Space Weather

RESEARCH ARTICLE

10.1029/2021SW002837

**Key Points:**

- We adopt a polarity inversion line (PIL) detecting algorithm to obtain PIL masks for the $B_r$ component 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

**Correspondence to:**

Y. Chen,
ychenang@umich.edu

**Citation:**

Sun, H., Manchester, W. IV, & Chen, Y. (2021). Improved and interpretable solar flare predictions with spatial and topological features of the polarity inversion line masked magnetograms. *Space Weather*, *19*, e2021SW002837. https://doi.org/10.1029/2021SW002837

Received 8 JUL 2021
Accepted 15 NOV 2021

# Improved and Interpretable Solar Flare Predictions With Spatial and Topological Features of the Polarity Inversion Line Masked Magnetograms

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

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

**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 quantities across active regions. 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 those using SHARP parameters (also along the PIL). We found that using the new features can improve the skill scores 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 magnetic field component, $B_r$, instead of all SHARP parameters, one can still derive strongly predictive features for flare classification.

**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 magnetograms, beyond those commonly adopted, aggregated quantities, can be helpful to improve flare predictability.

## 1. Introduction

Solar flares occur in regions of strong magnetic fields possessing concentrated free energy, which are 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 and well-defined polarity inversion lines (PILs), and complex flux distributions (e.g., Barnes et al., 2007; Falconer et al., 2002, 2003, 2006; Schrijver, 2007). Vector magnetogram observations reveal that 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) to more extreme cases of flare production in delta spot active regions, which contain both polarities within single penumbra of uniform polarity. A well-known example is active region numbered 10,486, 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. Similar 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 magnetic flux and the length of the PIL, which have proven useful for flare prediction. In the case of Falconer (2001), 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), where HMI is the Helioseismic and Magnetic Imager for the Solar Dynamics Observatory (SDO), which we refer to as HMI/SDO hereafter. 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; C. Liu et al., 2017; Nishizuka et al., 2018; J. 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 build-up of energy necessary for flare occurrences such as the active region (AR) size as denoted by unsigned magnetic flux and total electric current, current helicity or total free energy itself. Numerous studies illustrates (e.g., Barnes et al., 2016; Leka & Barnes, 2003b; Leka et al., 2019a, 2019b; 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 much 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 HMI/SDO vector magnetograms 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; H. Liu et al., 2019; Muranushi et al., 2016; Nishizuka et al., 2018, 2021). The success of using the SHARP parameters in solar flare forecasting showcases the importance of these parameters to the triggered solar eruptions. Despite this success, the SHARP parameters may not capture the full information of the HMI/SDO magnetogram images, which is relevant for flare forecasting. To directly analyze the magnetograms, there are efforts using deep neural network methods, which directly employ HMI/SDO magnetogram images to predict solar eruptions (e.g., the Long Short Term Memory network adopted by Y. Chen et al. [2019] and H. Liu et al. [2019]). The neural network models, however, suffer from poor interpretability, making it difficult to learn new insights of the underlying physics/mechanism of flare eruptions from those models. In an effort to obtain an 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 was shown that features derived from a deep neural network (autoencoder) can give similar performance to the SHARP parameters when adopted to solar flare predictions. However, one cannot conclude that the SHARP parameters contain all useful magnetogram information. Flare predictions for general-purpose image/video feature reduction techniques typically result in high-dimensional, highly correlated features, many of which are redundant for flare predictions. To alleviate this issue, efficient dimension reduction and feature selection techniques are adopted in Y. Chen et al. (2019). However, the amount of random noise 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 magnetograms, we can achieve similar flare prediction performance when using only the SHARP parameters. This possibility makes it promising to further investigate alternative feature construction methods that result in information complementary to those already contained in the SHARP parameters.

More recently, some promising results have already been shown along this direction: in Deshmukh, Berger, Meiss, and Bradley (2021), topological features are derived for the $B_r$ component, which are shown to improve the performance of flaring active region predictions on top of SHARP parameters. More precisely, they count the number of "loops" formed by high-$B_r$ pixels from the $B_r$ component as a flare predictor, a.k.a. the topological features referred therein.

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 images and the distribution of SHARP quantities within active regions. 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 calculated from the three-component, two-dimensional HMI magnetic field data of active regions. 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 some information of the spatial distribution is the SHARP parameters weighted by the polarity inversion line (PIL) mask, as given by J. Wang et al. (2020). We refer to these features as the PIL-weighted SHARP parameters in the following text. 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 smaller 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, but merely be selective to information based on position. In this section, we introduce two sets of features constructed with the aim of augmenting the SHARP parameters by considering local spatial distribution of the magnetic field components around the polarity inversion line: one based on persistence homology in topology and the other is based on concepts from the field of spatial statistics. We will apply these newly constructed features to flare classification and develop their utility in context with existing literature on related topics, for example, the weighted SHARP parameters by J. Wang et al. (2020) and the topological data analysis enabled feature construction by Deshmukh, Berger, Bradley, and Meiss (2020).

### 2.1. Motivating Example and PIL Mask Construction

To motivate our feature construction machinery, we start with a concrete example: comparing a strong and a weak flare that occurred in HARP region 377. In Figure 1, we show two rows of images with HMI magnetograms of HARP region 377, and two derived quantities. The top row is for an M class flare and the bottom for a B class flare. All images are collected 1 hr 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; J. Wang et al., 2019). We also focus on the PIL regions when distinguishing strong and weak flares.

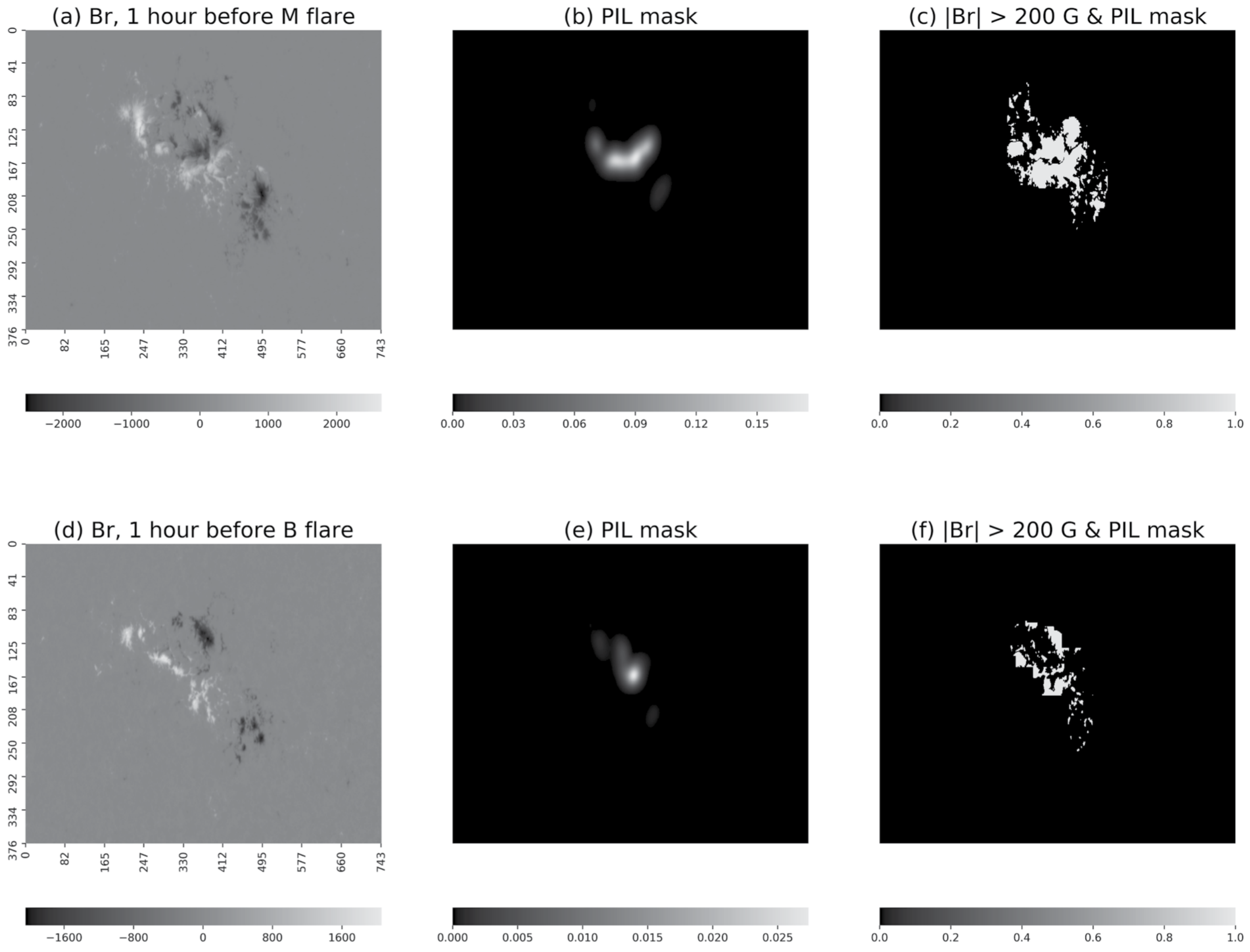

**Figure 1.** Two HMI magnetograms and derived images of HARP 377. The two rows consist of three related images: (left) the Br component, (middle) the PIL weight mask, and (right) bitmap labeling pixels with absolute values of $B_r$ over 200 G, which locate the PIL region. The top row is collected 1 hr before the M-class flare peaked at 2011.02.13 17:38:00 while the bottom row is collected 1 hr 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. Panels (c and f) appear to have a larger mask region when compared to the PIL mask shown in panels (b and e), which is due to the fact that many pixels that are also part of the PIL mask but have near-zero PIL weights, leading to their invisibility in the PIL masks.

