From the Sun to Earth
Exploring the Multifaceds of Solar Eruptive Events

Mohamed Nedal

A thesis submitted for the degree of
Doctor of Philosophy

Institute of Astronomy and National Astronomical Observatory,
Bulgarian Academy of Sciences
Solar and Space Weather Research
December 24, 2023

Supervisor: Assoc. Prof. Kamen Asenov Kozarev

Abstract

The following are the 3 abstracts of my 3 main papers. Merge them together in a single abstract with a narrative story ...

We present a comprehensive characterization of 61 Coronal Mass Ejection (CME)-driven compressive waves known as Coronal Bright Fronts (CBFs) observed in the low solar corona between 2010 and 2017. These CBFs have been found to be associated with Solar Energetic Particle (SEP) events near Earth, indicating their importance in understanding space weather phenomena. The aim of this study is to analyze and describe the early dynamics of CBFs using a physics-based heliospheric SEP forecasting system known as the Solar Particle Radiation Environment Analysis and Forecasting - Acceleration and Scattering Transport (SPREAdFAST) framework. This framework utilizes a chain of data-driven analytic and numerical models to predict SEP fluxes at multiple locations in the inner heliosphere by considering their acceleration at CMEs near the Sun and subsequent interplanetary transport. To estimate the time-dependent plasma and compression parameters of the CBFs, we utilized sequences of base-difference images obtained from the Atmospheric Imaging Assembly (AIA) instrument on board the Solar Dynamics Observatory (SDO) satellite, and measurements of the height-time profiles of the coronal waves obtained from the Large Angle and Spectrometric COronagraph (LASCO) instrument on board the Solar and Heliospheric Observatory (SOHO) satellite. We employed kinematic measurements and plasma model results to derive these parameters. The SPREAdFAST framework facilitated the analysis and correlation of these observations with SEP events near Earth. Our analysis yielded statistical relations and distributions for both the shocks and plasma parameters associated with the 61 CBFs investigated. By combining the observations from the AIA and LASCO instruments, as well as the data products from the SPREAdFAST framework, we obtained a comprehensive understanding of the early dynamics of CBFs, including their temporal evolution, plasma properties, and compressional characteristics. These findings contribute to the growing body of knowledge in the field and have implications for space weather forecasting and the study of SEP events.

This study aims to investigate the ambiguous source and the underlying physical processes of the type III radio bursts that occurred on April 3, 2019, through the utilization of multi-wavelength observations from the Low-Frequency Array (LOFAR) radio telescope and the Parker Solar Probe (PSP) space mission, as well as incorporating results from a Potential Field Source Surface (PFSS) and magnetohydrodynamic (MHD) models. The primary goal is to identify the spatial and temporal characteristics of the radio sources, as well as the plasma conditions along their trajectories. We applied data preprocessing techniques to combine high- and low-frequency observations from LOFAR and PSP between 2.6 kHz and 80 MHz. We then extracted information on the frequency drift and speed of the accelerated electron beams from the dynamic spectra. Additionally, we used LOFAR interferometric observations to image the sources of the radio emission at

multiple frequencies and determine their locations and kinematics in the corona. Lastly, we analyzed the plasma parameters and magnetic field along the trajectories of the radio sources using PFSS and MHD model results. We present several notable findings related to type III radio bursts. Firstly, through our automated implementation, we were able to effectively identify and characterize 9 type III radio bursts in the LOFAR-PSP combined dynamic spectrum and 16 type III bursts in the LOFAR dynamic spectrum. We found that the frequency drift for the detected type III bursts in the combined spectrum ranges between 0.24 and 4 MHz s^{-1}, while the speeds of the electron beams range between 0.013 and 0.12 C. Secondly, our imaging observations show that the electrons responsible for these bursts originate from the same source and within a short time frame of fewer than 30 minutes. Finally, our analysis provides informative insights into the physical conditions along the path of the electron beams. For instance, we found that the plasma density obtained from the magnetohydrodynamic algorithm outside a sphere (MAS) model is significantly lower than the expected theoretical density.

Solar energetic particles are mainly protons and originate from the Sun during solar flares or coronal shock waves. Forecasting the Solar Energetic Protons (SEP) flux is critical for several operational sectors, such as communication and navigation systems, space exploration missions, and aviation flights, as the hazardous radiation may endanger astronauts', aviation crew and passengers' health, the delicate electronic components of satellites, space stations, and ground power stations. Therefore, the prediction of the SEP flux is of high importance to our lives and may help mitigate the negative impacts of one of the serious space weather transient phenomena on the near-Earth space environment. Numerous SEP prediction models are being developed with a variety of approaches, such as empirical models, probabilistic models, physics-based models, and AI-based models. In this work, we use the bi-directional long short-term memory (BiLSTM) neural network model architecture to train SEP forecasting models for 3 standard integral GOES channels (>10 MeV, >30 MeV, >60 MeV) with 3 forecast windows (1-day, 2-day, and 3-day ahead) based on daily data obtained from the OMNIWeb database from 1976 to 2019. As the SEP variability is modulated by the solar cycle, we select input parameters that capture the short-term, typically within a span of a few hours, and long-term, typically spanning several days, fluctuations in solar activity. We take the F10.7 index, the sunspot number, the time series of logarithm of the x-ray flux, the solar wind speed, and the average strength of the interplanetary magnetic field as input parameters to our model. The results are validated with an out-of-sample testing set and benchmarked with other types of models.

Acknowledgments

For my family ...

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1   Background and Motivation

The Sun, an ordinary main-sequence star situated at the center of our Solar System, exhibits various forms of activity and variability on multiple spatial and temporal scales One of the main manifestations of solar activity relevant to space weather research are transient energetic eruptive phenomena such as flares, Coronal Mass Ejections (CME), and wide-ranging emissions of electromagnetic radiation and energetic particles (**??**). These eruptive events originate due to the sudden release of free magnetic energy stored in complex, twisted or sheared magnetic field structures in the solar atmosphere (**????**). The energetic phenomena are driven by the rapid dissipation of magnetic energy via magnetic reconnection which can accelerate large numbers of electrons to relativistic energies and heat plasma to tens of million Kelvin (**??**).

The eruptive solar events drive major disturbances in the near-Earth space environment and planetary environments across the heliosphere, collectively termed space weather (**??**). Enhanced fluxes of solar energetic particles (SEPs), plasma ejecta, and electromagnetic radiation emitted during solar eruptions can impact the geomagnetic field, radiation belts, ionosphere, thermosphere, and upper atmosphere surrounding the Earth (**??**). Adverse effects range from disruption of radio communications to damage of satellites, power grid failures, aviation hazards due to radiation risks for airline crew and passengers, and increased radiation exposure for astronauts (**?**). The societal dependence on space-based infrastructure has increased exponentially, escalating the vulnerability to space weather disturbances. Recent studies estimate a severe space weather event could lead to trillion-dollar economic damages in the US alone (**?**). Besides the near-Earth space environment, solar eruptive transients also drive adverse space weather effects across the Solar System impacting activities such as deep space exploration and astronomy (**?**).

Therefore, advancing our understanding of the origins and propagation characteristics of solar eruptive phenomena, as well as quantifying their impacts on geospace and planetary environments, has become an extremely important pursuit for nations worldwide Fundamental research seeks to uncover the physical processes involved using observations coupled with theory and modeling. Concurrently, significant efforts are underway to develop next-generation space environment modeling and forecasting capabilities for predicting the impacts of solar variability. The field combining these research and predictive aspects related to Sun-Earth connections is broadly termed heliophysics (Schrijver and Siscoe, 2010). It encompasses understanding the fundamental solar, heliospheric and geospace plasma processes; coupling across multiple spatial and temporal scales; quan-

tifying the impacts on humanity's technological systems and space-borne assets; and utilizing this knowledge to prevent/mitigate adverse effects (**??**). NASA's Living With a Star program and the National Science Foundation's Space Weather activities exemplify strategic efforts to advance scientific understanding and predictive capabilities across the interconnected domains of heliophysics (**?**).

The present thesis focuses on studying several important phenomena related to solar eruptive activity and its impacts from the perspective of heliophysics research and space weather. The specific topics investigated include: (1) The propagation and evolution characteristics of large-scale coronal disturbances termed EUV waves that are triggered by solar flares and CMEs; (2) The generation, propagation and plasma characteristics of solar radio bursts emitted by accelerated electron beams traveling along open magnetic field lines in the corona; (3) The forecasting of gradual solar energetic particle (SEP) events which constitute one of the major components of space radiation hazards at Earth.

These diverse topics are united by the common theme of seeking to uncover the origins and propagation mechanisms of key transient phenomena resulting from solar eruptions, utilizing observational data, analytical theory and modeling, and data science techniques. The phenomena have been studied for several decades using observations from multiple space missions, but gaps persist in our understanding of their underlying physics and space weather impacts. The thesis aims to provide new insights that help address some of the outstanding questions, guided by the overarching goals and framework of heliophysics research. The following sub-sections elaborate on the background, significance, observational challenges and knowledge gaps pertaining to each of the research topics investigated. The following subsection provide a concise overview of key literature related to the research topics investigated in the thesis. A detailed review is presented in each chapter specific to the respective phenomenon.

1.1.1 Coronal Waves

Coronal waves, or Coronal Bright Fronts (CBFs), also known as Extreme Ultra-Violet (EUV) waves, are large-scale arc-shaped bright fronts or disturbances observed propagating across significant portions of the solar corona following the eruption of CMEs and flares (**??????**). They are best observed in EUV and white-light coronal emission, as well as in radio wavelengths, spanning distances of up to several 100 Mm with speeds ranging from 100-1000 $km\,s^{-1}$, faster than the local characteristic speed in the solar corona, transforming into shock waves (**????**). These structures consist of piled-up plasma with higher density, making them appear brighter in white-light images.

The discovery of coronal waves dates back to observations obtained with the EIT instrument on SOHO launched in 1995, appearing as bright propagating fronts in 19.5 nm wavelength imaging of Fe XII emission lines formed at ~ 1.5 MK plasma (**?**). Subsequent studies based on SOHO/EIT and TRACE imaging found correlations between waves and CMEs, favoring an interpretation as fast-mode MHD waves driven by CME lateral expansions (**?**).

Since 2010, the initiation and evolution of coronal waves are being exquisitely observed with unprecedented resolution by the Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO) instrument (**?**) across multiple EUV passbands sensitive to a wide temperature range (**?**). Alternatively, shock waves can be indirectly observed through the detection of type II radio bursts, which are commonly associated with shock waves in the solar corona **?**. The AIA instrument has provided valuable insights into the

dynamics of the low solar corona over the past decade, thanks to its exceptional spatial and temporal resolution. Equipped with telescopes observing the solar disk in bands 193 and 211 Å, the AIA instrument has demonstrated its ability to distinguish compressive waves in the lower corona **???**. These observations offer valuable information about the kinematics and geometric structure of CBFs. To accurately study the evolution of the wave's leading front, observations off the solar limb are preferred to mitigate projection effects, which may introduce ambiguities in estimating time-dependent positions and the global structure of the wave **?**.

In situ observations of shock waves have revealed their classification into quasi-parallel, quasi-perpendicular, sub-critical, and super-critical shocks based on the angle between the wavefront normal vector and the upstream magnetic field lines **?**. Quasi-parallel shocks have an angle (θ_{BN}) smaller than 45°, while quasi-perpendicular shocks have θ_{BN} greater than 45°. Supercritical shocks, often associated with accelerated particles, are promising candidates for generating type II radio bursts **?**. However, obtaining accurate estimates of shock strength and obliquity solely from remote observations is challenging.

Coronal waves exhibit diverse morphology and kinematics ranging from circular fronts to narrow jets or expanding dome-like structures (**?**). A taxonomy of wave properties based on extensive observational surveys can be found in papers by **?** and **?**. However, despite being observed for over two decades since their serendipitous discovery, fundamental questions remain regarding the physical nature and drivers of coronal waves (**???**). The debate centers around two competing interpretations - the wave versus pseudo-wave (or non-wave) models. The wave models envisage coronal waves as fast-mode MHD waves or shocks that propagate freely after being launched by a CME lateral over-expansion or an initial flare pressure pulse (**??**). The pseudo-wave models interpret them as bright fronts produced by magnetic field restructuring related to the CME lift-off process rather than a true wave disturbance (**??**). Extensive observational and modeling studies have been undertaken to evaluate the two paradigms (**??**), but a consensus remains elusive. Addressing these outstanding questions related to the nature and origin of coronal waves is imperative, since they are being incorporated into models as a primary agent producing SEP events and geomagnetic storms during CMEs (**??**). Their use as a diagnostic tool for CME and shock kinematics predictions in these models requires discriminating between the different physical mechanisms proposed for their origin.

The present thesis undertakes an extensive statistical analysis of coronal EUV wave events observed by SDO to provide new insights into their kinematical properties and relationship to CMEs. We focus on analyzing their large-scale evolution as a function of distance and direction from the source region, leveraging the extensive EUV full-disk imaging capabilities of SDO spanning nearly a decade. Statistical surveys to date have mostly focused on initial speeds and morphological classifications rather than large-scale propagation characteristics. Our study aims to uncover systematic trends in their propagation kinematics using a significantly larger sample compared to previous works. We also comprehensively evaluate associations with CME and flare parameters in order to discriminate between wave and pseudo-wave origins. The results have important implications for incorporating coronal waves into predictive models of CMEs and SEP events for future space weather forecasting.

1.1.2   Solar Radio Bursts

Solar radio emissions have been the subject of extensive study and research due to their connection with solar activity and their potential impact on Earth's atmosphere and technology. One area of particular interest is solar radio bursts, which are intense bursts of electromagnetic radiation originating from the Sun. These bursts can be classified into different types based on their characteristics and associated phenomena. Solar radio bursts, including Type III bursts, serve as remote diagnostics for the study of energetic electrons within the solar corona. These bursts result from transient energetic electron beams injected into the corona, which then propagate along interplanetary magnetic field (IMF) lines (**???**). As these electron beams traverse the corona, they induce plasma waves, also known as Langmuir waves, which subsequently transform into radio emission at the local plasma frequency or its harmonic components (**?**). The frequency of the radio emission is directly linked to the plasma density, making Type III bursts a valuable tool for investigating the inner heliosphere and understanding the underlying processes that drive solar active phenomena, such as solar flares and coronal mass ejections (**??**). These bursts offer insights into the acceleration of energetic electrons in the corona and their transport along magnetic field lines (**?**). The generation of electromagnetic emission at radio frequencies through plasma emission mechanisms is a key aspect of solar radio bursts, shedding light on the dynamic interplay between non-thermal electron distributions and the ambient plasma (**?**).

In radio spectrograms, type III bursts manifest as intense enhancements of radio flux over background levels, exhibiting rapid frequency drifts over timescales ranging from seconds to minutes, signifying plasma dynamics (**?**). These bursts are observable across a wide range of frequencies, spanning from GHz to kHz, and wavelengths extending from metric to decametric (**???**). This phenomenon is detectable by ground-based instruments on Earth as well as various spacecraft within the heliosphere, underscoring the significance of plasma dynamics in their manifestation.

Pioneering observations of solar radio bursts were made in the 1940s leading to their classifications (**?**). Subsequent spectrographic studies uncovered emission mechanisms, source regions and particle diagnostics (**?**). Magnetic reconnection models of flares provided theoretical explanations for particle acceleration generating radio bursts (**?**). Radio imaging spectroscopy using interferometric imaging arrays coupled with high time/frequency resolution spectrometers enables tracking radio sources as a function of frequency and position on the Sun, yielding particle acceleration locations and trajectories through the corona into interplanetary space (**???**). This provides a unique diagnostic of energetic particle transport from the Sun to the Earth which is crucial for improving SEP forecasting models.

Different types of bursts are observed, classified based on their spectral characteristics as documented in radio burst catalogs (**?**). The present thesis focuses on detailed analysis of solar type III radio bursts and their associated phenomena (**?**). Type III bursts appear as intense rapidly drifting emissions from high to low frequencies over seconds, corresponding to the propagation of energetic electron beams from the low corona to beyond 1 AU along open field lines. They signify the initial escape of flare-accelerated electrons into interplanetary space, making them an important precursor signature of SEP activity (**??**). Investigating their source locations, plasma environments, and beam kinematics based on multiwavelength observations coupled with plasma emission theory is therefore vital for improved understanding of coronal particle acceleration and transport processes

relevant for SEP forecasting models.

While type III bursts have been studied for over 50 years since their initial discovery by **???**, gaps persist in our understanding of their exciter beams and emission mechanisms. Key outstanding questions pertain to the detailed electron acceleration and injection sites, beam configurations and energy spectra, drivers of burst onset and duration, and the role of density fluctuations in propagating beams (**???**). Recent work combines imaging and spectral data with modeling to constrain radio burst exciters in unprecedented detail (**??**). Key challenges remain in reconciling emission models with observations and predicting radio diagnostics. Advancing our knowledge of these aspects through coordinated observations and modeling can help constrain the predictions of energetic electron properties based on radio diagnostics. The present work undertakes detailed investigation of a solar type III burst combining imaging and radio spectral data to derive electron beam trajectories and coronal densities, and models the emission sources. The results provide insights into the corona plasma environment and energetic electron transport relevant for SEP forecasting applications.

1.1.3   Solar Energetic Particle (SEP) Forecasting

Solar Energetic Protons (SEP) are high-energy particles that are believed to be originated from the acceleration of particles in the solar corona during coronal mass ejections (CMEs) and solar flares (**?????**). They are typically characterized by their high energy levels - with some particles having energies in the relativistic GeV/nucleon range - and their ability to penetrate through spacecraft shielding, causing radiation damage (**??**). The fluence and energy spectrum of SEP are influenced by several factors, including the strength of the solar flare or CME that produced them, and the conditions of the interplanetary environment (**????????**).

SEP exhibit a strong association with the solar cycle, with the frequency and flux of SEP events peaking during the maximum phase of the solar cycle (**?**). This is thought to be due to the increased activity of the Sun during this phase, which leads to more frequent and powerful flares and CMEs. Previous studies have shown a relationship between the occurrence frequency of SEP and the sunspot number (SN; **?**, **?**; **?**, **?**). However, the exact relationship between the solar cycle and SEP is complex and not fully understood. Hence, more work is needed to better understand this connection, as previous studies have reported intense SEP events during relatively weak solar activity (**??**).

