Predicting Solar Proton Events of Solar Cycles 22-24

Aatiya Ali[1], Viacheslav Sadykov[1], Alexander Kosovichev[2] ,
Alin Paraschiv[3], Sarah Gibson[3]

[1]Georgia State University, [2]New Jersey Institute of Technology, [3]High Altitude Observatory

Georgia State University

INTRODUCTION

- Solar energetic particles (SEPs) propagating through the heliosphere and interactions with Earth's magnetosphere and atmosphere are damaging to numerous aspects of life and technology, including astronauts' health, satellites orbiting the Earth, etc.
- Our efforts conclude building catalogs using data from Solar Cycles (SCs) 22-24 :
  (1) >10 MeV >10 pfu Solar Proton Events (SPEs)
  (2) daily statistics of proton & soft X-ray (SXR) data

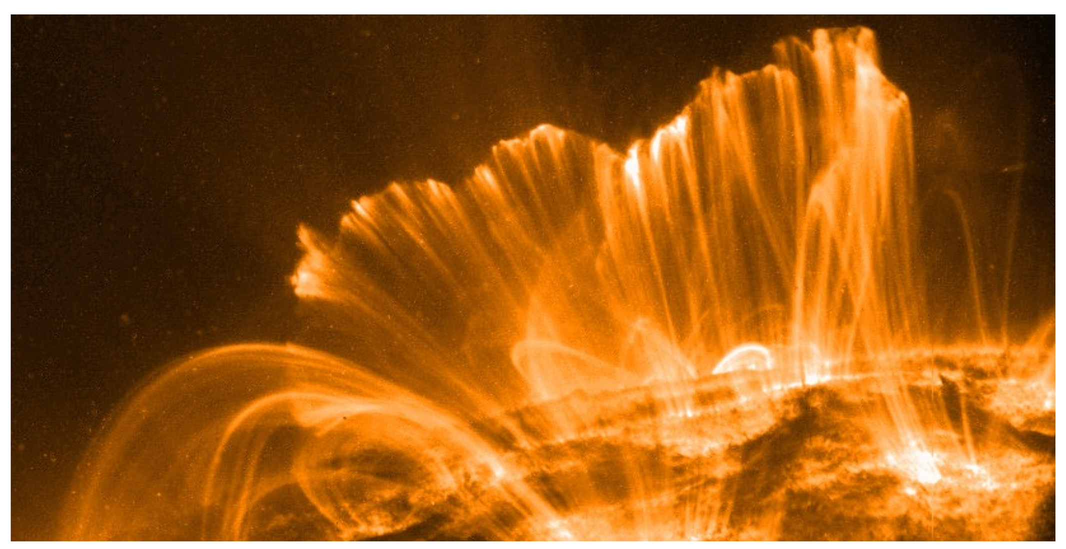

Figure 1: Post-eruptive loops in the wake of SEPs propagated during a solar flare (Photo Credits: NASA).

- Using data from previous SCs available in operational regimes may provide a more robust (validated on longer timescales) prediction model that can be operationalized.

MOTIVATION

- The rare nature of SPEs makes it crucial to consider data from long timescales to develop the ability to reliably predict them.
- Our previous study (*Sadykov et al. 2021*[1]) highlights proton and SXR parameter significance when constructing an "all-clear" forecast.
- The Sun's dynamic nature complicates the standardization of input for reliable predictions.