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 derive the positive and negative masks using a density-based clustering algorithm (Ester et al., 1996). Finally, we 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 J. Wang et al. (2019).

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 $B_r$ values. But if one focuses on the spatial distributions of regions of high $B_r$ values, 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 to 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 (J. Wang et al., 2020) that can potentially create noises for the derived features.

**Table 1**
*SHARP Parameter Maps, Where the Formulas in the Middle Column Are Applied to Every Pixel of the HMI Magnetogram*

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

*Note*. 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).

Besides the $B_r$ component, we also derive eight additional 2D SHARP quantity maps as follows. We still 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. And we refer to these maps as **SHARP parameter maps** hereafter. Table 1 gives a summary of all nine SHARP parameter maps we collect/construct. We only include these nine SHARP parameter maps because they: (a) can be localized to every pixel; (b) can be thresholded with easy interpretations; (c) cover most of the "important features" based on many machine learning studies.

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 PIL mask. For the topological features alone, our work using all nine SHARP parameter 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 PIL-weighted SHARP parameters as a benchmark feature set, which has been shown to improve flare predictions as compared to unweighted SHARP parameters (J. Wang et al., 2020).

We formulate the machine learning task in this study as strong (M/X) versus weak (B) flare binary classification. We exclude C-class flare samples to create a strong contrast between the positive and negative classes, thus making the post-hoc interpretation more meaningful. We also exclude the non-flaring samples as it is hard to decide the amount of such "quiet time" samples in the training and testing set. An excessive amount of non-flaring samples would boost the final classification metrics. Also, including these samples would make the model interpretation harder because the signals picked up by the features can be a hybrid of discriminators against quiet times and against weak flares. We limit our sample sets to M/X and B flares only to make the interpretation straightforward. In Appendix A, we give results when C-class flares are also included as an extension.

### 2.2. Topological Features

The derivation of our topological features is based on the topological data analysis (TDA) in Deshmukh, Berger, Bradley, and Meiss (2020) and Deshmukh, Berger, Meiss, and Bradley (2021). They take the $B_r$ component from both flaring ARs and non-flaring ARs, apply a series of $B_r$ threshold values to every pixel and count the number of "loops" formed by the high-flux pixels above the thresholds. Loops are defined under the notion of cubical complex, which will be developed shortly. The number of loops under each threshold is used as a flaring region predictor.

Compared to the previous works, we have these extensions:

1. We expand the image set to be analyzed from just the $B_r$ mask to multiple other SHARP parameters maps.
2. We restrict our TDA application to the PIL region only to remove noises in feature construction.
3. 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

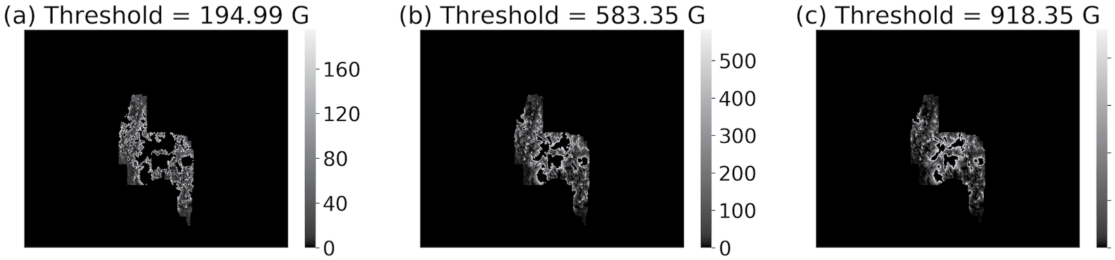

**Figure 2.** $B_r$ component of HARP 377, 2011.02.13 17:38:00. Panels (a–c) only keep pixels whose $|B_r|$ value is below the threshold (194.99, 583.35, 918.35 G) and within the PIL region. The three thresholds are the 30-th, 75-th, and 90-th percentile of the $B_r$ values, at pixel-level, across all data.

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, … dimension. In 2D, $\beta_0, \beta_1$ represents the number of connected components and loops. These loops are closed curves formed at similar levels of $B_r$ attained in the magnetogram rather than the more familiar coronal loops observed in the extreme ultraviolet. In our analysis, we just focus on the $\beta_1$ Betti number, counting the number of loops within the 2D SHARP parameter maps.

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 $B_r$ component 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.

One can notice that around the center of each image, there are several closed loops coming into existence when one increases the threshold from 194.99 to 583.35 G; and that some of the loops disappear while others remain but shrink in size when the threshold level is further increased to 918.35 G. All these thresholds are the 30-th, 75-th, and 90-th percentile of the pixel-level $B_r$ across all data. We will elaborate on how to choose these thresholds empirically shortly. Intuitively, these loops are regions of high $|B_r|$ values within the PIL region, or in other words, they surround 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 $\tau$ as shown in Figure 2, we generate a corresponding bitmap. Each bitmap sets all pixels that are: (a) below the threshold $\tau$, (b) 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. Figures 3a and 3b 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 Figure 3a, 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 Figure 3b), some remain but shrink in size (loop 1 in both Figures 3a and 3b) and some disappear (loop 3 in Figure 3a).

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* (Ghrist, 2008; Kaczynski et al., 2004; Munch, 2017; Zomorodian, 2012). 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° line in the persistence diagram. In Figure 4, we show the persistence diagrams for all loops of the $B_r$ component of three flares coming from HARP region 377. One can see that compared to the two M flares, the persistence diagram of the B flare has loops that are way off the 45° line, indicating that there are fewer intense $B_r$ regions prior to the B flare.

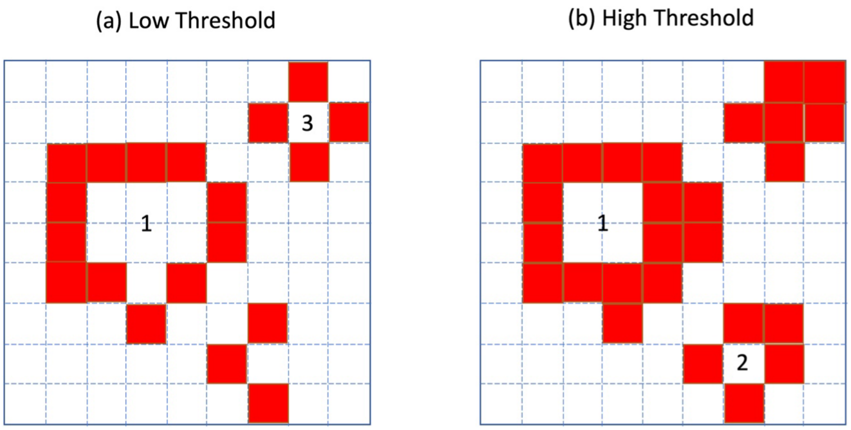

**Figure 3.** Example of loops formed by cubical complexes. In a cubical complex, two pixels are connected if they share a vertex or an edge. Panel (a) at a low threshold, there are two loops formed, labeled as 1 and 3. Panel (b) at a higher threshold, more pixels are selected (marked red), and loop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

In order to extract features, namely the Betti number $\beta_1$, 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%, …, 90%, 95% across all pixels of our data set. For the $B_r$, we do not distinguish between the positive and negative pixels but take the absolute value before calculating the Betti numbers. The percentiles for $|B_r|$ are: 19.4, 42.8, …, 918.3, 1,165.8 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 (Maria et al., 2014). 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 summarize the shape information of the spatial distribution of the corresponding SHARP parameter maps. Using the Betti number $\beta_1$ 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 $B_r$ pixels and the spatial variation of a SHARP quantity, that cannot be summarized 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 this 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).

