Flare-forecasting Algorithms Based on High-gradient Polarity Inversion Lines in Active Regions

Domenico Cicogna$^{1,2}$, Francesco Berrilli$^{1}$, Daniele Calchetti$^{1}$, Dario Del Moro$^{1}$, Luca Giovannelli$^{1}$, Federico Benvenuto$^{3}$, Cristina Campi$^{3}$, Sabrina Guastavino$^{1}$, and Michele Piana$^{3}$

1 Department of Physics, University of Rome Tor Vergata, Via della Ricerca Scientifica, 1, I-00133, Rome, Italy; francesco.berrilli@roma2.infn.it
2 NEXT Ingegneria dei Sistemi S.p.A., Rome, Italy
3 Dipartimento di Matematica, Università di Genova, via Dodecaneso 35, I-16146 Genova, Italy

Received 2020 December 13; revised 2021 April 11; accepted 2021 April 21; published 2021 July 1

Abstract

Solar flares emanate from solar active regions hosting complex and strong bipolar magnetic fluxes. Estimating the probability of an active region to flare and defining reliable precursors of intense flares are extremely challenging tasks in the space weather field. In this work, we focus on two metrics as flare precursors, the unsigned flux $R$, tested on Michelson Doppler Imager/Solar and Heliospheric Observatory data, and a novel topological parameter $D$, representing the complexity of a solar active region. In greater detail, we propose an algorithm for the computation of the $R$ value, which exploits the higher spatial resolution of Helioseismic Magnetic Imager maps. This algorithm leads to a differently computed $R$ value, whose functionality is tested on a set of solar cycle 24 flares. Furthermore, we introduce a topological parameter based on the automatic recognition of magnetic polarity inversion lines in identified active regions and are able to evaluate its magnetic topological complexity. We use both a heuristic approach and a supervised machine-learning method to validate the effectiveness of these two descriptors to predict the occurrence of X- or M-class flares in a given solar active region during the following 24 hr period. Our feature ranking analysis shows that both parameters play a significant role in prediction performances. Moreover, the analysis demonstrates that the new topological parameter $D$ is the only one, among 173 overall predictors, that is always present for all test subsets and is systematically ranked within the top 10 positions in all tests concerning the computation of the weights with which each predictor impacts the flare forecasting.

Unified Astronomy Thesaurus concepts: Solar flares (1496); Solar activity (1475); Solar physics (1476); Solar active regions (1974); Solar active region magnetic fields (1975)

1. Introduction

Solar flares represent the most energetic explosive events in our solar system. They consist of sudden and powerful coronal events that are triggered by plasma instabilities and energized by reconnections of the magnetic field present in coronal loops (e.g., Priest & Forbes 2002), which in turn are controlled by their foot point’s advection by plasma flows in the photosphere (e.g., Viticchié et al. 2006; Reale 2014; Caroli et al. 2015; Kuridze et al. 2019). Their most impressive manifestation is the sudden release of large amounts of energy, mainly emitted in the high-frequency regions (e.g., above extreme ultraviolet (EUV)) of the solar electromagnetic spectrum (see Fletcher et al. 2011; Toriumi & Wang 2019 and references cited therein). These bursts of energy, in addition to producing a consistent increase in the high-energy radiative flux, are often accompanied by intense fluxes of solar energetic particles, mainly electrons and protons, and sometimes by coronal mass ejections (CMEs; e.g., Chen 2017). Energetic particles, CMEs, frozen in solar magnetic field, and high-energy radiation cross the interplanetary space and interact with the magnetosphere and upper atmosphere of the planets or with artificial satellites that are in their path, generating the physical processes associated with circumterrestrial and planetary space weather (e.g., Dorman 2005; Plainaki et al. 2016).

Indeed, solar energetic particles and high-energy radiation pose severe threats to satellites operating in higher orbits than low-Earth orbits, and more importantly, to astronauts operating outside the Earth’s protective magnetosphere, especially if involved in extravehicular activity (e.g., Narici et al. 2018; Walsh et al. 2019). Moreover, the increased complexity of today’s society and its dependence on space technology represents a relevant risk factor, as many of the technologies used, in space or ground-based assets, are potentially vulnerable to solar flares (e.g., Cannon 2013; Berrilli et al. 2014; Di Fino et al. 2014). Operators of satellites, airlines, railways, and power grids are continuously confronted by a number of challenges, ranging from mitigating the effects of solar flares and CMEs on their assets to the growing need of carefully predicting these events. For this reason, many national and international institutions have developed forecasting services within large space weather projects (e.g., Kumar 2010; Crown 2012; Rodriguez et al. 2019; Vadakke Veettil et al. 2019; Plainaki et al. 2020), and research groups have worked on targeted flare or multi-flare-forecasting tools (e.g., Barnes et al. 2007; Georgoulis & Rust 2007; Schrijver 2007; Falconer et al. 2011; Korsós et al. 2015; Romano et al. 2015; Florios et al. 2018; McCluskey et al. 2018; Berrilli et al. 2019; Falco et al. 2019; Giovannelli et al. 2019; Lim et al. 2019; Cinto et al. 2020; Korsós et al. 2020; Nishizuka et al. 2020; Yi et al. 2020).

Usually, flare-forecasting algorithms described in the literature use an approach based on the physics of the processes necessary to trigger solar flares. In practice, those algorithms analyze line-of-sight (LOS) or vector magnetograms, or other maps of physical quantities such as the LOS velocity or spectral intensity, of the recognized active regions (ARs) to estimate single value quantities (i.e., descriptors) on the basis of several image analysis techniques (e.g., Campi et al. 2019). Forecasts
based on the extracted descriptors usually provide the probability of the occurrence of a flare of a given class within 24 hr and require a robust statistical analysis on a large number of events (or non-events), flaring and nonflaring ARs, and for a long time period. For this reason, different predictive statistical techniques have been applied, including machine-learning (ML) techniques (e.g., Barnes et al. 2016; Benvenuto et al. 2018, 2020; Leka et al. 2019a, 2019b; Campi et al. 2019; Piana et al. 2019; Park et al. 2020).

