



## HOCHSCHULE NEUBRANDENBURG GEODESY AND GEOINFORMATICS

Illustration of different effects in the correlation of solar and ionosphere parameters in time scales of several days to months

Submitted By  
**FREDY DAVIS**

A thesis presented for the degree of  
Master of Engineering

Supervisors:  
**Prof. Dr.-Ing. Andreas Wehrenpfennig**  
**Dr. Claudia Borries**

Date of Submission: 18 January 2023

URN: urn:nbn:de:gbv:519-thesis2022-0263-8

# Abstract

The activities of the Sun have a significant impact on the ionosphere and its variations. Solar irradiance, approximated by F10.7 solar flux index, drives the ionization processes, increasing the electron density in the ionosphere. Solar wind deposits energy in the high-latitude ionosphere through convection electric and Joule heating. This study investigates different effects in the correlation of solar and ionosphere parameters in time scales of several days to months. It inspects the temporal, latitudinal and local time dependency of the correlation analysis between total electron content in the ionosphere against F10.7 solar flux and solar wind speed from 1998 to 2020 using Pearson correlation coefficient with a window size of 90 days. The correlation analysis is performed on the relative differences of the parameters ( $\Delta F_{10.7,rel}$ ,  $\Delta V_{SW,rel}$ ,  $\Delta TEC_{rel}$ ). The result shows, at low latitudes, there is a significant positive correlation between  $\Delta TEC_{rel}$  and  $\Delta F_{10.7,rel}$ , but as latitude increases, the correlation decreases. F10.7 solar flux and total electron content are positively correlated across all regions and seasons. Total electron content and F10.7 solar flux have no distinguishable pattern of seasonal correlation. However, the degree of positive correlation between summer and winter fluctuates slightly. The correlation between TEC and F10.7 solar flux is highest during daytime hours than nighttime hours and the local time dependency is most pronounced in the equatorial region. Results of correlation between  $\Delta TEC_{rel}$  and  $\Delta V_{SW,rel}$  shows positive correlation at the low latitude region during solar minimum conditions and at high latitude regions there it shows a significant negative correlation. In contrast to correlation between F10.7 solar flux, solar wind speed correlation with TEC shows strong seasonal dependency. Changes in electric fields caused by variations in the solar wind speed are assumed to modify the plasma transport in the ionosphere and

thus modify TEC. These dependencies of TEC on solar wind in the range of days needs deeper investigation to understand the coupling processes and quantify the impact.

# Acknowledgement

I would first like to thank my thesis supervisors, Prof. Dr.-Ing. Andreas Wehrenpfennig of the Department of Landscape Sciences and Geomatics at Hochschule Neubrandenburg and Dr. Claudia Borries of the Department for Solar-Terrestrial Coupling at the Deutsches Zentrum für Luft- und Raumfahrt (DLR), for their time, effort, and understanding in helping me succeed in my studies. Their extensive knowledge and expertise have inspired me throughout my studies. In addition, I'd like to thank Dr. Samira Tasnim for her valuable comments on this thesis and her assistance with my work. I want to thank everyone at the Deutsches Zentrum für Luft- und Raumfahrt (DLR) and Hochschule Neubrandenburg. Thanks to their generosity and encouragement, my time spent studying and living in Germany has been truly rewarding. To conclude, I'd like to thank my parents and other family members. It would have been impossible to finish my studies without their unwavering support over the past few years.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Space Weather . . . . .	1
1.2	Goal of the thesis . . . . .	2
<b>2</b>	<b>The Sun, Solar Wind and Ionosphere</b>	<b>5</b>
2.1	The Sun . . . . .	5
2.1.1	Structure of the Sun . . . . .	5
2.1.2	Sunspot & Solar cycle . . . . .	9
2.1.3	Coronal Holes . . . . .	11
2.2	Solar Wind and Interplanetary magnetic field . . . . .	12
2.3	The Atmosphere . . . . .	15
2.4	Ionosphere . . . . .	17
2.4.1	Structure of Ionosphere . . . . .	17
2.4.2	Ionosphere - Chemical process . . . . .	20
2.4.3	Solar variability effects in the ionosphere . . . . .	23
<b>3</b>	<b>Observations and Solar Parameters</b>	<b>26</b>
3.1	F10.7 Solar Flux . . . . .	26
3.2	Solar Wind Speed . . . . .	28
3.3	Total Electron Content (TEC) . . . . .	30
3.4	IGS TEC maps . . . . .	33
<b>4</b>	<b>Analysis</b>	<b>36</b>
4.1	Data Collection and Pre-processing . . . . .	36
4.2	Mathematical Background of the Correlation Analysis . . . . .	38
4.3	Software Tool . . . . .	41

4.3.1	Software Requirements and Design . . . . .	41
4.3.2	Related Packages . . . . .	44
4.3.3	Software Functionalities . . . . .	45
<b>5</b>	<b>Results</b>	<b>52</b>
5.1	Correlation between TEC and F10.7 Solar Flux . . . . .	52
5.1.1	Temporal changes of the Correlation between TEC and F10.7 Solar Flux . . . . .	52
5.1.2	Latitudinal dependency of the Correlation between TEC and F10.7 Solar Flux . . . . .	53
5.1.3	Local time dependency of the Correlation between TEC and F10.7 Solar Flux . . . . .	56
5.1.4	Discussion of TEC and F10.7 Solar Flux Correlation .	57
5.2	Correlation between TEC and Solar Wind Speed . . . . .	59
5.2.1	Temporal changes of the Correlation between TEC and Solar Wind Speed . . . . .	59
5.2.2	Latitudinal dependency of the Correlation between TEC and Solar Wind Speed . . . . .	60
5.2.3	Local time dependency of the Correlation between TEC and Solar Wind Speed . . . . .	63
5.2.4	Discussion of TEC and Solar Wind Speed Correlation .	64
<b>6</b>	<b>Summary and Conclusion</b>	<b>66</b>

# List of Figures

2.1	Structure of the Sun . . . . .	7
2.2	International Space Environmental Services(ISES) Solar Cycle Sunspot Number Progression. . . . .	10
2.3	Schematic representation of shape of the Earth's magnetosphere. . . . .	13
2.4	The artist's rendition of the heliospheric current sheet's distortion. . . . .	15
2.5	Vertical regions of the Earth's atmosphere. . . . .	16
2.6	Illustration of F2, F1, E and D layers of the ionosphere. . . . .	19
2.7	Cross-correlation coefficients and time delays between the global TEC and F10.7 according to Vaishnav et al. (2019) . . . . .	24
2.8	Correlation between TEC and between TEC and Mg II index for the years 1999 to 2017 . . . . .	25
3.1	F10.7 Solar flux values from 1998 to 2020. . . . .	28
3.2	Solar Wind Speed values from 1998 to 2020. . . . .	29
3.3	Zonal mean TEC values in the ionosphere at latitude 60N, 30N and 0N from 1998 to 2020. . . . .	33
3.4	IGS TEC map Date: 2003-07-15 . . . . .	34
4.1	The spectrum of the correlation coefficient (-1 to +1) . . . . .	40
4.2	Window size test with mean absolute correlation between $\Delta TEC_{rel}$ , $TEC$ , $\Delta V_{SW_{rel}}$ , $V_{SW}$ , $\Delta F_{10.7_{rel}}$ , $F_{10.7}$ at location 12.5N 0E . . . . .	41
4.3	Use case diagram of the software system. . . . .	43
4.4	Process flow of the software package for the analysis. . . . .	44
4.5	Workflow of data load module . . . . .	47
4.6	Workflow of data cleaning & formatting module . . . . .	48
4.7	Detailed workflow of computation module . . . . .	50

---

LIST OF FIGURES VII

---

4.8	Workflow of visualisation module. . . . .	51
5.1	Correlation between $\Delta F_{10.7_{rel}}$ and $\Delta TEC_{rel}$ with a 90 days window at 12.5 N latitude from 1998 to 2010 . . . . .	53
5.2	Latitudinal variation of correlation coefficient with 90 days box window; F10.7 radio Flux and Total Electron Content from 1998 to 2020. . . . .	54
5.3	Correlation between F10.7 radio Flux and Total Electron Content; seasonal mean from 1998 to 2020 . . . . .	55
5.4	Local time variation of correlation coefficient between F10.7 solar flux and TEC. . . . .	57
5.5	Correlation between solar wind speed and Total Electron Content with 90 days window at 12.5 N latitude from 1998 to 2020	60
5.6	Latitudinal variation of correlation coefficient with 90 days box window; Solar wind speed and Total Electron Content from 1998 to 2020 . . . . .	61
5.7	Correlation between solar wind speed and Total Electron Content; seasonal mean from 1998 to 2020 . . . . .	63
5.8	Local time variation of correlation coefficient; Solar wind speed and Total Electron Content from 1998 to 2020 at locations 60N/15E, 30N/15E and 0N/15E . . . . .	64

# List of Tables

2.1 Statistical Properties of the Solar Wind at 1 AU . . . . .	13
--	----

# **Chapter 1**

## **Introduction**

### **1.1 Space Weather**

Space weather is a modern topic of interest. It provides new insights into the complex interactions and influences of the Sun and other cosmic sources on interplanetary space, thermosphere, ionosphere and the Earth's magnetosphere, on space- and ground-based technology systems, and life on Earth. Spacecraft's sophisticated equipment have given us unparalleled views of the Sun's surface and outer atmosphere (Bothmer & Daglis, 2007). Space weather is very relevant to today's society because everyday's life depends to a large extend on technological infrastructure which is vulnerable to space weather impact. As the term implies, 'space weather' is variable weather conditions on the Sun and in space which mainly caused by activities on the Sun. Extreme solar eruptions can create large concentrations of magnetized plasma and/or high-energy particle radiation that potentially engulf Earth. Spacecrafts and satellites must be prepared to withstand these occurrences, but the ejected particles can also penetrate Earth's magnetic fields and atmosphere, reaching the surface. These impacts can disturb and occasionally harm a wide range of daily technology systems that are frequently vital to the normal operation of modern civilizations. Plasma processes can destabilize technology. They can cause time-varying electric currents in the upper atmosphere, altering Earth's magnetic field. Changes in the magnetic fields can cause geo-electric fields which produce excess electric currents in

electrical network, affecting electricity supply. These plasma processes may deposit significant amount of thermal energy in the upper atmosphere region, especially in polar areas, changing the global circulation, density, and composition. Changes in the ionosphere's global morphology might alter radio signal transmission across that area. Space weather hazards have increased as our reliance on modern technologies has grown. Reports of detrimental impacts from significant space-weather events from 1967 till date (see Dang et al. (2022), Doherty et al. (2004), Knipp et al. (2018), Knipp et al. (2016), and Moran et al. (2014), Wik et al. (2009)) have reinforced the risk's significance. These occurrences have helped us better understand space weather and form a community of scientists, engineers, economists, and policymakers (Hapgood, 2017).

There are several solar processes that might cause space weather storms. Massive explosions on the Sun cause Solar Flares and Coronal Mass Ejections (CME) that propel space weather storms across our solar system. Additionally, the Sun generates a continual stream of charged particles that form the plasma of the solar wind. Sunspots increase and decrease in 11-year solar cycles. Solar Minimum describes years with few sunspots; Solar Maximum has many. When sunspot numbers are high, the Sun is particularly active, although major storms can occur anytime. Sunspots reveal where the Sun's magnetic field is building up and might release to create solar flares and CMEs. Solar maximum raises solar radiation and this extra energy modifies Earth's upper atmosphere. Activity on the Sun can produce space weather events that influence Earth. Solar storms affect GPS, satellites, and electrical networks etc. The US National Weather Service carries out studies in solar-terrestrial physics and offers real-time monitoring, prediction of solar activities and issues updates and warnings which may affect technology infrastructure and people in space (NOAA, n.d.).

## 1.2 Goal of the thesis

Variations in solar activity, as measured by sunspot numbers, have been recorded since earlier times, but how they impact earth's climate is questioned. The necessity to distinguish between natural and man-made climate change causes has pushed the topic back to the forefront of meteorological study (Haigh, 2007). As the Earth's ionosphere is formed largely by solar ion-

izing flux, classical study implies that substantial fluctuations in ionospheric electron density are connected to solar and geomagnetic activity. While these phenomena are the principal causes of ionospheric variability, studies also show large fluctuations in electron density related to impacts of gravity waves, tides, and planetary waves. Ionosondes have been used for decades to monitor the Earth's ionosphere, although they can only study local conditions. In the last two decades, Total Electron Content (TEC) measurements from ground-based Global Navigation Satellite System (GNSS) receivers have advanced our understanding of the ionosphere's spatial and temporal development (Goncharenko et al., 2022). Various scientific journals, magazines, and newspapers reported on the impacts of solar activity on Earth and in space, like satellite damage, radiation risks to astronauts and airline passengers, telecommunication challenges, power failures, and the significance of solar and cosmic factors for the evolution of Earth's climate (Bothmer & Daglis, 2007).

Variations in solar rotation have an effect on daily fluctuations in extreme ultraviolet (EUV) radiation from the Sun. Ionization and the heating of the upper atmosphere are both primarily powered by the energy that is provided by EUV irradiation. Therefore variations in the EUV radiation emitted by the Sun have a substantial impact on the ionosphere. The most prominent EUV fluctuations are the 11 year solar cycle and 27-day solar rotation variations which are greatly affect properties of ionosphere including electron density (Chen et al., 2018). Ionospheric disturbances may also be brought on by geomagnetic storms, which are caused by solar wind. The geomagnetic field act as a shields from the plasma impacts of the solar wind. Despite this, research has shown that its effects are transferred to the ionosphere. It is not only accountable for the disruptions but for some anomalies and also for certain irregularities in high latitude regions of lower magnitude (Piddington, 1964).

The main objective of this thesis is to investigate the correlation between solar forcing parameters in the ionosphere. The study investigates correlation between total electron content in the ionosphere against F10.7 solar flux and solar wind speed from 1998 to 2020. The purpose of correlation is to examine the degree of association between two variables under study. The correlation coefficient quantifies the degree of link between variables. It also investigates the seasonal, latitudinal, local time and solar activity dependence of the

variables in the 22 year span. It inspects the temporal, latitudinal and local time dependency of the correlation analysis between TEC with F10.7 solar flux and TEC with solar wind speed.

This work is structured as introduction giving overview about the space weather followed by chapter 2 discusses the Sun, solar wind and ionosphere. Chapter 3 describes about solar forcing parameters which the study is carried out. Solar forcing parameters including  $10.7\text{cm}$  Solar Radio Flux, solar wind speed are discussed in this chapter. Chapter 4 describes the mathematical background of correlation analysis, data pre-processing and also about the development of software tool for the analysis. Chapter 5 discusses the correlation between Total Electron Content in the ionosphere against  $10.7\text{cm}$  solar Flux and solar wind speed. The summary and conclusions of the analysis are described in Chapter 6.

# Chapter 2

## The Sun, Solar Wind and Ionosphere

### 2.1 The Sun

#### 2.1.1 Structure of the Sun

The Sun is the most important celestial body on our planet. It's a ball of hot plasma that's been heated to incandescence by the nuclear fusion processes at its core. The majority of the energy it emits as visible light, ultraviolet light, and infrared radiation. It is the primary source to the energy requirements of all living creatures on Earth. The Sun provides around 99.998% of the entire energy input to the Earth's climate; the little remaining is given by geothermal source of heat. The Sun's gravity also contributes to one-third of the tides. The region of the electromagnetic spectrum where the Sun emits the greatest energy corresponds to the portion of this spectrum that is visible to humans and the vast majority of other species (Arnold, 2006; Stix, 2012).

The Sun comprises of distinct parts with a bowl-shaped structure, and though the boundaries between them are not precisely defined. Figure 2.1 shows a schematic illustration of the Sun's structure by NASA.

**Core:** All of the Sun's energy comes from a place inside it called the “core”.

From the center, this core goes out to about a quarter of the diameter of the Sun's visible surface. Even though the core is only 1.6% of the Sun's total volume, it contains approximately 50% of the Sun's mass. Matter in the Sun is in the form of a plasma at a temperature to around 15.6 million K. Hydrogen nuclei combine to make helium nuclei in a process called the proton-proton reaction. This creates gamma radiation and electron neutrinos. The helium nuclei that are made have a little more energy because of the binding energy, less mass than the sum of the original nuclei of hydrogen (mass defect). Using the formula  $E = m \times c^2$ , the difference in mass is turned into energy (per proton-proton fusion = 27 MeV). In the center of the Sun, 700 million tonnes of hydrogen are turned into 695 million tonnes of helium every second. This gives off about 400 quadrillion watts of power, or  $4 \times 10^{26}$  W (Arnold, 2006).

**Radiation zone:** “radiation zone” is the area around the Sun’s core. It takes up about 70% of the Sun’s radius. Gamma photons move across space at the speed of light because space is empty. Density inside the Sun is higher so that photons keep running into plasma particles, getting drawn in, and then being sent out again. They move in a random way and spread out until they reach the Sun’s surface. According to statistics, it takes about 170,000 years for a photon to move through the radiation zone. This explains that the light from the Sun that we see today was made a very long time ago. Every time a photon strikes in the radiation zone, its energy drops and its wavelength grows. X-rays are made from the gamma radiation. Unlike photons, neutrinos move through the Sun’s layers almost unhindered because they rarely interact with matter. Since neutrinos move at the speed of light, they reach the earth in just eight minutes and move through it almost without being stopped. One square centimeter of the Earth’s surface is hit by about 70 billion neutrinos every second (Arnold, 2006).

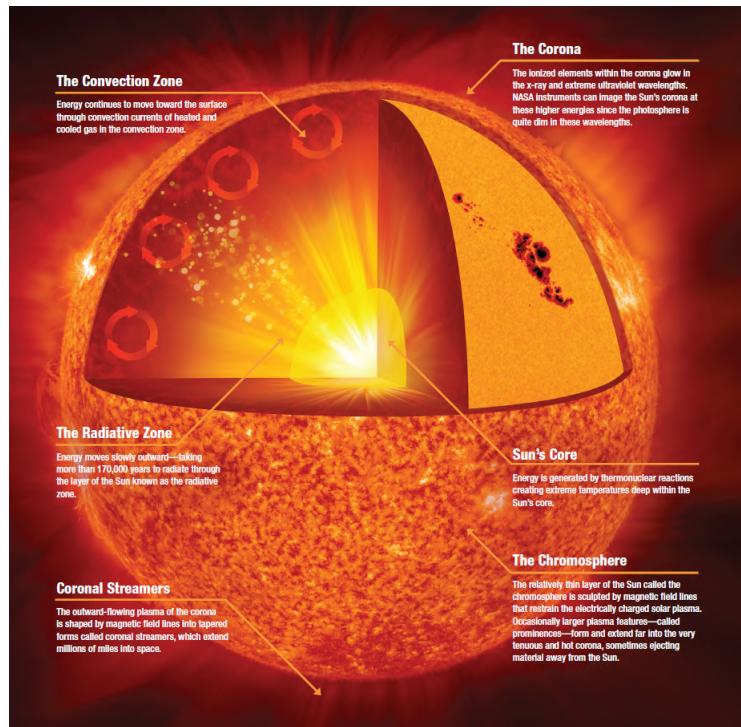


Figure 2.1: Structure of the Sun (NASA, 2011).

**Convection zone:** In the convection zone, energy is no longer emitted by radiation but by plasma convection. Hot stuff rises to the Sun's surface, cools, and sinks back within. Since rising plasma is hotter and brighter than descending plasma, convection cells appear as granulation on the Sun's surface.

**The Chromosphere:** The chromosphere is located above the photosphere and has an average thickness of around 1500 kilometers. During solar eclipses, it may be seen as a crimson light obscured by the photosphere. In this part of the Sun, the temperature gradually rises to reach 10,000 K (Engvold et al., 2018). The chromosphere differs from the underlying photosphere in terms of less density and higher temperature. The formation of shocks which causes the chromospheric heating is triggered by the upward movement of acoustic waves that are formed in the turbulent flow field of the upper convection zone. In addition to that, the magnetic field also plays a significant part in

the heating process. Large areas of horizontal magnetic field are seen in magnetograms taken on chromospheric lines close to the solar limb (Solanki & Hammer, 2001). Due to huge vortices and changing magnetic fields (source of sunspots), the surface is not smooth. SOHO, TRACE, or CHANDRA observations can be digitally processed to resemble hard, slowly moving material. The turbulence on the surface is also caused by the electric current. The coronal plasma gets frozen into the magnetic field due to the high conductivity of the Sun's surrounding environment (Arnold, 2006; Solanki et al., 2000).

One of the most dynamic areas of the solar environment is where the chromosphere and corona intersect which is called transition region. It distinguishes two distinct temperature ranges where the energy balance between heating and cooling processes behaves differently (Solanki & Hammer, 2001).

**Solar Corona and coronal streamers:** Above the chromosphere is the corona, whose density is  $10^{-4}$  to  $10^{-19} g/cm^3$ . The inner corona spans 1-2 solar radii into interplanetary space, depending on the solar cycle. Solar radiation, shock waves, and other mechanical or magnetic interactions heat the corona to 2 million Kelvin. These heating processes' causes are unknown. Acoustic waves and microflares might provide energy. At the corona's lower edge, where density declines faster than energy can be removed, the temperature gradient is very significant. Within a few hundred kilometers of altitude, the temperature of the kinetic gas goes up by a million degrees and "becomes air" as the extra heating energy releases as solar wind.

A dynamo process inside the Sun creates the magnetic field that can be seen on the Sun's surface. A small amount of this field goes beyond the Sun's corona and spreads through space. The field shows up at the surface as bipolar regions with a wide range of size and magnetic flux. The largest of these are called active regions, and the lowest are called ephemeral regions. The flow of electrically conductive gases is what makes the Sun's magnetic fields so strong. Plasma inside the Sun has the same ability to move electricity as copper at room temperature. There are electric currents in the Sun that are on the order of  $10^{12} \text{ Ampere}$ . So, the inside of the Sun works like a huge dynamo, turning the motion of an electrical conductor into electricity and a magnetic field. Sunspots and prominences are two things that can be seen as effects of the magnetic fields. Sunspots are areas of the Sun's atmosphere

that are relatively low temperature. Their temperature ranges from 3,700 K to 4,500 K (Arnold, 2006; Solanki et al., 2000).

