



**UNIVERSITÀ
DEGLI STUDI
DI TRIESTE**

UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVI CICLO DEL DOTTORATO DI RICERCA IN FISICA

TSRS 2.0 AND SOLAR RADIO DATA FOR GNSS SPACE WEATHER APPLICATION

Specific Subject Area : FIS\01 EXPERIMENTAL PHYSICS

**PH.D. STUDENT
ADRIANA MARCUCCI**

**PH.D. PROGRAM COORDINATOR
PROF. FRANCESCO LONGO**

**THESIS SUPERVISOR
PROF. MAURO MESSEROTTI**

**THESIS CO-SUPERVISOR
PROF. ANNA GREGORIO**

ACADEMIC YEAR 2023/2024

*To my parents,
for their endless love*

Abstract

Solar radio emission, a key indicator of solar activity, is crucial for monitoring and predicting Space Weather, particularly for its impact on technological infrastructures, such as global navigation satellite systems (GNSS). The Trieste Solar Radio System 2.0 (TSRS 2.0) is a new radio spectropolarimeter dedicated to operational Space Weather monitoring and forecasting, installed in Basovizza (Trieste, Italy) that will provide both solar radio indices and synoptic images of the solar disk. While awaiting the availability of data from the newly-installed TSRS 2.0, other solar radio data sources have been utilised for synoptic images and time series analysis. Dedicated algorithms for time series prediction, such as F10.7 index, have been developed and tested. Additionally, K-band radio images from the INAF SunDish project can probe the chromospheric magnetic field of the Active Regions (AR) through the detection of gyro-resonance spectral components related to flare events. Through correlation statistics analysis between sporadic anomalous AR spectrum and strong flares we discuss the relevant outcomes and the effectiveness of the developed methodologies. In the framework of INAF National Solar Radio Network, the upcoming Solaris Project of smart Solar monitoring system at high radio frequencies, based on single-dish imaging techniques, will be the only Solar facility offering continuous monitoring at 100 GHz, nearly 20h/day in Antarctica with optimal sky opacity.

Contents

Introduction	1
1 Solar Radio Emissions and Space Weather Effects on GNSS	7
1.1 Basics of Solar Radio Emissions	8
1.1.1 Free-Free Emission	13
1.1.2 Gyromagnetic Emission	14
1.1.3 Active Regions and Quiet Sun	16
1.1.4 Space Weather Phenomena from Eruptive Events	17
1.1.5 Solar Radio Bursts (SRB)	19
1.2 Space Weather Effects on GNSS	22
1.2.1 SRB Direct Effects	24
1.2.2 Ionospheric TEC Variability	25
1.2.3 Ionospheric Scintillation	26
2 The Trieste Solar Radio System 2.0	29
2.1 The TSRS 2.0 Architecture	30
2.2 The TSRS 2.0 Observation Modes and Data Products	35
2.3 Commissioning Test Planning	40
2.4 TSRS 1.0 Heritage and Design of the TSRS 2.0 Archive	42
2.5 Perspective for Automatic SRB Analysis and Detection	45
3 Time Series Forecasting for Space Weather Applications	50
3.1 Time Series Theoretical Background	51
3.2 Time Series Forecasting Models	52
3.2.1 Long-Short Time Memory Neural Networks	53
3.2.2 Self-Attention and Transformers	55
3.3 From Empirical Mode Decomposition (EMD) to Fast Iterative Filtering (FIF)	58

CONTENTS

3.4	Forecasting of Solar Radio F10.7 Index	60
3.4.1	FIF-LSTM-MHA Univariate Model	63
3.4.2	Multivariate NN-base Model	77
3.4.3	Long-Term Solar Activity Forecasting Model	82
3.5	External Drivers Effects for Ionospheric TEC Model	87
3.5.1	Feature Ranking and ML Algorithms	88
3.5.2	Ranking Results and Model Performances	94
4	Diagnostics by Synoptic Solar Radio Imaging	103
4.1	Single-Dish Solar Radio Imaging (SunDish)	105
4.1.1	Active Regions Spectral Index	108
4.1.2	Flare Statistical Correlation Analysis	111
4.1.3	Case Studies Discussion	122
4.2	SOLARIS: Solar Imaging System for Continuous Monitoring and Space Weather Applications	130
	Conclusions	137
	Appendix A	140
	Bibliography	157
	Acknowledgements	158

Introduction

Space weather refers to the dynamic and ever-changing conditions in space, particularly in the Earth's magnetosphere and ionosphere, driven by the activity of the Sun and interactions between the solar wind and the Earth's magnetic field. Much like terrestrial weather, space weather encompasses a wide range of phenomena, from solar flares and Coronal Mass Ejections (CMEs) to geomagnetic storms and auroras [176].

The core of space weather is the Sun, a dynamic and variable star that emits a continuous stream of charged particles known as the solar wind. Solar wind, along with occasional eruptions of highly energetic material such as CMEs, interacts with the Earth's magnetosphere, triggering a cascade of complex physical processes.

Solar flares, sudden releases of magnetic energy in the Sun's atmosphere, and CMEs can produce intense bursts of radio emission across a broad range of frequencies, providing valuable insights into the dynamics of solar eruptions and their potential impacts on the Earth's environment.

This range of frequencies offers valuable diagnostic potential because two of the natural frequencies of the atmosphere of the Sun, the electron plasma frequency and the electron gyrofrequency, belong to the radio band.

Solar radio emissions, one of the key indicators of solar activity, play a crucial role in monitoring and understanding space weather [146].

By studying solar radio emissions, we can track the evolution of solar activity, try to predict the occurrence of space weather events, and assess their potential impact on technological infrastructures, [140, 20, 102] in particular on positioning systems. Radio emissions are among the earliest observed signatures of an eruptive event in the corona. Warnings of major solar disturbances and their impact on Earth is an important element in order to mitigate space weather hazards [174].

Global Navigation Satellite Systems (GNSSs), such as the Global Positioning System (GPS), Galileo, and GLONASS, rely heavily on satellite-based navigation and timing signals for a wide range of applications, including aviation, transportation,

INTRODUCTION

agriculture, and telecommunications. However, GNSS signals are susceptible to disturbances in the Earth's ionosphere caused by space weather events. Variations in electron density in the ionosphere can cause signal delays, errors in positioning, and disruptions to GNSS-based services [137].

Therefore, understanding and predicting space weather is crucial for mitigating these effects and ensuring the reliability and accuracy of GNSS-based systems. Space weather forecasting provides advance warning of ionospheric disturbances, allowing operators to take proactive measures to mitigate their impact on GNSS operations [169]. By incorporating real-time space weather data into GNSS algorithms and receiver designs, the resilience of GNSS systems to space weather effects can be improved enhancing their performance in challenging environments [164, 101].

The investigation of the Sun environment and its emission mechanisms in the radio domain is of great importance for the advancement of our understanding of some of the complex solar phenomena, their origin and connections. This research has applications in both operational space weather and fundamental physics.

Solar radio emissions serve as a robust diagnostic instrument for the examination of coronal plasma and for the prediction of the direct influence of solar radio noise on the quality and reliability of technological systems, including radio-based communications and services.

With the advancements in observational instruments and development of forecasting models, as addressed in this thesis, it is possible to gain deeper insights into solar radio emissions. This paves the way for a more profound comprehension of the Sun-Earth connection and its effects on our technological society. [59]. Indeed, to better understand the involved physical phenomena and their implications on technological infrastructures, constant monitoring of the solar radio emission is fundamental. Moreover, the development of automated systems for the recognition and characterisation of solar radio events, as sources of interference and as precursors of perturbative solar phenomenologies, is essential for operational space weather purposes.

Usually, for the limited availability of dedicated solar radio instruments and their observational features, only nowcasting has been possible to date.

The original project of this PhD work was the study of solar radio data for Space Weather applications, in particular for radio frequency interferences with GNSS receivers, by the use of the new Trieste Solar Radio System 2.0 (TSRS 2.0): a dedicated solar radio spectropolarimeter expected to provide synoptic radio im-

INTRODUCTION

ages of the Sun at 18 GHz, in each of the two circular polarisation components (Left-Handed (LHCP) and Right-Handed Circular Polarisation (RHCP)); solar radio spectra in the L-Band (1-2 GHz) and solar radio indices at selected frequencies, in order to identify in advance solar radio sources that can be candidate as GNSS receiver jammers or as precursors of perturbative solar phenomenology.

TSRS 2.0, has been recently installed in the Basovizza Observing Station but, unfortunately, in addition to significant delays, we are still facing serious hardware issues, it is not fully operational and we could not begin the calibration yet.

Thanks to its peculiar and unique operational capabilities, the system will be an interesting diagnostic tool and will play a fundamental role in the Solar Radio Weather framework.

While awaiting the availability of TSRS 2.0 data, other sources of solar radio data have been utilised in the present work, as a test-bed for the development of tools for space weather applications.

The need for an instrument like TSRS 2.0 arises from the absence of other instruments dedicated to solar observation, for both operational and research purposes, that provide real-time L-band radio fluxes with polarization information. Most of the available solar radio data, even when considering an international framework, come from sporadic observations (as they are not dedicated to the Sun, such as Medicina or Sardinia Radio Telescope) or lower (such as LOFAR) and higher (e.g. Iberian Solar Radio Telescope - SSRT) frequency bands, or provide only complementary information to the indices, such as solar imaging and dynamic spectra (e.g. Callisto Solar Radio Spectrometer).

The existing systems already in operation, such as the RSTN (Radio Solar Telescope Network) operated by the US Air Force Weather Agency (US AFWA), does not provide polarisation information, which instead are fundamental for GNSS Space Weather application.

Moreover, TSRS 2.0 solar radio indices will be made freely available to the scientific community in near-real-time while, although providing alerts in case of solar radio bursts (SRBs) to the users, RSTN raw data have a latency period of several days.

This thesis is intended to provide knowledge and methodological tiles towards a preliminary forecasting pipeline, by the use of different data, such as solar radio indices from different sources and solar radio images from other promising single-dish antenna contributing to the National Radio Space Weather Network.

Dedicated models for time series analysis and forecasting have been implemented

INTRODUCTION

by the use of machine learning techniques. In particular, new forecasting algorithms for the F10.7 solar radio index will be presented and compared, in order to develop operational tools that will be ready for use once the TSRS 2.0 data become accessible. Indeed, F10.7 index, crucial for monitoring solar activity, holds significance across various fields, as it is also the input driver for several operational models, e.g orbital determination of LEO satellites and debris.

Leveraging datasets such as TSRS 2.0 promises to enhance the accuracy of our predictions, thereby amplifying the usefulness of these operational tools in a variety of applications.

In our results, by the use of feature importance technique and Neural Network (NN)-based models, the significance of F10.7 prediction has been confirmed and highlighted, as a main driver for operational forecasts of ionospheric vertical Total Electron Content (vTEC) models.

These global vTEC models are already employed in operational Space Weather forecasts, further emphasizing the importance of accurate and timely data inputs like those by TSRS 2.0.

TSRS 2.0 is also supposed to provide full disk radio images of the low solar corona, on a daily basis, as it will be a dedicated instrument for solar activity operational monitoring.

High frequency radio band surprisingly lacks of Solar observations, which could, however, provide valuable information on characterizing the solar Active Regions (ARs) and their spectra characterisation.

K-band solar radio mapping can probe the chromospheric magnetic field of the Active Regions through the detection of gyro-resonance spectral components anticipating flare events.

To fill this research gap, in this work, synoptic solar radio maps performed in the K-band (18-26 GHz) with a non-dedicated radio telescope in Medicina (Bologna, Italy), have been analysed. Solar images, performed from 2018 to 2023, have been obtained in the framework of the Sundish project, that helped to establish the INAF Italian Radio Telescope Network as a non-dedicated solar radio imaging facility, filling a frequency gap, that presently exists in the worldwide solar radio monitoring scenario.

Enhanced AR magnetic fields determine a spectral flattening compared to pure free-free emission due to the addition of a steeper gyro-resonance component.

A statistical correlation analysis between gyro-resonance enhancement, spectral indices and several features of more than 120 solar Active Regions (ARs) have been

INTRODUCTION

deeply investigated.

The studies of free-free and gyro-magnetic emission components present in most solar ARs, provide possible applications to solar flare forecast, contributing to the Space Weather warning/alert network. The solar disk maps acquired by Medicina, showed a significant fraction of AR spectra with relatively flatter profiles compared to the thermal bremsstrahlung emission, suggesting the potential existence of gyro-magnetic emission components that peaks in this range or in neighbourhood frequencies. In the analysis, we considered several parameters for each AR, such as the maximum brightness temperature, the magnetic complexity and the flare intensity that has been originated. V-Stokes polarimetric information have been also taken into account, as high degree of polarization is expected in sporadic ARs dominated by a strong gyro-magnetic component, helpful in disentangling gyro-magnetic contributions from free-free emission.

Developing solar flare prediction is one of the essential methods to prevent possible storm-like geo-effectiveness and related hazards for GNSS users, a core research in the perspective of operational Space Weather research.

In the last decade, the number of published approaches to flare forecasting using photospheric observations has proliferated. Although a lot of effort has been devoted to flare prediction [151, 55, 56, 192], operational near-real-time flare forecasting systems remains a challenge.

The study of the dynamical phenomenology of the solar atmosphere would require a multi-wavelengths and continuous monitoring to follow the evolution of the Active Regions, Coronal Holes and other peculiar features, as well as energetic radio burst emissions, flares and Coronal Mass Ejections. To further expand our studies on solar observations, we are planning to observe at higher frequencies in the near future, thanks to the Solaris project, which has the aim to observe the Sun up to 100 GHz in Antarctica in a permanent observatory.

TSRS 2.0, SunDish and Solaris are offering tools for Space Weather forecast and to contribute to prompt response in the frame of national and international networks.

The content of this thesis is structured as follows:

The first Chapter of this work gives an introduction on solar radio astronomy, focusing on the emission mechanisms in the radio domain and polarisation and their application for Space Weather, although there is still a gap of observation of the Sun at radio frequencies and many of the physical phenomena are not clear yet.

In the second Chapter we describe the Trieste Solar Radio System 2.0 (TSRS

INTRODUCTION

2.0), fully dedicated to the solar radio surveillance and soon in operation under the management of INAF – Astronomical Observatory of Trieste.

The third Chapter is therefore dedicated to the description of time series forecasting models that we have implemented, by the use of machine learning techniques.

In the fourth Chapter, we present an original statistical analysis investigating the potential correlation between ARs spectral features in the K-band and the occurrence of flares within a 30-hour time window. Through correlation statistics, we investigate and discuss the sensitivity and robustness of this flare forecast method, for the future perspective of coupling it with other multi-messenger Space Weather proxies.

In summary, the present work is intended to provide knowledge and methodological tiles towards a preliminary forecasting pipeline, by the use of different data, such as solar radio indices from different sources and solar radio images from other promising single-dish antenna contributing to the INAF National Radio Space Weather Network.

Chapter 1

Solar Radio Emissions and Space Weather Effects on GNSS

Space Weather, a multifaceted domain of study, delves into the intricate interplay between the Sun and the Earth's space environment. Governed by the dynamic behaviour of the Sun, Space Weather encompasses a broad spectrum of phenomena, from solar flares and coronal mass ejections (CMEs) to solar wind variations and geomagnetic storms. These solar activities phenomena can exert profound effects on the Earth's magnetosphere, ionosphere, and atmosphere, with far-reaching implications for technological infrastructures, satellite operations, and even terrestrial climate patterns.

At the heart of understanding Space Weather lies the study of solar radio emissions. These emissions, originated by the Sun, provide a wealth of information about the energetic processes occurring in the solar atmosphere. It serves as a crucial diagnostic tool for monitoring solar activity and forecasting Space Weather events. Solar radio emissions are generated through various mechanisms, including thermal bremsstrahlung radiation, gyroresonance emission, and non-thermal processes involving accelerated particles. Each mechanism offers unique insights into the dynamics of the solar atmosphere and contributes to our understanding of solar eruptive events and their impact on the space environment.

One of the key arenas where solar radio emissions exert their influence is the Earth's ionosphere. Situated in the upper atmosphere, the ionosphere is ionised by solar radiation and interacts with solar radio emissions. These emissions can ionise the neutral atmosphere, leading to enhancements in ionospheric electron density. Such enhancements can, in turn, influence the propagation of radio waves through the

ionosphere, affecting communication and navigation systems on Earth and in space. Understanding the intricate relationship between solar radio emissions and the ionosphere is paramount for advancing our knowledge of Space Weather phenomena. By unravelling the complexities of these interactions, scientists can improve Space Weather forecasting capabilities and develop more robust mitigation strategies to safeguard critical infrastructure and assets in space and on Earth.

In the following sections, we will delve deeper into the mechanisms of solar radio emissions, explore their effects on the ionosphere, and discuss their broader implications for Space Weather applications, with an overview of the profound impacts of solar activity on our technologically reliant society.

1.1 Basics of Solar Radio Emissions

Solar radio emissions encompass a wide range of electromagnetic radiation emitted by the Sun across a broad spectrum of radio frequencies, therefore they offer a unique opportunity to probe the structure and dynamics of the solar corona, the outermost layer of the Sun's atmosphere. These emissions serve as a crucial diagnostic tool for probing the physical processes occurring in the Sun's atmosphere, the density, temperature, and magnetic field strength of the corona, providing valuable constraints for theoretical models of solar activity and magnetohydrodynamic simulations [204], as a powerful diagnostic tool for studying the Sun's atmosphere and understanding solar dynamics and Space Weather phenomena [23].

Radio astronomy developed during the last fifty years, that is relatively early compared to X-ray, ultraviolet and infrared techniques. The reason lies in the fact that most radiation in the radio band penetrates down to the ground level.

In Fig. 1.1, we show how the Earth atmosphere affects the incoming radiation from space as a function of its wavelength. We see that the atmosphere is completely transparent for waves between a few centimeters and 10 m. Below 0.5 mm the radiation is almost totally absorbed. At the other end of the radio spectrum, radiation with wavelengths greater than about 20 m are totally blocked by the Earth's ionosphere. To collect radio photons and concentrate them, radio telescopes are used [167]. The radio emission received by a radio telescope is measured in terms of its intensity I_ν , which is the power received per square meter, per frequency unit Hz, and per solid angle unit. Integrating I_ν over the solid angle subtended by the

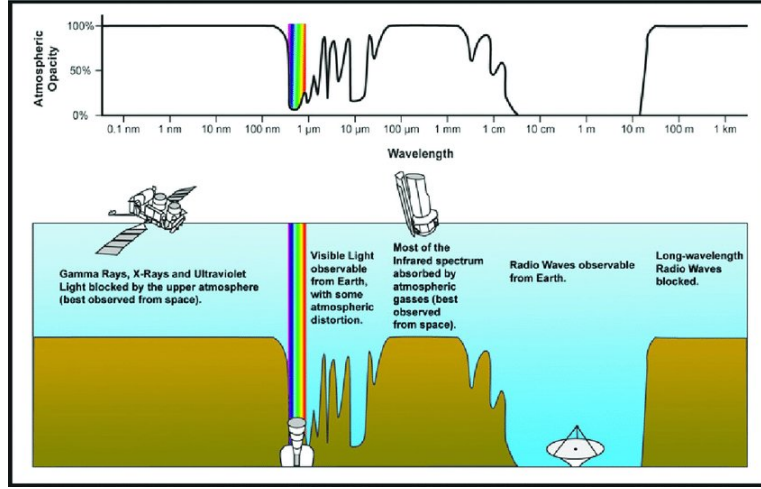


Figure 1.1: Atmospheric opacity as a function of wavelength [47]

emitting source we get a quantity, S_ν , called the spectral flux density:

$$S_\nu = \int I_\nu d\Omega \quad (1.1)$$

The unit of $1 \text{ W m}^{-2} \text{ Hz}^{-1}$ is quite large for most of the radio emitting astrophysical objects, and a new flux density unit has been defined and called Jansky (Jy), with $1 \text{ Jy} = 10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$.

In solar radio astronomy, and since the radio signal from the Sun is quite strong (few orders of magnitude stronger than the brightest non-solar radio sources in microwaves), we commonly use solar flux units (SFU) define as $1 \text{ SFU} = 10^4 \text{ Jy} = 10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$.

The expression of I_ν radiated by a black-body at temperature T , at the frequency ν , is given by the Planck's law:

$$I_\nu = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{k_B T}} - 1} \sim \frac{2\nu^2 k_B T}{c^2} \quad (1.2)$$

where we have used the Raleigh–Jeans approximation for radio photons, i.e., $h\nu \ll k_B T$, where h is the Planck's constant, k_B is the Boltzmann constant and c is the speed of light in vacuum.

Another quantity largely used in radioastronomy is the brightness temperature T_B , which is the equivalent temperature a black-body should have to radiate an intensity

I_ν . We thus have a relation between T_b and S_ν given by:

$$S_\nu = \frac{2k_B\nu^2}{c^2} \int T_b d\Omega \quad (1.3)$$

where Ω is the angle subtended by the emitting source.

Although the brightness temperature has been defined for thermal emission of a black-body we can also use it for non-thermal emitting sources. Non-thermal emission may be due to accelerated particles with energy E , and in this case, we shall consider T_b as an effective temperature T_{eff} given by the kinetic temperature E/k_B . The nonthermal emissions are generally associated with solar flares and originate from all levels of the solar atmosphere between the lower chromosphere (millimetre and centimetre waves) and the outer corona to heights of several solar radii (metre and decametre waves).

Solar radio emission produced from electrons can be either incoherent or coherent. An incoherent radiation mechanism is a process where particles behave independently of each other. There is no phase relation between the emitted photons, that is no coherence. In this case, the resulting brightness temperature cannot exceed the effective temperature T_{eff} , that is the source temperature for a thermal process, or the kinetic temperature for a non-thermal process. Thus, for an incoherent non-thermal emission, due to e.g. MeV electrons as it is often observed during solar flares, T_b is limited by the mean particle energy, such that $T_b \leq E/k_B \cdot 10^{10}$ K. During a coherent process emitting particles can behave as a whole, generating photons that are in phase [146].

Thus, a coherent emission is not the result of individual particle emission, but rather a collective process for example triggered by waves when certain instabilities develop in magnetized plasma. The brightness temperature resulting from a coherent emission mechanism can thus largely exceed the mean particle energy, and can reach a few 10^{15} K as for some decimetric solar radio bursts.

A fundamental parameter in plasma physics that must be introduced is the plasma frequency for electrons (f_{pe}) that characterises the natural oscillation frequency of charged particles in a plasma. It is defined as the frequency at which the plasma exhibits collective oscillations due to the restoring force provided by the electromagnetic field. It represents the frequency at which the combined effects of the charged particles and the electromagnetic field lead to resonant behaviour. The

plasma frequency for a certain plasma species is given by:

$$f_{pe} = \frac{1}{2\pi} \sqrt{\frac{n_e e^2}{m_e \epsilon_0}} = 9 \times 10^{-3} \sqrt{n_e [cm^{-3}]} MHz \quad (1.4)$$

where n_e is the number density of electrons, e is the elementary charge, ϵ_0 is the vacuum permittivity, m_e is the electron mass.

The plasma frequency depends on the electron density of the plasma and is typically expressed in units of hertz (Hz). It serves as a critical parameter in understanding various plasma phenomena, including wave propagation, dispersion, and absorption. Plasma emission is the most known coherent process to produce radio emission.

In the presence of a magnetic field B , another characteristic frequency is the gyrofrequency: the frequency with which an electron gyrates around magnetic field lines due to the Lorentz force, often referred to Ω_{be} in the literature ($\Omega_e = 2\pi f_{be}$), can be expressed using

$$f_{be} = \frac{eB}{m_e} = 2\pi 2.8 \times 10^6 B[G] Hz \quad (1.5)$$

Thermal Bremsstrahlung radiation and non-thermal processes are the two main mechanisms responsible for continuous emissions in various astrophysical contexts. The former results from the thermal motion of electrons in the solar corona. As electrons interact with ions, they emit radiation across a broad range of frequencies, producing a continuum spectrum. The intensity and spectral characteristics of thermal emissions provide key information about the temperature, density, and composition of the solar plasma [211]. The latter, such as gyrosynchrotron emission, arise from the acceleration of energetic particles in regions of intense magnetic activity, such as solar flares and coronal mass ejections (CMEs). These emissions exhibit complex spectral features and are typically observed at lower radio frequencies [49].

The observed spectral features of a solar radio emission are determined by both the mechanism responsible for the emission and the propagation conditions of electromagnetic waves in the solar corona. Thus, the observed dynamic spectra of the solar radio emissions (e.g time-frequency representation of radio signals emitted by the Sun) reveal essentially the physical conditions of the solar corona during the emission periods.

The magnetic field significantly influences all radiative processes, making its quan-

titative description crucial for accurate measurement. While qualitative information on the corona's magnetic field is readily obtainable through EUV or soft X-ray images, the most reliable method for measuring the magnetic field in the corona is through its influence on the radio emission [12].

In most cases, the corona can be defined as a cold magnetized plasma, and the magnetoionic theory is used to describe the propagating electromagnetic waves. Due to the presence of a magnetic field an electromagnetic wave propagates in different modes because the magnetised plasma is birefringent: the extraordinary mode (X) and the ordinary (O) mode, z-mode, and whistler mode. The x- and o-modes are of observational interest because they can propagate from the source to infinity. The z- and whistler modes are prevented from doing so by stopbands in the refractive index [19].

For most applications in solar radiophysics, the propagation of the x- and o-mode waves can be described by either the quasi-longitudinal (QL) or the quasi-transverse (QT) approximation (propagation almost parallel and almost perpendicular to the magnetic field, respectively). The detected radiation is said to be polarised when the emission in one of the two modes dominates that in the other [12].

For observational purposes, it is easier to describe radiation using the Stokes parameters I, Q, U, and V. Under conditions of QL propagation, we get $Q = U = 0$, and the x- and o-mode waves are circularly polarised in opposite senses. Thus, we obtain $V = I_x - I_o$, and the polarisation is circular having the sense of the dominant mode. In the case of QT propagation, we obtain $Q = I_x - I_o$ and $U = V = 0$, and the polarisation is linear [146]. As highlighted in the paper Alikandrakis et al. [12] it is important to remind that the polarisation of the two modes depends only on the properties of the medium in which they propagate and not on the emission mechanism. Therefore, the polarisation characteristics are expected to change along the path of the waves, reflecting the local values of the plasma parameters (related to the magnetic field and the plasma density), and the angle between the magnetic field and the direction of wave propagation. This is true as long as the geometrical optics approximation is valid, where the two modes propagate independently of each other (weak coupling) and each mode retains its identity as it propagates toward the observer. There is, however, a region along the path where the coupling of the modes becomes strong and the polarisation characteristics lock and change no further; this leads to the concept of limiting polarisation.

The observed polarisation of the radio emission is determined by two factors: the intensity difference between the oppositely polarised extraordinary and ordinary

modes, (in terms of brightness temperature) and the conditions of propagation until the region of limiting polarisation is reached. As a consequence, the observed polarisation can be quite different from that at the source [12].

1.1.1 Free-Free Emission

Free-Free is the principal emission mechanism for thermal plasma in the absence of gyroresonance emission and dominates the radio emission of the quiet Sun, originating from electrons that are diverted and accelerated by the Coulomb force of ambient ions. The term “free-free” is due to the state of the electrons; they are free both before and after the interaction.

This mechanism arises when accelerated charged particles, predominantly electrons, collide with ions in the solar plasma. These collisions result in the emission of radiation spanning a wide range of frequencies, with radio wavelengths being a prominent component. As electrons undergo random collisions with ions, they emit photons with energies corresponding to the changes in their kinetic energies. The intensity and spectral characteristics of thermal bremsstrahlung emission serve as vital diagnostics, offering insights into the physical conditions prevailing in different regions of the solar atmosphere.

In the classical limit, the radiation of free accelerated charged particles is described by Larmor’s formula:

$$\frac{dP}{d\Omega} = \frac{q^2 a^2}{4\pi c^2} \sin^2 \theta \quad (1.6)$$

where P is the power emitted within the solid angle $d\Omega$ by a particle of charge q , mass m , and acceleration a in the direction θ relative to the acceleration vector. The total radiated power for radio emission processes is obtained after integration over solid angle.

$$P = \frac{2q^2 a^2}{3c^2} \quad (1.7)$$

Interaction between identical charges also does not produce much radiation because radiation power is proportional to the second derivative of the dipole moment of the system of charged particles, which does not change when two identical particles interact. Consequently, only electron-ion collisions are relevant, and significant radiation is produced by the electrons only.

Many radio diagnostics of the magnetic field are based on the polarisation of the emission.

In the free-free emission the absorption coefficient, and therefore the opacity of the plasma, is slightly different for the two wave modes, hence the net effect will be weakly polarised emission. This is a powerful diagnostic of the magnetic field, because we can immediately obtain qualitative information.

polarised emission reveals the presence of magnetic field and its sense gives the direction of the field with respect to the line of sight: right hand circular polarisation corresponds to positive magnetic field, left hand circular to negative one. We should note however that, far from the centre of the disk, the observed circular polarisation may be highly influenced by propagation effects.

Quantitative magnetic field information is harder to extract. The simplest case is that of an optically thin uniform slab above a uniform background. In this case in terms of Stokes parameters I (total intensity, here measured above the background) and Stokes V (circular polarisation) the fractional polarisation, ρ is:

$$\rho = \frac{V}{I} = 2 \frac{f_p}{f} \cos \theta$$

Substituting numerical values and inverting the formula, we have that that longitudinal component of the magnetic field is given by: [12]

$$B[G] \cos \theta \simeq 5400 \frac{\rho}{\lambda[cm]} \quad (1.8)$$

1.1.2 Gyromagnetic Emission

Gyromagnetic emission is generated when free electrons are accelerated or/and change their velocity direction in a magnetic field due to the influence of the magnetic component of the Lorentz force. An electron with velocity components $v_{\parallel}$ and $v_{\perp}$ parallel and perpendicular to the magnetic field, respectively, will be accelerated perpendicular to both $v_{\perp}$ and B . At low, non-relativistic energies ($\gamma - 1 \ll 1$) the electron velocity distribution in the corona is thermal and the resulting gyroresonance emission is the primary emission mechanism above sunspots with strong magnetic fields at microwaves. Gyroresonance (g-r) emission is produced by thermal electrons gyrating around the lines of force of the magnetic field. It is strong in regions where the observing frequency, f , is a low order harmonic (2nd to 4th) of the electron gyrofrequency,

$$\omega_{ce} = \frac{eB}{m_e c}$$

thus, for a given harmonic s , the following numerical relation holds between the wavelength of observation and the magnetic field [12]:

$$B[\text{G}] = \frac{10700}{s \cdot \lambda[\text{cm}]} = \frac{360 \cdot f[\text{GHz}]}{s} \quad (1.9)$$

Consequently, a fairly high magnetic field is necessary (e.g., 600 G for third harmonic emission at $6 \text{ cm} - \lambda$). Although gyroresonance radiation is emitted at discrete frequencies, it generally gives rise to a continuous spectrum due to the variation of the magnetic field with height. The characteristics of gyrosynchrotron (g-s) emission from mildly relativistic electrons, trapped in flaring loops, depend strongly on the magnetic field. Gyrosynchrotron radiation from electrons that gyrate in the magnetic field with energies of tens to hundreds of keV is the basic emission mechanism at microwaves.

Gyroresonance emission, in particular will be discussed in the last Chapter of this work for the Sundish ARs statistical analysis, as it is relevant to study magnetic field structure above solar active regions. Indeed, due to their significant magnetic field strength, sunspots are a prime location to observe gyroresonance emission.

Sunspots are sources of localized microwave emission, generated in thin layers around iso-Gauss surfaces where the magnetic field strength is such that the observing frequency is equal to a harmonic of the local gyro-frequency.

Regarding polarisation measurements, the gyroresonance opacity is a complex function influenced by temperature, density, magnetic field intensity, wave mode, and the direction of the field relative to the line of sight [91]. This opacity is null when the magnetic field and line of sight are parallel. It exhibits a significantly higher value in the extraordinary mode compared to the ordinary mode, because the opacity is greater for the second harmonic than for the third one. In atmospheric layers above sunspots, conditions typically render the third harmonic opaque in the extraordinary mode and transparent in the ordinary mode within the microwave range, while it remains opaque for the second harmonic in both modes [12].

When the photospheric field strength is sufficiently weak (or the frequency sufficiently high), both the third and second harmonic layers lie below the Transition Region (TR), resulting in no significant sunspot-associated emission. However, with stronger fields or lower frequencies, the third harmonic enters the TR while the second remains in the chromosphere, leading to pronounced emission, highly polarised in the extraordinary mode [180]. With even higher field strengths, the second harmonic also enters the TR, resulting in strong emission in both the ordinary

and extraordinary modes, with reduced polarisation.

1.1.3 Active Regions and Quiet Sun

Active regions are the totality of observable phenomena in a 3D volume represented by the extension of the magnetic field from the photosphere to the corona, revealed by emissions over a wide range of wavelengths from radio to X-rays and γ -rays (only during flares) accompanying and following the emergence of strong twisted magnetic flux through the photosphere into the chromosphere and corona. In the photosphere the presence of a strong magnetic field is manifested by the appearance of dark sunspots or pores and bright faculae representing concentrated and dispersed magnetic fields, respectively, inhibiting convection. These magnetic fields can be thousands of times stronger than Earth's magnetic field and often form pairs with opposite polarities.

In the chromosphere arch filament systems connect opposite polarity magnetic concentrations, filaments form along the magnetic inversion line and the bright regions that appear above dispersed fields are called plages. In the transition region and corona bright, hot, dense loops connect the opposite magnetic polarities.

Coronal loops, also known as solar loops or magnetic loops, are structures of hot, ionised gas (plasma) following the Sun's magnetic field lines. They are visible in the Sun's outer atmosphere, the corona, and can extend for tens of thousands of kilometers above the solar surface.

ARs they are crucial for understanding solar dynamics and their influence on Space Weather and Earth's environment, as they are the principal source of a broad range of solar activity phenomena: ranging from small-scale brightenings and jets to the largest flares and coronal mass ejections (CMEs) [199].

Active regions can persist for days to weeks, evolving as magnetic fields interact and rearrange. They tend to occur more frequently during periods of high solar activity, such as solar maximum, but can also appear during quieter periods. Understanding active regions is crucial for predicting solar flares, which can impact satellite communications, power grids, and even endanger astronauts during space missions.

Studying solar active regions provides valuable insights into the Sun's magnetic dynamo, the process responsible for generating the Sun's magnetic field, as well as its variability over time.

The term "Quiet Sun" refers to the relatively calm and stable state of the Sun's surface when it is devoid of significant solar activity, such as sunspots, solar flares,

or coronal mass ejections.

The Quiet Sun (QS) brightness temperature spectrum has been extensively modeled in various studies (e.g., [103, 155]). Its radio emission predominantly originates from thermal bremsstrahlung in Local Thermodynamic Equilibrium (LTE), rendering it comprehensively understood and modeled compared to impulsive emissions in the other frequency bands. Consequently, it serves as a robust diagnostic tool for discerning physical conditions and parameters across a wide range of atmospheric layers.

However, accurate and calibrated measurements of the brightness temperature of the QS component in the K-band (18-26 GHz) remain challenging due to the difficulty in distinguishing it from the contribution of Active Regions (ARs).

The ARs radio emission mechanism is more difficult to understand than the QS since there is an entanglement of thermal (free-free and gyro-resonance) and non-thermal emission depending on the observing frequency. In the radio domain, the gyro-resonance has been detected up to 17 GHz [146], and has its emission peak in the 2-5 GHz range.

Additionally, in the chromospheric layers, QS emission exhibits slight deviations from simple thermal bremsstrahlung, primarily due to the logarithmic dependence of the Gaunt factor on frequency. As a result, the flux density spectral index (α) is approximately 1.9, deviating from the expected value of 2 in the Rayleigh-Jeans regime.

There is evidence suggesting spectral changes in QS emission from lower to higher frequencies, with indications of a potential flattening around 10 GHz, as proposed by Landi and Chiuderi Drago [103].

Taking into consideration the ARs flux density spectra [183], the K-band seems to be a transition region between different emission mechanisms. This points out the importance of further refining observational techniques to enhance our understanding of QS emissions in the K-band and their implications for solar atmospheric dynamics.

1.1.4 Space Weather Phenomena from Eruptive Events

Space Weather phenomena encompass a broad array of dynamic and often unpredictable disturbances in the space environment surrounding Earth. These phenomena originate from the Sun's activity, including solar flares, coronal mass ejections (CMEs), and solar wind, which interact with Earth's magnetic field and atmosphere. The resulting interactions can produce disturbances in Earth's ionosphere and mag-

netosphere, leading to various effects on technological systems and infrastructures. CMEs consist of large-scale structures containing magnetised plasma ejected from the solar corona into the interplanetary medium at velocities exceeding millions of kilometers per hour, [203, 63] CMEs (Fig. 1.2) and their interplanetary counterparts (ICMEs) are known to be associated with a variety of other important phenomena, such as solar energetic particle (SEP) events, interplanetary (IP) shocks, geomagnetic storms (GS), which form essential ingredients of the current Space Weather. These eruptions, often originating from regions of heightened magnetic activity such as solar active regions and sunspots, can span vast distances and possess very high kinetic and magnetic energies.

The complex dynamics governing CMEs involve the rapid reconfiguration and expulsion of the Sun's magnetic field lines, often facilitated by magnetic reconnection processes [61].

This expulsion of magnetized plasma, laden with charged particles, constitutes a formidable solar storm that, when directed towards Earth, interacts with our planet's magnetosphere, inducing geomagnetic disturbances and triggering auroral lights at high latitudes and low latitudes in strong events.

Solar Flares, intense bursts of electromagnetic radiation emanating from localised regions of magnetic instability on the Sun's surface, represent another facet of solar eruptive activity [22]. These events, characterized by their abrupt release of energy across multiple wavelengths, from radio to gamma rays (Fig. 1.3), bear witness to the rapid acceleration of charged particles within the Sun's magnetized atmosphere [181]. The underlying mechanisms driving solar flares is the magnetic reconnection, a process where magnetic field lines rearrange and release energy. During this process oppositely oriented components of the magnetic field annihilate, the magnetic field reconfigures to a lower-energy state, and the liberated free energy of the magnetic field in the plasma is efficiently converted into particle kinetic energy through acceleration and plasma heating [141].

Solar flares occur when magnetic flux emerges from the lower regions of the Sun onto its surface and are believed to be at least partially responsible for heating the Sun's corona [21].

Typical flares last from minutes to hours and emit radiation at mostly short wavelengths. However, large flares often produce electromagnetic radiation at all observable wavelengths including the visible spectrum, in which case the flare is known as a "white light flare".

Radio emission from solar flares offers a number of unique diagnostic tools to

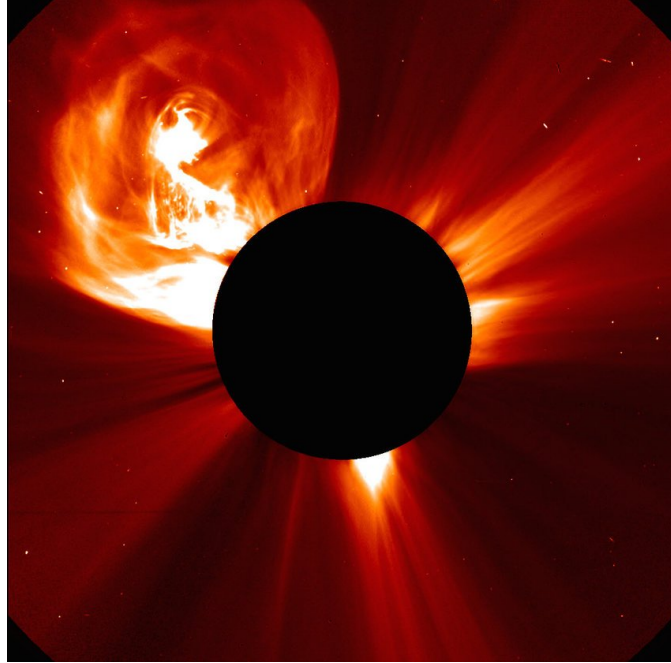


Figure 1.2: A coronal mass ejection imaged by the SOHO/LASCO C2 coronagraph on 4 January 2002. The dark disk is the result of the occulting disk, which has a radius of $1.7 R_{sun}$. [63] Credits: NASA

address long-standing questions about energy release, plasma heating, particle acceleration, and particle transport in magnetized plasmas. At millimetre and centimetre wavelengths, incoherent gyrosynchrotron emission from electrons with energies of tens of keV to several MeV plays a dominant role. These electrons carry a significant fraction of the energy released during the impulsive phase of flares. At decimetre and metre wavelengths, coherent plasma radiation can play a dominant role [19].

Solar flares are classified by their soft x-ray peak energy into several categories: A, B, C, M, and X, with A being the least intense and X being the most intense [87]. The Table 1.1 lists different classifications of solar flares, along with their associated X-ray fluxes [43], usually based on Geostationary Operational Environmental Satellite (GOES) measurements.

1.1.5 Solar Radio Bursts (SRB)

Non-continuous solar radio emissions, commonly referred to as radio bursts, are transient enhancements in radio emission observed across a range of frequencies.

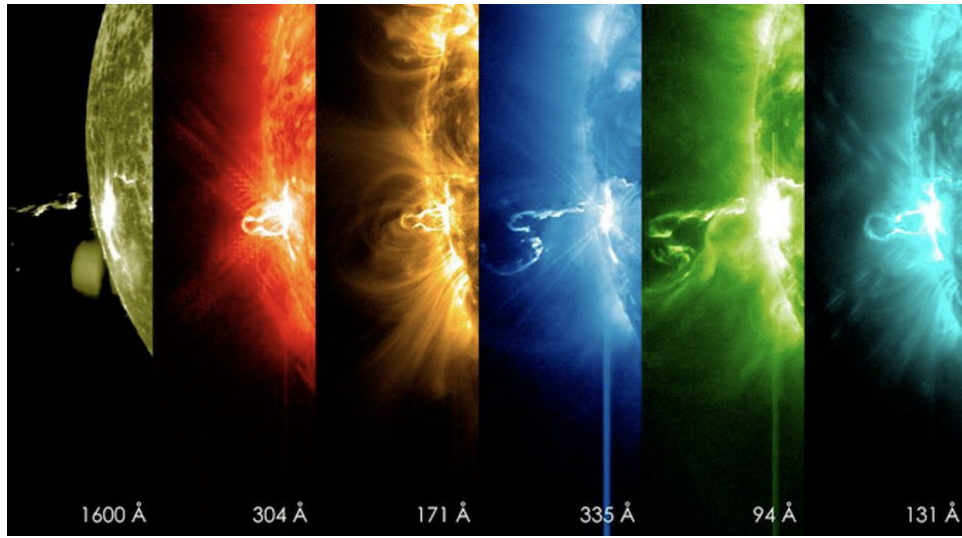


Figure 1.3: A solar flare seen in multiple wavelengths of ultraviolet light. Note that since flares are EM radiation events. Image Credit: NASA/SDO.

Solar Flare Class	Associated X-ray Flux I [$W \cdot m^{-2}$]
X	$I \geq 10^{-4}$
M	$10^{-5} \leq I < 10^{-4}$
C	$10^{-6} \leq I < 10^{-5}$
B	$10^{-7} \leq I < 10^{-6}$
A	$I < 10^{-7}$

Table 1.1: Description of solar flare classes where the definition of the flare classes is based upon their peak flux in soft X-rays of 0.1–0.8 nm [43].

