Relationships between Characteristics of the Line-of-sight Magnetic Field and Solar Flare Forecasts

Viacheslav M. Sadykov1,2,3 and Alexander G. Kosovichev1,2,3

1 Center for Computational Heliophysics, New Jersey Institute of Technology, Newark, NJ 07102, USA; vsadykov@njit.edu
2 Department of Physics, New Jersey Institute of Technology, Newark, NJ 07102, USA
3 NASA Ames Research Center, Moffett Field, CA 94035, USA

Received 2017 April 11; revised 2017 September 29; accepted 2017 October 2; published 2017 November 9

Abstract

We analyze the relationship between the flare X-ray peak flux, and characteristics of the polarity inversion line (PIL) and active regions (ARs), derived from line-of-sight (LOS) magnetograms. The PIL detection algorithm based on a magnetogram segmentation procedure is applied for each AR with 1 hr cadence. The PIL and AR characteristics are associated with the AR flare history and divided into flaring and nonflaring cases. Effectiveness of the derived characteristics for flare forecasting is determined by the number of nonflaring cases separated from flaring cases by a certain threshold, and by their Fisher ranking score. The Support Vector Machine (SVM) classifier trained only on the PIL characteristics is used for the flare prediction. We have obtained the following results: (1) the PIL characteristics are more effective than global characteristics of ARs, (2) the highest True Skill Statistics (TSS) values of 0.76 ± 0.03 for >M1.0 flares and 0.84 ± 0.07 for >X1.0 flares are obtained using the “Sigmoid” SVM kernel, (3) the TSS scores obtained using only the LOS magnetograms are slightly lower than the scores obtained using vector magnetograms, but significantly better than current expert-based predictions, (4) for prediction of >M1.0 class flares 74.4% of all cases, and 91.2% for >X1.0 class, can be pre-classified as negative with no significant effect on the results, (5) the inclusion of global AR characteristics does not improve the forecast. The study confirms the unique role of the PIL region characteristics in the flare initiation process, and demonstrates possibilities of flare forecasting using only the LOS magnetograms.

Key words: methods: statistical – Sun: activity – Sun: flares – Sun: magnetic fields

1. Introduction

Usually lasting from several minutes to several hours, solar flares can release more than $10^{32}$ erg of energy, and cause harmful effects to the terrestrial environment. The only possible source to accumulate such large amounts of energy is the magnetic field of active regions (ARs). Emslie et al. (2012) demonstrated for a sample of 38 flares that the free (nonpotential) energy of magnetic field was sufficient to explain the flare energy release including coronal mass ejections (CMEs), energetic particles, and hot plasma emission and dynamics. For understanding the flare physical mechanism and developing flare prediction methods, it is important to find critical magnetic field characteristics that are linked to the flare initiation and strength.

There are two types of such studies. The first approach is to focus on global characteristics of ARs, and the second approach is to search for local critical properties of magnetic fields. For instance, in the first type, Mandage & McAteer (2016) demonstrated the difference between the magnetic field power spectrum slopes of flaring and nonflaring ARs. Korsós et al. (2014) found several promising preflare signatures using the SOHO/MDI-Debrecen Data sunspot catalog. Korsós et al. (2015) introduced the weighted horizontal magnetic gradient, $W_G$, which allowed them to predict the onset time for $\geq M5.0$ class flares, and conclude whether or not a flare is likely be followed by another event in the next 18 hr. The daily averages of $W_G$ together with a separation parameter $S_{\varphi, i}$ of magnetic polarities were used by Korsós & Erdélyi (2016) to obtain some conditional probabilities of flare and CME characteristics. Bobra & Couvidat (2015), Bobra & Ilonidis (2016), Nishizuka et al. (2017), and Liu et al. (2017) have used vector magnetograms from the Space-weather HMI Active Region Patches (SHARP) and applied machine-learning techniques (Support Vector Machine (SVM), Random Forest, and Nearest-Neighbor classifiers) for flare and CME predictions. Also, a recent study by Rabonik et al. (2017) used the Zemke moments as characteristics of the AR magnetic field for flare prediction.

Many observational studies of the second type found that the magnetic field polarity inversion line (PIL) in regions of the strong field plays an important role in the flare activity (e.g., Severny 1964; Hagyard et al. 1990; Wang et al. 1994; Falconer et al. 1997; Kosovichev & Zharkova 2001; Jing et al. 2006; Schrijver 2007; Kumar et al. 2015; Barnes et al. 2016; Schrijver 2016; Sharykin et al. 2017; Bamba et al. 2017; Toriumi et al. 2017; Zimovets et al. 2017). Kusano et al. (2012) demonstrated from three-dimensional magnetohydrodynamic simulations that flares eruptions can be initiated by the emergence of certain small magnetic structures near PIL, as is evident from observations. Toriumi et al. (2013, 2014) pointed out the important role of the highly sheared magnetic field in the vicinity of PILs in the flare development process. Guennou et al. (2017) found from simulations that the PIL parameters measuring the total nonpotentiality of ARs present a significant ability to distinguish between eruptive and noneruptive cases. From magnetograms, one can extract several descriptors representing the local field in the PIL vicinity. For example, Falconer et al. (2003) showed that the length of the PIL with a strong field gradient and sheared transverse field correlates with the CME and flare productivity. Mason & Hoeksema (2010) introduced the Gradient-Weighted PIL length as a characteristic for solar flare forecasts. Falconer et al. (2011, 2012, 2014) found that this characteristic is a good
proxy for the free magnetic energy. Leka & Barnes (2003a, 2003b, 2007) suggested to use a shear angle between the observed and reconstructed magnetic fields. Chernyshov et al. (2011) used the PIL length, the area of strong magnetic field in the PIL vicinity, and the total flux in this area, as well as the rates of change of these characteristics.

