Draft version March 10, 2023
Typeset using LaTeX default style in AASTeX63

Predicting Solar Proton Events of Solar Cycles 22-24 using GOES Proton & Soft X-Ray Flux Statistics

Aatiya Ali,[1] Viacheslav Sadykov,[1] Alexander Kosovichev,[2,3] Irina N. Kitiashvili,[3] Vincent Oria,[4] Gelu M. Nita,[2] Egor Illarionov,[5,6] Patrick M. O'Keefe,[4] Fraila Francis,[4] Chun-Jie Chong,[4] Paul Kosovich,[2] and Russell D. Marroquin[7]

[1] *Physics & Astronomy Department, Georgia State University, Atlanta, GA 30303, USA*
[2] *Physics Department, New Jersey Institute of Technology, Newark, NJ 07102, USA*
[3] *NASA Ames Research Center, Moffett Field, CA 94035, USA*
[4] *Computer Science Department, New Jersey Institute of Technology, Newark, NJ 07102, USA*
[5] *Department of Mechanics and Mathematics, Moscow State University, Moscow, 119991, Russia*
[6] *Moscow Center of Fundamental and Applied Mathematics, Moscow, 119234, Russia*
[7] *Department of Physics, University of California San Diego, La Jolla, CA 92093, USA*

ABSTRACT

Solar Energetic Particle (SEP) events and their major subclass, Solar Proton Events (SPEs), can result in unfavorable consequences to numerous aspects of life and technology, making them one of the most prevalent and harmful effects of solar activity. Garnering knowledge leading up to such events by studying proton and soft X-ray (SXR) flux data to alleviate the burdens they cause is therefore critical for their forecasting. Our previous SEP prediction study (Sadykov et al. 2021) indicated that it may be sufficient to utilize only proton and SXR parameters for SPE forecasts considering a limited data set from Solar Cycle (SC) 24. In this work we report the completion of a catalog of ≥10 MeV ≥10 particle flux unit (pfu) SPEs observed by Geostationary Operational Environmental Satellite (GOES) detectors operated by the National Oceanic and Atmospheric Administration (NOAA), with records of their properties spanning through SCs 22-24. We report an additional catalog of daily proton and SXR flux statistics. We use these catalogs to test the application of machine learning (ML) for the prediction of SPEs using a Support Vector Machine (SVM) algorithm. We explore how previous SCs can train and test on each other using both earlier and longer data sets during the training phase, evaluating how transferable an algorithm is across different time periods. Validation against the effects of cross-cycle transferability is an understudied area in SEP research, but should be considered for verifying the cross-cycle robustness of an ML-driven forecast.

Keywords: Solar energetic particles (1491) — Space weather (2037) — Solar Cycle (1487)

1. INTRODUCTION

Solar Energetic Particle (SEP) events are enhanced fluxes of high-energy particles ejected by the Sun. The occurrence rates of such events are greatest during the maxima of ∼11-year Solar Cycles (SCs). Solar Proton Events (SPEs), a subclass of SEPs, encompass a wide range of energies from KeVs up to multiple GeVs (Anastasiadis et al. 2019), ejected into the heliosphere. SPEs are defined by a threshold of proton detections ≥10 MeV energy, while also exceeding 10 particle flux unit (pfu) counts. Energetic protons can harm satellites, navigation, and communication systems, as well as technological grids and other equipment. High-energy charged particles in the magnetosphere can also manipulate output signals of electronic devices. Not only causing spacecrafts' built-in atmospheric effect calibration systems to fail, when these charged particles strike a critical device, the instrument may fail entirely. High-energy electrons during geomagnetic storms add further complications as they are able to penetrate shielding aboard satellites and spacecraft

Corresponding author: Aatiya Ali
aali87@student.gsu.edu

as they pile up in short amounts of time; leading to discharging accumulated energy and an eventual small-scale 'lightning strike'. Radio-wave dependent communication systems are also vulnerable to such events: changes in the ionosphere alter how radio waves travel, and ultimately delay their arrival time from satellites to ground-based global positioning systems (GPS), mislabeling positions by a few meters. While seemingly insignificant, this poses issues for aviation, robotics, military, transportation, and other industries' operations (Metha 2022).

SPEs also motivate research to pursue an additional understanding of the physics behind solar processes related to their generation. Further, by enhancing the levels of radiation in interplanetary space (Langford 2022), SPEs may also be responsible for the atrophy of astronaut health. Onorato et al. (2020) expand on this, highlighting the risks to astronauts developing cancer, experiencing central nervous system decrements, and even exhibiting degenerative tissue effects. More recently, such concerns have caught the attention of administrations dealing with commercial airlines and space tourism. Studies (e.g., Beck et al. 2005) conclude that combined effects of magnetic field disturbances and solar particle fluence due to solar storms can show up to a 70% variation in radiation exposure at typical flight altitudes. Collins (2006) also notes the main difference in safety during space tourism between lunar and orbital travel being the higher exposure to cosmic radiation; which would only be enhanced during the propagation of SEPs. Further, Naito et al. (2020) conclude that composite materials (i.e. carbon fiber reinforced plastic, silicon-carbon plastic) are promising for spacecraft shielding considering the factors of resistance to shock and heat, and reduced fragmentation cross-sections. Still, these are only promising materials and none have thus far been discovered to completely withstand the variations in a radiative environment one would expect during SPEs.

Given the multitude of hazards presented by these events, it is critical to develop reliable forecasting capabilities to predict and allow a sufficient period of time for astronauts and even equipment used to conduct scientific experiments to be brought to safety. External SEP detectors may be mounted aboard such spacecraft to act as warning and defense systems, coupled with an appropriate predictive algorithm. **Our overarching goal is to explore a predictive algorithm's performance using only proton and soft X-Ray (SXR) flux data, and assessing variations when considering timescales longer than a single SC during the training phase.**

1.1. *The problem of Solar Proton Event prediction*

Approaches to building predictive models for SPEs become arduous given their rare nature and reliance on indirect measures of volatile SC activity. If using measures like forecasting accuracy to quantify the success of a predictive model, the rarity of SPEs permits an algorithm to be able to miss most, if not all, incoming SPEs while still sustaining high-scoring prediction rates. The need for balanced data sets instead to reflect accurate scores is emphasized by Martens & Angryk (2017). More specifically, the class imbalance in our data set shows 11946 negative cases (days with no SPEs), in vast contrast to only 545 positive cases (days with SPEs). This imbalance in positive cases overshadowed by the prevalence of quiet periods remains the overarching problem encountered in this work, and the disparity is visualized in Fig. 1. **Variance in the Sun's global magnetic activity and consequential changing event frequency across SCs brings forward the question of how transferable an algorithm built based on previous SCs is when considering future SCs of unknown activity levels.**

1.2. *Current & previous results and limitations*

In recent years, there has been a plethora of research projects contributing to the effort of predicting SPEs in an attempt to mitigate their detrimental effects. In our previous study (Sadykov et al. 2021) we use SXR wavelength ranges (long (0.1 - 0.8 nm) and short (0.05 - 0.4 nm)), along with $\geq$ 10 MeV proton flux data observed by the Geostationary Operational Environmental Satellite (GOES) series. From the various products obtained by GOES, we retrieve and use SXR flux data with 1-minute cadences, and $\geq$ 10 MeV proton fluxes with a 5-minute cadence. The data has been made publicly available by the National Oceanic & Atmospheric Association (NOAA) National Center for Environmental Information (NCEI [1]). The success of utilizing derivatives of these data products alone- specifically an event's preceding proton flux, presents the most valuable input for prediction, allowing us to continue exploring these parameters in depth in this work. Sadykov et al. (2021) also discuss the lack of performance loss when excluding daily regional, ground, and space-based radio characteristics of Active Regions (ARs) and type II, III, IV radio bursts when generating predictive algorithms, although acknowledging the brevity of the considered data set. Proton and SXR flux characteristics alone were comparable with predictions based on the addition of both AR characteristics, and

[1] www.ncei.noaa.gov/data/goes-space-environment-monitor/access/avg/

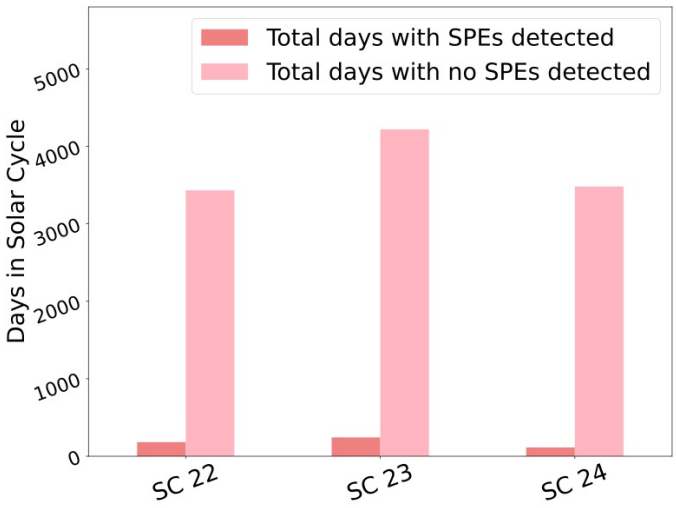

Figure 1. The drastic imbalance in days where SPEs are observed per SC, compared to days they are not observed.
Note: SC 22 is dated from September, 1986 to August 1996, SC 23 from August, 1996 to December, 2008, and SC 24 from December, 2008 to December, 2019.*

the inclusion of radio burst counts. When proton flux data alone was compared with the addition of SXR data as the second most valuable parameter, it proved to enhance prediction accuracy. Our goal here is therefore to emphasize the features of proton and SXR fluxes most important towards building a predictive algorithm.