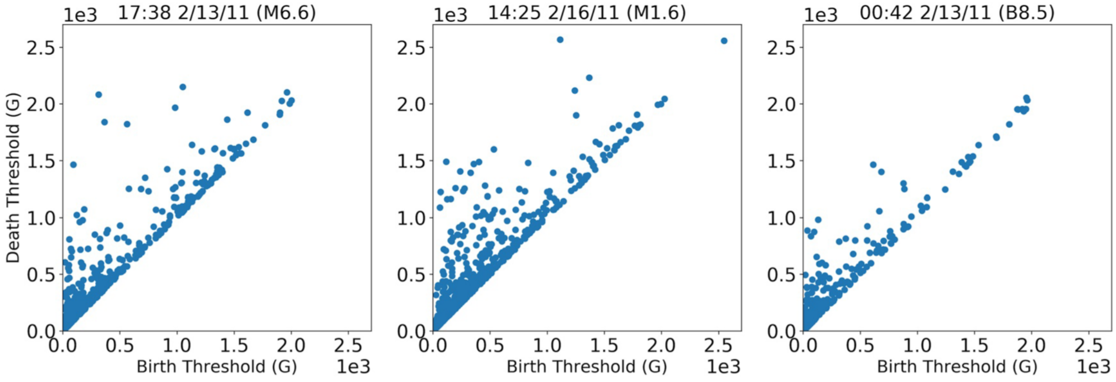

**Figure 4.** Examples of persistence diagrams for two M flares and a B flare of HARP region 377. Every subplot has the flare peak time, flare class and flare intensity in the title. One can see that compared to the M flares, the B flare has less loops that are persistent, indicating that there are fewer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

### 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 Figures 5a and 5b we show two identical point clouds, generated from a 2D uniform distribution, located in areas of different sizes. In Figure 5a, the red bounding box shows the area considered when calculating the Ripley's K function for the point cloud. In Figure 5b, 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 Figure 5d, are very different. The Ripley's K function for pattern A and B are relatively straight, meaning that points are nearly uniformly and randomly distributed in the area. When all points are randomly distributed, the proportion of points within radius $d$ grows linearly with the size the circle $\pi d^2$, so $L(d) \sim O(d)$ and the slope is related to $\sqrt{A}$. In Figure 5c, we show another point pattern coming from a Gaussian mixture distribution. Its corresponding Ripley's K function is plotted in Figure 5d 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 a 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 $|B_r|$ above a threshold $\tau$. In some active regions, there are many pixels with extremely large $|B_r|$ values, leading to many pixels being selected above the threshold $\tau$.

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 $|B_r|$. 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 $B_r$ image 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 thresholds being 400, 1,000, and 1,600 G (on the left, middle, right columns) for the $B_r$ component collected 1 hr 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 $\tau$, the point clouds are more and more clustered into several small clusters, and the Ripley's K function deviates more from a straight line (straight line corresponds to a random distribution pattern) toward 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 $B_r$ 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.

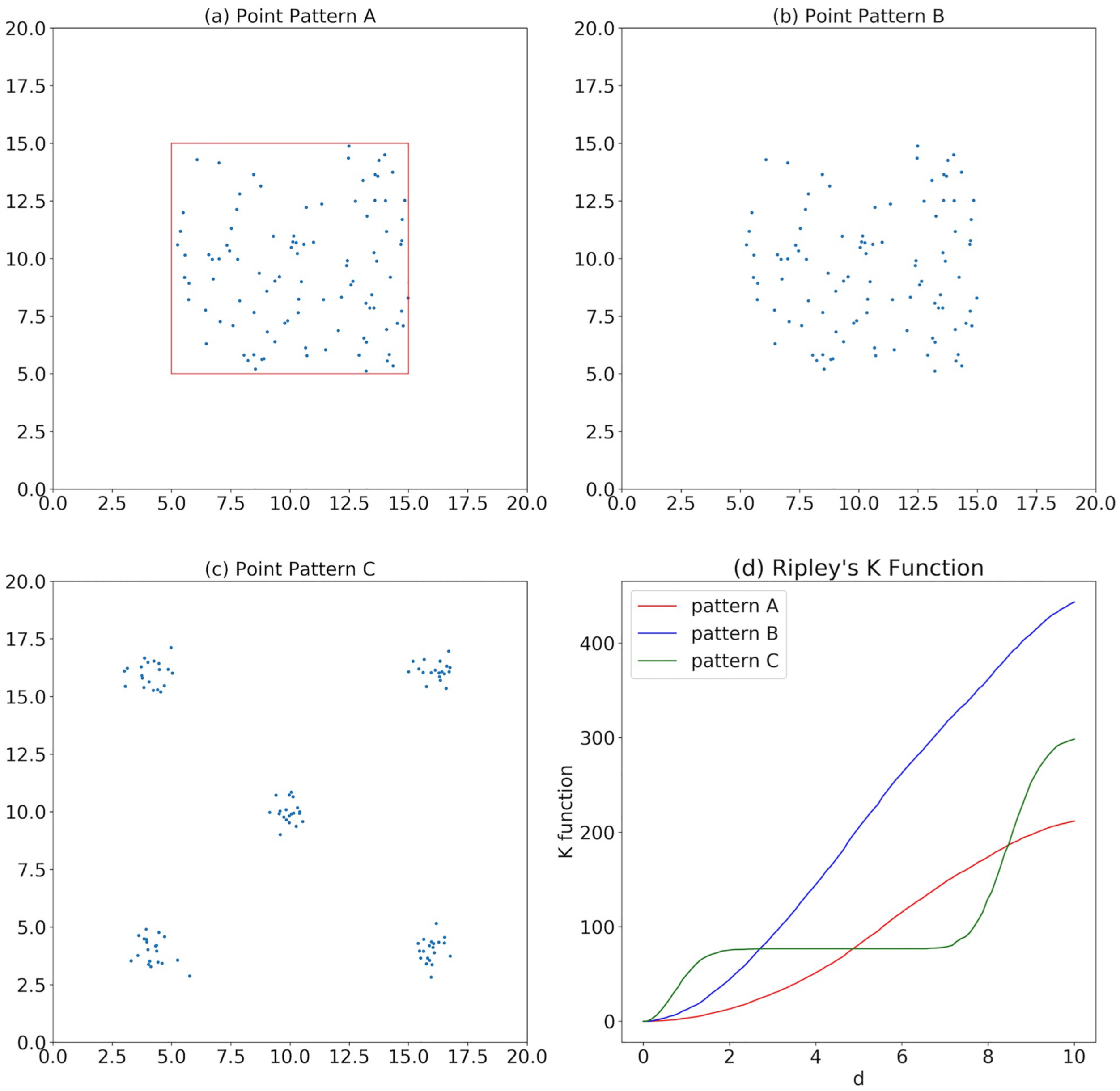

**Figure 5.** Schematic examples of Ripley's K function. (a) A uniformly distributed point cloud in a small area. (b) Same point cloud but in a larger area, looking more like clustered in the middle. (c) A non-uniform point cloud. (d) Ripley's K function for all three point patterns. The function is 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.

In practice, we implement the Ripley's K function calculations for the $B_r$ component as follows. For every $B_r$ map, we use 11 thresholds at 0, 200, 400, 600, 800, 1,000, 1,200, 1,400, 1,600, 1,800, 2,000 G, for generating the point clouds. We choose these thresholds because they can filter high $|B_r|$ value pixels at various magnitudes. 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 at 100 (pixel scale) to guarantee that it works for all $B_r$ images. Additionally, most PIL regions do not exceed 25,000 pixels, so analyzing a neighborhood of radius 100 for every pixel covers a sub-region large enough to within the PIL region for a meaningful sample while still remaining localized. Thus in total we have 11 Ripley's K functions and each function has 100 functional values,