In this paper, we perform a critical analysis of various line-of-sight (LOS) magnetic field characteristics (derived for the entire AR and for the PIL vicinity), their relationship to the flaring activity, and the importance for flare forecast. Such an analysis based on the LOS magnetograms is important because these observations can be performed more easily and accurately than the full vector magnetic field measurements in near-real time by various space-based and ground-based observatories. In Section 2, we describe automatic procedures for identification of PIL, calculation of various magnetic field characteristics, association of the derived characteristics with flare events, and construction of “train” and “test” data sets. In Section 3, we estimate the effectiveness in the separation of flaring and nonflaring cases for different LOS characteristics. Section 4 describes the application of the SVM classifier for the prediction of M- and X-class flares. The results are summarized in Section 5. The comparison with previous results, expert-based scores, and our conclusion are presented in Section 6.

2. Data Preparation

2.1. Magnetogram Segmentation

For analysis, we used the LOS magnetograms of ARs, obtained by the Helioseismic and Magnetic Imager on board the Solar Dynamics Observatory (SDO/HMI, Scherrer et al. 2012). The AR data were represented in the form of 30° × 30° data cubes with 1 hr cadence, remapped onto the heliographic coordinates using Postel’s projection, and tracked with the solar differential rotation during the whole passage of ARs on the solar disk, employing the standard SDO software. To avoid projection effects, following Bobra & Couvidat (2015), we consider ARs only when they are located within ±68° from the disk center.

By definition, the PIL is the line where the LOS magnetic field changes its sign. For the automatic robust detection of the PIL of strong fields in ARs, we use the algorithm initially introduced by Chernyshov et al. (2011) and Laptev (2011). This algorithm is based on a magnetogram segmentation process formulated as an optimization task. The goal is to divide the magnetogram into regions with strong positive field (“positive” segments), strong negative field (“negative” segments), or weak field (“neutral” segments). We describe the algorithm in detail in Appendix A. An example of the segmentation and PIL detection for AR 11158 is illustrated in Figure 1.

To isolate the AR area, we use the following two algorithms. The first one is based on the segmentation result: we apply one morphological dilation (inclusion of neighboring pixels) to the positive/negative segments (see Appendix A), combine them, choose the largest segment containing the AR center, and determine the minimum bounding box around it. The second algorithm is implemented following the procedure of Stenflo & Kosovichev (2012). The magnetogram is smoothed, and for each strong magnetic field island the bounding box with a margin of fixed width (18°) on all sides is defined. Then, the intersecting bounding boxes are replaced by a larger bounding box. The solution represents the largest bounding box intersecting the center of the data cube (the center of AR).

We have found that by applying both algorithms and selecting the smallest bounding box, almost all ARs can be effectively separated from their neighbors. The bounding box extracted for AR 11158 is presented in Figure 1.

2.2. Derivation of PIL and AR Characteristics

After performing the segmentation and bounding procedures, we calculate the following descriptors (characteristics) using the derived PIL and the tracked and remapped magnetogram.
1. The PIL length defined as the number of pixels occupied by the PIL.
2. The PIL area obtained after 10 morphological dilations of the PIL.
3. The unsigned magnetic flux in the PIL area.
4. The unsigned horizontal gradient in the PIL area defined as the sum of \( \nabla B_z = \sqrt{\left(\frac{\partial B_z}{\partial x}\right)^2 + \left(\frac{\partial B_z}{\partial y}\right)^2} \) over the PIL area pixels.
5. The maximum gradient of the LOS magnetic field across the PIL.
6. The gradient-weighted PIL length (Mason & Hoeksema 2010) calculated as the sum of the PIL pixels multiplied by the unsigned horizontal gradient in each pixel.
7. The R-value (Schrijver 2007) representing the unsigned magnetic flux weighted with the inverse distance from the PIL.
   Also, we calculate the following characteristics of the entire AR (“global” characteristics):
8. The AR area defined as the total area of the positive and negative segments.
9. The unsigned magnetic flux in the AR area.
10. The maximum strength of magnetic field in AR.
11. The unsigned horizontal gradient in the AR area.

2.3. Definition of Positive and Negative Classes, and Construction of “Train” and “Test” Data Sets

The next important step is to associate the magnetic field characteristics derived for each AR with the flare events detected by the GOES satellite. Following Nishizuka et al. (2017), we classify a set of magnetic field characteristics as a “positive” case if a \( \geq \text{M}1.0 \) flare occurred in the corresponding AR within 24 hr after the last field measurement. This means that for each flare there can be 24 positive cases (sets of measured LOS magnetic field characteristics) or less. For the period from 2010 April to 2016 June, 521 M-class and 31 X-class flares were associated with at least one positive case.

Ahmed et al. (2013) introduced two ways to determine the negative cases, described by so-called “operational” and “segmented” associations of AR characteristics and flares. According to the operational association, the negative cases are defined to be exactly opposite to the positive cases, i.e., they are assigned if there was no flare of \( \geq \text{M}1.0 \) X-ray class within 24 hr after the magnetic field measurement. For the segmented association, the case is defined as negative if no flares occurred 48 hr before and after the case time moment. In the following, we will use the operational association for the “test” subset while keeping the segmented association for the “train” subset. The segmented association better separates the positive and negative cases (by neglecting negative cases occurring very close to the flare time), while the operational association is needed for real-time predictions. The same procedure was applied also for \( \geq \text{X}1.0 \) class flares.