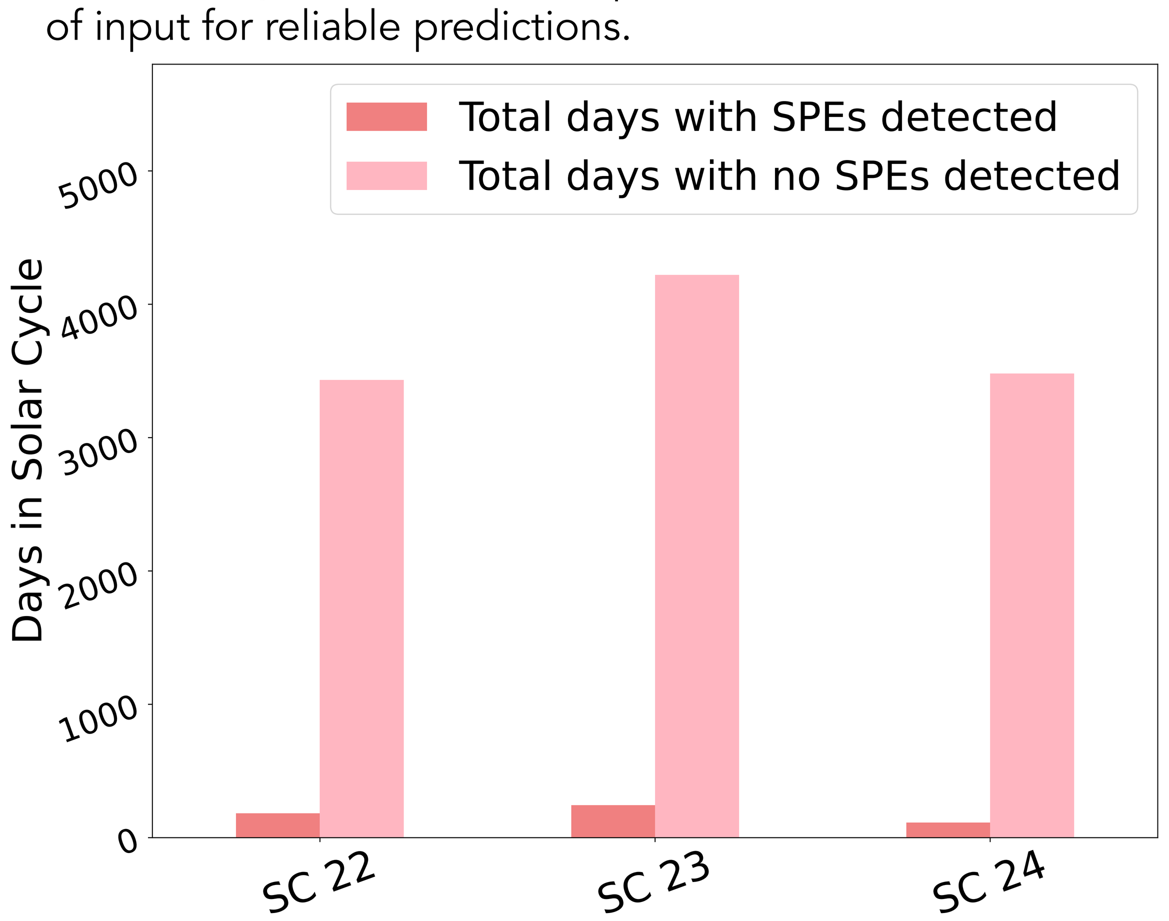

Figure 2: The drastic imbalance in number of days SPEs are observed per SC, compared to days they are not.

PROCESSING DATA

- Predictions are done using proton & SXR data from SCs 22-24 taken by Geostationary Operational Environmental Satellites launched by the NOAA.
- >10 MeV >10 pfu SPEs were identified and recorded using proton flux data. This first catalog was successfully cross-checked with that produced by NOAA's Space Environment Center for validation.

