8-2024

Novel Short-term Predictions of Solar Energetic Particle Events Using Machine Learning Time Series Classifiers

Sumanth Rotti
Georgia State University

Novel Short-term Predictions of Solar Energetic Particle Events Using Machine Learning Time Series Classifiers

by

Sumanth Ananda Theertha Rotti

Under the Direction of Petrus C.H. Martens, Ph.D.

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2024

ABSTRACT

Solar energetic particles (SEPs) are one of the most crucial aspects of space weather, predominantly constituting intense proton beams from the Sun. Those surpassing the Space Weather Prediction Center's S1 threshold of a solar radiation storm have severe technological and biological implications for space missions outside the Earth's magnetic field. Predicting their arrival in near-Earth space strongly depends on various factors, including the solar eruptions, such as flares and coronal mass ejections (CMEs). Therefore, we follow a strategic path toward SEP forecasting by leveraging machine learning (ML) principles for 'research to operations' demands. For efficient applications, there are two aspects in this dissertation: (1) the development of a benchmark data set of SEP events and (2) the design of a robust methodology for short-term predictions of SEPs using time-series-based ML classifiers.

The first outcome of this research is the Geostationary Solar Energetic Particle (GSEP) events data set, a comprehensive collection of over 400 SEP events spanning solar cycles 22 - 24. Each event in this data set has been meticulously analyzed for its spatiotemporal properties, including the characteristics of the associated source eruptions, such as flare magnitudes, locations, rise times, speeds and widths of CMEs, and radio bursts.

Next, we present an ensemble learning approach that merges the results from univariate time series of solar protons and X-ray fluxes. Furthermore, we demonstrate a proof-of-concept of using feature-based classifiers to distinguish SEP events from non-events. Here, non-events constitute those SEPs below the S1 threshold and "SEP-quiet" periods. We do a comparative analysis of the models and extensively evaluate the feasibility of our data-driven approach by establishing confidence in the predictions utilizing multiple forecasting metrics. Lastly, we present our results for lead times of up to 60 minutes that display great potential for the implementation of our methodology in near-real-time operations for short-term SEP event predictions.

INDEX WORDS: Solar Activity, Solar Eruptions, Solar Energetic Particles, SEP Events Forecasting, Space Weather Forecasting, Multivariate Time Series Classification

Novel Short-term Predictions of Solar Energetic Particle Events Using Machine Learning Time Series Classifiers

by

Sumanth Ananda Theertha Rotti

Committee Chair: Petrus C.H. Martens

Committee: Manolis K. Georgoulis

Rafal Angryk

Vadym Apalkov

Fabien Baron

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2024

DEDICATION

I dedicate this work to the two most important people in my life who strongly motivated me to pursue PhD, and unfortunately, I lost them during this journey:

(1) Dr. Bannanje Govindacharya (my illustrious guru and an enigma)

(2) Mr. Ananda Theertha Rotti (my fearless father)

ACKNOWLEDGMENTS

I was about two years old when my father got me two shaped-board books containing gorgeous and captivating illustrations and pictures: one about dinosaurs and the other about space. They impacted my mind so much that I either wanted to become a paleontologist or an astrophysicist. I zeroed in on the latter in high school. Today, I am satisfied to have made it possible primarily because I have been very serious in my pursuits all my life.

During my time at Georgia State University (GSU), several people have helped me acquire the critical skills necessary to make a career in scientific research. First and foremost, I want to thank my advisor, Prof. Piet Martens, for his continual support and guidance as a non-pareil mentor over the last five volatile years. He always allowed me to work independently and taught me to focus on the physical implications of any scientific analysis. I thank him for everything he taught me about heliophysics and his kindness at crucial junctures.

Many thanks to GSU for funding the initial part of my research. I am most grateful to NASA, which later supported my research, publication and travel via its FINESST grant. I thank the graduate advisors, Profs. Murad Sarsour (Physics) and Russel White (Astronomy) for their encouragement and support during my initial structuring stage at GSU. The many discussions with Prof. Manolis Georgoulis during the early phase of my PhD immensely helped in clarifying several aspects of my work. Prof. Rafal Angryk, Drs. Berkay Aydin, Dustin Kempton, and Azim Ahmadzadeh at the Data Mining Lab have always encouraged my involvement in our interdisciplinary projects. I am grateful to them for teaching me the

language of a computer scientist. I thank Dr. Justin Cantrell and Jeremy Simmons for their support with computing hardware and troubleshooting. I also thank Felicia Watts and Alicia Rice for always helping with the paperwork for refunds and every administrative need.

I had some fantastic graduate colleagues who have helped me in multiple ways. Dr. Sushant Mahajan was very encouraging when I joined GSU and provided valuable suggestions on how to get accustomed to living in Atlanta. Dr. Beena Meena was an earnest source of support as a mentor under the AstroPal program. Dr. Aparna Venkataramanasastry has greatly supported me throughout my time at GSU, and I am grateful to her for helping me find the right resources, among other things. I will cherish all my memories with my colleagues Dr. Aranyo Mitra, Dr. Neva Agarwala, Dr. India Jackson, Sayed Hossaini, Uttam Acharya, and Suresh Gnawali. Some of my junior colleagues and desk-mates have given me a great time of conversation and collaboration whenever necessary: Becky Flores, KhaDeem Coumarbatch, Varun Chaturmutha, Fallon Konow, Nikhil Gupta, Akhil Arya, and Anli Ji.

I am immensely thankful to all the homeowners who provided me shelter: Dr. Sunil Rao, Mr. Palimar Gopal Rao, and Mr. Ramakrishna Joshi. None of this would have been possible without the encouragement and support of my family and friends, for whom I am deeply ever grateful. I extend my sincere gratitude to the following: Mrs. Bharathi Anand (my hardworking mother), Mrs. Soumya Ananth (my mighty little sister) and Mrs. Tejaswini M.K. (my mentor and guide who has been there for me at every vital phase). I heartily acknowledge my interactions with Mr. S. L. Rao which have profoundly impacted my life.

I acknowledge the use of X-ray and proton flux data from the GOES missions made

available by NOAA. I thank all the anonymous reviewers for their constructive comments that have improved the contents of the papers. The works presented in this dissertation have been supported by NASA's SWR2O2R grant 80NSSC22K0272 and FINESST grant 80NSSC21K1388. The acknowledgments for specific chapters are as follows:

- For the work presented in Chapter 2, I thank the teams behind the catalogs, namely PSEP, CDAW-SEP, and NOAA-SEP, for the opportunity to utilize their work. S.R. thanks (1) Dr. Hazel Bain of NOAA for information on the GOES primary and secondary observations and (2) Dr. Steve Johnson of NASA-SRAG for discussing much of his work in detail and agreeing to merge the efforts. The explanations on the events of different characters were crucial to classify and flag the SEP events.

- For the work presented in Chapter 3, I acknowledge the use of observations from SOHO and SDO missions and thank the team for the availability of the data online. SOHO is a project of international cooperation between ESA and NASA. SDO is the first mission to be launched for NASA's Living With a Star (LWS) Program. I also thank the CDAW for the opportunity to utilize their SEP, CME, and type II radio burst catalogs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR	Active Region
AIA	Atmospheric Imaging Assembly
AUC	Area Under the ROC Curve
BOSS	Bag of SFA Symbols
BS	Brier Score
CCMC	Community Coordinated Modeling Center
CDAW	Coordinated Data Analysis Workshop
CME	Coronal Mass Ejection
CNO	Carbon-Nitrogen-Oxygen
CV	Cross Validation
DFT	Discrete Fourier Transform
DT	Decision Tree
DTW	Dynamic Time Warping
EIT	Extreme-ultraviolet Imaging Telescope
EPEAD	Energetic Proton, Electron, and Alpha Detector
EPS	Energetic Particle Sensor
ESP	Energetic Storm Particle
EUV	Extreme Ultraviolet
EVA	Extra-vehicular Activity

FINESST	Future Investigators in NASA Earth and Space Science and Technology
FN	False Negative
FORSPEF	Forecasting Solar Particle Events and Flares
FP	Flase Positive
FPR	False Positive Rate
GCR	Galactic Cosmic Ray
GOES	Geostationary Operational Environmental Satellite
GSEP	Geostationary Solar Energetic Particle
GSS	Gilber Skill Score
HSS	Heidke Skill Score
IMP	Interplanetary Monitoring Platform
IP	Interplanetary
IQR	Interquartile Range
KNN	K-Nearest Neighbor
LASCO	Large Angle and Spectrometric Coronagraph
LR	Logistic Regression
MCB	Multiple Coefficient Binning
MCC	Matthew's Correlation Coefficient
MK	Million degrees Kelvin
ML	Machine Learning

MVTS	Multivariate Time Series
NASA	National Aeronautics and Space Administration
NN	Neural Networks
NOAA	National Oceanic and Atmospheric Administration
NRT	Near-real-time
QF	Quality Factor
RB	Radio Burst
RF	Random Forest
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SDO	Solar Dynamics Observatoty
SEM	Space Environment Monitor
SEP	Solar Energetic Particle
SEPEM	Solar Energetic Particle Environment Modelling
SEPVAL	SEP Validation
SF	Solar Flare
SFA	Symbolic Fourier Approximation
SOHO	Solar and Heliospheric Observatory
SPE	Solar Proton Event
SRAG	Space Radiation Analysis Group
SSC	Summary Statistic Classifier

SSN	Sunspot Number
STSF	Supervised Time Series Forest
SVM	Support Vector Machine
SWPC	Space Weather Prediction Center
SWx	Space Weather
TN	True Negative
TP	True Positive
TPR	True Positive Rate
TSF	Time Series Forest
TSS	True Skill Statistic
XGB	Extreme Gradient Boosting
XRS	X-Ray Sensor
XT	Extremely Randomized Trees

CHAPTER 1
Introduction

1.1 The Sun

Our sky is filled with stars, as evidenced by their twinkling at night. Amongst them, one brightly burning star dominates the Earth's neighborhood: the Sun. Etymologically, the word 'Sun' has been derived from the Latin root - Sol. We can find several variants of the word 'Sun' in many languages, such as 'zon' or 'zonne' in Dutch and 'sunna' in Gothic and Old Norse. The old German word 'sunne' in feminine gender refers to the "heavenly body" (Barnhart 1995). The ancient Greeks embodied the Sun as a deity named Helios, a descendant of the Titans. The ancient Romans worshiped the Sun as a god but replaced the Greek word Helios with the Latin Sol. In the Indian sub-continent, an architectural marvel in the form of a temple dedicated to the Sun god (Surya) was built in 1250 AD by the Ganga dynasty in Konark, situated in the eastern state of present-day Odisha.

As cultures evolved in different parts of the world, the inspiration to study the Sun remained intact. With the development of the telescope, western scientific thinkers have contributed vastly to observing the Sun over the past four centuries, paving the way for the evolution of the field of heliophysics. In the early 17th century, Galileo Galilei observed sunspots on the solar disc. In 1802, English chemist William Hyde Wollaston discovered absorption lines in the solar spectrum. In the mid-19th century, the discovery of sunspot activity by German astronomer Samuel Heinrich Schwabe and the observation of a solar flare by British astronomers Richard Carrington and Richard Hodgson prefaced toward under-

standing the dynamics of the Sun. American astrophysicist George Ellery Hale's discovery of the Sun's magnetic activity in 1908 (Hale 1908) was prominent in analyzing its influence on Earth's magnetic field. Since then, the success of increased knowledge in solar magnetism has shown that the Sun's magnetic activity increases and decreases based on the periodic variability of the number of sunspots, which are sites of strong emerging magnetic fields.

Figure 1.1 shows the annual average number of sunspots observed in the last 65 years. The pattern follows a variation both in magnitude and duration (Usoskin 2017). Solar magnetism associated with the number of sunspots has an $\approx$ 11-year cycle. A low sunspot number (SSN) corresponds to a solar minimum, and a higher SSN indicates a solar maximum. In each cycle, the Sun's magnetic field completely flips, which means that the polarity of the Sun's north and south poles switch places. Similarly, it takes about another solar cycle to reverse the Sun's polarity. The conventional numbering of solar cycles was initiated by Swiss astronomer Johann Rudolph Wolf in 1852 using the historical sunspot data. Solar cycle 1 is assigned to have started in 1755-56, while the ongoing cycle 25 started in December 2019.

1.2 Solar Activity

It had become apparent that there is a connection between the Sun's magnetic activity and the magnetic disturbances on the Earth, which was not through electromagnetic radiation. In 1958, Parker proposed the concept of a steady stream of solar plasma material, what is now called the "solar wind", to be flowing out from the Sun into the solar system (Parker 1958). The existence of solar wind was confirmed by the Soviet Union's Luna I and Venus

Figure 1.1 The yearly averaged international sunspot number since 1960 obtained from the Solar Influences Data Analysis Center.

I spacecraft in 1959. Thanks to swift advancements of technology in the last three decades, we have been fortunate enough to conduct deeper investigations of the Sun and near-Earth space. Figure 1.2 shows space-based observatories that have continuously observed the Sun since the 1960s. Apparently, the Sun is the only star that can be observed in great detail in a wide range of wavelengths (or energies), as its existence and properties are critical to us on Earth. Solar observations, for example, have revealed that solar wind properties, including the speed, density, strength of its magnetic field, and composition, change with solar activity.

The solar wind is composed of solar plasma materials containing $\approx 95\%$ ionized hydrogen (electron and protons), 4% α-particles, and about 1% other heavy nuclei consisting of carbon, nitrogen, oxygen, neon, magnesium, silicon, and iron (Bame et al. 1968). Typically, solar winds are categorized into two types based on their speed: fast ($\approx$750 km.s^{-1}) and slow ($\approx$350 km.s^{-1}). They carry the Sun's magnetic field while flowing into interplanetary (IP) space called the interplanetary magnetic field. Interaction of the Earth's magnetosphere

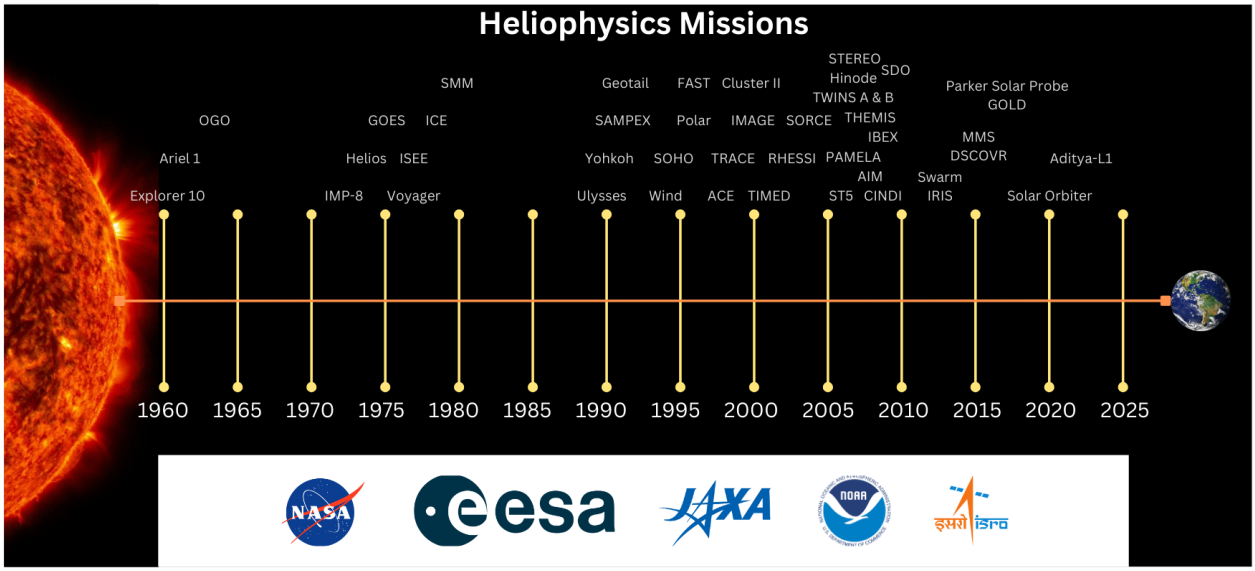


Figure 1.2 A list of heliophysics missions based in space starting from 1960 until the recent Aditya-L1 launched by India.

with the solar wind causes it to compress on the Sun-facing side and stretch at the other. Relative to the background solar wind, large amounts of energy, radiation, and highly charged particles are emitted from the Sun that may reach near-Earth space. While the release of mass and energy in the form of flares and coronal mass ejections (CMEs) have broader impacts on our technological and biological systems, a particular case of interest regarding the effects of space radiation constitutes solar energetic particle (SEP) events. Because SEPs are associated with solar activity, an overview of the internal processes leading to the emergence of powerful solar bursts is relevant. In Section 1.3, I will briefly introduce the Sun's structure based on the physical understanding obtained through mathematical models and observations. Later in Section 1.4, I will explain the eruptive phenomena on the Sun. Lastly, in Section 1.5, I will elaborate on the SEP events, their impacts, and the relevance of forecasting them, which form the heart of this dissertation.

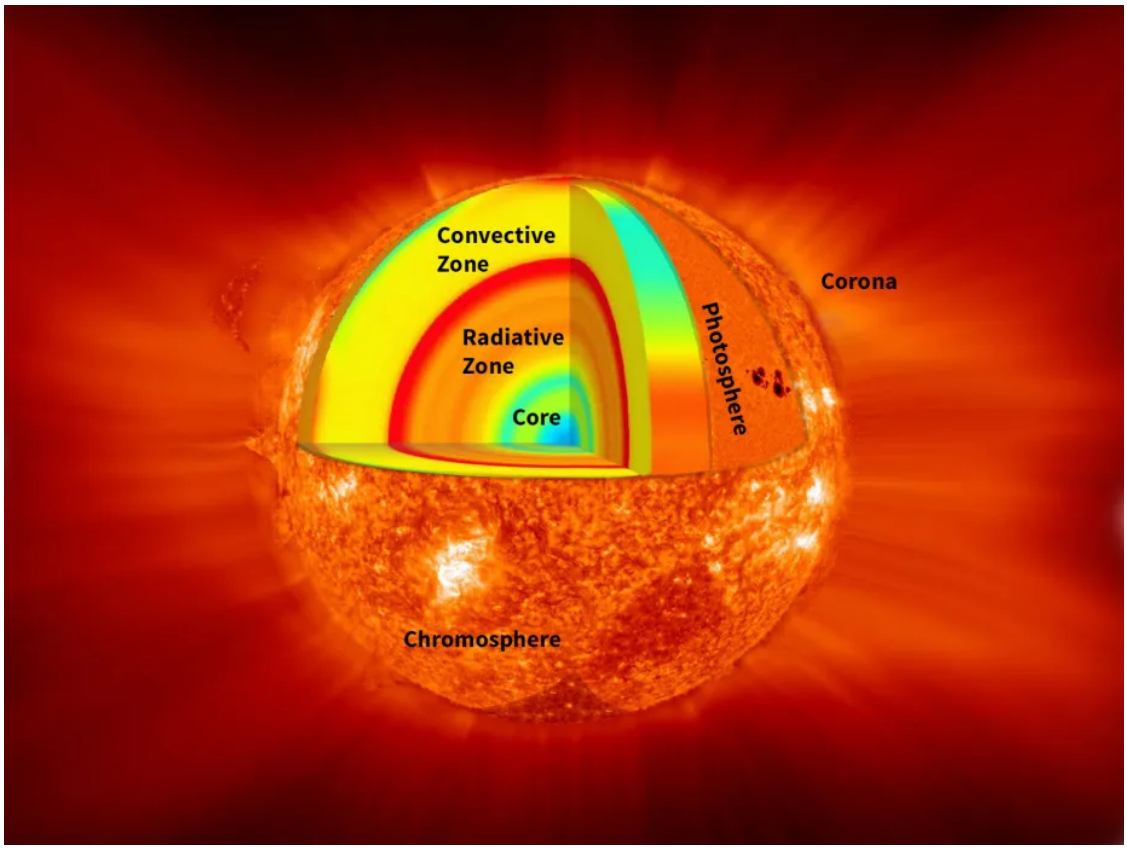


Figure 1.3 A cross-sectional illustration of the Sun's primary radial structure from the core to the corona that contains evaporating solar wind.

1.3 Structure of the Sun

The Sun is considered a main sequence star in a typical astronomer's jargon that is located at a distance of 151.66×10^6 km from Earth. It has a diameter of 1.392×10^6 km and a mass of 1.989×10^{33} g. Figure 1.3 shows different layers below the observable solar atmosphere constituting the solar interior: a central core, a radiative zone, and an outer convective zone. The Sun fuses hydrogen to helium in its core in two distinct fusion reactions: the proton-proton chain reaction and the carbon-nitrogen-oxygen (CNO) cycle. In the inner core of the Sun, the temperature goes up to 15 million degrees Kelvin (MK), where some of

the protons gain enough energy to tunnel the Coulomb barrier of the nuclear charge. The energy released in nuclear fusion is radiated and reabsorbed as it diffuses outward across the solar radiative zone. At the outer edge of the radiative zone lies a thin region called the tachocline extending until the base of the convective zone. In the tachocline, the Sun rotates (from East to West) like a rigid body. The solar rotation is differential throughout the convective zone, i.e., faster at the equator than at the poles. The sidereal period of solar rotation at the equator is ∼24.47 days, but it is 25% longer at latitudes above 60°. The Sun also rotates differentially in the radial direction.

Azimuthal surfaces of constant rotation speed run radially through the convection zone until the tachocline, forming conical shells about the rotation axis. Circulation of the hot and dense plasma (ions and electrons) via convective motion across the solar convection zone transfers energy to the photosphere. Nonetheless, the overlying material in the photosphere is too thin to absorb radiation or prevent its escape into space. Moreover, it is understood that the radiation of thermal energy from sunspots cools the photosphere to ∼4500 K in sunspots. In the region above the photosphere called the chromosphere, which is up to a height of about 2 Mm, the temperature remains up to 20,000 K at the top. However, above the chromosphere extending outward about another solar radius in the transition region up to the corona, the temperatures rapidly rise to over 1 MK (e.g., Aschwanden 2005).

The Sun's inner convection and differential rotation of plasma generate a magnetic field. The field lines emerge from the photosphere at one sunspot and reenter at a nearby spot. The largest magnetic fields measured in the sunspots are about 3000 gauss (G). A group

of sunspots on the solar photosphere are surrounded by active regions (ARs) of moderate field strength (about 100 G). As the ARs are hotter than their surroundings, they appear brighter. Observations have revealed that ARs tend to occur at mid-latitudes on the Sun, where the effect of differential rotation on magnetic field generation is most significant.

1.4 Space Weather

Solar magnetic field, which is very complicated in time and space (Sheeley 2005; Parker 2009), is the fundamental factor governing solar activity. Increasing SSN and ARs indicate a rise in solar activity, which leads to a variety of eruptions, such as flares, CMEs and SEPs (Cohen 2006; Mewaldt et al. 2006). Figure 1.4 shows an illustration of a variety of solar phenomena resulting from the Sun's magnetic field that substantially affect the Earth and its magnetic field. This stringent influence of solar activity broadly on the solar system is called space weather (SWx).

1.4.1 Solar Flares

Solar flares are localized emission of electromagnetic radiation in the Sun's atmosphere across the electromagnetic spectrum, from radio waves to γ-rays. They appear as an intense flash, typically in an AR, occurring over a few minutes when observed in the Hα line. They are the most efficient mechanisms for releasing magnetic energy up to 6×10^{26} joule (J) during the magnetic reconnection. X-rays and extreme ultraviolet (EUV) radiation from solar flares are absorbed in the Earth's atmosphere and can thus be observed only in space. Solar X-rays measured by the Geostationary Operational Environmental Satellites (GOES; Sauer 1989;

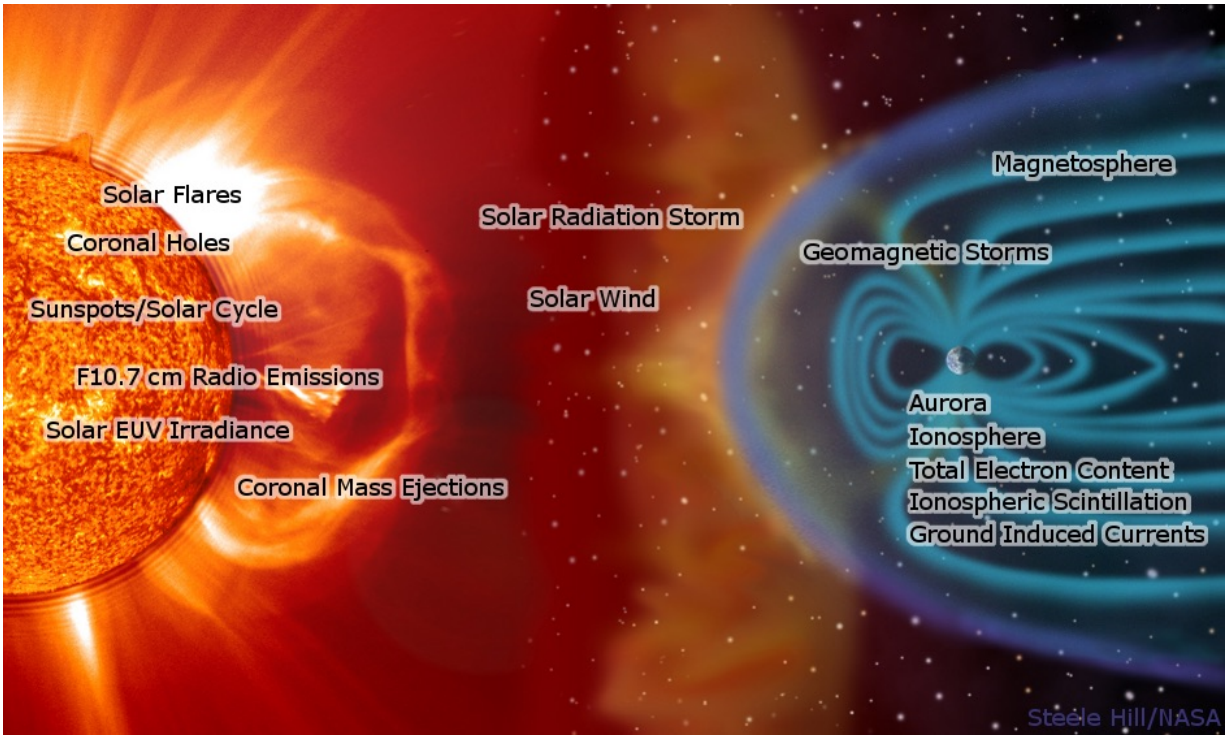


Figure 1.4 Illustration of different solar phenomena such as flares, CMEs and SEPs that form the basis of space weather.
Note: Due to the continuously flowing solar wind, the Earth's magnetosphere (shown by the blue-toned lines) gets compressed on the Sun-side and stretched on the night side while plasma and solar particles stream around it. Image courtesy of NASA.

Bornmann et al. 1996) in 1 to 8 Å are used to classify solar flares based on the peak X-ray flux. The intensity is indexed as follows: A (10^{-8} - 10^{-7} Wm^{-2}), B (10^{-7} - 10^{-6} Wm^{-2}), C (10^{-6} - 10^{-5} Wm^{-2}), M (10^{-5} - 10^{-4} Wm^{-2}), and X ($>10^{-4}$ Wm^{-2}). Each category (from A to X) is ten times more intense than others in the specified order.

1.4.2 Coronal Mass Ejections

CMEs are large chunks of dense plasma material leaving the Sun into the heliosphere. They travel at speeds of less than 250 $km.s.^{-1}$ to larger than 2000 $km.s.^{-1}$ relative to the background solar wind. If directed towards Earth, the fastest CMEs can reach us anywhere

between 15 to 18 hours. Typical estimations of ejected masses are in the range of 5 $\times$ 10^{12} - 5 $\times$ 10^{13} kg, with the angular sizes extending to 40° - 50°. The energy carried by a CME is of the order of 10^{24} J, mainly in the kinetic energy of the plasma cloud and less in high-energy particles. In certain cases, a filament structure visible in the corona for days is suddenly destabilized and erupts within a CME, disappearing from its former position. An Hα brightening may form a classic double-ribbon pattern along the filament channel with slight heating and X-ray emission, but no flaring occurs. Such events can produce a fast CME, a shock wave, and a substantial gradual SEP event without the need for a flare or even an AR. Nonetheless, CMEs pose great observational challenges despite their enormous size. Specialized instruments called coronagraphs, which artificially eclipse the visible disc of the Sun, are used to observe CMEs in white light. The observed light is produced by Thomson scattering of the solar photons from the electrons in the cloud.

1.5 Solar Energetic Particle Events

SEP events represent non-equilibrium distributions of energetic particles from the Sun constituting protons, electrons, and heavier nuclei up to iron (Schmelz et al. 2012). This phenomenon is different from the steady emission of plasma in the form of solar wind that radially expands out from the Sun. The solar wind, which blows continuously at speeds of 300 to 800 km.s^{-1}, corresponds to proton energies of 0.5 to 3 kilo-electronvolt (keV), while SEPs are accelerated from the Sun from a few keV to GeV energies. Solar protons with energy E $\geq$10 MeV may increase relative to the background flux by four to five orders of

magnitude for a period of several hours to a few days. Nonetheless, the shapes of the energy spectra, as well as the total fluence, vary considerably from event to event.

SEPs occur more often near solar maximum, but the correlation between event frequency and solar conditions is not precise (Shea & Smart 1990). The generation of an SEP event is typically associated with solar flares and the IP shocks driven by CMEs (Cane at al. 2003; Vlahos et al. 2019). Those accelerated at flares are known as impulsive SEP events, and particle populations accelerated by near-Sun CME shocks are termed gradual SEP events (Reames 1999). This two-class paradigm of SEP events is only a general structure, as more sensitive measurements of numerous events show a mixture of the two types (Kocharov & Torsti 2002; Kallenrode 2003). That is, a combination of solar eruptions leads to SEP events because most M- and X-class flares are observed to be followed by a CME. However, the relative role of flares in the processes of particle energization is still under debate (Reames 2021). Some enhancements in the particle fluxes can be seen at the passage of a CME shock near Earth. These are known as energetic storm particle (ESP) events (Bryant et al. 1962; Desai & Giacalone 2016).

Energetic particles exist only when the ambient plasma is very rarefied, so that Coulomb collisions, which act to thermalize the distributions, are negligible. During solar flares, accelerated electrons produce X-rays due to collisions with ambient ions (bremsstrahlung emission), while streams of accelerated ions produce γ-rays due to nuclear collisions at the dense layers above the solar surface. Hence, remote sensing observations of X-ray and γ-ray result in direct information on particle acceleration at the Sun and provide complementary

diagnostics to the escaping energetic charged particles observed as SEPs (Cliver et al. 1989; Kahler 1994). The relevant force acting on charged particles is the Lorentz force, which dominates the gravitational force. Relativistic effects occur when particle energies are very high (e.g., above a few GeV for protons and a few MeV for electrons). As the magnetic field lines are tied to the surface of the rotating Sun, the expansion of solar wind causes the magnetic field to take a spiral shape like the skirt of a ballerina (Parker 1958; Rosenberg & Coleman Jr 1969; Owens & Forsyth 2013). The energetic particles leaving the Sun are constrained in their path along the spiraling magnetic field lines, also called Parker's spiral. Thus, SEPs originating from the western hemisphere of the solar disc have a higher probability of reaching near Earth (Papaioannou et al. 2016).

1.5.1 Effects of SEP Events

Human space travel beyond the Earth's magnetic field, for example, to the Moon and Mars, is currently in preparation by space agencies worldwide. For space travel beyond the low-Earth orbit, there are three primary sources of space radiation: galactic cosmic rays (GCRs), SEPs, and particles trapped in the Earth's magnetic field (Benton & Benton 2001; Bourdarie & Xapos 2008; Nelson 2016). GCRs are a continuously present low-intensity background flux of high-energy (several GeV/nucleon) ions from hydrogen to nickel and beyond. Conversely, SEPs are mostly protons from the Sun that can cause a significant radiation environment within a short time period. SEPs dominate over GCRs during higher solar activity, and it has long been recognized that SEPs are of more than scientific interest (Reames 2021).

SEPs of very high intensities can lead to both technological and biological effects on

space travel outside of the Earth's magnetic field as radiation levels from SEPs vary both with solar activity and its connectivity to Earth. SEP storm can directly affect satellites (Stassinopoulos & Raymond 1988), while it can also cause acute and chronic effects on biological systems (Chancellor et al. 2014). Such effects can be in-mission and functional depending on the magnitude of the radiation absorbed, the type of particle, and the matter affected (Schrijver & Siscoe 2010). The physical characteristics of the material define how it may be relatively easily degraded by radiation or stay resistant. The threshold of radiation damage to biological systems, such as the crew on long-term missions, is far lower than spacecraft electronics (Tribble 2010). The amount of radiation dose absorbed by living tissues is defined by the SI unit 'gray' (Gy), where 1 Gy is 1J of radiation energy absorbed by one kg of tissue/matter. To factor how energetic a source of radiation is, the gray is multiplied by a "quality factor" (QF). QF is an experimentally determined value that defines excess damage as a proportion of γ-ray damage for the same dose (with γ-rays having QF = 1.0) to produce a value indicating the relative biological effectiveness of a radiation dose. This value, dose equivalent, is measured in sieverts (Sv; Guetersloh & Zapp 2010).

1.5.2 Operational Requirements During SEP Events

It is important to assess all potential risks, including radiation exposure, that astronauts and spacecraft equipment may experience during a space mission where the impact of dose rate can occur both during IP flight and extravehicular activities (EVAs; Cucinotta et al. 2001; Kennedy 2014; Cucinotta et al. 2014). In the case of SEP events, the total dose effects may significantly rise rapidly within hours or days. Here, the energies of solar protons are critical

because the degree of penetration of these protons is directly proportional to energy. Technologically, the effects of radiation dosage leading to the degradation of electronic systems is taken into consideration during design. Biologically, most of the SEP events observed in the last five decades have been relatively harmless to human health. In these scenarios, the radiation dose rates are below 10 mGy, demanding minimal protection that can be achieved via shielding. In other words, solar protons at energies <10 MeV per nucleon can be relatively easily shielded by spacecraft hulls to reduce exposure by attenuating the radiation and reducing the dose rates. However, the high density of protons with E≥10 MeV can be a particular concern for external operation beyond the Earth's magnetic field (Cucinotta et al. 2010; Xapsos et al. 2012). SEPs of E≥30 MeV, that have a high fluence of particles, can be of concern for thinly shielded habitats, e.g., during EVAs (Kim 2011). Most of the radiation risk to humans in space from SEPs is due to proton fluxes of E ≥50 MeV. At this energy, incoming solar protons (called 'soft' radiation) begin to penetrate the skin of spacecraft and the spacesuits of astronauts. Protons with E ≥150 MeV are considered 'hard' radiation since there are no possible measures to provide shielding (Reames 2021).

Biological responses to sudden and intense SEP radiation make protecting humans in space challenging (Barth et al. 2003). For future IP missions, such as NASA's mission to Mars, operational protocols will be crucial to any radiation safety program as the spacecraft will spend most of its time outside of the Sun-Earth line (Campa et al. 2009; Patel 2020; Ramos et al. 2023). Therefore, SEP predictions that apply to the Sun-Earth-Moon system are of limited value when the spacecraft makes a large angle with the Earth-Sun direction.

In this regard, operations on a human spacecraft going to Mars would involve measuring (1) the indicators of incoming energetic solar radiation and (2) the actual radiation dosage levels of the astronauts. The latter can only be achieved by installing one or more active dosimeters on the personnel or on nearby equipment, such as a rover or robotic assistant (Guetersloh & Zapp 2010), while the former requires a scaled-down version of the GOES-like instruments to produce warnings of intense solar radiation storms. Therefore, when there is a likelihood of an SEP event occurring, the EVA will require specific contingency plans such as seeking shelter or using available shielding. This is also applicable during the launch or other hazardous vehicle operational situations. Typically, a surface EVA on the Moon or Mars is estimated to be as long as eight hours. However, quick access (possibly less than an hour) to shielding is necessary to avoid possible excessive radiation exposure (Schrijver & Siscoe 2010). Hence, highly accurate forecasts of SEPs are essential to support operational requirements and towards safety of the astronauts and spacecraft equipment.

1.6 SEP Event Forecasting

Efforts in SEP event forecasting using predictive models generally focusing on solar protons have been undertaken for a long time (e.g., Smart & Shea, 1979). Nonetheless, the achievement of a highly reliable forecasting tool is currently far from reality. A recent review article on existing SEP forecasting models is presented by Whitman et al. (2023), which summarizes several models that are being developed in the SWx research community and categorizes them as physics-based, empirical, and machine learning (ML)-based models

depending on the methods implemented.

Physics-based models (e.g., Arge & Pizzo 2000; Aran et al. 2006; Schwadron et al. 2010; Strauss & Fichtner 2015; Marsh et al. 2015; Hu et al. 2017; Zhang & Zhao 2017; Linker et al. 2019; Tenishev et al. 2021; Li et al. 2021; Wijsen 2021; Kozarev et al. 2022) are developed based on the current understanding of the physical processes that lead to SEP events, including particle acceleration and transport at the Sun and in IP space. There are models providing their forecasts on NASA's SEP scoreboard[1] (e.g., Luhmann et al. 2007), while some are being deployed to the Community Coordinated Modeling Center (CCMC; e.g., Sokolov et al. 2004, Borovikov et al. 2015). Operationally, physics-based models cannot produce real-time efforts due to limited resources such as poorly constrained parameters, seed population spectra, and identifying the specific magnetic field line to predict the particle flux at the observer.

Empirical models use observational data to identify and characterize patterns between parameters related to SEP events and other precursive eruptions (e.g., Posner 2007; Richardson et al. 2018). Forecasts from these models are typically fast due to established relationships from historical data and are easily incorporated into operations on real-time observations. Certain models (e.g., Balch 2008; Bruno and Richardson 2021) focus on forecasting the SEP parameters such as peak flux, event fluence, and onset or peak of the event. For providing probabilistic forecasts (e.g., Falconer et al. 2011; Núñez 2011, 2015; Núñez et al. 2017; Papaioannou et al. 2018), typically, the goal is to predict the proton intensities crossing a certain threshold in the prediction window. Similarly, "All-Clear" forecasters (e.g., Sadykov

[1]NASA/CCMC SEP scoreboard: https://ccmc.gsfc.nasa.gov/scoreboards/sep/

et al. 2021) implement a specific decision threshold to provide a yes/no forecast.

ML approaches (e.g., Boubrahimi et al. 2017; Engell et al. 2017; Aminalragia-Giamini et al. 2021; Lavasa et al. 2021; Kasapis et al. 2022; Ali et al. 2024) investigate a variety of observational parameters of SEP precursors to obtain faster and more accurate forecasts. Similar to empirical models, these approaches use ML techniques on exisiting data to identify patterns or relationships between parameters of SEP events and other source eruptions. Typically, ML architectures do not consider the physics involved during modeling. The goal is to discover hidden patterns in the data that might not otherwise be evident from a simple statistical analysis. Models that predict the probability of occurrence and the expected peak proton fluxes of SEP events consider different energy ranges with different particle flux thresholds—starting from E$\geq$10 MeV (e.g., Stumpo et al. 2021; Torres et al. 2022) to E$\geq$50 (e.g., Kahler et al. 2017) and E $\geq$ 100 MeV (e.g., Papaioannou et al. 2022). These models are built to produce a rapid forecast with information accessible before or at the onset of an SEP event. Nonetheless, an ML framework generally requires reliable and high-quality data sets for model development, validation, and efficient comparison with new models. Such approaches have been undertaken in the heliophysics community to develop publicly available ML-ready benchmark data sets (Martens et al. 2017; Angryk et al. 2020) but are lacking in the area of SEP events primarily due to the scarcity of homogenous data.

