Improved and Interpretable Solar Flare Predictions with Spatial & Topological Features of the Polarity-Inversion-Line Masked Magnetograms

Hu Sun[1], **Ward Manchester**[2], **Yang Chen**[1]

[1]Department of Statistics, University of Michigan, Ann Arbor
[2]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor

Key Points:

- We adopt a polarity inversion line (PIL) detecting algorithm to obtain PIL masks for the B_r channel and several SHARP parameter maps.
- We construct two sets of spatial statistics features and a collection of topological features based on the PIL masks.
- Our newly constructed features, by itself or joint with topological features, can significantly improve flare predictions than using SHARP parameters only.

Plain Language Summary:

Our research is targeted at improving the accuracy of solar flare classification by training machine learning models with new interpretable features beyond well-known physics-based predictors. We count the number of closed loops and calculate multiple summary statistics of the spatial distribution of high-resolution magnetic field images of solar active regions to boost the classification result of strong and weak flares. Our results reveal that the spatial distribution of local physical quantities derived from the magnetic field images, beyond those commonly adopted, aggregated quantities, can be helpful to improve flare predictability.

Corresponding author: Yang Chen, ychenang@umich.edu

Abstract

Many current research efforts undertake the solar flare classification task using the Space-weather HMI Active Region Patch (SHARP) parameters as the predictors. The SHARP parameters are scalar quantities based on spatial average or integration of physical quantities derived from the vector magnetic field, which loses information of the two-dimensional spatial distribution of the field and related quantities. In this paper, we construct two new sets of spatial features to expand the feature set used for the flare classification task. The first set uses the idea of topological data analysis to summarize the geometric information of the distributions of various SHARP parameters. The second set utilizes tools coming from spatial statistics to analyze the vertical magnetic field component B_r and summarize its spatial variations and clustering patterns. All features are constructed within regions near the polarity inversion lines (PILs) and classification performances using the new features are compared against the SHARP parameters (also along PIL). We found that using the new features can improve the skill of the flare classification model and new features tend to have higher feature importance, especially the spatial statistics features. This potentially suggests that even using a single channel B_r, instead of all SHARP parameters, one can still derive strongly predictive features for flare classification.

1 Introduction

Solar flares occur in magnetic regions of concentrated free energy, which is susceptible to spontaneous reconnection that rapidly heats the plasma producing flare emission. At the photosphere these fields typically take the form of active regions that are characterized by strong horizontal field gradients, long, well defined polarity-inversion lines (PILs), and complex flux distributions (Barnes, Leka, Schumer, & Della-Rose, 2007; Falconer, Moore, & Gary, 2002, 2003, 2006). Vector magnetogram observations reveal the the magnetic field runs nearly parallel to the PIL, where the field is in a configuration that possesses significant free energy to drive flares. In addition to the free energy content, the other critical feature related to flare onset is the magnetic field complexity represented by the intermixing of opposing magnetic flux. At the simplest level, such complexity is found in strong gradients along the PIL (Falconer et al., 2003; Schrijver, 2007) and in more extreme cases from flare production from delta-spot active regions, which contain both polarities within single penumbra of uniform polarity. A well known example is AR 10486, which produced a series of X-class flares in the fall of 2003.

The question naturally arises as to whether there are unique features which may serve as strong discriminators or predictors of flares. Schrijver (2007) found that the amount of unsigned flux within a distance of 15 Mm of the PIL was such a discriminator. Phenomenological studies have revealed a number of physical discriminators for flare production based solely on the observed line-of-sight component of the field (e.g. Barnes et al., 2007; Falconer, 2001; Falconer et al., 2002, 2006; Leka & Barnes, 2003a, 2003b). These works describe scalar quantities that are measures of the unsigned flux and the length of the PIL, which have proven useful for flare prediction. In the case of Falconer (2002), four measures of magnetic free energy were derived based on 3-component vector magnetograms: the length of the PIL exhibiting strong shear, vertical electric current, total unsigned flux, and the current helicity. These works were followed by more elaborate flare prediction studies employing more than 20 scalar quantities derived from Space Weather HMI Active Region Patches (SHARPs). These SHARP variables are derived from largely-uninterrupted high-resolution high-cadence full disk vector magnetograms, which provide sufficient data for successful application of machine learning for flare prediction (e.g. (Bobra & Couvidat, 2015; Bobra et al., 2014; Y. Chen et al., 2019; Florios et al., 2018; Jiao et al., 2020; Liu, Deng, Wang, & Wang, 2017; Nishizuka, Sugiura, Kubo, Den, & Ishii, 2018; Wang et al., 2020)).

Statistical analysis typically strives to determine discriminant functions which may be linear or nonlinear functions of variables constructed from the observed vector magnetic field, which are related to distributions of free energy in the system. Flares follow from two processes, first the buildup of free magnetic energy and second the release of that energy by magnetic reconnection. For the first component there are meaningful measures of buildup of energy necessary flares such as the AR size as denoted by unsigned magnetic flux and total electric current, current helicity or total free energy itself. Numerous studies illustrate the predictive power of these quantities (e.g. Y. Chen et al., 2019). However, whether it is sufficient to produce a flare involves much more subtlety as it relates to the complexity of the magnetic field which ties into the formation of current sheets, which initiates the flare energy release. Because of this complexity of the reconnection process, there can be a wide range of free energy densities found in flaring ARs, which emphasizes the challenging aspect of predicting flares.

In recent years, data driven flare forecasting has caught a lot of attention in the field of space sciences. Many machine learning algorithms have been adopted for solar flare prediction, either with or without operational forecasting in mind (Barnes et al., 2007). Bobra et al. (2014) introduced the Space-weather HMI Active Region Patch (SHARP) parameters, which are derived from the magnetograms of the HMI/SDO images and have been used by many solar flare prediction models in recent years (e.g. Barnes et al., 2016; Bobra & Couvidat, 2015; Camporeale, 2019; Y. Chen et al., 2019; Florios et al., 2018; Jiao et al., 2020; Leka & Barnes, 2018; Leka et al., 2019a, 2019b; Liu, Liu, Wang, & Wang, 2019; Muranushi et al., 2016; Nishizuka, Kubo, Sugiura, Den, & Ishii, 2021; Nishizuka et al., 2018). The success of using the SHARP parameters in solar flare forecasting showcases the importance of the SHARP parameters in triggering solar eruptions. Despite this, researchers suspect that the SHARP parameters may not capture the full information that the HMI/SDO magnetogram images contain which are relevant for flare forecasting. There are efforts of using the deep neural network methods which directly takes the HMI/SDO magnetogram images directly to predict solar eruptions (e.g. the Long Short Term Memory network adopted by Y. Chen et al. (2019) and Liu et al. (2019)). The neural network models, however, suffer from poor interpretability and it is hard to learn new insights on the underlying physics/mechanism of flare eruptions from those models. In an effort to obtain interpretable solar flare forecasting model, while at the same time pushing the frontiers of discovering new physics of solar eruptions, we present our new results and findings on constructing spatial and topological features that are important for solar flare predictions, from the magnetograms of HMI/SDO.

In Y. Chen et al. (2019), it is shown that the derived features from using deep neural network (autoencoder) give similar performance as compared to the SHARP parameters when adopted to solar flare predictions. This result, however, does not conclude that the SHARP parameters already contain all useful information from magnetograms for flare predictions, due to the fact that general-purpose image/video feature reduction techniques typically result in high-dimensional, highly correlated features, many of which are redundant for flare predictions. Despite that efficient dimension reduction and feature selection techniques are adopted in Y. Chen et al. (2019) to alleviate this issue, the amount of random noises brought in with this high-dimensional feature construction with limited amount of training data still makes the performance not ideal. Meanwhile, it is important to realize that even in this unfavorable situation of directly extracting features from magnetogram, we can achieve similar flare prediction performance as when using only the SHARP parameters. This possibility makes it promising to further investigate alternative feature construction methods that result in information complimentary to those contained already in the SHARP parameters. More recently, some promising results are already shown along this direction: in Deshmukh, Berger, Meiss, and Bradley (2020), topological features are derived for the B_r channel, which are shown to improve the performance of solar flare predictions. Here, we extend this approach by combining both spatial and topological features of the polarity inversion line masked magnetograms

(HMI/SDO active regions) with the SHARP parameters for solar flare predictions. We show that the results of the prediction models are better when we use all three sets of features, and that the spatial statistics features are of key importance. Furthermore, we point out the interpretations of the spatial statistics features and relate it back to the underlying mechanism of flare formation, thus using these features to inform the discovery of new physics.

The remainder of this paper is organized as follows. In Section 2, we first give a motivating example to show the potential of extracting shape information from HMI images for flare classification, then we describe the methodology and details of constructing topological features and spatial statistics features out of HMI/SHARP images. In Section 3, we present the data used for empirical study and the prediction results based on newly constructed features and corresponding interpretations. Section 4 concludes.

2 Feature Construction

The SHARP parameters, as introduced in (Bobra et al., 2014), contain several essential physical quantities. They are calculated from the three-component two-dimensional HMI magnetic field data of an active region. It is important to note that all the SHARP parameters are scalar values summarizing physical quantities of the *whole* active region, which are calculated based on spatial averages (e.g. the MEANGAM) or spatial integration (e.g. the TOTUSJH). As a result, the SHARP parameters do not contain any information regarding the spatial distribution of the physical quantities. A refined version of the SHARP parameters taking into account of some information of the spatial distribution is the SHARP parameters weighted by the polarity inversion line (PIL) mask, as given by Wang et al. (2020). The PIL mask is a weighted mask which puts high weight on the region where the vertical component of the magnetic field is zero ($B_r = 0$) between regions of strong opposite polarity. Empirically, it is not practical to accurately pinpoint all pixels with $B_r = 0$, so adjacent regions are also given small weights and overall the PIL mask is giving high weights to regions around the PIL and near-zero weight to other regions. However localized, the PIL-weighted SHARP parameters still cannot reflect the spatial characteristics of the magnetic field. In this section, we introduce two sets of features constructed with the aim of augmenting the SHARP parameters to consider local and global spatial distribution of the magnetic field components around the polarity inversion line: one based on persistence homology in topology and the other based on concepts from the field of spatial statistics. We will apply these newly constructed features to flare classification and prediction and develop their utility in context of existing literature on related topics, e.g. the weighted SHARP parameters by Wang et al. (2020) and the topological data analysis enabled feature construction by Deshmukh, Berger, Bradley, and Meiss (2020).