These bursts can be classified into several types based on their duration, spectral characteristics, and associated solar phenomena:

- **Type I Bursts:** Type I bursts are short-lived, narrow band emissions typically observed at frequencies below 300 MHz. These bursts are associated with thermal electron plasma emissions and are often observed in regions of quiet Sun [224].
- **Type II Bursts:** Type II bursts are indicative of shock-accelerated electrons in the solar corona and are often associated with the eruption of coronal mass ejections (CMEs). These bursts exhibit a characteristic spectral signature, with a slowly drifting frequency (~ -30 MHz/s) and a rapid intensity decrease [145].
- **Type III Bursts:** Type III bursts are produced by energetic electron beams

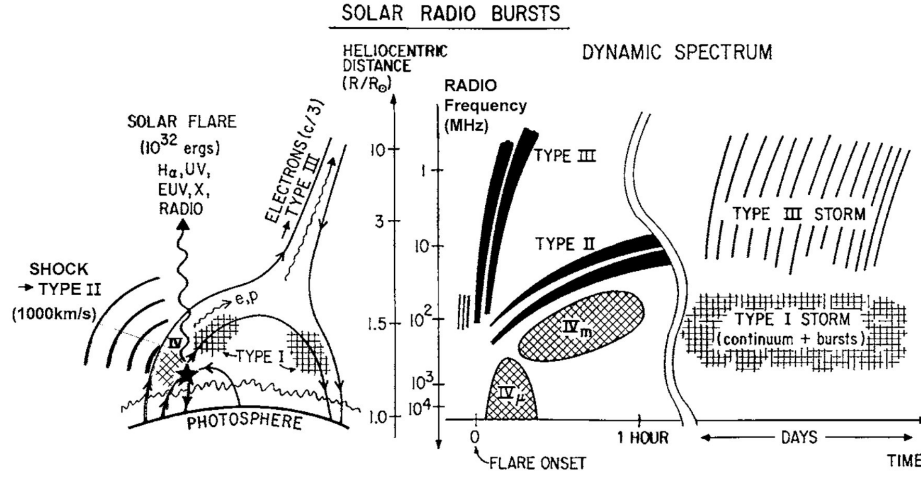


Figure 1.4: Synopsis of a solar radio event. A typical helmet streamer structure is shown on the left. The corresponding radio events are shown in the dynamic spectrum on the right. [94]

propagating along open magnetic field lines in the solar corona. These bursts exhibit high frequency drifts (~ 100 MHz/s) and are often observed in association with solar flares and energetic particle events [112].

- **Type IV Bursts:** Type IV bursts are characterized by broad-band continuum emission and are typically associated with eruptive events such as solar flares and CMEs. These bursts often exhibit long-duration emissions spanning a wide range of frequencies [32].

Solar radio emission provides valuable insights into the physical processes occurring in the solar atmosphere, including magnetic reconnection, particle acceleration, and plasma dynamics, as sketched in Fig. 1.4. By studying the properties and evolution of radio bursts, we can investigate the mechanisms driving solar activity and variability, shedding light on fundamental processes such as solar flares and CMEs [131], therefore they play a crucial role in Space Weather prediction and forecasting. The occurrence and properties of radio bursts provide valuable information about solar eruptive events, which can impact Earth's magnetosphere, ionosphere, and technological infrastructure. Monitoring solar radio emissions allows forecasters to predict and mitigate the effects of Space Weather events on satellite communications, GPS navigation, and power grids [136].

1.2 Space Weather Effects on GNSS

The ionosphere is the ionised component of the atmosphere comprising free electrons and positive ions, generally in equal numbers, in a medium that is electrically neutral. Extending from approximately 50 to 600 kilometers above the Earth's surface it plays a basic role in long-distance communication and it holds significant importance in global navigation satellite systems (GNSS) as well as in ionospheric research [97].

The ionosphere exhibits both temporal and spatial variations and its effects on GNSS signal propagation is one of the main error sources which limits the accuracy and the reliability of GNSS applications.

GNSS, comprising constellations of satellites such as GPS, GLONASS, Galileo, and BeiDou, utilise signals transmitted from satellites to determine positions on Earth's surface. The signals beamed from satellites must pass through the ionosphere on their way to the receiver on Earth, therefore its variability can be one of the main error sources which limits the accuracy and reliability of GNSS applications.

Free electrons, as the most massive particles in the ionosphere affect the propagation of the signal, changing their speed, direction and shape of the signal path.

The accuracy of GNSS positioning depends on the precise determination of the travel time needed by the radio signal to propagate from the satellite transmitter to the receiver [126].

When radio waves, such as those emitted from GPS satellites pass through the ionised path, there are two effects: the trajectory of the beam is bent and the signal comes to a destination with a delay.

In a magnetized plasma the refractive index becomes a complex quantity due to the interaction between the waves and the charged particles in the plasma, which Edward Appleton has described in his magneto-ionic theory. He demonstrated first that when going through the magnetised plasma, a plane polarised wave splits into two circularly polarised waves rotating in opposite directions. The change in refractive index as a function of frequency in a plasma medium is described by the Appleton-Hartree formula [166]:

$$n(\nu) = \sqrt{1 - \frac{f_p^2}{f^2}} \quad (1.10)$$

where f_p is the plasma frequency while f is the frequency of the electromagnetic wave (Eq. 1.10).

Solar radio bursts, for instance, directly affect GNSS by emitting bursts of radio waves that can interfere with GNSS signals at the receiver, causing disruptions in navigation and timing services.

Solar Radio Bursts are proxies of the occurrence of eruptive events, such as CMEs and flares, that can induce physical variations in the ionosphere.

Ionospheric irregularities, known as plasma density irregularities or ionospheric bubbles, can scatter and refract GNSS signals, causing rapid fluctuations in signal strength and phase, while increase electron density of the ionospheric layers can induce additional delays (Fig. 1.5).

Scintillation refers to rapid fluctuations in the amplitude and phase of GNSS signals

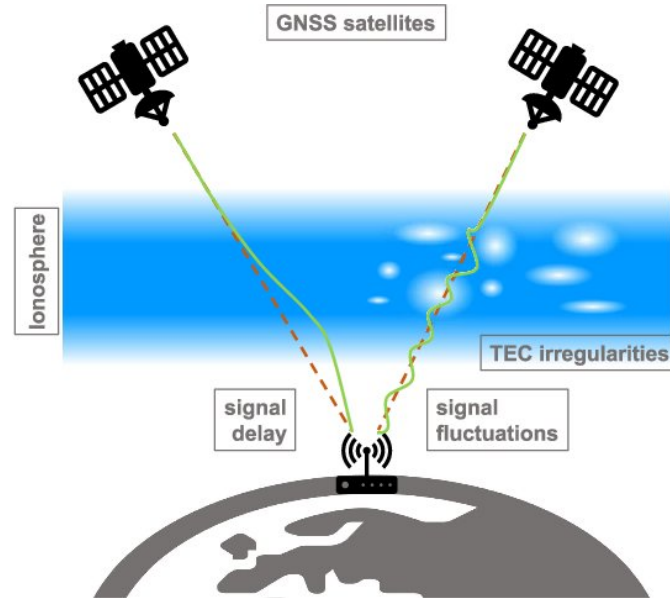


Figure 1.5: Ionospheric delay and disturbances due to scintillation effects. The red dashed lines are the line-of-sight signal paths from the GNSS satellites to the receiver on earth; the green continuous signal accounts for propagation distortions [113].

caused by irregularities in the ionosphere, particularly prevalent in equatorial and high-latitude regions during periods of increased solar activity. These fluctuations can lead to signal fading and degradation, affecting GNSS accuracy and reliability. On the other hand, enhanced TEC values can introduce additional unexpected signal delays leading to larger errors in GNSS positioning and navigation.

To mitigate these impacts, algorithms and techniques for detecting and correcting

signal disturbances caused by scintillation and TEC variations are fundamental. Understanding and monitoring Space Weather phenomena are crucial for mitigating their potential impacts on GNSS and other technological systems. By developing strategies to anticipate and respond to Space Weather events, we can ensure the reliable operation of critical infrastructure in diverse operational environments.

1.2.1 SRB Direct Effects

Solar radio noise can increase by several orders of magnitude during increased solar activity and it can persist at high levels for minutes to hours. Solar radio bursts, originating from highly energetic solar events such as solar flares and coronal mass ejections (CMEs), can directly influence Global Navigation Satellite Systems (GNSS) signals. The intensity of these emissions is increased with respect to the solar radio emission background and the occurrence of an intense SRB can reduce the carrier-to-noise power spectral density ratio (C/N) of the receivers located in the sunlit hemisphere of the Earth [130]. Consequently this results into a degradation of the positioning and navigation accuracy and in the worst cases to the total interruption.

The threshold value of SRB flux value that could result in a significant effect on GNSS signals is believed to be between 1000–10,000 SFU in L band [216].

A SRB will mainly affect stations located in the sunlit hemisphere during radio flux enhancement, while the strength of the influence depends on the solar incidence angle, the antenna pattern, the tracking algorithm, and some other factors. Indeed, solar interference occurs when specific conditions are met:

- Geometrical condition where the Sun is in the antenna beam of the communication system for a minimal time. It is determined by seasonal and geographical aspects (Sun's position in the sky) as well as by technical aspects (antenna beam width and system sensitivity);
- The solar radio noise is significantly high at the relevant frequencies. This conditions depends on the solar activity cycle;
- Technical condition where the acceptable Signal-to-Noise Ratio (SNR) is high.

In the worst scenario a complete loss of lock of GNSS signals can occur and may persist also for a significant period of time, causing the complete loss of receiver tracking.

With regard to that, we stress the importance of adequate observations of SRBs flux density with high time resolution and accurate polarisation measurements (as GPS antennas are typically Right-Hand circularly polarised) to provide the required information for a significant correlation analysis. If a solar radio outburst occurs with the enhancement in L band radio flux, it could influence GNSS signals through direct radio wave interferences, resulting in reduction of signal-to-noise ratio (SNR) and instantaneous or long-period loss of lock (LOL) on GNSS signals.

1.2.2 Ionospheric TEC Variability

When signals from satellites pass through the ionosphere to reach receivers on the Earth's surface, they experience refraction, due to the changing refractive index. This refraction affects the phase and group velocities of the waves, causing delays and distortions that can be characterised by parameters like the total number of free electrons in a vertical column of the Earth's ionosphere (VTEC).

Assuming that the ionosphere is compressed in the infinitesimally thin layer around the Earth at an altitude of 350-400 km, at the intersection of the direction of propagation of the signal and ionospheric layer we define the pierce point (IPP - Ionospheric Pierce Point).

Each satellite used for positioning transmits two carrier frequencies, for example GPS constellation uses L1 and L2, where $L1 = 1,575.42MHz$, $L2 = 1,227.60MHz$. When the GPS signals simultaneously propagate through the ionosphere, the relative ionospheric delay of the two signals is proportional to the total amount of electrons along the ray path. Time delay measurements of L1 and L2 frequencies can, therefore, be converted to the STEC value, defined as the total number of free electrons along the raypath T_x between a satellite and a receiver on the Earth's surface.

$$STEC = \int_{T_x} n dl \quad (1.11)$$

Where n is the location-dependent electron density, measured in *electrons* · m^{-3} and the l-axis stands for the receiver-to-satellite direction.

Fig. 1.6 depicts the conversion of STEC into VTEC:

$$VTEC = sTEC \cos(\chi) \quad (1.12)$$

$$\cos(\chi) = \sqrt{1 - \left(\frac{R_E}{R_E + h} \cos(\theta)\right)^2} \quad (1.13)$$

where χ is the zenith angle, R_E is the Earth's radius, equal to $\sim 6378\text{km}$, h is the height of the ionospheric layer, assumed at 350 km , and θ is the elevation angle of the satellite [184]. The satellite-to-ground communications are influenced by

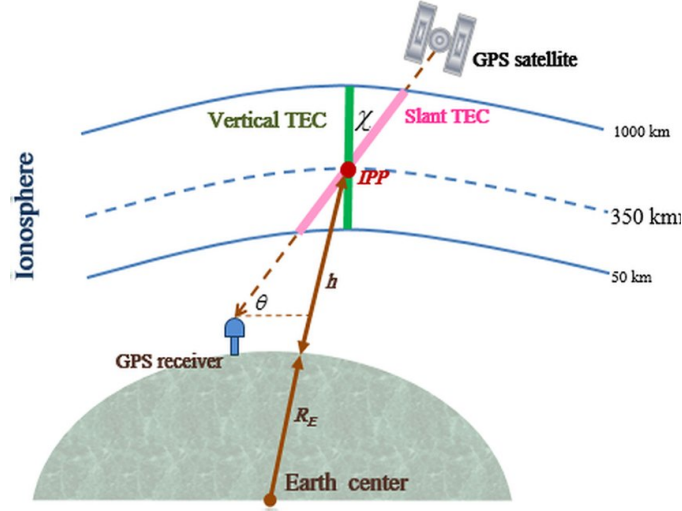


Figure 1.6: Vertical value of total electron content (VTEC) obtained from STEC. Ionospheric Pierce Point (IPP) refers to the point in the ionosphere where the line of sight from the satellite intersects with the Earth's ionosphere. (Source: Thai GNSS and Space Weather Information Data Center) [184].

ionospheric plasma which varies depending on solar and geomagnetic activities as well as regions and local times. With the expansion of use of the space, continuous monitoring of the ionospheric plasma has become an important issue. The ionospheric delay, which is proportional to ionospheric TEC along the propagation path, is the largest error in signal propagation.

Estimating VTEC accurately is essential for various applications, including satellite communication, navigation, and Space Weather monitoring [71].

1.2.3 Ionospheric Scintillation

Radio waves can be scattered by irregularities in the density of the ionosphere, such as those produced by a variety of plasma instabilities. The irregularities can act as diffraction screen producing interference patterns in radio waves that pass through the ionosphere [65]. Ionospheric scintillation is the rapid and random fluctuation in amplitude and phase of trans-ionospheric radio waves caused by electron density

irregularities, caused by small-scale irregular structures in the ionisation density via forward scattering and diffraction of the GNSS signal [86].

The most commonly used parameter characterizing the intensity fluctuations is the scintillation index S_4 .

It is defined as the standard deviation of signal intensity fluctuations normalized by the mean signal intensity:

$$S_4 = \frac{\sqrt{\langle I^2 \rangle} - \langle I \rangle}{\langle I \rangle} \quad (1.14)$$

where $\langle I \rangle$ represents the mean signal intensity, and $\langle I^2 \rangle$ denotes the mean square signal intensity [171]. It provides a measure of the relative magnitude of intensity fluctuations compared to the mean signal intensity.

On the other hand, the phase scintillation index, denoted as σ_ϕ , quantifies the rapid fluctuations of the phase carried by the GNSS signal, caused by ionospheric irregularities. It is defined as the standard deviation of phase fluctuations:

$$\sigma_\phi = \sqrt{\langle \sigma^2 \rangle - \langle \sigma \rangle^2} \quad (1.15)$$

Due to the differences in the generation mechanism of ionospheric scintillation in the low latitude and the polar regions, amplitude scintillations often occur in the low latitude region, while phase scintillations dominate in the polar region [223]. This time period of the ensemble averaging is nominally 60s [99]. The dynamics of the formation of plasma irregularities in the low-latitude ionospheric region is a unique phenomenon. Typically, the plasma density becomes inhomogeneous at the bottom side of the F region of the ionosphere in the equatorial and low latitude region during post-sunset hours which is called the equatorial spread F. At the E-region, the higher recombination rate results in low electron density at the same occurs after local sunset hours. This produces a vertical gradient of plasma density between the high plasma density F-region and the low plasma density E-region in the opposite direction of gravity. This vertical gradient of plasma density induces Rayleigh-Taylor instability in the plasma between the F and E regions, which in turn produces the prereversal enhancement of the eastward electric field [9]. When the enhancement of the eastward electric field becomes strong enough, the vertical $\mathbf{E} \times \mathbf{B}$ drives the plasma irregularities in the upward direction. During its motion towards the peak of the F-region, the irregularities in the plasma density grow in size which results in the formation of large plasma-depleted structures, called equatorial plasma bubbles (EPB). The ionospheric irregularities move downward along the

geomagnetic field lines to the higher latitude regions (0° to $\pm 15^\circ$) after reaching a higher altitude of about 1000 km. This phenomenon is known as the Fountain effect which produces two anomaly crest regions around $\pm 15^\circ$ geomagnetic latitudes where the electron density is larger than the trough region between the two crest regions [120].

In conclusion, strong scintillation events, due to significant Space Weather events, can degrade the accuracy and availability of GNSS signals, particularly in equatorial and high-latitude regions where ionospheric irregularities are more prevalent.

Chapter 2

The Trieste Solar Radio System

2.0

The Trieste Solar Radio System 2.0 (TSRS 2.0) is the project for a new, state-of-the-art solar radio spectropolarimeter dedicated to the continuous surveillance of the solar radio emission in the range 1 – 18 GHz.

TSRS will be one of the observing nodes of the forthcoming INAF National Space Weather Service Network (INAF NSWSN) consisting of several INAF observing assets involved in different Space Weather fields, whose development is in progress or has just started. In addition, it will be part of the INAF Trieste Solar Radio Weather Centre (TSRWC), a centre dedicated to both Solar Radio Physics and to the monitoring, modelling and nowcasting/forecasting of radio events. The antenna system is definitely smaller compared to large radio facilities distributed worldwide, but it has peculiar operational features that make it unique with respect to most of them.

Firstly, it has been designed to perform diachronic (i.e. continuous) near-real-time solar radio observations of the calibrated radio flux density at the observing frequencies. Particularly important in this context are the measurements in the L band (1.1 to 1.6 GHz), the band relevant to GNSS systems, and the monitoring of the solar 10.7 cm flux (2800 MHz), universally used as a proxy indicator of solar activity. It will also provide accurate Circular Polarization sense and percentage data, a kind of measurement that state-of-the-art operational radio facilities (e.g. RSTN) do not support.

Polarisation measurement is an added value for twofold reasons. On one hand, it provides Solar Radio Physicists with diagnostic tools for interpreting plasma

characteristics at the radio emission source and radio emission mechanisms and propagation effects such as plasma structures, reflections, diffusive processes, Faraday rotation and so on. On the other hand, polarisation information allows to discriminate between RHCP and LHCP radio events and, therefore, to identify potential impact to GNSS-based services. In order to set everything ready for the forthcoming TSRS 2.0 data we have conducted a study aiming to establish a new data archive for the old Trieste Solar Radio System (TSRS 1.0) data, tailored to the needs of the Space Weather community. This part of the project serves as a testbed for optimizing the forthcoming TSRS 2.0 data infrastructure and will be discussed in a dedicated Section in this Chapter.

The TSRS 1.0 monitored solar radio emissions from 1969 to 2010 but the data collected are no longer accessible to the scientific community. To address this issue, the project proposes a comprehensive data curation process along with the development of a new data model and interfaces which meet principles of findability, accessibility, interoperability, and reusability (FAIR principles). This upgrade will significantly improve the accessibility and usability of the archive, enabling advancements in the understanding of solar radio physics, post-event studies and Space Weather phenomena.

2.1 The TSRS 2.0 Architecture

Although more advanced, in terms of technology in use, than the previous instrumental set, TSRS 1.0 (see Fig 2.1), the design of TSRS 2.0 is the result of many years of experience and lessons learned from the use of this no-longer-operational facility. Figure 2 depicts the main components of the TSRS 2.0 architecture, that are described hereafter. The 3.7-meter parabolic antenna, with full sky coverage motorised mount, azimuth over elevation configuration, has been installed on the roof of the Basovizza Observing Station’s main building. The main environmental and mechanical specifications of the receptor are summarised in Table 2.1, Tables 2.2 and 2.3 define the characteristics of the horn and LNA (POAM Electronics [2]), while an example of a parabolic dish of the same manufacturer (POAM Electronics [2]) is depicted in Figure 2.3 (left panel).

The antenna positioning is controlled by the Antenna Control Unit (ACU) and the Antenna Drive Unit (ADU) that can be interfaced with the dedicated scheduling software installed in the workstation. An ultra wide-band 1-18 GHz dual polarised horn antenna is installed in the antenna secondary focus (see Fig 2.1). The ultra



Figure 2.1: The TSRS 1.0 system. It was a set of two multi-channel radio polarimeters, which were detecting the solar radio emissions in the metric band by a 10-meter diameter parabolic dish (left panel), and in the decimetric one by a 3-meter parabolic dish (right panel), at the following frequencies 237, 327, 408, 610, 1420, 2695 MHz, with very high temporal resolution (1 ms standard) and accurate circular polarisation measurements. It is not longer operational due to a lightning stroke that irreparably compromised its operations in 2010.

Features	
Antenna diameter	3,7 m
Reflector material	Aluminium
Antenna type	Focus ring antenna
Mount type	Az over El
Antenna travel range	Az = 0 to 360° El = 0 to 90°
Acceleration (each axis)	Constant
Velocity (each axis)	0.5°/s
Tracking accuracy	< 0.1°
Pointing accuracy	0.01°
Surface accuracy	0,5 mm
Operating temperature	-40°C to 55°C
Outdoor humidity	0 ~ 100
Operating wind speed	80 km/h, gusting 85 km/h
Survival wind speed (park position at zenith)	210 km/h Gusting 230 km/h
Operational ice resilience	13 mm
Survival ice resilience	25 mm

Table 2.1: Environmental and mechanical characteristics.

Features	
Operating frequency	900 MHz 19 GHz
Typical gain	7 to 16 dBi
Voltage Standing Wave Ratio (VSWR)	1,25 : 1 on average
Max Power handling	100 W CW
Cross polarization	-35 dB on average
Cross port isolation	-35 dB on average
Noise temperature	34 K at 10° EI
First sidelobe	< -14 dB
Feed material	Aluminium
Operational wind loading	80 km/h
Survival wind loading	160 km/h
Operational Temperature	-40° to +70°C
Survival Temperature	-50° to +80°C

Table 2.2: Wideband Horn.

Parameter	Min	Typical	Max	Units
Frequency Range	0.9		19	GHz
Gain	26	28	30	dB
Gain Flatness		±1	±2	dB
Gain Variation Over Temperature (-45°C ~ +85°C)		±1.0		dB
Noise Figure	0.7	1.3	1.5	dB
Input VSWR		1.3	1.5	:1
Output VSWR		1.5	2.0	:1
Output 1dB Compression Point (P1dB)	8	12		dBm
Saturated Output Power (P _{sat})		15		dBm
Output Third Order Intercept (OIP ₃)		18		dBm
Supply Current (Vcc = +15V)		100	130	mA
Isolation S ₁₂		-55		dB

Table 2.3: LNA Specifications

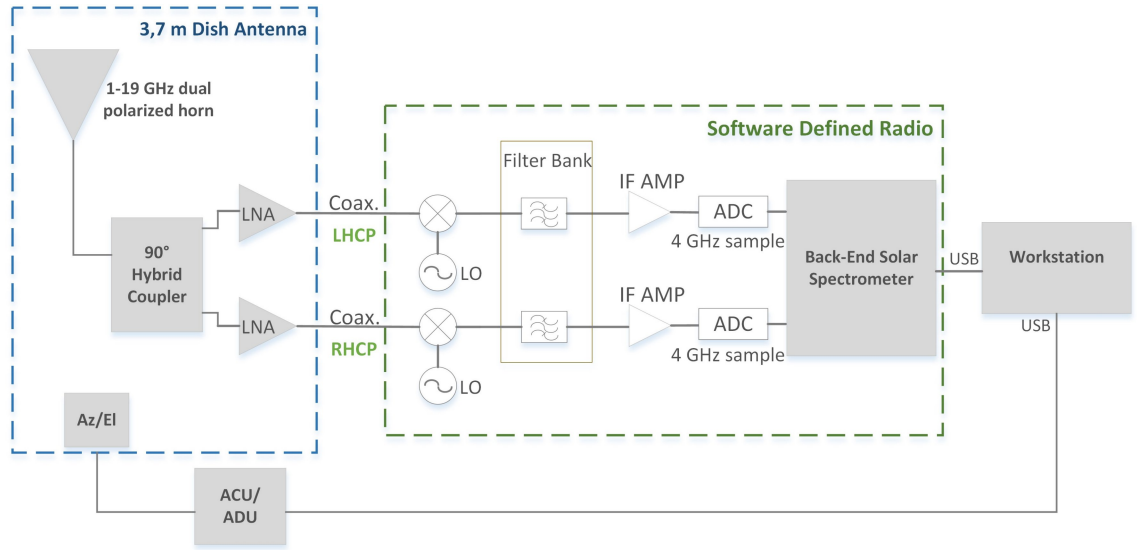


Figure 2.2: The TSRS 2.0 architecture.

wide-band hybrid coupler will split the input signal with a resultant 90 degree phase shift between the output ports, separating the two circular polarized signals into RHCP and LHCP. The signals are then conveyed to the Software Defined Radio (SDR) receiver through coaxial cables, after the wide-band Low Noise Amplifier (LNA). The SDR is a receiver unit for continuum and spectral line observation operating in selected frequency bands in the range 1 to 18 GHz such as the L (1-1.6 GHz), 2800 MHz, C (4-8 GHz) and Ku (12-18 GHz) bands with high resolution digitiser. The main features of the receiver are listed in Table 2.1. The frequency

Features	
Spectrometer for spectral line and continuum observations	
Dual channel 4 GHz bandwidth	
SNR	> 100 dB
Min. frequency resolution (RBW)	10 Hz
Low level signal detection	-150 dBm

Table 2.4: Receiver main features.

of the receiver is selected by changing the frequency of the Local Oscillator (LO) and then filtering the signal to a 4 GHz band. Then the Analog to Digital Converter (ADC) samples at 8 Gsps rate. A dual channel back-end spectrometer analyses the sun spectra.



Figure 2.3: The left panel shows 3-m antenna by POAM. The right panel depicts the mechanical structure.



Figure 2.4: Dual polarisation horn antenna.

Fig. 2.1 shows the 19-in rack mount controllers (ACU and ADU) on the top and SDR receiver on the bottom. The SDR will be linked via USB interface to the TSRS 2.0 workstation. A general overview of the main software components foreseen are depicted in Fig.2.1. A set of API for programmatic discovery and access to the data collections contents, as well as harvesting interfaces to connect to external services and collaborations are at the present in implementation phase. Archive discovery, access and interfacing solutions have been designed to follow the open science and FAIR-principles-driven interoperability adopting global technological standards and architectures [88].



Figure 2.5: POAM Electronics ACU and ADU controllers (first two elements on the top) and SDR receiver (last element on the bottom).

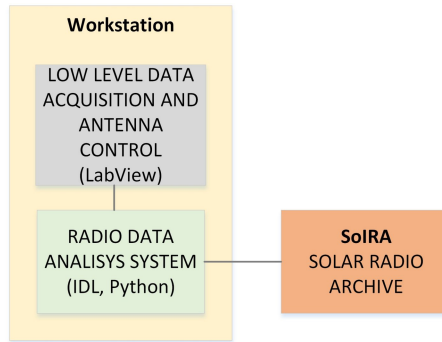


Figure 2.6: General overview of the software architecture components.

2.2 The TSRS 2.0 Observation Modes and Data Products

TSRS 2.0 will carry out two kind of observations, namely frequency scan and spatial scan, see Fig. 2.8. A scheduling system will allow to switch from one to the other observation mode and produce different collections of raw data.

In frequency scan mode, high time resolution 2.5D spectra (i.e. records of flux variations versus frequency and time) will be produced with spectral coverage from 1 to 18 GHz and with instantaneous bandwidth of 4 GHz.

In the other mode, the raw data consists of 2D synoptic images at selected frequencies with low time resolution by spatial scanning. The two raw collections feed independent reduction pipelines. In particular, the processing of the spectra results in full time resolution spectropolarimetric time series defined at selected frequencies or bands. From these time series, lower resolution indices will be derived as sum, mean and max values of the radio flux evaluated in a given time range for both the RHCP and LHCP components.

Fig 2.7 depicts the multi-channel synoptic graphs of the radio emission of the per-

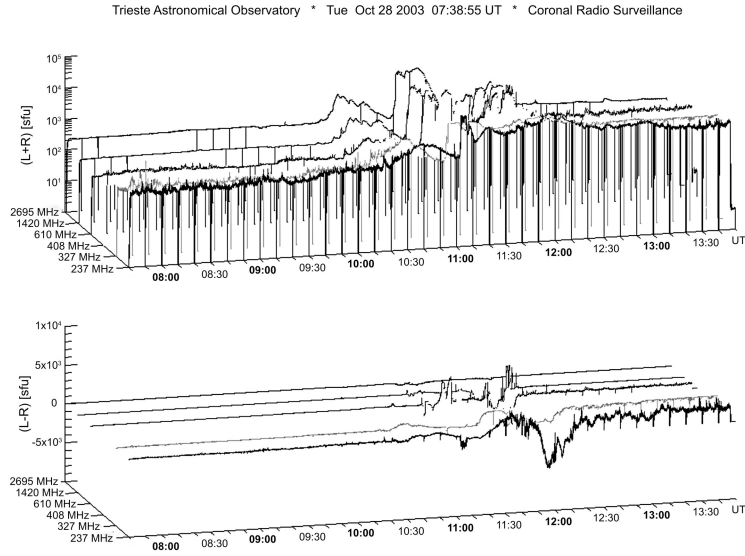


Figure 2.7: Multi-channel synoptic graph of the radio emission of the perturbed corona observed by TSRS 1.0. during the 28 October 2003 event (Halloween Storm). On the top panel the total intensity $L+R$ is depicted, while on the circular polarisation component $L-R$ is shown on the bottom panel.

turbed corona measured by TSRS 1.0, as an example. From the comparison of the relative polarisation contributions, it is evident that the detected SRB was a fully RHCP event. Radio indices are suitable metrics used as proxy of solar activity and are useful inputs to a variety of nowcasting and forecasting models in Space Weather prediction. Among these, the most widely used is the 10.7 cm (2800 MHz) radio flux, also known as F10.7 index.

All raw and derived data products will be stored in the Solar Radio Archive in different formats (i.e., FITS, ASCII, and binary) with provenance linking to track the logical connections among the various data sets. It will be therefore possible for a given radio event to point to the relevant data sets of indices, series and spectra, but also, if available, to the relevant radio source.

Data and indices will be also published in near-real-time on a dedicated web site along with a selection of quick-look graphs updated in near-real-time. As these products will become available, all the Space Weather Science and forecast community, web applications, client tools, etc. will be set up as consumers of the archive's API.

Given the unique characteristics of TSRS 2.0, it will produce useful and robust data products and indices for monitoring purposes. One of the purposes of this thesis

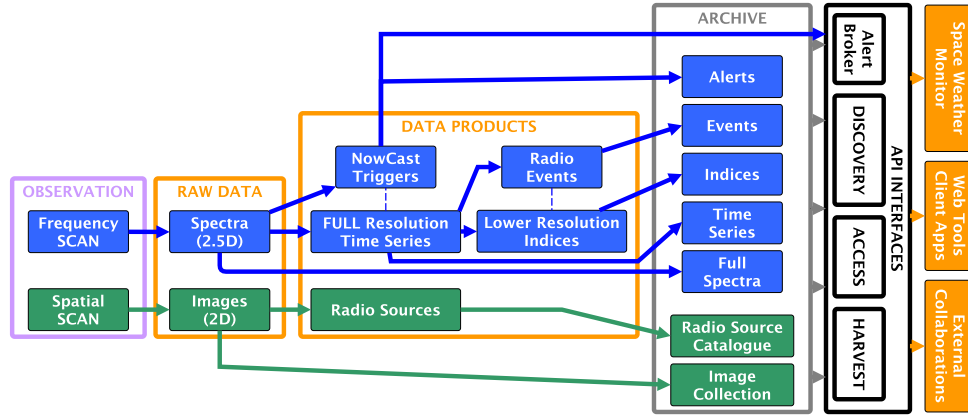


Figure 2.8: Data flow for the TSRS 2.0 radio spectropolarimeter.

work is to develop useful tools ready to be used as forecasting aids.

TSRS 2.0 solar observations will overlap with the development and the operation of new generation ground-based and space-borne instruments that (will) observe the Sun in different spectral ranges and with different observation modes, integrated into the National Space Weather Service Network.

They will share some scientific objectives, and combining their data will result in an added value, giving an interesting opportunity to produce a more comprehensive picture of all the involved solar phenomenology over an extended frequency range. In this collaborative approach, TSRS 2.0 is a small and flexible instrument supposed to operate in synergy with the aforementioned large radio facilities. In particular, thanks to its high time resolution, it could issue alerts about SRBs occurrence when solar radio flux exceeds the reference threshold and triggers, consequently, their solar radio observations.

The better angular, spectral and temporal resolution as well as sensitivity of these large instruments could not only provide significant insights into solar physics, but also contribute to improve the diagnostic capabilities of TSRS 2.0 itself.

The installation of the radio telescope at the observing station of the Trieste Astronomical Observatory in Basovizza began in September 2020 with the structural reinforcement of the building roof, on which the antenna has been installed (see Fig 2.9). The activity culminated with the creation of a basement on the roof of the observing station right above the control room of the telescope. The specific position has been chosen in order to: (i) maximise the observing time by excluding possible obstacles on the line of sight during the daily and annual movement of the Sun across the sky at the specific site; (ii) minimize possible dispersion and delays



Figure 2.9: Picture from the installation of the TSRS 2.0 antenna system in Basovizza (OAT), Trieste, Italy.

in the data transport from the antenna to the data processing and real time archiving facilities; (iii) minimise incidents such as fiber cuts due to ground work. Even though all the items were delivered to the observatory in early September 2020, the installation of the radio system could not be completed until 2022. Unfortunately we are still facing serious hardware issues. Training sessions are also planned, once we are ready for the "first light", to allow the personnel in Trieste to proficiently operate the radio telescope, understand possible malfunctions and identify problems and failures when they occur.

In our team, about 11 specialists of INAF-Trieste with expertise in different areas such as astronomy, Space Weather physics, radio signals and instrumentation, control software, data archiving, interfaces and management collaborate to guarantee a full 24/7 service to the scientific community, other radio telescopes and potentially other external stakeholders, such as GNSS systems, aviation and others. To understand how to better serve the scientific community, preliminary analyses of the solar radio data products currently available through the ESA SSA Space Weather Service Network [4] and SWPC (Space Weather Prediction Server)[3] portals have been performed. Data collected by TSRS 2.0 can be used for refining models used to study the evolution of CMEs and SEPs (by for example by improving the accuracy of the predicted CME arrival time on Earth [100]). They can constitute useful information for civil avionics[39] and GPS-based services (by identifying radar disturbances relevant to the Air Traffic Control cause by solar activity and by

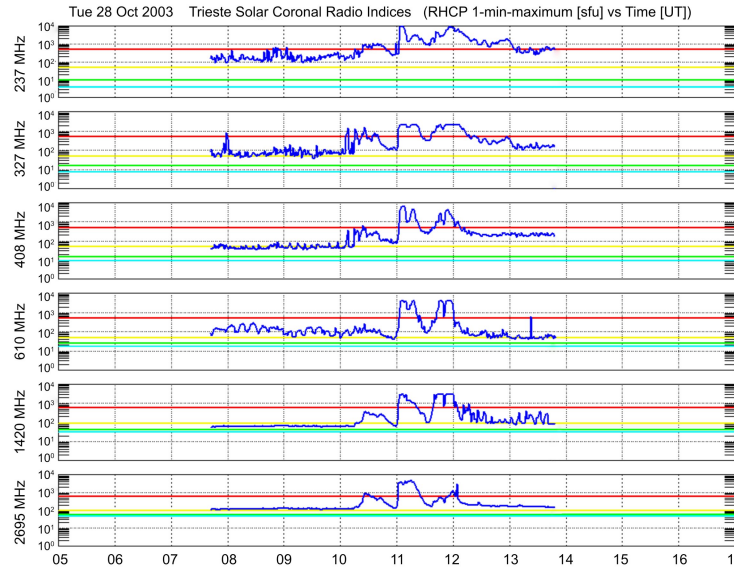


Figure 2.10: Time evolution of the TSRS 1.0 averaged (LHCP+RHCP) coronal radio indices for 28 October 2003. Different threshold levels are emphasised by colour-based coding and are associated with different levels of warning.

helping identifying the cause of reduced SNR and Loss of Lock for GNSS systems). Moreover, satellite tracking orbital decays models include solar radio emission data at 1 GHz and 2.880 GHz, as an input to calculate drag temperature along satellite orbits.

TSRS 2.0 data use cases are both commercial and scientific users, whose consider fundamental the reliability of their radio systems for communications or positioning. We will provide the possibility to the GNSS users to register to a mailing list in order to receive warnings once a threshold on radio is overcome, SFU in the L-band RHCP. For this purpose, we can leverage the experience already gained with TSRS 1.0, where such functionality has already been implemented, as shown in Fig 2.10 which reports, just as an example, the time evolution of the total radio indices for the set of frequencies used by the previous TSRS 1.0 system. The sudden onset in the flux measured in SFU is well correlated with the occurrence of the massive X17.1 flare on 28 October 2003.

Software prototype for automatic mailing alert system has already been set up and tested by the use of different data, currently available in near-real-time, such as the x-rays flux from GOES satellite.

Users will also provide feedbacks on specific solar radio events when they occur, so that a tuning process to establish operational thresholds could be refined.

The algorithms implemented in the following Chapter for solar indices forecasting are already trained and tested, therefore they are now ready to be fed with real time TSRS 2.0 data. The predicted values can be used to send warnings when enhanced solar radio emission is anticipated.

2.3 Commissioning Test Planning

The commissioning phase commenced immediately following the installation of the instrument in 2022, as this is a crucial step in ensuring the functionality, calibration and observational capabilities of the system are verified and tested. This is essential to guarantee the system is fully operational and able to perform radio observations without any performance issues, and to achieve the desired scientific outcomes. Following the installation of the antenna, the initial acquisition tests have been initiated using the Radio Telescope Control (RTC) Software, the graphical user interface (GUI) of which is illustrated in Fig.2.3. The RTC Software version 11.A.1 is a standalone application aimed at controlling the radio telescope and relevant data acquisition. It has been installed on a Windows 10 (64-bit) workstation after ensuring the proper installation of the antenna and the related drivers of the two receivers.

The anticipated outcomes of the commissioning phase comprise an evaluation of



Figure 2.11: GUI of the Radio Telescope Control Software

the pointing precision and the necessary radio calibration, the detection and characterization of the RFI environment, the characterization of the system temperature as a function of elevation position for both circular polarisations, and the definition of the gain curve.

The maximum time resolution was found to be 0.2 s in continuum acquisition mode, which is deemed to be an acceptable timescale for the purposes of this study, particularly when compared to the 1 s time resolution of the US Radio Solar Telescope Network (RSTN). Moreover, the United States does not provide circular polarisation measurements, but only the total radio flux density. During the commissioning phase, all observing modes must be tested and calibrated with the identified calibration sources.

Figure 2.3 illustrates the initial acquisition in continuum mode, with the antenna oriented towards the Sun, during the calibration attempt. Unfortunately, during

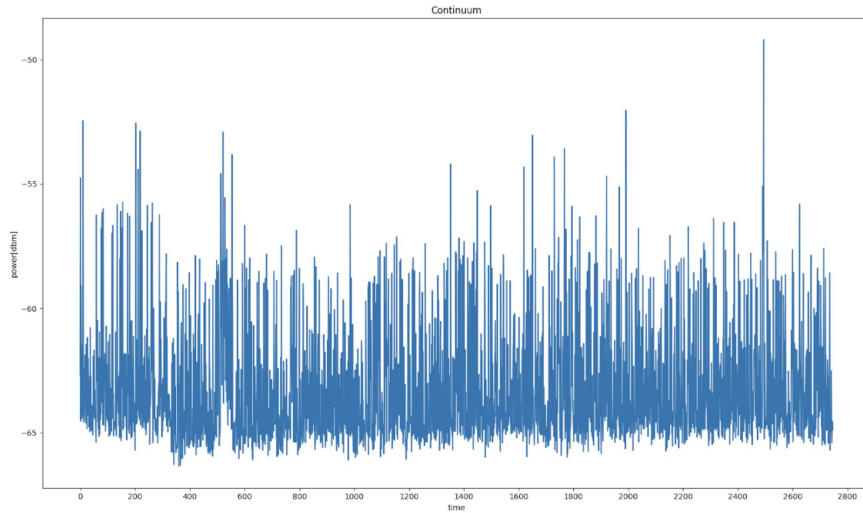


Figure 2.12: Example of the continuum power for TSRS 2.0 Ch2 (LHCP) acquired during our first tests.

first tests, we noticed that the response of the system does not change whether the antenna is pointed at the Sun or to the galactic source 3C218 or to the sky region with no (astrophysical) radio sources. This may indicate an issue in all the receiving chains or that the system has insufficient sensitivity to detect solar radio signals.

Other serious issues arose during our testing activities: the receiver associated with the left polarization seems to be noisier and to produce a signal higher than the left one and the file saving mechanism provided by the RTC Software does not allow us

to use the system for nowcasting purposes or to get alerts. This is a strong limitation of the system, which does not meet one of our most stringent requirements.

For monitoring purposes, which is the main goal of this instrument, the control software should allow a programmable operating mode, capable of providing daily unattended observations and data storage with real-time access.

To overcome all the software limitations and address the most important requirements for the system, we are still working with POAM Electronics to fix the software and hardware issues.

2.4 TSRS 1.0 Heritage and Design of the TSRS 2.0 Archive

In the framework of the forthcoming new TSRS 2.0 solar radio maps and indices, we conducted a study to refurbish and modernise the digital component of the Trieste Solar Radio System Heritage Archive (TSRS-HA), which will also host our TSRS 2.0 data.

This archive holds a trove of observational data collected by the TSRS 1.0 instruments from 1969 to 2010 under the management of the INAF - Astronomical Observatory of Trieste. It comprises decimetric and metric coronal radio measurements, providing high-time resolution and circular polarisation information, which is notably absent in most past and present solar radio instruments [135].

The TSRS-HA archive has played a crucial role in advancing research in solar radio physics, contributing to significant discoveries such as the identification of fine patterns known as "fibers" and "zebras" [122], alongside correlations between solar radio bursts and disruptions in the Global Navigation Satellite System (GNSS) [138]. Fig. 2.13, showing a SRB observed from TSRS 1.0 in 2005, is an example of the wide variety of radio burst profiles in indices time series and the solar radio flux measured for each circular polarization component can be different. The degree of polarization with high time resolution plays a fundamental role for Space Weather applications.

Based on past TSRS 1.0 experience and inspired by new technological approaches, fundamental work has been done to study the main use cases and to design the new archive structure. The new archive can be used for our forthcoming data but also to organize archival data, which still have the potential to provide important physical information.

In this Section, this analysis study is presented for the refurbishment of the Trieste Solar Radio System Archive, adopting state-of-the-art approaches [119].

