularly important in this context is the work of Ryan et al. (2012) which, with the aim of deriving flare thermal properties, introduces an automated temperature and emission measure-based background subtraction method.
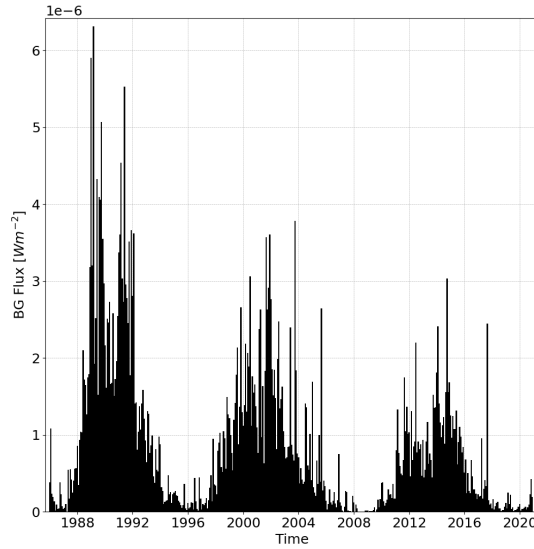
Fig. 3. The plot reports the monthly average of the background flux f_{BG} calculated for all the events identified by our procedure in the period from 1986 to 2020.

In our algorithm the calculation of the flux level background f_{BG} is computed from the starting minimum t_i preceding the event candidate (maximum). The figure 3 shows the monthly average of the f_{BG} calculated for all the events identified by our procedure in the period from 1986 to 2020. Its dependence on solar activity and the 11-years cycle is evident.

Since each minimum point is considered as a possible flare starting time t_i, a "pre-flare" background flux f_{BG}, obtained as the mean flux in a time interval $[t_i - 60s, t_i]$, is calculated and associated to each starting time t_i. Once the value of f_{BG} has been defined, the flux f_{end} associated with the end point of the event is defined by the relation:

$$f_{end} \leq f_{BG} + f_{noise} \tag{3}$$

where $f_{noise} = 2 \cdot 10^{-8} Wm^{-2}$ is the noise level for the typical GOES fluxes and is calculated during quiescent time periods (Aschwanden & Freeland, 2012).

When the condition described by the relationship 3 is not satisfied within the analyzed 24-hour batch, it means that the decay phase of the flare has not finished within the analyzed interval. In this case we assume the end of the analyzed batch as the end of the event. The effects of this simplification, which will be taken into account in a new version of the algorithm, were discussed at the beginning of the section.

At the end of these operations a list of candidate flaring intervals is produced. The procedure iterates through the list and looks for the maximum flux f_{max} within the considered interval. The condition for the definition of f_{max} is:

$$f_{max} > f_{BG} + f_{threshold} \tag{4}$$

where the background flux f_{BG} is referred to the starting minimum and $f_{threshold}$ is an adjustable parameter. The term f_{BG} in condition 4 accounts for the soft X-ray flux level background and guarantees that only events with a positive absolute intensity $I_F = f_{max} - f_{BG} > 0$ are considered. It is important to underline that this definition allows to identify class A and B flares even during the phases of maximum solar activity.

As for the choice of the $f_{threshold}$ parameter, we set $f_{threshold} = f_{noise}$. The $f_{threshold}$ parameter influences the identification of events. As its value increases, the procedure selects more energetic events, on the contrary by decreasing the value too much we
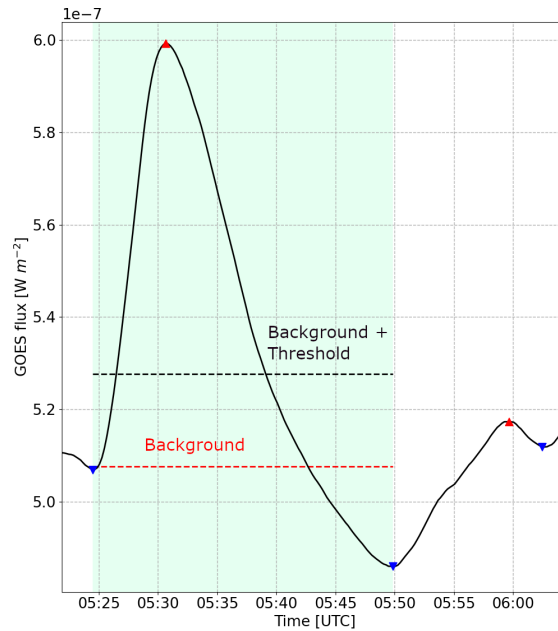
Fig. 4. Detail about the definition of flare event. The background value f_{BG} (red dashed line) is computed from the starting minimum t_i (first blue downward triangle). The threshold value $f_{threshold}$ is added to background value and the ending minimum f_{end} is chosen (second blue downward triangle). Since the maximum (first red upward triangle) meets Condition 4, a flare is detected in the highlighted area. This figure has been obtained using $f_{threshold} = f_{noise} = 2 \cdot 10^{-8} Wm^{-2}$. Data refer to 1st January 2001.