For the operational-type real-time flare forecasts, the classifier is defined for future cases based on the previously observed classified cases. To simulate the real-time operational forecast, we constructed the “train” and “test” data sets to be sequential in time. We assign all the cases belonging to ARs with the NOAA numbers 11059-12158 to the “train” data set, and AR 12159-12559 to the “test” data set. The ratio of the “train” and “test” data sets is approximately 70%-30% (following Bobra & Couvidat 2015; Nishizuka et al. 2017). We also assume that we have just one attempt to classify a “test” data set for prediction of \( \geq \text{M}1.0 \) or \( \geq \text{X}1.0 \) flares, which means that the classifier tuning should be done on the “train” data set only.

3. Effectiveness of Characteristics

In this section, we analyze the effectiveness of the derived magnetic field characteristics to separate the positive and negative (flaring and nonflaring) cases. One of the simplest ways to illustrate the separation ability of magnetic field characteristics is to construct combined histograms for positive and negative cases. The examples of such histograms are presented in Figure 2. The upper two panels correspond to two PIL characteristics: the unsigned magnetic flux in the PIL area and the gradient-weighted PIL length; and the lower two panels correspond to two AR characteristics: the unsigned magnetic flux in the AR area and the unsigned horizontal gradient in the AR area.

One can notice that, for the PIL characteristics, there are more flaring than nonflaring cases in the tails of the histograms (light colored areas). We found such a situation for all PIL characteristics that we computed. For the global AR characteristics, we found a slight dominance of positive cases in the distribution tail only for the unsigned magnetic flux, and did not observe it for the other three characteristics.

There is one common feature in the histograms. The positive cases occur only if the characteristics reach some critical (threshold) value. For some LOS characteristics the existence of the critical values is more prominent in the normal-scaled histogram, but for others in the logarithmic-scaled histogram. This feature is used to simplify the classification (prediction) problem by reduction of the amounts of data considered for the classification. The red dashed (for \( \geq \text{M}1.0 \) flares) and green dashed (for \( \geq \text{X}1.0 \) flares) lines in Figure 2 represent the threshold values, above which 95% of positive cases are observed. Note that the threshold values are determined using the “train” data set. At the same time, the mean values of the positive cases are shown by solid lines of the same color.

The threshold and mean values for the positive cases, as well as the mean value for the negative cases, are summarized in Table 1.

There are many ways to quantitatively determine which characteristics are most effective for a classification problem. The inclusion of characteristics that are not discriminative leads to a high computational cost without improvement of the result, and may even decrease the performance of the SVM (Bobra & Couvidat 2015). Breiman (2001) proposed to evaluate feature importance by using the Random Forest classification, which was also used by Nishizuka et al. (2017). Al-Ghraibah et al. (2015) employed the univariate True Skill Statistics (TSS) score as a measure of feature importance. Ahmed et al. (2013) used the Correlation-Based Feature-Selection (CFS) and Minimum Redundancy Maximum Relevance methods. Leka & Barnes (2003b) suggested the Mahalanobis distance between classes and Hotelling’s \( T^2 \)-test to measure statistical differences between flaring and nonflaring cases. Bobra & Couvidat (2015) calculated the Fisher Ranking score (or F-score) as a measure of a univariate effectiveness of the separation ability.

In this work, we calculated two simple univariate scores for the obtained magnetic field characteristics. First, for each characteristic, we derived the threshold separating 5% of the
positive cases. As seen from Table 1, these threshold values (for both ≥M1.0 and ≥X1.0 flares) are comparable or even greater than the mean values for the negative cases for most characteristics. Thus, the fraction of negative cases that could be cut off by this threshold is used as a measure of effectiveness of characteristics in separating the “train” and “test” data sets. Second, we calculate the Fisher ranking score (or F-score, Chang & Lin 2008; Bobra & Couvidat 2015):

$$F(i) = \frac{\overline{x}_i^+ - \overline{x}_i^-}{\frac{1}{n^+ - 1} \sum_{k=1}^{n^+}(x_{ik}^+ - \overline{x}_i)^2 + \frac{1}{n^- - 1} \sum_{k=1}^{n^-}(x_{ik}^- - \overline{x}_i)^2},$$

where $\overline{x}_i$ is the mean value of characteristic $i$; $\overline{x}_i^+$ and $\overline{x}_i^-$ are the mean values of characteristic $i$ for the positive and negative cases; and $n^+$ and $n^-$ are the total numbers of the positive and negative cases. We calculated the F-score for all the characteristics for the train data set. Sometimes, the F-score is higher if calculated for the logarithms of the parameters. Therefore, we also calculated the F-scores of decimal logarithms of each parameter and used it if the score was higher than the one for the normal-scaled characteristic.

The results for both estimates of effectiveness are combined and summarized in Tables 2 and 3 for the ≥M1.0 and ≥X1.0 class flares respectively. The cases for which the logarithmic scale was used in the F-score calculation are labeled as (log) in Tables 2 and 3. The SVM training and testing were also done in the logarithmic scale for such parameters. One can notice from Tables 2 and 3 that for every considered univariate test the PIL characteristics have higher scores than the global AR parameters.