One of the most prominent evidence of solar activity are coronal streamers, often known as Helmet streamers. It is observed in white and UV light as bright, huge, and persistent structures that extend several solar radii of the solar corona. Coronal streamers seem to have a structure like a pointed helmet or an arch that extends into interplanetary space eventually fading. During the solar maximum, it may be seen in any latitudes, whereas during the solar minimum, it is usually observed on the solar equator (Parenti et al., 2000). The network of magnetic field lines that links the sunspots in the active regions causes the coronal streamers to emerge from the Sun. The closed magnetic field lines are lead to the formation of the structure that contains electrically charged particle. The movement of the solar wind flowing from the Sun into the gaps between the streamers is responsible for the formation of the pointed structure.

### 2.1.2 Sunspot & Solar cycle

Sunspots are comparatively cool regions that look as dark spots on the surface of the Sun. It is also know that a sunspot is a region with a high magnetic field that prevents the convective transmission of heat and, as a result, keeps this spot cooler than the surface that is around it (Ruzmaikin, 2001). Sunspots are being employed extensively throughout the field of space weather research as a measure for the solar activity. The Zürich or Wolf sunspot number has been the most important measure of the Sun's activity (now known as the International sunspot number). It is the oldest way to track how active the Sun has been over time (Hathaway et al., 2002).

Figure 2.2 from the National Oceanic and Atmospheric Administration's Space Weather Prediction Center displays the observed and expected sunspot number. The black line depicts monthly averaged data, whereas the purple line depicts a 13-month, smoothed version of monthly mean data. The red line is the prediction for the upcoming solar cycle. The 23rd solar cycle started in late 1996 and ended in 2008 and solar cycle 24 began in early 2009 and ended in 2019. The smoothed monthly value during cycle 23's solar minimum is approximately 11.2, while the value observed during the solar maximum is 180.3. Considering solar cycle 24, the smoothed monthly value

during solar minimum is 2.2, and during solar maximum it is 116.4. The Sun seems to be less active in solar cycle 24 compared to solar cycle 23. It is anticipated that the solar maximum of solar cycle 25 will take place during November 2024 and March 2026. Moreover, there is a higher possibility that the actual number of sunspots will be higher than the numbers that were predicted.

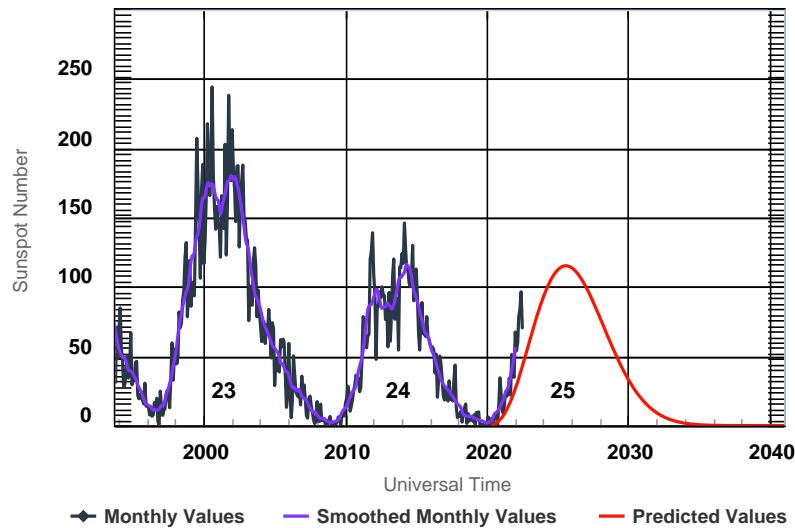


Figure 2.2: International Space Environmental Services(ISES) Solar Cycle Sunspot Number Progression. (Space Weather Prediction Center ([www.swpc.noaa.gov](http://www.swpc.noaa.gov)))

Many facts that can be seen help us figure out where sunspots come from. The most important ones, especially those that have to do with magnetic fields are spots show up in a latitude that is about 20 degrees wide. In a given solar cycle, the first spots show up at about 30 degree in both hemispheres. Each spot or group of spots lasts only a few days or a few weeks, but the average latitude of where they appear moves toward the equator over an 11-year period (the butterfly diagram). Spots come in pairs, with preceding (*p*) and following (*f*) elements having opposite magnetic polarities. Hale's law says that the *p* – *f* polarities are different in the northern and southern

hemispheres, and they switch every 11 years. The *f* spot is farther from the equator than the *p* spot. As latitude goes up, the angle between the line that connects these points and the east-west direction gets bigger. Sunspots are made when bright and dark spots (faculae and pores) join together. Faculae are similar to small magnetic flux tubes, and pores are similar to middle-sized flux tubes. Spots tend to come back in places where people have been before. This makes sunspots group together and creates clusters of activity that last for a long time.

Observations of individual sunspots are completely in line with the classic idea of a magnetic loop forming, the feet of which make up the opposite-polarity part of a new sunspot pair. In addition to it, the investigations demonstrate that a sunspot is just not made by only one flux emergence but by the assembly of many loops that have emerged, and that sunspots tend to group together. Observations of sunspots as a group show that there is a subphotospheric magnetic field with a regular direction in the azimuth direction that changes its polarity every 11 years (Ruzmaikin, 2001).

### 2.1.3 Coronal Holes

Coronal holes are the areas of the Sun that are seen to be the darkest and less active. These zones can be found on both the solar disk and over the solar limb. Coronal holes are considered to be open magnetic fields that are increasing at a rapid rate and also induces the high speed solar wind. Coronal holes are low-density plasma with magnetic fields that open into the heliosphere. Coronal holes tend to behave as collisionless plasma in the outer solar environment due to their low density. Open magnetic fields carry ions and electrons into the heliosphere (Cranmer, 2009b). Coronal holes are common during periods of low solar activity and will remain for several solar rotations (Kutiev Ivan et al., 2013). The North Pole and South Pole of the Sun are the most common and stable locations for coronal holes, although these polar holes may develop and spread further to equatorial regions of the Sun (Cranmer, 2009a).

It has also been discovered that coronal holes are closely connected to high speed solar wind streams which is capable to generate mild geomagnetic storms (Kutiev Ivan et al., 2013; Nolte et al., 1976). During the solar minimum conditions, the influence of solar irradiance on the ionosphere is coun-

teracted by frequent geomagnetic storms caused by coronal holes (Kutiev Ivan et al., 2013).

Solar plasma is pushed out from the rotating Sun along the open magnetic field lines radially, reaching high velocity and generating the high speed solar wind (Hofmeister, Stefan J. et al., 2022).

## 2.2 Solar Wind and Interplanetary magnetic field

The Solar Wind is a phenomenon that can be found through out the interplanetary space. It exists as a result of the expansion of the solar corona, the hot outer atmosphere of the Sun, at supersonic speeds. The solar corona is plasma, which is a charged gas. Most of the solar wind is made up of electrons and protons, but there are also small amounts of alpha particles and numerous other ionic particles. At 1 astronomical unit (AU) from the Sun, the average density and flow speed, and temperature of the solar wind are about 8 protons per cubic centimeter,  $440 \text{ km/s}$ , and  $1.2 \times 10^5 \text{ K}$ , respectively. However, the solar wind is quite variable in both space and time. Some low-energy cosmic rays are kept out of the solar system by a weak magnetic field in the plasma of the solar wind, and energetic particles from the Sun are guided into the heliosphere by the same field. The solar wind plays a key role in shaping the Earth's magnetosphere like a comet's tail (Gosling, 2014).

The solar wind's interaction with planets and other solar system bodies is also very dynamic, depending on each body's magnetic field. For the Venus and the Moon, the solar wind hits the atmosphere or solid surface immediately. The solar wind is deflected due to the presence of magnetic field on Earth and Jupiter. As a result of solar wind's supersonic expansion, a shock wave or bow shock forms at its interface with the magnetosphere or ionosphere. The Earth's magnetosphere extends  $\sim 12$  planetary radii sunward direction and Jupiter's 50–100 radii. Solar wind ions can escape into planetary magnetospheres at the poles, causing auroras on Earth. Solar wind's interaction with planetary magnetospheres creates massive magnetotail formations that often stretch interplanetary distances (See figure 2.3) (Gosling, 2014).

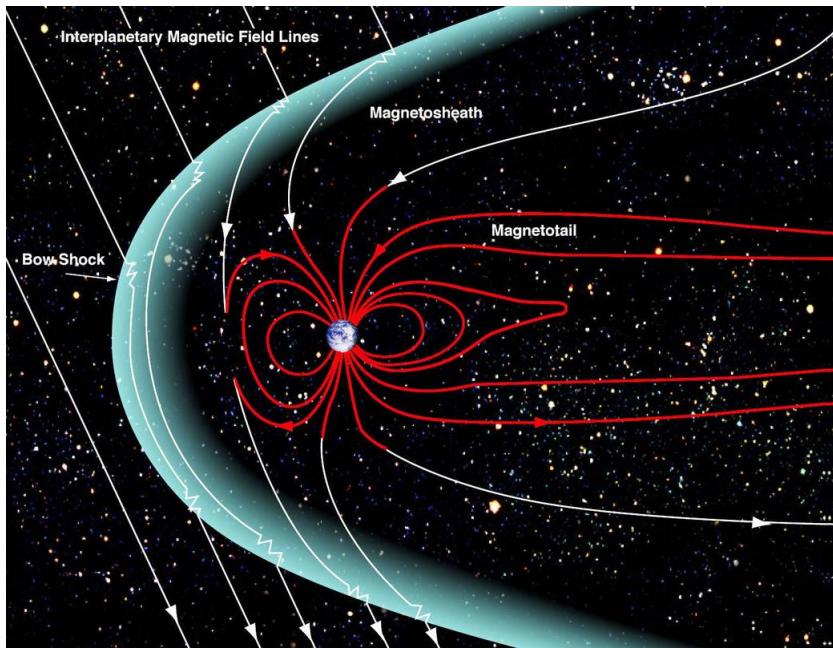


Figure 2.3: Schematic representation of shape of the Earth's magnetosphere (Zell, 2018).

Parameter	Mean	STD
$n (cm^{-3})$	8.7	6.6
$V_{sw} (km/s)$	468	116
$B (nT)$	6.2	2.9
$A(He)$	0.047	0.019
$T_p (\times 10^5 K)$	1.2	0.9
$T_e (\times 10^5 K)$	1.4	0.4
$T_\alpha (\times 10^5 K)$	5.8	5.0
$T_e/T_p$	1.9	1.6
$T_\alpha/T_p$	4.9	1.8

Table 2.1: Statistical Properties of the Solar Wind at 1 AU

Typically, two types of plasma flows are recognized in solar wind: slow solar wind with velocities below 400 km/h and fast solar wind above this speed. Proton density in slow and fast solar wind is  $8.3 \text{ cm}^{-3}$  and  $2.5 \text{ cm}^{-3}$  at Earth

orbit, respectively. Fast solar wind is connected with plasma ejection from coronal holes, with speeds vary from 450 to 800 km/s and maximum solar wind speed is linearly correlated with coronal hole size. Slow solar wind may originate from the edges and tips of closed-field structures on the Sun (Wurz, 2005).

The solar wind carries the coronal magnetic field into the solar system, generating the interplanetary magnetic field (IMF) (Owens & Forsyth, 2013). The solar wind's influence on the magnetic field structure causes magnetic field lines to expand out from the Sun in a radial manner (Hofmeister, Stefan J. et al., 2022). The IMF's structure and dynamics are crucial to understanding and predicting space weather because it connects the Sun to planetary magnetospheres and channels solar and cosmic energetic particles (Owens & Forsyth, 2013).

The IMF geometry may be understood by assuming a steady-state idealized solar wind with a radial outflow of steady velocity regardless of radial and latitudinal location. Magnetic field line footpoints are considered to be stationary in the photosphere and spin with the Sun. This magnetic field is thought to be frozen in solar wind plasma but have no effect. Parker (1958) anticipated that such circumstances would twist the heliospheric magnetic field like an Archimedean spiral in the solar equatorial plane (Owens & Forsyth, 2013). A schematic picture of the Parker Spiral is demonstrated in Figure 2.4

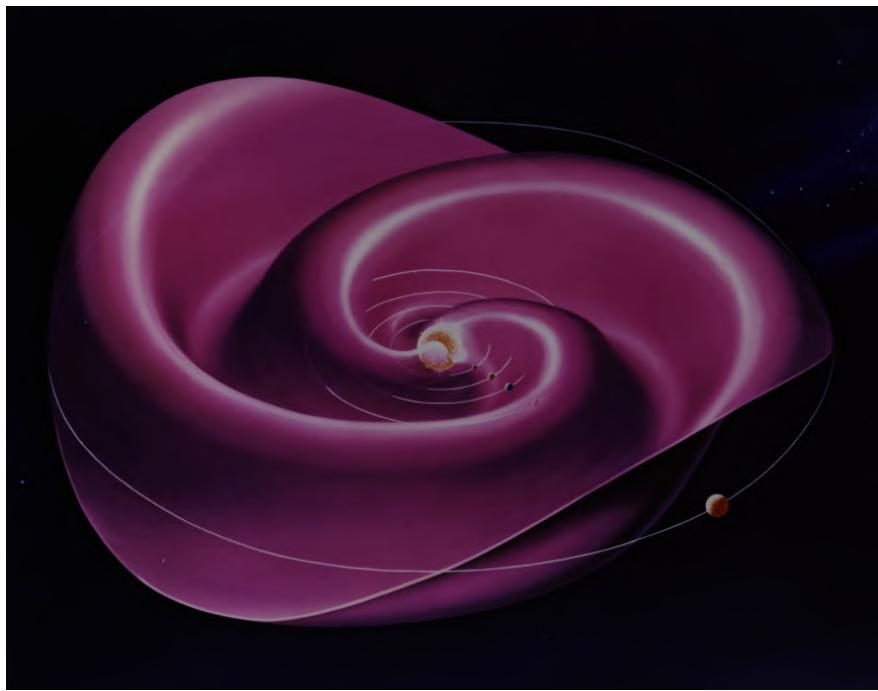


Figure 2.4: The artist's rendition of the heliospheric current sheet's distortion (Wilcox et al., 1980).

## 2.3 The Atmosphere

As seen in Fig 2.5, Earth's atmosphere is divided into five distinct layers, each with an increasing height. Where the temperature changes, new layers begin to form.

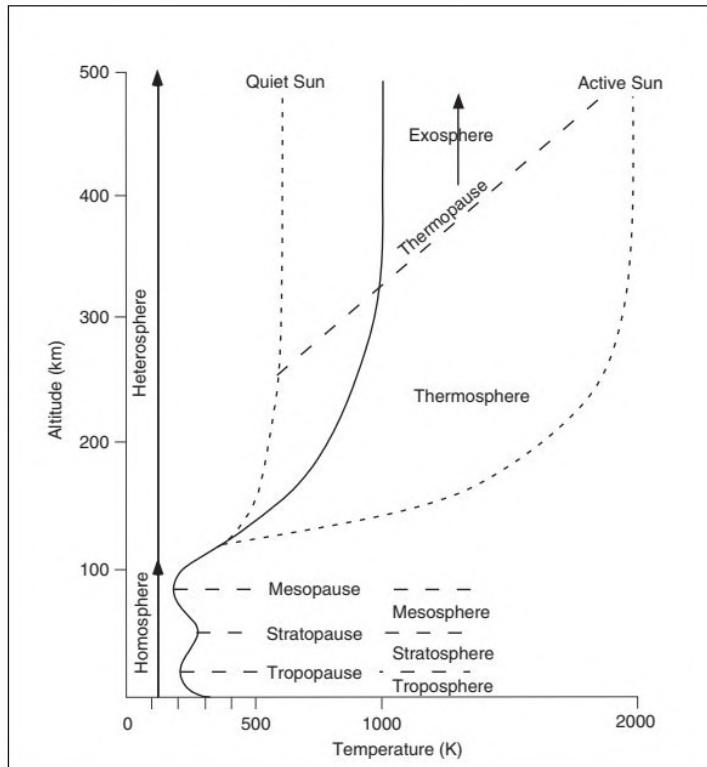


Figure 2.5: Vertical regions of the Earth’s atmosphere, shown schematically (Catling & Kasting, 2017).

The troposphere is the area between the Earth’s surface and the tropopause, which is about  $11\text{ km}$  ( $0.2\text{ bar}$ ) altitude. The tropopause can be as low as  $8\text{ km}$  at the poles or as high as  $17\text{ km}$  in the tropics. The average pressure at the tropopause is  $0.16\text{ bar}$ , but it varies from  $0.1$  to  $0.3\text{ bar}$  from the tropics to the poles. *Tropos* is the ancient Greek word for “turning” and “pause” is the Greek word for “stop”.

The stratosphere is the layer that lies above the troposphere. The temperature in the stratosphere stays constant approximately  $10\text{ km}$  above the tropopause, and this area is known as the isothermal layer. Temperature increases with height above the isothermal layer as a result of ozone’s absorption of UV light ( $O_3$ ).

At an altitude of around 85 kilometers, the stratopause marks the beginning of the mesosphere. At this level, the air pressure is approximately 1–0.1 Pa. The temperature in the mesosphere decreases with altitude. At the mesopause, the temperature is about  $-84^{\circ}\text{C}$ . The atmosphere's composition is rather homogeneous up to mesospheres from the surface. The term “homosphere” refers to the whole atmospheric layer that contains the troposphere, the stratosphere, and the mesosphere.

The “heterosphere” is the region above the “homosphere” in which the composition of the atmospheric gases is not uniform. The thermosphere is the layer of the atmosphere between approximately 85 and 500 km in height that contains the ionosphere. The thermosphere's temperature rises with altitude. The absorption of solar radiation by  $\text{N}_2$  and  $\text{O}_2$  causes the temperature characteristics in the region. The thermopause is the area that separates the thermosphere and the exosphere. The height of the thermopause varies depending on the amount of UV radiation coming from the Sun. After passing over the thermopause, the atmosphere returns to its normal, constant temperature.

The exosphere is the layer of the atmosphere that extends beyond the thermopause and into interplanetary space. In contrast to the other levels, which are delineated by the temperature profile, the exosphere is the layer in which collisions between molecules are so uncommon that they are, for the most part, able to be ignored. The exobase is the lowest point in the exosphere and is located very close to where the thermopause is located.

Molecular nitrogen ( $\text{N}_2$ ) and molecular oxygen( $\text{O}_2$ ) dominate the lower and middle thermosphere. Meanwhile, in the upper thermosphere, due to the impact of photodissociation and molecular diffusion at high altitude, atomic oxygen (O) is the major gas (Solomon & Roble, 2015).

## 2.4 Ionosphere

### 2.4.1 Structure of Ionosphere

The ionosphere is defined as the region of the atmosphere where free thermal (1 eV) electrons and ions are distributed in large numbers. The free elec-

trons and ions are formed from the ionization of neutral particles by strong ultraviolet light from the Sun and interactions with energetic particles. Once generated, charged particles are subject to a variety of processes, including chemical changes, diffusion, wave disruptions, plasma instabilities, and movement due to magnetic and electrical fields (R. Schunk & Nagy, 2009).

Due to comparatively lower air density in the upper ionosphere, charged particles do not recombine quickly following ionization, leads continuous ions and free electrons. The distribution of freely moving charged particles in the ionosphere is a combination of ionization and recombination, with ionizing radiation intensity and gas density being particularly important. In addition to photochemical ionization and recombination, neutral wind and electromagnetic fields also affect local electron concentrations (B. McElroy, 2012).

The ionosphere-thermosphere system is a very complex and dynamic medium that changes substantially. These changes are especially potent during geomagnetic storms and substorms. It is constantly hit by changing UV/EUV radiation from the Sun, and it transfer energy, mass and momentum with the lower atmosphere and magnetosphere. At high latitudes, the main ways that the ionosphere-thermosphere system gets energy and momentum are through plasma convection, Joule heating and particle precipitation (R. W. Schunk, 2014).

The free electron density, which is a measure of the degree of ionization, is one measurement that can be used to determine the structure of the ionosphere. The electron density continuity equation is crucial for comprehending ionospheric variability. It portrays the rate of change in density of electrons in the ionosphere over time at each altitude as

$$\frac{\partial N_s}{\partial t} = Q_s - L_s - d_s \quad (2.1)$$

Where  $Q_s$  represent rate of electron production,  $L_s$  represent rate of electron loss and  $d_s$  represents effects of transport processes.

The Figure 2.6 shows the heights that are occupied by the layers of the

ionosphere that go by the labels D (below 90 kilometers), E (between 90 and 130 kilometers), and F (above 130 km). It has been seen that the structure of the layer changes throughout the day and night, as well as at different height ranges, and that during the night, the D layer appears. F1 and F2 layer mix to form one layer F. In each successive layer, not only the electron density but also the various ion densities are distinct from one another (B. McElroy, 2012; Ratcliffe, 1972).

In the E region, the radiations with wave lengths less than 14 nm and between 80 and 102.7 nm that ionize  $N_2$ ,  $O_2$ , and  $O$  create the greatest number of electrons. In the F region, the radiations with wave lengths between 14 and 80 nm that ionize  $N_2$  and  $O$  produce the greatest number of electrons. The main electron production peak is located close to 150 km (F layer) and belongs to the ionization of  $O$  and  $N_2$ . At higher altitudes, electrons are produced at a rate proportional to the concentration of each gas (Ratcliffe, 1972).