TSRS-HA aims to enable scientific advancements across multiple domains:

- **Advancement in Solar Radio Physics:** by analyzing untapped data and leveraging high-time resolution and circular polarization information (Fig.2.13).
- **Reinterpretation of Historical Events:** by reexamining and refining interpretations of radio signatures and characterising underlying physical mechanisms.
- **Improvement in Space Weather Predictions Models:** through post-event and case studies analysis, establish robust criteria for assessing the potential impact of solar activity.
- **Promotion of International Collaborations and Cross-disciplinary Research:** Open access to the TSRS 1.0 archive will enable researchers to correlate TSRS 1.0 data with complementary ground-based and space-borne observations (Fig. 2.4).
- **Development of Advanced Forecasting Algorithms:** by harnessing TSRS 1.0 data for the development of advanced Deep Learning algorithms for multivariate solar radio time series forecasting, enabling automatic analysis and near-real-time nowcasting services
- **Understanding the SRBs effects on GNSS:** Leveraging circular polarization information enables to study the correlation of disturbances in the GNSS and radar signal, mobile communication with historical solar radio events registered by TSRS 1.0 and customise alert threshold [139]

The TSRS-HA houses approximately 2 TB of compressed data within the INAF-IA2 tape archive facility for long-term preservation. The data is organised into two main sections: the Radio Archive (RARC) and the SOLar Radio Archive (SOLRA). RARC contains raw data in a proprietary binary format, encompassing LHCP/RHCP time series at six distinct frequencies (237, 327, 408, 610, 1420, and 2695 MHz) along with quick-look images representing the 10-minute average flux per frequency. SOLRA holds calibrated radio polarimetric data in FITS binary tables organised into daily blocks representing millisecond radio flux for each 10-minute interval, along with solar indices derived from the maximum and average flux per second.

For the modernisation process, we conducted a state-of-the-art analysis. Our project aims to optimize data management processes and improve user experience by implementing advanced data preservation strategies and incorporating FAIR

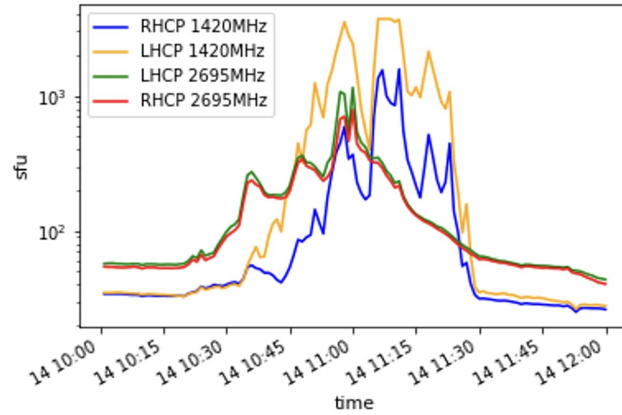


Figure 2.13: 14 July 2005 TSRS 1.0 1-min Solar Radio Indices at 1420 MHz and 2695 MHz (frequencies of interest when dealing with GNSS communication) in both Left and Right Hand Circular Polarization Component.

principles into the archive architecture. This includes addressing inadequate directory structures and data models by introducing new standardised metadata aligned with specific requirements and recommendations outlined by adopted standards. The EPN-TAP framework, the Heliophysics Application Programming Interface (HAPI), and the Space Physics Archive Search and Extract (SPASE) Metadata Model are key frameworks under consideration.

The EPN-TAP framework, developed by the Europlanet Research Infrastructure, enables efficient data discovery and retrieval through standardised interfaces based on the widely adopted IVOA [62] Table Access Protocol (TAP) standard in astronomy [50]. Furthermore, HAPI, which is utilised by numerous organisations including NASA, provides specific functionalities for the management of time series data, thereby facilitating enhanced data discovery and access [208]. The SPASE Metadata Model provides another standardised approach for describing and organising metadata related to heliophysics data. Adoption of this framework could enhance the consistency and interoperability of metadata across different datasets, particularly given its widespread acceptance by institutions such as NASA and ESA [170]. It is then necessary to devote particular attention to the process of data curation. This encompasses the assessment of data quality and the implementation of flagging mechanisms to detect and address instances of invalid or corrupted data. Furthermore, image-driven inspection tools that employ vector databases (e.g. Pinecone, Chroma-DB, or the PostgreSQL vector extension) and AI have been investigated as a means of facilitating event identification.

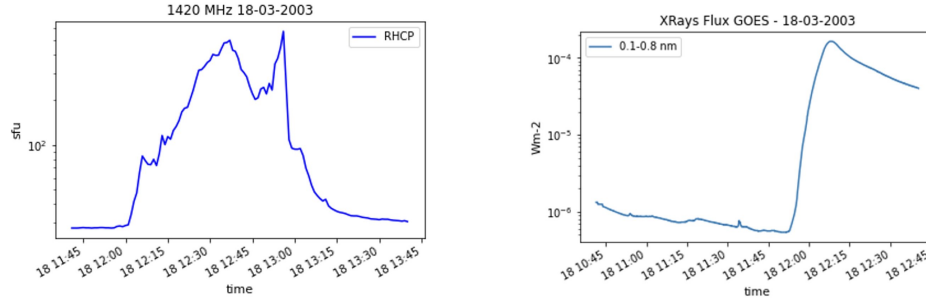


Figure 2.14: Solar radio burst from TSRS 1.0 and the correspondent solar flare measured from GOES satellite for two events: 18 March 2003 and 3 November 2003

The time series forecasting techniques that have been tested and explored in Chapter 2 of the present work can also be used for TSRS-HA data once the catalogue is ready and easily consultable.

2.5 Perspective for Automatic SRB Analysis and Detection

From an operational monitoring perspective, a neural network has been developed for the automatic detection of solar radio bursts in the 1420 MHz time series, utilising both circular polarisation components. This system is capable of detecting the presence of a solar radio burst in the last two hours. The proposed methodology can be employed as a valuable tool for the automatic monitoring of radio bursts, utilising temporal moving windowing to scan the entire data set in real-time.

To emphasise the significance of a comprehensive and accessible archive, we have implemented a preliminary prototype to assess this methodology, utilising the 1-m average indices TSRS 1.0 dataset. A small number of events, both with and without solar radio bursts, have been manually selected for training the Fully Connected Recurrent Neural Network.

The network details are reported in Fig. 2.5; the reached accuracy for the test dataset is now 0.947 but still with a validation loss of 0.29. Therefore, further improvement will be implemented with fine-tuning of the hyperparameters and also selecting a larger dataset for the training and validation processes. When applying this process to the accessible archive it can also help to automatically create a SRB catalogue. In particular, for this purposes, the newly implemented automatic bursts detection network could be trained and tested by the use of e-Callisto time series data, although its frequency range goes from 45 MHz to 870 MHz [75].

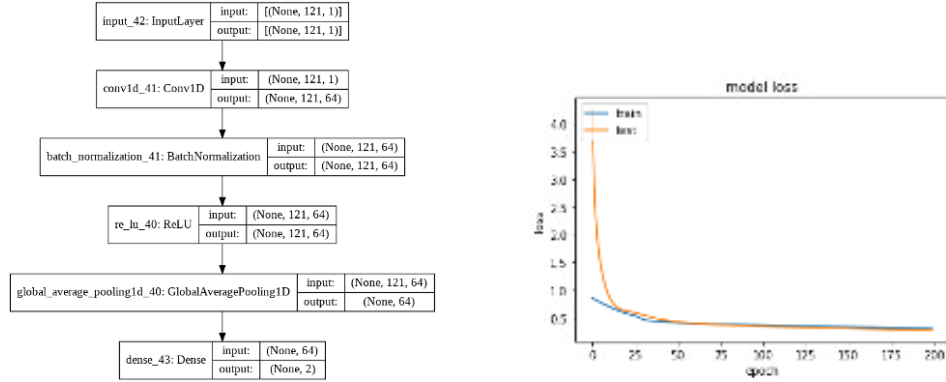


Figure 2.15: Neural Network Architecture for the automatic detection prototype. Loss=0.29, Accuracy=0.947

Furthermore, methodologies will be developed to address inconsistencies within the dataset resulting from periodic calibration procedures, thereby ensuring a comprehensive and uninterrupted dataset. Robust procedures will be established to facilitate the seamless integration of data from disparate formats and proprietary software into a unified, standardised and open-source solution. The project will culminate in the delivery of a user-friendly and lightweight service, specifically tailored to the requirements of the heliophysics community. The service will be based on a microservice architecture deployed as a containerised framework, thereby facilitating seamless integration of various community standards and applications (such as Jupyter Notebook). The service will be subjected to a series of tests to ensure that it can be accessed by a carefully selected subset of TSRS-HA data, which will cover specific frequencies and time periods. It will offer basic functionalities for data exploration and analysis. The service will be designed with modularity and scalability as key priorities, in order to facilitate future enhancements based on user feedback.

The long high time resolution TSRS-HA data can be used to identify peculiar characteristics of the signals, as solar radio bursts exhibit complex temporal behaviour and show a wide variety of temporal profiles. It is of interest to undertake a comparative study of events in order to identify whether there are any similarities, differences or common features among them. For example, statistical studies may be employed in this regard. The Gaussian fit parameters of the time profile, the cumulative integration, the local slope of the curves, and the second derivative have been This approach could facilitate the clustering of solar radio events.

Furthermore, the Hurst exponent, which is not directly related to plasma physical

processes, may offer insights into the underlying dynamics and behaviour of certain plasma systems, potentially assisting in the classification of solar radio bursts. The Hurst exponent (H) can be computed using the rescaled range (R/S) analysis method, a statistical measure to evaluate randomness in a data set.

The general procedure for the Hurst exponent H computation of a time series dataset $\{X_t\}$ (as in the example analyzed in Figure 2.5) is summarised in the following. For each segment length n , the sample mean $\bar{X}_n$ and the standard deviation S_n of the data within the segment are calculated.

The rescaled range R_n is then computed, as the difference between the maximum and minimum values within the segment, normalized by the standard deviation:

$$R_n = \max(X_t) - \min(X_t) \quad \text{where} \quad t \in [1, n]$$

$$R_n = \frac{R_n}{S_n}$$

The rescaled ranges R_n is averaged across all segments of length n . R_n is plotted against n , where n is the segment length, in a log-log graph in order to perform linear regression and estimate the slope, which corresponds to the Hurst exponent (H), as shown in the example depicted in Figure 2.5. If $\log(R_n)$ is regressed against $\log(n)$, the linear regression equation is:

$$\log(R_n) = \log(C) + H \cdot \log(n)$$

where C is a constant. The aforementioned technique has been used for Gamma

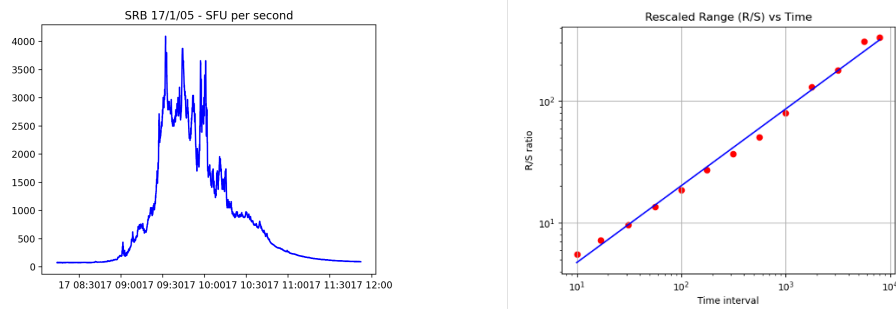


Figure 2.16: Solar Radio Burst (17th January 2005), from San Vito observatory (RSTN) and Rescaled Range for each of the sub-segment of the SRB, for the computation of the Hurst exponent.

Ray Burst ([121]) statistical analysis and the authors obtained that the mean Hurst exponent for the short GRBs is significantly smaller than that for the long GRBs. In Table 2.5 we report the early analysis for few SRBs from the Radio Solar Telescope Network (RSTN) observatories, that will be very interesting to continue when full access to TSRS 1.0 data will be possible. It seems that when radio bursts last few minutes, they have a lower H , while long SRBs have greater Hurst exponent values.

It would be useful as a discriminant feature also for SRBs classification, but it will still need a further investigation, with a larger dataset, for a meaningful statistic, that we will obtain with the catalogue of high time resolution TSRS-HA. The physical

Type II SRB	Event Duration	H
18-03-2003	11:50-13:20	0.60
17-01-2005	09:10-11:00	0.55
23-08-2005	15:17-17:15	0.54
01-08-2010	08:50-0.70	0.70
09-09-2005	10:04-10:06	0.38
18-11-2003	08:23-08:35	0.49

Table 2.5: The Hurst exponent obtained for each event, from a preliminary study, using Range Scale Analysis.

differences between long and short SRBs arise from the underlying processes driving the emission. They exhibit complex temporal behavior, and understanding their characteristics is crucial for Space Weather forecasting and solar physics research. SRBs, whether long or short, can have distinct effects on Space Weather due to their different characteristics and underlying physical mechanisms.

Long bursts, such as Type II and IV bursts, are often associated with the eruption of coronal mass ejections (CMEs). These bursts, typically longer in duration, can trigger geomagnetic storms when they interact with the Earth's magnetosphere. These storms are closely linked to turbulent physical processes in the solar corona and interplanetary space. CMEs are driven by the release of magnetic energy stored in the solar atmosphere, leading to the rapid expansion of magnetized plasma into the solar wind. The interaction of CME-driven shock waves with pre-existing solar wind structures can generate turbulence, magnetic reconnection events, and particle acceleration, contributing to the observed radio emission during long bursts, particularly in the L-band frequency range.

On the other hand, short bursts, like Type III and Type I bursts, are often associated with solar flares. Although shorter in duration and less intense, can still disrupt

radio communications sporadically, particularly in the high-frequency (HF) band, due to their rapid onset and potential for strong radio emission.

The physical processes driving short bursts involve localized energy release events in the solar corona, such as magnetic reconnection and particle acceleration. In the case of Type III bursts, energetic electrons accelerated during magnetic reconnection events travel along open magnetic field lines, producing intense radio emission at decimeter wavelengths, including the L-band. Type I bursts, on the other hand, are associated with the rapid release of energy during solar flares, leading to enhanced radio emission across a broad frequency range, including the L-band.

While short bursts may not produce as intense SEP events as long bursts, they can still contribute to the overall radiation environment in space during periods of heightened solar activity.

Understanding these differences between long and short solar radio bursts and their turbulent physical processes, particularly in the L-band frequency range, is crucial for Space Weather forecasting and mitigating potential impacts on technological systems and infrastructure.

Chapter 3

Time Series Forecasting for Space Weather Applications

Time series are sets of measurements performed with a certain cadence over a period of time. They have many applications in various fields, from science to economics and medicine, providing a means to understand the mechanisms underlying the observed phenomena.

In the work presented in this Chapter, after a short review on the basic concepts of time series and their forecasting techniques, several tools based on different algorithms will be exploited and compared, particularly for solar radio index forecasting. These tools, appropriately tuned on different data, such as the radio index data measured by TSRS 2.0, can be a useful means that we have also set up for nowcasting and operational forecasting, ready to be re-trained as soon as new data will become available.

A multivariate Deep Learning model that combines the Fast Iterative Filtering (FIF) algorithm, the Long-Short Term Memory (LSTM) network and a Multi-Attention module for short-term forecasting of F10.7 index will be also presented.

Three methods for F10.7 time series forecasting will be presented: FIF-LSTM-MHA model and its comparison with classic LSTM, its upgrade by the use of Multivariate FIF-LSTM-MHA algorithm and a long term forecasting as a proxy solar activity forecasting.

Most of the results are also illustrated in our recent published papers [89, 125].

An important section of this chapter is also dedicated to studying the different significance of input driver for ionospheric global TEC prediction models, currently utilized in operational contexts for disturbances on GNSS following Space Weather

events. The results show that solar radio data are fundamental for this applications and it highlights the crucial role of continuous and reliable observed and predicted F10.7 data.

Prediction models of TEC that take into account forecasted indices as external drivers, to include the effects of solar activity and geomagnetic disturbances, require reliable input forecasts.

3.1 Time Series Theoretical Background

The formal definition of a time series X can be as follows [124]:

$$X = \vec{x}^1, \vec{x}^2, \dots, \vec{x}^n$$

where each point $x(t) \in \mathbb{R}^m$ in the time series is an m-dimensional vector

$$x_1(t), x_2(t), \dots, x_m(t)$$

, whose elements correspond to the input variables. From a statistical point of view, time series can be described as stationary or non-stationary, where a strictly stationary time series is defined by the property that all its statistical moments remain constant over time.

It involves the use of a set of methods and techniques to predict future values based on past observations or data points that are collected over time. This capability can be extremely valuable for organizations as it facilitates data-driven decision-making, aids in planning processes, optimizes resource allocation and addresses challenges associated with Big Data.

Time series forecasting is a rapidly evolving field of research that finds applications in various disciplines [217]. By leveraging patterns and trends identified in past data, time series forecasting techniques enable accurate predictions across diverse domains such as economics, engineering, social studies, natural sciences, climate and weather modeling [175].

In particular, this work focuses on the prediction of solar radio flux, which is a relevant example of an application within the Space Weather framework.

The simplest method for time series forecasting is univariate time series forecasting, which is the procedure for computing future values of a single time series variable based on its past values [157]. Let's consider a sequence of observations for a single

time series

$$y = y(t - D), \dots, y(t - 2), y(t - 1), y(t)$$

y_t represents the value of the time series variable at time t , and D represents the number of historical data values available. Our goal is to forecast the future terms y_{t+H} , where H represents the time horizon and can take values from 1 to h . Multivariate time series forecasting: In this case, multivariate time series involves multiple variables and can be expressed in matrix form as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} y_1(t-D) & \dots & y_1(t-2) & y_1(t-1) & y_1(t) \\ y_2(t-D) & \dots & y_2(t-2) & y_2(t-1) & y_2(t) \\ \dots & \dots & \dots & \dots & \dots \\ y_n(t-D) & \dots & y_n(t-2) & y_n(t-1) & y_n(t) \end{pmatrix}$$

where $y_i(t-j)$ identifies the set of time series, with $i = 1, 2, \dots, n$, being $m = 0, \dots, D$ the historical data. In multivariate time series forecasting, models take multiple variables along with their respective historical data as inputs to predict specific values of the target time series [205]. To capture relationships and dependencies among different variables, different strategies, such as feature selection, can be employed. These techniques identify the optimal set of features that impacts the target variable, thereby enhancing the performance of deep learning models. Additionally, they help prevent the risk of over-fitting and reduce computational costs. Moreover it is important to introduce the difference between two approaches in time series forecasting: one-step and multi-step prediction.

The former refers to the use of a model to make a prediction one-step ahead in time, i.e. estimating the value of x_{t+1} denoted by $\tilde{x}_{t+1}$, by minimizing the error represented as $\tilde{x}_{t+1} - x_{t+1}$. The latter refers to the process of forecasting multiple future steps ahead in time in a time series [34]. Specifically, in this work, we present a multistep F10.7 time series forecasting method for up to 25 days.

3.2 Time Series Forecasting Models

The field of time series forecasting has seen the development of a plethora of models over the years. Initially, in the 1970s, traditional statistical models such as AutoRegressive (AR), Moving Average (MA), AutoRegressive Moving Average (ARMA) and AutoRegressive Moving Integrated Average (ARIMA) were prevalent [123]. These models, based on the Box-Jenkins methodology, were effective in capturing

linear dependencies in stationary time series but they faced limitations when dealing with large-scale data sets or complex patterns.

When stationarity and linearity can be assumed, a plethora of well-understood methods, such as autoregressive and moving average techniques along with their combinations and flavours, provide effective ways to describe the signal and to forecast its future values. A general definition of a linear process is as follows: X_t is said to be linear if it has the representation $X_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}$, where ϵ_t are uncorrelated random variables with zero mean and unitary variance. A thorough description of how to analyse linear or non-linear stationary time series can be found in [29].

Unfortunately, the majority of real-world phenomena, e.g solar radio emission, are generated by non-linear processes and show time-changing statistics and transient conditions. The time series that describe them are therefore both non-stationary and non-linear and require a different set of techniques to be understood and forecasted. With the increasing data availability and computing power, Machine Learning and Deep Learning techniques have emerged as a crucial component of next-generation time series forecasting models [197].

Machine learning and deep learning methods offer an automatic learning ability from data without any explicit programming needs and are able to discover the hidden causal relationships between input and output. These methods exhibit relevant features particularly in the time series forecasting framework, because they have excellent generalization performance and are able to capture non-linearity. The physics-based models are generally expensive and often impossible to run in real-time; on the contrary, in data-driven models, the training is an expensive process depending on the data size but the execution is very fast.

These models outperform traditional approaches by leveraging their ability to learn temporal dynamics in a data-driven manner without the need for explicit programming. They employ Artificial Neural Networks (ANN) such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM), which significantly improve prediction performance by identifying causal relationships between input and output while capturing non-linearities within the data set.

3.2.1 Long-Short Time Memory Neural Networks

In the realm of sequence modeling and time series analysis, Long Short-Term Memory (LSTM) networks have emerged as a powerful tool for capturing long-range dependencies and patterns within sequential data.

They were first introduced in 1997 [76] and have since been refined to imitate human behaviour in handling sequential data [57].

LSTM models, which belong to the family of RNNs, have gained popularity in time series forecasting applications, addressing the limitations in retaining information over extended time intervals.

Several papers [116, 227] have highlighted the suitability of LSTM models for F10.7 forecasting.

LSTM networks are specifically designed to address the Vanishing Gradient Problem, which is a common issue for RNNs that hinders their ability to learn long-term dependencies in sequential data sets, resulting in slow and ineffective training.

LSTMs overcome this problem by utilizing three gates: the forget gate, the input gate and the output gate. These gates allow for better control over gradient flow, preserve important long-term information and discard irrelevant input data.

The LSTM model has a chain-like structure (see Fig. 3.1) consisting of repeating basic units, called *memory cells*. In each memory cell, the three gates (represented

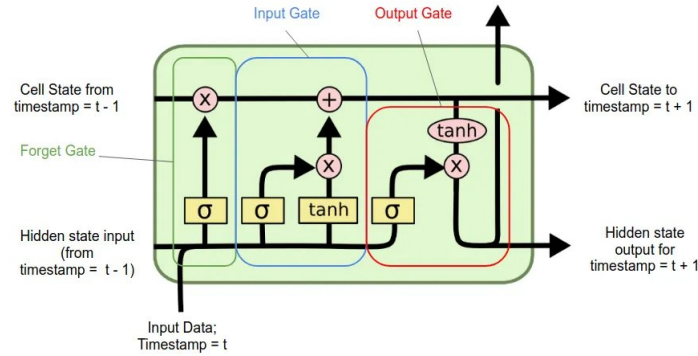


Figure 3.1: The repeating module in an LSTM contains three interacting layers: a sigmoid **forget gate layer** f_t to decide what information we're going to throw away from the cell state; a sigmoid layer i_t called **input gate layer** decides which values we'll update; a tanh layer creates a vector of new candidate values, C_t , that could be added to the state. In the next step, we'll combine these two to create an update to the state, the **output gate layer** o_t . The $\times$ and $+$ operators mean an element-wise vector multiplication and sum.

as f_t , i_t and o_t , respectively, in Fig. 3.1) work together to selectively update and utilize the cell state C_t , which is the activation term holding the long-term information from previous inputs.

The input gate i_t regulates the flow of input data in the cell state of the LSTM unit, by determining which information in the current cell state should be updated and what information is relevant for accurate predictions. The forget gate f_t is respon-

sible for discarding the useless information from the previous step and updating the cell state. Finally, the output gate o_t determines the output value to be passed to the next hidden unit.

The equations governing the gates in LSTM are as follows:

$$\begin{aligned} f_t &= \sigma(w_f[h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(w_i[h_{t-1}, x_t] + b_i) \\ o_t &= \sigma(w_o[h_{t-1}, x_t] + b_o) \end{aligned}$$

where σ represents the sigmoid activation function, which computes the gate values as binary outputs. A value close to 0 means that the past information is forgotten, while a value close to 1 means the information is kept for future use. w_x and b_x denote respectively weights and biases, that govern the behaviour of the gates. The equations for the cell state, candidate cell state, and the final output are, respectively, as follows:

$$\begin{aligned} \tilde{c}_t &= \tanh(w_c[h_{t-1}, x_t] + b_c) \\ c_t &= i_t * \tilde{c}_t + f_t * c_{t-1} \\ h_t &= o_t * \tanh c_t, \end{aligned}$$

where $\tanh$ represents the hyperbolic tangent activation function and $*$ denotes element-wise multiplication of vectors.

3.2.2 Self-Attention and Transformers

While LSTM networks have proven to be highly effective in capturing temporal dependencies within sequential data, recent advancements in deep learning have introduced alternative architectures that offer complementary benefits.

Despite the memory function of the LSTM model and its ability to retain time series information, it still possesses certain limitations due to its utilization of the conventional encoder-decoder structure. While processing the time series, the Encoder initially encodes the input sequence into a fixed-length implicit vector and assigns uniform weight to the vector. However, as the length of the time series increases, the average weight distribution diminishes the ability to distinguish elements within the series. Consequently, crucial time-series information may be overlooked during the model training process, thereby impacting the accuracy of predictions. In this study, the Attention module is employed to enhance the performance of the Encoder-Decoder structural model.

Transformers are a type of neural network architecture introduced by Vaswani et al. [200], marking a departure from traditional recurrent and convolutional neural network architectures by relying solely on self-attention mechanisms for capturing dependencies across input and output sequences.

The Attention-based architecture represents a significant breakthrough in deep learning, drawing inspiration from human cognitive processes [17]. When seamlessly integrated into an LSTM architecture, the Attention module enhances the model's adaptability and effectiveness.

The capacity of the system to prioritise the elements relevant to the input, effectively suppressing irrelevant information, has been exploited for solar radio time series forecasting. Its primary advantage lies in its ability to consider both global and local connections in a single step, enabling parallel computation, a particularly crucial aspect for processing large-scale data. This approach allows the model to focus more effectively on important parameters, ultimately boosting its performance.

Few details of the attention mechanism and the encoding-decoding process in transformers architecture are explained in the following.

To enhance representational capacity and capture different aspects of the input sequence, Transformers employ multi-head attention. Multi-head attention consists of multiple parallel self-attention mechanisms, or "heads". The outputs of the multiple heads are concatenated and linearly transformed to obtain the final multi-head attention output. The use of this technique permits the different heads to focus on different correlations in data and enhance different values in the sequence. The transformer model is constituted by encoder and decoder blocks. Each block feeds on the result of the block before (Fig.3.2). The idea around the model is that the encoder layers will have the job to encode in the vectors all the relations found in the data. The encoders get as input the encoding from the precedent encoder layer and the first one gets the input data. Encoder also returns the encoded matrix in which all the gathered relations are written. The decoder does the opposite task. It receives as input the result of the encoding stack and decodes the matrix to predict the actual result. Both encoder and decoder are composed of multiple layers, each containing self-attention mechanisms and feed-forward neural networks.

The encoder module consists of a stack of repeated encoder units. Each unit is designed in the same way. Each encoder layer has two sub-layers. First there's a multi-head self-attention module, as the one described above, the second is a fully connected feed-forward layer. Each sub-layer is followed by a normalization layer that normalizes the computed sum between the sub-layer output and the input. Each

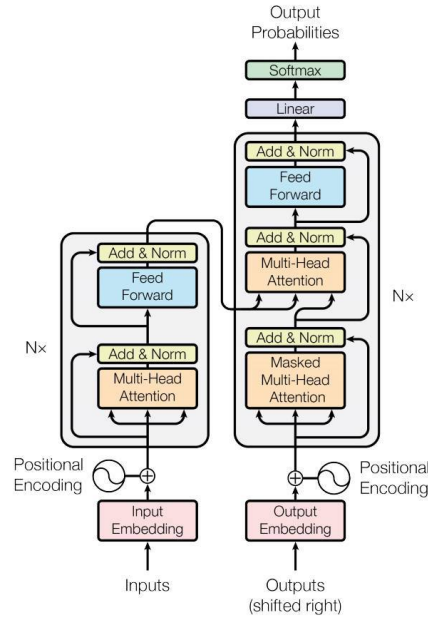


Figure 3.2: General schema of a transformer model architecture [200].

encoder should focus on different features in the input and aims to encode in the sequence all the important relations that will contribute to the prediction.

The decoder module has a similar structure. Each unit consists of three sub-layers: the first and second one are attention modules, the first is a multi-head self-attention on the output of the decoder before, the second one performs a cross-attention between the output of the first sub-layer and the output of the encoder staff. This step will decode all the data encoded before and elaborate the decoded matrix ready to make the prediction. The third sub-layer is a feed-forward layer. As mentioned before in the encoder, the output of all the sub-layers will be summed with the input and then normalized.

In the network there are different feed-forward layers: two layers in each encoder, two layers in each decoder and a small network after the decoders. The aim of the layers inside the encoders and decoders is to elaborate all the correlations learned by the attention modules. In this case the layers are very small as they don't need to massively change the results of the attentions but just help the attention layers in encoding and decoding the gathered information. The network at the end of the model is the one actively forecasting the results using the output from the last decoder. As in our work we focused on the application of the transformers for solar

radio time series forecasting, more details on their mechanisms can be found in literature [200, 190].

3.3 From Empirical Mode Decomposition (EMD) to Fast Iterative Filtering (FIF)

Signal decomposition techniques play a pivotal role in extracting meaningful information from complex signals, such as those encountered in time series analysis, biomedical signal processing, and environmental monitoring. Among the plethora of methods available, Empirical Mode Decomposition (EMD) and its derivative, Fast Iterative Filtering (FIF), have garnered significant attention for their effectiveness in decomposing non-linear and non-stationary signals into their constituents. In order to capture features in non-linear and non-stationary signals, it is necessary for the analysis methods to be local, adaptive and stable.

One approach to coping with the complexity of the F10.7 time series is firstly to decompose the signal into simpler components and then analyse them separately, according to established decomposition methods [186, 185].

The experimental results have demonstrated that decomposition methods coupled with artificial neural networks offer a significant enhancement in the field of time series prediction, allowing chaotic characteristics of the original data to be reduced. A number of techniques have been so far in the literature: the simplest method to analyse any signal is by Fourier Transform [185]. However, since the coefficients represent the amplitude of the signal as a function of frequency, the time information is lost and this method is therefore only applicable to stationary signals. Wavelet transform [186] is also a powerful variant that has been widely used in this field but it is a non-adaptive technique too since it is still a linear decomposition method and basis functions must be provided a priori.

A very successful empirical and data-driven signal decomposition method, devoted to the analysis of non-stationary and nonlinear signals and based on the Hilbert-Huang transform, was proposed by Norden E. Huang in the late 1990s [82].

EMD has been widely used in literature for physical applications. In particular, EMD provides a very good method to extract periodic components of solar activity cycles, based on SSN, highlighting characteristics of isolated features on various time scales [107]. Regarding Space Weather research, EMD has been also exploited to investigate time variability of solar activity indices, such as sunspot areas (SAs)[8], to deal with properties of the solar wind [10], to the relativistic electron

flux fluctuations during magnetic storms [165].

Recently, within the Space Weather framework, some scholars used a hybrid model based on the use of the Empirical Mode Decomposition (EMD) and the LSTM model to predict respectively the F10.7 [115], SSN values [108] and geomagnetic storm index (Dst) [218]. It requires two steps: the first one, called Empirical Mode Decomposition (EMD), divides the signal into orthogonal oscillatory components called Intrinsic Mode Functions (IMFs). The original time series $x(t)$ can be denoted as the sum of the intrinsic mode functions $c_i(t)$ and a residue $Res(t)$:

$$x(t) = \sum_{i=1}^n c_i(t) + Res(t), \quad (3.1)$$

where n and $c_i(t)$ represent the number and the i_{th} intrinsic mode function, respectively. The residue $Res(t)$ represents the overall trend of the mean value of the original time series data.

The second one, Hilbert Spectral Analysis (HSA), calculates the contribution to the total amplitude of each IMF. This is achieved by applying the Hilbert-Huang Transform to the IMFs, which results in the Hilbert spectrum.

The EMD algorithm proposed by Hung and his group suits non-linear and non-stationary data thanks to the locality and adaptivity of the base functions that are derived directly from the data.

Unlike other popular approaches, such as the Fourier and wavelet transforms, EMD allows for the decomposition of the original signal into simpler components without making any assumptions or selection of the basis function to be used in the process. The drawbacks are that the convergence of the used sifting algorithm and the orthogonality of the IMFs cannot be mathematically proved, and the high level of adaptivity can result in the instabilities of the algorithm.

Various researchers tried to address the instability of EMD and its tendency to suffer from mode mixing by developing a set of EMD-based methods [215, 225, 104].

Lin et al. in 2009 [111] developed and presented a complementary approach denominated Iterative Filtering (IF). The key idea behind this decomposition technique is separating simple oscillatory components contained in a signal $s(x)$, $x \in R$, the so called IMFs, by approximating the moving average of s and subtracting it from s itself [37]. The approximated moving average is computed by convolution of s with a window/filter function w : a non-negative and even function in $C^0([-L, L])$, $L > 0$, and such that

$$\int_R w(z)dz = \int_{-L}^L w(z)dz = 1 \quad (3.2)$$

The idea of iteratively subtracting the moving average is similar to the EMD method, where the moving average was computed as a local average between an envelope connecting the maxima and one connecting the minima of the signal under study. The moving average is computed as a convolution of the original signal and a low-pass filter. The use of envelopes in an iterative way is the reason why the EMD algorithm is still lacking a rigorous mathematical framework. The technique is also resilient to perturbations. An in-depth analysis of the algorithm and its convergence can be found in literature [38]. The IF algorithm shows the same structure of EMD but, compared with EMD-based methodology, it enjoys the following advantages: it is able to guarantee a priori convergence and stability; it is faster, and thus computationally more efficient than EMD; it can a priori estimate the errors while in the EMD-based model it is difficult.

Cicone et al. [38, 35] realised the opportunity to improve the performance of the IF method by computing each IMF via the Fast Fourier Transform (FFT).

To use the FFT, the signal needs to be periodically extended at the boundaries, and the number $N_0 \in \mathbb{N}$ of steps required to meet a certain accuracy needs to be pre-computed. Thanks to the adoption of FFT, the FIF algorithm outperforms the regular IF by a factor of 100 [38].

The FIF technique was developed with the objective of automating the data analysis process, thereby overcoming some of the shortcomings of the EMD. However, it has been observed that the IMFs produced by FIF decomposition can be more challenging to interpret in physical applications, potentially resulting in the lack of truly empirical basis functions. Nevertheless, in the context of an operational space weather service, the employment of FIF can facilitate the streamlining and automation of the analysis and forecasting process.

3.4 Forecasting of Solar Radio F10.7 Index

The solar radio flux density at 10.7 cm (2800 MHz), commonly referred as the F10.7 index, is arguably one of the most suitable and widely used solar indices to describe the evolution of the solar activity cycle at different time scales. The F10.7 index is defined as the total power flux density emitted by all sources on the full solar disk at the wavelength of 10.7 cm within a bandwidth of 100 MHz centered on 2800 MHz [193, 189]. The F10.7 value is expressed in solar flux units (sfu), where $1 \text{ sfu} = 10^{-22} \text{ W m}^{-2} \text{ Hz}^{-1}$. It results from thermal ionisation at the boundary between the photosphere and the chromosphere and from magnetic

resonance above sunspots and plages [196]. In general, it is used as a proxy for the full-disk flux at the ultraviolet (UV), extreme ultraviolet (EUV), X-ray, Ca II and Mg II wavelengths, and for the total solar irradiance.

Along with the Sunspot Number (SSN), F10.7 index is characterised by one of the longest observational running records which is obtained from observations made daily from 1947 first at Ottawa, Ontario and now by the Canadian National Research Council's Herzberg Institute of Astrophysics at its ground-based Dominion Radio Astrophysical Observatory located in Penticton, British Columbia, Canada.

Solar radio flux is traditionally decomposed into three components: a constant and low-intensity background that originates from the quiet Sun and evolves on times scales of months to years; a slowly varying moderate intensity component (S-component) that originates in active regions and is rotationally modulated; and, finally, a highly variable and high-intensity component associated with radio burst emissions which is frequently used to diagnose energetic particle acceleration processes on time scales of days and beyond [160].

The daily value is given for an observation made near local noon. Typical values range from approximately 64 sfu to more than 300 sfu over the solar cycle [193]. Since the F10.7 radio flux can be measured with high accuracy and radiometric stability on a daily basis from the Earth's ground surface and in any weather conditions, in some applications the F10.7 index is preferred over other solar indices - such as the SSN, Lyman α irradiance or the Mg II core-to-wing index. The F10.7 index has long been known to correlate with the SSN but, unlike the SSN, it can be considered an objective index that does not need a homogenization procedure of its values. Moreover, F10.7 is a good proxy for determining other solar quantities and it has many applications spanning solar physics, climate modelling, geophysics, meteorology, radio communications, and satellite systems. In particular, the daily F10.7 flux is thought to be a good proxy for the solar extreme ultraviolet (EUV), ultraviolet (UV) and X rays emissions [73, 189, 48], that are the primary drivers of variations in the density of the terrestrial upper atmosphere and ionosphere [202], main source of enhanced GNSS errors in timing and positioning.

For these reasons, F10.7 serves as the prime solar input in many climate and atmospheric models which predict and reconstruct parameters of the upper atmosphere and ionosphere, such as the atmospheric density in low Earth orbit and density distribution of the ionospheric TEC. The International Reference Ionosphere (IRI) [26], NeQuick [143], JB2008 [27], the NRLMSISE-00 models [161] are some examples of relevant empirical ionospheric and thermospheric models which use

F10.7 to infer atmospheric density and temperature. Consequently, the availability of a reliable prediction of F10.7 [105] is very important and has fundamental implications under the Operational Space Weather perspective [156] where it is crucial to estimate in advance the solar effects on Space Weather applications and services. Moreover, the forecasting of F10.7 is relevant, especially at medium and short time scales, which are related to different physical processes. Long-term forecasts, which refer to timescales of solar cycles (about 11 years) and deal with the prediction of the very general features of a solar cycle, such as its amplitude, peak time and period, typically influence the planning and prediction of missions and estimation of spacecraft lifetime. A medium-term forecast of solar activity gives predictions from one to several months in advance. Short-term predictions typically refer to forecasting several days (up to 25-30 days; i.e., up to one solar rotation) in advance and are important for spacecraft operations, such as orbit determination and conjunction analysis (collision avoidance maneuvers [132] re-entry, and orbital debris predictions [198]). Anyway, due to its daily variability, the high-accuracy forecasting of F10.7 in the short and mid-term is often challenging to accomplish. In recent literature, various F10.7 index forecasting methods have been proposed. According to the prediction taxonomy, these works can be divided into three main categories. The first group comprises the so-called precursor methods which are usually used for long-term forecasting and are based on the identification of solar activity indices and correlations among them in order to predict the maximum strength of the next solar cycle [95, 30, 67, 148].

The second group makes use of physics-based predictions which combine models constructed from observed data with models based on physical laws [66, 114, 72, 214].

A third group comprehends a variety of statistical methods, which do not require any knowledge of the physical phenomena involved and can be applied for short and medium-term forecasting. It comprehends modern machine learning and deep learning models, which are based on neural network techniques and outperform models based on traditional approaches, due to their ability to learn temporal dynamics in a purely data-driven manner.

Although the use of these techniques is not yet the standard in solar proxy forecasting since many available forecasts rely on traditional approaches (precursor, physics-based or statistical methods), in recent times, with the increasing data availability and computing power, they are getting a vital part of the next generation time series forecasting models and present some advantages [31, 110].

Beneath its promising facade, time series prediction ML-models face significant challenges, particularly related to the quality and quantity of training data.

Indeed, these models require large volumes of high-quality data that accurately represent the temporal patterns of interest. However, obtaining such data can be problematic, as it often involves dealing with incomplete, noisy, or inconsistent datasets. The lack of sufficient historical data can also hinder the model's ability to capture complex temporal dynamics and patterns, leading to less reliable forecasts. Moreover, ensuring data quality through rigorous preprocessing steps is essential but can be resource-intensive and time-consuming. These data limitations underscore the importance of careful data collection and preparation to enhance the robustness and accuracy of time series forecasting models.

Another drawbacks of ML algorithms, that are often referred to as "black boxes", is the lack of transparency and interpretability.

A number of deep learning architectures characterized by different performance and fitting abilities have been developed for the daily short-term F10.7 forecasting [80, 116, 188, 219] but the prediction accuracy is still expected to be enhanced.

In order to further improve the prediction effect, some researchers also applied decomposition approaches by extracting the signal characteristics of the original data with signal processing methods and then predicting it with machine learning techniques.

In the following sections we propose deep learning approaches for the daily forecasting of the adjusted F10.7 solar radio flux, using a family of LSTM-based models combined with a decomposition method. In particular, we apply two approaches that, to our best knowledge, have never been adopted before in this specific forecasting framework.

All the proposed architecture have been implemented in Python language taking advantage of the Tensorflow library, which is dedicated to various machine learning tasks, and Keras, a neural network library.

3.4.1 FIF-LSTM-MHA Univariate Model

In this section, we present a comparative study among three neural network architectures specifically designed for the univariate prediction of the solar radio index F10.7: LSTM, adopted as baseline; LSTM model used in combination with FIF as decomposition algorithm (hereafter referred to as FIF-LSTM); and FIF-LSTM-MHA model, with the additional use of Multi-Head-Attention architecture.

Comparative analysis of these different forecasting approaches is evaluated in terms

of their performance and forecasting accuracy, in order to highlight the advantages and disadvantages of one model rather than the other.

The results of the original work proposed in this Chapter have been recently published in Jerse, Marcucci, 2024 [89].

As depicted in Figure 3.3, for each model we adopted a methodology process that encompass data preparation, model training, hyperparameter tuning and rigorous performance evaluation, with the ultimate objective of building a robust neural network to achieve precise forecasting. Before inputting the data into our neural

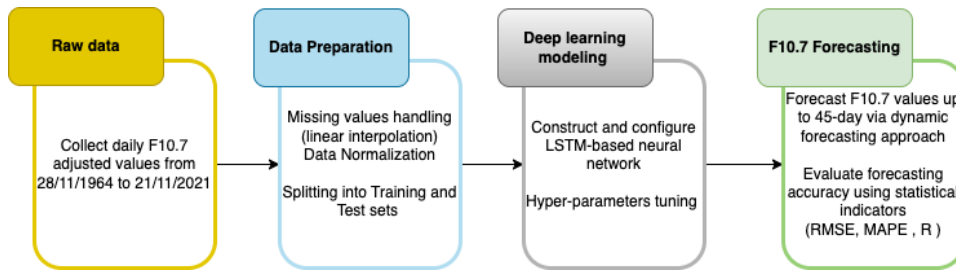


Figure 3.3: Overview of flow-chart of the deep learning-based forecasting framework.

network models, we conducted several essential preprocessing steps for the LSTM model, as depicted in Figure 3.4 and detailed in the following:

- **Data Collection** We utilized daily F10.7 data, which is publicly accessible through NASA's OMNIWeb Data Explorer website [58]. This website is maintained by the National Space Science Data Center (NSSDC) of NASA. F10.7 data is available in two forms: observed values and adjusted values. The observed values vary throughout the year due to the changing Sun-Earth separation, whereas the adjusted values are observations that have been standardized to 1 Astronomical Unit (1 AU). For our research, we used the adjusted data, which is commonly used in several thermosphere models, as noted in [198]. The dataset we used spans from November 1963 to November 2021, covering more than 6 solar cycles. It consists of 21,180 records, with F10.7 index values ranging from 63 to 367 SFU.
- **Missing Values Handling** Occasionally, the dataset is affected by data gaps, which we addressed by filling missing values with linearly interpolated data. According to the input data used in the present work, missing values issue resulted to be a side effect, with no statistical relevance.
- **Data Normalization** We normalized the input data to mitigate issues such as

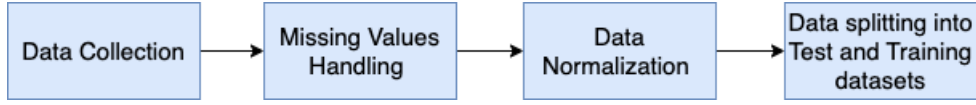


Figure 3.4: Flow chart of data preparation for LSTM model

exploding gradients and improve the numerical stability of our models. We used the min-max method for data normalization, linearly transforming the data into the [0,1] range using the following formula:

$$x_{norm} = \frac{x - min}{max - min} \quad (3.3)$$

where x_{norm} represents the normalized value, x is the original value to be processed, min and max are, respectively, the minimum and maximum values in the dataset.