In this study we aim at assessing the flare-forecasting performances of two descriptors estimated by the analysis of the LOS magnetograms associated with a set of ARs. These ARs are both flaring and nonflaring within 24 hr and are acquired by the Helioseismic Magnetic Imager on board the Solar Dynamics Observatory (SDO/HMI) during solar cycle 24 (SC24).

The first predictor is a suitable generalization of the morphological metric \( R \), originally proposed by Schrijver (2007) for Solar and Heliospheric Observatory (SOHO)/Michelson Doppler Imager (MDI) LOS magnetograms acquired during solar cycle 23 by the MDI on board SOHO. The second feature is a new topology-based metric \( D \) that estimates the magnetic complexity of the AR by counting the number of polarity inversion lines (PILs) above a suitable magnetic field threshold. This metric, as we will discuss in detail in the following sections, represents an infrequent use of PILs, as it focuses on the analysis of the magnetic topology, rather than merely the morphology, of potentially flaring regions. In fact, as discussed in, e.g., Antiochos (1998), the complexity of the photospheric magnetic field, defining the right topology for magnetic breakout, is an essential element for strong activity such as large flares.

In the application of this analysis, the paper fulfills two main goals. The first one is to recalibrate the Schrijver (2007) algorithm for the calculation of \( R \), originally calibrated on LOS magnetograms with the resolution of SOHO/MDI, on LOS magnetograms at the full resolution of SDO/HMI. This predictor, which is scaled to have the same range as \( R \), and therefore to be easily compared to the original Schrijver’s feature, referred to as \( R' \) in the following. The second goal is to compare the performances of the two predictors, namely, \( R' \) and \( D \), by flaring and nonflaring scenarios in order to quantify their predictive capability and reliability.

The content of the paper is as follows. Section 2 describes the SDO/HMI data set of flaring and nonflaring regions, and the selection procedure. Section 3 overviews the recalibration process of Schrijver’s \( R \) feature, explaining in detail the numerical procedure to compute the \( R \) value for SDO/HMI full resolution magnetograms and the calibration of the original log \( R \) and the new log \( R' \). Section 4 introduces the new topological descriptor \( D \), shortly explains the conceptual background of this new feature, and explains in detail the numerical procedure to compute it. Section 5 contains the results of the forecasting analysis using \( R' \) and \( D \), while our conclusions are offered in Section 6.

2. The SDO/HMI Data Set of Flaring and Nonflaring Regions

The first task to perform for testing and calibrating \( R' \) and \( D \) was to create the magnetogram data set of nonflaring and flaring regions during SC24. A catalog of all the solar flares detected by the Geostationary Operational Environmental

Satellites (GOES) is available from the Space Weather Prediction Center at the National Oceanic and Atmospheric Administration (NOAA). The GOES flare catalog provides, among other information, the flare class, the flare start time, and the NOAA AR number. Flares are detected in X-ray flux in the 0.1–0.8 nm channel of GOES \( F \), and their class is determined by the peak in such curve \( f_{\text{peak}} \). We selected from the NOAA catalog the list of all M and X flares, from 2010 to 2018 June. The magnetogram data set was selected from the HMI Active Region Patches (HARPs) available from the Joint Science Operations Center (JSOC)\(^5\), which provides LOS magnetograms (MHARPs) from the SDO/HMI instrument (Schou et al. 2012) (specifically, we used the hmi.sharp_720s_data stream). The HARP pipeline identifies and tracks ARs in the solar photosphere, generating a time series of magnetogram patches for each AR.

The data set has been created by adopting the same criteria reported in Schrijver (2007). In greater detail, we selected LOS magnetograms within 45° from the disk center for each day from 2010 to 2018 June. Limiting the data set to near the disk center reduces the influence of projection effects on LOS magnetograms. We then compared the magnetogram data set with the GOES flare list to label magnetograms as flaring or nonflaring.

On the one hand, the flaring set contains only regions hosting events with peak flux greater than \( 10^{-5} \) W m\(^{-2} \) (M or greater class flares). We considered only ARs hosting a single flare since in the case of ARs flaring multiple times over a period of days, \( R \) remains high as magnetic flux continues to emerge (see, e.g., Piersanti et al. 2017). This set is composed of 100 magnetograms of 100 different HARPs hosting an M- or X-class flare in the next 24 hr. On the other hand, the nonflaring set is composed by a collection of HARPs that did not produce any M- or X-class flare during their lifetime. This set can contain more than one magnetogram per HARP, separated by at least 24 hr. This control group includes 745 total magnetograms. This number of magnetograms, belonging to nonflaring regions, keeps the ratio between flaring and nonflaring regions equal to that used to calibrate \( R \) in Schrijver (2007). It is worth noting that the total statistics of the analyzed regions in this work is however different from that of Schrijver (2007), due to the different level of activity between cycle 23 and cycle 24. The properties of the data set are listed in Table 1.

In the following, we use this data set for the computation of \( R' \) and \( D \) and to assess their forecasting capability. Moreover, we also compare the two new descriptors with the properties

\(^4\) https://www.ngdc.noaa.gov/stp/satellite/goes/index.html

\(^5\) http://jsoc.stanford.edu/jsocwiki/HARPSeries
developed by the FLARECAST Consortium (http://flarecast.eu) using Near-Realtime Space Weather HMI Archive Patch (SHARP) data (Bobra et al. 2014).

3. SDO/HMI $R^*$ Value Algorithm

Among the physics-based descriptors, i.e., among the descriptors that estimate the probability of the occurrence of a flare on the basis of physical properties derived from maps of the selected AR, the $R$ value (Schrijver 2007) is of great significance. The computation of $R$ is based on the hypothesis that the emergence of electrical currents embedded in the magnetic flux is a key ingredient in the triggering of flares (e.g., Wheatland 2000; Schrijver et al. 2005). The presence of these currents is strongly linked to the presence of PILs, which are regions in which there is an intense vertical component of the photospheric magnetic field ($B_z$) of opposite polarity. Magnetograms from ground-based instruments (e.g., Caivallini et al. 2002; Caivallini 2006; Scharmer et al. 2008; Steward et al. 2011; Puschmann et al. 2012; Tadesse et al. 2013; Del Moro 2002; Cavallini 2006; Scharmer et al. 2008; Steward et al. 2011) or on board satellites (e.g., Scherrer et al. 1995; Suematsu et al. 2017; Peter et al. 2012; Scherrer et al. 2012; Berrilli et al. 2015) can be used to detect PILs. The $R$ value is defined in Schrijver (2007) and is a measure of the unsigned photospheric magnetic flux close to selected PILs, where an appropriate weighting map is applied for the SOHO/MDI instrument.