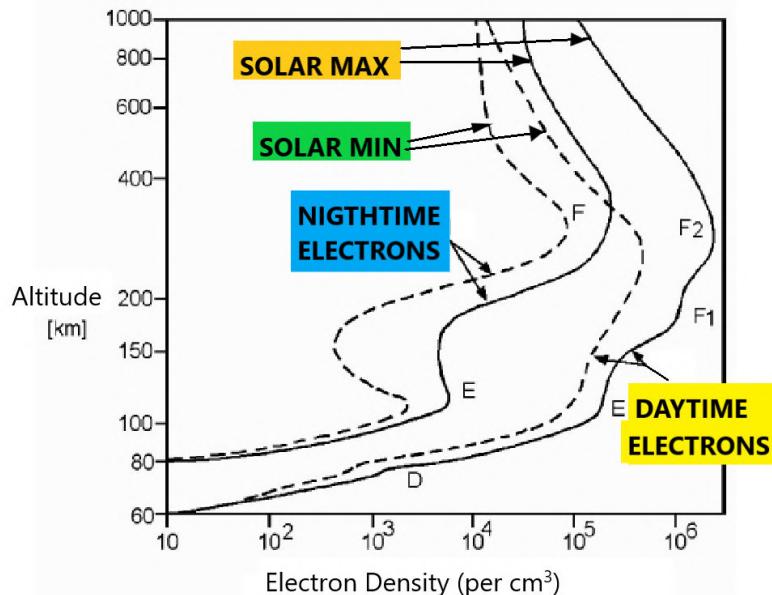


Figure 2.6: Electronic density based on altitude during the day, night, solar maximum, and solar minimum. Illustration of F2, F1, E and D layers of the ionosphere (Naval Postgraduate School, 2003).

The electron and ion density in the F-region is highly variable to altitude, latitude, longitude, and local time. Due to the dynamical behavior of the ionosphere, the spatial and temporal changes of ionosphere at mid latitudes are small in comparison to those at the equator and low-latitudes. The dynamical structure of the ionosphere is significantly observable in the low latitude regions (Aggarwal et al., 2012). The equatorial ionization anomaly (EIA) and the equatorial electrojet (EEJ) are two ionospheric characteristics that are distinctive to geomagnetic low latitudes (Appleton, 1946). The EIA is distinguished by the presence of two plasma crests on edges of the magnetic equator, with the magnetic equator exhibiting a lower electron density. The equatorial plasma fountain and upward  $E \times B$  plasma drift is responsible for the generation of the EIA. This plasma fountain transports plasma from the magnetic equator to higher altitudes, where it diffuses along magnetic field lines to higher latitudes. Also plasma is removed from the region surrounding the equator as a result of the upward  $E \times B$  drift resulting in two ionization crests on either side of the magnetic equator (Aggarwal et al., 2012; Balan et al., 2018). The Equatorial Electrojet is enhanced by the eastward electrostatic field produced by the E region dynamo. The EEJ is a diurnal small band of enhanced eastward current moving within the magnetic equator's  $\pm 3^\circ$  latitude range at 100-200 Km altitude (Acharya et al., 2010).

The winter anomaly, which is defined as a larger F2 layer electron density in the winter hemisphere than in the summer hemisphere during the solstices, is another significant characteristics of the ionosphere. It is primarily caused by shifts in the winter to summer proportion of  $O/N_2$ , as well as changes in the recombination rate as the solar cycle's temperature changes (Burns et al., 2014).

### 2.4.2 Ionosphere - Chemical process

Solar extreme ultraviolet (EUV) emissions that originate in the chromosphere and corona are the primary contributors to the formation of the ionosphere. There is some degree of dissynchronicity in the development of these emissions. The ionosphere is regulated by a variety of chemical, dynamical, and electrodynamical processes, all of which are modified to varying degrees by the activity of the Sun (Liu et al., 2011). The ions that are made then react chemically with the neutrals, recombine with the electrons, spread out to higher or lower altitudes, or are moved by the effects of the neutral wind.

But the Earth's magnetic field, which is dipolar at ionospheric altitude, has a big effect on the way particles move and spread out (R. Schunk & Nagy, 2009).

In order to keep the ionosphere, which is a combination of generating free electrons and their removal, the upper atmosphere is home to a number of chemical reactions that take place constantly. The ionosphere's composition may also change as a result of intermediate reactions like photodissociation and charge-exchange reactions (R. Schunk & Nagy, 2009). In each of the generic reactions that follow, the  $\gamma$  represent a photon of electromagnetic energy (UV, EUV, or X-ray), the uppercase letters symbolize a molecule, and  $e^-$  represent an electron.

Electron production reaction:



The primary cause of ionization in the majority of ionospheres is the photoionization, which results in the production of free electron-ion pairs. When the energy of an ionizing photon is greater than a certain threshold value of the gas particle, the excess energy is converted either into the kinetic energy of the electrons or into the excitation of the resultant ion. The maximum production rate occurs when the Sun is above, and it rises as the solar zenith angle decreases (R. Schunk & Nagy, 2009).

Intermediate reactions:

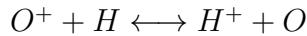


Atmospheric molecules may change to excited states and then undergo photodissociation as a result of photon absorption in the visible and ultraviolet regions of the spectrum. Molecular oxygen ( $O_2$ ) absorbs the majority of the UV energy from the Sun and photodissociates into two oxygen atoms ( $O$ ). The lower ionosphere's photochemistry highly depends on photodissociation.

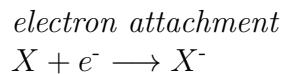
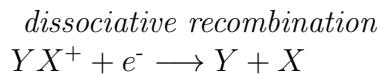
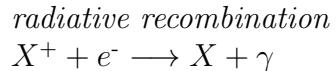
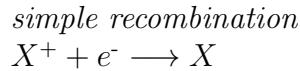
Dissociative recombination and collisional dissociation of oxygen molecules are two other events that may lead to the production of atoms in various states. These reactions take place when oxygen molecules come into contact with rapidly moving charged particles in the atmosphere (Golubkov et al., 2020).

Around 95 km in altitude is typically where one will find the highest  $O$  concentration. At a distance of less than 90 kilometers, the influence of quickly reacting hydrogen species starts to become substantial for the loss of odd oxygen (Osepian et al., 2008).

Reversible reactions are reactions that can take place in either direction; the charge exchange reaction, which is described below, is an example of a reversible reaction (R. Schunk & Nagy, 2009).



Electron loss reactions:



An illustration of electron loss reaction can be seen in the dissociative recombination of  $O_2^+$  with an electron.



Dissociative recombination, which occurs when an electron comes into contact with a molecular ion, results in the formation of an unstable neutral molecule. This process is far more efficient. It takes only a very short period of time for the atoms that make up this complex to travel away a sufficient distance to make it unlikely that the free-electron state will be restored. Therefore, this kind of recombination has the potential to be exceedingly effective (Donahue, 1968).

### 2.4.3 Solar variability effects in the ionosphere

The continuously changing levels of solar radiation that happen over the course of a solar cycle contribute to the extremely dynamic nature of the thermosphere-ionosphere system. The response of ionosphere to solar radiations has been investigated using ground-based measurements, space-based observations, numerical models, and empirical models (see Afraimovich et al. (2008), Jakowski et al. (1991), Ren et al. (2018), and Schmöller et al. (2018)). These findings show an ionosphere variation delay of one to two days. This delay may be owing to the delayed changes in atomic oxygen density caused by  $O_2$  photodissociation. According to research conducted by Vaishnav et al. (2018), transport processes could play an important role in the ionosphere delay.

Similar research was conducted by Vaishnav et al. (2019), which investigates long-term variations in the ionospheric response to solar EUV. The study utilizes 18 years (1999-2017) of global mean TEC and various solar EUV proxies to examine the temporal and spatial characteristics of the ionosphere to solar activity. According to this study, since solar activity was lower during solar cycle 24 than it was during solar cycle 23, the zonal mean TEC characteristic is greatly influenced by solar activity. The photoionization of atomic and molecular neutrals caused by solar EUV radiation, together with recombination at various altitudes and solar zenith angles, determines the majority of the free electron concentration in the ionosphere. A stronger correlation can be detected between the variables nearer to the equator (Vaishnav et al., 2019).

The ionosphere is affected by the solar wind by electrodynamic and hydro-magnetic drift, which are brought by solar wind pressure and friction on the geomagnetic cavity lying above the ionosphere. The Earth's rotation and the solar quiet ( $S_q$ ) both contribute to other magnetospheric movements (Piddington, 1964). The collision between the solar wind and the magnetosphere of the Earth creates a window through which the energy of the solar wind may enter into the magnetosphere. A portion of the energy is transferred into the ionosphere as electromagnetic energy and energetic particle precipitation. Because of the relative flow between both plasma and the neutrals, the majority of the electromagnetic energy is transformed into Joule heating in the ionosphere (Cai et al., 2014).

Vaishnav et al. (2019) contains investigation of long-term variability in the ionosphere response to solar EUV variations. The study examines the temporal and spatial response of the ionosphere to solar activity using total electron content (TEC) and 12 solar proxies, including F10.7 solar flux. Also, the study was conducted using data from 1999 to 2017, a duration of 18 years. The various solar proxies are analyzed using cross-wavelet and Lomb-Scargle periodogram (LSP) techniques to see how they affect the global mean TEC (GTEC). This study shows that the characteristics of the zonal mean TEC is highly dependent on solar activity. The photoionization of atomic and molecular neutrals caused by solar EUV radiation, together with recombination at various heights and solar zenith angles, determines the majority of the free electron concentration in the ionosphere (Vaishnav et al., 2019).

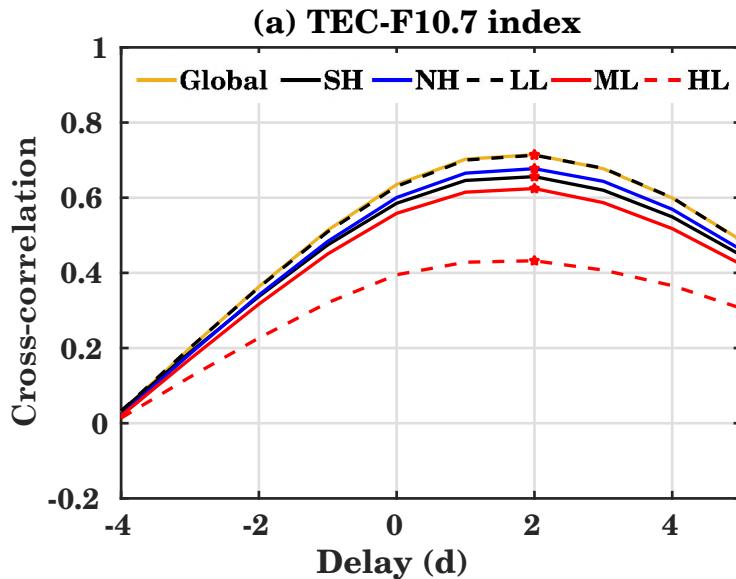


Figure 2.7: Cross-correlation coefficients and time delays between the global, Northern Hemisphere (NH), and Southern Hemisphere (SH) as well as low-latitude (LL,  $\pm 30^\circ$ ), midlatitude (ML,  $\pm (30\text{--}60^\circ)$ ), and high-latitude (HL,  $\pm (60\text{--}90^\circ)$ ) TEC with F10.7 during the years 1999 to 2017 for a different lag. A positive lag means that solar flux variations are heading TEC ones. The maximum correlation is indicated by a red star (Vaishnav et al., 2019).

Figure 2.8 depicts the cross-correlation coefficients and time delays between

F10.7 solar flux and global, Southern Hemisphere (SH), and Northern Hemisphere (NH), as well as low-latitude (LL,  $\pm 30^\circ$ ), midlatitude (ML,  $\pm (30\text{--}60^\circ)$ ), and high-latitude (HL,  $\pm (60\text{--}90^\circ)$ ) TEC which are indicated using the colors yellow, black, blue, dotted black, red and dotted red lines respectively. The highest correlation is observed for both GTEC and LL TEC, with correlation values of around 0.7 over a time delay of approximately 2 days. HL has observed the lowest correlation, with a maximum correlation value of 0.42 with a time delay of about 2 days (Vaishnav et al., 2019).

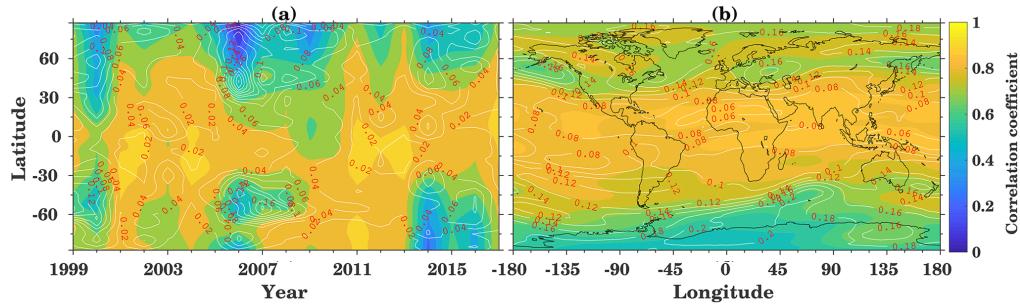


Figure 2.8: (a) Zonal mean (b) long term mean correlation coefficient between TEC and Mg II index for the years 1999 to 2017. The white contour lines indicate the respective standard deviations (Vaishnav et al., 2019).

Figure 2.8 illustrates zonal mean and long term mean correlation between TEC and Mg II index for the years 1999 to 2017. X axis of the Figure 2.8 (a) represent year and Y axis represent latitudes while the Y axis of the Figure 2.8 (b) represent longitude. The correlation values are indicated using a color gradient. At low latitudes, a maximum correlation of around 0.9 is seen during periods of high solar activity. Figures 10a and 10b demonstrate that the correlation drops as latitude increases. The fact that particle precipitation also varies with solar wind dynamics presumably describes why the correlation in the NH is lowest.

# Chapter 3

## Observations and Solar Parameters

### 3.1 F10.7 Solar Flux

The F10.7 solar flux index is a measurement of the microwave radiation emitted by the Sun at a wavelength of  $10.7\text{cm}$ , or  $2800\text{MHz}$ . As an indicator of solar activity, it is commonly used in many models of thermosphere and ionosphere. There is a strong connection between these microwave emissions and other solar emissions such as X-rays, extreme ultraviolet (EUV), and ultraviolet (UV) (Taylor et al., 2021).

The F10.7 cm solar flux is measured in solar flux unit (sfu), where 1 sfu is equivalent to  $10^{-22} \text{ W m}^{-2} \text{ Hz}^{-1}$ . Two small radio telescopes, called flux monitors, are used to measure the flux (Tapping, 2013). This 10.7 cm solar flux database has two parts: readings of the 10.7cm Flux and daily archives of the output of the flux monitor. For each readings of the 10.7cm Solar Flux, three values are given: the observed value, the adjusted value, and the URSI Series D value. When studying the Sun, it is not beneficial that the Earth's distance from the Sun changes every year, which changes the 10.7cm Solar Flux. But the distance between the Sun and the Earth comes out of the calculations for the ephemeris that are needed for the solar flux monitors to find and follow the Sun. So, we make an extra number that is adjusted

for changes in the distance between the Earth and the Sun and given for the average distance. This number is known as the “adjusted value” (Natural Resources Canada, 2021). This adjust value from the data source is used in this study for further analysis.

Figure 3.1 presents the timeseries data of adjusted value of F10.7 solar flux from 1998 to 2020. The X axis indicates the time in days from 1998 to 2020, and the Y axis includes the solar flux in SFU. Solar maximum is distinguished by a 3 year moving average of F10.7 solar flux values more than 100 sfu, and solar minimum is described by values less than 100 sfu through out the analysis. According to the three-year moving average, January 2005 to March 2011 and April 2016 to December 2020 are solar minimum periods, whereas January 1998 to January 2005 and March 2011 to April 2016 are solar maximum periods. It is evident from the F10.7 solar flux graph that the solar maximum of solar cycle 23 is stronger than the solar maximum of solar cycle 24.

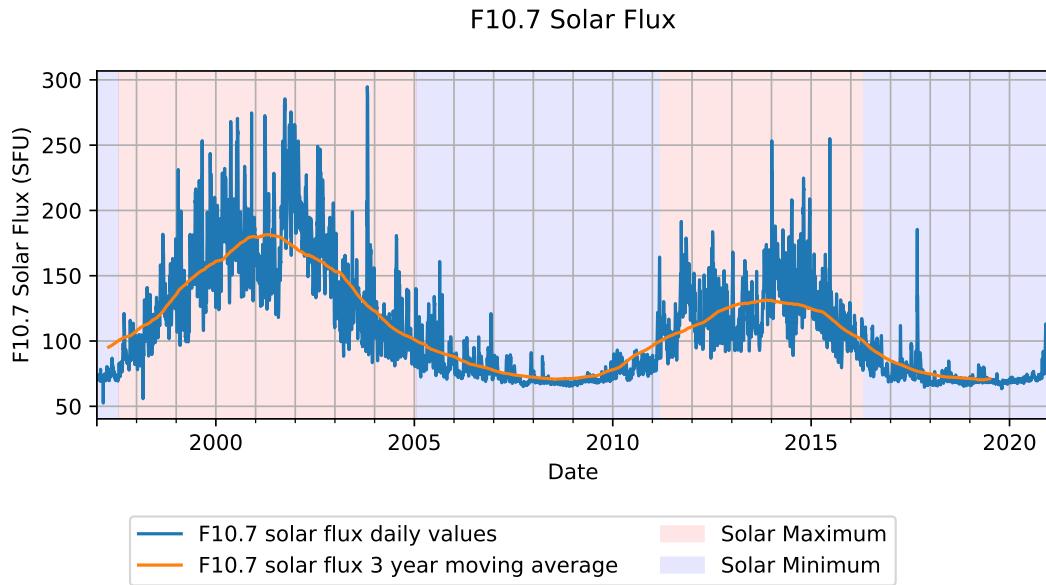


Figure 3.1: F10.7 Solar flux values from 1998 to 2020. The blue line represents the daily adjusted F10.7 solar flux readings provided by Dominion Radio Astrophysical Observatory and Natural Resources Canada, while the orange line represents the three-year moving average value. Red and blue shading represent the solar maximum and solar minimum, respectively.

## 3.2 Solar Wind Speed

The Solar Wind Electron Proton Alpha Monitor (SWEPAM) instrument is responsible for providing the Advanced Composition Explorer (ACE) with the majority of the solar wind information. This measurements offer direct examination of several solar wind events, like interplanetary shocks, coronal mass ejections, rates of solar wind plasma electron and ion fluxes flow, and solar wind structure, using modern, three-dimensional plasma instrumentation. They also provide a perfect data set for multi-spacecraft studies of the heliosphere and magnetosphere, in which they may be utilized in combination with other parallel investigations from spacecraft like Ulysses. All those data offer minute-by-minute insight into the environments of the solar wind as well as the solar system's internal state. SWEPAM level 2 data consists of ion data and electron data which include proton density, proton temper-

ature, proton speed in Geocentric Solar Ecliptic (GSE), Radial Tangential Normal (RTN) and eocentric Solar Magnetospheric (GSM) coordinates. In this analysis, the solar wind speed is represented by the proton speed ( $V_p$ ), which is given in kilometers per second (McComas et al., 1998).

Figure 3.2 presents the timeseries data of solar wind speed from 1998 to 2020. The X axis indicates the time in days from 1998 to 2020, and the Y axis includes the speed of solar wind in  $km/s$ . The blue line indicates the ACE spacecraft's SWEPPAM daily averaged solar wind speed data, while the orange line indicates the three-year moving average value. Red shading represents the solar maximum, while blue shading represents the solar minimum. The 3-year moving average illustrates that the solar wind speed increases during solar maximum and drops during solar minimum.

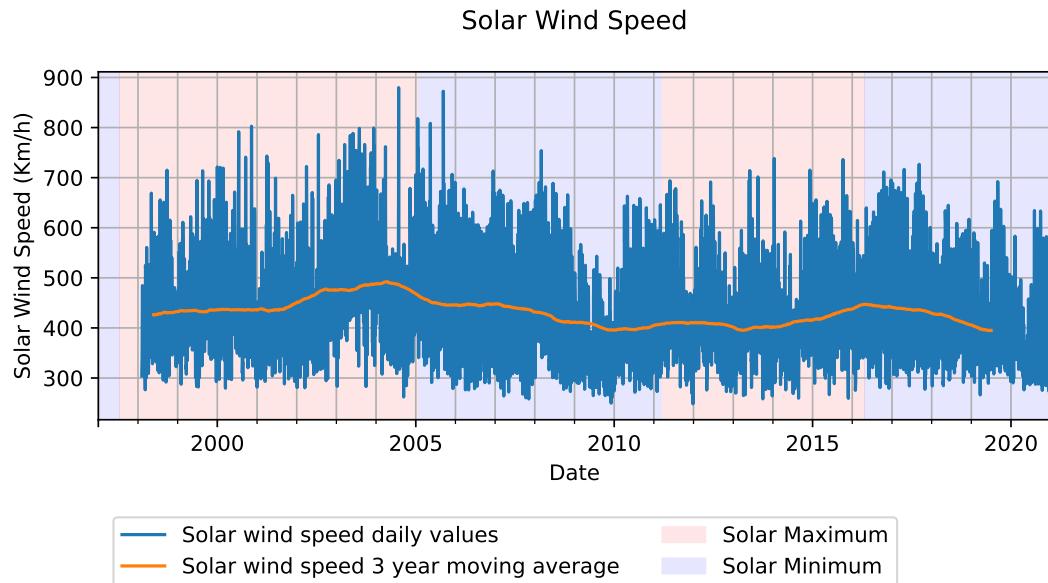


Figure 3.2: Solar Wind Speed values from 1998 to 2020. The blue line represent the SWEPPAM daily averaged solar wind speed data from ACE spacecraft, while the orange line represents the three-year moving average value. Red and blue shading represent the solar maximum and solar minimum, respectively.

### 3.3 Total Electron Content (TEC)

The total electron content, known as TEC, is an indicator that is used to quantify the status of the Earth's ionosphere, which has an effect on the electromagnetic waves that travel through it. The study of ionospheric structure requires the measurement of the TEC of the ionosphere.

TEC can be described as the line integral of the electron density along a ray-path.

$$TEC = \int_L N_e(l) dl \quad (3.1)$$

Where  $N_e$  refers to the distribution of the electron density. TEC is measured in units called TECU, where  $1\text{TECU} = 10^{16}\text{el/m}^2$ . Due to the high fluctuation of the ionosphere, electron density distribution and TEC are extremely dynamic in both space and time. Modelling of TEC one of the most important roles in detailed investigation and analysis of the near Earth space science and space weather (Nayir et al., 2007).