risk confusing stochastic fluctuations of the signal for low energy events. Our choice, i.e. $f_{threshold} = f_{noise}$, reduces the probability that fluctuations produced by noise are confused with flares and places a limit on the intensity of the identifiable flares which is equal to about an A2.0 class. The Figure 4 shows an example of flare identification. The main thresholds used by the algorithm are also shown.

3.4. Overlapping flare events

The final list can contain flares with overlapping time intervals because the same ending minimum (flare endpoint) can be matched by more than one starting point. Every time that a flare has not yet reached its ending point before another flare starts, the duration of the first flare is truncated at the beginning of the second one. Aschwanden & Freeland (2012) already pointed out how flare duration can be underestimated by this truncation process. To solve this problem we introduce a system of labels which can preserve the information about the original ending point. Every time that the duration of the first flare is truncated at the beginning of the second one, the two (or more) events are labelled with the same *Multiple ID*. In this way, while performing data analysis, these events can be considered both individually or as multiple events. The total duration of the multiple event can be obtained using the group ID and considering the starting point of the first flare of the group and the end point of the last event of the group. This labeling system can add a layer of complexity and open new possibilities for the analysis of flare characteristics. Figure 5 shows how the time truncation and labelling procedures work.

3.5. Event attributes and descriptions

The catalogue consist of records of 10 fields, each of which represented an attribute of the identified event. Table 1 reports a list of the event attributes and a short description of the attribute.
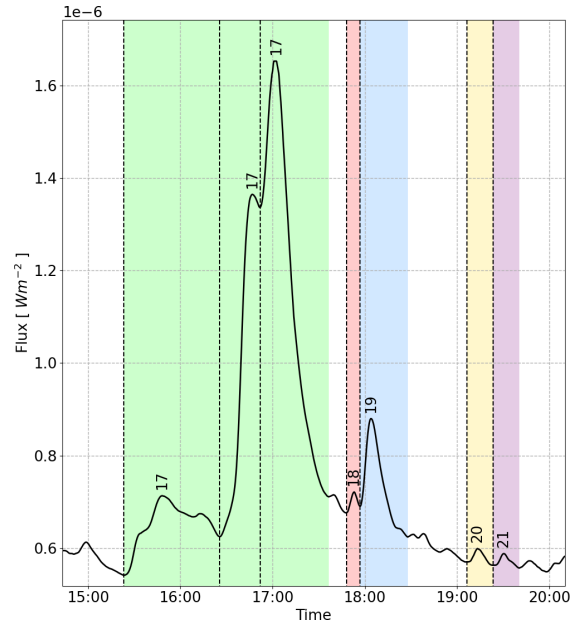
Fig. 5. Dashed vertical lines represent flare starting times, while highlighted areas refer to the duration of each event. Different colors have been used to group flares with the same label. The first two flares in the green time interval do not reach an ending minimum point that satisfies Condition 3, therefore they are associated with the last flare of the group, with the same *Multiple ID*. Conversely, the ending point of flare 18, even if representing the starting point of flare 19, meets Condition 3, so these two flares are associated with two different *Multiple ID*. Data refer to 1st January 2001.

Event attributes and descriptions	
Attribute	Description
Event ID	Unique ID to identify single flare events
Start time [UTC]	Estimated event start time
End time [UTC]	Estimated event end time
Peak time [UTC]	Estimated event maximum (peak) flux time
Peak Flux [Wm^{-2}]	As in the GOES catalog without subtracting the associated background flux
Flare class	As in the GOES catalog, i.e., A, B, C, M, X
Background Flux [Wm^{-2}]	The evaluated background level (f_{BG}) of the GOES X-ray emission close to the revealed event
Multiple ID	A unique ID for overlapping flares. This ID can be used to group flaring events and test how flare characteristics change when grouping the flare events. In principle, the same Multiple ID does not mean that the flares show a homologous behavior but simply that their signal is identified overlapping a previous event and most likely they come from the same active region
Total flux [Wm^{-2}]	Integral of the x-ray flux during the flare, as a further estimate of the total energy of an event

Table 1. Event attributes and short description of the various attributes.