However, in addition to these parameters, different approaches to SPE prediction include physics-based and empirical models that take into account parameters of solar magnetograms, optical imaging, extreme ultraviolet (EUV) imaging, CMEs from single or multiple vantage points, in situ energetic proton and electron observations, particle acceleration and transport, and in situ measurements of solar wind density, temperature, and magnetic fields (Whitman et al. 2022). Most of these empirical models are vastly computationally expensive, limiting their implementation into current forecasting workflows as "real time" predictions; making it crucial to minimize the computational intensity required for such algorithms- which is what we seek to do when employing only certain features of our data sets into our machine-learning (ML) algorithm. This is further discussed in Section 3.2.

Sadykov et al. (2021) also discuss how GOES' recently deployed Energetic Proton, Electron and Alpha Detector (EPEAD) East and West detectors capture different populations of protons given their orientation and alignment with the interplanetary magnetic field structures (Parker 1958). He & Rodriguez (2018) further discuss this, concluding that differences in proton flux measurements between the different GOES detectors (when applicable) are due to the effect of magnetic field configuration variation with geomagnetic longitude. These studies together lead to the fact that simply taking the average of these detectors' fluxes is a misrepresentation of SEP propagation towards the geostationary orbit, leading us to take into account the effects of the Parker spiral during the detection of particles hurling towards GOES. We therefore must consider the data obtained in the EPEAD series separately from GOES' earlier Energetic Particle Sensor (EPS) detections, as discussed in Section 2.1.

Studies comparing results between ML-driven models with those used as daily operational forecasts by NOAA's Space Weather Prediction Center (SWPC) for SPEs may identify combinations of observed parameters preceding SPEs that could improve such forecasts (Sadykov et al. 2021). To achieve this, we aggregate a catalog of statistical parameters (discussed in Section 3.2), which are minimized by meticulously selecting features per SC to find combinations of flux statistics relevant to predicting SEP propagation during unpredictable levels of solar activity across proton, and short and long SXR irradiance channels. Additionally, Whitman et al. (2022) expand on a number of current SEP prediction efforts discussing typical model inputs, the employment of empirical, physics-based, ML-based, and Multi-Module Forecasting Systems, and elaborates on the model validation diagnostics utilized per project. In parallel to this, fundamental parameters required by the model may be poorly characterized without a complete understanding of the underlying physics associated with SEP ejection and acceleration. SEP models have thus been motivated by both research aimed at understanding the physical processes related to SEPs and by operational forecasting needs. These current complex models show promise of modeling the complete distribution of SEP events with time in 2- or

3-dimensional space. ML approaches are therefore still being investigated in the hope of yielding a new class of SEP models to produce fast, reliable forecasts (Whitman et al. 2022). Our work here reflects a different approach to the problem.

1.3. *Scope of this paper*

Our work in this project focuses on the analysis of an ML-driven algorithm for the prediction of SPEs using previous SCs (22-24) to build an understanding of cross-cycle transferability of a preestablished model. We explore this using long-term data acquired using only NOAA's GOES series for consistency. Starting with detecting SPEs from these SCs using the GOES series, event parameters are stored in the first catalog built in this effort. In summary, the catalog includes:

- Dates for: Start of an event, an event's peak flux detection, end of an event
- Peak event fluxes detected in energy channels: 1 MeV, 5 MeV, 10 MeV, 30 MeV, 50 MeV, 60 MeV, 100 MeV
- Fluence of SPEs in energy channels: 1 MeV, 5 MeV, 10 MeV, 30 MeV, 50 MeV, 60 MeV, 100 MeV

Using both data during SPEs and 'quiet' periods, we form an additional catalog of daily statistical features of SXR and $\geq$10 MeV proton fluxes that works as the input for our Support Vector Machine (SVM) algorithm.

We begin by applying default arguments (discussed in Section 3) to our SVM across our available data in different training and testing splits (discussed in Section 3.1), then use Gini Importance and Fisher-score standards to explore which supplied statistical features are of significant importance when working towards predictions. The success of the algorithm is then measured using True Skill Statistic (TSS) and Heidke Skill Scores (HSS_2); the latter of which was developed by Bobra & Couvidat (2015). Analyzing different arguments associated with a SVM, we attempt to standardize our algorithm using only a *radial basis function (RBF)* kernel to get our forecasts as accurate as possible. Raschka & Mirjalili (2017) indicate that if complete model selection using the *RBF* kernel has been conducted in isolation, that there is no need to consider *linear* SVM. This is further discussed in Section 3.

2. GOES DATA PREPARATION AND PRODUCTS

2.1. *Querying Data*

Our prediction algorithm is built using $\geq$10 MeV proton and SXR flux data queried from NOAA's NCEI through years 1986-2019 to encompass SCs 22-24. The data are obtained by different GOES launched by NOAA as a series from GOES-05 to GOES-15 during the period of interest. The primary operational purpose of GOES is to support forecasting operations, providing real-time access to X-ray flux and proton flux measurements (Aminalragia-Giamini et al. 2021) from the geostationary orbit- an altitude of $\sim$36,000 km above Earth's equator; making this instrumental series a clear candidate for data collection for the purposes of this project. Before NOAA's primary and secondary scheme[2] (a configuration where at each time, multiple satellites in the GOES series provided real-time data where one was usually assigned to be primary and others as secondary) was compiled in 1974, Aminalragia-Giamini et al. (2021) retained measurements from the highest numbered GOES satellite. Since 1986- the beginning of SC 22 and therefore the start of our period of interest, marks the initiation of individually considering operational satellite data.

GOES undergo a change from a mounted EPS to an EPEAD in 2011 starting with GOES-13. The satellite series operates from two primary locations: the East detector at 75.2° W, and the West detector at 137.2° W. Additionally, NOAA maintains an on-orbit spare satellite in the event of mechanical failure or instrumental errors of either detector. GOES observations often overlap one another, calling for a method to streamline the timeline of data available to analyze, which was done by manually choosing a 'primary' instrument for every month across SCs 22-24. By our definition, a primary instrument refers to that reflecting a higher proton flux count when two or more satellites are capturing data at the same time. When EPEAD satellite data is introduced into our data set, both primary instruments and primary detectors (labeled as East or West) are selected and recorded for use. This ordering of confidence in acquired data was influenced by the work presented by Rotti et al. (2022) exploring integral proton flux data intensity profiles for space weather predictions also using GOES data. The primary instrument selection used for SCs 22-24 is shown in Fig. 2, and a visualization comparing proton populations captured by the primary versus