Multi-module approaches involve coupled systems of different forecasts feeding predictive output from one module to another (e.g., Ji et al. 2021). These approaches mainly deal with linking the AR parameters to the likelihood of a large flare that could be followed by a

CME leading to a potential SEP event. Such systems have an advantage when they combine predictions from physics-based, empirical, and ML models, harnessing the powers of each approach to build a broader predictive tool. Common practices of predicting SEP parameters or providing an "All-Clear" forecast suffer from high challenges in achieving robustness. To better complement such models, short-term forecasts are relevant to put more emphasis on rightly predicting the possibility or absence of a strong SEP event. For this reason, it is critical that high-cadence, reliable, and accurate data streams relevant to SEPs are publicly available for the development and deployment of forecasting models.

1.7 Outline

The 'research to operations' aspect of SEPs forms the prime focus of the works undertaken for this dissertation. The main questions addressed are:

1. Can we develop an ML-ready benchmark data set of SEP events?

2. What solar source eruptions are triggering SEP events? Can we identify and statistically quantify these solar sources for all SEP events observed near Earth?

3. Can ML models that strongly rely on quality data be used to provide highly accurate predictions?

4. Other than long-term "All-Clear" forecasts, can we build a novel and competitive ML architecture to provide short-term forecasts?

The structure in the following chapters of this dissertation is presented below:

1.7.1 Chapter 2: Integrated Geostationary Solar Energetic Particle Events Catalog: GSEP

In Chapter 2, I present the development of a benchmark data set for SEP events covering three solar cycles from 1986 to 2018. Here, each SEP has been identified with a solar source eruption wherever possible. For this, I have utilized existing catalogs of flares, CMEs and SEPs. To cross-check and validate the source associations, I have manually examined observational data from multiple space-based missions such as GOES, SOHO and SDO.

1.7.2 Chapter 3: Analysis of SEP Events and Their Possible Precursors Based on the GSEP Catalog

In Chapter 3, I review and extend the GSEP data set by (1) adding "weak" SEP events that have proton enhancements below 10 pfu in the GOES E≥10 MeV channel and (2) improving the associated solar source eruptions information. Furthermore, I analyze the spatiotemporal properties of parent eruptions, such as flare magnitudes, locations, rise times, and speeds and widths of CMEs. I check for the correlation of these parameters with peak proton fluxes and event fluences. Based on the evaluation of our best models, I find both the flare and CME parameters are requisites to predict the occurrence of an SEP event.

1.7.3 Chapter 4: Short-term Classification of Strong Solar Energetic Particle Events using Multivariate Time Series Classifiers

In Chapter 4, I develop a framework using time-series-based ML models with the aim of developing robust short-term forecasts by classifying SEP events. For this purpose, I introduce an ensemble learning approach that merges the results from univariate time series of three

proton channels (E $\geq$10, 50, and 100 MeV) and the long-band X-ray flux (1–8 Å) channel from the GOES missions and analyze their performance. This study also focuses on understanding and developing confidence in the predictive capabilities of the models. Therefore, I utilize multiple evaluation techniques and obtain promising results.

1.7.4 Chapter 5: Precise and Accurate Short-term Forecasting of Solar Energetic Particle Events with Multivariate Time Series Classifiers

In Chapter 5, I examine the time-series ML framework by adding "SEP-quiet" intervals in the data set to introduce a natural class-imbalance between events and nonevents. I consider lead times from five min up to 60 min and extensively evaluate the feasibility of the methodology. To establish confidence in the predictive capabilities of the best model, I implement multiple forecasting metrics. The results from this work display great potential for transferring the modeling framework to near-real-time operations of short-term SEP event predictions to support deep space missions such as NASA's Artemis.

1.7.5 Chapter 6: Conclusion and Future Path

Lastly, in Chapter 6, I summarize the results of this work. I provide an outlook of how much more we can benefit from the use of ML for forecasting SEP events and discuss the many future paths awaiting exploration.

CHAPTER 2

Integrated Geostationary Solar Energetic Particle Events Catalog: GSEP

The contents of this chapter have been published in the Astrophysical Journal Supplement Series, Volume 262, Issue 29.

2.1 Introduction

Solar energetic particle (SEP) events are radiation storms of particle fluxes comprising electrons, protons, and heavier ions from the Sun. SEP events are known to originate in large eruptions such as solar flares (SFs) and coronal mass ejections (CMEs) (Reames 1999, 2013; Desai & Giacalone 2016). The number of SEP events occurring in any solar cycle (SC) varies and is much less than the number of SFs and CMEs that occur because of the acute directionality effects of SEPs and the fact that they are only detected in situ (Klein & Trottet 2001; Klein & Dalla 2017; Anastasiadis et al. 2019). The time intensities of particle fluxes are used to define and characterize SEP events. Enhancement above a nominal background level is considered to indicate a possible event. Such time profiles can be used to distinguish the source event as the temporal behavior appears to be different.

The energy of particles in strong SEP proton events can reach GeV in some instances (Reames 2001; Bruno et al. 2018), and these events can last from a few hours to several days (Kallenrode 2003; Klein & Psoner 2005; Kahler 2005; Cane & Lario 2006). These events have the capacity to disrupt spacecraft operations (Smart & Shea 1992; Pulkkinen 2007), and pose a hazard of radiation exposure to astronauts and aircraft traveling in polar routes where protection/shielding is limited (Beck et al. 2005; Schrijver & Siscoe 2010; Schwadron et al.

2010; Jiggens et al. 2019). Understanding the origin and propagation of SEPs is a formidable scientific challenge, and of crucial importance to space weather (SWx) research (Jackman & McPeters 1987; Gopalswamy 2003). In addition, as these hazards impose significant constraints on space-based activities for both humans and electronic equipment, predicting the event occurrences along with a sufficient advanced warning time is of vital importance to operations.

Multiple space and ground-based missions currently obtain in situ solar particle composition and energy spectra fluxes. Researchers have prepared catalogs of SEP events using the available in situ particle data. Flux measurements beyond the Earth's magnetosphere have been used to populate these catalogs. For example, Wind spacecraft data are used by Kahler (2005) and Miteva et al. (2018). Solar and Heliospheric Observatory (SOHO) mission data are used by Cane et al. (2010) and Paassilta et al. (2017). SEPServer (Vainio et al. 2013) uses data from Wind, SOHO and the Advanced Composition Explorer.

Solar proton event catalogs based on near-Earth observations such the Geostationary Operational Environmental Satellite (GOES) and Interplanetary Monitoring Platform (IMP-8) are of interest in this study. In Table 2.1, a list of existing SEP event catalogs utilizing near-Earth satellites is presented. Researchers who do not often work on data-processing issues can benefit enormously from a carefully integrated data set in testing theoretical or working hypotheses. Whether we consider data from active regions or time-series measurements or images of the full solar disk, cleaned and organized data sets are crucial when it comes to building SWx forecasting systems. To implement machine learning (ML) methodologies,

cleaned data sets are vital during the research phase because the unavailability of feasible data creates a shortfall to research-oriented approaches. Hence, careful integration of observational evidence backed up with theoretical reasoning is necessary while developing data sets. In addition, it is necessary to identify and correct errors, shortcomings, and caveats in the measurements and corresponding metadata because (1) data quality can impact the research output; and (2) it can mislead both model and data-driven analysis. To bridge the gap, comparisons and integration of data catalogs are critical for improving the performance of event predictions and the outputs of comparative scientific studies.

Table 2.1: The Consulted List of SEP Catalogs Based on the GOES Data.

Catalog	**Period (Events)**	**Threshold**		**Solar source**		
		Channel (MeV)	**Intensity (pfu)**	**SF**	**CME**	**AR**
Kurt et al. (2004)	1970 - 2002 (253)	>10	>10	Y	N	Y
Belov et al. (2005)	1975 - 2003 (1144)	>10	>0.1	Y	Y	Y
Gerontidou et al. (2009)	1996 - 2006 (368)	>10	>0.1	Y	Y	N
Dierckxsens et al. (2015)	1997 - 2006 (90)	>10	>0.1	Y	Y	N
Papaioannou et al. (2016)	1984 - 2013 (314)	>10	>0.5	Y	Y	Y
PPS (Kahler et al. 2017)	1986 - 2016 (138)	>50	>1.0	Y	N	N
CDAW-SEP [a]	1998 - 2017 (152)	>10	>10	Y	Y	Y
NOAA-SEP [b]	1976 onwards (266)	>10	>10	Y	Y	Y
RF-SPE [b]	1970 - 2019	>10	>1.0	Y	Y	Y

[a] https://cdaw.gsfc.nasa.gov/CME_list/sepe/
[b] https://umbra.nascom.nasa.gov/SEP/
[c] http://www.wdcb.ru/stp/solar/solar_proton_events.html

The Astroinformatics cluster at Georgia State University pursues data-driven research with particular solar physics applications (Angryk et al. 2020; Rotti et al. 2020a). One of the areas is SEP event forecasting. The tasks include integrating reference data sets, constructing metadata with well-defined statistical parameters derived from the measurements, and postprocessing. Efforts on SEP event predictions using ML have been ongoing over the last decade (Laurenza et al. 2009; Falconer et al. 2011; Engell et al. 2017; Stumpo et al. 2021; Kasapis et al. 2022). The GSEP data set to be discussed in this work supports the SEP predictions research area in two key perspectives, namely, in providing:

- metadata for the source active regions, associated flares, CMEs and radio bursts;
- time-series subsets of proton fluxes for the SEP event duration with an observation window of 12 hr.

This paper aims to bring together available SEP event catalogs based on GOES data as explained in Section 2.3. We have integrated a comprehensive list of SEP events with reference to their parent SFs and CMEs. The database comprises 341 SEP events, extending from 1986 to 2017. Section 2.4 describes the processes undertaken in data retrieval, preprocessing the GOES data, and the generation of the catalog under discussion. In Section 2.5, the results with observational details and minor differences between the catalogs are summarized. The purpose of this work is to provide the largest possible base for experimenting with statistical and ML models on SEPs and their solar-source (SF and CME) properties. Source eruptions can then be correlated with photospheric magnetic field and

metadata thereof to complement and physically/statistically connect SEP events with their solar, low-atmospheric progenitors.

2.2 Background

The National Oceanic and Atmospheric Administration (NOAA) continuously monitors the near-Earth space environment through GOES in geostationary orbit (Sauer 1989; Bornmann 1996). The GOES satellites record the solar activity and the in situ radiation environment. They usually operate in pairs with one satellite over the west coast and another over the east coast of the United States in geostationary orbit. NOAA classifies the two GOES satellites making parallel measurements as the "primary" and the "secondary" one. Over the three SCs from 1986 to 2017, 11 different GOES satellites have been launched and commissioned.

The GOES series carries various instruments, including the Space Environment Monitor (SEM; Grubb 1975). One of its constituent detectors is called the Energetic Particle Sensor (EPS; Onsager et al. 1996) on GOES-05 to 12. The twin EPS system on GOES-13 to 15 is called the Energetic Proton, Electron, and Alpha Detector (EPEAD). There are seven proton channels in the EPS/EPEAD taking in situ differential measurements with characteristic energies spanning from a few up to several hundred MeV (Sandberg et al. 2014). Furthermore, these channels are binned to seven nominal integral energies: P1 (>1 MeV), P2 (>5 MeV), P3 (>10 MeV), P4 (>30 MeV), P5 (>50 MeV), P6 (>60 MeV), and P7 (>100 MeV). However, GOES-09 and 14 missions, and channels P6 and P7 on GOES-12 have failed (Rodriguez et al. 2014). Nonetheless, measurements are available from as many

as nine GOES satellites, from GOES-05 to GOES-15.

NOAA's Space Weather Prediction Center (SWPC) provides radiation storm products based on proton intensity levels as observed by SEM's particle sensors (Rodriguez et al. 2014; Kress et al. 2020). The severity of the proton events is measured using the NOAA Solar Radiation Storm Scale (S-scale). SWPC's S-scale relates to biological impacts and effects on technological systems. The S-scale relies on the $\geq$10 MeV integral peak proton flux that characterizes an SEP event's size or intensity, although different peak fluxes logarithmically define different event sizes. The base threshold, associated with an S1 storm, corresponds to a GOES 5 minutes averaged $\geq$10 MeV integral proton flux exceeding 10 particle flux units (1 pfu = 1 particle/cm^2 sr s) for at least three consecutive readings (Bain et al. 2021). As can be seen in Table 2.1, many studies do not always conform to this definition because multiple enhancements or rises in the proton flux are considered in one SEP event. Differences in event definition occur due to different needs in research and operations, making it hard to achieve a harmonized data treatment.

2.3 Source Catalogs

We consider three SEP event catalogs developed using GOES data as sources: PSEP (Papaioannou et al. 2016), CDAW-SEP[1] and NOAA-SEP[2]. We classify the former two as "primary" and the latter as "reference" data.

[1]CDAW-SEP List: https://cdaw.gsfc.nasa.gov/CME_list/sepe/
[2]NOAA-SEP List: ftp://ftp.swpc.noaa.gov/pub/indices/SPE.txt

2.3.1 PSEP Catalog

Papaioannou et al. (2016) developed a catalog of 314 well-defined SEP events by statistically studying the relationship between SEP events and possible source eruptions such as flares and CMEs. For each event, they calculated the SEP onset times per event and per channel using the so-called σ *method* (for details, see Papaioannou et al. (2014)). This catalog is based on cleaned differential proton fluxes[3] from EPS made available directly by the Solar Energetic Particle Environment Modelling (SEPEM) Team (Crosby et al. 2015). The cleaned EPS data set spans over 40 yr (1974-2016) and has been cross calibrated by Sandberg et al. (2014) with data from the Goddard Medium Energy instrument on IMP-8. Papaioannou et al. (2016) define an SEP event based on the following threshold parameters:

1. a threshold of 0.01 particles/cm^2 sr s MeV (differential flux) above which a possible enhancement was marked;

2. a minimum peak of 0.5 pfu of the candidate event;

3. a waiting time of 2 hr between two consecutive candidate events;

4. a minimum event duration of 2 hr.

2.3.2 CDAW-SEP Catalog

The series of Coordinated Data Analysis Workshops (CDAW) was organized to analyze the set of all major SEP events (>10 MeV protons crossing the ≥10 pfu threshold) detected by NOAA's GOES spacecraft (Gopalswamy et al. 2002; 2003b). The CDAW-SEP list has 152

[3]SEPEM Reference Data Set (RDS): http://sepem.eu/help/SEPEM_RDS_v2-01.zip

events from 1997 to 2017 identified using integral proton data. The only criterion used for event selection was the peak proton flux crossing 10 pfu in the 10 MeV channel following the NOAA S1 standard. Each SEP event in the CDAW-SEP list, the associated flares and CMEs, and their properties are identified when available (Gopalswamy et al. 2003a; 2015). All the information is compiled and extended from an earlier report by Gopalswamy (2003; 2012). SEP events from SC 23 & 24 are studied by Gopalswamy et al.(2004a; 2004b; 2014), Mäkelä et al. (2015), Thankur et al. (2016) and Xie et al. (2016).

2.4 GSEP Events List

The preliminary data-processing and work structure in the integration and development of the catalog is illustrated in Figure 2.1.

2.4.1 GOES Data

We utilize the integral proton fluxes measured on board GOES-05 to 15 that are archived on the NOAA website[4]. Lower energy fluxes corresponding to the P1 channel were not used because of their high sensitivity to interplanetary disturbances. We performed a visual inspection of GOES data to understand flux enhancements and identify the more accurate observational sources. Although the design of EPS and EPEAD on board GOES has not changed, some variations in the measurements have occurred between satellites. As the instruments were built with passive shielding, measurements are affected by significant side- and rear-penetration effects, i.e., particles can pass through the shielding from any direction

[4]NOAA-SEM Data: https://www.ngdc.noaa.gov/stp/satellite/goes/index.html

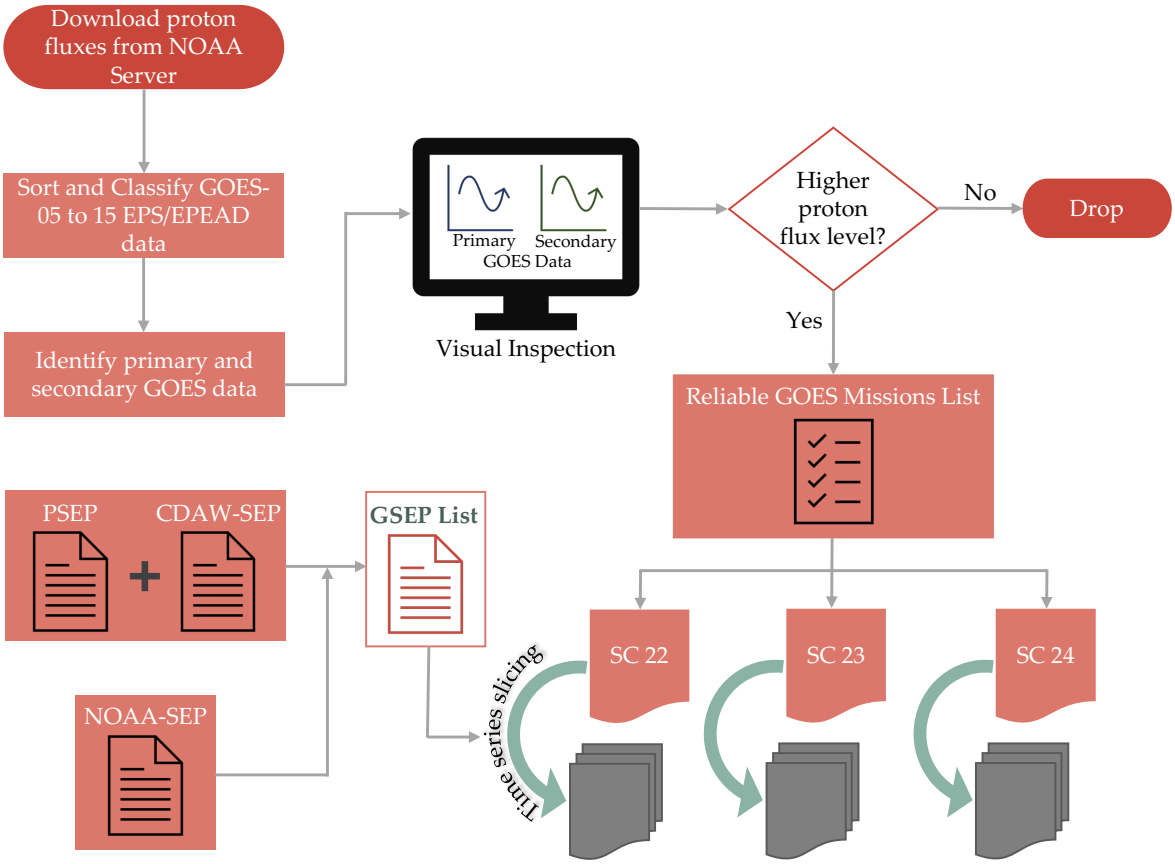


Figure 2.1 Process flow diagram of the background work.
Note: Indicating the GOES data inspection, verification, and slicing implemented in this work. The time-series slices are generated from the GSEP metadata. For details, see text.

and be counted as though they had entered through the nominal detector entrance aperture (Posner 2007; Bruno 2017). This is crucial as the differences in sensor data could impact the identification of an event and its timing in cases where the flux levels are near the event threshold. For instance, in the intercalibration of GOES-08 to 15 solar proton detectors by Rodriguez et al. (2014), it was reported that the relative responses between GOES primary and secondary agree to within ±20% while varying during a significant event. Based on a calibration study of the EPSs onboard GOES-05, -07, -08, and -11, Rodriguez et al. (2017)

validated the derived cross-calibrated energies by comparison with the STEREO data. In their study, they utilize the integrated proton fluxes calculated using the algorithm developed by R. Zwickl in 1989. See the Appendix of Rodriguez et al. (2017) for the details of the algorithm.

In light of this significant/known instrument-to-instrument variation, we carefully identified reliable missions by comparing the time-intensity plots of primary and secondary GOES instruments for all the observed event periods. To illustrate this, Figure 2.2 shows the difference in the flux-level enhancements between primary (GOES-06) and secondary (GOES-05) satellites for an SEP event during SC 22. Therefore, it is not a straightforward option to utilize data from the primary GOES satellite. The differences in the measurements of solar proton fluxes between the GOES primary and secondary are due to the geomagnetic cutoff, i.e., the effect of variation of the magnetic field configuration with geomagnetic longitude (Rodriguez et al. 2014). Therefore, we consider the strongest proton signal as the best for two reasons: (1) the corrected fluxes have been checked with intercalibration and (2) the peak values of the strongest signals closely match with those reported in the CDAW-SEP and NOAA-SEP lists. We performed additional data processing that ensured a compromise with imputable data gaps on better sensory responses of the instruments. That is, if the primary (secondary) GOES has a better response, but with more data gaps than the secondary (primary) GOES, then we consider the GOES primary (secondary) as a reliable data provider. We impute all the data gaps with linear interpolation. In addition, the EPEAD data were inspected for differences in enhancements between the “East” and “West” chan-

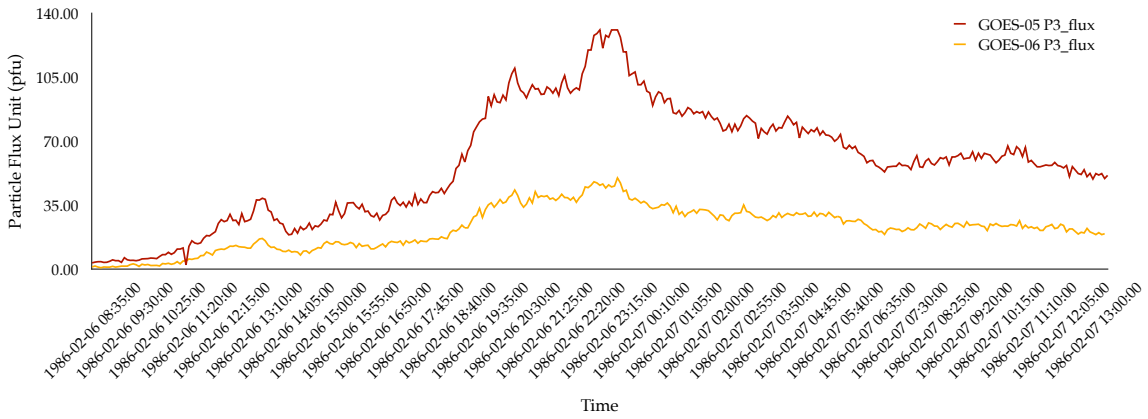


Figure 2.2 Time series of an SEP event showing the variation in the fluxes captured by the GOES-05 ("secondary") and GOES-06 ("primary") satellites.

nels. According to Rodriguez et al. (2010), the East-West effects are more relevant at lower energies. We have examined all SEP event temporal profiles and observed up to ±190% (in SC 22), ±90% (in SC23) and ±30% (in SC 24) differences between primary and secondary source energy channels.

2.4.2 Integration of Catalogs

The SEP events from the PSEP and the CDAW-SEP catalogs are integrated into Geostationary Solar Energetic Particle (GSEP) events list of this paper. In addition, we utilize the NOAA-SEP list as a reference catalog. That is, each event in the GSEP list is cross-checked with the reference catalog. Furthermore, a binary secondary source verification indicator is given in the metadata, where 0 represents that no source was found in the NOAA-SEP list, changing to 1 if found.

The PSEP and CDAW-SEP catalogs contain many valuable parameters related to temporal characteristics, integrated flux information, and solar-source metadata. Details on the

SEP events' start time, peak time, and peak-flux value in the >10 MeV channel are provided. They also report associations of SEP events with a parent solar eruption. Information such as the event coordinates about the associated flare and CME is provided as well. This information is used to determine if the entries in the PSEP, CDAW-SEP, and NOAA-SEP catalogs represent the same event, notwithstanding the minor differences in temporal characteristics, i.e., if they happened simultaneously with the same enhancements or if they vary and are eventually different events.

2.4.3 Challenges

Key challenges we had to address in integrating multiple catalogs were overlapping events, repetitions, and different criteria in event start time, peak time, and the corresponding peak fluxes. Different catalogs implemented different data calibration methods as well. To illustrate with the example of different onset criteria, a time-series plot of GSEP event 211 is shown in Figure 2.3. Here, PSEP considers the event onset prior to an M4.4 class solar flare and >29 hr ahead compared to the CDAW-SEP (2002 January 14 00:30:00 and 2002 January 15 05:35:00, respectively). The associated flare had a rise time of 58 minutes and is followed by a CME erupting behind the western limb. We take into consideration the start time as reported in CDAW-SEP as it accounts for the SWPC threshold of a significant SEP event. The event peak is observed on 2002 January 15 at 20:00:00 to reach a maximum of 15 pfu in the >10 MeV channel.

In the above example, although both source catalogs refer to the same event, the difference in start time is due to the criterion (like the event threshold) in considering a distinct onset.

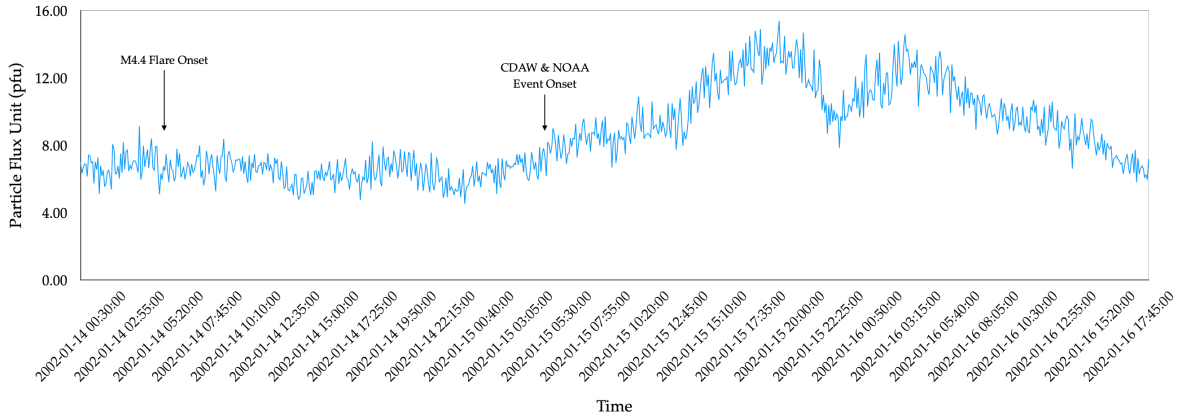


Figure 2.3 Time profile of an SEP event occurring between 2002 January 14 and 16. Note: The time series begins with the event start time (2002 January 14 00:30:00) according to the PSEP catalog. A solar flare of magnitude M4.4 occurs at 05:30:00 followed by a CME 6 minutes later from the western limb. The second arrow points to the onset of the SEP event (2002 January 15 05:35:00) as considered by the CDAW-SEP and NOAA-SEP catalogs. The event start time differs by >29 hr compared to PSEP. For this event, both catalogs report the same solar source (flare and CME). We retain the event onset as reported in the CDAW-SEP list.

Such dissimilarities between catalogs have been verified with the time profiles. Also, plots available from NOAA[5] were used to cross-verify our time-series plots and conclude whether an event occurred or crossed the NOAA threshold on a specific date.

2.4.4 Description of the Catalog

Our integrated catalog gathers SEP event records from multiple sources and provides relevant metadata useful for SWx research. The headers in the GSEP list and their descriptions are presented in Table 2.2. The majority of the SEP events have been captured by the primary GOES instrument. However, a total of 19 events were observed by the secondary instrument. Among them 13 events are in SC 22, and three for each of SCs 23 and 24. The final catalog

[5]NOAA Daily Plots: https://satdat.ngdc.noaa.gov/sem/goes/data/plots/

has observed source instrument flag “P” or “S” indicating whether the event was measured in the primary or secondary GOES instrument, respectively.

Table 2.2: Header Description in the GSEP List.

Header	**Description**
sep_index	Index for the GSEP events list
pp_id	Event identifier in the PSEP catalog
cdaw_sep_id	Event identifier in the CDAW-SEP list
timestamp	Start time of the event in PSEP
cdaw_start_time	Start time of the event in CDAW-SEP
cdaw_max_time	Event peak time in CDAW-SEP
cdaw_evn_max	Event peak flux in CDAW-SEP
cme_id	Identifier of the CME in LASCO CME catalog
cme_launch_time	Start time of the CME
cme_1st_app_time	First appearance time of the CME
lasco_cme_width	Width of the CME in Lasco Catalog
p_cme_width	Width of the CME in PSEP
lasco_linear_speed	CME velocity reported by LASCO
p_cme_speed	CME velocity reported by PSEP
fl_id	Autogenerated unique flare identifier
fl_start_time	Start time of the flare
fl_peak_time	Time of Flare Maximum
fl_rise_time	Time taken to reach peak
fl_lon	Longitude of flaring region
fl_lat	Latitude of flaring region
fl_goes_class	GOES Flare classification
noaa_ar	NOAA Active region number
noaa_ar_uncertain	Flag for NOAA active region number uncertainty
harpnum	HMI Active Region Patch (HARP) number corresponding to the originating NOAA AR
noaa_pf10MeV	Peak flux in the NOAA-SEP >10 MeV channel
ppf_gt10MeV	Peak flux in the PSEP >10 MeV channel
ppf_gt30MeV	Peak flux in the PSEP >30 MeV channel

Table 2.2—*Continued*

Header	**Description**
ppf_gt60MeV	Peak flux in the PSEP >60 MeV channel
ppf_gt100MeV	Peak flux in the PSEP >100 MeV channel
fluence_gt10MeV	Peak fluence in the PSEP >10 MeV channel
fluence_gt30MeV	Peak fluence in the PSEP >30 MeV channel
fluence_gt60MeV	Peak fluence in the PSEP >60 MeV channel
fluence_gt100MeV	Peak fluence in the PSEP >100 MeV channel
gsep_pf_gt10MeV	Peak flux in the GSEP >10 MeV channel
gsep_max_time	Event peak time in GSEP
m_type2_onset_time	metric type II radio burst start time
dh_type2_onset_time	decameter-hectometric (DH) type II radio burst start time
noaa-sep_flag	1 if present in NOAA-SEP; 0 otherwise
Inst_category	GOES Instrument category: P for Primary; S for Secondary
Comments	Retained from PSEP
Notes	Retained from PSEP
Fe_e_p_shock_notes	Retained from PSEP
gsep_notes	Data observational notes, if any
slice_start	Start time of the slice
slice_end	End time of the slice

2.4.5 Time Series Slices

The plots of time series slices from the GSEP list consist of:

- electron fluxes (channels E2 and E3, i.e., >2.0 MeV and >4.0 MeV);
- proton fluxes (channels P2 to P7, i.e., from >5.0 MeV to >100.0 MeV).

Here, the integral fluxes are derived from the observations of GOES/EPS from 1986 to 2012 and GOES/EPEAD from 2013 to 2017. Each value in time-series data represents a 5 minute interval. The length of each time profile denotes the events' start and end times.

These characteristics, which help describe the flux evolution and the data quality, provide visual information for selecting events for further analysis.

To summarize, we downloaded the GOES integral fluxes and classified the data into respective instruments. We visually inspected the primary and secondary observations to understand the overlaps, data gaps, and intensity variations. We merged the data in series for each SC. Finally, we sliced the GOES particle fluxes with reference to the onset/start time and the observed end time of each SEP event as defined in the GSEP list. The identification metadata is encoded in the file names of time-series data instances. We use the SEP event initiation as a default reference time. The best and simplest form we opted for contains the event date and time that correspond to the time stamp of the event onset; for example, 2017-09-10_04-25.csv). We also assign and maintain the correspondence between our index and the indices in the primary source catalogs.

2.5 Results

Integrating the primary catalogs, we have obtained 335 unique events. However, after comparing with the reference one, six more events were included. In the course of this work, we discussed with NASA's Space Radiation Analysis Group (SRAG) to validate our SEP events list. Hence, a total of 341 events are available in the GSEP catalog from 1986 to 2017. The time-series profile for each event has been visually inspected to confirm the event definition. Of these, 96 events fall under the weak enhancement category (peak flux <10 pfu at >10 MeV), while 245 events achieve a peak flux >10 pfu in the >10 MeV channel over the past

three SCs. In Table 2.3, the number of events is provided according to the flux enhancements in different levels of the NOAA solar radiation storm scale. Here, S0 is a custom scale used to denote a subevent of proton fluxes below 10 pfu. As the PSEP catalog has no records beyond 2013, the number of sub-events in the GSEP catalog for SC24 is lesser as compared to the prior two cycles.

Table 2.3: Number of SEP Events with Respect to the NOAA S-Scale Across the Last Three SCs.

Scale (pfu)	**SC22**	**SC23**	**SC24**
S0 (<10)	48	34	14
S1 (≥ 10 to $<10^2$)	49	61	31
S2 ($\geq 10^2$ to $<10^3$)	21	31	10
S3 ($\geq 10^3$ to $<10^4$)	14	13	06
S4 ($>10^4$)	03	06	00
Total	135	145	61

Note: S0 is a custom label to indicate a subevent.

An important property we want to address is the timescales of SEP events. We retained the criteria followed by the source catalogs PSEP and CDAW-SEP. During SC22, all events' start times are based on PSEP. For cycles 23 and 24, we choose CDAW-SEPs' criteria and switch to PSEP or NOAA-SEP when discrepancies occur. That is, for an event, whenever there is a CDAW-SEP onset available, it is visually verified to see if the temporal profile matches the event definition. If yes, then the CDAW-SEP onset is used. Else, we prefer

PSEP. In Figures 2.4(a-c), we present rise-time distributions for both weak enhancements and strong events (S1 and above) for the three latest SCs. It can be seen that the majority of the events last more than 6 hr over the rising phase, and several events take more than 24 hr to reach peak proton fluxes.

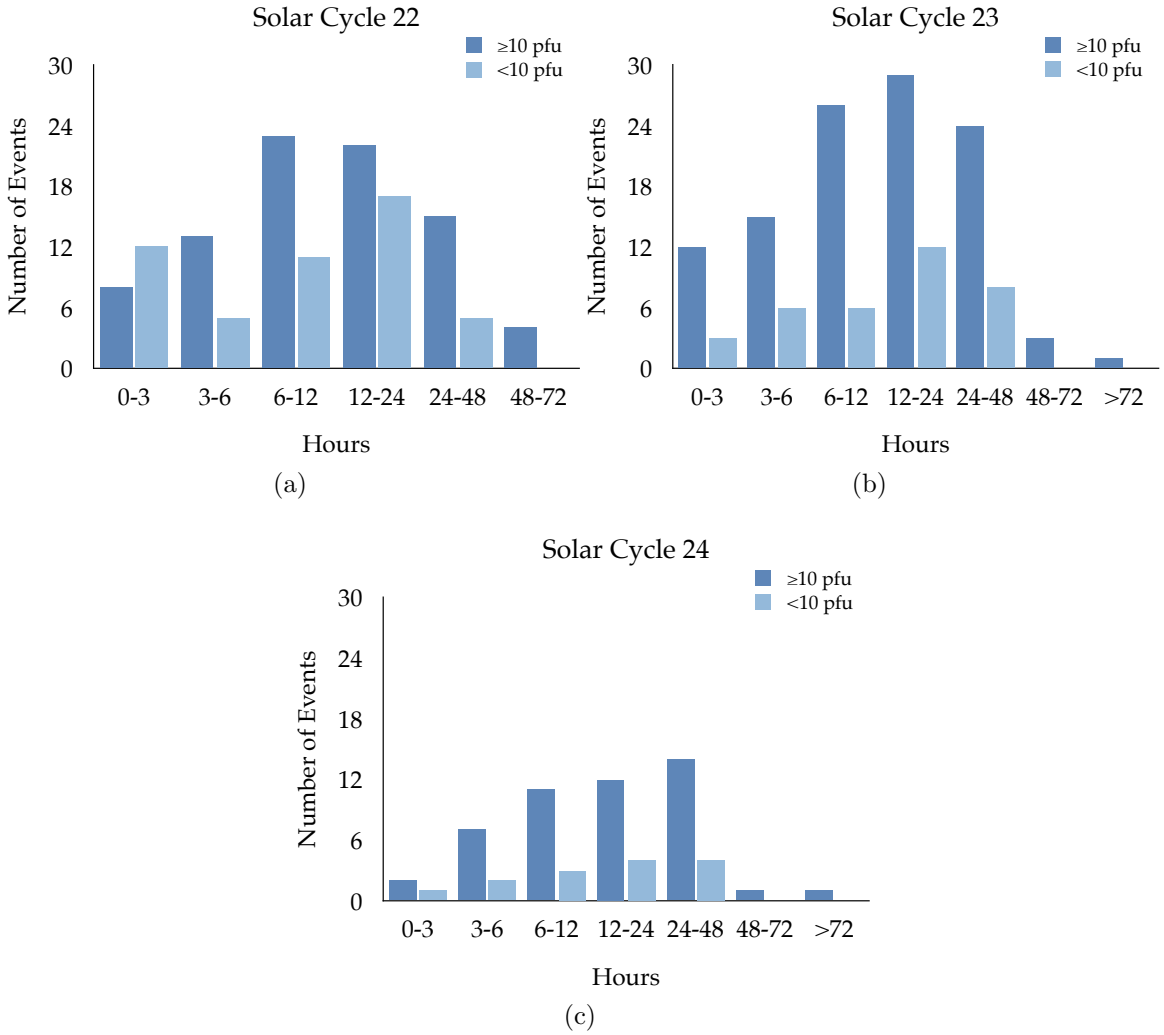


Figure 2.4 Histograms of rise times for the GSEP events.
Note: Rise times are times elapsed between onset and peak flux. Shown above are numbers of SEP events vs. their rise times, classified in bins of 0 - 3, 3 - 6, 6 - 12, 12 - 24, 24 - 48, 48 - 72, and >72 hr. Weak enhancements (<10 pfu for protons >10 MeV) and events above the NOAA S1 scale are included in (a-c) for SCs, 22, 23 and 24, respectively.

A great number of events have the rising phase predominant anywhere from 24 to 48 hr while some events take more than 48 hr to reach peak proton fluxes. Interestingly, this trend has reduced with the SC. In addition, a few strong events in the last two SCs take more than 72 hr to reach peak fluxes. In some cases with complicated SEP event temporal profiles, the peak flux occurs after an initial, or a pair of, peaks. Although initial peaks could directly reflect the parent (SF or CME) properties, later peaks may be due to particle transport effects (Kihara et al. 2020). Regardless, it is the proton fluence (i.e., the time-integrated flux) that determines doses which are crucial when Six effects are considered.

In the GSEP list, the NOAA active region number is available for 297 SEP events. Full source information (i.e., a flare and a CME) exists for 164 events. Nonetheless, 145 SEP events are associated with flares only, and 24 SEP events are associated with CMEs only. Because the necessary CME data is unavailable for SC 22, the numbers for CME–SEP association are less. For 309 SEP events where a source flare could be determined, 84 events are weak and 225 events are strong. There are two SEP events with no recorded flare peak time, five SEP events with no recorded GOES flare class, and 12 SEP events (nine in SC 22 and three in SC 23) with no reference to the flare location. There are eight SEP events (three in SC22 and five in SC23) in the GSEP list that do not have any source association. Three of these events have weak proton enhancements. Among the five strong SEP events, three are flagged as probable energetic storm particle events by SRAG.

