Improved and Interpretable Solar Flare Predictions with Spatial and Topological Features of the Polarity-Inversion-Line Masked Magnetograms[1]

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

Aug 9, 2021

[1]Our submitted manuscript can be viewed at https://www.essoar.org/doi/abs/10.1002/essoar.10507540.1

Overview

1. Background
2. Data
3. Feature Engineering
   - Topological Feature
   - Spatial Feature I: Ripley's K Function
   - Spatial Feature II: Variogram
4. Main Results
   - Prediction Performance
   - Interpretation
5. Conclusions

Flare Prediction with HMI Magnetograms

- Bobra, Sun, et al. (2014) introduced the Space-weather HMI Active Region Patch (SHARP) parameters, which are derived from the HMI/SDO magnetograms and have been used frequently for solar flare prediction models in recent years (e.g. Bobra and Couvidat, 2015; Florios et al., 2018; Chen et al., 2019; Camporeale, 2019; Jiao et al., 2020).

Flare Prediction with HMI Magnetograms

- Bobra, Sun, et al. (2014) introduced the Space-weather HMI Active Region Patch (SHARP) parameters, which are derived from the HMI/SDO magnetograms and have been used frequently for solar flare prediction models in recent years (e.g. Bobra and Couvidat, 2015; Florios et al., 2018; Chen et al., 2019; Camporeale, 2019; Jiao et al., 2020).
- There are efforts using deep neural network methods, which directly take the HMI/SDO magnetogram images to predict solar eruptions (e.g. the Long Short Term Memory network adopted by Chen et al. (2019) and Liu et al. (2019)).

Flare Prediction with HMI Magnetograms

- Bobra, Sun, et al. (2014) introduced the Space-weather HMI Active Region Patch (SHARP) parameters, which are derived from the HMI/SDO magnetograms and have been used frequently for solar flare prediction models in recent years (e.g. Bobra and Couvidat, 2015; Florios et al., 2018; Chen et al., 2019; Camporeale, 2019; Jiao et al., 2020).
- There are efforts using deep neural network methods, which directly take the HMI/SDO magnetogram images to predict solar eruptions (e.g. the Long Short Term Memory network adopted by Chen et al. (2019) and Liu et al. (2019)).
- Recent efforts (Deshmukh, Berger, Bradley, et al., 2020; Deshmukh, Berger, Meiss, et al., 2020) leverage the shape information contained in HMI magnetograms to construct interpretable and predictive new parameters for flare prediction.

Highlights of Our Work

1. Expand the feature set derived from the HMI magnetograms for flare prediction using tools from both *topological data analysis* and *spatial statistics*.

Highlights of Our Work

1. Expand the feature set derived from the HMI magnetograms for flare prediction using tools from both *topological data analysis* and *spatial statistics*.
2. Derive features not only from the PIL-masked HMI magnetograms but also from SHARP parameter masks.

Highlights of Our Work

1. Expand the feature set derived from the HMI magnetograms for flare prediction using tools from both *topological data analysis* and *spatial statistics*.
2. Derive features not only from the PIL-masked HMI magnetograms but also from SHARP parameter masks.
3. Marginally but steadily improved the skill score of the classification model of strong vs. weak solar flares.

Dataset

- We use the Geostationary Operational Environmental Satellites (GOES) flare list spanning 2010/12 - 2018/06 for collecting flare events, leading to 399 M/X class flares and 1,972 B class flares coming from 487 HARP regions.

Dataset

- We use the Geostationary Operational Environmental Satellites (GOES) flare list spanning 2010/12 - 2018/06 for collecting flare events, leading to 399 M/X class flares and 1,972 B class flares coming from 487 HARP regions.
- 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 soft X-ray flux.

Dataset

- We use the Geostationary Operational Environmental Satellites (GOES) flare list spanning 2010/12 - 2018/06 for collecting flare events, leading to 399 M/X class flares and 1,972 B class flares coming from 487 HARP regions.
- 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 soft X-ray flux.
- For each flare, at all four time points, raw data of the B_r, B_p, B_t components of the magnetic field are collected.