[2] www.ngdc.noaa.gov/stp/satellite/goes/doc/GOES_XRS_readme.pdf

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1986	G6	G6	G6	G6	G6	G6	G6	G6	G6	G6	G6	G6
1987	G6	G6	G7	G7	G7	G6	G6	G6	G6	G6	G7	G7
1988	G7	G6	G7	G6	G6	G7	G7	G6	G7	G7	G7	G7
1989	G7	G6	G6	G7	G7	G7	G6	G7	G6	G7	G7	G7
1990	G6	G6	G7	G7	G7	G7	G7	G7	G7	G7	G7	G6
1991	G7	G7	G7	G7	G7	G7	G7	G7	G7	G7	G7	G7
1992	G7	G7	G7	G7	G7	G7	G7	G7	G7	G7	G7	G6
1993	G6	G7	G7	G7	G7	G7	G7	G7	G6	G7	G7	G7
1994	G7	G7	G7	G7	G7	G7	G7	G7	G6	G7	G7	G7
1995	G8	G8	G8	G7	G7	G7	G7	G7	G7	G8	G8	G7
1996	G8	G8	G8	G8	G8	G8	G8	G8	G8	G8	G8	G8
1997	G8	G8	G8	G8	G8	G8	G8	G8	G8	G8	G8	G8
1998	G8	G8	G8	G8	G8	G8	G8	G10	G10	G8	G8	G10
1999	G8	G8	G8	G8	G8	G8	G8	G8	G8	G8	G8	G8
2000	G8	G8	G8	G8	G8	G8	G8	G8	G8	G8	G8	G8
2001	G8	G8	G8	G8	G8	G8	G10	G10	G10	G10	G10	G10
2002	G10	G10	G10	G8	G8	G8	G10	G10	G10	G10	G10	G10
2003	G10	G10	G10	G10	G10	G10	G11	G11	G11	G11	G11	G11
2004	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11
2005	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11
2006	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11
2007	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11
2008	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11
2009	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11
2010	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11	G11
2011	G11	G11	G15 E	G15 W	G15 W	G15 W	G15 W	G15 W	G13 E	G13 E	G15 E	G15 E
2012	G15 E	G15 E	G15 E	G13 W	G15 W	G15 W	G15 W	G15 W	G15 W	G13 E	G15 E	G15 E
2013	G15 E	G15 E	G15 E	G13 W	G15 W	G15 W	G15 W	G15 W	G13 W	G15 E	G13 E	G13 E
2014	G13 E	G13 E	G13 E	G15 W	G15 W	G15 W	G15 W	G15 W	G15 W	G15 E	G15 E	G15 E
2015	G13 E	G13 E	G15 E	G15 W	G15 W	G13 W	G13 W	G13 W	G13 W	G15 E	G13 E	G13 E
2016	G15 E	G15 E	G15 E	G15 W	G15 W	G15 W	G15 W	G15 W	G15 W	G15 E	G15 E	G15 E
2017	G15 E	G15 E	G15 E	G15 W	G15 W	G15 W	G13 W	G15 W	G15 W	G15 E	G15 E	G15 E
2018	G15 E	G15 E	G15 E	G15 W	G15 W	G15 W	G15 W	G15 W	G15 W	G15 E	G15 E	G15 E
2019	G15 E	G15 E	G15 E	G15 W	G15 W	G15 W	G15 W	G15 W	G15 W	G15 E	G15 E	G15 E

Figure 2. Timeline of primary GOES instruments and detectors (when applicable) used to streamline data used for SCs 22-24. *G* simply represents 'GOES', with succeeding numbers representing the satellite of interest from the GOES series, and *E* and *W* indicating East or West detectors having detected the higher proton flux.

secondary detectors is presented in Fig. 3, reflecting the minimal- but still distinct differences in proton counts each detector records.

It is also important to note changes in GOES directionality after launch for those interested in recording instruments/detectors with no data gaps, and also to account for the detectors capturing the largest SEP populations. As specified by Rodriguez (2012) and the National Geophysical Data Center, GOES 13, 14 & 15 all undergo a yaw flip, where the satellite rotates about the axis pointed toward the center of the Earth. The EPEAD telemetry channels labeled 'East' in the data files look westward when the spacecraft is upright, and eastward when the spacecraft is inverted. Likewise, the EPEAD telemetry channels labeled 'West' look eastward when the spacecraft is upright and westward when the spacecraft is inverted. The inversions in directionality over time are as follows:

- GOES-13 (2006-2018): only upright during its operational period between May 2010 and September 2012.
- GOES-14 (2009-2020): upright from its launch date in 2009, and inverted during a SPE in early September 2012. The satellite has not corrected itself ever since.[3]
- GOES-15 (2010-2018): experiences a flip twice a year at every equinox. This maneuver usually lasts under an hour; during which data is not recorded.

These inversions are considered for each satellite by their specific guidelines, and orientation labels are corrected for in the catalogs presented in Sections 2.2 and 2.3 for our prediction efforts. During yaw flips and days where proton and SXR flux data were not recorded (during orientation changes, instrumental corrections, etc.), we interpolate fluxes from the time before and after the data gap in order to attain continuous flux records for the timeline of interest. Proton and SXR flux data, as well as our resultant catalogs and visuals of the timeseries' are fetched onto a database[4] organized as a part of the NASA ESI project.

2.2. *Generating the SPE Catalog*

Considering the definition of an SPE as an observation of a $\geq$ 10 MeV $\geq$10 pfu event, this is the threshold used throughout the process of generating the first catalog resulting from this work. This threshold being the baseline of a SPE, defines the weakest ranking storm in the NOAA S-scale[5] hierarchy of progressively damaging solar events spanning from S1 (minor) to S5 (extreme). Specifically, the foundational limits placed on what can be considered an SPE qualify as only an S1 event- the weakest of the S-scales, affecting only high-frequency radio propagation in the polar regions. Higher scales (S2, S3, S4, & S5) are associated with events of much higher pfus (10^2, 10^3, 10^4 & 10^5 respectively). Events qualifying higher than an S1 ($\geq$ 10 pfu) on the scale are then essentially subsets of S1, and by predicting the most standard SPEs we are also predicting more harmful events. The scaling of S1 through S5 also reflects the increasing intensity of damage these solar storms cause on biological and instrumental functions on both Earth and through interplanetary space.

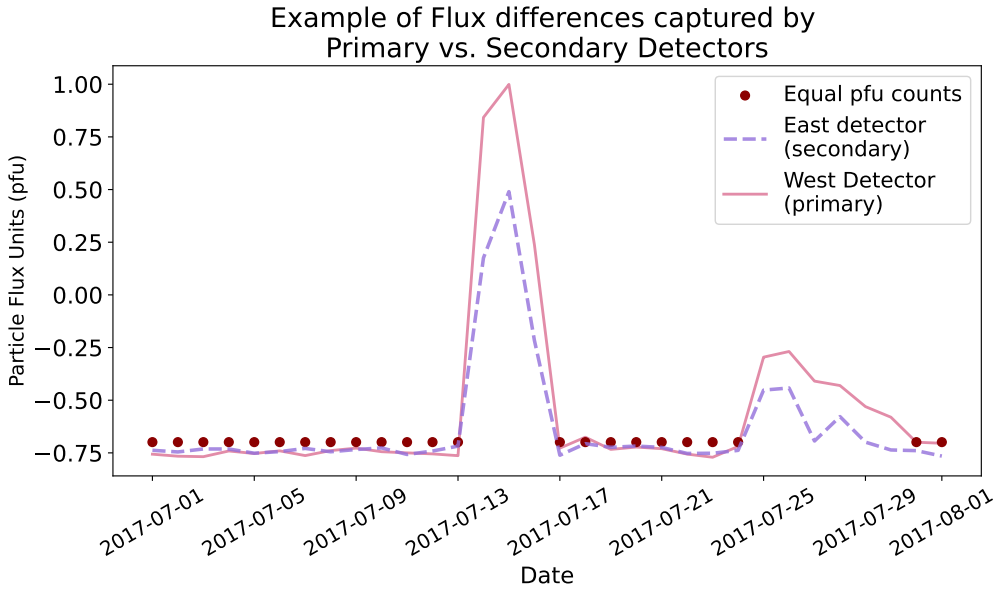

Figure 3. An example of the stark difference in the daily median flux values captured by the East and West detectors of GOES-13. The consistently higher fluxes registered by the West detector compared to the East during the same observation period allows it to be considered the primary instrument for the month. *Note, fluxes rounded to the nearest decimal point for readability.*

[3] Powered off in 2020, this satellite can be called back into service if needed (NOAA 2019).