4. Methodology of Flare Prediction

Currently, most operational flare forecasts are based on expert decision. However, many recent works (Bobra & Couvidat 2015; Hada-Muranushi et al. 2016; Shin et al. 2016; Anastasiadis et al. 2017; Liu et al. 2017; Nishizuka et al. 2017; Raboonik et al. 2017) demonstrated that the machine-learning algorithms can be successfully applied for flare prediction. In this section, we test if it is possible to forecast ≥M1.0 and ≥X1.0 flares, using machine-learning algorithms based solely on the LOS magnetic field characteristics. Our approach is to utilize the SVM (Cortes & Vapnik 1995) classifier for flare forecasting using the Python module

![1D-histograms of (a) unsigned magnetic flux in the PIL area, (b) gradient-weighted PIL length, (c) unsigned magnetic flux in the AR area, and (d) unsigned horizontal gradient in the AR area. The negative cases are shown in gray, the positive ≥M1.0 class cases are shown in red, and the positive ≥X1.0 class cases are shown in green. The darker areas represent the intersections of the histograms. The red and green solid lines represent the average values of the positive ≥M1.0 and ≥X1.0 cases, the corresponding dashed lines show the thresholds corresponding to 5% of positive cases.](image-url)
Table 1
Relationship between Magnetic Field Characteristics and Solar Flares of the GOES X-Ray Classes Greater Than M1.0 and X1.0

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Average Value (Negative Cases)</th>
<th>Average Value (Positive $\geq$ M1.0 Cases)</th>
<th>Average Value (Positive $\geq$ X1.0 Cases)</th>
<th>5% Threshold (Positive $\geq$ M1.0 Cases)</th>
<th>5% Threshold (Positive $\geq$ X1.0 Cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIL length (m)</td>
<td>$(1.7 \pm 1.8) \cdot 10^1$</td>
<td>$(6.8 \pm 4.0) \cdot 10^4$</td>
<td>$(8.3 \pm 3.6) \cdot 10^7$</td>
<td>$1.7 \cdot 10^4$</td>
<td>$3.2 \cdot 10^4$</td>
</tr>
<tr>
<td>PIL area ($m^2$)</td>
<td>$(7.3 \pm 5.5) \cdot 10^{14}$</td>
<td>$(9.3 \pm 7.9) \cdot 10^{14}$</td>
<td>$(21.8 \pm 7.6) \cdot 10^{14}$</td>
<td>$7.9 \cdot 10^{14}$</td>
<td>$9.6 \cdot 10^{14}$</td>
</tr>
<tr>
<td>Unsigned magnetic flux in the PIL area (G m$^{-2}$)</td>
<td>$(1.14 \pm 1.15) \cdot 10^{17}$</td>
<td>$(4.68 \pm 2.98) \cdot 10^{17}$</td>
<td>$(5.96 \pm 2.98) \cdot 10^{17}$</td>
<td>$1.20 \cdot 10^{17}$</td>
<td>$1.57 \cdot 10^{17}$</td>
</tr>
<tr>
<td>Unsigned horizontal gradient in the PIL area (G m)</td>
<td>$(0.81 \pm 0.75) \cdot 10^{11}$</td>
<td>$(2.93 \pm 1.67) \cdot 10^{11}$</td>
<td>$(3.40 \pm 1.34) \cdot 10^{11}$</td>
<td>$0.89 \cdot 10^{11}$</td>
<td>$1.47 \cdot 10^{11}$</td>
</tr>
<tr>
<td>Maximum gradient across the PIL (G m$^{-1}$)</td>
<td>$(3.8 \pm 2.3) \cdot 10^{-4}$</td>
<td>$(9.0 \pm 4.4) \cdot 10^{-4}$</td>
<td>$(10.3 \pm 3.4) \cdot 10^{-4}$</td>
<td>$3.7 \cdot 10^{-4}$</td>
<td>$5.3 \cdot 10^{-4}$</td>
</tr>
<tr>
<td>Gradient-weighted PIL length (m G m$^{-1}$)</td>
<td>$(3.1 \pm 4.1) \cdot 10^{3}$</td>
<td>$(19.4 \pm 16.4) \cdot 10^{3}$</td>
<td>$(24.1 \pm 13.2) \cdot 10^{3}$</td>
<td>$2.8 \cdot 10^{3}$</td>
<td>$6.2 \cdot 10^{3}$</td>
</tr>
<tr>
<td>R-value (G m$^2$)</td>
<td>$(2.4 \pm 3.2) \cdot 10^{15}$</td>
<td>$(14.2 \pm 11.7) \cdot 10^{15}$</td>
<td>$(19.1 \pm 10.7) \cdot 10^{15}$</td>
<td>$2.0 \cdot 10^{15}$</td>
<td>$4.8 \cdot 10^{15}$</td>
</tr>
<tr>
<td>AR area ($m^2$)</td>
<td>$(4.8 \pm 4.0) \cdot 10^{15}$</td>
<td>$(10.1 \pm 4.9) \cdot 10^{15}$</td>
<td>$(11.9 \pm 4.7) \cdot 10^{15}$</td>
<td>$3.2 \cdot 10^{15}$</td>
<td>$3.7 \cdot 10^{15}$</td>
</tr>
<tr>
<td>Unsigned magnetic flux in the AR area (G m$^2$)</td>
<td>$(7.7 \pm 7.1) \cdot 10^{17}$</td>
<td>$(21.1 \pm 13.0) \cdot 10^{17}$</td>
<td>$(29.2 \pm 13.0) \cdot 10^{17}$</td>
<td>$5.6 \cdot 10^{17}$</td>
<td>$7.1 \cdot 10^{17}$</td>
</tr>
<tr>
<td>Maximum strength of magnetic field in the AR (G)</td>
<td>$(1.31 \pm 0.41) \cdot 10^{3}$</td>
<td>$(1.66 \pm 0.48) \cdot 10^{3}$</td>
<td>$(1.84 \pm 0.52) \cdot 10^{3}$</td>
<td>$1.06 \cdot 10^{3}$</td>
<td>$1.20 \cdot 10^{3}$</td>
</tr>
<tr>
<td>Unsigned horizontal gradient in the AR area (G m)</td>
<td>$(6.1 \pm 5.4) \cdot 10^{11}$</td>
<td>$(13.4 \pm 7.4) \cdot 10^{11}$</td>
<td>$(16.5 \pm 8.2) \cdot 10^{11}$</td>
<td>$3.9 \cdot 10^{11}$</td>
<td>$4.3 \cdot 10^{11}$</td>
</tr>
</tbody>
</table>