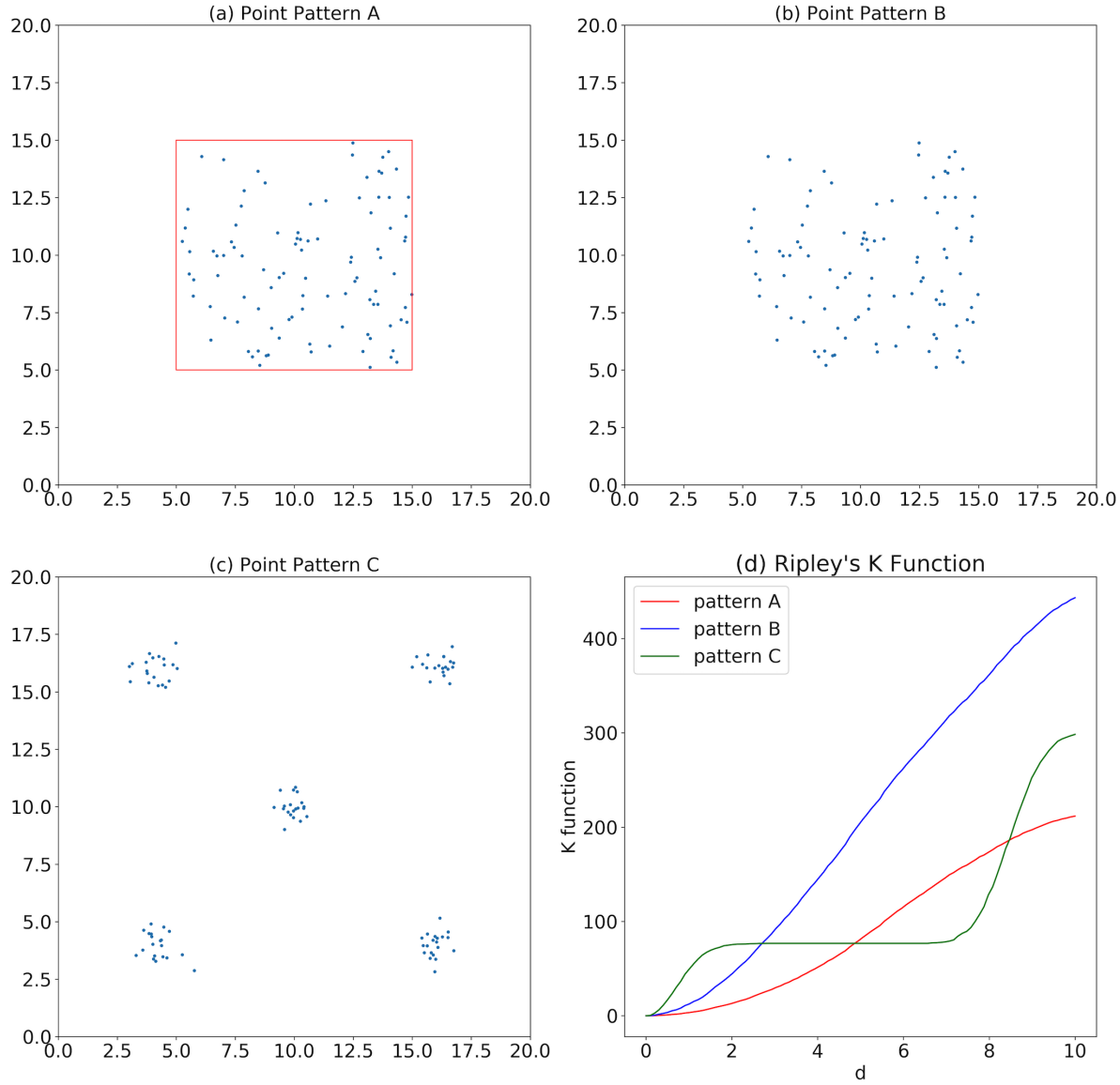
Derive SHARP Parameter Maps

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	$\arccos\left(\frac{\mathbf{B}_x^{POT}\times\mathbf{B}_x+\mathbf{B}_y^{POT}\times\mathbf{B}_y+\mathbf{B}_z^2}{\sqrt{\mathbf{B}_x^{POT\,2}+\mathbf{B}_y^{POT\,2}+\mathbf{B}_z^2}\sqrt{\mathbf{B}_x^2+\mathbf{B}_y^2+\mathbf{B}_z^2}}\right)$	Degree