[4] sun.njit.edu/SEP3/datasets.html

[5] www.swpc.noaa.gov/noaa-scales-explanation

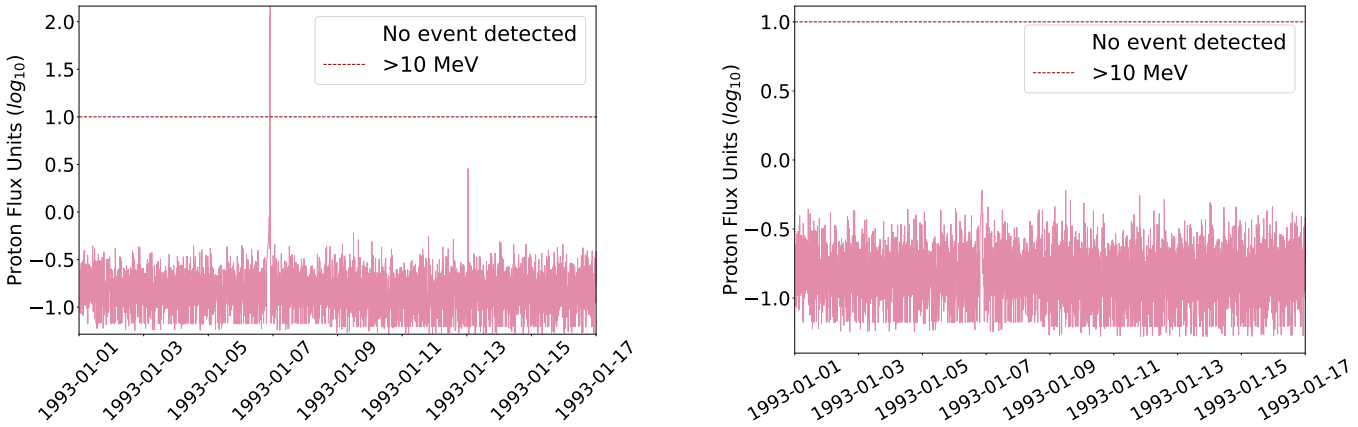

Figure 4. Examples of (left) instrumental effects producing a spike in the data, and (right) how a corrected spike is presented, mirroring a quiet period of the Sun.

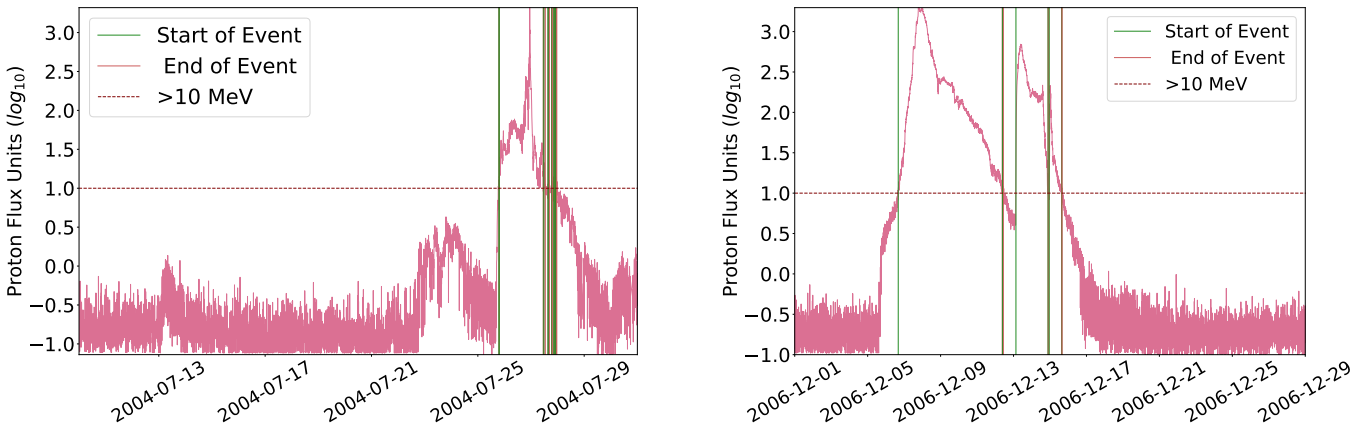

Figure 5. Examples of (left) oscillations around the 10 MeV threshold during an event, and (right) a clearer representation of a SPE, as they are recorded in our catalog.

To build our catalog of SPE statistics, GOES data was cleaned for instrumental effects prior to recording any event parameters- most of which present themselves as a spike in the data, an example of which is shown in the left panel of Fig. 4. These spikes are corrected for by taking the mean value of the entire month's flux in *log scale*, and simply replacing the heightened flux value correspondent to the spike, as seen in the right panel of Fig. 4. This allows for the quiet period of the Sun to be presented more accurately, and for the avoidance of recording a SPE when the falsely amplified flux value crosses the $\geq$10 pfu threshold. Further, given that the threshold can be repeatedly crossed during an event- left panel of Fig. 5 being an example of this, these oscillations are handled by checking if the time between an ongoing event's end and the start of the next is within 30 minutes of each other. We decided on this arbitrary length of time, given our familiarity with general SPE properties. Given that oscillations less than half an hour apart are presented in rapid succession and recorded as multiple events over a short period of time, they are instead stitched together as one event in the resulting catalog. These quick oscillations may truly represent proton flux counts, or are side effects of flux detections made by the instrument. Opposing these situations, a more prevalent example of a SPE usually has a more gradual rise and/or fall. The end product of this part of data processing is shown in the right panel of Fig. 5, reflecting clear start and end dates of a SPE, barring any undesirable flux representations. Spikes and oscillations encountered when analyzing GOES data are therefore corrected by taking the aforementioned steps, and the cleaned data set is used for developing our first catalog, and for the remainder of the project itself.

As a summary of the catalog, we record observed SPE instruments, detectors (when applicable), start date & time, peak flux count date & time, and end date & time. Lastly, fluences are calculated as the integral of flux detected during

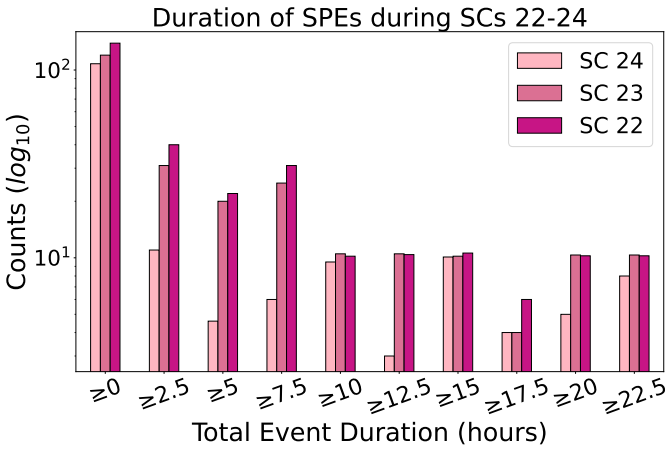

Figure 6. SPE event duration statistics across all SCs.

Averaged SPE Statistics	**SC 22**	**SC 23**	**SC 24**
Duration (∼hh:mm)	11:19 ± 25:22	10:31 ± 23:50	10:47 ± 22:16
Peak Flux (∼pfu)	432 ±3147	544 ±3245	156 ± 729
Fluence (∼10^4 pfu)	3 ± 19	7 ± 42	3 ± 13
Days with Observed Events	182	243	113
Duration > 24 hours	42	77	32
Previous day's flux ≥10 pfu	5	10	4

Table 1. SPE statistics considering proton flux data. We summarize (and include the standard deviation when applicable), the average duration, peak flux, and fluence per event for each SC, among other characteristics. We can see here how much weaker SC 24 activity was compared to SCs 22 & 23.

the entirety of an event. These are documented for energy channels : 1, 5, 10, 30, 50, 60, and 100 MeV. Examples of data products derived from this catalog are shown in Table 1 and Figure 6.

2.3. *Generating the Daily Feature Statistical Catalog*

The second catalog produced during this project supplies the input data for our ML forecasting algorithm. This consists of numerous features of daily flux data acquired by GOES in both proton and SXR channels. The ranges of all feature vectors of a data set are typically transformed into a unified range before the training phase starts. Optimization in SVMs occur by minimizing the decision vector. The classifier's optimal hyperplane is influenced by the scale of input features, requiring data to be logarithmically scaled prior to model training. In accordance to this, we apply a mean-standard deviation normalization transformation to the entire data set. This is primarily done to allow our models to distinguish between patterns and structures independent of the physical units and dynamical ranges of each parameter (Ahmadzadeh et al. 2021). Namely, the features generated for the catalog are (for each-proton, SXR short wavelength, and SXR long wavelength channels):

- Instrumental data[6]: Date of observation, EPS/EPEAD satellite used (with primary East or West labels when appropriate)
- Daily aggregated flux data & statistics: mean, median, minimum, maximum, standard deviation, last measured flux of the day, skewness, and kurtosis

[6] This is added for timeseries records only, and not used in the forecast itself

Both generated catalogs are publicly available at the Solar Energetic Particle Prediction Portal webpage (SEP[3]) page: https://sun.njit.edu/SEP3/datasets.html