Over the past several years, dual frequency signals from the Global Positioning System (GPS) have been popular for estimating global TEC values. The presence of a great network of GPS satellites located at an altitude of 20,000 kilometers, worldwide coverage, and receivers that are accessible commercially, all reliable to use GPS signals for TEC estimation. Because of the sufficient high frequency signals that are utilized in the GPS system, these signals are only marginally influenced by the ionospheric absorption and the magnetic field of the Earth. The time delay caused by the signal's propagation to Earth-based receivers can be used to calculate TEC. GPS satellites send two coherent frequencies at  $f_1 = 1575.4\text{ MHz}$  ( $L1$ ) and  $f_2 = 1227.6\text{ MHz}$  ( $L2$ ) in the L-band. The  $L_1$  frequency contains a Coarse/Acquisition code (C/A) which is available for public use with an effective wavelength of 300m. It also has a restricted precision code (P or Y) with a wavelength of about 30m that is usually used for the military purposes. In addition to these, navigation data is also embedded in the signal (Jakowski et al., 1996).

To compute the distance between the satellite and receiver, the time-delay of the pseudorandom-noise code sequences of the received signal is measured. The calculated ranges are called pseudoranges  $p$ , which are expressed as:

$$p = \rho + c(dt - dT) + d_I + d_T + d_{MP} + dq + dQ + \varepsilon_p \quad (3.2)$$

where  $\rho$  is the geometric distance between satellite and receiver,  $c$  represents speed of light in vacuum,  $dt$  and  $dT$  stands for offsets of satellite and receiver clock,  $d_I$  and  $d_T$  are the ionospheric and tropospheric delay,  $d_{MP}$  is the multipath error, satellite and receiver instrumental group delay is represented as  $dq$  and  $dQ$ , a random error is represented as  $\varepsilon_p$ .

When considering the ionospheric plasma's refractive index at L-band frequencies, the ionospheric delay  $d_I$  can be expressed as:

$$d_I = \frac{K}{f^2} \int n_e ds \quad (3.3)$$

where  $n_e$  represents the density of electrons along the path  $s$  and  $K = 40.3 m^3 s^{-1}$ .

Substituting equation 3.1 in equation 3.3 gives:

$$d_I = \frac{40.3}{f^2} TEC \quad (3.4)$$

Difference between code pseudo range for two different frequencies  $f_1$  and  $f_2$  gives:

$$P_1 - P_2 = d_{I1} - d_{I2} \quad (3.5)$$

The difference in ionospheric delay of L1 and L2 of the GPS frequencies can be calculated from the GPS signal.

The sender receiver signals are commonly assumed to take the identical path through the ionosphere. As a result (substituting equation 3.4 in equation 3.5), the group delay is then calculated as:

$$P_1 - P_2 = 40.3 TEC \left( \frac{1}{f_2^2} - \frac{1}{f_1^2} \right) \quad (3.6)$$

In which  $P_1$  and  $P_2$  are the group pathway lengths that correspond to the GPS frequencies  $f_1$  and  $f_2$ , respectively.

then TEC can be expressed as:

$$TEC = \frac{1}{40.3} \left( \frac{f_1 f_2}{f_1 - f_2} \right) (P_2 - P_1) \quad (3.7)$$

TEC can be classified into two types Vertical TEC (TECv) and Slant TEC (TECs). The TEC of the ionosphere along the path between the satellite and the receiver is defined by slant TEC. Even though TECs is observed at various elevation angles, TECv is often modelled instead. The mapping of TEC on the Earth's surface is made possible by TECv (Ya'acob et al., 2010).

Three timeseries of zonal mean TEC values at different latitude from 1998 to 2020 are shown in Figure 3.3. This has been taken into account for the analysis. The figure contains data from latitudes 60N, 30N and 0N latitudes. The X axis is the time (days from 1998 to 2020) and Y axis contains TEC values in TEC unit. Red and blue shading represent the solar maximum and solar minimum, respectively. The zonal mean TEC value indicates that it is increasing as solar activity increases. A 3 year mean of the zonal mean TEC is indicated using orange line.

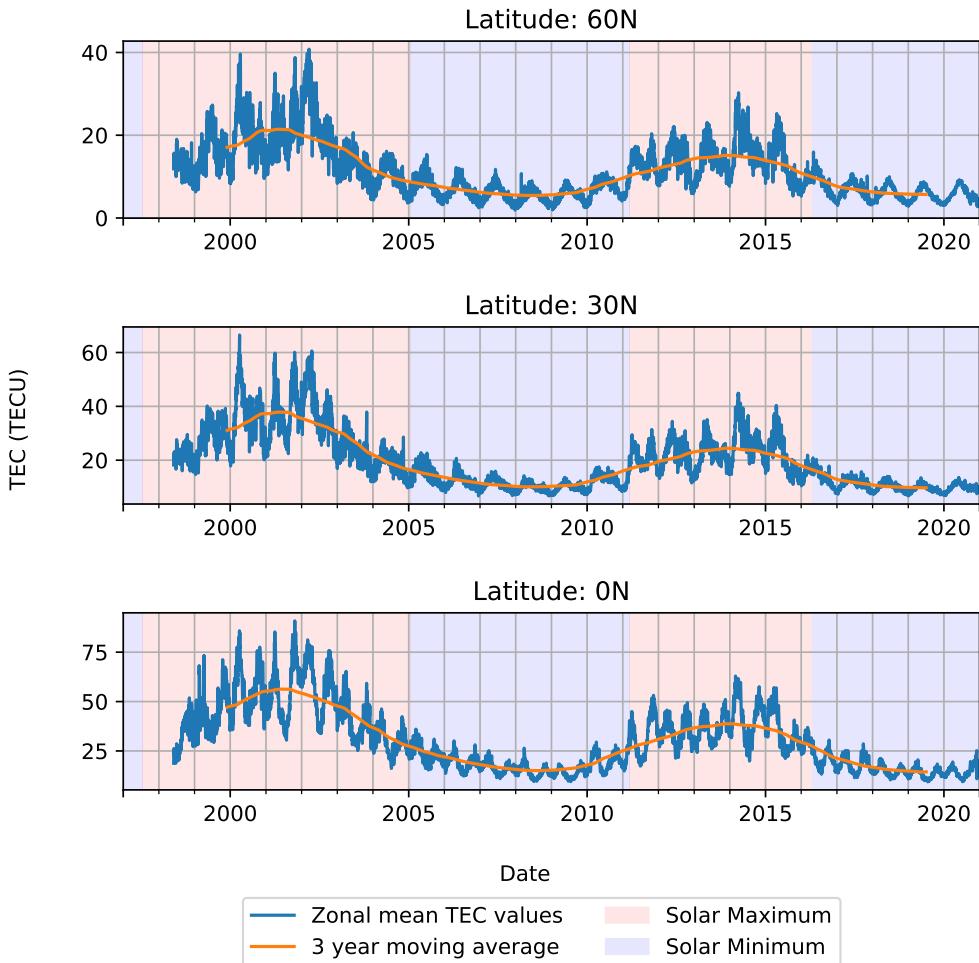


Figure 3.3: Zonal mean TEC values in the ionosphere at latitude 60N, 30N and 0N from 1998 to 2020. Red and blue shading represent the solar maximum and solar minimum, respectively.

### 3.4 IGS TEC maps

In June 1998, the Ionosphere Working Group started developing Ionosphere Vertical Total Electron Content maps in association with five IGS Ionosphere Associate Analysis Centers (IAACs). The analysis centers are ESOC (*European Space Operations Center of ESA, Darmstadt, Germany*), NRCan/EMR

(Natural Resources Canada, Ottawa, Ontario, Canada), JPL (Jet Propulsion Laboratory, Pasadena, California, U.S.A), UPC (Technical University of Catalonia, Barcelona, Spain), and CODE (Center for Orbit Determination in Europe Astronomical Institute, University of Berne, Switzerland). There are several ways in which the data they provide may be put to use in the domains of science and technology, including the prediction of space weather events using empirical models and the enhancement of navigation systems for end users. IGS Working Group produces reliable products like global VTEC maps that are produced independently by multiple analysis centers, rated, and integrated with the appropriate weight to distinct analysis centers (Hernández-Pajares et al., 2009; Pajares, 2003).

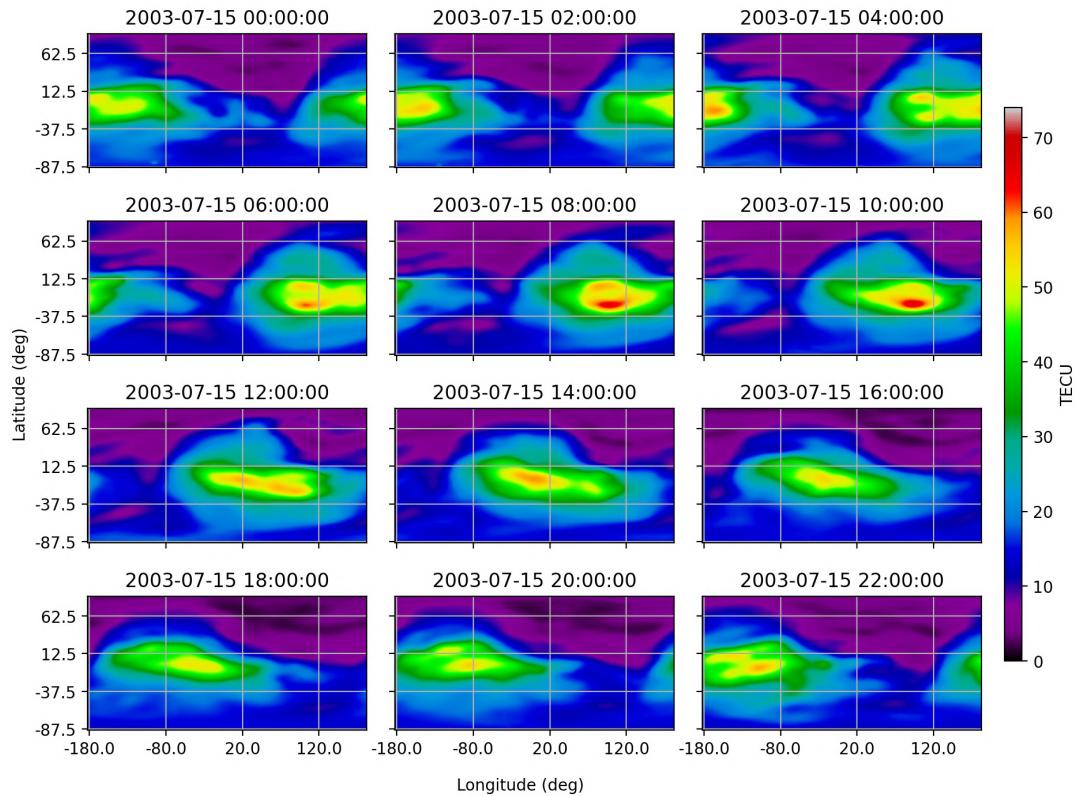


Figure 3.4: A 2-hour resolution TEC map for 2003-07-15 created using data from the IGS Ionosphere Group.

Figure 3.4 is an example IGS TEC map. The graph was created using data from the 2003-07-15. The X axis represents longitude, whereas the Y axis represents latitude. Color gradients are used to illustrate TEC readings. Each graph in Figure 3.4 reflects a 2-hour resolution. In terms of geographic longitude and latitude, data set has a resolution of  $5^{\circ}$  and  $2.5^{\circ}$  respectively ( $73 \times 71$ ). According to the figure, the value of the TEC in the equator is greater than it is in the high latitude regions. The high concentration of TEC values can be seen moving along the magnetic equator. The effect of EIA is visible in the graph at 6:00 UT and 8:00 UT as two crests of high concentrations of TEC values at low latitude regions. Additionally, the TEC values are greater during the daytime hours compared to the nighttime hours.

The Crustal Dynamics Data Information System (CDDIS) of NASA, which is responsible for the distribution of the data on behalf of the IGS Iono Group, was the source from which the data were obtained for this analysis. The data is offered on a daily basis in files that are of the Common Data Format (CDF) format. Each file contains the data collected by all five of the analysis facilities. The data source includes information with three distinct levels of temporal resolution, namely 15 minutes, 1 hour, and 2 hours.

# **Chapter 4**

## **Analysis**

### **4.1 Data Collection and Pre-processing**

Accuracy of the analysis performed is highly depended on the quality of data. Raw data consist of data from different sources, duplicates, corrupted data or incorrect data. In this step raw data files are read from the specified location and making it ready for operations. We make the data easier to analyze and use by subjecting it to some preliminary processing. This process removes gaps or duplicates in the data, which would otherwise have the potential to negatively impact the accuracy of the model. Additionally, the pre-processing of data ensures that there are no incorrect or missing values as a result of human error or bugs in the system.

Time series data contain a huge amount of information, but it is typically hidden until it is extracted using different analytics methods. Incomplete timestamps, missing data values, outliers in the data, and noise in the data are typical issues encountered in time series data. The handling of incomplete data is the most complex of the issues described. The first step in getting a clean and continuous database is to check all three data sets that will be used for the analysis for these problems and then attempt to find solutions. Due to the fact that the date format of the various data sets is different, the date format of the time series must be defined at the beginning. Initially, the values were treated as if they were not defined, and later, values were

discovered through the use of interpolation algorithms to fill in possible gaps in the data.

The raw data that will be used for the analysis can be found in a variety of file formats, including comma-separated values (CSV), text files, and other forms that could potentially slow down the reading process. During this step of the procedure, the data files are read and then transformed to a binary file format known as the Hierarchical Data Format (HDF). This makes the data handling and computations go much more quickly.

The data set is organized in a table-like structure with a two-dimensional, labeled data structure containing columns of various data types. The F10.7 solar flux data and solar wind speed data is structured as a time series having frequency of daily values and the data is indexed with date. The TEC data is structured and indexed with date and time, latitude and longitude having a resolution of  $2.5^\circ$ ,  $5^\circ$  and 1 hour respectively.

When it comes to the data regarding the total electron content, the data can be obtained from five different analysis centers in three different levels of temporal resolution. These maps have a frame that is fixed to the Earth. The geographic longitude and latitude ranges between  $-180^\circ$  and  $180^\circ$  with  $5^\circ$  resolution and  $-87.5^\circ$  and  $87.5^\circ$  with a resolution of  $2.5^\circ$ , respectively. In order to carry out this analysis, we require data with a temporal resolution of one hour for the period being analyzed. Therefore, in order to fill missing data in the files with a temporal resolution of one hour, we also used data from files with a resolution of two hours and fifteen minutes. After that, the data points with a temporal resolution of one hour were averaged at each universal time, and finally, the daily average was calculated.

Interpolation is a method that is often used in the process of filling missing values in time series. It is helpful in estimating the data gaps by using the two data points that are surrounding it that are already known. This approach is simple and straightforward to comprehend. The utilization of imputation methods is one strategy that can be utilized to address issues regarding incomplete data (Glas, 2010). As a result, the method of linear interpolation was chosen for this research to replace data gaps.

In order to begin the analysis, a smoothing of data points was performed

on the three data sets (10.7cm Solar Flux, solar wind speed, and total electron content) by taking a 27-day moving average based on the sun's average rotation to reveal underlying trends. In addition to the moving average for the 27 days, relative difference is computed to determine the percentage of actual values to the 27 days mean for each data set. All computations for the analysis are carried out on the basis of the relative difference of the data set. The relative difference of the data set is computed in order to remove the long-term effects of the solar cycle. Equation 4.1 is used to compute the relative difference where,  $X$  denotes the recorded value of the parameters and  $X_{mean}$  denotes the 27 day moving average of the parameter.  $\Delta F_{10.7,rel}$ ,  $\Delta V_{SW,rel}$ ,  $\Delta TEC_{rel}$  are used to denote relative difference of F10.7 solar flux, solar wind speed and TEC values respectively.

$$\Delta X_{rel} = \frac{X - X_{mean}}{X_{mean}} \times 100 \quad (4.1)$$

The TEC values are then averaged further by taking into account each UT hour as well as latitude-wise independently. In order to obtain the TEC values required for future computations, we finally averaged the TEC values across each day. The workflow of the pre-processing phase of the software package is outlined in section 4.3.3.

For the local time dependency analysis, the solar wind speed and F10.7 solar flux data are available in daily resolution values. The data have been interpolated to hourly values in order to get smooth results.

## 4.2 Mathematical Background of the Correlation Analysis

Numerous researches perform correlation analysis to determine the degree of relationship between variables. Analysis of linear correlation is a method for depicting the proximity between two variables. Almost all studies employ correlation analysis for the same objective, which is to investigate the association between independent and dependent variables. For this reason, the variation in the magnitude of one parameter in correlated data is connected with a variation in the magnitude of another parameter, either in the same

direction or in the opposite direction. To put it another way, when one variable has greater values, this is typically related with either greater or lower values of the second parameter, and vice versa. Pearson's Product Moment Correlation Coefficient and Spearman's Rank Correlation Coefficient are the two most common correlation coefficients often used in applications. This study focuses on the applications of Pearson's Simple Linear Correlation to investigate the relationship between variables. Hauke and Kossowski (2011) agree that the Pearson's Product Moment Correlation Coefficient ( $R$  or  $r$ ) is a scale for figuring out how strongly two variables are linked linearly. As it measures how linearly two variables are related, interval or ratio variables should be taken into account, but only if the variables in study have a normal distribution (Senthilnathan, 2019).

Pearson's formula for quantifying the level of correlation ( $R$ ) between two variables, specifically  $X$  and  $Y$ , is as follows:

$$R = \frac{n(\sum XY) - (\sum X) * (\sum Y)}{\sqrt{n(\sum X^2) - (\sum X)^2} * \sqrt{n(\sum Y^2) - (\sum Y)^2}} \quad (4.2)$$

Where  $n$  is the number of observations,  $X$  is the measure of variable 1 and  $Y$  is the measure variable 2. Depending on the direction, the degree of correlation can be classified as positive, zero, or negative. In practice, perfect zero correlation coefficients between variables are uncommon; consequently, positive and negative correlations can be classified identically in studies.

If the pattern of a variable is positive and nearly identical to that of another variable, there is a possibility of positive association between the variables, and such an association can result in a positive correlation coefficient; conversely, if the trend of a variable is positive and nearly opposite to that of another variable, there is a possibility of negative association between the variables, and such an association can result in a negative correlation coefficient (Edwards, 1976; Gogtay & Thatte, 2017).

Fundamentally, the correlation coefficient  $R$  will vary between  $-1$  and  $+1$ , or  $-1 \leq R \leq +1$ . The correlation coefficient cannot be interpreted in a particular way. Figure 1 illustrates, according to Gogtay and Thatte (2017), how the correlation coefficient might be interpreted depending on its value.

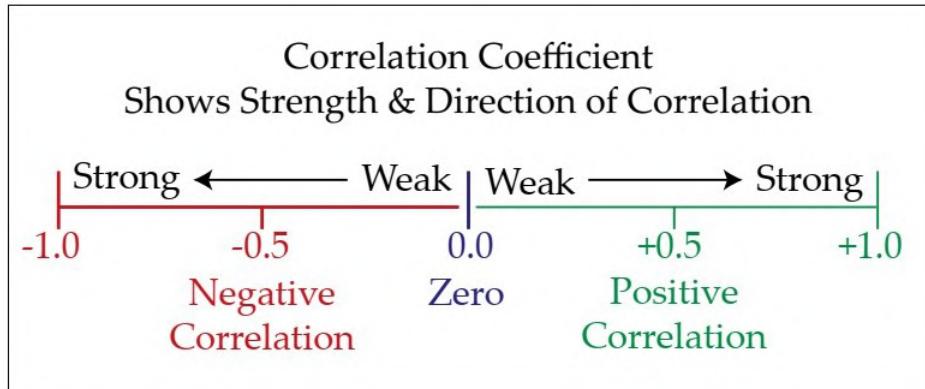


Figure 4.1: The spectrum of the correlation coefficient (-1 to +1). (Gogtay & Thatte, 2017)

Calculations of correlation are carried out using a box window that covers a period of 90 days. To study the seasonal effects of the correlation between TEC, F10.7 solar flux and solar wind speed, the window size should be small enough to detect the seasonal effects and large enough to have a sufficient number of values to correlate. The window size was checked by finding the mean absolute correlation between  $\Delta TEC_{rel}$ ,  $TEC$ ,  $\Delta V_{SW_{rel}}$ ,  $V_{SW}$ ,  $\Delta F_{10.7_{rel}}$ ,  $F_{10.7}$ . When looking at the correlation between TEC and F10.7 solar flux as well as TEC and solar wind speed, a window size of 90 days produces the optimum results. Figure 4.2 show the test result of mean absolute correlation between  $\Delta TEC_{rel}$ ,  $TEC$ ,  $\Delta V_{SW_{rel}}$ ,  $V_{SW}$ ,  $\Delta F_{10.7_{rel}}$ ,  $F_{10.7}$  with window size ranging from 27 days to 11 years (4015 days) at location 12.5N, 0E . Figure 4.2a depicts a window size test for the mean absolute correlation between  $\Delta TEC_{rel}$  and  $\Delta V_{SW_{rel}}$ ,  $V_{SW}$ ,  $\Delta F_{10.7_{rel}}$ ,  $F_{10.7}$ , whereas Figure 4.2b depicts a window size test for the correlation between  $TEC$  and  $V_{SW}$ ,  $\Delta F_{10.7_{rel}}$ ,  $F_{10.7}$ . The X axis shows window sizes, while the Y axis shows mean absolute correlation value. A red vertical line indicates 90 days window.