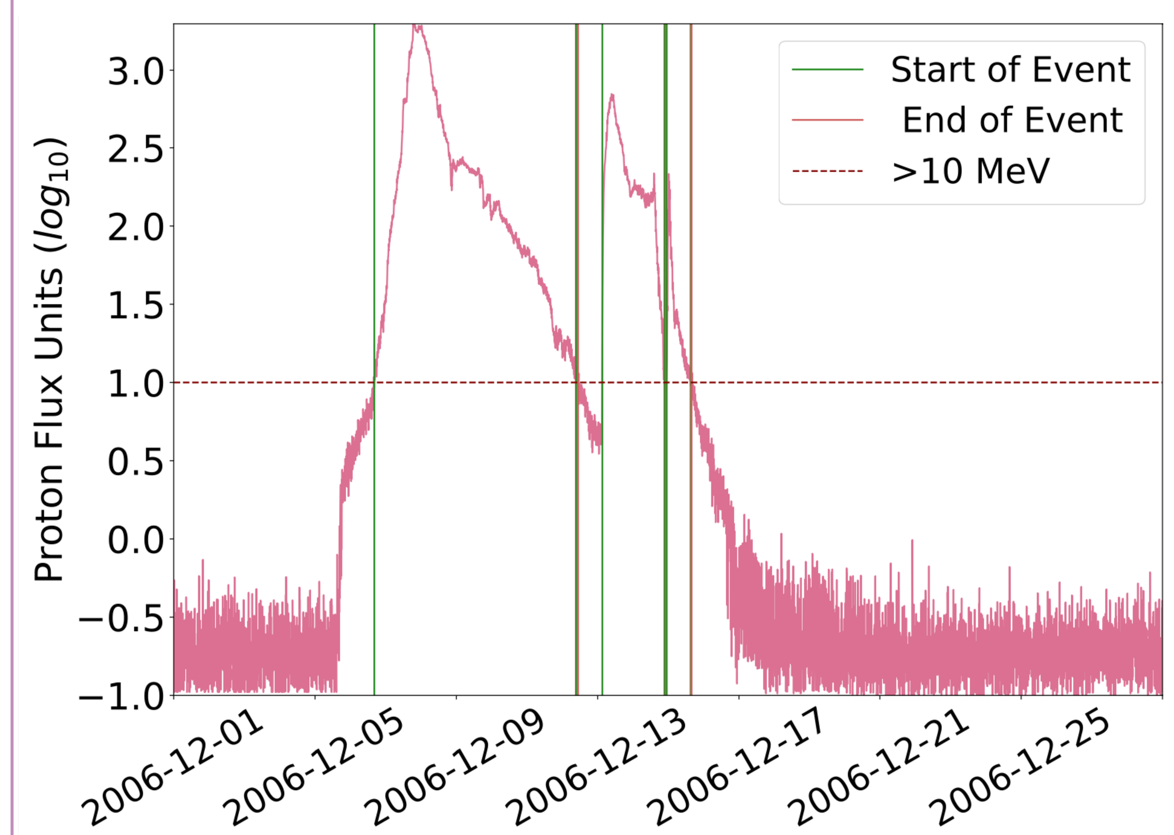

Figure 3: Example of the classification of a Solar proton event recorded in Catalog 1.