- **Data splitting into Test and Training Dataset** To evaluate the performance of our neural network models, we divided the data into Training and Test sets. The Training set was used for adjusting the weights and biases of the network to minimize a loss function by updating the weights. The Test set consists of data that the model had not encountered during training, allowing us to assess its performance on new, unseen data. Specifically, we allocated 80% of the total values to the Training set, with the remaining 20% serving as the Test dataset. This arrangement covers almost the entirety of Solar Cycle 24, as illustrated in Figure 3.5. The Training set spans from 28 November 1961, to 26 April 2010 while the Test set from 27 April 2010 to 21 November 2021.

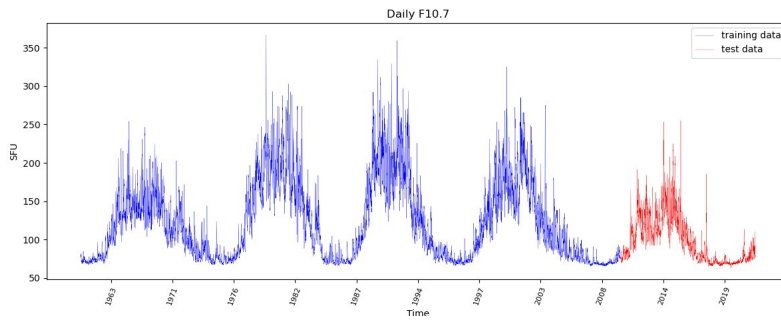


Figure 3.5: Splitting of the F10.7 time series into Training (blue) and Test (red) datasets. F10.7 daily data used in our study covers the time range from November 1961 to November 2021 over the past six solar cycles.

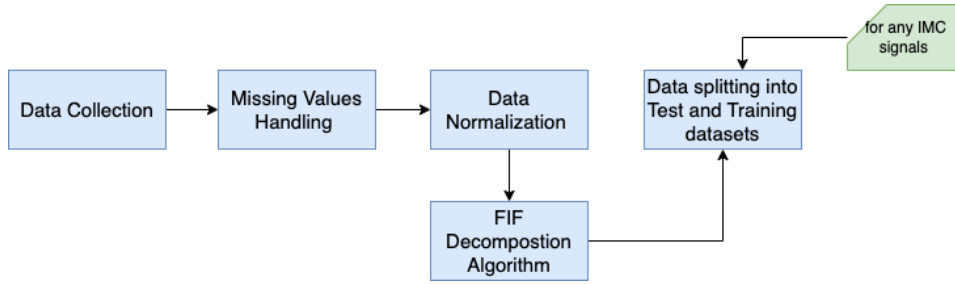


Figure 3.6: Flow chart for data preparation for hybrid LSTM models

In the context of time series forecasting using deep learning models, a critical step is transforming the time series into a supervised learning problem. We achieved this using the Sliding Window approach. In this approach, input features are constructed using historical values within a specified lookback window, and the target variable is the value at the next time step. This transformation allows the model to learn patterns and dependencies based on past observations.

Diverging from the conventional LSTM model, the data preparation process for the hybrid networks, illustrated in Figure 3.6, introduces a distinctive step: FIF decomposition performed just prior to the splitting of the time series into Test and Training datasets. In this case, we applied the FIF algorithm to the same F10.7 data used in the basic LSTM model. This application of FIF led to the generation of a set of intrinsic mode functions (IMFs), which you can observe in Figure 3.7 in juxtaposition to the original F10.7 time series signal. Each IMF was then individually split into Test and Training sets and fed into dedicated LSTM neural networks for further analysis. These IMFs demonstrate time-dependent behavior: the lower-order IMFs exhibit chaotic and random characteristics with added noise, while the higher-order IMFs capture lower-frequency components that replicate oscillations on common physical time scales. These scales include the Sun's rotation period, the Schwabe cycle (11 years), the Hale 22-year magnetic cycle, and the Gleissberg cycle, which spans approximately a century. This approach allowed us to explore the intricate patterns within the F10.7 data, enhancing our understanding of its underlying dynamics. The process of proper configuration and optimization of the network through training is fundamental to developing an effective model capable of making accurate predictions on new data.

In order to achieve this goal we have recursively followed the steps shown in Figure 3.8.

After an initial estimation of the hyperparameters, we iteratively run the model and

F10.7 FIF decomposition

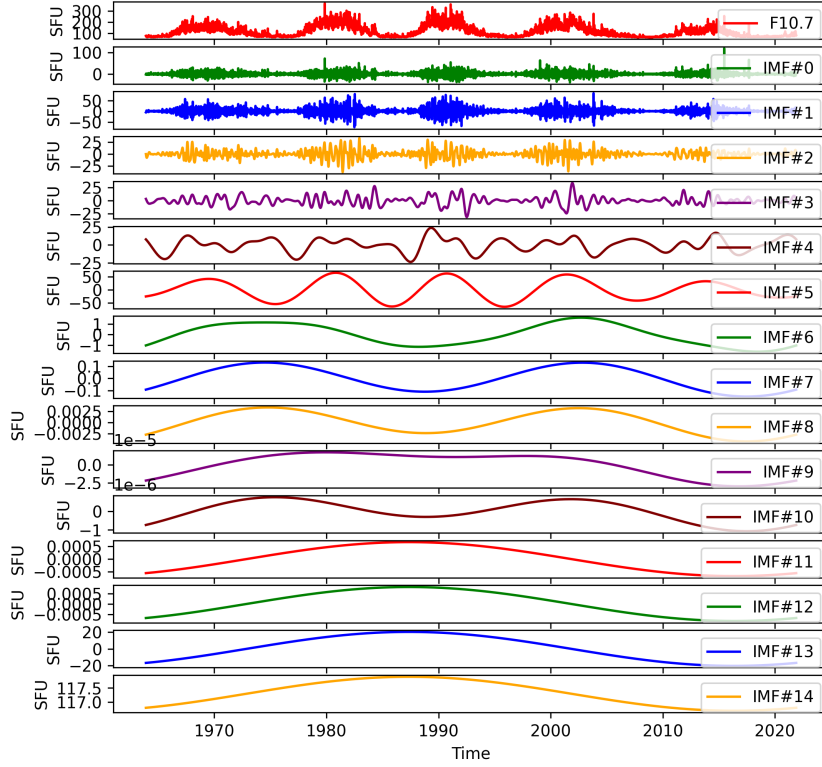


Figure 3.7: Total decomposition in 15 Intrinsic Mode Functions of the F10.7 time series obtained using FIF algorithm.

evaluate the results to refine our choices. This iterative process allows us to make adjustments and fine-tune the hyperparameters to improve the model's performance. Based on the evaluation results, we make adjustments to the hyperparameters and repeat the training and evaluation process. This iterative approach enables us to explore different hyperparameter configurations and assess their impact on the model's performance. By iteratively running the model and adjusting our choices based on evaluation results, we strive to develop a highly effective model that achieves accurate predictions on new data. The selected parameters for this study are summarized in Table 3.1. This iterative adjustment for the hyperparameters aims to minimize a loss function that quantifies the difference between the model's predictions and the true values in the training data.

The training process involves both forward propagation, where input data is passed

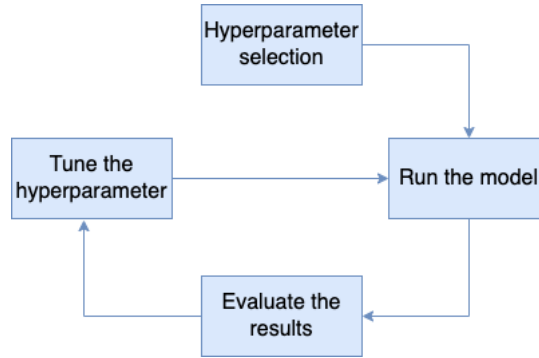


Figure 3.8: Hyperparameter tuning process.

through the model, and backward propagation, where gradients of the loss function are computed to update the parameters. This iterative process continues until the model converges to a point where further parameter updates do not significantly reduce the loss.

As an optimizer, we opted for the Adam algorithm, selected for its compatibility with problems involving substantial data or parameters, computational efficiency, and the capability to handle low signal-to-noise ratio signals with minimal memory requirements [98].

The model training is performed using the sliding look-back window mechanism on the training dataset. The lookback parameter refers to the historical data or preceding time steps used as input to forecast the next time step. In this work, a 3-day lookback window has been chosen, but the effect of different lookbacks on model performance should be studied.

The number of iterations (epochs) for training is not set in advance. The loss on the training set is monitored to detect the point at which the model starts to overfit and an early-stopping schema is employed, which stops the training when the model begins to overfit.

The LSTM model architecture, as shown in Fig.3.9, comprises a three-layer neural network, with each layer containing 24 neurons, along with a dense fully connected layer.

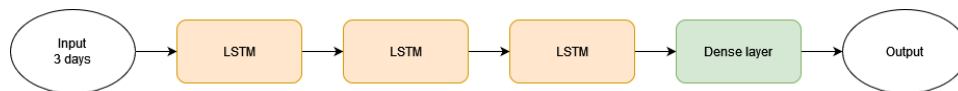


Figure 3.9: Schematic description of the LSTM Network used in this work: 3 layers with 24 neurons each and a final dense layer. A lookback window of 3 days in input has been set.

Parameter	Value
Number of layers	3
Number of hidden neurons	24
Lookback window	3 days
Loss function	RMSE
Activation function	tanh
Optimizer	Adam
Epochs	Early stopping
Batch size	48
Train epochs	100

Table 3.1: An empirical fine-tuning process was employed to determine the optimal hyperparameter settings for the LSTM architecture.

The two hybrid LSTM-based models, specifically the FIF-LSTM and FIF-LSTM-MHA models, integrate the LSTM neural network with the FIF algorithm and the Multi-Head Attention (MHA) module, respectively. The structure of the FIF-LSTM neural network employed in our research is depicted in Figure 3.10. In this approach, we initially apply the FIF algorithm to decompose the F10.7 time series into distinct oscillation . Subsequently, each of these is fed into an LSTM network following the same architecture as the pure LSTM model.

The FIF-LSTM-MHA model consists of three LSTM layers, each containing 40 nodes, and an attention layer with six heads, as displayed in Figure 3.11. To predict future F10.7 values, the model relies on a 20-day lookback window. The

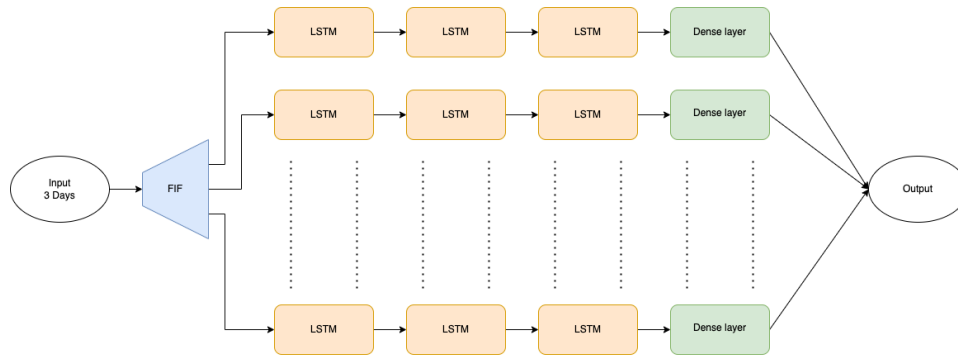


Figure 3.10: Schematic diagram of the hybrid FIF-LSTM neural network.

primary distinction between the FIF-LSTM and FIF-LSTM-MHA models lies in their treatment of the intrinsic mode functions generated by the FIF algorithm. In the FIF-LSTM model, each intrinsic mode function is considered an individual univariate time series. These separate time series are then individually input into

each LSTM network, and the final prediction is obtained by summing the output from all these. Conversely, in the FIF-LSTM-MHA model, each intrinsic mode function is regarded as a single element within a multivariate time series. The MHA module directly provides the predicted value, eliminating the need for additional computations, thereby streamlining the process.



Figure 3.11: FIF-MHA based LSTM neural network used in this paper is composed of 3 LSTM layers of 40 nodes each and 6 heads in the attention layer, respectively. A lookback window of 20 days has been set to predict future F10.7 values.

Once the models are trained, they are used to predict the future trend of the F10.7 index for up to 45 days. The direct method for multistep forecasting generate predictions for the entire lead time at once. Each time step in the forecast horizon is treated as a separate prediction task, and the model is trained to produce these predictions concurrently.

To comprehensively evaluate and compare the models, we employ three performance metrics that capture different aspects of model performance: the Root Mean Squared Error (RMSE), which measures the degree of deviation between predicted and observed values; the Mean Absolute Percentage Error (MAPE), a scale-free metric commonly used in time series forecasting; and the Pearson linear correlation coefficient (R). The formulas for these metrics are as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y'_i - y_i)^2}, \quad (3.4)$$

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y'_i - y_i}{y_i} \right|, \quad (3.5)$$

$$R = \frac{\sum_{i=1}^N (y'_i y_i)}{\sqrt{\sum_{i=1}^N (y'_i)^2} \sqrt{\sum_{i=1}^N (y_i)^2}}, \quad (3.6)$$

where y_i and y'_i are the i_{th} observed and predicted values, respectively, and N is the number of data points in the testing time series. To provide a more granular analysis, we present forecasting results at three selected reference lead times: 1 day (very short-term forecasting), 25 days (approximately the duration of one solar rotation), and 45 days (used, for example, by the U.S. Air Force, USAF) [3].

In Figure 3.12, RMSE and MAPE values are depicted as a function of the time step (from 1 to 45 days) for the classic LSTM, FIF-LSTM, and FIF-LSTM-MHA models. The results indicate that the RMSE increases significantly for all models during the first 1-7 forecasting days. Notably, models based on the FIF decomposition technique demonstrate greater stability over time compared to the others. The classic LSTM model exhibits worse performance after approximately 5 lags compared to the other models, as evidenced by higher RMSE and MAPE values. Additionally, it appears to exhibit a period strongly correlated with the 25-day solar rotation.

The stability of models based on the FIF decomposition technique is attributed to the nature of the algorithm itself, as analyzed mathematically in several papers [187, 226]. The algorithm's convergence to a meaningful decomposition can be guaranteed apriori for any signal. In Table 3.2 we summarize the statistical results

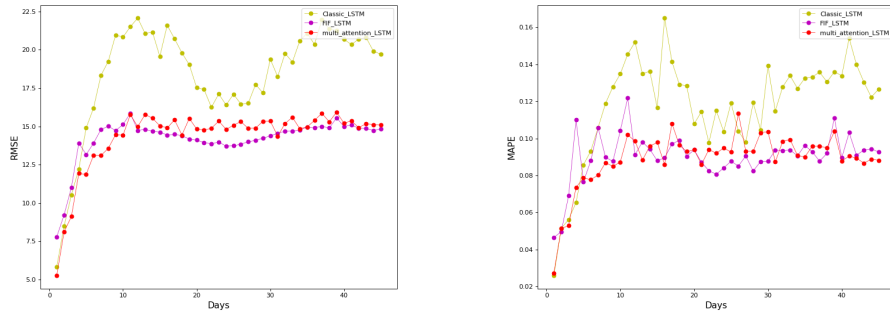


Figure 3.12: Compared model performances: RMSE (left panel) and MAPE (right panel) for lag time of 1-45 days.

obtained for each model from 1 to 45 days of lead time, focusing on the selected times.

We have also investigated the different trends in F10.7 forecasting based on solar activity levels. To facilitate a more effective performance comparison at different solar cycle phases, we selected two time ranges: during the solar maximum (2014-2015) and solar minimum (2018-2019).

In Figure 3.13, 3.14 and 3.15, we represent the ratio between the predicted and observed values for both solar activity levels and for lead times of 1, 25, and 45 days ahead. The observed bumps in the error (Fig. 3.13) are likely correlated with days when there are unexpected changes in the number of active regions on the solar disk. For a more quantitative analysis, Figure 3.4.1 present the results of

	Classic LSTM			FIF-LSTM			FIF-LSTM-MHA		
	RMSE	MAPE	R	RMSE	MAPE	R	RMSE	MAPE	R
1-day	5.81	0.025	0.98	7.77	0.046	0.96	5.25	0.027	0.98
2-day	8.47	0.052	0.96	9.21	0.049	0.95	8.10	0.051	0.97
3-day	10.53	0.056	0.94	11.01	0.069	0.93	9.11	0.052	0.95
4-day	12.21	0.065	0.91	13.91	0.110	0.91	11.88	0.073	0.93
5-day	14.89	0.086	0.88	13.10	0.076	0.89	11.89	0.078	0.92
6-day	16.19	0.093	0.86	13.90	0.087	0.88	13.10	0.079	0.90
7-day	18.32	0.110	0.83	14.70	0.101	0.87	13.09	0.080	0.89
8-day	19.24	0.120	0.81	15.01	0.089	0.86	13.50	0.086	0.89
9-day	20.96	0.120	0.78	14.70	0.087	0.86	14.40	0.084	0.87
10-day	20.83	0.130	0.77	15.11	0.093	0.86	14.40	0.087	0.87
15-day	19.56	0.110	0.77	14.62	0.087	0.86	15.03	0.097	0.87
20-day	17.52	0.110	0.82	14.10	0.093	0.88	14.82	0.093	0.86
25-day	17.10	0.120	0.84	13.72	0.086	0.88	15.07	0.092	0.88
30-day	19.38	0.130	0.82	14.38	0.087	0.87	15.34	0.100	0.86
40-day	20.69	0.130	0.76	14.98	0.089	0.86	15.20	0.087	0.86
45-day	19.71	0.130	0.79	14.81	0.092	0.86	15.09	0.088	0.86

Table 3.2: RMSE, MAPE and R correspondent to each model, for increasing forecasting time steps.

performance comparison. Each histogram in these figures displays MAPE values for different phases of the 24th solar cycle, averaged over 13 months. The observed F10.7 values are superimposed on the plots (right scale).

In the context of 1-day ahead forecasting, the hybrid FIF-LSTM-MHA and classic LSTM models demonstrate the best performance. They exhibit the lowest MAPE values during the maximum, minimum, and ascending phases of the solar cycle. During the descending phase, the classic LSTM model seems to outperform the others, but only for very short-term predictions. This is due to its sensitivity to short-term F10.7 daily oscillations compared to other models. Moving to longer lead times, the classic LSTM model performs poorly. On the contrary, although the FIF-LSTM model is the weakest in 1-day ahead prediction, its performance substantially improves with increasing lead times. It is also evident that the hybrid methods based on FIF decomposition are very promising, particularly compared to the classic LSTM during solar maximum, the descending phase, and the minimum when the forecasting horizon is extended.

It is worth noting the computational time constraints for each model, which play a role in the evaluation of different techniques. In Table 3.3, we present the mean time spent on training the three model architectures. Notably, the FIF-

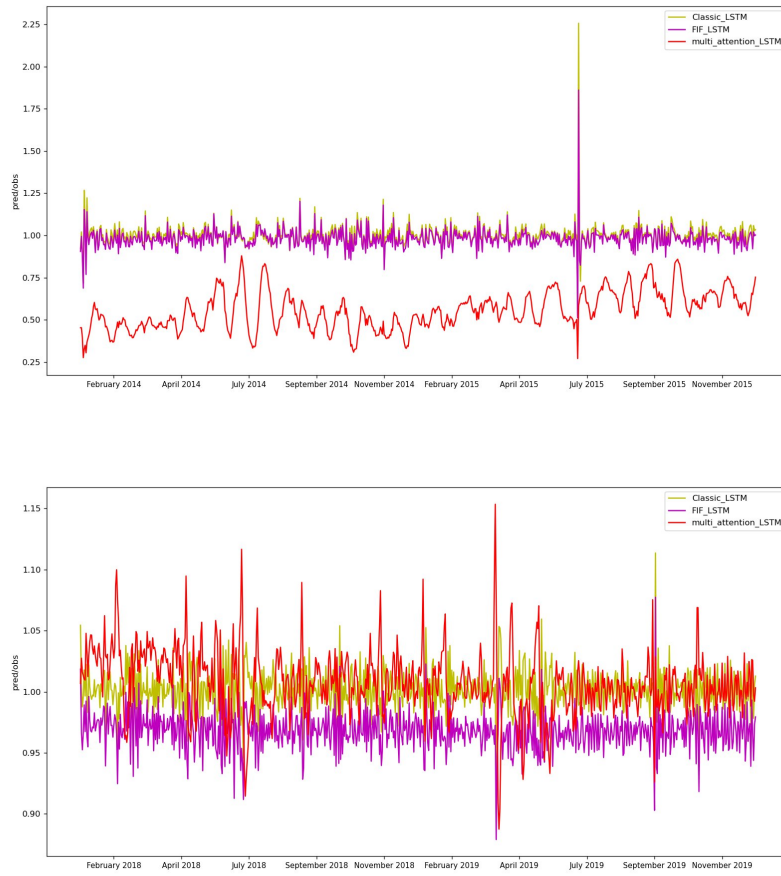


Figure 3.13: Ratio between the predicted and observed F10.7 is represented for 1-day ahead during solar maximum (2014-2015) in the top panel and solar minimum (2018-2019) in the bottom panel.

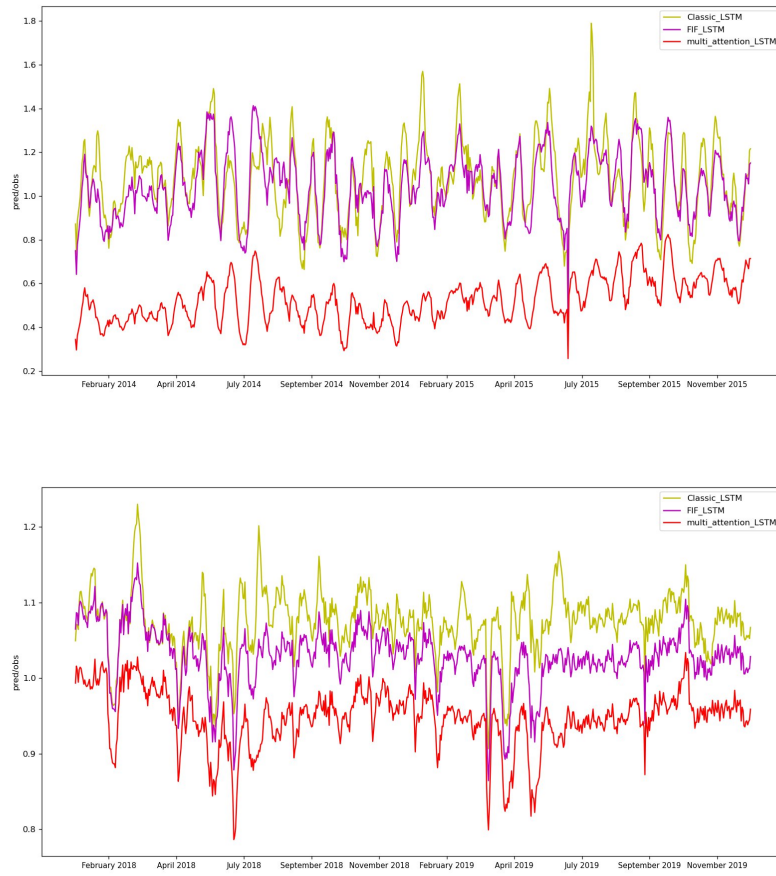


Figure 3.14: Ratio between the predicted and observed F10.7 is represented for 25-day ahead during solar maximum (2014-2015) in the top panel and solar minimum (2018-2019) in the bottom panel.

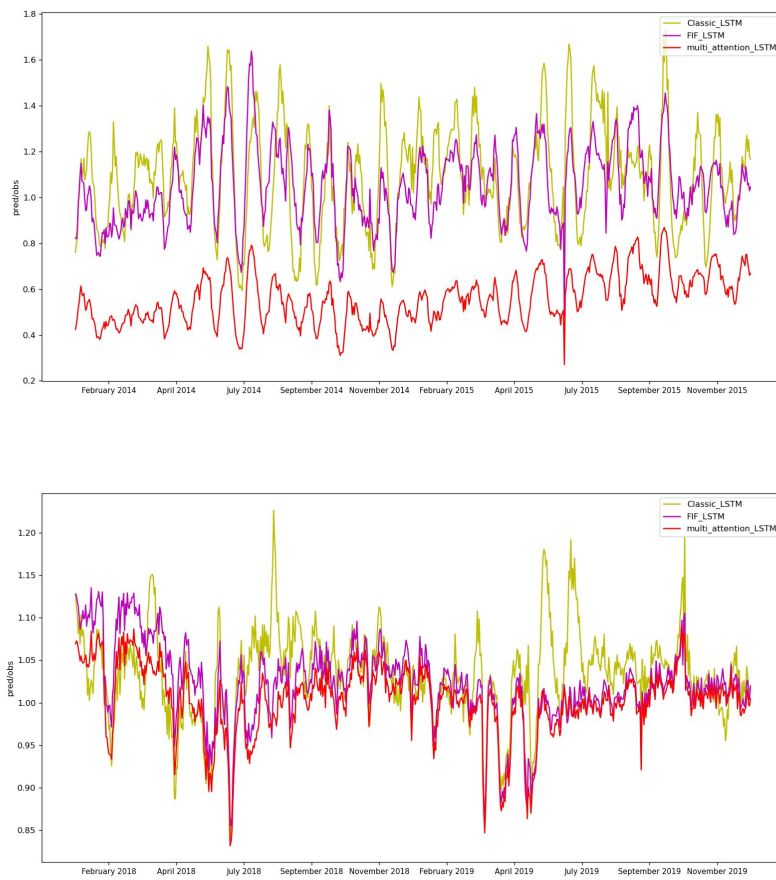


Figure 3.15: Ratio between the predicted and observed F10.7 is represented for 45-day ahead during solar maximum (2014-2015) in the top panel and solar minimum (2018-2019) in the bottom panel.

TIME SERIES FORECASTING FOR SPACE WEATHER APPLICATIONS

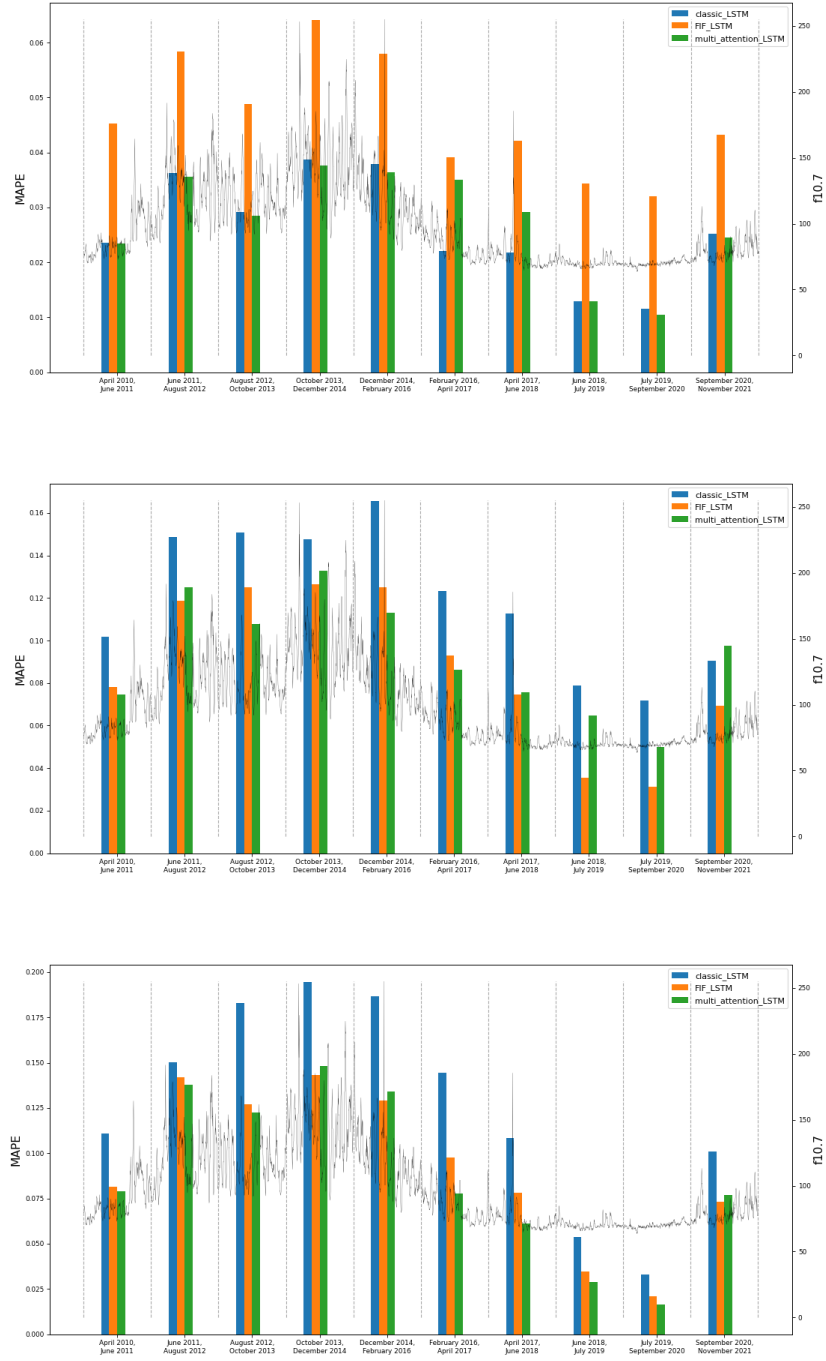


Figure 3.16: MAPE respect to the solar cycle phase for 1-day ahead (top panel), 25-day ahead (mid panel), and 45-day ahead (bottom panel)

LSTM-MHA architecture generally requires less computational time compared to the others, making it particularly advantageous for extended time series and multi-head forecasting. Classic LSTM training is quicker than the hybrid approaches based on FIF decomposition methods but performs less favorably at longer lead times. It is important to highlight that the training of the FIF-LSTM-MHA model can be carried out on a standard personal computer without issues.

Model Architecture	t_{min}
Classic LSTM	400 s
FIF-LSTM	1100 s
FIF-LSTM-MHA	235 s

Table 3.3: Average training time for each analyzed model.

3.4.2 Multivariate NN-base Model

As from the comparative analysis LSTM-FIF-MHA was globally outperforming, we decided to adopt it for our additional implementations: the addition of a multivariate input solar radio indices for the F10.7 forecasting.

Radio waves at different frequencies are emitted at different heights in the solar chromosphere and low corona by plasma layers that have decreasing electron density with increasing altitude: short-wavelength emissions originated at lower altitudes, whereas long-wavelength ones originated at higher altitudes. Hence, we considered exploiting the “tomography” of the solar plasma layers to enhance the performances in solar activity forecasts with a multivariate approach.

The architecture combines LSTM and the MHA modules. Our approach follows the same standard sequence of steps depicted in Figure 3.3, including data preparation, model training, hyperparameter tuning, and rigorous evaluation, with the ultimate goal of achieving accurate forecasting.

The multivariate data set includes six features, consisting of five solar radio fluxes and the International Sunspot Number. The data spans from November 1, 1951, to July 25, 2023 (Fig. 3.17). Five radio fluxes emitted at 3.2, 8, 10.7, 15 and 30 cm are measured by the radio telescopes situated in Toyokawa (historical data), Nobeyama (recent data), Ottawa (historical data) and Penticton (recent data) and their value is given in solar flux units (SFU).

To prepare the data set for our study, we relied on the Collecte Localisation Satellites (CLS) tools: an external service [77] that focuses on modelling the upper

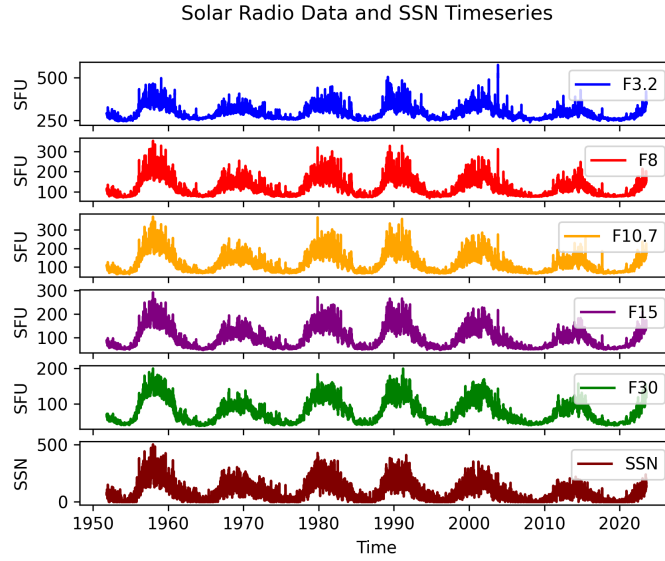


Figure 3.17: Solar Radio Fluxes and Sunspot Number (SSN) time series from 1 November 1951 to 25 July 2023. Data are given by the LASP Interactive Solar IRradiance Datacenter (LISIRD) portal [54]. The service is maintained by the Laboratory for Atmospheric and Space Physics (LASP) of the University of Colorado, Boulder [78]. F3.2, F8, F10.7, F15 and F30 refer, respectively, to the Solar Radio Index at 3.2, 8, 10.7, 15 and 30 cm, expressed in solar flux units (SFUs).

atmosphere to predict low Earth satellite orbits. It performs the initial processing of the data, filling potential gaps and replacing values with residual errors using an expectation-maximization algorithm and an auto-regressive model.

The International Sunspot Number time series is derived from the sunspot index and long-term solar observations provided by the SILSO World Data Center [79] of the Royal Observatory Belgium. To bring out relevant features in the data, we applied the FIF decomposition algorithm to the input time series.

As an example, Fig. 3.18 shows the decomposition into IMFs obtained for the F8 index. To enhance the learning capabilities of the neural network, we selected the IMFs that best correlated with F10.7 and included them as additional inputs to the training dataset. The computation of Pearson's correlation coefficients enabled the selection. Figure 3.19 panel shows the correlation coefficients resulting from the correlation study for the F8 index, while the selected IMF is displayed together with the original signal in panel.

The representation and quality of the data set is the first and foremost success factor in the effectiveness of a machine learning process. With this in mind, we added a Feature Selection step to the data preparation, which allows us to understand

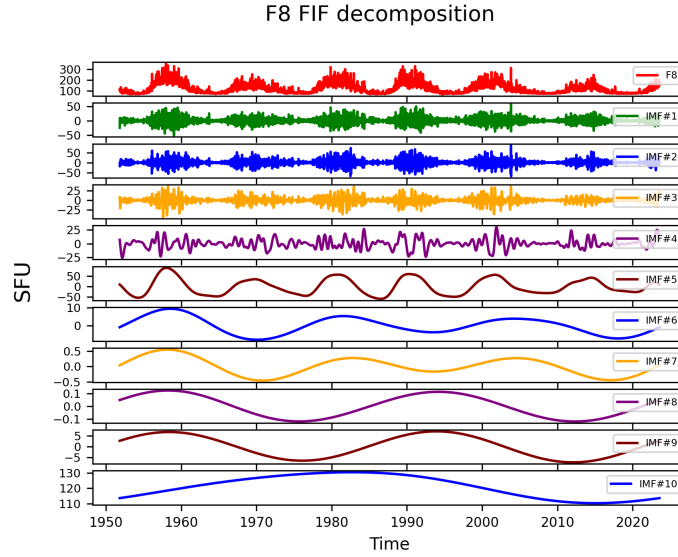


Figure 3.18: Results of FIF decomposition technique applied to the F8 radio index. A set of 10 IMFs is plotted together with the F8 original signal (red).

whether it is necessary to use all the time series (5 radio indices, SSN and the most correlated IMF for each of them) in the network.

Feature Selection is defined as the process of selecting a minimal-size subset of the features that contain all predictive information necessary to produce an optimally predictive model for a target outcome of interest. This reduces the dimensionality of the data and may enable forecasting models to operate faster and more effectively. In fact, knowledge discovery during the training phase becomes more difficult in the presence of irrelevant and redundant information or noisy and unreliable data [96]. In this work we chose to apply the permutation importance technique which allows us to assess the importance of a feature, estimating the effect of permuting the values of a particular feature on the model's performance.

More details on the permutation importance algorithm will be provided in the Section 3.5.1

Figure 3.20 depicts an example of the feature importance obtained for the time series used in the F10.7 forecasting at 1 and 25-day predictive horizon.

For very short-time forecasting, F8 and F15 carry most of the information as they are generated by physical processes related to those underpinning the F10.7 emission. While the contribution of F15 remains important, F8 importance decreases in favour of F30 for a longer forecasting horizon (Fig. 3.20). Overall, the IMFs contribute more to longer time window forecasts than they do for very short timestamps.

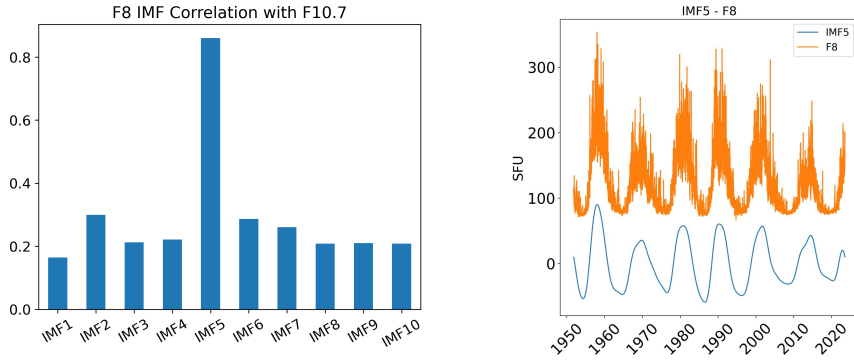


Figure 3.19: Correlation coefficients (left panel): Different IMFs are identified by progressive numbers from 1 to 10. IMF5 shows the highest correlation with F10.7 for the F8 index. In the right panel IMF5 (blue) plotted along with the original F8 time series (orange)

For each time step we have discarded the three features with the lowest importance according to our results, while all the others will be given as input to the multivariate LSTM network.

The model architecture used in this work (Fig. 3.21) consists of three LSTM

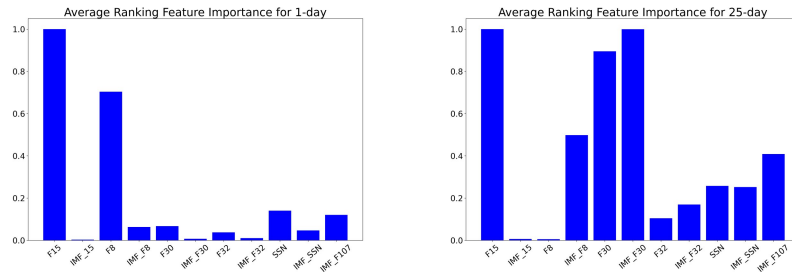


Figure 3.20: Average ranking of the feature importance for 1-day (left) and (right) 25-day ahead F10.7 forecasting.

layers, each with 12 nodes, followed by an Attention layer with 8 heads and a dense layer. The hyperparameters of the model, which were refined during the training phase are summarised in Table 3.4.

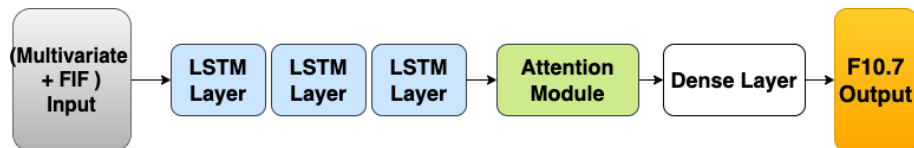


Figure 3.21: Multi-Head Attention LSTM Model structure diagram.

Parameter	Value
Number of layers	3
Number of hidden neurons	12
Lookback window	50 days
Loss function	RMSE
Activation function	sigmoid
Optimizer	Adam
Epochs	Early stopping
Batch size	48
Train epochs	100

Table 3.4: Hyperparameter settings for the LSTM architecture.

The network’s training utilises 80% of the data. It exploits the Adam optimisation algorithm, which is suitable for handling large data sets and stable even for signals with a low signal-to-noise ratio [98]. An early-stopping schema interrupts the training when the model starts to overfit.

The network’s performance is evaluated by comparing the predicted values with the remaining 20% of the data by the RMSE.

Figure 3.22 in the left panel shows the observed and predicted values of F10.7 index up to 25-day forecast horizon while in the right panel the relative error is plotted, expressed as

$$RE = \frac{|F10.7_{pred} - F10.7_{obs}|}{F10.7_{obs}}$$

The model performs better during solar maximum, while the difference between the observed and predicted values is higher during solar minimum, underling the dependence on the solar activity level. In particular, during the minimum phase, the model tends to slightly overestimate the values.

The plot presented in Fig. 3.23 compares the performance of different model architectures and input data. The red, green, and purple lines represent univariate models that are fed only with the F10.7 time series. The blue and orange lines refer to the multivariate models, that incorporate additional solar indices along with the selected IMFs. The plot demonstrates that the multivariate architectures outperform the univariate models due to the wider information provided to the network. Additionally, the difference between univariate models with and without the MHA (Multi-Head Attention) layer is minimal. However, the adoption of FIF results in clear improvements (green line compared to red line). Moreover, the Feature Ranking selection further reduces the deviation between the observed and predicted values (blue line compared to orange line). Overall, the proposed model

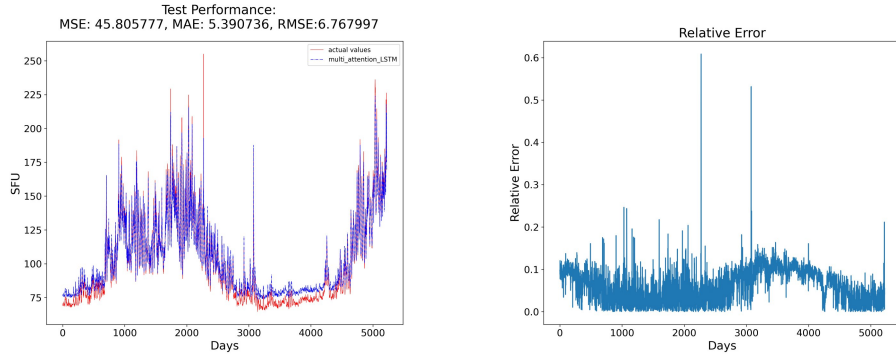


Figure 3.22: Observed and predicted F10.7 values (left panel) and relative error (right panel) by the proposed multivariate model for 25-day ahead.

exhibits a stable behaviour in terms of lag size if compared to the univariate models, where the RMSE rapidly increases in the first 10 days and then stabilises. It also maintains a relatively low RMSE value (as shown in Fig. 3.23) that does not exceed 5 SFU, indicating overall effectiveness in predicting the F10.7 index compared to other models.