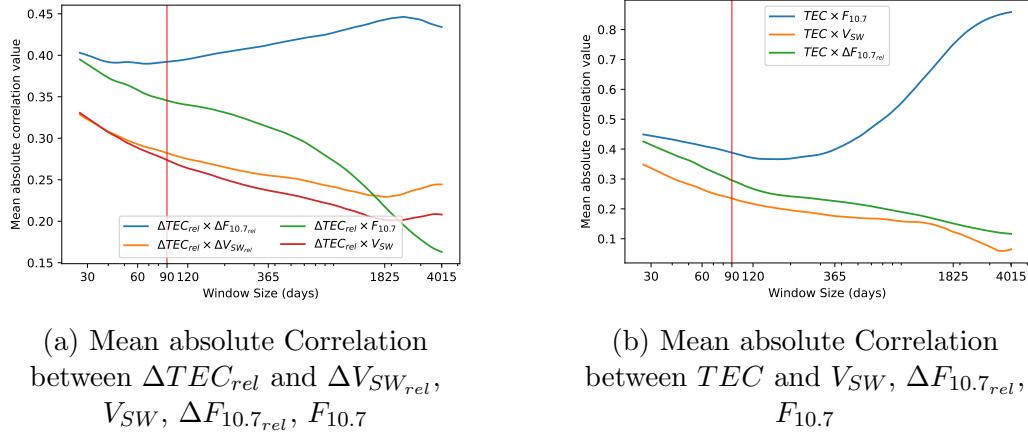


Figure 4.2: Window size test with mean absolute correlation between  $\Delta TEC_{rel}$ ,  $TEC$ ,  $\Delta V_{SW_{rel}}$ ,  $V_{SW}$ ,  $\Delta F_{10.7_{rel}}$ ,  $F_{10.7}$  at location 12.5N 0E. Red vertical line indicates 90 days window size.

## 4.3 Software Tool

### 4.3.1 Software Requirements and Design

Modern software tools are required to analyze large data-sets. The data set can be raw data or unstructured data, which has to be formatted according to the requirements of the analysis. Data scientists extract information from these data sets for data driven decision-making. The accuracy of data is critical because inaccurate data leads to faulty conclusions. Frequent use of the same analytics for multiple scientific researches brings the requirement of a software tool to save time and cost of the research.

Figure 4.3 shows a use-case diagram highlighting the key features of the system. The general users of the system, who are able to perform a variety of operations, is the actors in the system. A general user does not need to have any prior experience in programming. Users with no prior knowledge of Python may use the software system by configuring attributes that drive the product's functionalities in a configuration file. It was one of the important requirements when designing the software system. Another important requirement was about the data and data file types. Since the software pack-

age is accountable for managing large sets of data, it is necessary for it to take into account the file format, loading time, the structure of the data, and the computation complexity of the data sets. The structure of the data is also determined depending on the simplicity of segmenting and segregating the data as per preferences. The requirements for computation were to figure out how to manage multiple computation operations and to independently manage the flow of the computations and results. Each computing operation in the system has an input and an output, which makes it independent. The way in which the output of one operation is used as the input for the subsequent operation controls the flow of operations. When it comes to requirements for result visualization, the system should be able to generate various types of plots and adjust the attributes of the plots. The software system is separated into four distinct modules. Modularization makes it simple to add new features to existing software program.

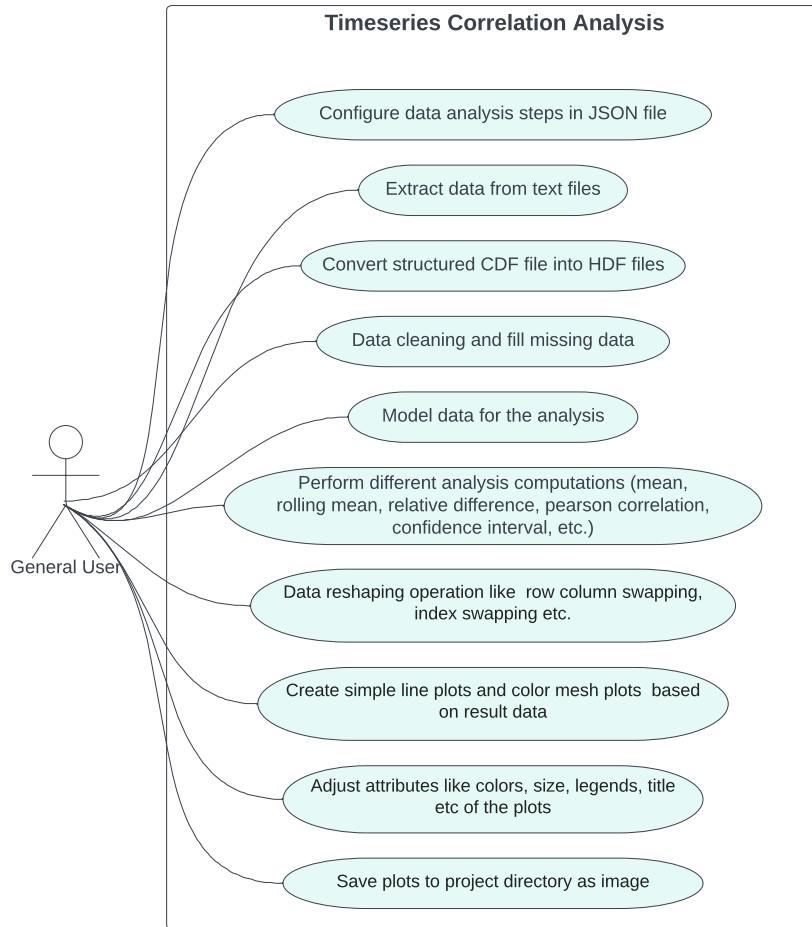


Figure 4.3: Use case diagram of the timeseries correlation analysis software system.

Setting up the project and loading the data (P1), data cleaning and formatting (P2), computations (P3), and visualization (P4) are the four separate processes that make up the analysis process (see figure 4.4). HDF file format serves as an interface between the various modules in the system. The process begins with setting up the project and loading the data. Each individual step in the process flow has its own data configurations to manage

the various data sets and operations that are required at that stage. The attributes in configuration files control how the program components function. The software can easily work with the configuration files because they are written in JSON format. There will be independent folder locations for each phase of the process, and the configuration files will be stored within these folders.



Figure 4.4: Process flow of the software package for the analysis.

### 4.3.2 Related Packages

A growing number of scientific applications, including those that have historically been dominated by MATLAB, R, Stata, and other research environments, are beginning to make use of Python. Python has become more user-friendly and approachable for a wide range of people as a direct result of improvements to its essential numerical libraries such as NumPy and SciPy, as well as improvements to the documentation's overall quality gives python more stability and dependability in the research community. In addition, the integration of `matplotlib` library creates an interactive environment for study and development, complete with a data visualization that is acceptable for the vast majority of users (McKinney et al., 2010). Python is utilized throughout the development process of this software package. Additionally, the software makes use of a variety of different packages, including `pandas` and `NumPy` for data management and computations, and the `matplotlib` library for data visualization.

The `pandas` package in Python handles data structure requirements to ensure that analysis functions efficiently. `DataFrame` in `pandas` is a two-dimensional data structure with columns that could be of different data types, similar to a spreadsheet or a SQL table. `DataFrame` makes it easier to work with large amounts of data in complex data structures. They are easier to query and take less time to process. `DataFrame` is a powerful, fast, and simple data

structure tool in Python as it is built on the `numpy` library. The `DataFrame` library is pre-packaged with a variety of operations that simplify the mathematical operations that are made available by the `math` library.

The processed data contains important information for the investigation, and visualization techniques make it simple to understand these important information. `Matplotlib` is one of the most popular libraries for data visualization in the Python programming language. `Matplotlib` support for a wide range of graphs, from simple line graphs to complex three-dimensional charts. In this investigation, we are attempting to visualize the data with both simple plot and colormesh. Line plots are used to represent individual latitude data, while colormesh is used to visualize temporal, seasonal, and local time dependencies.

### 4.3.3 Software Functionalities

The operations on the data have been designed as functions, which will make it much simpler to carry out a number of operations on the data by making use of the computation configuration file. The configuration is extendable with a variety of functions at every stage, beginning with the initial step of setting up the project and continuing through the step of data loading, cleaning, computations and visualization. The software is able to read data from a variety of file formats, including text files, files with commas separated values (csv), files in the CDF file format. After all files have been read, they are then transformed into HDF files so that subsequent read operations can take place more quickly.

It is necessary to have a clear understanding of the information structure of a text file, as well as its organization and the location of its storage, before one can successfully extract any data from within that file. The attributes of the data being processed can be obtained easily with the assistance of the metadata that is associated with the data files. It is information such as the structure of the data, date format, information about data values in the data file, among other things. The project uses the start position and length of the data value to pull information out of text files. This information will be written in the metadata file.

The software package has been designed so that it can be configured by

using a data file; more specifically, the software package accepts JSON files as input that explain what to do and how to set the software. Information such as data locations, data operations (including what should be done with the data), data types, and specific data properties for operations are included in the JSON configuration file.

Figure 4.5 depicts the workflow of the data loading module, where the configuration file specifies the location of the raw data, the frequency of the file, the file type, the start and end positions of the data, the analysis period, the attribute name, and the index value. The data is extracted from text and CSV files depending on their position. On the other hand, data in CDF format is more organized and simply requires conversion to HDF format for further processing.

The first module of the software is data load module (P1). The primary function of the module P1 is to input raw data and extract relevant information from whole data sets. It accepts input in a variety of data types, including text files, CSV files, and CDF files. By indicating the start and end positions of the data in the configuration JSON file, information from text files and CSV files is retrieved depending on position. Attributes in configuration files are formatted as an array of JSON objects, with each element pointing to a configuration for F10.7 solar flux, solar wind speed, and TEC data. The raw data for F10.7 solar flux and solar wind are in text file format. `DataFrame` contributes to the organization of the data, as was discussed in section 4.3.2. Each entry in the data format contains the solar flux value as well as the date. Thus, values as well as the timestamp from data files must be fetched based on their start and end positions. This stage also includes the primary indexing of the data that was retrieved. When retrieving data, indexing is managed by setting an attribute in configuration labeled ‘`isIndex`’ to true. All retrieved data is then formatted into a table and stored as an HDF file for use in other modules.

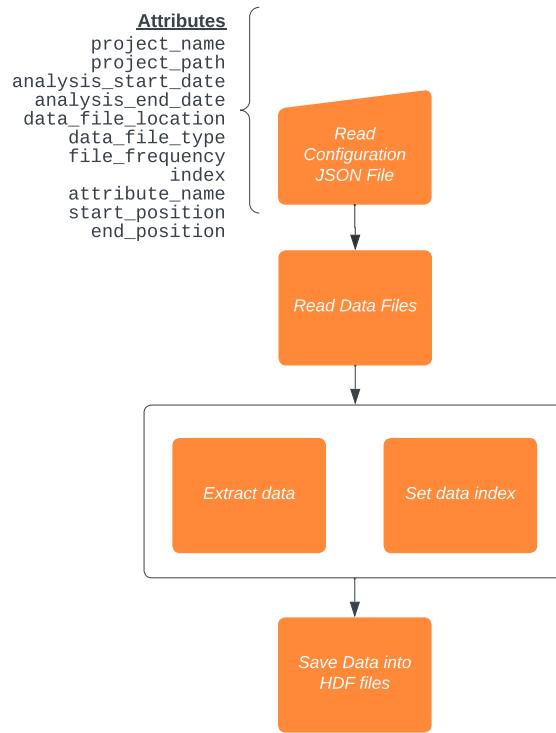


Figure 4.5: Detailed workflow of data load module (P1) of the timeseries analysis software

In the data cleaning and formatting module, the data is filled in if there are any missing values by utilizing interpolation techniques or by pulling values from other files if they have been provided. This component of the software system takes input in HDF file format, which is further formatted and saved in the same format for usage in the other two modules. Simple interpolation is used to manage missing values in F10.7 solar flux and solar wind data. As explained in section 3.4, the TEC data-set contains data from five distinct analysis centers that are combined to fill in data gaps. It also checks to make sure there are no missing time indexes depending on the frequency of the time series. The data cleaning and formatting process flowchart is shown in Figure 4.6.

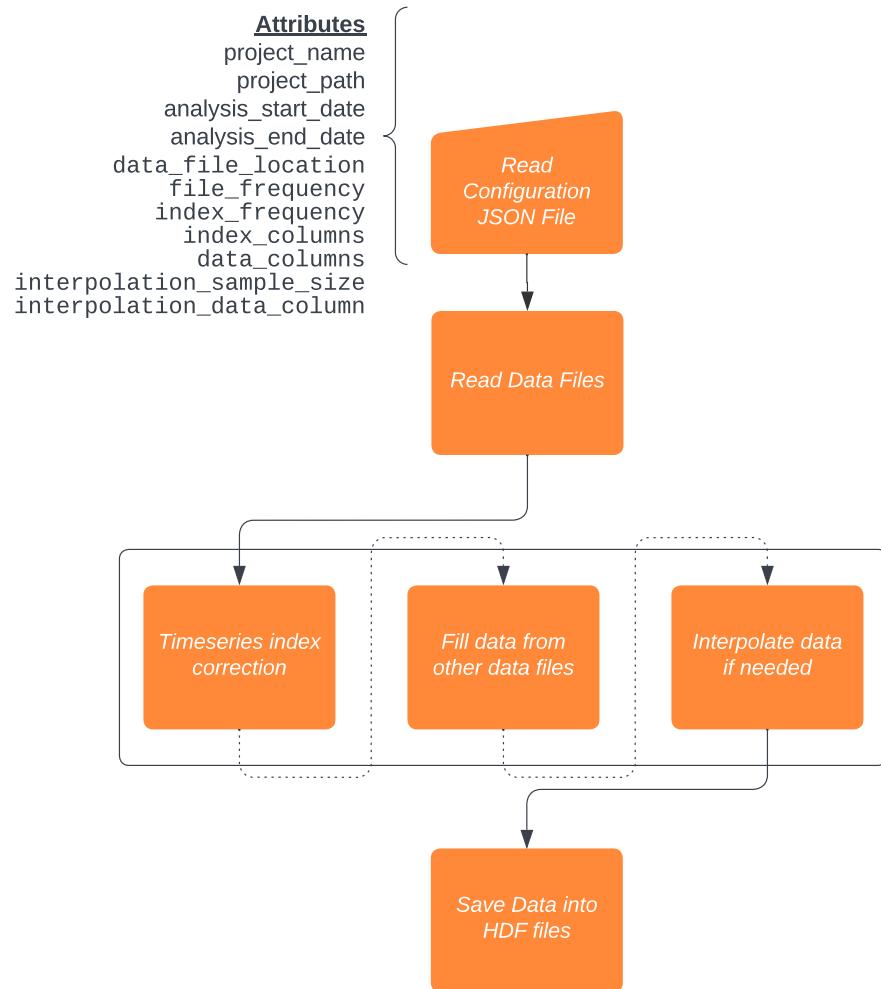


Figure 4.6: Detailed workflow of data cleaning & formatting module (P2) of the timeseries analysis software.

Calculating the Pearson correlation coefficient, determining the confidence interval, finding the mean, calculating the moving average, and calculating the relative difference are the primary computational tasks covered by the computation modules. In addition to these computation steps, there are other operations that can be performed, such as extracting data from a time

series that falls within a specific time period or columns, utilizing interpolation techniques, modifying the multi-index levels, or exchanging the column levels of the data. To get the final correlation result, a series of computations are performed. The computation's control flow is as follows: first compute the 27-day moving average, then compute the relative difference between recorded value and 27 day moving average of each data-sets. In the case of TEC data, each latitude's mean operation is carried out in order to get the zonal mean data of the relative difference TEC data. The data in the TEC data file has a temporal resolution of 1 hour and must be converted to daily average values by performing another mean operation on zonal mean TEC data. The data now contains daily average values for each latitudes. Pearson correlation is computed with daily values of relative differences of F10.7 solar flux and relative differences of daily average TEC values. The correlation between relative differences of solar wind speed and TEC is also computed, same as the correlation between F10.7 solar flux and TEC. In order to get seasonal variability, the data obtained from the correlation coefficient are further grouped and averaged depending on the day of the year. The upper and lower confidence interval can also be computed using the correlation coefficient results. Before computing correlation for local time dependency, basic interpolation is done since the TEC data with a resolution of 1 hour has some gaps in the data. It is possible to retrieve specific data, such as data for a certain time period or latitude, by performing operations such as extracting columns, swapping columns and indexing columns. The results of each computing operation performed by the module are returned in HDF file format and stored within the project directory. Figure 4.7 is a workflow diagram of the computation module.

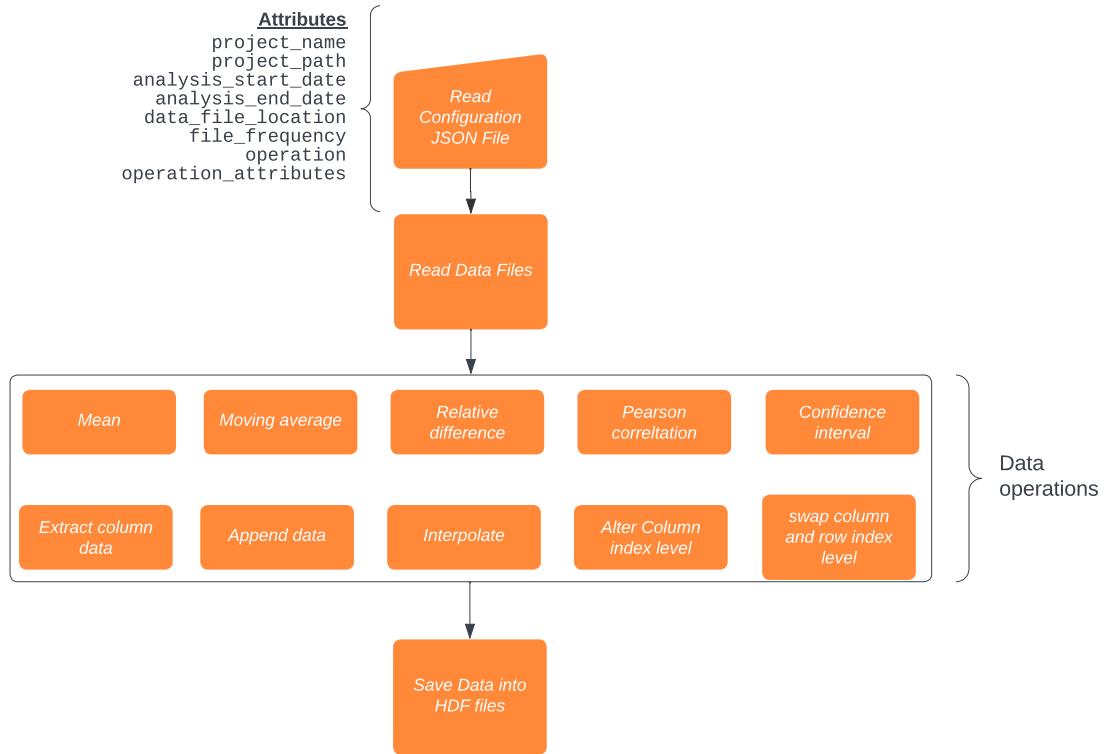


Figure 4.7: Workflow of computation module (P3) of the timeseries analysis software.

The software application is capable of generating line graphs and color mesh graphs in accordance with the specifications provided. The configuration file include the information that determines things like the type of graph, its colors, titles, labels, legends, and X & Y axes. The generated plot is then saved as an image in the project directory.

Figure 4.8 is a workflow diagram of the visualisation module. The plots are made by loading the results from the computation module into the visualization module. Line graphs are used to show results like a specific latitude, while 2D color mesh plots are used to show how the correlation analysis varies with the seasons and latitude.

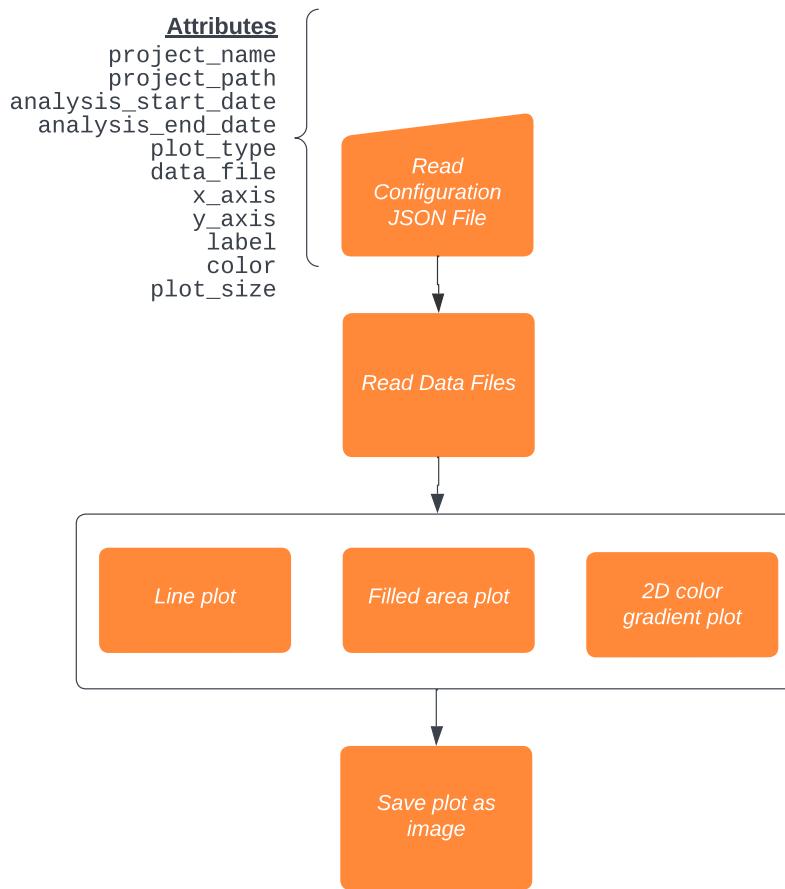


Figure 4.8: Detailed workflow of visualisation module (P4) of the timeseries analysis software.

# Chapter 5

## Results

### 5.1 Correlation between TEC and F10.7 Solar Flux

#### 5.1.1 Temporal changes of the Correlation between TEC and F10.7 Solar Flux

The results of a correlation between  $\Delta F_{10.7rel}$  and  $\Delta TEC_{rel}$  at 12.5N latitude is displayed in Figure 5.1. This correlation was performed with a 90-day window from 1998 to 2020. The years 1998 through 2020 are plotted along the X axis, while along the Y axis are correlation coefficient values ranging from -1 to +1. On top of the correlation graph, a plot of the 3 year moving average of F10.7 solar flux is displayed so that information about the activity of the Sun during the period of time can be seen. On the graph of the correlation coefficient, the red line represents the correlation coefficient using a box window of 90 days. A green line represents the 3 year moving average of the correlation coefficients. The confidence interval for the computation of the correlation coefficient at the 90-day box window is indicated by the use of gray shading. The correlation values indicate significant fluctuation over the analysis period, as seen in Figure 5.1.

