Investigating the link between Active Region photospheric characteristics and their eruptive potential

F. Berrilli[1], D. Cicogna[1], D. Del Moro[1], R. Foldes[1-2], L. Giovannelli[1], E. Pietropaolo[1]

[1]Università degli Studi di Roma "Tor Vergata", Via della Ricerca Scientifica, 1, 00133 Rome, Italy
[2]Università dell'Aquila, Via Vetoio, 48, 67100 Coppito L'Aquila, Italy
Correspondence to: delmoro@roma2.infn.it

Forecasting the probability of a solar Active Region (AR) to flare is a challenging and pursued topic in the Space Weather field. The R value, developed by [3] on MDI data, is among the most used descriptors of the photospheric magnetic field in ARs for flare forecasting applications. We propose **a modified computation for the R value**, to exploit the higher spatial resolution of HMI (cross-calibrated with MDI data to maintain compatibility) and verify its functionality on the full set of cycle 24th solar flares. Furthermore, we propose **a new parameter D**, based on the automatic recognition of magnetic polarity-inversion lines in the identified AR, to parametrize the AR magnetic topological complexity. We use both R and D parameters to train a logistic regression method to predict the occurrence of X- or M- class flares in a given solar active region during the following 24 hours period. The results of our statistical analysis show that **both parameters are good descriptors of the flaring proneness of an AR** and possible tools for flare forecasting. The D parameter will be included in the flare forecasting product of the Space WEeatherR TOr vergata university (SWERTO), presently based on the R value.

The Dataset

We used HMI line-of-sight magnetograms of ARs **from June 2010 to September 2018** from the tracked AR patch data product on the Joint Science Operation Center database (http://jsoc.stanford.edu), dividing them into two different sets, depending of whether they represent flaring active regions (positive class) or non-flaring active regions (negative class).

Positive class

- 100 active regions
- Active regions with X- or M-class flares within 24 hours
- Flares occurring in active regions within 45 degrees from the center

Negative class

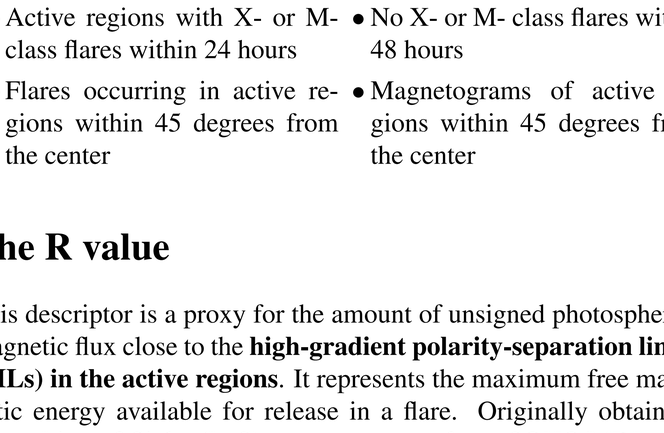
- 745 active regions
- No X- or M- class flares within 48 hours
- Magnetograms of active regions within 45 degrees from the center

The R value

This descriptor is a proxy for the amount of unsigned photospheric magnetic flux close to the **high-gradient polarity-separation lines (PILs) in the active regions**. It represents the maximum free magnetic energy available for release in a flare. Originally obtained using Line-of-Sight (LoS) magnetograms from SOHO/MDI, we modified the R value algorithm [3] to work with the LOS magnetograms produced by SDO/HMI. These differ in pixel size and spatial resolution and use a different spectral line to to compute the magnetograms. The proposed algorithm to compute R value from SDO/HMI LoS magnetograms uses:

- different dilatation kernels;
- different Gaussian kernels (41 HMI pixels FWHM) to convolve the PIL mask and compute the weighted map;
- a rescale of the area to take into account HMI pixel size and to obtain R in Maxwells.

Finally, the same statistical calibration, including the ratio between flaring and non-flaring regions, was applied to the whole flare dataset observed by SDO/HMI.