The MDI (Scherrer et al. 1995) instrument was active from 1996 to 2011 April on board the SOHO satellite. As a legacy of the MDI instrument, since 2010 SDO/HMI Scherrer et al. (2012) have been providing full vector solar magnetograms, with higher spatial and temporal resolution. Indeed, the cadence of the tracked Active Regions Patches for HMI is 12 minutes and the images have a spatial resolution of $1''$, while MDI had a 96 minute cadence and a spatial resolution equal to $4''$. The pixel scales for HMI and MDI are $0.5''$/pixel and $2''$/pixel, respectively. The spectral line used for the computation of the vector magnetic field maps are different. MDI used the Ni 676.8 nm line, while HMI uses the Fe i 617.3 nm line (Norton et al. 2006). The line-forming heights are substantially similar, the models suggesting 125 km for the Fe i 617.3 nm line and 100 km for the Ni 676.8 nm line (Fleck et al. 2011). Clearly, the differences in the experimental setup between the two instruments affect the magnetograms produced. Their use for the calculation of the descriptors therefore requires an accurate calibration. The comparison of the LOS magnetograms acquired during the overlapping period of the two missions revealed that the LOS magnetic signal inferred by MDI is larger than that derived from HMI by a factor of 1.40 (Liu et al. 2012). The same authors provided a recipe to convert HMI images to MDI-like images, adapting the temporal cadence, spatial resolution, and calibration of the magnetic field strength. In the present study, we adapt the original (Schrijver 2007) algorithm, for the $R$ value calculation, so that it can be applied to the full spatial resolution HMI magnetograms, and we benchmark its performance. We name the new descriptor $R^*$. The procedure and the tests to compute and benchmark $R^*$ are as follows. First of all, we appropriately rescale the dilation kernels necessary to identify the high-gradient polariety-separation lines, maintaining the correct spatial scales. The bitmaps of HMI magnetograms, both for positive and negative magnetic flux densities, are dilated with kernels of $9 \times 9$ pixels (see Figure 1). To create the HMI bitmaps we keep the same threshold in flux densities (i.e., 150 G) originally used by Schrijver (2007), despite the difference in the magnetic sensitivity of the MDI and HMI instruments discussed above. The reason for this is that we want to consider only high-gradient regions.

The two dilated bitmaps are then overlapped to find the pixels where their product is nonzero, showing the high-gradient polarity-separation lines. Then, the bitmap containing the polarity-separation lines is convolved with a Gaussian function with an FWHM of $D_{\text{ap}} = 15\ Mm$ to calculate a weighting map. The value of $15\ Mm$ is the average maximum distance between the occurring flare in EUV images of the corona and the nearest point of any PILs in the region, according to Schrijver (2007). This weighting map is then multiplied by the absolute value of the flux in the magnetograms.

Finally, in order to estimate the new value $R^*$, we take into account the difference in pixel area of MDI and HMI. Indeed, the pixel areas are $\approx 2.2 \ Mm^2$ pixel$^{-1}$ for MDI and $\approx 0.14 \ Mm^2$ pixel$^{-1}$ for HMI, respectively. The $R^*$ value is eventually obtained from the weighted unsigned flux summed over all pixels.

There are various reasons why we think it is of interest to generalize the morphological metric $R$. First of all, we can potentially take advantage of full resolution of the SDO/HMI magnetograms, but especially we can study the behavior of this parameter when the algorithm is tuned on instruments that have different spatial and spectral characteristics. This fact widens the applicability of the algorithm and is useful to more easily...
intercalibrate metrics derived by magnetograms produced by different instruments. Another important aspect is the replication of the statistical analysis, according to the procedure described in Schrijver (2007), and the complementary supervised ML analysis, in order to verify the behavior of the new metric on a second solar cycle. Applying the modified descriptor to a new solar cycle means extending the reliability analysis and confirming its forecasting capabilities.

A first validation of the $R^*$ descriptor has been performed by comparing the log $R^*$ values with the values of log $R$ recorded in the Joint Science Operations Center (JSOC) data header, log $R_{\text{JSOC}}$, and with the estimate scaled to the resolution of SOHO/MDI magnetograms according to the recipe presented in Liu et al. (2012), log $R_{\text{MDIL}}$. In the left panel of Figure 2, we show a scatter plot of the values of log $R_{\text{JSOC}}$ and log $R_{\text{MDIL}}$ compared to log $R^*$ for the magnetograms of 100 different HARPs hosting an M- or X-class flare in the next 24 hr described in Section 2. Combining the log $R$ values from these three different methods, we find a strong correlation between log $R_{\text{JSOC}}$ and log $R_{\text{MDIL}}$ and the estimations from our algorithm, log $R^*$.

4. $D$ Value Algorithm

It has been known for a long time that the probability of an occurrence of a flare is higher in magnetically complex ARs, e.g., Mount Wilson type $\gamma$ spot groups, compared to magnetically simple ones (e.g., Giovanelli 1939; Howard 1964; Antiochos 1998). For this reason, for example, prediction algorithms based on a morphological measure of roughness, as the fractal or multifractal dimension, of the ARs have been proposed. However, these geometric properties have proved more useful in studying the scale properties of solar magnetism than in characterizing ARs in terms of eruptive ability or flare forecasting (e.g., Georgoulis 2012). The algorithms that measure morphological quantities related to the absolute magnetic vertical-field component $B_s$ (approximated by the LOS field component near disk center) have turned out to be much more efficient in terms of predicting flares. Among these, in addition to Schrijver’s $R$ value (Schrijver 2007), that we have extensively discussed in the previous section, we find the WL$_{SS}$ and WL$_{SG}$ parameters of Falconer et al. (2009) or the procedures based on gradient-weighted inversion-line length (Mason & Hoeksema 2010). As reported in Georgoulis (2013), not all of these parameters weigh heavily on PILs, although with different specifics.