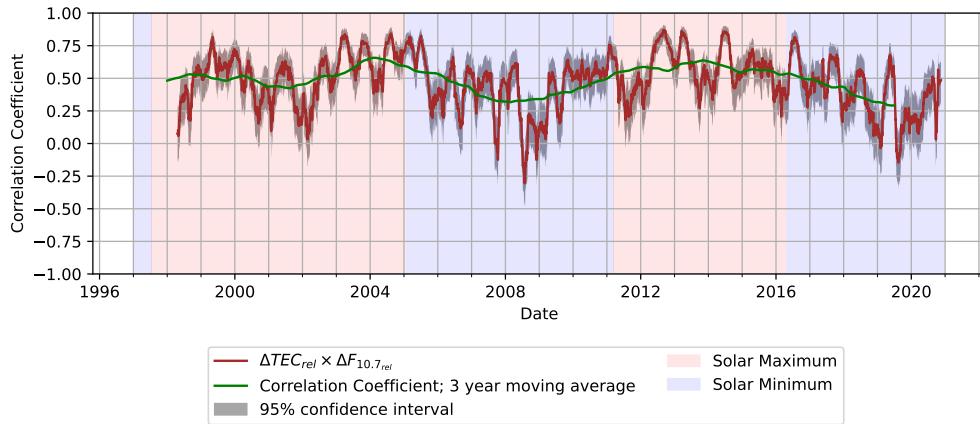


Figure 5.1: Correlation between  $\Delta TEC_{rel} \times \Delta F_{10.7_{rel}}$  with a 90 days window at 12.5 N latitude from 1998 to 2010. The red line represents the Pearson correlation coefficient values, and the green line represents the 3-year moving average of the correlation result. Red and blue shading represent the solar maximum and solar minimum periods, respectively.

When looking at Figure 5.1, the correlation coefficients during a low solar activity period were substantially lower than the correlation values during a high solar activity period. Over the time period of the study, the correlation values vary from 0.8, indicating a significant positive correlation, to -0.4, indicating a negative correlation.

In spite of the fact that there is a substantial amount of variability throughout the analysis period in the correlation graph, a higher level of correlation can be seen throughout period of greater solar activity which is shaded as red in Figure 5.1. On the other hand, low solar activity is shaded as blue; during these periods, the correlation between the two variables is also low.

### 5.1.2 Latitudinal dependency of the Correlation between TEC and F10.7 Solar Flux

The latitudinal change of the correlation coefficient throughout the investigated time period is illustrated in Figure 5.2 A. The study period's dates are on the X axis, while the latitude is on the y axis. The values of the

correlation coefficients are displayed as a blue to red color gradient bar with values ranging from -1 to +1, respectively. The figure shows clearly that there is a positive correlation trend across all latitudes during the study period. As stated earlier, it is clear that there is a higher positive correlation when solar maximum conditions, however the correlation values are low in solar minimum conditions.

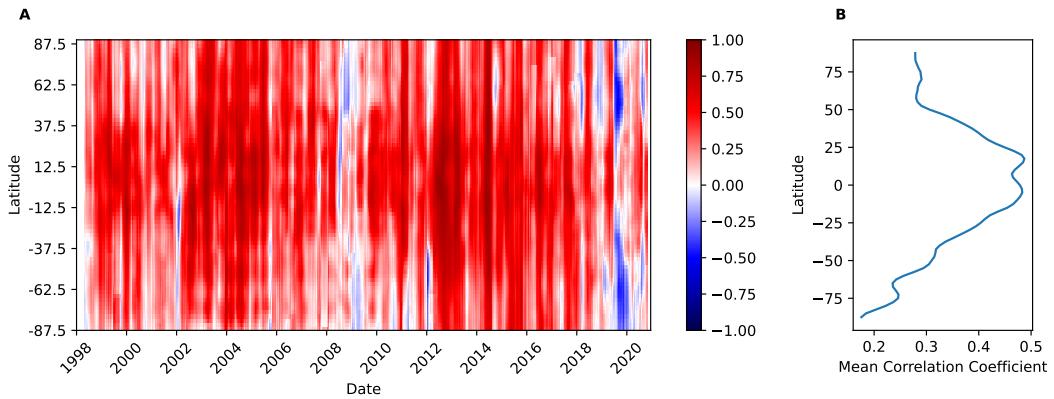


Figure 5.2: A: Latitudinal variation of correlation coefficient between  $\Delta F_{10.7,rel}$  and zonal mean  $\Delta TEC_{rel}$  with 90 days box window from 1998 to 2020. Red gradient represents magnitude of positive correlation coefficient and blue gradient represents negative correlation coefficient values.  
B: Latitude mean correlation coefficient  $\Delta F_{10.7,rel}$  and  $\Delta TEC_{rel}$  from 1998 to 2020.

The strong positive correlation is stronger at latitudes that are closer to the equator, which are  $\sim \pm 25^\circ$  latitudes. Higher correlation values at the equator generate a structure with two crests at the equator between -20 and +20 degrees latitude. The correlation coefficient gradually drops as it moves closer to the polar regions which is visible in the Figures 5.2. Nevertheless, there are very few areas where it has negative values which are quite rare.

In the equatorial regions, the average correlation coefficient is more than 0.5 during solar maximum conditions (taking into consideration the years 2001 to 2003 and 2012 to 2015 as solar maximum). On the other hand, the average correlation coefficient at the equatorial region when solar minimum conditions (considering 2006 to 2009 and 2018 to 2020 as solar minimum)

is 0.2. Among these, the years 2009 and 2020 have the lowest correlation coefficient.

Figure 5.2 B shows the mean correlation between  $\Delta F_{10.7,rel}$  and  $\Delta TEC_{rel}$  at latitudes over the period of the study. The mean correlation values at latitudes were determined from the data obtained across the study period (refer Figure 5.2 A) is shown in this figure. The X axis represents mean correlation values, while the Y axis indicates latitudes. Figure 5.2 B clearly shows the pattern of correlation value changes along latitude, as mentioned previously. The mean correlation coefficient is at its highest in regions closer to the equator, and as it moves towards the polar regions, the correlation coefficient declines.

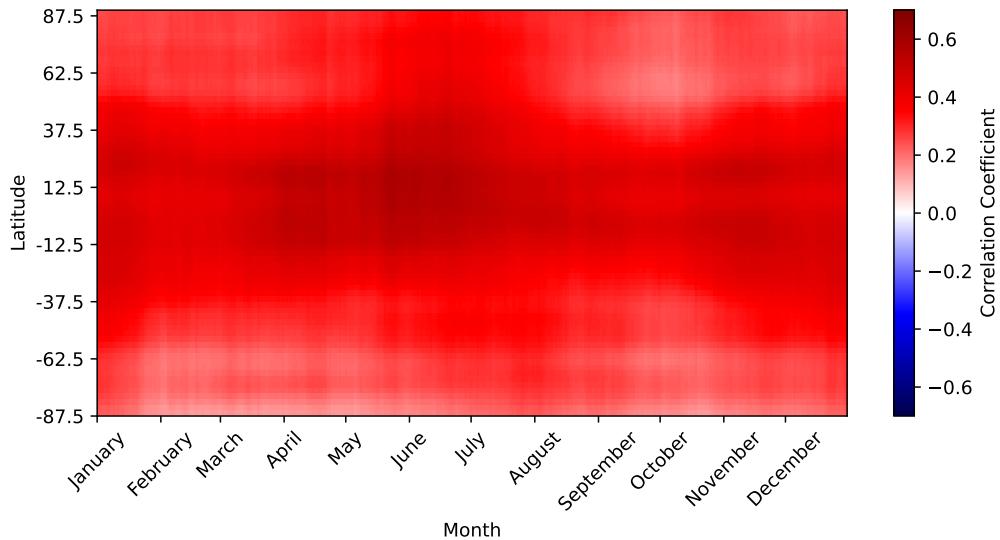


Figure 5.3: The correlation computation between  $\Delta F_{10.7,rel}$  and  $\Delta TEC_{rel}$  averaged over the course of the year in the period 1998 to 2020. Red gradient represents magnitude of positive correlation coefficient and blue gradient represents negative correlation coefficient values.

The changes in correlation values across the seasons are illustrated in Figure 5.3. The X axis of the figure contains the months, and the Y axis contains the latitude. The  $\Delta F_{10.7,rel}$  and  $\Delta TEC_{rel}$  correlation coefficient values, which range from -0.7 to 0.7, are illustrated by a color gradient moving from blue

to red respectively. The  $\Delta F_{10.7_{rel}}$  and  $\Delta TEC_{rel}$  have been shown to have a positive correlation, which holds true across all areas and during all four seasons. There is no distinguishable pattern of seasonal dependence in the relationship between the  $\Delta F_{10.7_{rel}}$  and  $\Delta TEC_{rel}$  solar flux. Nevertheless, there are small shifts in the degree to which a positive correlation exists between the months of summer and winter.

Figure 5.3, the seasonal dependency graph of  $\Delta F_{10.7_{rel}}$  and  $\Delta TEC_{rel}$  correlation shows a significant positive correlation close to the equator, but it decreases as it moves to the polar regions.

### 5.1.3 Local time dependency of the Correlation between TEC and F10.7 Solar Flux

Figure 5.4 illustrates the dependence of local time on the correlation between  $\Delta F_{10.7_{rel}}$  and  $\Delta TEC_{rel}$  for locations 60N 15E, 30N 15E, and 0N 15E. The mean correlation values for each hour were computed using data from the whole study period at the locations highlighted in this figure. On the X axis of the graph are the time ranging from 0 to 23, and on the Y axis are the mean correlation coefficient values. The colors blue, orange, and green correspond to the 60N 15E, 30N 15E, and 0N 15E locations, respectively.

This chart also illustrates that the magnitude of correlation values is highest during the daytime hours, while correlation values are lowest during the nighttime hours. From 6:00 UT to 16:00 UT, the correlation values are very high, which is easy to see in Figure 5.4 at the 0N, 15E latitude. As one gets closer to the equator, it is easy to see that the time dependency becomes more pronounced. At high latitude regions, the amplitude of the correlation is greatest during the day, while the time period gets shorter.

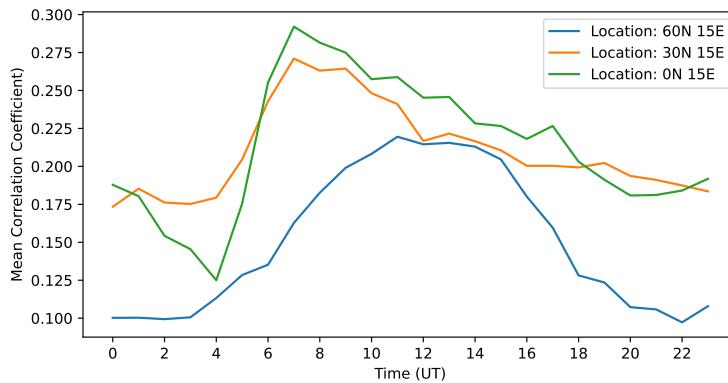


Figure 5.4: Local time variation computed by superposition of correlation between  $\Delta F_{10.7,rel}$  and  $\Delta TEC_{rel}$  at each UT during the period 1998–2020. Locations 60N/15E, 30N/15E and 0N/15E are represented using colours blue orange and green respectively.

### 5.1.4 Discussion of TEC and F10.7 Solar Flux Correlation

This study shows that there is a significant positive correlation between the TEC and F10.7 solar flux what is in good agreement with common knowledge. The correlation is particularly strong in the equatorial region, as shown in e.g. Vaishnav et al. (2019) (c.f. 2.8). This agrees with theory, because the ionization efficiency depends on the solar zenith angle and the solar zenith angle is highest in the equatorial region. Accordingly, the correlation coefficient values becomes much weaker with increasing latitude because we have low solar zenith angle in high latitudes and even polar night conditions without ionization during winter in polar regions. Ionization is the primary factor that determines TEC.

Due to the EIA, higher correlation values at the equator generate a structure with two crests between -20 and +20 degrees magnetic latitude. This phenomenon, which is also referred to as the Appleton anomaly and the fountain effect, is permanently seen in the ionosphere at low latitudes during daytime.

A high level of solar activity is characterized by a rise in the number of sunspots as well as an increase in the magnitude of solar irradiance in the

frequency bands causing ionisation in the Earth's atmosphere. This leads to stronger impact of solar irradiance on ionosphere variability during solar maximum conditions. Since TEC also depends on the magnitude of the solar irradiance, and we can see a solar cycle effect in TEC (c.f. Fig. 3.3), also the correlation between  $\Delta F_{10.7,rel}$  and  $\Delta TEC_{rel}$ . Accordingly, we observe during solar minimum, the magnitude of ionizing irradiance is lower and other factors like atmosphere coupling and solar wind coupling may compete in driving the ionosphere variability. Thus, the correlation between TEC and F10.7 is lower during solar minimum. The findings of this investigation are clearly consistent with the statement made by Hamzah and Homam (2015)

When comparing the findings of this research to those of Vaishnav et al. (2019), both show a significant positive connection between F10.7 and TEC. The studies in this paper are conducted using relative differences of the parameter and a 90 day window size for the correlation calculation. Since this window size is different to the 365 days window in Vaishnav et al. (2019), the actual correlation values differ slightly.

In the seasonal variation of correlation between  $\Delta F_{10.7,rel}$  and  $\Delta TEC_{rel}$ , there are small shifts in the degree to which a positive correlation exists between the months of summer and winter. Because of the increased amount of solar irradiance that occurs during the summer months, the positive correlation between the two variables is significantly higher during the summer months than it is during the winter months. As a result of the winter anomaly in the ionosphere, winter months have also demonstrated a second maximum of positive correlation. During equinox, when TEC is usually largest, the correlation with  $\Delta F_{10.7,rel}$  is lowest.

X-ray and UV light from the Sun mainly impacts the production term of the continuity equation. This impact is only during daytime hours. During night, there is no production of ionisation due to solar irradiance. This explains why the correlation with  $\Delta F_{10.7,rel}$  is lower during night.

## 5.2 Correlation between TEC and Solar Wind Speed

### 5.2.1 Temporal changes of the Correlation between TEC and Solar Wind Speed

The results of the correlation computations on  $\Delta V_{SW,rel}$  and  $\Delta TEC_{rel}$  at 12.5N latitude from 1998 to 2020 with box window of 90 days is shown in Figure 5.5. Dates are plotted on the X axis, and correlation coefficient values ranging from -1 to +1 are displayed on the Y axis. The red line graph represents correlation calculation with a box window of 90 days. The green line graph represents the 3 year moving average of the correlation coefficients. A gray shading indicates the confidence interval of the correlation calculation with the box window. A plot of 3 year moving average of F10.7 solar flux is provided above the correlation graph to identify the level of solar activity at each time.

Even though it exhibits a very high degree of fluctuation in the correlation coefficient, there is a pattern visible from solar maximum to solar minimum conditions. The figure demonstrates that the correlation becomes very low and often negative during solar maximum conditions near the equator region. Whereas the correlation coefficient values increases when solar conditions are minimum. The degree of positive and negative correlation during solar maximum and minimum varies according to the strength of solar activity. Strength of correlation is higher in solar cycle 23 compared to cycle 24.

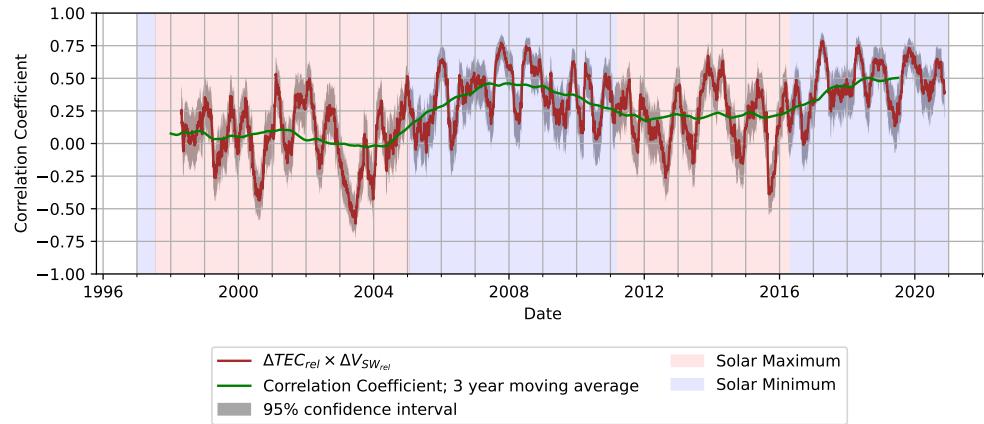


Figure 5.5: Correlation between  $\Delta V_{SW_{rel}}$  and  $\Delta TEC_{rel}$  with a 90 days window at 12.5 N latitude from 1998 to 2010. The red line represents the Pearson correlation coefficient values, and the green line represents the 3-year moving average of the correlation result. Red and blue shading represent the solar maximum and solar minimum periods, respectively.

### 5.2.2 Latitudinal dependency of the Correlation between TEC and Solar Wind Speed

The latitudinal change of the correlation coefficient between  $\Delta V_{SW_{rel}}$  and  $\Delta TEC_{rel}$  throughout the investigated time period is illustrated in Figure 5.6 A. The study period's dates are on the X axis, while the latitude is on the Z axis. The values of the correlation coefficients are displayed as a blue to red color gradient bar with values ranging from -1 to +1, respectively.

The correlation coefficient values range from as low as -0.775 to as high as 0.778, indicating a strong positive and negative correlation in particular latitudes over particular time periods. A strong positive correlation is more prevalent in places that are closer to the equator, whereas a negative correlation is more prevalent in regions that are closer to high latitudes. In addition, there is a pattern that can be seen in the correlation coefficient map. This pattern indicates that the correlation values at the equator are at their best during the period of solar minimum, whereas the correlation coefficient values drop to negative values during the solar maximum. In the polar regions, there is a significant negative correlation that exists during

the summer months of the years, and there is only a very slight positive correlation when solar conditions are minimum.

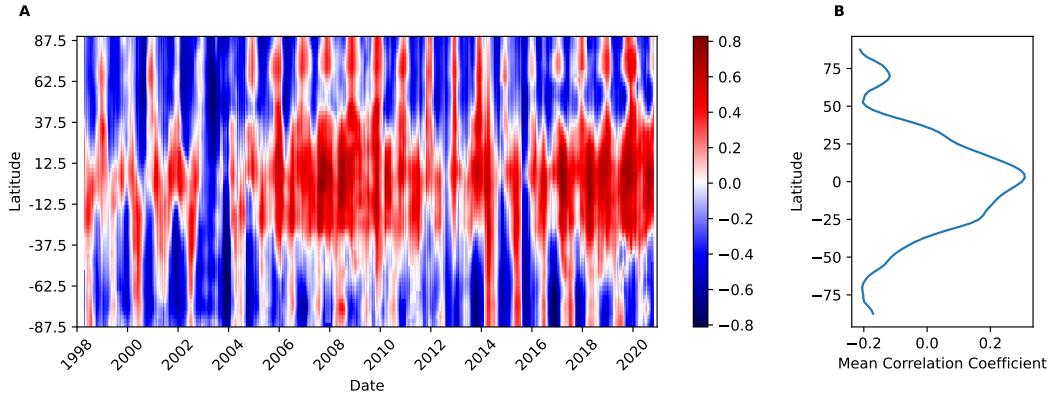


Figure 5.6: A: Latitudinal variation of correlation coefficient between  $\Delta V_{SW,rel}$  and  $\Delta TEC_{rel}$  with 90 days box window from 1998 to 2020. Red gradient represents magnitude of positive correlation coefficient and blue gradient represents negative correlation coefficient values.  
B: Latitude mean correlation coefficient  $\Delta V_{SW,rel}$  and  $\Delta TEC_{rel}$  from 1998 to 2020.

Figure 5.6 B shows the mean correlation between solar wind speed and TEC at latitudes over the period of the study. The mean correlation values at latitudes were determined from the data obtained across the study period (c.f. Figure 5.6 A) is shown in this figure. The X axis represents mean correlation values, while the Y axis indicates latitudes.

As indicated before, the pattern of correlation value changes with latitude is clearly shown in Figure 5.6 B. Furthermore, the mean correlation coefficient is highest in the equator and decreases as this goes to the poles. The mean correlation graph indicates that there is a difference between the North Pole and the South Pole. The North Pole has a greater positive correlation on the high latitudes than the South Pole. There is a secondary maximum of the positive correlation during winter in high latitudes. This is particularly noticeable between the latitudes between 50N and 90N.

The average seasonal correlation between  $\Delta V_{SW,rel}$  and  $\Delta TEC_{rel}$  with a 90-

day box window is illustrated in Figure 5.7. The months are on the X axis, and the latitudes are on the Y axis. The correlation coefficient values that correspond to -0.7 to 0.7 are depicted here using a color gradient that goes from blue to red. Based on this diagram, seasonal changes in correlation coefficient values across latitude can be identified.

Changes in season and latitude affect a significant change in the correlation between  $\Delta V_{SW_{rel}}$  and  $\Delta TEC_{rel}$ . There is a positive correlation that exists around the equator during all seasons. In comparison, the strength of positive correlation at the equator during the summer months is lower than it is during the winter months. During the winter, the positive correlation at the polar region extends to the North Pole, while the southern polar region is having a negative correlation influence. During the summer, a positive correlation persists in the southern polar region, while a negative correlation exists in the northern polar regions.

The highest mean seasonal correlation of +0.303 is found at a latitude between 10S and 10N in the month of December. Similarly, the correlation value between 75S and 90S in the month of February is -0.415, which is the lowest mean correlation value across seasons.