Table 1. SHARP parameter mask, formula 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 parameters can be found in Table 3 of Bobra et al. (2014).

Topological Feature: Betti Numbers

Spatial Feature I: Ripley's K Function



Spatial Feature I: Ripley's K Function

- For the thresholded B_r mask, we randomly pick 500 pixels, with sampling probability proportional to $|B_r|$, to construct a point cloud. Each picked pixel has a pair of (x, y) pixel coordinates in the 2D pixel grid.
- Ripley's K function:

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

where $k_{i,j} = 1$ if the i-th and j-th pixel are within distance d, and $n = 500$ in our case. A is the area size and is defined as the number of PIL pixels in our study.

Spatial Feature I: Ripley's K Function

- For the thresholded B_r mask, we randomly pick 500 pixels, with sampling probability proportional to $|B_r|$, to construct a point cloud. Each picked pixel has a pair of (x, y) pixel coordinates in the 2D pixel grid.
- Ripley's K function:

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

where $k_{i,j} = 1$ if the i-th and j-th pixel are within distance d, and $n = 500$ in our case. A is the area size and is defined as the number of PIL pixels in our study.

Spatial Feature I: Ripley's K Function

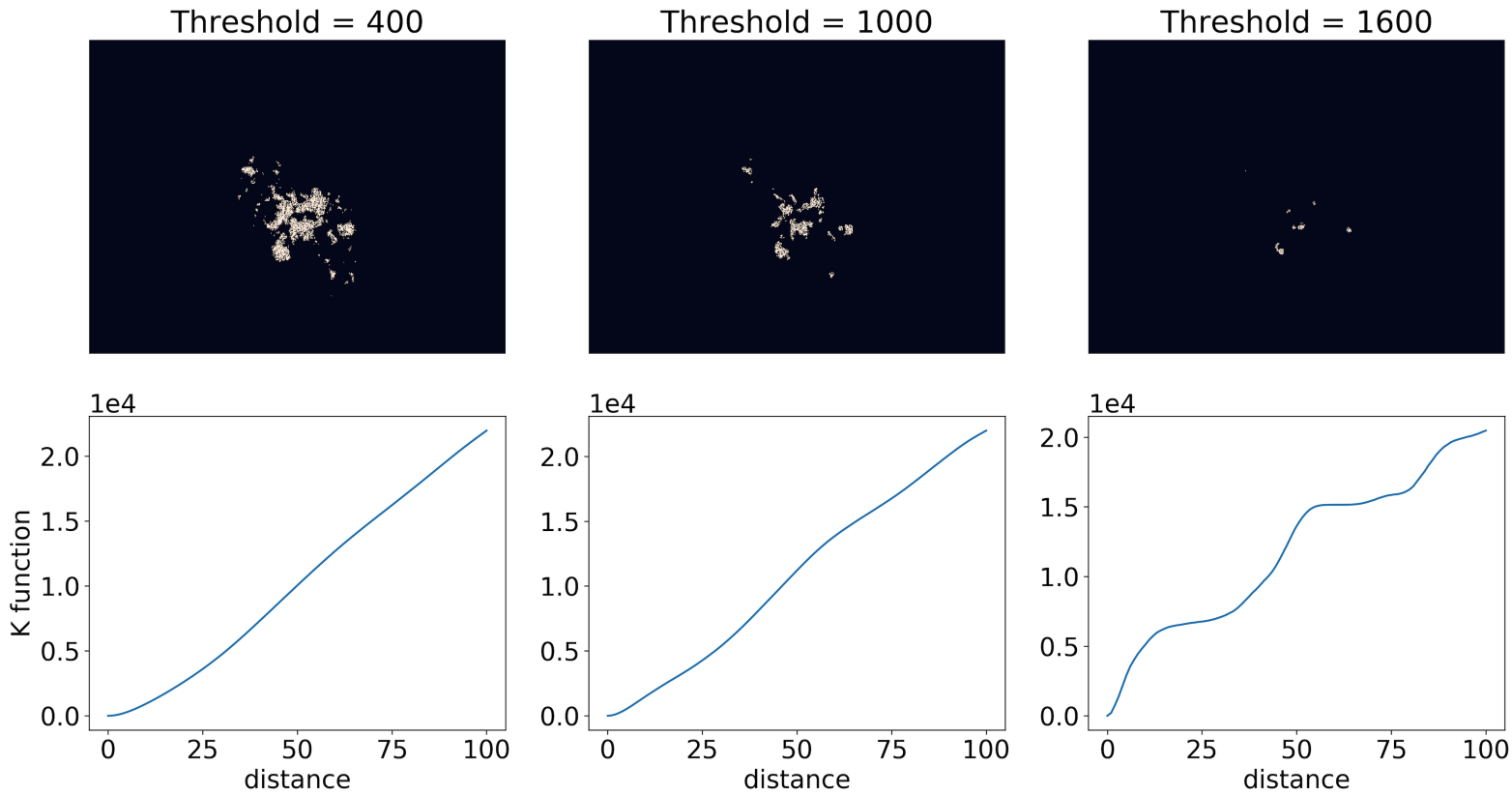


Figure: Point cloud and the corresponding Ripley's K function for the B_r mask collected from HARP 377, 1 hour before the M flare peaked at 2011.02.13 17:38. The top row includes 3 point clouds generated by 3 thresholds at 400G, 1000G, 1600G. The bottom row shows the 3 corresponding Ripley's K functions.

Spatial Feature II: Variogram

- With the same point cloud as in Ripley's K function calculation, the Variogram is:

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

  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. And $z(.)$ yields the B_r value at a pixel.
- In practice, it is hard to find multiple pairs of pixels separated exactly by distance d. Pairs of pixels will be put into disjoint bins of Euclidean distance for estimating the Variogram.
- Variogram is measuring the variation of B_r at two spatial locations separated by an arbitrary distance d.