SEP have been a subject of interest and research in heliophysics for decades. It is hypothesized that shock waves generated in the corona can lead to an early acceleration of particles. However, SEP have sufficient energy to propagate themselves by *surfing* the interplanetary magnetic fields (IMF), and therefore, the expanding CME is not necessary for their transport (**????**). While this theory has gained acceptance, there is an ongoing debate among scientists over the specific mechanisms and conditions responsible for SEP production and acceleration.

The creation, acceleration, and transport mechanisms of SEP are complex and involve a combination of magnetic reconnection, shock acceleration, and wave-particle interactions (**???**). The specific mechanisms responsible for SEP production and acceleration can vary depending on the type and strength of the solar event that triggered them. Further research is imperative to better understand the processes involved in the production and transport of SEP in the heliosphere. This will facilitate the development of more precise models that assist in minimizing the impact of SEP on astronauts and

space-based assets.

The arrival of solar energetic particles (SEPs) in the near-Earth space environment constitutes one of the major components of adverse space weather (**??**). SEPs consist primarily of protons (and some heavy ions), accelerated to very high energies by CME-driven shock waves during large solar eruptive events. The gradual SEP events, so called due to their long durations from several hours to a few days, involve protons accelerated to energies above ~ 10 MeV which can penetrate Earth's magnetic field and atmosphere posing radiation hazards to humans and equipment in space and at polar regions (**?**).

Initial SEP forecasting models were based on empirical correlations between proton intensity profiles and CME or flare properties (**?**). The complex physics of CME shock acceleration combined with modeling the transport of SEPs through turbulent interplanetary magnetic fields presents major challenges for first-principles based SEP forecasting models (**??**). As an alternative approach, empirical and data-driven models based on statistical/machine learning techniques applied to historical SEP event data have shown considerable promise for operational forecasting over the past decade (**???**). This motivates detailed investigation of data-driven SEP forecasting models using state-of-the-art machine learning algorithms which can outperform conventional empirical methods.

The emergence of d learning techniques has enabled application of sophisticated machine learning models to SEP forecasting, yielding improved predictions since they can capture complex nonlinear relationships between parameters which has been leveraged for diverse space weather applications recently (**??**). Opportunities exist for novel forecasting approaches utilizing deep learning algorithms and expanded input parameters. However, applications to SEP forecasting problems are still limited, presenting an important research gap which this thesis aims to address. In the present work, we develop a deep neural network model for predicting the intensity profile of >10 MeV gradual SEP proton events utilizing near real-time solar wind plasma measurements as model inputs. The developed model is trained and tested on a database of historical SEP events spanning solar cycles 23 and 24, with the goal of producing SEP flux forecasts over an hour in advance of particle arrivals near Earth. Such capability can provide actionable information for mitigating radiation effects from extreme SEP events. The study demonstrates the potential of state-of-the-art machine learning algorithms to achieve significant enhancement of SEP forecasting capabilities building upon conventional empirical methods.

1.2   Objectives and Scope

The primary objectives and research questions addressed through the investigations carried out in this thesis include:

1. Characterize the large-scale propagation kinematics of coronal EUV waves over distances of hundreds of Mm from the eruption source location. Compare observed spatial and temporal variations in speeds with analytical CME-driven wave/shock models.

2. Conduct a comprehensive statistical analysis correlating properties of EUV waves with associated CME and flare parameters utilizing a large event sample. Discriminate between wave and pseudo-wave models based on observational evidence.

3. Analyze coordinated observations of a solar type III radio burst across imaging and

radio spectral domains to derive coronal density profiles, electron beam kinematics and emission source models.

4. Develop a deep neural network model for forecasting the intensity profile of >10 MeV gradual SEP proton events using real-time solar wind data as inputs. Evaluate model performance and forecast accuracy over different lead times.

The scope of the thesis encompasses key phenomena related to solar eruptions and their space weather impacts that align with the outstanding questions and challenges highlighted in the background discussion. While expansive in scope, some limitations exist that bound the present work:

- The studies rely primarily on remote sensing observations of the Sun and heliosphere, limited by measurement capabilities and resolution.

- Analytical modeling utilizes simplified theory and assumptions which cannot account for all complexities.

- Machine learning models have dependencies on data coverage and uncertainties in input parameters.

- Findings are constrained by the event samples studied and applicability to the broader population.

These factors imply appropriate care and diligence in interpretation of results and their generalizability. Nevertheless, the present work establishes an important foundation for future advances that can build upon these limitations.

1.3 Methodology Overview

The research presented in this thesis employs a synergistic methodology combining analytical theory, numerical modeling, and data science techniques. Both observational case studies and statistical analysis approaches are utilized for gaining new insights from application of these tools. The data sources, models, and algorithms employed in each of the investigations are concisely summarized below.

Coronal waves: The study utilizes an event database of ~ 60 coronal EUV waves observed by SDO/AIA since 2010, tracking kinematics to $30\,R_\odot$ distances. Evolution trends are compared with analytical CME-driven wave propagation models. Statistical associations with CME and flare parameters provide corroboration for physical interpretation.

Solar radio bursts: Multi-wavelength observations of type III bursts from the LOFAR stations and Parker Solar Probe (PSP) are analyzed. Beam trajectories, densities, and emission sources are characterized by combining imaging data, plasma emission theory and coronal density models.

SEP forecasting: A database of >10, >30, and >60 MeV SEP integral fluxes during the last four solar cycles is generated using the GOES database. Deep neural network models are developed using solar wind data time-series as inputs. Model training, testing and validation is performed to evaluate forecast accuracy over different lead times.

This triangulation between data analysis, physics-based modeling and data-driven modeling provides confidence in the results obtained. Details of the methodological approaches are elaborated in their respective chapters.

1.4   Main Contributions

The primary contributions arising from the research presented in this thesis include:

- New large-scale kinematical characterization of coronal EUV waves propagating to distances over 100 Mm. Derived velocity and acceleration trends challenge steady-wave behavior assumed in models.

- Statistical analysis correlating EUV wave and CME/flare properties using a significantly larger event sample compared to prior studies. This enables stronger discrimination between competing initiation models.

- Novel methodology combining radio and EUV observations with analytical modeling to reconstruct plasma environments and electron beam trajectories for a solar type III radio burst.

- Deep learning forecasting model for intense SEP events using an expanded input parameter space based on solar wind data. This demonstrates cutting-edge artificial intelligence capabilities for space weather applications.

- Synergistic approach leveraging analytical theory, numerical modeling and data science techniques to gain new insights on long-standing problems in heliophysics research related to solar eruptions and their space weather impacts.

These contributions provide advances over prior state-of-the-art in the respective areas. They have implications for improving models used in operational space weather monitoring and forecasting systems, besides progressing fundamental physics understanding of solar and heliospheric phenomena. The results validate the merit of cross-disciplinary studies combining traditional analytical techniques with modern statistical and machine learning methods to enable discoveries from application of these synergies.

1.5   Outline

This thesis is divided into the following five chapters:

Chapter 1 - Introduction: Provides a background to the research topics, motivation and context of the work, summary of literature, overview of methodology, and the structure of the thesis.

Chapter 2 – Propagation and Drivers of Coronal EUV Waves: Presents a statistical analysis of the kinematics and physical interpretation of coronal waves using EUV imaging observations and analytical models.

Chapter 3 – Plasma Environment and Energetics of a Solar Type III Radio Burst: Details a multi-wavelength observational case study of a type III burst combining data analysis and modeling to probe the radio emission physics.

Chapter 4 – Deep Learning Approach for Forecasting Intense SEP Events: Describes the development and evaluation of a neural network model for predicting SEP properties using solar wind data.

Chapter 5 – Conclusions and Future Outlook: Summarizes the key findings, implications, and limitations of the research studies. Discusses future extensions building on the present work.

The core chapters 2 through 4 present the major research investigations carried out. The multi-faceted phenomena are studied by tailoring the methodology to leverage their key observational signatures. Together they provide new insights on different aspects of solar eruptions and space weather. Each chapter is structured to be reasonably self-contained, with relevant background and literature specific to the phenomenon under

study. The findings are synergistic and united by the common thread of employing heliophysics principles to address outstanding questions using cutting-edge analytics.

Chapter 2

Characterization of Coronal Bright Fronts

2.1   Introduction

2.2   Observations

2.3   Data Analysis and Methods

2.4   Results and Discussion

2.5   Conclusions

Chapter 3

Multi-Viewpoint Solar Radio Observations: Integrating Space-based and Ground-based Data for Coronal Diagnostics

3.1 Introduction

Type III radio bursts are manifestations of transient energetic electron beams injected into the solar corona, propagating along the interplanetary magnetic field (IMF) lines **???**. As these beams traverse the corona, they trigger plasma waves (also known as Langmuir waves) that are then transformed into radio emission at the local plasma frequency or its harmonic components **?**. In the radio spectrograms, type III bursts are usually observed as intense emissions that drift in frequency over timescales of several seconds to minutes and over a wide range of frequencies, from metric to decametric wavelengths **???**, making them detectable by ground-based instruments on Earth and various spacecraft within the heliosphere. The frequency of the radio emission is directly related to the plasma density, making type III bursts a valuable diagnostic tool for examining the inner heliosphere and the processes that drive solar active phenomena, such as solar flares and coronal mass ejections **??**.

The electron beams follow open magnetic field lines and can persist well beyond 1 astronomical unit (AU) (e.g., **??**), offering in situ insights into the burst and ambient conditions of the heliosphere, including electron density, radio frequency drift, speed of the electron beams and even potential direct detection of Langmuir waves (see **??** and **?** and references within). In addition, tracing the path of type III bursts provides a map of the density structure of the heliosphere, serving as a foundation for developing and testing density models. Since radio observations below ~10 MHz cannot be accomplished from the ground, it is important to combine high- and low-frequency observations from ground-based and space-borne instruments. In this work, we perform a study of several type III radio bursts that occurred in close succession on April 3, 2019. We use remote observations of type III radio bursts detected by the Low-Frequency Array (**?**, LOFAR) ground-based radio telescope and the Parker Solar Probe (**?**, PSP) spacecraft during Encounter 2 to study the sources of these radio emissions and to investigate the physical conditions responsible for their generation. Additionally, we incorporate results of two steady-state models of the solar corona: the potential field source surface (PFSS) model

?? and the magnetohydrodynamic algorithm outside a sphere (MAS) model **?**, to gain a better understanding of the coronal magnetic environment and its role in the acceleration of electrons. The ground-based LOFAR imaging observations provide valuable insight into the actual location of the burst sources. This research aims to expand upon current knowledge of the electron beams responsible for triggering type III radio bursts and the coronal conditions they experience. Gaining a deeper insight into this aspect is vital in comprehending other solar phenomena, such as solar energetic particles and solar wind, and how they influence the near-Earth space environment.

A number of recent studies investigate the physical mechanisms responsible for the generation of solar type III radio bursts. For example, **?** investigated the association of type III bursts with flaring activities in February 2011, via combined multiwavelength observation from the Solar Dynamic Observatory (SDO) instruments, as well as Wind/WAVE and ground-based instruments. They found that the SDO measurements indicated that type III emission was correlated with a hot plasma (7 MK) at the extreme ultraviolet (EUV) jet's footpoint. By using a triangulation method with the Wind and the twin STEREO spacecraft, **?** reported the first measurements of the beaming characteristics for two type III bursts between 2007-2008, assuming the source was located near the ecliptic plane (see also **?**). They concluded that the individual type III bursts have a broad beaming pattern that is roughly parallel to the Parker spiral magnetic field line at the source. **?** conducted a study on almost 10,000 type III bursts observed by the Nancay Radioheliograph between 1998 and 2008. Their analysis revealed discrepancies in the location of type III sources that may have been caused by a tilted magnetic field. Additionally, they found that the average energy released during type III bursts throughout a solar cycle could be comparable to the energy produced by non-thermal bremsstrahlung mechanisms in nano-flares. **?** utilized LOFAR data to investigate the statistical characteristics of over 800 type III radio bursts within an eight-hour period on July 9, 2013. They discovered that the drift rates of type III bursts were twice that of type S bursts and plasma emission was the primary emission mechanism for both types.

? introduced a statistical overview of type III radio bursts during the first two PSP solar encounters. While the first encounter in November 2018 revealed a small number of bursts, the second encounter in April 2019 exhibited frequent type III bursts, including continuous occurrences during noise storms. They reported the characteristics of type III bursts with spectral and polarization analysis.

? performed a statistical survey of 30 type III radio bursts detected by PSP during the second encounter in April 2019 and estimated their decay times, which were used to estimate the relative electron density fluctuations in the solar wind. They localized radio sources using a polarization-based-radio triangulation technique, which placed the sources near the modeled Parker spiral rooted in the active region AR12738 behind the plane of the sky as seen from Earth.

? explored correlations between type III radio bursts and EUV emission in the solar corona. Using coordinated observations from PSP, SDO, and Nuclear Spectroscopic Telescope Array (NuSTAR) on April 12, 2019, they identified periodicities in EUV emission correlated with type III burst rates. The findings suggested impulsive events causing heating and cooling in the corona, possibly nano-flares, despite the absence of observable flares in X-ray and EUV data, which implies periodic non-thermal electron acceleration processes associated with small-scale impulsive events.

? explored the origin of the type III radio bursts we are tackling in this paper and found that electron beams that triggered radio bursts may have emanated from the

periphery of an active region that showed significant blue-shifted plasma. More recently, **?** observed a distinct type III radio burst using the PSP and LOFAR between 0.1 and 80 MHz on April 9, 2019, around 12:40 UT, six days after the occurrence of the event analyzed in our study. While no detectable flare activity was linked with the event, a type III noise storm was ongoing during the PSP encounter 2. The authors determined the type III trajectory and reconstructed its source using observations from Wind and STEREO spacecraft, as well as measuring related electron enhancement in situ.

In the last few years, we have witnessed the emergence of modern instruments, such as LOFAR and PSP, that have allowed for the observation of solar radio emissions with higher sensitivity from a better vantage point. Although type III bursts have been extensively studied **?**, there are still some unresolved issues regarding the exact mechanism of type III emissions. For example, it is not yet clear how the electrons are accelerated to the high energies required to generate type III radio bursts or what role the coronal magnetic field plays in this process. Furthermore, there are inconsistencies between the observations and the models, which need to be resolved in order to gain a more complete understanding of the dynamics of the solar corona. Examples of these inconsistencies are the origin of the type III radio bursts and the discrepancy between the estimated plasma densities from the models and the observations. This paper aims to address these unresolved challenges by using new observations from LOFAR and PSP and models of the solar corona to study the physical mechanisms responsible for the generation of type III bursts. The data analysis includes a combination of radio spectroscopy and imaging techniques to study the frequency, temporal and spatial variations of the radio bursts.

The paper is organized as follows. In Section **??**, we describe the observations of type III radio bursts made with LOFAR and PSP. In Section **??** we explain the data analysis and modeling techniques used to study these events. In Section **??**, we present the results of our analysis, including an investigation of the potential physical mechanisms responsible for the generation of type III radio bursts and a comparison of the observations with models of the solar corona. Finally, in Section **??**, we summarize our findings and discuss their implications.

3.2 Observations

A number of studies focused on observing the solar radio emissions during the second encounter of the PSP in late 2019 **?????**. In this study, our primary emphasis is directed towards investigating a set of type III radio bursts that took place on April 3, 2019, during the time interval spanning from ∼12:10 to 12:50 UT. This period coincided with the presence of two distinct active regions (ARs) on the Sun, denoted as AR12737 and AR12738. AR12737 was situated on the Sun's near side at coordinates E12^{o}N06^o. Notably, this region had eight sunspots and exhibited a β magnetic configuration according to the Hale magnetic classification **?**. On the other hand, AR12738 was positioned on the solar far side at coordinates E140^{o}N02^o. Due to its remote location, detailed observations of the magnetic configuration and activity within AR12738 were unattainable in this time frame.

We observed a group of intense type III radio bursts by four instruments (Wind/WAVES, PSP/FIELDS, STEREO-A/SWAVES, and LOFAR/LBA) via a regular survey. In Figure **??**, we show the first type III burst within the time of this study as observed by the four instruments. By taking the second derivative of the light curve at a specific

frequency channels, we determined the start time of the burst, which is denoted by the vertical red dashed line. The frequency bands used for obtaining the start time at each instrument are as follows: 6.97 MHz (Wind), 7.03 MHz (STEREO), 5.03 MHz (PSP), and 40.16 MHz (LOFAR).

We checked the relative orientations of the instruments with respect to Earth (Fig. **??**). Since the PSP and STEREO spacecraft were almost aligned (close in an angular sense) with the Sun, the STEREO/EUVI image could be taken as what PSP would see (Fig. **??**). Figure **??** shows how the solar disk looks like from the Earth perspective (using the SDO/AIA instrument) and from the eastern side where the PSP and STEREO were located at that time (using the STEREO/EUVI instrument). The right panel shows a closer view of AR12737 with the contours of the photospheric magnetic field obtained from the Helioseismic and Magnetic Imager (HMI) on board SDO. From the GOES-15/XRS and SDO/EVE instruments in the panels below, they also confirm that there is no flaring activity at that time.

The solar disk was quiet, including only one AR that is visible with no X-rays and no EUV transient emissions over this period. Nevertheless, the very sensitive LOFAR telescope detected a number of bursts close to noon. We checked PSP data, and we found bursts there as well. Meanwhile, from the EUVI and AIA images, we see that there are numerous small localized regions of relatively higher intensity (i.e., likely small-scale coronal brightenings spots or campfires; see **???**). In the next subsections, we introduce the PSP and LOFAR instruments and their observations of the radio bursts.