By adopting a multivariate approach, the training dataset expands to encompass multiple solar indices and a carefully curated set of Intrinsic Mode Functions (IMFs). This richer data set provides a wealth of information, potentially strengthening the model's forecasting abilities. The proposed model has consistently demonstrated its accuracy and stability by boosting a relative error under 10% and an RMSE of approximately 5 sfu, across various forecasting horizons ranging from 1 to 25 days.

3.4.3 Long-Term Solar Activity Forecasting Model

In this section, we focus on the prediction of the activity level over a period of 11 years corresponding to one solar cycle. The result can be used as input to various atmospheric models that characterise the physical state of the Earth's upper atmosphere, such as the thermosphere and ionosphere, and are used for satellite orbit determination, re-entry services, collision avoidance manoeuvres and modelling of the evolution of space debris [158]. Variations in the thermospheric density, in fact, cause variations in the atmospheric drag affecting orbiting satellites, which need to be carefully taken into account when planning spacecraft operations. Moreover, the abundance in the heliosphere of highly energetic particles, known as galactic cosmic rays, shows a periodicity that anticorrelates with solar activity. Cosmic rays

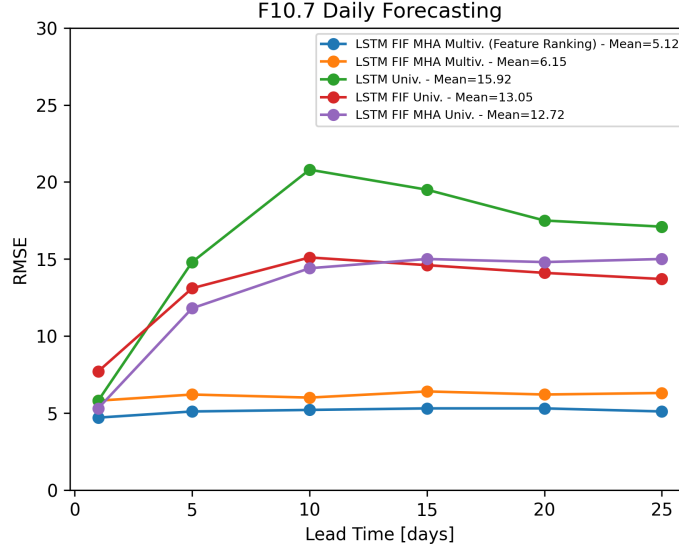


Figure 3.23: RMSE values of the proposed model compared to classical approaches, as a function of lag size. In green the Univariate Classic LSTM, in red the Univariate LSTM FIF, in purple the Univariate LSTM FIF MHA, in orange the Multivariate LSTM FIF MHA without feature selection and in blue the Multivariate LSTM FIF MHA with time series selection based on feature ranking. The average RMSE for all the analyzed models is also reported in the legend

constitute a hazard to space instruments and human missions. Hence, an increase in the precision at which we can predict their occurrence is critical.

Historically, long-term forecasts of solar activity have been based on the solar sunspot number (SSN), the weighted sum of the numbers of sunspots and sunspot groups on the solar disk at one time, present in records since 1600. They are good descriptors of almost all other features on the Sun, including active regions, plagues, flares, prominences, and, to some extent, changes in the evolution of coronal and solar wind features [41]. The results discussed in this paragraph are published in Marcucci et al. 2023 [125] and have been presented at the 9th International conference on Time Series and Forecasting, therefore data reported are updated with information available up to 2023.

The decision to average the data over a bi-weekly interval comes from the need to balance computer performances with accurate long-term predictions.

The model architecture used in this work is the same as the one adopted in the previous Section (Figure 3.21). Figure 3.24 shows the RMSE of the model at different lead time forecasts. The model performance shows a very stable behaviour

in lag size with an RMSE value (see Fig. 3.24) not exceeding 25%, which proves the overall model prediction effectiveness. As an example, the actual and predicted values of F10.7 index corresponding to a lead time of 80 weeks are plotted in Figure 3.24. Our model tends to smooth the prediction curve due to the peculiarity of the attention module to discard less relevant features of the time series and highlight the global trend. The forecasting result of F10.7 index for the solar cycle 25 are

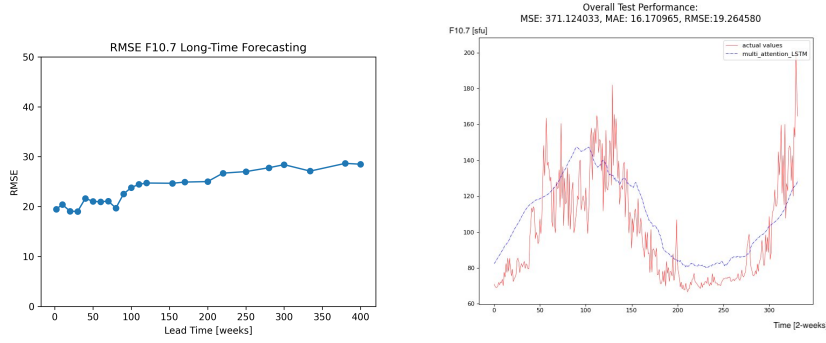


Figure 3.24: (left panel) Performance evaluation of the model at different lead times; (right panel) Observed and predicted F10.7 values for 80 weeks ahead.

given in Figure 3.25. In order to characterize the amplitude and peak time of the predicted solar cycle, we fitted the shape of the predicted curve with a function with four free parameters [68, 46] of the form:

$$f(t) = a(t - t_0)^3 \cdot \exp\left[\frac{-(t - t_0)^2}{b^2}\right] - c, \quad (3.7)$$

where parameter a represents the amplitude and is directly related to the rate of rise from minimum; b is related to the time in months from minimum to maximum; c gives the asymmetry of the cycle; and t_0 denotes the starting time. This well-known function reproduces both the rise and decay part of the solar cycle and has a more rapid decline after maximum. The curve fitting is given in Figure 3.26 with the following optimal values: $a = 124.33712279$, $t_0 = 16.79123391$, $b = 32.81850839$, $c = 0.6408711$. From the fitted curve, we derived that the predicted F10.7 maximum amplitude for the solar cycle 25 is about 145 ± 25 , peaking in the first half of 2024.

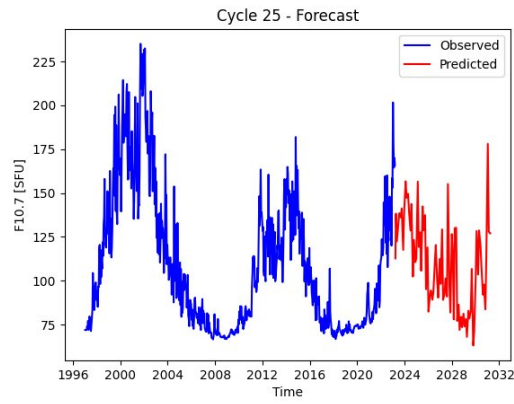


Figure 3.25: The forecast of F10.7 index for the solar cycle 25 (red). The Observed values for the previous solar cycles are shown in blue.

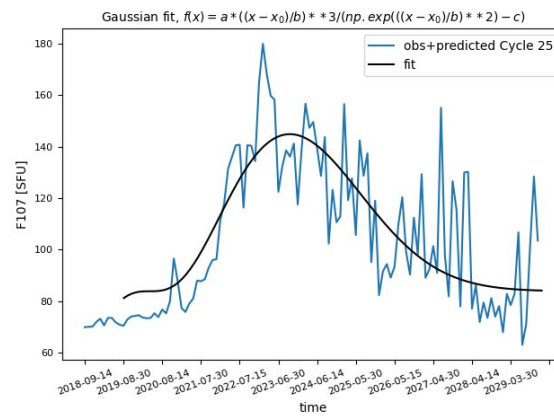


Figure 3.26: Fitting curve on predicted F10.7 values from 2019 and 2030.

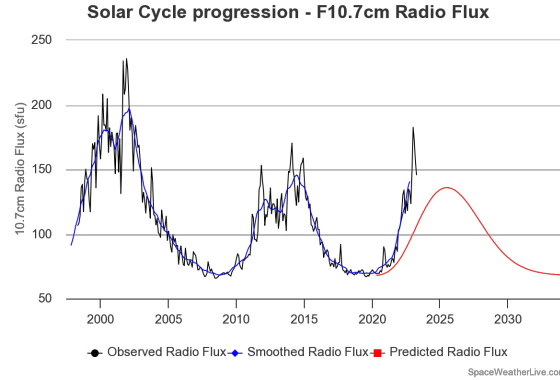


Figure 3.27: ISES Solar Cycle F10.7 cm Radio Flux Progression, [6]

We compared our solar cycle 25 forecasting with the official predictions released by the Solar Cycle 25 Prediction Panel and the International Space Environmental Services (ISES) in 2019 April [25].

The Solar Cycle 25 Prediction Panel is an international group of experts co-chaired by the National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), and International Space Environmental Services (ISES), whose objective is to predict the amplitude of solar cycle 25. These predictions are a synthesis of a variety of prediction methods coming from the scientific community ranging from physical models, precursor methods to statistical inference, machine learning, and other techniques. The Prediction Panel predicted that solar cycle 25 will be similar to solar cycle 24 and will reach a maximum in July 2025 [6] with a peak F10.7 value of 135 ± 10 SFU (Fig. 3.27). This prediction is in line with the current general agreement in the scientific literature, which holds that solar cycle 25 will be weaker than average, even if observed values from 2020 to 2022, the first three years of the cycle, are significantly higher than the predicted values. It is evident that our predictions are consistent with the official ones in the peak F10.7 value but our model predicts the 25th solar cycle maximum about 20 months before.

In addition to the most widely used F10.7 we have also considered a new hybrid index (TSAI, Trieste Solar Activity Index), which is based on the modified index of the solar radio flux at 2800 MHz augmented by a linear combination of the occurrence of soft X-ray flares of class C, M, and X, weighted by coefficients that take into account their occurrence probability [118]:

$$TSAI = S_{2800} \cdot M + 0.5 \cdot C + M + 10 \cdot X$$

Indeed, SSN and F10.7 are global descriptor of Solar Activity, expressing the potentiality of occurrence of energetic phenomena, but they do not provide any information of which ones have to be expected or about the kind of energetic phenomena which will be associated with the Sunspot Groups. The preliminary analyses indicate that TSAI can match the Space Weather requirements and fill the gap in adequately describing Solar Activity and therefore it could substitute F10.7 in our algorithms for a further dedicated study.

3.5 External Drivers Effects for Ionospheric TEC Model

The ionosphere's behavior is influenced by various factors, including solar radiation, geomagnetic activity, and local time, as widely discussed in Chapter 1. Consequently, TEC exhibits diurnal, seasonal, and solar cycle variations. Understanding the ionosphere's electron density distribution, crucial for mitigating these effects, relies on measuring TEC [159].

To generate global ionospheric maps and assess ionospheric conditions, GNSS stations worldwide continuously measure TEC and these measurements, integrated over time and space, provide insights into the ionosphere's spatial and temporal variability. Data from multiple GNSS satellites and receivers are combined and interpolated to create global maps of TEC known as Global Ionospheric Maps (GIMs), representing spatial electron density distributions of TEC values over the entire Earth's surface.

The dissemination of ionospheric data is facilitated through formats like the IONosphere map EXchange (IONEX) files [74]. IONEX files contain TEC values, satellite and receiver positions, and other metadata, enabling researchers to analyze and model ionospheric behavior. Global ionospheric TEC data are provided by the Ionospheric GNSS Service (IGS) in a geographic grid map, in the so called IONEX (ionosphere exchange) format, with a spatial resolution of $2.5^{\circ}(\text{lat}) \times 5^{\circ}(\text{lon})$.

Monitoring these variations through GIMs aids in Space Weather monitoring, ionospheric modeling, and GNSS applications.

The amount of contributions from different external drivers to the ionospheric state is not well understood yet but it would be essential for successfully forecasting the ionospheric conditions and the knowledge of TEC could lead to the improvement of positioning accuracy.

The purpose of the work presented in this section is to implement methodologies to investigate the correlation between external drivers and ionospheric Vertical Total

Electron Content with the use of several machine learning techniques (ANN, CNN, ELM and RF) at different latitudes and solar activity conditions. When CME, fast solar wind, coronal hole high speed stream hit the Earth, the coupled conditions among solar wind, magnetosphere and ionosphere may lead to several effects both at high and low latitudes.

Indeed, it has the aim to understand the most relevant features (input parameters) for the TEC prediction task.

Solar radio flux index, proxy of solar UV radiation, is the feature most impacting on the global ionosphere conditions therefore its prediction results very useful as it has been shown in the present chapter, using different ML approaches.

Daily files span in the range between 2006 and 2017 (in order to include a whole 11-years solar cycle), they have been downloaded from the CDDIS-NASA website and multiple time series have been extracted, with a timestep of 2 hours. [147]

The ionospheric TEC, in addition to the climatological oscillations, has also a high variability related to Space Weather forcing conditions, that we want to investigate in our work.

As a starting point, a single grid point of the Global Ionospheric Map (GIM) have been chosen to extract and reconstruct the time series from the daily IONEX map files.

External drivers, selected as descriptors of Space Weather conditions, are available in the OMNIWeb NASA-website [58] with 1-hour time resolution. In particular, for our work, we have selected observations from the ACE satellite, in L1, regarding the solar wind, such as: - the three of the geomagnetic field B (B_x, B_y, B_z); - the proton density (ρ), temperature and speed vector (T, v); - the solar sunspot number R ; - the solar radio flux at the wavelength of 10.7 cm (F10.7); - the geomagnetic indices Dst, ap, and Kp; - the auroral indices AU, AL, and AE.

3.5.1 Feature Ranking and ML Algorithms

Machine Learning has been extensively exploited in different fields in order to help and support decision making, but it is important to understand why a particular output is produced according to the input (features) used to train the models. In order to ascertain the most influential external drivers, the permutation technique is employed, which is a feature importance algorithm [13].

The adopted feature ranking technique is called permutation of feature importance and it measures the predictive value of a feature for a regressor. It allows to evaluate how the prediction error increases in a model score when a single feature timeseries

is randomly shuffled. Breaking the relationship between the feature and the target, the drop in the model score is indicative of how much that particular model depends on the feature.

The permutation importance of a feature is calculated as follows. Let X be the data set used to train the model that includes the feature of interest. First, you start by training the model on X and evaluate the baseline performance by computing the performance metric. Next, the values of the selected feature are shuffled, while keeping the other features unchanged, and the performance is evaluated again, training the model on the permuted data set. The permutation importance of the feature is defined by comparing the baseline performance with the performance on the permuted data [154]. The greater the difference, the more important that feature is.

To avoid obtaining a model-dependent result, we implemented, trained and tested many different linear and non-linear neural network models for each forecasting lead time. More specifically we took advantage of the following Python libraries, KNN, Multiple Linear regression, Lasso, Support Vector Regression, Gradient Boost, Voting regression, Random Forest, Adaptive Boost, Extreme Gradient Boost and Decision Tree. We considered the effective feature ranking as the average of the permutation importance across all the considered models.

Ensembling these models together ensures robustness by reducing the variance, since there is no specific Machine Learning model which is always optimal (no free lunch theorem for Machine Learning [213]).

Before using the permutation importance technique, we first checked for multicollinearity between the external drivers and removed the correlated ones, for example AE and AL/AU are highly correlated with one another in regression models: AE can be linearly predicted from the others. This algorithm is then applied to the remaining set of external drivers.

To determine the importance of the external drivers, we used 12 Machine Learning models including both linear (the first three models described in the following) and non-linear model:

- Multiple regression, a statistical technique used to analyze the relationship between a quantitative dependent variable (such as Total Electron Content, TEC) and multiple independent variables (features). The goal is to estimate the coefficients that represent the strength and direction of the relationship between the dependent variable and each independent variable. The cost function, also known as the loss function, is the Mean Squared Error (MSE).

[172]

- Support Vector Regression a regression technique based on Support Vector Machines (SVMs), where the ϵ -insensitive loss function is used, where the error allowance ϵ determines the width of the tube around the regression line within which errors are not penalized. [16]
- In LASSO (Least Absolute Shrinkage and Selection Operator) regression, the cost function includes a penalty term that encourages sparsity in the coefficients of the regression model. The penalty term is a regularization term that penalizes the absolute values of the coefficients, thereby shrinking some of them towards zero and effectively performing feature selection. The cost function in LASSO regression is defined as:

$$\text{Cost} = \text{MSE} + \lambda \sum_{j=1}^p |\beta_j|$$

where λ is the regularization parameter, which controls the strength of the penalty term; $\sum_{j=1}^p |\beta_j|$ represents the sum of the absolute values of the coefficients β_j for each feature j and p is the number of features. LASSO regression allows for a different coefficient (penalty) for each feature, which means that the penalty applied to each coefficient can be weighted differently based on the importance or relevance of the corresponding feature. [195]

- Decision Tree operates by recursively splitting the dataset into subsets based on the feature that best separates the data according to a chosen criterion, such as Mean Squared Error (MSE) for regression. Each internal node of the tree represents a decision based on a feature, and each leaf node represents the prediction for the target variable. Decision trees aim to minimize the MSE by making splits that reduce the variance of the target variable within each subset.
- Random Forest, an ensemble learning method that combines an ensemble of decision trees, each trained on a random subset of the features and a random subset of the training data (bootstrapped samples). During prediction, each tree in the forest independently makes a prediction, and the final prediction is determined by averaging (for regression) or voting (for classification) over all the trees.

- K-Nearest Neighbor (KNN), a non-parametric and instance-based learning algorithm used for classification and regression tasks. It makes predictions based on the similarity between the input data point and its k nearest neighbors in the training dataset. For regression tasks, the prediction is the local interpolation (e.g., mean or weighted average) of the target values of the k nearest neighbors.
- Gradient Boosting, an ensemble learning technique that combines multiple weak learners, typically decision trees, in a sequential manner to build a strong predictive model. It works by iteratively fitting new trees to correct the errors of the previous ones. Each new tree is trained to predict the negative gradient of the loss function with respect to the ensemble's predictions so far.
- Extreme Gradient Boosting (XGBoost), highly optimized implementation of Gradient Boosting that includes built-in regularization techniques, advanced tree construction algorithms, efficient handling of sparse data
- Adaptive Boosting (AdaBoost), an ensemble learning method that combines multiple weak learners, typically decision trees, to create a strong predictive model. It sequentially trains a series of weak learners on weighted versions of the training data, with each subsequent learner focusing more on the instances that were misclassified by the previous ones.
- Extremely Randomized Tree (Extra Tree), an ensemble learning method that extends the idea of Random Forests. Like Random Forests, Extra Trees builds multiple decision trees on random subsets of the training data and features.
- Bagging, ensemble learning technique that improves the performance of machine learning models by training multiple instances of the same base model (Support Vector Machine) on different subsets of the training data. Finally, the predictions of all the base models are aggregated through averaging to make the final prediction.
- Voting, combine the predictions of multiple base models (Adaptive Boosting, Extra Tree, Gradient Boosting, Extreme Gradient Boosting) independently on the same training data to make a final prediction aggregating the individual predictions through a voting mechanism.

In order to build our predictive models, we required input dataset including the two best-ranked external features for different hours with 24 hours and VTEC times

series. We took the previous 24 hours of the VTEC and the two best external drivers times series as the features of each instant of the training dataset. [51] The different ML algorithms used in this work are:

- Artificial Neural Networks (ANNs), computational models inspired by the structure and function of biological neural networks in the human brain. ANNs consist of interconnected nodes organized into layers, including input, hidden, and output layers. Each node, or neuron, performs simple mathematical operations on its inputs and generates an output signal.

Training an ANN involves adjusting the weights and biases of connections between neurons to minimize prediction errors. This process typically employs optimization algorithms such as gradient descent and backpropagation, where errors are propagated backward through the network to update the parameters.

ANNs are versatile models capable of learning complex patterns from input data, making them suitable for a wide range of tasks, including regression, classification, and function approximation. They have been successfully applied in various domains, including image recognition, speech recognition, and financial forecasting.

Despite their effectiveness, ANNs have limitations, such as the need for large amounts of labeled data for training, susceptibility to overfitting, and difficulty in interpreting model decisions, especially in deep architectures.[69]

- Convolutional Neural Networks (CNNs), deep learning models specifically designed for processing structured grid-like data, such as images and audio. CNNs consist of convolutional layers, pooling layers, and fully connected layers, allowing them to learn hierarchical representations of input data.

In CNNs, convolutional layers apply filters to input data, extracting features such as edges, textures, and shapes. Pooling layers reduce the dimensionality of feature maps, improving computational efficiency and increasing translational invariance. Fully connected layers perform classification or regression based on the learned features. CNNs excel in image classification, object detection, and image segmentation tasks, thanks to their ability to capture spatial hierarchies and translational invariance. They have been successfully applied in domains such as computer vision, medical imaging, and autonomous driving.

Despite their effectiveness, CNNs require large amounts of labeled data for training and are computationally intensive, especially in deep architectures.

Additionally, interpreting the learned features and decisions of CNNs can be challenging due to their complex hierarchical structure. [106]

- **Extreme Learning Machines (ELM)**, single-hidden-layer feedforward neural networks that employ random initialization of hidden layer weights and analytically determine output weights. Unlike traditional gradient-based learning algorithms, ELM trains the output weights directly using linear algebra techniques. ELM offers several advantages, including fast training times, excellent generalization performance, and simplicity of implementation. It is particularly suitable for scenarios with large-scale datasets and real-time processing requirements, where efficiency and speed are critical. ELM has found applications in regression, classification, and time series forecasting tasks, including financial prediction, medical diagnosis, and pattern recognition. Its simplicity and efficiency make it an attractive alternative to more complex and computationally intensive machine learning techniques. Despite its advantages, ELM may lack the flexibility and adaptability of other machine learning models, especially in tasks with complex data distributions and nonlinear relationships. Additionally, the random initialization of hidden layer weights may lead to variability in model performance across different runs. [81]
- **Random Forests (RF)**, an ensemble learning technique based on decision trees. RF constructs multiple decision trees during training by randomly selecting subsets of features and data samples. Each tree in the forest makes a prediction, and the final prediction is obtained through averaging or voting over all trees. RF is known for its robustness against overfitting and its ability to handle high-dimensional data, nonlinear relationships, and noisy inputs. It can be used for both classification and regression tasks and provides built-in measures of feature importance, making it suitable for feature selection. RF has found applications in various fields, including bioinformatics, remote sensing, and finance. It is widely used for data mining, predictive modeling, and anomaly detection tasks due to its scalability and ease of use. Despite its advantages, RF may suffer from computational inefficiency when dealing with large datasets and high-dimensional feature spaces. Additionally, the interpretability of RF models may be limited compared to simpler models such as decision trees. [28]

3.5.2 Ranking Results and Model Performances

The grid point of the IGS maps has been chosen at mid latitudes (lat 37.5, lon 15.0). Single global maps in IONEX format have been extracted from 2006 to 2017 and a time series has been reconstructed for the selected location.

To understand which external drivers are suitable for our ionospheric TEC forecasting model, we ranked the external drivers in order of importance with regard to the effect they have on TEC values.

NN Supervised Regression Models have been built by using the external drivers as input multivariate variable and TEC values as output at different hours, with regard to the hour of interest. The ranking has 2 hours time resolution and goes up to 24 hours ahead.

Since the lag between the solar and geomagnetic forcing and the response of the ionosphere is not fully known understood, we performed 12 different feature ranking at each forecasting horizon and then averaged their results.

The overall ranking of the features (external drivers) are depicted in Fig. 3.28

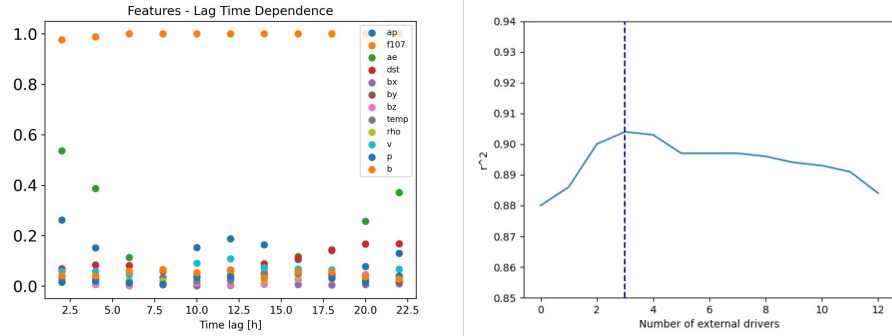


Figure 3.28: Ranking of external drivers. (Left) F10.7 was ranked first throughout the 24 hours. Correlation coefficient variability as function of the number of drivers, added one by one with decreasing importance (ANN model). Best choice of the number of external input is 3. (Right)

F10.7 is much more important than other drivers for TEC forecasting, at any lead-time, because the photoionization of atomic and molecular neutrals mainly due to solar EUV radiation. Auroral Electrojet (AE) exhibits a consistent ranking as the second most reliable indicator: indeed, the purpose of AE is to offer a comprehensive, numerical assessment of magnetic activity within the auroral zone, which is generated by intensified ionospheric currents occurring both below and within the auroral oval. Ideally, the concept refers to the comprehensive extent of variation at a given moment in time from the baseline values of the horizontal

magnetic field in the vicinity of the auroral oval. The ap index, planetary magnetic activity index, is also an effective driver, providing average values of the irregular disturbance levels in the horizontal field components, observed at selected magnetic observatories worldwide.

The consistency between by feature ranking results, obtained with neural network models, and the expected physical behavior instills greater confidence in our model, even in their utilization as predictive models for Total Electron Content (TEC).

The grid point used in the present analysis is close to the GNSS receiver installed in Noto station (NOT1: lat: 36.53, lon: 14.59) and the results are in accordance to the ones presented in [51], by the use of single station 30-second GNSS data from IGS regional data collections centers. Although they have a higher time resolution, they could only train the network by the use of 1 year time series (2018), not appreciating the differences of the solar cycle phases, and they had to face the missing value issue. This could be the reason why, by their training for just 1 year of the solar minimum could not appreciate the influence of geomagnetic indices, due to the lack of disturbed events. In their ranking, in addition to the F10.7 and AE they appreciated the total magnitude of the interplanetary magnetic field.

In order to build our predictive model, we required input dataset including the three best-ranked external features for different hours with 24 hours and VTEC times series. We took the previous 24 hours of the VTEC and the three best external drivers times series as the features of each instant of the training dataset. The reason why three is the our choice is in Fig. 3.28: trying to add one by one the other parameters, with decreasing importance, the performance model begin to decrease. It is reasonable to limit to three external drivers because increasing the number of them causes an increment in the trainable parameter size of the neural network model which can lead to overfitting and also affects the training time.

The coefficient of determination denoted as R^2 , is a statistical measure used to assess the goodness of fit of a regression model. It represents the proportion of the variance in the dependent variable that is explained by the independent variables in the model and it is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Where:

- y_i is the observed value of the dependent variable for observation i .
- $\hat{y}_i$ is the predicted value of the dependent variable for observation i obtained

from the regression model.

- $\bar{y}$ is the mean of the observed values of the dependent variable.
- n is the number of observations.

The R^2 value ranges from 0 to 1: higher values of R^2 indicate a better fit of the regression model to the data, suggesting that the independent variables are good predictors of the dependent variable. In Fig. 3.29 the model performance is shown, for each hourly prediction. The F10.7 solar index is an effective proxy of the external driver which influences the most the ionospheric TEC.

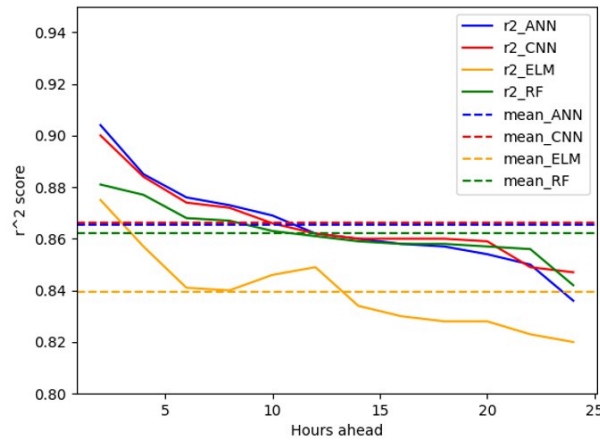


Figure 3.29: Coefficient of determination of each hourly prediction within 24 hours, with 2-hours time resolution, for ANN, CNN, ELM and RF.

In fact, the photoionisation of atomic and molecular neutrals is mainly due to solar EUV radiation originated by solar activity phenomena. On the contrary, solar wind parameters such as its density, speed and temperature seem to be less descriptive in this context. In Tab. 3.5.2 the mean R^2 values are reported for each model. The performance of the model decreases when the time lapse increases and the ELM, although has a training time much faster than the training time of ANN, CNN, has the lowest performance at all the timescales.

As a further step it is interesting to test the different response of the models during disturbed or quiet Space Weather conditions.

We have selected two different periods, both in the month of August 2018, which has not been used during the training and testing of the network. The first one (6th-10th) is a quiet period, where no significant flares or geomagnetic disturbances

Model	Mean R^2
ANN	0.87
ELM	0.84
RF	0.87
CNN	0.87

are ongoing, as shown in the Kyoto Dst plot in Fig. 3.5.2.

The second one (25th-31st) is a disturbed period, where a G3 storm is ongoing, as shown in the Kyoto Dst plot in Fig. 3.31.

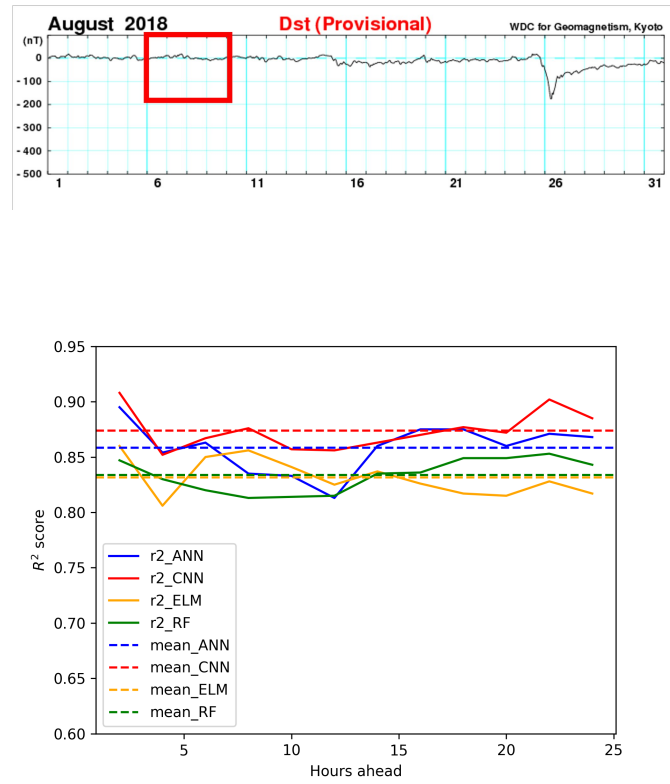


Figure 3.30: Model performance for VTEC forecasting during quiet geomagnetic conditions (bottom) with the corresponding Dst values in the red box used to test the model (top). 6-10 August 2018.

It is also interesting the model behaviour when a grid point of the IGS-GIM is selected at a different latitude.

We have then tried to repeat the same training and testing with the same NN models but at low and high latitudes in order to investigate the possible different ionospheric response. The results are plotted in Fig. 3.32

At high latitudes auroral indices are relevant as signature of the Disturbance

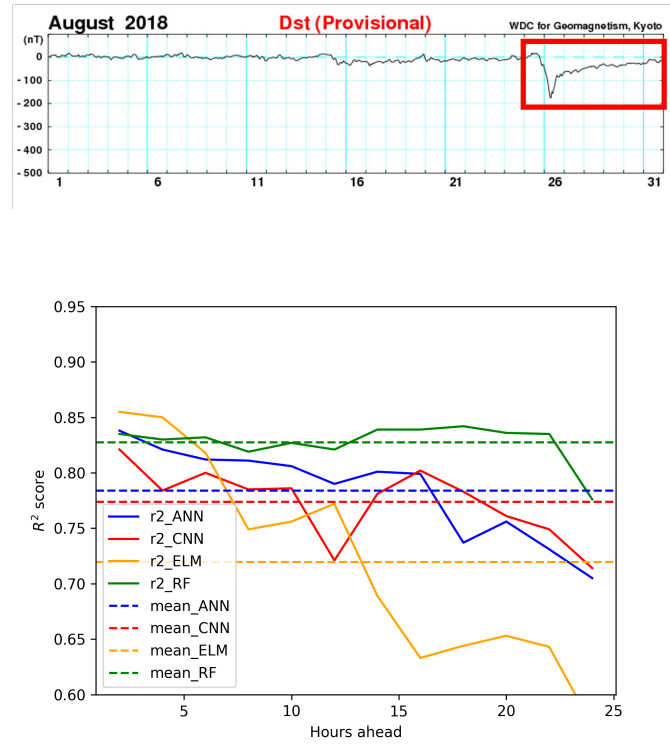


Figure 3.31: Model performance for VTEC forecasting during G3 geomagnetic storm with respective Dst values in the red box used to test the model. 25-31 August 2018.

Dynamo Electric Fields (DDEF) consequence of the energy input, resulting in Joule heating and collisional interactions in the high latitude ionosphere. At low latitudes auroral indices are relevant as signature of the Prompt Penetration Electric Fields (PPEF): penetration of interplanetary E during magnetic substorms, causing disturbances. The delayed evidence of the disturbed condition effects at lower latitudes, is due to the gravity waves, generated during geomagnetic disturbances in the auroral region by the auroral electrojet and Joule heating, that propagate from high latitude to midlatitude or even to the equatorial latitudes [84].

The Dst index measures the strength of geomagnetic storms caused by disturbances in the solar wind interacting with Earth's magnetosphere. It represents the deviation of the horizontal component of the Earth's magnetic field at the geomagnetic equator from its quiet-time level. High latitudes, particularly near the polar regions, experience stronger geomagnetic disturbances during solar storms and substorms and the Dst index specifically measures the strength of geomagnetic storms caused by disturbances in the solar wind interacting with Earth's magnetosphere. Moreover

TIME SERIES FORECASTING FOR SPACE WEATHER APPLICATIONS

Model	Quiet (6-10 Aug 2018)	Disturbed (25-31 Aug 2018)
ANN	0.86	0.78
CNN	0.87	0.77
RF	0.83	0.83
ELM	0.83	0.72

Table 3.5: Average of the model performance for quiet and disturbed geomagnetic conditions.

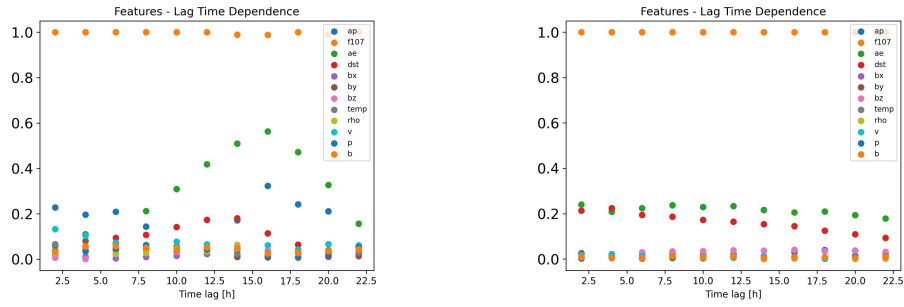


Figure 3.32: Low (left) and High (right) latitudes feature importance.

at high latitudes (85°N) F10.7, AE and Dst are the most important external drivers throughout the 24-hours of our forecasting horizon. Results are confirmed also for different longitudes, as shown in Fig 3.33

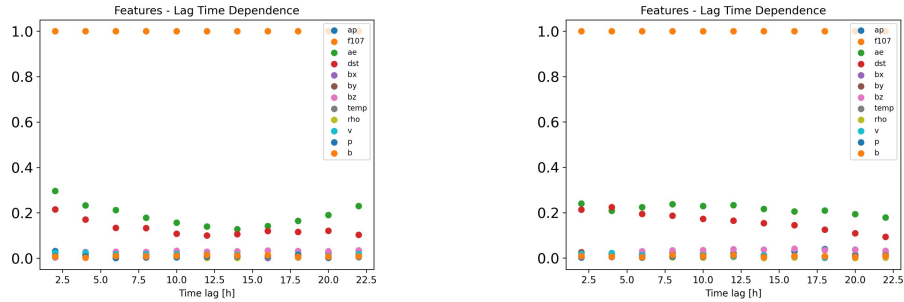


Figure 3.33: High latitudes feature importance: (85°N,0°E) and (85°N,120°E)

We have also observed (Fig. 3.34) that at low latitudes (20°N) the model response is much more reliable, compared to mid and high latitudes: at 20°N all the models are able to predict better the VTEC, while at 85°N all the model performances decrease more rapidly.

An alternative strategy can be to employ a model ensemble, which combines

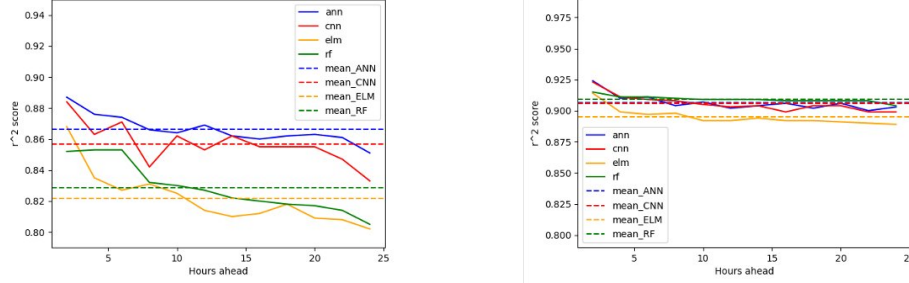


Figure 3.34: High (85°N,120°E) and low latitudes model performance (20°N,120°E)

the outputs of multiple individual models together has been presented in [212]. Ensemble learning reduces the risk of choosing a poor individual model, thus improving on the model selection procedure. An ensemble method called Stacked generalisation, more often referred to as Stacking, has been widely used across multiple domain areas [168] to improve predictive performance. It consists of feeding the outputs of multiple base models into a secondary layer model, or so-called meta-learner, which produces the final output.

In the quest of improving the performance of a predictive model, an ensemble approach called stacking, based on hierarchy of the models, has been also implemented and tested for our grid point data.

The results, after the tuning of hyperparameters and the model architecture, are shown in Fig. 3.35.

In conclusion, our results presented in the current section, demonstrated that in addition to the F10.7 index, a reliable prediction of the geomagnetic AE index, which has a hourly time resolution, would be very useful as effective driver for TEC forecasting problems.

Indeed, a further implementation that can be tested is the application of the transformer technology to a forecasting problem for AE.

Cesaroni et al. [33] in 2020 presented an empirical model designed to forecast TEC globally, 24 hours in advance, which runs operationally every 2-hours. An example of the global TEC map provided in output, in json format, is shown in Fig. 3.36. This model follows a two-step approach: VTEC single point prediction by training a Nonlinear Autoregressive Exogenous (NARX) model at selected 18 GIM grid points. As external input they currently use is the geomagnetic index Kp, available several hours in advance (up to 3 days) from NOAA (<https://www.swpc.noaa.gov>). In principal, the algorithm is able to use other parameters (Dst, AE, AU, AL) as

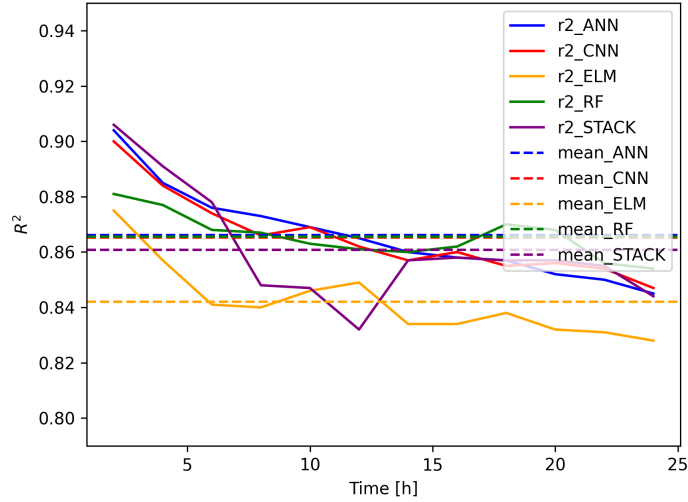


Figure 3.35: Comparison of ANN, CNN, RF, ELM and Stacked model.

external input but, so far, Kp is the only index available 24 h in advance.

In the second step, the forecast is propagated from local to global levels, facilitated by the combined use of NeQuick2 [144], suitably supplemented by an effective sunspot number R12, determined through minimizing the root mean square error (RMSE) between NARX output and NeQuick2 predictions at the same GIM grid points.

Although F10.7 index resulted to be the most effective index, it is a daily value, therefore it can not be used operational for this purposes.

As a future perspective, a further upgrade for our model could be the substitution

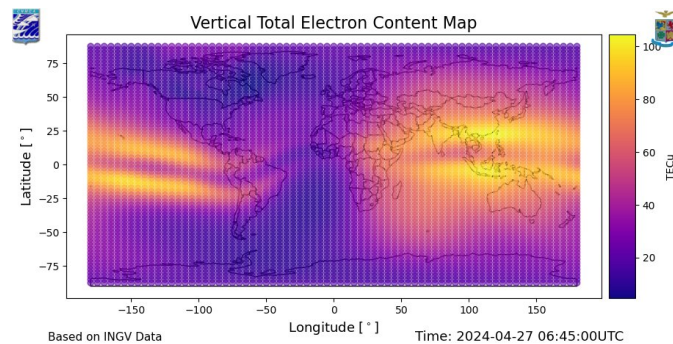


Figure 3.36: Global map of VTEC. Sample graph of the json output of the INGV model. [33]

of the loss function with a physics-informed loss function could be implemented, as proposed recently in literature [15], properly designed for the task of forecasting the ionospheric VTEC, which seems very promising.

Chapter 4

Diagnostics by Synoptic Solar Radio Imaging

The study of the Sun's atmosphere through ground-based and space monitoring programs at different wavelengths represents a major topic of present astrophysics. However, some radio frequency bands surprisingly lack solar observations, despite being relevant for investigating many astronomical and astrophysical phenomena at large.

All the plasma processes, such as magnetic reconnections, shocks, particle acceleration [134]) contribute to a complex and not fully understood radio emission picture needing spatially-resolved and time-resolved scientific data in order to fully explore their nature. These solar radio data are crucial to complement the wealth of existing information in the UV, Optical, IR and X-ray domains.

Compared to the emission in other spectral ranges (e.g. the EUV), the radio quiet Sun has the advantage of being understood as originating mostly from thermal bremsstrahlung in local thermodynamic equilibrium, with the addition of sporadic and variable gyro-magnetic emission in active regions [49].

Long-term diachronic observations of the solar disk at high radio frequencies represent an effective tool to characterise the vertical structure and physical conditions of the solar chromosphere both for quiet and active regions, during their evolution at different phases of the solar cycle.

Radio Sun can thus be used as a powerful diagnostic of the physical conditions in a wide range of atmospheric layers, including the problem of the heating of the chromosphere and the corona [179].