In terms of a sensory response to measure protons, the peak flux recorded by the primary GOES instrument is better than the secondary in the majority of the cases. Nonetheless, the

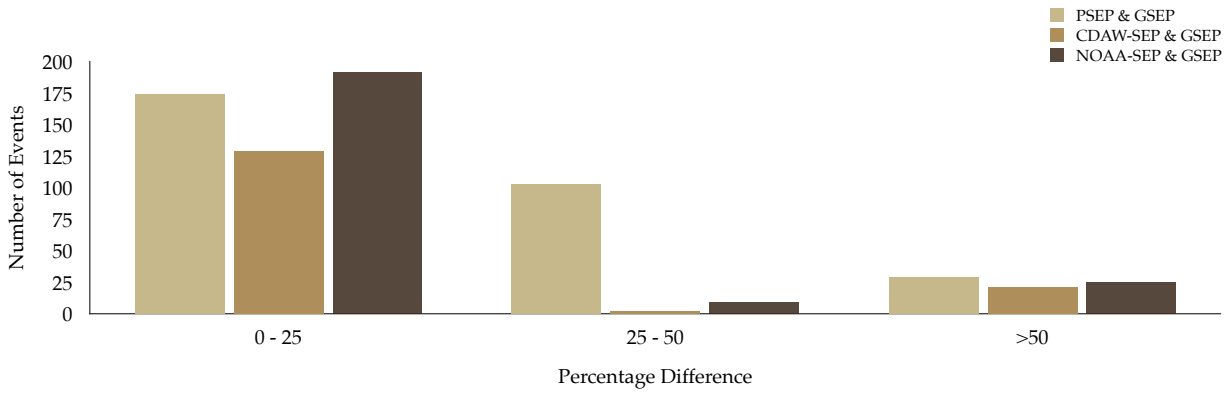


Figure 2.5 Distribution of the absolute percentage differences.
Note: The histograms are classified in bins of 0 - 25, 25 - 50, and >50% in the peak proton fluxes between GSEP and (a) PSEP, (b) CDAW-SEP, and (c) NOAA-SEP.

peak values reported by PSEP, CDAW-SEP, and NOAA-SEP differ from the GSEP metadata on several occasions. In Figure 2.5, the distribution of percentage difference comparing the GSEP list with the PSEP and CDAW-SEP catalogs is shown. A comparative summary between the GSEP and the PSEP and CDAW-SEP catalogs is given below:

- GSEP and PSEP lists:

    1. 280 out of 303 events from the PSEP catalog are within ±50% difference in the peak flux enhancements with respect to GSEP.

    2. There are nine events where PSEP records a peak flux of <10 pfu, but GSEP records >10 pfu. Out of these, five events are close enough with fluxes between 9 to 12 pfu, while differences in the remaining four events are significant. All these events are listed in Table 1 in Appendix A.

    3. PSEP event 185 (**psep185**) reports same proton fluxes as event 186 (**psep186**). This could be a possible computational or human error because the episode ap-

pears to be entangled. It is a relatively weak event and of short duration. However, we did not merge the two events because they are associated with distinct solar sources, both flares and CMEs. (See Appendix B).

- GSEP and CDAW-SEP lists:

  1. 126 out of 150 events have peak flux enhancements agreeing within ±20%.
  2. Slight discrepancies exist for extremely strong events (with peak proton fluxes at least >1500 pfu).
  3. Due to the variation in the event identification criteria, some of the event peaks have been missed by the CDAW-SEP.

- GSEP and NOAA-SEP lists:

  1. 189 events are within ±20% difference in the peak flux enhancements.
  2. Seven events with higher differences correspond to extremely strong events (of peak proton fluxes at least >1500 pfu).

2.6 Conclusions

We present an integrated Geostationary Solar Energetic Particle Events Catalog (GSEP) created from a set of available SEP event catalogs based on the particle fluxes of GOES missions from 1986 to 2017. We homogenized the SEP events from two primary catalogs (Papaioannou et al. (2016) and CDAW-SEP) by filtering all events, i.e., removing overlapping and repetitive episodes. Then we cross-checked the SEP events with the reference,

i.e., the NOAA-SEP list. Every entry in the catalog is assigned a new index for SEPs with reference to the indices of the source catalogs. The metadata provides an association of an SEP event to the corresponding source solar eruption, where available. The main summary of the paper is as follows:

1. There are 341 SEP events in the GSEP list. In that, 245 events have peak proton fluxes >10 pfu in the >10 MeV channel.

2. The particle fluxes of each event are visually inspected for errors and variations by parallel comparison of time profiles.

3. The fluxes are further sliced with respect to the event start and end times as reported in the GSEP metadata.

4. The headers in the GSEP list describe physical descriptors (both those stored in the source catalogs and calculated by us) and carry relevant indicators (data quality, observed GOES instrument, and parallel reports.)

5. The time-series slices are published as a data set to implement ML or other statistical analysis for experimenting on SEP event forecasting.

This work provides a catalog from which users can explore SEP events with parameters of interest for various statistical studies and ML exercises. Also, it provides a reference to various parameters for each event, allowing researchers to understand if the event satisfies the criteria for case studies. Our approach is to contribute to the SEP research community

with a combined database and present additional data for each event. The integrated GSEP catalog provides a one-stop database for researchers to study SEP events using an extensive, long-term data archive. Our GSEP dataset is available at Harvard Dataverse: https://doi.org/10.7910/DVN/DZYLHK. The plots and statistics presented in this study is based on version 4.0 of the dataset.

CHAPTER 3

Analysis of SEP Events and Their Possible Precursors Based on the GSEP Catalog

The contents of this chapter have been published in the Astrophysical Journal Supplement Series, Volume 267, Issue 40.

3.1 Introduction

High-energy particles are ejected out of the solar corona by shock waves (Gopalswamy et al. 2001) and due to the release of significant amounts of magnetic energy in solar active regions during the reconnection process (Lin & Hudson 1976). One aspect of such ejections is the solar energetic particles (SEPs) that predominantly constitute of protons along with electrons and other ions. SEPs are accelerated from the Sun to the interplanetary medium by eruptive phenomena such as solar flares (SFs) and shocks followed by coronal mass ejections (CMEs) (Cane et al. 1986; Kahler 1992; Reames 1999). SFs can cause radio blackouts, and CMEs heading toward Earth with a southward-pointed magnetic field (in the opposite direction of Earth's field at the Sun-facing side) can lead to geomagnetic storms (Pulkkinen 2007; Aparna & Martens 2020). Although SEP events are rare in numbers compared to SFs and CMEs (Klein & Trottet 2001), they have significant impacts in terms of space weather (SWx), including different levels of technological disruptions (Smart & Shea 1992) and biological perils on various economic scales (Schrijver & Siscoe 2010). Typical health risks include severe radiation hazards to astronauts taking spacewalks and airline travels on polar routes (Beck et al. 2005; Schwadron et al. 2010; Jiggens et al. 2019).

According to the Space Weather Prediction Center (SWPC), proton intensities ≥10 pfu (1 pfu = 1 particle per cm^2.s.sr) in the E>10 MeV energy channel are termed as strong SEP events with regards to causing significant SWx effects (Bain et al. 2021). With great advancements in space engineering and technology, space-based satellites have acquired near-continuous observations of SFs and CMEs in multiple wavelengths since 1997. For SEPs, in situ measurements show characteristic behaviors with energetic proton flux profiles rising to peak values in a few minutes to several hours (Kallenrode 2003, Kahler 2005, Klein & Posner 2005, Cane & Lario 2006). We will show further in our discussions how these flux profiles depend on the energy of protons, and the Sun–Earth connection. All these solar data have provided insights to infer a wide range of SWx consequences from solar eruptions (Gopalswamy et al. 2003b). One important aspect of analyzing solar data is to advance operational capabilities by mitigating SWx effects on our technological systems (Jackman & McPeters 1987). This requires the development of a robust tool to predict eruptive event occurrences.

Statistical analyses of SEP event properties have been carried out for decades by several researchers (Van Hollebeke et al. 1975; Laurenza et al. 2009; Dierckxsens et al. 2015; Papaioannou et al. 2016). A principal component analysis considering a combination of six solar variables (namely, the CME width/size and velocity, the SF longitude, duration, and rise time, and the logarithm of the SF magnitude) was presented in Papaioannou et al. (2018). The association of SEP events to a parent flare or CME has given us insights into particle acceleration and propagation in the interplanetary medium. Because SEP

events depend on source solar eruptions, gathering statistical evidence of their causation and acceleration is vital for the forecasting of SWx consequences.

One operational system that uses parameters of both SFs and CMEs to predict the occurrences of SEP events is the Forecasting Solar Particle Events and Flares (FORSPEF) tool (see Anastasiadis et al. 2017 for details.) In addition, great efforts are being made toward implementing the proven abilities of machine learning (ML) techniques to predict solar eruptive events (Camporeale 2019). Efforts on SEP event forecasting based on ML have been of a crucial focus globally with upcoming human missions to the Moon and Mars (Whitman et al. 2022). Best practices in ML comprise building robust models on high-quality data sets, referred to as benchmarks. The advantage of a carefully curated data set is that, with the right computational tools, it provides clues to understand and explore eruptive activities, during various phases (rise, peak, and descent) of the solar cycle.

Existing SEP event forecasting tools exploit the characteristics of source solar events (Posner 2007). Studies based on flare and CME parameters individually or a combination of both have been made in the last two decades (Kahler et al. 2007; Núñez 2011; Falconer et al. 2011; Núñez 2015). The earliest evidence of two distinct processes of particle acceleration leading to SEP events was obtained from radio observations (Wild et al. 1963). Over the last three decades, several researchers have put efforts into exploring SEP-associated slow drift rate type II and fast drift rate type III radio bursts statistically (Nicholson et al. 1978; Gopalswamy et al. 2002; Cliver et al. 2004; Balch 2008; Cliver & Ling 2009; Laurenza et al. 2009; Winter & Ledbetter 2015; Alberti et al. 2017).

In this work, we analyze and describe statistical correlations of SEP events with associated solar source eruptions. We explore the properties of SEP events from solar cycles 22, 23, and 24. We consider the times and peak intensities of flare fluxes, CME speed, and angular width, as they are considered to influence the SEP event populations. The longitudinal dependence of particle propagation from the solar surface is also studied. Although real-time radio data are unavailable, we briefly highlight its relevance as a vital element in forecasting efforts. Nonetheless, we defer our analysis of radio data to future work. We study feature importance of solar source parameters using two tree-based classifiers, namely, random forests and extreme gradient boosting models. We also evaluate the performance of our best models in a supervised binary classification schema on an undersampled subset of our data. This ML experiment forms a vital step to our future efforts in building robust models for SEP event forecasting. Our focus in the present study is to (1) provide an updated data set of SEP events and (2) explore the combined effects of including flare and CME parameters toward forecasting SEP events using a carefully curated benchmark data set. In this work, we consider the term "SEP events" analogous to solar proton events. The data set used in this study is described in Section 3.2. We discuss the methodologies and techniques used for source–event association in Section 3.3. We analyze the properties of solar source eruptions associated to SEP events in Section 3.4. We present our findings and results in Section 3.5, discussing the SEP event properties, correlations of SEP and source parameters, the feature importance study based on tree-based ML models, and the binary classification techniques implemented in our work. Last, in Section 3.6, we present the conclusions of our work.

3.2 Data Sources

The Space Environment Monitor (SEM) suite (Grubb 1975) on board the Geostationary Operational Environmental Satellite (GOES; Sauer 1989; Bornmann et al. 1996) obtains in situ measurements of SEPs, and these are made publicly available by the National Oceanic and Atmospheric Administration (NOAA) in their online database.[1]. We follow the definition of NOAA for a significant SEP event as crossing the 10 pfu threshold in the P3 channel. We consider classifying the SEP events into two categories based on the NOAA-SWPC threshold. That is, based on the integral proton flux (I_P) recorded by P3, we define as strong those events for which $I_P \geq 10$ pfu and weak events those for which $0.5 < I_P < 10$ pfu.

Our recently published geostationary solar energetic particle (GSEP) events catalog[2] has been the fundamental source of the present work. The GSEP data set stands as a large database of SEP events for solar cycles 22, 23, and 24 to the solar physics community. It was developed with multiple aspects: (1) integration of SEP events from existing databases, (2) identification and association of source solar eruptions with each SEP event, (3) including metadata that consists of physical parameters relevant to source eruptions, and (4) providing cleaned time series slices for each SEP event. In this work, we present the updates and improvements that were necessary to be made in the GSEP catalog in all above aspects, which we discuss further.

The primary part of the current work has been identifying and adding "weak" events to the GSEP data set. The description of our extended data set development follows the

[1]NOAA/GOES data archive: https://www.ngdc.noaa.gov/stp/satellite/goes/index.html
[2]GSEP Harvard Dataverse doi: https://doi.org/10.7910/DVN/DZYLHK

methodology of Rotti et al. (2022). The integrated $E \geq 10$ MeV proton fluxes corresponding to the P3 channel in the GOES-SEM have been used in this study. In the Appendix C, Table 2 lists the GOES series from 1986 to 2017 considered in the development of the GSEP data set. Here, the choice of the GOES missions depends on continuous data availability with fewer gaps and errors. In most cases, we considered the GOES proton data that recorded the largest intensity. However, exceptions are for inaccurate/invalid data where we visually verified measurements from each SEM instrument making parallel observations. In addition to proton fluxes, we utilize the one-minute averaged GOES soft X-ray (1–8 Å) fluxes measured by the X-Ray Sensor (XRS) on board GOES. The archived data is available online from NOAA's website.[3]

When there are uncertainties in solar source associations for SEP events, we examine observations ranging from extreme ultraviolet to X-ray images and data from the Solar and Heliospheric Observatory (SOHO) mission's coronagraphs (Brueckner et al. 1995). To examine and associate source solar eruptions, we utilize: (1) the GOES flare catalog,[4] (2) the CDAW CME catalog,[5] and (3) the CDAW type-II radio burst catalog.[6] The respective CDAW catalogs have been available since 1997 only. Hence, no data on CMEs and radio bursts have been included for cycle 22 in our data set. Also, we have not surveyed radio events observed from the ground.

[3]GOES-XRS: https://www.ncei.noaa.gov/data/goes-space-environment-monitor/access/avg/

[4]GOES Flare List: https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/xrs/

[5]CDAW-CME Catalog: https://cdaw.gsfc.nasa.gov/CME_list/

[6]CDAW-RB II Catalog: https://cdaw.gsfc.nasa.gov/CME_list/radio/waves_type2.html

3.3 Data Analysis

The severity of the proton events is measured using the NOAA Solar Radiation Storm Scale (S-scale)[7] that relates to biological impacts and effects on technological systems. The S-scale relies on the $E \geq 10$ MeV integral peak proton flux that characterizes the intensity of an SEP event. The base threshold, designated as "S1," corresponds to a GOES 5 minute averaged $E \geq 10$ MeV integral proton flux exceeding 10 pfu for at least three consecutive readings 2021. Further scales from "S2" to "S5" logarithmically increase from one another, therefore defining different event sizes.

Our data set consists of 433 events, wherein 244 (189) are strong (weak) in terms of SEP event intensity. Some events within data gaps of GOES observations have been verified by cross-referring SEP events from other catalogs (see Table 1 in Rotti et al. 2022) and interpolating the missing data. In certain scenarios related to SEPs, a passing interplanetary shock causes energetic storm particle (ESP) acceleration (Cane 1995). Nine events in our data set during the solar maximum are ambiguous in particle production, primarily due to the injection of fluxes from previous events. By visual inspection of the GOES proton data, a rise in the fluxes is noticed after the passage of a shock. The most reliable convention was suggested by the Space Radiation Analysis Group (SRAG) to consider such follow-up events as ESPs (S. Johnson, private communication).

Figure 3.1 shows the distribution of all the SEP events in our data set from 1986 to 2017. It can be noticed that the amplitude of the cycles in terms of SEP events has decreased over

[7]NOAA S-scale: https://www.swpc.noaa.gov/noaa-scales-explanation

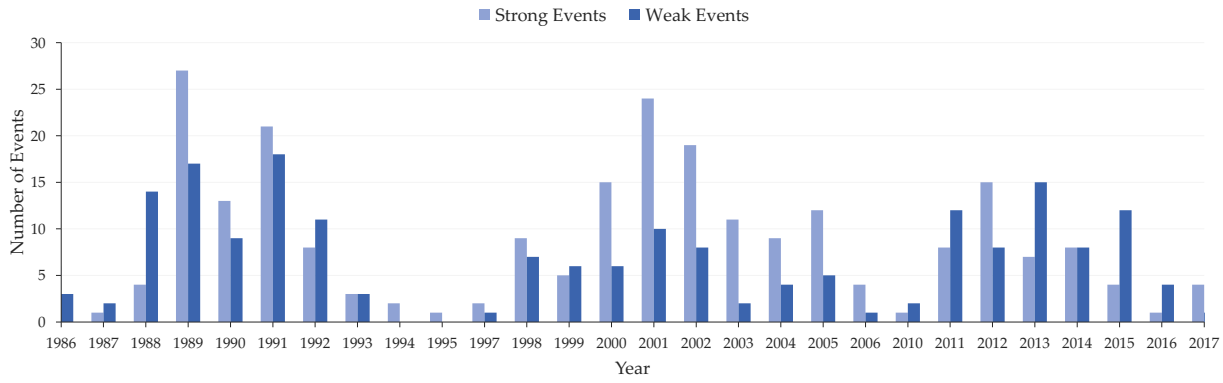


Figure 3.1 The strength of solar activity in terms of SEP events between 1986 and 2017. Note: A total of 163, 160, and 110 events have been identified in cycles 22, 23, and 24, respectively. In the legend, "strong" events are defined as those crossing the SWPC threshold of 10 pfu in the $E \geq 10$ MeV channel, and "weak" events are ≥ 0.5 pfu but <10 pfu.

the last three cycles, although cycles 22 and 23 are almost equal in strength while SC24 is less active. The percentage ratio between strong and weak events is as follows:

1. SC22 – 52:48
2. SC23 – 68:32
3. SC24 – 44:56

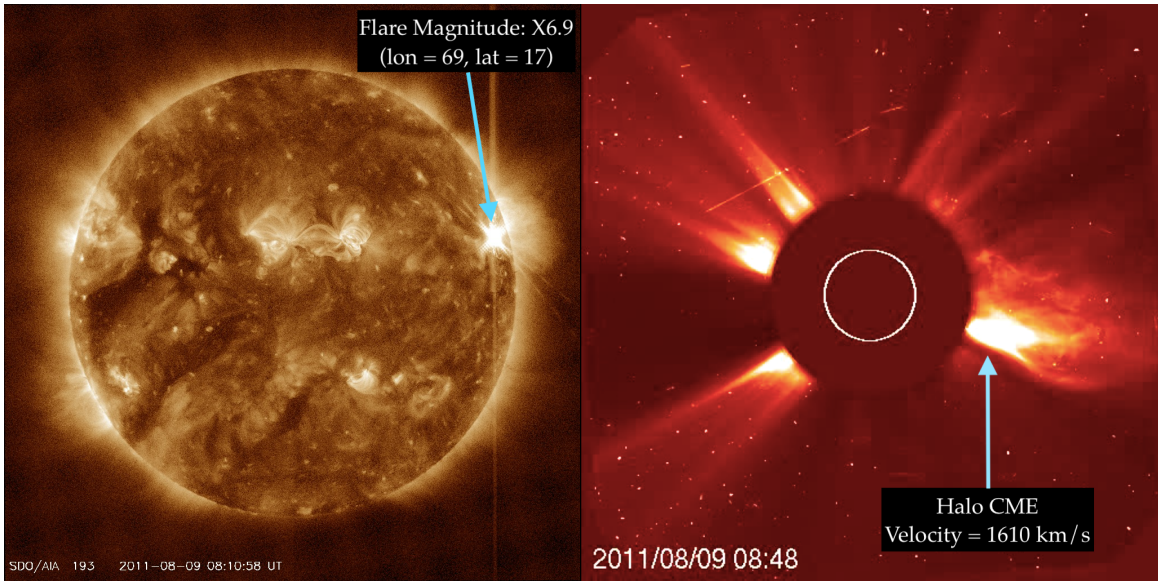
Because of NOAA's definition, there are a few events in our data set that are very close to being considered strong SEP events. We observe eight such events to have peak proton fluxes fluctuating between 10 and 12 pfu in the $E \geq 10$ MeV channel. Similarly, there are 12 events with peaks oscillating between 12 and 13 pfu. The proton measurements in all the scenarios are above the 10 pfu threshold for three consecutive points. However, on estimating the fluence levels, we notice they are below the threshold of a radiation hazard. Hence, we consider these 20 cases as "weak" SEP events.

We have put effort into identifying, for each SEP event in our data set with at least one eruption, either a flare or CME. As we will show later on in the paper, a probable precursor occurs within 12 hr of SEP onset for many events in our data set. The temporal profiles of SEP events positively correlate with longitude, with a few exceptions. Eastern hemisphere events typically have gradually rising proton profiles, whereas western events reach peak fluxes within a few hours of the parent flare eruption. We iteratively identified consistent source associations for exceptional SEP events. However, nine SEP events in our data set could not be associated with any solar source information. This probably constitutes a few far-side events that lack observational evidence to match with a source eruption.

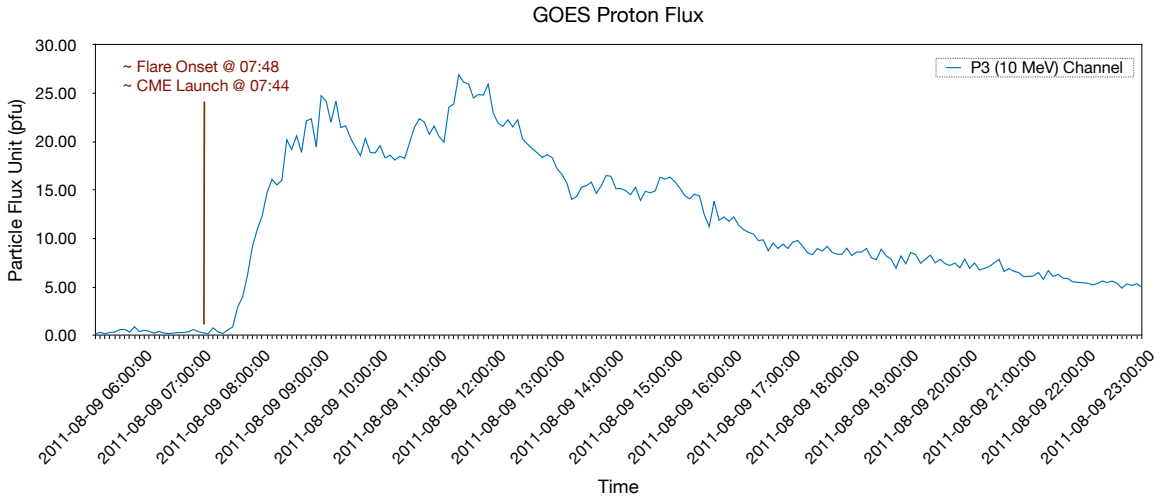
3.3.1 Solar Source Selection

For most cases in our data set, the association of an SEP event with a source is straightforward. Many active regions giving rise to a flare or CME are distributed on the visible disk of the Sun, while a few are on the backside. For those beyond limb and backside events, an estimate of the likely source active region has been made in reference to existing methods. A simple technique was to follow the heliographic longitudes of active regions on the visible disk that were previously sources of intense flares and SEPs.

Figure 3.2 shows the parent eruptions and associated and SEP event on 2011 August 9 indexed as **gsep**_292 in the GSEP data set. In Figure 3.2(a) is the image of the Sun showing an instance of the source flare eruption taken in 193Å by the Atmospheric Imaging Assembly (AIA) on board the Solar Dynamics Observatory (SDO; Lemen et al. 2012). The flare originated from active region 11263 (lon = 69, lat = 17) and had a magnitude



(a) (b)

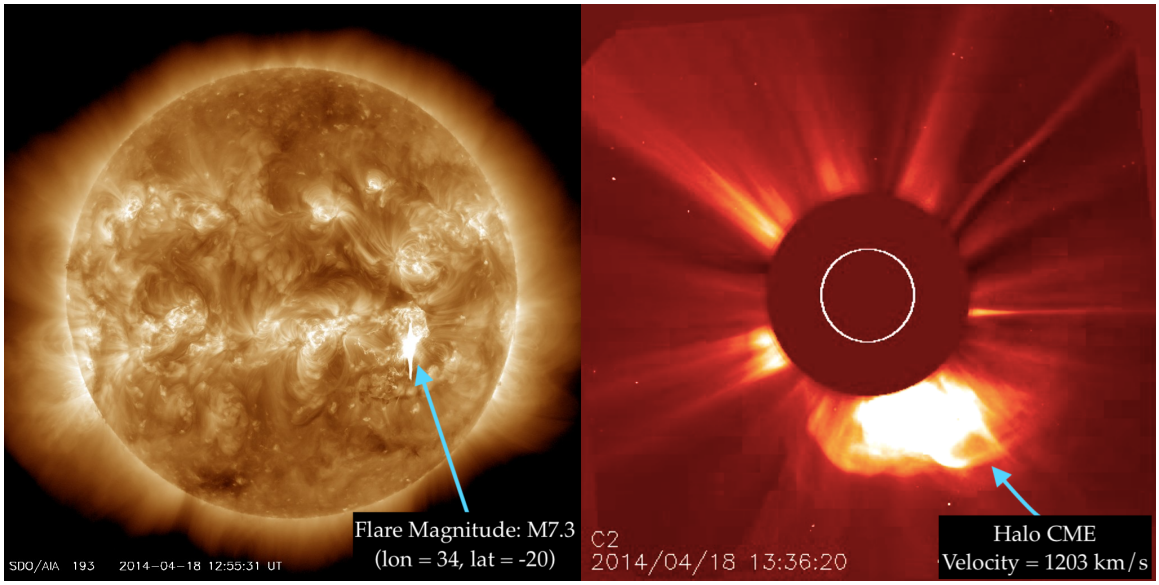


(c)

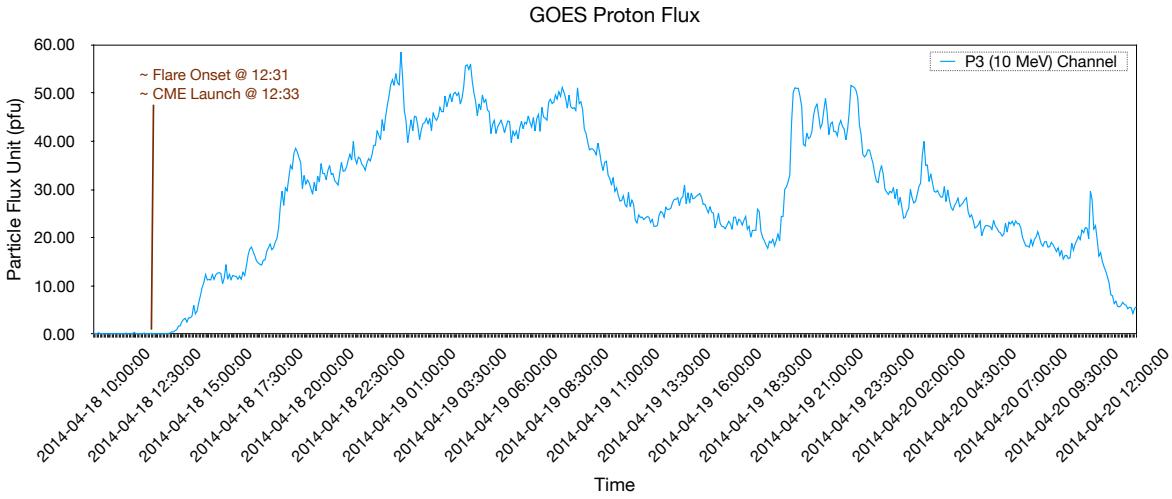
Figure 3.2 Plots of parent eruptions followed by an SEP event on 2011-08-09 indexed as **gsep**_292 in the GSEP data set.
Note: (a) A solar flare of magnitude X6.9 seen in 193Å erupted from active region 11263 with spatial extents: lon = 69, lat = 17. Image courtesy of SDO/AIA. (b) A halo CME traveling at a mean speed of $\approx$ 1610 km.s^{-1} was launched less than four minutes after the parent flare. Image courtesy of SOHO/LASCO-C2. (c) Proton intensity profile plot of the SEP event with a peak flux of 26 pfu in the $E \geq 10$ MeV integral channel measured by the GOES-SEM suite.

of X6.9 as measured by the GOES/XRS instrument. Following the flare was a halo fast CME propagating with a velocity of ≈1610 km.s^{-1} shown in Figure 3.2(b) as observed by LASCO/C2 instrument several minutes after the launch time. In the frame shown, we can see the evolving CME cloud (bright region) out of the solar corona and the shock-front is visible at the side of the CME. There was also a DH type II radio burst observed ∼35 minutes after the CME eruption. In Figure 3.2(c) is the time series plot of the subsequent strong SEP event with a peak flux of 26 pfu in the $E \geq 10$ MeV channel measured by the GOES-SEM instrument. The vertical line overlaid in the plot indicates an approximate start time of the parent eruptions. The onset of this SEP event occurs at 08:45 with a rise time of 3 hr and 25 minutes. The event lasted for more than 8 hr until it subsided below the threshold of 10 pfu in the P3 channel.

Figure 3.3 shows the parent eruptions and associated SEP event on 2014 April 18 indexed as **gsep**_330 in the GSEP data set. Figure 3.3(a) is the image of the Sun showing an instance of the source flare eruption taken in 193 Å by the SDO/AIA instrument. The flare erupted less than 3 hr before the SEP event onset from active region 12033 (lon = 34, lat = -20) and had a magnitude of M7.3 as measured by the GOES/XRS instrument. Following the flare was a halo fast CME propagating with a velocity of ≈ 1203 km.s^{-1} shown in Figure 3.3(b) as observed by the LASCO/C2 instrument several minutes after the launch time. In the frame shown, we can see the evolving CME cloud (bright region) out of the solar corona and the shock front is visible around the CME. There was also a DH type II radio burst associated with this CME. Figure 3.3(c) shows the time series plot of the subsequent strong



(a) (b)



(c)

Figure 3.3 Plots of parent eruptions followed by an SEP event on 2014-04-18 indexed as **gsep**_330 in the GSEP data set.
Note: (a) A solar flare of magnitude M7.3 erupted from active region 12033 with spatial extents: lon = 34, lat = -20. Image courtesy of SDO/AIA. (b) A halo CME traveling at a mean speed of $\approx$ 1203 km.s^{-1} was launched less than three minutes after the parent flare. Image courtesy of SOHO/LASCO. (c) Proton intensity profile plot of the SEP event with a peak flux of 58 pfu measured by the GOES-SEM instrument in the $E \geq 10$ MeV integral channel.

SEP event with peak flux of 58 pfu in the $E \geq 10$ MeV channel measured by the GOES-SEM instrument. The vertical line overlaid in the temporal plot approximately indicates the start time of the parent eruptions. The onset of this SEP event occurs at 15:25 (UT) with a rise time of 9 hr and 40 minutes.

3.3.2 Metadata

The metadata in the updated GSEP catalog consist of corrected and carefully associated source eruptions. In the latest version, proton fluence estimates are provided for each SEP event. We also include a "Flag" column to indicate "1" or "0" representing "strong" or "weak" events, respectively. Complete source solar eruption information is available for as many as 144 SEP events in our data set. In Table 3.1, we present the total number of possible solar sources for each SEP event.

Table 3.1: Total SEP Events and Their Solar Sources in the GSEP Catalog.

	Strong Events (≥10pfu)			Weak Events (<10pfu)		
	SC22	**SC23**	**SC24**	**SC22**	**SC23**	**SC24**
SEP Events	86	110	48	77	50	62
ARs	78	97	46	65	43	52
Flares	84	97	43	73	45	44
CMEs	-	97	48	-	41	61
Type-II RBs	-	75	35	-	21	16

A summary of our data set is as follows:

1. We obtain 386 cases of flare information for associated SEP events.

2. There are 16 flaring events without an associated CME and active region data.

3. We notice 18 CMEs with no supportive observational relations to flares and active regions.

4. Three CME events have no information about their widths and speeds.

5. There are 147 type II radio bursts associated with SEP events.

6. We cross-verified the possible ESP events in our data set with the list by Huttunen-Heikinmaa & Valtonen (2009) and find three (**gsep**_167, **gsep**_216 and **gsep**_223) "confirmed" and two (**gsep**_208 and **gsep**_209) "probable" ESP events. The remaining four events (**gsep**_146, **gsep**_164, **gsep**_168 and **gsep**_180) are not reported in the former list. Nonetheless, we retain all nine of these events in our catalog but indicate them as "ESP" events under the "*gsep_notes*" column.

7. Two strong and weak SEP events each have no observational data on source solar eruptions.

8. No flare has been reported on 1991 May 20 by NOAA. Hence, no source association is available for the weak SEP event observed that day.

3.4 Parent and SEP Event Properties

The initiation of an SEP event depends on either a flare or CME eruption or both. We observe two scenarios here: (1) there is at least a solar flare as a precursor to the SEP event or (2) there is a CME only as a precursor to the SEP event. The onset of a flare, its magnitude, and its location may influence the extent of the rise of proton fluxes and the respective fluences near Earth. Some SEPs connected with a flare and a CME typically show large fluxes near Earth even when they originate from the eastern side of the Sun. This could be partially due to the magnetic field lines getting stretched by the CMEs while the shock fronts sweep the particles and accelerate them in the interplanetary region.

We find 41 SEP events with a parent flare that erupted before 12–24 hr of SEP event onset. In a general situation where there are two or more precursors, a flaring event is always initiated before the launch of the corresponding CME. Temporal measurements of CMEs such as launch time and their first appearance are subject to errors. However, pertinent data were considered in our analysis based on observational evidence along with those reported in the CDAW catalog.

3.4.1 Solar Flares

We utilize the NOAA-GOES flare list that provides data on the flare's onset, peak, end time, peak X-ray flux level (labeled as GOES class), coordinates of the flares and associated active regions based on in situ soft X-ray (1–8 Å) measurements by the GOES/XRS. Using the start and peak times of the flaring events reported in the NOAA-GOES flare list, we

estimate the flare rise time, i.e., the number of minutes taken to reach maximum X-ray intensity. In the present work, erroneous flare associations were overcome based on the event temporal information, i.e., flares must occur before an SEP event. In addition, the proton flux enhancements had to be observed during the soft X-ray emissions, and the flare events were generally of long duration and intensity. Exceptions were cross verified by visually checking the observational data.

3.4.1.1 Active Region Locations

Measurements of SEP events near Earth depend on the spatial region of source eruption on the Sun. Generally, it is understood that the eruptions at the western side of the Sun have a higher probability of energetic particles reaching near-Earth space due to the spiral structure of the interplanetary magnetic field lines, popularly known as the Parker spiral (Parker 1965; Reames 1999). Nonetheless, many SEP events have been detected arising from widely spread locations on the Sun. Figure 3.4 shows a cylindrical projection of flare coordinates that have given rise to strong SEP events between 1986 and 2017. The location values are plotted here as a function of the NOAA S-scale to aid in understanding the source location distribution and the corresponding SEP event peak intensity observed at Earth. It can be seen here that S1 and S2 level proton event origins appear concentrated on the western hemisphere but also are widely spread across the solar disk. A few S3-level SEP events originate from the eastern hemisphere and a few more from the western limb and regions beyond. Interestingly, three S4-level SEP events have been reported in SC22, six in SC23, and none in SC24. For these events, the minimum rise time is roughly 16 hr, with a mean of 26 hr to reach peak

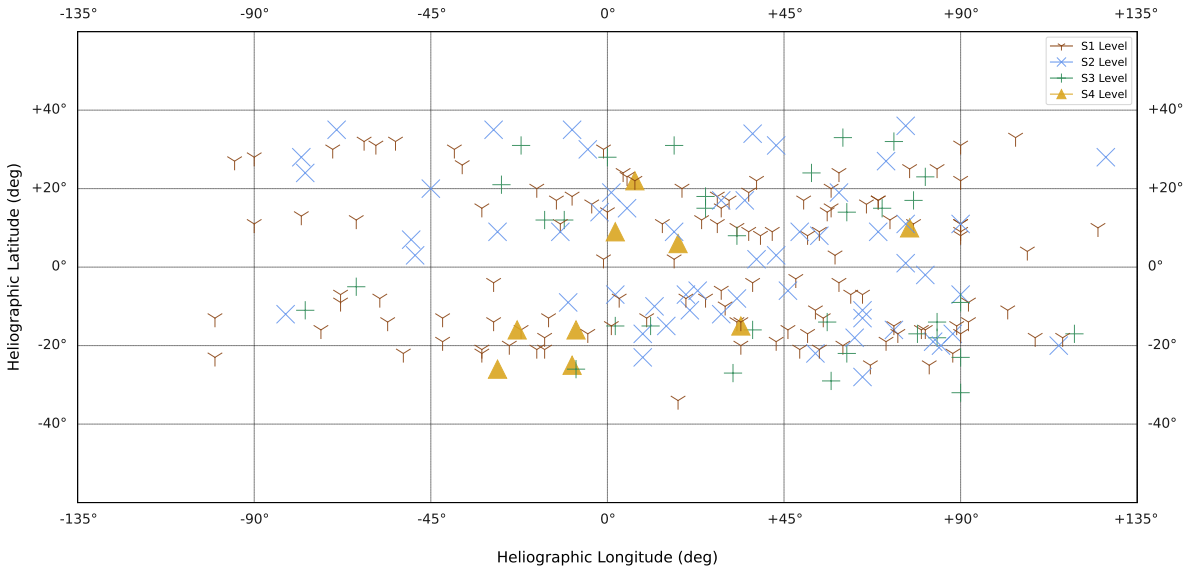


Figure 3.4 Spatial variations of all the strong SEP events in the GSEP dataset across the solar disk.
Note: The location values are plotted here as a function of the NOAA S-scale.

fluxes. All nine events originated from magnetically well-connected regions between the Sun and Earth. However, the high intensity of S4 events and their large temporal extents here contradicts the commonly believed notion of SEP events from the western side displaying fast rise and short duration compared to eastern events.

For missing location values, we made efforts to gather information from additional sources such as other catalogs and images/movies from observations as mentioned earlier in Section 3.2. However, when there are inconsistencies and illogical conclusions, we refrain from including such erroneous information in our data set. For example, an SEP event on 1992 March 7 at 15:50 occurred over the eastern limb and is associated with a C5.0 flare with no report of an accurate flare onset. Another strong SEP event on 1991 March 29 at 21:20 has no location information for a possible flare in the GOES flare list.

A weak SEP event on 2002 November 1 at 18:40 arises from the far side and lacks observational evidence to match with a source eruption. Figure 3.5 shows a scatter plot of solar flare coordinates associated with weak SEP events between 1986 and 2017. Here, the location magnitudes are plotted as a function of the GOES flare class to understand the dependence of SEP peak fluxes at Earth on X-ray flare peak intensities. We note that the differences in sizes of the tick marks correspond to the GOES flare level. Apart from a few C-class flares in the far eastern hemisphere, most of the flaring regions are present from -30°to the far western side. A similar trend can be noticed for M-class flares, except for those few beyond -45°. However, X-class flare locations are widely spread across the Sun. Despite the significant flare intensities, associated proton flux levels remain <10 pfu near Earth. On comparing the two scatter plots in Figures 3.4 and 3.5, it appears that there are several factors from the upper solar atmosphere to the corona, including the footpoints, that dominate the release and acceleration of SEPs from the Sun.

Figure 3.6(a) shows the distribution of longitudes for 383 SEP-associated flaring events in our data set. The seven bins are calculated based on the number of events between a pair of longitudinal extents, namely, [< -90], [≥-90 to < -60], [≥-60 to < -30], [≥-30 to < 0], [≥0 to < 30], [≥30 to < 60], [≥60 to < 90], and [≥90]. There is an increase in the number of events as a function of flare longitude with strong preferences towards the western hemisphere.