3.2.1   PSP Observations

Parker Solar Probe (PSP) is a pioneering spacecraft with cutting-edge technologies, launched on August 12, 2018, aimed at helping to resolve key questions about solar corona and solar wind (**?**). To study the radio bursts, we used the level-2 data of the radio dynamic spectrum obtained from the FIELDS instrument suite (**??**), which can be downloaded from this website[1]. The data file is in CDF format and the unit of the data values is converted from V^2/Hz to dB units using the formula

$$I_{dB} = 10 \times log_{10}(I/10^{-16})$$ (3.1)

The minimum power spectral density (PSD) of 10^{-16} V^2/Hz is used as a threshold for radio bursts according to **?** for converting to decibels. Then, both the High-Frequency Receiver (HFR: 1.3 – 19.2 MHz) and the Low-Frequency Receive (LFR: 10.5 kHz – 1.7 MHz) data are combined into a single dynamic spectrum as shown in Figure **??** with a full frequency range between 10.5 kHz - 19.2 MHz. The mean intensity value at each timestep over the full frequency range is subtracted from each frequency channel to clean the spectrum and minimize the noise level.

3.2.2   LOFAR Observations

The LOw Frequency ARray (LOFAR) radio telescope (**?**) is a powerful tool for studying the Sun at low radio frequencies ranging between 10 and 240 MHz. Its high sensitivity and high time resolution have enabled the detection of various solar phenomena, including radio bursts and CMEs, and the study of dynamic processes in the solar atmosphere

---

[1]PSP FIELDS data products: `http://research.ssl.berkeley.edu/data/psp/data/sci/fields/`

Figure 3.1: Radio dynamic spectra for a single burst obtained from multiple instruments. The top-left panel is from the LOFAR/LBA instrument, the top-right is from the PSP/FIELDS instrument, the bottom-left is from the STEREO/SWAVES instrument, and the bottom-right is from the Wind/WAVES. The vertical red dashed line denotes the start time of the burst.

Figure 3.2: Top view of the spacecraft positions in the ecliptic plane at 12:15 UT on April 3, 2019, with the Sun-Earth line as the reference point for longitude. The Earth's location is representative of the positions of LOFAR, Wind/WAVES, and GOES-15/XRS instruments. The spacecraft were connected back to the Sun by a 400 km/s reference Parker Spiral. The black arrow represents the longitude of AR12737 and the blue arrow represents the longitude of the AR12738. The gray dotted lines are the background Parker spiral field lines. The black dashed spiral shows the field line connected to the AR12737, and the blue dashed spiral is connected to the AR12738. The figure is generated using the Solar MAgnetic Connection Haus (Solar-MACH) tool **?**.

on timescales of milliseconds. The LOFAR dynamic spectrum from the beamformed radio observations is obtained by the Low-Band Antenna (LBA: 10 – 90 MHz) and can be downloaded from the LOFAR long-term archive (LTA)[2]. The High-Band Antenna (HBA: 110 – 190 MHz) data are not available for that timeframe. For this day under study, the LOFAR data are available between 11:42 – 13:27 UT. To clean the spectrum, background subtraction is performed, which flattens the sensitivity (response) with the frequency of the LBA antennas. Basically, the mean spectrum along each frequency band is calculated and subtracted from the whole frequency band, the same applied to the PSP spectrum. This operation effectively removes the constant background from the spectrum. Then a Gaussian smoothing filter is applied to the spectrum using the **scipy.ndimage.gaussian_filter** function with a sigma value of 1.5, which helps to reduce noise and variations in the data. After that, the PSP and LOFAR spectra are combined together in a single plot within the same time interval. The bursts' signals observed by the PSP occur earlier than those at LOFAR. This is due to the fact that the PSP spacecraft is much closer to the Sun and hence it detects the radio emissions earlier than LOFAR because of the shorter travel time of radio signals from the Sun. Therefore, the PSP dynamic spectrum must be shifted with respect to the LOFAR observations based on a calculation of the relative time travel of the radio emission from the Sun to PSP and to LOFAR. In addition, the time cadence of the PSP observations changes according to its distance from the Sun. On that day, the PSP data cadence was 7 seconds, while LOFAR's is 1 second. Therefore, the LOFAR dynamic spectrum was downsampled to 7 seconds to match the time resolution of the PSP. Figure **??** shows the resulting combined LOFAR-PSP spectrum on a logarithmic y-axis. The LOFAR LBA frequency ranges between 19.82 – 80.16 MHz and for the PSP is 10.55 kHz – 19.17 MHz.

In order to detect the type III radio bursts automatically from the combined dynamic spectrum, we applied the **?** algorithm based on the probabilistic Hough transformation that detects vertical bright edges in images, within a certain degree of deviation from the vertical direction.

3.3 Methods

3.3.1 Imaging of radio sources

As part of our task, we developed an automated pipeline consisting of several modules that not only preprocessed and calibrated the LOFAR interferometric data to produce cleaned images of the Sun in the radio band (**?**), but also utilized the resulting data to find the trajectory of the radio sources and sample the magnetic field and plasma parameters at their respective locations through modeling and simulations in subsequent modules.

First, we ran the burst detection algorithm (**?**) [3] on the combined dynamic radio spectrum of LOFAR and PSP (Fig. **??**) to find the characteristics of each type III burst. We converted the spectrum into a binary map to isolate the bursts from the background. Then we applied the Hough transformation to get line segments of the features. For each type III burst, the line segments are grouped together into one group. To account for the interplanetary component within radio dynamic spectra, we employed the Parker electron-density model (**?**) assuming a fundamental emission. This model enabled map-

---

[2]LOFAR LTA: `https://lta.lofar.eu/`
[3]Detection algorithm repository: `https://github.com/peijin94/type3detect`

ping between the time and frequency indices for each type III burst and subsequently converted electron densities into radial distances. Finally, a least-squares fitting method was applied to derive both the frequency drifts and the speed of the electron beams.

After this step, we did the same for the LOFAR dynamic spectrum only (Fig. **??**) to find the (f, t) pairs for every type III burst. Then we took snapshot frequencies for each burst defined by a list of 60 central frequencies between $\sim$20 – 80 MHz from LOFAR LTA for the interferometric imaging. We obtained the interferometric data from LOFAR core and remote stations at the snapshot frequencies for all type III bursts. We used the concurrent observations of the radio source Tau-A to calibrate the interferometric observations. For that, we used the default preprocessing pipeline (**?**, DP3) for preliminary processing and calibrating the measurement sets (MS). Finally we obtained the cleaned images of the radio sources by using w-stacking clean (WSClean) algorithm (**?**) only at the time indices in the MS files that are equivalent to the snapshot frequencies.

Table 3.1: Characteristics of the type III bursts detected via the automatic algorithm from the combined spectrum.

Burst ID	Start Time (UT)	End Time (UT)	Start Frequency (MHz)	End Frequency (MHz)	Frequency Drift (MHz s^{-1})	Beam Speed (c)
1	12:18:45	12:22:42	76.44	1.57	0.892	0.044
2	12:34:05	12:36:31	41.24	0.86	0.241	0.119
3	12:34:40	12:34:56	54.44	26.54	3.992	0.046
4	12:37:14	12:38:09	66.03	10.02	4.006	0.046
5	12:38:17	12:40:54	76.92	1.57	0.77	0.066
6	12:39:34	12:40:11	78.86	11.93	3.192	0.062
7	12:40:28	12:40:40	45.34	22.9	3.21	0.067
8	12:41:39	12:43:06	78.21	2.13	1.555	0.093
9	12:43:53	12:44:15	59.07	42.13	2.424	0.013

After processing and cleaning the interferometric measurements of LOFAR, we explored the observations of each burst individually. Out of the 60 frequency bands in the LOFAR LTA, we chose 54 frequency bands that have unique integer numeric, between 19.92 - 80.08 MHz. For each burst, at each timestamp, the nearest frequency of the fit model to the list of chosen frequencies is picked as the snapshot frequency at that particular timestamp. This process was repeated for all the 16 type III bursts detected in the LOFAR dynamic spectrum in order to obtain snapshot images for each type III burst (Fig. **??**). For each type III burst, we applied persistence imaging in order to create a continuous display of the radio emissions (**?**).

Persistence imaging enables the creation of a clearer and more informative image. In the context of a time-ordered series of images, a method of persisting pixel values can be employed as follows: for each image, compare the value of each pixel to its corresponding value in the previous persistence image in the series. If the pixel value in the current image is brighter than its corresponding pixel in the previous image, replace the previous value with the current one; otherwise, retain the previous value. This process generates a new image, referred to as the current persistence image, which serves as the basis for the subsequent evaluation of the next image in the series. This evaluation involves a pixel-by-pixel comparison between the current image and its associated persistence image, allowing for the identification of any changes or patterns that may have occurred over time. The mathematical background is explained in Appendix **??**.

In order to estimate the locations of the type III sources in 3D space, we combined observations with modeling. We used magnetogram data from the Global Oscillation

Network Group project (GONG) (**?**). We constructed a grid of footpoints on the GONG map over two longitudinal belts around the two active regions AR12737 and AR12738, which are the two potential candidates source regions for the group of type III bursts under study. These points are used as the seed points for tracing the coronal magnetic field lines using pfsspy python package[4], which is a robust implementation in python of the PFSS model developed by **?**. Using the major and minor axes of the beam size, we estimated the radius of the radio source using Equation **??**, which was used to approximate the source size. Since we already obtained the (x, y) positions of the type III sources in the plane of the sky (POS) through LOFAR observations, now it is necessary to determine their corresponding z position to have an overall understanding of their spatial distribution. Therefore, we employed **?**'s approach here, assuming that the type III bursts were from harmonic emission. Firstly, we found the radial distance of the radio source from the Sun in the POS (Eq. **??**). Secondly, we calculated the sources' radial distance (r_{model}) using the 2.5×Newkirk electron-density model (**??**). The 2.5 fold factor is taken to incorporate the effects of scattering and overdensity (streamers) beyond the nominal Newkirk quiet Sun model. The MAS model results (Fig. **??**) show streamers above the eastern limb, supporting the inclusion of such a factor. Lastly, we estimated the z location of the type III sources (Eq. **??**). We proceeded with the $+z$ solution because the theory precludes emission behind POS in this region of high-density gradients (i.e., the emission would be absorbed by passing through the high-density regions of the corona). More details are explained in Appendix **??**.

$$r_{source} = \sqrt{(b_{major}^2 + b_{minor}^2)}, \tag{3.2}$$

$$r_{pos} = \sqrt{(x^2 + y^2)}, \tag{3.3}$$

$$z = \sqrt{(r_{model}^2 - r_{pos}^2)}. \tag{3.4}$$

The result of the deprojection of the type III sources for the sixth burst are shown in Figure **??** with 70% contours made for ten frequencies on the extrapolated magnetic field lines. The red dashed line is a spline fitting curve that represents the trajectory of the centroids of the radio sources. The black arrow points toward the Earth's line of sight (LOS). It is worth to mention that the axes direction in the POS of LOFAR images are different in the 3D space. The (x, y) coordinates in the POS are translated into (y, z) in the 3D space, and z in the POS is translated into x in the 3D space.

3.3.2   Modeling

To explore the characteristics of the coronal plasma environment during the studied events, we used the Predictive Science Inc. (PSI) standard coronal solutions from magnetohydrodynamic (MHD) simulations originating from the MAS code (**?**). The data are available on the PSI's data archive[5]. We obtained the PSI MAS coronal solution (a thermodynamic-with-heating MHD model) on April 3, 2019, at 12:00 UT, with the following simulation result ID[6]. Initially, we calculated the angle between the burst's source radial vector and the LOS. Moreover, we calculated the complement angle, which is the

---

[4]Pfsspy tool: `https://pfsspy.readthedocs.io/`
[5]Predictive Science Inc.: `https://www.predsci.com/mhdweb/home.php`
[6]Simulation result ID: `hmi_med-cor-thermo2-std01_med-hel-poly-std01`

separation angle between the burst's radial vector and POS from the Earth's perspective. Subsequently, we utilized the complement angle to derive the Carrington longitude (**?**), facilitating the extraction of a longitudinal segment from the MAS datacube, as if it were in the POS. Following this, the selected data slice was fed into the FORWARD model, a toolset responsible for generating synthetic coronal maps of observable quantities describing the plasma state. For extracting the longitudinal slices from the MAS data, we utilized the psipy python package[7]. The MAS datacube is specifically defined on a spherical grid and represents a steady-state MHD model. Owing to the inherent attributes of this datacube, the utilization of the FORWARD toolset proves more practical and advantageous for our objective. In Figure **??** we show the first radio contour of the sixth type III burst on top of the equivalent 2D maps for six plasma parameters, as an example. The plasma parameters are, from left to right and from the top to bottom: plasma density, plasma temperature, magnetic field strength, plasma beta parameter, the total plasma pressure, and the Alfven speed, By taking the value of these physical plasma quantities at the centroids' coordinates of the type III sources at each frequency band, we obtained estimates of local plasma conditions shown in Figure **??** for the sixth type III burst, as an example.

3.4   Results and discussion

3.4.1   Detection and characterization of type III radio bursts

We found that the radio waves arrived at STEREO one minute before they arrived at Wind (Fig. **??**). However, the difference between the $+z$ and $-z$ positions of the burst this close to the Sun in terms of light travel time is $\sim$10 seconds ($\sim$4 $R_\odot$), which is within the time resolution of the observations (1 min time resolution). Thus, we cannot confidently conclude whether the emission arrived at one spacecraft first and the other second.

Figure **??** shows the combined dynamic spectrum from both LOFAR and PSP. The free parameters of the auto-detection algorithm do not have the same values as for detection the type III bursts in the LOFAR spectrum alone. Upon visual examination, we observed that the detection algorithm effectively identified type III bursts in the LOFAR dynamic spectrum (Fig. **??**), but it had limitations in terms of detecting type III bursts in the combined spectrum of the LOFAR and PSP, as well as missing segments of the detected bursts and a few bursts entirely. This could be due to the increased frequency drift and dispersion of the radio bursts at lower frequencies, which made it a challenging task for the detection algorithm. We captured nine type III bursts from the combined dynamic spectrum and their characteristics are reported in Table **??**. However, the detection algorithm performed better on the LOFAR dynamic spectrum only and we traced 16 type III bursts.

3.4.2   Imaging of radio emission sources

Figure **??** shows the persistence imaging for the 16 type III bursts in the LOFAR dynamic spectrum (Fig. **??**). The observation frequencies and timestamps of the snapshot images used to produce the persistence image are shown at the top-right corner of each image.

---

[7] Psipy repository: `https://github.com/predsci/PsiPy`

From a visual inspection of Figure **??**, it seems that all the type III emissions originated from the same quadrant in the images (south-east direction on the solar disk), although there was no active region presented at that location except for a single active region nearby the central meridian (Fig. **??**). Based on the imaging data presented in Figure **??**, we chose one representative type III burst (No. 6) for a single-burst analysis in this paper, as it shares similarities in extent and location with other bursts. To determine the spatial connection between the sources of radio emissions and the coronal magnetic field, a three-dimensional (3D) projection of the radio source contours onto the extrapolated coronal magnetic field via the PFSS model was employed (Fig. **??**). The result indicates a discernible south-eastward propagation of the radio sources relative to the Earth's perspective, with no open field line crossing the radio sources. In Figure **??**, we performed an extrapolation only over the two active regions presented on the solar surface at that time. However, when we extrapolated the magnetic field over the entire solar surface, we noticed that the radio sources are aligned with the lower part of large-scale closed field lines and are placed onto the open field lines emanating from the southern coronal hole. No open field lines crossing the radio sources are observed. We note that the PFSS modeling is limited by the fact that AR12738 is behind the limb on April 3 as observed from Earth. Consequently, the magnetic data available to us could be around two weeks old or more. This might limit the reliability of PFSS extrapolation for that region during that specific timeframe.

From Figure **??**, the results suggest several potential origins of these type III radio emissions: 1) they could be triggered in a closed-field lines structure such as large-scale coronal loops, given that the radio sources are aligned to closed-field lines geometry in the southern hemisphere; 2) they could be triggered by electron beams that are accelerated from an open-field active region (**?**). However, from the PFSS model, we found no evidence for magnetic connectivity from both ARs on the Sun at that time; 3) they may result from electron beams that are accelerated in the corona due to expanding magnetic fields from plasma upflows in the active region (**??**). Our findings indicate a notable inverse relationship between imaging quality and the level of solar radio emission brightness (e.g., for type III bursts 10 and 13, for instance). This observation is due to the leakage of solar radio emission into the side lobes of the calibrator beam, which disrupts the accuracy of calibration solutions.

3.4.3   Plasma diagnostics and magnetic field analysis

Considering the observed alignment of radio sources in Figure **??** and the case depicted in Figure **??**, it becomes evident that radio sources at higher frequencies (indicating proximity to the Sun) align with a streamer-like structure near the equator within the coronal model. This structure is characterized by elevated plasma beta, reduced coronal temperature, and diminished Alfven speed. The coronal plasma density was relatively homogeneous with no prominent structures, probably due to the model resolution.

The location of radio sources of all the bursts were in the same quadrant as seen from Earth. Therefore, we assumed that the former description applies for all bursts. We also found that the radio sources were confined between the equatorial sheet and the southern coronal hole and moving along that boundary. Figure **??** shows the variability of the coronal plasma quantities at the radio sources' centroids, taken from FORWARD maps in Figures **??**, at different frequencies for the sixth burst. To estimate the error bars, we initialized random centroids, within the limits of the 70%-contours of the radio

emissions, to sample the plasma quantities at those locations. Then the standard error (SE) is calculated using Equation **??**, where σ is the standard deviation, and n is the number of points.

$$SE = \frac{\sigma}{\sqrt{n}} \tag{3.5}$$

The coronal temperature was increasing with radial distance, which implies there may have been some heating locally. The behavior of the coronal magnetic field, the plasma total dynamic pressure, and the Alfven speed were decreasing over distance as expected. Finally the value of plasma beta parameter started increasing sharply around 40 MHz, which implies that the plasma pressure became more dominant than the magnetic pressure around that distance from the Sun (for a 2.5×Newkirk model, it is 1.89 $R_\odot$ assuming a fundamental emission, or 2.57 $R_\odot$ assuming a harmonic emission).