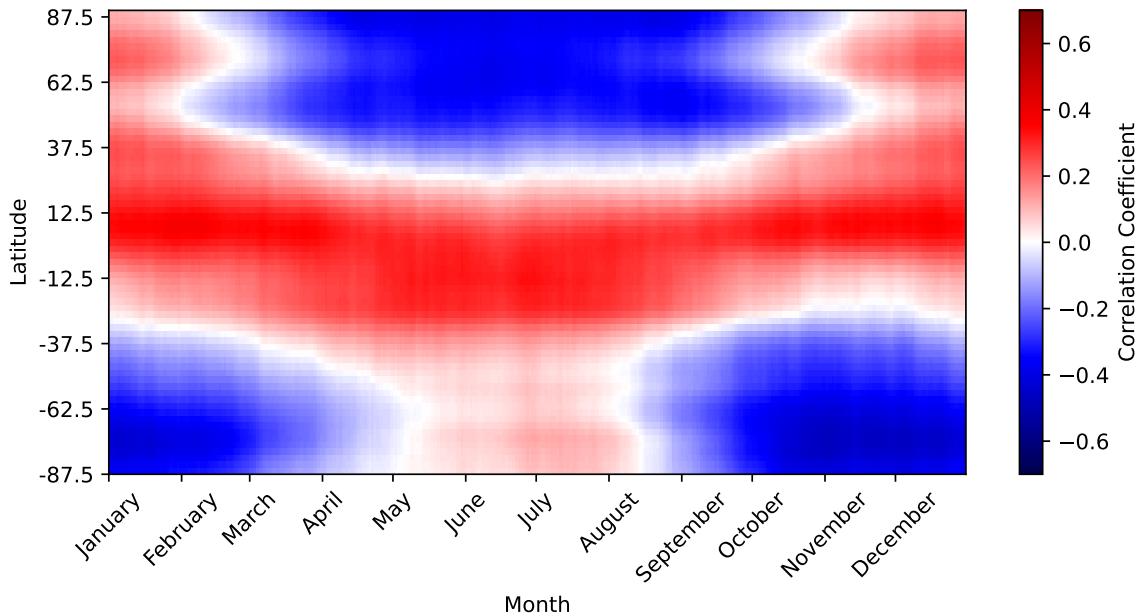


Figure 5.7: The correlation computation between  $\Delta V_{SW_{rel}}$  and  $\Delta TEC_{rel}$  averaged over the course of the year in the period 1998 to 2020. Red color represents magnitude of positive correlation coefficient and blue color represents negative correlation coefficient values.

### 5.2.3 Local time dependency of the Correlation between TEC and Solar Wind Speed

Figure 5.8 illustrates the dependence of local time on the correlation between  $\Delta V_{SW_{rel}}$  and  $\Delta TEC_{rel}$  for locations 60N/15E, 30N/15E, and 0N/15E. The mean correlation values for each hour were calculated using correlation calculation data from the whole study period at the locations 60N/15E, 30N/15E, and 0N/15E. On the X axis of the graph are the time ranging from 0 to 23, and on the Y axis are the mean correlation coefficient values. The locations at 60N/15E, 30N/15E, and 0N/15E are shown by blue, orange, and green, respectively.

The local time dependency graph also shows some characteristics, similar to those seen in the variation of correlation across seasons and latitude. At the

equator, correlation values increase throughout the day and decrease during the night, as depicted in the figure. As it goes to the higher latitudes, the magnitude of mean correlation decreases as well as the period also decreases. However, correlation values are still higher during the day.

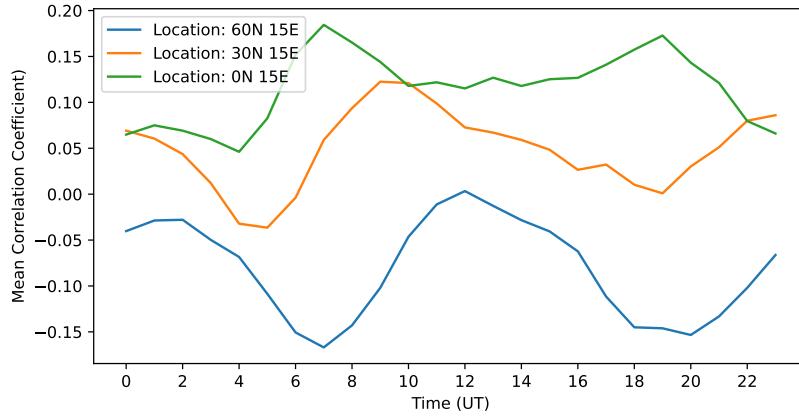


Figure 5.8: Local time variation computed by superposition of correlation between  $\Delta V_{SW_{rel}}$  and  $\Delta TEC_{rel}$  at each UT during the period 1998–2020. Locations 60N/15E, 30N/15E and 0N/15E are represented using colours blue orange and green respectively.

#### 5.2.4 Discussion of TEC and Solar Wind Speed Correlation

The correlation between  $\Delta V_{SW_{rel}}$  and  $\Delta TEC_{rel}$  varies with the solar activity. On average, a positive correlation between TEC and solar wind speed is visible during the solar minimum of the solar cycle. When solar irradiance decreases, the effect of solar wind on TEC variations becomes more visible. The correlation between solar wind speed and TEC changes with the seasons. In the winter, there is a significant positive correlation, whereas in the summer, there is a significant negative correlation that can be seen in the mid and high latitude regions. These correlated variations are likely to be caused by the combined action of the solar wind imposing an electric field onto the regular ionosphere electric fields and the to seasonally changing neutral winds in the thermosphere.

During the solar minimum, the equatorial region exhibits a positive correlation, whereas the mid-latitudes and polar regions frequently exhibit a negative correlation. A clear seasonal dependency in the correlation between  $\Delta V_{SW_{rel}}$  and  $\Delta TEC_{rel}$  can be observed, which is most prominent in mid-latitudes. Because the solar wind does not change with season, it must be a change in the thermosphere conditions (circulation, temperature, ...), which causes the seasonal variation of the correlation with the solar wind. The relevant electrodynamic process must be the ionosphere dynamo, which is generated through the joint action of winds and background electric field in a magnetic field environment (Richmond, 1989; Yamazaki & Maute, 2017). An electric field is produced inside of the inner magnetosphere as a result of the impact of the solar wind through the Earth's open magnetic field lines. The strength of this field varies depending on the speed of the solar wind and solar wind makes a significant contribution to the strengthening of the electric field causes increase in upward drift of the equatorial plasma. In equatorial latitudes, this electric field impacts the plasma fountain effect, leading to a change (enhancement or depletion) in the EIA. In high-latitudes, the magnetosphere convection electric field has a direct impact on plasma transport, due to the  $E \times B$  drift. It causes horizontal plasma drifts across the polar cap, which are normally directed from the dayside to the night side. If the plasma drifts are strong enough (especially during storms), the collision between plasma and neutrals can cause a thermosphere wind. In mid-latitudes, the electric field impacts plasma transport processes connected to the solar quiet current system, which is driven by the ionosphere dynamo.

# Chapter 6

## Summary and Conclusion

The aim of this thesis was to illustrate and characterize correlation between the solar forcing parameters in the ionosphere at time scales of several days to months. Temporal, latitudinal, seasonal, and local time variations of the correlation between solar forcing parameters in the ionosphere were investigated in this thesis. Based on quantitative evaluation of TEC, F10.7 solar flux, and solar wind speed data, it can be concluded that there is a significant relationship between these parameters in terms of time, latitude, and season.

TEC and F10.7 solar flux correlation result indicates that, there exist a significant positive correlation in the equatorial region. Because of the high solar zenith angle in the equatorial region, which leads to high ionization rates, more ionisation occurs, resulting in high TEC concentration. This is evident in temporal, latitudinal, and seasonal variations. In the time scales of days to months, the correlation between F10.7 and TEC is in average 0.35, which is significantly less than in the time scale of years. This shows that in the time scale of days to months, there are other factors like atmosphere variability and geomagnetic activity, which contribute to the TEC variability. The presented results agree well with the state-of-the art knowledge about the correlation between TEC and F10.7 solar flux.

Correlation studies between TEC and solar wind have not yet been published. The results presented here show that in the time scale of days, solar

wind is obviously impacting TEC. The correlation is slightly less than for F10.7. While the correlation of TEC with F10.7 is constantly positive, the correlation with solar wind depends on season and location. The correlation can be even negative. It can be assumed that the solar wind modify electric fields in the ionosphere, which modify the plasma transport and ionosphere currents.

These dependencies of TEC on solar wind in the range of days needs deeper investigation to understand the coupling processes and quantify the impact. Changes in electric fields caused by variations in the solar wind speed are assumed to modify the plasma transport in the ionosphere and thus modify TEC. Studies with numerical models can generate further insight into the coupling processes. Numerical and empirical modelling can be used to quantify the contribution of solar wind to TEC variabiltiy.

# Bibliography

- Acharya, R., Roy, B., Sivaraman, M., & Dasgupta, A. (2010). An empirical relation of daytime equatorial total electron content with equatorial electrojet in the indian zone. *Journal of Atmospheric and Solar-Terrestrial Physics*, 72(9), 774–780. <https://doi.org/10.1016/j.jastp.2010.03.023>
- Afraimovich, E. L., Astafyeva, E. I., Oinats, A. V., Yasukevich, Y. V., & Zhivetiev, I. V. (2008). Global electron content: A new conception to track solar activity. *Annales Geophysicae*, 26(2), 335–344. <https://doi.org/10.5194/angeo-26-335-2008>
- Aggarwal, M., Joshi, H., Iyer, K., Kwak, Y.-S., Lee, J., Chandra, H., & Cho, K. (2012). Day-to-day variability of equatorial anomaly in gps-tec during low solar activity period. *Advances in Space Research*, 49(12), 1709–1720.
- Appleton, E. V. (1946). Two anomalies in the ionosphere. *Nature*, 157(3995), 691–691.
- Arnold, D. (2006). *Sonnensystem: Entstehung, forschung, entwicklung*. Zenodot-Verl.-Ges.
- B. McElroy, M. (2012). Ionosphere and magnetosphere - mechanisms of ionization — britannica. [www.britannica.com/science/ionosphere-and-magnetosphere/Mechanisms-of-ionization](http://www.britannica.com/science/ionosphere-and-magnetosphere/Mechanisms-of-ionization)
- Balan, N., Liu, L., & Le, H. (2018). A brief review of equatorial ionization anomaly and ionospheric irregularities. *Earth and Planetary Physics*, 2(4), 257–275. <https://doi.org/10.26464/epp2018025>
- Bothmer, V., & Daglis, I. A. (2007). *Space weather: Physics and effects*. Springer Science & Business Media.

- Burns, A. G., Wang, W., Qian, L., Solomon, S. C., Zhang, Y., Paxton, L. J., & Yue, X. (2014). On the solar cycle variation of the winter anomaly. *Journal of Geophysical Research: Space Physics*, 119(6), 4938–4949. [https://doi.org/https://doi.org/10.1002/2013JA019552](https://doi.org/10.1002/2013JA019552)
- Cai, L., Aikio, A., & Nygrén, T. (2014). Solar wind effect on joule heating in the high-latitude ionosphere. *Journal of Geophysical Research: Space Physics*, 119. <https://doi.org/10.1002/2014JA020269>
- Catling, D. C., & Kasting, J. F. (2017). *Atmospheric evolution on inhabited and lifeless worlds*. Cambridge University Press.
- Chen, Y., Liu, L., Le, H., & Wan, W. (2018). Responses of solar irradiance and the ionosphere to an intense activity region. *Journal of Geophysical Research: Space Physics*, 123(3), 2116–2126. <https://doi.org/10.1002/2017JA024765>
- Cranmer, S. R. (2009a). Coronal holes. *Living Reviews in Solar Physics*, 6(1), 1–66.
- Cranmer, S. R. (2009b). Coronal holes. *Living Reviews in Solar Physics*, 6(1), 3. <https://doi.org/10.12942/lrsp-2009-3>
- Dang, T., Li, X., Luo, B., Li, R., Zhang, B., Pham, K., Ren, D., Chen, X., Lei, J., & Wang, Y. (2022). Unveiling the space weather during the starlink satellites destruction event on 4 february 2022. *Space Weather*, 20(8), e2022SW003152.
- Doherty, P., Coster, A. J., & Murtagh, W. (2004). Space weather effects of october–november 2003. *GPS Solutions*, 8(4), 267–271.
- Donahue, T. (1968). Ionospheric composition and reactions: Our present knowledge of what ions are in the ionosphere, and why, is summarized. *Science*, 159(3814), 489–498.
- Edwards, A. L. (1976). An introduction to linear regression and correlation. *The Correlation Coefficient*, 33–46.
- Engvold, O., Vial, J., & Skumanich, A. (2018). *The sun as a guide to stellar physics*. Elsevier Science. <https://books.google.de/books?id=WKR6DwAAQBAJ>
- Glas, C. (2010). Missing data. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education (third edition)* (Third Edition, pp. 283–288). Elsevier. [https://doi.org/https://doi.org/10.1016/B978-0-08-044894-7.01346-4](https://doi.org/10.1016/B978-0-08-044894-7.01346-4)
- Gogtay, N. J., & Thatte, U. M. (2017). Principles of correlation analysis. *Journal of the Association of Physicians of India*, 65(3), 78–81.

- Golubkov, G., Maslov, T., Bychkov, V., Borchevskina, O., Adamson, S., Dyakov, Y. A., Lushnikov, A., & Golubkov, M. (2020). Atomic oxygen in the ionospheric e layer. *Russian Journal of Physical Chemistry B*, 14(5), 853–861.
- Goncharenko, L. P., Harvey, V. L., Randall, C. E., Coster, A., Zhang, S.-R., Zalizovski, A., Galkin, I., & Spraggs, M. (2022). Observations of pole-to-pole, stratosphere-to-ionosphere connection.
- Gosling, J. (2014). Chapter 12 - the solar wind. In T. Spohn, D. Breuer, & T. V. Johnson (Eds.), *Encyclopedia of the solar system (third edition)* (Third Edition, pp. 261–279). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-12-415845-0.00012-8>
- Haigh, J. D. (2007). The sun and the earth's climate. *Living reviews in solar physics*, 4(1), 1–64.
- Hamzah, S., & Homam, M. (2015). The correlation between total electron content variations and solar activity. *Asian research publishing network (ARPN). ARPN J. Eng. Appl. Sci*, 10, 20.
- Hapgood, M. (2017). *Space weather*. IOP Publishing. <https://doi.org/10.1088/978-0-7503-1372-8>
- Hathaway, D. H., Wilson, R. M., & Reichmann, E. J. (2002). Group sunspot numbers: Sunspot cycle characteristics. *Solar Physics*, 211(1), 357–370.
- Hauke, J., & Kossowski, T. (2011). Comparison of values of pearson's and spearman's correlation coefficient on the same sets of data.
- Hernández-Pajares, M., Juan, J., Sanz, J., Orus, R., Garcia-Rigo, A., Feltens, J., Komjathy, A., Schaer, S., & Krancowski, A. (2009). The igs vtec maps: A reliable source of ionospheric information since 1998. *Journal of Geodesy*, 83(3), 263–275.
- Hofmeister, Stefan J., Asvestari, Eleanna, Guo, Jingnan, Heidrich-Meisner, Verena, Heinemann, Stephan G., Magdalenic, Jasmina, Poedts, Stefaan, Samara, Evangelia, Temmer, Manuela, Vennerstrom, Susanne, Veronig, Astrid, Vrsnak, Bojan, & Wimmer-Schweingruber, Robert. (2022). How the area of solar coronal holes affects the properties of high-speed solar wind streams near earth: An analytical model. *A&A*, 659, A190. <https://doi.org/10.1051/0004-6361/202141919>
- Jakowski, N., Fichtelmann, B., & Jungstand, A. (1991). Solar activity control of ionospheric and thermospheric processes. *Journal of atmospheric and terrestrial physics*, 53(11-12), 1125–1130.

- Jakowski, N., Sardon, E., Engler, E., Jungstand, A., & Klähn, D. (1996). Relationships between gps-signal propagation errors and eiscat observations. *14*(12), 1429–1436.
- Knipp, D. J., Fraser, B. J., Shea, M., & Smart, D. (2018). On the little-known consequences of the 4 august 1972 ultra-fast coronal mass ejecta: Facts, commentary, and call to action. *Space Weather*, *16*(11), 1635–1643.
- Knipp, D. J., Ramsay, A., Beard, E., Boright, A., Cade, W., Hewins, I., McFadden, R., Denig, W., Kilcommons, L., Shea, M., et al. (2016). The may 1967 great storm and radio disruption event: Extreme space weather and extraordinary responses. *Space Weather*, *14*(9), 614–633.
- Kutiev Ivan, Tsagouri, Ioanna, Perrone, Loredana, Pancheva, Dora, Mukhtarov, Plamen, Mikhailov, Andrei, Lastovicka, Jan, Jakowski, Norbert, Buresova, Dalia, Blanch, Estefania, Andonov, Borislav, Altadill, David, Magdaleno, Sergio, Parisi, Mario, & Miquel Torta, Joan. (2013). Solar activity impact on the earth's upper atmosphere. *J. Space Weather Space Clim.*, *3*, A06. <https://doi.org/10.1051/swsc/2013028>
- Liu, L., Wan, W., Chen, Y., & Le, H. (2011). Solar activity effects of the ionosphere: A brief review. *Chinese Science Bulletin*, *56*, 1202–1211. <https://doi.org/10.1007/s11434-010-4226-9>
- McComas, D., Bame, S., Barker, P., Feldman, W., Phillips, J., Riley, P., & Griffee, J. (1998). Solar wind electron proton alpha monitor (SWEPAM) for the advanced composition explorer. *The Advanced Composition Explorer Mission*, 563–612.
- McKinney, W., et al. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, *445*(1), 51–56.
- Moran, D., Lenzen, M., Cairns, I., Steenge, A., et al. (2014). How severe space weather can disrupt global supply chains. *Natural Hazards and Earth System Sciences*, *14*(10), 2749–2759.
- NASA. (2011). Anatomy of the sun. *NASA*. <https://www.nasa.gov/mission-pages/sunearth/science/solar-anatomy.html>
- Natural Resources Canada, G. o. C. (2021). About the solar flux data. *spaceweather.gc.ca*. Retrieved June 30, 2022, from <https://spaceweather.gc.ca/forecast-prevision/solar-solaire/solarflux/sx-3-en.php>
- Naval Postgraduate School, U. N. (2003). Hf and lower frequency radiation - the ionosphere. *www.met.nps.edu*. [https://www.met.nps.edu/~psguest/EMEO\\_online/module3/module\\_3\\_2.html](https://www.met.nps.edu/~psguest/EMEO_online/module3/module_3_2.html)

- Nayir, H., Arikан, F., Arikан, O., & Erol, C. B. (2007). Total electron content estimation with reg-est. *Journal of Geophysical Research: Space Physics*, 112(A11). <https://doi.org/https://doi.org/10.1029/2007JA012459>
- NOAA. (n.d.). Space weather storms from the sun space weather storms from the sun. [https://www.swpc.noaa.gov/sites/default/files/images/u33/swx\\_booklet.pdf](https://www.swpc.noaa.gov/sites/default/files/images/u33/swx_booklet.pdf)
- Nolte, J. T., Krieger, A. S., Timothy, A. F., Gold, R. E., Roelof, E. C., Vaiana, G., Lazarus, A. J., Sullivan, J. D., & McIntosh, P. S. (1976). Coronal holes as sources of solar wind. *Solar Physics*, 46(2), 303–322. <https://doi.org/10.1007/BF00149859>
- Osepian, A., Tereshchenko, V., Dalin, P., & Kirkwood, S. (2008). The role of atomic oxygen concentration in the ionization balance of the lower ionosphere during solar proton events. *Annales Geophysicae*, 26(1), 131–143. <https://doi.org/10.5194/angeo-26-131-2008>
- Owens, M. J., & Forsyth, R. J. (2013). The Heliospheric Magnetic Field - Living Reviews in Solar Physics [[Online; accessed 2022-07-20]].
- Pajares, M. (2003). Performance of igs ionosphere tec maps.
- Parenti, S., Bromage, B., Poletto, G., Noci, G., Raymond, J., & Bromage, G. (2000). Characteristics of solar coronal streamers. element abundance, temperature and density from coordinated cds and uvcs soho observations. *Astronomy and Astrophysics*, 363, 800–814.
- Parker, E. N. (1958). Dynamics of the interplanetary gas and magnetic fields. *The Astrophysical Journal*, 128, 664.
- Piddington, J. H. (1964). Some ionospheric effects of the solar wind. *IETE Journal of Research*, 10(8), 285–291. <https://doi.org/10.1080/03772063.1964.11485057>
- Ratcliffe, J. A. (A. (1972). *An introduction to the ionosphere and magnetosphere*. Cambridge [Eng.] University Press. Retrieved August 9, 2022, from <https://archive.org/details/introductiontoio0000ratc/page/38/mode/2up?view=theater>
- Ren, D., Lei, J., Wang, W., Burns, A., Luan, X., & Dou, X. (2018). Does the peak response of the ionospheric f2 region plasma lag the peak of 27-day solar flux variation by multiple days? *Journal of Geophysical Research: Space Physics*, 123(9), 7906–7916. <https://doi.org/https://doi.org/10.1029/2018JA025835>
- Richmond, A. (1989). Modeling the ionosphere wind dynamo: A review. *pure and applied geophysics*, 131(3), 413–435.