3. ML APPLICATION TO SPE FORECASTING

Introduced in 1963, SVMs are supervised classifiers with roots in the theory of statistical learning, with the ability to learn from nonlinear decision surfaces with the application of different kernels. In ML, hyperplanes are defined as decision boundaries that help classify data points. Data points falling on either side of the hyperplane can be attributed to different classes- features relevant and irrelevant to the forecasting of SPEs in our case (Bishop 2006). Black-box machines like a SVM aim to generate optimal hyperplanes with dimensions mirroring that of the number of features set as input data, determining feature space (the n-dimensions where input variables live) dimensionality (Bishop 2006). Different kernels available for use with SVMs allow for better data transformations as it searches for the optimal plane in the descriptor space to separate positive and negative cases (Sadykov & Kosovichev 2017). The arguments we apply to the support vector classifier (SVC; used interchangeably with SVM) supplied by the *scikit-learn*[7] library include the *RBF* kernel, regularization parameter C, a 'scaled' γ, and a *balanced* class weight. In order, the *RBF* kernel presented to be more effective compared to a linear kernel when evaluated under TSS and HSS_2. Following SWPC's formulation in 2007 (Mason & Hoeksema 2010), these are defined as:

$$TSS = \frac{TP}{TP+FN} - \frac{FP}{FP+TN}, \tag{1}$$

$$HSS_2 = \frac{TP+TN-E}{P+N-E}, \tag{2}$$

where TP = true positives, FN = false negatives, FP = false positives, TN = true negatives, E is the expected number of correct forecasts based solely on the probability of an event occuring, P = TP + FN (total positives), and N = TN + FP (total negatives). The TSS ranges from -1 to +1, where +1 indicates perfect agreement to the training set. Any misclassification, both of positive or negative examples, reduces this score accordingly. Values of zero or less indicate a performance no better than a purely random forecast (Ahmadzadeh et al. 2021). HSS_2 ranges from -∞ to +1, where an algorithm of complete accuracy obtains a score of +1, an algorithm being no better than random guesses obtaining a negative score, and a score of 0 attributed to either no events forecasted or observed (NOAA 2022). The standard *HSS* scoring presented by Bobra & Couvidat (2015) is not considered as a metric of success in our work, given that it strongly depends on class-imbalance ratio $\frac{P}{N}$ of the testing set- which as stated before, is drastic and the major obstacle for being able to easily predict SPEs. An advantage of using TSS to validate our algorithm comes from the fact that it is unbiased towards the class-imbalance ratio- the score itself does not depend on the inequality in trials. Bloomfield et al. (2012) and Manzato (2005) echo this in their respective work, agreeing that TSS is an adequate measure of the overall classifier quality, and should be the standard to use in comparisons of the performance of various classifiers for flare forecasts, weather forecasting, and the forecasting of rare events, like SPEs (Ahmadzadeh et al. 2021).

The HSS_2[8] measures the improvement of the forecast over a random forecast, and exhibits only some dependence on the class-imbalance ratio of the testing set (Bobra & Couvidat 2015). Overall, only TSS does not at all depend on the testing-set imbalance ratio, while HSS_2 converges toward zero for large $\frac{P}{N}$ case ratios.

In order, the default parameters used with the SVM are as follows. For SVMs, a kernel refers to a method allowing the application of classifiers to non-linear problems by mapping non-linear data into a higher-dimensional space; a space not visited during the SVM application given that this is automated, again by *scikit-learn*. In practice, *linear* kernels perform well when the number of features is large (e.g. there is no need to map to an even higher dimensional feature space). *Linear* kernels may be considered first given its low computation power needs, allowing it to train and test quickly. Considering non-linear kernels after these initial results, if the accuracy increases, then it is intuitively the better pick to use with the SVM- it is simply important to note that this is not always the case. Overall, even in our case the *RBF* kernel proved to be the better choice at the end of the algorithm's results and is the only kernel

[7] scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

[8] It is important to note the exclusion of using HSS_1 scoring (following the equation $HSS_1 = \frac{TP+TN-N}{P}$, Bobra & Couvidat 2015); given that its dependence on the class-imbalance is magnified and far more difficult to overcome when compared to TSS and HSS_2 SVMs. These differing dependencies on the imbalance are concisely represented in Fig. 2 by Bobra & Couvidat (2015).

employed and discussed from this point on. The results of *RBF* performing drastically better when compared to multiple kernels is also reflected in Cristianini & Shawe-Taylor (2000).

Because our primary goal is not to parametrize our algorithm with minute detail, we leave the regularization parameter C (unique to each SC tested on), as 1- the default value implemented by *scikitlearn*. This leads to a SVM building both a hyperplane with the largest minimum margin, and a hyperplane separating as many instances as possible. The applied C parameter decides how to prioritize enhancing the latter. For large values of C-as those produced by linear kernels, the SVC optimization chooses a smaller-margin hyperplane when it efficiently classifies most- if not all training points correctly. However, the disadvantage to this is the disregard of data points far from the majority; which is detrimental given the imbalanced nature of our positive versus negatively classed data set. Conversely, smaller values of C dictates the classifier to look for a larger-margin separating hyperplane. Given that the regularization parameter affects different testing data sets in different ways, there is no absolute of whether larger or smaller values lead to more appealing results.

The γ = *'scaled'* argument is automated when called, and is simply the kernel coefficient for the *RBF*, and is passed using $\frac{1}{(feature\#)(X.var())}$ as the value of γ. In simpler terms, the parameter declares how far the influence of a single training data point reaches (Buitinck et al. 2013). Automation by Buitinck et al. (2013) of scaling this parameter to the data alleviates the need to manually process this between a small or large value; where the former considers data farther from the plausible separation line compared to the latter. Smaller γ values therefore allow for the consideration of data points farther from where most clump together when defining the separation line. The automated scaling searches for the best fit between higher and smaller values to achieve the best possible fit for the provided data set, so as to not overlook data points outside the highest populated locations, working as almost a proxy to bridge the gap of our imbalance.

Lastly, the *class weight* = *'balanced'* argument sets up the SVC to automatically adjust weights inversely proportional to their respective frequencies as weight = $\frac{sample\#}{(class\#)(bincount(y))}$ (Buitinck et al. 2013). This allows us to optimize the weights assigned to the minority class (in our case, days where SPEs are observed; labeled as '1' in our data set) and lower weights to the majority class (days where SPEs are not observed; labeled as '0' in our data set). To the same effect, we tested an oversampling technique as discussed in Section 4.3, which operates in leiu of the *class weight* parameter by inflating positive case relevance in the data. Overall, the results of both approaches were comparable.

Adequacy of the resulting TSS and HSS_2 scores validating our algorithm and the future investigation of possible dependence of predictions on physical parameters of the solar atmosphere are also discussed later in the paper. The use and results of these scores using our data set are discussed in Section 4.1.

3.1. *Separation to Solar Cycles*

During the development of ML models, data sets are divided into "training" and "testing" sets. Respectively, these are *learning phases* to tune the parameters of a model, and testing sets are what the model is used against to check its accuracy (Bishop 2006). In our case, we have divided our data by years corresponding to our different SCs of interests (i.e. SC 22: 1986-1996, SC 23: 1996-2008, SC 24: 2008-2019). The SVM looks to a given time-interval to generate a prediction for a different time interval. Once these classifiers are trained on one part of the benchmark data set (usually of the order of 2/3 of the data), they are tested for their forecasting capability on the remaining part of the benchmark data (Martens & Angryk 2017). TSS and HSS_2 then quantify the success of prediction based on these traning and testing data sets. The TSS measures the difference between the probability of detection (TP) and that of false alarms (FP). Following the requirements of SVM data set splits, our algorithm is validated on how well it trains itself from SC 23 & 24 data separately, and how accurately it can reproduce the observed events of our prediction window of SC 22. SC 23 is then trained and tested on SC 22 & SC 24 separately, and finally SC 24 trains and tests on SC 22 & SC 23. Additionally, we stacked training data sets as well to explore if longer timeseries' allow for differing or more precise results- testing SC 22 on both SC 23 & 24, SC 23 on both SC 22 & 24, and finally SC 24 on both SC 22 & 23. The results of our algorithm's accuracy in both cases are discussed in Section 4.1.