The proposed $D$ value, although based on the identification of PILs, is not based on morphological measures such as the partial or total measurement of the length of the PILs or a $B_s$ weighted estimate in the AR. Indeed, $D$ is not a morphological descriptor, e.g., based on $B_s$ shape or distribution. On the contrary, it is a topological descriptor, therefore linked to quantitative measurement of magnetic features (the total number of PILs) in an AR. The $D$ value, although very simple to calculate as we will discuss below, has similarities with the connectivity matrix, which is sensitive to flux partitioning, and which is based on a flux-tessellation scheme used for topological analyses (Georgoulis 2013) or with the connection and persistence diagram analysis used to evaluate the information about the spatial scales of the topological features in ARs (Deshmukh et al. 2020).

As with other complexity metrics, the topological complexity increases with the number of loops (e.g., Schwarz 2013), i.e., of PILs. If we also consider that the amount of magnetic flux near a PIL gives us an estimate of the reservoir of magnetic energy potentially available for the flare, it is natural to combine the two quantities and introduce a new parameter based on the number of PILs present in an AR. With this motivation, we define this new metric to estimate the probability of triggering such an energy release. Our assumption is that a complex and fragmented magnetic region could be more effective in accumulating free energy and triggering the flare process.

Therefore, the proposed topological descriptor $D$ counts the number of separate PIL fragments present in the ARs, i.e., the number of different polarity inversion lines in the $B_{i\text{LOS}}$ magnetogram of the selected AR. Our implementation to calculate $D$ exploits a fast thresholding and labeling procedure. Starting from the $B_{i\text{LOS}}$ magnetogram $M(x, y)$ of the selected AR, indicated as magnetogram in Figure 3, we apply an absolute thresholding technique (e.g., Berrilli et al. 2005; Schrijver 2007) to prepare three different bitmap images: $I_+$, $I_-$, and $I_U$. 

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Left panel: scatter diagram of the correlation between log $R_{\text{JSOC}}$ as obtained from the JSOC magnetograms header (red diamonds) and log $R_{\text{MDIL}}$ as obtained from reprocessing SDO/HMI magnetograms in MDI-like magnetograms, according to the recipe presented in Liu et al. (2012) and applying the Schrijver (2007) algorithm (black crosses). The points actually shown in this plot are 99 since an outlier with log $R_{\text{MDIL}} = 0$ and log $R = 1.8$ was excluded. Right panel: scatter diagrams of the correlation between $D$ and the features wslg_blos/value_int (top left), mpil_blos/tot_length (top right), mpil_blos/max_length (bottom left), and mpil_blos/tot_usflux (bottom right). The definition of these descriptors can be found in Table 2.}
\end{figure}
The three bitmap images (see Figure 3) are

\[
I_+(x, y) = \begin{cases} 
1 & \text{if } M(x, y) \geq 150 \text{ G} \\
0 & \text{otherwise}
\end{cases},
\]

\[
I_-(x, y) = \begin{cases} 
1 & \text{if } M(x, y) \leq -150 \text{ G} \\
0 & \text{otherwise}
\end{cases},
\]

\[
I_U(x, y) = \begin{cases} 
1 & \text{if } |M(x, y)| \geq 150 \text{ G} \\
0 & \text{otherwise}
\end{cases}.
\]

Eventually, a classical labeling technique identifies and counts the number of compact flux density structures (composed of at least two connected pixels) in the three images, \(N_U, N_+,\) and \(N_-,\) respectively. The difference \(D = (N_+ + N_-) - N_U\) counts the number of PIL fragments in the magnetogram.

5. Flare-forecasting Benchmarks

5.1. Flare Forecasting with \(R^*\) and \(D:\) A Heuristic Approach

In order to test the effectiveness of \(R^*\) and \(D\) as statistical classifiers for flare forecasting, we first considered their distributions for all the 845 AR magnetograms, see Figure 4. In the left panel of Figure 4, we show the distribution of the number of events as a function of log \(R^*\). Almost no M- or X-class flare occurred within 24 hr from the calculation of \(R^*\) in a region with log \(R^* < 3.2\), which includes about 40% of nonflaring ARs. Only about 15% of the regions with log \(R^* > 3.2\) showed an M- or X-class flare in the next 24 hr. Among the 100 flaring ARs, 37% of them have log \(R^* \approx 5\) or greater.

The scatter plot in Figure 5 correlates the GOES flux density with the value of log \(R^*\); we notice the same behavior that was found for solar cycle 23 by Schrijver (2007), with a clustering of flares in the lower right corner. This result suggests that \(R^*\) measures the maximum energy available for flares in a given AR given by \(F_{\text{max}} = 1.2 \times 10^{-8} R^*\). As already noted by Schrijver (2007), it also indicates that ARs reduce their energy...
through multiple flares, with greater probability of smaller flares; even at very high \( R^* \) values, M-class events are far more frequent than X class. Comparing the results of this analysis with those shown in Schrijver (2007) is of course an interesting exercise since the latter ones were obtained from the analysis of ARs appeared during the 23rd solar cycle, while those used in the present study came from the 24th one, which was far less active and with a much smaller number of ARs. However, despite these differences between the two cycles, the general trend seems to be very similar (see the exact values in Table 3).

Namely, there is an almost complete absence of flaring ARs for \( \log R^* < 3 \); about 39% of the ARs with \( \log R^* \approx 4.5 \) flares in the next 24 hr; the flaring fraction is 93% for those ARs with \( \log R^* \approx 5 \) or greater.

As far as descriptor \( D \) is concerned, the right panel of Figure 4 shows that about 10% of the magnetograms in the flaring AR set (with a flaring probability less than 2%) and 70% of the magnetograms in the nonflaring set have \( D = 0 \). About 41% of the ARs with \( D \geq 1 \) have a major flare in the next 24 hr. For \( D = 6 \), the probability of having a major flare in the next 24 hr grows to 60%, and for \( D \geq 10 \) it almost reaches 100% (see the exact values in Table 3). We completed this preliminary analysis by checking the possible correlation between these two descriptors: Figure 6 shows a clear trend of flaring events (represented with red diamonds in the plot) for higher values of the two features.