2.1 Motivating Example & PIL Mask Construction

To motivate our feature construction machinery, we illustrate the SHARP parameters with a concrete example: comparing a strong and a weak flare that occurred in HARP region 377. In Figure 1, we show two series of HMI magnetograms and some derived images of HARP region 377, one for an M class flare (top row) and the other for a B class flare (bottom row), all images are collected 1 hour prior to the moment of peak flare intensity. Focusing on (a) and (d) first, one can see that before the M flare, the high magnetic flux pixels of opposing polarities are more clustered and concentrated in close proximity to one another. This distribution coincides with the formation of a well-defined polarity inversion line (PIL) (where $B_r = 0$) that separates the regions of opposing polarity. The PIL is picked up by the corresponding polarity inversion line mask shown in (b) and (e). PILs are the epicenter of flare activity and have been used as a discriminator of flare regions in prior studies (Falconer et al., 2006; Wang et al., 2019). Similarly,

we concentrate on the PIL regions by calculating the PIL in order to distinguish strong and weak flares.

We derive the PIL mask based on the method given in Schrijver (2007), which follows three steps. First, we produce two bitmaps, one for the positive field where pixels with $B_r > 200$G are set to 1 and 0 otherwise, one for the negative field where pixels with $B_r < -200$G are set to -1 and 0 otherwise. Second, we need to derive the positive and negative masks using a density-based clustering algorithm (Ester, Kriegel, Sander, & Xu, 1996). Finally, we need to multiply the two masks after convolving each mask with a Gaussian kernel with a bandwidth of 10 pixels. Details of the derivations can also be found in Wang et al. (2019).

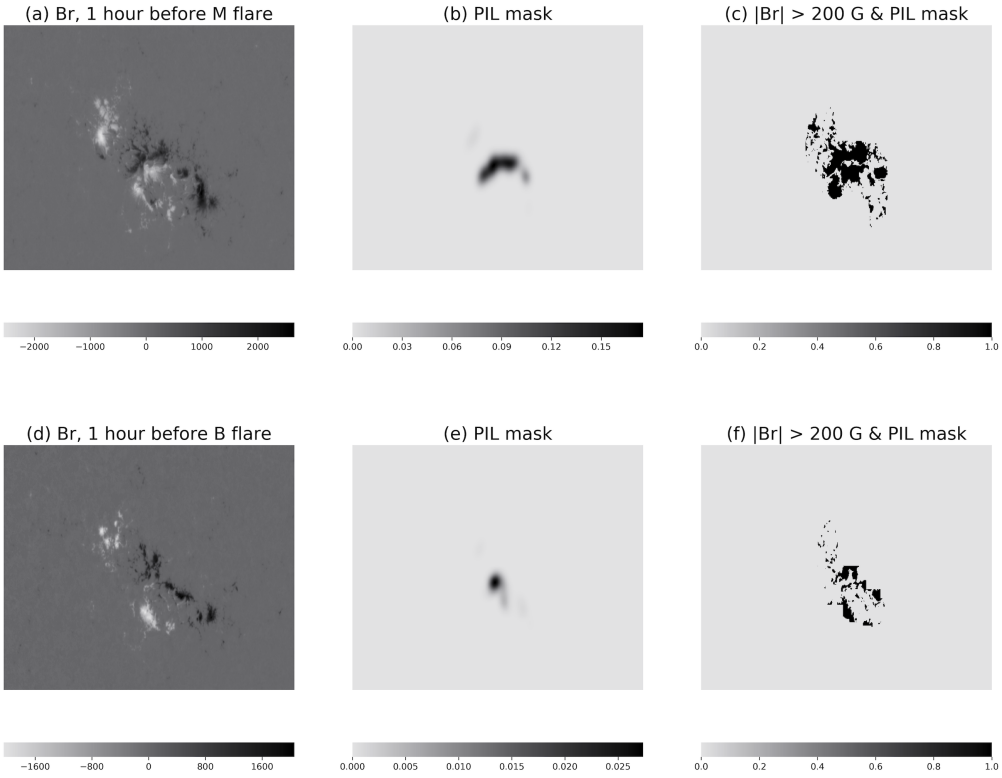

Figure 1. Two HMI magnetograms (vertical component) and derived images of HARP 377. The two rows consist of 3 related images: (left) the Br channel, (middle) the PIL weight mask, (right) bitmap labeling pixels with absolute flux over 200 G and locate in the PIL region. The top row is collected 1 hour before the M-class flare peaked at 2011.02.13 17:38:00 while the bottom row is collected 1 hour before the B-class flare peaked at 2011.02.13 00:42:00. Note the more energetic flare occurs at a later time when more flux has emerged and the PIL is significantly longer. Different scales are used to bring out structure in B_r and the PIL mask.

If one only compares the SHARP parameters, such as the total unsigned flux, the difference between the M flare and B flare is not necessarily significant because both flares have regions with relatively high flux values. But if one focuses on the spatial distributions of the high flux regions, such as how clustered these regions are and what geometric shapes they form jointly and individually, then extra discriminating power for telling the flares apart would emerge. In the following two sub-sections, we will focus on the HMI magnetograms (the B_r component in particular) restricted in the PIL region (non-zero

PIL weight region), for which we derive topological and spatial features. We concentrate on the PIL region since it removes many irrelevant regions outside of the flaring region (Wang et al., 2020), which can create potential noises on the derived features.

Besides the B_r channel, we also derive eight additional 2D SHARP related parameter maps as follows. Here, we calculate the SHARP intensive physical quantities for every pixel, but instead of taking the averages or spatially integrating them to form the scalar SHARP parameters, we retain the 2D distribution of each of these maps and multiply it by the same B_r PIL mask. Table 1 gives a summary of all nine SHARP parameter masks we collect/construct.

Channel	Formula	Unit
Br	$\mathbf{B}_z$	G
GAM	$\arctan\left(\frac{\mathbf{B}_h}{\mathbf{B}_z}\right)$	Degree
GBT	$\sqrt{\left(\frac{\partial \mathbf{B}}{\partial x}\right)^2+\left(\frac{\partial \mathbf{B}}{\partial y}\right)^2}$	G$\times$ Mm^{-1}
GBH	$\sqrt{\left(\frac{\partial \mathbf{B}_h}{\partial x}\right)^2+\left(\frac{\partial \mathbf{B}_h}{\partial y}\right)^2}$	G$\times$ Mm^{-1}
GBZ	$\sqrt{\left(\frac{\partial \mathbf{B}_z}{\partial x}\right)^2+\left(\frac{\partial \mathbf{B}_z}{\partial y}\right)^2}$	G$\times$ Mm^{-1}
USJZ	$\lvert\left(\frac{\partial \mathbf{B}_y}{\partial x}-\frac{\partial \mathbf{B}_x}{\partial y}\right)\rvert$	A
USJH	$\lvert \boldsymbol{J}_z \times \mathbf{B}_z \rvert$	G^2 m $^{-1}$
POT	$\left((\mathbf{B}_x-\mathbf{B}_x^{POT})^2+(\mathbf{B}_y-\mathbf{B}_y^{POT})^2\right)$	erg cm^{-3}
SHR	$\arccos\left(\frac{\mathbf{B}_x^{POT}\times\mathbf{B}_x+\mathbf{B}_y^{POT}\times\mathbf{B}_y+\mathbf{B}_z^2}{\sqrt{\mathbf{B}_x^{POT2}+\mathbf{B}_y^{POT2}+\mathbf{B}_z^2}\sqrt{\mathbf{B}_x^2+\mathbf{B}_y^2+\mathbf{B}_z^2}}\right)$	Degree

Table 1. SHARP parameter masks, where the formulas in the middle column are applied to every pixel of the HMI magnetogram. Here, $\mathbf{B}_x, \mathbf{B}_y, \mathbf{B}_z$ are the x, y, z components of the magnetic field and $\mathbf{B}_x^{POT}, \mathbf{B}_y^{POT}$ the potential field components respectively. Detailed definition of the SHARP parameters can be found in Table 3 of Bobra et al. (2014).

The same feature construction procedure is then applied to all nine SHARP parameter maps, which are created by multiplying by the 2-D SHARP distributions by the B_r mask. For the topological features alone, our work using all nine PIL maps can be seen as an extension to the topological data analysis in Deshmukh, Berger, Bradley, and Meiss (2020), which mainly focused on the original B_r component when constructing topological features. For spatial statistics features, which have not yet been adopted by others in literature as far as we are aware of, we only focus on the B_r component for the ease of interpretation. All of the features, both topological and spatial statistics features, are derived only for the PIL region. As a comparison, we use the SHARP parameters weighted by the PIL mask as a benchmark feature set, which has been shown to improve flare predictions as compared to unweighted SHARP parameters (Wang et al., 2020).

2.2 Topological Features

The derivation of our topological features is based on the TDA analysis in Deshmukh, Berger, Bradley, and Meiss (2020), Deshmukh, Berger, Meiss, and Bradley (2020), but we have these extensions:

- we expand the image set to be analyzed from just the B_r mask to multiple other SHARP parameters masks,
- we restrict our TDA application to the PIL region only to remove noises in feature construction,
- we adopt the topological features for strong (M/X) versus weak (B) flare classification purpose.

In this subsection, we briefly review the methodology first and then highlight the differences of our data pipeline compared to existing literature.

Topological data analysis, a.k.a. TDA, (see Wasserman (2018) for a brief review) is a field investigating the mathematical shape of a point cloud or a function. It establishes the concept of topological invariance between objects that share the same topological property, such as a ball and cube, or a cup and a donut. TDA can be applied to analyze the 2D SHARP parameter maps since all SHARP parameters can be viewed as a 2D function with each pixel's x and y coordinate being the functional input. It is not practical to summarize all shape information of a function or a point cloud, and a very simple summary statistics in TDA about the topology property is the Betti number (Munch, 2017). Betti numbers, typically denoted as $\beta_0, \beta_1, \beta_2, \ldots$, counts the number of holes of $0, 1, 2, \ldots$ dimension. In 2D, β_0, β_1 represents the number of connected components and loops. In our analysis, we just focus on the β_1 Betti number, counting the number of loops within the 2D SHARP masks.