Class	log R ≈ 3.0 (%)	log R ≈ 3.5 (%)	R log R ≈ 4 (%)	log R ≈ 4.5 (%)	log R ≈ 5 (%)
M1	∼ 0	4	8	33	92
X1	0	0	∼ 0	2	17

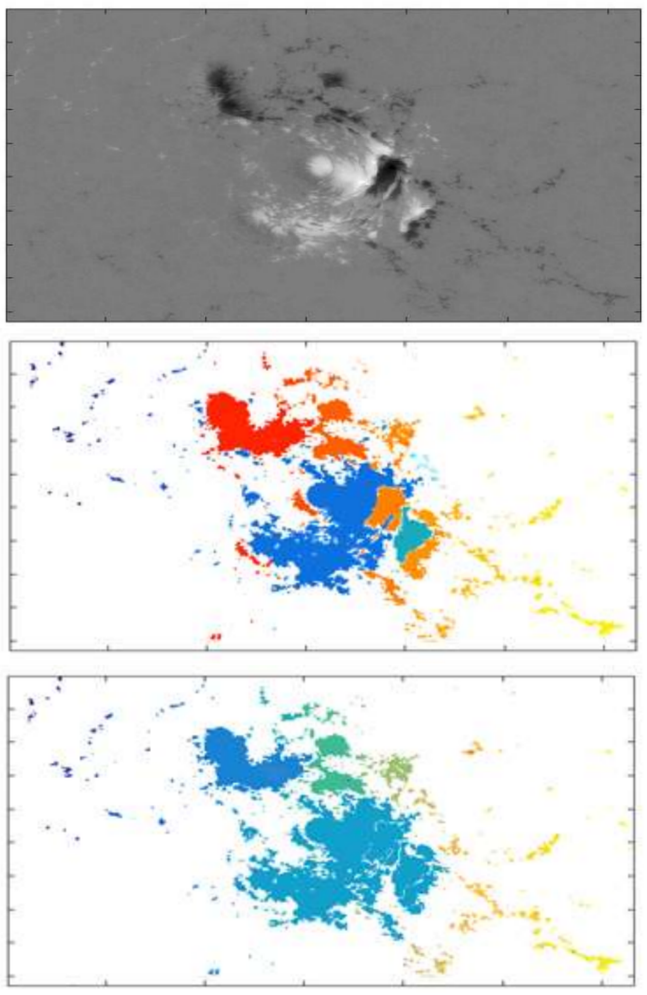
Class	D ≈ 1 (%)	D ≈ 3 (%)	D D ≈ 5 (%)	D ≈ 7 (%)	D ≥ 10 (%)
M1	9	22	36	70	96
X1	0	∼ 0	7	20	25

R value results: a) Scatter diagram of peak flare flux densities in the 1-8 $\AA$ band in W m^{-2} vs $log\ R$. The straight line $F = 1.2 \cdot 10^{-3}\ R$ has all flares below it. **b)** Likelihood of X- or M- class flare within 24 hours from the determination of the two parameters.

The D parameter

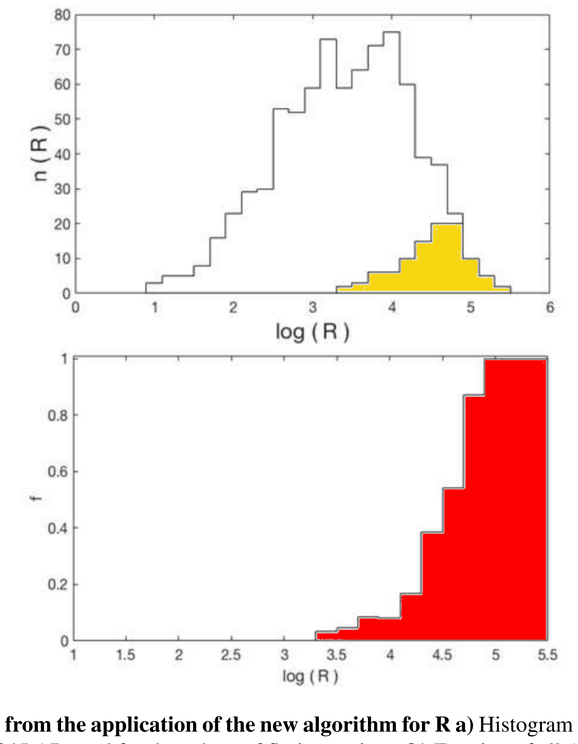
The D parameter is a simple, computationally inexpensive feature to evaluate the complexity of PILs. **It is a topological parameter**, which represents the number of fragments of PIL present in an AR, both in terms of different lines, as well as fragmentation of a single one. The algorithm steps to obtain D are:

- two binary masks (PBMs) are created for each polarity from the HMI LoS magnetogram;
- a label is assigned to every connected region (fragment) in each PBM;
- f_t is the sum of the positive and negative fragment counts;
- from the LoS magnetogram a single unsigned binary mask (UBM) is created;
- the labelling procedure is applied to the UBM, to count the number of connected regions, f_u;
- D= f_t - f_u counts the number of regions/fragments containing PILs.