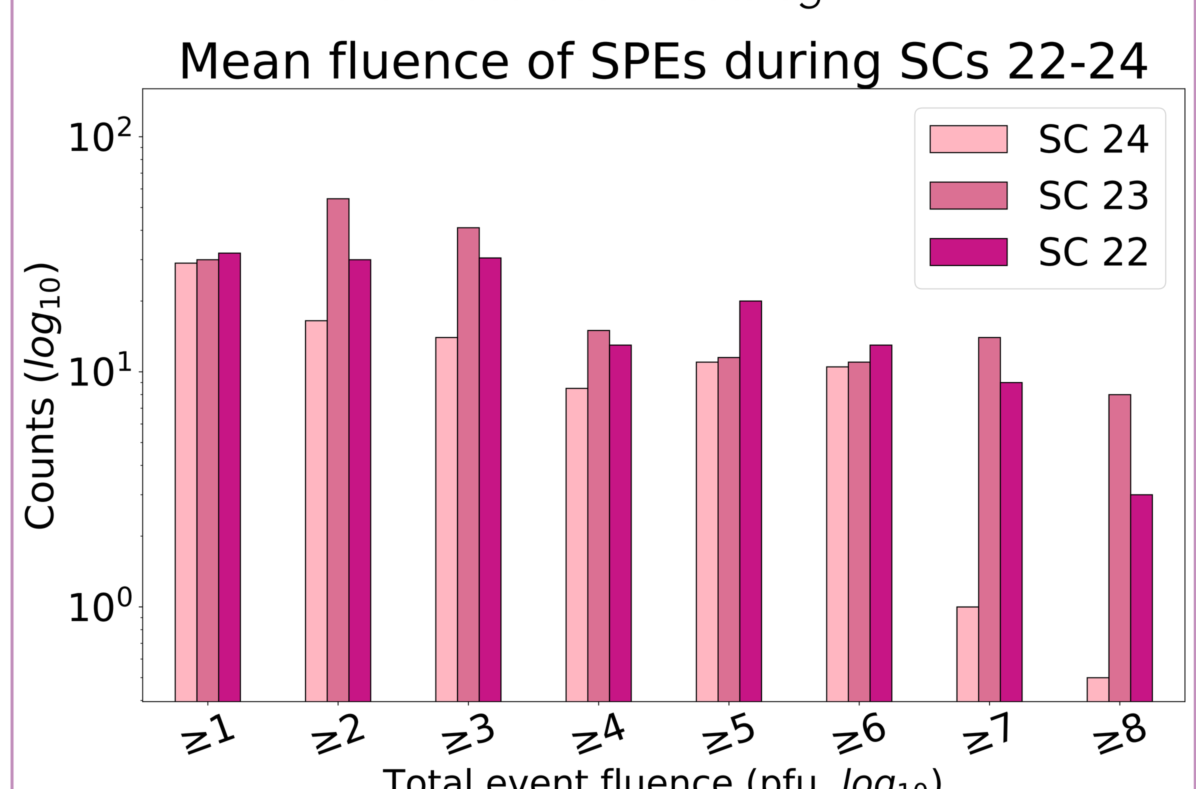

Figure 4: Example of a data product created using Catalog 1.

- Revised SXR data according to NOAA scaling corrections, we built a second catalog presenting daily SXR and proton flux statistics (mean, median, kurtosis, etc).
- Data recorded during overlapping observations are compared, and the instrument recording higher fluxes is considered the *primary* for a given date. The resulting dataset is merged, providing a continuous timeseries.
- Proton and SXR flux data, our resultant catalogs, and visualizations have been fetched to a database[2] organized as a part of the NASA ESI project.

PREDICTION EFFORTS

- Catalog 1 supplies daily feature vectors indicating if a SPE is in progress at any time the next day. Catalog 2 is used to determine features important (Fig. 5) for forecasting, reducing computational costs and predictive inaccuracy when accounting for irrelevant data.

