Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

GEOFISICA

Ciclo XXXI

**Settore Concorsuale:** 02/C1

**Settore Scientifico Disciplinare:** FIS/06

# A three-dimensional regional assimilative model of the ionospheric electron density

**Presentata da:** Alessio Pignalberi (matricola n. 0000772781)

**Coordinatore Dottorato:**

Prof.ssa Nadia Pinardi

**Supervisore:**

Prof. Rolando Rizzi

**Co-supervisore:**

Dr. Michael Pezzopane

**Esame finale anno 2019**

## Abstract

The focus of this thesis is on the development, implementation, and validation of a three-dimensional regional assimilative model of the ionospheric electron density.

Empirical climatological models, like the International Reference Ionosphere (IRI) model (Bilitza et al. 2017), cannot predict the whole ionospheric variability, specifically under disturbed magnetic conditions. The model presented in this work has the purpose to improve the IRI description by implementing a data assimilation procedure, based on ionospheric measurements collected by several ground-based or satellite-based instruments.

The first phase of the development of the model, called IRI UPdate (IRI UP), is devoted to update the IRI model by ingesting effective indices ($IG_{12eff}$ and $R_{12eff}$) calculated after assimilating F2 layer characteristics values, measured by a network of ionosondes or derived by vertical total electron content values measured by a network of Global Navigational Satellite Systems receivers. The ingestion of effective indices in the IRI model allows to significantly improve the F2 layer peak density and height description. Being the F2 layer peak an anchor point for the whole IRI's vertical electron density profile, such procedure allows to update the whole profile.

The second phase of the development of the model is devoted to improve the modeling of the topside part of the ionospheric vertical electron density profile by making use of the IRI UP method and in-situ measurements collected by Swarm satellites.

Finally, a procedure called IonoPy, embedding the two aforementioned steps, assimilates the whole bottomside electron density profile measured by an ionosonde, thus further improving the ionospheric plasma description in the bottomside ionosphere.

All the procedures described in this thesis have been tested and validated by comparing them with other similar models or with independent datasets, for both quiet and disturbed conditions.

# Table of contents

# Introduction

The upper part of the Earth's atmosphere, called the ionosphere, is characterized by the presence of free electrons and ions whose concentration affects the propagation of electromagnetic waves. The study of the ionosphere played a substantial role in the past for radio-communication purposes, and, more recently, for telecommunications via satellite. Modern society relies on technologies strongly dependent on the ionosphere and near-Earth environment conditions. The malfunction of radars, navigation systems based on Global Navigational Satellite Systems (GNSS) such as the Global Positioning System (GPS) satellites, power distribution networks, the increased corrosion of pipelines, and the disruption of radio communications, can be directly attributed to phenomena occurring in such regions. This is why, over recent decades, the study of the complex physical atmosphere-ionosphere-plasmasphere-solar wind system, a subject usually called Space Weather, has increasingly gained importance.

Notably, Space Weather aims to describe the physical conditions characterizing the region of space surrounding the Earth and the wide variety of phenomena occurring in it (geomagnetic storms, ionospheric perturbations, geomagnetically induced electric currents at the Earth's surface). The principal driver of such phenomena is the Sun through its electromagnetic (above all Ultra-Violet and X ray emissions) and corpuscular radiation which can considerably affect the solar wind and, in turn, the geomagnetic field. Furthermore, the ionospheric plasma is coupled with the neutral component of the atmosphere so that, particularly in the lower part of the ionosphere, the vertical distribution of neutral molecules and the structure of neutral winds can strongly affect the plasma distribution. This is why, the forecast of the solar activity, the geomagnetic activity, and neutral atmosphere changes are essential to perform reliable ionospheric predictions.

The complex interactions occurring between the atmosphere, the ionosphere, and the plasmasphere, make difficult to describe the spatial and temporal variations of the ionospheric plasma, for both quiet and, particularly, magnetically disturbed conditions. Ionospheric empirical climatological models, like the International Reference Ionosphere (IRI) model (Bilitza et al. 2017), are not able to predict the whole ionospheric variability, specifically under disturbed magnetic conditions. Therefore, as established by the IRI community (Bilitza et al. 2011, 2017), real-time space-sparse ionospheric data should be used in conjunction with climatological models in order to get a picture of the ionospheric plasma variability as close as possible to the real conditions.

This thesis tries to fill this gap implementing methods to update the IRI model, through the assimilation of ionospheric measurements collected by different measurement techniques. The data assimilation approach here proposed aims to improve a climatological background model to obtain a "weather" model, as is operationally done for meteorological forecasts. This approach is currently made possible also in the ionospheric field thanks to the increasingly availability of real-time measurements taken by ground-based and satellite-based ionospheric instruments. Data assimilation is a process of merging measurements with a model to improve the estimate of the ionospheric conditions over the area covered by the model, also where direct measurements are not available. Thus, by means of the data assimilation, it is possible to spatially expand the effectiveness of limited measurements using the model and, at the same time, to increase the accuracy of model estimates. To accomplish this task, in the framework of this thesis several data assimilation methods have been developed, which will be carefully described.

Chapter 1 contains a brief introduction about the ionosphere. Its main features are introduced, as well as the physical principles behind its formation and the presence of several ionospheric layers. The main ionospheric sounding techniques, whose measurements will be employed in this work, are also described.