Spatial Feature II: Variogram

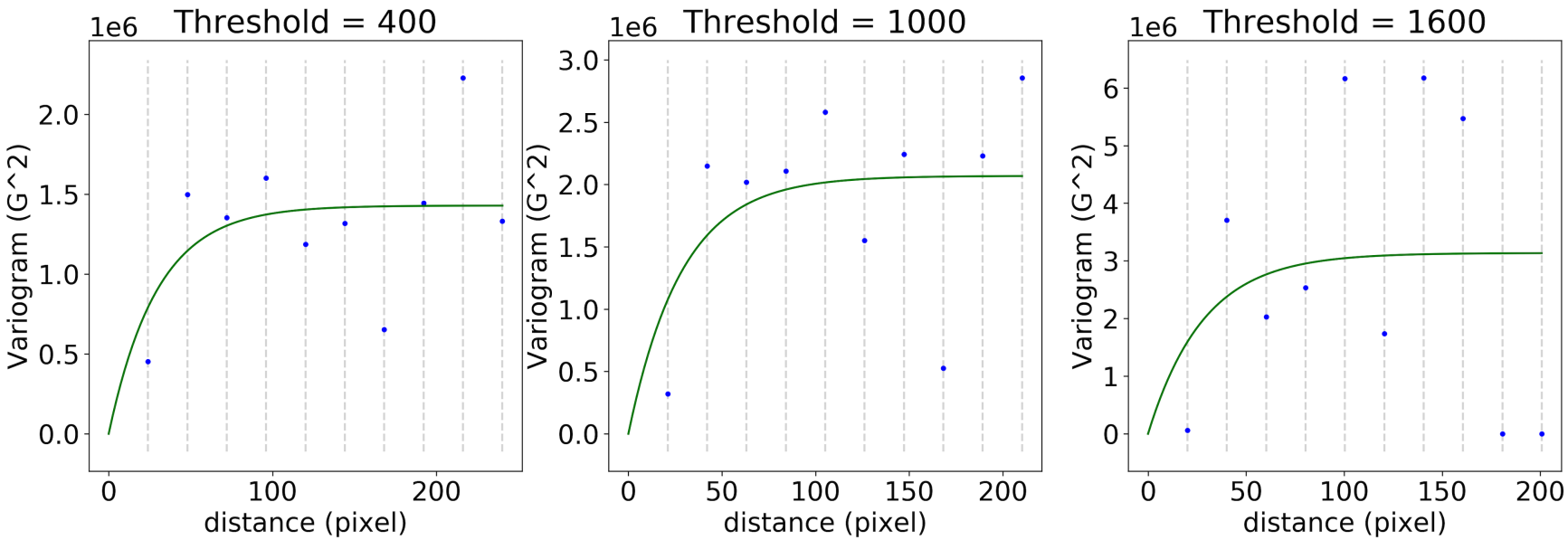


Figure: Variogram examples. Vertical dashed line show the center of each distance interval, and the scatter points are the semi-variance (see equation 1) of B_r values for all pairs of pixels separated by the distance within the interval. The blue line is the fitted curve for the variogram estimates. Note that the scales of x,y axes are different across the three graphs.

Feature Overview

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: A brief overview of the feature sets considered for flare classification.

True Skill Score based on Fitted Xgboost Model

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**

Figure: True skill score (TSS) based on Xgboost model fitted with different combinations of feature sets. 20 train-test-split are used to evaluate the average performance. Boldface numbers mean that the TSS is statistically significantly better than the baseline (use SHARP parameter only).