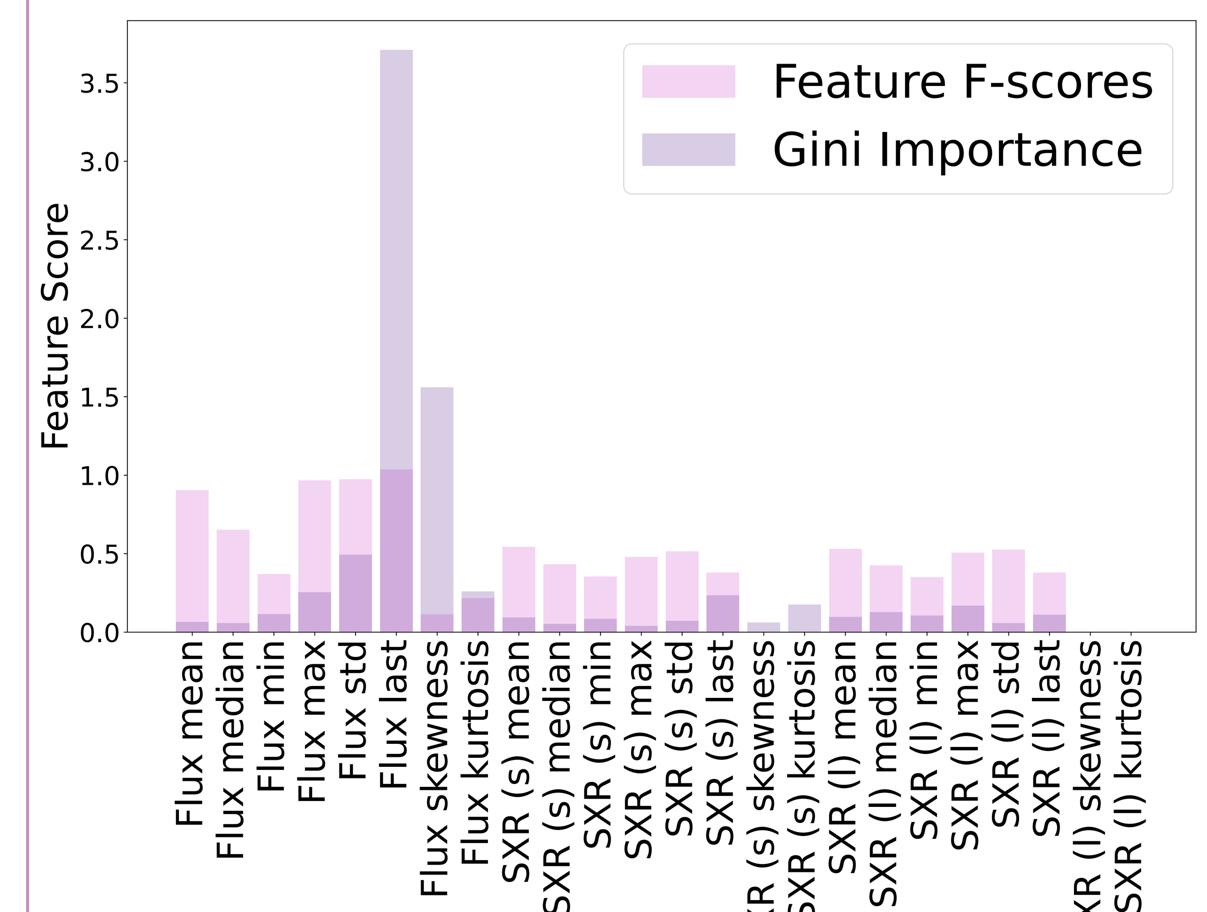

*Figure 5: Feature selection for oversampled SC 23 using F-score and Gini Importance (*8 for clarity) parameters.*

- Using the 12 highest-scoring features per SC, we use a Support Vector Machine (SVM) for predictions using *balanced* weights, and a *radial basis function* kernel.
- Prediction success is measured using True Skill Statistic (TSS) and Heidke Skill Scores (HSS_2):

	SC22 (Training cycle)	SC23 (Training cycle)	SC24 (Training cycle)	Training on 2 Remaining Cycles
SC22 (Testing cycle)		0.74	0.66	0.73
SC23 (Testing cycle)	0.74		0.45	0.76
SC24 (Testing cycle)	0.69	0.66		0.70

Table 1: TSS scores for a RBF SVM.

	SC22 (Training cycle)	SC23 (Training cycle)	SC24 (Training cycle)	Training on 2 Remaining Cycles
SC22 (Testing cycle)		0.52	0.22	0.38
SC23 (Testing cycle)	0.68		0.09	0.51
SC24 (Testing cycle)	0.52	0.54		0.56

Table 2: HSS_2 for a RBF SVM.

- We compare this method to an oversampled dataset, with no substantial differences seen feature selection ordering or predictive scores.

RESULTS

- Prediction accuracy loosely favors longer timescales (training on two SCs instead of one), resulting in slightly enhanced TSS and HSS_2 results across all SCs.
- Prediction accuracy also loosely favors oversampling, with TSS and HSS_2 changes on a range of ~1-5 %. However, an outlier in the oversampled HSS_2 exists with a decrease of ~34% yet to be explored.
- Predictive scores averaging around 70% suggest that proton and SXR flux data alone may not being sufficient to produce reliable forecasts.

FUTURE WORK

- We plan to expand our analysis of SEP prediction by including Hale and McIntosh class properties of Solar active regions (ARs) in our forecasting model.
- We aim to explore transitory variations of coronal plasma and magnetic field effects on SEP-quiet and SEP-active regions. Any relations revealed here will be explored.
- SolarSoft package *FORWARD* provides us with coronal observations and magnetohydrodynamic models. It allows us to model coronal distributions for density, temperature, magnetic configurations, etc. provided by PSImas.