The top-left panel of Figure **??** shows a comparison between the density profiles of the MAS model, the 2.5×Newkirk model and the theoretical expected density profiles under the fundamental and harmonic assumptions. Although the Newkirk density model provided a useful approximation for determining the height of radio sources in the corona, it is not entirely accurate due to a number of its underlying assumptions; for instance, the assumption of a steady state and the spherical symmetry of the corona, which do not always apply. Therefore, we tried to use the MAS density values to estimate the depth along the LOS of the radio source, since it is supposed to give a more realistic result. We found that the plasma density obtained from the MAS and FORWARD modeling results were significantly lower compared with the 2.5×Newkirk density model and the theoretical expected density obtained from the classical relation in Equation **??**, where f_p is the plasma frequency (in MHz) and n_e is the electron density (in cm^{-3}):

$$n_e = \left(\frac{f_p}{8.98 \times 10^{-3}} \right)^2 \tag{3.6}$$

The required density from the fitted Newkirk model is much higher (about ten times) than what is obtained from the MAS model, even after accounting for the 2.5× enhancement already applied to the standard Newkirk model. This implies the discrepancy cannot be fully explained by the density enhancement factor alone. Furthermore, the imaging places the radio sources near a streamer which is an overdense region in the MAS model, so it seems unlikely the source's apparent location in the model is wrongly attached to a less dense feature, as there are not denser options available. The apparent source positions from the imaging are likely too high, possibly due to scattering effects (**???**), which could lead to fitting an overly dense Newkirk model. Another potential explanation is that there could be a stealth CME that pushed the coronal magnetic field outward, allowing the plasma to follow it to be perceived as having a higher density than expected, and there was not enough time for the magnetic field relaxation to occur (private communication with J. Magdalenić). However, scattering alone does not seem to fully explain the large density discrepancy. While further investigation is certainly needed regarding scattering and propagation effects on the radio waves, it is interesting to report this significant discrepancy between the model and observations, as it highlights limitations in the current modeling and suggests the need for additional physics to properly characterize the density distribution. Resolving this discrepancy could lead to important insights into the true nature of the corona.

3.5   Summary and conclusions

In this work we analyse the characteristics of a series of type III bursts that occurred on April 3, 2019, during the second near-Sun encounter period of PSP. The bursts were observed in dynamic spectra taken with the PSP/FIELDS (2.6 kHz – 19 MHz) instrument, as well as in interferometric imaging with the LOFAR (20 – 80 MHz) ground-based telescope, as part of a coordinated observing campaign. The series of 16 separate weak bursts were observed over the span of ~20 minutes, during an otherwise relatively quiet period. The solar disk as observed from Earth was dominated by a single active region near its centre. We combined the dynamic spectra for the LOFAR frequency range and the PSP frequency range to study the solar radio emissions within the frequency range of 2.6 kHz – 80 MHz.

For the study, we developed a semi-automated pipeline, which allowed us to obtain the exact times and frequencies of the bursts. These we used to align the PSP to the LOFAR observations and to generate interferometric images between 20 and 80 MHz. We performed data pre-processing of the PSP and LOFAR dynamic spectra to resample and shift the data based on the relative location of the spacecraft with respect to the Sun and Earth, and found an excellent temporal match between the two sets of observations. Thus we automatically traced the type III bursts in the dynamic spectra algorithmically and estimated frequency drift and the electron beam speeds. We found that frequency drifts remained relatively uniform between the high-frequency (LOFAR) and low-frequency (PSP) observations, as well as among the bursts, suggesting that they are related.

In addition, we imaged the type III emission at multiple frequency bands using the interferometric observations from LOFAR to determine the locations of the sources in the solar corona. The type III emissions observed were all found to occur in the same general region off the southeast limb of the Sun, leading us to conclude that they shared a single source of electron beams low in the corona. The potential origins of these emissions are varied and include: 1) small-scale impulsive events such as nano-flares (**???**); 2) plasma upflows from the active region (**?**); 3) coronal closed-loop structures (**?**); 4) electron beams accelerated from interchange reconnection (**?**); 5) high-frequency Alfven waves and/or magnetic reconnection in the outer corona (**??**).

Our magnetic extrapolation shows that there is no open potential field to either AR12737 or AR12738, which is consistent with **?**. Our findings are in line with the conclusions of **?**, who proposed that the likely origin of these type III bursts is the AR12737 region. The type III radio bursts in **?** occurred between April 1 and 4, are temporally aligned with the emergence of AR12737 near the eastern limb of the solar disk.

While potential field source surface models provide valuable insight into the large-scale magnetic topology, their reliability decreases near active regions where the field can deviate significantly from a potential configuration. Therefore, the lack of open field connectivity directly to AR12737 suggested by the PFSS model should be viewed with some caution.

This work complements those results by precisely locating the burst sources in the middle of the corona. We used the Newkirk density model to estimate the height of the radio sources from the Sun of one of the type III bursts as representative of all. Combining this with PFSS magnetic modeling, we found good agreement between the centroids of the radio sources and the location of the southern open field lines in the corona, which would be required to produce radio emissions at interplanetary wavelengths in general. On the other hand, this location does not seem to be well connected to the AR itself,

according to the PFSS model.

We attempted to correct the radial distance of the radio sources from the Sun by replacing the Newkirk model with more realistic MHD results from the MAS model, but we found that there is a significant discrepancy between the Newkirk model profile fitted to the observations and the MAS density. This could result from scattering lensing the apparent burst location to a higher altitude, thus, overestimating the height of radio sources in the corona. The presence of type III radio sources at relatively high distances in the corona, with plasma density higher than expected from the MAS model, suggests that there may be missing information in the modeling. One possibility is the existence of a stealth CME that pushed the coronal magnetic field outward, causing the plasma to appear denser than expected (see **?**) — or other non-obvious changes in large-scale coronal magnetic topology. These findings demonstrate that scattering and propagation effects play a significant role in determining the location and directionality of solar radio bursts (**???**). Therefore, the discrepancy between the observed and modeled density profiles could potentially be attributed to scattering and lensing effects that make the radio sources appear higher in the corona than their true location. Further investigation is required to disentangle these effects from limitations in the density models themselves. Overall, accounting for scattering and refraction will likely lead to improved modeling of the corona and solar radio bursts. In future work, we will also employ the time delay of arrival (TDoA) technique (**?**) to estimate the radio burst source positions from multi-instrument observations and compare that with the current methodology in this paper. Solar Orbiter observations shall also be included.

High-fidelity interferometric radio imaging in metric-decametric wavelengths provides a powerful method to characterise solar eruptive events. It is also becoming increasingly important for studying relatively quiet periods, during which there may be elevated levels of in situ particle fluxes. The ability to observe and image faint radio bursts such as those presented in this work, which may be related to episodes of reconnection on the solar surface, and potentially to episodes of solar wind release, is a testament to LOFAR's power as a space weather instrument. In a future work, we will automate and use our method for studying hundreds of faint bursts observed with LOFAR and will investigate their relation to small-scale activity on the solar surface.

Through a novel combination between the LOFAR imaging and MAS model results, we found that the type III radio bursts experienced a weakening background magnetic field, decreasing solar wind dynamic pressure and Alfven speed, increasing plasma beta and coronal temperature, and plasma rarefaction. The radio sources appeared at larger radial distances than the models predicted, which suggests scattering and density fluctuations are important when attempting to interpret the actual burst trajectory. The discrepancies between the observed and modeled radial distances of the radio sources suggest refinements are needed in the models to fully explain the radio imaging and modeling results. Overall, comparing the LOFAR imaging and MAS modeling for these type III bursts motivates further analysis on additional radio bursts to improve our understanding of the physical conditions that influence the propagation of radio emissions in the corona.

chapter3/figs/aia_sta_cutout.pdf

chapter3/figs/xrs_eve.pdf

Figure 3.3: Exploring the X-ray and extreme ultraviolet (EUV) emissions from the Sun. The top panel showcases a cutout region of the SDO/AIA 193 Åimage of the solar disk along with the STEREO-A EUVI 195 Åpoint of view. The white curve is the limb of the solar disk as seen by AIA from the right side. The red and blue colors are the contours of the line-of-sight magnetogram from the SDO/HMI instrument. The levels are (50, 100, 150, 300, 500, 1000) Gauss. The middle panel shows the X-ray flux from the GOES-14 spacecraft shows minimum activity. The bottom panel shows the time series of the ESP Quad band from the SDO/EVE instrument, which shows the solar irradiance in the extreme ultraviolet (EUV) band.

Figure 3.4: Automatic detection of type III radio bursts from the combined radio dynamic spectrum of LOFAR and PSP instruments. The dashed horizontal lines separates the LOFAR frequency range (top) and the PSP frequency range (bottom).
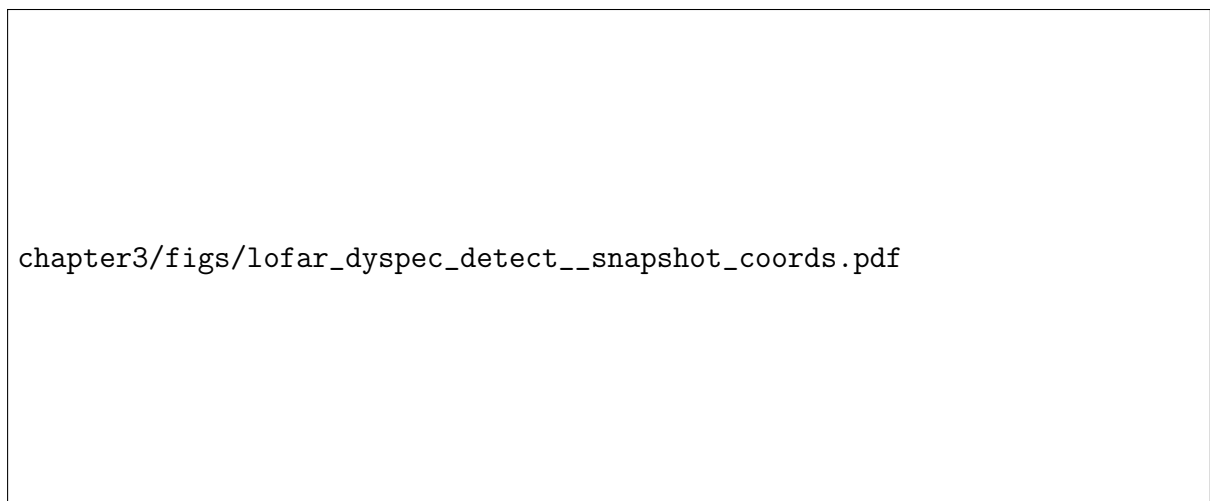

Figure 3.5: Automatic detection of type III bursts observed by LOFAR. The red symbols along the fit lines are the (f, t) coordinates of the image snapshots shown in Figure **??**.

chapter3/figs/subplots.pdf

Figure 3.6: Persistence imaging for the 16 type III bursts detected in the LOFAR dynamic spectrum. The label shows the observation frequencies in MHz and times in (minutes:seconds from 12:00:00 UT). Here, the color coding is not absolute, but rather each panel has its own color code.
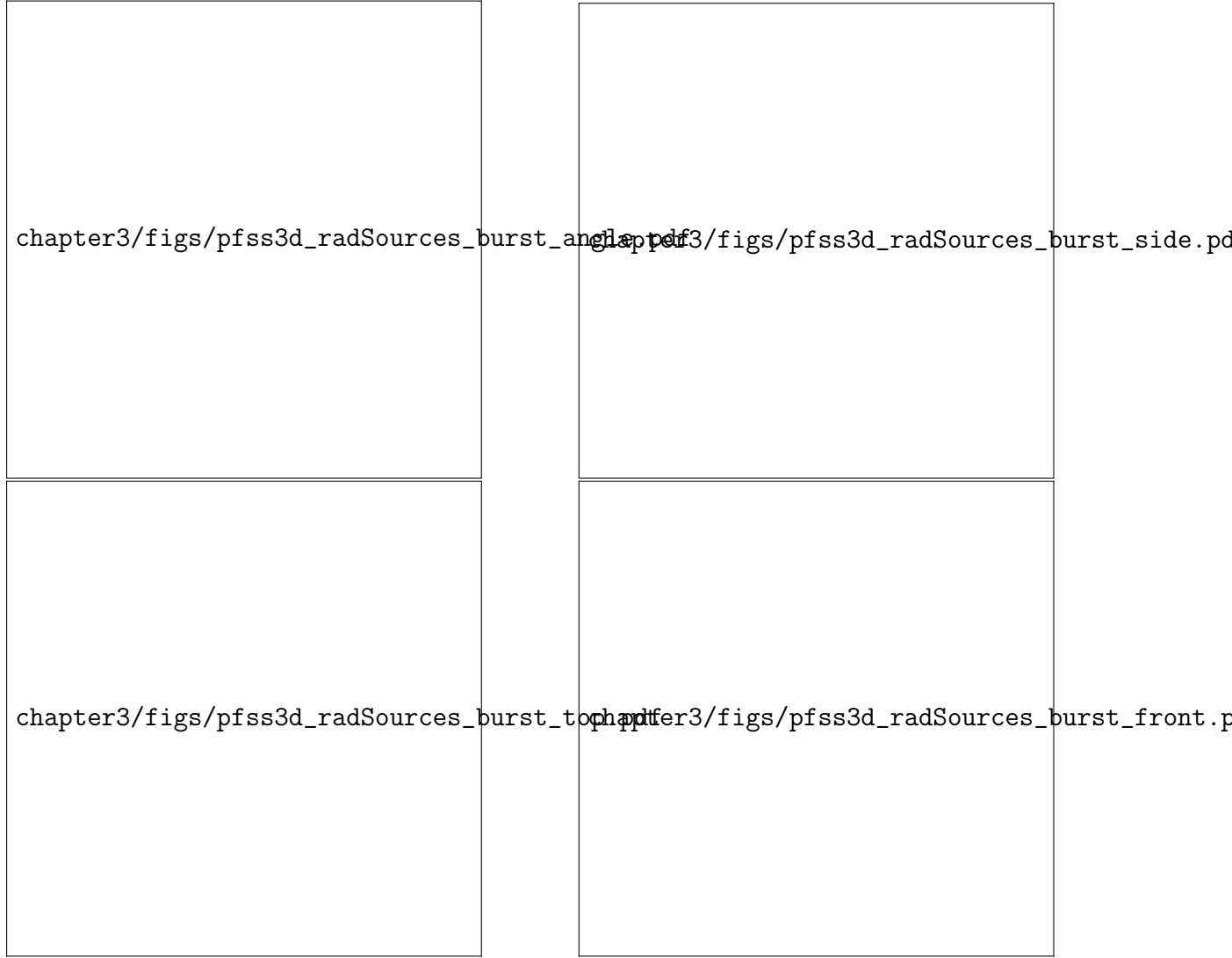
chapter3/figs/pfss3d_radSources_burst_angle.pdf chapter3/figs/pfss3d_radSources_burst_side.pd

chapter3/figs/pfss3d_radSources_burst_top.pdf chapter3/figs/pfss3d_radSources_burst_front.p

Figure 3.7: Different viewing angles for the de-projection of the radio sources of the sixth burst using the 2.5×Newkirk electron-density model on the PFSS solution. The black arrow points toward the Earth's LOS. The yz plane is the POS as seen from the Earth. The red dashed line is a spline curve fit for the sources' centroids. The red, black, and blue curves are the open northern, closed, and open southern field lines, respectively. The opacity of the closed field lines is decreased for better visualization.

Figure 3.8: Synthesized maps of plasma parameters obtained using the FORWARD toolset, with the 70%-contour of radio emission of the sixth burst at the first timestamp (12:34:06.8 UT) at the frequency of 72.26 MHz depicted on top of the 2D POS cuts. The left column represents, from top to bottom, plasma density, magnetic field, and the total plasma dynamic pressure. The right column represents, from top to bottom, the temperature, plasma beta, and the Alfven speed.

Figure 3.9: Coronal plasma parameters sampled from the 2D maps by the source centroids. The top panel shows (from left to right) the plasma density profiles from the MAS model, 2.5×Newkirk model, and the theoretical densities under the fundamental and harmonic assumptions, plasma temperature, and magnetic field. The bottom panel shows, from left to right, the total plasma dynamic pressure, Alfven speed, and plasma beta. The x-axis is inverted to demonstrate a progression of increasing radial distance from the Sun as the observer moves towards the right.

Chapter 4

Modeling and Forecasting of Solar Energetic Protons (SEP)

4.1   Introduction

Several models are available, or under development, for forecasting SEP, which use diverse approaches and serve different objectives. These models comprise computationally complex physics-based models, quick and simple empirical models, Machine Learning (ML)-based models, and hybrid models that combine different approaches and produce different types of outputs, including deterministic, probabilistic, categorical, and binary. Deterministic models always generate the same output without any randomness or stochastic components, such as predicting the SEP flux at a specific moment or the arrival time of SEP. On the other hand, probabilistic models provide a probability value that reflects the likelihood of an SEP event occurring. However, replicating SEP fluxes at a specific time is still a significant challenge for current models.

An excellent review on SEP models and predictive efforts was recently published by ?, which summarizes the majority of the existing models. For instance, ? introduced the Probabilistic Solar Particle Event Forecasting (PROSPER) model, which is incorporated into the Advanced Solar Particle Event Casting System (ASPECS)[1]. The PROSPER model utilizes a Bayesian approach and data-driven methodology to probabilistically predict SEP events for 3 integral energy channels >10, >30, and >100 MeV. The model's validation results indicate that the solar flare and CME modules have hit rates of 90% and 100%, respectively, while the combined flare and CME module has a hit rate of 100%. ? developed an empirical model to predict the peak intensity and spectra of SEP at 1 AU between 10 and 130 MeV, using data from multiple spacecraft. The model is tested on 20 SEP events and shows good agreement with observed values. The spatial distribution of SEP intensities was reconstructed successfully, and they found a correlation between SEP intensities and CME speed. ? extended the Particle Acceleration and Transport in the Heliosphere (PATH) model to study particle acceleration and transport at CME-driven shocks. They showed that the model can be used to obtain simultaneous calculations of SEP characteristics such as time-intensity profiles, instantaneous particle spectra, and particle pitch angle distributions at multiple heliospheric locations. Overall, results resemble closely those observed in situ near the Earth but also yield results at other places of interest, such as Mars, making it of particular interest to Mars missions. SPREAd-

---

[1]http://phobos-srv.space.noa.gr/

FAST (**??**) is a physics-based, data-driven framework that utilizes EUV observations and models to simulate SEP fluxes at 1 AU and to estimate energetic particle acceleration and transport to various locations in the inner heliosphere. It generates time-dependent histograms and movies distributing them through an online catalog. The accuracy and efficiency of the model were encouraging, but the highest energy fluxes showed disagreement with in situ observations by the SOHO/ERNE instrument. However, the framework has great potential for space weather science and forecasting.