4. XRS flare detection and catalogue population

Our catalogue lists 334123 events, identified in the GOES XRS data in period from 1986 to 2020, used to populate our catalogue. The GOES catalogue reports 61375 events for the same period. The first check we carried out is a comparison between the entries
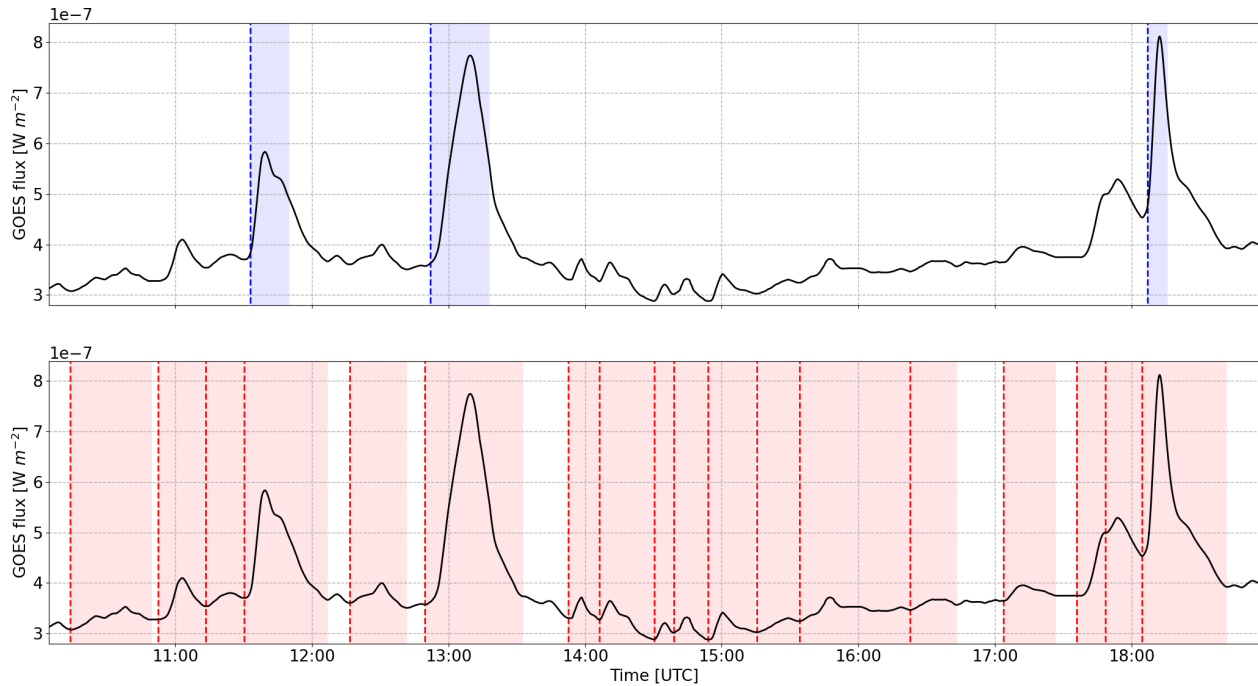
Fig. 6. Comparison between the two flare catalogs. Dashed lines are used to mark the starting points of events, while the highlighted areas represent the duration of the event. The upper panel, marked with blue, refers to the GOES catalog, while the lower panel and the red color refer to the catalog produced in this study. Data refer to 3rd March 2001.

of the official GOES catalogue and those produced in output by our procedure. GOES catalogue data have been retrieved using the Python module *sunpy* (Mumford et al., 2015). Since our list has been produced using the GOES SXR flux data scaled appropriately to obtain true fluxes, while the official GOES catalogue, currently available, does not implement this correction, before carrying out the comparison of the events we have appropriately rescaled the SXR flux data of GOES. Therefore, we repeated the scaling procedure mentioned in the GOES SXR Operational Data document on the data of the GOES catalogue. After this operation we have produced two lists of events with the same scaling and therefore correctly scaled between them for a final comparison.

4.1. Comparison with GOES catalogue

We performed various types of comparisons between the two catalogs, both qualitative and quantitative. First of all, we made several qualitative comparisons between the two lists. We compared the light curves and flares listed by GOES soft-X data and those identified by our procedure. In Figure 6 we show a typical example of these comparisons. It is evident how our procedure identifies a greater number of flares compared to those listed by GOES.

