Calibration of statistical solar flare forecast parameters for images from SDO/HMI space instrument.

Luca Giovannelli, Francesco Berrilli, Domenico Cicogna, and Dario Del Moro

Department of Physics, University of Rome Tor Vergata, ITALY (https://www.fisica.uniroma2.it/~solare/en/)

Photography encouraged

Abstract

Forecasting the probability of a solar active region to flare is a challenging and pursued topic in the Space Weather field. The R value, developed by Schrijver (2007), is one of the most used descriptors of the photospheric magnetic field in active regions for flare forecasting applications. The R value method was calibrated on the magnetograms obtained from the Michelson Doppler Imager (MDI) instrument on board the Solar and Heliospheric Observatory (SOHO) between 1998 and 2006, during solar cycle 23. Since 2010, the Helioseismic and Magnetic Imager (HMI) on board the Solar Dynamics Observatory (SDO) satellite is operative, collecting solar magnetograms with a spatial resolution 4 times higher than MDI, while the observation program of MDI was terminated on April 2011.

We recalibrate the R value algorithm to adapt it to the higher spatial resolution of HMI. We also compare the statistical analysis of a sample of cycle 24th solar flares following the statistical approach performed by Schrijver (2007). Furthermore, we propose a new parameter D, aimed to identify and count the number of magnetic polarity-inversion lines within the same active region, to parametrize the magnetic active region topological complexity. We use R and D parameters as features to train a logistic regression method to predict the occurrence of X- or M- class flares in a given solar active region during the next 24 hours period, obtaining relatively good performances.

Flare Forecasting Features:

R value

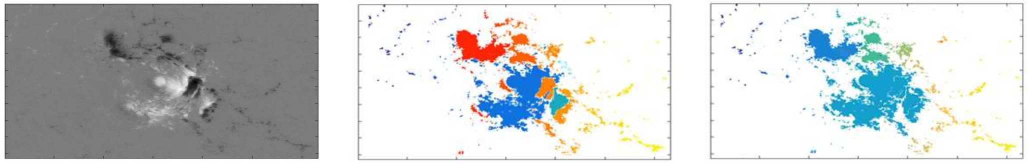
This descriptor is a proxy for the ammount of unsigned photospheric magnetic flux close to the high-gradient polarity-separation lines (PILs) in the active regions. It represents the maximum free magnetic energy available for release in a flare. Originally obtained using Line-of-Sight (LoS) magnetograms from the Michelson Doppler Imager, we modified the original R value algorithm (developed by Schrijver in 2007) to work with the line-of-sight magnetograms produced by Helioseismic and Magnetic Imager, which differs in pixel size and spatial resolution and use a different spectral line to derive all its data products. The algorithm for the R value calculation still follows what developed by Schrijver with few differences like:

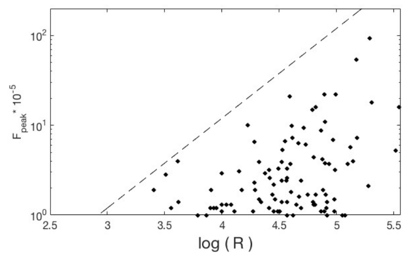
- Different dilatation kernels
- The binary mask with the PILs is colvolve with a gaussian with FWHM of 41 HMI pixels to obtain the weighted map
- To obtain the value of R in maxwells, we use HMI's pixel area and rescale the value to adapt it with the other works which used the R value.