Chapter 2 is devoted to the IRI empirical model. A careful description of this model is needed not only because such model has been used as background, but also because many of the choices made in the development of the data assimilation methods were driven by definite IRI's features. The analytical formulations used by IRI to model each part of the whole vertical electron density profile of the ionosphere are described.

Chapter 3 introduces the main ideas behind the development of the IRI UPdate (IRI UP) method, a new data assimilation method to update the IRI model and obtain a more reliable representation of the ionospheric plasma distribution. The procedure employed to calculate maps of effective indices $IG_{12eff}$ and $R_{12eff}$, by assimilating values of *fo*F2 and *M*(3000)F2 recorded by a network of reference ionospheric stations, is described in detail.

Chapter 4 contains an implementation of the IRI UP method based on the assimilation of *fo*F2 and *M*(3000)F2 values collected by a European network of ionosonde stations. The collection and filtering of ionosonde measurements, the effective indices calculation and their mapping by means of the Universal Kriging method, the choice of the best variogram model, and the ingestion of effective indices maps in IRI to produce updated F2 layer characteristics, are described. A validation of the method on 30 periods embedding geomagnetic storms occurred between 2004 and 2016 is reported.

Chapter 5 introduces another implementation of the IRI UP method, based on the assimilation of vertical total electron content (vTEC) measurements from a South-African GNSS receiver network. An original mathematical procedure developed to obtain *fo*F2 and *M*(3000)F2 values from assimilated vTEC measurements is described. This allows the calculation of effective indices which are used to update the IRI model. A validation of the method on several quiet and disturbed days in 2017 is presented. Such IRI UP variation is particularly useful when applied on regions poorly covered by ionosonde stations but with a good GNSS receivers' coverage.

Chapter 6 deals with the topside ionosphere modeling. It is first demonstrated that the IRI model has difficulty in modeling such ionospheric region, by comparing its output with measurements collected by Swarm satellites. Then, an empirical method to model this part of the ionosphere, by using Swarm Langmuir probe measurements and the IRI UP method, is proposed. The procedure used to model the topside effective scale height as a function of *fo*F2 and *hm*F2 characteristics is described and statistically tested.

Chapter 7 describes the IonoPy model. This model, as a first step, embeds the IRI UP method to update the climatological background IRI model; then, as a second step, it updates the topside ionosphere description by means of the topside model described in chapter 6. Furthermore, it includes a procedure able to assimilate the whole bottomside electron density profile measured by an ionosonde. Such procedure is first mathematically described and then its application for a very disturbed period is shown. A comparison between the IonoPy model and other similar assimilation models, for the Mediterranean region, during some intense and severe geomagnetic storms, is also presented and discussed.

Each chapter of the thesis begins with a brief description which has to be intended as an abstract of the chapter. Chapters from 3 to 7, at the end, contain also a summary which summarizes the main results and findings described in the chapter.

# 1. Introduction on the ionosphere

The term *ionosphere* identifies the upper part of the Earth's atmosphere in which free electrons and ions exist under the control of the gravity and magnetic field of the planet, and whose concentration is sufficient to affect the propagation of electromagnetic waves (Rishbeth and Garriott 1969; Ratcliffe 1972; Kelley 2009; Zolesi and Cander 2014). This definition of the ionosphere (the one adopted by IEEE Std 211-1997 (IEEE 1998)) highlights the strong connection between the knowledge about the ionosphere and the development of the electromagnetic wave propagation techniques.

Notably, the evolution of the ionospheric science has followed the development of radio propagation measurements and studies. While the existence of ionized layers in the upper atmosphere had been postulated during the eighteenth century, it was only in 1901 that the first transoceanic radio link realized by Guglielmo Marconi provided an experimental proof confirming that long distance wireless communication was possible in spite of the obstacle caused by the curvature of the Earth. Afterwards, in 1902 two scientists, Oliver Heaviside and Arthur E. Kennelly, independently proposed the same explanation of the Marconi's experiment: the upper atmosphere is characterized by the presence of charged particles so that radio waves can be transmitted across the Atlantic Ocean not by the line of sight but within a conducting cavity between this particular region and the Earth (Zolesi and Cander 2014). The vertical structure of this region was finally described during the 1920s thanks to the systematic experiments and theoretical studies of Edward V. Appleton that led to the establishment of the magneto-ionic theory which describes the propagation of electromagnetic waves in the ionospheric medium (Ratcliffe 1962). This was possible also thanks to the technological developments of the radio wave propagation technique made by Gregory Breit and Merle A. Tuve that are even now the cornerstone of the vertical ionospheric sounding performed by ionosondes (Hunsucker 1991). The establishment of the magneto-ionic theory and of the vertical ionospheric sounding by ionosondes marked the beginning of systematic ionospheric measurements.

In this chapter, a brief review of the principal ionosphere's properties is presented. The basic Chapman theory is introduced to explain the ionospheric layers formation, and their main properties are briefly described. Finally, the principles behind the most spread ionospheric sounding technique, i.e. the vertical ionospheric sounding by ionosonde, are introduced, along with other techniques of which measurements have been used in this work.