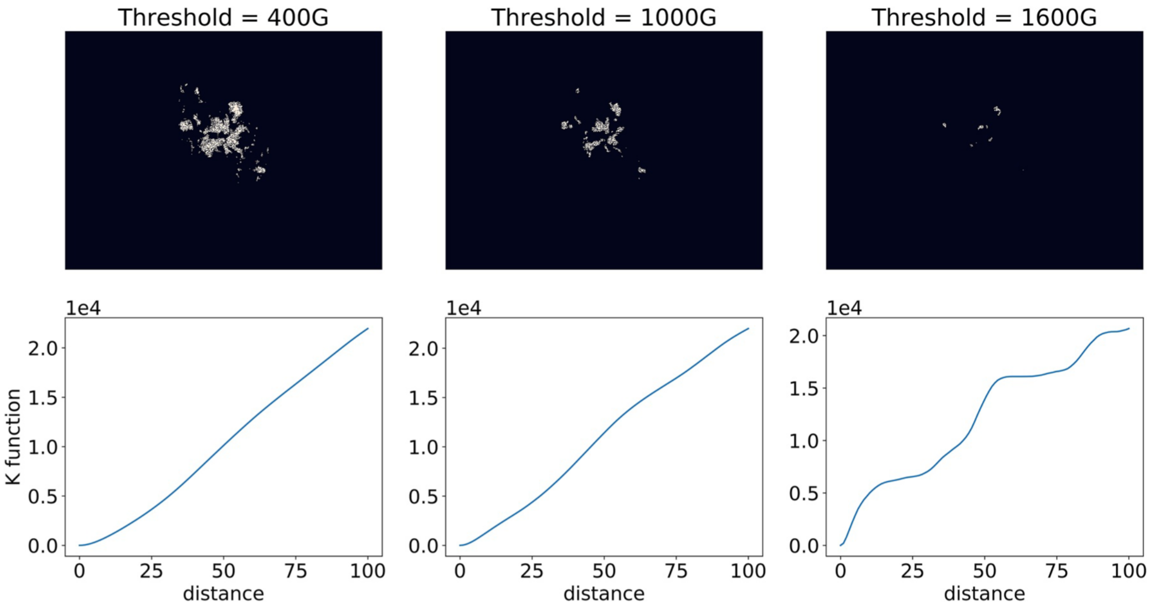

**Figure 6.** Point cloud and the corresponding Ripley's K function for the $B_r$ component collected from HARP 377, 1 hr before the M flare peaked at 2011.02.13 17:38:00. The top row includes three point clouds generated by three thresholds at 400, 1,000, 1,600 G. The bottom row shows the three corresponding Ripley's K functions.

leading to 1, 100 raw features per $B_r$ map. When constructing the final feature, **we divide all Ripley's K functions 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 $|B_r|$ 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 $B_r$ 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 & 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 of the variogram, we use the same point cloud for Ripley's K function calculation. In Figure 7, we show the corresponding variogram for the 3 point clouds shown in Figure 6.

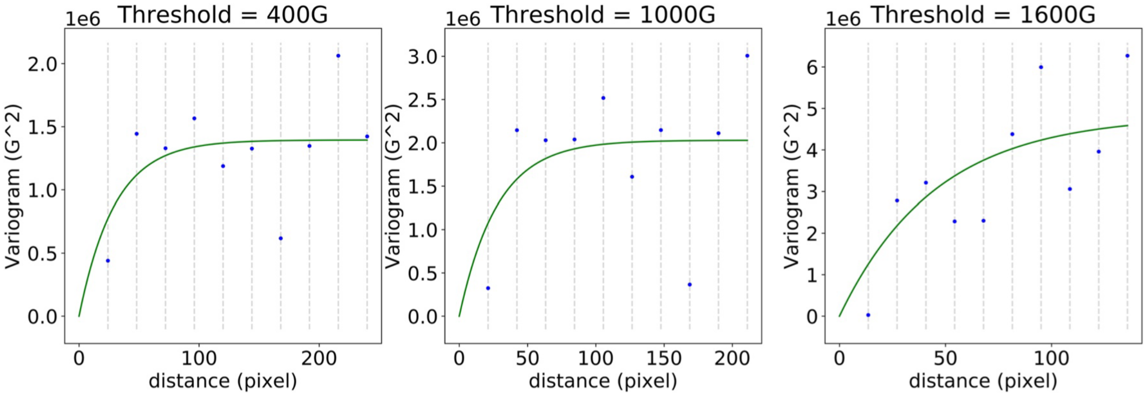

**Figure 7.** Variogram estimates for point clouds in Figure 6. Vertical dashed lines show the center of each distance interval, and the scatter points are the semi-variance (see Equation 2) of $B_r$ values for all pairs of pixels separated by the distance within the interval. The green line is the fitted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

In each variogram, there are multiple distance bins whose centers are indicated by a 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:

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

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 fit the exponential parametric model above with least squares and take the fitted 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 two 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 1,100 Ripley's K features. Variogram is also applied only 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 that do not belong to any of the categories above, so we group all of them, including the width and height of the $B_r$ map, the size of the PIL region and the sum of PIL weights, as the auxiliary features. Figure 8 shows a workflow summary of all features constructed. The topology features and Ripley's K functions have high dimensions as shown in Figure 8, so we also use principal component analysis (PCA) and functional principal component analysis (FPCA) to reduce their dimensions. And we use the reduced dimensions as features for classification tasks. See the next section for details of the data preparation for prediction models.

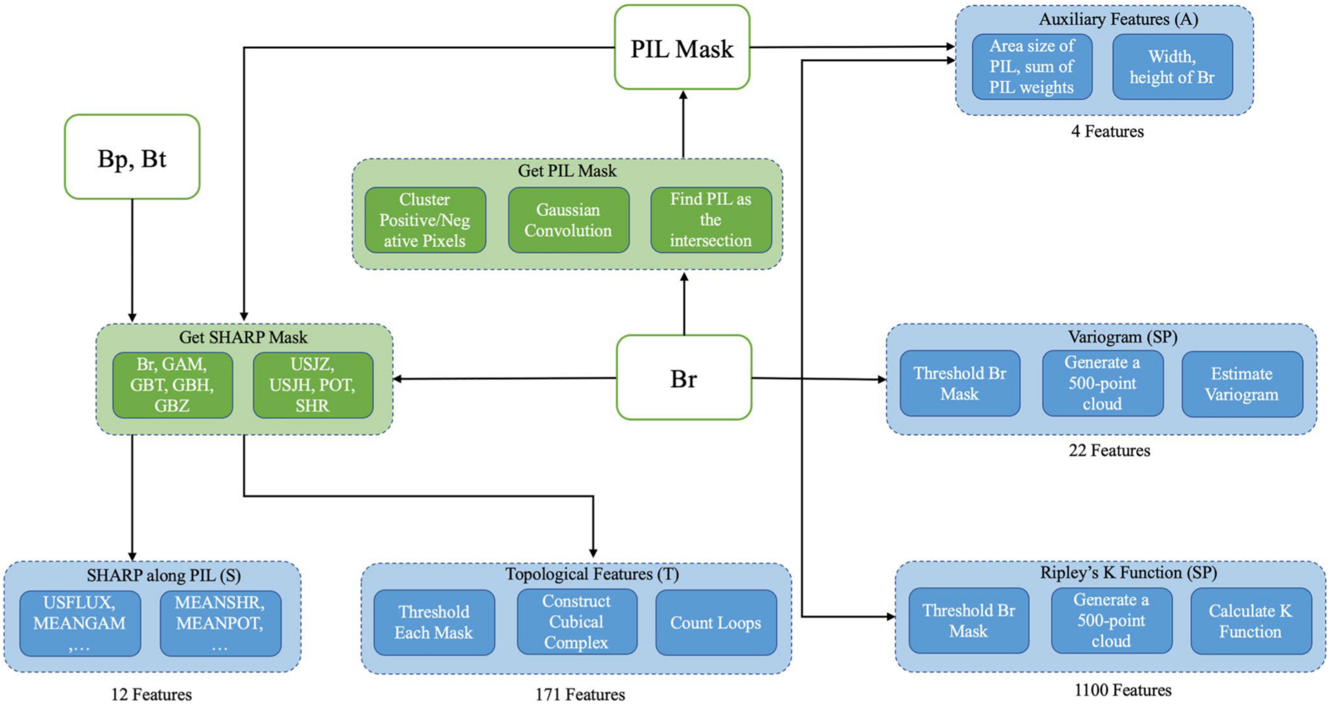

**Figure 8.** A Workflow Illustration of Feature Construction. The workflow chart shows all five sets of features constructed based on the HMI magnetograms. All source data and data used in the intermediate steps are colored in green and all output features are in blue. Small oval boxes include either the names of the features/masks or the steps for deriving the features/masks. Both of the topological features and Ripley's K function have rather high dimensions, so we apply dimension-reduction procedures to compress information before inputting them into the flare classification model.

## 3. Data Description and Prediction Results

In this section, we first introduce the data set used for feature construction. Then we elaborate on the model we use for flare classification and the data preprocessing steps for dimension reductions on the feature sets. Finally, we demonstrate the prediction performances 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 four time points: 1, 6, 12, and 24 hr prior to the peak soft X-ray 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 four time points, especially for 12 and 24-hr data due to the lack of records at exactly 12 or 24 hr before the flare peak time, so 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 four datasets.