Relying on the results of this preliminary analysis, we investigated the forecasting power of the two descriptors, both separately and jointly.

As a first heuristic approach, we first considered different thresholds of \( \log R^* \) and \( D \) to create deterministic forecasts for M-class flares or more intense in the next 24 hr. We measured the performance of such binary classifiers by means of commonly used classification evaluation metrics. Specifically, as in Bobra & Couvidat (2015), we constructed confusion

---

### Table 2

| Descriptor | Definition |
|------------|------------|
| wls_blos/value_int and wls_br/value_int | \( B_{\text{los}} \) and \( B_{\text{radial}} \) Falconers' WLSG |
| mpil_blos/tot_length | Total length of all MPILs |
| mpil_blos/max_length | Maximum length of a single MPIL |
| mpil_blos/tot_usflux | Total unsigned flux around all MPILs |
| sfunction_blos/zq | Inertial range index |
| sharp_kw/snetjpp/total | Sum of absolute value of net currents per polarity |
| sharp_kw/us/hstddev | Standard deviation of the unsigned vertical current helicity |
| decay_index_br/maxl_over_hmin | Maximum ratio of MPIL length to minimum height of critical decay index |
| helicity_energy_bvec/abs_tot_dedt_in | Absolute value net vertical Poynting flux |

---

**Figure 4.** Left panel: histograms of \( \log R^* \) for all the 845 AR magnetograms. The regions with \( R^* = 0 \) are not shown. Right panel: histogram of \( D \) for the total 845 AR magnetograms.

**Figure 5.** Scatter diagram of the peak flux density in the GOES 0.1–8 nm band \( (\text{W m}^{-2}) \) vs. \( \log(R^*) \) for all flaring regions. The red dashed line at \( F_{\text{max}} = 1.2 \cdot 10^3 R^* \) defines the green region where all the flares appear. We notice that flares from the 24th solar cycle show a similar behavior to those from the 23rd one (compare this figure to Figure 3 in Schrijver 2007).

As a first heuristic approach, we first considered different thresholds of \( \log R^* \) and \( D \) to create deterministic forecasts for M-class flares or more intense in the next 24 hr. We measured the performance of such binary classifiers by means of commonly used classification evaluation metrics. Specifically, as in Bobra & Couvidat (2015), we constructed confusion.
The bin size for log $R^*$ is 0.5 with the exception of the first and the last bin.

Note. The three columns in the box identified by log $R^*$ refer to the prediction individually built on $R^*$. The three columns in the box identified by $D$ refer to the prediction individually built on $D$. The three columns in the box identified by log $R^*$ and $D$ refer to the prediction built with the joint use of log $R^*$ and $D$.

The Astrophysical Journal, 915:38 (12pp), 2021 July 1
Cicogn et al.

Table 3
Likelihood of M or X Flare Within 24 hr from the Determination of the log $R^*$ and $D$ Parameters, Similarly to Table 1 in Schrijver (2007)

| Class  | log $R^*$ $< 3.25$ (%) | $3.25 \leq$ log $R^*$ $< 3.75$ (%) | $3.75 \leq$ log $R^*$ $< 4.25$ (%) | $4.25 \leq$ log $R^*$ $< 4.75$ (%) | log $R^*$ $\geq 4.75$ (%) |
|-------|-----------------------|----------------------------------|----------------------------------|----------------------------------|--------------------------|
| $>$M1 | ~0                    | 3                                | 9                                | 39                               | 93                       |
| $>$X1 | 0                     | 0                                | 0                                | 1                                | 23                       |

Table 4
TPR, TS, and HSS Scores for Different M-class Flare-forecasting Algorithms Created by Setting Different $R^*$ and $D$ Thresholds

| Threshold | log $R^*$ $>$ 4, $D$ $\geq 5$ | log $R^*$ $>$ 4.5, $D$ $\geq 7$ | log $R^*$ $>$ 5, $D$ $\geq 10$ |
|-----------|-------------------------------|---------------------------------|-------------------------------|
| log $R^*$ | TPR                           | TSS                             | HSS                           |
|           | 0.85                          | 0.66                            | 0.42                          |
|           | 0.62                          | 0.59                            | 0.62                          |
|           | 0.12                          | 0.12                            | 0.19                          |

Figure 6. Scatter plot of log $R^*$ vs. $D$ parameter. The flaring ARs are shown with diamonds, while the nonflaring ARs with dots. ARs with $R^* = 0$ are not shown.

matrices according to the following scheme: we classified as true positives (TPs) all flaring ARs that have been correctly predicted as flaring; as true negatives (TNs) all nonflaring ARs that have been correctly predicted as nonflaring; as false negatives (FNs) all flaring ARs that have been incorrectly predicted as nonflaring; and as false positives (FPs) all nonflaring ARs incorrectly predicted as flaring. From these quantities, we could compute the following skill scores to assess the classifying performances of the two descriptors. The true positive rate (TPR) is defined as

$$TPR = \frac{TP}{TP + FN} \quad (1)$$

and is the ability of the classifier to identify all of the positive events (in brief, a TPR = 1 means that the classifier will recognize all the ARs that had a flare in the next 24 hr). We also consider the true skill statistics (TSS)

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN} \quad (2)$$

and the Heidke skill score (HSS)

$$HSS = \frac{2(TP \cdot TN - FN \cdot FP)}{(TP + FN)(TN + FN) + (TP + FP)(TN + FP)} = \frac{TP + TN - E}{TP + TN + FP + FN - E}, \quad (3)$$

with

$$E = \frac{(TP + FN)(TN + FN) + (TP + FP)(TN + FP)}{TP + TN + FP + FN} \quad (4)$$

the expected number of correct random forecasts (either a hit by chance $\frac{TP + FN}{TP + FN + FN} + \frac{FP}{FP + FP} + \frac{FP}{FP + FN}$) or correct rejection by chance $\frac{TP}{TP + TN + FN} + \frac{FP}{FP + FN}$. These two scores range from −1 to 1 and provide 1 in the case of perfect forecasts, while for constant forecasts they are equal to 0. TSS tends to TPR when the number of TNs tends to infinity. Both TSS and HSS can assume negative values if the forecast is worse than a random one.