## 1.1 The Earth's ionosphere

The ionosphere is a plasma, i.e. is a weakly ionized gas consisting of free electrons and ions (Rishbeth and Garriott 1969; Ratcliffe 1972; Kelley 2009; Zolesi and Cander 2014). In fact, only a small fraction of the neutral gas density of the atmosphere is ionized (Prolss 2004). Typical ratios of electron to neutral gas density are $10^{-2}$ at the top of the ionosphere (at a height of about 1000 km), $10^{-3}$ at the height of the ionospheric absolute electron density maximum (located roughly between 200 and 450 km of height), $10^{-8}$ at the bottom of the ionosphere (at a height of about 100 km). Because such ionized gas is permeated by the geomagnetic field, it is called a *magnetoplasma*. The geomagnetic field has a strong effect on the ionospheric plasma distribution because a plasma can be assimilated to an electrically charged bi-component fluid made by electrons and ions. Anyway, strong electric fields inside the plasma make it globally neutral.

For what concerns ionospheric studies, the most important physical quantity to be studied is the electron density; in particular, the vertical electron density profile. The electron density assumes considerably high values (from a radio wave propagation point of view) in a region located between 60 and 1000 km of height. These limits mark the boundaries of the ionosphere but, however, they must not be considered fixed. They are highly variable in both space and time, above all the upper limit which marks the separation between the underlying ionosphere and the overlying protonosphere. A schematic ionospheric vertical electron density profile is shown in Fig. 1.1. Referring to Fig. 1.1, the ionosphere is characterized by an absolute maximum in the electron density (located roughly between 200 and 450 km of height) which divides the ionosphere in two well separated regions; the *bottomside* part below the absolute maximum, and the *topside* part above the absolute maximum. This sharp distinction between the bottomside and the topside ionosphere is not only a matter of nomenclature, but reflects very distinct physical and chemical processes happening in the two regions, and it also affects the experimental techniques employed to sound them (Rishbeth and Garriott 1969; Ratcliffe 1972).

Above about 100 km of height the molecular species composing the neutral atmosphere are vertically distributed depending on their molecular weight. This is why heaviest molecules ($O_2$, $N_2$, and NO) are more present at low ionospheric heights (60-250 km), lightest atoms (H and He) dominate the upper ionosphere (above about 600 km), and O atoms dominate the region around the absolute maximum (200-600 km). The ionizing effect of the solar radiation, in the Extreme Ultra-Violet (EUV) and X-rays spectral frequency range, is the main driver that leads to the formation of free electrons and ions in the ionosphere (Ratcliffe 1972). Minor drivers, like the precipitation of energetic particles, can be neglected to a first approximation. The combined effect of the ionizing solar radiation and the vertical distribution of molecular and atomic species, sets up several maxima in the electron density (being the ionospheric plasma globally neutral, one can refers to electron or ion density at the same time). Such ionospheric maxima identify corresponding *ionospheric layers*. The basic principles that lead to the formation of the ionospheric layers and their features are described in the next section.

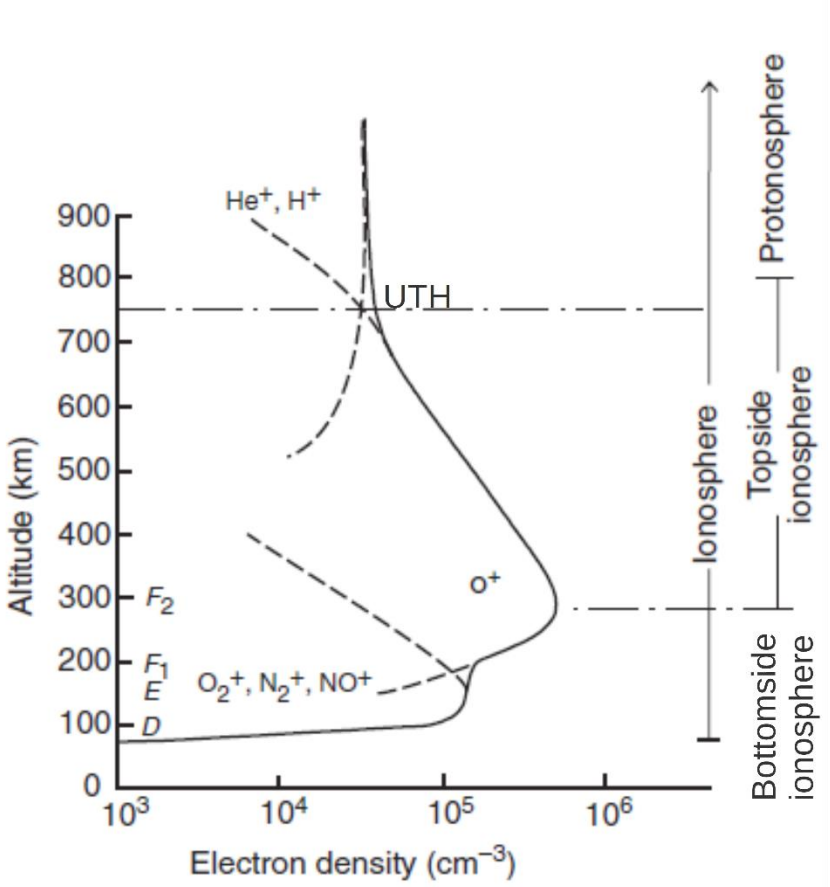


**Figure 1.1:** Schematic representation of the ionospheric vertical electron density profile. Molecular and atom ion species characterizing each altitude range are highlighted. The Upper Transition Height (UTH) is the height between the oxygen ion and hydrogen ion dominated plasma regions.