Feature Importance: Fisher Score

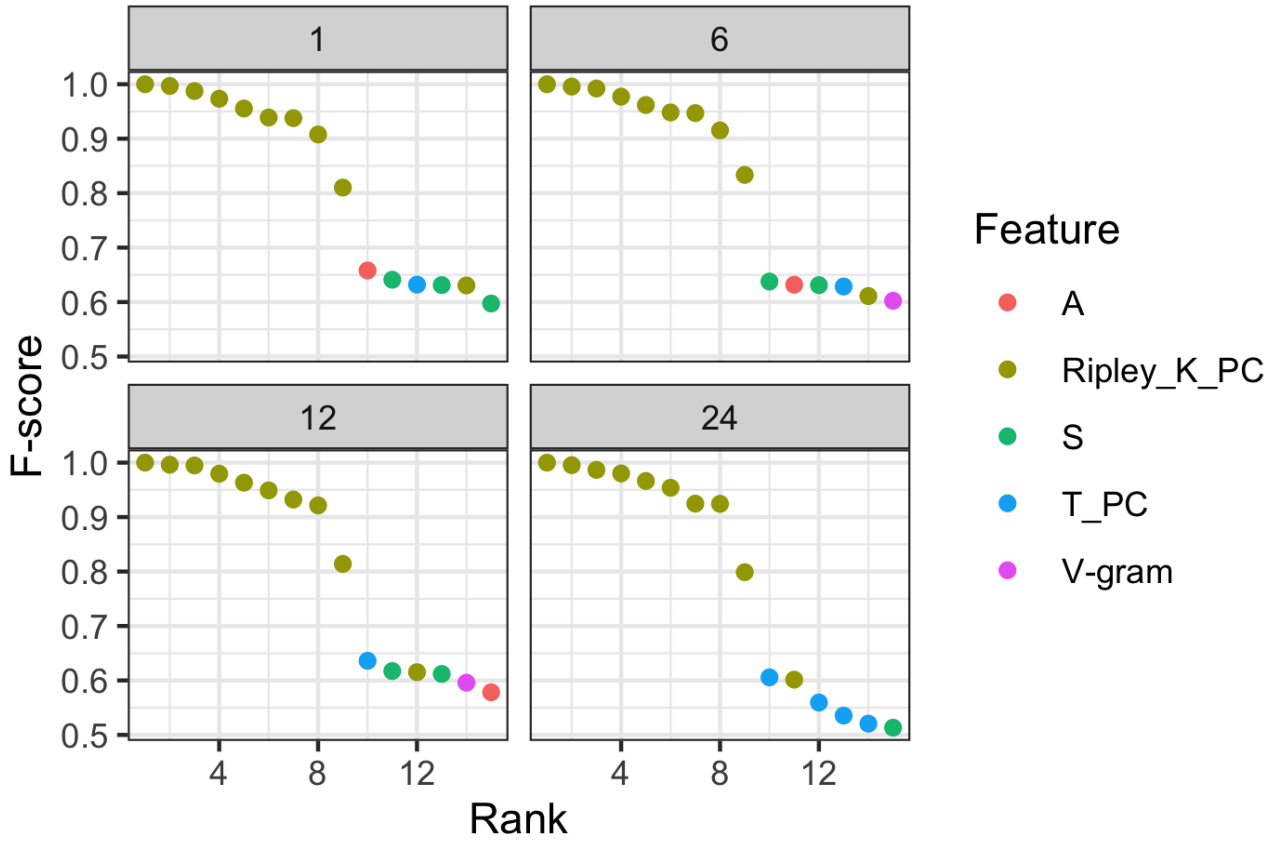
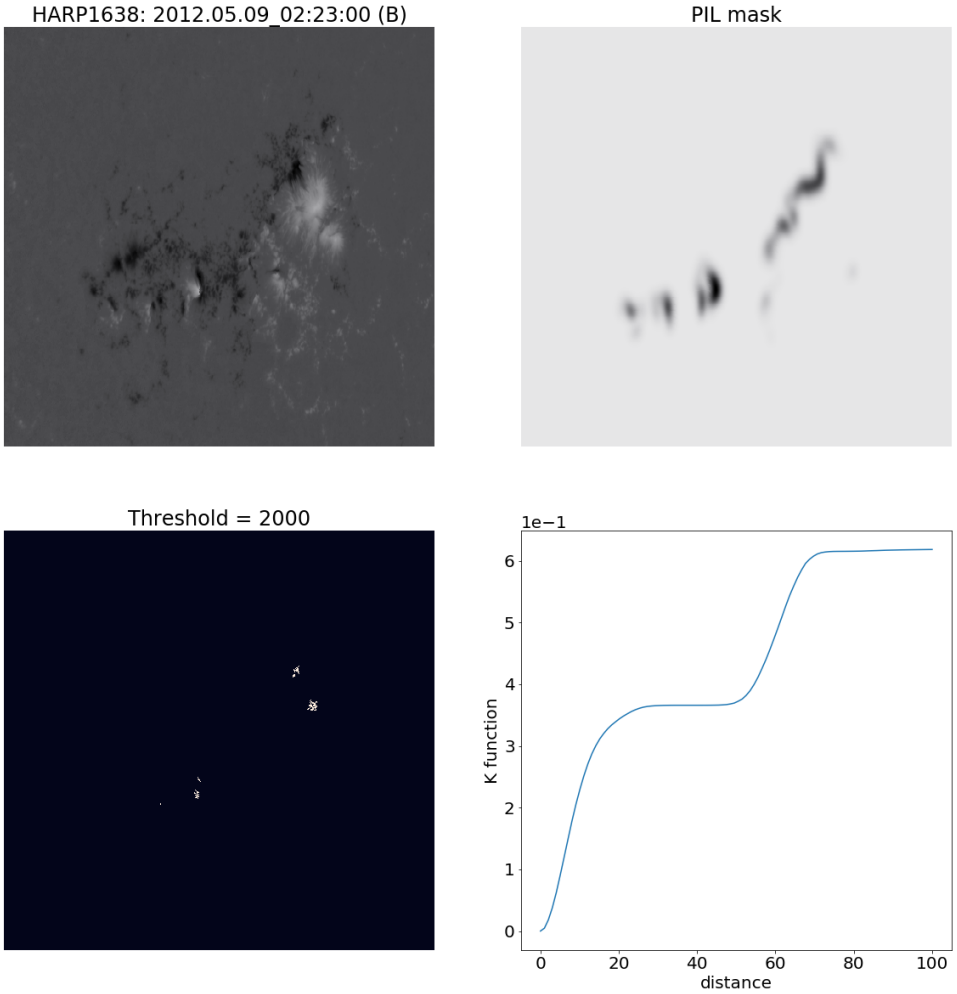