In **?**, they used neural networks to provide probabilities for the occurrence of SEP based on soft X-rays data from 1988 to 2013. They obtained >85% for correct SEP occurrence predictions and >92% for correct no-SEP predictions. **?** described a consistent approach to making a binary prediction of SEP events using ML and conventional statistical techniques. The study evaluated various ML models and concluded that random forests could be the best approach for an optimal sample comprising both flares and CMEs. The most important features for identifying SEP were found to be the CME speed, width, and flare soft X-ray fluence. **?** employed ML techniques to anticipate the occurrence of a SEP event in an active region that generates flares. They utilized the Space-Weather MDI Active Region Patches (SMARP) dataset, which comprises observations of solar magnetograms between June 1996 and August 2010. The SMARP dataset had a success rate of 72% in accurately predicting whether an active region that produces a flare would result in a SEP event. Moreover, it provided a competitive lead time of 55.3 min in forecasting SEP events.

? introduced the Space Radiation Intelligence System (SPRINTS), a technology that uses pre- and post-event data to forecast solar-driven events such as SEP. It integrates automatic detections and ML to produce forecasts. Results show that SPRINTS can predict SEP with an 56% probability of detection and 34% false alarm rate. Nevertheless, the HESPERIA REleASE tools provide real-time predictions of the proton flux at L1 by using near-relativistic electrons as a warning for the later arrival of protons and have been set to operation(**?**). Historical data analysis indicates high prediction accuracy, with a low false alarm rate of approximately 30% and a high probability of detection of 63% (**?**).

Forecasting SEP is a critical task that serves operational needs and provides insight into the broader field of space weather science and heliophysics. As emphasized in previous works, a high precision forecasting model is urgently required to predict SEP flux within a period of time, given the risks associated with these events. This highlights the critical requirement for a dependable forecasting system that can mitigate the risks associated with SEP.

Scientists have been using physics-based and empirical models for decades to forecast SEP. However, these models have certain limitations. Physics-based models require accurate input data and underlying physical assumptions. In addition, the complexity of the physics involved and incorrect parameters may introduce uncertainties that can lead to inaccurate predictions. On the other hand, empirical models rely on historical data to make predictions. While they can be accurate sometimes, they may be unable to account for changes in physical conditions related to the acceleration and propagation of SEP, which can influence prediction accuracy. ML models, however, provide a different approach to SEP forecasting. These models can analyze vast amounts of data, learning patterns from the data that are used, and connections that may not be obvious to experts. Additionally, ML models can adapt to changes in underlying physical conditions, resulting in more accurate predictions as more data is collected; they also provide relatively rapid forecasts, which allows for incorporation into a real-time forecasting workflow.

In the upcoming sections, we will explore the limitations in accuracy that arise from dealing with an imbalanced dataset and low-resolution data. Specifically, the presence of intrinsic outliers in the time series data pertaining to SEP flux poses a significant challenge in modeling. These outliers correspond to occurrences of SEP events and, consequently, have an impact on the accuracy of predictions. Notably, they often lead to an underestimation of the SEP fluxes, primarily due to the predominance of relatively low values throughout the majority of the time interval.

In this paper, we present advanced deep learning models to forecast the daily integral flux of SEP over a 3-day forecasting window by using bi-directional long short-term memory (BiLSTM) neural networks, for 3 energy channels ($>$10, $>$30, and $>$60 MeV). Our models can forecast the time-dependent development of SEP events in different energy domains, which can be used to model the space radiation profiles using frameworks such as BRYNTRN **?** and GEANT4 (**?**). We describe the data selection and pre-processing in Section **??**. We present an overview on the analysis methods and the models implemented in Section **??**. Then we show the forecasting results in Section **??**. Finally the summary and implications are presented in Section **??**.

4.2   Data preparation

In this section, we describe the physical quantities, the types of inputs and their sources, as well as the outputs we are forecasting. Some of the technical terms used in this study are explained further in the appendices.

In order to capture the variability of solar activity, which modulates the SEP flux, we selected input physical quantities that describe both the interplanetary medium and solar activity. These input features can be categorized into two groups: remote signatures and in-situ measurements. The remote signatures consist of the F10.7 index, as well as the long-wavelength (X_L) and short-wavelength (X_S) x-ray fluxes. The F10.7 index represents the flux of solar radio emission at a wavelength of 10.7 cm, measured in solar flux units (sfu). To obtain the x-ray fluxes, we utilized 1- and 5-minute averaged data from the Geostationary Operational Environmental Satellite (GOES) database[2], specifically at long wavelengths (1 - 8 Å) and short wavelengths (0.5 - 4.0 Å).

The in-situ measurements encompass the near-Earth solar wind magnetic field and plasma parameters. These include the solar wind speed (in km s^{-1}), average interplanetary magnetic field (IMF) strength (in nT), and the integral SEP fluxes at three energy channels: $>$10, $>$30, and $>$60 MeV, which correspond to the GOES channels (in 1/cm^2 sec ster). These SEP fluxes were obtained from multiple spacecraft stationed at the first Lagrange point (L1) throughout the study period. In particular, the IMF and plasma data in the OMNI database are obtained from the IMP, Wind, and ACE missions, while the energetic particle fluxes are obtained from the IMP and GOES spacecraft[3].

To ensure a comprehensive dataset, we acquired hourly-averaged data covering a timeframe from December 1976 to July 2019, which spans the past four solar cycles. These data were sourced from the Space Physics Data Facility (SPDF) OMNIWeb database[4], hosted by the Goddard Space Flight Center. This database provides a wealth of information, including integral proton fluxes, as well as an extensive range of solar wind plasma

---

[2]`https://satdat.ngdc.noaa.gov/sem/goes/data/avg/`
[3]`https://omniweb.gsfc.nasa.gov/html/ow_data.html`
[4]`https://omniweb.gsfc.nasa.gov`

and magnetic field parameters. Lastly, the daily data on sunspot numbers were obtained from the Sunspot Index and Long-term Solar Observations (SILSO) archive[5], maintained by the World Data Center.

Figure ?? shows a plot for the timeseries data of all features. The top 3 panels are the logarithms of the SEP integral flux at the 3 energy channels (log_PF10, log_PF30, and log_PF60), then the sunspot number, the F10.7 index (F10_idx), the logarithms of the x-ray fluxes (log_Xs and log_Xl), the solar wind speed (Vsw), and the average magnitude of the IMF (avg_IMF). Throughout this paper, we adopt the convention that "log" refers to the common logarithm with a base of 10. The gray shades refer to the timespan of solar cycles. The blue, orange, and gold colors refer to the training, validation, and test sets, respectively. The data split method will be explained shortly.

Since the input SEP data have been compiled from various spacecraft, it may have artifacts even after processing. In particular, there are occasional jumps in the background level. There are also several day-long gaps in the OMNI solar wind parameters from the early 1980s to mid-1990s where only IMP 8 data are available and this spacecraft spent part of each orbit in the magnetosphere. We are reasonably confident that these issues do not influence the overall analysis significantly.

In deep learning applications, the dataset is split into 3 sets; namely the training set, the validation set, and the test set. The training set is usually the largest chunk of data that is used to fit the model. The validation set is a smaller chunk of data used to fine-tune the model and evaluate its accuracy to ensure it is unbiased. The test set is the out-of-sample data exclusively used to assess the final model when performing on unseen data (?).

After inspecting the correlation between the solar wind indices and the SEP integral fluxes in the OMNIWeb database, we chose the top-correlated features with the SEP flux. The correlations were made between the SEP fluxes and the individual parameters. Hence we took only timeseries of logarithms of the protons' integral flux at 3 energy channels (>10, >30, and >60 MeV), the timeseries of logarithm of the X-ray fluxes, the F10.7 index, the sunspot number, the solar wind speed, and the average strength of the IMF as input parameters to our model. The log of the SEP flux was used across the whole study. The correlation matrices for the training, validation, and test sets are shown in Figure ??. The X-ray and proton fluxes were converted into the logarithmic form because it was more convenient than the original form of data since the time series data were mostly quiet and had numerous sharp spikes, which correspond to solar events. Based on a previous experience with NNs (?), we found that training separate models for each target (output) feature can lead to better results. This is because a dedicated model for each output feature can more easily learn the interrelationships between input features and make more accurate predictions. Therefore, in our current study, we trained 3 separate models, each one targeting the logarithm of the protons integral flux at a specific energy channel.

In order to ensure consistency across all features, all durations of the time series data of the physical quantities were matched to be within the same time range. Subsequently, the dataset was resampled to obtain daily averaged data, resulting in a significant reduction of the dataset size by a factor of 24. This reduction facilitated expeditious training and yielded prompt results.

There were missing data values in the original dataset; for the B_{avg} (~10.7%), V_{sw} (~10.5%), F10.7-index (~0.08%), short-band x-ray flux (~8%), long-band x-ray flux
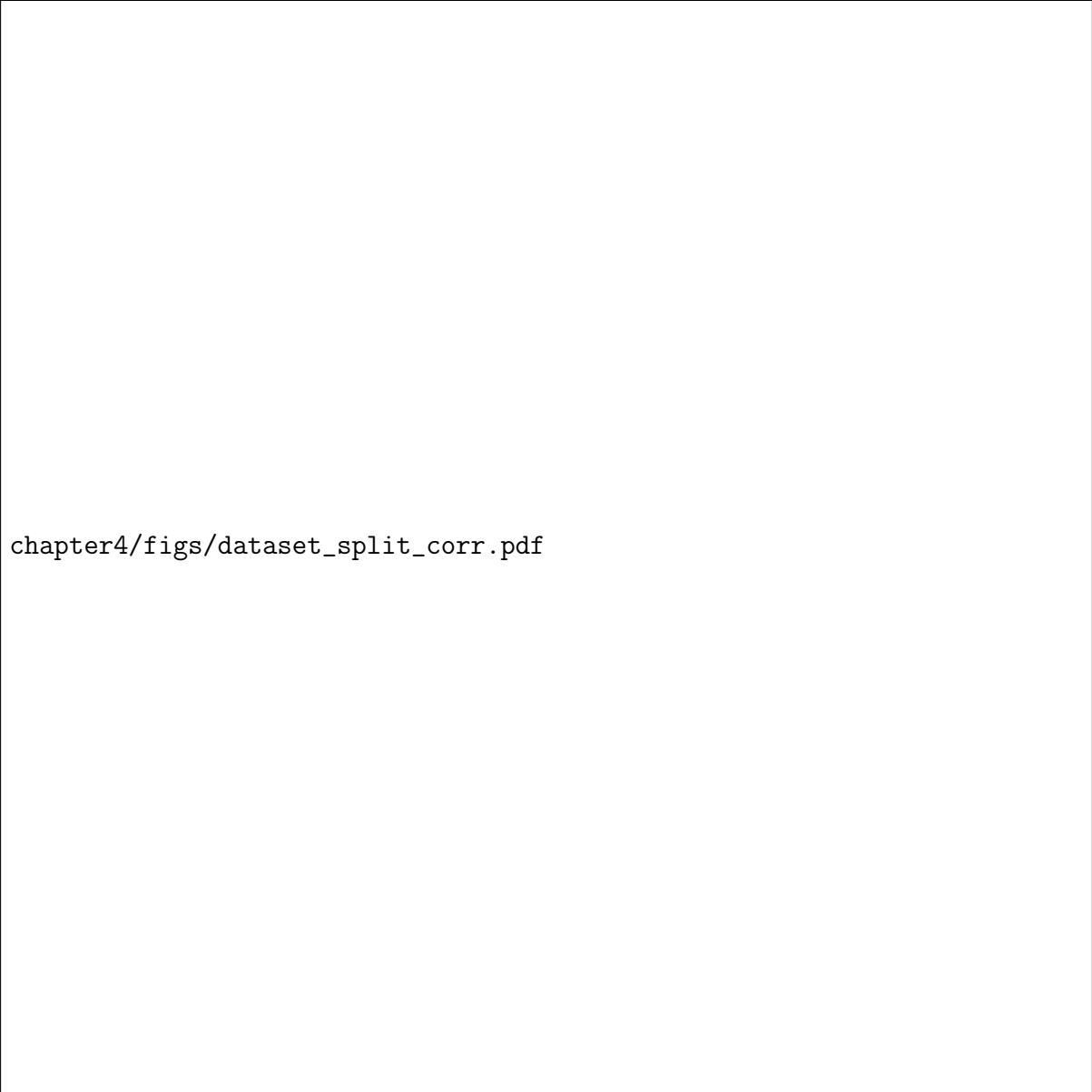
---

[5]https://www.sidc.be/silso/home

($\sim$9.8%), and proton fluxes ($\sim$4.3%). The data gaps were linearly interpolated.

Figure 4.1: Data splitting for all input features, showing the training, validation, and testing sets. Daily data from 1976-12-25 00:00 to 2019-07-30 00:00. The gray shading labels the solar cycles from SC21 to SC24.

In timeseries forecasting, it is a common practise to take a continuous set of data points from the main dataset to be the validation set and another smaller chunk of data to be the test set, for instance in ????. From our experiments, we got descent results when we applied the same data split method, but the results were a bit biased toward the end of the solar cycle 24 and the testing set was biased towards a quiet period. So, we adopted the 9-2-1 strategy, that is taking from each year 9 months to be added in the training set, 2 months to be added in the validation set, and 1 month to be added in the test set. This is applied over the $\sim$43 years of data (Fig. ??), which yields 74.29% of the data for the training set, 16.2% for the validation set, and 9.51% for the testing set. By doing so, we eliminated the need to do cross-validation and hence, made the training more efficient. It is worth to mention that the timeseries data must not be shuffled as that will break temporal and logical order of measurements, which must be maintained.

Figure 4.2: Correlation matrices show the correlation between the features in the training, validation, and test sets.

4.3 Method

In this section, we introduce the data analysis methods used in this work. We start with explaining the model selection phase, followed by a discussion of the bidirectional long-short-term memory (BiLSTM) neural network architecture. The technical terminologies are described in the appendices.

4.3.1 The Bi-LSTM Model

Recurrent neural networks (RNNs) that support processing input sequences both forward and backward are known as Bidirectional Long Short-Term Memory (BiLSTM) neural networks (**?**). Regular RNNs (**??**) depend on the prior hidden state and the current input to determine the output at a given time. The output of a BiLSTM network, on the other hand, is dependent on the input at a given moment as well as the previous and future hidden states. As a result, the network is able to make predictions using contexts from the past as well as the future. Hence, accuracy is improving. Each BiLSTM layer consists of two LSTM layers; a forward layer that processes the input sequences from the past to future, and a backward layer that processes the input sequences from the future to the past, as illustrated in Figure **??**, to capture information from both past and future contexts. The output from each layer is concatenated and fed to the next layer, which can be another BiLSTM layer or a fully connected layer for final prediction.

BiLSTM networks are advantageous than traditional LSTM networks in a variety of aspects (**???**). First, as we demonstrate in this study, they are excellent for tasks like timeseries forecasting, as well as speech recognition and language translation (**?????**) because they can capture long-term dependencies in the input sequence in both forward and backward directions. Second, unlike feedforward networks, BiLSTM networks do not demand fixed-length input sequences, thus being able to handle variable-length sequences better. Furthermore, by taking into account both past and future contexts, BiLSTM networks can handle noisy data. However, BiLSTM networks are computationally more expensive than regular LSTM networks due to the need for processing the input sequence in both directions. They also have a higher number of parameters and require more training data to achieve good performance.

The final dataset has 7 features, including the target feature, from December 25th 1976 to July 30th 2019, with a total of 15,558 samples (number of days). The training set has 11,558 samples, the validation set has 2,520 samples, and the test set has 1,480 samples.

The input horizon of 270 steps (30 days $\times$ 9 months) was used. A data batch size of 30 was used, which is the number of samples processed that result in one update to the model's weights (Appendix **??**). The model consists of 4 BiLSTM layers with 64 neurons each, and an output dense layer with 3 neurons, representing the output forecasting horizon. The total number of trainable parameters is 333,699. The number of training epochs was set to 50 because from experiments, the model stopped improving remarkably after almost 50 epochs. Thus, there was no need to waste time and computational resources to train the model for more than 50 epochs.

The *ModelCheckpoint* callback function was used to register the model version with the minimal validation loss. The *EarlyStopping* callback function was used to halt the model run when detecting overfitting, with a *patience* parameter of 7. *ReduceLROnPlateau* callback function was used to reduce the learning rate when the validation loss stops
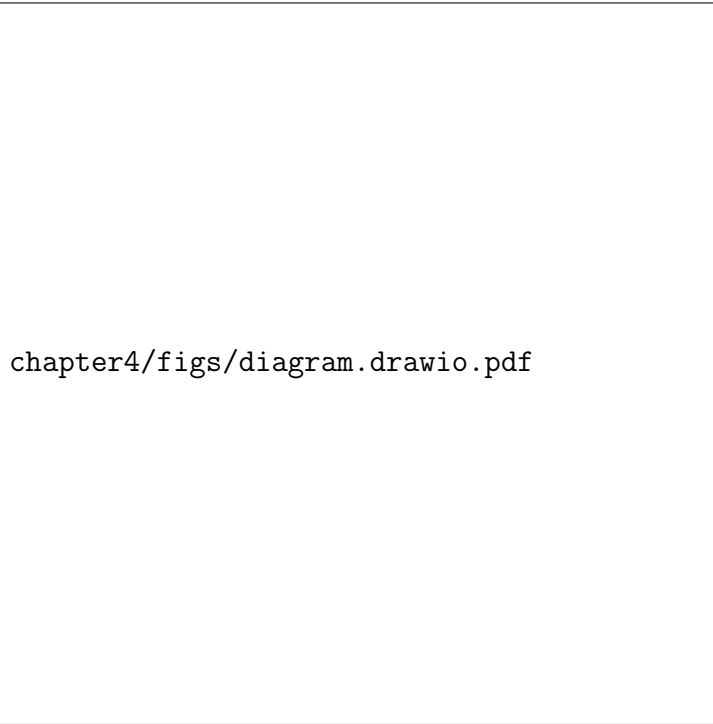
```
chapter4/figs/diagram.drawio.pdf
```

Figure 4.3: Architecture of a single BiLSTM layer. The blue circles at the bottom labeled by $(x_0, x_1, x_2, ..., x_i)$ are the input data values at multiple time steps. The purple circles, on the other hand, are the output data values at multiple time steps labeled by $(y_0, y_1, y_2, ..., y_i)$. The dark green and light green boxes are the activation units of the forward layer and the backward layer, respectively. The orange and yellow circles are the hidden states at the forward layer and the backward layer, respectively. Both the forward and backward layers composes a single hidden BiLSTM layer. The figure is adopted from **?**

improving, with a *patience* parameter of 5, a reduction factor of 0.1 and minimal learning rate of $1e^{-6}$.