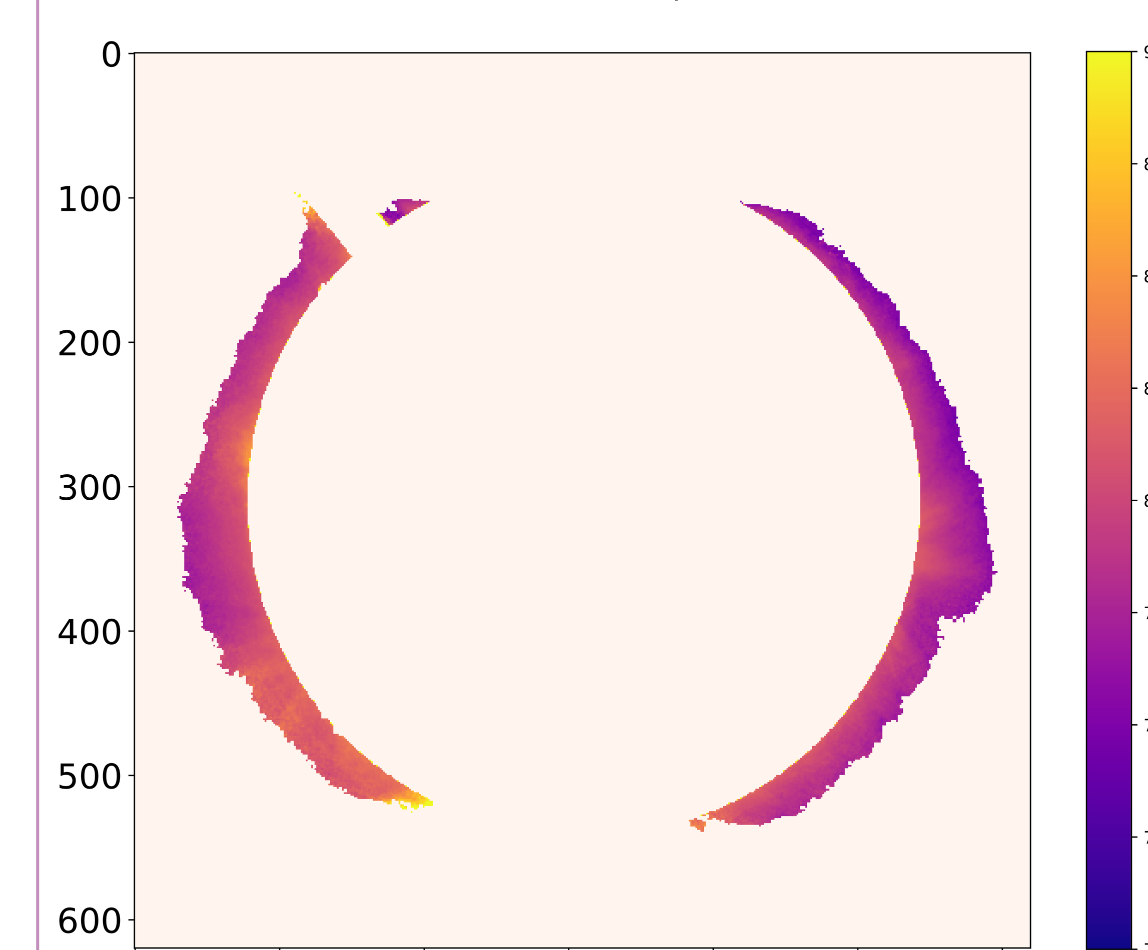

Figure 6: Example of a FORWARD-generated density map.

ACKNOWLEDGEMENTS & REFERENCES

This research is supported by NASA ESI program grant 80NSSC20K0302, NSF FDSS grant 1936361, & NSF EarthCube grant 1639683.

[1] Sadykov et al. 2021, "Prediction of Solar Proton Events with Machine Learning: Comparison with Operational Forecasts and "All-Clear" Perspectives"

[2] *SEP Prediction Portal (SEP[3])* https://sun.njit.edu/SEP3/contact.html