3.2. *Feature importance, selection & SC dependence*

As we discussed in Sadykov et al. (2021), the inclusion of all features available does not necessarily lead to an increase of the forecast performance, and may even result in a notable decrease. In the presence of multiple irrelevant or redundant features, learning methods tend to over-fit contributions and become less interpretable, or produce results entirely inadequate for usage. A common way to resolve this problem is using feature selection, reducing the

(a) True Skill Statistics

Testing Cycle	SC 22 (Training Cycle)	SC 23 (Training Cycle)	SC 24 (Training Cycle)	Training on 2 Remaining Cycles
SC 22		0.74	0.66	0.73
SC 23	0.74		0.45	0.76
SC 24	0.69	0.66		0.70

(b) Heidke Skill Scores$_2$

Testing Cycle	SC 22 (Training Cycle)	SC 23 (Training Cycle)	SC 24 (Training Cycle)	Training on 2 Remaining Cycles
SC 22		0.52	0.22	0.38
SC 23	0.68		0.09	0.51
SC 24	0.52	0.54		0.56

Table 2. Our current predictive scores for a *RBF*-based SVM using *'balanced'* weights.

dimensionality of the supplied data by selecting only a subset of features from the input feature set. This also works to reduce associated computational costs and removes irrelevant features for problems with multi-dimensional data (Gu et al. 2012). We therefore use Gini Importance and Fisher-score as measures to determine the contributions of different features to the predictive algorithm. Both used as general indicators of feature relevance, Gini Importance is computed from a Random Forest structure, focusing on providing relative rankings of input features indicating how often specific features are selected during node splitting (referring to how a selected feature is used to decide how to divide data into two separate sets with similar responses within nodes of the Decision Tree). Gini importance quantifies how much a particular feature plays in partitioning data into the defined classes, or how large the sum of required node splits reach during classification in a recursive feature elimination scheme (Menze et al. 2009).

On the other hand, the primary use of the Fisher-score is related to the maximum likelihood estimation and is one of the most popular supervised univariate feature selection methods; following the equation:

$$F = \frac{(mean(p) - mean(pn))^2 + (mean(n) - mean(pn)^2}{\sigma(p)^2 + \sigma(n)^2} \tag{3}$$

where p = total positive cases, pn = the probability of a random draw between positive and negative case, n = total negative cases, and σ = standard deviation. Fisher-scoring aims to find a subset of features from all those provided, where distances between data points in different classes are as large as possible, while the distances between data points in the same class are as small as possible (Gu et al. 2012). In simpler terms, this estimation looks for the cleanest branching off of feature dependence in the algorithm in use, assigning high relevance to features in the same class compared to low relevance values assignments to features from different classes.

It is clear from Fig. 7 that the normalized, logarithmic flux value from the previous day dominates in importance compared to the rest of the features in both feature selection measures. This is almost intuitive given that a rise of SEP intensity on the previous day can easily be used as an early predictor that detected SEPs may cross the 10 pfu threshold the next day. Still, this may not always be the case as shown in the left panel of Fig. 5 where an event may begin extremely abruptly without any indication of rising proton intensity the days leading up to it. Only the trailing features (lowest ∼8-10 scoring features) show differences in ranking when comparing Gini Importance and Fisher-scoring.

4. RESULTS & DISCUSSION

4.1. *Resulting forecast accuracy*

Following our conclusions in Sadykov et al. (2021), while our statistical catalog (the input data set for forecasting) generates a total of 24 features to select from, we dwindle this down to half, using only the *top 12 scoring features* according to the Fisher and Gini Importance scores. Compiling the SVC now with the aforementioned arguments in Section 3, result in the forecasting scores shown in Table 2.

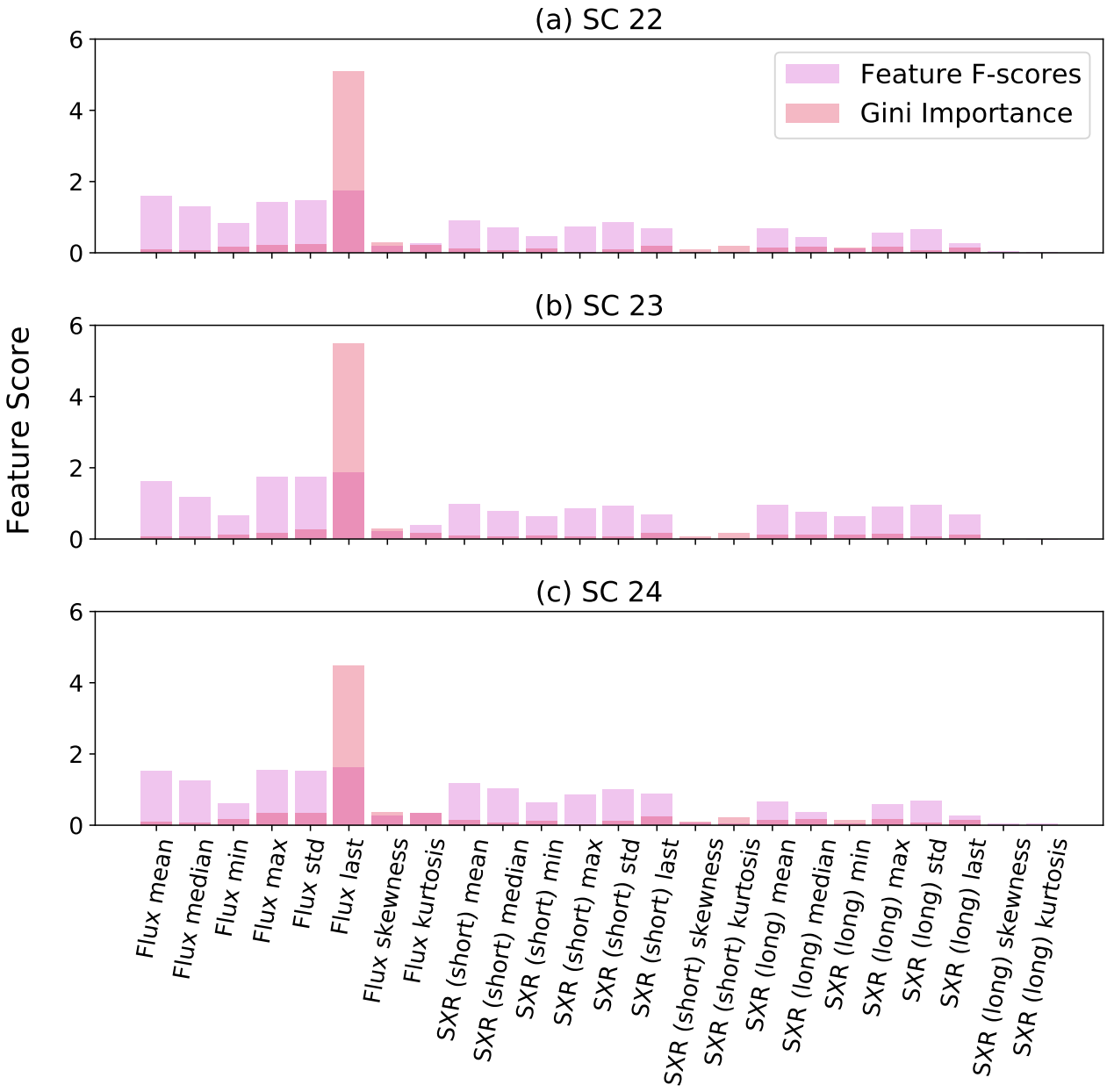

Figure 7. The differences in feature scores between SCs 22-24 using both Gini Importance and Fisher-scores.
**Note: Gini importance scores have been multiplied by 8 for readability.*

4.2. *Exploring how transferable prediction forecasts are between SCs*

The model's performance clearly varies between the three SCs given the large spread of scores: TSS ranging between 43-77% and HSS_2 between 8-69%, as shown in the first three columns in both panels of Table 2. These low scores make it clear that a static algorithm using only proton and SXR data is inadequate to accurately capture the dynamic flux ejections from the solar atmosphere and predict what a different cycle may look like in terms of SPE eruptions based on differing levels of solar activity. This also hints at the fact that other physical parameters of the Sun may be important to consider to develop better forecasts. Lastly, on average, considering longer timelines of previous SCs

showed more favorable TSS and HSS_2 for most cases, as shown in the last column of both panels in Table 2. This may suggest using data from multiple SCs instead of just one during the training phase being favorable for predictions. The slight increases in scores for testing done on SC 24 is worth noting given its inherent low activity- initial scores from training on single SCs was lacking, and using longer timescales could potentially enhance them.