In particular, mapping the brightness temperature of the free-free continuum radio

emission in the centimetre and millimetre range is an effective tool to characterise physical parameters of the chromosphere [60]. As the opacity increases with the wavelength of observation, the effective height of formation moves from the temperature minimum (at sub-millimeter waves) to the low corona (mostly visible at meter waves). Among these extreme wavelengths, the chromosphere is visible in our range and it hosts most of active solar features linking the photosphere to the corona with a temperature rise of still unclear origin. Solar atmospheric models are critically constrained by actual temperature brightness in K-band (18 – 27 GHz) and above [178], therefore it is crucial to have dedicated reliable observations at this frequencies.

This portion of the spectrum is suitable for detailed measurements of the chromospheric brightness temperature of the QS and to assess the rich thermal vs. non-thermal content of AR emission.

Multi-wavelength observations could provide insights into the formation process of the ARs hosting major flares and determining the slowly varying component of the Sun over the 11-year cycle. In particular, observations of strong gyro-magnetic emission over sunspots groups can be used to test in detail the magnetic field model extrapolated from measurements at the photosphere. In this framework, it is important to enlarge the sample of measurements of intensity, spectra, and size of the bright features on the solar chromosphere as a function of the frequency.

Synoptic solar radio maps from TSRS 2.0 at 18 GHz are still not available, therefore we have analysed the images provided by different single-dish antennas, operating in the framework of National Space Weather Radio Service, at the same or slightly higher frequencies in the K-band.

In this Chapter the state-of-the-art of the projects involving our INAF solar radio team in the Space Weather framework will be presented.

High-frequency radio observations with large single-dish radio telescopes of the INAF network provided ~400 solar images since 2018 (SunDish project), useful to monitor the vertical structure and physical conditions of the solar chromosphere both for quiet and active regions, during their evolution at different phases of the solar cycle.

Solar radio mapping in K-band (18-26 GHz) can probe the chromospheric magnetic field of the Active Regions through the detection of gyro-resonance spectral components related to flare events. In particular, we detected peculiar spectral variations in the active regions anticipating flare events, deserving further investigations as a probe for magnetic fields in the chromosphere and a possible valuable forecasting

method for the Space Weather hazard.

From SunDish observations we noted that enhanced magnetic fields (up to $\sim 1.5\text{--}2 \cdot 10^8$ nT) determine a spectral flattening ($\alpha < 1.5$) in the ARs compared to pure free-free emission ($\alpha \sim 1.9$) due to the addition of a steeper gyro-resonance component also associated to circular polarisation up to $\sim 40\%$.

When this sporadic anomalous AR spectrum is detected, the probability of a strong flare occurrence within 1-2 days is $>80\%$, further rising at $>90\%$ also requiring that AR brightness temperature exceeds $\sim 50\%$ the quiet Sun level (10000 K).

We present several examples of 18-26 GHz radio images showing peculiar AR spectral and polarisation configurations anticipating or following flare events. Through correlation statistics analysis, we discuss the sensitivity and robustness of this flare forecast method in the perspective of coupling it with other multi-messenger Space Weather proxies.

4.1 Single-Dish Solar Radio Imaging (SunDish)

SunDish is a solar radio imaging system implemented through the upgrade of the large single-dish telescopes of the Italian National Institute for Astrophysics (INAF), not originally conceived for solar observations. This project is devoted to radio imaging and monitoring of the solar atmosphere at high radio frequencies (at present between 18 and 26 GHz) through single-dish observations and it is a valuable tool to fill the observational gap in the field of solar imaging at these frequencies.

The INAF single-dish radio telescopes network involved in SunDish project, started in 2018, includes the 64-m Sardinia Radio Telescope (SRT) and two 32-m antennas at Medicina and Noto.

INAF radio telescopes required additional setups and accurate electromagnetic and thermal assessments to guarantee instrumentation safety during solar observations and ensure a linear response to the strong solar input signal impacting the systems' amplification chains. After a first development phase aimed at defining and optimising solar observing configurations and techniques for the radio telescopes and calibration issues, the system is providing early science results. These include the accurate measurement of the brightness temperature of the radio quiet Sun component, the solar radius and outer coronal emission that has been poorly explored in the 18-26 GHz range to date, constraining atmospheric models in term of density and temperature.

In K-band, the solar brightness ($\sim 5000 \text{ Jy/arcmin}^2$) is over three orders of magnitude higher than typical radio-astronomical calibration sources. The implementation of variable filters for additional signal attenuation, included in the receivers' amplification chain, was crucial to avoid electronic saturation in signal response and possible instrumental damage. After equipping the telescopes with a suitable signal attenuation setup, the typical variability of solar phenomenology at these frequencies is compatible with the available instrumental dynamic range. In fact, apart from strong flare episodes, most of the enhanced emission associated with ARs in the 18 – 26 GHz range does not typically exceed the QS brightness level by orders of magnitudes, as it happens instead at lower radio frequencies in the MHz-GHz domain.

Accurate measurements of the brightness temperature of the QS component are lacking in the 20 – 26 GHz range. Present information suggests a possible spectral change, such as flattening, in the quiet solar chromosphere at these frequencies, but the available measurements are scattered due to the difficulty of separating the pure QS component from the contribution by the ARs.

Spectral variations in the observed spectrum of solar ARs appear might be used as important proxy in predicting strong flares [155]. Radio monitoring campaigns, focusing on at least two frequencies at the low/high edge of the K-band, can provide spectral index estimates for the ARs and detect anomalous spectral variations (e.g., sudden flattening) as a signature anticipating the flaring process.

Most of the images that have been studied in the present work (2018 – 2023) are performed by the 32-m Medicina radio telescope and few sessions by the larger SRT. These receiver can observe the targets one at a time, with a quick frequency switch when necessary. In recent years, a new control system (DISCOS) was designed and produced by INAF in order to enhance the performance of this telescope – originally conceived for VLBI observations – as a single-dish instrument.

Fast On-The-Fly (OTF) mapping is now one of the built-in observing modes; its sky coverage efficiency is doubled in the case of the K-band receiver, as it is a dual-feed device. The overall K-band observable portion is 18 – 26.5 GHz, with an instantaneous configurable bandwidth of 2 GHz maximum. The output of each feed consists of two separate lines: LHCP and RHCP, that reach an analog total-power back-end. The frequency and bandwidth actually employed may slightly vary with time according to the incidence of RFI (Radio Frequency Interference).

Observing schedules are produced using a custom generator, in turn, relying on solar ephemeris computed with the NASA JPL Horizons web tool [142], as the ob-

servation of Solar system targets is currently not offered among the built-in options of the Medicina control system.

Obtained solar images have typical resolutions in the 0.7–2 arcmin range and a brightness temperature sensitivity < 10 K. Accurate calibration adopting the Supernova Remnant Cas A as a flux reference provided typical errors $< 3\%$ for the estimation of the quiet-Sun level components and active regions flux measurements. Flux density calibration is achieved by observing calibration sources in either cross-scan mode or mapping one. Mostly, the Supernova Remnant Cas A is observed producing 40×40 arcmin maps, with 2.0 scan/beamsize, in order to obtain the conversion factor from arbitrary counts to flux density units, expressed in units of Jy. The estimate of the atmospheric opacity – one for each frequency – is possible thanks to the Skydip observing mode, which measures the sky brightness at a range of elevations. Skydip acquisitions require additional 12 minutes to complete the data set.

In order to acquire spectral information, observations are carried out at the boundaries of the available radio frequency band: maps are acquired first at 18 GHz, then at 26 GHz. The two maps are obtained by scanning the desired area in Right Ascension (RA) and Declination (Dec), respectively. This choice reflects the need to reduce the total observing time while permitting to investigate of weak disk features that, in a single map, might be altered – or even created – by striping effects, usually due to fast-variable weather conditions differently affecting the many sub-scans composing the map (RFI might also contribute to this issue) [155].

Image production and calibration are performed using the SRT Single-Dish Imager (SDI), which is an IDL (Interactive Data Language) tool designed to perform continuum and spectropolarimetric imaging, optimized for OTF scan mapping, and suitable for most receivers/back-ends available for INAF radio telescopes [128, 127]. The control system in use at the INAF radio telescopes provides celestial coordinates for each measurement performed during the OTF scans. These scan measurements are then binned through an ARC-tangent projection using pixel sizes of about 1/4 of the HPBW, which corresponds to the effective resolution of the images. FITS images are then produced and ready for scientific analysis, such as image RMS/sensitivity, dynamic range, and brightness profiles.

In order to convert the raw maps (where intensity is expressed in “machine counts”, directly proportional to the input signal level) to brightness temperature images, calibration observations are required.

The raw images are obtained in counts, an electronic measurement given by the

back-end in use proportional to the flux density. In order to obtain a map expressed in a physical quantity (flux density or brightness temperature), there is the need to observe a calibration source – a celestial object with a flux density, or brightness temperature, well known and studied in the literature – to calculate the conversion factor from counts to a physical unit of measurements for each observing session, since the counts value depends on various factors (e.g. the attenuation set up, the sky opacity, or the meteorological conditions). A Gaussian fit of the image histogram (counts distribution among pixels) provides an accurate estimate of the average counts from the QS.

The brightness reference for the QS emission comes from the work of Landi and Chiuderi Drago [103], although no QS measurements were reported in the 18 – 30 GHz range in their work. The observed spectrum of the QS is characterized by a break at about 10 GHz, which corresponds to the brightness temperature of about 12000 K referred to the center of the solar disk. Above this frequency value, the spectrum is fitting ($\chi^2 = 0.032$) a logarithmic linear relation between the brightness (T_b , in units of kelvin) and the frequency (ν , in units of Hz):

$$\log_{10}(T_b) = a + b \cdot \log_{10}(\nu) \quad (4.1)$$

where $a = 6.43$ and $b = -0.236$, although an error estimation is not provided. In the 18-26 GHz interval, QS is predicted to have a value between 10,227-9,377 K.

4.1.1 Active Regions Spectral Index

The identification of the ARs and the definition of their parameters – e.g., the calculation of their fluxes and spectra – is a non-trivial task.

For the observations where AR brightness (including lower and upper limits) is available at two frequencies (characterized by high/medium quality), spectral index ($S_\nu \sim \nu^\alpha$) measurements are provided.

In this Section we present the methodology and the original results that we have obtained from a statistical study, that aims to investigate the correlation between AR spectra at high radio frequencies and solar flare occurrences.

For ARs parameters identification purposes, a dedicated code developed in Python called SUNDARA (SunDish Active Region Analyser), has been used for a preliminar automatic screening of the maps, based on several Python packages. SUNDARA receives as input the solar images (in standard FITS format) provided by the data processing pipeline and provides automatically the identification and a complete

analysis of the solar disk and ARs with physical information such as brightness temperatures, fluxes and spectral indices, with respective errors [129].

The first analysis of data images, acquired from 2018 to 2020, is published in [155]. The AR spectral distribution peaked between the Rayleigh-Jeans (RJ) and the QS spectrum ($\alpha_{QS} \sim 1.75$ in our frequency range), suggesting the presence of emission significantly deviating from simple thermal bremsstrahlung in the optically thick regime ($\alpha_{RJ} = 2$).

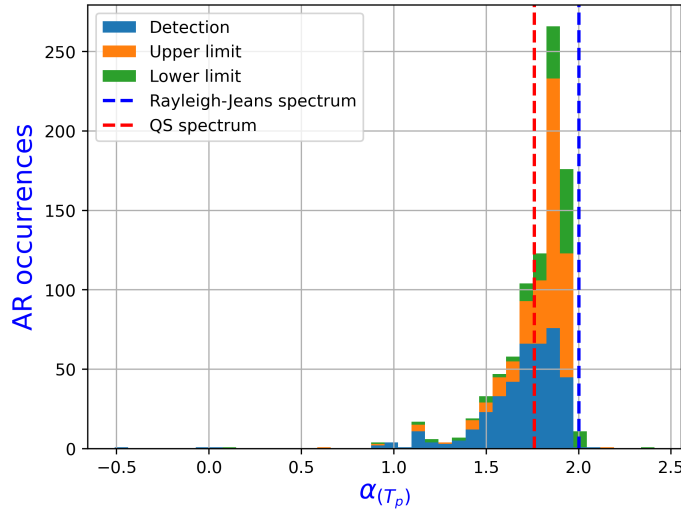


Figure 4.1: Histogram of the flux density spectral index values (2018-2023). Blue counts indicate detections orange and green counts show upper limits and lower limits, respectively. Dashed lines indicate the spectral indices of Rayleigh-Jeans (blue) and of the QS-level as estimated by the work of Landi and Chiuderi Drago.

Radio monitoring campaigns from ARs spectral indices estimation continued and an updated analysis of the solar maps acquired up to the end of 2023 in the same frequency range of 18–26 GHz. Data confirmed important clues on anomalous spectral variations (e.g., sudden flattening) as a signature anticipating the flaring process. In Figure 4.1, we report the spectral index distribution obtained by comparing AR peak brightness temperatures (including the QS component) at different frequencies. Out of the 888 ARs (~ 400 is only detection at both frequencies are considered) analysed in our 2018 – 2023 data set, this distribution peaks at the expected spectral index ($\alpha_{RJ} \sim 1.9$) for thermal bremsstrahlung in the optically thick regime. On the other hand, $\sim 13\%$ of the the ARs spectra is significantly softer ($\alpha < 1.5$) and is not compatible (within 4σ level) with a pure free-free thermal

Yr	$N_{\alpha < 1.5}$	$N_{\alpha > 2}$
2018	1/16 (6.3%)	0/16 (0%)
2019	0/10 (0%)	0/10 (0%)
2020	5/33 (15.2%)	0/33 (0%)
2021	9/96 (9.4%)	1/96 (1%)
2022	24/148 (16.3%)	2/148 (0.7%)
2023	20/96 (16.7%)	0/90 (0%)
TOT	59/399 (14.8%)	3/399 (0.5%)

Table 4.1: Rate of the ARs anomalous spectral index for each year. All the identified ARs, both on the limb and on the solar disk, have been considered. Upper limits and lower limits have been excluded, as only detection have been considered.

emission ($\alpha=1.9\pm0.1$).

In Table 4.1 the rate of the anomalous spectral indices is shown, for each year from 2018 to 2023. Nearly 400 solar maps have been acquired by SunDish in both two frequencies. Significant spectral variations and/or the formation of a local maximum in the microwave spectrum of solar ARs already appeared to be an important factor in predicting powerful flares (within 24 – 48 hours from the event). In addition to thermal bremsstrahlung, the concurring presence of sporadic gyro-magnetic components in the AR emission could contribute to this spectral flattening. We could explain these spectra of peculiar AR events by considering the contribution of a steep high-frequency tail provided by gyro-resonance emission peaking slightly below our frequency range. It is important to note that the measured spectral indices represent an average value within the beam and thus differ from the actual values rising from the smaller-scale unresolved regions that emerge from the diffuse free-free emission.

Gyro-resonance emission at radio frequencies has been extensively studied by various researchers aiming to estimate that the peak goes up to ~ 20 GHz.

Studies have indicated that the gyro-resonance emission for radio frequencies > 10 GHz is produced in regions of magnetic fields with intensities of kG (located close to the transition region) and that at 17 GHz, the lower limit for the magnetic field intensity to produce this emission (due to the third harmonic) is $\sim 2 \cdot 10^8$ nT at the photospheric level. Moreover, the enhancement of gyro-magnetic effects when leaving the minimum of solar activity could cause a significant increase in the AR average brightness up to 10^5 K and higher.

An enhancement in gyro-magnetic emission might serve as a precursor to flaring gyrosynchrotron emission originating from regions undergoing anomalous energy release, characterized by sub-relativistic plasma in moderately strong magnetic

fields [179].

In active regions, at microwaves, gyroresonance opacity is competing with free-free opacity. Free-free emission is ubiquitous in the corona but whenever sunspot-associated microwave sources of coronal brightness temperature or/and high degree of circular polarisation appear, they can be safely attributed to gyroresonance emission. In these cases, the sunspot's photospheric magnetic field should be strong enough to bring harmonic layers of the gyrofrequency above the base of the TR. On the other hand, microwave free-free emission is spatially extended, its brightness temperature is smaller (because it is optically thin) and its degree of polarisation is low [146]. Imaging observations above 30 GHz reported in the literature are not conclusive about the possible presence of gyro-resonance spectral contributions in addition to free-free emission.

Few of the observed spectral indices are also higher than QS values and this could suggest a thermal gradient within the vertical structure of typical ARs that differs from QS emission models. The presence of peculiar hard spectra ($\alpha_T > 2$) could provide clues of gyroresonance lines peaking slightly above our frequency range (rising part of the spectral component). In particular, Table 4.1 shows that, considering only spectral index detections (where ARs are identified at both frequencies), 3 events exhibit spectra harder than RJ emission. In 2 of those cases (maps from Medicina of 30 August 2021 and 23 May 2022) the value is still compatible with a thermal spectra within the error ($\alpha_T = 2.04$ and $\alpha_T = 2.05$ respectively).

On the other hand, in the case of AR 3017 (observed on 22 May 2023) $\alpha_T = 2.41$, therefore it exceeds the error margin by over 4σ , indicating it is not consistent with a thermal spectrum. A possible explanation for the high value of the brightness temperature observed at 25.8 GHz could be related to the C5 flare occurrence during the acquisition of the higher frequency map. The observed spectra hardening, also by the inclusion in the statistics of lower limit spectral indices requires a further investigation and it will be analysed in a dedicated future work.

4.1.2 Flare Statistical Correlation Analysis

The behavior of the chromospheric footpoint plasma during the impulsive phase of solar flares remains ill-understood theoretically but the magnetic reconfiguration that allows the rapid release of stored magnetic energy in a flare is generally agreed to occur at the chromospheric heights.

The primary energy release occurs the impulsive-phase of the solar flare and it is concentrated at the chromospheric endpoints of the magnetic field involved in the

magnetic reconnection. Thus, the radiation produced in the flaring chromosphere is an important, yet not fully understood, source of diagnostic potential of physical processes [53].

The prediction of major solar disturbances and their impact on Earth is one element to mitigate Space Weather hazards. It is presently not possible to predict the Space Weather impact of a solar event before the first signatures of the solar event can be observed. Radio emission is among the earliest observed signatures of an eruptive event in the corona [174].

The aim of the work is to understand if changes in the magnetic field configuration of the ARs and the eruptive events that have taken place one or two days prior in the same region could be statistically correlated. The rearrangement of the ARs magnetic field could lead to a change in the balance between the thermal and non-thermal emission component, with a prevalence of the gyro-magnetic emission that, for example, can produce spectral index inversion, from positive to negative values. We have selected the ARs of solar radio maps associated to peculiar low peculiar spectral indices ($\alpha < 1.5$) and significant flares that occurred within 30 hours from a SunDish observation.

Regarding the flare selection, we have considered all the events in the catalogue by Plutino et al. [162], based on the long channel (1–8 Å) in the soft X-ray (SXR) detected by the GOES satellite (NOAA/NASA) constellation. Their catalogue has been extended¹ lately, up to April 2023.

For the last 7 months still missing we have checked manually, one by one, the presence of flares above our thresholds, from GOES data.

In order to study the correlation between intense flares and the characteristics of the sunspot that produced them, we needed to define a flare intensity threshold.

Since the analyzed images (from 2018 to 2023) cover both the period of solar minimum and the rising phase of the solar cycle, we decided to rank the flares in ascending order and consider in our study, for each semester, only the 3% strongest flares of that epoch. In Tab. 4.2 the threshold chosen for each semester from 2018 to 2023, for the purpose of this study. We have then defined three different categories for each of the considered event:

- TP, true positive cases, where AR with a spectral index lower then 1.5 has been identified and a flare coming from the same AR has been detected in the previous or following 30 h

¹Ref: <https://zenodo.org/records/10560189>

Semester	Threshold Class
2018 Jan-Jun	C1
2018 Jul-Dec	B4
2019 Jan-Jun	C2
2019 Jul-Dec	B2
2020 Jan-Jun	B7
2020 Jul-Dec	C1
2021 Jan-Jun	C2
2021 Jul-Dec	C4
2022 Jan-Jun	C6
2022 Jul-Dec	C7
2023 Jan-Jun	C8
2023 Jul-Dec	C8

Table 4.2: Thresholds for the flare intensity class, defined for each semester from 2018 to 2023.

- FP, false positive cases, where AR with a spectral index lower then 1.5 has been identified but no flare coming from the same AR has been detected in the previous or following 30 h
- FN, false negative cases, where a flare has been detected within the 30 hours preceding or following our observation but the associate spectral index is greater then 1.5

We have selected a total number of 36 solar radio maps, including 59 different ARs. The Table that summarize all the 124 solar observations (99 TP, 13 FN and 12 FP) that has been used in the present analysis, is given in Appendix 4.2.

In all the remaining cases not reported in the Table 4.2, we have True Negative (TN) events ($1.5 < \alpha_T < 2$), which is consistent with thermal emission, indeed no flare occurred within 30 hours from the observation. The total number of cases, when also considering the upper and lower limits of spectral indices, are 816, which represents $\simeq 91.9\%$ of the events.

The combination of relevant AR features for flare prediction can improve the forecast performance, therefore the parameters that we took into account in this study are:

- the maximum T_p brightness temperature observed for each AR,
- Spectral Index α_T from the 2 observing frequencies at ~ 18 GHz and ~ 25 GHz,
- the magnetic complexity based on the Mount Wilson Magnetic Classification

[64], assuming 0 for low (α and β) and 1 for higher classes of magnetic complexity,

- the V-stokes, maximum percentage of circular polarisation component (LHCP and RHCP),
- Flare class, for flares occurring within 30-hours from solar maps provided by SunDish observation
- Time delay Δt between the flare and the image, assuming positive when the flares occurs after the observation and positive when the flare occurred before our SunDish observation

The general result of the correlation analysis shows that events with anomalous spectral index-flare association (True Positive, flag TP) are $\sim 80\%$ of total entries in Appendix 4.2, while significant spectral flattening not associated to strong flare occurrences (False Positive, flag FP), and intense flare without anomalous spectrum association are both $\sim 10\%$ (False Negative, flag FN) of the events in the Table in Appendix 4.2.

Therefore when an anomalous spectral index is detected there is $\sim 89\%$ probability (i.e tool precision $TP/(TP+FP)$) of an intense flare (for the considered class threshold for that epoch) occurring within ~ 30 hours. With the application of our proposed tool based on the spectral index, the sensitivity, expressed as $TP/(TP+FN)$ is $\sim 88\%$, while the false alarm rate $FP/(TN+FN)$ is 3.5% .

Only $\sim 10\%$ intense flares are missed by this forecast method (miss rate given by $FN/(TP+FN)$), although most of them still present AR spectral index close to our $\alpha = 1.5$ threshold.

Table 4.3 presents the confusion matrix for the classification tool, obtained from our statistical analysis of the flare/no-flare spectral index correlation.

AR spectral estimation close to the solar disk limb can be affected by observational

	Predicted Positive	Predicted Negative
Actual Positive	TP = 99	FN = 13
Actual Negative	FP = 12	TN = 816

Table 4.3: Tool Performance Evaluation: Confusion Matrix showing total number of cases for each category (True Positive, False Negative, False Positive and True Negative.)

biases as for example the lack of visibility of the entire AR.

In the context of our analysis we define an AR on the limb if its distance from

the center of the Sun exceeds 90% of the solar disk radius. If we exclude the 24 AR-flare limb events from our statistics, the sensitivity, or recall, increases from 89% to 95%.

Although the statistics dropped, the F-score of the potential tool, defined as the harmonic mean of the model's precision and sensitivity, goes from 88% up to ~90%.

In Figure 4.2 we plotted the spectral index against the peak brightness temperature for each event, the Pearson correlation coefficient, taking into account all the events, is $C_t = -0.74$.

When the temperature increases, the spectral index decreases, suggesting that the enhancement in T_{peak} is a clue for observing gyroresonance emission. Solar flares

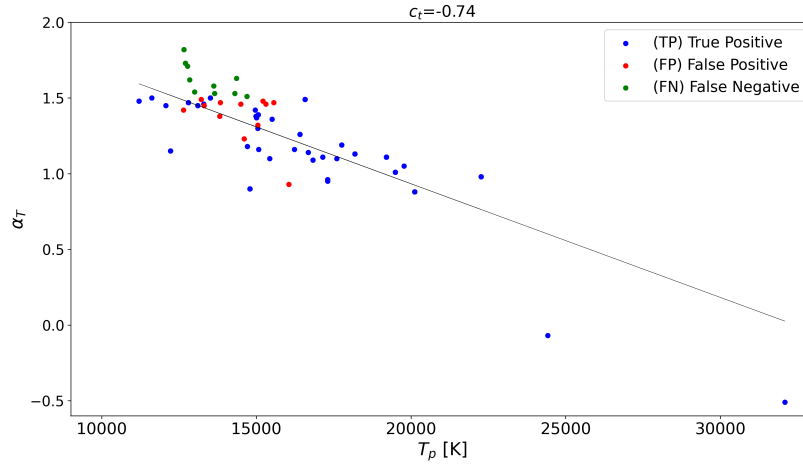


Figure 4.2: ARs Spectral Index correlation with T_p . In blue, red and green respectively the three categories TP, FP and FN. Pearson correlation coefficient C_t is shown.

are commonly observed at the ARs with different magnetic field complexities. The magnetic complexity is determined following the classification scheme of Mount Wilson [64], where [70]:

- α is a unipolar sunspot group
- β is a bipolar sunspot group, with the simple and distinct division between opposite polarities
- γ is a complex sunspot group with irregular polarities
- δ is a sunspot group with umbra having opposite polarities within a penumbra and spans less than

- γ - δ is a sunspot group containing one or more sunspots
- β - γ is a bipolar sunspot group with more than one continuous polarity reversal line
- β - δ is a sunspot group containing one or more sunspots
- β - γ - δ is a sunspot group containing one or more sunspots

Associations between flare productivity and AR properties have been discussed extensively in previous studies and it is generally believed that intense flares preferentially originate from the ARs containing large sunspot groups with strong magnetic fields and complex magnetic field configurations [52].

Our K-band observations specifically probe the magnetic field in the chromosphere where reconnection process responsible of flares are expected to occur. The link between chromospheric fields with those in deeper layers of the solar atmosphere deserves further investigation. Figure 4.3 shows the histogram of the number of ARs magnetically complex or not, for each of the three categories we defined.

The true positive class is the only one comprehending most of the AR with high degree of magnetic complexity. Therefore, introducing the high AR magnetic com-

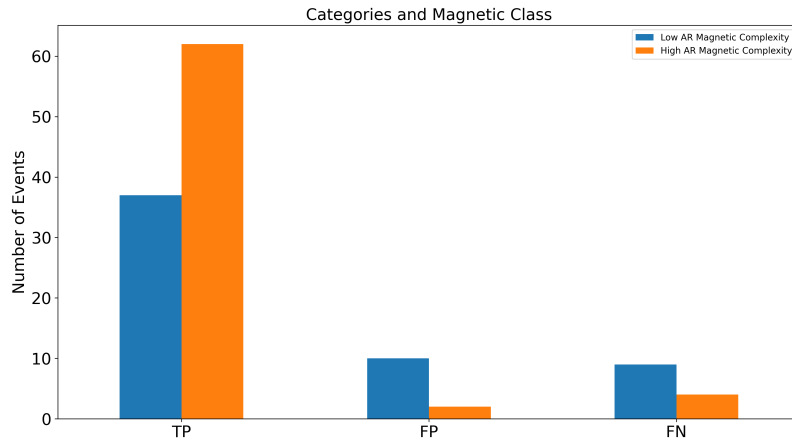


Figure 4.3: Histogram of the number of ARs with low and high magnetic complexity for each of the three categories (TP, FN, FP)

plexity as an additional flare proxy parameter accordingly with literature [149], it improves the true positive rate, despite this parameter is more related to the photosphere conditions.

Similarly we took into analysis the peak brightness temperature of the considered AR. Most of the events when a spectral softening is related to the flare occurrence

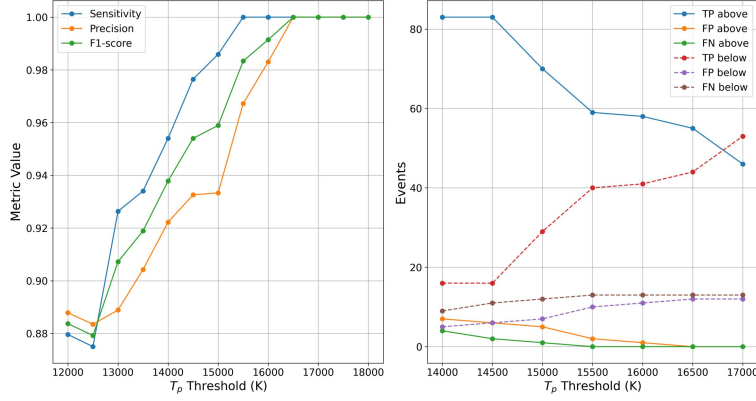


Figure 4.4: Statistical performance (left) for different metrics as a function of temperature threshold and (right) number of events for categories, by varying the tool T_p threshold.

have also an enhanced brightness temperature. On the other hand, false positive or false negative events are associated to T_{peak} lower then 15500K.

A temperature threshold of 15500K can be a reasonable choice from the plots in Figure 4.4. Indeed, it is a good compromise between the need to maximize the number of TP (14500 K) and also the F1 score (16500 K). Moreover, there are no events flagged as FN or FP when AR brightness temperature is above this value (Fig. 4.2).

When the sporadic anomalous AR spectrum ($\alpha < 1.5$) is detected and we also require AR peak brightness exceeds $\sim 50\%$ the quiet Sun level (> 15500 K), the precision rises but at the expenses of a much higher flare miss rate (FN).

The observed correlation is also confirmed in a work by Kallunki et al., who recently stated that intense maximum radio brightening is needed in ARs to cause solar flare, from their statistical study by the use of data from Metsähovi Radio Observatory (MRO) of Aalto University [92].

In Figure 4.5 the histogram of estimated excess brightness temperature respect to the QS ($\sim 10000K$) above flaring ARs. It results to be $\sim 5000K$. The time delay (positive or negative) between flare detection and our solar observation of flaring AR spectral flattening has been also investigated. From the beginning of the first SunDish observing session and the end of the second one, for the two different K-band frequencies that allow the spectral index computation, approximately two hours are required. Even taking this time lag into consideration, from our prelimi-

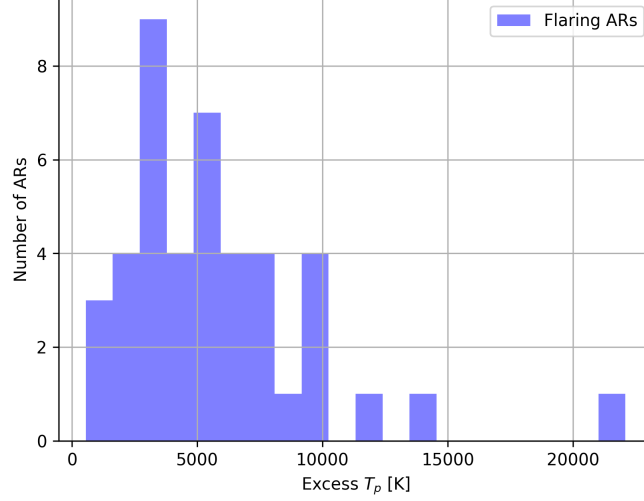


Figure 4.5: Histogram of the excess temperature T_p respect to the quiet Sun for flaring ARs (TP and FN categories).

nary analysis in $\sim 49\%$ of the cases soft spectra AR are observed after the solar flare (negative time delay). Therefore in our study we did not show a preference respect to the anomalous spectra observed before flares (positive time delay).

AR spectral observations very close to flare events could be more likely associated to a intense gyro-magnetic component, but the present statistics cannot prove that, although very flat spectrum are seen close to flare in different cases: e.g. $\alpha_T < 0$ within ~ 1 hour of a C8.6 flare at MJD 59960. On the other hand, most of detected AR spectra with $\alpha_T < \sim 1$ are associated with strongest flares (Fig. 4.6), mostly with M and X-class flares.

As we observed a dependence of the spectral index on temperature (Fig. 4.2), we have defined the new index T_p/α_T , that seems to be promising as flare proxy from our correlation statistical study based on SunDish.

In the histograms in Figure 4.7 we show the difference of its mean value depending on the three categories of TP, TN and FP, by the gaussian fit. True positive real flare events peak at higher value then the missed cases.

The additional investigation on polarimetric mapping of solar disk emission features could allow us to correlate magnetic information with AR spectra, helpful in disentangling gyro-magnetic contributions from free-free emission.

The eventual inclusion of the Stokes V in a more complex index could be considered, although the circular polarisation intensity is also related to geometrical and

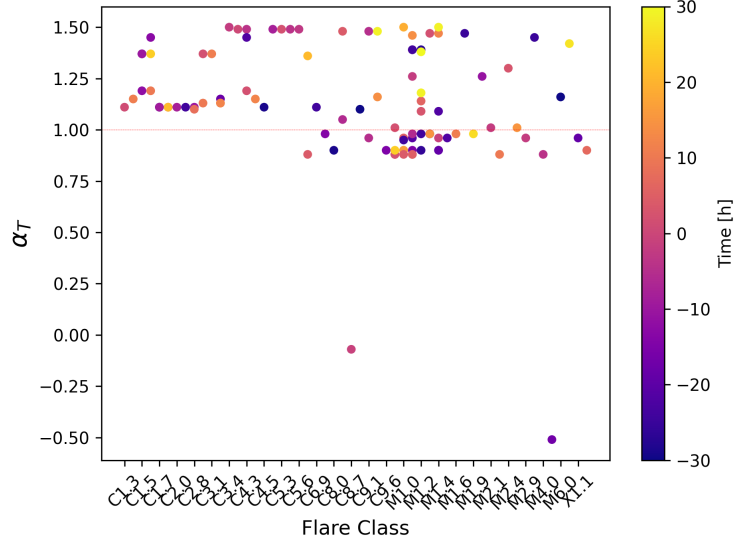


Figure 4.6: Spectral index of the flaring AR (TP and FN categories) and corresponding flare class. Colors indicate the time lag between the flare occurrence and the map. The red horizontal line is referred to $\alpha_T=1$.

line-of-sight issues.

Medicina station allows dual-polarisation observing mode and for the first time we have tried to study the obtained maps. Most of the disk exhibit a QS component which is almost circularly unpolarised, with a $V \leq 0.1\%$.

In active regions, the observed values of circular polarisation are typically higher than the values reported for the QS. Higher values of polarisation degree in AR can be explained by stronger magnetic field in sunspots than in the QS, therefore indirect magnetic field measurements in active regions will be feasible.

In Appendix 4.2 we reported, for each event, the RMS of a background region where no sunspots have been identified on the photosphere, which shows $V_{RMS} \approx 0.05\%$ for most of the maps.

If considering only the ARs with a circular polarisation emission at 18 GHz greater than 1% we have that they are about 75% of the total number of events in our statistics and > 80% of them belong to the TP category. AR producing significant flares with a spectral index not compatible with thermal emission (true positive events) in 76% of the cases are polarized (79% if we set a greater threshold $V > 2\%$). On the other hand, only 33% of the ARs exhibit polarisation percentage >1% when no flares have been detected.

In the Figure 4.8 the correlation analysis between the spectral indices of the consid-

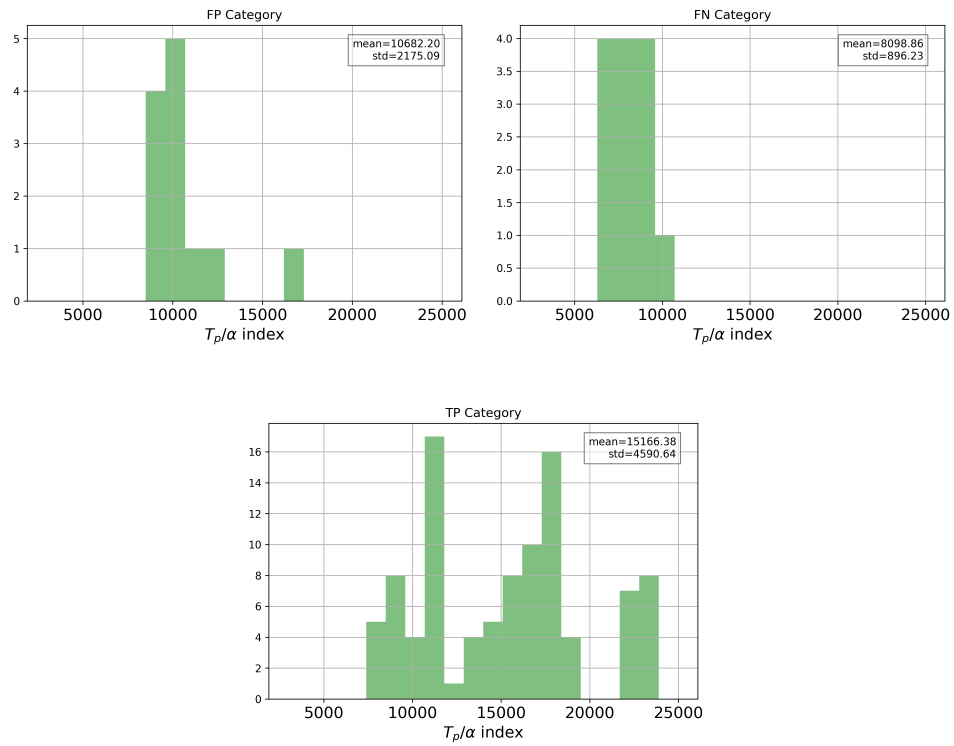


Figure 4.7: T_p/α_T index distribution. Parameters from Gaussian Fit, for each of the three categories (False Positive, False Negative and True Positive) are also shown.

ered events α_T and the absolute value of the V Stokes, e.g the maximum percentage of circular polarisation of the corresponding AR.

Indeed, strong anti-correlation obtained in the plots for high degree of polarisation is likely referred to sporadic ARs dominated by a strong gyro-magnetic component. The goal of our study is to set the basis for a useful tool that could be used

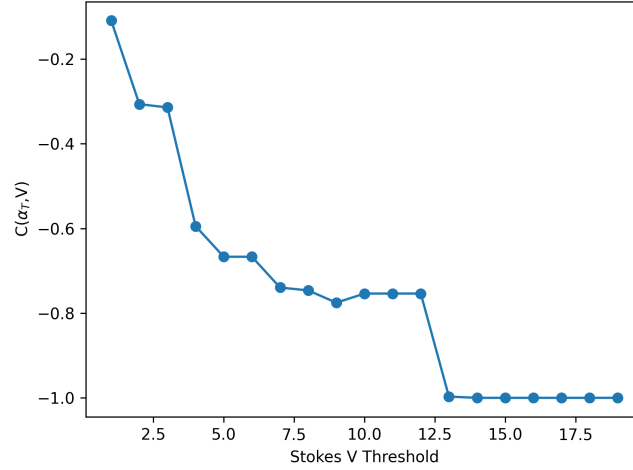


Figure 4.8: Correlation coefficient between α_T and absolute value of V Stokes for increasing polarization threshold.

to build up an operational flare forecasting system. Although for the moment we are exploiting exclusively SunDish radio images, the preliminary results from our statistical correlation analysis are promising and will be reported in the paper under preparation [173].

Moreover, through the analysis of the circular polarisation in our maps we could constrain the strength of the magnetic field directly in the chromosphere. Due to the complexity of the phenomenological scenario we provide a tentative explanation for few selected case studies.

Two acquisitions of the solar map have been also performed by Medicina Radio Telescope at two frequencies in the K-band during the Mother's Day Storm occurred in May 2024. The detailed analysis is still ongoing but we observed a very bright source in the radio spectrum and a spectral index of the AR 3664 $\alpha_T \simeq 1$ and a very high $T_p > 41000K$ few hours before the occurrence of the X3 flare.

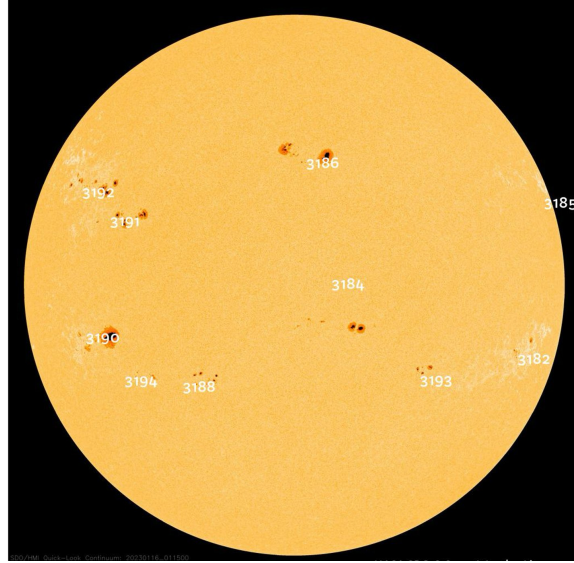


Figure 4.9: Intensitygram from Helioseismic Magnetic Imager (HMI) aboard the Solar Dynamics Observatory (SDO) of the solar photosphere, 16 January 2023. AR numbers on the solar disk are assigned by NOAA.

4.1.3 Case Studies Discussion

Among the various ARs listed in the Appendix 4.2 that have been analysed for statistical correlation analyses aimed at identifying flare precursors, we discuss more in detail few case studies.

The solar disk visible on 16 January 2023, in Figure 4.9, displays a multitude of ARs. Both the Stokes I and Stokes V parameters have been observed from Medicina Radio Telescope (Fig.4.10) and AR 3190 was found to be particularly noteworthy. Indeed, among all the data we analysed, AR 3190 in the solar map on 16 January 2023 exhibited the highest polarisation percentage $> 36\%$ and the highest peak brightness temperature $T_B > 32000K$. An eruptive event of M4.8 flare took place few hours (< 20 hours) earlier from the same region. This could lead to a change in the balance between the thermal and non-thermal emission component and this can explain the inversion from positive to negative values in the spectral index ($\alpha_T = -0.5$). Indeed, AR 3190 represents a typical example of AR classified as a β - γ - δ , therefore highly magnetically complex, and with very soft spectrum showing high circularly polarised radio emission, that is related to an M-class flare emission.

High levels of circular polarisation also indicate the presence of strong magnetic

fields, leading to gyroresonance and gyrosynchrotron emission processes. Regard-

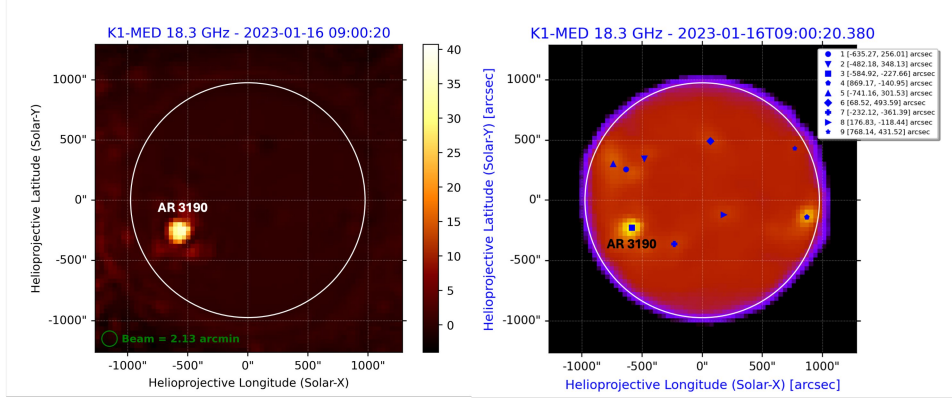


Figure 4.10: Total intensity map of the solar disk (Stokes I, right) with all the identified ARs in the K-band, and circular polarisation percentage (Stokes V, left), both obtained with the Medicina Radio Telescope on January 16th 2023 at 18.3 GHz.

ing the other ARs present on the disk, during the same observation: both AR 3192 and AR 3182 exhibit a peculiar spectrum and they produced high C-class flares only few hours later. They both have a low circular polarisation degree $< 1.5\%$, however AR 3182 is close to the west-limb therefore electromagnetic waves can be affected by additional propagation effects through a thicker solar corona layer.