In Table 4, we report the values of TPR, TSS, and HSS for the forecast algorithms created using $R^*$ or $D$ alone, and using both $R^*$ and $D$, varying the threshold settings used to predict whether there will be an M1 or greater flare in the next 24 hr. Commenting on the results in this table we find the following.
1. Predictor individually built on $R^*$. When the threshold used to forecast whether an AR will release a flare in the next 24 hr is $\log R^* > 4$, 85% of the positive cases are correctly predicted (TPR); however, TSS and HSS present smaller values. The performance improves with higher threshold, and for $\log R^* \geq 4.5$, it is optimal for an M-class flare predictor.

2. Predictor individually built on $D$ alone. Again, for a threshold set at $D \geq 5$, the predictor guesses 58% of the flaring ARs, with a relatively small number of false positives (TSS = 0.55). With a threshold of $D \geq 7$, the skill scores decrease slightly; even further increasing the $D$ threshold, the classifier correspondingly increases its effectiveness in selecting only flaring ARs, but it recognizes only 26% of all the flaring ARs. Thus, its overall performances decrease to TSS = 0.26 and HSS = 0.38.

3. Predictor built on both $R^*$ and $D$. For thresholds set at $\log R^* \geq 4$ and $D \geq 5$, 58% of the flaring ARs are correctly recognized. Apparently, the algorithm is not as successful in retrieving all the flaring ARs, since its TPR values are smaller than or at most equal to the TPR values of the classifiers built on $R^*$ or $D$ alone, for the same threshold values. Nevertheless, the high values of the TSS and HSS scores demonstrate that this predictor performs highly satisfactorily, particularly for the thresholds $\log R^* \geq 4$ and $D \geq 5$.

5.2. Flare Forecasting with $R^*$ and $D$: An ML Approach

The previous approach to the computation of the prediction effectiveness has the advantage of assessing the role of the two predictors while tuning the threshold. However, the way these threshold values are chosen is arbitrary and therefore it is necessary to perform an optimized calculation of parameters by means of an ML approach that allows an automatic computation of the scores. Therefore, we have studied the performances of a hybrid Lasso supervised algorithm (Benvenuto et al. 2018) in which the classification is obtained by means of the application of a fuzzy clustering technique on the outcomes of the Lasso regression step. More specifically, the fuzzy clustering algorithm automatically computes the threshold separating the regression outcomes into two classes, corresponding to a flare/no-flare prediction. The regression method is therefore transformed into a classifier whose binary outcomes are used to compute the contingency tables. In order to realize the training phase we have considered the same data sets of ARs used in Campi et al. (2019). Specifically:

1. We have exploited the SDO/HMI archive in the time range between 2012 September 14 and 2016 April 30.
2. The corresponding magnetograms have been grouped into four subsets belonging to the four issuing times 00:00, 06:00, 12:00, and 18:00 UT. Of course, it may happen that an AR lasted more than 1 day: in that case, each one of the four subsets will contain as many samples associated to that AR, as the days the AR lasts.
3. For each subset, i.e., for each issuing time, we randomly extracted two-thirds of ARs in order to construct the training set using the samples contained in that subset, while the remaining one-third was used to populate the corresponding test set.
4. Each sample in each training set has been labeled by using GOES data in such a way that label “1” corresponds to an event occurrence within the 24 hr from the issuing time.

This process was repeated 100 times, so that, at the hand, we had 100 realizations of the training set and 100 realization of the test set, for each issuing time. From each AR we extracted $D$ and $\log R^*$ and fed the ML algorithm with the corresponding two-dimensional feature vectors. In order to assess the forecasting performances we computed the same skill score as above and the results we obtained are in Table 5. Differently than that done for Table 4, Table 5 is populated by means of a neural network that has been applied against a test set with no intersection with the training set. Further, the training set used to optimize a hybrid Lasso is significantly more imbalanced against M flares with respect to the data set utilized to generate Table 4. However, the score values automatically obtained in Table 5 are in several cases comparable with the ones obtained by heuristically tuning the thresholds on $\log R^*$ and $D$.

We quantify the impact on the forecasting performance of the two descriptors introduced in this study when used in combination with many more predictors. The performed feature ranking analysis relies again on a hybrid Lasso, and follows the approach described by Campi et al. (2019). That paper used as properties the 171 ones computed within the Horizon 2020 FLARECAST project (http://flarecast.eu) and arranged them again in the four training sets corresponding to the four issuing times (00:00, 06:00, 12:00, and 18:00 UT); then, we applied a hybrid Lasso machine supervised ML technique (Benvenuto et al. 2018) in order to both realize binary prediction within the next 24 hr for each issuing time and identify the features that mostly impact such prediction. In the present paper we applied the same algorithm to the same data set, this time enriched with descriptors $R^*$ and $D$. Coherently with the previous sections, the task was the prediction of M1 or more intense events within the next 24 hr.