**Note.** Columns 2 and 3 show the average values of the parameters for the $\geq$M1.0 and $\geq$X1.0 class flares correspondingly. Columns 4 and 5 show the thresholds, above which 95% of All $\geq$M1.0 and $\geq$X1.0 class flares were observed.
Sadykov & Kosovichev

Table 2
Importance of Magnetic Field Characteristics for the Forecast of ≥M1.0 Class Solar Flares

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Fraction of Negative Cases Below Threshold, %</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIL length (log)</td>
<td>0.63</td>
<td>1.41</td>
</tr>
<tr>
<td>PIL area</td>
<td>0.60</td>
<td>1.46</td>
</tr>
<tr>
<td>Unsigned magnetic flux in the PIL area (log)</td>
<td>0.63</td>
<td>1.41</td>
</tr>
<tr>
<td>Unsigned horizontal gradient in the PIL area (log)</td>
<td>0.64</td>
<td>1.48</td>
</tr>
<tr>
<td>Maximum gradient across the PIL (log)</td>
<td>0.56</td>
<td>1.15</td>
</tr>
<tr>
<td>Gradient-weighted PIL length (log)</td>
<td>0.62</td>
<td>1.45</td>
</tr>
<tr>
<td>R-value (log)</td>
<td>0.61</td>
<td>1.35</td>
</tr>
<tr>
<td>AR area (log)</td>
<td>0.44</td>
<td>0.66</td>
</tr>
<tr>
<td>Unsigned magnetic flux in the AR area (log)</td>
<td>0.49</td>
<td>0.86</td>
</tr>
<tr>
<td>Maximum strength of magnetic field in the AR (log)</td>
<td>0.29</td>
<td>0.30</td>
</tr>
<tr>
<td>Unsigned horizontal gradient in the AR area</td>
<td>0.44</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Table 3
Importance of Magnetic Field Characteristics for the Forecast of ≥X1.0 Class Solar Flares

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Fraction of Negative Cases Below Threshold, %</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIL length</td>
<td>0.84</td>
<td>2.68</td>
</tr>
<tr>
<td>PIL area</td>
<td>0.71</td>
<td>2.36</td>
</tr>
<tr>
<td>Unsigned magnetic flux in the PIL area</td>
<td>0.74</td>
<td>2.51</td>
</tr>
<tr>
<td>Unsigned horizontal gradient in the PIL area</td>
<td>0.83</td>
<td>2.81</td>
</tr>
<tr>
<td>Maximum gradient across the PIL</td>
<td>0.79</td>
<td>2.46</td>
</tr>
<tr>
<td>Gradient-weighted PIL length (log)</td>
<td>0.84</td>
<td>2.62</td>
</tr>
<tr>
<td>R-value (log)</td>
<td>0.84</td>
<td>2.47</td>
</tr>
<tr>
<td>Total AR area</td>
<td>0.51</td>
<td>1.32</td>
</tr>
<tr>
<td>Unsigned magnetic flux in the AR area (log)</td>
<td>0.60</td>
<td>1.91</td>
</tr>
<tr>
<td>Maximum strength of magnetic field in the AR area (log)</td>
<td>0.41</td>
<td>0.68</td>
</tr>
<tr>
<td>Unsigned horizontal gradient in the AR area (log)</td>
<td>0.49</td>
<td>1.29</td>
</tr>
</tbody>
</table>

“Scikit-Learn” (Pedregosa et al. 2011). The description of SVM can be found in Bobra & Couvidat (2015) and in Appendix B.

The computational cost of the SVM classifier scales with the number of cases in the “train” data set and the number of features (characteristics, descriptors) as $O(N^2 \times M)$ if $N \gg M$. On one hand, a large number of training samples should positively affect the classifier performance. On the other hand, the SVM classifier has many parameters that should be optimized, and the computing time quadratically increases with the size of the “train” data set. Thus, any possibility to reduce the number of cases that need to be classified should be utilized. In the previous section, we have found that the flaring cases mostly occur if a specific characteristic exceeds a certain threshold. We have also obtained that the PIL descriptors are more effective in the separation of the positive and negative cases. Thus, we first performed the classification based on the PIL characteristics only. We automatically classified a case as negative if any of its PIL characteristics were below the corresponding threshold. It was found that this procedure allows us to reduce the amount of data for the SVM classification by 74.4% (leaving about one-fourth of all cases) for the ≥M1.0 class flares and by 91.2% for the ≥X1.0 class flares. Only about 11.6% of positive cases for the ≥M1.0 and 14.0% for the ≥X1.0 class flares were misclassified as negative at this stage. To check the validity of this approach, we repeated the training procedure with the threshold values decreased by a factor of two, which led to the exclusion of 52.2% of cases (two times more cases need to be classified) for the ≥M1.0 class and 72.8% (three times more cases need to be classified) for the ≥X1.0 class cases. We have also checked how the inclusion of the global AR parameters (AR area, unsigned magnetic flux, maximum strength of magnetic field, and unsigned horizontal gradient) affect the forecasting result by repeating the training procedure with all 11 parameters.