- Ruzmaikin, A. (2001). Origin of sunspots. *Space Science Reviews*, 95(1), 43–53.
- Schmöller, E., Berdermann, J., Jakowski, N., Jacobi, C., & Vaishnav, R. (2018). Delayed response of the ionosphere to solar euv variability. *Advances in Radio Science*, 16, 149–155. <https://doi.org/10.5194/ars-16-149-2018>
- Schunk, R. W. (2014). Ionosphere-thermosphere physics: Current status and problems. *Modeling the Ionosphere–Thermosphere System*, 3–12.
- Schunk, R., & Nagy, A. (2009). *Ionospheres: Physics, plasma physics, and chemistry*. Cambridge university press.
- Senthilnathan, S. (2019). Usefulness of correlation analysis. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3416918>
- Solanki, S. K., Schüssler, M., & Fligge, M. (2000). Evolution of the sun's large-scale magnetic field since the maunder minimum. <https://www.nature.com/articles/35044027#citeas>
- Solanki, S. K., & Hammer, R. (2001). The solar atmosphere. In J. A. M. Bleeker, J. Geiss, & M. C. E. Huber (Eds.), *The century of space science* (pp. 1065–1088). Springer Netherlands. [https://doi.org/10.1007/978-94-010-0320-9\\_45](https://doi.org/10.1007/978-94-010-0320-9_45)
- Solomon, S., & Roble, R. (2015). Thermosphere. In G. R. North, J. Pyle, & F. Zhang (Eds.), *Encyclopedia of atmospheric sciences (second edition)* (Second Edition, pp. 402–408). Academic Press. <https://doi.org/10.1016/B978-0-12-382225-3.00408-4>
- Stix, M. (2012). *The sun: An introduction*. Springer Science & Business Media. Retrieved June 7, 2022, from [https://books.google.de/books?id=oSrqCAAAQBAJ&printsec=frontcover&source=gb\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.de/books?id=oSrqCAAAQBAJ&printsec=frontcover&source=gb_ge_summary_r&cad=0#v=onepage&q&f=false)
- Tapping, K. (2013). The 10.7 cm solar radio flux (f10. 7). *Space weather*, 11(7), 394–406.
- Taylor, H., Vreugdenburg, M., Sangalli, L., & Vincent, R. (2021). Rmcsat: An f10. 7 solar flux index cubesat mission. *Remote Sensing*, 13(23), 4754.
- Vaishnav, R., Jacobi, C., & Berdermann, J. (2019). Long-term trends in the ionospheric response to solar extreme-ultraviolet variations. *Annales Geophysicae*, 37(6), 1141–1159. <https://doi.org/10.5194/angeo-37-1141-2019>
- Vaishnav, R., Jacobi, C., Berdermann, J., Schmöller, E., & Codrescu, M. (2018). Ionospheric response to solar euv variations: Preliminary re-

- sults. *Advances in Radio Science*, 16, 157–165. <https://doi.org/10.5194/ars-16-157-2018>
- Wik, M., Pirjola, R., Lundstedt, H., Viljanen, A., Wintoft, P., & Pulkkinen, A. (2009). Space weather events in july 1982 and october 2003 and the effects of geomagnetically induced currents on swedish technical systems. *Annales Geophysicae*, 27(4), 1775–1787.
- Wilcox, J., Hoeksema, J., & Scherrer, P. (1980). Origin of the warped heliospheric current sheet. *Science*, 209(4456), 603–605.
- Wurz, P. (2005). Solar wind composition. 600, 44.
- Ya'acob, N., Abdullah, M., & Ismail, M. (2010). *Gps total electron content (tec) prediction at ionosphere layer over the equatorial region*. IntechOpen.
- Yamazaki, Y., & Maute, A. (2017). Sq and eej—a review on the daily variation of the geomagnetic field caused by ionospheric dynamo currents. *Space Science Reviews*, 206(1), 299–405.
- Zell, H. (2018). Earth's magnetosphere. NASA. Retrieved December 3, 2022, from [https://www.nasa.gov/mission\\_pages/sunearth/multimedia/magnetosphere-unlabeled.html](https://www.nasa.gov/mission_pages/sunearth/multimedia/magnetosphere-unlabeled.html)

# Appendix

## Timeseries Correlation Anlysis Software Tool - User Guide

### Overview

This software is a time series analysis tool that streamlines the data loading, data cleaning and formatting process as well as calculations and data visualization. The four independent modules (see figure 4.4) of the software tool work based on the configurations written in JSON format. The software tool is written in Python programming language.

### Requirements

The software tool requires the following version of python and some other python packages to run.

- Python version 3.9
- Latest version of spacepy
- Latest version of NumPy
- Latest version of Pandas
- Latest version of Matplotlib

## Repository

The source code of the software tool is stored in the following repository on GitHub. Clone the repository locally.

```
git https://github.com/fredythekkekara/correlation_app.git  
./correlation_app
```

Now change directory to the software tool's directory to execute further commands.

```
cd ./correlation_app
```

The structure of the repository looks like as following:

```
correlation_app  
└── computation.py  
└── data_cleaning_and_formatting.py  
└── init.py  
└── load_data_files.py  
└── plot_data.py
```

With reference to Figure 4.4, the files `load_data_files.py`, `data_cleaning_and_formatting.py`, `computation.py`, and `plot_data.py` correspond to the modules data load (P1), data cleaning and formatting (P2), computations (P3), and data visualisation (P4), respectively.

## Data Source

This analysis uses F10.7 solar flux readings provided by the Dominion Radio Astrophysical Observatory and Natural Resources Canada, SWEPAM daily averaged solar wind speed data from the ACE spacecraft, and TEC data from the IGS Ionosphere Group for the correlation study. Download and store data from the above sources to prepare configuration files for the execution.

F10.7 solar flux data source:

[https://www-app3.gfz-potsdam.de/kp\\_index/Kp\\_ap\\_Ap\\_SN\\_F107\\_since\\_1932.txt](https://www-app3.gfz-potsdam.de/kp_index/Kp_ap_Ap_SN_F107_since_1932.txt)

Solar wind speed data source:

[https://izw1.caltech.edu/ACE/ASC/level2/lvl2DATA\\_SWEPAM.html](https://izw1.caltech.edu/ACE/ASC/level2/lvl2DATA_SWEPAM.html)

TEC data source:

<https://cdaweb.gsfc.nasa.gov/pub/data/gps/>

## Execution

### Initialising project

The following project initializes the project and the project folder directory.

```
python init.py project_name project_location
```

The above project initialization command returns the following directory and JSON configuration files created in the project location:

```
project_name
├── computations
│   └── computations_config.json
├── data_formatting
│   └── data_formatting_config.json
├── load_data
│   └── load_data_config.json
└── plot_data
    └── plot_config.json
```

The JSON files in each module directory define what to do with the software. The following steps explain what are the attributes, parameters, and their functions in the system.

## Data loading

Data load module (P1) of the software system fetches raw data values from the data sources, structures them into a Pandas DataFrame, and saves them into HDF files. The configuration for the modules is stored in .../project\_name/load\_data/load\_data\_config.json location. The function of this step is to extract data from text files based on character location. Attributes, possible input formats, and their functionalities are defined as follows:

Attribute	Input Format	Functionality
name	string	Name of the attribute column for the data which is going to be fetched.
fileLocation	path	Location of the data file
fileType	.txt .csv .cdf	Type of the input data file
attrName	string	Name of the attribute column for the data which is going to fetch.
isIndex	true or false	'true' sets the attribute column as index for the DataFrame.
startPosition	integer	Start position of the fetching attribute in the text data file.
endPosition	integer	Start position of the fetching attribute in the text data file.
maxThresholdValue	integer	Replaces data value by threshold value if it is greater.

An example of the JSON object for data loading in configuration file is given below.

```
{  
    "name": "f10_7",  
    "fileLocation": "C:\\\\Users\\\\davi_fr\\\\Documents\\\\Thesis  
        Project Final\\\\data\\\\RAW DATA FILES\\\\f10_7",  
    "fileType": ".txt",  
    "attributes": [  
        {  
            "attrName": "date",  
            "isIndex": true,  
            "startPosition": 0,  
            "endPosition": 10  
        },  
        {  
            "attrName": "f10_7",  
            "isIndex": false,  
            "startPosition": 149,  
            "endPosition": 156,  
            "maxThresholdValue": 300  
        }  
    ],  
    "indexAttributes": []  
}
```

The data loading module can be executed using the following command, passing input as the configuration file of the module.

```
python load_data.py ..\\\\project_name\\\\load_data\\\\  
load_data_config.json
```

## Data cleaning and formatting

The data cleaning and formatting module (P2) of the software system is used mainly to fill the data gaps and adjust the indexes of the timeseries. Following are the attributes of the corresponding configuration file:

## Appendix

---

Attribute	Input Format	Functionality
name	string	Name of the attribute column for the data which is going to be fetched.
fileFrequency	yearly	To read files recursively from a folder location based on analysis period. Name of the file should be year.
indexFrequency	single	To read single data file
	hourly	Index frequency of the data points in the timeseries. This will be used to correct missing indexes.
dataColumns	daily	
	string array	column names for the data columns in the DataFrame.
dataFiles	array of JSON object	Details of the data file locations and data columns are defined as JSON object. Merge data columns to missing values from other files if mentioned if more than 1 JSON object is there.
location	file path	Location of the data file
mergeColumns	string array	Array of columns names to be merged.

An example of the JSON object for data cleaning and formatting from the configuration file is given below.

```
{  
    "name": "tec",  
    "fileFrequency": "yearly",  
    "indexFrequency": "hourly",  
    "dataColumns": ["tec"],  
    "dataFiles": [  
        {  
            "location": "C:/Users/davi_fr/Documents/Thesis  
                Project Final/tets folder main/test/load_data  
                /tec_1hr",  
            "mergeColumns": ["tecUHR", "tecEHR", "tecCOR", "  
                tecCOD"]  
        },  
        {  
            "location": "C:/Users/davi_fr/Documents/Thesis  
                Project Final/tets folder main/test/load_data  
                /tec_15min",  
            "mergeColumns": ["tecUQR"]  
        },  
        {  
            "location": "C:/Users/davi_fr/Documents/Thesis  
                Project Final/tets folder main/test/load_data  
                /tec_2hr",  
            "mergeColumns": ["tecCOD", "tecCOR", "tecESA", "  
                tecESR", "tecIGR", "tecIGS", "tecJPL", "  
                tecJPR", "tecUPC", "tecUPR"]  
        }  
    ]  
}
```

The data cleaning and formatting module can be executed using the following command, passing input as the configuration file of the module.

```
python data_cleaning_and_formatting.py ..\project_name\  
    data_formatting\data_formatting_config.json
```

## Computations

The computation module (P3) of the software system performs all the computations for the correlation analysis. Each computation in the module is treated as an independent function that accepts data input as well as individual parameters for the operation. The operations receive input as a DataFrame and return a result as a DataFrame, which will be stored inside the computations folder in the project directory. Following are the operations in the computation modules and their configurations.

### Moving average

In a time series, a moving average is just the mean across a predetermined number of prior periods. Moving average in DataFrames are computed for each columns. Following are the attributes for the moving average calculations.

Attribute	Input Format	Functionality
operation	moving_average	operation computes moving average of the data set
name	string	name of the data set
fileLocation	file path	Location of the data file
fileFrequency	'yearly'	Data stored in yearly files which fetches based on analysis period recursively from a file location.
windowSize	'single'	Data stored in single file
minimumPeriod	integer	Size of the moving window.
groupBy	'hour'	Groups data values based on each hour in the time index.

An example of the JSON object for moving average from the computation module's configuration file is given below.

```
{
    "operation": "moving_average",
    "name": "tec",
    "fileLocation": "C:\\\\Documents\\\\correlation_project\\\\
                    data_formatting\\\\tec_og",
    "fileFrequency": "yearly",
    "windowSize": 27,
    "minimumPeriod": 20,
    "groupBy": "hour",

}
```

## Relative difference

Relative difference aims to assess the variability between two measurements. Attributes for the relative difference calculation are listed in the following table.

Attribute	Input Format	Functionality
operation	relative_difference	operation computes moving average of the data set
name	string	name of the data set
fileFrequency	'yearly'	Data stored in yearly files which fetches based on analysis period recursively from a file location.
	'single'	Data stored in single file
newValueFile	file path	Location of the initial value data set
referenceValueFile	integer	Location of the final value data set

An example of the JSON object for relative difference from the computation module's configuration file is given below.

```
{
    "operation": "relative_difference",
    "name": "tec",
    "newValueFile": "C:\\Documents\\correlation_project\\\
        computations\\moving_average\\tec",
    "referenceValueFile": "C:\\Documents\\\
        correlation_project\\data_formatting\\tec_og",
    "fileFrequency": "yearly",

}
```

## Mean

Mean operation is used to find the average of a set of data values along different axes, either row-wise or column-wise. The list of attributes for configuring mean operation is listed below.

Attribute	Input Format	Functionality
operation	relative_difference	operation computes moving average of the data set
name	string	name of the data set
fileLocation	file path	Location of the data file
fileFrequency	'yearly'	Data stored in yearly files which fetches based on analysis period recursively from a file location.
groupBy	'single' column_name 'date' 'season'	Data stored in single file Compute group by mean of the data set based on a column in the DataFrame Computes mean by grouping date index column. Computes mean by grouping day of the year.

An example of the JSON object for mean operation from the computation

module's configuration file is given below.

```
{
    "operation": "mean",
    "name": "tec_lon",
    "fileLocation": "C:\\\\Documents\\\\correlation_project\\\\
                    computations\\\\relative_difference\\\\tec",
    "fileFrequency": "yearly",
    "groupBy": "lat",

}
```

## Append

Operator to combine multiple DataFrame with same index columns into one singe DataFrame file. The list of attributes for configuring append operation is listed below.

Attribute	Input Format	Functionality
operation	append	Combine multiple DataFrame in a file location.
name	string	name of the data set
fileLocation	file path	Location of the data file
fileFrequency	'yearly'	Data stored in yearly files which fetches based on analysis period recursivly from a file location.

An example of the JSON object for append operation from the computation module's configuration file is given below.

```
{
    "operation": "append",
    "name": "tec",
    "fileLocation": "C:\\Documents\\correlation_project\\\
                    computations\\mean\\tec_ut",
    "fileFrequency": "yearly",

}
```

## Correlation Coefficient

Computation to find correlation coefficient using Pearson correlation coefficient method. following are the attributes and inputs for the correlation computation.

Attribute	Input Format	Functionality
operation	correlation	Compute pearson correlation coefficient of two data sets
name	string	name of the data set
data_1	file path	Location of the first data file
data_2	file path	Location of the second data file
data_1Attributes	string	Specific column name of the DataFrame or 'Null' to compute correlation for every column in the first dataset.
data_2Attributes	string	Specific column name of the DataFrame or 'Null' to compute correlation for every column in the second dataset.
windowSize	integer	Size of the rolling window for calculating correlation coefficient.

An example of the JSON object for correlation computation from the computation module's configuration file is given below.

```
{
    "operation": "correlation",
    "name": "tec_f10_7",
    "data_1": "C:\\\\Documents\\\\correlation_project\\\\
computations\\\\append\\\\tec\\\\tec.h5",
    "data_2": "C:\\\\Documents\\\\correlation_project\\\\
computations\\\\relative_difference\\\\f10_7\\\\f10_7.h5",
    "data_1Max": 100,
    "data_2Max": null,
    "data_1Attributes": null,
    "data_2Attributes": "f10_7",
    "windowSize": 90

},
```

## Confidence interval

Computation function to find 95% confidence interval of data values in a DataFrame column wise. Following is the attributes for the confidence interval computation.

Attribute	Input Format	Functionality
operation	confidence_interval	Compute confidence interval for each data value in the columns of the DataFrame.
name	string	Name of the data set
fileLocation	file path	Location of the data file

An example of the JSON object for confidence interval computation in the configuration file is given below.

```
{
    "operation": "confidence_interval",
    "name": "conf_int_corr_tec_f10_7",
    "fileLocation": "C:\\Documents\\correlation_project\\\
                    computations\\correlation\\tec_f10_7\\tec_f10_7.h5",
}
```

## Interpolation

Fill missing values values using linear interpolation method. The operation is applied column wise. Following are the list of attributes for the interpolation operation.

Attribute	Input Format	Functionality
operation	interpolate	Fill missing values by interpolation method.
name	string	Name of the data set
fileLocation	file path	Location of the data file

An example of the JSON object for interpolation operation in the configuration file is given below.

```
{
    "operation": "interpolate",
    "name": "tec_location_60_0_interpolated",
    "fileLocation": "C:\\Documents\\correlation_project\\\
                    computations\\extract_column\\tec_location_60_0\\\
                    tec_location_60_0.h5"
}
```

## Extract Column

Function to extract data values of a specific column from the DataFrame using name of the column. Attributes for the operation is listed below.

Attribute	Input Format	Functionality
operation	extract_column	extract specific column from the DataFrame and save as a file
name	string	name of the data set
fileLocation	file path	Location of the data file
fileFrequency	'yearly'	Data stored in yearly files which fetches based on analysis period recursively from a file location.
columns	'single' string	Data stored in single file Name of the column in the DataFrame to be extracted.

An example of the JSON object for extract column operation in the configuration file is given below.

```
{
    "operation": "extract_column",
    "name": "tec_location_60_0",
    "fileLocation": "C:\\\\Documents\\\\correlation_project\\\\
                    computations\\\\relative_difference\\\\tec",
    "fileFrequency": "yearly",
    "columns": "(60,0)"
}
```

## Alter data levels

Operator to reshape DataFrame by giving level name of the column or row. It stack or unstack given levels from row to column or vice versa. The attributes for the operation is listed below.

Attribute	Input Format	Functionality
operation	alter_data_level	Pivot data index levels of the DataFrame.
name	string	name of the data set
fileLocation	file path	Location of the data file
fileFrequency	'yearly'	Data stored in yearly files which fetches based on analysis period recursively from a file location.
alterOperation	'single'	Data stored in single file
	'stack'	Stack the prescribed level from columns to index.
	'unstack'	Unstack the prescribed level from index to column.
dataLevel	string	Specific index column name from multi-indexes of the DataFrame

An example of the JSON object for alter data level operation in the configuration file is given below.

```
{
    "operation": "alter_data_level",
    "name": "tec_15E_13UT",
    "fileLocation": "C:\\\\Documents\\\\correlation_project\\\\
                    computations\\\\interpolate\\\\tec_ut_13_interpolate\\\\
                    tec_ut_13_interpolate.h5",
    "fileFrequency": "single",
        "alterOperation": "unstack",
        "dataLevel": "ut",
}
```

The computation module can be executed using the following command, passing input as the configuration file of the module.

```
python computation.py ..\project_name\computations\  
computations_config.json
```

## Data Visualisation

Data visualisation module is to visualise data in the form of graphs using Matplotlib library of Python. The module takes DataFrames as input files through configuration files to generate different types of plots. The plots generated are saved as image in the plot directory.

### Line plots

Create a line graph showing Y vs X depending on the data that was provided. Following are the attributes for generating line plots.

## Appendix

---

Attribute	Input Format	Functionality
name	string	File name of the plot being generated
title	string	Title of the plot.
xLabel	string	X axis label
yLabel	string	Y axis label
width	integer	Width of the plot
height	integer	Height of the plot
dataFiles	array of JSON object	Data files and other attributes of the data.
filePath	file path	Data file path of the data being plotted
xAxis	column name	Name of X axis column from the DataFrame
yAxis	column name	Name of Y axis column from the DataFrame
color	string	Color indication of the data being plotted. Accepts all Python color codes
label	string	Legend label for the data.
type	'line'	Plots a line graph based on the data column given.
	'fillBetween'	Plots two data lines and fills the inbetween area to show confidence interval. Data files are attributed as 'lowerLimitFile' and 'upperLimitFile' instead of 'filePath'.

---

An example of the JSON object for line plot in the configuration file is given below.

```
{  
    "name": "Cross correlation tec x f10",  
    "title": "Cross correlation tec x f10 at latitude 60N",  
    "dataFiles": [  
        {  
            "filePath": "C:\\Documents\\correlation_project  
                \\\computations\\correlation\\tec_f10_7\\  
                tec_f10_7.h5",  
            "xAxis": "date",  
            "yAxis": 60.0,  
            "color": "b",  
            "label": "Cross correaltion with box window",  
            "type": "line"  
        },  
        {  
            "lowerLimitFile": "C:\\Documents\\  
                correlation_project\\computations\\  
                confidence_interval\\conf_int_corr_tec_f10_7  
                \\\conf_int_corr_tec_f10_7_lowerConfInterval.  
                h5",  
            "upperLimitFile": "C:\\Documents\\  
                correlation_project\\computations\\  
                confidence_interval\\conf_int_corr_tec_f10_7  
                \\\conf_int_corr_tec_f10_7_upperConfInterval.  
                h5",  
            "xAxis": "date",  
            "yAxis": 60.0,  
            "color": "lightgray",  
            "label": "95% confidence interval",  
            "type": "fillBetween"  
        }  
    ],  
    "xLabel": "Day of year",  
    "yLabel": "Correlation",  
    "width": 12,  
    "height": 4  
}
```

## Color mesh

Generates 2D raster image from the input DataFrame using given color scheme.

Attribute	Input Format	Functionality
name	string	File name of the plot being generated
title	string	Title of the plot.
xLabel	string	X axis label
yLabel	string	Y axis label
zLabel	string	Color bar label
width	integer	Width of the plot
height	integer	Height of the plot
dataFiles	array of JSON object	Data files and other attributes of the data.
filePath	file path	Data file path of the data being plotted
numXTicks	integer	Number of tick points on the X axis
numYTicks	integer	Number of tick points on the Y axis
colorMap	string	Colormap name of the data being plotted. Accepts all Python colormap names. Default colormap is 'seismic'.
label	string	Legend label for the data.
type	'colorBar'	Display data as a 2D regular raster image using a color scheme.

An example of the JSON object for color mesh plot in the configuration file is given below.

```
{  
    "name": "Correlation map tec x f10 with gaussian window  
        ",  
    "title": "Correlation map tec x f10 with gaussian  
        window",  
    "dataFiles": [  
        {  
            "filePath": "C:\\Documents\\\\correlation_project  
                \\\computations\\\\mean\\\\  
                seasonal_mean_norm_corr_tex_x_f10_7\\\\  
                seasonal_mean_norm_corr_tex_x_f10_7.h5",  
            "type": "colorBar",  
            "numXTicks": 12,  
            "numYTicks": 10  
        }  
    ],  
    "xLabel": "Month of year",  
    "yLabel": "Latitude",  
    "zLabel": "Correlation",  
    "width": 12,  
    "height": 8  
}
```

The data visualisation module can be executed using the following command, passing input as the configuration file of the module.

```
python plot_data.py ..\\project_name\\plot_data\\plot_config.  
json
```