Because every 2D SHARP parameter map is a digitized image, it is not a continuous but a discretized surface. Therefore, we need a proper definition of loop in these 2D images. Take the B_r mask as an example, if one thresholds the B_r image at various levels and only keeps pixels below the threshold and within the PIL region, forcing all other pixels as 0, one obtains distributions as found in Figure 2 shown below.

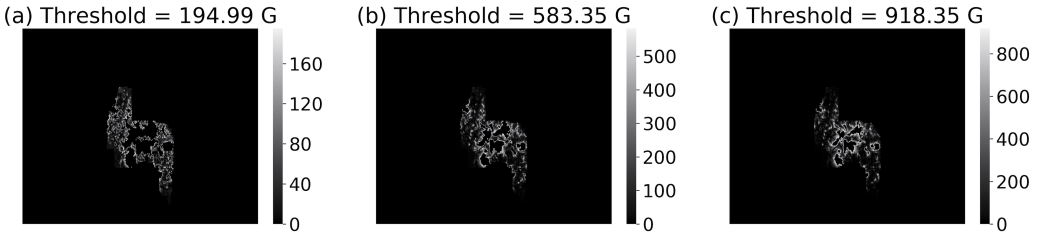

Figure 2. B_r mask of HARP 377, 2011.02.13 17:38:00. (a), (b), (c) only keep pixels whose $|B_r|$ value is below the threshold (194.99G, 583.35G, 918.35G) and within the PIL region.

One can notice that around the center of each image, there are several closed loops coming into existence when one increases the threshold from 194.99G to 583.35G; and that some of the loops disappear while others remain but shrink in size when the threshold level is further increased to 918.35G. Intuitively, these loops are regions of high $|B_r|$ values within the PIL region, or in other words, they represent the peaks/valleys of the spatial distribution of B_r. This sequence of thresholded images forms a *filtration*, where the set of all nonzero pixels of one image is a subset of the nonzero pixels of the next image with a higher threshold.

To give these kinds of loops in a digitized image a rigorous definition, we invoke the concept of *cubical complex*, as suggested by Deshmukh, Berger, Bradley, and Meiss (2020). For every image with threshold τ as shown in Figure 2, we generate a corresponding bitmap. Each bitmap sets all pixels that are: 1) below the threshold τ, 2) within the PIL region, as 1, and 0 otherwise. The loops are then defined for these bitmaps.

As a schematic example, in Figure 3 we show two bitmaps with red pixels being 1 and white pixels being 0. Subfigures (a) and (b) show two bitmaps generated from a low and a high threshold value, respectively. In a cubical complex, two pixels are connected if they share an edge or vertex and a loop is a closed, hollow area surrounded by a series of connected pixels. In (a) we have two loops, labeled as 1 and 3. As we increase the threshold, those pixels above the low threshold but under the high threshold become red and some new loops are formed (loop 2 in (b)), some remain but shrink in size (loop 1 in both (a) and (b)) and some disappear (loop 3 in (a)).

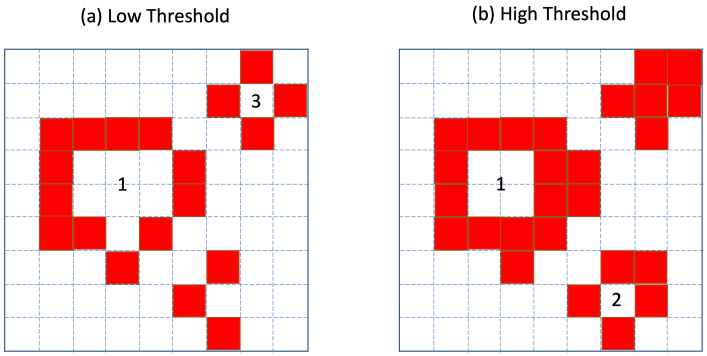

Figure 3. Example of loops formed by cubical complexes. In a cubical complex, two pixels are connected if they share a vertex or an edge. (a) at a low threshold, there are two loops formed, labeled as 1 and 3. (b) at a higher threshold, more pixels are selected (marked red), and loop 3 is closed, loop 2 comes into existence and loop 1 remains alive.

Every loop in this cubical complex has its own life-cycle. Each has a birth time and a death time, represented by the threshold values at which the loop emerges and disappears. A graph showing the birth-death pair of all loops is called the *persistence diagram* (Munch, 2017). When a loop has its death time much larger than the birth time, this loop indicates a strong peak/valley in the 2D distribution of a SHARP parameter map. Such loops appear as scattered points away from the 45 degree line in the persistence diagram. In Figure 4, we show the persistence diagrams for all loops of the B_r mask of 3 flares coming from HARP region 377. One can see that compared to the two M flares, the persistence diagram of the B flare has less loops that are way off the 45 degree line, indicating that there are fewer high flux regions prior to the B flare.

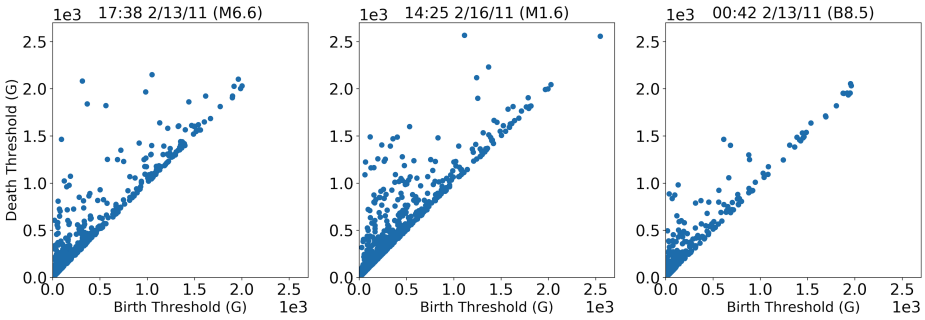

Figure 4. Examples of persistence diagrams for two M flares and a B flare of HARP region 377. Every subplot has the flare peak time, flare class and flare intensity in the title. One can see that compared to M flares, B flare has less loops that are persistent, indicating that there are fewer high flux regions in the PIL region. All B_r masks analyzed are collected 1 hour prior to the flare time. Both axes in each plot have the unit Gauss.

In order to extract features, namely Betti number β_1, from such diagrams, we do the following. For every SHARP parameter map in Table 1, we first calculate the pixel-level percentiles at $5\%, 10\%, \ldots, 90\%, 95\%$ across all pixels of our dataset. For the B_r, these percentiles are: $19.4\text{G}, 42.8\text{G}, \ldots, 918.3\text{G}, 1165.8\text{G}$. Other SHARP parameters have their corresponding percentiles. Then we derive the persistence diagram for every SHARP parameter map using the procedure above. Finally, we count, at each percentile, how many loops are still alive. Because we have 19 percentiles, we have 19 Betti number features, and since we have 9 SHARP maps, in total we obtain $19 \times 9 = 171$ topological features for a single flare.

2.3 Spatial Statistics Features

In the previous subsection, we reviewed the TDA analysis on SHARP parameter maps, which is able to summarise the shape information of the spatial distribution of corresponding SHARP parameter maps. Using the Betti number β_1 as a summary statistics, we can locate the peaks/valleys of a spatial distribution of any SHARP parameter. However, there are other aspects of the spatial distribution, such as the density of high flux pixels and the spatial variation of a SHARP parameter, that cannot be summarised by the TDA. *Spatial statistics* techniques (see Gelfand, Diggle, Guttorp, and Fuentes (2010) and references therein), on the other hand, provides natural tools to extract these information from spatial distributions. In our spatial statistics analysis, we focus on the B_r maps only for the sake of interpretability. The two relevant spatial statistics tools that we adopt here are the Ripley's K function (Ripley, 1976) and the variogram (Cressie & Hawkins, 1980; Oliver & Webster, 2015).

2.3.1 Ripley's K Function

We briefly review the definition and interpretation of the Ripley's K function here. Interested readers can refer to Ripley (1976) for more details. The Ripley's K function is a functional summary of the density of a point cloud at various scales. Consider a set of points with coordinates $\{(x_i, y_i)\}_{i=1}^n$, which are all located within an area of size A. The Ripley's K function is a function of the scale parameter d, defined as

$$L(d) = \sqrt{\frac{A \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} k_{i,j}}{\pi n(n-1)}}, \tag{1}$$

where $k_{i,j} = 1$ if point (x_i, y_i) and point (x_j, y_j) are within Euclidean distance d. Intuitively, the Ripley's K function calculates the proportion of pairs of points that are within distance d. Since d can take on any arbitrary value, the Ripley's K function reflects the clustering and dispersion properties of the points in the point cloud.

The area size A is an important parameter in the calculation. In Figure 5 (a) and (b) we show two identical point clouds, generated from a 2D uniform distribution, located in areas of different sizes. In (a), the red bounding box shows the area considered when calculating the Ripley's K function for the point cloud. In (b), the area is the whole region $[0, 20]\times[0, 20]$. With the same point cloud, one can see that their Ripley's K function, plotted in (d), are very different. The Ripley's K function for pattern A and B are relatively straight, meaning that points are nearly uniformly randomly distributed in the area. Since when all points are fully randomly distributed, the proportion of points within radius d grows linearly with the size the circle πd^2, so $L(d) \sim O(d)$ and the slope is related to $\sqrt{A}$. In Figure 5 (c),we show another point pattern coming from a Gaussian mixture distribution. Its corresponding Ripley's K function are plotted in (d) in green. With a relatively small d, the Ripley's K function is concave and indicates a clustering pattern of points. As d becomes larger, the function becomes convex and indicates that at relatively large scale, distribution of points are dispersed (scattered in five clusters).