### 3.2. Data Preprocessing and 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. The XGBoost model is implemented using the *xgboost* package in R, with the maximum depth of a tree set to 2, the learning rate set to 1 and the maximum number of rounds of boosting to be 100. We do not tune the parameter to optimize the performances to leave a fair ground for the comparison of features. We train the model with several different combinations of features from the feature set detailed in Figure 8.

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 five PCs in place of the original 19 features. We have confirmed that by choosing the top five PCs, we are able to explain 97.67% ∼ 99.61% variation of the topology features, depending on the SHARP masks. For every Ripley's K function, which is itself 100-dimensional, we carry out a functional principal component analysis (FPCA, Hall et al. [2006]) and keep top five PCs in place of the original 100 features. We can also confirm that the choice of top five PCs is able to summarize >99% of the functional variation for each Ripley's K function across all samples. After both steps, we reduce the topological features from 171 to 45 dimensions, and the Ripley's K function features from 1,100 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 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 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 four datasets, but the flares in the train/test set, in each iteration, is the same for all four 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 2, we show the mean test-set TSS for different feature sets across 20 iterations. In Appendix A, we include three extra tables for readers' references. In Table A1, we show the TSS with the same sets of features but trained with a totally random train-test split procedure. In Table A2, we also show the mean test-set TSS with the same sets of features for a strong (M/X) versus weak (B/C) flare classification task. In Table A3, we show the Heidke Skill Score (HSS) for the binary classification (M/X vs. B) models.

We define the model fitted with SHARP parameters as the benchmark model, and compare all other feature combinations against it. We can see 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 2, we highlight the TSS in boldface when the TSS of a certain feature combination is significantly higher than the benchmark, that is, we reject the null hypothesis with 95% confidence. Since the sample size is small and we split the train and test set based on HARP regions, the uncertainty can be large. Given the high uncertainty of the TSS for every feature set, it is more informative to look at the pairwise hypothesis

**Table 2**
*Average True Skill Score (TSS), Based on 20 Train-Test Splits, for Classifying Strong (M/X) Versus Weak (B) Flares Using Different Sets of Features*

| | Prediction time (hour) | | | |
|---|---|---|---|---|
| Feature combination | 1 | 6 | 12 | 24 |
| S | 0.553 (0.075) | 0.555 (0.071) | 0.539 (0.068) | 0.489 (0.077) |
| T | 0.548 (0.069) | 0.575 (0.071) | **0.561** (0.063) | **0.525** (0.069) |
| SP | 0.558 (0.066) | **0.578** (0.076) | 0.546 (0.071) | **0.528** (0.072) |
| S + T | **0.578** (0.071) | **0.581** (0.072) | 0.554 (0.057) | **0.536** (0.052) |
| S + SP | 0.56 (0.059) | **0.58** (0.073) | 0.538 (0.078) | **0.533** (0.074) |
| S + T + SP | **0.586** (0.077) | **0.599** (0.068) | 0.558 (0.08) | **0.57** (0.06) |
| S + T_PC + SP_PC | 0.554 (0.075) | 0.561 (0.077) | 0.53 (0.082) | **0.533** (0.076) |
| S + T + SP + A | **0.587** (0.071) | **0.605** (0.063) | 0.551 (0.077) | **0.55** (0.059) |
| S + T_PC + SP_PC + A | **0.578** (0.068) | 0.561 (0.071) | 0.533 (0.076) | **0.521** (0.089) |

*Note*. Standard errors calculated based on a bootstrap procedure are given in brackets. 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 PIL-weighted SHARP parameters, across the 20 iterations. The column names labeled by 1, 6, 12, and 24 stands for the four datasets collected at 1, 6, 12, and 24 hr before the flare peak time.

testing result (pairing TSS score based on each train/test split), which suggests the presence of performance improvement when adding new features.

As one can see that when using the topological and spatial statistics features, we can improve the TSS of flare classification 12–24 hr before the flare peak time, and further improves when it is getting close to the flare peak time (i.e., 1 and 6 hr before). When combining all features together, we can do significantly better than only using the SHARP parameters. When we reduce the dimension of the topological and spatial statistics features, we can still maintain a margin of skill improvement in the 24 hr prediction time scenario. The improvement is even more salient for the Heidke Skill Score, as shown in Table A3.

Different from our topological features, in Deshmukh, Berger, Bradley, and Meiss (2020) and Deshmukh, Berger, Meiss, and Bradley (2021), the authors split the $B_r$ mask into a positive component and a negative component and calculate the Betti numbers separately for the two components. We tried the same method and replace the topology features of the $B_r$ channel with two sets of features corresponding to its positive/negative component but did not get any significantly different results. We contemplate that the reasons are: (a) other SHARP masks have taken the sign of flux into account and (b) the feature space is very high-dimensional. So we still stick to our original topology features when we report our results.

As discussed in Barnes et al. (2016) and subsequent related works on flare forecasting, it is impossible or unfair to compare skill scores directly with published work in literature when the data preparation procedures are conducted differently. The same story goes for our case here. However, with the same data preparation adopted for various prediction models that we train and test, the conclusions that we make within the scope of this paper are based on fair comparisons. Our ultimate goal of this work is not to give the "best" prediction model for solar flare events, but to construct and interpret meaningful features that are useful for flare forecasting. Interested readers can refer to Barnes et al. (2016), Leka et al. (2019a), and X. Wang et al. (2020) for more thorough comparisons of results for flare forecasting models published in literature.

The performance improvement introduced by the new features is incremental and 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 et al., 2001], as adopted by Bobra et al. (2014) and 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

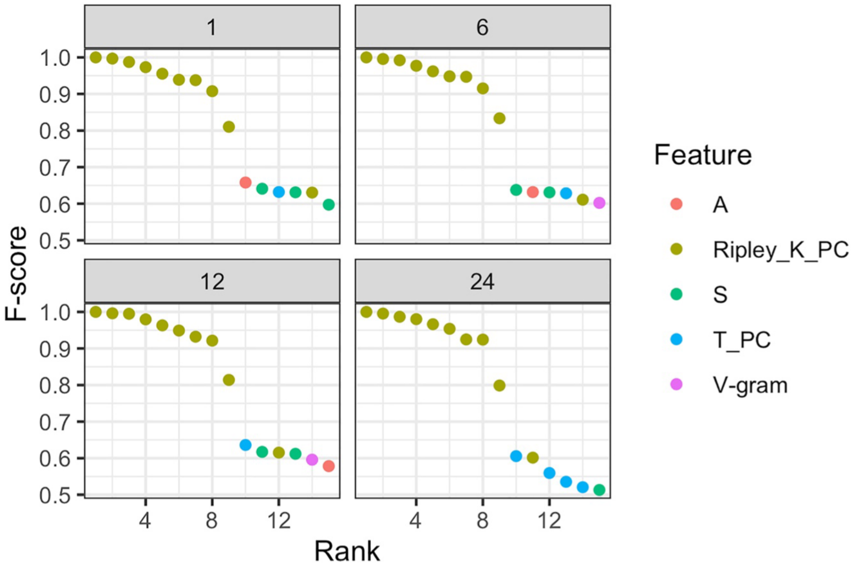

**Figure 9.** Normalized Fisher Score for selected features. Four panels correspond to the 1, 6, 12, and 24 hr data set. In each panel, top 15 features in terms of F-score are plotted. Among all four datasets, the top features are always the Ripley's K function's principal component score. Some features from other categories also rank among the 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.

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

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 principal component scores (**T_PC**), Ripley's K functional principal component scores (**Ripley_K_PC**), variogram (**V-gram**) and auxiliary features (**A**) based on their F-score, for the 4 datasets separately. All scores are calculated using the full data set. In Figure 9, 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 first PC score of Ripley's K function generated based on different threshold values.

To visualize the discriminating power of Ripley's K function, we choose four representative cases, two for B flare and two for M flare, and show their $B_r$, PIL mask, sampled point cloud at threshold 2,000 G and the corresponding Ripley's K function (divided by the number of PIL pixels) in Figure 10. The most distinctive difference between the derived Ripley's K functions, as one can see by comparing the four plots, is the level of the function value. There are two factors that create the level difference: PIL area size and concentration of high-$B_r$ 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 examples have longer, wide-spread PIL regions, as suggested by the PIL mask, but only have small scattered pieces of high-$B_r$ regions along the PIL. Both of these facts would lead to a low functional value. The M flares, on the contrary, have smaller PIL areas, but the high-$B_r$ regions are close to each other and clustered into relatively bigger chunks. Both of these lead to higher functional values.