3.4.1.2 Flare Magnitudes

Out of 433 events in our data set, there are 386 flare-associated SEP events. Here, 383 (220 strong and 163 weak) events have GOES class information available. Of them, 121 (32%)

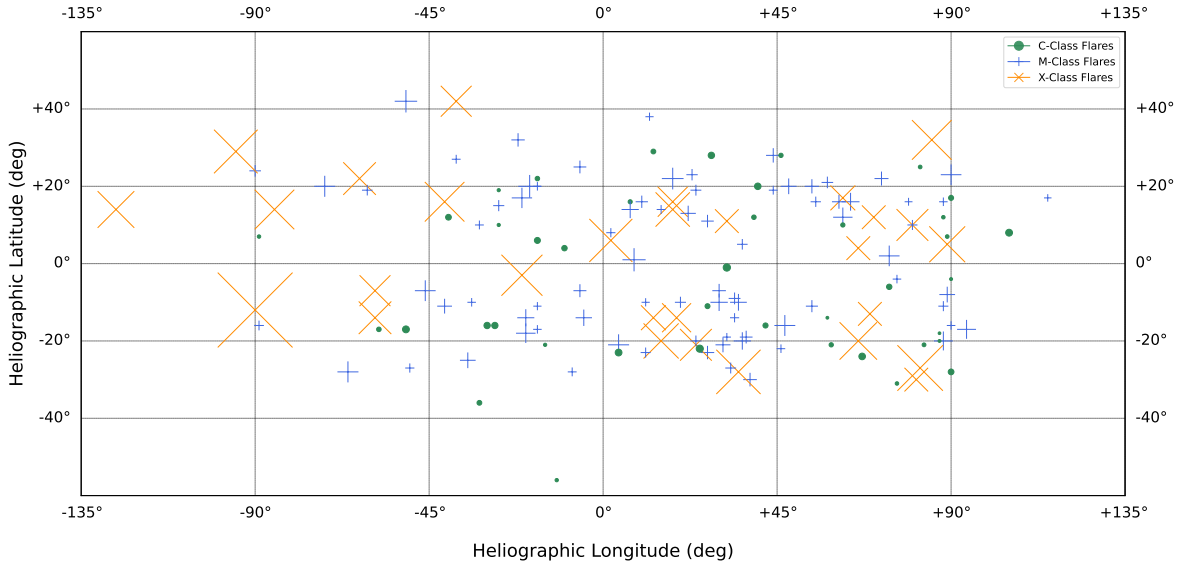
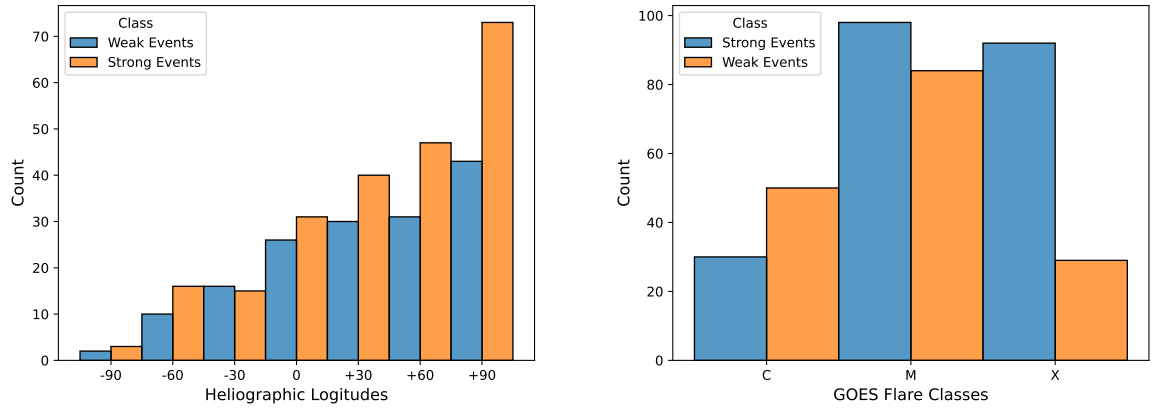


Figure 3.5 Spatial variations of all the weak/sub-SEP events in the GSEP dataset across the solar disk.
Note: The location values are plotted here as a function of peak solar flare intensity. The X-ray flare measurements are obtained from the GOES missions.



(a) (b)

Figure 3.6 Distribution of SEP-associated solar flare parameters
Note: Shown are (a) longitudinal variation of solar flare locations that are associated with SEP events, and (b) soft X-ray flare peak intensities based on GOES flare classification for flares associated with SEP events. The two classes of SEP events (strong and weak) are shown in the legend.

are associated with X-class flares, while 182 (48%) and 80 (20%) are associated with M- and C-class flares, respectively. Figure 3.6(b) shows the distribution of flare-associated SEP events where bin sizes correspond to the GOES flare classification. Weak SEP events show a closely normal distribution for source flare magnitudes. However, strong events show a higher probability, with M-class flares followed by X-class ones. strong SEP events have 92 X-, 98 M-, and 30 C-class flares associated, where the median flare class is M2.0. Weak SEP events have 29 X-, 84 M-, and 50 C-class flare associations.

3.4.2 Coronal Mass Ejections

The CDAW provided a list of CMEs as observed by the SOHO coronagraphs. Estimations and descriptions of various parameters, including CME speed and angular size have been made by Gopalswamy et al. (2009) and are available from the official CDAW website. Since 1997, LASCO has observed that a CME preceded every SEP event except for 18 cases, apart from five more events when there was sustained activity in 1998 (Yashiro et al. 2004).

3.4.2.1 Source CME properties.

In our data set, 247 SEP events have been identified with precursor CMEs. Here, 162 CMEs are halo with a median speed of 1366 km.s^{-1}. About 144 flares have been observed to associate with these halo CMEs, and 114 of them resulted in strong SEP events as measured near Earth. SEP events associated with non-halo CMEs are 85 in number, with a median speed of 767 km.s^{-1}. We obtain the CME speeds from the CDAW CME catalog publicly available online.

In Figure 3.7, we show the variance of logarithmic peak proton flux with respect to linear CME speed for (a) strong and (b) weak SEP events in our data set. The distribution consists of a linear regression trend fit on the data and shows a 22% and 1% variance (R^2), respectively. In addition, logarithmically transformed proton fluences of strong and weak events are shown in subsets (c) and (d), respectively. Here, the linear regression fit on the distribution shows ≈30% and 3% variance (R^2), respectively. As evident from the figures, CME speeds cannot be fully deterministic about SEP events, especially the weak class.

Shocks from halo CMEs are dominant features, as they stretch the interplanetary magnetic field, thereby expanding the horizon of field lines, which in turn could cause the path of accelerating particles to stay away from Earth (Desai & Giacalone 2016). Strong SEP events (≈70%) in our data set are predominantly associated with halo fast CMEs. Figure 3.8(a) shows the distribution of the speeds of CMEs that are associated with SEP events. Weak SEP events have a median (mean) speed of ≈905 (835) km.s^{-1}, while strong events have 1522 (1444) km.s^{-1}. It is worthwhile to mention here that the measurements of CME speeds suffer from projection effects. Nonetheless, several researchers have made efforts to infer as much accuracy as possible (Yashiro et al. 2004; Gopalswamy et al. 2009). The number of SEP events depending on the associated CME widths is shown in Figure 3.8(b). The bins correspond to ¡60, 60–120, 120–180, 180–240, 240–300, and the halo (360) measured in degrees. A median (mean) of 281°(246°) is noted for weak events. In our study, we find most (66%) of the CMEs associated with SEP events are halo in angular extents and fast with a median speed of 1200 km.s^{-1}.

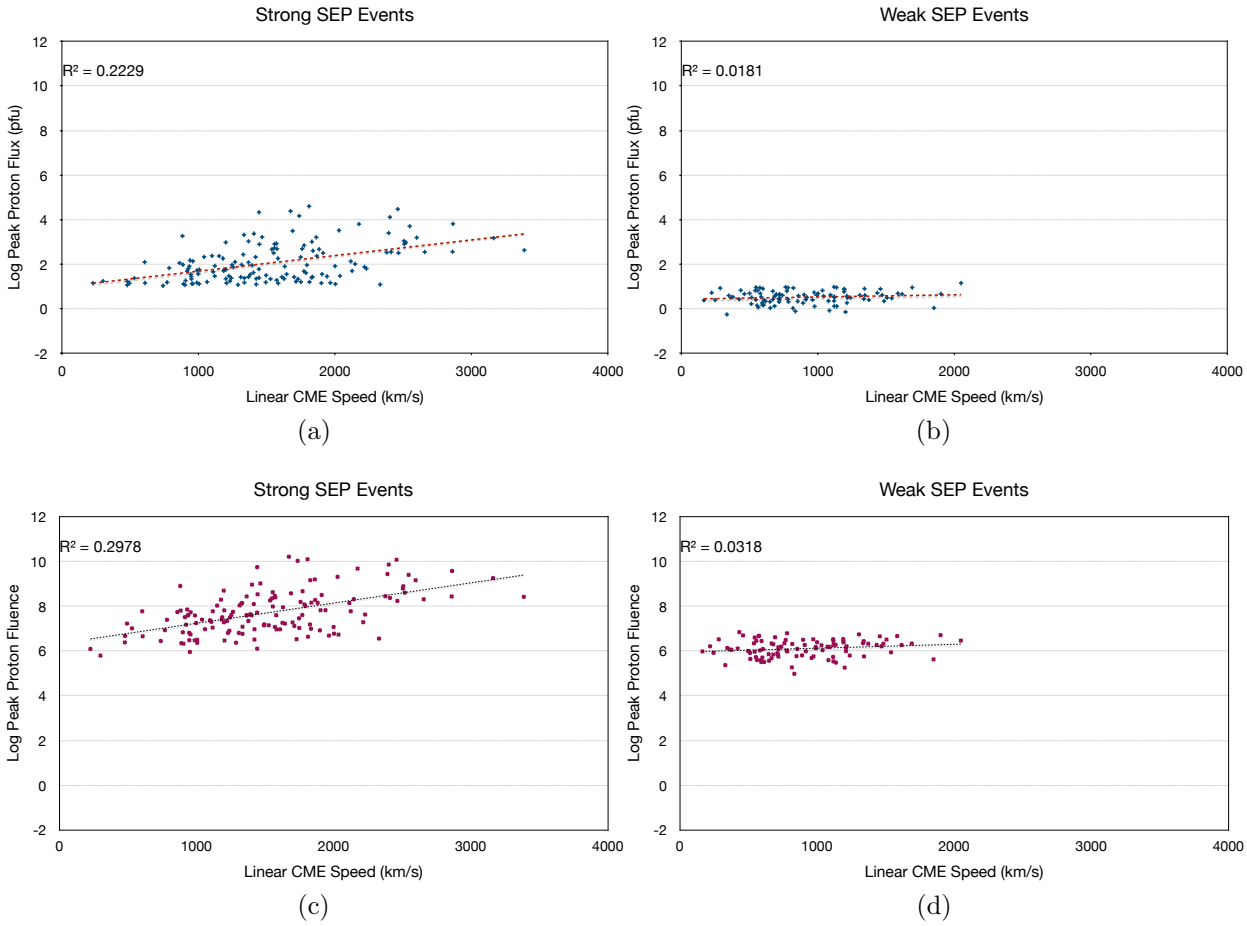
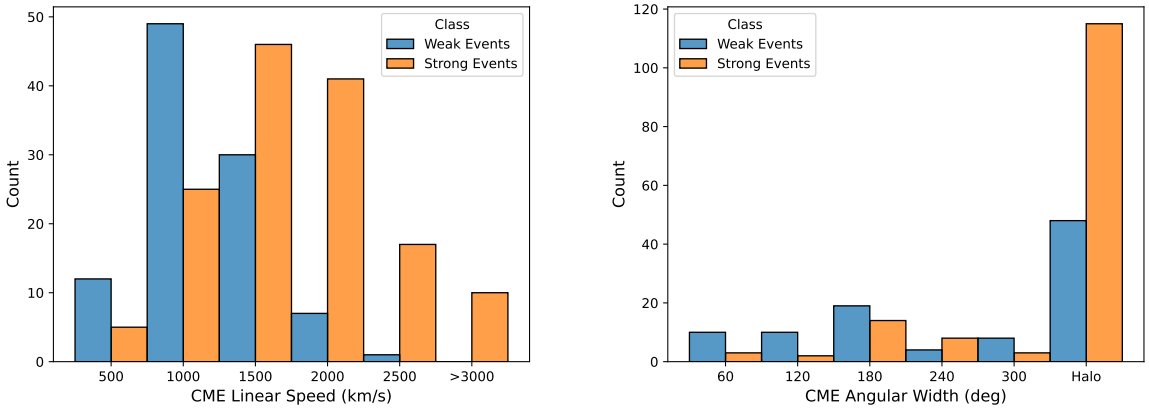


Figure 3.7 Scatter plot of SEP-associated CME parameters.
Note: The log-transformed peak proton fluxes for (a) strong and (b) weak SEP events. Similarly, we show a scatter plot of proton event fluences with respect to CME speeds for (c) strong and (d) weak SEP events. A linear regression trend is fit to the data, showing the respective variance (R^2) of the distribution as an inset within the corresponding figure.

3.4.3 Type II Radio Bursts

We consult and integrate the CDAW's type II radio bursts catalog into our data set. The CDAW list constitutes radio data measurements from the WAVES experiment (Bougeret et al. 1995) on board the Wind spacecraft (Acuña et al. 1995). Type II radio bursts have been largely associated with fast CMEs and are widely considered to be possible signatures of



(a) (b)

Figure 3.8 Distribution of 247 SEP-associated CME parameters in the GSEP data set. Note: Shown are (a) linear speed and (b) angular width of CMEs that are associated with SEP events. The two classes of SEP events (strong and weak) are shown in the legend. It can be noticed that halo fast-CMEs appear to dominate the occurrence of strong SEP events.

SEP events. It has also been observed that each type II burst follows a CME eruption in the upper coronal region (Gopalswamy et al. 2019). For a total of 270 SEP events in our data set (for SCs 23 and 24 only), 147 have been associated with a type II radio burst. More than half of these bursts occur within 3 hr before the SEP event onset with the dominant feature of that day's spectral plot. In our data set, one radio signature is uncertain, and another is reported by CDAW to be caused by a CME–CME interaction. Three radio bursts occur prior to SEP events originating behind the eastern limb and 14 behind the western limb. In these scenarios, radio emissions may be occulted. In the CDAW-SEP list, seven more SEP events are associated with type II radio bursts in solar cycle 23 but are not reported in CDAW's Wind/ WAVES type II burst catalog. Hence, we have omitted them from our data set. We note that we have not examined individual frequency radio profiles, as that is

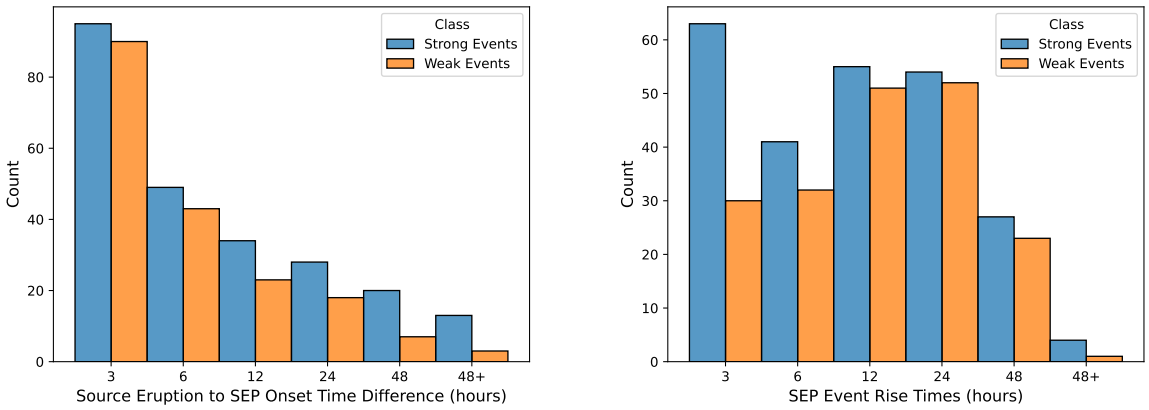
a complex and time-consuming task beyond this work's scope.

3.5 Results

3.5.1 SEP Event Temporal Properties.

In addition to substantial variability in their elemental abundances, SEP events also exhibit variation in their time profiles. Hence, temporal variations between source solar eruption and the SEP onset become crucial in understanding the propagation of particles toward Earth. Figure 3.9(a) shows a histogram of the time difference between a source (flare or CME) eruption and SEP event onset. As can be seen, 79% (335 out of 423) of SEP events have a precursor eruption within 12 hr. Forty-five source eruptions occur between 12 and 24 hr, 28 between 24 and 48 hr, 13 between 48 and 72, and two SEP events have a source solar eruption over 72 hr before the SEP event onset. The median time delay between a halo-CME launch and SEP event onset is about 3.5 hr. At a median (average) of 26 (31) minutes after a CME is launched, a type 2 radio burst is initiated. The median (average) non-halo CME- RB time difference is 35 minutes (43 minutes).

Figure 3.9(b) shows histograms of rise times (i.e., the time between the event onset and peak) in our data set, classified in bins of 0–3, 3–6, 6–12, 12–24, 24–48, and >48 hr. The overall median (average) rise time SEP events is 7.5 hr (>11.5 hr). For halo-CME-associated events, the median (average) rise time is ∼9 hr (13 hr). For non-halo associations, the median (average) SEP event rise time reduces to 7 hr (∼9.5 hr).



(a) (b)

Figure 3.9 Distribution of temporal parameters in the GSEP data set.
Note: Shown are (a) time difference between source solar eruption (flare/CME) and SEP event onset. Up to 79% of SEP events arise from a precursor within 12 hr; (b) SEP event rise time, i.e., the time taken by the proton fluxes to reach peak intensities.

3.5.2 Source and SEP Events Correlation

We study the spatiotemporal dependence of SEP events on source eruptions by calculating the Pearson correlation coefficient of features in our data set. As discussed earlier, the peak of an SEP event, its fluence, and the rise time can be studied in relation to different source eruption features such as flare magnitude, flare rise time, corresponding active region longitude, CME speed, and width. To examine the strengths between each of these variables, we construct a heatmap of the correlation coefficients as shown in Figure 3.10.

One should take note that the logarithm of peak proton fluxes and proton event fluences have a good correlation with the logarithm of CME speed. There is a moderate correlation with flare level (X-ray magnitude) and a low correlation with the logarithm of CME width. There is also a good correlation between the logs of CME speed and width. In addition,

the flare magnitude and log of CME speed show a low correlation coefficient. But features such as flare rise time and flaring location (longitudes) are poorly correlated with the rest of the variables. A possible scenario for the correlations of CME parameters not to be strong could be due to interacting CMEs. Interactions occur when a fast CME erupts immediately right behind a relatively slower CME or occurs too close in time to preceding events. Such back-to-back interacting CMEs allow an additional flux of particles to be accelerated in the interplanetary medium. However, analyzing these cases one by one is highly complex and is out of the scope of the current work.

3.5.3 Supervised Machine Learning Implementation

In ML, one of the preprocessing techniques to improve the model performance is the reduction of dimensionality of the data set by identifying and choosing the most essential features (Khalid et al. 2014; Ghareb et al. 2016). Selected highly relevant features from a data set can then be used as a core subset in training models (Wei et al. 2020). This step has multiple purposes: improving the performance of the models by eliminating unimportant features; providing robust predictors, thereby reducing computational costs; and offering better interpretability to the underlying physical process that generated the data model (Guyon & Elisseeff 2003).

In this section, we present our efforts in examining feature importance in our data set. All our computational experiments are performed using the Python programming language. In the previous section, we examined the correlations of solar source parameters with SEP events and obtained some insights on possible features that can be relevant to SEP predictions.

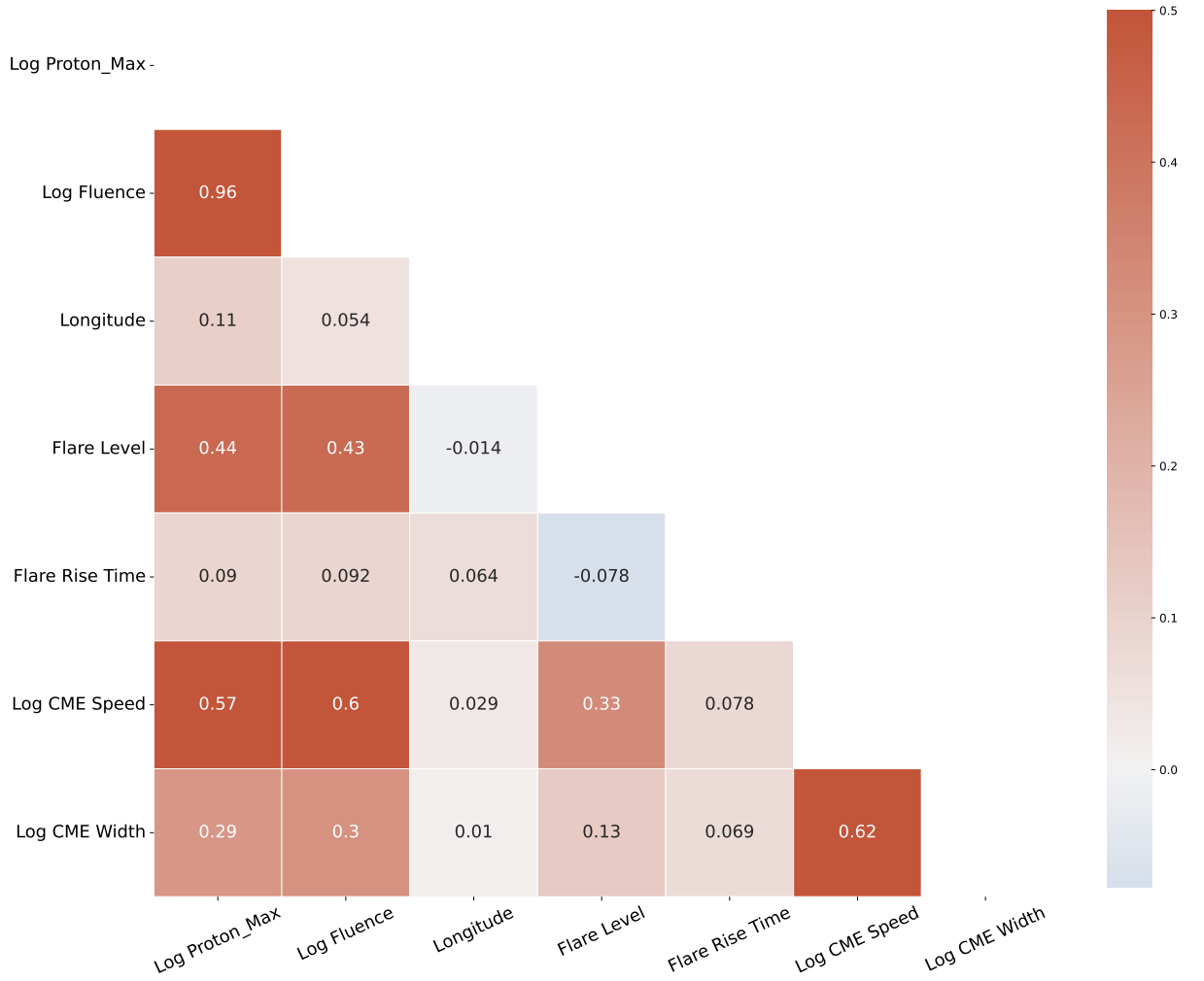


Figure 3.10 Heat map of the correlations coefficients.
Note: Shown are five parameters of source solar eruptions with respect to peak proton fluxes and fluences of SEP events. Logarithmic transformations of the CME and SEP variables are considered here. CME Speed shows a good correlation with SEP events, while flare intensities are moderately correlated.

However, another way to confirm the existing relationships of those relevant features can be implemented using tree-based ML models by extracting information "gain" on each feature (Janitza et al. 2016; Prasetiyowati et al. 2020). For this purpose, we utilize random forest (RF; Breiman 2001), and extreme gradient boosting (XGBoost; Chen & Guestrin 2016)

classifiers. These two models are popular tree-based learners that are extensively used in many areas of active research (Pal 2005; Fawagreh et al. 2014; Sarica et al. 2017; Tyralis et al. 2019; Can et al. 2021; Korsós et al. 2021; Lavasa et al. 2021; Osman et al. 2021).

RF is a collection of decision trees on subsets of the data where the average of the prediction from each tree is obtained, and based on a majority vote, the final output is predicted. RF follows the bagging technique (parallel building) to improve its performance. XGBoost builds on the gradient boosting model, that is, implementing the gradient decent algorithm in parallel and improving its performance in each step.

3.5.3.1 Model Generation

We consider strong SEP events as a "positive" class and weak events as a "negative," thereby designing the problem as a binary classification task in this work. The six features considered are magnitude, rise times, and location coordinates of flares, log-transformed speeds, and widths of CMEs, and the time difference between source eruption and SEP event onset. Because we want to consider both the SF and CME parameters to classify the SEP events, we drop all rows with any null values in our desired feature columns. Doing that, the size of the data set has been undersampled to 146 instances, preserving a lower number of (37) weak SEP events. We standardize our features by implementing the **StandardScaler** module from scikit-learn. We hold 40% of the data as the test set to evaluate the performance of the models.

Traditionally, a third validation set is partitioned from the data to tune the hyperparameters and evaluate the model. However, this results in further reduction of sample size,

to help the model learn the pattern in the data. Hence, we implement a procedure called k-fold cross validation (CV) where the training data is split into k number of smaller sets or folds (Mosteller & Tukey 1968; Stone 1977). The model is first trained on k-1 folds, and the resulting model is validated on the held-out fold. The overall performance of k-fold CV is evaluated as the average of the values computed in each step. In our work, we choose $k = 2$.

We determine the best parameters of the models based on grid search hyperparameter optimization techniques. For our tree-based models, we experimented with the maximum depth of the tree and the number of estimators. The best-case scenario for both the models resulted in 100 estimators and a maximum depth of four. We handled the class imbalance ratio between positive and negative instances by assigning appropriate parameters to automatically adjust and balance the weights based on class frequencies.

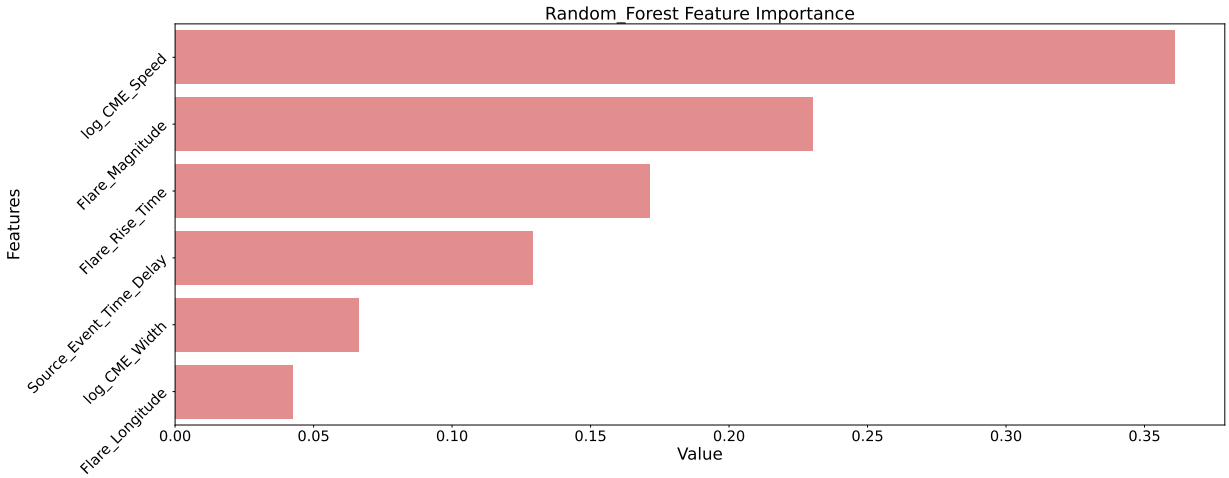
3.5.3.2 Feature Importance

Statistical findings from earlier works (such as Cane et al. 2010; Trottet et al. 2015; Papaioannou et al. 2016; Anastasiadis et al. 2019) emphasize the importance of SF and CME features leading to SEP events. In addition, recent work by Lavasa et al. (2021) also implemented model-based feature importance and finds the CME speed, width, and soft X-ray fluence as prominent features relevant to SEP event prediction.

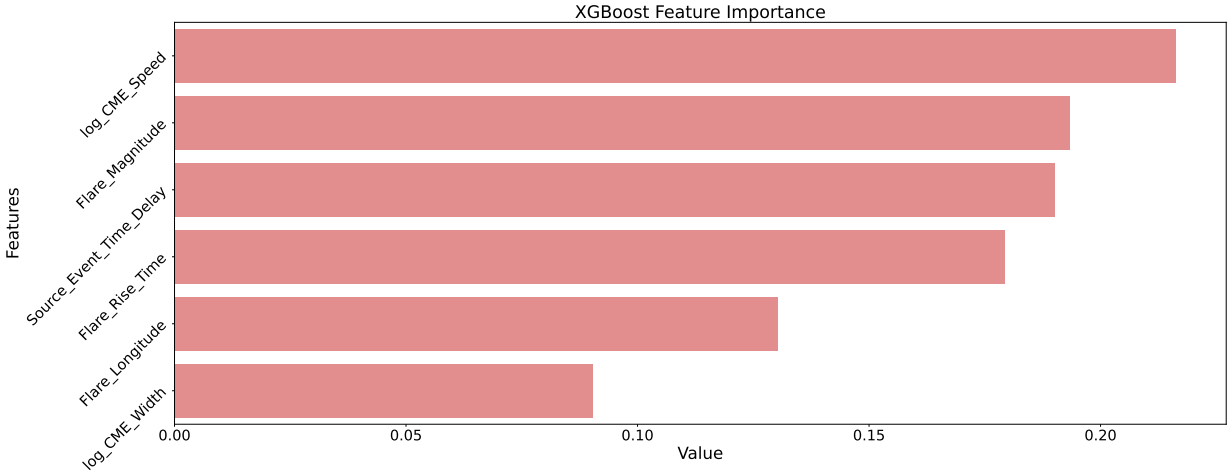
Both the tree-based models used in our study have a built-in feature_importance attribute provided by the scikit-learn library[8] (Pedregosa et al. 2011) that can be used to understand the splitting criterion within our models. Each feature is assigned a score that indicates how

[8]Scikit-learn: https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html

informative it is toward predicting our target variable. Therefore, higher-scoring features are more important. Using this technique, we extracted feature importance from our data set, and we show the rank of our models in Figure 3.11.



(a)



(b)

Figure 3.11 Ranked results of feature importance in our data set.
Note: Shown are (a) random forest and (b) XGBoost models. Both models build on both flare and CME parameters where the logarithm of CME speed and the flare magnitude are the top two highly significant features.

Here, we can see that there is a close match in the results between models. Both RF and XGBoost have used "log_CME_speed" and "flare magnitude" as the top two important features. Nonetheless, the slight difference in the ranking of the rest of the features arises from the fact that RF randomly selects the features for each tree while sampling the data set. After that, the importance of a feature is estimated as a total reduction in the criterion due to that feature (Pedregosa et al. 2011). On the other hand, XGBoost uses features that have higher correlation between them to split the trees. The chosen top feature then remains constant throughout the model. Nonetheless, we conclude from our feature importance results that both the SF and CME parameters (CME speed and flare intensity, respectively) are indispensable in understanding and predicting the SEP event occurrence.

3.5.3.3 SEP Event Classification

In Table 3.2, we list all the available studies on SEP event predictions based on ML models that have been developed in the recent past. Each study here implements a variety of popular models such as K-nearest neighbors (KNN), logistic regression (LR), decision trees (DT), extremely randomized trees (XT), support vector machines (SVM), and neural networks (NN) to obtain promising results. The models in these studies have been developed focusing on a combination of variety of solar parameters such as AR features to fluxes of radio, X-rays, and protons exhibiting high correlations between each other. In addition, the time period considered in these studies depend on the availability of their desired data.

In our work, we approach the SEP event prediction problem from a different perspective based on the GSEP data set. As there is no one-to-one correspondence between the data and

the methods implemented, we do not compare our results with earlier studies. Nonetheless, we present below the results of our best models to classify the SEP events as "strong" and "weak" in our data set. In addition to RF and XGBoost, we explore the performances of LR (Cox 1958) and SVM (Cortes & Vapnik 1995) classifiers in our work. We emphasize that the best-performing hyperparameters were chosen for our models to obtain robust results.

Table 3.2: List of Recently Developed ML-based Models to Predict SEP Events.

Source	**Period**	**Solar Features**				
		AR	**SF**	**CME**	**RB**	**Protons**
Boubrahimi et al. (2017)	1997 - 2013	N	Y	N	N	Y
Engell et al. (2017)	1986 - 2018	N	Y	N	N	Y
Lavasa et al. (2021)	1988 - 2013	N	Y	Y	N	Y
Aminalragia-Giamini et al. (2021)	1988 - 2013	N	Y	N	N	Y
Stumpo et al. (2021)	1995 - 2017	N	Y	N	Y	Y
Sadykov et al. (2021)	2010 - 2019	Y	Y	N	Y	Y
Kasapis et al. (2022)	1996 - 2010	Y	Y	N	N	N

Note: Binary values in columns (3)-(7) indicate whether the corresponding solar event feature was considered in the corresponding study. That is, "Y" means yes, and "N" otherwise.

3.5.3.4 Evaluation

A 2 × 2 contingency table is implemented in our work for a binary classification task (strong/weak SEP events). This table constitutes the following scores: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Here, TP indicates the number of correctly predicted strong SEP events (positive class) by a model while TN rep-

resents the number of rightly predicted weak SEP events (negative class). FP corresponds to the number of weak events predicted as strong (false alarms) while FN corresponds to the number of strong events predicted as weak (misses). Subsequently, the aim of our best model should be to reduce incorrect results represented by both FP and FN.

We use F1 score, Matthews correlation coefficient (MCC; Matthews 1975), true skill statistics (TSS; Woodcock 1976; Daan 1985) and Heidke skill score (HSS; Heidke 1926) as our evaluation metrics to estimate the performance of the models on training and test sets. Focusing on the importance of positive classes, we consider the F_1-score that can be estimated as the harmonic mean of "precision" (TP/(TP + FP)) and "recall" (TP/(TP + FN)) as shown in Equation 3.1.

$$F_1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{3.1}$$

Because we have an imbalanced data set, we prefer the following metrics that take into account the variation in the sizes of the two classes of target variable. Using the confusion matrix, TSS is defined to account for the false-positive rate as shown in Equation 3.2.

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN} \tag{3.2}$$

However, accounting the true negatives to assess the performance for a binary class problem is important in our context. Hence, we also choose MCC and HSS as defined in Equations 3.3 and 3.4, respectively.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TP + FP) \times (TN + FN)}} \tag{3.3}$$

$$HSS = \frac{2 \times ((TP \times TN) - (FP \times FN))}{(TP + FN) \times (TN + FN) + (FP + TN) \times (FP + TP)} \tag{3.4}$$

The results of our best-performing models on the test set are presented in Table 3.3. The models are ranked based on the F_1-score where XGBoost performs higher than the rest of the models. On the other hand, LR and SVM models perform better in terms of TSS, MCC, and HSS compared to the trees-based ones. These results give us hope for improvement based on further feature extraction and experimentation with a variety of ML models.

Table 3.3: Performance of Supervised Classifiers on the GSEP Data Set.

Model	$\mathbf{F_1}$	**TSS**	**MCC**	**HSS**
XGBoost	0.85	0.41	0.43	0.43
RF	0.80	0.40	0.38	0.38
LR	0.77	0.58	0.52	0.46
SVM	0.71	0.58	0.52	0.46

Note: Class metrics are presented here for the best models implemented on the test set.
Models: XGBoost - extreme gradient boosting; RF - random forest; LR - logistic regression; SVM - support vector machine.
Metrics: TSS - true skill statistics; MCC - Matthews correlation coefficient; HSS - Heidke skill score.

3.6 Conclusions

Signatures of solar activity constitute many transient events, including solar flares, coronal mass ejections, and solar energetic particles. These are the main drivers of SWx. Toward predictive efforts of SEP events, we have built and analyzed the GSEP data set first presented

in Rotti et al. (2022) and further discussed in this paper. Our extended data set consists of 433 events, out of which 244 cross the 10 pfu threshold in the $E \geq 10$MeV proton channel. We rely on several sources such as existing event catalogs of solar flares, CMEs, radio bursts, and SEP events, and relevant observational data from SOHO and SDO missions to associate each SEP event in our data set with a flare and/or CME wherever possible. We conduct statistical analysis of SEP events along with the parent solar eruptions: namely, estimating the relationships of flare magnitude, its rise time, the flaring locations, speed, and width of CMEs, and the temporal variations of sources with SEP events.

With respect to SEP peak fluxes and event fluences, we find a good correlation for CME parameters, and moderate correlations are observed for X-ray flare intensities. Most of the strong SEP events in our data set are associated with intense flares ($\geq$M2.0). Also, we find most (66%) of the CMEs associated with SEP events are halo in angular extents and fast with a median speed of 1200 km.s^{-1}. In summary, as shown in previous works such as Cane et al. 2010 and Papaioannou et al. 2016, the SEP event intensities increase with increasing CME speed and flare strength. The distribution of event origins shows an increase in proportionality with source eruptions occurring on the visible disk up to western hemisphere. Exceptions are noticed for very large flare sources ($\geq$X1.0) occurring over the eastern and beyond-limb locations due to poor magnetic connectivity for the particles to reach Earth.

We focus on extracting important features in our data set by implementing tree-based machine-learning models, namely, random forest and XGBoost classifiers. We find both the

CME speed and flare intensity/magnitude to be the top contributing features toward the corresponding SEP event peak fluxes. In a recent study, Lavasa et al. (2021) undertake a feature extraction process and find the CME speed and width and the flare soft X-ray fluence as the most important features for identification of SEP events.

In this work, we also implement logistic regression and support vector machines in the framework of a binary classification problem. We consider strong SEP events as the positive class and weak ones as the negative. We perform hyperparameter tuning and implement a twofold cross-valida- tion technique to optimize our models for robustness. We use the F_1-score, MCC, TSS, and HSS to assess the model performance. The results in this paper show that all our models perform moderately well. For tree-based models, we see XGBoost perform slightly better than RF, while LR and SVM are close in their outcomes. There are many areas in which SEP event prediction efforts can be improved, and we emphasize undertaking further study on feature engineering in future work. The GSEP data set is available at Harvard Dataverse repository at doi: 10.7910/DVN/DZYLHK. The statistical plots and analysis presented in this study are based on version 5.0 of the data set.

CHAPTER 4

Short-term Classification of Strong Solar Energetic Particle Events Using Multivariate Time Series Classifiers

The contents of this chapter have been published in the Astrophysical Journal, Volume 966, Issue 165.

4.1 Introduction

Solar energetic particle (SEP) events are manifestations of solar activity that constitute the emission of energetic electrons, protons and heavier ions from the Sun. These events are usually associated with parent solar eruptions, namely solar flares (SFs) and shock fronts of coronal mass ejections (CMEs; Cane et al. 1986; Kahler 1992; Reames 1999; Gopalswamy et al. 2001). Generally, it is understood that the eruptions at the western side of the Sun have a higher probability of SEPs reaching near-Earth space due to the spiral structure of the interplanetary magnetic field lines, known as the Parker spiral (Parker 1965; Reames 1999).