## 1.2 Basic principles on the ionosphere formation

### 1.2.1 The formation of the ionospheric layers

When the solar radiation is absorbed by the atmosphere it heats it, dissociates its molecules, and liberates free electrons. The rate at which dissociation or ionization are produced (the *rate of production*), at any level, is proportional to the product of the gas concentration and the intensity of the radiation. At the top of the atmosphere the rate of production is small because the concentration is small; as the radiation penetrates downwards the concentration increases and with it the rate of production. However, the strength of the radiation decreases by absorption in the upper atmospheric layers and its rate of decrease downwards is greater than the rate of increase of the concentration. There is thus a height where the rate of production reaches a maximum.

Such simple physical arguments led S. Chapman (1931) to derive an analytical expression for the production of an ionospheric layer, since then named *Champan layer*, valid for a simplistic atmosphere constituted by only one type of molecules (having all the same absorption cross-section independent of the wavelength), horizontally stratified and vertically exponentially distributed with a constant scale height (because of the assumption of constant temperature), dissociated by a monochromatic solar radiation through photo-ionization, with the electron loss caused by recombination, and neglecting any dynamics and diffusion mechanism.

The rate of production $q$ for a Chapman layer, on the basis of these assumptions, is (Ratcliffe 1972)

$$\begin{cases} q(z,\chi) = q_0 \exp\left[1 - z - \sec\chi \exp(-z)\right] \\ q_0 = \dfrac{\sigma n_0 I_\infty}{e} \\ z = \dfrac{h - h_0}{H} \\ h_0 = h - H\log(\sigma n_0 H) \end{cases}, \tag{1.1}$$

where $q_0$ is the production rate at the height $h_0$ (for $\chi$=0) of the maximum of production rate. $\sigma$ is the absorption cross-section, $n_0$ the concentration of the atmosphere at the height $h_0$, $I_\infty$ the monochromatic solar irradiance outside the atmosphere, $H$ the atmosphere scale height, $\chi$ the zenith angle of incidence of the solar radiation (the angle between the local vertical and the Sun position), and $e$ the Napier's constant.

From Eq. (1.1) one obtains (Ratcliffe 1972)

$$\begin{cases} q_M(\chi) = q_0 \cos\chi \\ z_M(\chi) = -\log(\cos\chi) \end{cases}, \tag{1.2}$$

where $q_M(\chi)$ is the absolute maximum of the rate of production, and $z_M(\chi)$ the corresponding reduced height, both function of the solar zenith angle $\chi$.

The shape of the function $q(z, \chi)/q_0$ (from Eq. (1.1)) for several values of $\chi$ is shown in Fig. 1.2. The coordinates of the maximum are those of Eq. (1.2).

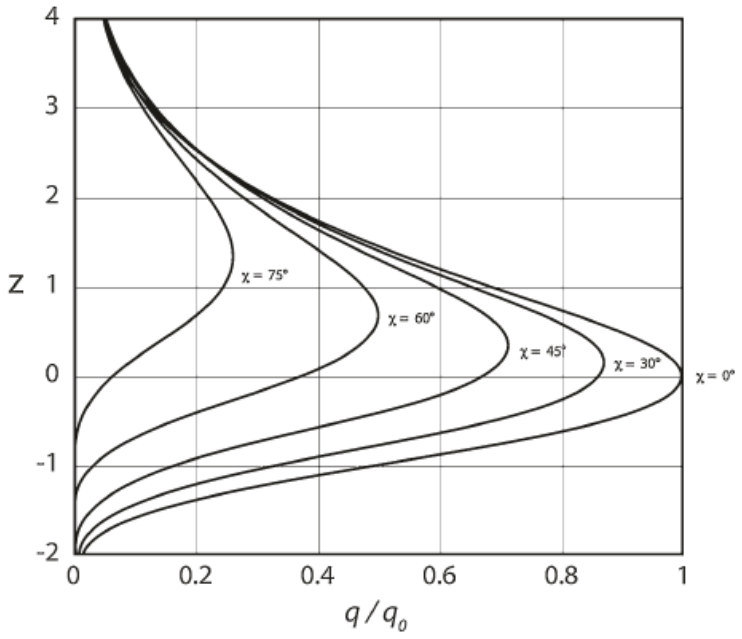


**Figure 1.2**: The function $q(z, \chi)/q_0$ for several values of $\chi$ (from Zolesi and Cander 2014).

Looking at Fig. 1.2 and Eqs. (1.1)-(1.2), it appears clear that:

1. The maximum rate of production depends only on the solar zenith angle $\chi$ and on the solar irradiance outside the atmosphere $I_\infty$ (having considered only one type of molecules with assigned absorption properties). The dependency on the solar irradiance leads to a strong solar activity dependence. Instead,

the dependence on the solar zenith angle $\chi$ causes a strong daily and seasonal dependence because, in a geocentric coordinates system, $\cos\chi = \sin\varphi\sin\delta + \cos\varphi\cos\delta\cos\omega$ (Zolesi and Cander 2014), where $\varphi$ is the geographic latitude , $\delta$ the solar declination on a given day of the year, and $\omega$ the hourly angle of the Sun;

2. The height of the maximum rate of production increases as the Sun goes down to the horizon, i.e. as $\chi$ increases.

The *continuity equation* (or the equation of electronic equilibrium) for the electron density $N_e$ can be written as

$$\frac{\partial N_e}{\partial t} = q - l + d \ , \tag{1.3}$$

where $q$ is the electron production rate defined in Eq. (1.1), $l$ is the *rate of electron loss*, and $d$ is a term that takes into account *movements of electrons* due for instance to diffusion, neutral winds, or vertical electromagnetic drift.