For the SVM training, we normalize the “train” data set following Nishizuka et al. (2017): $Z = (X - \mu) / \sigma$, where $X$ is a non-normalized data set, $\mu$ is the mean, and $\sigma$ is the standard deviation. We use the same $\mu$ and $\sigma$ parameters to normalize the “test” data set. To find the optimal SVM kernel (among the Linear, RBF, Polynomial, and Sigmoid available in the Python Scikit-Learn package) and its parameters, we perform a cross-validation procedure on the “train” data set: divide it into two subsets (one simulating the train data set, and another simulating the test data set) 10 times, and then average the SVM results. As a measure of the SVM performance, we use the TSS metrics defined as

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN},$$

where TP is the true positive prediction (the number of positive cases predicted as positive), TN is the true negative prediction (number of negative cases predicted as negative), FP is the false positive prediction (number of negative cases predicted as positive), FN is the false negative prediction (number of positive cases predicted as negative). The TSS score is not sensitive to the class imbalance ratio (the relative number of positive and negative cases), and is zero for a pure negative prediction (when all cases are predicted as negative). The standard deviation of the TSS was estimated from the scores obtained during the cross-validation procedure with the optimal parameters.

5. Results

In Section 3, it was pointed out that the PIL characteristics separate flaring and nonflaring cases more effectively than the global (integrated) characteristics obtained for all of the ARs. The results in Tables 2 and 3 demonstrate that all PIL characteristics give approximately the same scores in both
Prediction of the global AR characteristics, the highest score is obtained for are only summarized in the second column of Table 4. For the exceed the scores for any PIL parameter. The Astrophysical Journal, (number of cases following classi-grams for each AR with 1 hr cadence. We estimated the results: TP = 194, TN = 42382, FP = 6654, and FN = 234 (including all cases in the test data set). For the flares, we obtained TSS = 0.84 ± 0.07 for the same “sigmoid” SVM kernel but with different parameters: C = 0.0001, γ = 10.0, and r = 0.0001, and the negative/positive class weights of 1/100. This TSS was derived from the following classification results: TP = 194, TN = 44991, FP = 6009, and FN = 8. Interestingly, the flare forecasts performed using only the PIL characteristics have almost the same TSS scores as the forecasts based on the full set of characteristics (including both the PIL and global AR characteristics). The TSS scores for the full set of characteristics are summarized in the third column of Table 4. For prediction of M1.0 solar flares, the inclusion of global characteristics even decreased the TSS score from TSS = 0.76 to TSS = 0.74. For prediction of X1.0 flares, we have obtained the same TSS = 0.84 score. The last column of Table 4 summarizes the results of the classification using the PIL parameters with the pre-classification threshold decreased by a factor of two. The 50% decrease of the threshold (which results in a smaller number of pre-classified samples) leads to an insignificant increase of TSS for the X1.0 flare prediction (from TSS = 0.84 to TSS = 0.85) and gives the same TSS for the M1.0 flare prediction. Thus, we can conclude that it is possible to pre-classify a significant number of cases (74.4% for the M1.0 class flares and 91.2% for the X1.0 class flares) by applying thresholds to the PIL parameters without a significant decrease of the prediction TSS score.

### 6. Discussion and Conclusion

In this paper, we have developed a machine-learning procedure solely based on the LOS magnetic field observations that are available in near-real time from space-based and ground-based observatories. The procedure is based on analysis of characteristics of the magnetic field PIL, which is automatically identified by performing the magnetogram segmentation formulated as an optimization task. The PIL characteristics were derived from the SDO/HMI magnetograms for each AR with 1 hr cadence. We estimated the effectiveness of these characteristics for forecasting M1.0 and X1.0 solar flares, and trained the SVM to maximize the TSS metrics. Interestingly, the univariate effectiveness scores are similar for all PIL characteristics, probably because the PIL characteristics (except, possibly, the Maximum gradient across PIL) correlate with each other (they depend on the same PIL length or the PIL area, which depends on the PIL length).

The obtained TSS scores TSS = 0.76 for prediction of M1.0 class flares, and TSS = 0.84 for prediction of X1.0 class flares, can be compared with the scores mentioned in other works. For example, Anastasiadis et al. (2017) reported TSS ≈ 0.5 for the prediction of C1.0 class flares, Shen et al. (2016) received a maximum of TSS = 0.371 for M1.0 class flares, Hada-Muranushi et al. (2016)—the TSS = 0.295 for M1.0 class flares, Liu et al. (2017)—TSS = 0.50 for M1.0 class flares. On the other hand, our TSS score for M1.0 is lower than those in the works of Bobra & Couvidat (2015, TSS = 0.817), Nishizuka et al. (2017, TSS = 0.88 for SVM classifier), and Raboonik et al. (2017, TSS = 0.856). Also, Nishizuka et al. (2017) reported a higher TSS score for X1.0 class flares (TSS = 0.88 for SVM classifier). Our results, solely based on the LOS magnetic field observations, are lower than those obtained with the use of vector magnetograms, but still comparable.