AR 3184 and AR 3186, both close to the center of the disk, have a spectral index > 1.5 and did not produce significant flares. It is worth noting, from Figure 4.11, the comparison between the magnetogram on the solar photosphere and the Stokes V parameter above the active regions. We note immediately that the sense of the circular polarisation corresponds to the sign of the longitudinal component of the photospheric magnetic field. In opposite hemispheres, ARs have the opposite leading polarity, and this also corresponds to opposite circular polarisation sense for the relevant leading sunspots. The two opposite (positive/negative) magnetic field polarities, correspond to the two (right/left) circular polarisation components.

Free-free, e.g bremsstrahlung, is the principal emission mechanism for thermal plasma in the absence of gyroresonance emission, as we can assume for AR 3186 and AR 3184.

From equation 1.8 we can estimate that a 1% polarisation at $\lambda \sim 16.3\text{cm}$ requires a longitudinal component of the magnetic field of $\sim 330\text{G}$, as in case of AR 3186.

On 28 November 2020, we observed a very soft spectrum in AR 2786 and just a few hours later a cluster of C-class flares occurred. It is also possible to distinguish

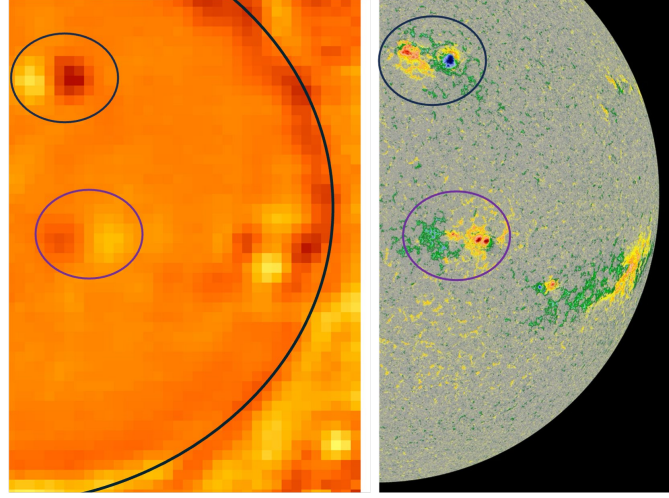


Figure 4.11: Zoom of the western hemisphere of the solar disk on 16 January 2023. Circular polarisation component (Stokes V, left) at 18.3 GHz, obtained with the Medicina Radio Telescope on and respective (right) HMI Colorized Magnetogram (SDO). AR 3186 and AR 3184 are highlighted.

another AR 2785, very close to the previous one. From the two Medicina maps, respectively at 18.3 GHz and 25.8 GHz (Fig. 4.12), we have computed the respective spectral indices: the one associated to the flaring region is $\alpha=1.37$ with a $T_p=15507$ K at 18.3 GHz, while the second one is $\alpha=1.60$, that is above our gyroresonance threshold, and the $T_p=13647$ K at 18.3 GHz. Indeed, no significant flare has been produced from AR 2785. After $\sim 11h$ from the images acquisition a C3.1 class flare

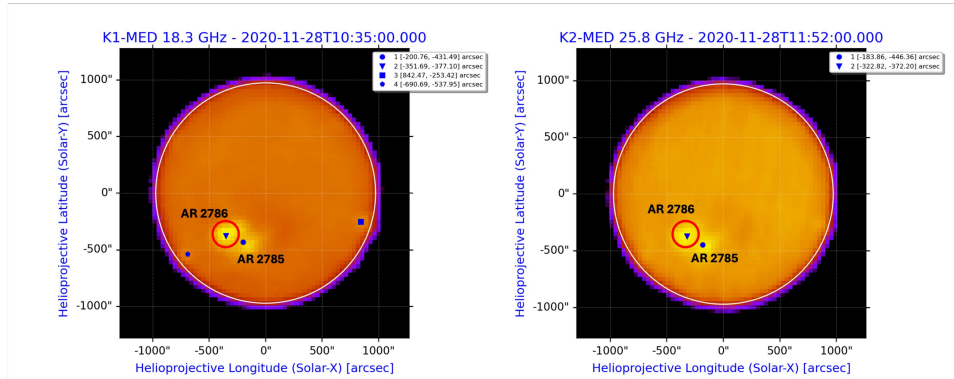


Figure 4.12: Total intensity map (Stokes I) obtained with the Medicina Radio Telescope on 28 November 2020, respectively at 18.3 (left) GHz and 25.8 GHz (right). Both AR 2786 and AR 2785 are visible and distinguishable.

was produced by AR2786. It is also worth noting that the polarisation image (Fig.

4.13) shows a Stokes V magnitude $> 8\%$ for AR 12786, while the AR 2785 is not distinguishable from the QS background, suggesting the gyro-resonance process ongoing above AR 2786, respect to the purely thermal emission of the AR 2785. The gyroresonance opacity is a complex function of several variables, including

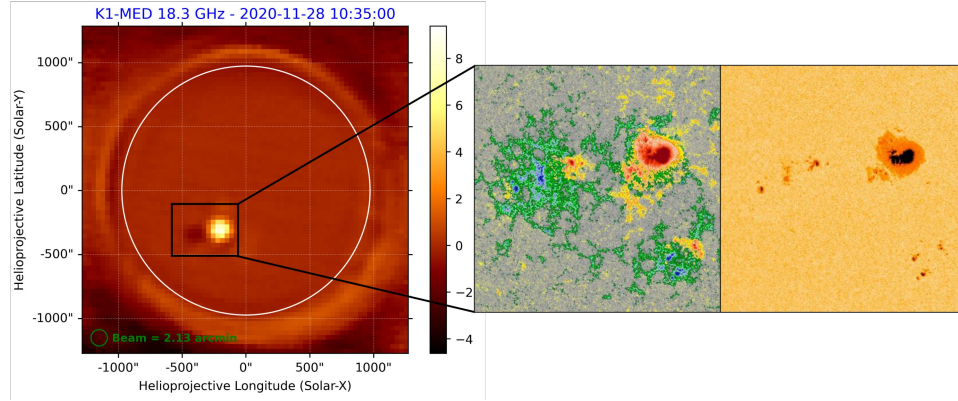


Figure 4.13: Circular Polarisation (Stokes V) map (28 November 2020) obtained by Medicina Radio Telescope at 18.3 GHz (left). Zoom of the AR 2786 HMI Intensitygram (center), with relative HMIBC Magnetogram from SDO (right).

temperature, density, magnetic field strength, and wave mode. Its value is strongly influenced by the direction of the field with respect to the line of sight, with zero opacity occurring when the fields are parallel. Emission is generated in thin layers above the base of the transition region (TR), where the magnetic field strength is relatively constant. This field strength is determined by the observing frequency, which must be equal to low harmonics of the gyrofrequency [146].

Opacity results much greater in the x-mode than in the o-mode, it is also much greater at the second harmonic than at the third; thus, under conditions prevailing in the sunspot atmosphere, the third harmonic is usually opaque in the extraordinary and transparent in the ordinary mode, while the second harmonic is opaque in both modes. If the photospheric field is weak enough, both the 3rd and the 2nd harmonic layers are below the TR and no strong sunspot-associated emission is expected. For stronger field, or lower frequency, the third harmonic enters into the TR while the second is still in the chromosphere; consequently strong emission is observed, highly polarised in the sense of the extraordinary mode [12].

Under the assumption of gyro-resonance source emission, by the use of the formula 1.9, we can derive an indicative magnitude of the chromospheric longitudinal magnetic field above the observed AR.

At 18.3 GHz we observe in Figure 4.13 strong emission from the leading spot of

AR 2786 in the extraordinary mode (LHCP) and no emission in the ordinary mode, which means that the third harmonic level ($\sim 2100G$) is already in the low TR; at the same frequency there is no o-mode emission from the leading sunspot, i.e., the second harmonic level ($\sim 3200G$) is still below the TR.

AR 3217, which is visible on the 12 February 2023 Medicinia Radio map, represents the region responsible for the most intense flare among the SunDish map series. This event occurred within 30 hours from the solar observation. During the period between 2018 and 2023, other cases of strong flares were observed (for example, an X1.8 flare from AR 3184 on 9 January 2023). However, these events were excluded from the present analysis due to the proximity of the ARs to the limb. The X1.1 produced by AR 3217 $\sim 19h$ before the observation is embedded in a cluster of M and C-class flares within our time threshold from the image acquisition. The spectral index of the magnetically complex (β - γ) AR is $\alpha_T < 1$ and the brightness temperature is $T_p > 17000K$.

From Figure 4.14 many ARs at both 18.3 and 25.8 GHz are distinguishable, as the 25 solar cycle is approaching its rising phase. Moreover, AR 3220 produced an

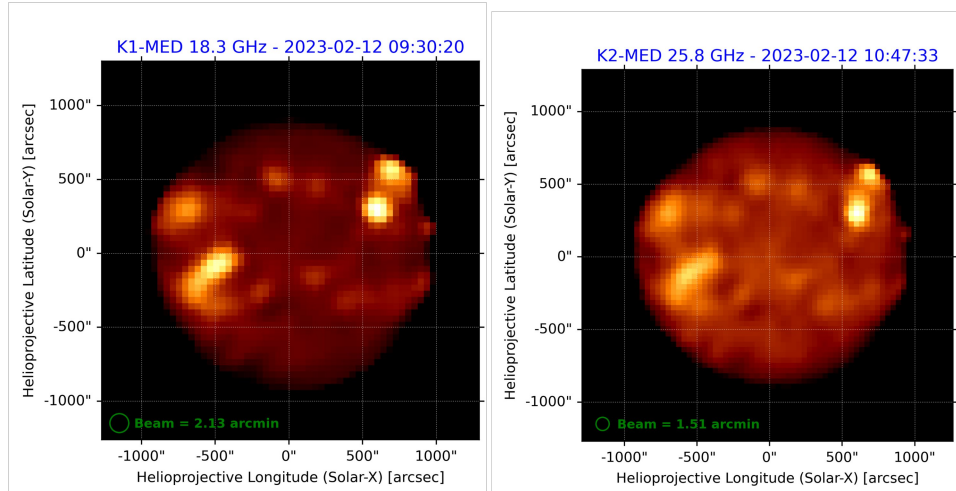


Figure 4.14: SunDish solar radio images (Medicina Radio Telescope) of the 12 February 2023, respectively at 18.3 GHz (left panel) and 25.8 GHz (right panel).

M1.0 flare and it has been classified as a true positive (TP), as its spectral index is below our gyroresonance threshold. As depicted in our maps (Fig. 4.14), in the present case, the two regions are very close and they are not distinguishable. From an operational point of view, our tool would have predicted the occurrence of both flares coming from the two separate ARs. Therefore, we associate the same

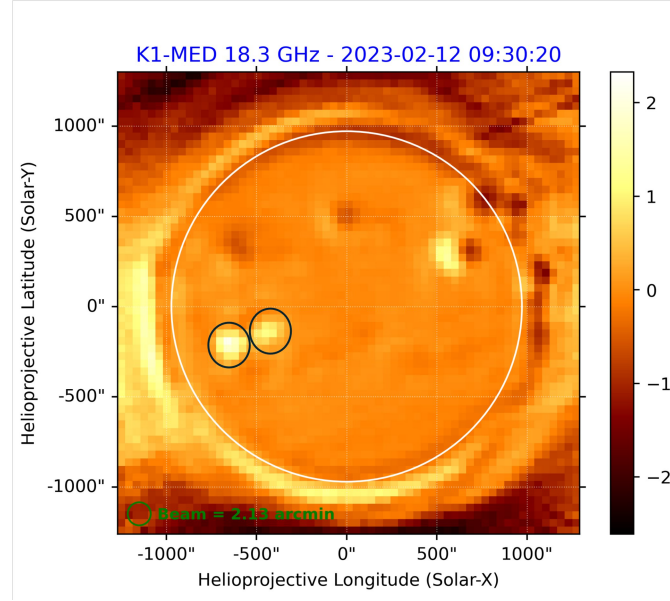


Figure 4.15: Stokes V solar radio map at 18.3 GHz by Medicina Radio Telescope, 12 February 2023. It is worth noting that AR 3217 and AR 3220, highlighted by the circles, are distinguishable in this polarisation map.

parameters to both AR 3217 and AR 3220, although we have considered them as two separated TP events (4.2). It is also worth noting that in the Stokes V image (Fig. 4.15) at 18.3 GHz AR 3217 and AR 3220 show two well separated polarization peaks: respectively $V=1.18\%$ and $V=2.17\%$.

Although it is not possible to distinguish the two ARs in the map in total intensity, thank to the polarization image we can identify two different peaks.

The SunDish project observes the Sun, on average, once per week. With this conditions, it is difficult to monitor the temporal evolution of the same ARs. However, we were able to identify the same AR both on the 19 April 2022 and on the 28 April 2022 (Fig. 4.16). Indeed, AR 2994 emerged on the 16 April and decayed on the 30 April of the same month, therefore we have been able to observe it during its rising and the declining phase.

On the 19 April 2022 the AR present a brightness peak temperature greater than 22200 K at 18.3 GHz with a spectral index of 0.98 with an extension in the photosphere of 930 millionths of a solar hemisphere (MH). After 11 and 15 hours two M-class flares emerged from the same AR. It exhibit β - γ magnetic complexity, with high negative polarization degree of -12.4% . After a week, on the 28 April 2022,

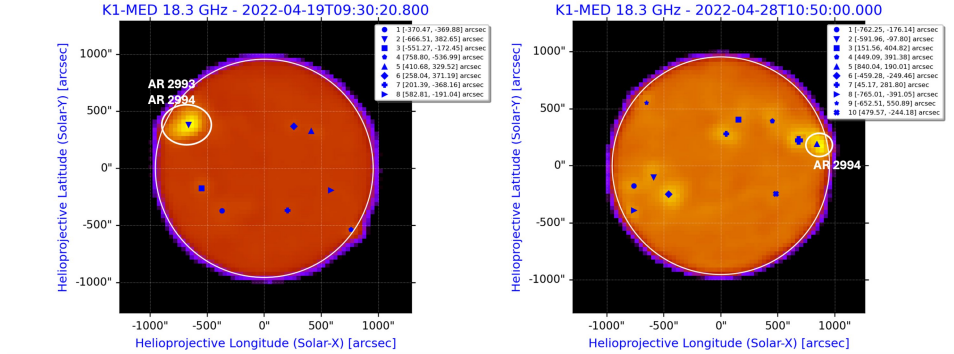


Figure 4.16: SunDish solar radio images (Medicina Radio Telescope, 18.3 GHz) of the AR 2994 respectively on the 19 April 2022 (left) and 28 April 2022 (right).

AR 2994 presents a peak temperature of ~ 15000 K at 18.3 GHz, an $\alpha = 1.38$ and an extension in the photosphere of 300 MH. The magnetic complexity decreased from β - γ to β and the polarization percentage is significantly lower and of opposite sign (+0.8%). After about 30 hours, another M-class flare emerged from the AR 2994.

All this data are consistent with the evolution of the AR: on the first observation it has a higher brightness temperature, extension and magnitude of percentage polarization, while the spectral index has a lower value. With the decay of the AR, the peak temperature dropped down and both its extension and magnitude of polarization percentage decreased, while the spectral index increased compared to the previous week. Although showing different parameters, in both cases the AR is classified as a TP event.

On the first map, AR 2294 is indistinguishable from AR 2993, indeed peak temperatures (see Fig 4.16) can not be separated. In the Table 4.2 we used only one case event to identify both ARs.

In the framework an operational tool for flare forecasting, we would have been able to forecast the upcoming flares from both AR 2993 and AR 2994, within our time delay threshold (30 hours).

Since in the corresponding polarization map they are also indistinguishable, we may be able to separate them by comparing our images with other frequencies.

By separating the ARs we obtain a set of parameters for each of them: $\alpha_T=1.04$, $T_p=20147$ K and $\alpha_T=0.96$, $T_p=22303$ K, respectively for AR 2993 and AR 2994 on the 19 April 2022.

On the 28 April 2022, AR 2993 rotated behind the solar limb and AR 2994 is the

only one visible on the limb from our map. From the comparison between the two Stokes V maps (Fig. 4.17), the geometric effect on the bipolar AR causes the polarization inversion of the limbward part of the AR 2994, as it rotates from the eastern to the western area of the solar disk. The characteristics of the observed

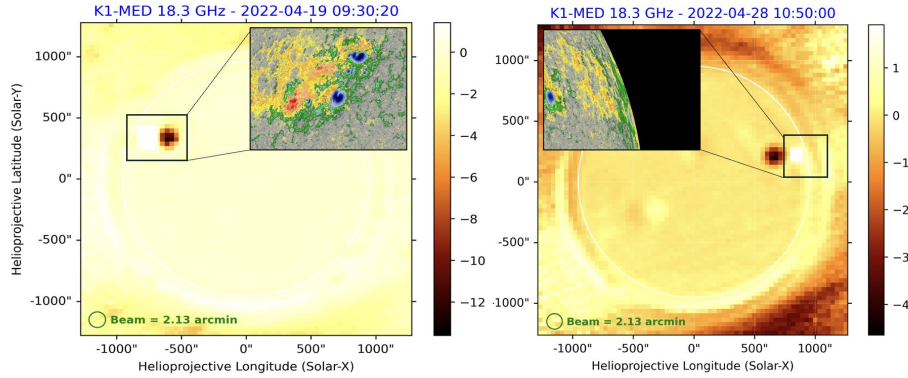


Figure 4.17: Stokes V SunDish solar radio images (Medicina Radio Telescope) respectively on the 19 April 2022 (left) and 28 April 2022 (right) at 18.3 GHz with the respective zoom of the AR magnetogram on the photosphere (SDO/HMII).

polarization of radio waves are determined both by emission mechanisms and propagation conditions in the solar corona. When the physical conditions change along the path of the wave, its polarization changes accordingly.

The Transverse Field Region (TFR) is the area where the magnetic field lines are oriented perpendicular to the line of sight from the observer's perspective. When the wave crosses a TFR, the sense of its polarization will change, since the sign of the longitudinal component of the magnetic field changes.

In case of weak coupling between the x-mode and o-mode waves, the polarization changes along the ray path, reflecting the local physical conditions, resulting in the inversion of the sense of circular polarization. When the wave modes are strongly coupled their polarization sense does not change, even if a TFR is crossed [11, 210]. Wave coupling in solar corona propagation has been widely studied in [18, 177, 40]. When the source is near the disk center the TFR is crossed high in the corona where the density and the magnetic field are low and therefore the coupling is strong. Consequently, the observed polarization is the same as the intrinsic. This is the case for the AR 2994 on the 19 April, where the leading spot shows a positive magnetic field (blue) and positive Stokes V (RHCP).

As the region moved toward the west limb, the radiation crossed the TFR at a lower

height, hence the coupling coefficient decreased. At a certain point the radiation from the east part of the source crossed the TFR under conditions of weak coupling, and the sense of its circular polarization was inverted.

Closer to the limb, as in the 28 April map, the radiation from the entire source crossed the TFR under weak coupling conditions, therefore only negative Stokes V (LHCP) can be observed [12].

In conclusion, AR radio observations will support empirical and physical models of spectral behavior that can anticipate an upcoming flare; the correlation between AR spectral index changes, due to incoming strong gyro-magnetic components in the K band, and subsequent flare occurrence in a few hours/days has been statistically investigated in the present work and it might represent a valuable forecasting probe for flare events contributing to the Space Weather warning/alert network.

Our results underline the need of continuous observation of the same active region, in order to follow in time the variations of its features that will allow us to define more reliable constraints on the correlation with solar flares. Moreover we highlight the need to increase the statistics of the analysed ARs of our study, which can be considered as a test-bed that we want to deepen with the forthcoming TSRS 2.0 images, in particular with its polarisation information.

The addition, multi-messenger correlation studies on eruptive events are needed to support this hypothesis, in particular to set constraints on the gyromagnetic occurrence above active regions. The close association of the gyroresonance emission with the magnetic field makes it a valuable tool for studying the atmospheric layers above sunspots and for magnetic field estimation. Although it was not explored in depth in this work, further dedicated studies on AR polarization, in radio K-band, appear very promising for investigating the relationship with chromospheric magnetic fields.

4.2 SOLARIS: Solar Imaging System for Continuous Monitoring and Space Weather Applications

The dynamic chromospheric activity lacks of continuous or frequent monitoring suitable for Space Weather applications. In particular, accurate measurements of the brightness temperature of the QS component are lacking and/or scattered above 30 GHz due to the difficulty to separate the pure QS component from the contribu-

tion of the ARs.

As widely discussed in the previous Section, disentanglement and quantification of the thermal (free-free and gyro-resonance) and non-thermal emission components is an open issue for radio ARs [107, 209], as well as for peculiar large-scale structures such as coronal holes, loop systems, filaments, streamers, and the coronal plateau, that would benefit from systematic long-term observations.

Solaris (a smart Solar imaging system at high radio frequency for continuous Solar monitoring and Space Weather applications) is a scientific and technological project aimed at the development of a smart Solar monitoring system at high radio frequencies based on single-dish imaging techniques. This opens for the identification and spectral analysis of Active Regions before, after and during the occurrence of solar flare.

The Solaris project [85] is a network of small solar radiotelescopes designed for operations in Antarctica and other polar/mountain sites, that will offer unique observing conditions (very low sky opacity and long Solar exposures) and unprecedented operational Solar monitoring in radio W-band.

The system features will allow Solaris to explore cutting-edge aspects of Solar Physics (e.g. chromosphere dynamic monitoring) and Space Weather applications (e.g. flare forecast).

Details of the Solaris project are illustrated in [7] and I am among the authors of the proceedings.

The project, has already been approved by PNRA (Italian National Plan for Research in Antarctica) as permanent observatory, exploiting existing on-site radiotelescopes and the first campaign in Antarctica is scheduled in February 2025.

It combines the implementation of dedicated and interchangeable high-frequency receivers on existing small single-dish radio telescope systems (1.5/2.6 m) to be adapted for Solar observations.

The scientific and technological background of Solaris relies on the innovative solar observations performed by our groups with large single-dish radiotelescopes at lower radio frequencies (SunDish) and the project mostly uses existing expertise and instrumentation yet available at the Universities and INAF laboratories of our collaboration.

Sporadic observations provided by SunDish project could be complemented and greatly empowered by continuous solar monitoring at higher frequency through a dedicated observatory.

A useful application of Solaris can be also the trigger for high-resolution follow-ups

with other facility of the Network, that are not dedicated to the monitoring of the Sun, such as Medicina or Sardinia Radio Telescopes, but can be used for specific observing sessions in order to obtain a multi-wavelengths analysis when peculiar features or particularly interesting active regions appear on the solar photosphere. The electromagnetic spectrum at mm wavelengths has been mostly unexplored until the latest years. Solar radio emission at very high frequencies is mainly due to the superposition of different synchrotron spectra or the emission from ultra-relativistic electrons. There are numerous examples of complex synchrotron spectra obtained during both large and small solar bursts showing flat or increasing fluxes with the observed frequency, suggesting a spectral maximum located above the highest observed frequency of 100 GHz [167].

Features on the Solar disk can evolve on very short timescales, thus a daily or even a more frequent monitoring is the optimal and typical choice for dedicated radio Solar facilities.

Very interesting single-dish solar solar imaging results are also obtained through regular monitoring, at least during summer, at 37 GHz by MRO [93].

An important factor in predicting powerful flares is the availability of continuous radio monitoring campaigns that provide spectral index estimates for the ARs, and detect anomalous spectral variations (e.g. sudden flattening).

Solaris can perform continuous Solar imaging observations nearly 20h/day during Antarctic summer, exploiting the observation facility OASI existing in East Antarctica at Mario Zucchelli Station, according to our simulations including overheads and accounting for local orography.

Solaris at MZS will be the only Solar facility offering continuous monitoring at 100 GHz and it will be able to collect and disseminate data in synergy with the existing national and international network of Space Weather facilities.

The main goals of the project focus both on solar physics science and operational Space Weather applications. In particular it includes accurate measurement of the brightness temperature of the quiet Sun component and diachronic monitoring of the slowly varying component around 100 GHz.

Moreover ARs flux density, spectral variability and dynamical features monitoring in the W-band will be of paramount importance, as well as solar flares detection, observations and understanding of eventual precursors.

Our automated data processing pipeline can extract from raw OTF scans the relevant scientific information in a few minutes after the observations, including FITS Solar images and measurements of AR parameters (identification, fluxes, spectra,

etc.). Raw data will be generally stored locally in portable hard disk drives at MZS for each entire summer campaign duration, while high-level scientific data can be compacted and easily exported in our dedicated online archives in Italy just after the observations. For these latter data products a daily telemetry load of 20 MB/day maximum is expected. A dedicated Solaris Observatory Archive (SOA) will contain all Solar images and their relevant scientific information. After a period of technical tests and an early science phase dedicated to checking the system performances and the definition of procedures for the scientific validation of the results, all data products will be made publicly available.

In perspective, the system could be also extended to the COCHISE facility at Concordia Station, Dome C, as well as at 78° N latitude in Norway (Ny Alesund CNR facility, Svalbard Islands), in order to offer a unique H24 solar monitoring for the whole year.

Solaris has two possible acquisition configurations: mapping OTF of the solar disk and single AR for flares detection.

The system, thank to the adaptability of the dynamic range ($10^3 - 10^6 K$) for both quiet and disturbed solar activity level, allows us to monitor precursors as well as gyro-synchrotron emission from non-thermal electrons during flares, avoiding saturation issues.

The expected duration of each OTF radio mapping covering the entire Solar disk and part of the corona ($1.5 \text{ deg} \times 1.5 \text{ deg}$ map, 5 arcmin/s scan speed, 4 scans/beam) is approximately 25 min. The expected brightness sensitivity is below 10 K at 100 GHz.

To convert the raw maps to brightness temperature images, calibration observations are required. A possibility is to compare the average raw counts in the image from the QS with literature data and published brightness information, to find a count-to-Kelvin conversion factor (self-calibration). A brightness reference can be obtained from Landi and Chiuderi Drago [103] that provides an estimate of about 7000 K for the QS level in our high frequency range.

It is possible to perform detailed measurements of the brightness temperature of the QS chromosphere and active regions, contributing to Space Weather monitoring networks and forecasts, taking advantage of observations not significantly affected by sky opacity, due to the very high solar radio brightness that can be detected with relatively small telescopes and not particularly sophisticated detectors.

A simulation of a typical Solaris image is presented in Figure 4.18.

For our scientific applications, smart single-dish radio mapping of the Solar disk is

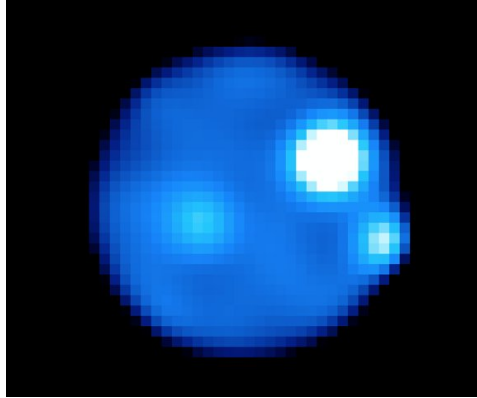


Figure 4.18: Solaris simulated image, assuming a beamsize of 4.5' BWHM, made the presence of three ARs of 1000/2000/3500 K respectively over a QS level of ~ 7000 K

more suitable than interferometric observations, especially at high-frequencies. In fact, synthesis images through interferometric networks cannot be easily obtained in the frequency range 20-100 GHz on relatively large sources (involving a variety of features at different spatial cases), with the exception of dedicated and expensive short-baseline facilities covering only up to K-band (for example the Nobeyama Radioheliograph) or above W-band (for example with ALMA) requiring non-trivial calibration and data analysis processes and performing only sporadic Solar observations.

The first step of Solaris project consists in the development of a small prototype in Milan. The 1.5 m antenna of this first test phase has been possible mostly re-adapting already existing instrumentation and the brightness profile of the solar disk has been observed in the 86.7-102.9 GHz frequency range.

The system has been integrated and carefully characterized, being now under careful validation at UNIMI in order to test the optical and mechanical systems and the autoguide pointing accuracy. The presence of the Solaris prototype in the courtyard of the University is also offering a direct and concrete educational, training and outreach perspective on solar physics, Space Weather and related technologies. On March 2024, the "first light" of Solaris Prototype was provided in the University of Milan (UNIMI, Italy) for the preliminary testing and validation phase, through transit observations of the Sun while crossing the radiotelescope beam (Fig.4.19). During the extreme "Mother's Day" event, occurred on May 10th 2024, an early test acquisition has been performed with Solaris prototype.

Although the telescope performances are still in the early test phase and the acquisition conditions in the University courtyard are not optimal, in Figure 4.20 we can

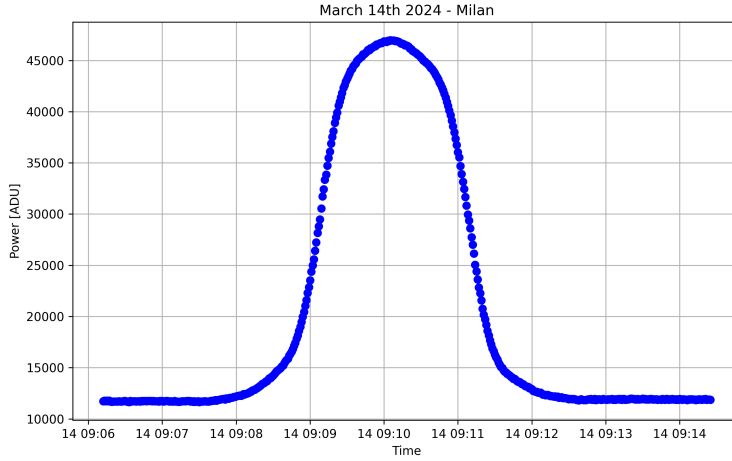


Figure 4.19: First Light Curve - Solaris Prototype 1.5m radiotelescope (University of Milan, Italy). Brightness profile of a section of the solar disk in the 86.7-102.9 GHz frequency range, obtained through a solar transit observation on 14 March 2024.

distinguish the enhanced high frequency emission from AR 3664 during the solar passage. The following step of Solaris project was the installation of Solaris Alpine, 2.6m antenna at 3500m a.s.l. on the Italian Alps ("Testa Grigia" (TG) station, Valle d'Aosta, Italy), to fully develop and test the final system with good sky opacity conditions and in an extreme mountain environment, though more accessible than remote Antarctic sites. At the time of writing, the Solaris TG station has just been installed and first light has been observed. It is now ready for scientific development, with observing campaigns starting in the coming months.

In perspective this additional station can offer mid-season solar coverage complementary to polar monitoring.

The long-term goal to have a h24 high radio-frequency solar monitoring for the whole year can be achieved through the development of a Solaris station also in the northern hemisphere. Feasibility studies for the implementation of Solaris in suitable Scandinavian/Arctic sites are ongoing. With the installation of Solaris North it will be the only facility offering radio coverage at 100 GHz for the whole year, nearly h24.

Indeed, the implementation of different Solaris stations both in the south (Antarctica) and in the north (Arctic regions) hemisphere paves the way for a global coverage, that can offer an interrupted solar monitoring.

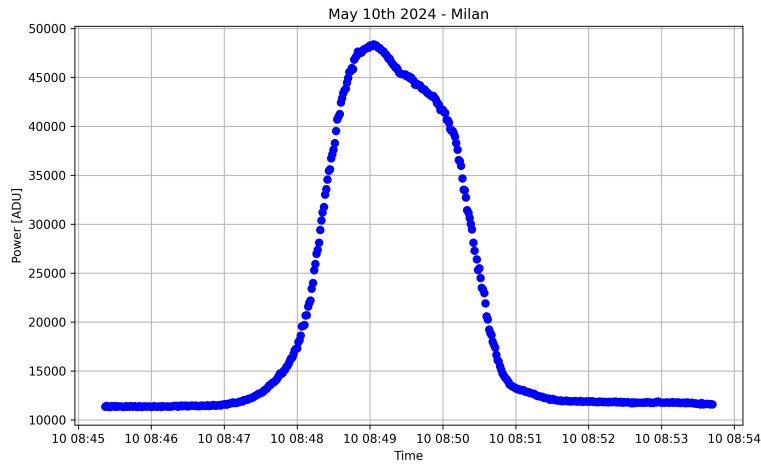


Figure 4.20: Brightness profile test of the Solaris Prototype (University of Milan, Italy), 10th May 2024. The observed bump in the brightness profile has an actual physical origin. The very powerful AR 3664 (Mother's Day Geomagnetic Storm) emission is clearly distinguishable from the background during the solar transit.

Conclusions

Solar radio data play a key role in space weather science but also in space weather operational applications, in particular with reference to GNSS effects. In this work we presented the analysis of solar radio images and indices, as a wide variety of applications and powerful tool well integrated in the perspective of the upcoming TSRS 2.0 new data, the INAF solar radio spectropolarimeter already installed in Basovizza (Trieste).

TSRS 2.0 assumes paramount importance as it bridges the critical gap in observational assets in the radio band, dedicated to furnishing polarization data and continuous monitoring of the Sun. It enables a better understanding of the physical processes involved in radio bursts but it also will be operationally exploited to send alerts to the users when specific warning thresholds on SFU are overcome, potentially affecting GNSS. Working closely with GNSS users is paramount for discerning the optimal thresholds and requirements to set.

In the present thesis we have presented the results of the F10.7 solar radio index by the use of modern univariate and multivariate LSTM-based models technique, which are still not widely used in the field of space radio weather. Previous research, such as [108], utilized EMD and LSTM hybrid models for Sunspot Number Time Series predictions. In our work, the FIF decomposition technique, employed as an alternative decomposition method with the use of LSTM and transformers, has paved the way for a reliable operational tool dedicated to solar radio data forecasting. From the comparison among different deep neural network models emerged that FIF-LSTM-MHA shows a promising accuracy respect to baseline models, although it tends to smooth the prediction curve due to the peculiarity of the Attention module, discarding less relevant features of the time series and highlighting the global trend. The complexity of the physical processes involved makes forecasting a challenging endeavour, with an increased likelihood of errors occurring during periods of high solar activity. However, our proposed model demonstrated an ability to handle information pertaining to different regimes of time scales and

CONCLUSIONS

activity levels.

The importance of reliable and continuous forecasting of radio data has been highlighted by the findings of the input ranking analysis for the TEC model forecasting [194]. Indeed, F10.7 is of significant interest as a primary driver of ionospheric disturbances affecting GNSS signal propagation, as demonstrated by our study.

From the perspective of scientific research, we focused on the case studies and statistical analysis of ARs present in SunDish images, mainly acquired with the INAF Medicina radio telescope from 2018 to 2023.

Enhanced AR magnetic fields (up to 1500-2000 Gauss) determine a spectral flattening compared to pure free-free emission due to the addition of a steeper gyro-resonance component [146]. Solar radio mapping in K-band (18-26 GHz) can efficiently probe the magnetic field and energy content of the active regions through the detection of variable gyro-resonance spectral components before and after flare events, in the chromospheric layers, where magnetic reconnection is expected to occur.

Indeed, our goal is to investigate a possible correlation between solar flares occurrence and observed parameters of active regions, such as peculiar spectral indices and circular polarization percentage (Stokes V). This study aims to understand the physical mechanism of solar flares, the presence of phenomena such as gyroresonance and free-free thermal emission.

Our research is specifically focused on the statistical study of AR spectral indices correlation, with the objective of identifying favourable conditions to intense flare emission. The increases in X-ray, UV, and EUV fluxes associated with the solar flare effect result in an increase in ionospheric density, which in turn impairs the ability to produce accurate navigation solutions. It is therefore crucial to develop a robust flare warning tool, which would be of critical importance for GNSS users.

Our correlation analysis results are extremely promising, as sporadic gyro-resonance emission has been observed, providing reliable clues on spectral softening as a precursor of flaring gyrosynchrotron emission.

Indeed, we discuss the sensitivity and robustness of this flare forecast method. When this sporadic anomalous AR spectrum is detected, the probability of a strong flare occurrence within 1-2 days is >80%, further rising at >90% also requiring that AR brightness temperature exceeds ~50% the quiet Sun level.

Although the quantitative results need further test and validation, we believe they are very promising, as no operational tool for reliable flare forecasting are currently available.

CONCLUSIONS

The integration of SunDish results into a multi-instrumental and multi-band studies will be crucial to further enrich the available information providing a more complete framework of flare emission. However, even using the single instrument, the obtained results are encouraging. Once images will be also acquired by TSRS 2.0 in both circular polarization components, it will be possible to increase the statistics and confirm the observed preliminary results.

Indeed, this work will not only contribute to basic scientific research in the field of solar activity, but may also have significant practical implications, such as in the development of tools and tailored forecasting models for predicting solar activity and, therefore, its effects on Earth.

Appendix A

Flare Peak [MJD]	Δ_T [h]	Flare Class	α_{TP}	MAPID ARID	Flag	Limb [Y/N]	T_{peak} [K]	Magn.Compl.	Stokes V [%]
59151	-12 +2h +10h	C1.5 C4.3 C1.6	1.19	SRT_201029 AR2779	TP	N	17797	β	V=-1 +0.67
59152	-13h -22h	C1.6 C4.3	1.45	MED_201030 AR2779	TP	N	13175	β	V=-0.4
59159	-28h -24h -10h -9h -8h +1h +6h +21h	C1.7 C2.3 C2.8 C1.7 C2.0 C1.3 C1.8 C1.8	1.11	MED_201106 AR2781	TP	N	19203	$\beta - \gamma$	V=1.4
59164	-16h	C1.3	1.53	MED_201111 AR2781	FN	N	13570	β	V=-0.6 V=+0.5
59176	-18h +11h +12h	C3.3 C1.4 C4.4	1.15	MED_201123 AR2785	TP	N	12227	β	V=4.3
59181	-19h -10h +2h +11h +24h	C1.5 C1.5 C2.9 C3.1 C1.6	1.37	MED_201128 AR2786	TP	N	15007	$\beta - \gamma$	V=8.52
59323	+14 h	M1.1	1.46	MED_210419 AR2816	TP	N	13297	β	V=-0.7
59360	+10 +11 +12 h	C2.9 C3.3 C3.3	1.13	MED_210526 AR2826	TP	N	18187	β	V=-6.3
59388	-2h	C3.4	1.5	MED_210623 AR2833	TP	N	11616	α	V=-0.5
59395	+8h	C2.8	1.1	MED_210630 AR2835	TP	N	15429	$\beta - \gamma - \delta$	V=2.18
59456	-24h	C6.0	2.04	MED_210830 AR2866	FN	N	15139	$\beta - \gamma$ V=-0.7+0.5	
59565	-29h -24 h	C4.5 C6.9	1.11	MED_211216 AR2907	TP	N	17144.5	$\beta - \gamma$	V=+0.68-0.6
59572	-8h -3h	C5.2 C5.6	1.49	MED_211224 AR2907	TP	N	16573	$\beta - \gamma$	V=-0.47
59572	-4h -4h +0.5h +2h	C5.4 C4.3 C4.0 C5.3	1.49	MED_211224 AR2908	TP	Y	16573	$\beta - \gamma$	V=-0.47
59572	-7h	C7.5	1.62	MED_211224 AR2917	FN	N	12845	β	V=-0.40
59610	+21 h	C6.0	1.36	MED_220131 AR2940	TP	Y	15509	β	V=-1.2
59610	-	-	1.42	MED_220131 AR2939	FP	Y	12642	β	V=1.49
59610	-	-	1.32	MED_220131 AR2936	FP	N	15049	β	V=-7.44
59619	-	-	1.46	MED_220209 AR2941	FP	N	14493	β	V=+0.8-0.8
59619	-	-	1.46	MED_220209 AR2940	FP	N	13720	β	V=+0.5-0.4
59638	-	-	0.93	MED_220228 AR2957	FP	Y	16052	β	V=-0.17
59653	-25h + 2h + 12 h	M1.7 M1.3 M1.4	1.47	MED_220315 AR2965	TP	N	12799.9	$\beta - \gamma$	V=-+1.01
59666	-4h +2h +2h +4h +4h +5h +10h +11h	M4 C9.7 C9.7 C6 M1 M1.1 M2.2 M1.1	0.88	MED_220328 AR2975	TP	N	20118	$\beta - \gamma$	V=-1.2

APPENDIX A

Flare Peak [MJD]	Δ_T [h]	Flare Class	α_{TP}	MAPID ARID	Flag	Limb [Y/N]	T_{peak} [K]	Magn.Compl.	Stokes V
59666			1.38	MED_220328 AR2976	FP	N	13813	$\beta - \delta$	V=+1.83
59688	-27h -24h -14h -6h +26h	M1.3 M1.2 C7.2 M1.1 M1.9	0.98	MED_220419 AR2993	TP	N	22262.5	$\beta - \gamma$	V=-12.4
59688	+11h +15h	M1.6 M1.3	0.98	MED_220419 AR2994	TP	N	22262.5	$\beta - \gamma$	V=-12.4
59697	+30 h	M1.2	1.38	MED_220428 AR2994	TP	N	14983	β	V=+0.8
59708	-	-	1.45	MED_220509 AR3004	FP	N	13312.8	α	V=-0.66
59708	+27 h	X1.0	1.42	MED_220509 AR3006	TP	N	14965	$\beta - \delta$	V=-0.89+0.79
59708	-15 h	C7.9	1.58	MED_220509 AR3007	FN	Y	13620.9	β	V=0.24
59715	-	-	1.47	MED_220516 AR3014	FP	Y	15563.6	$\beta - \delta$	V=+0.4
59715	+3 h	M2.4	1.30	MED_220516 AR3017	TP	Y	15044.2	$\beta - \gamma$	V=0.3
59715	+25 h	C9.8	1.51	MED_220516 AR3010	FN	N	14696.8	β	V=+0.77
59878	-	-	1.47	MED_221026 AR3130	FP	N	13833	β	V=+0.57
59895	-27h -24h -10h	M1.2 M1.1 M1.1	1.39	MED_221112 AR3141	TP	N	15057	$\beta - \gamma$	V=-1.46 +0.97
59940	-11 h -2 h	M2.0 M1.1	1.26	MED_221227 AR3176	TP	N	16407	β	V=-1.5
59940	+6 h	M1.2	1.14	MED_221227 AR3169	TP	N	16680	$\beta - \gamma$	V=-1.6
59946	-4h	C8.7	1.54	MED_230102 AR3176	FN	N	13004	β	V=-0.6 +0.5
59946	-	-	1.46	MED_230102 AR3180	FP	N	15307.1	β	V=+0.8
59953	-2h + 2h + 15 h	M2.1 C9.5 M2.7	1.01	MED_230109 AR3181	TP	N	19494.6	$\beta - \gamma - \delta$	V=1.29 -0.96
59953	-30 h	C9.0	1.1	MED_230109 AR3182	TP	N	17597.6	$\beta - \gamma - \delta$	V= 1.23
59953	-29h -26h -25h -19h -15h -15h +7h +16h +24h +30h	C8.0 M1.2 M1.4 M1.4 C9.6 M1.1 X1.8 M1.0 C9.7 M1.0	0.9	MED_230109 AR3184	TP	Y	14794.3	α	V=3
59960	+14h	C9.3	1.16	MED_230116 AR3192	TP	N	16226.1	$\beta - \gamma$	V=0.5
59960	-17h	M4.8	-0.51	MED_230116 AR3190	TP	N	32086	$\beta - \gamma - \delta$	V=36.99
59960	-1h	C8.7	-0.07	MED_230116 AR3182	TP	Y	24421.1	β	V=1.18
59960	-30h	M6.0	1.27	MED_230116 AR3191	TP	N	15075	β	V=-0.5
59960	-	-	1.23	MED_230116 AR3194	FP	N	14608	β	V=8.17
59987	-22h +3h	M1.4 M1.2	1.09	MED_230212 AR3222	TP	Y	16825	β	V=-1.2
59987	-24h	M1.0	0.95	MED_230212 AR3220	TP	N	17303	$\beta - \gamma$	V=2.17
59987	-22h -21h -19h -5h -2h -1h +5h	M1.5 M1.1 X1.1 C9.1 M2.9 M1.4 M1.0	0.95	MED_230212 AR3217	TP	N	17303	$\beta - \gamma$	V=1.18
59987	+19h +29h	M1.0 M1.4	1.5	MED_230212 AR3226	TP	N	13512	$\beta - \gamma$	V=-0.59
59987	-6h	C8.2	1.05	MED_230212 AR3214	TP	N	19773	β	V=+1.4 -0.9
60108	-	-	1.49	MED_230613 AR3326	FP	N	13215	α	V=0.3
60108	-	-	1.48	MED_230613 AR3327	FP	N	15211	α	V=+0.5 -0.4
60165	-16h	C8.7	1.53	MED_230809 AR3394	FN	N	14301	$\beta - \gamma$	V=3.14-0.3
60165	-25 h	M3.6	1.45	MED_230809 AR3387	TP	N	12073	β	V=0.2
60258	+29h	M1.2	1.18	MED_231110 AR3477	TP	N	14705	β	V=19.9
60259	+6 h	M1.2	1.71	MED_231111 AR3477	FN	N	12773	β	V=13.4
60276	+ 4h + 5h	M3.48 M9.82	1.82	MED_231128 AR3500	FN	N	12655	$\beta - \gamma$	V=13
60290	-5h +4h +29h	C9.1 C8.2 C9.3	1.48	MED_231212 AR3511	TP	Y	11207.1	β	V=3.4
60290	-6h +4 h	C9.1 C8.2	1.73	MED_231212 AR3514	FN	N	12705	β	V=9.9
60306	+21h	C9.7	1.63	MED_231228 AR3533	FN	N	14355.8	β	V=13.2

Bibliography

- [1] CelesTrak. <https://celestrak.org/SpaceData/>.
- [2] POAM Electronics. <https://www.poamelectronics.com/>.
- [3] Space Weather Prediction Center (SWPC). <https://www.swpc.noaa.gov/>.
- [4] SSA Space Weather Activities. <https://swe.ssa.esa.int/ssa-space-weather-activities>.
- [5] LSTM Networks. Available online, 2023. Accessed on 10 May 2023.
- [6] Solar cycle progression. Available online, 2023. Accessed on 20 May 2023.
- [7] Pellizzoni A., Cavaliere F. Potenza M., Villa F., Boria E., Buttu M., Bruni I., Cuttaia F., De Petris M., Della Torre S., Egron E., Iacolina M.N., Limonta A., and Marcucci A. et al. The solaris observatory: a smart solar imaging system at high radio frequency for continuous solar monitoring and space weather applications. In *SPIE*, 2024. in press.
- [8] Vecchio A., D'Alessi L., Carbone V., Laurenza M., and Storini M. The empirical mode decomposition to study the quasi-biennial modulation of solar magnetic activity and solar neutrino flux. *Advances in Adaptive Data Analysis*, 04(01n02):1250014, 2012.
- [9] M. A. Abdu. Equatorial ionosphere–thermosphere system: Electrodynamics and irregularities. *Advances in Space Research*, 35:771–787, 2005.
- [10] T. Alberti, G. Consolini, V. Carbone, E. Yordanova, M.F. Marcucci, and P. De Michelis. Multifractal and chaotic properties of solar wind at mhd and kinetic domains: An empirical mode decomposition approach. *Entropy*, 21(3):320, 2019.
- [11] C. E. Alissandrakis. Magnetic field diagnostics from microwave circular polarization inversion. In B. Schmieder, A. Hofmann, and J. Staude, editors, *Third Advances in Solar Physics Euroconference: Magnetic Fields and Oscillations*, volume 184 of *ASP Conference Series*, pages 23–27. Astronomical Society of the Pacific, 1999.
- [12] C. E. Alissandrakis and Dale E. Gary. Radio measurements of the magnetic field in the solar chromosphere and the corona. *Frontiers in Astronomy and Space Sciences*, 7, 2021.