An ideal model would not need to be fine-tuned after the retrieval of recent data, and would be able to make accurate predictions from real-time input. The issue of overfitting slightly plays into our model because it is too closely fit to a limited set of data points and therefore does not generalize well (Ahmadzadeh et al. 2021). This is obvious when looking at both TSS and HSS_2 when SC 24 alone is used as the training cycle. Fig. 1 shows the highest and lowest SPE activity presenting in SC 23 (with SC 22 close to this level of activity) and SC 24 respectively. SCs 22 and 23 have scores close to one another- most apparent in the TSS, which may be due to testing from both a low and high SPE generating cycle whereas SC 24- the cycle with substantially lower SPEs associated with it, has only higher SPE generation SCs to train from. Proton and SXR flux alone do not account for the low levels of solar activity during SC 24 when compared to SC 22 and 23- the resulting lack of SPE eruptions associated with this SC is therefore not accounted for when forming predictions. The physical parameters of the Sun and their variances is the biggest hurdle to overcome in the problem of predicting SPEs, because it is still not well understood which aspects of the solar atmosphere most directly contribute to the generation of SEPs that evolve to SPEs.

4.3. *Considering an Oversampled data set*

Following the classical definition of oversampling, we duplicate our positive cases of days with SPEs observed on the multiplicity of reaching an imbalance ratio $\frac{P}{N}\sim 1$. Doing this allows us to evaluate how the SVM performs when considering a data set with two classes of approximately the same weights instead of the drastically imbalanced (original ratios following : $\frac{P}{N}$ cases $\sim$ 5.3%, 5.8%, and 3.2% for SCs 22, 23, & 24 respectively) data set we have used thus far.

While this seems repetitive given the *balanced* argument used when calling the classifier (Section 3), oversampling provides different results due to its intrinsic methodologies. Consequently, while both methods work to favor predictions for the minority class, the data sets are split and trained differently, yielding slightly different prediction results. Discussed in Bishop (2006), a loss function is a method of evaluating how well an algorithm models its supplied featured data set; i.e. an optimal algorithm minimizes its loss function. The class weight = *'balanced'* argument directly modifies an algorithm's inherent loss function by varying penalties assigned to classes with different weights. When using this, there is essentially a trade in the ability to predict the majority class by purposely biasing the model to favor more accurate predictions of the minority class- by assigning higher importance to them (Abadi et al. 2015). In lieu of this, oversampling calls for synthetic repeated observations of the minority class (days with SPEs), until its frequency in the data set is comparable to that of the majority class (days without SPEs). Doing this therefore brings equal representation of both classes, as well as an artificially extended data set. While the *'balanced'* argument is called using the original number of positive and negative cases supplied, the oversampling technique occurs prior to training, adding significantly more cases for the SVM to learn from.

The results of feature importance also varied when accounting for the oversampled data set, and are shown in Fig. 8. The greatest change in Gini Importance ranking was the *SEP flux skewness* parameter for SC 24- an increase of a factor of $\sim$ 5 when compared to the unsampled data set. Changes of lower magnitudes ($\sim$10-30%) were seen for the parameters: *SEP Flux maximum, SEP Flux skewness,* & *SEP Flux kurtosis* across all SCs of interest. In comparison, the Fisher scores consistently saw a decrease in scores by $\frac{1}{2}$ when considering our previous results in Fig. 7.

Using these reordered features based on updated Fisher and Gini Importance rankings and again dwindling 24 features down to 12 as discussed in Section 3.2, both panels of Table 3 show the resulting TSS and HSS_2 using this oversampled data set. These can be compared to the results using just the class weight = *'balanced'* argument shown in Table 2. The variations by percentage of the oversampled data set forecasting scores compared to the original data set's results are not significant. The changes here loosely favor oversampling, with 10 scores increasing across TSS and HSS_2, 3 scores undergoing no change, and 5 scores slightly dropping in predictive abilities. For TSS scores these changes are within a range of $\sim$1-5 %, and while HSS_2 generally follows the same trend, 2 scores are outliers with changes of $\sim$12% and $\sim$34%.

Overall, the main difference between these oversampling and the class weight = *'balanced'* parameter is that when using the latter, positive cases are represented with sporadic large weights applied across the supplied features. Con-

(a) True Skill Statistics

Testing Cycle	SC 22 (Training Cycle)	SC 23 (Training Cycle)	SC 24 (Training Cycle)	Training on 2 Remaining Cycles
SC 22		0.74	0.63	0.70
SC 23	0.72		0.43	0.76
SC 24	0.68	0.66		0.70

(b) Heidke Skill Scores$_2$

Testing Cycle	SC 22 (Training Cycle)	SC 23 (Training Cycle)	SC 24 (Training Cycle)	Training on 2 Remaining Cycles
SC 22		0.49	0.21	0.25
SC 23	0.69		0.08	0.53
SC 24	0.54	0.52		0.56

Table 3. Predictive scores for a RBF-based SVM using an oversampled data set.

versely, oversampling represents positive cases with their original low weights at a much higher frequency due to feature duplication, and therefore considers their relevance during the training period differently.

In our case, no substantial difference between oversampling and a simple *'balanced'* parameter was seen, even when considering longer timescales by training on two SCs instead of one. Although our data set was balanced manually, the recent work of Stumpo et al. (2021) utilizes a similar approach- a Synthetic Minority Oversampling TechniquE (SMOTE) accounting for the maximum-likelihood estimation (MLE) for SPE prediction (Korstanje 2021). The input for this work includes SXR and radio fluence, as well as flare heliolongitudes to take into account particle propagation from the solar source (Stumpo et al. 2021). Using a synthetic oversampled data set produced better model scores compared to other techniques: basic MLE, and weighted MLE. Oversampling on both (a) the entire data set and (b) using heliolongitudes only $\geq 20°$ resulted in the probability of SPE detections of scores 0.80 and 0.83 respectively. This work and ours suggest that using longer timescales opposed to individuals SCs may be favorable for SPE forecasting.

5. SUMMARY AND CONCLUSIONS

In this work, we have investigated the problem of the prediction of Solar Proton Events (SPEs) using machine learning (ML) and the cross-cycle transferability of the developed forecasts. In particular:

- We explored the significance of different proton and SXR flux features on SPE predictions specifically during SCs 22-24, and conclude that these features alone are not sufficient to produce reliable forecasts.
- We completed two catalogs built in this effort using data from SCs 22-24, one recording fundamental features of all SPEs detected during this time, and another containing daily statistical features that serves as the input data set for our static ML predictive algorithm.
- We tested our algorithm considering both long (using two cycles for training) and short (using a single cycle for training) timescales. **Predictions seem to favor longer timescales, considering slightly enhanced TSS and HSS$_2$ results across all SCs.**
- In the same vein, we compared a *'balanced'* class weight to an oversampled version of our original data set. We observed slight increases in $\sim\frac{2}{3}$ of our TSS and HSS$_2$ results, possibly indicating that **synthetically oversampling the minority class of SPEs may lead to more accurate predictions compared to using only *'balanced'* weights.**

5.1. *Future Work*

While TSS and HSS$_2$ have shown slight increases across SCs when changing the aforementioned input, there still remains a large disparity when aiming for a reliable forecasting algorithm. Our next step is to investigate any relations between McIntosh and Hale classifications of ARs in our ML input catalog. We hypothesize significant contributions to better predictive scores using data considering solar magnetic fields, areas of SEP generation, eruption location