To quantify the difference in identified flares we classified the events in the usual way and we obtained the histogram shown in Figure7. The comparison is limited to the events identified in the period 1998-2020. This because before May 1997 (Swalwell et al., 2018) GOES flare list does not show the same event timing definition used in the following period. In the GOES list soft X-ray profiles are used to define the flare event duration only after May 1997, while previous listed events make use of $H\alpha$ data to obtain start and end times (Veronig et al., 2002). In this period the proposed list contains 205989 flares, against the 38517 listed by GOES catalog. Although we do not believe that this difference has a real impact on the determination of intensities, but only in the timing of events, we prefer to restrict the analyzes to a homogeneous data period.
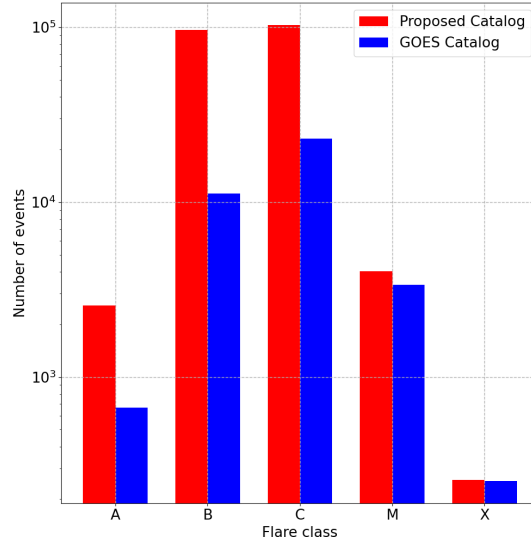
Fig. 7. Histogram of the number of flares by classes for the period from 1998 to 2020. The catalogue produced by our procedure increases the statistics of the events of all classes with respect to the events listed in the GOES catalogue. The increase of events in the different classes can be appreciated by observing the number of events listed in our catalogue (red bars) compared to those associated with the GOES catalogue (blue bars).

4.2. Peak flux distributions

As mentioned in Section 1 GOES catalog has already been used to analyse the distribution of solar flare peak fluxes. We compare this distribution for the GOES catalog with our list of flares for the period from 1998 to 2020.

To make this comparison we treat the flares peak flux as a random variable X and consider the Probability Density Function (PDF) $p(x)$, which is related to the probability of obtaining a flux value between F_{min} and F_{max} through the formula:

$$Pr(F_{min} \leq X \leq F_{max}) = \int_{F_{min}}^{F_{max}} p(x)dx \tag{5}$$

The expected shape for a power law PDF is the following:

$$p(x) = \frac{\alpha - 1}{x_{min}} (\frac{x_{min}}{x})^{\alpha} \tag{6}$$

with α index of the power-law and x_{min} the minimum value of flux for which the power-law holds. The PDF obtained using GOES list and the one obtained using our list can be seen in Figure 8 (*Panel a* and *b*). The list of events obtained with our algorithm is missing flares below the A2.0 class (with a peak flux lower than f_{noise}) because of the built-in Condition 4.

In Figure 8 different distributions of peak fluxes are shown both from the GOES catalogue and from ours. Relevant are panels c) and d) which show our results and take into account the background flux contribution. We hypothesized power law distributions (red lines) and validated the hypothesis by means of the Kolmogorov-Smirnov (K-S test) statistics. This test gives us a measure of the goodness-of-fit of the data (Goldstein et al., 2004). We used the Python package *powerlaw* (Alstott & Bullmore, 2014) to perform the K-S test and estimate the power law region, including its lower (x_{min}) and upper (x_{max}) bounds (which must hold more decades to have statistical significance). This package returns the bound values and the value of power law index α by means of a linear fit.

In more detail, the results obtained can be listed as follows:

- GOES list (without correction for f_{BG}): we find a power law index $\alpha = -2.16 \pm 0.02$ and $x_{min} = 7.4 \cdot 10^{-6}$ Wm^{-2}. The fit (red line) is shown in Figure 8 *Panel a)*.
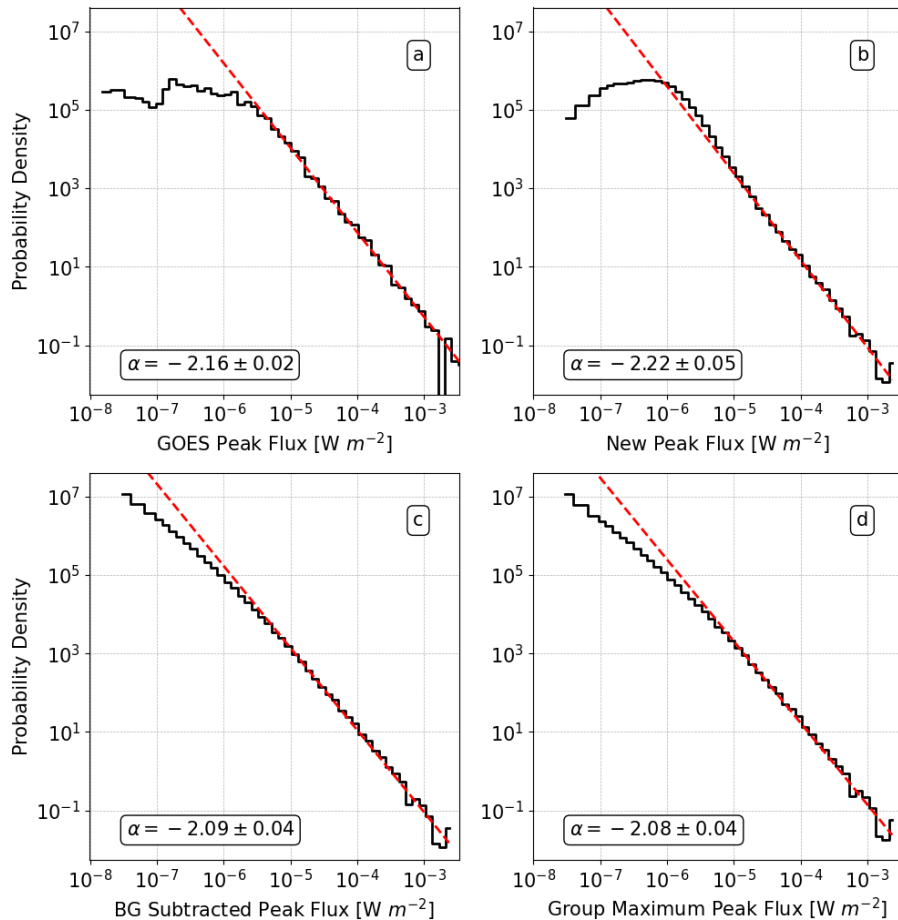
Fig. 8. In this figure we use GOES list and the proposed one to obtain peak flux probability density functions for the period 1998-2020. Each distribution is obtained with a different criterium: *Panel a)* peak fluxes listed in the GOES catalog; *Panel b)* peak fluxes listed in the catalog proposed in this study; *Panel c)* and *Panel d)* show peak fluxes obtained from our catalog once the f_{BG} is subtracted. *Panel c)* is obtained considering all 205989 events listed in our catalog, while *Panel d)* is computed using the most intense event flux within a sequence recognized as multiple (130925 events).

- Our catalog (without correction for f_{BG}): we find a power law index $\alpha = -2.22 \pm 0.05$ and $x_{min} = 1.8 \cdot 10^{-5}$ Wm^{-2}. The fit (red line) is shown in Figure8 *Panel b)*.

- Our catalog (f_{BG} subtracted from peak fluxes): we find a power law index $\alpha = -2.09 \pm 0.04$ and $x_{min} = 5.5 \cdot 10^{-6}$ Wm^{-2}. The fit (red line) is shown in Figure8 *Panel c)*.

- Our catalog (f_{BG} subtracted from peak fluxes), we only show the maximum flux in the flare set with the same multiplicity flag: we find a power law index $\alpha = -2.08 \pm 0.04$ and $x_{min} = 5.9 \cdot 10^{-6}$ Wm^{-2}. The fit (red line) is shown in Figure8 *Panel d)*.

The results found demonstrate that our power-law indices, calculated taking into account the f_{BG}, agree with the values reported in the literature (see the brief discussion in the next section). For completeness, we underline that the algorithm used assumes that there is no upper bound to the empirical power-law distribution.

5. Conclusions and access to the catalogue

In this paper we present a catalogue of SXR events with a whole coverage over about three solar cycles based on GOES XRS data. A total of 334123 events have been identified and analyzed. If we limit the analysis to the period 1998-2020, taking into

account that NOAA did not use a standard definition for flare timing prior to May 1997, we then identify 205989 events, compared to the 38517 events listed in the official GOES catalog during the same period. In both cases we have an increase of about a factor 5 in the number of identified events. Obviously not all classes are incremented by the same factor. But the key point is that the statistics of the sample to be analyzed is greatly increased.

This result allows us to carry out a preliminary statistical analysis on our data. An analysis was performed on the 1998-2020 events in order to study the distributions of peak fluxes under different conditions. In particular, we carried out both a direct comparison of the distributions of the peak fluxes between our catalog and GOES list (see Figure 8 *Panels a, b*), and a comparison when we subtract from the peak flux the estimated background level of the GOES X-ray emission associated to the identified event.