We need to define two key elements in order to calculate the Ripley's K function that contains the spatial information we desire from the SHARP parameter maps: an area and a point cloud within the area. In our implementation, we take the PIL region, as the area, and the number of pixels of the PIL region as the size A. Within the PIL region, we select all pixels with unsigned flux $|B_r|$ above a threshold τ. In some active regions, there are many pixels with extremely large $|B_r|$ values, leading to many pixels being selected above the threshold τ. Based on the formula of Ripley's K function, one can see that it involves the pairwise distance matrix of all points. Empirically, we found that having over 500 points in the point cloud would make the computation of Ripley's K very time consuming. Therefore, if there are more than 500 such pixels, we randomly select only 500 of them where the probability of each pixel getting selected is proportional to its unsigned flux value (per unit area defined by the pixel) $|B_r|$. When there are fewer than 500 pixels before the sub-sampling, we also sample with replacement to make the final point clouds having 500 points. The pixels selected are the point clouds and their (x, y) coordinates in the B_r mask specify their locations. The distance between any two pixels is the Euclidean distance between their coordinates. Since we are using the Cylindrical Equal Area (CEA) version of the B_r magnetogram, each pixel has the same physical size, so it makes sense to use the pixel coordinate instead of the true latitude and longitude as the pixel location.

In Figure 6, we show the point clouds generated with threshold being 400G, 1000G and 1600G (on the left, middle, right columns) for the B_r map collected 1 hour prior to an M flare of HARP 377, without any sub-sampling for comparison purpose. The bottom row gives the corresponding Ripley's K functions for all pixel distances within the range $[0, 100]$. As we increase the threshold value τ, the point clouds are more and more clustered into several small clusters, and the Ripley's K function is getting more deviated from a straight line (which corresponds to a random distribution pattern) towards the line pattern indicating a clustering pattern of points at various scales. This is exemplified also in the top row of Figure 6.

Similar to the topological features, the Ripley's K function also locates the clusters of high flux pixels. What sets it apart from the topological features is that it indicates at which scale (in terms of pixel distance) the clusters appear. In other words, the Ripley's K function provides additional information of the size of the loops, whereas topological features only captures the numbers of the loops.

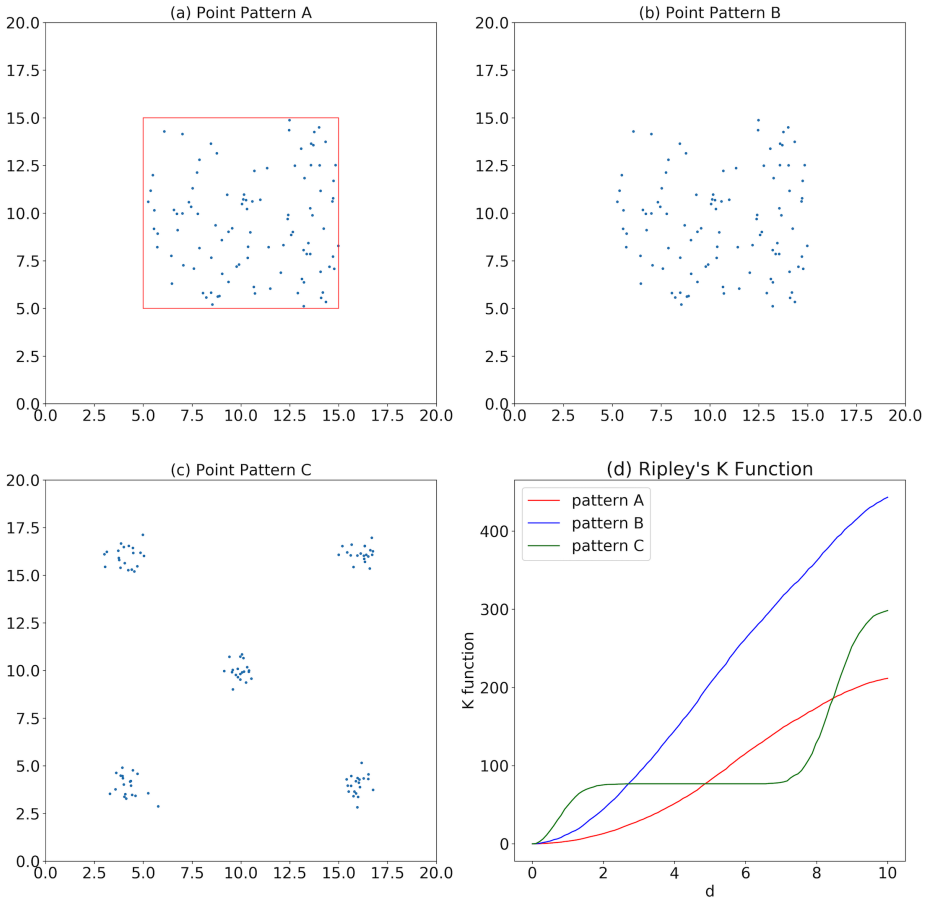

Figure 5. Schematic example of Ripley's K function. (a) a uniformly distributed point cloud in a small area. (b) same point cloud but in a larger area, looking more like clustered in the middle. (c) a non-uniform point cloud. (d) Ripley's K function for all three point patterns. The function is nearly a straight line for both pattern A and B, indicating that points are uniformly randomly distributed in the area. But the different area sizes drastically scale the level of the function. Concave function indicates that points are more clustered than random while convex function indicates that points are more dispersed than random.

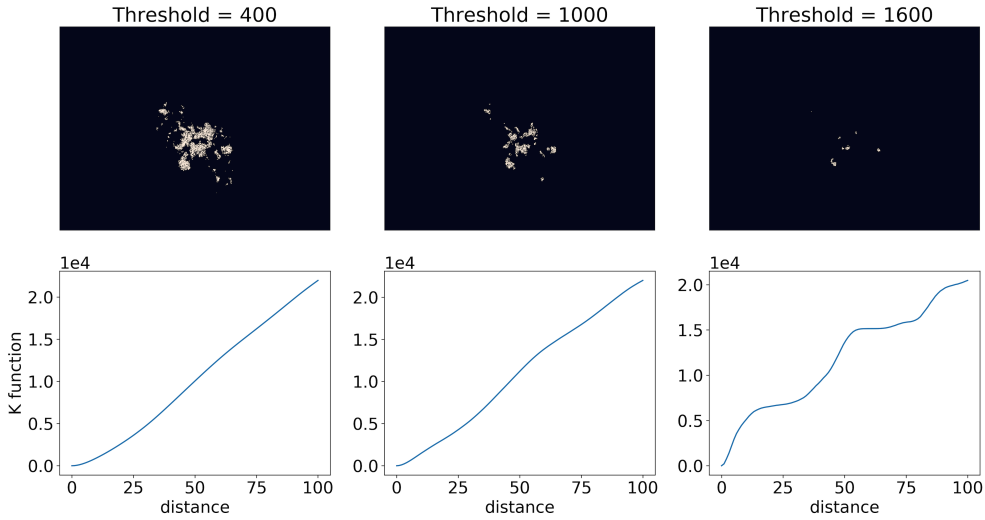

Figure 6. Point cloud and the corresponding Ripley's K function for the B_r mask collected from HARP 377, 1 hour before the M flare peaked at 2011.02.13 17:38. The top row includes 3 point clouds generated by 3 thresholds at 400G, 1000G, 1600G. The bottom row shows the 3 corresponding Ripley's K functions.

In practice, we implement the Ripley's K function calculations for the B_r map as follows. For every B_r map, we use 11 thresholds at 0G, 200G, 400G, 600G, 800G, 1000G, 1200G, 1400G, 1600G, 1800G, 2000G, for generating the point clouds. We choose these thresholds because they can filter high $|B_r|$ value pixels at various magnitudes and there are some considerable differences among the point clouds selected based on these thresholds so that different Ripley's K functions do not have too much overlapping information. For every point cloud, we calculate the Ripley's K function and evaluate the function on the distance grid $\{1, 2, \ldots, 99, 100\}$. We restrict the upper bound of the distance scale below 100 (pixel scale) to guarantee that it works for all B_r images. Additionally, most PIL regions do not exceed $25,000$ pixels, so by analyzing the a neighborhood of radius 100 pixels of every pixel covers sub-region that is large enough within the PIL region while at the same time still remains localized. Thus in total we have 11 Ripley's K functions and each function has 100 functional values, leading to $1,100$ raw features per B_r map. When constructing the final feature, **we divide all Ripley's K function by the number of PIL pixels**, namely A, so finally the level of Ripley's K function is inversely proportional to $\sqrt{A}$.

2.3.2 Variogram

Both the topological features and the Ripley's K function on the B_r maps are aimed at analyzing the clustering patterns of high flux pixels. Here we introduce another spatial statistics quantity, called the variogram, which is a description of the spatial continuity of the data. The variogram of the B_r map is a functional summary of the spatial variation of flux values at various scales. Again, we briefly introduce the definition and interpretations here, and interested readers shall refer to Omre (1984) for details. For any point cloud $(x_i, y_i)_{i=1}^n$, denote their B_r values as $z(x_i, y_i)$. At (Euclidean) distance d, the variogram is defined as

$$\gamma(d) = \gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) = \frac{1}{2}\mathrm{Var}[z(\boldsymbol{s}_i) - z(\boldsymbol{s}_j)], \tag{2}$$

where $\boldsymbol{s}_i = (x_i, y_i), \boldsymbol{s}_j = (x_j, y_j)$ are two arbitrary points in the point cloud that has a Euclidean distance d in-between, and Var denotes the variance of a random variable.

The variogram measures the variance of the differences of B_r values between any two arbitrary pixels that have a distance d in between. The variogram is similar to the structure function, which has been applied to active regions to measure the fragmented nature of the magnetic network and its relationship to flares Uritsky and Davila (2012). In the B_r map, typically one would expect that when two pixels are close, their B_r values should be more similar. When two pixels are getting more distant, their B_r values can be more different as well. The relationship of local similarity and correlation of B_r and the scale d is captured by the variogram, which can provide extra spatial information on top of the topological features and the Ripley's K function.

In practice, it is hard to find a large number of pairs of pixels that are separated exactly by distance d. To estimate the empirical variogram, one would set a few disjoint bins of distance, and any pair of pixels would fall into one of the bins, depending on their distance. The variance is then calculated for all pairs of pixels belonging to the same bin. In our implementation, we use the same point cloud for Ripley's K function calculation for the calculation of variogram. In Figure 7, we show the corresponding variogram for the 3 point clouds shown in 6.

