# A Study of Extensive Radio Occultation Data from 2007-2015 to Understand and Model Short and Long-term Variability in Ionospheric Response to Solar Forcing

Project Report Submitted for the Partial Fulfilment of the Degree of

Master of Science 2017 in Physics of

Presidency University


## Arka Mitra

M.Sc. Physics, PG Year-2,
Registration No.- 12120911023,
The Department of Physics,
Presidency University, Kolkata.


**Project Guide**


## Dr. Tulasi Ram Sudarsanam

Upper Atmospheric Division,
Indian Institute of Geomagnetism (IIG), Mumbai.


**Institutional Guide**


## Dr. Debasish Datta,

The Department of Physics,
Presidency University, Kolkata.


05/05/2018

# Table of Contents

## Section 1. Data Analysis



## Section 2. Global Ionospheric Model

## Section 1.

# Data Analysis

Research on the Sun – Earth interactions (i.e., electromagnetic radiation and transient and recurrent solar wind emissions from Sun and their impact on Earth's magnetosphere – ionosphere – atmosphere) is vital for forecasting disturbances in the Earth's near space environment that has direct effect on the well-being of astronauts and space missions, as well as communication systems and ground-based power grids. For our present interest, the ionosphere is however, the most important layer.

# 1. The Ionosphere

The ionosphere is the upper region of the Earth's atmosphere where a small but significant number of the neutral atoms are ionized, resulting in free electrons and ions (a plasma). The ionization levels in this near-Earth space plasma are controlled by solar extreme ultraviolet (EUV) radiation and particle precipitation. It lies 75-1000 km (46-621 miles) above the Earth. (The Earth's radius is 6370 km, so the thickness of the ionosphere is quite tiny compared with the size of Earth.) Because of the high energy from the Sun and from cosmic rays, the atoms in this area have been stripped of one or more of their electrons, or "ionized," and are therefore positively charged. The ionized electrons behave as free particles. The Sun's upper atmosphere, the corona, is very hot and produces a constant stream of plasma and UV and X-rays that flow out from the Sun and affect, or ionize, the Earth's ionosphere. Only half the Earth's ionosphere is being ionized by the Sun at any time. During the night, without interference from the Sun, cosmic rays ionize the ionosphere, though not as strongly as the Sun.

The ionosphere has major importance to us because, among other functions, it influences radio propagation to distant places on the Earth, and between satellites and Earth. For the very low frequency (VLF) waves that the space weather monitors track, the ionosphere and the ground produce a "waveguide" through which radio signals can bounce and make their way around the curved Earth.

The ionosphere varies in systematic ways because the main source of ionization − solar UV and X-ray intensity − depends on the position of the Sun in the sky at a particular location on Earth and on the Sun's absolute output. When the Sun is directly overhead, the intensity of sunlight reaching the upper atmosphere is greatest. As the observer either moves towards the poles or to the day−night terminator, the intensity decreases because the angle the Sun makes with the upper atmosphere is more oblique. As the observer moves into the dark or night side hemisphere of Earth, the amount of sunlight goes to zero and production due to photoionization ceases. The rotation and curvature of Earth therefore give rise to variations in the ionospheric structure.

In addition, the Sun's output of energy is not constant in time. It changes rapidly (especially at the high-energy end of the electromagnetic spectrum) due to solar flares and over the solar cycle. During solar minimum, there is little X-ray emission, while at solar maximum the Sun's atmosphere emits large amounts of X-rays. This gives rise to a solar cycle variation in the intensity of ionization of the ionosphere. During solar storms, the ionospheric structure can be drastically

modified by the energy input from the Sun. Therefore, during geomagnetic storms the ionosphere becomes most disturbed and the most space weather impacts are noted.

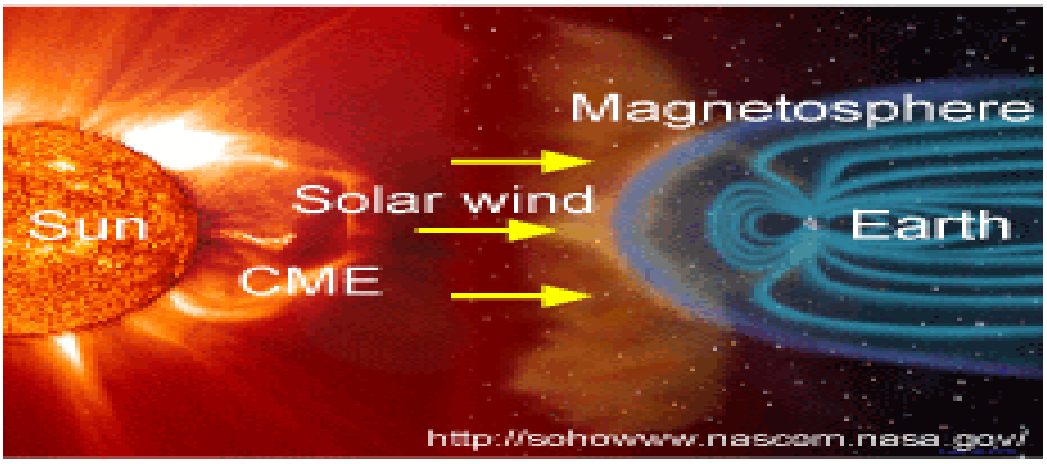



*Figure 1. Representative Diagram of the earth's geo-space. Coronal Mass Ejections (CMEs), Coronal holes and solar flares are sources of high-energy particles and geomagnetic disturbances. Solar effects have terrestrial effects when they interact with the Earth's ionosphere via Ionosphere-magnetosphere coupling. (Courtesy: NASA)*

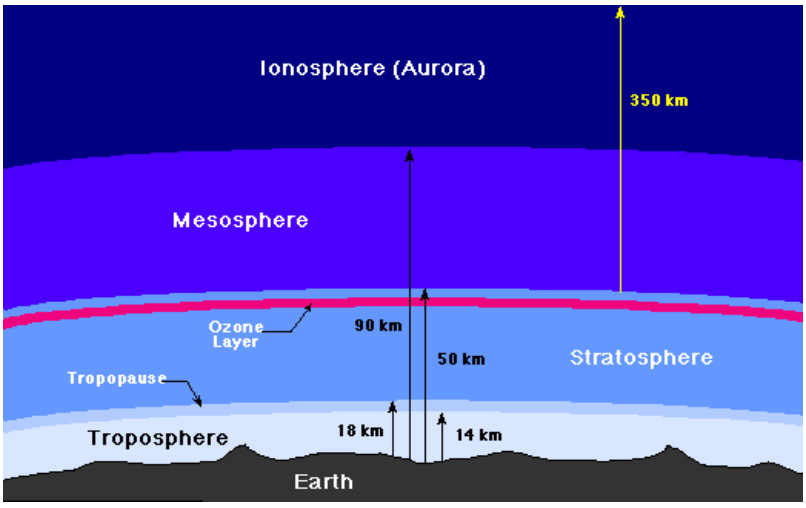



*Figure 2. Representative diagram of the extent of the layers of the Earth's atmosphere. The outer extent of the Earth's ionosphere is very poorly defined and varies wildly with location and time to an upper extent of nearly 1000km. From the top of the ionosphere to a distance of about 10,000km is a region known as the exosphere. (Courtesy: The University of Rochester, Department of Physics)*

The ionosphere is composed of three main parts, named for obscure historical reasons: The **D, E, and F regions**. The electron density is highest in the upper, or F region. The D region disappears

during the night compared to the daytime, and the E region becomes weakened. The F region exists during both daytime and nighttime. During the day, it is ionized by solar radiation, during the night by cosmic rays. For our present purposes, we shall only focus on the F region.

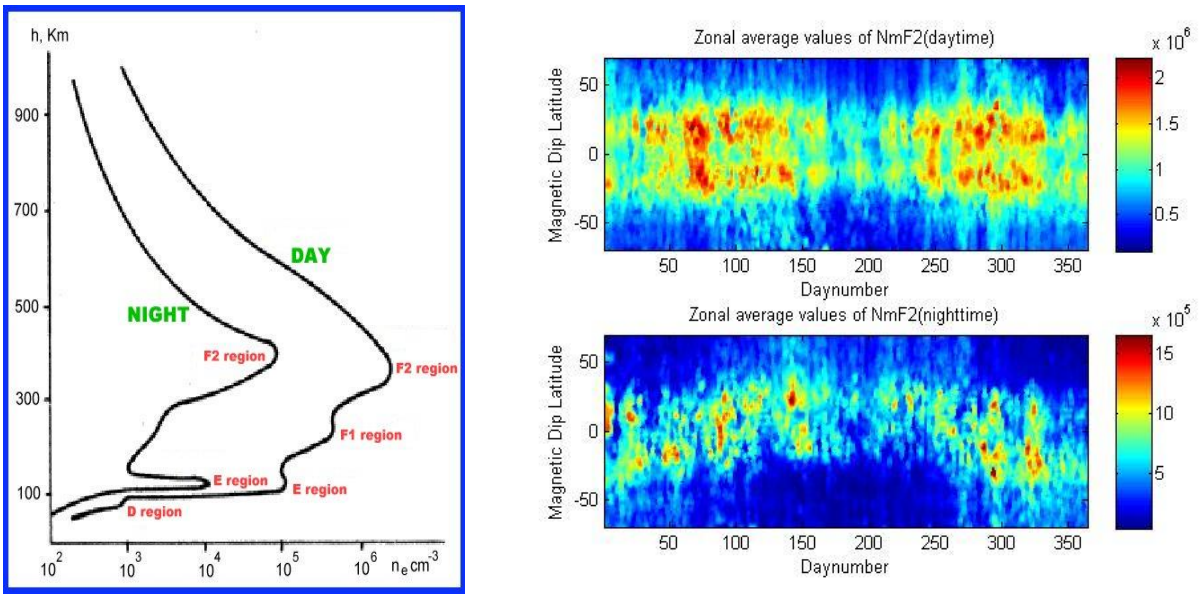



*Figure 3: (Left) Diagram showing the most distinctive structure of the layers of the ionosphere during the day and nighttime. Most distinctive and noteworthy are the splitting of the F-region into the F1 and F2 regions during the day due to increased ionization and the disappearance of the D-region during the night due to recombination. (Courtesy: University of Texas, Dallas).*

*(Right) A typical zonally-averaged contour plot of NmF2 values during the day- and nighttime over the course of one year (2012), as prepared in MATLAB (all values are in electrons per cc). The most distinctive feature of the daytime ionospheric plot are the high values of NmF2on both sides of the equator (a phenomenon known as the Equatorial Ionospheric Anomaly, or the EIA) and the most distinctive feature of the nighttime NmF2 plot is a distinctive bow-like structure.*

The location of these layers varies by day and night and is shown in the figure above. The lower most region of the ionosphere extending from about 50 km to 90 km is the D-region, which principally absorbs radio waves. Above the D layer is the E-region extending from 90 km to 150 km. The peak in the E region during day time is seen near 110 km. Above the E region is the F region consisting of two parts: the lower $F_1$ region between 150 km and 180 km and the $F_2$ region from 180 km and above. Note, the daytime densities are much larger than the nighttime densities. At night, recombination can result in the loss of the D region.

The region of Earth receiving the Sun's direct rays is the equator. So, the incoming solar flux is high at the equator than other latitudes. So, the maximum production of charged particles is at the equator. But there are interesting and important tropical ionospheric effects arising from equatorial

electrodynamics. Global wind circulations are due to the non-uniform heating of sun. Solar radiation varies with latitude as well as longitude. So, there is temperature gradient followed by pressure gradient always present globally and hence, there is always a flow of global winds from dusk to dawn. Response of electrons and ions are different to global winds leading to zonal electric field which is eastward during day and westward during night.

Here, two important concepts need to be introduced as they will be indispensable to our present discussion- the highest electron density value at a given time and location in the $F_2$ region ($\mathbf{N_mF_2}$) and the height above the Earth's surface where this highest electron density occurs ($\mathbf{H_mF_2}$). Changes in these two quantities shall be used in this work as representative of changes in the electron profile of the ionosphere.

## 1.1. Solar Effects on the Ionosphere

Astronomically, the Sun is a rather ordinary star, however it is this very fact that has led to Earth's stable atmosphere, availability of water in all three phases and hence, led to the existence of life as we know it. Long-term changes in the lower strata of the atmosphere has been studied (such as the change to modern times from the Ice Ages), however natural short-term changes to the troposphere and stratosphere are much less drastic and seasonal. However, for the upper atmosphere, where most of the more energetic solar radiations are stopped and which is heated by them, is much more responsive to solar activity variations in general, as well as to the short-lived, intense and localized outbursts of solar flares.

We investigate solar activity under the purview of two main phenomena- namely, the 11-year solar cycle in sun's activity that comes with rise and fall in the level of solar radiation (represented in our work through the **f10.7 solar flux index**) and energetic activity. We specifically focus on the solar minima, which is associated with the manifestation of Coronal Holes (CHs) on large areas of the solar surface. CHs are observed to be dark patches on X-ray and EUV images of the sun and are unipolar regions that have been shown [*Krieger et al.*, 1973] to be the source of high-speed solar wind streams (HSSs), which in turn have been shown to force geomagnetic and ionospheric properties [*Tulasi Ram* et al, 2010]. The impact of solar wind speeds on the ionosphere is represented through the Global Geomagnetic Storm index, also known as the **Kp-index**.

The ionosphere is embedded within the Earth's magnetic field (magnetosphere) and thus is constrained by interactions of the ionized particles with the magnetic field. At middle and low latitudes, the ionosphere is contained within a region of closed field lines, whereas at high latitudes the geomagnetic field can reconnect with the interplanetary magnetic field and thus open the ionosphere to the driving force of the solar wind. Solar storms ultimately have their major terrestrial impact when they encounter the ionosphere through magnetosphere/ionosphere coupling. The ionosphere responds to magnetospheric feedback in a number of different ways, with changes in electron and ion temperature, electron and ion and neutral density and the mixing with the neutral atmosphere resulting in changes to the ionic species. All of these are important physical parameters, but arguably, the most important is the electron density because it governs all of the effects on radio signals and electron profiles are what we focus on.

# 2. Radio Wave Propagation Through the Ionosphere

Subjected to an external electric field from a radio signal, the free electrons and ions (produced through ionization from solar radiation) will experience a force and be pushed into motion. However, since the mass of the ions is much larger than the mass of the electrons, ionic motions are relatively small and will be ignored here.