Neglecting the effect of dynamics ($d$=0), two different Chapman's layer can be deduced from the continuity equation assuming $q$ from Eq. (1.1) and considering two different electron loss processes:

- The α-Chapman layer, for which the electron loss process considered is the *recombination* where free electrons recombine with positive ions to produce neutral atoms. The rate of electron loss is $l_r = \alpha N_e^2$ , where $\alpha$ is the recombination coefficient. In the simplifying hypothesis of photo-chemical equilibrium $\partial N_e / \partial t = 0$ , from Eqs. (1.1) and (1.3) one obtains

$$N_e(\chi, z) = \sqrt{\frac{q_0}{\alpha}} \exp\left[\frac{1 - z - \sec\chi \exp(-z)}{2}\right]. \tag{1.4}$$

  From Eq. (1.4) the maximum of electron density and the corresponding reduced height are

$$\begin{cases} N_M(\chi) = \sqrt{\dfrac{q_0}{\alpha}\cos\chi} \\ z_M(\chi) = -\log(\cos\chi) \end{cases} . \tag{1.5}$$

  The electron density for an α-Chapman layer has the same behavior of the production rate $q$ shown in Fig. 1.2;

- The β-Chapman layer, for which the electron loss process considered is the *attachment* of electrons to neutral atoms. The rate of electron loss in this case is $l_a = \beta n N_e = \beta n_0 \exp(-z) N_e$ , where $\beta$ is the attachment recombination coefficient, and $n$ the neutral density of the atmosphere which is exponentially distributed along the vertical. In this case, one obtains

$$N_e(\chi, z) = \frac{q_0}{\beta n_0} \exp\left[1 - \sec\chi \exp(-z)\right]. \tag{1.6}$$

  The function $N_e$ of Eq. (1.6) is monotonically increasing with a maximum located at the top of the atmosphere.

### 1.2.2 The vertical electron density profile

Chapman's theory of ionospheric layers formation can explain only some of the features of the vertical electron density profile shown in Fig. 1.1. As stated before, the vertical distribution of atmospheric molecules and atoms must be considered, considering that each of these species has a different chemical and atomic behavior depending on the spectral frequency of the incident solar radiation. Briefly, the altitude for which the absorption of EUV is

maximum is reported in Fig. 1.3, where species responsible for such absorption are also indicated. Clearly, the primary ions which form the ionosphere are produced in the bottomside ionosphere by photo-ionization. Secondary ions are produced by chemical reactions of primary ions. Because the photo-ionization is very effective in producing ions at low ionospheric heights (60-170 km), the presence of a maximum of the electron density located well above such heights, and the presence of ions in the topside ionosphere, have to be explained by other physical phenomena. Specifically, the term *d* of Eq. (1.3) plays a significant role and cannot be neglected. In fact, the ambipolar diffusion of ions and electrons along the geomagnetic field lines (ions and electrons diffuse in the same direction at the same rate due to electric fields arising to maintain the quasi-neutrality of the plasma), the effect of neutral atmosphere changes, and dynamics effects caused by both neutral winds and the electromagnetic drift, come into play to explain the electron density absolute maximum formation and the topside profile.

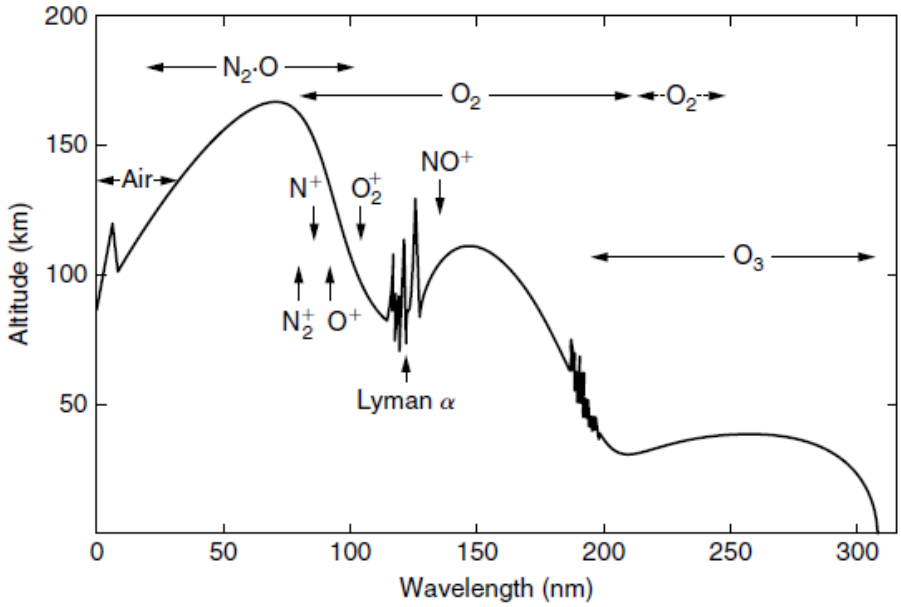


**Figure 1.3:** Altitude where the absorption of solar radiation due to different atmospheric species is maximum.

The *ionospheric layers* (see Fig. 1.1) are identified with different letters. Historically, the first layer discovered by Appleton was that named *E layer*, because of the electric field reflected by the ionosphere. Subsequently, the layer above was called *F layer*, and the one below the *D layer*. During day, the F layer is splitted in two layers named *F*1 and *F*2.