The results of binary classification are shown in Table 6. Comparing Tables 5 and 6, one can easily note that, on the one hand, the use of more descriptors allows ML to produce higher scores. On the other hand, comparing the two blocks of columns within Table 6 suggests that the use of the two new descriptors does not significantly increase the score values. However, the results of the feature ranking analysis illustrated in Figure 7 shows that the descriptor $D$ plays a significant role in the realization of the prediction performances. In particular, the histograms in the figure describe the predictors in the training set to which Lasso assigns the highest weights for the prediction task. In this ranking, descriptor $D$ is the only one among the 173 overall predictors, which is always present for
ISSUING TIME  TPR  TSS  HSS  TPR  TSS  HSS
00:00:00UT  0.66 ± 0.16  0.56 ± 0.14  0.27 ± 0.06  0.45 ± 0.21  0.57 ± 0.14  0.34 ± 0.10
06:00:00UT  0.76 ± 0.06  0.67 ± 0.05  0.35 ± 0.04  0.74 ± 0.07  0.65 ± 0.06  0.34 ± 0.04
12:00:00UT  0.74 ± 0.07  0.66 ± 0.06  0.38 ± 0.04  0.73 ± 0.07  0.65 ± 0.06  0.38 ± 0.04
18:00:00UT  0.71 ± 0.08  0.64 ± 0.07  0.39 ± 0.04  0.70 ± 0.08  0.63 ± 0.07  0.38 ± 0.04

Table 6
Average TPR, TSS, and HSS Values and Corresponding Standard Deviations for the Outcomes of a Hybrid Lasso When Applied to 100 Random Realizations of the Training/Test Sets at Four Specific Forecast Issuing Times if the 171 Predictors of Campi et al. (2019) (left panel) and the 171 Predictors Plus log $R^*$ and $D$ (right panel) Are Used

Figure 7. Histograms counting the number of times predictors are in the top 10 rankings, on average over the 100 random realizations of the training and test sets. Descriptors from f1–f6 are function_blos/zq, sharp_kw/snetjzpp, total, sharp_kw/ushz/stddev, decay_index_br/maxl_over_hmin, and helicity_energy_bvec/abs_tot_dedt_in, and wslg_br/value_int, respectively, and their descriptions can be found in Table 2.

all four issuing times. In order to quantitatively assess the contribution of $D$ to a multi-predictor forecast, we trained a hybrid Lasso on 100 random realizations of the training and test sets according to the following process: first, the training step relies just on $D$ and then we added the other features one by one, according to the order given by the ranking. The resulting skill scores are then compared to the ones obtained by training the network with training sets that do not contain $D$ and in which the features are added one by one, again according to the ranking. The scores averaged over the 100 random realizations, are shown within the panels of Figure 8, with the red solid line denoting when $D$ is included in the training set as a first descriptor, and with the blue dashed line denoting when $D$ is not in the training set. In this figure we also considered the values of the accuracy ACC, which is defined as

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}.$$  \hspace{1cm} (5)

We notice that, for most cases, when $D$ is used for the training, either the outcomes lead to higher skill score values, or the saturation of the skill score profiles is reached in advance.

As far as descriptor $R^*$ is concerned, we found that its position in the averaged feature ranking ranges between rank 20 and rank 30 for the four issuing times, while in the analysis in Campi et al. (2019) the two Schrijver $R$ properties (one for the $B_{\text{LOS}}$ and one for $B_{\text{radial}}$ magnetograms, Guerra et al. 2018) were ranked between the 80th and 110th position among the 171 features considered in that study. We interpret this relevant ascent in the ranking as the result of the larger information content of $R^*$ compared to that in $R$, obtained by exploiting the full resolution of HMI magnetograms.

6. Conclusions

This study introduced two novelties concerning the computation and exploitation of AR descriptors for use with prediction purposes. First, we developed a new release of the algorithm for the computation of the Schrijver $R$ value. In fact, among AR descriptors, this value plays a well-established role, this parameter being a measure of the magnetic flux associated to PILs in the solar photosphere. The original algorithm for the computation of $R$ was designed for MDI magnetograms and unreliablely adapts to data gathered by the new-generation HMI cameras. In this paper, we modified this algorithm in order to account for the differences in the design of the two instruments and therefore to perform an accurate computation of the descriptor. To test whether the rescaled algorithm, designed to work on the full resolution magnetograms acquired by SDO/HMI, estimated the flare occurrence probabilities in a way comparable that of Schrijver (2007), we selected 845 magnetograms (100 of flaring regions and 745 of nonflaring regions) and repeated the same statistical analysis. We found that the occurrence rates of M- or X-class flares for the new $R^*$ value is consistent with that reported in Schrijver (2007).

The second contribution of the paper relied on the well-established observation that topologically complex ARs are strongly correlated to flaring emissions. Therefore, we
proposed a topological descriptor that counts the number of separated PILs fragments in the ARs and whose computation relies on a fast thresholding labeling process.

The rest of the study was devoted to benchmarking these two descriptors, and more specifically, to assess their impact on the flare prediction process by using both a heuristic thresholding procedure and an ML approach. We considered SDO/HMI magnetograms in the time range between 2012 September 14 and 2016 April 30 and grouped them into four subsets belonging to the four time stamps 00:00, 06:00, 12:00, and 18:00 UT. For each subset we created a set made of two-thirds of randomly extracted ARs that represented the training set for the supervised ML method we utilized for the analysis. In particular, the use of a sparsity-enhancing technique for classification allowed the identification of the AR descriptors that mostly impact the flare-forecasting process. This analysis showed that $D$ is the only one among almost 200 predictors that achieves the top 10 ranking for all four issuing times considered, and that its use in the training set implies either a steeper increase of the skill score profiles or the achievement of higher skill scores values. This result is a quantitative confirmation of phenomenological analyses of the correlation.

**Figure 8.** Skill scores obtained by using just the seven features with the best rankings in decreasing order, from 1–7, for both ML methods and all four issuing times. Features are added one at a time. The red solid line refers to the case where $D$ is added as first descriptor, the blue dashed line refers to the case without $D$. 

Cicogna et al.
between AR complexity and the probability of flare emission, and also, a nice validation of the reliability of ML-based feature ranking procedures.

In the case of $R^2$, the outcome of the feature ranking process showed that this new way to compute this descriptor analogously increases its position in the ranking with respect to the same descriptor computed according to the approach proposed in Schrijver (2007). However, this improvement is not as significant as for $D$ and this is the reason why the use of $R^2$ does not lead to higher skill score values.

Descriptors $R^2$ and $D$ are in the process of being implemented in an updated version of the Space WEatheR TOr vergata university (SWERTO) service (Berrilli et al. 2018, 2019; Del Moro et al. 2018; Napolietano et al. 2018) along with other space weather event predictors.