Radio waves below 40 MHz are significantly affected by the ionosphere, primarily because radio waves in this frequency range are effectively reflected by the ionosphere. The E and F layers are the most important for this process. For frequencies beyond 40 MHz, the wave tends to penetrate through the atmosphere versus being reflected. The major usefulness of the ionosphere is that the reflections enable wave propagation over a much larger distance than would be possible with line-of-sight or even atmospheric refraction effects. This is shown graphically in Figure 4 (Left). The skip distance $d_{max}$ can be very large, allowing very large communication distances. This is further enhanced by multiple reflections between the ionosphere and the ground, leading to multiple skips. This form of propagation allows shortwave and amateur radio signals to propagate worldwide. Since the D layer disappears at night, the best time for long-range communications is at night, since the skip distance is larger as the E, and F regions are at higher altitudes.

Where does the reflection come from? The reflections from the ionosphere are actually produced by refraction as the wave propagates through the ionosphere. The ionosphere is a concentrated region highly charged ions and electrons that collective form an ionized gas or plasma. This gas has a dielectric constant that is a function of various parameters, including the electron concentration and the frequency of operation. We now derive the dielectric constant of a plasma. The electric field will produce a force on a given electron and displace it along a vector $\vec{r}$ as shown

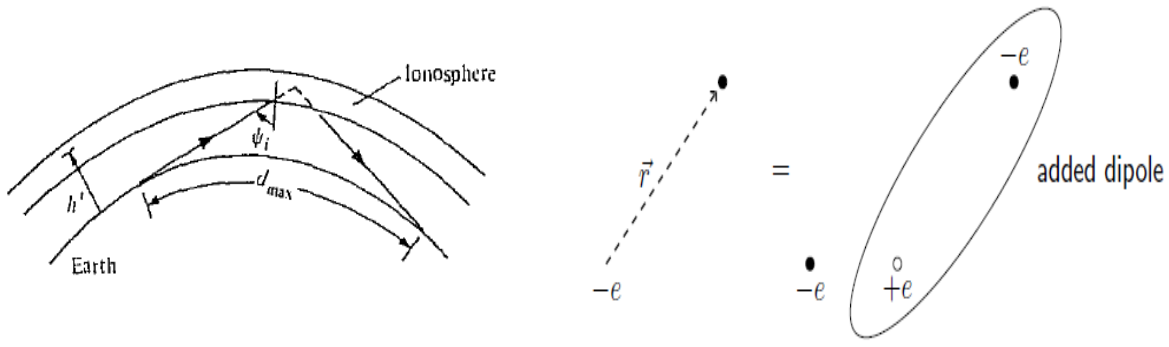



*Figure 4. (Left) A single skip of a radio wave in the ionosphere. (Right) Moving electron as a dipole (Courtesy: Prof Sean Victor Hum, ECE422)*

on the left side of Figure 4 (Right). The displacement of an electron along this path can be modelled in an equivalent situation where the original electron remained stationary and an equivalent electric dipole is added, as shown in the right half of the figure. The dipole moment of this dipole is then equal to $-e\vec{r}$. If there are N electrons per unit volume, each displaced by $\vec{r}$ on average, then the volume polarization is

$$\vec{P} = -Ne\vec{r} \qquad (1.1)$$

The equation of motion for a single electron of mass me = $9.109 \times 10^{31}$ kg, charge e =$1.6021 \times 10^{19}$ C, with velocity $\vec{v} = \frac{d\vec{r}}{dt}$, and acted upon by an electric field $\vec{E}$, is

$$m_e \frac{dv}{dt} = -e\vec{E} \qquad (1.2)$$

The electron also experiences a frictional force resulting from collisions with neutral molecules. This force is added to the electric field force above, yielding

$$m_e \frac{dv}{dt} = -e\vec{E} - \upsilon m_e \vec{v} \qquad (1.3)$$

where υ is the electron collision frequency. Re-writing the equation in terms of r⃗,

$$m_e \frac{dv}{dt} = -e\vec{E} - \upsilon m_e \frac{d\vec{r}}{dt} \qquad (1.4)$$

We know that for sinusoidal fields, we can write equations in terms of phasors and replace d/dt with $j\omega$. Hence, we can write the equation of motion on the electron in terms of phasors as

$$-\omega^2 m_e \vec{r} = -e\vec{E} - j\omega\upsilon m_e \vec{r} \qquad (1.5)$$

or,

$$\vec{r} = \frac{e\vec{E}}{\omega^2 m_e - j\omega\upsilon m_e} = \frac{e\vec{E}}{m_e\omega^2\left(1 - j\frac{\upsilon}{\omega}\right)} \qquad (1.6)$$

Substituting this into (1.1),

$$P = \frac{Ne\vec{E}}{m_e\omega^2\left(1 - j\frac{\upsilon}{\omega}\right)} \qquad (1.7)$$

The electric flux density can then be found as

$$\vec{D} = \epsilon_0 \vec{E} + \vec{P} = \epsilon_0 \vec{E} - \frac{Ne\vec{E}}{m_e\omega^2\left(1 - j\frac{\upsilon}{\omega}\right)} \equiv \epsilon_r \epsilon_0 \vec{E} \qquad (1.8)$$

The effective relative dielectric constant of the plasma is then

$$\epsilon_r = 1 - \frac{Ne^2}{m_e\omega^2\epsilon_0\left(1 - j\frac{\upsilon}{\omega}\right)} \qquad (1.9)$$

An *angular plasma frequency* can be defined such that

$$\omega_P^2 = \frac{Ne^2}{m_e\epsilon_0} \approx 3183N \qquad (1.10)$$

which is purely a function of electron density N. Then,

$$\epsilon_r = 1 - \frac{\omega_P^2}{\omega^2\left(1 - j\frac{\upsilon}{\omega}\right)} \qquad (1.11)$$

We see that in the presence of electron collisions, the dielectric constant can in general be complex. If we ignore collisions for the moment, then

$$\epsilon_r = 1 - \frac{\omega_P^2}{\omega^2} \approx 1 - \frac{80.6N}{f^2} \qquad (1.12)$$

From this result, we can make several important observations. Since the propagation constant of a wave travelling in a plasma is $\sqrt{\epsilon_0}\ k_0$

- For frequencies $\omega > \omega_p$, the effective dielectric constant is less than unity but the propagation constant is real. Hence, the wave will be refracted by the plasma according to the variation of $\epsilon_r$ with altitude.

- For frequencies $\omega < \omega_p$, we get a negative value for the dielectric constant, which leads to an imaginary propagation constant. Hence, a plane wave in the medium will decay exponentially with distance. It is not absorbed (we have ignored losses/electron collisions here), but instead becomes evanescent, like a waveguide in cutoff. A wave incident on a medium with this propagation constant would be totally reflected.

- For frequencies $\omega > \omega_p$, the effective dielectric constant is essentially 1. Practically this happens at VHF frequencies and above. The waves simply pass through the plasma without significant refraction, but there can be other effects, especially if the plasma is magnetized by the Earth's magnetic field and the medium becomes anisotropic. Waves at these frequencies undergo Faraday rotation by the ionosphere, whereby the polarization vector is rotated as the wave passes through the atmosphere.

# 3. Data

Earlier methods like ionosonde or radar imaging had the prime disadvantage of being able to map electron content only of a much-localized region. However, a newer, novel approach called the **GPS Radio Occultation Technique** gives us continuous, global coverage of total ionospheric electron content (TEC) data. In this method, GPS receivers onboard the low earth orbiting (LEO) satellites can receive the occulted signals from the GPS satellites, in medium Earth orbits. During an occultation measurement, a GPS receiver in LEO observes the rise or set of GPS signals behind the earth's "limb" while signal slices the earth's atmosphere and ionosphere. The GPS receiver in LEO records the change of the delay of the signal between the GPS and the LEO satellite that is related by slowing and bending of the signal path. The bending of the radio waves is determined through precise measurements of the phase changes and time delay and used to compute bending angle, refractivity and other products at high vertical resolution. The refractivity allows for reconstruction of the electron density in the ionosphere (by a process known as **Abel Transformation**).

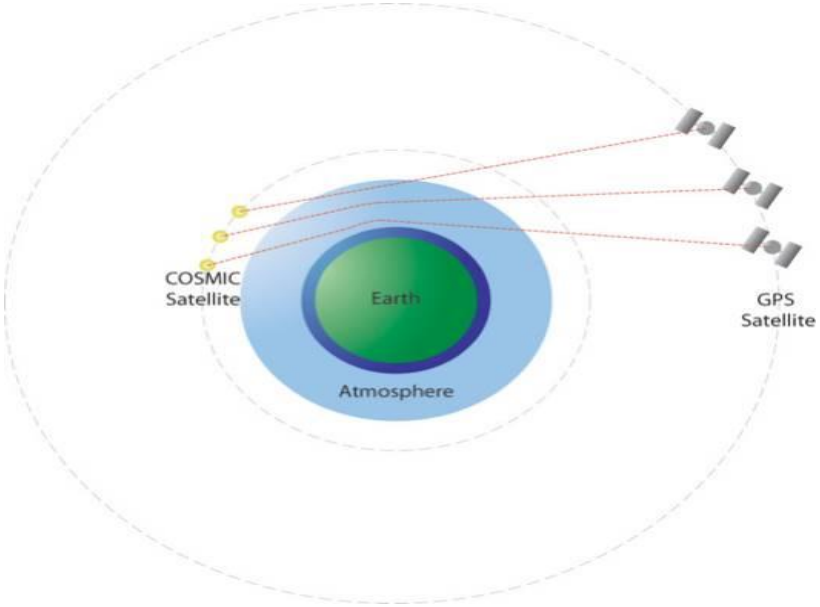



*Figure 5.* ***Radio Occultation imaging technique.*** *As the Low Earth Orbiting (LEO) satellite carrying a GPS receiver rises or sets behind Earth, a series of scans of Earth's atmosphere is obtained. (Courtesy: University of Colorado, Center for Astro-dynamics)*

## 3.1. Brief Description of the Radio Occultation technique and calculation of Ionospheric Total Electron Content (TEC)

COSMIC satellite is provided with four antennas each, with which ionospheric electron density measurements are taken. Two antennas are used- one for rising and one for setting occultation during this period. The GPS satellite receiver onboard LEO can track up to 13 GPS satellites through the two antennas and measure the difference GPS phase data on $L_1$ and $L_2$ every second. The bend angle, $\alpha$ (of the satellite or radio signals) obtained through Abel inversion method is used to compute the refractivity. The radio occultation electron density profile retrieval approach is based on a few assumptions and approximations such as straight line signal propagation, spherical symmetry of electron density, circular LEO orbit and first order estimation of orbital electron density. Even though radio occultation technique is characterized with high precision, provides global profiles and maps of the atmospheric boundary layer, no satellite to satellite bias, self-calibration, independent of processing center; yet, the spherical symmetry assumptions used in Abel inversion is known to be the most significant source of error. TEC calculation involves the integration of electron density along the straight line (signal path) between the LEO receiver and GPS transmitter (as demonstrated in Equation (11)). Measuring the phase delay of radio waves at $L_1$ and $L_2$ frequencies from GPS as they are occulted by the earth atmosphere is illustrated in **Figure 1(a)** and note that the sample rate per second is 1 Hz. The simulated STEC (Slant TEC) is then inverted into electron density profile along the tangent points by the same Abel inversion software package used for CDAAC electron density profiles retrieval.

### 3.1.1. Ionospheric Bending and Refractivity

The refractivity of the radio signals passing through the ionosphere depends majorly on the amount of electron density and ions, time of the day and carrier frequency. It is expressed as shown in Equation (1.12) in which the refractive index is proportional to the electron density and inversely proportional to the square of the carrier frequency:

$$n^2 - 1 = -80.6 \times 10^6 N / f^2 \qquad (1.13)$$

where the index of refraction is denoted with $n_2$, $N$ is the electron density (el $cm^{-3}$) and $f$ is the signal frequency (Hz).

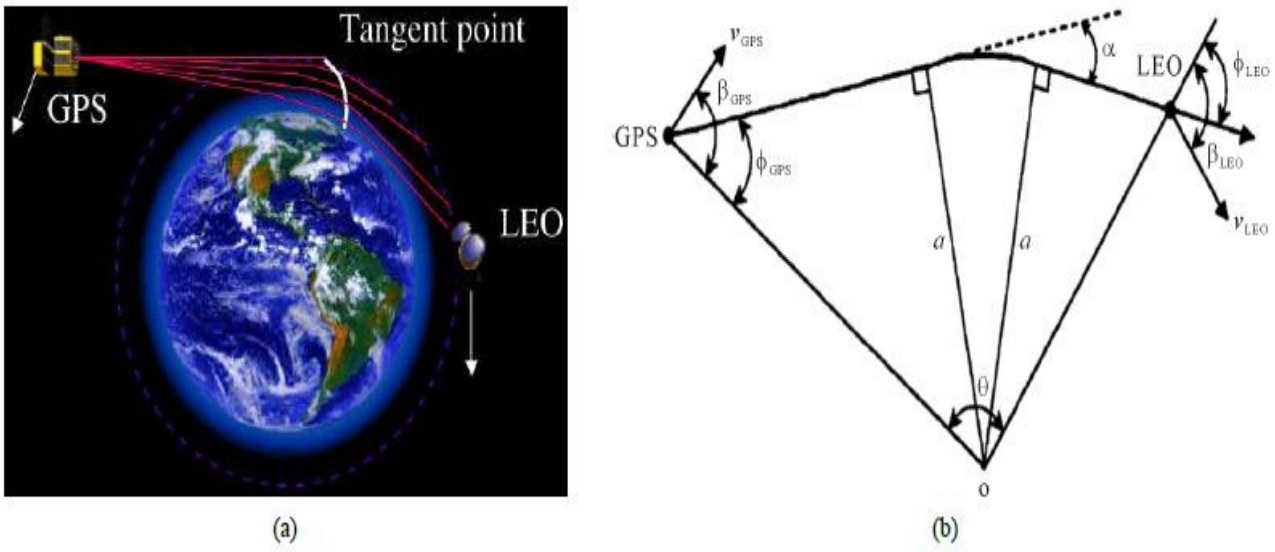



*Figure 5. Schematic diagram illustrating radio occultation, geometric of ray, bending angle α and impact parameter a*

As mentioned earlier, the constraint problem when constructing a 3-D refractivity field from 1-D observational data is the assumption; spherical symmetry of refractivity around the extended region of the perigees of the sounding rays. The Snell's law is obtained when the 3-D ray equations are integrated under this assumption.