4.3.2 Model Selection

To determine the most suitable model for our objective and provide justifiable reasons, we conducted the following analysis. First we examined the naive (persistence) model, which is very simplistic and assumes that the timeseries values will remain constant in the future. In other words, it assumes that the future value will be the same as the most recent historical value. That was the baseline. Next we examined the moving-average model, which calculates the future values based on the average value of historical data within a specific time widow. This gives a little bit lower error.
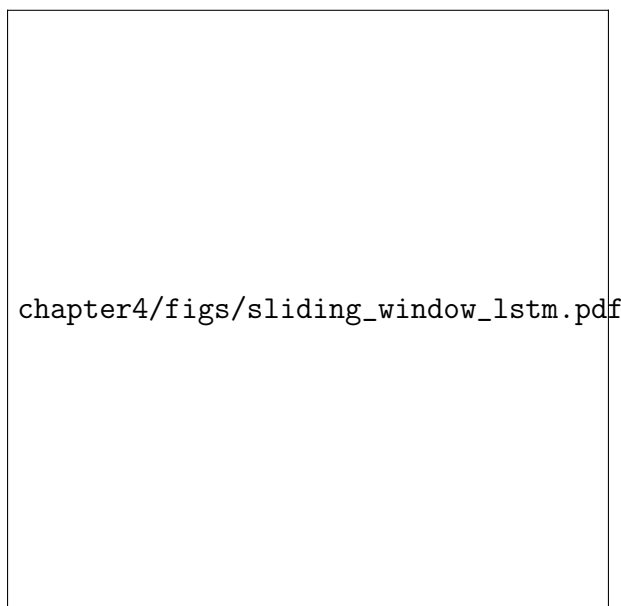

Figure 4.4: Illustration of the sliding window technique for a sample of 10 timesteps, where each number denotes a distinct time step. As an example here, the input horizon (blue color) length is 4 timesteps and the output horizon length is 3 timesteps. The input window slides 1 time step at a time across the entire data sequence to generate 4 distinct input and forecast horizon pairs. The purple, orange, and green colors of the output horizon represent 1-day, 2-day, and 3-day ahead forecasting, respectively. The timesteps of 1-day ahead forecasting across the data sequences are then concatenated into a single timeseries list that is called 1-day ahead prediction. The same for 2-day and 3-day ahead.

After that, we went towards the machine learning (ML)-based models. For all the ML models, we chose the Adaptive moment estimation (Adam) optimizer (**?**) as the optimization algorithm due to its minimal memory requirements and high computational efficiency as it is well-suited for applications that involve large number of parameters or large datasets. As a rule of thumb, we set the optimizer's learning rate to be 0.001 as it is usually recommended (**?**).

In order to prepare the data in a readable format to the ML models, we created a windowed dataset with an input horizon of 365 steps representing 1 year of data and an output horizon of 3 steps representing the forecast window of three days. We call this windowing method as Multi-Input Multiple Output (MIMO) strategy, in which the entire

output sequence is predicted in one shot. The MIMO strategy adopts the sliding window method that was mentioned in **?** in which each sequence is shifted by one step with respect to the previous sequence until reaching the end of the available data (Fig. **??**). This approach minimized the imbalance of active days, with high SEP fluxes, and quiet days.


```
chapter4/figs/models_benchmark.pdf
```

Figure 4.5: Benchmarking of 10 models, shows the Huber loss for the validation and test sets.

After experiments with different loss functions and evaluate their performance on our dataset, we chose the Huber function **??** as the loss function and the Mean Absolute Error (MAE) is used as the metric function to monitor the model performance. We used the Huber function because it is robust and combines the advantages of both Mean Squared Error (MSE) and MAE loss functions. It is less sensitive to outliers than MSE, while still being differentiable and providing gradients, unlike MAE. Since our data is noisy and contains outliers that may negatively impact the model's performance, the Huber loss function is a good choice.

We examined various neural network models to determine the optimal architecture for our task. Initially, we started with a simple linear model comprising of a single layer with a single neuron. However, this model did not yield satisfactory results. We then explored a dense ML model consisting of two hidden layers, each with 32 neurons and a *RelU* activation function. Next, we experimented with a simple RNN model with the same number of hidden layers and neurons. To find the optimal learning rate, we utilized the *LearningRateScheduler* callback function and discovered that a rate of $1.58e^{-4}$ under the basic settings minimized the loss. We proceeded to examine stateful versions of RNN, LSTM, and BiLSTM models with three hidden layers, each with 32 neurons and a learning rate of $1.58e^{-4}$. In addition, we explored a hybrid model that consisted of a 1-dimensional convolutional layer with 32 filters, a kernel size of 5, and a *RelU* activation function. We combined this with a two-hidden layer LSTM network with 32 neurons each and a learning rate of $1.58e^{-4}$. We experimented with *Dropout* layers but did not observe any significant improvement in the results. Finally, we evaluated a BiLSTM model with five hidden layers, 64 neurons each, and a learning rate of 0.001. Based on the evaluation of all the models on both the validation and test sets (Fig. **??** and Table **??**), we selected the BiLSTM model for further refinement. More details on the final model architecture and hyperparameters are explained in the Appendix **??**. Figure **??** presents a comparative analysis of the Huber loss within the validation and testing sets across the ten aforementioned models. We used several evaluation measures to assess our models since each metric provides valuable insights into the accuracy and performance of the forecasts (Appendix **??**), helping to identify areas for improvement and adjust the forecasting models accordingly.

4.4   Results and discussion

The benchmarking in Figure **??** showed that, in general, the ML-based methods were not much different. On the other hand, the persistence model and moving average model resulted in the highest errors compared with the ML-based models, and their results were close to some extent. As we see, the BiLSTM model performed the best over both the validation and test sets compared with the other models.

We developed and trained 3 BiLSTM models to forecast the integral flux of SEP, one model per energy channels. After the training was completed, we evaluated the performance of the models from the loss curve (Fig. **??**) using the Huber loss (the left panel) and the metric MAE (the middle panel). During the training, the learning rate was reduced multiple times via the *LearningRateScheduler* callback function (the right panel). The left panel quantifies the discrepancy between the model's predictions and the true values over time. It shows how the Huber loss function changes during the training iterations (Epochs) for the training and validation sets for the three energy channels so that each channel has one color. The middle panel shows how the model's metric MAE changes with training epochs. It is used to evaluate the performance of the trained model by measuring the average absolute difference between the model's predictions and the true values, providing a single numerical value that indicates the model's error at a given epoch. The right panel shows how the learning rate of the model's optimizer changes with epochs via the *LearningRateScheduler* callback function, which changes the learning rate based on a predefined schedule to improve training efficiency and convergence. The learning rate refers to the rate at which the model's parameters are updated during the

training process. We noticed that at the epochs where the learning rate has changed, there were bumps in the loss curves across all the energy channels, which is expected. This highlights the boundaries within which the learning rate yields better performance.

chapter4/figs/loss_curve_allenergies.pdf

Figure 4.6: *Left Panel* - The Huber loss vs. the number of training epochs for the BiLSTM model for the validation and test sets, for the 3 energy channels. *Middle Panel* - The mean absolute error (MAE); the model's metric vs. the number of training epochs. *Right Panel* - Shows how the learning rate of the Adam optimizer changes over the number of epochs.

From experimentation, we found that the batch size and the optimizer learning rate are the most important hyperparameters that have a strong influence on the overall model's performance (**?**). In addition, adding *dropout* layers as well as varying the number of hidden layers and hidden neurons resulted in only marginal improvements to the final model performance, while substantially increasing training time and requiring greater computational resources.

The term *batch size* refers to the number of data sequences processed in one iteration during the training of a ML model (**?**). Initially, a batch size of 64 was selected, however, we observed that the model produced better results when a batch size of 30 was used instead. This could be related to the Carrington rotation, which lasts for ∼27 days. There were ∼570 Carrington rotations between December 25th 1976 and July 30th 2019. Therefore, updating the model's weights after every Carrington rotation could be a reasonable choice for improving its performance. Figure **??** shows how good the model predictions are (on the y-axis) compared with the observations of the validation set (on the x-axis). The blue, orange, and gold colors refer to 1-day, 2-day, and 3-day ahead predictions, respectively. The top panel is for the >10 MeV channel, the middle panel is for the >30 MeV channel, and the bottom panel is for the >60 MeV channel. The left column is for the entire validation set, while the right column is for the observations points ≥10 proton flux units (pfu). That is the threshold value of proton flux as measured by the National Oceanic and Atmospheric Administration (NOAA) GOES spacecraft to indicate severity of space weather events caused by SEP.

We found that, overall, the models performed very well. The R correlation was >0.9 for all points of the validation set across the forecasting windows for the 3 energy channels. The R correlation was >0.7 for the observations points ≥10 pfu as well. However, the correlation between the modeled data and the observations exhibited a decline as the forecast horizon increased, in accordance with the anticipated result. To confirm the validity of the models, we performed the same correlation analysis between the modeled data and the observations of the out-of-sample test set (Fig. **??**), which was not given to the model. Again, we found a high correlation across the forecasting windows for the 3 energy channels. The points were more dispersed between 1 and 1.5 on the x-axis, which reflected in a bit lower correlation. This might be a limitation in the current version of the model between that range of SEP fluxes since the models underestimated the flux values within that range across all energy channels, possibly due to the relatively smaller training samples with fluxes above 10 pfu compared with the majority of the data.

In order to see the temporal variation of the correlation between the modeled data and the observations, we applied a rolling window of 90 steps (3 months × 30 days/month = 1 season) that shows the seasonal variation of the correlation, as shown in Figure **??**. Here, we show only the 1-day ahead predictions for the test set, for the 3 energy channels. We observe drops in the correlation factor synchronized with the transition between solar cycles (e.g., particularly between ∼1995 - 2000, which represents the declining phase of the solar cycle 22 and the rising phase of the solar cycle 23). This could be related to the fact that the low SEP fluxes during quiet times are more random and thus more difficult to forecast (**????**).

During periods of low solar activity, the forecasting of low SEP fluxes becomes more challenging due to their increased randomness. This difficulty arises from the reduced occurrence of conventional SEP drivers, such as solar flares and CMEs. Studies have suggested that the most significant solar eruptions tend to happen shortly before or after the solar cycle reaches its maximum (**?**). Additionally, sporadic increases in solar activity have been observed (**?**), which might contribute to the diminished correlations observed in our research. There is clearly some factor that is influencing the correlation during certain periods where there are no or only weak SEP events. However, it is not obvious which physical phenomena are the cause rather than, for instance, some artifact of the data. Understanding the interplay between these factors and their influence on SEP fluxes during periods of reduced solar activity remains a critical area of research. It would be
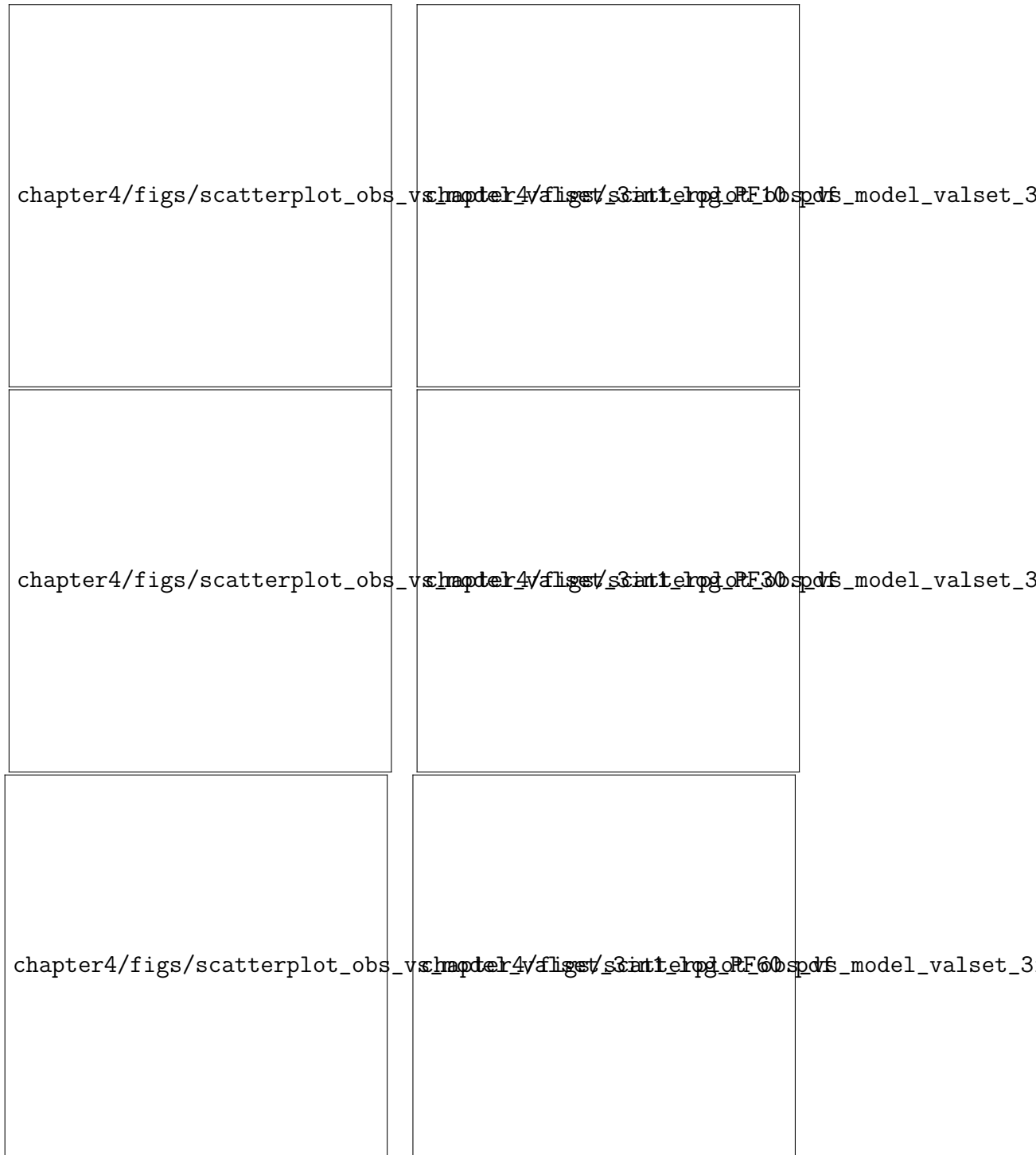
Figure 4.7: Correlation between the model predictions and observations for 1-day, 2-day, and 3-day ahead for >10 MeV (top panel), >30 MeV (middle panel), and >60 MeV (bottom panel). The panels in the left column represent all the points of the validation set, those in the right column represent all the observations points with daily mean flux ≥10 pfu.

chapter4/figs/scatterplot_obs_vs_model_tstset_3chapter4/figs/scatterplot_obs_vs_model_tstset_3

chapter4/figs/scatterplot_obs_vs_model_tstset_3chapter4/figs/scatterplot_obs_vs_model_tstset_3

chapter4/figs/scatterplot_obs_vs_model_tstset_3chapter4/figs/scatterplot_obs_vs_model_tstset_3

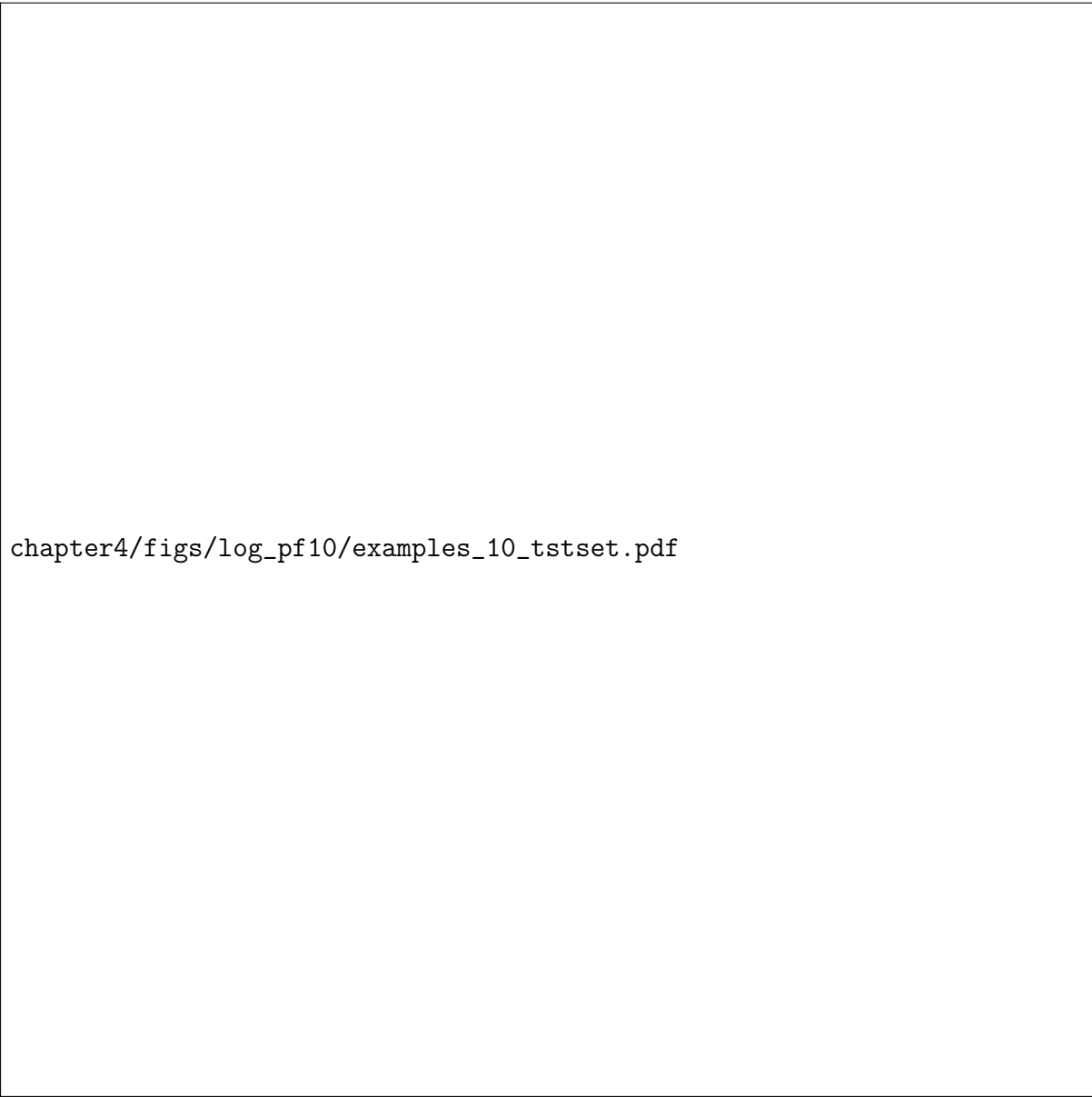Figure 4.8: Same as Figure **??** but for the test set.

interesting to find what is reducing the correlations, thus more investigation is needed.