Apart from the differences of the level of the functions, one may also notice that the shape of the functions differ. The B flares have two jumps in the function while the M flares have a constantly increasing trend prior to the plateau. This exactly corresponds to the scattered small pieces of high-$B_r$ regions shown in both B flare cases.

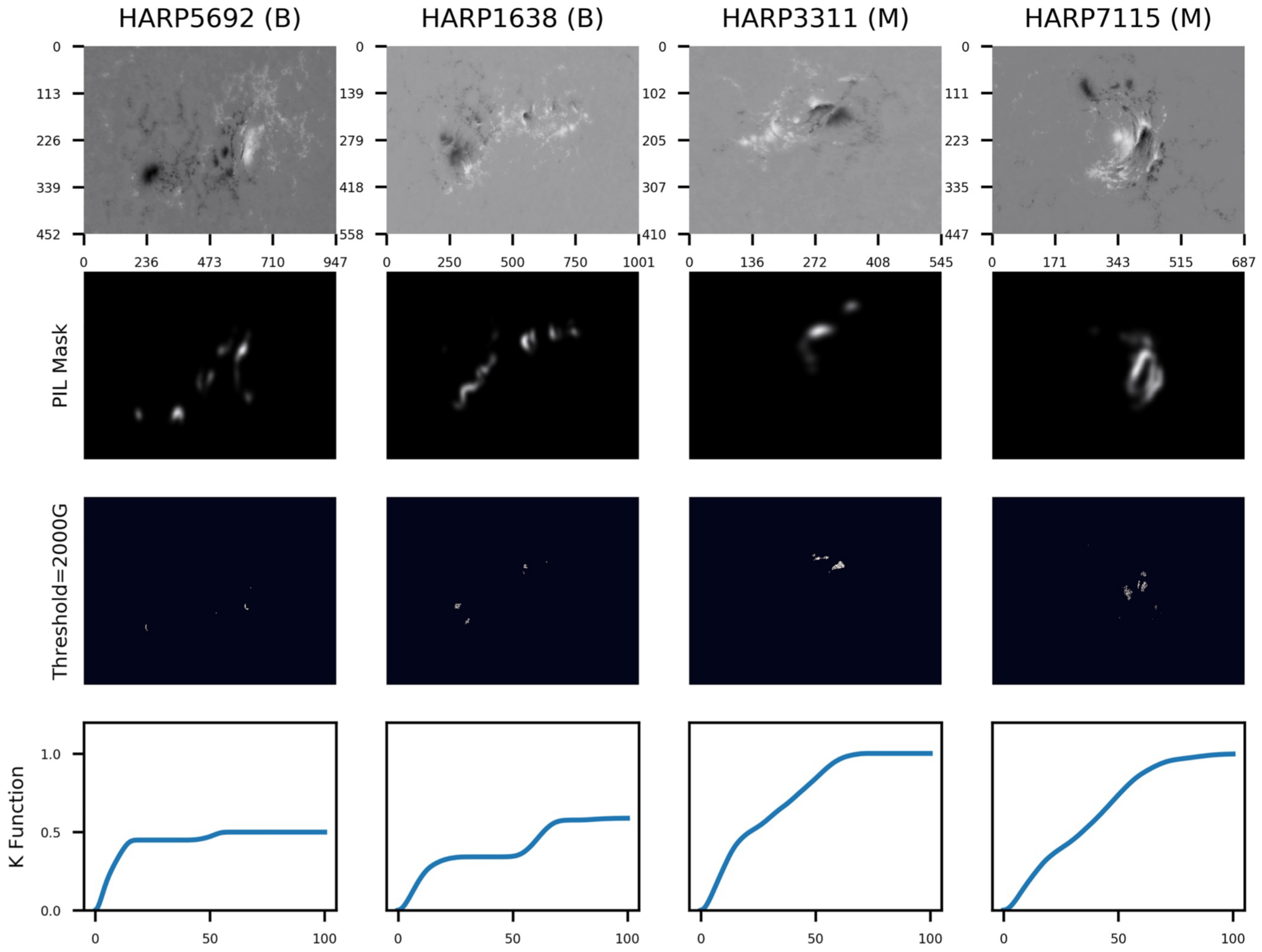

**Figure 10.** Four flare examples (columns 1–4): B6.1 from HARP 5692 peaked at 04:36, 26 June 2015; B5.3 from HARP 1638 peaked at 02:23, 9 May 2012; M1.0 from HARP 3311 peaked at 19:53, 26 October 2013; M1.0 from HARP 7115 peaked at 03:51, 5 September 2017. The four rows correspond to their $B_r$ values, PIL masks, point clouds with $B_r > 2{,}000$ G within each PIL region, and the Ripley's K functions respectively. The K function differs, in terms of level and shape, between two M flares and two B flares. The main reason is that there are only scattered small clusters of high-$B_r$ regions for the B flares. On the contrary, M flares have a subregion full of high-$B_r$ pixels.

Possibly some of the scattered small clusters of high-$B_r$ pixels for the B flare locate outside of the flaring area. The functions for the M flare have an increasing trend overall, suggesting that there is a region full of high-$B_r$ pixels (i.e., there are such pairs of high-flux pixels separated by an arbitrary distance within range 0 ~ 80). The flare discriminator, captured by the Ripley's K function, is essentially the degree of concentration of high-flux regions of an active region, along the polarity inversion line. An M-flare tends to have a larger proportion of high-flux $B_r$ pixels clustered spatially in the flaring area, leading to a different Ripley's K function in terms of both functional level and shape, as compared to a B-flare.

The variogram effective range parameters also differ significantly between the B and M flares. For instance, the variogram of the B flare from HARP 1638 has a range parameter at 1.96 while the M flare from HARP 3311 has it at 75.95. This fact suggests that most of the $B_r$ 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 $B_r$ values 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: (a) small PIL area; (b) large clusters of high-$B_r$ pixels not locating far from each other; (c) large $B_r$ value variation between clusters (e.g., between clusters of strongly positive and strongly negative flux). Such a pattern is not exhaustive for all strong flares, but it covers the majority of the cases and leads to features with high discriminating power.

## 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 (2021). We extend the scope of HMI images from just the $B_r$ component to multiple SHARP parameter 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 high $B_r$. The Variogram analyzes the spatial variation of the $B_r$ 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.

One major finding, besides the prediction performance gains, is that by focusing on the $B_r$ component 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 the $B_r$ component and topological features of the $B_r$ component. This result suggests that the spatial and shape information we derived from the $B_r$ component alone is a more powerful discriminator than the SHARP quantities derived from the horizontal magnetic field components. One of our future research goals is to analyze the spatial correlations of the $B_r$ component 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), Schrijver (2007), and Schrijver et al. (2005) who respectively found the gradient 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.

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 et al., 1998; Toriumi & Hotta, 2019). Simulations of flux emergence also show the development of strong shear along forming PILs (Archontis & Török, 2008; Fan, 2001; Fang et al., 2010; Manchester, 2001, 2007; Török et al., 2014), as well as producing eruptive behavior (Manchester et al., 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 related way.

The analysis of the $B_r$ component at multiple 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 et al., 2005). They found that typically M/X flares have higher 2D fractal dimensions, indicating a larger flaring areas of high $B_r$. The key difference between the spatial statistics tools we adopted and the fractal dimensions is that, after we choose $B_r$ areas above a threshold, we measure the spatial clustering pattern of high $B_r$ pixels, rather than 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 strong $B_r$ areas at multiple scales, and our result suggests that there are more discriminating features to be found in the shape (i.e., clustering pattern) of the strong $B_r$ areas, beyond their size.

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 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 time-series of HARP regions magnetograms to see if there are temporal trends of such spatial information that can further benefit flare predictions.