Therefore, the bending angle $\alpha$ and excess phase $s$ are given as follows:

$$\alpha(a) = -a\left[\int_a^{xGPS} + \int_a^{xLEO}\right] \frac{\frac{dn}{dx}}{n(x)\sqrt{x^2-a^2}}\, dx \qquad (1.14)$$

$$S(a) = \left[\int_a^{xGPS} + \int_a^{xLEO}\right] \frac{x\left(1-xn^{-1}\frac{dn}{dx}\right)}{\sqrt{x^2-a^2}}\, dx - L_{GL} \qquad (1.15)$$

where $a$ is defined as the impact parameter ($a = \rho n(\rho)$), $\rho$ is the distance from the ray tangent point to the center of sphericity of the refractivity field, $x = rn(r)$ represent the fractional radius ($r$ is radius) and $L_{GL}$ denotes the distance between the GPS and LEO satellites. Figure 5(b) shows the bending angle $\alpha$ and impact parameter $a$ that are used in Equations (1.14) and (1.15) as well as the ray geometry.

Ionospheric refractivity is negligible at GPS orbit altitudes in the first term of Equation (1.14) while the second term (LEO) is not. The bending angle does not depend on the positions of the GPS and LEO satellite, excess phase and ray separation from a straight line connecting the satellites. Calculations based on Figure 5(b) were done by numerical integration of Equations (1.14) and (1.15) where $L_1$ frequency ($f = f_1$) and $L_2$ ($f = f_2$).

## 3.1.2. Abel Inversions

In this section under the assumption of spherical symmetry, the formulations for the construction of electron density profile through the use of bending angle data and excess phase (TEC) data was conducted.

## 1) Abel Inversion Obtained from Bending Angle Data

To determine the bending angle from excess phase, the relationship between Doppler shift of the carrier frequency, $f_d = -f_c^{-1}$ (d*S*/*dt)* and the projections of satellite velocities at the GPS and LEO positions is as follows:

$$f_d = f\left[\frac{c - n_{LEO}v_{LEO}\cos(\beta_{LEO} - \varphi_{LEO})}{(c - n_{GPS}v_{GPS}\cos(\beta_{GPS} - \varphi_{GPS}))}\right] \qquad (1.16)$$

For GPS $v'_{GPS}$ and LEO $v'_{LEO}$ represent 2-D projections of GPS and LEO satellite 3-D velocities respectively on the occultation plane, *c* is defined as the velocity of light in a vacuum, angles $\beta$ and $\varphi$ are defined in Figure 5(b) while $n_{LEO}$ and $n_{GPS}$ denotes the indices of refraction. It is observed that Equation (1.16) has insufficient information to solve for $\varphi_{GPS}$ and $\varphi_{LEO}$ as it contains only Doppler data. Based on the assumption made, Snell's law is incorporated as shown below to complement Equation (1.16):

$$r_{GPS}n_{GPS}\sin(\varphi_{GPS}) = r_{LEO}n_{LEO}\sin(\varphi_{LEO}) \qquad (1.17)$$

Solving both Equations (1.16) and (1.17) iteratively to obtain the impact parameter, *a* and the bending angle, $\alpha$ with the aid of Equation (1.18):

$$\alpha = \varphi_{GPS} + \varphi_{LEO} + \theta - \pi \qquad (1.18)$$

Solving Equations (1.17) and (1.16) causes an error due to the assumption of the refractivity at GPS and LEO positions to be unity.

## 2) Abel Inversion Obtained from Total Electron Content Data

Total electron content is a measure of propagation delay time of the radio signal transmitted from the satellite to the receiver. The total electron content (TEC), *T* along the signal path can be related to electron density *N*, excess phase *S* and index of refraction *n* as:

$$T = \int Ndl = -\frac{f}{40.3\times10^6}\int(n-1)dl = -\frac{f^2S}{40.3} \qquad (1.19)$$

*S* is measured in metres (m) and *T* in (TEC (0.1 TECu).

If bending is neglected, TEC can be obtained from $S_1 - S_2$ ($L_1 - L_2$):

$$T = -\frac{(S_1 - S_2)f_1^2 f_2^2}{40.3(f_1^2 - f_2^2)} \qquad (1.20)$$

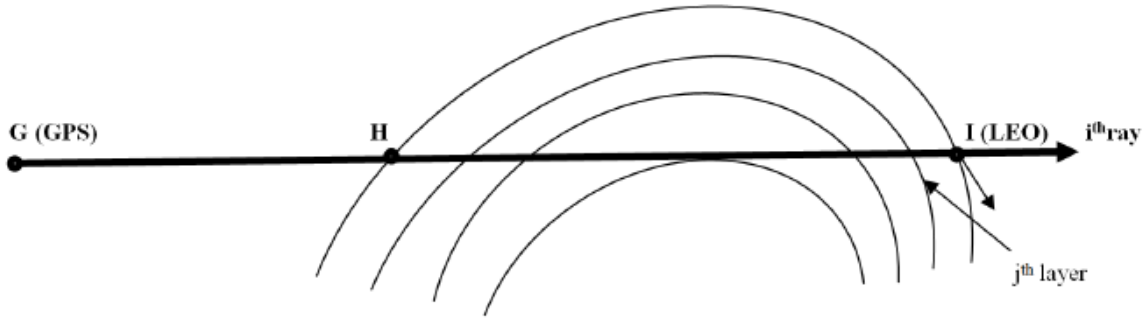



*Figure 6. Total electron content and bending angle calibration geometry diagrammatic illustration.*

It is advantageous using $(S_1 - S_2)$ or $(L_1 - L_2)$ to estimate TEC as it automatically eliminates the orbit and clock errors from the difference while the disadvantage is that when anti-spoofing is activated $L_2$ noise is introduced which may degrade the TEC inversion results. Just like the inversions through bending angles using the assumption of spherical symmetry, it is thus adopted in electron density for TEC inversions. Then, TEC in Equation (1.19) is related to electron density as shown:

$$T(r_0) = [\int_{r_0}^{r_{GPS}} + \int_{r_0}^{r_{LEO}}] \frac{r\,N(r)}{\sqrt{r^2 - r_0^2}} dr \qquad (1.20)$$

$r_0$ is defined as the impact distance of the straight line connecting the GPS and LEO satellites. Hence, TEC data calibration is done along the ray path as demonstrated in Figure 6 and expressed as:

$$\dot{\mathrm{T}}\,(r_0) = T_{HI} = T_{GI}(r_0) - T_{GH}(r_0) \qquad (1.21)$$

The calculation of $r_0$ (the straight-line impact distances) is carried out for all observational data with respect to the point where there is maximum impact distance. Then, the interpolation of uncalibrated TEC as a function of impact distance with the aids of cubic splines. Calibration of TEC is the next stage, *i.e.* $\dot{\mathrm{T}}\,(r_0)$ . Now, the calibrated TEC Equation becomes:

$$\dot{\mathrm{T}}\,(r_0) = 2\int_{r}^{r_{LEO}} \frac{r\,N(r)}{\sqrt{r^2 - r_0^2}} dr \qquad (1.22)$$

## 3.2. COSMIC Data

Ionospheric Electron Content Data for our present study has been downloaded from the UCAR COSMIC Data Analysis and Archive Center (CDAAC) site (http://cdaac-www.cosmic.ucar.edu/cdaac/products.html). COSMIC is a constellation of six micro satellites orbiting around 800 km altitude, 72° inclination and 30° separation in longitude. Each F3/C satellite has GPS occultation observations in the ionosphere that provides ~600 vertical electron density profiles per day, distributed uniformly around the globe.

The ionospheric level 12 operational data products are made available within 24 hours by UCAR/CDAAC. The data products output used in this paper are ionospheric profiles of electron density (ionprf) whose accuracy is generally about 104 - 105 $cm^{-3}$ and in NetCDF format. Observation shows that, cycle slips may affect some of these profiles [6]. UCAR/CDAAC is the only primary data processing centre particularly dedicated for COSMIC mission. This data is made available to the public both ftp and http server through CDAAC.

One of the most important radio occultation products is the electron density profile for ionosphere. It is achieved from the calculation of excess phase file for each occultation with the aid of Abel inversion (described in the previous section).

Unpacking the NetCDF files (which amounted to somewhere between 300 to 800 a day, depending on space weather and satellite conditions), we subjected the data through multiple conditional filters in MATLAB to obtain our initial database after restricting the values of $H_mF_2$ to values within 200 and 500 km (to prevent excessive fluctuations from becoming a factor).

All latitudes are converted to their magnetic inclination angle counterparts (commonly called dip latitudes) using the "igrf11magm" method in MATLAB. Whenever required, we have calculated the zonal and meridional winds using Horizontal Wind Model (HWM-14) and the Inclination/Declination values using the IGRF14 model. All Universal Time values (UT, as given in original data) are converted to their corresponding Local Time values (LT) using the relation: $\boldsymbol{LT = UT + Latitude/15.0}$. We made sure that after the application of the conditional filters, the remaining number of 'good" profiles were sufficient enough to carry forward our investigation.

Daily averaged solar data (Solar Wind speeds, Kp-index and f10.7 solar flux index data) has been downloaded and collated from NASA's NSSDC OMNIWeb interface (http://nssdc.gsfc.nasa.gov/omniweb/).

# 3. Methodology, Results and Proposed Further Work

**(a)** As expected from physical considerations, when zonally averaged values of daytime (6.0<LT<18.0) $N_mF_2$ and $H_mF_2$ were plotted with respect to time in contours of geographical longitudes and magnetic dip latitudes [Fig 7], the higher values of either quantity correspond to the period 2011-2015, which also corresponds to the period of solar maximum, while the period from 2007-2010 is reflected in the contours plots as low values of $N_mF_2$ and $H_mF_2$. However, the same solar minimum period (as can be seen from the bottommost plot) is also associated with very high values of solar wind speeds. As documented in previous literature, this phenomenon is due to the prevalence of coronal holes during solar minima [*Tulasi Ram et al., 2008*].

**(b)** Daytime values of $N_mF_2$ from within a dip-latitude bin of ±5º on both sides of the zero dip-latitude (i.e., equatorial $N_mF_2$) were daily-averaged for each day under consideration and each year's $N_mF_2$ values were then subjected to the Lomb-Scargle (LS) periodogram technique, and the LS amplitude spectra of NmF2 for each year were, in turn, contrasted with the LS amplitude spectra for f10.7 solar flux index and solar wind speed values [Fig 8]. While the predominant periodicity is of 27 days (which corresponds to the rotation of the Sun around its own axis), the other periods of 13.5 days, 6.8 days (sub-harmonic periodicities of the 27-day period) and the 9 day-periodicity (most prominent in the 2008 spectrum) are the most noteworthy, and is pertinent to further study.

**(c)** There are traces of sub-harmonic periodicities of periodicities greater than the solar periodicity (i.e., greater than 27 days) in the periodograms obtained, which lead to spurious peaks. To get rid of these effects, the equatorial $N_mF_2$ signal is subjected to a low-pass filter by subtracting its 30-day running mean from the overall signal, which removes all the periodicities greater than 30 days. [Fig. 9]

**(d)** This truncated signal is then subjected to a Morlet Wavelet transform to check the epochs in which particular periodicities are most predominant and how such changes relate to the solar cycle. In the Morlet wavelet periodogram, the 27-day periodicity is constantly the most prominent, often showing a sharper and a less intense peak in the same year, which on comparing with the corresponding time-series plot of 27-day solar flux signal, shows a one-to-one correspondence. We find that over the course of a solar cycle, the annual twin peaks (maxima) in both the signals shows a continuous uptick as well as a marked increase in the magnitude difference between the peaks, themselves. The peaks themselves are flipped in phase over 2009 and again during 2013, which is correlated with the solar forcing behind it. [Fig. 10]

**(e)** Clearly, to understand the physics of ionospheric peak parameters and hence of the ionization itself, one has to understand the physics of the solar cycle itself, which unfortunately, is not so. However, one way to effectively predict future ionospheric conditions based on epoch, solar activity and geographic location is worked on and presented in the next section.

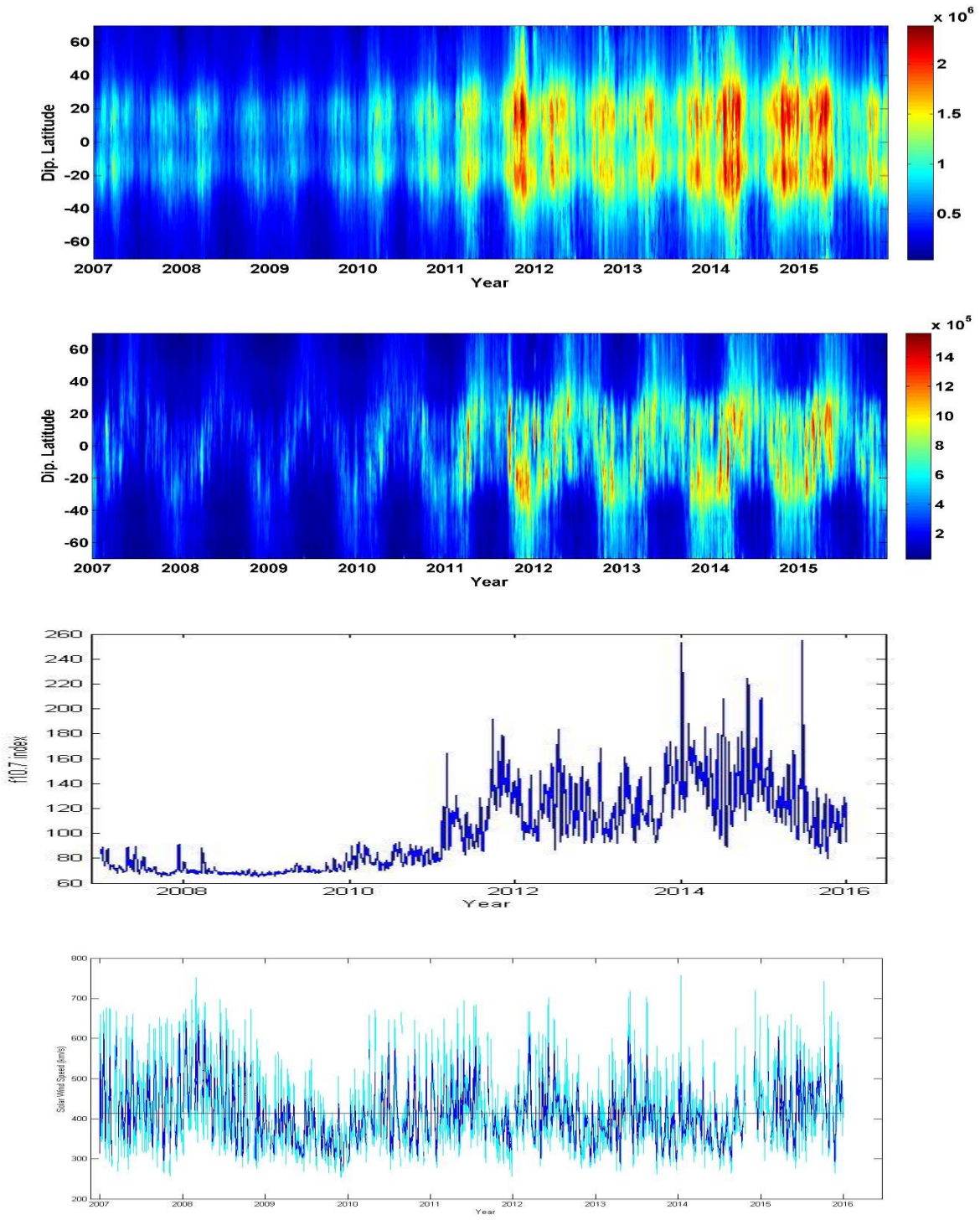



*Figure 7: Contour plots of zonally averaged values of NmF2 (electrons per cc) during the (Topmost) daytime and (Second from top) Nighttime. Grids are of resolution 2.5-degree latitude × 2.5-degree longitude. The relatively low values of NmF2 during the periods of 2007-10 both in the daytime and nighttime profiles coincides with the minima of the 11-year solar cycle as shown in the plot of f10.7 flux values, given in solar flux units (2nd from the Bottom). However, the solar minimum period is also associated with higher values of Solar Wind speeds, as shown by the uptick in the 27-day running mean above the overall average value of 440km/hr (Bottommost).*