If we limit the analysis to the latter case, in which we take into account the estimate of the background level, we observe (see Figure 8, *Panels c, d*) that the power law index is substantially the same if we consider all the 205989 identified events ($\alpha = -2.09 \pm 0.04$) or we consider only the most intense events (130925 events) in the various sequences of overlapping events ($\alpha = -2.08 \pm 0.04$). But above all it must be emphasized that distributions of all the 205989 identified events show a clear power-law behavior over more than three decades with a power law index $\alpha = -2.09 \pm 0.04$. The value of α we found is compatible within the errors with the power-law slope of $\alpha_F = -1.98 \pm 0.11$ for the (background-subtracted) soft X-ray peak fluxes distribution reported by Aschwanden & Freeland (2012).

This preliminary analysis and the comparison with similar works reported in the literature (e.g. Veronig et al., 2002; Yashiro et al., 2006; Aschwanden & Freeland, 2012) makes us confident of the goodness of the algorithm developed. The next step will consist in using the other physical parameters calculated for the events in order to study other statistical properties of the flares such as the estimated total energy distribution, waiting times, duration or these properties associated with the multiplicity of events. These analyses, which have already begun, will be the subject of future work.

The whole list of identified flares and all the outputs of the procedure are available through the web page `https://github.com/nplutino/FlareList` together with Python scripts to retrieve selected data that match user-defined criteria.

6. Acknowledgments

References

Alstott, J., & Bullmore, D. P. (2014). powerlaw: a python package for analysis of heavy-tailed distributions. *PloS one*, *9*(1).

Aschwanden, M. J., & Freeland, S. L. (2012). Automated solar flare statistics in soft x-rays over 37 years of goes observations: The invariance of self-organized criticality during three solar cycles. *The Astrophysical Journal*, *754*(2), 112.

Berrilli, F., Casolino, M., Cristaldi, A., Del Moro, D., Forte, R., Giovannelli, L., Martucci, M., Mergé, M., Napoletano, G., Narici, L., Pietropaolo, E., Pucacco, G., Rizzo, A., Scardigli, S., & Sparvoli, R. (2019). Introducing SWERTO: A regional space weather service. *Nuovo Cimento C Geophysics Space Physics C*, *42*(1), 47. doi:`10.1393/ncc/i2019-19047-4`.

Berrilli, F., & Giovannelli, L. (2022). The Great Aurora of 4 February 1872 observed by Angelo Secchi in Rome. *Journal of Space Weather and Space Climate*, *12*, 3. doi:`10.1051/swsc/2021046`. arXiv:`2201.01171`.

Bobrinsky, N., & Del Monte, L. (2010). The space situational awareness program of the European Space Agency. *Cosmic Research*, *48*(5), 392–398. doi:`10.1134/S0010952510050035`.

Boffetta, G., Carbone, V., Giuliani, P., Veltri, P., & Vulpiani, A. (1999). Power laws in solar flares: self-organized criticality or turbulence? *Physical review letters*, *83*(22), 4662.

Charbonneau, P., McIntosh, S. W., Liu, H.-L., & Bogdan, T. J. (2001). Avalanche models for solar flares (invited review). *Solar Physics*, *203*(2), 321–353.

Cicogna, D., Berrilli, F., Calchetti, D., Del Moro, D., Giovannelli, L., Benvenuto, F., Campi, C., Guastavino, S., & Piana, M. (2021). Flare-forecasting algorithms based on high-gradient polarity inversion lines in active regions. *The Astrophysical Journal*, *915*(1), 38.

Crosby, N. B., Aschwanden, M. J., & Dennis, B. R. (1993). Frequency distributions and correlations of solar x-ray flare parameters. *Solar Physics*, *143*(2), 275–299.

Crown, M. D. (2012). Validation of the NOAA Space Weather Prediction Center's solar flare forecasting look-up table and forecaster-issued probabilities. *Space Weather*, *10*, S06006. doi:10.1029/2011SW000760.

Fletcher, L. (2005). The observational motivation for computational advances in solar flare physics. *Space science reviews*, *121*(1), 141–152.

Giuliani, P., Carbone, V., Veltri, P., Boffetta, G., & Vulpiani, A. (2000). Waiting time statistics in solar flares. *Physica A Statistical Mechanics and its Applications*, *280*, 75–80. doi:10.1016/S0378-4371(99)00620-2.

Goldstein, M. L., Morris, S. A., & Yen, G. G. (2004). Problems with fitting to the power-law distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, *41*(2), 255–258.

Ishii, M. (2017). Japanese space weather research activities. *Space Weather*, *15*(1), 26–35. doi:10.1002/2016SW001531.

Kashapova, L. K., Broomhall, A.-M., Larionova, A. I., Kupriyanova, E. G., & Motyk, I. D. (2021). The morphology of average solar flare time profiles from observations of the Sun's lower atmosphere. *Monthly Notices of the Royal Astronomical Society*, *502*(3), 3922–3931. doi:10.1093/mnras/stab276. arXiv:2102.02596.