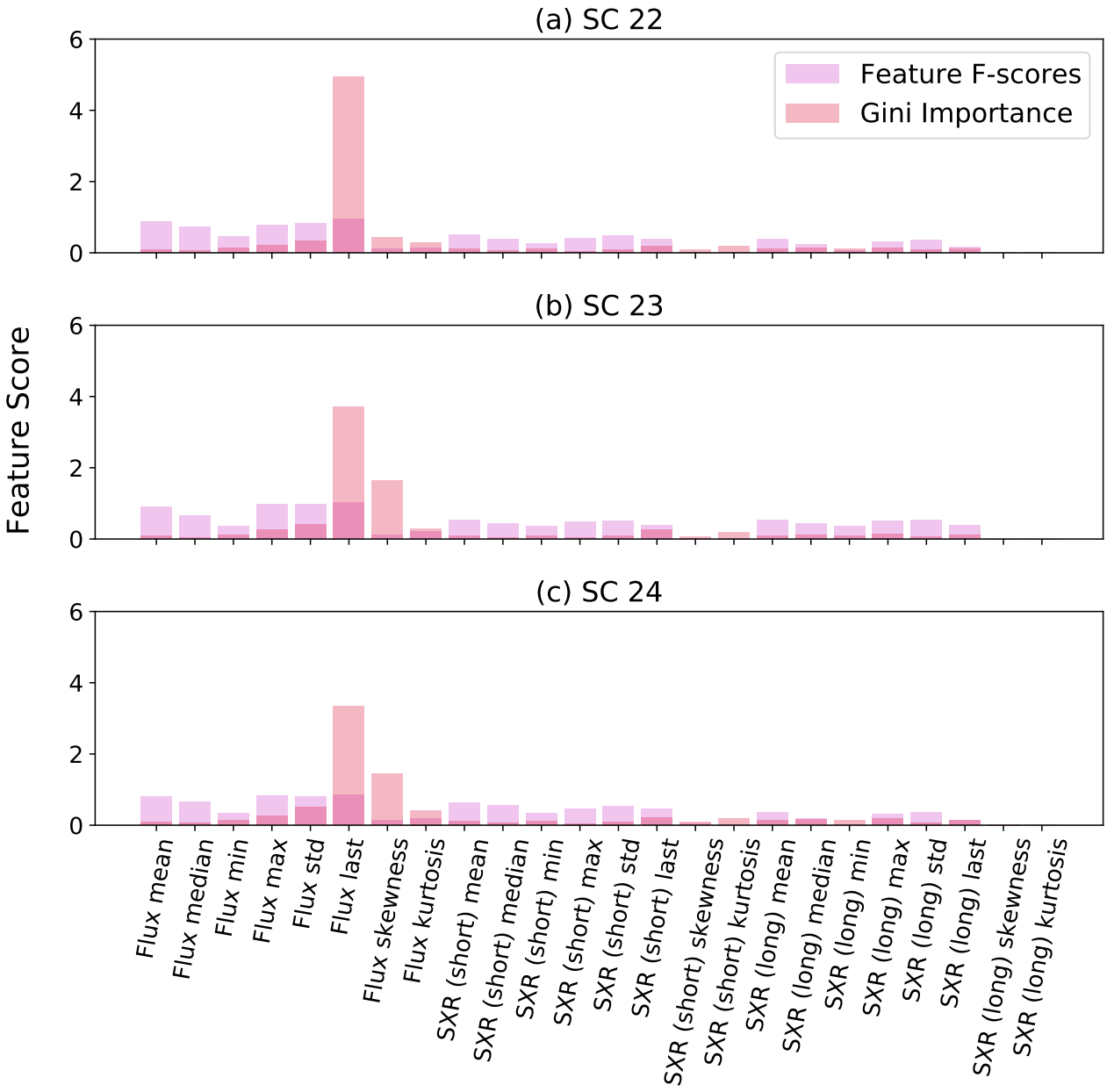

Figure 8. The differences in feature scores between SCs 22-24 using both Gini Importance and Fisher-scores when considering an oversampled data set.
**Note: Gini importance scores have been multiplied by 8 for readability.*

along the Parker spiral (Richardson et al. 2018), properties of sunspots varying during SCs, and dynamic features of the solar corona (Gibson 2015).

ACKNOWLEDGMENTS

The research was supported by NASA Early Stage Innovation program grant 80NSSC20K0302, NASA LWS grant 80NSSC19K0068, NSF EarthCube grant 1639683, and NSF grant 1835958. VMS acknowledges the NSF FDSS grant 1936361 and NSF grant 1835958. EI acknowledges the RSF grant 20-72-00106.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/

Ahmadzadeh, A., Aydin, B., Georgoulis, M. K., et al. 2021, The Astrophysical Journal Supplement Series, 254, 23, doi: 10.3847/1538-4365/abec88

Aminalragia-Giamini, S., Raptis, S., Anastasiadis, A., et al. 2021, Journal of Space Weather and Space Climate, 11, 59, doi: 10.1051/swsc/2021043

Anastasiadis, A., Lario, D., Papaioannou, A., Kouloumvakos, A., & Vourlidas, A. 2019, Philosophical Transactions of the Royal Society of London Series A, 377, 20180100, doi: 10.1098/rsta.2018.0100

Beck, P., Latocha, M., Rollet, S., & Stehno, G. 2005, Advances in Space Research, 36, 1627, doi: 10.1016/j.asr.2005.05.035

Bishop, C. M. 2006, Pattern Recognition and Machine Learning (Springer)

Bloomfield, D. S., Higgins, P. A., McAteer, R. T. J., & Gallagher, P. T. 2012, The Astrophysical Journal, 747, L41, doi: 10.1088/2041-8205/747/2/l41

Bobra, M. G., & Couvidat, S. 2015, The Astrophysical Journal, 798, 135, doi: 10.1088/0004-637X/798/2/135

Buitinck, L., Louppe, G., Blondel, M., et al. 2013, in ECML PKDD Workshop: Languages for Data Mining and Machine Learning, 108–122

Collins, P. 2006, Advances in Space Research, 37, 116, doi: 10.1016/j.asr.2005.05.107

Cristianini, N., & Shawe-Taylor, J. 2000, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods (Cambridge University Press), doi: 10.1017/CBO9780511801389

Gibson, S. 2015, in Polarimetry, ed. K. N. Nagendra, S. Bagnulo, R. Centeno, & M. Jesús Martínez González, Vol. 305, 245–250, doi: 10.1017/S1743921315004846

Gu, Q., Li, Z., & Han, J. 2012, arXiv e-prints, arXiv:1202.3725. https://arxiv.org/abs/1202.3725

He, J., & Rodriguez, J. V. 2018, Space Weather, 16, 245, doi: 10.1002/2017SW001743

Korstanje, J. 2021, Smote, Towards Data Science. https://towardsdatascience.com/smote-fdce2f605729

Langford, M. 2022, Space Radiation Analysis Group (SRAG) web site, NASA. https://srag.jsc.nasa.gov/spaceradiation/what/what.cfm

Manzato, A. 2005, Weather and Forecasting, 20, 918, doi: 10.1175/WAF899.1

Martens, P. C., & Angryk, R. A. 2017, Proceedings of the International Astronomical Union, 13, 344–347, doi: 10.1017/S1743921318000510

Mason, J. P., & Hoeksema, J. T. 2010, The Astrophysical Journal, 723, 634, doi: 10.1088/0004-637x/723/1/634

Menze, B., Kelm, B., Masuch, R., et al. 2009, BMC bioinformatics, 10, 213, doi: 10.1186/1471-2105-10-213

Metha, P. 2022, Solar storms can destroy satellites with ease, Space. www.space.com/solar-storms-destroy-satellites

Naito, M., Kodaira, S., Ogawara, R., et al. 2020, Life Sciences in Space Research, 26, 69, doi: https://doi.org/10.1016/j.lssr.2020.05.001

NOAA. 2019, NOAA readies GOES-15 and GOES-14 for orbital storage, NESDIS. https://www.nesdis.noaa.gov/news/noaa-readies-goes-15-and-goes-14-orbital-storage

—. 2022, Forecast verification glossary - space weather prediction center, NOAA. https://www.swpc.noaa.gov/sites/default/files/images/u30/Forecast%20Verification%20Glossary.pdf

Onorato, G., Di Schiavi, E., & Di Cunto, F. 2020, Frontiers in Physics, 8, doi: 10.3389/fphy.2020.00362

Parker, E. N. 1958, ApJ, 128, 664, doi: 10.1086/146579

Raschka, S., & Mirjalili, V. 2017, Python machine learning - second edition. https://www.packtpub.com/product/python-machine-learning-second-edition/9781787125933

Richardson, I. G., Mays, M. L., & Thompson, B. J. 2018, Space Weather, 16, 1862, doi: 10.1029/2018SW002032

Rodriguez, J. 2012, Note on GOES 13-15 Yaw Flip, NOAA. https://www.ngdc.noaa.gov/stp/satellite/goes/doc/Note%20on%20GOES%2013-15%20Yaw%20Flip.pdf

Rotti, S., Aydin, B., Georgoulis, M. K., & Martens, P. C. 2022, ApJS, 262, 29, doi: 10.3847/1538-4365/ac87ac

Sadykov, V., Kosovichev, A., Kitiashvili, I., et al. 2021, arXiv e-prints, arXiv:2107.03911. https://arxiv.org/abs/2107.03911

Sadykov, V. M., & Kosovichev, A. G. 2017, ApJ, 849, 148, doi: 10.3847/1538-4357/aa9119

Stumpo, M., Benella, S., Laurenza, M., et al. 2021, Space Weather, 19, e2021SW002794, doi: 10.1029/2021SW002794

Whitman, K., Egeland, R., Richardson, I., et al. 2022, Advances in Space Research, doi: 10.1016/j.asr.2022.08.006