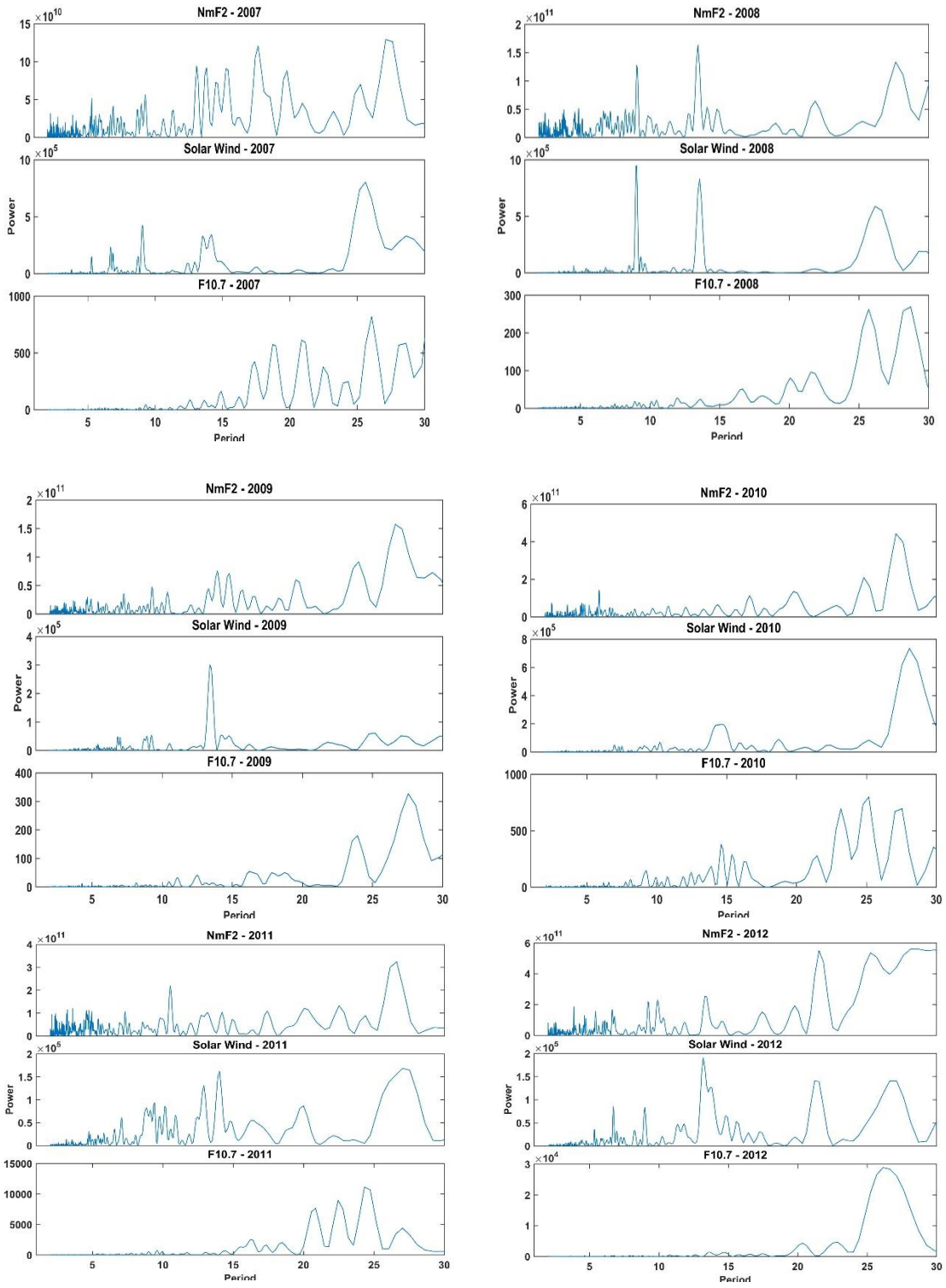



*Fig 8: Lomb-Scargle (LS) periodogram amplitude spectra of NmF2, Solar f10.7 flux index values and Solar wind speeds of each separate year from 2008-2012. The 27-day peak is associated with the rotation of the Sun and 13.5 day peak is associated with its sub-harmonic frequency. The 9-day peak seen in 2008-2009 is associated with solar winds from isotropically located CHs on the solar surface.*

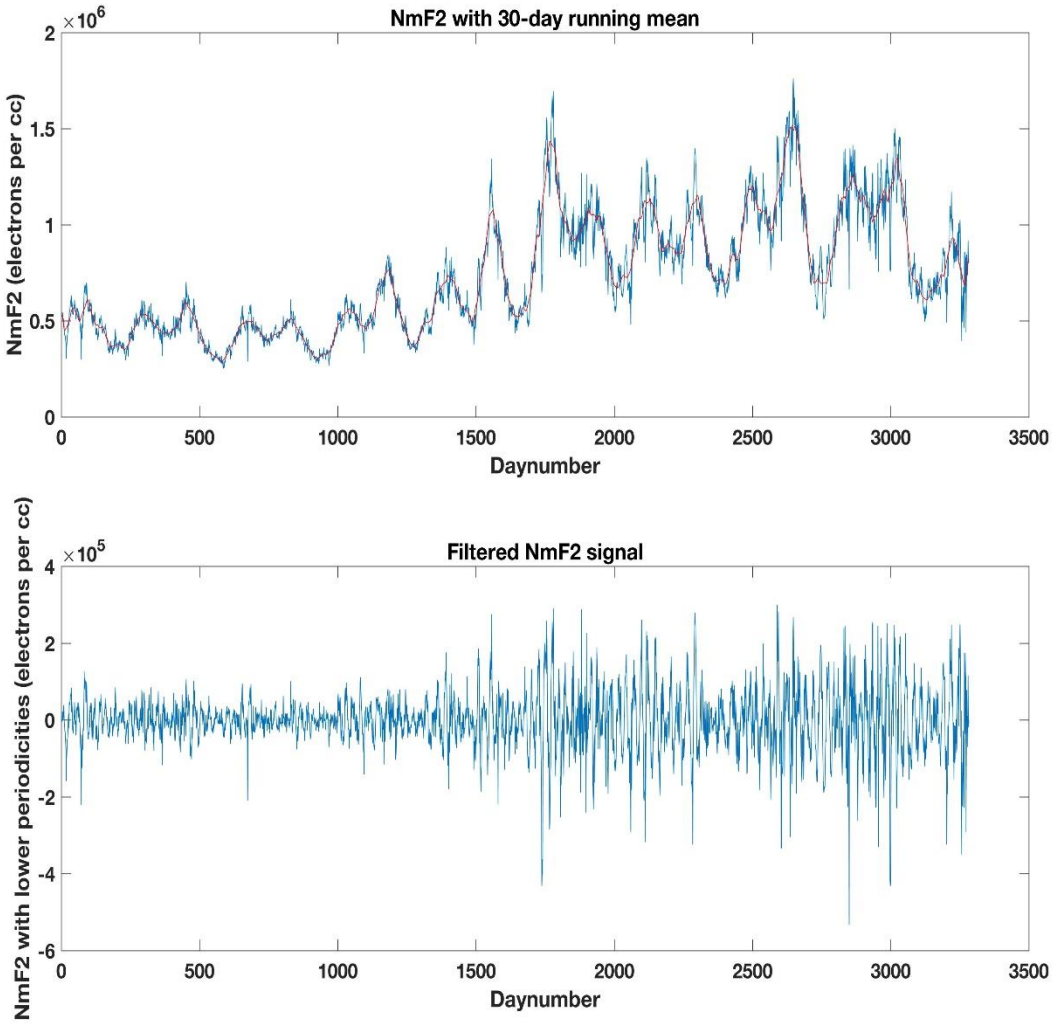
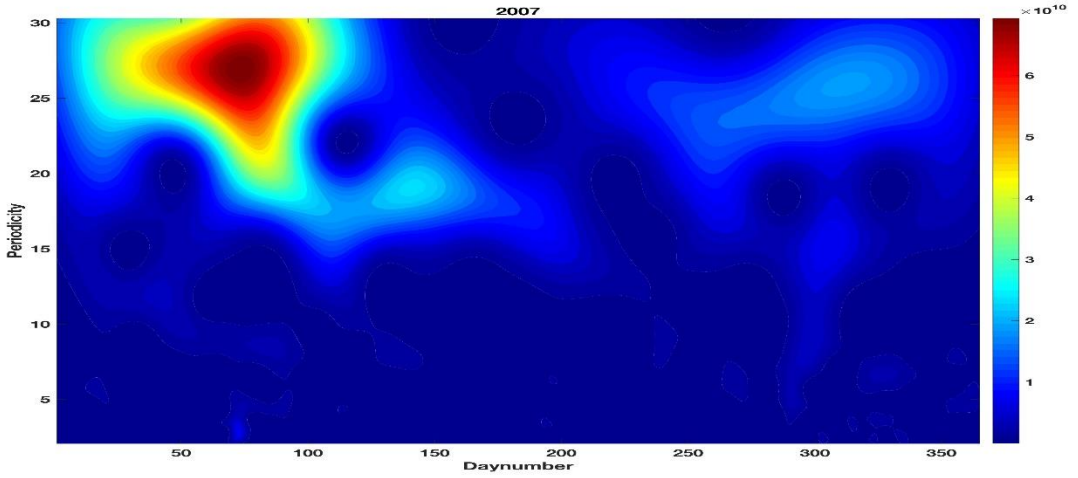
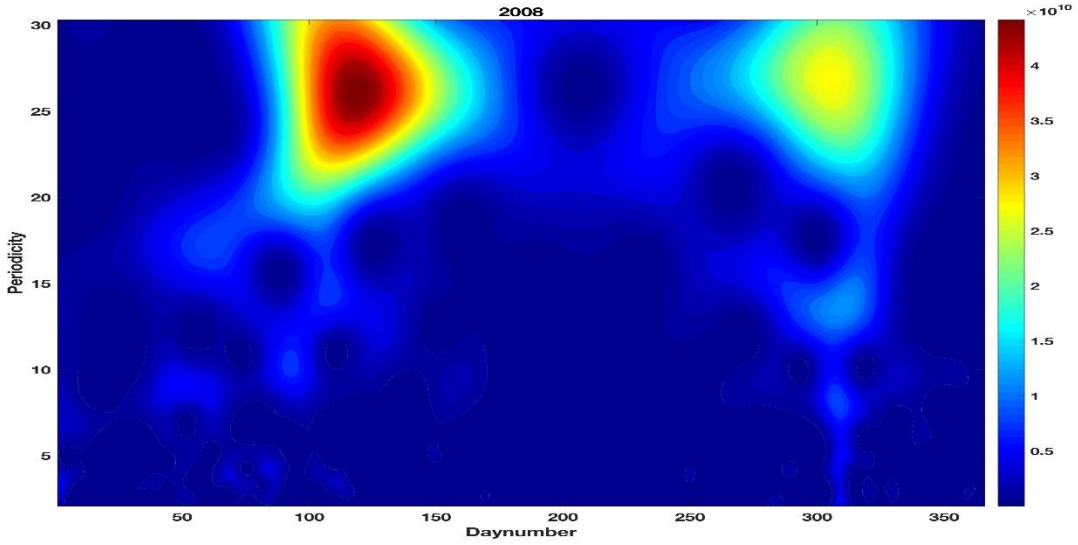
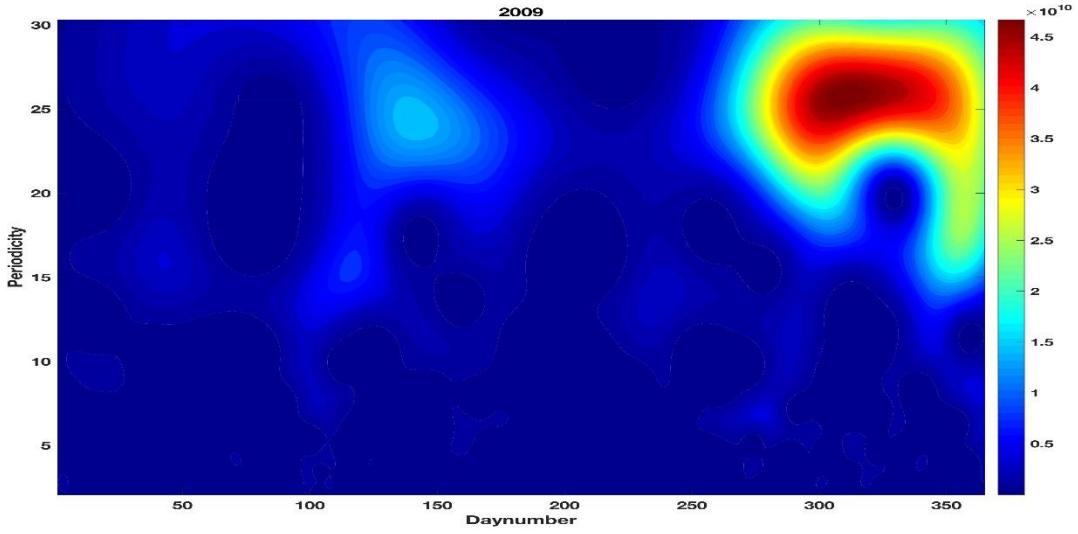
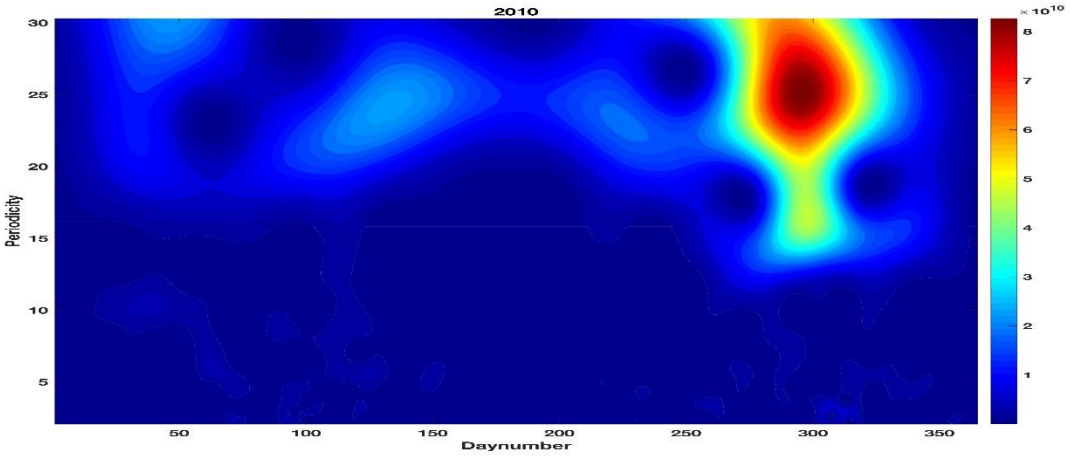