Li, Y.-P, Feng, L., Zhang, P., Liu, S.-M., & Gan, W.-Q. (2016). On the power-law distributions of x-ray fluxes from solar flares observed with goes. *Research in Astronomy and Astrophysics*, *16*(10), 161.

Lin, R., & Hudson, H. (1976). Non-thermal processes in large solar flares. *Solar Physics*, *50*(1), 153–178.

Lu, L., Feng, L., Li, D., Ying, B., Li, H., Gan, W., Li, Y., & Zhou, J. (2021). Catalog and statistical examinations of lyα solar flares from goes/euvs-e measurements. *The Astrophysical Journal Supplement Series*, *253*(1), 29.

McKinney, W. (2011). pandas: a foundational python library for data analysis and statistics. *Python for high performance and scientific computing*, *14*(9), 1–9.

Mumford, S. J., Christe, S., Pérez-Suárez, D., Ireland, J., Shih, A. Y., Inglis, A. R., Liedtke, S., Hewett, R. J., Mayer, F., Hughitt, K. et al. (2015). Sunpy—python for solar physics. *Computational Science & Discovery*, *8*(1), 014009.

Murray, S. A., Bingham, S., Sharpe, M., & Jackson, D. R. (2017). Flare forecasting at the Met Office Space Weather Operations Centre. *Space Weather*, *15*(4), 577–588. doi:10.1002/2016SW001579. arXiv:1703.06754.

Park, S.-H., Leka, K., Kusano, K., Andries, J., Barnes, G., Bingham, S., Bloomfield, D. S., McCloskey, A. E., Delouille, V., Falconer, D. et al. (2020). A comparison of flare forecasting methods. iv. evaluating consecutive-day forecasting patterns. *The Astrophysical Journal*, *890*(2), 124.

Plainaki, C., Antonucci, M., Bemporad, A., Berrilli, F., Bertucci, B., Castronuovo, M., De Michelis, P., Giardino, M., Iuppa, R., Laurenza, M., Marcucci, F., Messerotti, M., Narici, L., Negri, B., Nozzoli, F., Orsini, S., Romano, V., Cavallini, E., Polenta, G., & Ippolito, A. (2020). Current state and perspectives of Space Weather science in Italy. *Journal of Space Weather and Space Climate*, *10*, 6. doi:10.1051/swsc/2020003.

Priest, E., & Forbes, T. (2002). The magnetic nature of solar flares. *The Astronomy and Astrophysics Review*, *10*(4), 313–377.

Priest, E. R. (1986). Magnetohydrodynamic Theories of Solar Flares. *Solar Physics*, *104*(1), 1–18. doi:10.1007/BF00159941.

Pucci, F., & Velli, M. (2013). Reconnection of quasi-singular current sheets: the "ideal" tearing mode. *The Astrophysical Journal Letters*, *780*(2), L19.

Reep, J. W., & Knizhnik, K. J. (2019). What determines the x-ray intensity and duration of a solar flare? *The Astrophysical Journal*, *874*(2), 157.

Ryan, D. F., Milligan, R. O., Gallagher, P. T., Dennis, B. R., Tolbert, A. K., Schwartz, R. A., & Young, C. A. (2012). The Thermal Properties of Solar Flares over Three Solar Cycles Using GOES X-Ray Observations. *The Astrophysical Journal Supplement Series*, *202*(2), 11. doi:10.1088/0067-0049/202/2/11. arXiv:1206.1005.

Sadykov, V. M., Kosovichev, A. G., Kitiashvili, I. N., & Frolov, A. (2019). Statistical Properties of Soft X-Ray Emission of Solar Flares. *The Astrophysical Journal*, *874*(1), 19. doi:10.3847/1538-4357/ab06c3. arXiv:1810.05610.

Schwenn, R. (2006). Space weather: The solar perspective. *Living reviews in solar physics*, *3*(1), 1–72.

Smart, D. F., Shea, M. A., & McCracken, K. G. (2006). The Carrington event: Possible solar proton intensity time profile. *Advances in Space Research*, *38*(2), 215–225. doi:10.1016/j.asr.2005.04.116.

Swalwell, B., Dalla, S., Kahler, S., White, S. M., Ling, A., Viereck, R., & Veronig, A. (2018). The reported durations of goes soft x-ray flares in different solar cycles. *Space Weather*, *16*(6), 660–666.

Temmer, M., Thalmann, J. K., Dissauer, K., Veronig, A. M., Tschernitz, J., Hinterreiter, J., & Rodriguez, L. (2017). On flare-cme characteristics from sun to earth combining remote-sensing image data with in situ measurements supported by modeling. *Solar physics*, *292*(7), 1–22.