Such ionospheric layers undergo large spatial and temporal variations. Their main characteristics are now briefly summarized.

**D layer**

The D layer lies at an approximate altitude of 50–90 km. The neutral species composing the D layer consist mostly of $N_2$, $O_2$, Ar, $CO_2$, He, and a highly variable quantity of $O_3$ and $H_2O$. The ionized component consists mostly of $NO^+$ as the major positive charge carrier, while electrons and $O_2^-$ are the negative charge carriers. Solar Lyman-alpha radiation ($\lambda$=121.6 nm) appears to be the most important for ionizing NO, solar X-rays ($\lambda$ < 0.8 nm) for ionizing $N_2$, $O_2$, and Ar, solar ultraviolet radiation ($\lambda$ < 111.8 nm) for ionizing unstable $O_2$, and galactic cosmic rays for ionizing all atmospheric constituents. Hence, the D layer is totally controlled by photo-chemical processes driven by the solar radiation and cosmic rays. D layer electron densities are typically around $10^8$–$10^9$ el/m$^3$. The D layer shows typical diurnal, seasonal, and solar cycle variations.

**E layer**

The E layer lies at an approximate altitude of 90–140 km. The ionization in the E layer is mainly caused by X-rays in the 8–10.4 nm range and the EUV radiation ranging from 80 nm to the Lyman-beta radiation (102.6 nm), producing the primary $NO^+$, $O_2^+$ , and secondary $O^+$, $N_2^+$ ion components. Again, the E layer is a purely photo-chemical layer. In this region, the electron density reaches average maximum values of ~$10^{11}$ el/$m^3$ at around 110 km. The E layer exhibits a very distinct α-Chapman layer behaviour during daytime with a daily maximum at local noon (when $\chi$=0), a seasonal maximum in summer, and a strong solar cycle dependence, as inferable from Eq. (1.4). The E layer departs from a pure α-Chapman layer in its nighttime behavior; in fact, it does not disappear completely at night and remains in a condition of weak ionization.

**F1 layer**

The F1 layer lies at an altitude of approximately 140–210 km. The ionized component of the F1 layer consists mostly of $NO^+$ and $O_2^+$ as primary, and $O^+$ and $N^+$ as secondary positive charge carriers. The main source of ionization is the EUV in the wavelengths $\lambda \approx 58.4$ and 30.4 nm. Like the E layer, the F1 layer is an α-Chapman layer, with the maximum electron density reaching about $2 \cdot 10^{11}$ el/$m^3$ at noon, exhibiting a dependence on both the solar zenith angle and the solar activity. Accordingly, the F1 layer is more pronounced in summer than in winter, it disappears at night and sometimes in winter even during the day. Under those conditions the F1 and F2 layers form a single ionospheric F region. The height of the F1 layer electron density maximum is located approximately between 160 and 180 km, and varies with the solar activity, the season, and the geomagnetic activity.

**F2 layer**

The F2 layer is always present under every solar-terrestrial condition, making it the most important layer of the ionosphere. The F2 layer has a different nature from the photo-chemically dominated lower layers. The main characteristic of the F2 layer is its high variability, on timescales ranging from the 11 years of the solar cycle, to a few minutes during magnetically severely disturbed conditions. In the F2 layer, the dominant charge carriers are $O^+$ ions with secondary $H^+$ and $He^+$. The main electron-ion source in the F2 region is the process of photo-ionization by EUV radiation in the wavelength range from 5 to 102.7 nm; anyhow, differently from the lower regions, the transport of plasma produced at lower regions plays a fundamental role in the formation of the F2 layer. In the lower F2 region (the region between F1 and F2 electron density maxima) the most effective electron loss process is the *attachment* of electrons to neutral atoms. This should lead to a β-Chapman layer behavior (Eq. (1.6)) in which the electron density increases with height. Obviously, this increase in the electron density cannot continue indefinitely because of the exponential reduction of the atmospheric neutral density with height. In fact, it can be demonstrated that, in the topside ionosphere the electron density should decrease with a plasma scale height (Ratcliffe 1972; Prolss 2004) due to ambipolar diffusion processes. The combined effect of the electron density increase in the lower F2 region and of the electron density decrease in the upper F2 region leads to a maximum in the electron density located at a height for which the scale times of these two opposing processes are comparable. The F2 layer has a maximum electron density between about $10^{11}$ and $8 \cdot 10^{12}$ el/$m^3$, at a height between 200 and 450 km. Because the F2 layer is a non-Chapman layer, one might think that there is no solar zenith angle dependence of the diurnal and latitudinal variations of the electron density. However, the Sun does affect the electron density of the F2 layer causing a rapid increase after sunrise, with maximum values occurring during the day. The replenishment from the protonosphere plays a significant role in maintaining the F region during the night. To explain the complex behaviour of the F2 layer, chemical changes in the neutral atmosphere, diurnal heating and cooling of the atmosphere, ambipolar diffusion along the geomagnetic field lines, winds in the neutral atmosphere, and electric fields generating electrodynamic drifts, should be concurrently be taken into account. The departures of the F2 layer from a Chapman layer behavior are usually named *anomalies*. All of this makes the prediction of the F2 layer characteristics a very difficult task.