Measurements of SEP events near Earth depend on the spatial region of source eruptions on the Sun. In the case of extreme SEP events, given the right conditions such as geomagnetic connectivity and enough seed population, they are often associated with fast CMEs (Marqué et al. 2006; Gopalswamy et al. 2008; Swalwell et al. 2017; Gopalswamy et al. 2017; Cliver & D'Huys 2018; Rotti & Martens 2023). The impacts of SEP events include severe technological (Smart & Shea 1992) and biological effects on various economic scales (Schrijver & Siscoe 2010). Although the Earth's magnetic field provides us a protective shield from the energetic particles and filters them out from reaching the ground, they can be fatal for space-based

missions and aircraft travel along polar routes (Beck et al. 2005; Schwadron et al. 2010). For instance, long-lasting strong SEP events pose a radiation hazard to astronauts and electronic equipment in space (Jiggens et al. 2019).

According to the Space Weather Prediction Center (SWPC) proton intensities $\geq$10 pfu (1 pfu = 1 particle per cm^2.s.sr) in the E>10 Mega electron-Volt (MeV) energy channel are termed as strong SEP events with regards to causing significant space weather (SWx) effects (Bain et al. 2021). In addition, the severity of the solar proton events is measured by SWPC using the Solar Radiation Storm Scale (S-scale)[1] which relates to biological impacts and effects on technological systems. The S-scale relies on the E$\geq$10 MeV integral peak proton flux from near-Earth observations of the Geostationary Operational Environmental Satellite (GOES) missions (Sauer 1989; Bornmann et al. 1996). The base threshold, associated with an S1 storm, corresponds to a GOES five minutes averaged $\geq$10 MeV integral proton flux exceeding 10 pfu for at least three consecutive readings. Further scales from 'S2' to 'S5' logarithmically increase from one another, therefore defining different event intensities.

With great advancements in space engineering and technology, we are fortunate to have near-continuous observations of solar activity from a fleet of space-based satellites over the last four decades. One important aspect of analyzing solar data is to advance operational capabilities by mitigating SWx effects on our human explorers and technological systems (Jackman & McPeters 1987). This urgently requires the development of robust tools to forecast eruptive event occurrences. With an SEP event prediction system we can forecast and send out warning signals before the event.

[1]NOAA S-scale: https://www.swpc.noaa.gov/noaa-scales-explanation

Several researchers have been focusing on implementing a variety of model-driven techniques for predicting SEP events. In this regard most scientific studies concentrate on predicting the peak fluxes. To predict event occurrences, many physics-based and data-driven statistical models have been designed based on the parameters of parent eruptions such as SFs and CMEs (Van Hollebeke et al. 1975; Posner 2007; Kahler et al. 2007; Balch 2008; Laurenza et al. 2009; Núñez 2011; Falconer et al. 2011; Dierckxsens et al. 2015; Winter & Ledbetter 2015; Núñez 2015; Anastasiadis et al. 2017; Alberti et al. 2017; Papaioannou et al. 2018; Ji et al. 2021). In the last decade machine learning (ML) methods have also been at the forefront of SEP event forecasting (Swalwell et al. 2017; Engell et al. 2017; Aminalragia-Giamini et al. 2021; Lavasa et al. 2021). ML-based algorithms have been rigorously explored by many teams across the globe due to their success in many other areas of research and operations (Camporeale 2019). Detailed descriptions of existing SEP event forecasting models can be found in Whitman et al. (2022).

We envision building low-risk, short-term predictive models as the first step towards building operationally driven, reliable SEP event forecasting systems. Therefore, we exploit the feasibility of multivariate time series (MVTS) data in this work. For this purpose, we utilize and compare the performances of three ML models. Two are interval-based algorithms: time series forest (TSF) and supervised time series forest (STSF)—lastly, a dictionary-based bag of SFA symbols (BOSS) model. Prior studies on SEP event forecasting using parent eruption features conclude that the tree-based model is viable (Boubrahimi et al. 2017). Both TSF and STSF implement a highly specialized random forest (RF) model and rely on

several interpretable statistical features extracted from the time series to feed into an ensemble of decision trees. We will discuss more on individual model architectures in the later part. The rest of the paper is organized as follows: Section 4.2 provides information about our data set and data preparation steps used in this work. Section 4.3 presents our research methodology including descriptions of the time series classifiers. Section 4.4 discusses the training phase of the models and presents the experimental evaluation framework. Lastly, Section 4.5 summarizes our work and future avenues.

4.2 Data

The SEP events are critical phenomena caused by SFs and CMEs. The parent eruptions are triggered by sudden, abrupt changes in the magnetic field, typically of active regions in the solar atmosphere. Thus, it is well expected to build predictive capabilities employing parameters of precursor events. Nonetheless, we do not consider any data related to CMEs, and restrict ourselves to use the one-minute averaged GOES X-ray (1–8Å) fluxes measured by the X-ray sensor (XRS) onboard GOES. The archived data is available online from the National Oceanic and Atmospheric Administration (NOAA)'s website[2]. In addition, we use the following integrated proton channels from GOES: (1) E≥10 MeV fluxes corresponding to P3, (2) E≥50 MeV fluxes corresponding to P5, and (3) E≥100 MeV fluxes corresponding to P7. Because SFs have characteristic durations from a few minutes to a few tens of minutes, we linearly interpolate the proton five-minute averaged fluxes to match with the one-minute cadence of the X-ray fluxes. We believe this interpolation is necessary to retain

[2]GOES-XRS: https://www.ncei.noaa.gov/data/goes-space-environment-monitor/access/avg/

the information on flaring peaks without altering the flare characteristics from X-ray fluxes.

4.2.1 GSEP Data Set

The Geostationary Solar Energetic Particle (GSEP) events data set (Rotti et al. 2022) is a recently introduced open-source[3] multivariate time series (MVTS) benchmark data set of SEP events covering solar cycles 22 to 24. The description of the data set and its development can be found in Rotti et al. (2022) and Rotti & Martens (2023). It was created using proton fluxes measured by the Space Environment Monitor (SEM) suite onboard GOES (Grubb 1975). This data set comprises a catalog of 433 (- 244 strong and - 189 weak) SEP events observed near Earth between 1986-2018. Each event is labeled a '1' or '0', indicating either a strong or weak SEP event, respectively, based on the event definition of NOAA-SWPC. Here, a strong SEP event corresponds to proton fluxes crossing 10 pfu in the GOES 'P3' channel. Whereas a weak SEP event has proton enhancements between $\geq$0.5 and $<$10 pfu. Furthermore, the data set consists of time series slices of GOES proton and X-ray fluxes of all the events. Each time series slice constitutes 12 hr fluxes prior to the onset of the event as an observation window and further, until the events cross the peak flux, finally falling to half that value.

As reported by Rotti & Martens (2023), $\approx$ 79% of SEP events have a precursor eruption within 12 hr prior. In other words, most SEP events' onset times are within 12 hr after the initiation of the parent flare eruption. Interestingly, most (53) events with a parent eruption more than 12 hr prior to SEP onset occur during solar maximum ($\pm$ one year). Many of

[3]The GSEP data set available on Harvard Dataverse: https://doi.org/10.7910/DVN/DZYLHK

these precursor eruptions occur more than a day before the onset of an SEP event. We consider 12 hr as an optimal span or observation window in the present work. However, limiting the observation window to 12 hr does not cause a huge limitation on our models. That is, the inclusion of X-ray fluxes is valuable but not trivial to short-term predictions of SEP events. Hence, we have considered 12-hr as an optimal window by including as much precursor (X-ray) data as possible. Increasing the window length to greater than 12-hr has the potential to induce noise such as additional and unrelated X-ray flux peaks in data. In addition, we omit five minutes of input data just before the SEP event onset. As we consider fluxes with one-minute cadences, our data set represents a 715-length soft X-ray and integral proton time series. A sample time profile for a strong SEP event in the GSEP data set is shown in Figure 4.1 that occurred on 2017-09-05T00:40 (UT) with a rise time of $\approx$ nineteen hours. The parent flare erupted about four hours before the SEP event onset from active region 12673 (solar lon = 12°, solar lat = -10°) and had a magnitude of M5.5 as measured by the GOES/XRS instrument. Following the flare there was a halo fast-CME propagating with a velocity of $\approx$1400 km.s^{-1}. The SEP event reached a peak flux of $\approx$210 pfu on 2017-09-05T19:30 (UT) in the E$\geq$10 MeV channel measured by the GOES-SEM instrument. The vertical dotted line overlayed in the plot indicates the event's start time while the horizontal dashed line indicates the SWPC S1 threshold. The shaded region shows the typical length of the time profile we utilize in our work.

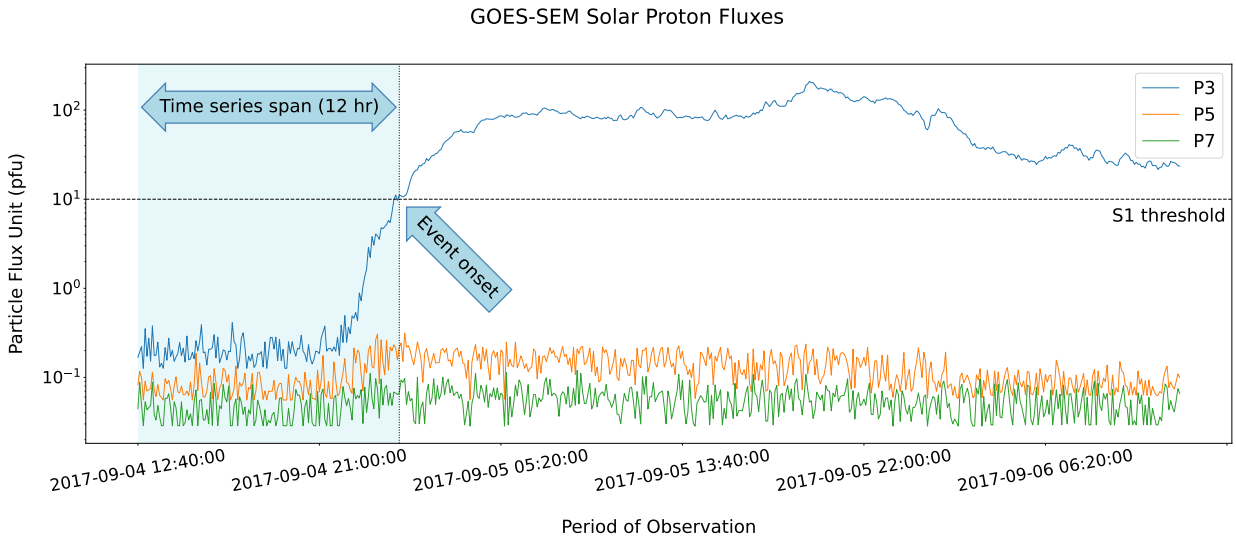


Figure 4.1 Time series plot of a strong SEP event that occurred on 2017-09-05T00:40 (UT). Note: Proton fluxes shown here on a log scale that reached a peak value of ≈210 pfu on 2017-09-05T19:30 (UT). The three fluxes in the legend correspond to GOES P3 (E≥10 MeV), P5 (E≥50 MeV) and P7 (E≥100 MeV) integral proton channels. The horizontal black dashed line indicates the SWPC threshold for a strong SEP event while the vertical black dotted line indicates the SEP event onset time. The shaded region shows the typical span of time series considered in our work. It corresponds to 12 hours of proton fluxes prior to the SEP event onset.

4.2.2 Data Labels

The work discussed here considers the term 'SEP events' analogous to solar protons events (SPEs). While variations exist, event labels are usually associated with the occurrence of strong SEPs based on the integral proton fluxes (I_P) recorded by P3 crossing the 10 pfu threshold. As mentioned earlier, the weak or sub-events are defined based on a threshold of $0.5 > I_P < 10$ pfu in the 10 MeV channel. If there are successive SEP events within 12 hours, then the observation window shall constitute fluxes prior to the former event onset. There are several events reported in the GSEP data set that have overlapping proton fluxes from the previous event. Due to the nature and characteristics of the SEP event, such overlapping

cannot be excluded. In these scenarios, when the proton fluxes in the 10 MeV channel are already above 10 pfu, the model outputs a "yes" label indicating a strong event. This "back-to-back events" situation is evident during solar maximum. In the GSEP data set, 23 (4) strong (weak) SEP events occur within the next 24 hours following the first event. There are only six successive events occurring within 12 hours, all of which are strong in nature, with a median rise time of $\approx$14 hours and a median event length of $>$48 hours.

Another critical threshold in terms of operational requirements concerning astronauts during extra-vehicular activities is one pfu in the E$\geq$100 MeV (P7) channel. Nonetheless, in the present work, we focus only on the SWPC 'S1' threshold and defer the former scenario to future work.

In the context of solar particle radiation a passing interplanetary shock causes energetic storm particle (ESP) acceleration (Cane 1995). Although ESPs are different kinds of particle events, they can still be brought under the "umbrella" of SEPs since the energetic particle fluences still determine the radiation exposure and dosage rate. Furthermore, it is relevant to minimize the total dosage rate of an astronaut during a space mission for their health and safety. Therefore, our focus has been a cumulative "solar particle event" prediction wherein we also include the nine ESPs reported in the GSEP catalog in our analysis.

4.3 Methodology

In this work, we attempt to address the grand problem of SEP event predictions from a time series classification perspective. This problem is constructed here in the framework

of a binary classification task. Here, the target labels are based on surpassing the proton flux threshold defined by NOAA-SWPC. Accordingly, the SEP event class labels that have proton enhancements above the threshold ($I_P \geq 10$ pfu) are "positive", else "negative". In this section, we describe a novel framework for classifying E≥10 MeV SEP events using time series-based ML models.

We use a column ensemble of univariate classifiers, a parameter-wise ensemble of columns in which individual classifiers are applied to every parameter (column). This is a homogeneous ensemble schema; an overview of it is shown in Figure 4.2. The ensemble estimator allows multiple feature columns of the input to be transformed separately. The statistical features generated by each classifier on samples of the original time series are ensembled to create a single output. Each feature is assigned a score that indicates how informative it is towards predicting the target variable (Hansen & Salamon 1990; Schapire 1990; Arbib 2003).

In our case of the GSEP data, we create a multivariate variant of univariate algorithms using the column ensemble method described above. We consider the long band X-ray (XL) and three proton channels (P3, P5, P7) as our input time series. We implement and compare the performances of three classifiers for strong/weak SEP event classifications. The prediction results from these individual column classifiers are then aggregated as a whole (with equal votes using prediction probabilities). The idea is to see if the observed time series span leads to a strong SEP event (positive class) or not (negative class). The negative classes here do not constitute SEP-quiet periods but are entirely weak SEP events. These sometimes behave almost as strong events but fall below the critical threshold. Identifying

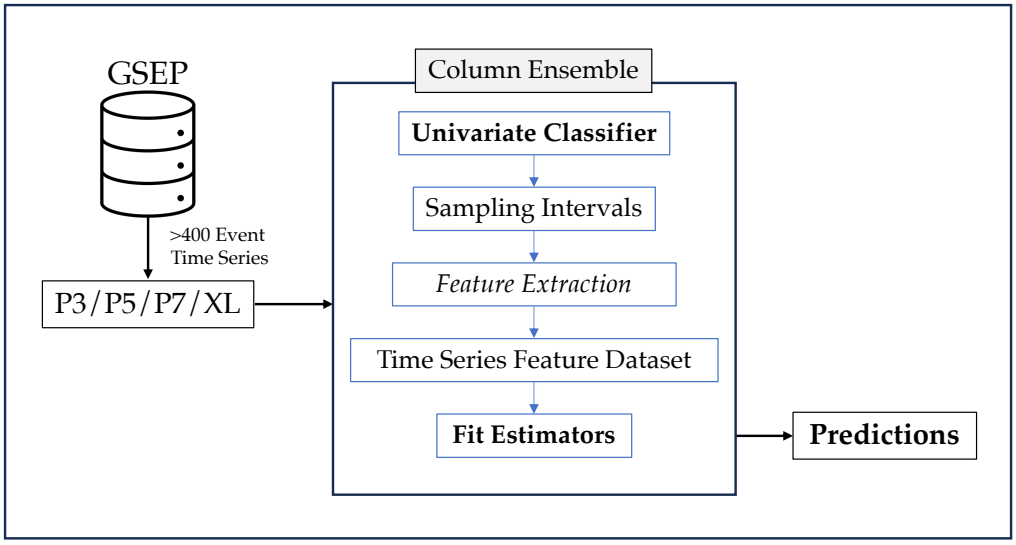


Figure 4.2 Schematic overview of the workflow.
Note: We consider three proton and the long band X-ray channels from the GSEP time series data set. In a column ensemble, we input the fluxes to our classifiers. Each univariate classifier subsamples the time series and extracts features from each interval to generate a feature data set. The classifier is trained to fit the input data and further tested on unseen data.

such patterns is relevant to reducing false alarms. In other words, the reason for choosing these two classes is that the models must pick up the incoming flux behavior of X-rays and earthward accelerating protons that may cross the SWPC event threshold, which requires mitigation measures in an operational context. The rationale is to explore the operationally relevant proton channels including those with the XL channel.

Regarding existing SWx forecasting methods, flare forecasters build models distinguishing between $\geq$M1.0 and $\leq$C9.9 classes (Ji et al. 2020). Similarly, we aim to provide an interpretable state-of-the-art time series ML model to classify strong and weak SEP events. Therefore, this method will provide a perspective to extend the univariate time series classifiers in an ensemble and build a prototype short-term SEP event prediction system that

optimizes the model based on forecast skill scores. Section 4.3.1 provides more details about these classifiers and their feature sets.

4.3.1 Time Series Classification

SWx practitioners and forecasters highly recommend using temporal features and time series analysis for better forecasting (Singer et al. 2001). In time series data, every timestamp is typically a vector or array of real values observed over time. It can be divided into univariate or multivariate such that an array of only one parameter is a univariate series and a set of univariate series forms a multivariate series (Ruiz et al. 2021). In time series-based ML, one of the techniques to improve model performance is the reduction of the dimensionality of the data set by identifying and choosing the most relevant features (Koegh et al. 2001; Cassisi et al. 2012).

Feature-based models extract highly relevant statistical features from the time series that are later used as a core subset in training models (Fulcher & Jones 2014). This step has multiple purposes, such as (1) optimizing the performance of the models by choosing relevant features, (2) providing robust predictors thereby reducing computational costs, and (3) offering better interpretability to the underlying physical processes that generated the data model. Time series classification uses supervised ML to analyze labeled classes of time series data and then to predict the class to which a new data set belongs. This is important in SWx predictions, where particle sensor data is analyzed to support operational decisions in near-real-time (NRT). The accuracy of classification is critical in these situations, and hence we must ensure that the classifiers are as accurate and robust as possible.

There are many algorithms that are designed to perform time series classification. Depending on the data, one type might produce higher classification accuracies than other types. This is why it is important to consider a range of algorithms when considering time series classification problems. In this work we experiment with interval-based and dictionary-based models on our data set.

Interval-based algorithms typically split the time series into multiple random intervals. Each temporal feature calculated over a specific time series interval can capture some essential characteristics. Therefore, the algorithm gathers summary statistics from each sub-series to train individual classifiers on their interval. Next, the most common classes are evaluated among the intervals and return the final class label based on equal voting for the entire time series (Bagnall et al. 2017). On the other hand, dictionary-based models implement the bag of words (Zhang et al. 2010) algorithm. In a broad structure a sliding window of length 'l' runs across a series of length 'n'. Then, all real-valued window lengths are converted into a symbolic string called a "word" through approximation and discretization processes. During this process, the possible representations are stored in a dictionary. At the end of the series length, the occurrence of each "word" from the dictionary in a series is counted and transformed into a histogram. Finally, histograms of the extracted words are used for the classification task of new input data (Faouzi 2022).

Amongst the univariate interval-based approaches, we consider Time Series Forest (TSF; Deng et al. 2013) and Supervised Time Series Forest (STSF; Cabello et al. 2020). From dictionary-based classifiers, we use the Bag of SFA Symbols (BOSS; Schäfer 2015) that

uses the Symbolic Fourier Approximation (SFA; Schäfer & Högqvist 2012) to transform and discretize subseries into words. We explain the model structure below. A brief summary of the model functions and parameters is presented in Table 4.1. All our computational experiments are performed using the Python programming language (Sanner et al. 1999). All the classifiers used in this study are from the sktime library (Löning et al. 2022).

Table 4.1: Summary Properties of the Models.

Model	**Sampling schema**	**Features**
TSF	Random intervals	μ, σ, m
STSF	Supervised intervals	μ, σ, m, median, IQR, min, max
BOSS	Sliding window	Word representations

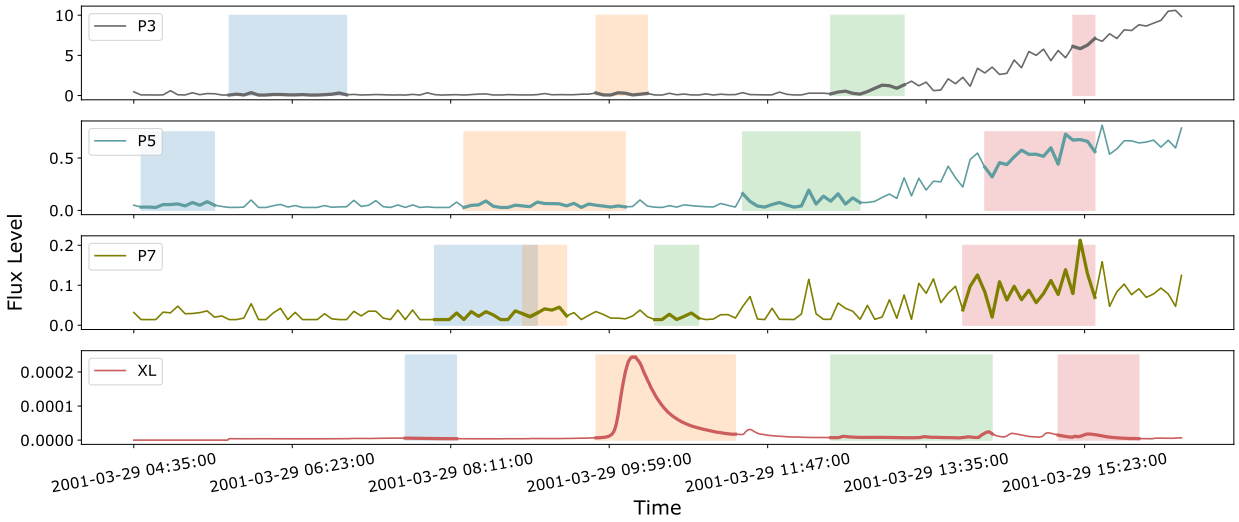
Note: Models: TSF - Time Series Forest; STSF - Supervised Time Series Forest; BOSS - Bag of Symbolic Fourier approximation Symbols.
Features: μ - Mean; σ - Standard deviation; m - Slope; IQR - Interquartile range; min - Minimum value; max - Maximum value.

4.3.1.1 Time Series Forest

One of the most commonly used and popular interval-based algorithms is the time series forest (TSF; Deng et al. 2013). This model implements a random forest approach where multiple decision trees are grouped. Each tree in this ensemble is trained using a subset of statistical features derived from randomly selected intervals, essential in reducing the dimensionality of high-dimensional feature spaces. The statistical features derived from random intervals are mean (μ), standard deviation (σ) and slope of the regression line (m).

Figure 4.3 illustrates the feature extraction process from random intervals in the TSF

algorithm. The process of obtaining statistical summaries of intervals is called flattening the vectors. Each decision tree classifier then assigns a target label to its interval of the data based on a majority vote of all trees. The voting process is needed since every single tree only evaluates a certain subseries of the time series.



(a)

	Interval 1			Interval 2			Interval 3			Interval 4		
Input	μ	σ	***m***	μ	σ	***m***	μ	σ	***m***	μ	σ	***m***
P3	1.274E-01	6.919E-02	2.061E-04	1.890E-01	9.116E-02	6.447E-04	7.030E-01	3.921E-01	2.221E-02	6.265E+00	3.977E-01	7.041E-02
P5	5.221E-02	1.472E-02	7.895E-04	4.907E-02	1.424E-02	-4.734E-05	8.469E-02	3.927E-02	6.045E-04	5.422E-01	1.003E-01	7.041E-02
P7	2.331E-02	6.833E-03	1.400E-04	3.345E-02	7.279E-03	4.635E-04	2.063E-02	5.132E-03	3.377E-04	9.323E-02	3.391E-02	7.041E-02
XL	4.842E-06	5.610E-07	-5.290E-08	8.113E-05	7.539E-05	-1.038E-06	8.938E-06	3.496E-06	4.309E-08	1.023E-05	4.178E-06	7.041E-02

(b)

Figure 4.3 Schematic overview of time series forest (TSF) model.
Note: (a) Random intervals are generated and the corresponding subsets from each time series are extracted. (b) Three statistical features are derived from each subintervals: mean (μ), standard deviation (σ) and slope (m).

4.3.1.2 Supervised Time Series Forest

Another interval-based model is the supervised time series forest (STSF; Cabello et al. 2020). Here, an ensemble of decision trees is built on intervals selected through a supervised process wherein the algorithm finds the discriminatory intervals. The ranking of the interval feature is obtained by a scoring function that indicates how well the feature separates a class of time series from the other classes. The final set of intervals is obtained in a top-down approach to represent the entire series.

STSF aims to improve the classification efficiency by selecting in a supervised fashion (based on their class-discriminatory capabilities) only a subset of the original time series. The algorithm uses three (time, frequency and derivative) representations of the time series as shown in Figure 4.4 and extracts seven features (μ, σ, m, median, interquartile range (IQR), minimum value and maximum value) from each interval. Finally, the feature set is concatenated to form a new data set upon which decision trees are built. The final output is based on majority voting of averaged probability estimates of the ensemble.

4.3.1.3 Bag-of-SFA-Symbols

The bag of symbolic Fourier approximation symbols or BOSS algorithm (Schäfer 2015) typically uses a sliding window to transform the time series into sequences of symbols to extract "words" and form a histogram. The final classification is made by determining the distribution of these "words" in the histogram. The intuition behind this method is that time series are similar, which means they are of the same class if they contain similar "words".

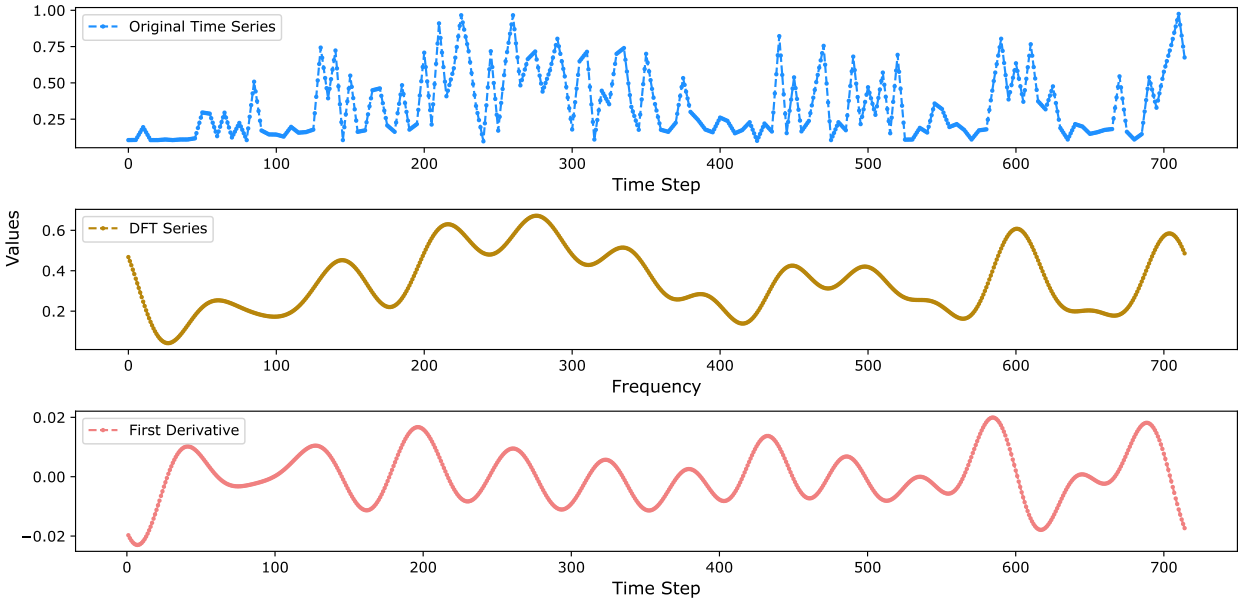
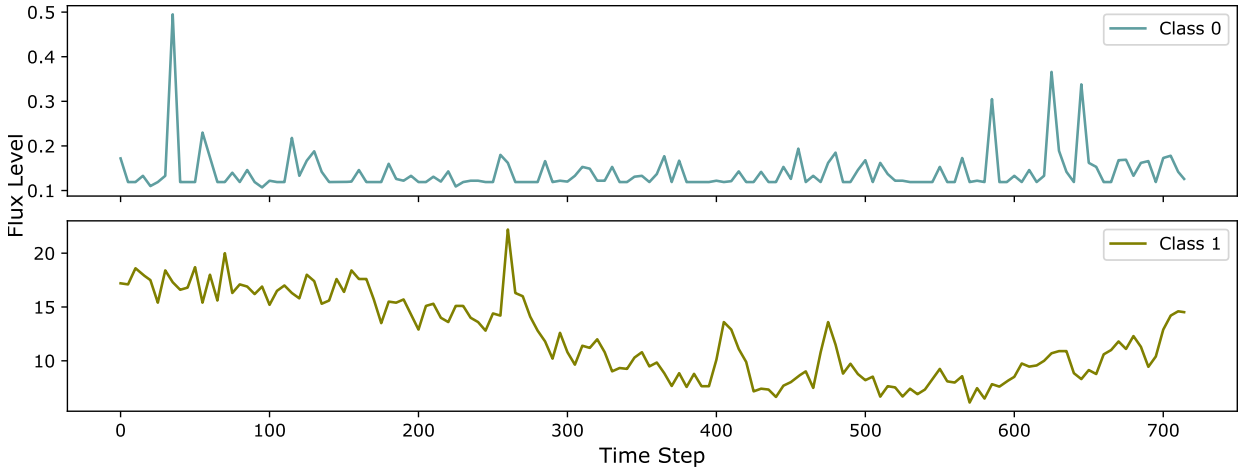
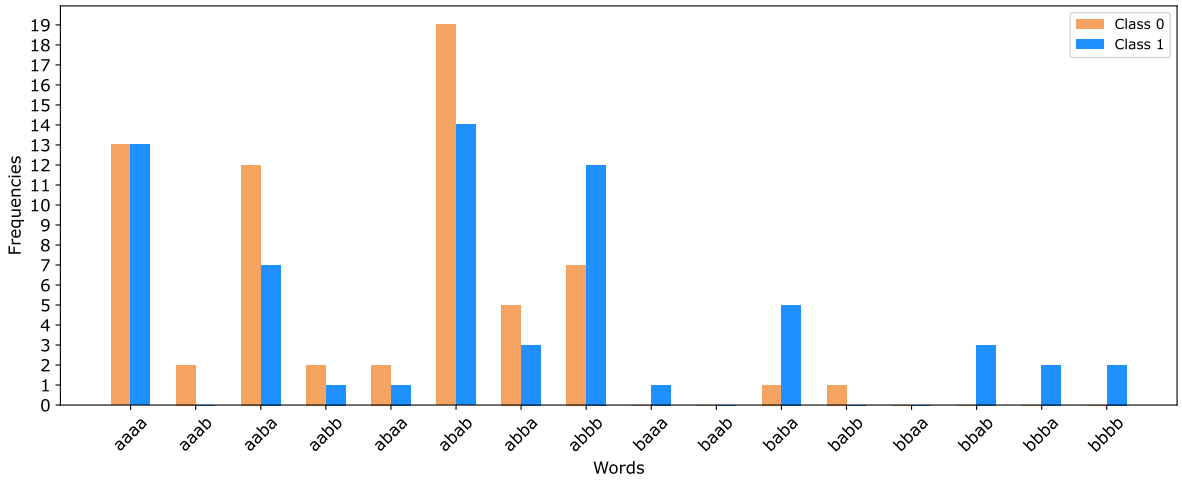


Figure 4.4 Schematic overview of time series representations of the STSF model.
Note: For a given original series, a periodogram representation derived from the discrete Fourier transom (DFT) and a first-order difference representation are generated to find candidate discriminatory intervals as a subset of the original time series. The discriminatory interval features constituting seven statistical parameters are obtained from all three (time, frequency, and derivative) domains prior to training the classifier.

Firstly, BOSS finds symbolic approximations using discrete Fourier transform (DFT). Then, it creates words and discretizes/vectorizes the input using words with multiple coefficient binning (MCB). This has the effect of reducing noise (Schäfer 2015). Finally, the algorithm uses a one-nearest neighbor over word frequency vectors and retains the estimators using the BOSS metric for best parameter training (Bagnall et al. 2017). Figure 4.5 illustrates these stages of the BOSS algorithm.



(a)



(b)

Figure 4.5 Schematic overview of bag-of-SFA symbols (BOSS) model.
Note: (a) Given a raw time series, a sliding window is applied to extract subsequences. Each subsequence is transformed into a word using the symbolic Fourier approximation (SFA) algorithm, and only the first occurrence of identical back-to-back words is retained. (b) Lastly, a histogram of the words is computed.

4.3.2 Data Partitions

For classification in a supervised setting where all the data has class labels, the data set is typically split into the training set and the test set (Hastie et al. 2009). The training set

is used to fit the data features on the parameters of the algorithms chosen to address the problem. The chosen algorithm is used to score the test set and determine the quality of the classifier. We partition our data into training-test sets with splitting criteria of 65-35 percent leading to 283 training samples and 150 test samples. A summary of the number of samples in each partition with respect to the target labels is presented in Table 4.2.

Table 4.2: Data Partitioning.

	Training	**Test**
Positive	167	77
Negative	116	73

Note: Number of instances in each partition corresponding to the binary target labels. Here, binary corresponds to a positive (strong SEP event) or negative (weak SEP event).

4.4 Results

In this work, we consider strong ($\geq$ S1) SEP events as a 'positive' class and weak events as a 'negative', thereby designing the problem as a binary classification task. The experiments are designed to fit a univariate model to a multivariate time series architecture for a short-term SEP event prediction system. We aim to demonstrate the robustness and compare the efficiency of time series classifiers towards generating short-term predictions during NRT operations. As explained in the previous section, the classifiers extract the features and data attributes from the input series. Because we want to aim at short-term predictions via SEP event classification, we consider 12 hours of observations minus five minutes before the SEP event onset. Here, the onsets are defined as follows: strong events crossing 10 pfu and

weak events surpassing 0.5 pfu in the P3 channel. We interpolate the five-minute proton time series to one minute to utilize the X-ray flux characteristics during flaring periods. The model hyperparameters considered are as follows: (i) Minimum interval length/window size is fifteen for TSF and BOSS, and (ii) Number of estimators is 200 for TSF and STSF.

4.4.1 Learning Curves

One of the essential tools in ML to trace the model performance is using learning curves. These curves visually indicate the sanity of a model for overfitting or underfitting during the training phase. They also help us to understand how the model performance changes as we input more training examples. In addition, these curves are useful to compare the performance of different algorithms (Perlich et al. 2003). Figure 4.6 shows the learning curves of the models in our consideration. Here, for providing a better performance estimate given the imbalanced nature of our data set, we use a 'weighted' average of F_1-scores (Manning et al. 2008) per class as defined in Equation 4.1.

$$F_{1_{weighted}} = \sum_{i=1}^{N} w_i \times F_{1_i} \tag{4.1}$$

$$F_1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{4.2}$$

As shown in Equation 4.2, F_1-score can be estimated as the harmonic mean of precision (Eq.4.3) and recall (Eq.4.4). Precision is used to evaluate the model's correct prediction

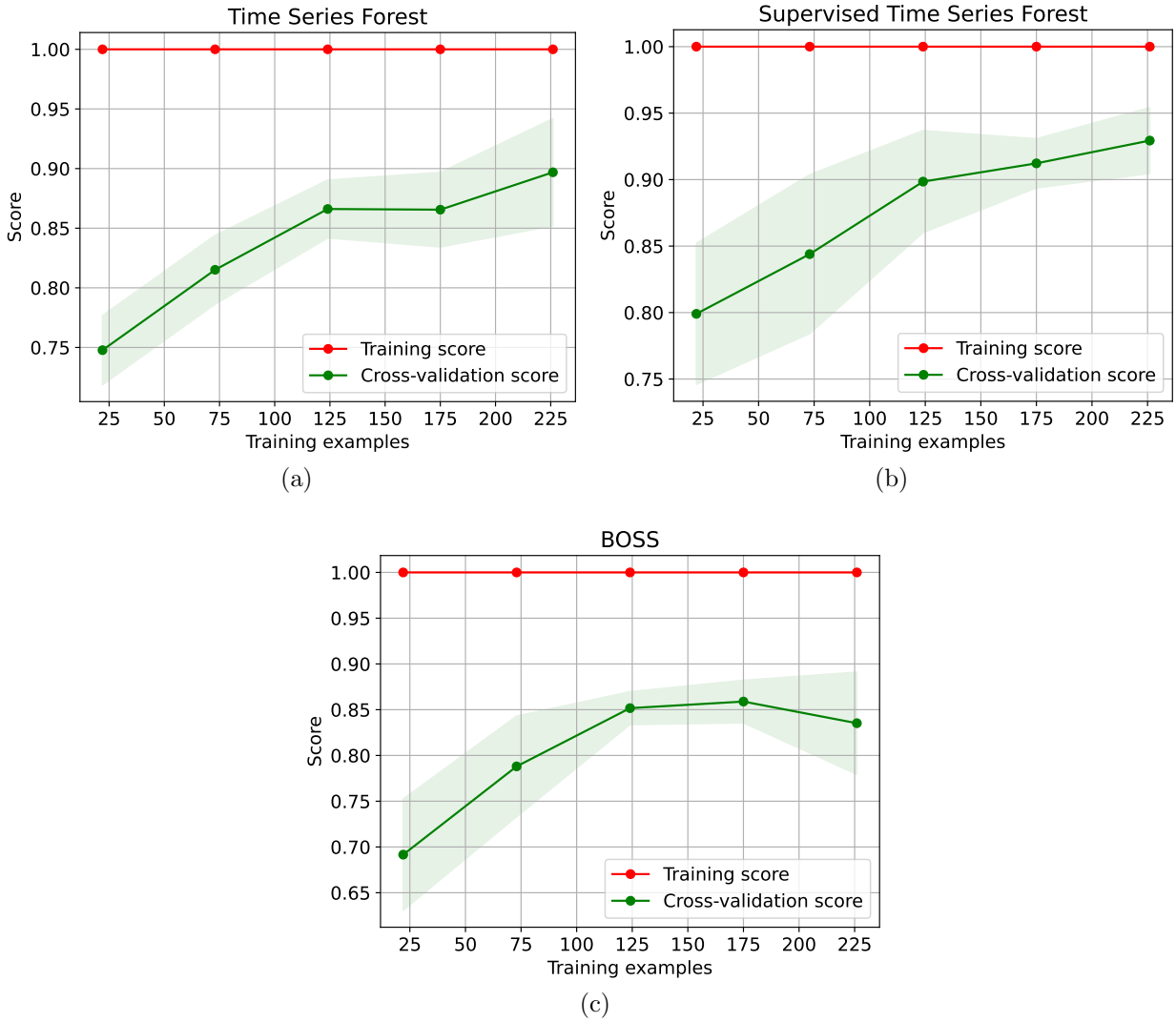


Figure 4.6 Learning curves for (a) time series forest (TSF); (b) supervised TSF (STSF); and (c) BOSS ensemble models.
Note: The weighted F_1-score has been used here as the scoring function. The red line represents the training score, while the green line shows the model estimations on validation. Here, the shaded region indicates the standard deviation of the validation score. The STSF model produces the best score ($\approx$0.925) at the end of cross-validation.

with respect to the false alarms. Recall characterizes the ability of the classifier to find all of the positive cases.

$$Precision = \frac{(TP)}{(TP + FP)} \tag{4.3}$$

$$Recall = \frac{(TP)}{(TP + FN)} \tag{4.4}$$

As we consider a 'weighted' average for the F_1-score, it computes the score for each target class and uses sample weights that depend on the number of instances in that class while averaging. The weight in the F_1-score is presented in Equation 4.5. Here, i is the number of target classes in the data set, which is two in the present work.

$$w_i = \frac{\text{Number of samples in class } i}{\text{Total number of samples}} \tag{4.5}$$

In our learning curves, the red line represents the training score, which evaluates the model on the newly trained data. The green line shows the estimations of the model on the samples used for validation. The shaded area represents the standard deviation of the scores after running the model multiple times with the same number of training data. It can be seen that the training score remains high for all models regardless of the size of the training set. In Figure 4.6(a), the steepness of the green line reaches a plateau between ≈ 125 to 175 samples but shows a small increment after 175 for TSF. On the other hand, the cross-validation score in Figure 4.6(b) for STSF greatly reduces after 125 samples. In Figure 4.6(c), the curve for BOSS model initially increases with the training size up to ≈125 but the slope reduces later, indicating that more training data is not helpful in the generalization

process. The STSF model achieves a high F_1-score (≈ 0.925) followed by TSF and then the BOSS model. Overall, the learning curves represent a satisfactory use of sample sizes to train the model efficiently. For TSF and STSF, we note that with more samples, this can be improved. In the remainder of this section, we present and discuss the implementation of several evaluation techniques to analyze the performances of the models on the test set.