The score for M1.0 class flares received in our work is higher than the known expert predictions quoted by Nishizuka et al. (2017): TSS = 0.50 for the NICT Space Weather Forecasting Center and TSS = 0.34 for the Royal Observatory of Belgium (Devos et al. 2014). It is also higher than the TSS = 0.53 of the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC) deduced from Table 4 of Crown (2012). For the X1.0 flares, again, our result is higher than the expert prediction with TSS = 0.21 (the NICT Space Weather Forecasting Center, Nishizuka et al. 2017) and with TSS = 0.49 (SWPC NOAA, deduced from Table 4 of Crown 2012). We can conclude that the accurately tuned machine-learning technique, even if it is solely based on the LOS magnetic field measurements, can compete with the expert-based predictions.

It is necessary to discuss the influence of the data set construction on the prediction results. First, the way of the division of the data set into the “train” and “test” subsets can change the prediction scores. For example, the shuffled division (when the “train” and “test” subsets are not consequent in time, but all cases from one AR are kept in one subset) reduces the scores from TSS = 0.76 to TSS = 0.70 for M1.0 class flares, and from TSS = 0.84 to TSS = 0.63 for X1.0 class flares. The strong difference in the TSS score for X1.0 class flares is caused by a low number of X-class flares in the data set. In this work, we relied on the NOAA AR detection and considered every case with the detectable PIL, which already causes the data set to be subjective to the PIL detection method. Nishizuka et al. (2017) used their own method to detect ARs, which definitely leads to another data set with a larger number of

---

**Table 4**

Comparison of TSS Scores for Different Methods of Prediction of M1.0 and X1.0 Class Solar Flares

<table>
<thead>
<tr>
<th>Prediction of M1.0 flares</th>
<th>PIL Characteristics Only</th>
<th>PIL + Global Characteristics</th>
<th>50% Decreased Cutoff Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.76 ± 0.03</td>
<td>0.74 ± 0.03</td>
<td>0.76 ± 0.03</td>
</tr>
<tr>
<td>Prediction of X1.0 flares</td>
<td>0.84 ± 0.07</td>
<td>0.84 ± 0.07</td>
<td>0.85 ± 0.04</td>
</tr>
</tbody>
</table>

**Note.** The standard deviations are estimated using a cross-validation procedure.
cases. Bobra & Couvidat (2015) reduced the actual data set by cutting out a randomly selected portion of negative cases. Thus, to guarantee the accurate comparison of different prediction methods, one should unify the starting data set and its division into the “train” and “test” subsets. Such attempts were done previously (Barnes et al. 2016), and hopefully will be continued in the future.

The important role of PIL in the flare development process was pointed out in many observations, simulations, and forecasts of solar flares. Generally, the PILs are characterized by highly sheared magnetic fields, strong field gradients, and complicated topology of neighboring magnetic field structures. These properties result in a substantial amount of free magnetic energy that can be released in flares. It is not surprising that many flares are developed locally in the PIL vicinity. Our study statistically confirms the importance of the PIL characteristics for flare forecasting. In particular, it demonstrated that the PIL characteristics obtained just from the LOS magnetic field component can be used to obtain flare predictions compatible with expert-based forecasts and comparable to the predictions that are based on full vector magnetic field observations. However, our results are accompanied by a significant number of false positive predictions. Generally, a more accurate comparison of machine-learning-based and expert-based predictions is required. Despite the promising results, we should always keep in mind that the prediction is metrics-dependent.

In this work, we maximize the TSS in a single parameter setup. Maximizing other metrics can result in other optimal SVM parameters and prediction scores (Bobra & Couvidat 2015). Further work is needed to develop algorithms for quantitative prediction of the flare class and physical properties (eruptive or noneruptive nature, geo-effectiveness etc.).

The authors thank the anonymous referee for a helpful and detailed review of the paper. The authors thank the GOES and SDO/HMI teams for the availability of high-quality scientific data. The authors also thank D. Laptev for valuable discussions of the magnetogram segmentation algorithm. The research was partially supported by the NASA Grants NNX14AB68G and NNX16AP05H.

Appendix A
Magnetogram Segmentation and the PIL Detection Algorithm

Suppose $B$ is a magnetic field strength map (magnetogram), $Z_i$ is a class of pixel $i$ of the magnetogram (i.e., “positive,” “negative” or “neutral”), $N$ is the total number of pixels in the magnetogram, and $\varepsilon(i)$ is a neighborhood (e.g., the closest 8 pixels) of pixel $i$. The magnetogram segmentation can be formulated as the following optimization procedure to maximize function $p(Z, B)$ for a given $B$ by finding optimal classification $Z_{\text{max}}$ (Laptev 2011):

$$p(Z_{\text{max}}, B) = \max_Z p(Z, B) \propto \prod_{i=1}^{N} \phi_i(Z_i, B_i) \prod_{j \in \varepsilon(i)} \phi(Z_i, Z_j).$$

Here $\phi_i(Z_i, B_i)$ and $\phi(Z_i, Z_j)$ are the scoring functions for each pixel depending on the magnetic field strength and assumed classes of pixels. The choice of the scoring function defines segmentation characteristics and, in fact, should do the following: separate the segments of positive and negative magnetic field polarity and avoid very small segments with weak field probably coming from noise in the data. We use the scoring functions suggested by Chernyshov et al. (2011):