**Topside ionosphere**

The topside ionosphere extends from the F2 layer maximum to the Upper Transition Height (UTH), the height between the oxygen ion and hydrogen ion dominated plasma regions (as depicted in Fig. 1.1). The location of the UTH (600-1000 km) is highly variable in both time and space. The plasma density distribution of this region is largely determined by field-aligned plasma flows and plasma transport processes (Rishbeth and Garriott 1969; Ratcliffe 1972). The topside ionosphere is characterized by values of the electron density monotonously decreasing with height, because of the ambipolar diffusion and other transport processes. The plasma scale height $H_\mathrm{p} = \frac{k_\mathrm{b} T_\mathrm{p}}{m_\mathrm{i} g}$ drives this decrease, where $k_\mathrm{b}$ is the Boltzmann constant, $T_\mathrm{p} = T_\mathrm{e} + T_\mathrm{i}$ is the plasma temperature ($T_\mathrm{e}$ and $T_\mathrm{i}$ are the electron and ion temperature, respectively), $m_\mathrm{i}$ is the ion mean molar mass, and $g$ is the acceleration due to gravity. From its definition, the plasma scale height is strongly connected with the ionospheric dynamics, plasma thermal structure and composition. Because this region cannot be probed by many ground-based ionospheric instruments, its spatial and temporal behavior is only partially known.

## 1.3 Experimental techniques for probing the ionosphere

### 1.3.1 Ionospheric sounding performed by an ionosonde

**The vertical incidence ionospheric sounding**

The traditional, but still the more widespread, method for probing the ionosphere is the vertical ionospheric sounding made by ionosondes, which can be considered the precursor of the radar technique.

Such method of measurement of the electron density of the ionosphere as a function of the altitude is based on the principle that when an electromagnetic wave vertically penetrates the ionospheric plasma, a reflection occurs at the level where the atmospheric *refractive index n* becomes zero.

According to the *magneto-ionic theory* (Ratcliffe 1962), the atmospheric refractive index of the ionospheric medium, in absence of the geomagnetic field and neglecting the absorption, is

$$n^2 = 1 - \frac{f_\mathrm{N}^{\,2}}{f^2} \ , \tag{1.7}$$

which is dependent on the *plasma frequency* $f_\mathrm{N}$

$$f_\mathrm{N} = \sqrt{\frac{N_\mathrm{e} e^2}{4\pi^2 \varepsilon_0 m}} \ , \tag{1.8}$$

where $N_\mathrm{e}$ is the electron density, $e$ the electron charge, $\varepsilon_0$ is the vacuum electric permittivity, and $m$ the electron mass; $f$ is the frequency of the emitted electromagnetic wave.

Because the reflection happens when the refractive index becomes null, from Eq. (1.7) the incident frequency $f$ must be equal to $f_\mathrm{N}$.

The ionosonde is an instrument designed to detect the electron density of the ionospheric plasma as a function of the height, emitting radio impulses scanning the transmitting frequency from 1 to 20 MHz, and measuring the time delay of electromagnetic echoes.

As the incident frequency increases, the radio impulses penetrate the ionosphere up to higher altitudes and are reflected by higher levels of electron density $N_\mathrm{e}$. However, according to the electron density profile shown in Fig. 1.4, the electron density $N_\mathrm{e}$ varies and increases with the altitude up to a relative maximum and even an absolute maximum *Nm*. The maximum electron density *Nm* corresponds to the maximum reflected incident frequency, called the critical frequency *fo* (notably, *fo* stands for the critical frequency of the ordinary mode of propagation, the meaning of which will be clear shortly), and the two quantities are related throught the following relationship

$$Nm = 1.24 \cdot 10^{10}\, fo^2 \ , \tag{1.9}$$

with *Nm* expressed in el/m$^3$ and *fo* in MHz. The (1.9) has been obtained by expliciting the physical constants in Eq. (1.8).

Therefore, the critical frequencies (*fo*E, *fo*F1, and *fo*F2) are associated to the relative and absolute maxima of the ionospheric vertical electron density profile (named *Nm*E, *Nm*F1, and *Nm*F2).

The time delay of the received echo $\Delta t$, for a definite emitted frequency signal, allows the calculation of the *virtual height* $h'$ of the point where the incident frequency is reflected through the following relationship

$$h' = c\frac{\Delta t}{2} \ , \tag{1.10}$$

where *c* is the velocity of the electromagnetic pulse in the free space.

$h'$ is called the virtual height because it is calculated assuming that the signal travels in the free space. Being the refractive index of the medium different from unity, the signal travels with a group velocity lower than *c*, thus the *real reflection height h* will be lower than the virtual one.

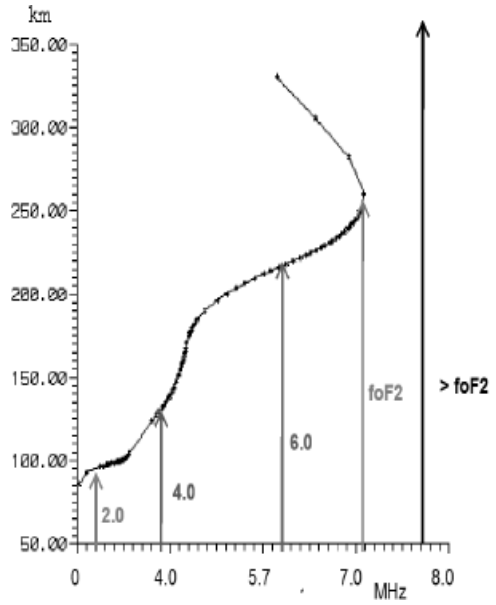


**Figure 1.4:** Schematic representation of the ionosonde vertical ionospheric sounding. The x-axis represents the frequency (in MHz) of the signal emitted by the ionosonde, the y-axis represents the real height (in km). Only signals emitted with a frequency lower than *fo*F2 are reflected by the ionosphere.