4.4.2 Reliability Curves

In ML, reliability curves/calibration plots are used to better understand a model's confidence intervals in its prediction probabilities. Models such as decision trees give the label of the event but do not support native confidence intervals. A simple decision tree is a hierarchical tree structure used to determine classes based on a set of rules (questions) about the attributes of the data points (Safavian & Landgrebe 1991). Here, every non-leaf node represents an attribute split (question), while all the leaf nodes represent the classification result. In short, if the decision tree model is input with a set of features and corresponding classes, it generates a sequence of criteria to identify a data sample's target class.

We can evaluate the models based on multiple tools to be confident in our predictions. One method is calibration plots that check whether the predicted class distributions are similar to the true ones. Calibration curves (Wilks 1990) visually aid us in comparing how well the probabilistic predictions of a binary classifier are calibrated. Figure 4.7 shows the predicted probability of a model in each bin on the x-axis and the fraction of the positive label in that bin on the y-axis. The calibration intercept seen in a black-dotted line is a best-fit assessment. Values under the curve suggest overestimation, whereas values above

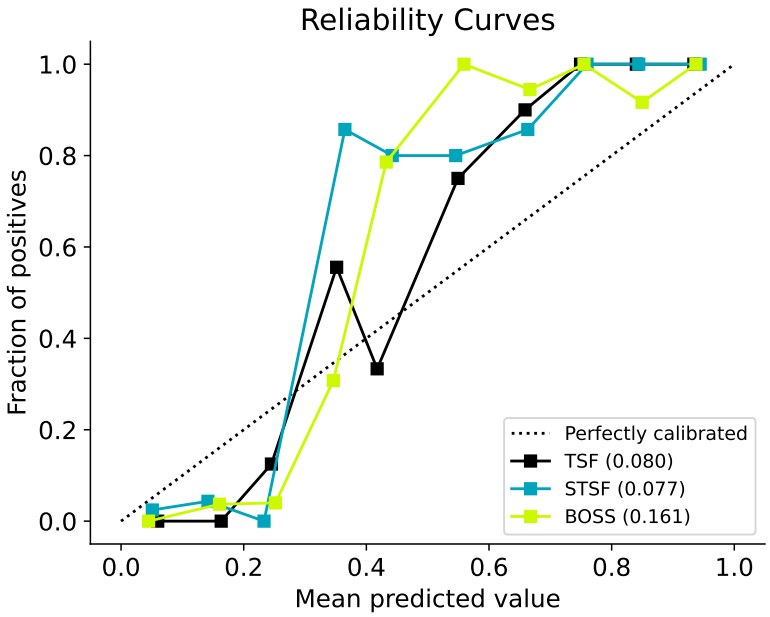


Figure 4.7 Reliability diagram or calibration plots of our models on the test set.
Note: The diagonal black dotted line shows the best fit. Data points above this line are underestimates, while those below it are overestimates. Shown in the legend are model names; TSF, STSF and BOSS with Brier score loss, respectively.

the curve suggest underestimation. TSF and STSF show close behavior in their average predictions over true values compared to the BOSS model. Nonetheless, all the models show underestimates of their predictive probability against the observed probability. In other words, this represents relatively lower confidence intervals in the model predictions. Hence, we use the Brier score (BS) loss (Murphy 1973) as defined in Equation 4.6 to evaluate the performance of the model.

$$BS = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{4.6}$$

Here, N is the number of data samples in the test set; y_i is the observed probability and

$\hat{y}_i$ denotes the prediction score (used as the estimated probability) of the i^{th} test sample. Brier score loss is strictly used to assess the calibration and discriminative power of a model, as well as the randomness of the data at the same time. The loss values range from 0 to 1, with 0 being a perfect score. In our case, TSF has **0.080**, STSF has **0.077** and BOSS has **0.161** as Brier score losses. Because of the low losses, our models indicate they are excellent predictors with more discriminatory power. Therefore, we further evaluate the model on the test set using popular metrics and compare their performances.

4.4.3 Evaluation

In Section 4.4.1, we have defined statistical metrics, such as precision and recall, that have been traditionally used to assess classifier performances. On a simple scale, accuracy (Eq.4.7) is another standard evaluation metric used to evaluate the quality of a classifier by counting the ratio of correct classification over total classifications.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{4.7}$$

Furthermore, we can focus on false negatives and measure the model performance using a receiver operating characteristic (ROC) curve. The ROC curve for the classifier is generated by plotting the true positive rate (TPR) against the false positive rate (FPR). The classifier predicts mean probabilities for each input instance belonging to the positive class, where the prediction score from the classifier is greater than a parametrized threshold. Then, a classification threshold (in the range 0 to 1) is used to assign a binary label to the predicted

probabilities. To find the optimal threshold that minimizes the difference between TPR and FPR of the classifier, we use the Youden Index (J; Youden 1950) defined in Equation 4.8. Here, sensitivity is the recall for the positive class and specificity is the recall for the negative class. We further explain our analysis on finding the optimal threshold in Appendix D.

$$J = Sensitivity + Specificity - 1 \tag{4.8}$$

The quality of the model is then assessed on the area under the ROC curve (AUC) for the positive class. The intuition behind this measure is that AUC equals the probability that a random positive sample ranks above a random negative sample. Ahmadzadeh et al. (2019) point out that the AUC is statistically consistent and more discriminating than accuracy. A measure of 1.0 for AUC signifies perfect classification, while a value of 0.5 means that the classifier cannot differentiate at all.

In Figure 4.8, we show the ROC curves for our models based on the TPR and FPR. Here, we indicate the optimal threshold of the classifiers in the upper-left corner of the ROC curve (as a blue star). Furthermore, the TSF has an ROC-AUC of 0.987, STSF has 0.981, and for BOSS, we get 0.966, indicating excellent discriminatory performance in all the classifiers. The skill scores and model evaluation discussed further are based on the specific chosen (that gives optimal results) threshold after our initial analysis: TSF = 0.40 (Fig. 4.8a), STSF = 0.39 (Fig. 4.8b), and BOSS = 0.59 (Fig. 4.8c). In Appendix D, we provide an evaluation of the influence of varying thresholds on the scores as shown in Figure 2.

A 2 $\times$ 2 contingency table constitutes the following elements: true positives (TP), true

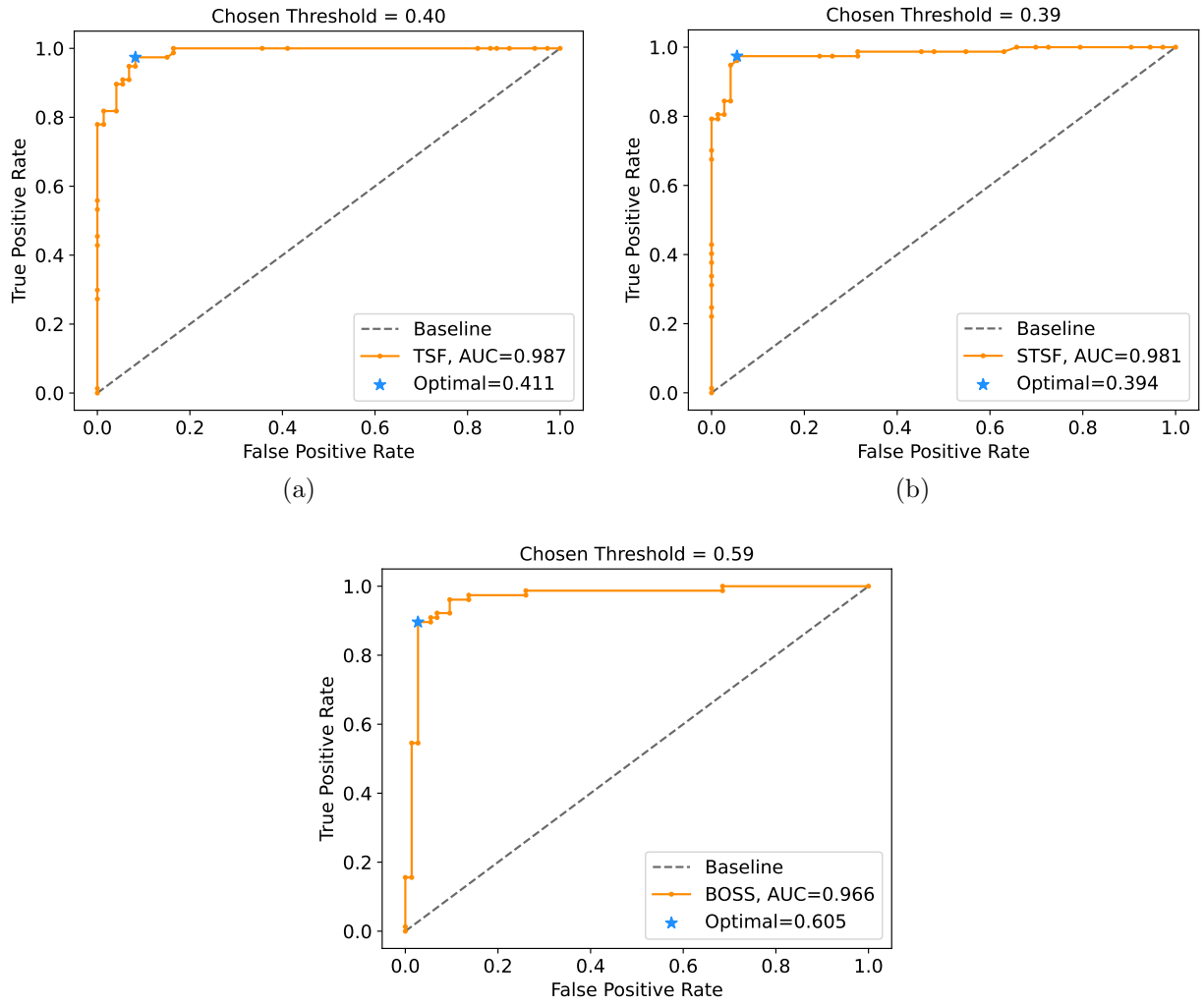


Figure 4.8 Receiver Operating Characteristic (ROC) curves for (a) TSF, (b) STSF and (c) BOSS models on the test set.
Note: Here, the area under the curve (AUC) is inset in the legend. The x-axis shows the false positive rate (FPR) and the y-axis shows the true positive rate (TPR) for the classifier. The dashed diagonal line indicates the ROC curve for a baseline or no-skill classifier. A starred point in blue color positioned at the top left of the plot indicates the optimal threshold value of the model. In addition, the chosen threshold to estimate the model skills is provided at the top of the plot for the model, respectively.

negatives (TN), false positives (FP), and false negatives (FN). Here, TP indicates the number

of correctly predicted strong SEP events (positive class) by a model, while TN represents

the number of correctly predicted weak SEP events (negative class). FP corresponds to the number of weak events predicted as strong (false alarms), while FN corresponds to the number of strong events predicted as weak (misses). Subsequently, the aim of our best model should be to reduce incorrect results represented by both FP and FN. In Table 4.3, we show the contingency tables based on the chosen classification threshold of our models on the test set. TSF and STSF indicate a relatively higher number of false alarms, but the BOSS model outputs a fairly close number of misses and false alarms.

Table 4.3: Contingency Tables for the Models on the Test Set.

		TSF		**STSF**		**BOSS**	
		Predicted		Predicted		Predicted	
		Strong	Weak	Strong	Weak	Strong	Weak
True	Strong	75	2	75	2	71	6
	Weak	6	67	4	69	5	68

Note: Truth tables based on the chosen classification threshold for all the models on the test set. The first column is a shared entry of true labels against predictive labels for each corresponding model. The elements indicate the number of predictions with respect to the actual occurrences in the test set.
Models: TSF - Time Series Forest; STSF - Supervised Time Series Forest; BOSS - Bag of SFA Symbols

Focusing on the importance of positive classes, we consider the F_1-score defined in Equation 4.1. It ranges between 0 and 1 such that scores closer to 1 indicate the model to be better. To account for the FPR, that is, compare the difference between the probability of

detection and the probability of false detection, we utilize true skill statistics (TSS; Woodcock 1976; Dann 1985) as shown in Equation 4.9. TSS ranges from -1 to +1, where the latter indicates a perfect score. TSS $\leq$0 indicates agreement no better than a random classification.

$$TSS = \frac{(TP \times TN) - (FP \times FN)}{(TP + FN) \times (FP + TN)} \tag{4.9}$$

Furthermore, the Heidke skill score (HSS; Heidke 1926) measures the improvement of the forecast over a random prediction as defined in Equation 4.10. HSS with 1 indicates perfect performance and 0 indicates no skill. A no-skill means the forecast is not better than a random binary forecast based on class distributions.

$$HSS = \frac{2 \times ((TP \times TN) - (FP \times FN))}{((TP + FN) \times (TN + FN)) + ((FP + TN) \times (FP + TP))} \tag{4.10}$$

The Gilbert Skill Score (GSS; Schaefer 1990) considers the number of hits due to chance, which is the frequency of an event multiplied by the total number of forecast events. This score formula is given by Equation 4.11. GSS ranges from -1/3 to 1. Here, 0 indicates no skill, while 1 is a perfect forecast.

$$GSS = \frac{TP - (\frac{(TP+FN)\times(TP+FP)}{TP+FP+TN+FN})}{(TP + FP + FN) - (\frac{(TP+FN)\times(TP+FP)}{TP+FP+TN+FN})} \tag{4.11}$$

However, accounting for the true negatives to assess the performance of a binary class problem is essential in our context. Hence, we also choose Matthew's correlation coefficient (MCC) as defined in Equation 4.12. MCC ranges from -1 to 1. Here, 0 indicates no skill,

while 1 shows perfect agreement predicted and actual values.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TP + FP) \times (TN + FN)}} \tag{4.12}$$

We approach the SEP event prediction problem from a time series classification perspective using the GSEP data set. The skill scores based on the respective chosen classification threshold for all our classifiers on the test set are presented in Table 4.4. One can see that the STSF model performs well compared to the TSF and BOSS models in terms of all the scores.

Table 4.4: Model Performances on the Test Set.

Model	**F_1**	**TSS**	**HSS**	**GSS**	**MCC**
TSF	0.947	0.892	0.893	0.807	0.894
STSF	**0.960**	**0.919**	**0.920**	**0.852**	**0.920**
BOSS	0.927	0.854	0.8533	0.744	0.853

Note: Class metrics are presented here for the best models implemented as an ensemble of univariate classifiers on the test set.
Models: TSF - Time Series Forest; STSF - Supervised Time Series Forest; BOSS - Bag of SFA Symbols.
Metrics: TSS - True skill statistics; HSS - Heidke skill score; GSS - Gilbert skill score; MCC - Matthews correlation coefficient.

As there is no one-to-one correspondence between the task, data set, and sampling implemented, we do not extensively compare our results with earlier studies. In Table 4.5, we list existing models that implement empirical or ML methods for predicting E≥10 MeV SEP events. The models in these studies have been developed focusing on a combination of various solar parameters, including solar flare X-ray fluxes and their properties. As can

be seen, the period considered in these studies varies depending on the availability of their desired data set. We include two common metrics; HSS and TSS (where available) used across these works in the table. HSS is an advanced metric and is highly dependent on the number of samples present in each binary class of a data set (Bobra & Couvidat 2015).

Table 4.5: List of Existing SEP Event Prediction Models That Consider Solar Protons, X-ray Flare Fluxes, and Their Properties as Input.

Model	**Period**	**Type**	**HSS**	**TSS**
Balch (2008)	1986 - 2004	Empirical	0.48±0.04	-
Laurenza et al. (2009)	1995 - 2005	Empirical	0.58	-
Winter & Ledbetter (2015)	1995 - 2005	Empirical	0.60	-
Alberti et al. (2017)	2004 - 2014	Empirical	0.55	-
Anastasiadis et al. (2017)	1984 - 2013	Empirical	0.37±0.011	0.5
Engell et al. (2017)	1986 - 2018	ML	0.58	-
Papaioannou et al. (2018)	1997 - 2013	Empirical	0.65	-
Lavasa et al. (2021)	1988 - 2013	ML	0.69±0.04	0.75±0.05
Aminalragia-Giamini et al. (2021)	1988 - 2013	ML	-	0.79
Sadykov et al. (2021)	2010 - 2019	ML	0.434±0.046	0.821±0.003

Note: HSS - Heidke skill score; TSS - true skill statistics; ML - Machine Learning

While we make short-term predictions, other works typically focus on forecasting SEP event onset hours and days ahead. Moreover, no previous work has focused on the classification task between strong and weak SEP events. Nonetheless, in addition to other evaluation methods demonstrated in this paper, our results show great performance potential in us-

ing column ensembles of the time series ML. The interval-based STSF model architecture demonstrated in this paper promises to be helpful to be implemented in NRT operations. In Appendix E, we show the effect of randomness in the TSF and STSF architectures on the optimal threshold for classification and further establish confidence and robustness in our predictions. Therefore, our future work will transform the capacity of the STSF model to provide short-term predictions on NRT data.

4.5 Conclusions

Solar energetic particle (SEP) events are one of the main elements of space weather, along with solar flares and coronal mass ejections. Towards predictive efforts of SEP events, we utilize the recently developed GSEP data set (Rotti et al. 2022) publicly available from Harvard Dataverse: https://doi.org/10.7910/DVN/DZYLHK. The data set constitutes *in situ* time series measurements from the NOAA-GOES missions for solar cycles 22 to 24. They are long band (1–8Å) X-ray measurements from the XRS instrument and proton fluxes (P3, P5, P7) from the SEM instrument. We use these parameters to evaluate the performance of our multivariate time series models. The target labels are defined based on integral proton fluxes (I_P) recorded by the GOES P3 channel. Positive labels are strong SEP events crossing the 10 pfu threshold; negative otherwise.

There are 433 SEP events in the GSEP data set, of which 244 are strong. We consider a fixed length of 12 hours minus five minutes of fluxes before the SEP event onset constitutes the observation window. Therefore, the total length for each time series corresponds to

715 instances. Our focus in the present work is to see whether the model can classify the P3 proton channel flux to be crossing the 10 pfu limit or not. In other words, if the 10 pfu limit is outset in the 10 MeV channel, then the model outputs a "true" or "yes" label indicating a strong event. If not, then it is a weak or sub-event. When implemented in NRT operation, the yes/no outputs from the models are in succession for the next few minutes of the prediction window.

Machine learning (ML) methods are at the forefront of the latest techniques in space weather forecasting. The crucial focus on implementing ML towards SEP event forecasting is for the upcoming NASA human missions to the Moon and Mars (Whitman et al. 2022). In this scenario, short-term forecasts become relevant and require distinct attention to precise and sensitive prediction of strong SEP event occurrences. This work implements time series-based ML models in a binary classification schema. Because no single algorithm always creates the best results, we want to experiment with multiple models and evaluate their performances.

Interval-based methods are based on splitting the time series into phase-dependent distinct intervals. Statistics are gathered from each interval to fit individual classifiers on the data. The final classification is assigned based on majority voting of the most common class generated by the individual classifiers. We consider two interval-based classifiers in our work. They are time series forest (TSF) and supervised time series forest (STSF). TSF is a collection of decision trees applied to the feature sets (mean, standard deviation and slope) extracted from the intervals. Here, the average prediction from each tree is obtained, and

based on a majority vote, the final output is predicted. STSF builds on the TSF model by implementing a metric to supervise the random sampling such that the subsamples represent the entire series. Statistical features such as mean, median, standard deviation, slope, min, max and interquartile range are extracted from each interval for three representations (time, frequency and derivative). The classifier then concatenates these extracted values to form a new dataset and builds a random forest model to make predictions. Another model we implement is the BOSS ensemble, a dictionary-based algorithm. In that, small intervals of length 'l' are transformed into "words" and stored as histograms for each input time series. The occurrence of the word during prediction is used to classify the series to a label on a weighted output.

The learning curves of our classifiers indicate sufficient data used during the training phase. On the test set, we estimate the confidence intervals of the predictions using reliability diagrams and use Brier score loss in our evaluation strategy. We construct the ROC curve for our models and identify the best classification threshold to transform the probabilistic decisions into binary labels. We use the area under ROC curve (AUC), F_1-score, true skill statistics (TSS), Gilbert skill score (GSS), Heidke skill score (HSS), and Matthews correlation coefficient (MCC) to further assess the performances of our models.

The results in this paper shows that the STSF classifier performs well compared to the TSF and BOSS models. Multiple evaluation schemes relatively indicate that our model obtains the best scores compared to existing methods but in the framework of SEP event classification. In addition, our work shows that interval-based classifiers have great potential

to improve short-term forecasts, and an ensemble model is a suitable predictor for use in an operational context.

The SEP prediction model we have developed in this paper is very high confidence. Our objective is to develop a short-term SEP event forecasting algorithm to predict whether the solar proton flux level will surpass the SWPC 'S1' threshold. In that respect, our approach is very different from the standard SEP prediction methods, which forecast the likelihood of an SEP storm in the coming 24 or 48 hours. Our model would allow for SEP warnings to be called off at the last minute and for high-level (E≥10 MeV) SEP event forecasts to be confirmed with high certainty or issued if there is no longer-term alert. Certainly, the latter case will be extremely valuable for Artemis astronauts in extra-vehicular activities (EVAs) or on the surface of the Moon. If reliable, our model will give the real-time forecasters at the Space Radiation Analysis Group (SRAG) a useful tool to help them decide whether to issue an alert. In an operational setting, we envisage our system to sit on top of forecasts with a much longer prediction horizon but lower precision, such as current forecasts[4].

More avenues can be explored for future work, which includes but is not limited to extending the analysis to (1) consider "no-SEP" phases i.e., SEP-quite periods following the occurrence of large (≥M1.0) flares, and (2) build different ensemble strategies.

[4]CCMC SEP-scoreboard: https://ccmc.gsfc.nasa.gov/scoreboards/

CHAPTER 5

Precise and Accurate Short-term Forecasting of Solar Energetic Particle Events with Multivariate Time Series Classifiers

The contents of this chapter have been submitted to the Astrophysical Journal.

5.1 Introduction

The activity of the Sun is considered the prime source of space weather (SWx) that constitutes different eruptive phenomena including solar flares (SFs) and coronal mass ejections (CMEs). SF is a sudden brightening in the solar atmosphere in a coronal soft X-ray source that is observed to have close relationships with CMEs (Feynman & Hundhausen 1994) and other observable phenomena such as jets and filament eruptions (Schrijver & Siscoe 2010). CMEs are clouds of plasma formed in the lower corona that often move faster (>1000 km.s^{-1}) than the ambient solar wind (Low 1996; Chen 2011). Another manifestation of solar activity constituting the emission of energetic electrons, protons, and heavier ions from the Sun is called a solar energetic particle (SEP) event (Klein & Dalla 2017). Typically, large flares and shock fronts of fast CMEs are known to accelerate SEPs and are often considered precursors or parent eruptions (Cane et al. 1986; Kahler 1992; Gopalswamy et al. 2001; Marqué et al. 2006; Gopalswamy et al. 2008; Swalwell et al. 2017; Gopalswamy et al. 2017; Cliver & D'Huys 2018; Rotti & Martens 2023). The Sun releases ions of about tens of Mega electron-Volt (MeV) and more in many SEP events. The motion of such highly charged particles is dictated by magnetic field lines (Reames 2013). To be geo-effective, SEPs should reach the near-Earth space on a magnetically well-connected path (Cane et al. 1988).

Generally, it is understood that the eruptions at the western side of the Sun have a higher probability of SEPs being geo-effective due to the spiral structure of the interplanetary magnetic field lines, known as the Parker spiral (Parker 1965; Reames 1999). Such Earth-bound SEPs are hazardous on many levels. The impacts include severe technological (Smart & Shea 1992) and biological effects on various economic scales (Schrijver & Siscoe 2010). Although Earth's magnetic field provides us a protective shield from the energetic particles and filters them out from reaching the ground, they can be life threatening for humans on missions outside of the Earth's magnetosphere and aircraft travels along polar routes (Beck et al. 2005; Schwadron et al. 2010). For instance, long-lasting strong SEP events pose a radiation hazard to astronauts and electronic equipment in space (Jiggens et al. 2019). In addition, the particles getting diverted to Earth's magnetic poles disturb the ionosphere's upper layers, causing disruption to high-frequency radio, GPS, and satellite communications (Desai & Giacalone 2016). Over the last four decades, great progress in space exploration has provided near-continuous observations of solar activity from a fleet of advanced space-based satellites (Usoskin 2017). These observational data should be analyzed in operational contexts to mitigate SWx effects on our human explorers and technological systems (Jackman & McPeters 1987). Therefore, we require robust tools and systems to forecast eruptive event occurrences such as SEPs and send warning signals before the event.

Researchers across the globe have been focusing on implementing a variety of model-driven techniques for predicting SEP events, mostly concentrating on predicting the peak flux-related characteristics. To forecast SEP event occurrences, physics-based and data-

driven statistical models have been designed based on the parameters of parent solar eruptions such as SFs and CMEs (Van Hollebeke et al. 1975; Posner 2007; Kahler et al. 2007; Balch 2008; Laurenza et al. 2009; Núñez 2011; Falconer et al. 2011; Dierckxsens et al. 2015; Anastasiadis et al. 2017; Alberti et al. 2017; Papaioannou et al. 2018).

Machine learning (ML) methods have been at the forefront of SEP event forecasting in the last decade due to their success in many other areas of research and operations (Camporeale 2019; Whitman et al. 2022). Our previous work (Rotti, Aydin & Martens 2024) implemented time series classifiers on a multivariate time series (MVTS) data set of SEP events and showed that the interval-based supervised time series forest (STSF) model is more efficient in the classification of strong SEP events. Possible areas of improvement here are to consider an actual class imbalance scenario, that is, including the phases of non-SEPs and comparing with baseline time series classification algorithms for the task at hand. Expanding on these ideas in the present work, we implement STSF, summary statistic classifier (SSC) and one-nearest neighbor (1NN) classifiers on an extended data set constituting SEP-quiet periods and analyze their performances. With that, we compare the results and establish a proof of concept of the model to be implemented in a near-real-time (NRT) scenario. That is, we aim to build a high fidelity (low-risk) and interpretable short-term (low lead time) predictive model for reliable SEP event forecasting systems suitable to operational standards.

SEPs observed in near-Earth space are rare events compared to major flares and CMEs. In addition, the occurrences of SEPs are dependent on precursor solar eruptions (Cliver et al. 2022; Rotti & Martens 2023). Thus, it makes physical sense to build SEP predictive

models based on the parameters of parent solar events. However, in the present work, we consider only SFs and defer to the use of CME data due to their low data quality, such as the projection effects and difficulty of tracking them as three-dimensional objects.

The rest of the paper is organized as follows: Section 5.2 provides information about our data set and data preparation steps used in this work. Section 5.3 presents our research methodology, including the descriptions of the time series classification models and data partitioning scheme. Section 5.4 discusses the training and validation phases of the model and presents the experimental evaluation framework. Lastly, Section 5.5 provides a summary of our work and future endeavors.

5.2 Data

In this work, we implement MVTS data integration, model training and selection pipeline toward building an efficient architecture for short-term SEP event forecasting. We consider a binary classification task between strong SEP and weak or non-SEP events. In the present framework, we consider two classes, "SEP" and "SEP-quiet," where SEP-quiet periods include smaller proton flux increases, as well as the flares that did not lead to SEP events near Earth. With that, we maintain a natural class imbalance between the occurrence and non-occurrence of strong SEP events. Our methodology for the creation of SEP-quiet samples will be presented in the next Section. The strong and weak SEP samples in our work come from the Geostationary Solar Energetic Particle (GSEP) events data set (Rotti et al. 2022), which is discussed later in Section 5.2.4.

5.2.1 SEP-quiet Periods

The National Oceanic and Atmospheric Administration (NOAA) has been operating the Geostationary Operational Environmental Satellite (GOES) series in geosynchronous orbits since 1976 (Sauer 1989; Bornmann et al. 1996). The space environment monitor (SEM) onboard GOES missions measures solar radiation in the X-ray and EUV region and the in-situ magnetic field and energetic particle environment (Grubb 1975). In this work, we utilize the GOES solar X-rays (Garcia 1994) and energetic proton fluxes (Onsager et al. 1996). We create our negative samples using "SEP-quiet" periods in two stages: (i) obtaining lists of SF and SEP events and (ii) gathering the corresponding GOES X-ray and proton flux measurements for each event.

NOAA-GOES has traced flaring activity since the beginning of 1974 and offers a catalog[1] with spatial and temporal specifications, flare magnitude, and associated active region (AR) information. However, some flare locations and/or AR numbers are missing from the GOES catalog or probably have known errors (Milligan & Ireland 2018; Angryk et al. 2020; Rotti et al. 2020b). NOAA categorizes SFs based on soft X-ray peak flux in the wavelength range 1–8 Å. Flare classes from least to most intense are labeled as A, B, C, M, and X, where, each category indicates an increase in flare intensity on a logarithmic scale.

5.2.2 Non-SEP Samples

We select our non-SEP samples such that there are no enhancements or relatively no enhancements in the GOES proton fluxes following flaring episodes. In other words, we identify

[1]https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-flares/x-rays/

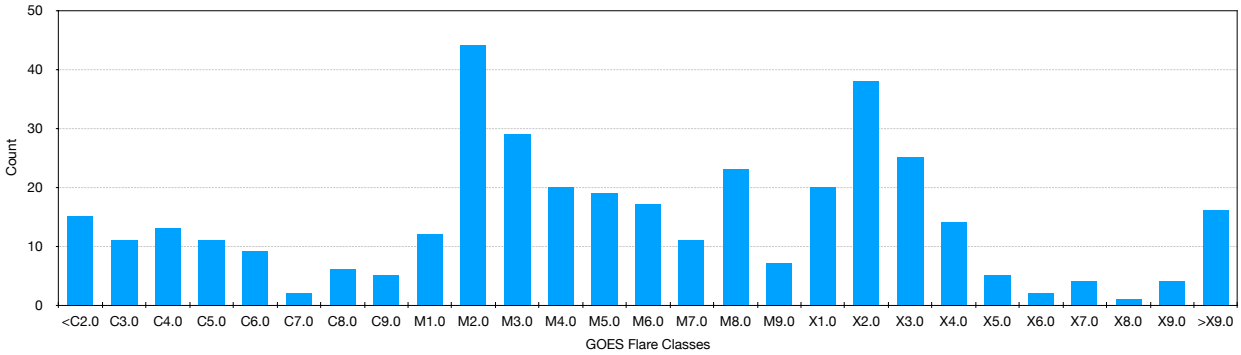


Figure 5.1 Distribution of soft X-ray flare peak intensities based on GOES flare classification for SEP-associated flares in the GSEP data set.
Note: There are a total of 383 SEP with flare class information, of which 59 events have flare magnitudes less than GOES-class C6.0.

large flares that do not lead to significant variations in the GOES $\geq$ 10MeV proton fluxes relative to the background levels. This step helps us increase the sample size over the previous study. At first, we chose M and X-class flares from the NOAA flare list between 1986 and 2018 that have not been associated with strong SEP events. Then, to add a few more negative samples, we extend our lower limit of a "non-SEP" flaring intensity using a data-driven approach.

The GSEP data set consists of carefully identified SF information for 383 associated SEP events. First, we inspect these SEP-associated flares as shown in Figure 5.1 and find that <C6.0 flares correspond to the bottom 15th percentile. Hence, we consider C6.0 as our bottom threshold for flares. Then, we add all flares from M1.0 to C6.0 to our earlier flare list and obtain a total of 7,981 flares. We filter the preliminary negative-sample flares in two steps: (1) we do not consider flares during an ongoing SEP event (either rise or decay) and (2) we remove all consecutive flares within 11 hr following the onset of the flare at consideration. With this filtering step, our final count of non-SEP flares with an intensity $\geq$C6.0 reduced

to 2,460. In Figure 5.2, we show the distribution of the number of "SEP-quite" flares in each sub-divisions of GOES (5.2a) C-, (5.2b) M-, and (5.2c) X-class.

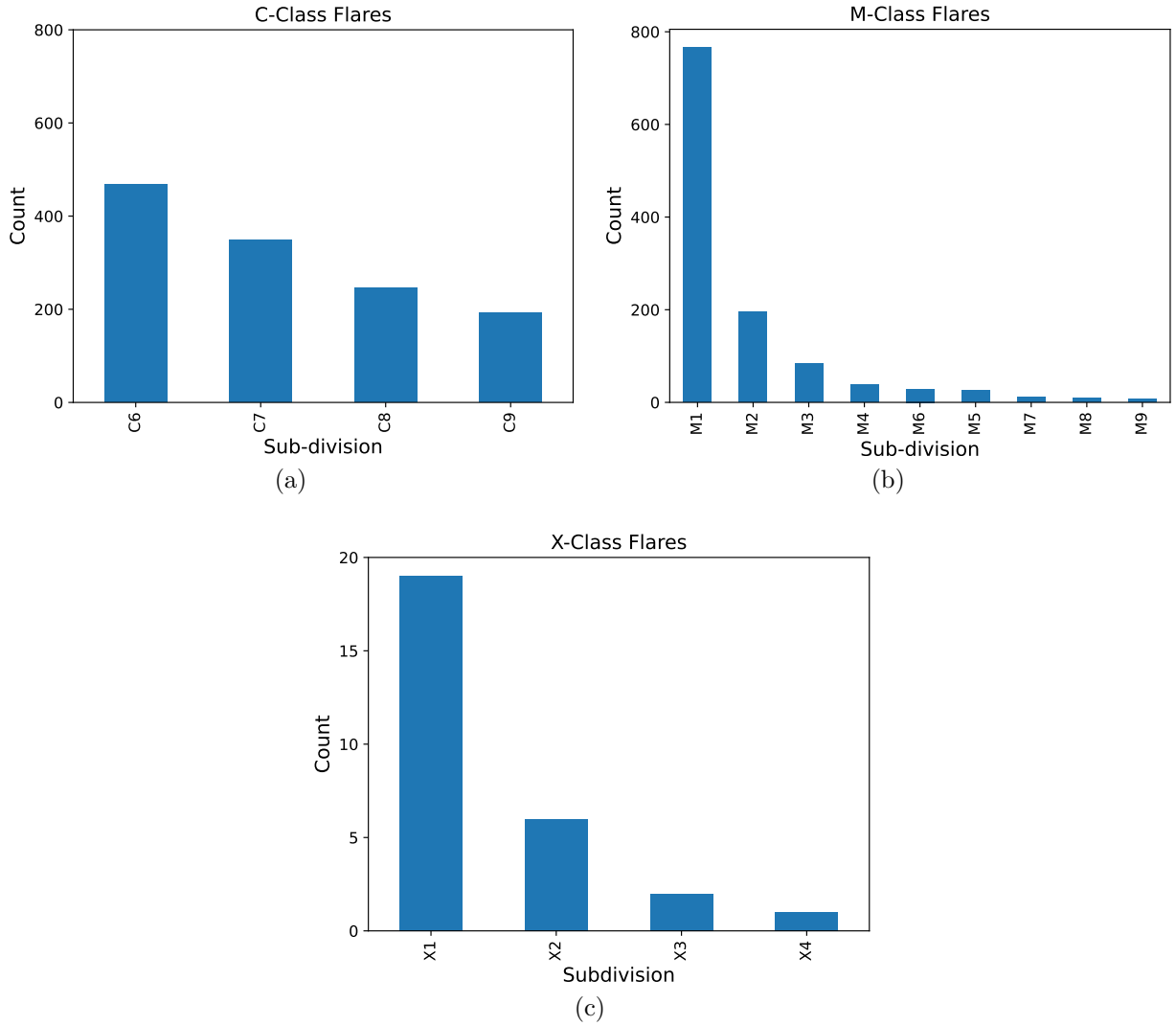


Figure 5.2 Distribution of the GOES subdivisions of (a) C- (b) M- and (c) X-class flares present on our "SEP-quiet" periods.

Whitman et al. 2022 recently organized an SEP model validation (SEPVAL) challenge in 2023[2]. For this, two lists of 33 SEP and 30 non-SEP events from solar cycles (SCs) 24 and

[2]SEPVAL Challenge: https://ccmc.gsfc.nasa.gov/community-workshops/ccmc-sepval-2023/

25 have been provided. In this regard, we cross-verified our non-SEPs with the SEPVAL's list of 15 non-SEP events for cycle 24. We do not include any (SEP or non-SEP) events from SC25 in the present work to make our analysis comparable with our earlier studies.

5.2.3 Time Series Slicing

To generate our dataset, we have interpolated and integrated particle and X-ray fluxe data from several GOES missions. First, we use the one-minute averaged X-ray (1–8 Å) fluxes measured by the X-ray sensor (XRS) onboard GOES. The archived data is available online from NOAA's website[3]. In addition to X-ray data, we use the following five-minute averaged integral proton fluxes measured by the SEM suite: (1) E≥10 MeV fluxes corresponding to P3, (2) E≥50 MeV fluxes corresponding to P5, and (3) E≥100 MeV fluxes corresponding to P7. We interpolate the proton fluxes to one-minute averages and retain the time stamp of every first measurement. The rationale here is to capture the flare dynamics that occur over a few minutes. In the present work, we follow the same procedure as discussed in Rotti et al. (2024) to generate our event slices for the additional negative samples and ensures that our analysis and models' predictive capability are relevant and comparable across different SEP event prediction tasks. We consider flares as the precursor to the onset of an SEP event.

In our earlier statistical study on strong and weak SEP events (Rotti & Martens 2023), we find that 70(±7)% of SEP event onsets occur within 11 hrs after the associated flaring eruption. However, sampling flares from the onset will likely lead the classifier to memorize the pattern of finding a spike at the start of the MVTS instance. A simple solution to

[3]GOES-XRS: https://www.ncei.noaa.gov/data/goes-space-environment-monitor/access/avg/

randomize the sampling is to include 1 hour of data prior to the flare onset because two to three flares may occur within an hour of observation in many instances. With that consideration, the classifiers become free from memorizing to find a spike at the start of the data or at the end of 60 data points corresponding to one hr. We accommodate both these aspects while creating our non-SEP time series sampling because our goal is to develop a highly accurate model predicting strong SEP events.