This paper is based on Master’s thesis research “The calibration of statistical solar flare methods for images from SDO/HMI space instrument” conducted by D.Ci. under the supervision of Prof. F. Berrilli, submitted to Physics Graduate Program at the University “Sapienza” of Rome, May 2018. D. Ci. thanks Prof. Silvia Masi for the guidance during the thesis work. The SWERTO space weather service has been financed by the Regione Lazio FILAS-RU-2014-1028 grant (2015 November–2017 October). This research is partially supported by the Italian MIUR–PRIN grant 2017APKP7T_003 on “Circumterrestrial Environment: Impact of Sun-Earth Interaction.” D.C. was supported by the Joint Research PhD Program in Astronomy, Astrophysics and Space Science between the universities of Roma Tor Vergata, Roma Sapienza, and INAF. F.B., C.C., S.G., and M.P. acknowledge the financial contribution from the agreement ASI–INAF n.2018-16-H.0.

The authors are grateful to the anonymous referee for the careful checking of the manuscript and for helpful comments that improved the paper.

**ORCID iDs**

Francesco Berrilli [https://orcid.org/0000-0002-2276-3733](https://orcid.org/0000-0002-2276-3733)

Daniele Calchetti [https://orcid.org/0000-0003-2755-5295](https://orcid.org/0000-0003-2755-5295)

Dario Del Moro [https://orcid.org/0000-0003-2500-5054](https://orcid.org/0000-0003-2500-5054)

Luca Giovannelli [https://orcid.org/0000-0001-7369-8516](https://orcid.org/0000-0001-7369-8516)

Federico Benvenuto [https://orcid.org/0000-0002-4776-0256](https://orcid.org/0000-0002-4776-0256)

Cristina Campi [https://orcid.org/0000-0003-2105-8554](https://orcid.org/0000-0003-2105-8554)

Sabrina Guastavino [https://orcid.org/0000-0001-7047-1148](https://orcid.org/0000-0001-7047-1148)

Michele Piana [https://orcid.org/0000-0003-1700-991X](https://orcid.org/0000-0003-1700-991X)

**References**

Antiochos, S. K. 1998, ApJL, 502, L181

Barnes, G., Leka, K. D., Schrijver, C. J., et al. 2016, ApJ, 829, 89

Barnes, G., Leka, K. D., Schumer, E. A., & Della-Rose, D. J. 2007, SpWea, 5, S09002

Benvenuto, F., Campi, C., Massone, A., & Piana, M. 2020, ApJL, 904, L7

Benvenuto, F., Piana, M., Campi, C., & Massone, A. M. 2018, ApJ, 853, 90

Berrilli, F., Casolino, M., Cristaldi, A., et al. 2019, NCimC, 42, 47

Berrilli, F., Casolino, M., Del Moro, D., et al. 2014, JSWSC, 4, A16

Berrilli, F., Casolino, M., Del Moro, D., et al. 2018, in IAU Symp. 353, Space Weather of the Heliosphere: Processes and Forecasts, ed. C. Foullon & O. E. Malandraki (Cambridge: Cambridge Univ. Press), 335

Georgoulis, M. K. 2013, Topology for Physicists (Amsterdam: Elsevier), 433

Nishizuka, N., Kubo, Y., Sugirii, K., Den, M., & Ishii, M. 2020, ApJ, 899, 150

Norton, A. A., Graham, J. P., Ulrich, R. K., et al. 2006, SpPh, 239, 69

Park, S.-H., Leka, K. D., Kusano, K., et al. 2019, ApJ, 890, 124

Peter, H., Abbo, L., Andretta, V., et al. 2012, ExA, 33, 271

Piana, M., Campi, C., Benvenuto, F., Guastavino, S., & Massone, A. M. 2019, NCimC, 42, 30

Piersanti, M., Alberti, T., Bemporad, A., et al. 2017, SpPh, 292, 169

Plainaki, C., Antonucci, M., Bemporad, A., et al. 2020, JSWSC, 10, 6

Plainaki, C., Liliensten, J., Radioti, A., et al. 2016, JSWSC, 6, A31

Priest, E. R., & Forbes, T. G. 2002, A&ARev, 10, 313

Puschmann, K. G., Denker, C., Kneer, F., et al. 2012, AN, 333, 880

Reale, F. 2014, LRP, 11, 4

Rodriguez, F., Ronchini, L. R., Di Rollo, S., et al. 2019, NCimC, 42, 45

Romano, P., Zuccarello, F., Giglielmino, S. L., et al. 2015, A&A, 582, A55

Scharmer, G. B., Narayan, G., Hillberg, T., et al. 2008, ApJ, 689, L69

Scherrer, P. H., Bogart, R. S., Bush, R. I., et al. 1995, SpPh, 162, 129

Scherrer, P. H., Schou, J., Bush, R. I., et al. 2012, SpPh, 275, 207

Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012, SpPh, 275, 229

Schröjer, C. J. 2007, ApJL, 655, L17

Schröjer, C. J., De Rosa, M. L., Title, A. M., & Metcalf, T. R. 2005, ApJ, 628, 501

Schwarz, A. S. 2013, Topology for Physicists (Berlin/Heidelberg: Springer Science & Business Media)
Steward, G. A., Lobzin, V. V., Wilkinson, P. J., Cairns, I. H., & Robinson, P. A. 2011, SpWea, 9, S11004
Suematsu, Y., Ichimoto, K., Katsukawa, Y., Tsuneta, S., & Shimizu, T. 2017, Proc. SPIE, 10566, 105662Z
Tadesse, T., Wiegelmann, T., Inhester, B., et al. 2013, A&A, 550, A14
Toriumi, S., & Wang, H. 2019, LRSP, 16, 3

Vadakke Veettil, S., Cesaroni, C., Aquino, M., et al. 2019, JSWSC, 9, A41
Viavattene, G., Calchetti, D., Berrilli, F., et al. 2020, NCintC, 43, 120
Viticchié, B., Del Moro, D., & Berrilli, F. 2006, ApJ, 652, 1734
Walsh, L., Schneider, U., Fogtman, A., et al. 2019, LSSR, 21, 73
Wheatland, M. S. 2000, ApJ, 532, 1209
Yi, K., Moon, Y.-J., Shin, G., & Lim, D. 2020, ApJL, 890, L5