$$\phi_i(Z_i, B_i) = e^{-C_1||B_0-B_i||}, \text{ for } Z_i \text{ “positive”}$$

$$\phi_i(Z_i, B_i) = e^{-C_1||B_0+B_i||}, \text{ for } Z_i \text{ “negative”}$$

$$\phi_i(Z_i, B_i) = e^{-C_1|B_i|}, \text{ for } Z_i \text{ “neutral”}$$

$$\phi(Z_i, Z_j) = e^{C_{\text{pair}}[Z_i=Z_j]},$$

where parameters $C_1 = 1.0, C_2 = 1.0, C_{\text{pair}} = 20$, and $B_0 = 1000$ G are chosen to obtain a stable segmentation of magnetic polarities in strong field regions. Here $[Z_i \neq Z_j]$ is equal to 1 if $Z_i \neq Z_j$, and zero otherwise. Following Laptev (2011), the function $p(Z, B)$ is interpreted as conditional probability density function $p(Z|B)$, and is approximated by the factorized probability density function $q(Z) = \prod_{i=1}^{N} q_i(Z_i)$. To measure how strongly the factorized distribution deviates from the actual, one can use the Kullback–Leibler (KL) divergence (Bishop 2006). In order to find the best approximating factorized distribution, $q(Z)$, one can minimize the KL divergence:

$$\min_{q(Z)} KL(q||p) = -\int q(Z) \log \frac{p(Z|B)}{q(Z)} dZ.$$

Here we keep the original notation for KL divergence $KL(q||p)$ between distributions $q$ and $p$ introduced in Bishop (2006). The optimal $q(Z)$ is given by solution of the equation (following Chernyshov et al. 2011):

$$q_i(Z_i) = \frac{1}{C} \exp(\log(\phi_i(Z_i))) - C_{\text{pair}} \sum_{t \in \varepsilon(i)} \sum_{j \neq i} q_i(Z_j),$$

which can be found iteratively:

$$q_{i,\text{new}}(Z_i) = \frac{1}{C} \exp(\log(\phi_i(Z_i))) - C_{\text{pair}} \sum_{t \in \varepsilon(i)} \sum_{j \neq i} q_{i,\text{old}}(Z_j).$$

Using this equation, one can calculate the factorized distribution multiplier $q_i$ for each pixel $i$ and its assumed class $Z_i$ (“positive,” “negative,” or “neutral”). Because the factorized distribution represents the product of multipliers for each pixel, one can simply maximize $q_i(Z_i)$ for each pixel $i$ separately and obtain $Z_{\text{max}}$.

For identification of PIL in ARs, we smooth the original HMI magnetogram using the Gaussian filter with width $\sigma = 1.5$, and apply the segmentation algorithm. Then, we apply a morphological dilation procedure separately for positive and negative segments (i.e., expand each segment to include neighboring pixels), and find the PIL as an intersection of the dilated positive and negative segments. Finally, we filter out all small islands of the PIL with the number of pixels less than 3% of the total number of pixels occupied by PIL. This approach is quite robust and allows us to automatically identify the PIL and calculate magnetic field properties.

Appendix B
Description of the SVM Classifier

The SVM (Cortes & Vapnik 1995) classifier is the widely used supervised-learning classification algorithm. The SVM finds a plane in the descriptor space, which optimally separates the positive and negative cases by solving the following

\[ \text{minimize } \sum_{i=1}^{N} \max(0, 1 - y_i (\mathbf{x}_i \cdot \mathbf{w} + b)) \]
functional minimization problem:
\[
\min_{\omega,\epsilon} L = \frac{1}{2}||\omega||^2 + C \sum_{i=1}^{m} W_i \epsilon_i, \quad \gamma(y_i(\langle \omega, x_i \rangle + b) \geq 1 - \epsilon_i, \quad \epsilon_i \geq 0,
\]
where \( \omega \) is a vector normal to the separating plane; \( i \) is case number in the “train” data set, varying from 0 to \( m \); \( C \) is a soft-margin parameter; \( W_i \) is the weight of the group, which the case \( i \) belongs to; \( \epsilon_i \) is a measure of misclassification of case \( i \); and \( y_i \) is a constant equal to 1 for positive cases, and \(-1\) for negative cases. After some transformations, this problem becomes a quadratic minimization problem: the functional depends only on scalar products of vectors of characteristics \( \langle x_i, x_j \rangle \). To achieve better separation between the positive and negative cases, very often the so-called Kernel trick is used. The scalar product of characteristics in the functional is replaced by a functional minimization problem:

\[
\min_{\omega,\epsilon} L = \frac{1}{2}||\omega||^2 + C \sum_{i=1}^{m} W_i \epsilon_i, \quad \gamma(y_i(\langle \omega, x_i \rangle + b) \geq 1 - \epsilon_i, \quad \epsilon_i \geq 0,
\]

where \( \gamma \), \( r \), and \( d \) are tuning parameters. The other SVM parameters are the soft-margin parameter and weights for both classes (multipliers of the soft-margin parameter). One needs to optimize all of these parameters during the cross-validation procedure.

**ORCID iDs**

Viacheslav M Sadykov @ https://orcid.org/0000-0002-4001-1295

**References**

Ahmed, O. W., Qahwaji, R., Colak, T., et al. 2013, SoPh, 283, 157
Breiman, L. 2001, Machine Learning, 45, 5
Devos, A., Verbeek, C., & Robbrecht, E. 2014, JSWSC, 4, A29
Falcohol, D., Baragouty, A. F., Khazanov, I., & Moore, R. 2011, SpWea, 9, S04003
Gagnon, C., Pariat, E., Leake, J. E., & Vilmer, N. 2017, JSWSC, 7, 17
Lapet, D. 2011, Specialist Dissertation “Search for Informative features on Solar Magnetoograms” (Moscow: Department of Computational Mathematics and Cybernetics, Moscow State University)
Shcherrer, P. H., Schou, I., & Bush, R. 1. 2012, SpWea, 275, 207