Based on our filtered list of 2,460 flares for negative samples, we consider 1 hour of data (all four fluxes mentioned above) before the flare onset and 11 hr afterward for each flaring event (as shown in Figure 5.3). Therefore, the total length of each MVTS corresponds to 12 hr. As our final data set comes with a cadence of one minute, each MVTS is a size-720 sequence of soft X-ray and integral proton fluxes. All time series data was cleaned for data gaps with linear interpolation and spike removal by replacing what would be considered an outlier (spike) with an interpolated value from moving averages. We note here that the fluxes are predominantly derived from primary GOES missions. Nonetheless, when there are ambiguous measurements in the primary satellite, we consider data from the secondary satellite of the corresponding period of observation. Because we are focusing on short-term predictions, we explore multiple prediction windows: 5-, 15-, 30-, 45-, and 60-min. In each case, we exclude the respective minutes of observations from the 720-size MVTS at the end by maintaining a constant 11 hr observation window as input to the model. Such a continuous and consistent observation window is needed for training a predictive model to be established in a near-real-time operational environment.

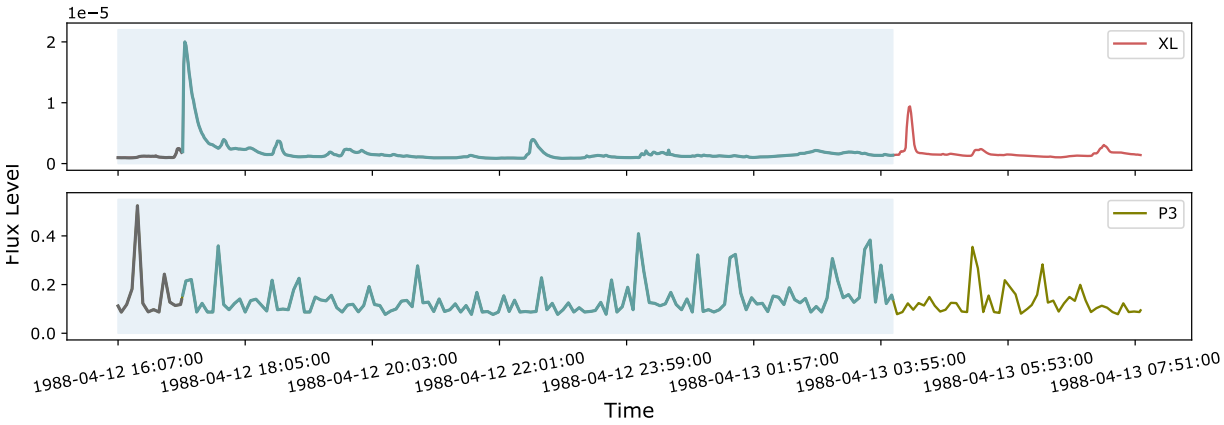


Figure 5.3 An example of the sampling technique for the non-SEP samples in our data set. Note: We show the XL and P3 fluxes for representation only. The blue-shaded region in the plot shows the 12-hour length of the input time series of our sample. Here, the grey-shaded line on the left corresponds to the one hour of fluxes prior to the (≥C6.0) flare onset, while the rest of the time series within the blue region covers 11 hours after the onset.

5.2.4 GSEP Data Set

Our primary source of SEP events comes from the recently published open-source GSEP data set[4]. It consists of a carefully vetted SEP events catalog with associated parent eruption metadata and time series slices. The data set comprises 433 (- 244 strong and - 189 weak) SEP events covering solar cycles 22 to 24. According to the NOAA Space Weather Prediction Center (SWPC), a strong SEP event is defined as having proton intensities I_P ≥10 pfu (1 pfu = 1 particle per cm^2.s.sr) in the GOES five minutes averaged E≥10 MeV integral energy channel for at least three consecutive readings (Bain et al. 2021). In the GSEP data set, "weak" SEP events correspond to an enhancement in proton intensities in E≥ 10 MeV channel whose peak flux measurements do not reach to 10 pfu. The description of the data

[4]The GSEP data set available on Harvard Dataverse: https://doi.org/10.7910/DVN/DZYLHK

set and its development can be found in Rotti et al. (2022) and Rotti & Martens (2023).

5.3 Methodology

We consider the term 'SEP events' analogous to solar protons events (SPEs). While variations exist, positive labels are usually associated with the occurrence of strong SEPs based on the integral proton fluxes recorded by the GOES-P3 channel crossing the 10 pfu threshold. As mentioned earlier, we have 189 weak SEP events in the GSEP data set and we add 2,460 non-SEPs to increase the negative samples to 2,649. For the positive class, we have 244 strong SEP events. The total size of the data set is 2,893 samples with a class-imbalance ratio of $\approx$ 1:11. Figure 5.4 shows the total number of samples in our data set from 1986 to 2018. Here, 'Events" are defined as those SEPs crossing the SWPC threshold of 10 pfu in the E$\geq$10 MeV channel, and "Non-Events" constitute both weaker enhancements of $<$10 pfu and no enhancements. In Section 5.3.1, we briefly explain the method of time series classification implemented in the present work.

5.3.1 Time Series Classification

We examine the performance of time series classifiers in the framework of a binary classification task for predicting SEP events. Time series classification uses supervised ML to analyze labeled classes of time series data and then predicts the class to which a new instance belongs. This is important in SWx predictions, where observational data is analyzed to support NRT operations and classification accuracy becomes critical. Furthermore, for short-term predictions, the algorithm must be highly accurate and robust to be useful.

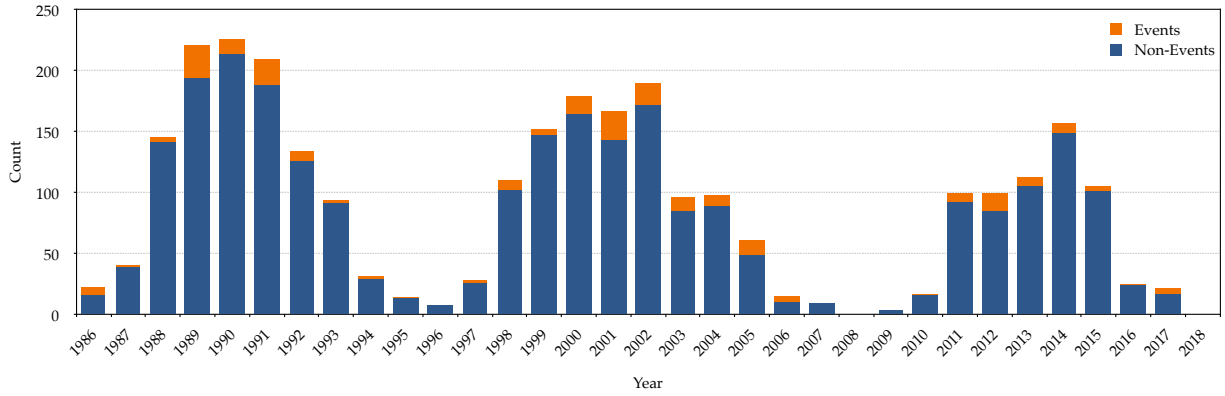


Figure 5.4 Distribution of SEP and non-SEP samples in our data set between 1986 and 2018. Note: We split our data set into three parts of non-overlapping years: Partition 1 = 1986 - 1992; Partition 2 = 1993 to 2002; and Partition 3 = 2003 to 2018. Here, each split consists of 998, 974 and 921 samples, respectively. The data set has a cumulative class-imbalance ratio of $\approx$ 1:11. In the legend, "Events" are defined as those SEPs crossing the SWPC threshold of 10 pfu in the E$\geq$10 MeV channel, and "Non-Events" consists of both weaker enhancements $<$10 pfu or no enhancements.

In our previous work (Rotti et al. 2024), we showed that an ensemble framework of interval-based classifiers achieved high performance in classifying between strong and weak SEP events. In the present work, we implement three time series classifiers and compare their performance to establish certainty in our SEP event prediction system. The motive is to see whether the input time series span leads to a strong SEP event (positive class) or not (negative class). We implement the same column ensemble schema presented in our previous work (Rotti et al. 2024), where the ensemble estimator allows multiple feature columns of the input to be transformed separately. The statistical features generated by each classifier on samples of the original time series are ensembled to create a single output. Each feature is assigned a score that indicates how informative it is towards predicting the target variable (Hansen & Salamon 1990; Schapire 1990; Arbib 2003). The predictions from individual

classifiers are then aggregated with equal votes using prediction probabilities. We will give a brief overview of our models in the following sections.

5.3.2 Summary Statistic Classifier

The feature-based summary statistic classifier is the first of the two baseline models we have used in this work. This classifier applies a time series feature transformation to an entire univariate time series using simple summary statistics and builds a random forest using these summary statistics. The summary statistics extracted are mean, standard deviation, minimum, maximum, and quantile values (at 0.25, 0.5, and 0.75, corresponding to first and third quartiles and median). The tabulated summary statistics are then fed to a random forest classifier for training a univariate time series classifier. For applying this method to our multivariate time series classification task, we employed a late fusion approach, where univariate time series classification models trained on individual parameters are ensembled based on a majority voting schema.

5.3.3 One Nearest Neighbor

The similarity-based (nearest neighbor) time series classification is one of the most popular classification methods due to its simplicity, nonparametric nature, potential for interpretability, and flexibility in utilizing different similarity schemes/metrics. It assigns a label for a new instance based on the target value of k-most similar instances in the training dataset. The similarity is assessed based on the inverse of a designated time series distance measure and utilization of elastic distance measures allows capturing the local and global similarity

patterns (Faouzi 2022). In this work, we implement a k-nearest neighbor (kNN, with k=1) classifier with Dynamic Time Warping (DTW) distance, which is a commonly used elastic measure (Sakoe & Chiba 1978). This 1NN is a baseline classifier for time series classification and similar to our earlier approaches, we use 1NN with DTW distance to issue predictions.

5.3.4 Supervised Time Series Forest Classifier

The supervised time series forest (STSF; Cabello et al. 2020) classifier comes under the category of interval-based models, which uses three representations (time, frequency, and derivative) of the input time series to extract features from a supervised selection of intervals. That is, for each representation, STSF builds an ensemble of decision trees on intervals selected through a supervised process wherein the algorithm finds the discriminatory intervals. STSF computes the region of interest (ROI) to highlight the location of discriminatory intervals, which are the intersected regions of such intervals. It extracts seven features, namely mean (μ), standard deviation (σ), slope (m), median, interquartile range (IQR), minimum value and maximum value, from each interval. The ranking of the interval feature is obtained by a scoring function that indicates how well the feature separates one class of time series from the other classes. The final set of intervals is obtained in a top-down approach to represent the entire series. The feature set is concatenated to form a new data set upon which decision trees are built. The final output is based on majority voting of averaged probability estimates of the individual estimators in the ensemble.

5.3.5 Data Partitions

As we are implementing a classification task with target labels, we split our data set into three sets: the training set, the holdout or validation set, and the test set. This partitioning schema is relevant to estimate the predictive capacity of the model on different aspects. For instance, the validation set is used in the present work to obtain the optimal threshold of classification for each model, estimate the skills scores and compare their performance. Furthermore, we use the unseen test set to sanity-check the best model's predictive capabilities.

We split our data set into three sets on non-overlapping years (1986-1992, 1993-2002 and 2003-2018) such that there are 998 samples for training, 974 samples for validation and 921 samples for testing the model. Here, each partition consists of a similar number of positive samples (≈ 80) and has similar class imbalance ratios, that is, $\approx$ 1:11. Table 5.1 shows the number of data samples in each partition in relation to the binary class labels.

Table 5.1: Partitioning Strategy of Our Data Set.

	Training	**Validation**	**Test**
Positive	80	80	84
Negative	918	894	837

Note: Number of samples in each partition corresponding to the binary target labels. Here, labels correspond to a yes (positive) or no (negative).

5.4 Results

In this work, we aim to demonstrate the robustness of the methodology for SEP event forecasting that was proposed in our previous work (Rotti et al. 2024) and provide a proof of concept on its efficiency in generating short-term predictions. For this purpose, we compare three classifiers (SSC, 1NN and STSF) and analyze their performances. We construct a binary classification task such that the 'positive' class consists of strong SEP events while the 'negative' class contains the rest.

We use the training set to train the model and perform a grid search for hyperparameter optimization of our classifiers. The best hyperparameters for STSF and SSC were on the default model settings with the number of estimators = 200. We explore the model capabilities for different short-term prediction windows: 5-, 15-, 30-, 45-, and 60-minutes. We use the following terms henceforth to identify different prediction windows that act as lead times for model predictions, respectively: T_5, T_{15}, T_{30}, T_{45} and T_{60}. In each experiment, we consider the respective prediction window consisting of a constant 11-hour observation window from our MVTS samples before training the model. We implement multiple metrics (see Section 5.4.1) to estimate the predictive/forecast skill of our models.

5.4.1 Metrics

We consider a 2 × 2 contingency table with the following elements: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Here, TP indicates the number of correctly predicted strong SEP events (positive class), while TN represents the number

of correctly predicted weak and non-SEP events (negative class). FP corresponds to the number of negative classes predicted as positives (false alarms), while FN corresponds to the number of positive class labels predicted as negatives (misses).

On a simple scale, statistical metrics such as precision (Eq. 5.1) and recall (Eq. 5.2) have been traditionally used to assess classifier performances. Precision for the positive class is used to evaluate the model's correct prediction with respect to the false alarms. Recall for the positive class characterizes the ability of the classifier to find all of the positive cases.

$$Precision = \frac{(TP)}{(TP + FP)} \tag{5.1}$$

$$Recall = \frac{(TP)}{(TP + FN)} \tag{5.2}$$

Focusing on the importance of positive classes, we consider the F_1-score for the positive class that can be estimated as the harmonic mean of precision and recall as shown in Equation 5.3. It ranges between 0 and 1 such that scores closer to 1 indicate the model to be better.

$$F_1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{5.3}$$

To understand the overall model performance, we use a 'weighted' average for the F_1-score (shown in Equations 5.4 and 5.5), which computes the score for each target class and uses sample weights that depend on the number of instances in that class while averaging. Here, i is the number of target classes in the data set, which is two in the present work.

$$F_{1_{weighted}} = \sum_{i=1}^{N} w_i \times F_{1_i} \tag{5.4}$$

$$w_i = \frac{\text{Number of samples in class } i}{\text{Total number of samples}} \tag{5.5}$$

We implement true skill statistics (TSS; Woodcock 1976; Dann 1985) to account for the false positive rate comparing the difference between the probability of detection and the probability of false detection as shown in Equation 5.6. TSS ranges from -1 to +1, where the latter indicates a perfect score. TSS $\leq$0 indicates it is worse than a random classification.

$$TSS = \frac{(TP \times TN) - (FP \times FN)}{(TP + FN) \times (FP + TN)} \tag{5.6}$$

Furthermore, we consider the Heidke skill score (HSS; Heidke 1926) that measures the improvement of the forecast over a random prediction as defined in Equation 5.7. HSS with 1 indicates perfect performance and 0 indicates no skill. A no-skill means the forecast is not better than a random binary forecast based on class distributions.

$$HSS = \frac{2 \times ((TP \times TN) - (FP \times FN))}{((TP + FN) \times (TN + FN)) + ((FP + TN) \times (FP + TP))} \tag{5.7}$$

The Gilbert Skill Score (GSS; Schaefer 1990) considers the number of hits due to chance, which is the frequency of an event multiplied by the total number of forecast events. This score formula is given by Equation 5.8. GSS ranges from -1/3 to 1. Here, 0 indicates no skill, while 1 is a perfect forecast.

$$GSS = \frac{TP - (\frac{(TP+FN)\times(TP+FP)}{TP+FP+TN+FN})}{(TP + FP + FN) - (\frac{(TP+FN)\times(TP+FP)}{TP+FP+TN+FN})} \tag{5.8}$$

Accounting for TN to assess the performance of a binary class problem is essential in our context. Hence, we also implement Matthew's correlation coefficient (MCC) as defined in Equation 5.9. MCC ranges from -1 to 1 where 0 indicates no skill and 1 shows perfect agreement between predicted and actual values. MCC is a robust metric that follows a comprehensive and balanced strategy for using the contingency table elements. That is, a high MCC score is obtained only when the predictions proportionately account for the size of both positive and negative samples in the data set. This is important because popular scores such as F_1 fall short to account for class imbalance.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TP + FP) \times (TN + FN)}} \tag{5.9}$$

5.4.2 Validation Set

As mentioned in Section 5.3.5, we use the second partition from our data set as a hold-out portion to estimate the optimal threshold for our classifiers. The classification threshold is the decision threshold that allows us to map the probabilistic output of a binary classifier to a binary category. In other words, it is a cut-off point used to assign a specific predicted class label for each sample. By default, the classification threshold in our models is 0.5. What that means is any prediction above 0.5 belongs to the positive class and that below 0.5 belongs to the negative class. However, 0.5 is not always optimal, especially when we

have an imbalanced data set. Therefore, we identify a reliable threshold for the classifier that better discriminates between the two class labels. The first step is to derive the prediction probabilities from the validation set and extract the best trade-off between true positive rate (TPR) and false positive rate (FPR). In general, any binary classification model predicts mean probabilities for each input sample belonging to the positive class, where the prediction score from the classifier is greater than a parametrized threshold. Then, a classification threshold (from 0.0 to 1.0) is used to assign a binary label to the predicted probabilities.

$$J = Sensitivity + Specificity - 1 \tag{5.10}$$

To find the optimal threshold that minimizes the difference between TPR and FPR, the Youden Index (J; Youden 1950) is used in the literature as defined in Equation 5.10. Here, sensitivity is the recall for the positive class and specificity is the recall for the negative class. J is the specialized version of TSS for the binary classification task. As a characteristic example, we show in Figure 5.5 the effect of "thresholding" on the performance of STSF by visualizing the variations in the skills due to changing thresholds. Here, the estimated optimal threshold of STSF for T_{30} on our data generation and sampling method is 0.28. We obtain optimal classification thresholds at all prediction windows for all the models and estimate the model skills. A comparison of the model performances is shown in Figure 5.6.

Computationally, the run time of SSC was the fastest, followed by STSF and then 1NN. In terms of performance, 1NN falls behind both models but maintains relatively close values for all the skill scores in each prediction window with a standard deviation (σ) of $\approx$0.04.

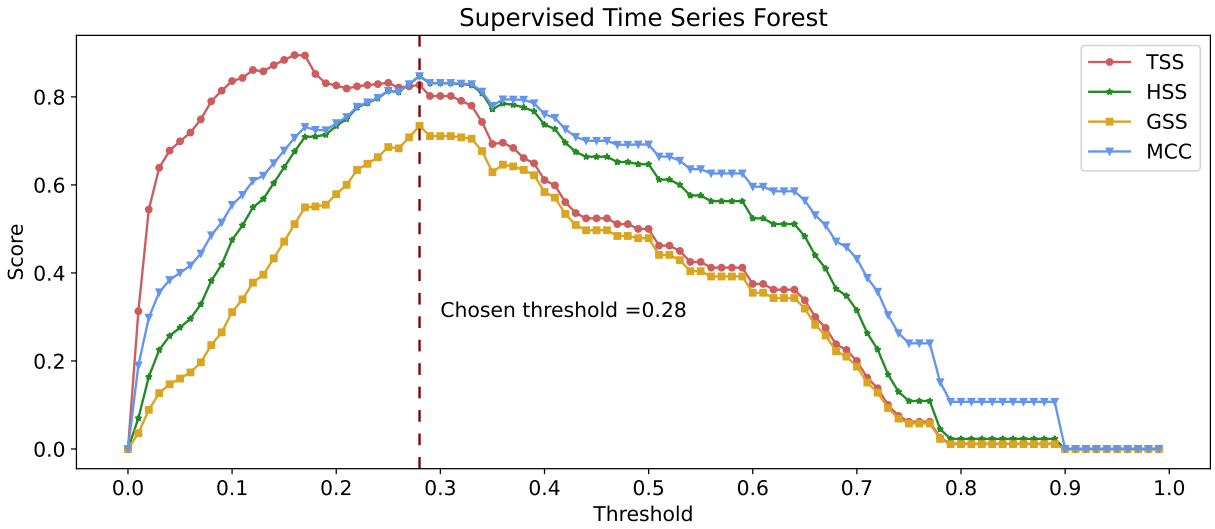


Figure 5.5 The variation in skills with respect to increasing the classification threshold for the STSF model on the validation/hold-out set.
Note: Shown are TSS, HSS, GSS and MCC scores at a prediction window of thirty minutes (T_{30}). The optimal threshold value for the model is inset in the plot.

On the other hand, SSC shows a similar trend in its skills with σ of $\approx$0.01 except for TSS ($\sigma \approx 0.07$). Although SSC was a little faster, the performance marginally lags on our data set compared to STSF. For example, the average TSS for SSC is $\approx$ 14% lower than that for STSF. The STSF model performs exceptionally well compared to the other two baseline classifiers at all parameters. Specifically, the high scores at T_{60} (for e.g., TSS = 0.769) provide satisfactory reliability to our methodology. Table 5.2 presents the contingency tables for all prediction windows corresponding to the optimal threshold of STSF on the validation set. The contingency tables for SSC and 1NN are provided in Tables 3 and 4, respectively, in the Appendix.

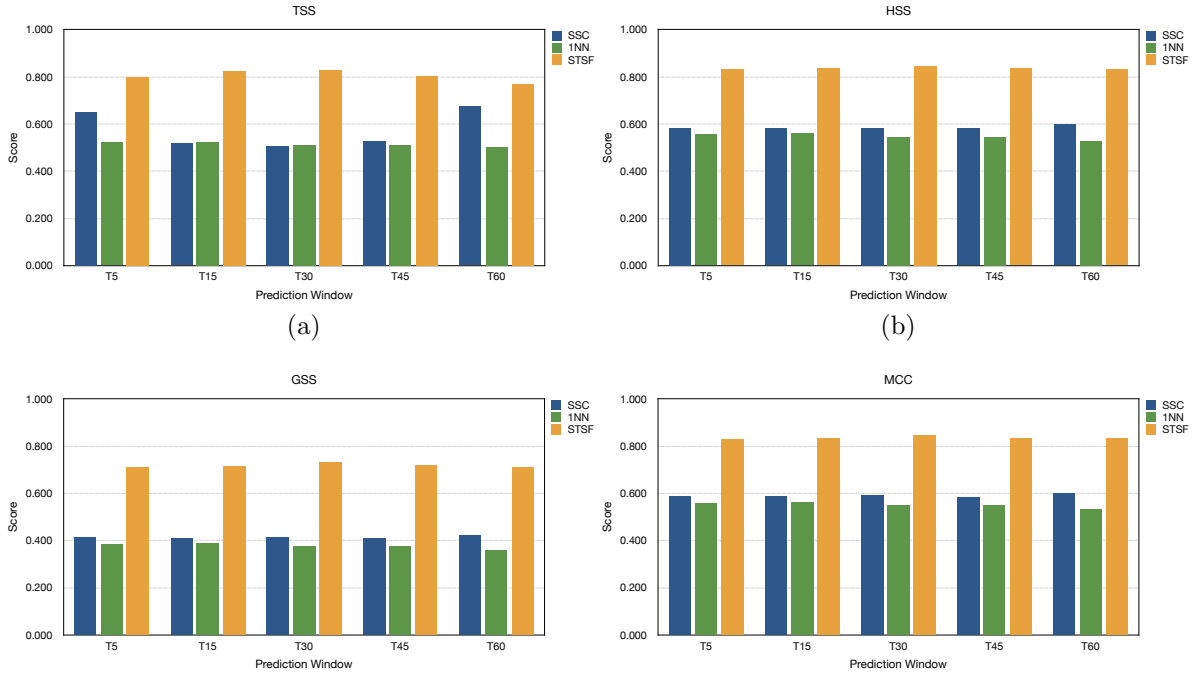


Figure 5.6 Comparison of the performances of SSC, STSF and 1NN classifiers.
Note: Shown here are with respect to variation in skills, namely (a) TSS, (b) HSS, (c) GSS and (d) MCC for prediction windows T_5, T_{15}, T_{30}, T_{45}, and T_{60} on the validation set.

Table 5.2: Contingency Tables for the STSF Model on the Validation Set.

		$\mathbf{T}_5$		$\mathbf{T}_{15}$		$\mathbf{T}_{30}$		$\mathbf{T}_{45}$		$\mathbf{T}_{60}$	
		Predicted		Predicted		Predicted		Predicted		Predicted	
		Strong	Weak	Strong	Weak	Strong	Weak	Strong	Weak	Strong	Weak
True	Strong	65	15	67	13	67	13	65	15	62	18
	Weak	9	885	11	883	9	885	8	886	5	889

Note: Truth tables for the STSF model for different prediction windows. The first column is a shared entry of true labels against predictive labels for each experiment. The elements indicate the number of predictions with respect to the actual occurrences in the validation set.
- Prediction windows are shown by a subscript to T. For example, T_5 indicates a prediction window of 5 min.
- Model name: STSF - Supervised Time Series Forest

5.4.3 Test Set

In our previous work (Rotti et al. 2024), we found that STSF is a highly efficient and optimal classifier to predict strong SEP events. STSF's approach to computing interval features in a supervised manner from three different time series representations by using robust statistics was a viable option for our problem specification. In this paper, we extended our modeling strategy by including non-SEP periods in our data set. Furthermore, we compared the performance of STSF and two baseline classifiers, namely SSC and 1NN. Based on the forecasting skill scores defined in Section 5.4.1, we used the validation set from our data partitions to estimate the model performances. This resulted in STSF as the best model under all comparative paramters. We use the test set to further assess the model's capabilities.

In Table 5.3, we show the contingency or the confusion matrix elements derived for the STSF classifier on the test set for all prediction windows (T_5 to T_{60}) considered in this work. The increasing prediction window shows a relatively consistent type II error (false negative rate) except for T_{30}. That is, the number of false positives is lower compared to false negatives. A possible reason for the situation here is that several weak SEP events have proton fluxes closely fluctuating in the vicinity of the SWPC threshold. Detecting such patterns at high precision becomes relevant to reducing FNs.

Table 5.3: Contingency Tables for the STSF model on the Test Set.

		$\mathbf{T}_5$		$\mathbf{T}_{15}$		$\mathbf{T}_{30}$		$\mathbf{T}_{45}$		$\mathbf{T}_{60}$	
		Predicted		Predicted		Predicted		Predicted		Predicted	
		Strong	Weak	Strong	Weak	Strong	Weak	Strong	Weak	Strong	Weak
True	Strong	71	13	73	11	75	9	73	11	72	12
	Weak	6	831	6	831	8	829	6	831	6	831

Note: Truth tables for the STSF model for different prediction windows. The first column is a shared entry of true labels against predictive labels for each experiment. The elements indicate the number of predictions with respect to the actual occurrences in the test set.
- Prediction windows are shown by a subscript to T. For example, T_5 indicates a prediction window of 5 min.
- Model name: STSF - Supervised Time Series Forest

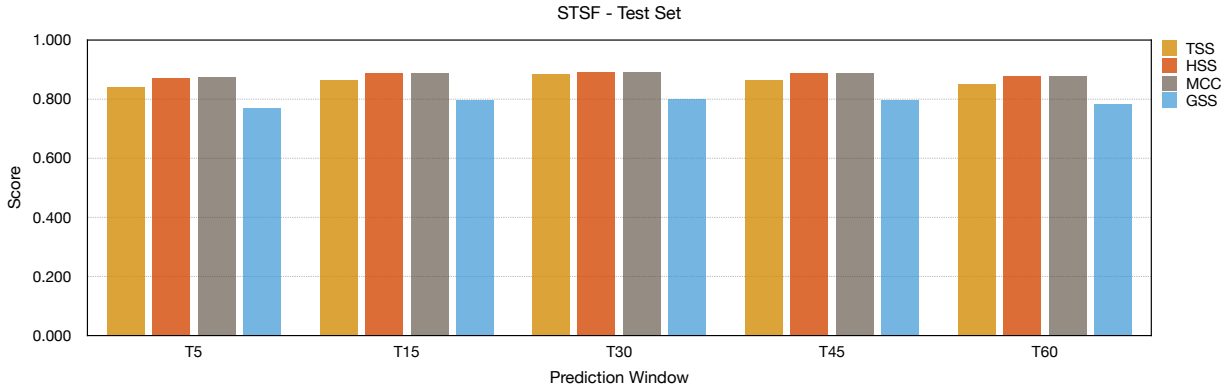


Figure 5.7 Comparison of skill scores of STSF model on the test set at prediction windows T_5, T_{15}, T_{30}, T_{45}, and T_{60}.

In Figure 5.7, we compare the skill scores of STSF with respect to the different prediction windows considered in this work. Overall, the average TSS is $\approx 0.86(\pm 0.02)$. Similarly, the HSS, MCC and GSS scores obtained are also high. There is a slight increment in the scores from T_5 to T_{30}, but reduces at T_{45} and T_{60}. This slight variation is due to subtle changes in the optimal classification threshold ($\approx 0.3(\pm 0.03)$) for STSF. The scores for T_{60} only have a marginal reduction ($<2\%$) compared to T_5. Furthermore, there is a decrease of $\approx 7\%(\pm 2\%)$ in skill scores compared to Rotti et al. (2024). This is expected due to extended lead times considered with a larger class imbalance in the data set. In summary, the classification and computational efficiencies we obtain for STSF are highly satisfactory and viable. This means that our work adequately provides a proof of concept of our modeling architecture for short-term predictive capabilities up to a 60-minute window and can be further transformed into implementation for NRT operations.

5.5 Conclusions

In this work, we approach the SEP event short-term prediction problem from a time series classification perspective. We primarily use the GSEP data set but extend our negative sample size by including "SEP-quiet" periods. That is, we consider ≥C6.0 flares that do not lead to SEP events observed near Earth. Our data set constitutes in-situ time series measurements from the NOAA-GOES missions for solar cycles 22 to 24. We utilize the long band (1–8Å) X-ray measurements from the XRS instrument and proton fluxes (P3, P5, P7) from the SEM instrument onboard GOES missions. Our data set has a total of 2893 samples, of which 244 are strong SEP events. We defined the target labels based on the SWPC 'S1' threshold for an SEP event. Positive labels are strong events crossing the 10 pfu threshold in the 10 MeV proton channel, and negative otherwise.

Our goal is to develop robust models that can successfully find discriminants between SEP and SEP-quiet patterns. In our recent study (Rotti et al. 2024), we found our methodology of using an ensemble of feature-based univariate time series classifiers to perform very well in classifying between strong and weak SEP events. Specifically, we found the supervised time series forest (STSF; Cabello et al. 2020) classifier to obtain optimal results. In the present work, we implemented our earlier methodology on an extended data set and compared the performances of three models, namely summary statistic classifier (SSC), one-nearest neighbor (1NN) and STSF. Furthermore, we consider a fixed observational window of 11 hours and use multiple prediction windows of prediction interval: 05-, 15-, 30-, 45-, and 60-min to leverage on high confidence for relatively low lead times.

In our model analysis phase, we used the Youden Index (J) to understand the trade-off between the true positive rate (TPR) and the false positive rate (FPR) at different classification thresholds. That is, we estimate an optimal threshold that provides a TPR with an acceptable FPR to make decisions using our classifiers. We considered the validation set to compare the performances of the models based on multiple skill scores, such as TSS, HSS, GSS and MCC. This resulted in STSF performing better compared to SSC and 1NN at all prediction windows. Furthermore, we use an unseen test set to obtain the classification report of the best model, including the contingency tables. For a 60-minute prediction window, we obtain the following scores for STSF: TSS = 0.850; HSS = 0.878; GSS = 0.783; MCC = 0.879. In summary, our results are promising and provide a gateway for our model architecture to be used in an operational context to offer short-term predictions on the arrival of strong SEP events up to 60 minutes. The data set and coding methodology of our model implementation have been made available on GitHub[5] under a Creative Commons license.

The applications we foresee for our predictive system to be useful to space agencies across the globe are two-fold. First, the system can serve as a convenient tool for forecasters, for example, at NASA's Space Radiation Analysis Group (SRAG) and NOAA-SWPC in assisting them in broadcasting highly accurate last-minute warnings to astronauts on the surface of the Moon or during extravehicular activities when outside the Earth's magnetosphere. Second, for future interplanetary missions, such as the missions to Mars that NASA is planning, the spacecraft will spend most of its time outside of the Sun-Earth line. Therefore, SEP predictions that apply to the Earth-Moon system are of limited value when the spacecraft-

[5]GitHub repository: https://github.com/sumanth-ra23/SEP-Predictions

Sun makes a large angle with the Earth-Sun direction. A crewed spacecraft on a journey to Mars would only have to carry a scaled-down version of the GOES-like instruments to produce short-term warnings of intense solar particle storms. Hence, we expect that our system can be a valuable addition to the tools that space agencies will need to protect astronauts on upcoming lunar and Mars missions.

CHAPTER 6

Conclusion and Future Path

The Sun is a sophisticated object in space exhibiting highly variable and complex magnetic activity dominating its neighborhood. Although driven by processes inside the Sun, much of the most fascinating solar activity occurs in the upper layers that influence the heliosphere. Events such as flares and coronal mass ejections (CMEs) increase with the increasing activity during a solar cycle and are often associated with solar radiation storms in the form of solar energetic particles (SEPs). These events majorly constitute intense proton beams in addition to electrons and alpha particles up to iron ions from the Sun. The combination of charge and mass of particles determines how quickly they lose energy when interacting with matter (Guetersloh & Zapp 2010). For instance, at equal energies, an electron penetrates farther into aluminum than a proton. Furthermore, the combination of mass, charge, and energy of the particles influences how much radiation exposure the systems and the crew may encounter depending on where the spacecraft is located beyond the Earth's protective magnetosphere. An intense SEP can arrive at Earth and be complete within hours or last for more than a week, during which bursts of radiation last a few hours. With significant advancements in computational powers, there are several opportunities to reduce the hazardous consequences of SEPs by robust and reliable forecasting. Methods involving artificial intelligence and machine learning (ML) have demonstrated promise in forecasting SEP events. A special focus on short-term predictions is relevant to facilitating operational requirements for space missions outside the Earth's magnetosphere.

In this regard, the objective of this dissertation can broadly be classified into two categories: (1) creating a large ML-ready benchmark data set of SEP events (chapters 2 & 3) and (2) developing a novel approach for robust short-term predictions (chapters 4 & 5). The Geostationary Solar Energetic Particle Events data set (GSEP) has been created from the particle and X-ray fluxes of GOES missions from 1986 to 2017. It consists of 433 events, of which 244 cross the 10 pfu threshold in the 10 MeV GOES proton channel. We rely on several sources, such as existing event catalogs of solar flares, CMEs, radio bursts, and SEP events, and relevant observational data from the SOHO and SDO missions to associate each SEP event in our data set with a parent solar eruption where possible. The GSEP data set serves as a one-stop database for heliophysics researchers to study SEP events using an extensive, long-term data archive publicly available on the Harvard Dataverse repository at: https://doi.org/10.3847/1538-4365/ac87ac.

We conduct a statistical analysis of SEP event parameters along with the parent solar eruptions, estimating the relationships of flare magnitude, rise time, flaring locations, speed, and width of CMEs, and the temporal variations of sources with SEP events. Concerning SEP peak fluxes and event fluences, we find a good correlation with CME parameters and moderate correlations are observed with X-ray flare intensities. Most (80%) of the SEP events in the GSEP data set are associated with $\geq$M2.0 flares. Also, most (66%) CMEs associated with SEP events are halo in angular extents and fast with a median speed of 1200 km.s^{-1}. The distribution of event origins shows an increase in proportionality with source eruptions occurring on the visible disk up to the western hemisphere. Exceptions are noticed for $\geq$X1.0

flares occurring over the eastern and beyond-limb locations due to poor magnetic connectivity for the particles to reach Earth. Toward predictive efforts of SEP events, we consider an ensemble strategy to classify strong SEP events (surpassing the SWPC 'S1' radiation storm threshold) using time series classifiers. After that the entirely data-driven methodology is further evaluated on an extended data set of "SEP-quiet" periods, introducing a natural class imbalance (≈11:1) between non-events and events. This methodology performs exceptionally well for an observation window of up to 60 minutes.

Because of the severity of SEPs, warning systems of high confidence are a great necessity, and we envision them as the final product of this research. Strategic operations of human space missions, as prioritized by multiple government and industrial entities, will require such products and is an immediate application we foresee. The immediate goal of this research is to test the model performance on NASA SRAG's SEPVAL challenge data set. This will ensure the achievement of a well-conditioned SEP event forecasting pipeline. Going further, we envisage keeping the GSEP data set updated by including events from solar cycle 25. That is, we will integrate information on flares, CMEs, and radio bursts with respect to SEP events from 2019 using data from multiple missions into the GSEP data set. This involves meticulously identifying strong and weak SEP events and associating possible solar sources for each event. Furthermore, we will add event-associated metadata and extend solar source parameters relevant to SEPs. Finally, we want to test our methodology for predicting E≥100 MeV SEP events and produce short-term forecasts. We anticipate that our forecasting system can sit on top of the current methods.

Appendices

A Weak Events in PSEP

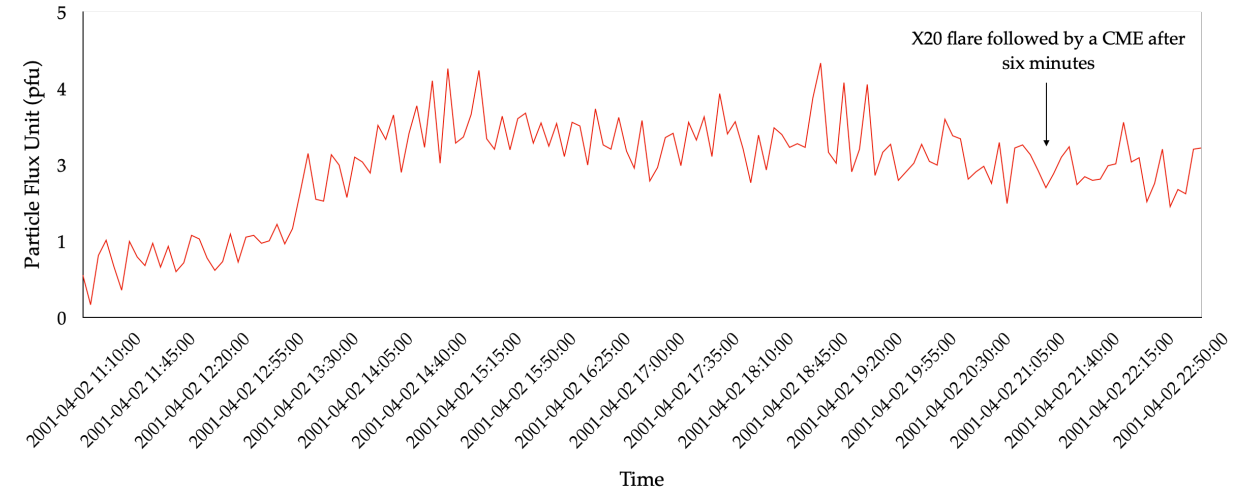
In Table 1 below, the nine events that are reported in PSEP with peak proton fluxes below 10 MeV are listed. The index refers to the event number in the GSEP list. The next column indicates the event onset followed by event maximum time stamp as reported in NOAA-SEP and GSEP. The last three columns show the peak proton fluxes (in pfu) from the PSEP, NOAA-SEP, and GSEP lists, respectively.

Table 1: SEP Events Reported in PSEP with Peak Fluxes <10 pfu but Observed to be >10 pfu in GSEP.