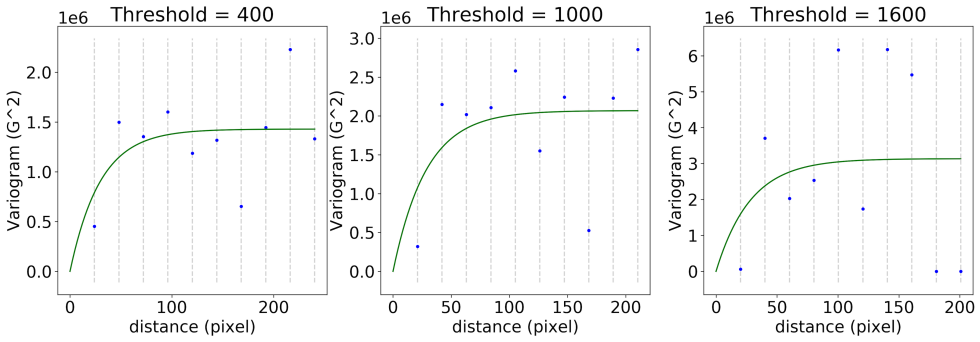

Figure 7. Variogram estimates for point clouds in Figure 6. Vertical dashed line show the center of each distance interval, and the scatter points are the semi-variance (see equation 2) of B_r values for all pairs of pixels separated by the distance within the interval. The blue line is the fitted curve for the variogram estimates. Note that the scales of x,y axes are different across the three graphs with distint thresholds.

In each variogram, there are multiple distance bins whose centers are indicated by the vertical dashed line. In each bin, there are many pairs of pixels whose distance lies in the distance bin. The scatter points are the variogram estimates based on all pairs of pixels in the same bin. The blue line is the fitted curve of the scatter points using an exponential parametric model:

$$\gamma(d|C_0, a) = C_0 \left(1 - e^{-\frac{d}{a}}\right) \tag{3}$$

Since different point clouds may have all sorts of pairs of pixels whose distance may vary in a wide range. Different variograms can have different distance bins, making it hard to vectorize the variogram as a fixed-length feature vector. As a partial solution, we use the fitted curve parameter C_0, a for each variogram as its features. The two parameters C_0, a are termed as the sill and range parameter of the variogram. Since we have

11 point clouds per B_r map, and each variogram has 2 parameters, we have 22 variogram features for a single B_r map.

2.4 Feature Set Summary

In our feature construction, we use tools from TDA and Spatial Statistics to analyze the spatial distribution of SHARP parameters. For every flare, we first derive its nine SHARP parameter maps (Table 1). For each magnetogram, we detect the active region polarity inversion line (PIL) region and focus only on the PIL region for feature construction. Topology features are derived for all nine SHARP parameter maps, and each map has 19 Betti numbers counting the numbers of loops in the 19 bitmaps. Ripley's K function is applied to the B_r map only and every B_r map has 11 Ripley's K functions out of 11 thresholds, and every function is evaluated on a 100-point grid, leading to 1100 Ripley's K features. Variogram is also applied to the B_r map and 11 variograms are calculated based on 11 thresholds, where each variogram only retains its sill and range parameter as features. In addition to all these features, there are some extra features which do not belong to any of the categories above, so we group all of them, including the width and height of the B_r maps, the size of the PIL region and the sum of PIL weights, as the auxiliary features. Table 2 gives a brief summary of all the features used later for flare classification.

Feature Category	Shorthand	Description	# of feature
SHARP	**S**	SHARP parameter in the PIL region	12
Topology	**T**	Betti Number for 9 SHARP masks	171
Spatial **(SP)**	**Ripley_K**	Ripley's K Function for B_r	1100
	V-gram	variogram sill and range parameter for B_r	22
Auxiliary Features	**A**	area of PIL, height/width of the masks, sum of PIL weights	4

Table 2. A brief overview of the feature sets considered for flare classification. We choose USFLUX, MEANGAM, MEANGBT, MEANGBH, MEANGBZ, TOTUSJZ, TOTUSJH, SAVNCPP, MEANSHR, MEANPOT, MEANALP, MEANJZD as the SHARP parameter features. The auxiliary features include the area of PIL (# pixels of PIL region), sum of PIL weights and height/width of PIL mask **(same as the SHARP parameter masks)**. In the result section, we introduce a dimension reduction on the topology features and the Ripley's K function features, and denote the dimension-reduced version of the features, in shorthand, as **T_PC**, **Ripley_K_PC**. And the spatial statistics feature category including the **Ripley_K_PC** and the variogram features as **SP_PC**.

3 Data Description and Prediction Results

In this section, we first introduce the dataset used for feature construction. Then we elaborate the model we use for flare classification and the data pre-processing steps for dimension reductions on the feature sets. Finally, we demonstrate the prediction per-

formances of different combinations of features and show the extra prediction gains of using topological and spatial statistics features on flare classification.

3.1 Description of Data

We use the Geostationary Operational Environmental Satellites (GOES) flare list spanning 2010/12 - 2018/06 for collecting flare events. There are originally 11,348 flares within this time range. For the purpose of flare classification, we only keep all B (weak), and M/X (strong) class flares. For each flare, we collect its corresponding high-resolution HMI magnetogram data from the JSOC at 4 time points: 1, 6, 12, 24 hours prior to the peak flux. These 4 time points lead to 4 datasets, and we perform feature construction and flare classifications separately for each of the 4 datasets. Some of the flares do not have available data for all 4 time points, especially for 12-hour and 24-hour data due to the lack of records at exactly 12 or 24 hours before the flare peak time, and we drop them to make sure that every flare appears in all four datasets, for sake of fair comparison across the four datasets. Finally, we have 399 M/X class flares and 1,972 B class flares coming from 487 HARP regions in each of the 4 datasets.

3.2 Data Pre-processing & Model Training

The prediction task is to classify strong (M/X) from weak (B) flares. For binary classification, we use the XGBoost model (T. Chen & Guestrin, 2016), which is an efficient gradient boosting (Friedman, 2002) method with a decision tree (Safavian & Landgrebe, 1991) base learner. We train the model with several different combinations of features from the feature set detailed in Table 2.

For topological features (**T**) and the Ripley's K function features (**Ripley_K**), we have very high dimensional feature spaces. To efficiently compress the information, we conduct dimension reduction on both sets of features. Specifically, for the 19 topology features of every SHARP map, we do a principal component analysis (PCA) and keep top 5 PCs in place of the original 19 features. For every Ripley's K function, which is itself 100-dimensional, we carry out a functional principal component analysis (FPCA, Hall, Müller, and Wang (2006)) and keep top 5 PCs in place of the original 100 features. After both steps, we reduce the topological features from 171 to 45 dimensions, and the Ripley's K function features from 1100 to 55 dimensions. We denote these dimension reduced features as **T_PC** (for topological features) and **Ripley_K_PC**, and the spatial statistics feature category as **SP_PC**, which includes the Ripley's K principal components and the original variogram parameter estimates.

To construct train and test sets for training and validating the classification model, we split the entire flare list according to their HARP regions. As a result, a flare from any HARP region appears in either train or test set only. On average, 70% of the HARP regions are assigned to the train set and the rest 30% go to the test set. The dimension reduction of topological and the Ripley's K features are performed using the train set only and the test set PCs are predicted using the PCA and FPCA results from the train set. Other features are standardized based on the mean and standard deviation of the train set only to avoid information leaking.

There is randomness in the process of deriving the topological and spatial statistics features because we always randomly pick a point cloud with 500 points from all candidate pixels. To guarantee the robustness of the features derived, we re-run the derivation of all these features 10 times and take the average as the final features. To make the feature comparison more robust to different train/test split, we re-run the model training/testing procedure 20 times, each time with a different train-test split, and then take the average performance for feature comparison. The model is trained separately for the 4 datasets, but the flares in the train/test set, in each iteration, is the same for all 4 datasets.

3.3 Model Evaluation and Discussion

To compare the discriminating power of different sets of features in the flare classification task, we define the M/X flares as the positive class and the B flares as the negative class. We use the test set True Skill Score (TSS), as the benchmark defined as follows:

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

where TP, TN are the positive (M/X), negative (B) samples that are classified correctly, and FP, FN are the negative (B), positive (M/X) samples that are classified wrongly as being positive and negative, respectively. The range of TSS is from -1 to 1, and it is designed such that a random classifier or an unskilled classifier (always predicting the majority class) has TSS 0. In Table 3, we show the mean test-set TSS for different feature sets across 20 iterations.

Feature Combination	1	6	12	24
S	0.496	0.487	0.455	0.390
T	0.487	0.521	**0.507**	**0.473**
SP	0.473	0.482	0.467	**0.459**
S+T	0.520	0.507	**0.495**	**0.491**
S+SP	0.507	**0.508**	**0.472**	**0.451**
S+T+SP	**0.539**	**0.528**	**0.515**	**0.505**
S+T_PC+SP_PC	0.505	0.502	0.483	**0.457**
S+T+SP+A	**0.544**	**0.540**	**0.515**	**0.505**
S+T_PC+SP_PC+A	0.510	0.505	**0.487**	**0.453**

Table 3. Average True Skill Score (TSS), based on 20 train-test split, for classifying strong (M/X) vs. weak (B) flares using different sets of features. Feature shorthand represents: SHARP parameter (**S**); Topology feature (**T**); Spatial Statistics feature (**SP**); Auxiliary feature (**A**); Topology feature principal component score (**T_PC**); Spatial statistics with functional principal component score for Ripley's K function (**SP_PC**). Boldface numbers indicate that the TSS is significantly higher than the benchmark model (**S**), which is fitted only with SHARP parameters in the PIL region, across the 20 iterations. The column names labeled by 1,6,12,24 stands for the 4 datasets collected at 1,6,12,24 hours before the flare peak time.

We define the model fitted with SHARP parameters as the benchmark model, and compare all other feature combinations against it. We show the average TSS differences between some combinations of features against the benchmark. To formally test whether

such TSS differences are significant, we apply a one-sided paired Student's t-test to test the null hypothesis that a certain feature combination has the same TSS as the benchmark for any arbitrary train-test split. In Table 3, we highlight the TSS in boldface when the TSS of a certain feature combination is significantly higher than the benchmark, i.e. we reject the null hypothesis with 95% confidence.