## Appendix A

### Appendix A1: TSS Based on Random Train-Test Split

Here are the results of TSS scores when we split train and test set randomly (Table A1):

**Table A1**
*Average True Skill Score (TSS), Based on 20 Random Train-Test Split, for Classifying Strong (M/X) Versus Weak (B) Flares Using Different Sets of Features*

| | Prediction time (hour) | | | |
|---|---|---|---|---|
| Feature combination | 1 | 6 | 12 | 24 |
| S | 0.613 (0.033) | 0.619 (0.04) | 0.588 (0.046) | 0.554 (0.037) |
| T | 0.621 (0.041) | 0.627 (0.037) | 0.601 (0.038) | **0.59** (0.05) |
| SP | 0.606 (0.039) | **0.642** (0.032) | 0.587 (0.046) | **0.593** (0.046) |
| S + T | **0.647** (0.044) | **0.653** (0.048) | **0.612** (0.034) | **0.587** (0.044) |
| S + SP | 0.624 (0.03) | **0.648** (0.033) | **0.605** (0.032) | **0.579** (0.039) |
| S + T + SP | **0.647** (0.037) | **0.66** (0.039) | **0.625** (0.042) | **0.608** (0.045) |
| S + T_PC + SP_PC | 0.616 (0.036) | 0.624 (0.048) | 0.6 (0.054) | **0.591** (0.039) |
| S + T + SP + A | **0.665** (0.04) | **0.66** (0.043) | **0.64** (0.042) | **0.615** (0.039) |
| S + T_PC + SP_PC + A | 0.628 (0.046) | 0.621 (0.043) | **0.608** (0.046) | **0.589** (0.043) |

*Note*. Standard error in brackets. 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, and 24 stands for the four datasets collected at 1, 6, 12, and 24 hr before the flare peak time.

### Appendix A2: TSS for M/X Versus B/C

The following table shows the result of TSS scores when we classify M/X flares against B/C flares. The TSS, when compared to our main result in Table 2, is much lower due to the inclusion of C-class flares. But the area under curve (AUC) for all feature combinations remain at a high level at around 0.69 ~ 0.78 and the statistical significance is also consistent with our main result: the new features can improve on top of the SHARP parameters, especially when one wants to predict flares 6, 12, and 24 hr ahead (Table A2).

**Table A2**
*Average True Skill Score (TSS), Based on 20 Train-Test Split, for Classifying Strong (M/X) Versus Weak (B/C) Flares Using Different Sets of Features*

| | Prediction time (hour) | | | |
|---|---|---|---|---|
| Feature combination | 1 | 6 | 12 | 24 |
| S | 0.234 (0.082) | 0.198 (0.069) | 0.182 (0.075) | 0.146 (0.089) |
| T | 0.224 (0.124) | 0.233 (0.097) | **0.310** (0.084) | **0.263** (0.109) |
| SP | 0.244 (0.086) | 0.192 (0.09) | 0.186 (0.047) | **0.200** (0.058) |

**Table A2**
*Continued*

| | Prediction time (hour) | | | |
|---|---|---|---|---|
| Feature combination | 1 | 6 | 12 | 24 |
| S + T | 0.266 (0.099) | 0.214 (0.072) | **0.302** (0.08) | **0.253** (0.11) |
| S + SP | 0.253 (0.096) | 0.188 (0.07) | 0.186 (0.086) | **0.209** (0.071) |
| S + T + SP | **0.276** (0.101) | 0.216 (0.098) | **0.302** (0.08) | **0.289** (0.081) |
| S + T_PC + SP_PC | 0.256 (0.074) | 0.231 (0.076) | 0.208 (0.078) | **0.236** (0.116) |
| S + T + SP + A | **0.276** (0.101) | 0.215 (0.098) | **0.303** (0.08) | **0.287** (0.083) |
| S + T_PC + SP_PC + A | 0.256 (0.074) | **0.231** (0.076) | 0.208 (0.078) | **0.236** (0.116) |

*Note*. 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, and 24 stands for the four datasets collected at 1, 6, 12, and 24 hr before the flare peak time.

### Appendix A3: Heidke Skill Score Results for Binary Classification

Here, as a supplemental result for Table 2, we show the Heidke Skill Score for the binary classification problem. The Heidke Skill Score (HSS) is calculated as:

$$\mathrm{HSS} = 2 \cdot \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FN} \times \mathrm{FP}}{(\mathrm{FN} + \mathrm{TP}) \times (\mathrm{FN} + \mathrm{TN}) + (\mathrm{FP} + \mathrm{TN}) \times (\mathrm{TP} + \mathrm{FP})}$$

where TP, TN, FN, and FP are the true positive, true negative, false positive, and false negative samples in the confusion matrix of the classification result (Table A3).

**Table A3**
*Average Heidke Skill Score (HSS), Based on 20 Random Train-Test Split, for Classifying Strong (M/X) Versus Weak (B) Flares Using Different Sets of Features*

| | Prediction time (hour) | | | |
|---|---|---|---|---|
| Feature combination | 1 | 6 | 12 | 24 |
| S | 0.473 (0.069) | 0.478 (0.067) | 0.467 (0.06) | 0.413 (0.073) |
| T | **0.501** (0.058) | **0.525** (0.071) | **0.504** (0.06) | **0.473** (0.061) |
| SP | **0.508** (0.058) | **0.529** (0.073) | 0.485 (0.068) | **0.478** (0.068) |
| S + T | **0.529** (0.072) | **0.526** (0.069) | **0.495** (0.056) | **0.478** (0.053) |
| S + SP | **0.514** (0.06) | **0.532** (0.076) | 0.481 (0.074) | **0.485** (0.075) |
| S + T + SP | **0.544** (0.071) | **0.548** (0.067) | **0.502** (0.069) | **0.521** (0.06) |
| S + T_PC + SP_PC | **0.525** (0.066) | **0.52** (0.067) | **0.486** (0.07) | **0.489** (0.067) |
| S + T + SP + A | **0.535** (0.066) | **0.545** (0.059) | **0.488** (0.065) | **0.496** (0.059) |
| S + T_PC + SP_PC + A | **0.542** (0.059) | **0.51** (0.064) | **0.488** (0.064) | **0.478** (0.075) |

*Note*. Standard error in brackets. 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 HSS 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, and 24 stands for the four datasets collected at 1, 6, 12, and 24 hr before the flare peak time.

## Data Availability Statement

All codes and data are included in our Github repository at https://github.com/husun0822/spat_feature_flare_pred.

**Acknowledgments**
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). Eruption of magnetic flux ropes during flux emergence. *Astronomy and Astrophysics*, *492*(2), L35–L38. https://doi.org/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), 530–543. https://doi.org/10.1086/524371

Aschwanden, M. J., Crosby, N. B., Dimitropoulou, M., Georgoulis, M. K., Hergarten, S., McAteer, J., et al. (2016). 25 years of self-organized criticality: Solar and astrophysics. *Space Science Reviews*, *198*(1–4), 47–166. https://doi.org/10.1007/s11214-014-0054-6

Barnes, G., Leka, K. D., Schrijver, C. J., Colak, T., Qahwaji, R., Ashamari, O., et al. (2016). A comparison of flare forecasting methods. I. Results from the "all-clear" workshop. *The Astrophysical Journal*, *829*(2), 89. https://doi.org/10.3847/0004-637x/829/2/89

Barnes, G., Leka, K. D., Schumer, E., & Della-Rose, D. (2007). Probabilistic forecasting of solar flares from vector magnetogram data. *Space Weather*, *5*(9). https://doi.org/10.1029/2007sw000317

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. https://doi.org/10.1088/0004-637x/798/2/135

Bobra, M. G., Sun, X., Hoeksema, J. T., Turmon, M., Liu, Y., Hayashi, K., et al. (2014). The helioseismic and magnetic imager (HMI) vector magnetic field pipeline: Sharps – Space-weather HMI active region patches. *Solar Physics*, *289*(9), 3549–3578. https://doi.org/10.1007/s11207-014-0529-3

Camporeale, E. (2019). The challenge of machine learning in space weather nowcasting and forecasting. *Space Weather*, *17*, 1166–1207. https://doi.org/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). ACM.