BIBLIOGRAPHY

- [13] A. Altmann, L. Tolosi, O. Sander, and et al. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347, 2010.
- [14] C. N. Arge, C. J. Henney, J. Koller, C. Compeau, and et al. Air force data assimilative photospheric flux transport (adapt) model. *AIP Conference Proceedings*, 1216(1):343–346, 2010.
- [15] E. N. Asamoah, M. Cafaro, I. Epicoco, et al. Physics-informed loss functions for vertical total electron content forecast. *Earth Science Informatics*, 2024.
- [16] M. Awad and R. Khanna. Support vector regression. In *Efficient Learning Machines*. Apress, Berkeley, CA, 2015.
- [17] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations, ICLR*, 2015.
- [18] R. Bandiera. Diagnostic of coronal magnetic fields from microwave polarization reversal. *Astronomy and Astrophysics*, 112(1):52–60, August 1982.
- [19] T. S. Bastian, A. O. Benz, and D. E. Gary. Radio emission from solar flares. *Annual Review of Astronomy and Astrophysics*, 36:131–188, 1998.
- [20] T. S. Bastian, A. O. Benz, and D. E. Gary. Radio emission from the sun and stars. *Physics Today*, 51(11):44–51, 1998.
- [21] Jason M. H. Beedle, Christopher E. Rura, David G. Simpson, Hale I. Cohen, and et al. A user’s guide to the magnetically connected space weather system: A brief review. *Frontiers in Astronomy and Space Sciences*, 8, 2022.
- [22] A. O Benz and Säm Krucker. Observations of solar flares: a review. *Space Science Reviews*, 101(1-2):153–196, 2002.
- [23] Arnold O Benz. Solar radio emission. *Living Reviews in Solar Physics*, 14(1):2, 2017.
- [24] S Bertocco, D Goz, L Tornatore, A Ragagnin, G Maggio, F Gasparo, C Vuerli, G Taffoni, and M Molinaro. Inaf trieste astronomical observatory information technology framework, 2019.
- [25] D.A. Biesecker and L. Upton. Solar cycle 25 consensus prediction update, 2019.
- [26] D Bilitza, LA McKinnell, B Reinisch, and T Fuller-Rowell. The international reference ionosphere today and in the future. *Journal of Geodesy*, 85:909–920, 11 2011.
- [27] B. Bowman and W.K. Tobiska. Jb2008: Empirical thermospheric density model. Astrophysics Source Code Library, record ascl:2007.021, July 2020.
- [28] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

BIBLIOGRAPHY

- [29] P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer science & business media, 2nd edition, 1991.
- [30] R. Cameron and M. Schüssler. Solar Cycle Prediction Using Precursors and Flux Transport Models. *Astrophysical Journal*, 659(1):801–811, April 2007.
- [31] E. Camporeale. The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather*, 17:1166–1207, 2019.
- [32] H V Cane, R G Stone, and D V Reames. Coronal mass ejections, type ii radio bursts, and solar energetic particle events. *Astrophysical Journal*, 303:624–629, 1986.
- [33] C. Cesaroni, L. Spogli, A. Aragon-Angel, M. Fiocca, and et al. Neural network based model for global total electron content forecasting. *J. Space Weather Space Clim.*, 10:11, 2020.
- [34] H. Cheng, PN. Tan, J. Gao, and J. Scripps. Multistep-ahead time series prediction. In WK. Ng, M. Kitsuregawa, J. Li, and K. Chang, editors, *Advances in Knowledge Discovery and Data Mining*, volume 3918. Springer, 2006.
- [35] A. Cicone. Iterative filtering as a direct method for the decomposition of nonstationary signals. *Numer Algor*, 85:811–827, 2020.
- [36] A. Cicone, J. Liu, and H. Zhou. Adaptive local iterative filtering for signal decomposition and instantaneous frequency analysis. *Applied and Computational Harmonic Analysis*, 41(2):384–411, 2016.
- [37] A. Cicone and H. Zhou. Fast iterative filtering: a new, fast and robust decomposition method for nonlinear and nonstationary processes. *EGU General Assembly 2021*, 2021.
- [38] A. Cicone and H. Zhou. Numerical analysis for iterative filtering with new efficient implementations based on fft. *Numer. Math.*, 147:1–28, 2021.
- [39] Civil Aviation Authority. Impacts of space weather on aviation. Technical Report CAP 1428, Aviation House, Gatwick Airport South, 2016.
- [40] M. H. Cohen. Magnetoionic mode coupling at high frequencies. *The Astrophysical Journal*, 131, 1960.
- [41] L.H. Deng, B. Li, Y.F. Zheng, and X.M. Cheng. Relative phase analyses of 10.7cm solar radio flux with sunspot numbers. *New Astron.*, 23–24:1–5, 2013.
- [42] M.W. Devlin, J. and Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

BIBLIOGRAPHY

- [43] S. L. Discola Junior. *Enhancing Solar Flare Forecasting: a Multi-Class and Multi-Label Classification Approach to Handle Imbalanced Time Series*. PhD thesis, Federal University of São Carlos, São Carlos, 2019.
- [44] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, and et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.
- [45] Y. Du. Forecasting the daily 10.7 cm solar radio flux using an autoregressive model. *Sol. Phys.*, 295:1–23, 2020.
- [46] Z.L. Du. The solar cycle: a modified gaussian function for fitting the shape of the solar cycle and predicting cycle 25. *Astrophys. Space Sci.*, 367:20, 2022.
- [47] L. Duband. Space cryocooler developments. *Physics Procedia*, 67:1–10, 12 2015.
- [48] T. Dudok de Wit and S. et al. Bruinsma. Synoptic radio observations as proxies for upper atmosphere modelling. *J. Space Weather Space Clim.*, 4:A06, 2014.
- [49] George A Dulk. Radio emission from the sun and stars. *Annual Review of Astronomy and Astrophysics*, 23:169–224, 1985.
- [50] S. Erard, B. Cecconi, P. Le Sidaner, M. Demleitner, and M. Taylor. Epn-tap: Publishing solar system data to the virtual observatory version 2.0. IVOA Recommendation, August 2022.
- [51] Asamoah E.N et al. A stacked machine learning model for the vertical total electron content forecasting. In *Proceedings of the URSI 2023 Conference*, Sapporo, Japan, August 2023. International Union of Radio Science (URSI).
- [52] Ya-Hui Yang et al. A statistical study of flare productivity associated with sunspot properties in different magnetic types of active regions. *ApJ*, 834, 2017.
- [53] L. Fletcher, B. R. Dennis, H. S. Hudson, et al. An observational overview of solar flares. *Space Science Reviews*, 159:19, 2011.
- [54] Laboratory for Atmospheric and Space Physics. Lasp interactive solar irradiance datacenter. <https://doi.org/10.25980/L27Z-XD34>, 2005.
- [55] Peter T. Gallagher, Yong-Jae Moon, and Haimin Wang. Active-region monitoring and flare forecasting – i. data processing and first results. *Solar Physics*, 209:171–183, 2002.
- [56] M. K. Georgoulis, D. Shaun Bloomfield, M. Piana, A.M. Massone, M. Soldati, and et al. The flare likelihood and region eruption forecasting (flarecast) project: flare forecasting in the big data & machine learning era. *J. Space Weather Space Clim.*, 11:39, 2021.
- [57] Felix A. Gers, J. Schmidhuber, and F. Cummins. Learning to forget: Continual prediction with lstm. *Neural Computation*, 12(10):2451–2471, 2000.

BIBLIOGRAPHY

- [58] Goddard Space Flight Center Space Physics Data Facility. SPDF - OMNIWeb Service. <https://omniweb.gsfc.nasa.gov/ow.html>, 2023. 2022/2023.
- [59] N. Gopalswamy. Solar eruptions and their space weather effects. *Space Science Reviews*, 124(1-4):145–168, 2006.
- [60] N. Gopalswamy. Proc. of ursi ap-rasc, seoul, 1075. In *2016 URSI Asia-Pacific Radio Science Conference (URSI AP-RASC)*, 2016.
- [61] Nat Gopalswamy. Coronal mass ejections: An overview. In *Coronal Mass Ejections*, pages 37–86. Springer, 2006.
- [62] M. Graham and et al. Vospace version 2.1. IVOA Recommendation, June 2018.
- [63] L. Green, B. Torok, T. and Vrsnak, Ward IV, and Astrid Veronig. The origin, early evolution and predictability of solar eruptions. *Space Science Reviews*, 214, 02 2018.
- [64] G. E. Hale, F. Ellerman, S. B. Nicholson, and A. H. Joy. The magnetic polarity of sun-spots. *The Astrophysical Journal*, 49:153, 1919.
- [65] M. Hapgood. Chapter 1 - linking space weather science to impacts—the view from the earth. In Natalia Buzulukova, editor, *Extreme Events in Geospace*, pages 3–34. Elsevier, 2018.
- [66] J. Harvey. An evolving synoptic magnetic flux map and implications for the distribution of photospheric magnetic flux. *Solar Physics*, 195, 2000.
- [67] David H. Hathaway, Robert M. Wilson, and Edwin J. Reichmann. A synthesis of solar cycle prediction techniques. *Journal of Geophysical Research*, 104(A10):22375–22388, October 1999.
- [68] D.H. Hathaway, R.M. Wilson, and E.J. Reichmann. The shape of the sunspot cycle. *Sol. Phys.*, 151:177–190, 1994.
- [69] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.
- [70] Yuanbo He, Yunfei Yang, Xianyong Bai, and et al. Research on mount wilson magnetic classification based on deep learning. *Advances in Astronomy*, 2021:5529383, 2021.
- [71] Win Zaw Hein, Yoshitaka Goto, and Yoshiya Kasahara. Estimation method of ionospheric tec distribution using single frequency measurements of gps signals. *International Journal of Advanced Computer Science and Applications*, 7(12), 2016.
- [72] C.J. Henney, W.A. Toussaint, S.M. White, and C.N. Arge. Forecasting f10.7 with solar magnetic flux transport modeling. space weather. *Space Weather*, 10(2), 2012.
- [73] C.J. et al. Henney. Forecasting solar extreme and far ultraviolet irradiance. *Space Weather*, 13:141–153, 2015.

BIBLIOGRAPHY

- [74] M. Hernández-Pajares, J. M. Juan, J. Sanz, and R. Orús. The igs vtec maps: A reliable source of ionospheric information since 1998. *Journal of Geodesy*, 83(3-4):263–275, 2009.
- [75] Yashan Hettiarachchi, Janaka Adassuriya, Chandana Jayaratne, Sasani Jayawardhana, and Christian Monstein. The analysis of type ii and type iii solar radio bursts: Gui for the e-callisto data. *New Astronomy*, 109:102194, 2024.
- [76] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [77] CLS Homepage. Cls homepage. <https://spaceweather.cls.fr>, Last updated 25 July 2023.
- [78] LASP Homepage. Lasp homepage. <https://lasp.colorado.edu/lisird/>, Last accessed 25 July 2023.
- [79] SILSO Homepage. Silso homepage. <https://www.sidc.be/silso/home>, Last updated 25 July 2023.
- [80] Cong Huang, Dan-Dan Liu, and Jing-Song Wang. Forecast daily indices of solar activity, f10.7, using support vector regression method. *Research in Astronomy and Astrophysics*, 9(6):694–702, may 2009.
- [81] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3):489–501, 2006.
- [82] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, and et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998.
- [83] Norden E. Huang and Zheng et al. Shen. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A.*, 454:903–995, 1998.
- [84] R. D. Hunsucker. Atmospheric gravity waves generated in the high-latitude ionosphere: A review. *Rev. Geophys.*, 20(2):293–315, 1982.
- [85] INAF. Solaris project website. <https://sites.google.com/inaf.it/solaris>, 2024.
- [86] N. Jakowski, C. Mayer, V. Wilken, and M. Hoque. *Ionospheric impact on GNSS signals*, volume 20, pages 11–25. 01 2008.
- [87] M. Janvier, G. Aulanier, and P. Démoulin. From coronal observations to mhd simulations, the building blocks for 3d models of solar flares (invited review). *Solar Phys.*, 290:3425–3456, 2015.

BIBLIOGRAPHY

- [88] G. Jerse, V. Alberti, M. Molinaro, V. Baldini, R. Cirami, I. Coretti, P. Dimarcantonio, and M. Messerotti. Solar radio diagnostic for space weather with the trieste solar radio system 2.0. In *2020 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE)*, pages 60–65, 2020.
- [89] G. Jerse and A. Marcucci. Deep learning lstm-based approaches for 10.7 cm solar radio flux forecasting up to 45-days. *Astronomy and Computing*, 46:100786, 2024.
- [90] S. Jin, Q. Wang, and G. Dardanelli. A review on multi-gnss for earth observation and emerging applications. *Remote Sensing*, 14(16), 2022.
- [91] T. Kakinuma and G. Swarup. A model for the sources of the slowly varying component of microwave solar radiation. *Astrophys. J.*, 136:975, 1962.
- [92] J. Kallunki. The prediction of solar flares using millimeter radio brightenings. *Latvian Journal of Physics and Technical Sciences*, 60(2):43–51, 2023.
- [93] J. Kallunki, M. Tornikoski, E. Oinaskallio, and et al. Solar observing system for radio frequencies 5–120 mhz. *Astronomische Nachrichten*, 339:656–660, 2018.
- [94] Y. Kamide and A. Chian. Solar radio emissions. In Y. Kamide and A. Chian, editors, *Handbook of the Solar-Terrestrial Environment*. Springer, Berlin, Heidelberg, 2007.
- [95] R.P. Kane. Size of the coming solar cycle 24 based on Ohl’s Precursor Method, final estimate. *Annales Geophysicae*, 28(7):1463–1466, July 2010.
- [96] M. Karagiannopoulos and D. et al. Anyfantis. Feature selection for regression problems. 2007.
- [97] M. C. Kelley. *The Earth’s Ionosphere: Plasma Physics and Electrodynamics*. Academic Press, 2009.
- [98] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR*, 2015.
- [99] P. M. Kintner, B. M. Ledvina, and E. R. de Paula. Gps and ionospheric scintillations. *Space Weather*, 5(9), 2007.
- [100] K. L. Klein and et al. Solar radio bursts as a tool for space weather forecasting. *Comptes Rendus Physique*, 19(1):36–42, 2018.
- [101] A. Krankowski, B. Zieba, L. Tomasik, and J. Krypiak-Gregorczyk. Space weather effects on gnss. *GPS Solutions*, 24(3):1–15, 2020.
- [102] M. R. Kundu. *Solar Radio Astronomy*. Wiley, 1965.
- [103] E. Landi and F. Chiuderi Drago. Radiative recombination rate coefficients for astrophysical and laboratory plasmas. *Apj*, 675:1629, 2008.
- [104] X. Lang and et al. Fast multivariate empirical mode decomposition. *IEEE Access*, 6:65521–65538, 2018.

BIBLIOGRAPHY

- [105] J. L. Lean, J. M. Picone, and J. T. Emmert. Quantitative forecasting of near-term solar activity and upper atmospheric density. *Journal of Geophysical Research: Space Physics*, 114(A7), 2009.
- [106] Y. LeCun, Y. Bengio and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [107] J. Lee. Radio emissions from solar active regions. *Space Sci. Rev.*, 133(1-4):73–102, December 2007.
- [108] T. Lee. Emd and lstm hybrid deep learning model for predicting sunspot number time series with a cyclic pattern. *Solar Physics*, 295(82), 2020.
- [109] Qiang Li, Jian Wu, Zheng wen Xu, and Jun Wu. Extraction of the periodic components of solar activity with the emd method. *Chinese Astronomy and Astrophysics*, 31(3):261–269, 2007.
- [110] B. Lim and S. Zohren. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194):20200209, feb 2021.
- [111] L. Lin, Y. Wang, and H. Zhou. Iterative filtering as an alternative algorithm for empirical mode decomposition. *Advances in Adaptive Data Analysis*, 1(4):543–560, 2009.
- [112] R P Lin. Radio bursts associated with solar flares. *Space Science Reviews*, 43(1-2):289–355, 1986.
- [113] N. Linty, A. Farasin, A. Favenza, and F. Dovis. Detection of gnss ionospheric scintillations based on machine learning decision tree. *IEEE Transactions on Aerospace and Electronic Systems*, 55:303–317, 2019.
- [114] C. Liu, X. Zhao, T. Chen, and H. Li. Predicting short-term f10.7 with transport models. *Astrophys. Space Sci.*, 363(266), 2018.
- [115] J. Luo, L. Zhu, H. Zhu, Wei C., and J. Liang. A new approach for the 10.7-cm solar radio flux forecasting: Based on empirical mode decomposition and lstm. *International Journal of Computational Intelligence Systems*, 14, 2021.
- [116] J. Luo and L. et al. Zhu. Forecasting the 10.7-cm solar radio flux using deep cnn-lstm neural networks. *Processes*, 10(2):262, 2022.
- [117] J. Luo and L. et al. Zhu. Forecasting the 10.7-cm solar radio flux using deep cnn-lstm neural networks. *Processes*, 10(2):262, 2022.
- [118] Messerotti M, Jerse G., Alberti V., and Marcucci A. The trieste solar activity index. In *Proceedings of the 4th URSI AT-RASC*, Gran Canaria, May 2024. in press.

BIBLIOGRAPHY

- [119] Molinaro M., Jerse G., Alberti V., Bertocco S., and Marcucci A. et al. Modernising discovery, access and interoperability for the trieste solar radio system heritage archive. In *SPIE*, 2024.
- [120] Mukulika M., Sanjay K., S. Banola, and A.K. Singh. Occurrence of ionospheric scintillations under different solar and geomagnetic conditions over low latitude station varanasi. *Advances in Space Research*, 73:3658–3674, 2024.
- [121] G. A. MacLachlan, A. Shenoy, E. Sonbas, R. Coyne, K. S. Dhuga, and et al. The hurst exponent of fermi gamma-ray bursts. *Monthly Notices of the Royal Astronomical Society*, 436(4):2907–2914, 2013.
- [122] J. Magdalenic and at al. Classification and properties of supershort solar radio bursts. *ApJ*, 642:L77, 2006.
- [123] Rumana Majid and S. Mir. Advances in statistical forecasting methods: An overview. *Economic Affairs*, 63:815–831, 2018.
- [124] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal. Long short term memory networks for anomaly detection in time series. In *ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges (Belgium), April 2015.
- [125] A. Marcucci, G. Jerse, V. Alberti, and M. Messerotti. A deep learning model based on multi-head attention for long-term forecasting of solar activity. *Engineering Proceedings*, 39(1):16, 2023.
- [126] M. Markovic. Determination of total electron content in the ionosphere using gps technology. *Geonauka*, 02:1–9, 02 2015.
- [127] M. Marongiu, A. Pellizzoni, M. Bachetti, S. Mulas, S. Righini, G. Murtas, S. Loru, E. Egron, and M.N. Iacolina. A dedicated pipeline to analyse solar data with inaf radio telescopes: Sunpit (sundish pipeline tool). Technical Report 137, OA@INAF, 2022.
- [128] M. Marongiu, A. Pellizzoni, E. Egron, and et al. Methods for detection and analysis of weak radio sources with single-dish radio telescopes. *Experimental Astronomy*, 49:159–182, 2020.
- [129] M. Marongiu, A. Pellizzoni, S. Mulas, and G. Murtas. A python approach for solar data analysis: Sundara (sundish active region analyser), preliminary development. 2021.
- [130] S.R. McKee, P.J Cilliers, S. Lotz, and C. Monstein. The effects of solar radio bursts on frequency bands utilised by the aviation industry in sub-saharan africa. *J. Space Weather Space Clim.*, 13:4, 2023.
- [131] D B Melrose. Plasma physics: Solar radio emission. *Plasma Physics*, 27(10):1067–1079, 1985.

BIBLIOGRAPHY

- [132] K. Merz, B. B. Virgili, and V. et al. Braun. Current collision avoidance service by esa's space debris office. *Proc. 7th European Conf. on Space Debris*, Year.
- [133] M. Messerotti. Space weather and space climate. In J. Seckbach, J. Chela-Flores, T. Owen, and F. Raulin, editors, *Life in the Universe. Cellular Origin and Life in Extreme Habitats and Astrobiology*, volume 7, page 37. Springer, 2004.
- [134] M. Messerotti. Radio science for space weather. In *2016 URSI Asia-Pacific Radio Science Conference (URSI AP-RASC)*, pages 459–462, 2016.
- [135] M. Messerotti, P. Zlobec, I. Coretti, M. Jurcev, and S. Padovan. Tsrs diagnostics for space weather effects analysis. *Memorie della Societa Astronomica Italiana Supplementi*, 3:36, 2003.
- [136] R. Miteva and K.L. Klein. Solar radio bursts as a space weather diagnostic. *Space Science Reviews*, 180(1-4):203–260, 2014.
- [137] Y. T. Morton and B. C. Low. Impact of space weather on gnss. *Journal of Geophysical Research: Space Physics*, 123(10):8569–8583, 2018.
- [138] B. Muhammad, V. Alberti, A. Marassi, E. Cianca, and M. Messerotti. Performance assessment of gps receivers during the september 24, 2011, solar radio burst event. *J. Space Weather Space Clim.*, 5:A32, 2015.
- [139] B. Muhammad, V. Alberti, A. Marassi, E. Cianca, and M. Messerotti. Performance assessment of gps receivers during the september 24, 2011, solar radio burst event. *J. Space Weather Space Clim.*, 5:A32, 2015.
- [140] Gopalswamy N. Introduction to solar radio astronomy and radio physics. *Cambridge University Press*, 2006.
- [141] S. Nagasawa, T. Kawate, N. Narukage, T. Takahashi, A. Caspi, and T. N. Woods. Study of Time Evolution of Thermal and Nonthermal Emission from an M-class Solar Flare. *The Astrophysical Journal*, 933(2):173, Jul 2022.
- [142] NASA. Horizons system. https://ssd.jpl.nasa.gov/horizons/app.html# /.
- [143] B. Nava, P. Coisson, and S. Radicella. A new version of the nequick ionosphere electron density model. *Journal of Atmospheric and Solar-Terrestrial Physics*, 70(15):1856–1862, December 2008.
- [144] B. Nava, P. Coisson, and S.M. Radicella. A new version of the nequick ionosphere electron density model. *Journal of Atmospheric and Solar-Terrestrial Physics*, 70(15):1856–1862, 2008.
- [145] G J Nelson. Solar type ii radio bursts. *Astrophysical Journal*, 288:808–819, 1985.
- [146] A. Nindos. Incoherent solar radio emission. *Frontiers in Astronomy and Space Sciences*, 7, 2020.

BIBLIOGRAPHY

- [147] Carey Noll. The crustal dynamics data information system: A resource to support scientific analysis using space geodesy. *Advances in Space Research*, 45(12):1421–1440, June 15 2010.
- [148] A. I. Ohl and G. I. Ohl. A new method of very long-term prediction of solar activity. In Richard F. Donnelly, editor, *NOAA Solar-Terrestrial Predictions Proceedings. Volume 2.*, volume 2, pages 258–263, December 1979.
- [149] J. Oloketuyi and et al. Investigating the associations between solar flares and magnetic complexity of active regions. *New Astronomy*, 100:101972, 2023.
- [150] A. Pankratz. *Forecasting with Univariate Box–Jenkins Models: Concepts and Case.* Wiley Series in Probability and Statistics, 1983.
- [151] B. Panos and L. Kleint. Real-time Flare Prediction Based on Distinctions between Flaring and Non-flaring Active Region Spectra. *The Astrophysical Journal*, 891(1):17, 2020.
- [152] E. Papini. Github repository. <https://github.com/Acicone/FIF>, 2021.
- [153] E. Papini. Github repository. Available online, 2022. Accessed on 23 March 2023.
- [154] F. Pedregosa, G. Varoquaux, and A. et al. Gramfort. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [155] A. Pellizzoni, S. Righini, M. N. Iacolina, and et al. Solar observations with single-dish inaf radio telescopes: Continuum imaging in the 18–26 ghz range. *Solar Physics*, 297:86, 2022.
- [156] W.D. Pesnell. Lessons learned from predictions of solar cycle 24. *J. Space Weather Space Clim.*, 10(60), 12 2020.
- [157] F. Petropoulos and E. Spiliotis. The wisdom of the data: Getting the most out of univariate time series forecasting. *Forecasting*, 3:478–497, 2021.
- [158] E. Petrova, T. Podladchikova, A.M. Veronig, S. Lemmens, B.B. Virgili, and T. Flohrer. Medium-term predictions of f10.7 and f30 cm solar radio flux with the adaptive kalman filter. *Astrophys. J. Suppl. Ser.*, 254:9, 2021.
- [159] X. Pi, A. J. Mannucci, U. J. Lindqwister, and C. M. Ho. Monitoring of global ionospheric irregularities using the worldwide gps network. *Geophysical Research Letters*, 24(18):2283–2286, 1997.
- [160] M. Pick and N. Vilmer. Sixty-five years of solar radioastronomy: flares, coronal mass ejections and Sun–Earth connection. *The Astronomy and Astrophysics Review*, 16, 10 2008.
- [161] J.M Picone, A. E. Hedin, D. P. Drob, and A.C. Aikin. NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. *JGR Space Physics*, 107(A12), 12 2002.

BIBLIOGRAPHY

- [162] N. Plutino, F. Berrilli, D. Del Moro, and L. Giovannelli. A new catalogue of solar flare events from soft x-ray goes signal in the period 1986–2020. *Advances in Space Research*, 71(4):2048–2058, 2023.
- [163] A. Prasad, S. Roy, A. Sarkar, S. Chandra Panja, and S. Narayan Patra. Prediction of solar cycle 25 using deep learning based long short-term memory forecasting technique. *Adv. Space Res.*, 69:798, 2022.
- [164] P. Prikryl, A. J. Coster, and T. J. Fuller-Rowell. Gnss and space weather. *Radio Science*, 46(2):RS2005, 2011.
- [165] Y. Qian, J. Yang, H. Zhang, C. Shen, and Y. Wu. An hourly prediction model of relativistic electrons based on empirical mode decomposition. *Space Weather*, 17:e2018SW002078, 2020.
- [166] J. A. Ratcliffe. *The Magneto-Ionic Theory and Its Applications to the Ionosphere*. Cambridge University Press, Cambridge, 1959.
- [167] J.-P. Raulin and A.A. Pacini. Solar radio emissions. *Advances in Space Research*, 35(5):739–754, 2005.
- [168] Ye Ren, Le Zhang, and P.N. Suganthan. Ensemble classification and regression-recent developments, applications and future directions [review article]. *IEEE Computational Intelligence Magazine*, 11(1):41–53, 2016.
- [169] P. Riley, J. A. Linker, and R. Lionello. Advances in space weather forecasting. *Space Weather*, 16(2):156–173, 2018.
- [170] D. A. Roberts and et al. The spase data model: A metadata standard for registering, finding, accessing, and using heliophysics data obtained from observations and modelling. *Space Weather*, 16:1899–1911, 2018.
- [171] R. Romero and G. Navarro. Impact of solar radio bursts on gnss signals. *Journal of Space Weather and Space Climate*, 9:A14, 2019.
- [172] Daniel L Rubinfeld. Reference guide on multiple regression. *Reference manual on scientific evidence*, 179:425–469, 2000.
- [173] Mulas S., Pellizzoni A., Marcucci A., and Marongiu M. Correlation between active regions’ spectra at high radio frequencies and solar flare occurrences. *Solar Physics*, 2024. in prep.
- [174] H. Salmane, R. Weber, K. Abed-Meraim, and et al. A method for the automated detection of solar radio bursts in dynamic spectra. *Journal of Space Weather and Space Climate*, 8:A43, 2018.
- [175] I.H. Sarker. Data science and analytics: An overview from data-driven smart computing, decision-making and applications perspective. *SN COMPUT. SCI.*, 2:377, 2021.

BIBLIOGRAPHY

- [176] R. Schwenn. Space weather: The solar perspective. *Living Reviews in Solar Physics*, 3:2, 2006.
- [177] S. E. Segre and V. Zanza. Evolution of polarization for radiation crossing a plasma layer of quasi-transverse propagation and the interpretation of radioastronomical measurements. *The Astrophysical Journal*, 554:408–415, June 2001.
- [178] C. L. Selhorst, A. V. R. Silva, and J. E. R. Costa. Solar atmospheric model with spicules applied to radio observation. *Astron. Astrophys.*, 433(1):365–374, April 2005.
- [179] K. Shibasaki, C. E. Alissandrakis, and S. Pohjolainen. Radio emission of the quiet sun and active regions (invited review). *Solar Phys.*, 273(2):309–337, November 2011.
- [180] K. Shibasaki, S. Enome, H. Nakajima, M. Nishio, T. Takano, and et al. A purely polarized s-component at 17 ghz. *Publ. Astron. Soc. Jpn.*, 46:L17–L20, 1994.
- [181] K. Shibata and T. Magara. Solar flares: magnetohydrodynamic processes. *Living Reviews in Solar Physics*, 8(6), 2011.
- [182] M. Shimojo, T. S. Bastian, S. Kamenó, and et al. Observing the sun with the atacama large millimeter/submillimeter array (alma): Polarization observations at 3 mm. *Solar Physics*, 299(2):20, 2024.
- [183] A. V. R. Silva, T. F. Laganá, C. G. Gimenez Castro, and et al. Diffuse component spectra of solar active regions at submillimeter wavelengths. *Solar Phys.*, 227(2):265–281, April 2005.
- [184] A. Silwal, S. Gautam, P. Poudel, and et al. Gps observations of ionospheric tec variations during the 15th jan 2010 and 21st june 2020 solar eclipse. *Radio Science*, 56, 05 2021.
- [185] P. Singh, S. D. Joshi, and et al. The fourier decomposition method for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2199):20160871, 2017.
- [186] Skander Soltani. On the use of the wavelet decomposition for time series prediction. *Neurocomputing*, 48(1):267–277, 2002.
- [187] Materassi M. Stallone A., Cicone A. New insights and best practices for the successful use of empirical mode decomposition, iterative filtering and derived algorithms. *Scientific Reports*, 10, 2020.
- [188] E. Stevenson, V. Rodriguez-Fernandez, E. Minisci, and D. Camacho. A deep learning approach to solar radio flux forecasting. *Acta Astronautica*, 193:595–606, 2022.
- [189] L. Svalgaard. Reconstruction of Solar Extreme Ultraviolet Flux 1740–2015. *Solar Physics*, 291(9), 08 2016.

BIBLIOGRAPHY

- [190] T. T. W. Haryono, A. U. Zailani, Y. M. Djaksana, N. Rosmawarni, and N. D. Arianti. Transformers in machine learning: Literature review. *Jurnal Penelitian Pendidikan IPA*, 9(9):604–610, 2023.
- [191] G Taffoni, U Becciani, B Garilli, G Maggio, F Pasian, G Umana, R Smareglia, and F Vitello. Chipp: Inaf pilot project for htc, hpc and hpda., 2020.
- [192] R. Tang, W. Liao, Z. Chen, and et al. Solar Flare Prediction Based on the Fusion of Multiple Deep-learning Models. *The Astrophysical Journal Supplement Series*, 257(2):50, 2021.
- [193] K. Tapping. The 10.7 cm solar radio flux (f10.7). *Space Weather*, 11:394–406, 2013.
- [194] Yekoye Asmare Tariku. Evaluation of the ssn and f10.7 indices as a cause for performance variation of the iri 2016 and iri-plas 2017 models. *Advances in Space Research*, 64(9):1751–1763, 2019.
- [195] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.
- [196] W.K. Tobiska, S.D. Bouwer, and B.R. Bowman. The development of new solar indices for use in thermospheric density modeling. *J. Atmos. Sol.-Terr. Phys.*, 70:803–819, 2008.
- [197] J. F. Torres and D. et al. Hadjout. Deep learning for time series forecasting: A survey. *Big Data*, 9(1):3–21, 2021.
- [198] D.A. Vallado and D. Finkleman. A critical assessment of satellite drag and atmospheric density modeling. *Acta Astronautica*, 95, 2014.
- [199] L. van Driel-Gesztelyi and L. M. Green. Evolution of active regions. *Living Reviews in Solar Physics*, 12:1, 2015.
- [200] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.
- [201] A.M. Veronig, S. Jain, T. Podladchikova, W. Pötzi, and F. Clette. Hemispheric sunspot numbers 1874–2020. *Astron. Astrophys.*, 652:A56, 2021.
- [202] A Vourlidas and S Bruinsma. EUV Irradiance Inputs to Thermospheric Density Models: Open Issues and Path Forward. *Space Weather*, 16(1):5–15, 1 2018.
- [203] A. Vourlidas and R.A. Howard. The role of coronal mass ejections in the structure and evolution of the solar wind. *Living Reviews in Solar Physics*, 7(3), 2010.
- [204] B. Vršnak, T. Žic, D. Vrbanec, and M. Temmer. Propagation of coronal mass ejections: The drag-based model. *Solar Physics*, 253(1-2):215–232, 2008.
- [205] Miss Lei Wang. Advanced multivariate time series forecasting models. *Journal of Mathematics and Statistics*, 14:253–260, 2018.

BIBLIOGRAPHY

- [206] Z. Wang, Q. Hu, Q. Zhong, and Y. Wang. Linear multistep f10.7 forecasting based on task correlation and heteroscedasticity. *Adv. Earth Space Sci.*, 5:863–874, 2018.
- [207] H.P. Warren, J.T. Emmert, and N.A. Crump. Linear forecasting of the f10.7 proxy for solar activity. *Space Weather*, 15:1039–1051, 2017.
- [208] R. S. Weigel et al. Hapi: An api standard for accessing heliophysics time series data. *Journal of Geophysical Research (Space Physics)*, 126, 2021.
- [209] S. M. White. Coronal magnetic field measurements through gyroresonance emission. In Dale E. Gary and C. U. Keller, editors, *Solar and Space Weather Radiophysics*, volume 314 of *Astrophysics and Space Science Library*. Springer, Dordrecht, 2004.
- [210] S. M. White, G. Thejappa, and M. R. Kundu. Observations of mode coupling in the solar corona and bipolar noise storms. *Solar Physics*, 138:163–187, March 1992.
- [211] J P Wild, G A Dulk, and D J Mclean. Observations of the spectrum of solar radio emission at meter wavelengths. *Australian Journal of Physics*, 7(2):159–169, 1954.
- [212] D. H. Wolpert. Stacked generalization. *Neural Networks*, 5(2):241–259, 1992.
- [213] D. H Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, 1997.
- [214] A.R. Yeates, D.H. Mackay, and A.A. van Ballegooijen. Modelling the global solar corona: Filament chirality observations and surface simulations". *Solar Physics*, 245:87–107, 2007.
- [215] J. R. Yeh, J. S. Shieh, and N. E. Huang. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Adv. Adapt. Data Anal.*, 2:135–156, 2010.
- [216] Xinan Yue, Weixing Wan, Limei Yan, Wenjie Sun, Lianhuan Hu, and William S. Schreiner. Chapter 22 - the effect of solar radio bursts on gnss signals. In Natalia Buzulukova, editor, *Extreme Events in Geospace*, pages 541–554. Elsevier, 2018.
- [217] G.P. Zhang. Neural networks for time-series forecasting. In G. Rozenberg, T. Bäck, and J.N. Kok, editors, *Handbook of Natural Computing*. Springer, Berlin, Heidelberg, 2012.
- [218] J. Zhang, Y. Feng, J. Zhang, and Y. Li. The short time prediction of the dst index based on the long-short time memory and empirical mode decomposition–long-short time memory models. *Applied Sciences*, 13(21):11824, 2023.
- [219] W. Zhang, X. Zhao, X. Feng, C. Liu, and et al. Predicting the daily 10.7-cm solar radio flux using the long short-term memory method. *Universe*, 8(30), 2022.
- [220] W. Zhang and X. et al. Zhao. Predicting the daily 10.7-cm solar radio flux using the long short-term memory method. *Universe*, 8(1):30, 2022.

BIBLIOGRAPHY

- [221] X. Zhang and et al. At-lstm: An attention-based lstm model for financial time series prediction. *2019 IOP Conf. Ser.: Mater. Sci. Eng.*, 569:052037, 2019.
- [222] Xuan et al. Zhang. At-lstm: An attention-based lstm model for financial time series prediction. *2019 IOP Conf. Ser.: Mater. Sci. Eng.*, 569:052037, 2019.
- [223] D. Zhao, W. Li, C. Li, and et al. Ionospheric phase scintillation index estimation based on 1 hz geodetic gnss receiver measurements by using continuous wavelet transform. *Space Weather*, 20(4):e2021SW003015, 2022.
- [224] V. Zheleznyakov and V. Zaitsev. Solar radio emission. *Nauka*, 29:29–43, 1970.
- [225] J. Zheng, J. Cheng, and Y. Yang. Partly ensemble empirical mode decomposition: An improved noise-assisted method for eliminating mode mixing. *Signal Processing*, 96:362–374, 2014.
- [226] H. Zhou and A. Cicone. Multivariate iterative filtering as a stable and fast alternative technique to multivariate emd for multidimensional nonstationary signals decomposition. In *Geophysical Research Abstracts*, volume 21, 2019.
- [227] Lifan Zhou and Siqin et al. Huang. The f10.7 solar radio flux prediction based on lstm neural network. In *China Automation Congress (CAC)*, 2021.

Acknowledgements

In the long and challenging journey that led to this thesis, the support of many people has been fundamental for me over the years, contributing in various ways to help me reach this point.

First of all, I would like to thank Prof. Mauro Messerotti who, with wisdom and kindness, guided me by leveraging his decades of experience and always believed in me, even in moments of uncertainty. I extend my deepest appreciation also to the PhD coordinator, Francesco Longo, my co-supervisor Prof. Anna Gregorio, Prof. Vincenzo Carbone and Dr. Andrea Csillaghy for their comments and suggestions that have helped me to improve my work.

I would like to express my sincere gratitude to Alberto Pellizzoni, for the trust he placed in me from the very beginning, for his passion and enthusiasm and for making me fully part of the projects and the Team. For this, I particularly thank Sara Mulas and Marco Marongiu for their invaluable collaboration.

I am deeply grateful to Giovanna Jerse and Valentina Alberti, who actively participated in my work and always offered their precious contributions.

I would like to extend my sincere thanks to Claudio Cesaroni and Eric Nana Asamoah for their invaluable help and for sharing their expertise with me.

I also feel the need to thank Lt.Col. Daniele Biron, without whom this journey might never have started, and Col. Francesco Zauli, who supported me in order to complete it.

I am also grateful to my friends, in a chaotic order, Roberto, Riccardo, Chiara, Marzia, Federico, Nicoletta and Norma for their encouragement, for the time spent together and for knowing that, somehow, they are there.