Toriumi, S., & Wang, H. (2019). Flare-productive active regions. *Living Reviews in Solar Physics*, *16*(1), 3. doi:10.1007/s41116-019-0019-7. arXiv:1904.12027.

Török, T., Kliem, B., & Titov, V. (2004). Ideal kink instability of a magnetic loop equilibrium. *Astronomy & Astrophysics*, *413*(3), L27–L30.

Verbeeck, C., Kraaikamp, E., Ryan, D. F., & Podladchikova, O. (2019). Solar flare distributions: Lognormal instead of power law? *The Astrophysical Journal*, *884*(1), 50.

Veronig, A., Temmer, M., Hanslmeier, A., Otruba, W., & Messerotti, M. (2002). Temporal aspects and frequency distributions of solar soft x-ray flares. *Astronomy & Astrophysics*, *382*(3), 1070–1080.

Viticchié, B., Del Moro, D., & Berrilli, F. (2006). Statistical properties of synthetic nanoflares. *The Astrophysical Journal*, *652*(2), 1734–1739.

Vlahos, L., Georgoulis, M., Kluiving, R., & Paschos, P. (1995). The statistical flare. *Astronomy and Astrophysics*, *299*, 897–911.

Wagner, W. J. (1988). Observations of 1-8 Å solar X-ray variability during solar cycle 21. *Advances in Space Research*, *8*(7), 67–76. doi:10.1016/0273-1177(88)90173-1.

Yashiro, S., Akiyama, S., Gopalswamy, N., & Howard, R. A. (2006). Different Power-Law Indices in the Frequency Distributions of Flares with and without Coronal Mass Ejections. *The Astrophysical Journal*, *650*(2), L143–L146. doi:10.1086/508876. arXiv:astro-ph/0609197.

Zhang, J., & Dere, K. (2006). A statistical study of main and residual accelerations of coronal mass ejections. *The Astrophysical Journal*, *649*(2), 1100–1109.

Zhang, J., Dere, K., Howard, R., Kundu, M., & White, S. (2001). On the temporal relationship between coronal mass ejections and flares. *The Astrophysical Journal*, *559*(1), 452–462.

Appendix A. Comparison between our algorithm and that of Aschwanden & Freeland (2012)

Aschwanden & Freeland (2012) Algorithm	Proposed Algorithm
Data rebinning to obtain a 12s resolution	Same technique
Elimination of data gaps and spike removal (Q parameter). The article does not specify how spikes are treated	We used the same criterium to find spikes. We tested data from year 1987, the year which presented the biggest number of spikes, and decided to remove portions of data affected by the problem, since they only represent a very small percentage of data
Definition of a minimum flare duration (5 bins or 60s)	We do not set a minimum flare duration, the maximum and minimum detection procedure used automatically solves this problem
Boxcar smoothing using a 21-points window. The article does not specify what kind of window has been used	We used a window of 21 bins and replaced the value at the center of the bin with the mean value over the window
All local maximum and minimum points obtained from the smoothed light curve are used as candidates for peak, start and end times	Even after the smoothing process, light curves still show small fluctuations. Not all maximum and minimum points are registered. A data point is considered maximum (minimum) only if its value is higher (lower) than the values of the two points on its left and of the two points on its right
The background is defined as the median flux in a time interval (60s) before the start of the event	Same definitions but we use the mean value instead of the median one
Definition of a noise parameter with a value $f_{noise} = 2 \cdot 10^{-8}$ W m^{-2}	Same definition
Threshold criteria to associate each starting point to an end point. The same ending point cannot be used for two different starting points. Intervals are therefore truncated at the start of another event	We allow the same ending point to be matched with different starting points and use the information to label multiple and single events. This is meaningful when considering the time duration of events and their energy
Given an interval the maximum value is registered as peak flux. All peak values are registered and sometimes the peak flux can be lower than the background, giving a negative absolute value $F = f_p - f_{BG} < 0$	We apply a threshold criterium to the maximum point too. Since we require $f_p > f_{BG} + f_{treshold}$, with $f_{treshold} = f_{noise}$ we are excluding extremely small events which can be associated with data noise. At the same time the condition $F = f_p - f_{BG} > 0$ is guaranteed
Event duration is truncated at the beginning of the following one	With the information about overlapping events the original duration of the event can be found. Each event can be interpreted as a single event or, if overlapping in time with another, as part of a group
No information about the energy associated to the event	Our algorithm registers the energy associated to each event. Using the Multiple Id, single event energies can be added together to get the total energy