As one can see that when using the topological and spatial statistics features, we can improve the TSS of flare classification 12-24 hours before the flare peak time, and further improves when it is getting close to the flare peak time (i.e. 1,6 hours before). When combining all features together, we can do significantly better than only using the SHARP parameters across all 4 datasets. When we reduce the dimension of the topological and spatial statistics features, we can still maintain a margin of skill improvement in the 12 and 24 hour lead time scenario.

The performance improvement introduced by the new features is marginal but significant based on our significance testing results. Besides measuring the contribution of all new features using the TSS, we also calculate the individual feature importance using the Fisher score (F-score (Stork, Duda, Hart, & Stork, 2001), as adopted by Bobra et al. (2014); Deshmukh, Berger, Bradley, and Meiss (2020)). The Fisher score measures the individual feature's discriminating power in a binary classification setting. For any feature, x_i, the Fisher score is calculated as

$$F(i) = \frac{(\bar{x_i}^+ - \bar{x_i})^2 + (\bar{x_i}^- - \bar{x_i})^2}{\frac{1}{n^+-1}\sum_{k=1}^{n^+}(x_{k,i}^+ - \bar{x_i})^2 + \frac{1}{n^--1}\sum_{k=1}^{n^-}(x_{k,i}^- - \bar{x_i})^2},$$

where $\bar{x_i}^+, \bar{x_i}^-$ are the subgroup mean of x_i for all positive, negative samples; n^+, n^- are the numbers of positive, negative samples; and $x_{k,i}^+, x_{k,i}^-$ are the k-th observation of the positive, negative class. The Fisher score is the ratio of between-class variation and within-class variation. A high Fisher score indicates that the feature shows high separability for the two classes. We rank all features including the SHARP parameters (**S**), topological feature PCs (**T_PC**), Ripley's K PCs (**Ripley_K_PC**), variogram (**V-gram**) and auxiliary features (**A**) based on their F-score, for the 4 datasets separately. In Figure 8, we show the top 15 features in their order of F-score (normalized to range $[0, 1]$), and color them based on their feature categories.

The feature ranking shows that the newly constructed features have equally good or even better individual discriminating power than the SHARP parameters. Among all new features, the Ripley's K function stands out as the features with the highest discriminating power. Specifically, the top features are the 1st PC score of Ripley's K function generated based on different threshold values.

To visualize the discriminating power of Ripley's K function and variogram parameters, we choose two representative cases, one for B flare and one for M flare, and show their B_r, PIL mask, sampled point cloud at threshold 2000G and the corresponding Ripley's K function (divided by the number of PIL pixels) in Figure 9 and Figure 10.

The fundamental difference between the derived Ripley's K function, as one can see by comparing the two plots, is the level of the function value. There are two factors that create the level difference: PIL area size and concentration of high-flux pixels. Since we divide each function by the number of PIL pixels, the function value tends to be small when the PIL area is large. The B flare example has a larger PIL region, as suggested by the PIL mask, but only have small scattered pieces of high-flux regions along the PIL. Both of these facts would lead to a low functional value. The M flare, on the contrary, has a smaller PIL area, but the high-flux regions are close to each other and clustered into relatively bigger chunks. Both of these lead to higher functional values.

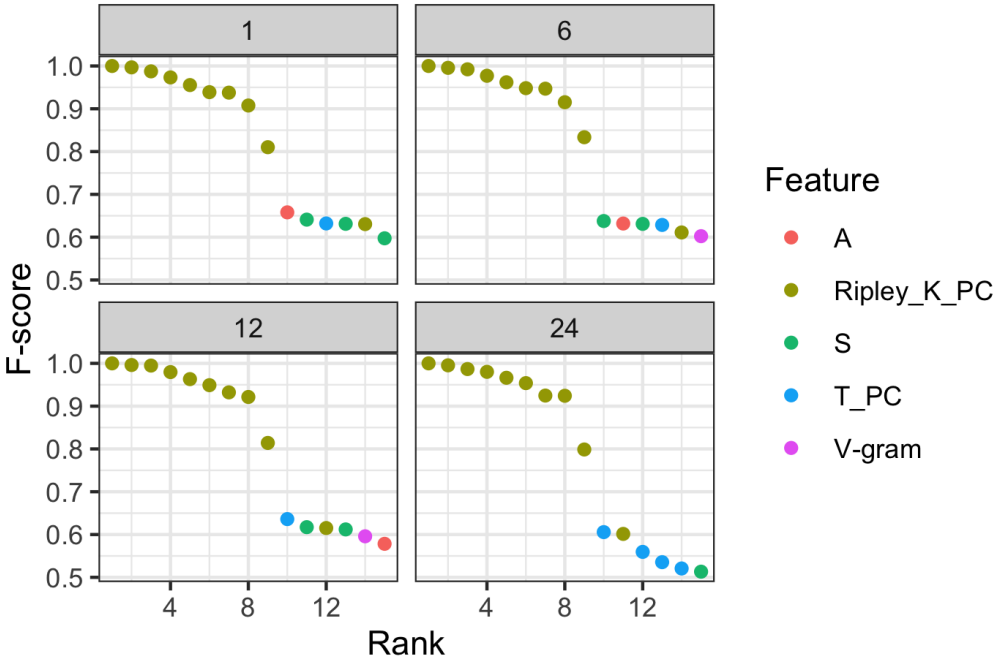

Figure 8. Normalized Fisher Score for selected features. Four panels correspond to the 1,6,12,24 hour dataset. In each panel, top 15 features in terms of F-score are plotted. Among all 4 datasets, the top features are always the Ripley's K function's principal component score. Some features from other categories also rank among top 15 features. The F-score ranking indicates that the spatial statistics features, especially the Ripley's K function, have greater individual discriminating power compared to SHARP parameters.

Apart from the differences of the level of the functions, one may also notice that the shape of the functions differ. The B flare has two jumps in the function while the M flare has a constantly increasing trend prior to the plateau. This exactly corresponds to the two pairs of high-flux regions shown in the lower left panel of the B flare example, suggesting that there are a few pairs of small clusters of high-flux pixels. The function for the M flare has an increasing trend overall, suggesting that there is a region full of high-flux pixels (i.e. there are such pairs of high-flux pixels separated by an arbitrary distance within range $0 \sim 80$).

The variogram effective range parameters also differ significantly between the B and M flare. The variogram of the B flare has range parameter at 1.96 while the M flare has it at 75.95. This suggests that most of the B_r value spatial variations are observed within ~ 2 pixel distances for the B flare point cloud and ~ 76 pixel distances for the M flare point cloud. This further consolidates our argument that the B flare has small, scattered clusters of pixels with high unsigned flux but the M flare has larger concentrations of such pixels.

In conclusion, the spatial statistics features we derived tend to predict an onset of a strong flare when observing the following B_r pattern: 1) small PIL area; 2) big clusters of high-flux pixels not locating far from each other; 3) has large B_r value variation between clusters (e.g. between clusters of strongly positive and strongly negative flux). Such pattern is not exhaustive for all strong flares, but it covers the majority of the cases and lead to features with high discriminating power.

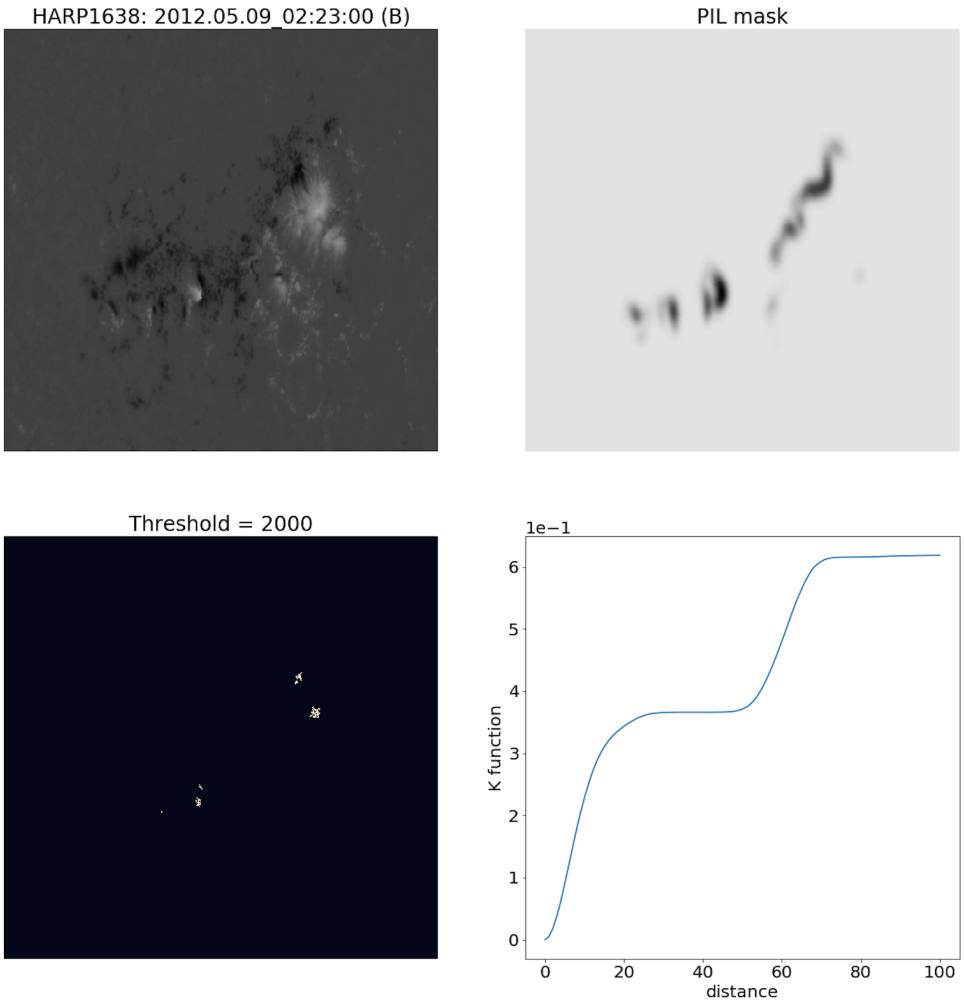

Figure 9. B5.3 flare for HARP 1638 peaked at 02:23, May 09, 2012. Top row shows the B_r and PIL mask, bottom row shows the sampled point cloud at threshold 2000G, and the Ripley's K function. All three subplots, except the Ripley's K function, are of size 559 × 1002. The variogram range parameter is 1.96 and sill parameter is 0.138.