Overall, the modeled data was correlated the most with observations at >60 MeV, then the second rank was for the >10 MeV channel, and the third rank was for the >30 MeV channel. That could be related to the relatively larger extent of drops in correlation at the >30 MeV channel. The decline in correlation at the >30 MeV channel is consistent with the findings of **?**. A summary of the performance results of the models for both the validation set and test set is presented in Table **??**.

chapter4/figs/comparison_crosscorr_tstset_1day_3channels.pdf
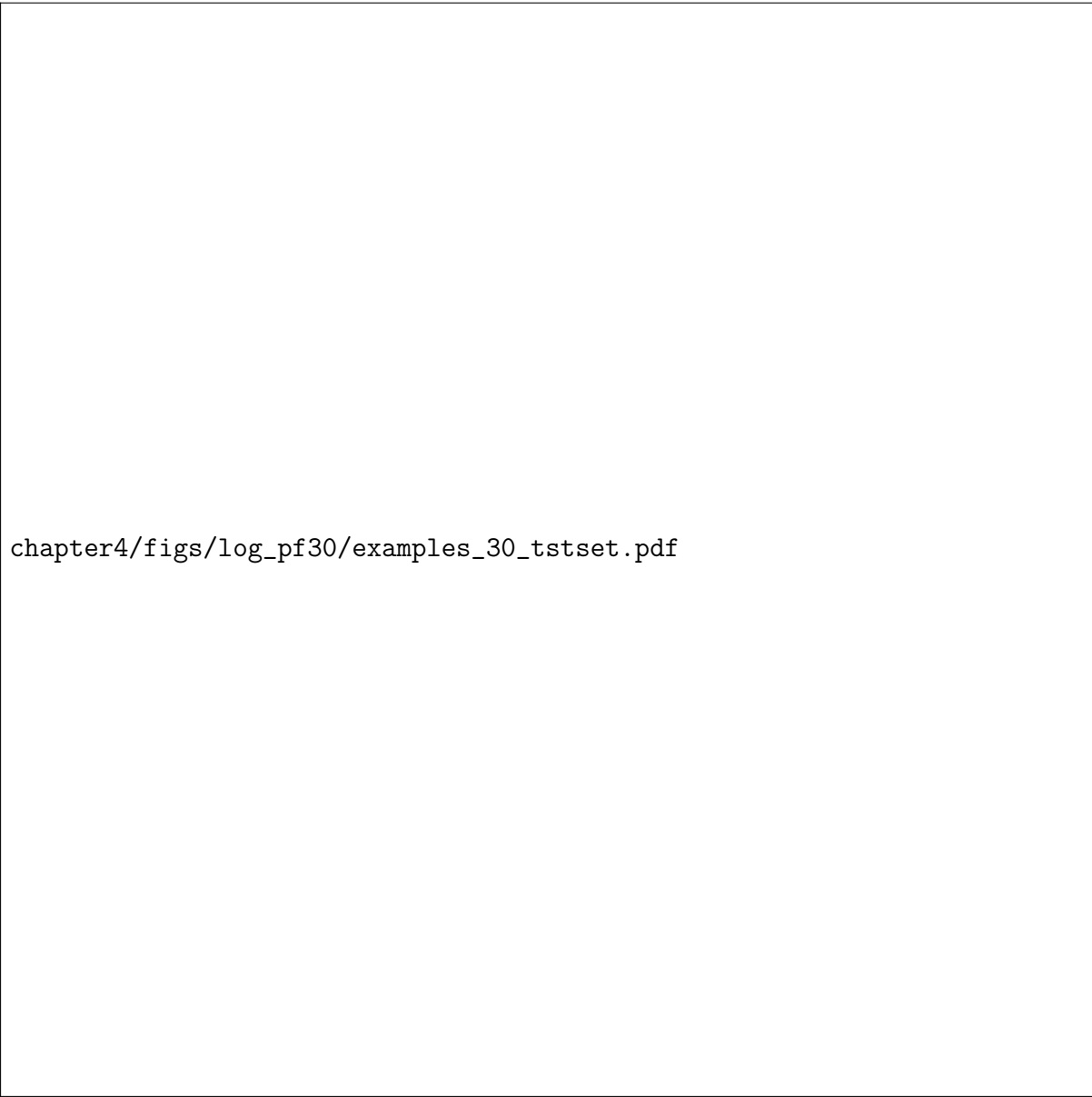
Figure 4.9: Comparison between the model outputs and observations of the test set for the 3 energy channels. In addition to the rolling-mean window correlation for 1-day ahead predictions.

From the visual inspection of the test set examples (Fig. **??**, **??**, and **??**), we found that the predicted onset time, the peak time, and end times of SEP events were highly correlated with those of the observations, which implies that the model captured the temporal variations, as well as the trends in SEP flux.

chapter4/figs/log_pf10/examples_10_tstset.pdf

Figure 4.10: The model's forecasts for the out-of-sample testing set for the >10 MeV channel are shown at forecast horizons of 1 day, 2 days, and 3 days ahead, using samples of data from December in selected years mentioned in the top-left side of the plots.

chapter4/figs/log_pf30/examples_30_tstset.pdf

Figure 4.11: The model's forecasts for the out-of-sample testing set for the >30 MeV channel are shown at forecast horizons of 1 day, 2 days, and 3 days ahead, using samples of data from December in selected years mentioned in the top-left side of the plots.
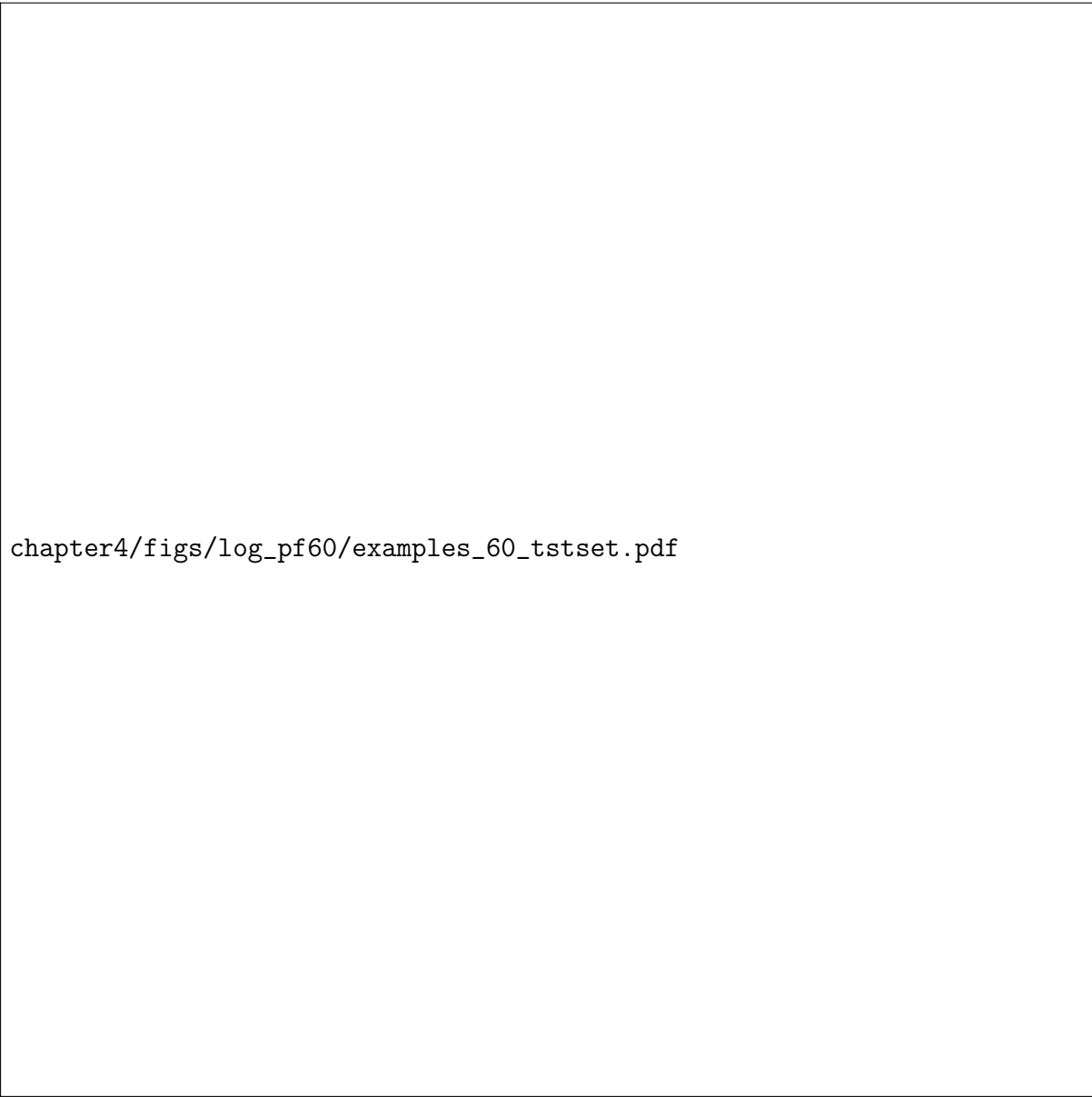
chapter4/figs/log_pf60/examples_60_tstset.pdf

Figure 4.12: The model's forecasts for the out-of-sample testing set for the >60 MeV channel are shown at forecast horizons of 1 day, 2 days, and 3 days ahead, using samples of data from December in selected years mentioned in the top-left side of the plots.

Table 4.1: Summary of the performance results of the models for the validation and test sets.

	Validation Set								
	log PF >10 MeV			log PF >30 MeV			log PF >60 MeV		
Model Loss	0.0016			0.0010			0.0009		
Model Metric	0.0329			0.0232			0.0218		
	1-Day	2-Day	3-Day	1-Day	2-Day	3-Day	1-Day	2-Day	3-Day
MAE	0.061	0.091	0.125	0.053	0.079	0.098	0.052	0.069	0.086
MSE	0.013	0.028	0.054	0.010	0.031	0.055	0.009	0.027	0.047
RMSE	0.114	0.168	0.233	0.098	0.176	0.234	0.097	0.164	0.217
MAPE	22.156	28.104	34.721	13.039	18.590	22.735	10.036	13.994	16.731
	Test Set								
	log PF >10 MeV			log PF >30 MeV			log PF >60 MeV		
Model Loss	0.0014			0.0011			0.0010		
Model Metric	0.0333			0.0283			0.0250		
	1-Day	2-Day	3-Day	1-Day	2-Day	3-Day	1-Day	2-Day	3-Day
MAE	0.072	0.099	0.125	0.053	0.088	0.107	0.045	0.066	0.081
MSE	0.015	0.030	0.050	0.009	0.029	0.048	0.007	0.020	0.034
RMSE	0.121	0.172	0.224	0.094	0.170	0.218	0.082	0.141	0.184
MAPE	30.135	37.498	48.139	20.599	34.300	40.803	12.358	20.504	25.305

To get further insight into the model's performance, we conducted an assessment of various skill scores, including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Additionally, skill score ratios such as Probability of Detection (POD), Probability of False Detection (POFD), False Alarm Rate (FAR), Critical Success Index (CSI), True Skill Statistic (TSS), and Heidke Skill Score (HSS). Detailed descriptions of these skill scores can be found in Appendix **??**. To extract individual SEP events from the test dataset, we implemented a threshold-based clustering algorithm. This algorithm uses the NOAA/SWPC warning threshold value of 10 pfu for the E $\geq$10 MeV channel. Upon analysis, we identified the number of detected SEP events for each output forecasting window and calculated the skill scores (Table **??**). In the true test set, we identified 12 SEP events.

The evaluation of the model revealed notable trends as the length of the output forecasting window increased. The POD and CSI exhibited a declining pattern, indicating a reduced ability of the model to accurately detect and capture positive events (SEP occurrences) as the forecasting horizon extended further into the future. This suggests that the model's performance in identifying and capturing true positive instances diminishes with longer forecasting windows. Moreover, the POFD demonstrated an increasing trend, indicating an elevated rate of false positive predictions as the forecasting horizon lengthened. The model's propensity to generate false alarms rose with the lengthening forecasting window, leading to incorrect identification of non-events as positive events. Consequently, the TSS and HSS exhibited decreasing values, signifying a deterioration in the model's overall skill in accurately capturing and distinguishing between positive and negative instances. Overall, our skill scores are comparable with those reported by previous studies (Table **??**). Although the UMASEP model does better than ours (i.e., has a higher POD), our FAR is much lower, thus, making fewer false alarms than the UMASEP model.

Table 4.2: Confusion matrix for the energy channel ≥10 MeV predictions in the test set.

E >10 MeV	No. events	TP	TN	FP	FN
1-day ahead	15	21	1441	2	13
2-day ahead	13	14	1441	2	20
3-day ahead	5	5	1443	0	29

Table 4.3: Comparing the skill scores with previous models. The dashed entries mean the data is unavailable (**?** for more details).

Model		POD	FAR	TSS	HSS	POFD	CSI	Accuracy	Precision
	1-Day	0.618	0.087	0.531	0.732	0.001	0.583	0.99	0.913
Our BiLSTM model	2-Day	0.412	0.125	0.287	0.553	0.001	0.389	0.985	0.875
	3-Day	0.147	0	0.147	0.252	0	0.147	0.980	1
UMASEP-10 (**?**)		0.822	0.219	—	—	—	—	—	—
PCA (**?**)		0.587	0.245	—	0.65	—	—	—	—
SPARX (**?**)		0.5	0.57	—	—	0.32	0.3	—	—
SPRINTS (**?**)		0.56	0.34	—	0.58	—	—	—	—
REleASE (**?**)		0.63	0.3	—	—	—	—	—	—

4.5 Conclusions

Forecasting the SEP flux is a crucial task in heliophysics since it affects satellite operations, astronaut safety, and ground-based communication systems. It is a challenging task due to its non-linear, non-stationary, and complex nature. Machine learning techniques, particularly neural networks, have shown promising results in predicting SEP flux. In this study, we developed and trained BiLSTM neural network models to predict the daily-averaged integral flux of SEP at 1-day, 2-day, and 3-day ahead, for the energy channels >10 MeV, >30 MeV, and >60 MeV. We used a combination of solar and interplanetary magnetic field indices from the OMNIWeb database for the past 4 solar cycles as input to the model. We compared the models with baseline models and evaluated them using the Huber loss and the error metrics in Appendix **??**.

The data windowing method we used, based on the MIMO strategy, eliminates the need to feed the output forecast as input back into the model and that allows to do forecasting relatively far into the future while maintaining decent results (e.g., the MSE is ranging between 0.007 and 0.015 for 1-day forecasting in the test set, compared to an MSE of 0.236 for a persistence model. See Table **??**). The results show that the model can make reasonably accurate predictions given the difficulty and complexity of the problem. The MSE was ranged between 0.009 and 0.055 for the validation set, and between 0.007 and 0.05 for the test set. The correlations between the observations and predictions were >0.9 for the validation and test sets (Fig. **??** and Fig. **??**). Nevertheless, the mean temporal correlation was ∼0.8 for the test set (Fig. **??**). Although our models performed well, we observed a relatively large discrepancy between the predictions and the observations in the >30 MeV energy band.

The findings of this study underscore the challenges encountered by the forecasting model in accurately predicting SEP data over longer time periods. As the length of the output forecasting window increased, the model's ability to detect true positives and its overall skill in differentiating positive and negative instances diminished. Additionally, the model displayed an elevated rate of false negative predictions, indicating an increased tendency to generate misses as the forecasting horizon extended. These results highlight

the importance of carefully considering the appropriate forecasting window length for SEP data to ensure the model's optimal performance. Our skill scores generally align with those from previous works (Table ??). There are variations in the metrics' values across different studies, highlighting the complexities and nuances associated with each study. Nevertheless, it is important to acknowledge that the statistical significance of the results in this study is limited due to data averaging. Future studies should consider incorporating hourly data, as this is likely to result in a greater number of identified events. The model can provide short-term predictions, which can be used to anticipate the behavior of the near-Earth space environment. These predictions have important implications for space weather forecasting, which is essential for protecting satellites, spacecraft, and astronauts from the adverse effects of solar storms.