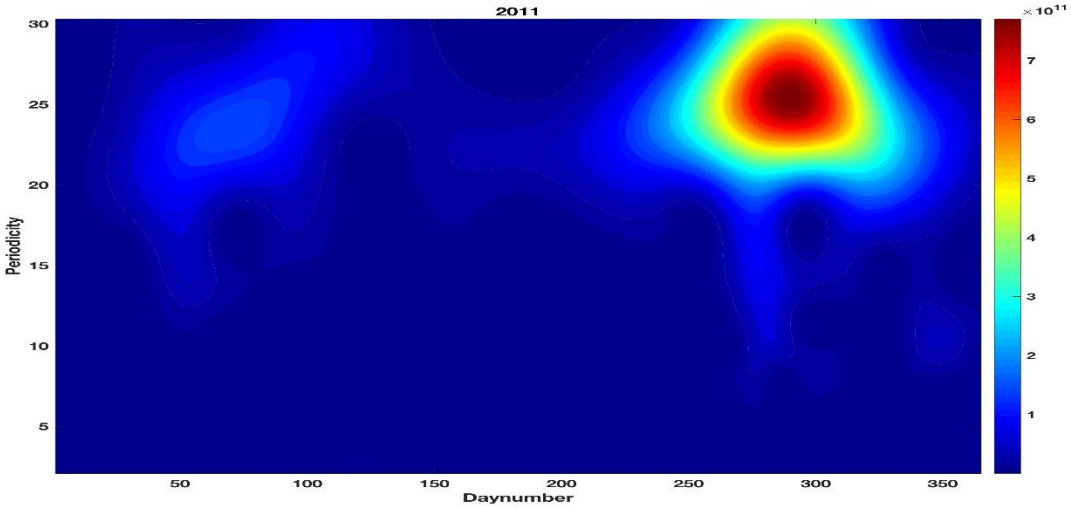

*Figure 9. (Top) The equatorial $N_mF_2$ signal with its 30-day running mean. (Bottom) The same signal with all periodicities greater than 30 days removed by subtracting the 30-day running mean from the above signal.*

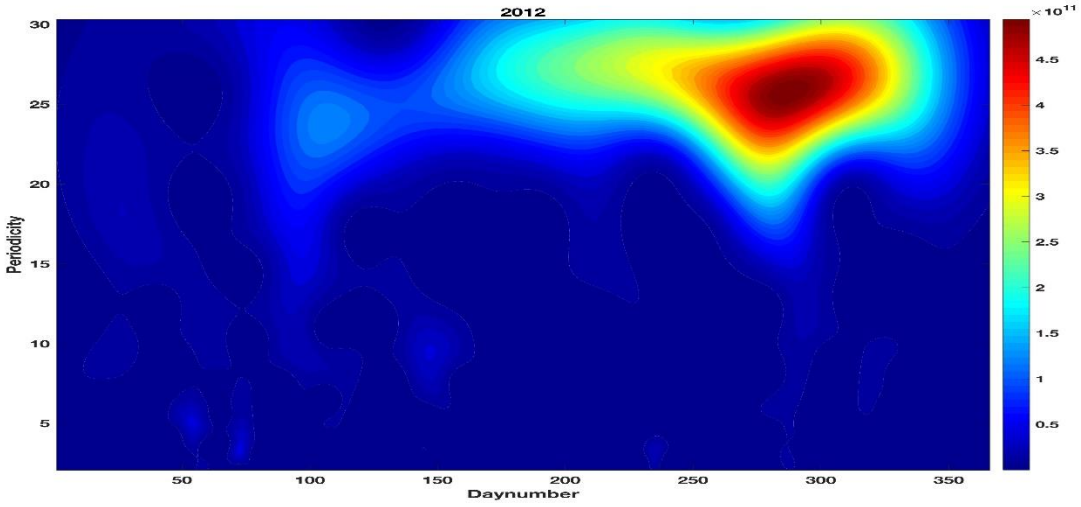

2007
Periodicity
Daynumber
×10^10


2008
Periodicity
Daynumber
×10^10


2009
Periodicity
Daynumber
×10^10


2010
Periodicity
Daynumber
30
25
20
15
10
5
50
100
150
200
250
300
350
×10^10
8
7
6
5
4
3
2
1


2011
Periodicity
Daynumber
30
25
20
15
10
5
50
100
150
200
250
300
350
×10^11
7
6
5
4
3
2
1


2012
Periodicity
Daynumber
30
25
20
15
10
5
50
100
150
200
250
300
350
×10^11
4.5
4
3.5
3
2.5
2
1.5
1
0.5

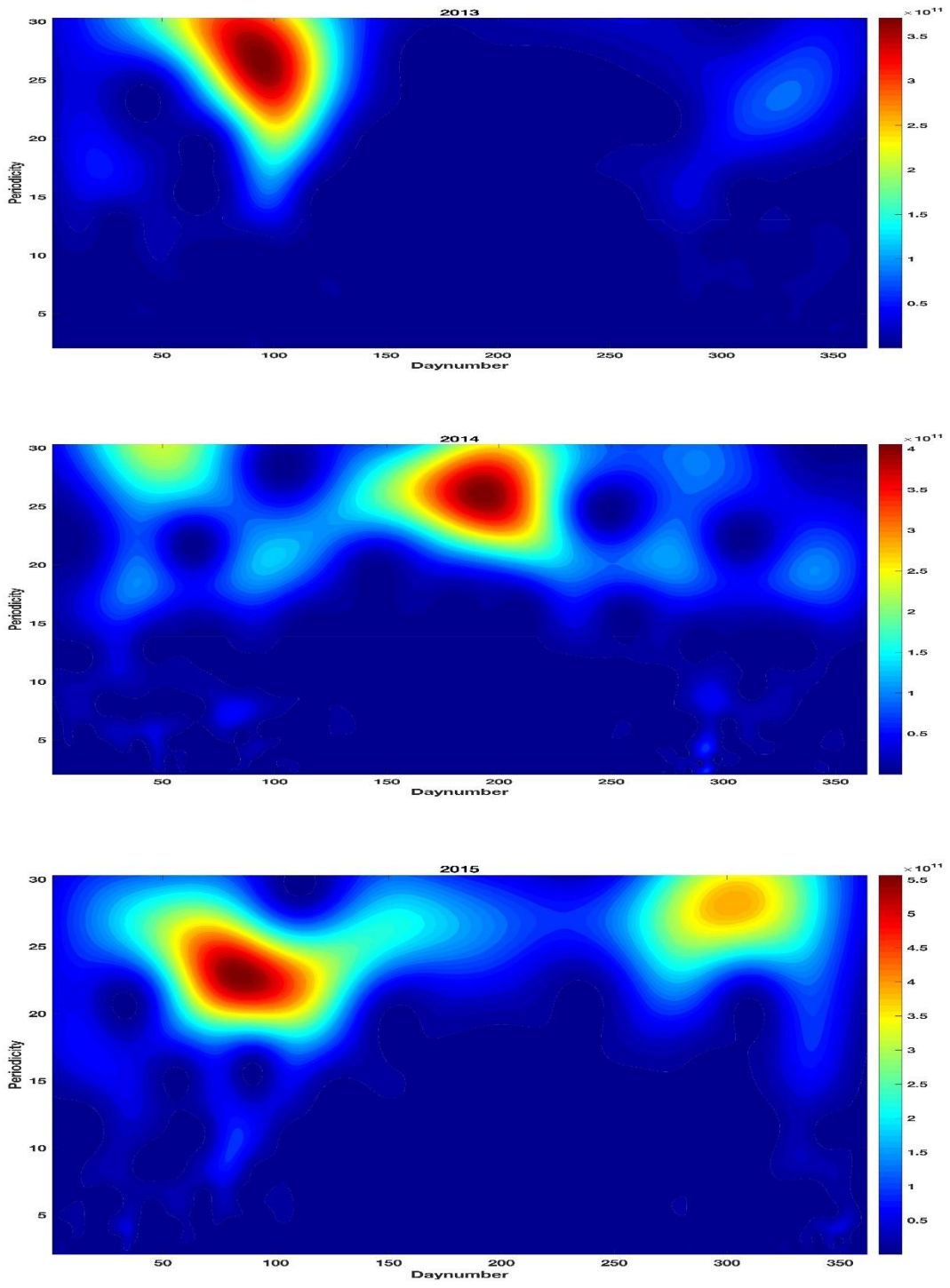



*Figure 10. Periodograms plotted with the contours of power spectral density from Morlet Wavelet Transform Analysis in grids of periodicity and the epoch for each year from 2007-2015, presented over the last three pages.*

# Section 2.

# Global Ionospheric Model

# 1. Using NN techniques for nonlinear regression problem

Conventional nonlinear regression techniques involve either *a priori* specification of the form of the regression equation with subsequent statistical determination of some undetermined constants, or statistical determination of the constants in a general regression equation, usually of polynomial form. The first technique requires that the form of the regression equation be known *a priori* or guessed. The advantages of this approach are that it usually reduces the problem to estimation of a small number of undetermined constants, and that the values of these constants when found may provide some insight to the investigator. The disadvantage is that the regression is constrained to yield a "best fit" for the specified form of equation. If the specified form is a poor guess and not appropriate to the data base to which it is applied, this constraint can be serious. Classical polynomial regression is usually limited to polynomials in one independent variable or low order, because high-order polynomials involving multiple variates often have too many free constants to be determined using a fixed number, *n,* of observations *(X ', Y').* A classical polynomial regression surface may fit the *n* observed points very closely, but unless *n* is much larger than the number of coefficients in the polynomial, there is no assurance that the error for a new point taken randomly from the distribution *f (x, y)* will be small.

The approach being considered here- Artificial Neural Networks, or simply Neural Network (NN)- uses a method that frees it from the necessity of assuming a specific functional form. This alternate technique more closely resembles a mapping *M*, between two vectors *X* (input vector) and *Y* (output vector) that can be symbolically written as:

$$Y = M(X); \quad X \in \Re^n, Y \in \Re^m \qquad (2.1)$$

where n and m are the dimensionalities of the input and output spaces correspondingly.

In recent years, NN techniques have been used to estimate these mappings, especially in geophysical, earth science and atmospheric studies such as in the work of Krasnopolsky et al. (2002) and Fox-Rabinovitz, Chalikov et al (2005). This new flavor of computational methods draws from the fact that predicting the high non-linearity and unpredictability of geophysical conditions require high adaptability.

NNs are a biologically inspired Artificial Intelligence (AI) technique. The name comes from the fact that they were initially modeled on the way neurons fire, with the accumulated firings of many neurons together determining the brain's response to any particular set of stimuli. The most common architecture used in NNs comprises three layers of neurons – an input layer, a layer of "hidden nodes" and a final output layer. Such an NN can represent any continuous function on a compact domain arbitrarily closely, even a nonlinear one, if it has enough hidden nodes – though choosing the optimal number of hidden nodes for a particular problem involves a certain level of arbitrariness. This category of NN architecture is called a feed-forward neural network (known for archaic reasons as Multilayer Perceptron or MLP).

The simplest MLP NN is a generic analytical nonlinear approximation or model for the mapping *M*. The MLP NN uses for the approximation a family of functions like:

$$y_q = NN\,(X,\, a,\, b) = a_{qo} + \sum_{j=1}^{k} a_{qj} \cdot t_j\,; \quad q = 1,\, 2,\, 3,\, \ldots,\, m \qquad (2.2)$$

where,

$$t_j = \Phi(\, b_{jo} + \sum_{i=1}^{n} b_{ji} \cdot x_i\,) \qquad (2.3)$$

and $x_i$ and $y_q$ are components of the input and output vectors, respectively; $a$ and $b$ are fitting parameters or NN "weights"; $\Phi$ is a so-called activation or "squashing" function (a nonlinear function, often specified as the hyperbolic tangent); $n$ and $m$ are the numbers of inputs and outputs, respectively; and $k$ is the number of the nonlinear basis function, $t_j$ (2.3), in the expansion (2.2). The expansion (2.2) is a linear expansion (a linear combination of the basis function $t_j$ (2.3)) and the coefficients $a_{qj}$ ($q$ = 1, …, $m$ and $j$ = 1, …, $k$) are the linear coefficients of this expansion.