4 Conclusion

In this paper, we investigate new features, on top of the SHARP parameters, for flare classification task. The first set of features is derived from persistence homology in topological data analysis, following the idea in Deshmukh, Berger, Meiss, and Bradley (2020). We extend the scope of HMI images from just the B_r channel to multiple SHARP maps when conducting the analysis, and pay specific attention to the polarity inversion line region (PIL). The second set of features come from spatial statistics concepts. The Ripley's K function analyzes the spatial clustering/dispersion patterns of pixels with strong flux. The Variogram analyzes the spatial variation of the flux at various distance scales. Both sets of features summarize some information regarding the spatial distribution of SHARP parameters, which adds additional information to the feature set that SHARP parameters themselves cannot provide. We demonstrate how the new features can improve skills of the prediction model and also show that new features, especially the Ripley's K functions, have great discriminating power.

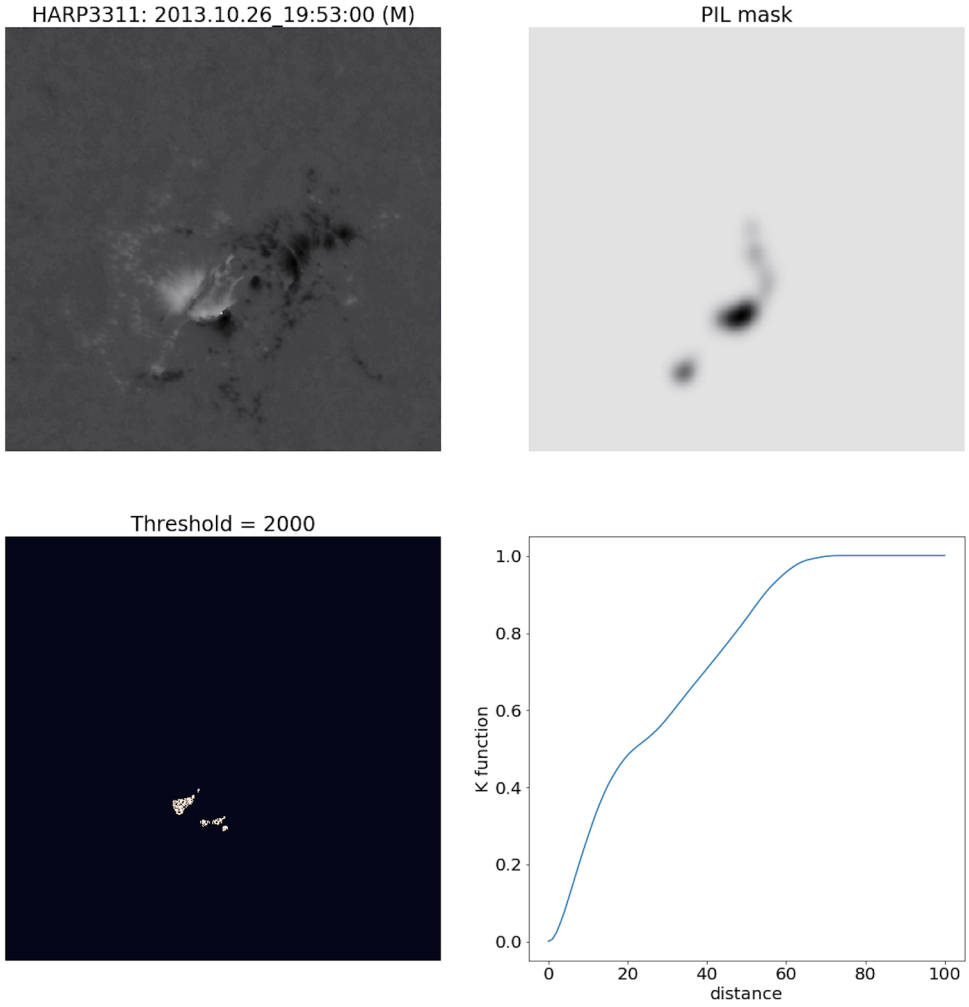

Figure 10. M1.0 flare for HARP 3311 peaked at 19:53, Oct 26, 2013. Top row shows the B_r and PIL mask, bottom row shows the sampled point cloud at threshold 2000G, and the Ripley's K function. All three subplots, except the Ripley's K function, are of size 411 × 546. The variogram range parameter is 75.95 and sill parameter is 0.143.

One major finding, besides the prediction performance gains, is that by focusing on the B_r channel only, one can still derive topological and spatial features that has equal or superior predictive power than the SHARP parameters. We find that the top features for discrimination are the Ripley's K functions based on B_r channel and topological features of the B_r channel. This result suggests that the spatial and shape information we derived from the B_r channel alone is a more powerful discriminant than the SHARP quantities derived from other SHARP maps. One of our future research goal is to analyze the spatial correlations of the B_r channel and the other SHARP quantities and check if there are systematic differences across different classes of solar flares.

We note that out findings of a strong correlation between the B_r spatial distribution and the flare productivity shows an inherent connection between the free energy buildup and release in solar flares that is related to the clustering and proximity of flux to the PIL. Such a connection has been established earlier by Falconer et al. (2003) and Schrijver (2007); Schrijver, De Rosa, Title, and Metcalf (2005) who respectively found the gra-

dient of and proximity of the magnetic flux (line-of-sight component) with respect to the PIL to be strongly correlated with flares and coronal mass ejections. Our work goes much further in establishing this relationship to such a degree that it has greater discriminating power for solar flares than the SHARP parameters derived from the full vector magnetic field.

The analysis of the B_r channel at multiple flux thresholds and multiple spatial scales can also be found in the thread of research on the fractal dimensions of flaring areas (Aschwanden & Aschwanden, 2008; Aschwanden et al., 2016; McAteer, Gallagher, & Ireland, 2005). They found that typically M/X flares have higher 2D fractal dimensions, indicating a larger flaring area with high flux values. The key difference between the spatial statistics tools we adopted and the fractal dimensions is that, after we choose the high flux areas above a threshold, we measure the spatial clustering pattern of the high flux areas, not the size of the area, since we always randomly sample 500 pixels in the high flux areas for analysis. We do believe that our work complements the researches on fractal dimensions to further quantify the shape complexity of high flux areas at multiple scale, and our result suggests that there are more discriminating features out of the shape (i.e. clustering pattern) of the high flux areas, beyond their sizes.

From a theoretical perspective, our work suggest that the necessary components for flares, complexity and also free energy are intrinsically related to one another through the photospheric B_r distribution. The relationship to B_r and coronal free energy is not immediately obvious. However, we may infer that the energized coronal field may result from the evolution of non-force-free fields in the convection zone that produce signatures in the B_r distribution detected by our statistical analysis. Such forces may affect the way individual flux tubes find cohesion from magnetic twist or emerge through the convection zone. A physical basis for this conjecture is found in simulations of flux emergence of kink-unstable flux ropes that form delta spots, which are highly prone to flaring (Fang & Fan, 2015; Linton, Dahlburg, Fisher, & Longcope, 1998; Toriumi & Hotta, 2019). Simulations of flux emergence also show the development of strong shear along forming PILs (Archontis & Török, 2008; Fan, 2001; Fang, Manchester, Abbett, & van der Holst, 2010; Manchester, 2001, 2007; Török et al., 2014), as well as producing eruptive behavior (Manchester, Gombosi, DeZeeuw, & Fan, 2004). Our spatial and topological analysis suggest a more subtle and universal process may be at work developing the photospheric magnetic field and coronal free energy in a relatable way.

With this paper, we also want to popularize the application of spatial statistics tools, which can provide both interpretable and predictive features for various machine learning tasks. The biggest advantage of the Ripley's K function and Variogram in our paper is that they can summarize the spatial patterns of pixels at various scales and give functional summary of the spatial distributions of the physical quantities. Focusing on various spatial scales can potentially uncover the multi-level patterns that are relevant for the initiation of flares. In the future, we want to apply the same or similar tools to analyze not just snapshots of HARP regions, but videos of HARP regions to see if there are temporal trends of such spatial information that can further benefit the flare prediction.

Acknowledgments

All codes and data are included in our Github repository at `https://github.com/husun0822/spat_feature_flare_pred`. This work is supported by NASA grant 80NSSC20K0600, NSF DMS 1811083. W. Manchester is also supported by NASA grant 80NSSC18K1208.

References

Archontis, V., & Török, T. (2008, December). Eruption of magnetic flux ropes dur-

ing flux emergence. *Astronomy and Astrophysics*, *492*, L35-L38. doi: 10.1051/0004-6361:200811131

Aschwanden, M. J., & Aschwanden, P. D. (2008). Solar flare geometries. i. the area fractal dimension. *The Astrophysical Journal*, *674*(1), 530.

Aschwanden, M. J., Crosby, N. B., Dimitropoulou, M., Georgoulis, M. K., Hergarten, S., McAteer, J., ... Uritsky, V. (2016). 25 years of self-organized criticality: solar and astrophysics. *Space Science Reviews*, *198*(1), 47–166.

Barnes, G., Leka, K., Schumer, E., & Della-Rose, D. (2007). Probabilistic forecasting of solar flares from vector magnetogram data. *Space Weather*, *5*(9).

Barnes, G., Leka, K. D., Schrijver, C. J., Colak, T., Qahwaji, R., Ashamari, O., ... Higgins, P. (2016). A comparison of flare forecasting methods. i. results from the "all-clear" workshop. *The Astrophysical Journal*, *829*(2), 89.

Bobra, M. G., & Couvidat, S. (2015). Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. *The Astrophysical Journal*, *798*(2), 135.

Bobra, M. G., Sun, X., Hoeksema, J. T., Turmon, M., Liu, Y., Hayashi, K., ... Leka, K. D. (2014, Sep 01). The helioseismic and magnetic imager (hmi) vector magnetic field pipeline: Sharps – space-weather hmi active region patches. *Solar Physics*, *289*(9), 3549–3578.