Chen, Y., Manchester, W. B., Hero, A. O., Toth, G., DuFumier, B., Zhou, T., et al. (2019). Identifying solar flare precursors using time series of SDO/HMI images and SHARP parameters. *Space Weather*, *17*(10), 1404–1426. https://doi.org/10.1029/2019sw002214

Cressie, N., & Hawkins, D. M. (1980). Robust estimation of the variogram: I. *Journal of the International Association for Mathematical Geology*, *12*(2), 115–125. https://doi.org/10.1007/bf01035243

Deshmukh, V., Berger, T., Meiss, J., & Bradley, E. (2021). *Shape-based feature engineering for solar flare prediction*. arXiv preprint arXiv:2012.14405. https://ojs.aaai.org/index.php/AAAI/article/view/17795

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. https://doi.org/10.1051/swsc/2020014

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 Proceedings* (Vol. 96, pp. 226–231). https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf

Falconer, D. A. (2001). A prospective method for predicting coronal mass ejections from vector magnetograms. *Journal of Geophysical Research*, *106*, 25185–25190. https://doi.org/10.1029/2000JA004005

Falconer, D. A., Moore, R. L., & Gary, G. A. (2002). Correlation of the coronal mass ejection productivity of solar active regions with measures of their global nonpotentiality from vector magnetograms: Baseline results. *The Astrophysical Journal*, *569*, 1016–1025. https://doi.org/10.1086/339161

Falconer, D. A., Moore, R. L., & Gary, G. A. (2003). A measure from line-of-sight magnetograms for prediction of coronal mass ejections. *Journal of Geophysical Research*, *108*, 1380. https://doi.org/10.1029/2003JA010030

Falconer, D. A., Moore, R. L., & Gary, G. A. (2006). Magnetic causes of solar coronal mass ejections: Dominance of the free magnetic energy over the magnetic twist alone. *The Astrophysical Journal*, *644*, 1258–1272. https://doi.org/10.1086/503699

Fan, Y. (2001). The emergence of a twisted Ω-tube into the solar atmosphere. *The Astrophysical Journal*, *554*(1), L111–L114. https://doi.org/10.1086/320935

Fang, F., & Fan, Y. (2015). $\delta$-Sunspot formation in simulation of active-region-scale flux emergence. *The Astrophysical Journal*, *806*(1), 79. https://doi.org/10.1088/0004-637X/806/1/79

Fang, F., Manchester, W., Abbett, W. P., & van der Holst, B. (2010). Simulation of flux emergence from the convection zone to the corona. *The Astrophysical Journal*, *714*(2), 1649–1657. https://doi.org/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). https://doi.org/10.1007/s11207-018-1250-4

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/s0167-9473(01)00065-2

Gelfand, A. E., Diggle, P., Guttorp, P., & Fuentes, M. (2010). *Handbook of spatial statistics*. CRC Press. https://doi.org/10.1201/9781420072884

Ghrist, R. (2008). Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, *45*(1), 61–75. https://doi.org/10.1090/s0273-0979-07-01191-3

Hall, P., Müller, H.-G., & Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, *34*(3), 1493–1517. https://doi.org/10.1214/009053606000000272

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. https://doi.org/10.1029/2020sw002440

Kaczynski, T., Mischaikow, K. M., & Mrozek, M. (2004). *Computational homology* (Vol. 3, No. 7). Springer. https://link.springer.com/book/10.1007/b97315

Leka, K. D., & Barnes, G. (2003a). Photospheric magnetic field properties of flaring versus flare-quiet active regions. I. Data, general approach, and sample results. *The Astrophysical Journal*, *595*(2), 1277–1295. https://doi.org/10.1086/377511

Leka, K. D., & Barnes, G. (2003b). Photospheric magnetic field properties of flaring versus flare-quiet active regions. II. Discriminant analysis. *The Astrophysical Journal*, *595*(2), 1296–1306. https://doi.org/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. https://doi.org/10.1016/B978-0-12-812700-1.00003-0

Leka, K. D., Park, S.-H., Kusano, K., Andries, J., Barnes, G., Bingham, S., et al. (2019a). 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. https://doi.org/10.3847/1538-4365/ab2e12

Leka, K. D., Park, S.-H., Kusano, K., Andries, J., Barnes, G., Bingham, S., et al. (2019b). A comparison of flare forecasting methods. III. Systematic behaviors of operational solar flare forecasting systems. *The Astrophysical Journal*, *881*(2), 101. https://doi.org/10.3847/1538-4357/ab2e11

Linton, M. G., Dahlburg, R. B., Fisher, G. H., & Longcope, D. W. (1998). Nonlinear evolution of Kink-unstable magnetic flux tubes and solar $\delta$-spot active regions. *The Astrophysical Journal*, *507*(1), 404–416. https://doi.org/10.1086/306299

Liu, C., Deng, N., Wang, J. T. L., & Wang, H. (2017). Predicting solar flares using SDO/HMI vector magnetic data products and the random forest algorithm. *The Astrophysical Journal*, *843*(2), 104. https://doi.org/10.3847/1538-4357/aa789b

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

Manchester, W., IV. (2001). The role of nonlinear Alfvén waves in shear formation during solar magnetic flux emergence. *The Astrophysical Journal*, *547*(1), 503–519. https://doi.org/10.1086/318342

Manchester, W., IV. (2007). Solar atmospheric dynamic coupling due to shear motions driven by the Lorentz force. *The Astrophysical Journal*, *666*(1), 532–540. https://doi.org/10.1086/520493

Manchester, W., IV., Gombosi, T., DeZeeuw, D., & Fan, Y. (2004). Eruption of a buoyantly emerging magnetic flux rope. *The Astrophysical Journal*, *610*(1), 588–596. https://doi.org/10.1086/421516

Maria, C., Boissonnat, J.-D., Glisse, M., & Yvinec, M. (2014). The gudhi library: Simplicial complexes and persistent homology. In *International Congress on Mathematical Software* (pp. 167–174). https://doi.org/10.1007/978-3-662-44199-2_28

McAteer, R. T. J., Gallagher, P. T., & Ireland, J. (2005). Statistics of active region complexity: A large-scale fractal dimension survey. *The Astrophysical Journal*, *631*(1), 628–635. https://doi.org/10.1086/432412

Munch, E. (2017). A user's guide to topological data analysis. *Journal of Learning Analytics*, *4*(2), 47–61. https://doi.org/10.18608/jla.2017.42.6

Muranushi, Y. H., Muranushi, T., Asai, A., Okanohara, D., Raymond, R., Watanabe, G., et al. (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. https://doi.org/10.1186/s40623-021-01381-9

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. https://doi.org/10.3847/1538-4357/aab9a7

Oliver, M. A., & Webster, R. (2015). *Basic steps in geostatistics: The variogram and kriging*. Springer. https://link.springer.com/book/10.1007/978-3-319-15865-5

Omre, H. (1984). The variogram and its estimation. In *Geostatistics for natural resources characterization* (pp. 107–125). Springer. https://doi.org/10.1007/978-94-009-3699-7_7

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, *13*(2), 255–266. https://doi.org/10.2307/3212829

Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, *21*(3), 660–674. https://doi.org/10.1109/21.97458

Schrijver, C. J. (2007). A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting. *The Astrophysical Journal*, *655*(2), L117–L120. https://doi.org/10.1086/511857

Schrijver, C. J., De Rosa, M. L., Title, A. M., & Metcalf, T. R. (2005). The nonpotentiality of active-region coronae and the dynamics of the photospheric magnetic field. *The Astrophysical Journal*, *628*(1), 501–513. https://doi.org/10.1086/430733

Stork, D. G., Duda, R. O., Hart, P. E., & Stork, D. (2001). *Pattern classification*. A Wiley-Interscience Publication. https://www.wiley.com/en-us/Pattern+Classification%2C+2nd+Edition-p-9780471056690

Toriumi, S., & Hotta, H. (2019). Spontaneous generation of $\delta$-sunspots in convective magnetohydrodynamic simulation of magnetic flux emergence. *The Astrophysical Journal*, *886*(1), L21. https://doi.org/10.3847/2041-8213/ab55e7

Török, T., Leake, J. E., Titov, V. S., Archontis, V., Mikić, Z., Linton, M. G., et al. (2014). Distribution of electric currents in solar active regions. *The Astrophysical Journal*, *782*(1), L10. https://doi.org/10.1088/2041-8205/782/1/l10

Uritsky, V. M., & Davila, J. M. (2012). Multiscale dynamics of solar magnetic structures. *The Astrophysical Journal*, *748*(1), 60. https://doi.org/10.1088/0004-637X/748/1/60

Wang, J., Liu, S., Ao, X., Zhang, Y., Wang, T., & Liu, Y. (2019). Parameters derived from the SDO/HMI vector magnetic field data: Potential to improve machine-learning-based solar flare prediction models. *The Astrophysical Journal*, *884*(2), 175. https://doi.org/10.3847/1538-4357/ab441b

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. https://doi.org/10.3847/1538-4357/ab7b6c

Wang, X., Chen, Y., Toth, G., Manchester, W. B., Gombosi, T. I., Hero, A. O., et al. (2020). Predicting solar flares with machine learning: Investigating solar cycle dependence. *The Astrophysical Journal*, *895*(1), 3. https://doi.org/10.3847/1538-4357/ab89ac

Wasserman, L. (2018). Topological data analysis. *Annual Review of Statistics and Its Application*, *5*, 501–532. https://doi.org/10.1146/annurev-statistics-031017-100045

Zomorodian, A. J. (Ed.). (2012). *Advances in Applied and Computational Topology: American Mathematical Society Short Course on Computational Topology, January 4–5, 2011, New Orleans, Louisiana* (Vol. 70). American Mathematical Society.