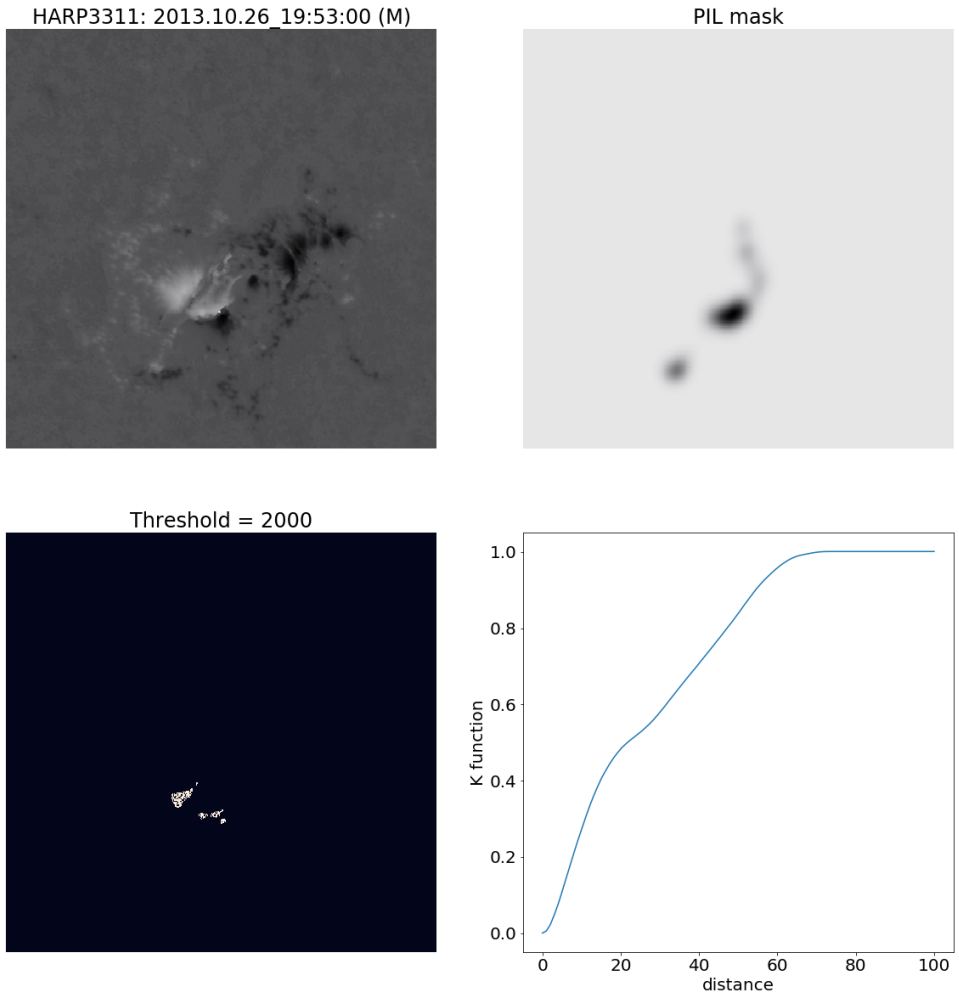