Multiple techniques exist for identifying the optimal combination of hidden layers and neurons for a given task such as empirical methods, parametric methods, and the grid search cross-validation method, which we will explore in future work. The observed reduction in correlation necessitates further investigation to determine its origin, whether stemming from tangible causal factors or potential aberrations within the model or data. We plan to expand upon this work by performing short-term forecasting using hourly-averaged data. This extension will involve integrating additional relevant features such as the location and area of active regions and coronal holes on the Sun.

BiLSTM networks are particularly useful for tasks involving sequential data such as timeseries forecasting. Given their capacity to handle input sequences in both directions in time and capture long-term dependencies, they are valuable in a broad range of applications. Nonetheless, one should carefully consider their data requirements and computational complexity before adopting them. Our results emphasize that the use of deep learning models in forecasting tasks in heliophysics are promising and encouraging, as pointed out by ?.

This work is a stepping stone towards real-time forecasting of SEP flux based on the public-available datasets. As an extension, we are currently working on developing a set of models that deliver near-real time prediction of SEP fluxes at multiple energy bands, multiple forecasting windows, with hourly-averaged data resolution, with a more sophisticated model architecture, as well as more features that address the state of solar activity more comprehensively. We plan to extend the analysis to include more recent data from solar cycle 25, in order to improve the accuracy of the models. In conclusion, our study highlights the potential of using BiLSTM neural networks for forecasting SEP integral fluxes. Our models provide a promising approach for predicting the near-Earth space environment too, which is crucial for space weather forecasting and ensuring the safety of our space assets. Our findings contribute to the growing body of literature on the applications of deep learning techniques in heliophysics and space weather forecasting.

Appendix A

Appendix

A.1 Terminologies

In this section, we introduce the main concepts related to machine learning which are presented in this paper.

- **Cross-validation**: A technique used to evaluate the performance of a machine learning model by dividing the data into subsets and assessing the model on different combinations of these subsets.

- **Input Horizon**: The number of previous time steps considered as input to a model for time series forecasting. It represents the length of the historical sequence used for predictions.

- **Batch Size**: The number of samples processed together in a single iteration of the training algorithm. It affects training speed and memory requirements.

- **Updating the Model's Weights**: The process of adjusting the parameters of a neural network based on training data to minimize the difference between predicted and true outputs. The model's weights represent the parameters that are learned during the training process.

- **Loss**: A function that quantifies the difference between predicted and actual outputs. It guides the optimization process during training.

- **Minimum Validation Loss**: The lowest value achieved by the loss function on a validation dataset during training. It indicates the most accurate predictions on unseen data.

- **Overfitting**: When a model performs well on training data but fails to generalize to unseen data due to memorizing training examples instead of learning underlying patterns.

- **Learning Rate**: A hyperparameter that determines the step size at each iteration of the optimization algorithm during training. It affects learning speed and convergence. A high learning rate can cause the training process to converge quickly, but it may also result in overshooting the optimal solution or getting stuck in a suboptimal solution. On the other hand, a very low learning rate can make the training process slow, and may struggle to find the optimal solution.

- **Reducing the learning rate when the validation loss stops improving**: This concept involves adjusting the learning rate dynamically during the training process. When the validation loss reaches a plateau or stops improving, it indicates a suboptimal point. By reducing the learning rate, the model can take smaller steps in weight space, potentially finding a better solution. This technique, known as learning rate scheduling or learning rate decay, is commonly used to fine-tune the model's performance.

- **Patience**: A parameter used in training to determine the number of epochs to wait for an improvement in validation loss before stopping the training process.

- **Patience Parameter of 7**: In the context of early stopping, training will be stopped if the validation loss does not improve for 7 consecutive epochs.

- **Adam Optimizer**: A popular optimization algorithm in deep learning that combines Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSprop) to achieve efficient optimization.

- **Optimal Architecture**: The best configuration of a neural network, including the number of layers, neurons, and other choices, for optimal performance on a specific task.

- **Hyperparameters**: Parameters set before training a model that control the learning algorithm's behavior, such as learning rate, batch size, and activation functions.

- **Layer**: A building block of a neural network that performs specific operations on input data. Includes input, hidden, output, fully connected, convolutional, recurrent, activation, and dropout layers. Here is a description for each layer:

  - **Input Layer**: The first layer of a neural network that receives raw input data. It passes the input to subsequent layers for further processing. The number of nodes in the input layer is determined by the dimensionality of the input data.

  - **Hidden Layers**: Intermediate layers between the input and output layers. They perform computations on the input data and capture higher-level representations or abstractions. Hidden layers are not directly exposed to the input or output.

  - **Output Layer**: The final layer of a neural network that produces model predictions or outputs based on computations from preceding layers. The number of neurons in the output layer depends on the problem being solved, such as regression or classification.

  - **Fully Connected Layer (Dense Layer)**: Each neuron in this layer is connected to every neuron in the previous layer. It allows information flow between all neurons, enabling complex relationships to be learned.

  - **Convolutional Layer**: Commonly used in Convolutional Neural Networks (CNNs) for analyzing grid-like data, such as images. It applies convolution operations using filters or kernels to learn spatial patterns or features.

– **Recurrent Layer**: Used in Recurrent Neural Networks (RNNs) to process sequential data. These layers have feedback connections that allow information to be passed from one step to the next, capturing temporal dependencies and maintaining memory of past inputs.

– **Activation Layer**: Applies a non-linear function to the output of a layer, introducing non-linearity into the neural network. Activation functions like Sigmoid, Hyperbolic Tangent (tanh), or Rectified Linear Unit (ReLU) determine neuron outputs based on weighted inputs.

– **Dropout Layer**: A regularization technique commonly used in deep learning models. It randomly sets a fraction of outputs from the previous layer to zero during training, preventing overfitting and improving generalization.

Layers play a crucial role in the information processing and learning capabilities of neural networks. The arrangement and combination of different layers determine the network's architecture and ultimately its ability to solve specific tasks.

- **Stateful**: A property of Recurrent Neural Networks (RNNs) where the hidden state is preserved between consecutive inputs, allowing the network to have memory.

- **Neuron**: A computational unit in a neural network that receives input, applies weights, and passes the result through an activation function to produce an output.

- **Hidden Neuron**: A neuron in a hidden layer of a neural network that performs intermediate computations.

- **Callback Function**: A function used during model training to perform specific actions at certain points or conditions, such as saving the best model, adjusting learning rates, or early stopping.

- *LearningRateScheduler* **Callback Function**: A function used in training to dynamically adjust the learning rate at specific points based on a predefined schedule or function. It improves training efficiency and convergence by allowing the model to make finer adjustments as it approaches the optimal solution.

A.2 Mathematical Representation of the LSTM NN Model

The computations inside one LSTM cell can be described by the following formulas (**?**):

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{A.1a}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{A.1b}$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{A.1c}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{A.1d}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{A.1e}$$

$$h_t = o_t \odot \tanh(C_t) \tag{A.1f}$$

where x_t is input data at time t. The input gate i_t determines which values from the updated cell states (candidate values) $\tilde{C}_t$ should be added to the cell state. It also takes

into account the current input x_t and the previous output h_{t-1}, and is passed through a sigmoid activation function. $\tilde{C}_t$ represent the candidate values that are added to the cell state at time t. The forget gate activation vector f_t at time step t, which determines how much of the previous cell state should be retained. The cell state C_t at time t is updated based on the forget gate, input gate, and candidate values. The output gate o_t at time t determines how much of the cell state should be output. The output vector h_t at time t is calculated based on the cell state and the output gate values. h_{t-1} is the output vector at the previous time step $t-1$. W_f, W_i, W_c, W_o are the weight matrices for the input vector x_t. U_f, U_i, U_c, U_o are the weight matrices for the output vector h_{t-1}. b_f, b_i, b_c, b_o are the bias vectors. The symbol $\odot$ denotes a pointwise multiplication. The sigmoid function σ is used as the activation function for the gate vectors, and the hyperbolic tangent function tanh is used for the candidate values and the output vector.

A.3 Evaluation Metrics

To evaluate the model performance, we used the following equations:

$$L_\delta(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2, & \text{if } |y - \hat{y}| \leq \delta, \\ \delta(|y - \hat{y}| - \frac{1}{2}\delta), & \text{otherwise} \end{cases} \tag{A.2a}$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{A.2b}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{A.2c}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{A.2d}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} |\frac{y_i - \hat{y}_i}{\hat{y}_i}| \tag{A.2e}$$

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{A.2f}$$

where y is the true value, $\hat{y}$ is the predicted value, and δ is a threshold in the Huber loss function that controls the trade-off between the mean squared error (MSE) and the mean absolute error (MAE). In this paper, it was set to 0.1, which was selected based on several experiments.

MSE is the mean squared error, which measures the difference between predicted and actual values by calculating the average of squared differences. It provides a measure of the average squared magnitude of the errors in your forecasts, which can be useful in penalizing larger errors more heavily than smaller errors.

MAPE is the mean absolute percentage error, which measures the difference between predicted and actual values by calculating the average of absolute differences. It provides a measure of the average magnitude of the errors, allowing to evaluate the overall accuracy of your forecasts.

RMSE is the root mean squared error, which measures the difference between predicted and actual values by taking the square root of the average of squared differences.

It provides a measure of the accuracy of the forecasts in the same units as the original data, allowing to evaluate the magnitude of errors in the same scale as the data.

MAPE is the mean absolute percentage error, which measures the accuracy of a forecast by calculating the average of absolute percentage errors. It provides a measure of the accuracy of the forecasts in percentage terms, allowing to evaluate the magnitude of errors relative to the actual values. MSE, MAE, RMSE, and MAPE are often used in regression analysis to assess the accuracy of the model's predictions.

Finally, R is the Pearson correlation coefficient, which measures the strength and direction of the relationship between two continuous variables, and can provide an indication of the extent to which changes in one variable may be related to changes in the other.

A.4 Model Configuration

The configurations for the ML models shown in Figure ?? and their performance on the validation set and the test set for the SEP integral flux ≥10 MeV are presented in Table ??. The batch size was set to be 64 and the number of training epochs was set to be 100. The *EarlyStopping* callback function, with a *patience* of 10, is used to help prevent overfitting during the training process by stopping training when the monitored metric has stopped improving for a certain number of epochs. The *patience* parameter controls how many epochs the training will continue without improvement before it is stopped. This is useful because if the validation loss stops getting better, the model has probably overfitted the training data and is not generalizing effectively to new data. By stopping the training early, we can avoid wasting time and resources on further training that is unlikely to improve the model's performance.

We used the *ModelCheckpoint* callback function to save the best weights of the model during training so that they can be reused later. The *LearningRateScheduler* callback function allows to dynamically adjust the learning rate of the model during training using a function passed to it that will be called at the beginning of each epoch, and it should return the desired learning rate for that epoch. It can be useful when training deep neural networks, as it allows for a higher learning rate in the early stages of training when the model is still far from convergence, and a lower learning rate as the model approaches convergence, which can help it to converge more accurately. The downside might be the longer training time.

Table A.1: Configuration of the ML model. (1) refers to the error value for 1-day forecasting. Same for (2) refers to 2-day forecasting, and (3) for 3-day forecasting. *In the 1D-CNN layer, 32 filters, a kernel size of 5, and strides of 1 were used.

Model Architecture	No. of Hidden Layers	No. of Hidden Neurons	Activation Function	Batch Size	Learning Rate	Epochs	Callbacks Functions	Validation Set				Testing Set			
								MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
Linear	–	–	–	64	0.001	100	EarlyStopping	0.312	0.141	0.376	87.883	0.143	0.045	0.211	60.689
Dense ML	2	32	ReLU	64	0.001	100	EarlyStopping	0.262 (1)	0.118 (1)	0.344 (1)	132.580 (1)	0.400 (1)	0.281 (1)	0.530 (1)	238.898 (1)
								0.275 (2)	0.138 (2)	0.372 (2)	132.004 (2)	0.395 (2)	0.286 (2)	0.535 (2)	234.704 (2)
								0.290 (3)	0.166 (3)	0.407 (3)	129.288 (3)	0.392 (3)	0.294 (3)	0.542 (3)	230.896 (3)
Simple RNN	2	32	Tanh	64	0.001	100	EarlyStopping ModelCheckpoint	0.143 (1)	0.035 (1)	0.187 (1)	70.990 (1)	0.178 (1)	0.052 (1)	0.228 (1)	69.624 (1)
								0.171 (2)	0.063 (2)	0.251 (2)	68.694 (2)	0.171 (2)	0.071 (2)	0.266 (2)	78.075 (2)
								0.264 (3)	0.118 (3)	0.343 (3)	72.505 (3)	0.200 (3)	0.084 (3)	0.289 (3)	67.416 (3)
Stateful RNN	3	32	Tanh	64	1.58e^{-4}	100	LearningRateScheduler EarlyStopping	0.203 (1)	0.060 (1)	0.244 (1)	56.390 (1)	0.155 (1)	0.039 (1)	0.197 (1)	59.988 (1)
								0.305 (2)	0.131 (2)	0.362 (2)	81.028 (2)	0.223 (2)	0.079 (2)	0.281 (2)	71.679 (2)
								0.349 (3)	0.173 (3)	0.416 (3)	82.819 (3)	0.223 (3)	0.084 (3)	0.289 (3)	64.159 (3)
Stateful LSTM	3	32	Tanh	64	1.58e^{-4}	100	EarlyStopping	0.095 (1)	0.021 (1)	0.146 (1)	40.335 (1)	0.098 (1)	0.020 (1)	0.141 (1)	41.781 (1)
								0.151 (2)	0.048 (2)	0.220 (2)	48.937 (2)	0.134 (2)	0.042 (2)	0.205 (2)	57.860 (2)
								0.174 (3)	0.076 (3)	0.275 (3)	55.662 (3)	0.166 (3)	0.071 (3)	0.267 (3)	68.025 (3)
Stateful Bi-LSTM	3	32	Tanh	64	1.58e^{-4}	100	EarlyStopping	0.149 (1)	0.043 (1)	0.207 (1)	58.151 (1)	0.170 (1)	0.049 (1)	0.221 (1)	71.059 (1)
								0.190 (2)	0.074 (2)	0.272 (2)	60.154 (2)	0.211 (2)	0.090 (2)	0.300 (2)	92.727 (2)
								0.249 (3)	0.120 (3)	0.347 (3)	67.988 (3)	0.229 (3)	0.108 (3)	0.329 (3)	87.049 (3)
1D-CNN LSTM	3	32 (5,1)*	ReLU Tanh	64	1.58e^{-4}	100	EarlyStopping	0.108 (1)	0.027 (1)	0.165 (1)	41.164 (1)	0.098 (1)	0.023 (1)	0.151 (1)	51.732 (1)
								0.146 (2)	0.051 (2)	0.226 (2)	47.512 (2)	0.138 (2)	0.047 (2)	0.217 (2)	68.376 (2)
								0.177 (3)	0.078 (3)	0.279 (3)	53.087 (3)	0.156 (3)	0.067 (3)	0.259 (3)	69.338 (3)

All the calculations and model runs were implemented under the framework of TensorFlow 2.3.0 (**?**) in Python 3.6.13. The models were executed on Ubuntu 20.04.1 LTS OS with $4 \times$ GPUs (NVIDIA GeForce RTX 2080 Ti, 11019 MiB, 300 MHz). According to the Keras API guide (**?**), the requirements to use the cuDNN implementation are the activation function must be set to *tanh* and the recurrent activation must be set to *sigmoid*. We also set the seed number to 7 across all the model runs to maintain reproducibility.

Stateful RNNs can be difficult to work with when using callbacks in Keras because their hidden state must be manually managed across mini-batch updates. When training a stateful RNN in Keras, the hidden state is carried over from the previous epoch and can cause problems with certain callbacks, such as *EarlyStopping* or *ModelCheckpoint*. To work around this issue, one can use stateless RNNs or manually reset the hidden state at the end of each epoch, but this can be complex and prone to errors.

A.5   Skill Scores

Skill scores and ratios are commonly used in evaluating the performance of classification models, particularly in binary classification tasks. They provide insights into the model's ability to correctly predict positive and negative instances. Here is a brief description of each skill score and ratio, along with their formulas:

- **True Positive (TP)**: The number of data points or intervals correctly identified as positive by the model. It represents instances where both the model and the ground truth indicate the presence of an event.

- **True Negative (TN)**: The number of intervals correctly identified as negative by the model. It represents instances where both the model and the ground truth indicate the absence of an event.

- **False Positive (FP)**: The number of intervals incorrectly identified as positive by the model. It occurs when the model predicts an event, but the ground truth indicates its absence.

- **False Negative (FN)**: The number of intervals incorrectly identified as negative by the model. It occurs when the model fails to detect an event that the ground truth indicates its presence.

- **Accuracy**: Represents the proportion of correct predictions out of total predictions.
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{A.3}$$

- **Precision**: Represents the proportion of positive predictions that are actually positive.
$$Precision = \frac{TP}{TP + FP} \tag{A.4}$$

- **Probability of Detection (POD) or Recall**: Represents the model's ability to correctly identify positive instances.
$$POD = \frac{TP}{TP + FN} \tag{A.5}$$

- **Probability of False Detection (POFD)**: Measures the model's tendency to falsely predict positive instances when the ground truth indicates their absence.

$$POFD = \frac{FP}{FP + TN} \tag{A.6}$$

- **False Alarm Rate (FAR)**: Indicates the ratio of false positive predictions to the total number of positive instances.

$$FAR = \frac{FP}{FP + TP} \tag{A.7}$$

- **Critical Success Index (CSI)**: Measures the model's ability to correctly predict both positive and negative instances.

$$CSI = \frac{TP}{TP + FP + FN} \tag{A.8}$$

- **True Skill Statistic (TSS)**: Takes into account both the model's ability to detect positive instances and its ability to avoid false alarms.

$$TSS = POD - FAR \tag{A.9}$$

- **Heidke Skill Score (HSS)**: Evaluates the model's performance by comparing it with random chance. It takes into account the agreement between the model's predictions and the observed data, considering both true positive and true negative predictions.

$$HSS = \frac{TP + TN - C}{T - C} \tag{A.10}$$

where

$$T = TP + TN + FP + FN$$
$$C = \frac{(TP + FP)(TP + FN) + (TN + FP)(TN + FN)}{T}$$

A.6

A.7