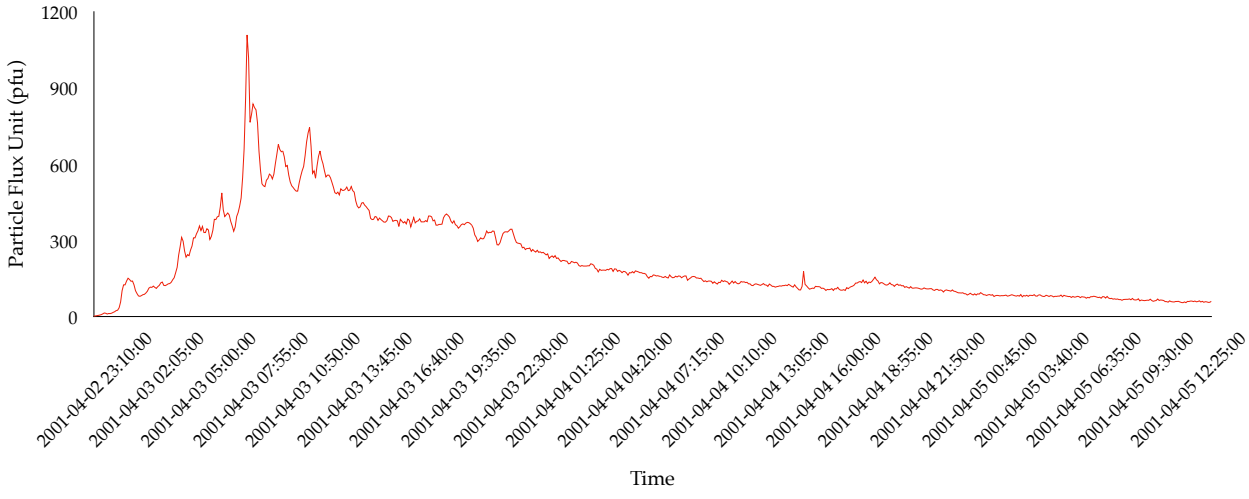
sep_index	event_start_time	noaa_max_time	gsep_max_time	ppf_gt10MeV (PSEP)	noaa_pf10MeV (NOAA-SEP)	gsep_pf_gt10MeV (GSEP)
gsep_034	1989-06-18 15:00:00	1989-06-18 19:10:00	1989-06-18 20:25:00	9.24	18	10.8
gsep_058	1989-11-15 07:05:00	1989-11-15 09:10:00	1989-11-15 09:05:00	4.85	71	38.3
gsep_062	1990-03-28 13:50:00	1990-03-29 10:05:00	1990-03-29 10:05:00	2.14	16	15.9
gsep_086	1991-03-31 21:25:00	–	1991-04-03 09:10:00	3.04	–	25.5
gsep_117	1992-03-16 04:35:00	1992-03-16 08:40:00	1992-03-16 09:00:00	9.11	10	10.4
gsep_130	1993-03-06 21:15:00	–	1993-03-07 07:10:00	9.73	–	10.8
gsep_195	2001-09-15 12:20:00	2001-09-15 14:55:00	2001-09-15 14:55:00	9.49	12	11.6
gsep_200	2001-10-19 17:45:00	2001-10-19 22:30:00	2001-10-19 22:30:00	9.53	12	11.7
gsep_295	2011-10-22 12:15:00	–	2011-10-23 15:35:00	7.88	–	13.1

B Entangled SEP Events

In Figure 1, the time profiles of two SEP events (182 and 183) in the GSEP list are shown. The latter is a very large event, while the former is apparently a weak event with peak proton flux of 4 pfu at >10 MeV. On 2001 April 2, the first SEP event appears to be associated with an X1.1 flare at 10:58:00, while an X20 flare at 21:32:00 leads to the second SEP event. The flaring active region (9393) is positioned at the western hemisphere of the Sun while erupting. Both the SEP events are associated with distinct CMEs, detected after the respective flares.



(a)



(b)

Figure 1 Time series profiles of GSEP event 182 in (a) and event 183 in (b).
Note: The first SEP event was due to an X1.1 flare and a CME (2001 April 2 10:58:00), while the second event was triggered due to an X20 flare (2001 April 2 21:32:00) and a CME erupted after 6 minutes. Both events originated from NOAA AR 9393.

C GOES Data

During the data collection and preparation stage, we also obtained a list of GOES primary and secondary missions from SWPC. The reported measurements of peak fluxes are not consistent over a few SEP events across the existing catalogs. Hence, we visually inspected all the GOES proton data from 1986 to 2017 to understand the differences and also the variations in measurements between parallel GOES missions. We compiled a list of the GOES missions from -5 to -15 during the last three solar cycles from which we considered the proton flux data for the present work. The list is presented here in Table 2. Here, 'G" stands for GOES and the corresponding number indicates the mission. For example, "G15" represents the GOES-15. The underlined text indicates a primary satellite from the GOES mission as identified by NOAA.

Table 2: List of GOES Missions Considered in the Present Work.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1986	G06	G05	G05	G06	G05	G06	G06	G06	G06	G06	G06	G06
1987	G06	G05	G06	G06	G06	G06	G06	G06	G06	G06	G06	G06
1988	G06	G07	G07	G07	G07	G07	G07	G06	G07	G07	G07	G07
1989	G07	G07	G07	G07	G06	G07	G07	G07	G07	G07	G07	G07
1990	G07	G07	G07	G07	G07	G06	G07	G07	G07	G07	G07	G07
1991	G07	G07	G07	G07	G07	G07	G07	G07	G07	G07	G07	G07
1992	G07	G07	G07	G07	G06	G07	G07	G07	G07	G07	G07	G07
1993	G07	G07	G07	G07	G07	G07	G07	G07	G07	G07	G07	G07

Table 2—*Continued*

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1994	G07	G07	G07	G07	G07	G07	G07	G07	G07	G07	G07	G07
1995	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08
1996	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08
1997	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08
1998	G08	G08	G08	G08	G08	G08	G08	G10	G10	G08	G08	G10
1999	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08
2000	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08	G08
2001	G08	G08	G08	G08	G08	G08	G10	G10	G10	G10	G10	G10
2002	G10	G10	G10	G08	G08	G08	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	G15	G15	G15	G15	G15	G13	G13	G15	G15
2012	G15	G15	G15	G13	G15	G15	G15	G15	G15	G13	G15	G15
2013	G15	G15	G15	G13	G15	G15	G15	G15	G13	G15	G13	G13
2014	G13	G13	G13	G15	G15	G15	G15	G15	G15	G15	G15	G15
2015	G13	G13	G15	G15	G15	G13	G13	G13	G13	G15	G13	G13
2016	G15	G15	G15	G15	G15	G15	G15	G15	G15	G15	G15	G15
2017	G15	G15	G15	G15	G15	G15	G13	G15	G15	G15	G15	G15

D Threshold Analysis

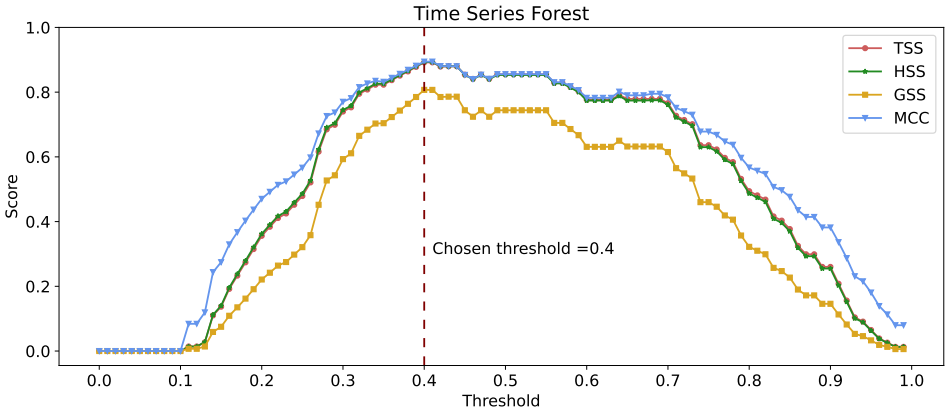
The classification threshold is the decision threshold that allows us to map the probabilistic output of a classifier to a binary category. In other words, it is a cut-off point used to assign a specific predicted class label for each sample. In our model analysis phase, we used the Receiver Operating Characteristic (ROC) curve, which is a diagnostic tool used to evaluate a set of probabilistic predictions made by a model. The ROC curve is useful for understanding the trade-off between true positive rate (TPR) and false positive rate (FPR) at different thresholds.

By default, the classification threshold in our models is 0.5. Any prediction above 0.5 belongs to the positive class, and that below 0.5 belongs to the negative class. However, 0.5 is not always optimal, and we identify a reliable threshold for the classifier that better splits between the two target classes. That is, we choose the threshold that provides a TPR with an acceptable FPR to make decisions using the classifier.

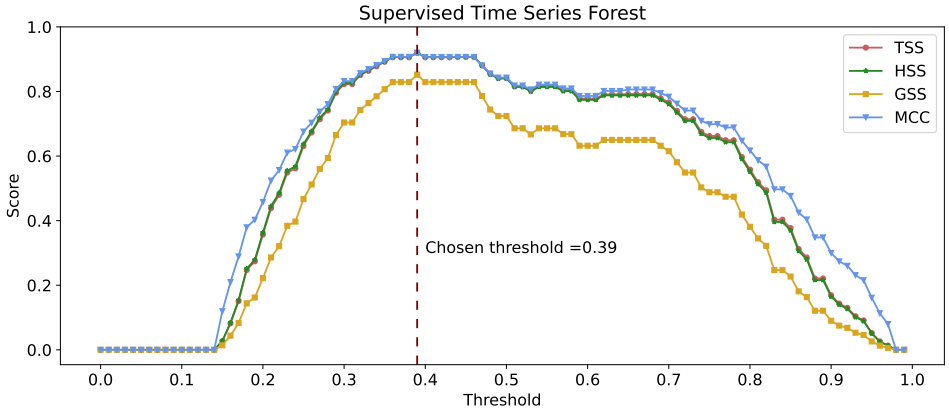
In the present work, we find the optimal threshold using the Youden Index (J; Youden 1950) defined in Equation 4.8. Here, sensitivity is TPR and specificity is (1-FPR). Therefore, by estimating TPR-FPR for each threshold, we obtain a maximum J as a cut-off point that optimizes classification between the two classes. The obtained best J-value gives us the optimal threshold of the classifier.

Furthermore, we demonstrate the effect of "thresholding" on the model performances by visualizing the variations in the skills due to changing thresholds. For this purpose, we used advanced metrics discussed in Section 4.4. We define a set of thresholds (from 0.0 to 1.0) and

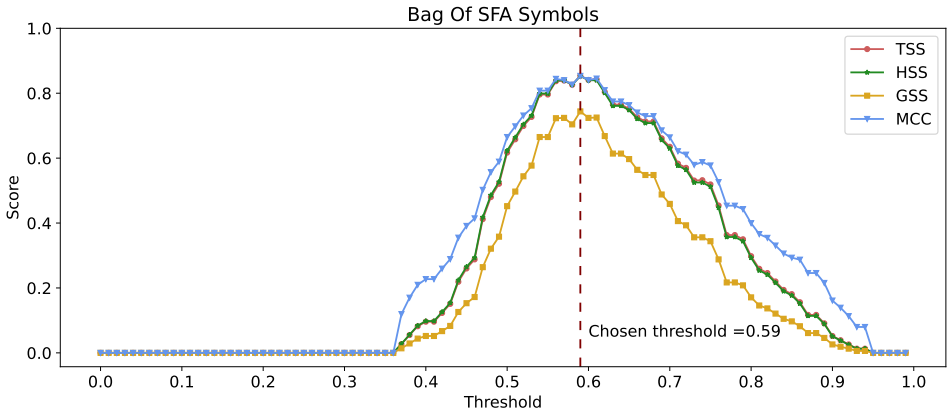
then evaluate predicted probabilities under each threshold. That is, we transform/binarize the predicted probabilities into labels for the respective threshold and estimate the skill scores in order to find and select the best threshold value. Figure 2 shows the influence of variation in the classification threshold for each model. The TSF (Fig. 2a) and STSF (Fig. 2b) have a very close optimal threshold that is less than 50%. The BOSS model (Fig. 2c) shows optimal performance at a threshold of $\approx 60\%$.



(a)



(b)



(c)

Figure 2 Variation in model skills on the test set.
Note: Shown here are TSS, HSS, GSS and MCC with respect to increasing the classification threshold for (a) TSF, (b) STSF, and (c) BOSS models. The optimal threshold value for each model is inset in the plot.

E Effect of Randomness

Of the three models considered in this work, TSF considers random intervals from the input time series and implements a random forest to fit the feature vectors and make predictions. Although STSF largely overcomes the randomization of interval selection, it consists of a tree-based random forest structure at its core. Because TSF and STSF models have random components in their architecture, we run both models multiple (10) times and find the variations in their respective optimal threshold values as shown in Figure 3. The median (mean) value for TSF is 0.412 (0.415), and for STSF it is 0.407 (0.412). Comparing the above values with the chosen thresholds (as shown in Figure 2) for the respective classifiers, we are confident of our model predictions and their capabilities to be further transformed for operational standards.

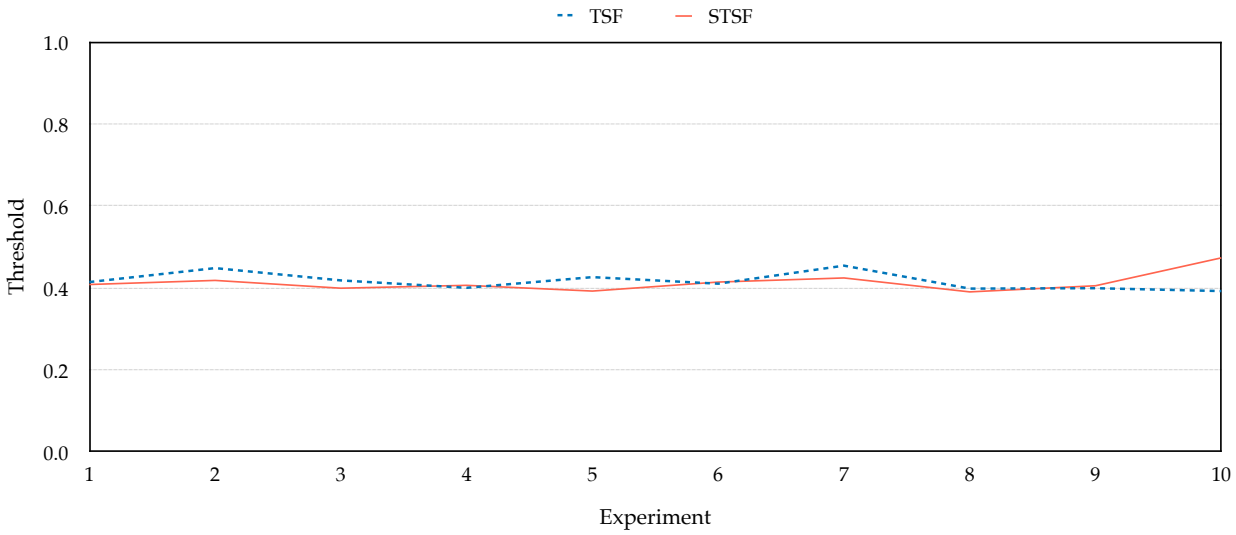


Figure 3 Experimantal evaluation of the impact of random components in the TSF and STSF model structures on optimal classification threshold.
Note: Here, the y-axis shows the thresholds (in the range of 0.0 to 1.0) and the x-axis shows the number of experiments. The median (mean) threshold for TSF is 0.412 (0.415) and for STSF it is 0.407 (0.412).

F Contingency Tables

The contingency table elements for SSC and 1NN on the validation set are provided in Tables 3 and 4, respectively, for each of our experiments on changing latency periods. The values from these tables are used to estimate the skill scores that are defined in Section 5.4.1.

Table 3: Contingency Tables for the SSC Model on the Test Set.

		$\mathbf{T}_5$		$\mathbf{T}_{15}$		$\mathbf{T}_{30}$		$\mathbf{T}_{45}$		$\mathbf{T}_{60}$	
		Predicted		Predicted		Predicted		Predicted		Predicted	
		Strong	Weak	Strong	Weak	Strong	Weak	Strong	Weak	Strong	Weak
True	Strong	56	24	43	37	42	38	44	36	58	22
	Weak	44	850	18	876	15	879	20	874	45	792

Note: Truth tables for the SSC model for different prediction windows. The first column is a shared entry of true labels against predictive labels for each experiment. The elements indicate the number of predictions with respect to the actual occurrences in the validation set.
- Prediction windows are shown by a subscript to T. For example, T_5 indicates prediction window of 5 min.
- Model name: SSC - Summary Statistic Classifier

Table 4: Contingency Tables for the 1NN Model on the Validation Set.

		$\mathbf{T}_5$		$\mathbf{T}_{15}$		$\mathbf{T}_{30}$		$\mathbf{T}_{45}$		$\mathbf{T}_{60}$	
		Predicted		Predicted		Predicted		Predicted		Predicted	
		Strong	Weak	Strong	Weak	Strong	Weak	Strong	Weak	Strong	Weak
True	Strong	44	36	44	36	43	37	43	37	43	37
	Weak	25	869	24	870	25	869	25	869	29	865

Note: Truth tables for the 1NN model for different prediction windows. The first column is a shared entry of true labels against predictive labels for each experiment. The elements indicate the number of predictions with respect to the actual occurrences in the validation set.
- Prediction windows are shown by a subscript to T. For example, T_5 indicates a prediction window of 5 min.
- Model name: 1NN - One-nearest Neighbor

REFERENCES

Acuña, M. H., Ogilvie, K. W., Baker, D. N., et al. 1995, SSRv, 71, 5

Ahmadzadeh, A., Hostetter, M., Aydin, B., et al. 2019, in 2019 IEEE Tnt. Conf. on Big Data (Big Data) (Piscataway, NJ: IEEE), 1423

Alberti, T., Laurenza, M., Cliver, E., et al. 2017, ApJ, 838, 59

Ali, A. et al. 2024, ApJS, 270, 15

Aminalragia-Giamini, S., Raptis, S., Anastasiadis, A., et al. 2021, JSWSC, 11, 59

Anastasiadis, A., Lario, D., Papaioannou, A., et al. 2019, RSPTA, 377, 20180100

Anastasiadis, A., Papaioannou, A., Sandberg, I., et al. 2017, SoPh, 292, 1

Angryk, R. A. et al. 2020, NatSD, 7, 227

Aparna, V., & Martens, P. C. 2020, ApJ, 897, 68

Aran, A., Sanahuja, B., & Lario, D. 2006, ASR, 37, 1240

Arbib, M. A. 2003, The Handbook of Brain Theory and Neural Networks (Cambridge, MA: MIT Press)

Arge, C., & Pizzo, V. 2000, JGR: Space Physics, 105, 10465

Aschwanden, M. J. 2005, ApJ, 634, L193

Bagnall, A., Lines, J., Bostrom, A., Large, J., & Keogh, E. 2017, Data Min. Knowl. Discov., 31, 606

Bain, H., Steenburgh, R., Onsager, T., & Stitely, E. 2021, SpWea, 19, e2020SW002670

Balch, C. C. 2008, SpWea, 6, S01001

Bame, S., Asbridge, J., Hundhausen, A., & Strong, I. 1968, JGR, 73, 5761

Barnhart, R. K. 1995, The H.W. Wilson Company

Barth, J. L., Dyer, C., & Stassinopoulos, E. 2003, IEEE Transactions on Nuclear Science, 50, 466

Beck, P., Latocha, M., Rollet, S., & Stehno, G. 2005, ASR, 36, 1627

Belov, A., Garcia, H., Kurt, V., et al. 2005, SoPh, 229, 135

Benton, E. R., & Benton, E. 2001, NIM-B, 184, 255

Bobra, M. G., & Couvidat, S. 2015, ApJ, 798, 135

Bornmann, P. L., Speich, D., Hirman, J., et al. 1996, Proc. SPIE, 2812, 291

Borovikov, D., Sokolov, I. V., & Tóth, G. 2015, JCP, 297, 599

Boubrahimi, S. F., Aydin, B., Martens, P., & Angryk, R. 2017, in 2017 IEEE Int. Conf. on Big Data (Big Data) (Piscataway, NJ: IEEE), 2533–2542

Bougeret, J. L. et al. 1995, SSRv, 71, 231

Bourdarie, S., & Xapsos, M. 2008, IEEE Transactions on Nuclear Science, 55, 1810

Breiman, L. 2001, MachL, 45, 5

Brueckner, G. E. et al. 1995, SoPh, 162, 357

Bruno, A. 2017, SpWea, 15, 1191

Bruno, A. et al. 2018, ApJ, 862, 97

Bruno, A., & Richardson, I. G. 2021, SoPh, 296, 36

Bryant, D. A., Cline, T., Desai, U., & McDonald, F. B. 1962, JGR, 67, 4983

Cabello, N., Naghizade, E., Qi, J., & Kulik, L. 2020, in 2020 IEEE Int. Conf. on Data Mining

(ICDM) (Piscataway, NJ: IEEE), 948

Campa, A. et al. 2009, Radiat. Res., 171, 438

Camporeale, E. 2019, SpWea, 17, 1166

Can, R., Kocaman, S., & Gokceoglu, C. 2021, Appl. Sci., 11

Cane, H. 1995, NuPhS, 39, 35

Cane, H., & Lario, D. 2006, SSRv, 123, 45

Cane, H., Reames, D., & Von Rosenvinge, T. 1988, JGR: Space Physics, 93, 9555

Cane, H., Richardson, I., & Von Rosenvinge, T. 2010, JGR: Space Physics, 115

Cane, H. V., McGuire, R. E., & von Rosenvinge, T. T. 1986, ApJ, 301, 448

Cane, H. v., Von Rosenvinge, T., Cohen, C., & Mewaldt, R. 2003, GRL, 30

Cassisi, C., Montalto, P., Aliotta, M., et al. 2012, Advances in Data Mining Knowledge Discovery and Applications (London: IntechOpen), 71

Chancellor, J. C., Scott, G. B., & Sutton, J. P. 2014, Life, 4, 491

Chen, P. 2011, LRSP, 8, 1

Chen, T., & Guestrin, C. 2016, in Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, KDD '16 (New York: Association for Computing Machinery), 785–794

Cliver, E., Forrest, D., Cane, H., et al. 1989, ApJ, 343, 953

Cliver, E. W., & D'Huys, E. 2018, ApJ, 864, 48

Cliver, E. W., Kahler, S. W., & Reames, D. V. 2004, ApJ, 605, 902

Cliver, E. W., & Ling, A. G. 2009, ApJ, 690, 598

Cliver, E. W., Schrijver, C. J., Shibata, K., & Usoskin, I. G. 2022, LRSP, 19, 2

Cohen, C. M. S. 2006, GMS, 165, 275

Cortes, C., & Vapnik, V. 1995, MachL, 20, 273

Cox, D. R. 1958, J. R. Stat. Soc. B, 20, 215

Crosby, N. et al. 2015, SpWea, 13, 406

Cucinotta, F., Manuel, F., Jones, J., et al. 2001, Radiat. Res., 156, 460

Cucinotta, F. A., Alp, M., Sulzman, F. M., & Wang, M. 2014, LSSR, 2, 54

Cucinotta, F. A., Hu, S., Schwadron, N. A., et al. 2010, SpWea, 8

Daan, H. 1985, Statistics and Decision Making in the Atmospheric Sciences, 379

Deng, H., Runger, G., Tuv, E., & Vladimir, M. 2013, Inf. Sci., 239, 142

Desai, M., & Giacalone, J. 2016, LRSP, 13, 1

Dierckxsens, M., Tziotziou, K., Dalla, S., et al. 2015, SoPh, 290, 841

Engell, A., Falconer, D., Schuh, M., et al. 2017, SpWea, 15, 1321

Falconer, D., Barghouty, A. F., Khazanov, I., & Moore, R. 2011, SpWea, 9

Faouzi, J. 2022, Machine Learning (Emerging Trends and Applications) (London: Proud Pen), https://inria.hal.science/hal-03558165

Fawagreh, K., Gaber, M. M., & Elyan, E. 2014, Syst. Sci. Control Eng., 2, 602

Feynman, J., & Hundhausen, A. J. 1994, JGR: Space Physics, 99, 8451

Fulcher, B. D., & Jones, N. S. 2014, IEEE Transactions on Knowledge and Data Engineering, 26, 3026

Garcia, H. A. 1994, SoPh, 154, 275

Gerontidou, M., Mavromichalaki, H., Belov, A., & Kurt, V. 2009, ASR, 43, 687

Ghareb, A. S., Bakar, A. A., & Hamdan, A. R. 2016, Expert Syst. Appl., 49, 31

Gopalswamy, N. 2003, GRL, 30

Gopalswamy, N. 2012, AIP Conf. Proc., 1500, 14, .Space Weather: The Space Radiation Environment: 11th Annual Int. Astrophysics Conf. (Melville, NY: AIP)

Gopalswamy, N., Lara, A., Yashiro, S., et al. 2001, JGR: Space Physics, 106, 29207

Gopalswamy, N., Mäkelä, P., & Yashiro, S. 2019, SunGe, 14, 111

Gopalswamy, N., Mäkelä, P., Yashiro, S., et al. 2015, JPhCS, 642, 012012

——. 2017, JPhCS, 900, 012009

Gopalswamy, N., Xie, H., Akiyama, S., et al. 2014, EP&S, 66, 104

Gopalswamy, N., Yashiro, S., Krucker, S., & Howard, R. 2004a, Proc. of the International Astronomical Union, 2004, 367

Gopalswamy, N., Yashiro, S., Krucker, S., et al. 2004b, JGR: Space Physics, 109

Gopalswamy, N., Yashiro, S., Lara, A., et al. 2003a, GRL, 30

Gopalswamy, N., Yashiro, S., Michałek, G., Kaiser, M., Howard, R., Reames, D., Leske, R., & Von Rosenvinge, T. 2002, ApJ, 572, L103

Gopalswamy, N., Yashiro, S., Michalek, G., et al. 2003b, in AIP Conf. Proc., 679, 608

——. 2009, Earth, Moon, and Planets, 104, 295

Gopalswamy, N., Yashiro, S., Xie, H., et al. 2008, ApJ, 674, 560

Grubb, R. N. 1975, The SMS/GOES Space Environment Monitor Subsystem, NOAA Technical Memorandum

Guetersloh, S., & Zapp, N. 2010, Heliophysics: Space Storms and Radiation: Causes and Effects (London: Cambridge Univ. Press), 359, ed. Schrijver, Carolus J. and Siscoe, George L.

Guyon, I., & Elisseeff, A. 2003, JMLR, 3, 1157

Hale, G. E. 1908, ApJ, 28, 315

Hansen, L. K., & Salamon, P. 1990, ITPAM, 12, 993

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. 2009, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Vol. 2 (Berlin: Springer)

Heidke, P. 1926, GeAnA, 8, 301

Hu, J., Li, G., Ao, X., et al. 2017, JGR: Space Physics, 122, 10

Huttunen-Heikinmaa, K., & Valtonen, E. 2009, AnGeo, 27, 767

Jackman, C. H., & McPeters, R. D. 1987, PhST, T18, 309

Janitza, S., Tutz, G., & Boulesteix, A.-L. 2016, Comput. Stat. Data Anal., 96, 57

Ji, A., Arya, A., Kempton, D., et al. 2021, in 2021 IEEE Third Int. Conf. on Cognitive Machine Intelligence (CogMI) (Piscataway, NJ: IEEE)

Ji, A., Aydin, B., Georgoulis, M. K., & Angryk, R. 2020, in 2020 IEEE Int. Conf. on Big Data (Big Data) (Piscataway, NJ: IEEE), 4218–4225

Jiggens, P., Clavie, C., Evans, H., et al. 2019, SpWea, 17, 99

Kahler, S. 1994, ApJ, 428, 837

Kahler, S. W. 1992, ARA&A, 30, 113

Kahler, S. W. 2005, ApJ, 628, 1014

Kahler, S. W., Cliver, E. W., & Ling, A. G. 2007, JASTP, 69, 43

Kahler, S. W., White, S. M., & Ling, A. G. 2017, JSWSC, 7, A27

Kallenrode, M. 2003, JPhG, 29, 965

Kasapis, S., Zhao, L., Chen, Y., et al. 2022, SpWea, 20, e2021SW002842

Kennedy, A. R. 2014, LSSR, 1, 10

Keogh, E., Chakrabarti, K., Pazzani, M., & Mehrotra, S. 2001, Knowledge and information Systems, 3, 263

Khalid, S., Khalil, T., & Nasreen, S. 2014, in 2014 Science and Information Conf. (Piscataway, NJ: IEEE), 372

Kihara, K., Huang, Y., Nishimura, N., et al. 2020, ApJ, 900, 75

Kim, M.-H. Y., De Angelis, G., & Cucinotta, F. A. 2011, Acta Astronautica, 68, 747

Klein, K.-L., & Dalla, S. 2017, SSRv, 212, 1107

Klein, K.-L., & Posner, A. 2005, A&A, 438, 1029

Klein, K.-L., & Trottet, G. 2001, SSRv, 95, 215

Kocharov, L., & Torsti, J. 2002, SoPh, 207, 149

Korsós, M. B., Erdélyi, R., Liu, J., & Morgan, H. 2021, FrASS, 7

Kozarev, K., Nedal, M., Miteva, R., Dechev, M., & Zucca, P. 2022, FrASS, 9, 801429

Kress, B. T., Rodriguez, J. V., & Onsager, T. G. 2020, in The GOES-R Series (Amsterdam: Elsevier), 243

Kurt, V., Belov, A., Mavromichalaki, H., & Gerontidou, M. 2004, AnGeo, 22, 2255

Laurenza, M., Cliver, E., Hewitt, J., et al. 2009, SpWea, 7

Lavasa, E., Giannopoulos, G., Papaioannou, A., et al. 2021, SoPh, 296, 107

Lemen, J. R. et al. 2012, SoPh, 275, 17

Li, G. et al. 2021, ApJ, 919, 146

Lin, R. P., & Hudson, H. S. 1976, SoPh, 50, 153

Linker, J. A., Caplan, R. M., Schwadron, N., et al. 2019, JPhCS, 1225, 012007

Low, B. 1996, SoPh, 167, 217

Luhmann, J., Ledvina, S., Krauss-Varban, D., et al. 2007, ASR, 40, 295

Löning, M., Király, F., Bagnall, T., et al. 2022, sktime: v0.13.4, doi: 10.5281/zenodo.7117735

Mäkelä, P., Gopalswamy, N., Akiyama, S., et al. 2015, ApJ, 806, 13

Manning, C. D., Raghavan, P., & Schütze, H. 2008, Introduction to Information Retrieval (Cambridge Univ. Press)

Marqué, C., Posner, A., & Klein, K.-L. 2006, ApJ, 642, 1222

Marsh, M., Dalla, S., Dierckxsens, M., et al. 2015, SpWea, 13, 386

Martens, P. C., & Angryk, R. A. 2017, Proc. of the International Astronomical Union, 13, 344

Matthews, B. W. 1975, Biochimica et Biophysica Acta, 405, 442

Mewaldt, R. A., Cohen, C. M. S., & Mason, G. M. 2006, GMS, 165, 115

Milligan, R. O., & Ireland, J. 2018, SoPh, 293, 18

Miteva, R., Samwel, S. W., & Costa-Duarte, M. V. 2018, SoPh, 293, 1

Mosteller, F., & Tukey, J. W. 1968, in Handbook of Social Psychology, Vol. 2 (Reading, MA: Addison-Wesley), ed. G. Lindzey & E. Aronson, 80

Murphy, A. H. 1973, JApMC, 12, 595

Nelson, G. A. 2016, Radiat. Res., 185, 349

Nicholson, D. R., Goldman, M. V., Hoyng, P., & Weatherall, J. C. 1978, ApJ, 223, 605

Núñez, M. 2011, SpWea, 9, S07003

——. 2015, SpWea, 13, 807

Núñez, M., Reyes-Santiago, P. J., & Malandraki, O. E. 2017, SpWea, 15, 861

Onsager, T., Grubb, R., Kunches, J., et al. 1996, Proc. SPIE, 2812, 281

Osman, A. I. A., Ahmed, A. N., Chow, M. F., et al. 2021, Ain Shams Eng. J., 12, 1545

Owens, M. J., & Forsyth, R. J. 2013, LRSP, 10, 5

Paassilta, M. et al. 2017, JSWSC, 7, A14

Pal, M. 2005, IJRS, 26, 217

Papaioannou, A., Anastasiadis, A., Kouloumvakos, A., et al. 2018, SoPh, 293, 100

Papaioannou, A., Sandberg, I., Anastasiadis, A., et al. 2016, JSWSC, 6, A42

Papaioannou, A., Souvatzoglou, G., Paschalis, P., et al. 2014, SoPh, 289, 423

Papaioannou, A., Vainio, R., Raukunen, O., et al. 2022, JSWSC, 12, 24

Parker, E. 1965, SSRv, 4, 666

Parker, E. N. 1958, ApJ, 128, 664

——. 2009, SSRv, 144, 15

Patel, S. 2020, IJC Heart & Vasculature, 30, 100595

Pedregosa, F. et al. 2011, JMLR, 12, 2825

Perlich, C., Provost, F., & Simonoff, J. 2003, JMLR, 211

Posner, A. 2007, SpWea, 5

Prasetiyowati, M. I., Maulidevi, N. U., & Surendro, K. 2020, in Proc. of the 9th Int. Conf. on Informatics, Environment, Energy and Applications, IEEA 2020 (New York: Association for Computing Machinery), 125

Pulkkinen, T. 2007, LRSP, 4, 1

Ramos, R. L., Carante, M. P., Ferrari, A., et al. 2023, IJMS, 24, 2328

Reames, D. V. 1999, SSRv, 90, 413

——. 2013, SSRv, 175, 53

Reames, D. V. 2021, Solar Energetic Particles. A Modern Primer on Understanding Sources, Acceleration and Propagation, Vol. 978 (Springer Nature)

Reames, D. V., Tylka, A. J., & Ng, C. K. 2001, in AIP Conf. Proc., Vol. 552, Space Technology and Applications International Forum - 2001 (Melville, NY: AIP), 1185

Richardson, I., Mays, M., & Thompson, B. 2018, SpWea, 16, 1862

Rodriguez, J., Krosschell, J., & Green, J. 2014, SpWea, 12, 92

Rodriguez, J., Onsager, T., & Mazur, J. 2010, GRL, 37

Rodriguez, J., Sandberg, I., Mewaldt, R., et al. 2017, SpWea, 15, 290

Rosenberg, R. L., & Coleman Jr, P. J. 1969, JGR, 74, 5611

Rotti, S., Aydin, B., Georgoulis, M., & Martens, P. 2022, GSEP Dataset, v5, Harvard Dataverse, doi: 10.7910/DVN/DZYLHK

Rotti, S., Aydin, B., Georgoulis, M. K., & Martens, P. C. 2022, ApJS, 262, 29

Rotti, S., & Martens, P. C. 2023, ApJS, 267, 40

Rotti, S. A., Aydin, B., & Martens, P. C. 2024, ApJ, 966, 165

Rotti, S. A., Martens, P. C., & Aydin, B. 2020a, ApJS, 249, 20

Rotti, S. A., Martens, P. C. H., & Aydin, B. 2020b, ApJS, 249, 20

Ruiz, A. P., Flynn, M., Large, J., et al. 2021, Data Min. Knowl. Discov., 35, 401

Sadykov, V., Kosovichev, A., Kitiashvili, I., et al. 2021, arXiv:2107.03911

Safavian, S. R., & Landgrebe, D. 1991, IEEE Transactions on Systems, Man, and Cybernetics, 21, 660

Sakoe, H., & Chiba, S. 1978, IEEE Transactions on Acoustics, Speech, and Signal Processing, 26, 43

Sandberg, I., Jiggens, P., Heynderickx, D., & Daglis, I. 2014, GRL, 41, 4435

Sanner, M. F., et al. 1999, J Mol Graph Model, 17, 57

Sarica, A., Cerasa, A., & Quattrone, A. 2017, Front. Aging Neurosci., 9, 329

Sauer, H. H. 1989, AIP Conf. Proc., High-energy Radiation Background in Space (Melville, NY: AIP), 186, 216

Schaefer, J. T. 1990, WtFor, 5, 570

Schäfer, P. 2015, Data Min. Knowl. Discov., 29, 1505

Schäfer, P., & Högqvist, M. 2012, in Proc. of the 15th Int. Conf. on Extending Database Technology, 516

Schapire, R. E. 1990, MachL, 5, 197

Schmelz, J., Reames, D., Von Steiger, R., & Basu, S. 2012, ApJ, 755, 33

Schrijver, C. J., & Siscoe, G. L. 2010, Heliophysics: Space Storms and Radiation: Causes

and Effects (London: Cambridge Univ. Press)

Schwadron, N. A., Townsend, L., Kozarev, K., et al. 2010, SpWea, 8

Shea, M., & Smart, D. 1990, SoPh, 127, 297

Sheeley, N. R. 2005, LRSP, 2, 1

Singer, H., Heckman, G., & Hirman, J. 2001, GMS, 125, 23

Smart, D., & Shea, M. 1979, in Solar-Terrestrial Predictions Proc., Vol. 1, NOAA, 406

Smart, D., & Shea, M. 1992, ASR, 12, 303

Sokolov, I., Roussev, I., Gombosi, T., et al. 2004, ApJ, 616, L171

Stassinopoulos, E., & Raymond, J. P. 1988, in Proc. of the IEEE, 76, 1423

Stone, M. 1977, Biometrika, 29

Strauss, R., & Fichtner, H. 2015, ApJ, 801, 29

Stumpo, M., Benella, S., Laurenza, M., et al. 2021, SpWea, 19, e2021SW002794

Swalwell, B., Dalla, S., & Walsh, R. W. 2017, SoPh, 292, 1

Tenishev, V., Shou, Y., Borovikov, D., et al. 2021, JGR: Space Physics, 126, e2020JA028242

Thakur, N., Gopalswamy, N., Mäkelä, P., et al. 2016, SoPh, 291, 513

Torres, J., Zhao, L., Chan, P. K., & Zhang, M. 2022, SpWea, 20, e2021SW002797

Tribble, A. 2010, Energetic Particles and Technology (London: Cambridge Univ. Press), 381, ed. Schrijver, Carolus J. and Siscoe

Trottet, G., Samwel, S., Klein, K.-L., et al. 2015, SoPh, 290, 819

Tyralis, H., Papacharalampous, G., & Langousis, A. 2019, Water, 11, 910

Usoskin, I. G. 2017, LRSP, 14, 3

Vainio, R. et al. 2013, JSWSC, 3, A12

Van Hollebeke, M. A. I., Ma Sung, L. S., & McDonald, F. B. 1975, SoPh, 41, 189

Vlahos, L., Anastasiadis, A., Papaioannou, A., et al. 2019, RSPTA, 377, 20180095

Wei, G., Zhao, J., Feng, Y., et al. 2020, Appl. Soft Comput., 93, 106337

Whitman, K., Egeland, R., Richardson, I. G., et al. 2022, ASR

——. 2023, ASR, 72, 5161

Wijsen, N., Samara, E., Aran, À., et al. 2021, ApJL, 908, L26

Wild, J. P., Smerd, S. F., & Weiss, A. A. 1963, ARA&A, 1, 291

Wilks, D. S. 1990, WtFor, 5, 640

Winter, L. M., & Ledbetter, K. 2015, ApJ, 809, 105

Woodcock, F. 1976, MWRv, 104, 1209

Xapsos, M. A., Stauffer, C. A., Jordan, T. M., et al. 2012, IEEE Transactions on Nuclear Science, 59, 1054

Xie, H., Mäkelä, P., Gopalswamy, N., & St. Cyr, O. 2016, JGR: Space Physics, 121, 6168

Yashiro, S., Gopalswamy, N., Michalek, G., et al. 2004, JGR: Space Physics, 109

Youden, W. J. 1950, Cancer, 3, 32

Zhang, M., & Zhao, L. 2017, ApJ, 846, 107

Zhang, Y., Jin, R., & Zhou, Z.-H. 2010, IJMLC, 1, 43