Figure: Normalized Fisher Score for selected features. Four panels correspond to the 1,6,12,24 hour dataset. Among all 4 datasets, the top features are always the Ripley's K function's principal component score. Some features from other categories are also ranked among top features.

Ripley's K function: B-flare Example



Ripley's K function: M-flare Example



Conclusion

In this project, we:

- Concentrate on SHARP parameter spatial distributions along the polarity inversion line regions.

Conclusion

In this project, we:

- Concentrate on SHARP parameter spatial distributions along the polarity inversion line regions.
- Engineered interpretable and predictive features summarizing the spatial variation, dispersion patterns of various SHARP quantities, especially the B_r component, using tools from TDA and spatial statistics.

Conclusion

In this project, we:

- Concentrate on SHARP parameter spatial distributions along the polarity inversion line regions.
- Engineered interpretable and predictive features summarizing the spatial variation, dispersion patterns of various SHARP quantities, especially the B_r component, using tools from TDA and spatial statistics.
- Obtained marginal but steady improvement on the solar flare classification task.

Conclusion

In this project, we:

- Concentrate on SHARP parameter spatial distributions along the polarity inversion line regions.
- Engineered interpretable and predictive features summarizing the spatial variation, dispersion patterns of various SHARP quantities, especially the B_r component, using tools from TDA and spatial statistics.
- Obtained marginal but steady improvement on the solar flare classification task.
- The spatial features derived solely from the B_r component are as good or better for flare prediction than full vector SHARP parameters. Theoretically interesting and important for future missions.

References I

Bobra, M. G., X. Sun, J. T. Hoeksema, M. Turmon, Y. Liu, K. Hayashi, G. Barnes, and K. D. Leka (Sept. 2014). "The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: SHARPs – Space-Weather HMI Active Region Patches". In: *Solar Physics* 289.9, pp. 3549–3578.

Bobra, Monica G and Sebastien Couvidat (2015). "Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm". In: *The Astrophysical Journal* 798.2, p. 135.

Camporeale, Enrico (July 2019). "The Challenge of Machine Learning in Space Weather Nowcasting and Forecasting". In: *Space Weather* 17. DOI: 10.1029/2018sw002061.

References II

Chen, Yang, Ward B Manchester, Alfred O Hero, Gabor Toth, Benoit DuFumier, Tian Zhou, Xiantong Wang, Haonan Zhu, Zeyu Sun, and Tamas I Gombosi (2019). "Identifying solar flare precursors using time series of SDO/HMI Images and SHARP Parameters". In: *Space Weather* 17.10, pp. 1404–1426.

Deshmukh, Varad, Thomas Berger, James Meiss, and Elizabeth Bradley (2020). "Shape-based Feature Engineering for Solar Flare Prediction". In: *arXiv preprint arXiv:2012.14405*.

Deshmukh, Varad, Thomas E Berger, Elizabeth Bradley, and James D Meiss (2020). "Leveraging the mathematics of shape for solar magnetic eruption prediction". In: *Journal of Space Weather and Space Climate* 10, p. 13.

References III

Florios, Kostas, Ioannis Kontogiannis, Sung-Hong Park, Jordan A Guerra, Federico Benvenuto, D Shaun Bloomfield, and Manolis K Georgoulis (2018). "Forecasting Solar Flares Using Magnetogram-based Predictors and Machine Learning". In: *Solar Physics* 293.2, p. 28. DOI: doi:10.1007/s11207-018-1250-4.

Jiao, Zhenbang, Hu Sun, Xiantong Wang, Ward Manchester, Tamas Gombosi, Alfred Hero, and Yang Chen (2020). "Solar flare intensity prediction with machine learning models". In: *Space Weather* 18.7, e2020SW002440.

Liu, Hao, Chang Liu, Jason T. L. Wang, and Haimin Wang (June 2019). "Predicting Solar Flares Using a Long Short-term Memory Network". In: *The Astrophysical Journal* 877.2, p. 121. DOI: 10.3847/1538-4357/ab1b3c. URL: https://doi.org/10.3847%2F1538-4357%2Fab1b3c.