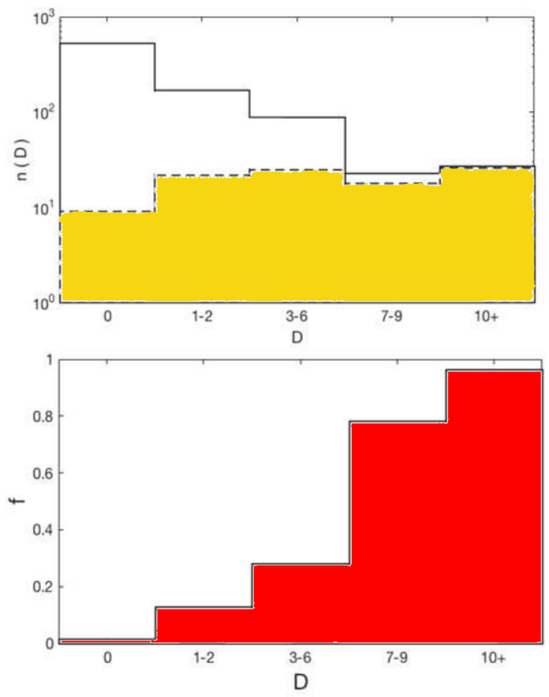


Example of mask to compute the D parameter a) Sample HMI LoS magnetogram **b)** Mask from the PBMs. Separate regions are labeled with different colors. Blue shades for the positive polarity regions, red shades for the negative polarity regions. **c)** Mask from the UBM. Separate regions are labeled with different colors

Results



Results from the application of the new algorithm for R a) Histogram of $log\ R$ for the 845 ARs and for the subset of flaring regions. **b)** Fraction of all ARs with at least one major flares within 24 hours as a function of $logR$.



Results from the application of the D parameter algorithm a) Histogram of D for the 845 ARs and for the subset of flaring regions. **b)** Fraction of all ARs with at least one major flare within 24 hours as a function of D.

	This work	Bobra [1]	Song [2]	Ahmed [3]
Time interval	**48 h**	48 h	24 h	48 h
Class-imbalance ratio	**7.46**	16.5	2.23	15.85
Accuracy	**0.937**	0.973	0.873	0.975
Precision positive	**0.694**	0.797	0.917	0.877
Precision negative	**0.977**	0.983	0.860	0.980
Recall positive	**0.833**	0.714	0.647	0.677
Recall negative	**0.951**	0.989	0.974	0.994
f1 positive	**0.758**	0.751	0.758	0.764
f1 negative	**0.964**	0.986	0.913	0.987
TSS - true skill statistic	**0.784**	0.703	0.620	0.512

Results of the logistic regression method and comparison with other works [1, 2, 4] that apply Machine Learning algorithms to forecast the behavior of a given active region. The active regions set was divided into a training set and a test set with ratio approximately 70% to 30%

Conclusions and Future Work

The R value and the D parameter have been applied to the whole set of selected HMI line-of-sight magnetograms of active regions. The R value has been calibrated from the original algorithm developed by Schrijver in order to work with the new instrument. The D parameter is a good proxy to forecast major solar flare in the following 24 hours.
By using the **logistic regression** method, R and D were used as input features to train a model to predict the activity of an active region. The **performance metrics** show the goodness both of the approach used and of the two features.
We are investigating the extension of this approach towards the prediction of Coronal Mass Ejection.
The new value of R presented here is **already in use in the SWERTO** operational Space Weather service to provide a forecast of flare for the next 24h and the implementation of a forecast based on the D parameter is in progress.

References

[1] Omar W Ahmed, Rami Qahwaji, Tufan Colak, Paul A Higgins, Peter T Gallagher, and D Shaun Bloomfield. Solar flare prediction using advanced feature extraction, machine learning, and feature selection. *Solar Physics*, 283(1):157–175, 2013.

[2] Monica G Bobra and Sebastien Couvidat. Solar flare prediction using sdo/hmi vector magnetic field data with a machine-learning algorithm. *The Astrophysical Journal*, 798(2):135, 2015.

[3] Carolus J Schrijver. A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting. *The Astrophysical Journal Letters*, 655(2):L117, 2007.

[4] Hui Song, Changyi Tan, Ju Jing, Haimin Wang, Vasyl Yurchyshyn, and Valentyna Abramenko. Statistical assessment of photospheric magnetic features in imminent solar flare predictions. *Solar Physics*, 254(1):101–125, 2009.

Acknowledgments

We thank the NASA/SDO and the AIA, EVE and HMI science teams for making the data products used in this work available. SWERTO is an operational Space Weather service operated on regional scale (Regione Lazio) based on data from space-based and ground-based instruments.