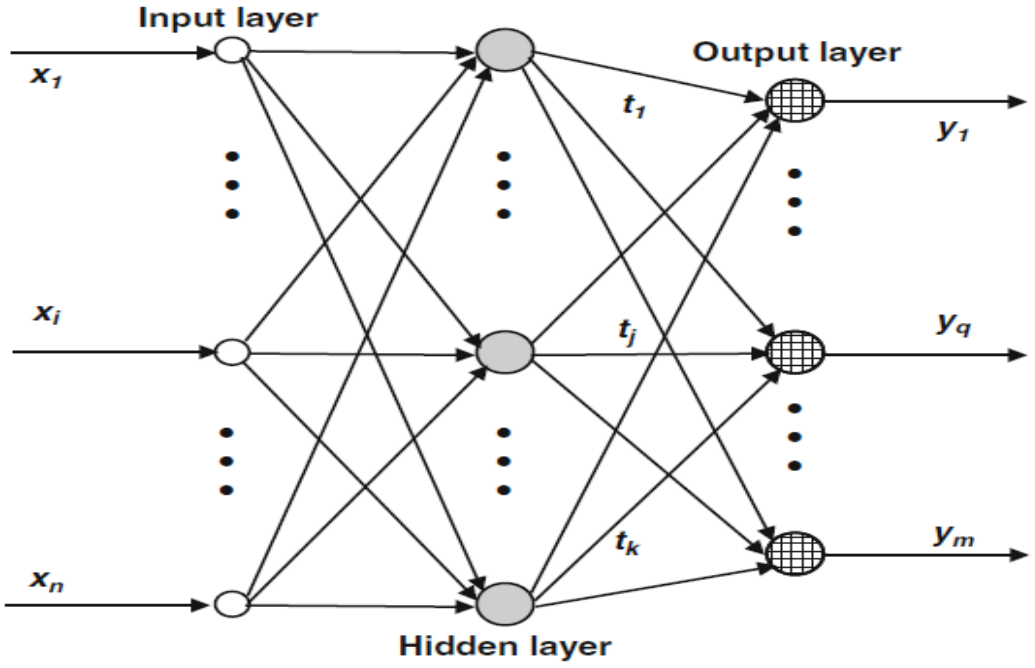



*Figure 1. The simplest MLP Neural Network architecture with one hidden layer and linear neurons in the output layer*

It must be noted that, if for each mapping output $y_q$ we construct a polynomial approximation, such a multidimensional polynomial of order $P$ would have $n^P$ unknown fitting parameters. Therefore, in the case of polynomial approximation, the computational complexity of the approximation for the entire mapping (2.1) is $m \times n^P$, which is a power law growth. The power law growth is slower than an exponential growth, but it is very fast and leads to the curse of dimensionality. Thus, the

polynomial approximation is of limited practical utility for multidimensional function and mapping approximations. NNs manage to address the curse of dimensionality and, due to the aforementioned linear dependence on the dimensionality of the input space, remain a practical approximation (or model) even for high-dimensional mappings.

# 2. Neural Network Architecture

In Fig. 1, a pictographic representation of the entire NN was introduced. The neurons are situated into *layers* inside the MLP NN. The input layer is a symbolic layer. Input neurons do not perform any numerical function; they simply distribute inputs to neurons in the following hidden layer. The hidden layer (there can be several) is usually composed of nonlinear neurons. The neurons in the output layer are usually linear. The *connections* (arrows) in Fig.1 correspond to the NN *weights*, the name used for the fitting parameters, *a* and *b* in NN terminology. Here we consider the simplest type of MLP NN that has one hidden layer and the output layer with linear neurons. Such architecture is sufficient for the approximation of any continuous (or almost continuous) mapping.

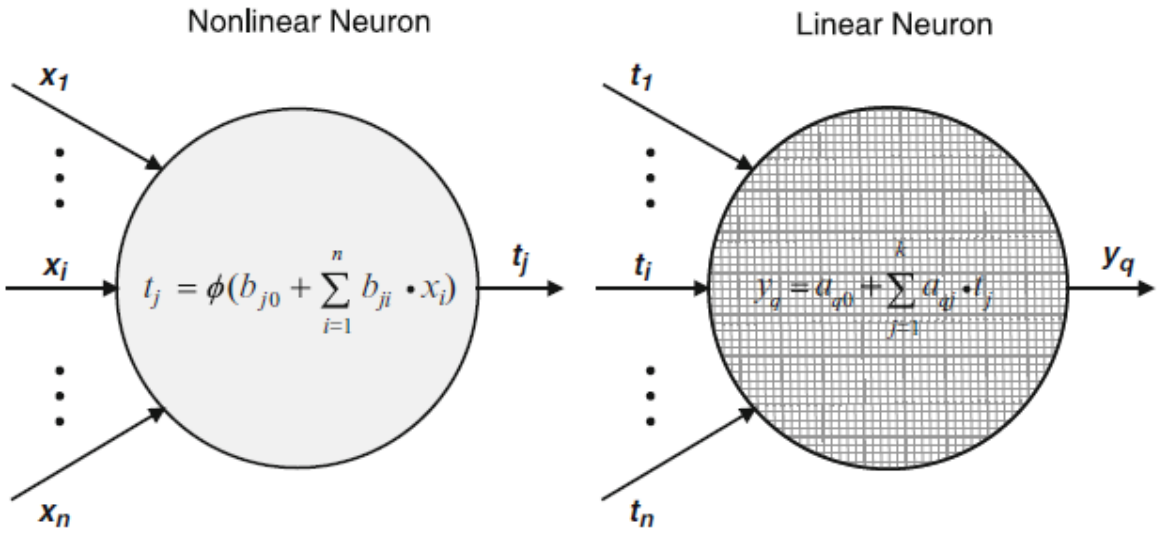



*Figure 1. The right figure shows linear (Eq. (2.2)) and the left – nonlinear (Eq. (2.3)) neurons.*

## 2.1. Training Set

In a practical application, a mapping (2.1) is usually represented and presented to the NN by a data or training set that consists of *N* pairs of input and output vectors, *X* and *Y:*

$$C_T = \{X_p, Y_p\}; \quad p = 1, \ldots, N \qquad (2.4)$$

where $Y_p = M(X_p) + \xi_p$, $X_p \in D$ and $Y_p \in R$, and $\xi_p$ represents any errors associated with the observations or calculation with a probability density function, $\rho(\xi)$. The set $C_T$ can also be considered as a combination of two rectangular matrices,

$$C_T = \{C_X, C_Y\} \qquad (2.5)$$

where $C_X$ is a matrix of dimensionality $n \times N$ composed of all input vectors $X$ and $C_Y$ is a matrix of dimensionality $m \times N$ composed of all output vectors $Y$. The training set is all that the NN "knows" about the target mapping that it is expected to approximate.

The training set represents the mapping (2.1) for the NN, and, therefore, it has to be *representative*. It means that the training set has to have a sufficient complexity to represent the complexity of the target mapping, allowing the NN to achieve the desired accuracy of the approximation of the target mapping. The set should have a sufficient ample size, $N$, of properly distributed data points that adequately resolve the functional complexity of the target mapping (2.1).

In our case, an extensive database of ionospheric, solar and wind data for the years 2007-2015 enables us with enough detail over nearly an entire solar cycle to call it a sufficient training set for an NN model that is aimed at predicting ionospheric conditions from inputs of solar and geomagnetic activity.

## 2.2. Normalization of the NN Inputs and Outputs

A degree of flexibility is provided by *normalization* of the NN inputs and outputs. NN inputs are usually normalized to an interval [-$a$, $a$], using a simple equation,

$$x_i'' = a_i \times \left[\frac{2.x_i - x_i^{max} - x_i^{min}}{x_i^{max} - x_i^{min}}\right] \qquad (2.6)$$

where $x_i$ is the $i$-th NN input before and $x_i''$ the same, after the normalization. We choose to normalize our data in the range [-1,1] for simplicity.

For a single NN with multiple outputs (similar to the ones we apply), the normalization of the outputs affects the approximation accuracy and NN performance more significantly than in the case of a single-output NN. Normalization for the case of multiple outputs can be written as:

$$y_q'' = \alpha . \frac{[y_q - y_q']}{\sigma_q} \qquad (2.7)$$

where $y_q'$ and $\sigma_q$ are the mean and SD of the $q$-th output, $y_q$. $\alpha<1$ is introduced to accelerate the training of the linear weights in the output layer of the NN. This normalization improves approximation accuracies for small outputs; however, if these outputs are noisy; it propagates the

noise to other outputs. Normalization (2.7) also reduces correlations that may exist between outputs.

## 2.3. Neural Network Training Process and the Levenberg-Marquardt (LM) algorithm

After the NN parameters (*n*, *k*, and *m)* are defined, the weights (*a* and *b*) can be found using the training set $C_T$ (2.5) and the maximum likelihood method (Vapnik 1995) by maximizing the likelihood functional:

$$L(a, b) = \sum_{i=1}^{N} ln\, \rho(\xi) \qquad (2.8)$$

with respect to the free parameters (i.e., the NN weights) *a* and *b*. Here, $\rho(\xi)$ is the probability density function for the approximation errors $\xi_i = y_i - NN(X_i, a, b)$ and the summation is performed over the *N* records in the training set. If the errors $\xi_i$ are normally distributed, Eq. (2.8) leads to the minimization of the mean-square error function (also called *cost function*) with respect to the NN weights given by *W* (*a* and *b*),

$$E(W) = \frac{1}{N}\sum_{i=1}^{N} \xi_i(W) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - Z_i)^2 \qquad (2.9)$$

where $Z_i = NN(X_i, W)$, *E* is the total error calculated over the entire training set (all *N* records of the training set are included), and $\xi_i = (Y_i - Zi)^2$ is an error corresponding to the *i*-th record in the training set. This procedure of minimization of the error function (2.9) is usually called *NN training*. Minimizing the error function is performed in the *W*-space (the space of NN weights or the training space), which has a dimensionality equal to the number of NN weights, $N_C$ given by:

$$N_C = k(n + m + 1) + m \qquad (2.10)$$

where, as before, *n* and *m* are the numbers of inputs and outputs, respectively; and *k* is the number of the nonlinear basis functions for the mapping. The NN complexity grows linearly with the growth of the dimensionalities of the input space (the number of inputs), *n,* and the output space (the number of outputs), *m*.

Optimal values for the weights are obtained by minimizing the error function (2.9); this task is a nonlinear minimization problem. A number of methods have been developed for solving this problem (Bishop 1995; Haykin 2008). The most popular and simplest one of them, a simplified version of the *steepest (or gradient) descent method* known as the back-propagation training algorithm. It was introduced as an NN training algorithm by Werbos (1982) and Rumelhart et al. (1986).

The back-propagation training algorithm is based on the simple idea that searching for a minimum of the error function (2.11) can be performed step by step iteratively, and that at each step we

should increment or decrement the weights in such a way as to decrease the error function. This can be done, for example, using the following simple steepest descent rule,

$$W^{(n+1)} = W^{(n)} - \eta \frac{\partial E\,(W^n)}{\partial W} \qquad (2.11)$$

where $W$ is either one of two weights ($a$ or $b$), $W^{(n+1)}$ is an adjusted or updated weight, $\eta>0$ is a so-called learning constant, and $W^{(n)}$ is the weight at the previous $n$-th iteration.

Geophysical data are generally stochastic in nature. In the case of training, the generally used criterion of minimum of the error function (2.11) should be substituted by the requirement that the error should not exceed the uncertainty $\varepsilon$ or

$$E\,(W) = \frac{1}{N} \sum_{i=1}^{N} (Y_i - M_{NN}(X_i))^2 \leq \varepsilon^2 \qquad (2.12)$$

Here, it is assumed that the stochastic error (that originates from the uncertainties in the training dataset) associated with the mapping is inherently additive in nature. All NNs that satisfy the condition (2.12) are valid emulations of the stochastic mapping (given as $Y = M(X) + \varepsilon$, where $M(X)$ is our original mapping given in (2.1)).

Simple gradient descent suffers from various convergence problems. Logically, we would like to take large steps down the gradient at locations where the gradient is small (gentle slope) and conversely, take small steps when the gradient is large, so as not to rattle out of the minima. With the update rule (2.11), we do just the opposite of this. Another issue is that the curvature of the error surface may not be the same in all directions. For example, if there is a long and narrow valley in the error surface, the component of the gradient in the direction that points along the base of the valley is very small while the component along the valley walls is quite large. This results in motion more in the direction of the walls even though we have to move a long distance along the base and a small distance along the walls. This situation can be improved upon by using curvature as well as gradient information, namely second derivatives. One way to do this is to use *Newton's method* to solve the equation $\nabla_w E(W^{(n)}) = 0$. Expanding the gradient of E using a Taylor series around the current state $W^{(n)}$, we get

$$\nabla_w E(W) = \nabla f(W^{(n)}) + (W^{(n+1)} - W^{(n)})^T \nabla^2 E(W^{(n)}) + \text{higher order terms} \qquad (2.13)$$

If we neglect the higher order terms (assuming $E(W^{(n)})$ to be quadratic around $W^{(n)}$), and solve for the minimum x by setting the left hand side of (2.13) to 0, we get the update rule for Newton's method:

$$W^{(n+1)} = W^{(n)} - \nabla^2{}_w E(W^{(n)})^{-1} \nabla_w E(W^{(n)}) \qquad (2.14)$$

The main advantage of this technique is rapid convergence. However, the rate of convergence is sensitive to the starting location (or more precisely, the linearity around the starting location). It can be seen that simple gradient descent and Gauss-Newton iteration are complementary in the

advantages they provide. Levenberg proposed an algorithm based on this observation, whose update rule is a blend of the above mentioned algorithms and is given as:

$$W^{(n+1)} = W^{(n)} - (H + \eta I)^{-1} \nabla_w E(W^{(n)}) \qquad (2.15)$$

where H is the Hessian matrix evaluated at $W^{(n)}$. [Hessian matrix or simply the Hessian is a square matrix of second-order partial derivatives of a scalar-valued function or scalar field. It describes the local curvature of a function of many variables.] This update rule is used as follows. If the error goes down following an update, it implies that our quadratic assumption on f (x) is working and we reduce λ (usually by a factor of 10) to reduce the influence of gradient descent. On the other hand, if the error goes up, we would like to follow the gradient more and so λ is increased by the same factor. The *Levenberg algorithm* is thus –

1. Do an update as directed by the rule above.
2. Evaluate the error at the new parameter vector.
3. If the error has increased as a result the update, then retract the step (i.e. reset the weights to their previous values) and increase λ by a factor of 10 or some such significant factor. Then go to (1) and try an update again.
4. If the error has decreased as a result of the update, then accept the step (i.e. keep the weights at their new values) and decrease λ by a factor of 10 or so.

The above algorithm has the disadvantage that if the value of η is large, the calculated Hessian matrix is not used at all. We can derive some advantage out of the second derivative even in such cases by scaling each component of the gradient according to the curvature. This should result in larger movement along the directions where the gradient is smaller, so that the classic "error valley" problem does not occur any more. This crucial insight was provided by Marquardt. He replaced the identity matrix in (2.15) with the diagonal of the Hessian resulting in the Levenberg-Marquardt update rule:

$$W^{(n+1)} = W^{(n)} - (H + \eta.diag[H])^{-1} \nabla_w E(W^{(n)}) \qquad (2.16)$$

Since the Hessian is proportional to the curvature of E, (2.16) implies a large step in the direction with low curvature (i.e., an almost flat terrain) and a small step in the direction with high curvature (i.e, a steep incline).