D parameter

D was derived looking for a feature that: *i)* could provide more information about the configuration of the PILs, *ii)* is simple and outright and *iii)* is computationally inexpensive. The D parameter is a morphology descriptor, which represents the number of fragments of lines present in an active regions, both in terms of different lines, as well as fragmentation of a single one. We use the LOS magnetograms from the HMI instrument. The algorithm to obtain D follows the next steps:

- We create two binary masks for each polarity from the los magnetogram
- We assign an identification number, a label, to every connected region in each mask
- The total number of positive and negative fragments are added in one index, f_t
- From the HMI LoS magnetogram we construct an unsigned binary mask
- The labelling procedure is applied to the unsigned binary mask, finding a second index, f_u
- $D = f_t - f_u$ the two indices differ in number due to the presence of the PILs that made two fragments of opposite sign in the unsigned binary mask indistinguishable





Scatter diagram of peak flare flux densities in the 1 – 8 Å band in W m^{-2} vs *log*R. The straight line at $F_{max} = 1.2 * 10^{-3} R$ has all flares below it.

Class	log R ≈ 3.0 (%)	log R ≈ 3.5 (%)	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 ≈ 5 (%)	D ≈ 7 (%)	D ≥ 10 (%)
M1	9	22	36	70	96
X1	0	~ 0	7	20	25

Likelihood of X- or M- class flares within 24 hours from the determination of the two paramets.

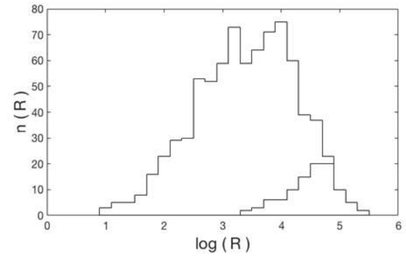
Acknowledgments. We would like to thank the NASA/SDO and the AIA, EVE and HMI science teams for producting the data products used in this work.

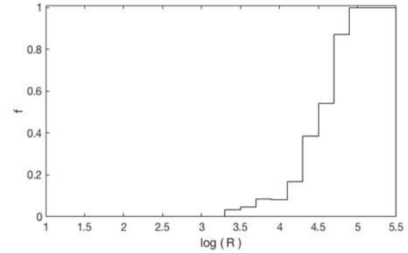
Dataset

For this work have been used only HMI line-of-sight magnetograms of active regions from June 2010 to September 2018 taken from the tracked active region 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)

Positve 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

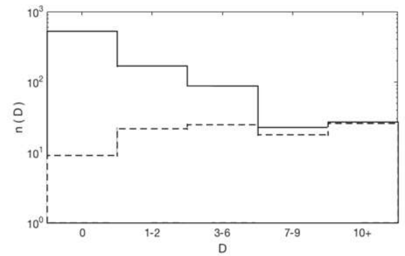
Negative class
- 745 active regions
- No X- or M- class flares within 48 hours
- Magnetograms of active regions within 45 degrees from the center

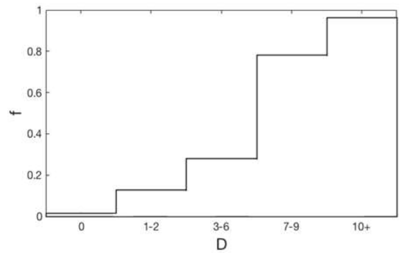




Results from the application of the new algorithm for R

Left: Histogram of logR for the 845 ARs and for the subset of flaring regions. Right: Fraction of all ARs with at least one major flare within 24 hours as a function of logR





Results from the application of the D parameter algorithm

Left: Histogram of D for the 845 ARs and for the subset of flaring regions. Right: Fraction of all ARs with at least one major flare within 24 hours as a function of D

Results of the logistic regression method and comparison with other works 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%

	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

[1] Bobra & Couvidat S., 2015, ApJ, 798, 135
[2] Song et al. 2009, SolPh, 254, 101
[3] Ahmed et al. 2013, SolPh, 283, 157

Conclusions

In this work, the R value and the D parameter have been applied to the whole set of selected HMI's 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. We verified that D parameter is a good proxy to forecast the insurgence of a major solar flare in the next 24 hours. By using the logistic regression method, R and D were used as input features to train the model to correctly predict the activity of an active region. The performance metrics show the goodness both of the approach used as well as of the two features.