Camporeale, E. (2019, jul). The challenge of machine learning in space weather nowcasting and forecasting. *Space Weather*, *17*. doi: 10.1029/2018sw002061

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Chen, Y., Manchester, W. B., Hero, A. O., Toth, G., DuFumier, B., Zhou, T., ... Gombosi, T. I. (2019). Identifying solar flare precursors using time series of sdo/hmi images and sharp parameters. *Space Weather*, *17*(10), 1404–1426.

Cressie, N., & Hawkins, D. M. (1980). Robust estimation of the variogram: I. *Journal of the International Association for Mathematical Geology*, *12*(2), 115–125.

Deshmukh, V., Berger, T., Meiss, J., & Bradley, E. (2020). Shape-based feature engineering for solar flare prediction. *arXiv preprint arXiv:2012.14405*.

Deshmukh, V., Berger, T. E., Bradley, E., & Meiss, J. D. (2020). Leveraging the mathematics of shape for solar magnetic eruption prediction. *Journal of Space Weather and Space Climate*, *10*, 13.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, pp. 226–231).

Falconer, D. A. (2001, November). A prospective method for predicting coronal mass ejections from vector magnetograms. *J. Geophys. Res.*, *106*, 25185-25190. doi: 10.1029/2000JA004005

Falconer, D. A., Moore, R. L., & Gary, G. A. (2002, April). Correlation of the Coronal Mass Ejection Productivity of Solar Active Regions with Measures of Their Global Nonpotentiality from Vector Magnetograms: Baseline Results. *Astrophys. J.*, *569*, 1016-1025. doi: 10.1086/339161

Falconer, D. A., Moore, R. L., & Gary, G. A. (2003, October). A measure from line-of-sight magnetograms for prediction of coronal mass ejections. *J. Geophys. Res.*, *108*, 1380. doi: 10.1029/2003JA010030

Falconer, D. A., Moore, R. L., & Gary, G. A. (2006, June). Magnetic Causes of Solar Coronal Mass Ejections: Dominance of the Free Magnetic Energy over the Magnetic Twist Alone. *Astrophys. J.*, *644*, 1258-1272. doi: 10.1086/503699

Fan, Y. (2001, June). The Emergence of a Twisted Ω-Tube into the Solar Atmosphere. *Astrophys. J. Lett.*, *554*, L111-L114. doi: 10.1086/320935

Fang, F., & Fan, Y. (2015, June). δ-Sunspot Formation in Simulation of Active-region-scale Flux Emergence. *Astrophys. J.*, *806*(1), 79. doi: 10.1088/0004

-637X/806/1/79

Fang, F., Manchester, W., Abbett, W. P., & van der Holst, B. (2010, May). Simulation of Flux Emergence from the Convection Zone to the Corona. *Astrophys. J.*, *714*, 1649-1657. doi: 10.1088/0004-637X/714/2/1649

Florios, K., Kontogiannis, I., Park, S.-H., Guerra, J. A., Benvenuto, F., Bloomfield, D. S., & Georgoulis, M. K. (2018). Forecasting solar flares using magnetogram-based predictors and machine learning. *Solar Physics*, *293*(2), 28. doi: doi:10.1007/s11207-018-1250-4

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, *38*(4), 367–378.

Gelfand, A. E., Diggle, P., Guttorp, P., & Fuentes, M. (2010). *Handbook of spatial statistics*. CRC press.

Hall, P., Müller, H.-G., & Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The annals of statistics*, 1493–1517.

Jiao, Z., Sun, H., Wang, X., Manchester, W., Gombosi, T., Hero, A., & Chen, Y. (2020). Solar flare intensity prediction with machine learning models. *Space Weather*, *18*(7), e2020SW002440.

Leka, K. D., & Barnes, G. (2003a, October). Photospheric Magnetic Field Properties of Flaring versus Flare-quiet Active Regions. I. Data, General Approach, and Sample Results. *Astrophys. J.*, *595*(2), 1277-1295. doi: 10.1086/377511

Leka, K. D., & Barnes, G. (2003b, October). Photospheric Magnetic Field Properties of Flaring versus Flare-quiet Active Regions. II. Discriminant Analysis. *Astrophys. J.*, *595*, 1296-1306. doi: 10.1086/377512

Leka, K. D., & Barnes, G. (2018). Solar flare forecasting: Present methods and challenges. In N. Buzulukova (Ed.), *Extreme events in geospace* (pp. 65 – 98). Elsevier. doi: 10.1016/B978-0-12-812700-1.00003-0

Leka, K. D., Park, S.-H., Kusano, K., Andries, J., Barnes, G., Bingham, S., ... Terkildsen, M. (2019a, aug). A comparison of flare forecasting methods. II. benchmarks, metrics, and performance results for operational solar flare forecasting systems. *The Astrophysical Journal Supplement Series*, *243*(2), 36. Retrieved from `https://doi.org/10.3847%2F1538-4365%2Fab2e12` doi: 10.3847/1538-4365/ab2e12

Leka, K. D., Park, S.-H., Kusano, K., Andries, J., Barnes, G., Bingham, S., ... Terkildsen, M. (2019b, aug). A comparison of flare forecasting methods. III. systematic behaviors of operational solar flare forecasting systems. *The Astrophysical Journal*, *881*(2), 101. Retrieved from `https://doi.org/10.3847%2F1538-4357%2Fab2e11` doi: 10.3847/1538-4357/ab2e11

Linton, M. G., Dahlburg, R. B., Fisher, G. H., & Longcope, D. W. (1998, November). Nonlinear Evolution of Kink-unstable Magnetic Flux Tubes and Solar δ-Spot Active Regions. *Astrophys. J.*, *507*(1), 404-416. doi: 10.1086/306299

Liu, C., Deng, N., Wang, J. T. L., & Wang, H. (2017, July). Predicting Solar Flares Using SDO/HMI Vector Magnetic Data Products and the Random Forest Algorithm. *Astrophys. J.*, *843*(2), 104. doi: 10.3847/1538-4357/aa789b

Liu, H., Liu, C., Wang, J. T. L., & Wang, H. (2019, jun). Predicting solar flares using a long short-term memory network. *The Astrophysical Journal*, *877*(2), 121. Retrieved from `https://doi.org/10.3847%2F1538-4357%2Fab1b3c` doi: 10.3847/1538-4357/ab1b3c

Manchester, W., IV. (2001, January). The Role of Nonlinear Alfvén Waves in Shear Formation during Solar Magnetic Flux Emergence. *Astrophys. J.*, *547*, 503-519. doi: 10.1086/318342

Manchester, W., IV. (2007, September). Solar Atmospheric Dynamic Coupling Due to Shear Motions Driven by the Lorentz Force. *Astrophys. J.*, *666*, 532-540. doi: 10.1086/520493

Manchester, W., IV, Gombosi, T., DeZeeuw, D., & Fan, Y. (2004, July). Eruption of

a Buoyantly Emerging Magnetic Flux Rope. *Astrophys. J.*, *610*, 588-596. doi: 10.1086/421516

McAteer, R. J., Gallagher, P. T., & Ireland, J. (2005). Statistics of active region complexity: A large-scale fractal dimension survey. *The Astrophysical Journal*, *631*(1), 628.

Munch, E. (2017). A user's guide to topological data analysis. *Journal of Learning Analytics*, *4*(2), 47–61.

Muranushi, Y. H., Muranushi, T., Asai, A., Okanohara, D., Raymond, R., Watanabe, G., ... Shibata, K. (2016). A deep-learning approach for operation of an automated realtime flare forecast. *CoRR*, *abs/1606.01587*. Retrieved from `http://arxiv.org/abs/1606.01587`

Nishizuka, N., Kubo, Y., Sugiura, K., Den, M., & Ishii, M. (2021). Operational solar flare prediction model using deep flare net. *Earth, Planets and Space*, *73*(1), 1–12.

Nishizuka, N., Sugiura, K., Kubo, Y., Den, M., & Ishii, M. (2018). Deep flare net (defn) model for solar flare prediction. *The Astrophysical Journal*, *858*(2), 113.

Oliver, M. A., & Webster, R. (2015). *Basic steps in geostatistics: the variogram and kriging*. Springer.

Omre, H. (1984). The variogram and its estimation. In *Geostatistics for natural resources characterization* (pp. 107–125). Springer.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of applied probability*, *13*(2), 255–266.

Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, *21*(3), 660–674.

Schrijver, C. J. (2007, jan). A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting. *Astrophysical Journal*, *655*(2), L117–L120.

Schrijver, C. J., De Rosa, M. L., Title, A. M., & Metcalf, T. R. (2005, July). The Nonpotentiality of Active-Region Coronae and the Dynamics of the Photospheric Magnetic Field. *Astrophys. J.*, *628*(1), 501-513. doi: 10.1086/430733

Stork, D. G., Duda, R. O., Hart, P. E., & Stork, D. (2001). Pattern classification. *A Wiley-Interscience Publication*.

Toriumi, S., & Hotta, H. (2019, November). Spontaneous Generation of δ-sunspots in Convective Magnetohydrodynamic Simulation of Magnetic Flux Emergence. *Astrophys. J. Lett.*, *886*(1), L21. doi: 10.3847/2041-8213/ab55e7

Török, T., Leake, J. E., Titov, V. S., Archontis, V., Mikić, Z., Linton, M. G., ... Kliem, B. (2014, February). Distribution of Electric Currents in Solar Active Regions. *Astrophys. J. Lett.*, *782*, L10. doi: 10.1088/2041-8205/782/1/L10

Uritsky, V. M., & Davila, J. M. (2012, March). Multiscale Dynamics of Solar Magnetic Structures. *Astrophys. J.*, *748*(1), 60. doi: 10.1088/0004-637X/748/1/60

Wang, J., Liu, S., Ao, X., Zhang, Y., Wang, T., & Liu, Y. (2019, oct). Parameters derived from the SDO/HMI vector magnetic field data: Potential to improve machine-learning-based solar flare prediction models. *Astrophysical Journal*, *884*(2), 175.

Wang, J., Zhang, Y., Webber, S. A. H., Liu, S., Meng, X., & Wang, T. (2020). Solar flare predictive features derived from polarity inversion line masks in active regions using an unsupervised machine learning algorithm. *The Astrophysical Journal*, *892*(2), 140.

Wasserman, L. (2018). Topological data analysis. *Annual Review of Statistics and Its Application*, *5*, 501–532.