It is to be noted that while the LM method is in no way optimal but it works extremely well in practice. The only flaw is its need for matrix inversion as part of the update. Even though the inverse is usually implemented using clever pseudo-inverse methods such as singular value decomposition, the cost of the update becomes prohibitive after the model size increases to a few thousand parameters. For moderately sized models (of a few hundred parameters) however, this method is much faster than say, normal gradient descent. The Levenberg-Marquardt (LM) algorithm is the most widely used optimization algorithm. It outperforms simple gradient descent and other conjugate gradient methods in a wide variety of problems.

The error function may be significantly different than the mean-square error (MSE) or cost function (2.9) (Liano 1996). For example, in geophysical systems, a parameterized variation of the mean-square error function has been proposed by Krasnopolsky et al (2006), given by:

$$Prmse(i) = \sqrt{\left(\frac{1}{L}\right)\sum_{j=1}^{L}[Y(i,j) - Y_{NN}(i,j)]^2} \qquad (2.17)$$

where $Y(i, j)$ and $Y_{NN}(i, j)$ are outputs from the original parameterization and its NN emulation, respectively, where $i = (latitude, longitude)$, $i = 1, \ldots, N$ is the horizontal location of a vertical profile; N is the number of horizontal grid points; and $j = 1, \ldots, L$ is the vertical index, and L is the number of the vertical model levels. A way to implement this modification and achieve higher resolution is clearly to divide the process into geographical grids with the standard form of MSE calculated in each individual grid, with an individual neural network for each individual grid. Both the standard process and the modified process are implemented with varying degrees of success, as is discussed in the following sections.

# 3. Single Neural Network Approach

The NN is aimed to create a multivariate regression model to predict the ionospheric parameters such as peak electron density ($N_mf_2$) and peak height ($H_mf_2$). We have used Levenberg-Marquardt (LM) backpropagation algorithm to train the neural network. To do so, initially we have conducted optimization studies to find the best set of input and target variables, best set of training parameters and the optimum hidden layers etc.

The input parameters under consideration with which ionospheric conditions are known to vary are:

1. Day number (epoch)
2. Time (Universal coordinates, UTC)
3. Latitude
4. Dip latitude
5. Longitude
6. Daily average F10.7 solar flux
7. Daily average Kp-index

We have conducted several initial test runs with varying sets of input parameter (while keeping the hidden layer NN architecture unchanged) to understand the regression relations that best describe the target variables in terms of the input variables. Nearly eighty percent of the global COSMIC/FORMOSAT-3 data over the time period from 2007-2015 is used to train, validate and test the neural network. For external testing of the neural network architecture, we use the remaining twenty percent data and the predicted values are plotted against the original values and the linear fit equations are noted. The results of these initial runs are tabulated below:

| Model | Input Variables | Target Variables | Number of Hidden Layers | Equation of linear fit for $N_mf_2$ | Equation of linear fit for $H_mf_2$ | Overall Regression Coefficient, R | Remarks |
|---|---|---|---|---|---|---|---|
| Test 1 | Day number, UTC, Latitude, Dip Latitude, Longitude | $N_mf_2$, $H_mf_2$ | 6 | 1.7*x + 4.3e4 | 0.13*x + 2.6e2 | 0.7236 | Poor model for both Nmf2 and Hmf2 due to lack of solar inputs. |
| Test 2 | Day number, UTC, Latitude, Dip Latitude, Longitude, F10.7 | $N_mf_2$, $H_mf_2$ | 6 | 1.8*x - 3.8e4 | 0.57*x + 1.4e2 | 0.8134 | Good regression, but poor fit between predicted and actual values. |
| Test 3 | Day number, UTC, Latitude, Dip Latitude, Longitude, F10.7, Kp-index | $N_mf_2$, $H_mf_2$ | 6 | 1.1*x – 8.8e4 | 0.68*x + 110 | 0.86069 | Good for $N_mf_2$ predictions but poor for $H_mf_2$ prediction |
| Test 4 | Day number, UTC, Latitude, Dip Latitude, Longitude, F10.7, Kp-index | $N_mf_2$ | 6 | 1.1*x – 4.8e4 | ___ | 0.77029 | Similar $N_mf_2$ as Test 3. However regression slightly deteriorated. |
| Test 5 | Day number, UTC, Latitude, Dip Latitude, Longitude, F10.7, Kp-index | $H_mf_2$ | 6 | ___ | 1.1*x - 14 | 0.64778 | Though fit is good, regression is poor, due to high spread of outliers |

From the above table, one can conclude that although the $N_mf_2$ prediction is improved over successive tests through the inclusion of solar forcing (F10.7 and Kp-index), but the $H_mf_2$ prediction is poor. We suspect this poor regression of $H_mf_2$ might be due to the absence of certain important input variables.

As we know, ionospheric peak parameters such as $N_mf_2$ and $H_mf_2$ depend on the production, loss and transport. In our previous tests, we have not taken the transport phenomena into account. Hence, we decided to modify the input dataset to include zonal and meridional winds. In fact, to give a complete set of input variables to predict ionospheric parameters, one must consider the fact that ionospheric transport occurs chiefly along prevailing magnetic field lines and hence, we decided to include magnetic field configuration parameters such as Inclination (I) and Declination (D). Zonal and meridional winds, with the combination of magnetic field configurations, can

produce effective winds either in upward or downward directions, which can significantly alter the recombination rates in the ionosphere and hence, the $N_mf_2$ and $H_mf_2$ predictions.

Considering all the above facts, we have calculated the zonal and meridional winds using Horizontal Wind Model (HWM-14) and the Inclination/Declination values using the IGRF14 model. Based on the new input variables, we re-train our neural network for $N_mf_2$ and $H_mf_2$ separately. Regression was found to improve significantly in both the cases (with R> 0.88).

## 3.1. Testing of the Single Neural Network Approach

The attributes of the neural networks (one each for Nmf2 and Hmf2) are saved after initial training for further use to predict the ionospheric parameters at a given set of inputs. Firstly, we want to test whether our model can be able to represent the local time and magnetic latitudinal variation of the well-known Equatorial Ionization Anomaly (EIA). Towards this purpose, the entire globe was divided into $5^o$ longitude X $2.5^o$ latitude bins and the $N_mf_2/H_mf_2$ values are calculated and then the zonal average is computed at a given time, season, solar flux and Kp-index to test the local time, latitudinal and seasonal variability.

Figure 1 shows the local time and magnetic latitudinal variation of zonally averaged $N_mf_2$ for March equinox at a solar flux of 120 sfu, under quiet geomagnetic conditions.

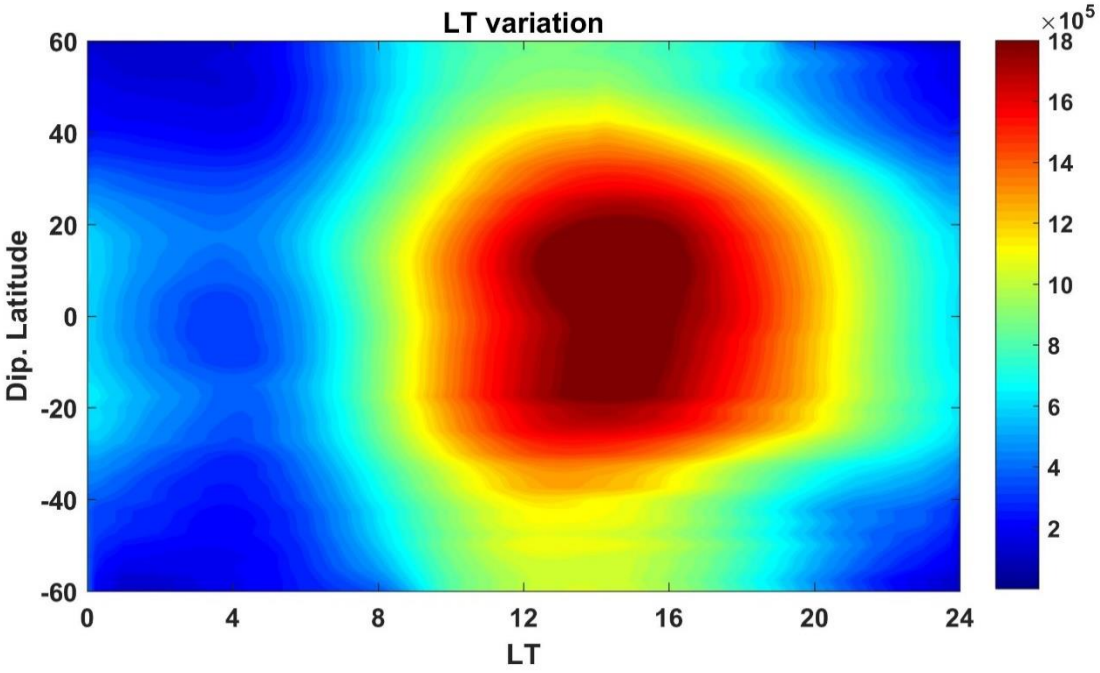



*Figure 2. Local Time and Magnetic Latitudinal Variation of zonally averaged $N_mf_2$ during March at 120 sfu and quiet geomagnetic conditions*

From the figure, one can see the diurnal variation of zonally averaged $N_mf_2$, with its peak around noon to afternoon hours. However, it can be noted that the separation between the two crests of

the EIA is absent. In fact, the value of zonally averaged $N_mf_2$ values are overestimated around the equatorial region. This might be due to the low resolution of single neural network approach, which limited our prediction of $N_mf_2$ and $H_mf_2$.

# 4. Gridded neural network approach

In the previous section, we discussed about the single neural network approach and its limitations. To increase the resolution of our model, we divided the entire globe into small grids of size 20$^o$ longitude x 10$^o$ latitude. Our basic idea is to get an individual trained neural network for each grid around the globe. Total data set is further separated according to the longitude and latitude grids. Then, the all input parameters including zonal, meridional winds and magnetic field configurations are given to each individual grid to train the network. We increased the number of hidden layers to 40 to optimize the network. We noticed that the overall regression values of the gridded neural network approach improved significantly over the single neural network approach. For example, figure 2 shows the regression of training, validation and testing of a grid around 0-10$^o$ latitude and 0-10$^o$ longitude. The thin black line with 45$^o$ slope represents the fit of the normalized target variable while the colored solid lines represents the fit of the corresponding predicted values. One can see that the regression values are around 0.95, which has significantly improved compared to the single neural network approach. Trained neural network for each individual grids is further saved for testing purpose.

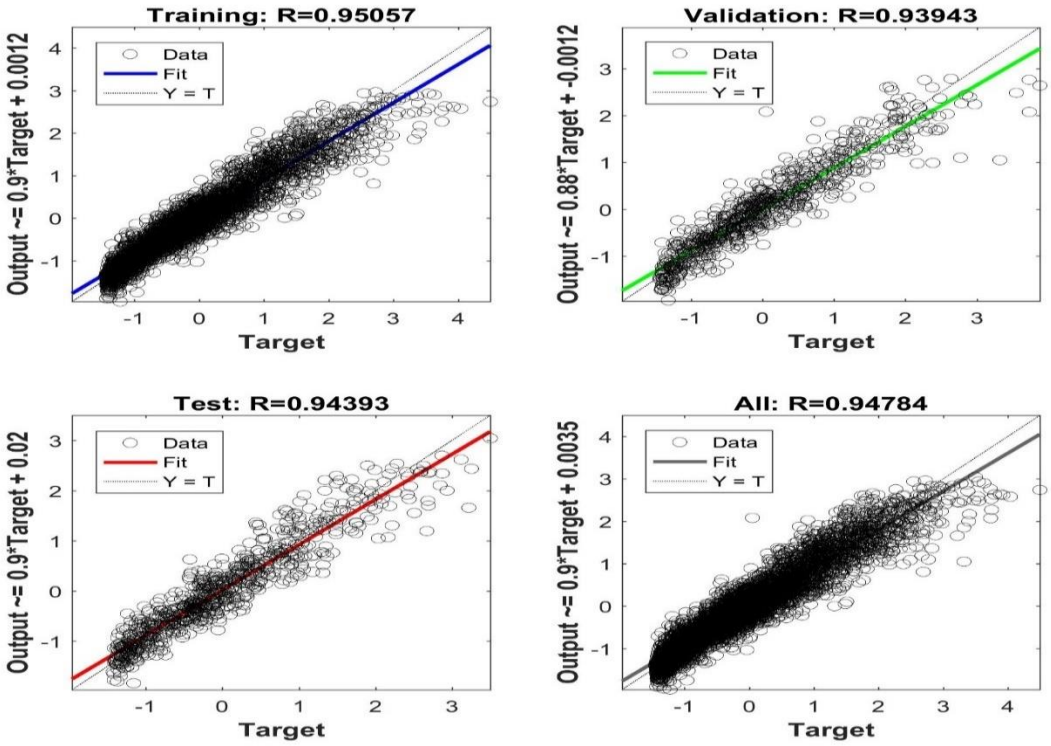



*Figure 3. Regression values of an example neural network grid around 0-10$^o$ latitude and 0-10$^o$ longitude*

## 4.1. Testing of the Gridded Neural Network Approach

To test the gridded neural network approach, we have again selected the EIA and its local time and latitudinal variability. $N_mF_2$ and $H_mF_2$ values are computed using the individual neural network grids at a given day number 80 (March equinox), time, solar F10.7 flux and geomagnetic conditions (Kp-index). Figure 3 shows the local time and magnetic latitude variation of $N_mF_2$ during March equinox at 120sfu and quiet geomagnetic condition. One can clearly see the EIA with its peaks around ±20$^o$ latitude and the trough around dip equator.

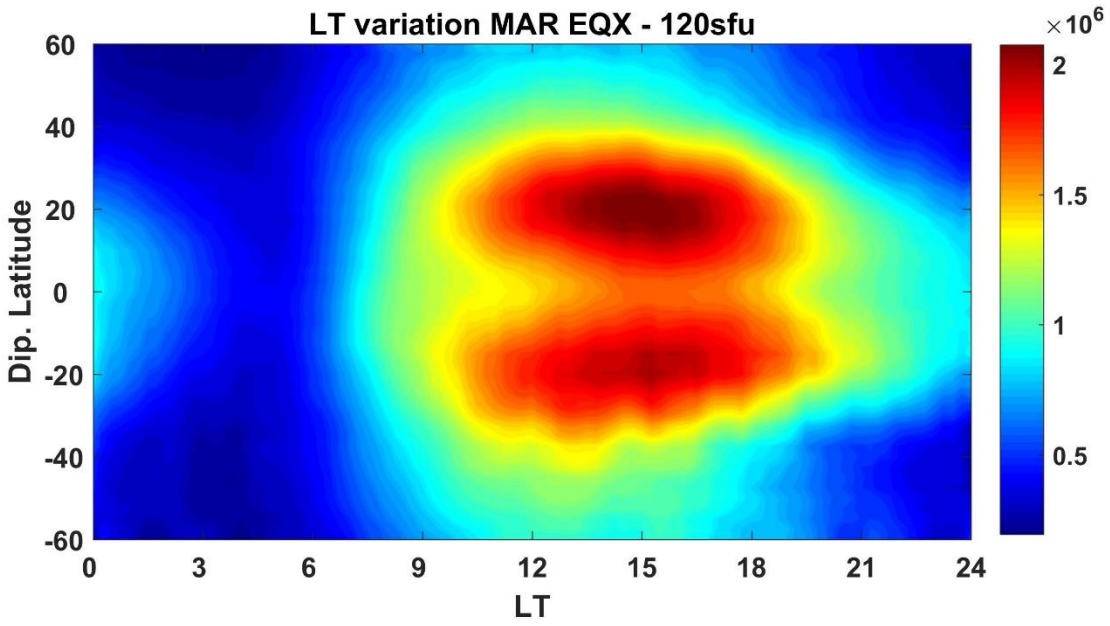



*Figure 4. Local time and magnetic latitudinal variation of zonally averaged $N_mF_2$ during March equinox (F10.7 - 120 and Kp - 3)*

EIA generally shows low values before sunrise, reaches its peak value around noon and these high values are maintained throughout afternoon hours. Further, the separation between the two crests slowly disappear after sunset, leading to a single crest. The latitudinal separation between the two crests are maximum during noon time due to strong E x B drift, and then the latitudinal extent slowly reduces during afternoon to pre-sunset hours. From the figure 3, it can be noted that the EIA is well represented from the gridded neural network approach with all the above discussed morphology.

Further pictorial comparison of global and model-predicted $N_mF_2$ are given in the following images:

## 5.1. Local Time variation of global $N_mF_2$ during December Equinox (120 sfu)

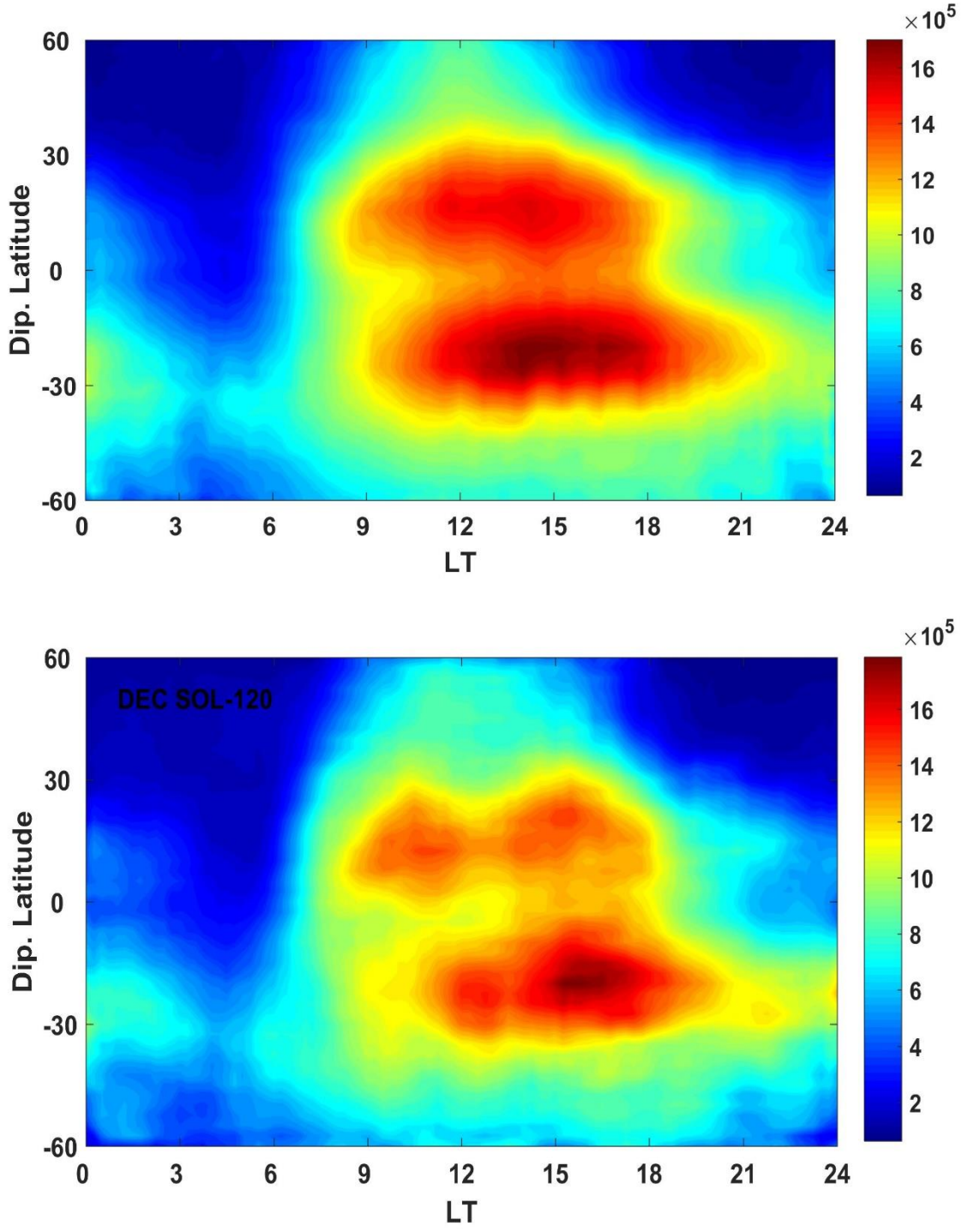



*Fig 4. LT variation of $N_mF_2$ from actual satellite data (above) and from gridded NN model predictions (below), both set to same scale.*

## 5.2. Local Time variation of global $N_mF_2$ during summer Solstice (120 sfu)

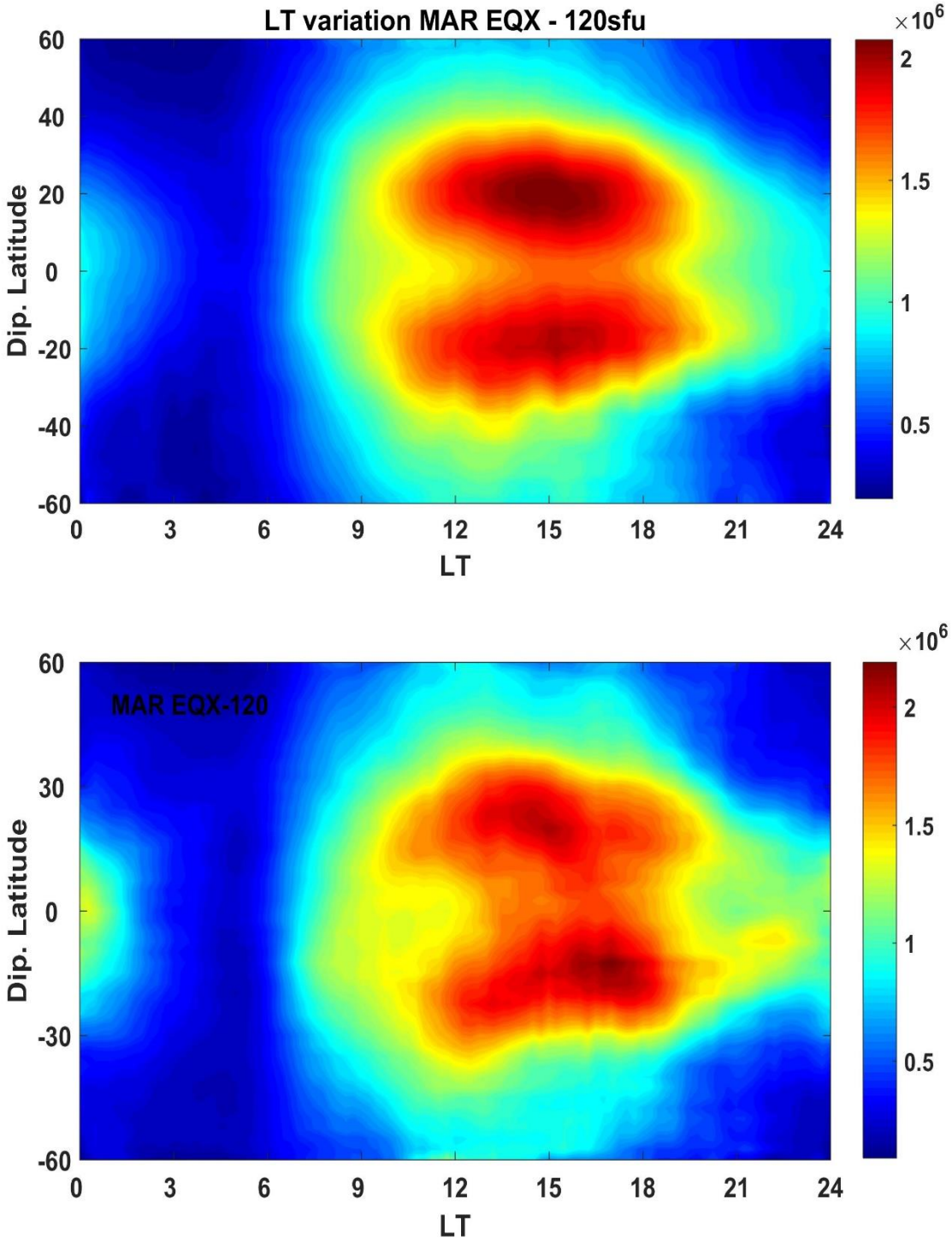



*Fig 4. LT variation of $N_mF_2$ from actual satellite data (above) and from gridded NN model predictions (below), both set to same scale.*

# Proposed Further Work

An effective ionospheric model should not be restricted to predicting the peak ionospheric parameters only, even though this is certainly a step in the right direction. Effectively predicting the $N_mF_2$ and $H_mF_2$ for all latitudes and longitudes means that the proposed Neural Network model is still restricted to being a 2-D model. For being a model that might have active applications in satellite applications and space sciences, the model must be modified so as to be able to predict ionospheric ionization (i.e, electron density) configurations.

Towards that effect, the same data from FERMOSAT-3/COSMIC mission for the years 2007-2015 may be used. However, for the altitudinal training of the neural network, instead of using only the ionospheric peak values, one needs to utilize all the data points generated by each sweeping scan of the ionosphere over a particular region on the Earth's surface. Since there are ~ 2000 scans per sweep and there are ~3.5 million sweeps over this time, the computational power required to train such a neural network will be considerable, especially if it is as high-resolution as our gridded approach. However, the success attained at adequately predicting ionospheric peak values in the present case may be considered as encouraging towards that endeavor.

# Acknowledgements

I am indebted to the authorities at IIG and above all, Dr. Tulasi Ram Sudarsanam for giving me a chance to work on the topic of my choice. Dr. Tulasi's tireless guidance and constant support made me overcome all the lack of knowledge in the field that I started with, at the beginning of the year. Even when I would take long to grasp a concept, his seemingly infinite source of motivation would never waver and I am grateful for having worked under his tutelage and I wish to work more in future with him.

Moreover, I am grateful to Sai Gowtham, a doctoral student of Tulasi Sir, at IIG, for taking me under his wings and generously offering me a place in his lab and letting me use his Wi-Fi connection and computer anytime I wanted. I am especially really grateful to him for his invaluable inputs and support at every stage of my research, specially showing me many new tricks in MATLAB.

I am also grateful to all the people at the IIG computer center and the Library for always being there to help either with their actions and an endless supply of free paper to print on as well as the people at the Canteen and Hostel for making my stay easier.

# References


[1] J. K. Hargreaves. *The Solar-Terrestrial Environment. Cambridge Atmospheric and Space Sciences Series*. Cambridge University Press. **1992**

[2] Jack B. Zirker. *Journey from the Center of the Sun*. Princeton University Press. **2002**

[3] S. Tulasi Ram, C. H. Liu and S. Y. Su. *Periodic solar wind forcing due to recurrent coronal holes during 1996–2009 and its impact on Earth's geomagnetic and ionospheric properties during the extreme solar minimum.* JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 115, A12340, **2010**

[4] S. Tulasi Ram, J. Lei, S.-Y. Su, C. H. Liu, C. H. Lin, and W. S. Chen. *Dayside ionospheric response to recurrent geomagnetic activity during the extreme solar minimum of 2008.* GEOPHYSICAL RESEARCH LETTERS, VOL. 37, L02101, **2010**

[5] Gary S. Bust and Kathryn N. Mitchell. *History, Current State, and Future Directions of Ionospheric Imaging*. American Geophysical Union. Reviews of Geophysics, 46, RG1003, **2008**

[6] S. Sripathi, Ram Singh, S. Banola, Dupinder Singh, and S. Sathish. *The response of the equatorial ionosphere to fast stream solar coronal holes during 2008 deep solar minimum over Indian region.* Journal of Geophysical Research: Space Physics, 121, 841–853, **2016**

[7] Shahab Araghinejad. *Data-Driven Modeling: Using MATLAB® in Water Resources and Environmental Engineering*. Springer. **2014**

[8] Kenneth Davies. Ionospheric Radio (IEEE Electromagnetic Wave Series). *The Institution of Engineering and Technology.* **1990**

[9] Sean Victor Hum. *Atmospheric Effects: Ionospheric Propagation. ECE422: Radio and Microwave Wireless Systems.* University of Toronto. **2017**

[10] Donald F. Specht. *A General Regression Network*. IEEE TRANSACTIONS ON NEURAL NETWORKS. VOL. 2 NO. 6, **1991**

[11] Sue Ellen Haupt, Antonello Pasini and Caren Marzban. *Artificial Intelligence Methods in the Environmental Sciences.* Springer. **2008**

[12] Vladimir M. Krasnopolsky. *The Application of Neural Networks in the Earth System Sciences: Neural Networks Emulations for Complex Multidimensional Mappings.* Atmospheric and Oceanographic Sciences Library 46. Springer. **2013**

[13] Vladimir M. Krasnopolsky, Michael S. Fox-Rabinovitzb, Hendrik L. Tolmanc, Alexei A. Belochitskib. *Neural network approach for robust and fast calculation of physical processes in numerical environmental models: Compound parameterization with a quality control of larger errors.* Neural Networks 21 (2008) 535–543, **2005**