**The ionosonde and the ionogram**

The ionosonde is a bistatic radar system. This means that the transmitting antenna is separated from the receiving antenna and that there are two separate energy paths. Fig. 1.5 schematically illustrates the main components of an ionosonde:

1. a synthesizer that generates the radio frequency to be transmitted and a transmitter that amplifies the electromagnetic pulse to be emitted from an antenna system;
2. a receiver apparatus with an antenna system that amplifies the received signal;
3. a control system able to program the emission time in accordance with the expected arrival of the signal at the receiver, and analyze and elaborate the signal received as a graphic trace.

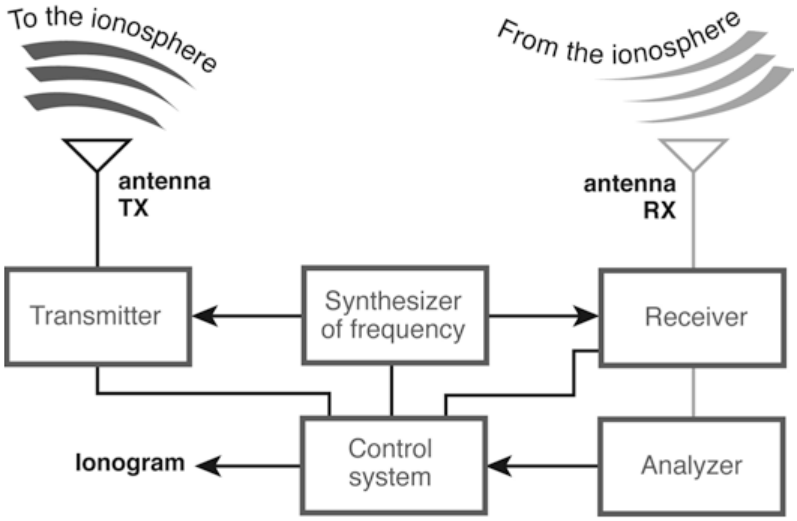


**Figure 1.5:** Simple scheme of the main components of an ionosonde (from Zolesi and Cander 2014).

In the last decades, the technological progress brought by the solid state electronics and computer systems, allowed the development of digital ionosondes, like the Digital Portable Sounder developed by the Lowell University (DPS, Bibl and Reinisch 1978), and the Advanced Ionospheric Sounder developed by the Istituto Nazionale di Geofisica e Vulcanologia (AIS-INGV, Zuccheretti et al. 2003). Modern digital ionosondes are fully programmable and use specific softwares for the automatic interpretation of digital ionograms and for the interfacing with the Internet for Space Weather applications.

The output of a vertical ionospheric sounding made by ionosondes is the *ionogram*. The ionogram is a graphical representation of the time delay of the radio signal (*y* axis in Fig. 1.6) emitted by the ionosonde and reflected by the ionosphere, for a specific frequency (*x* axis in Fig. 1.6). Notably, the ionosonde emits radio impulses scanning the transmitting frequency from 1 to 20 MHz, and measures the time delay of the electromagnetic echo. If the emitted radio signal is reflected by the ionosphere, the virtual height is calculated with Eq. (1.10) and a point is placed in the ionogram for the specific frequency emitted. The sequence of each reflected signal makes up the ordinary (red trace in the right panel of Fig. 1.6) and the extraordinary traces (green trace in the right panel of Fig. 1.6). The presence of two reflected signals in the ionogram is due to the geomagnetic field which makes the ionospheric medium anisotropic, causing a birefringence phenomenon. In the left panel of Fig. 1.6, a sketch of an ideal daytime ionogram is represented, with the main ionospheric characteristics reported. In this graph, the ordinary trace is a solid line while the extraordinary one is dashed. The scaling of the ionospheric characteristics operation is nowadays automatically performed by specific softwares, like the Automatic Real-Time Ionogram Scaler with True height (ARTIST, Galkin et al. 2008; Galkin and Reinisch 2008) for the DPS ionosonde, and Autoscala (Scotto 2009; Scotto et al. 2012) for the AIS-INGV ionosonde. Nevertheless, many research institutes, like INGV, invests time and efforts in the manual scaling of ionograms which is performed by trained personal.
The ionospheric characteristics scaled from an ionogram typically are:

- The highest frequencies reflected by the main ionospheric layers (the so-called critical frequencies as stated above), for the ordinary trace: *fo*F2, *fo*F1, and *fo*E;
- The minimum virtual height of the main ionospheric layers, for the ordinary trace: $h'$F2, $h'$F, and $h'$E;
- The highest reflected frequency *fx*I associated to the extraordinary trace;
- The lowest frequency *fmin* recorded in the ionogram;
- The ionospheric characteristics of the *Sporadic E layer*, which is an ionospheric layer that occurs randomly and only in particular conditions. In particular, its ordinary highest reflected frequency *fo*Es, blanketing frequency *fb*Es, and minimum virtual height $h'$Es.

The critical frequencies are associated to the cusps for the E and F1 layers (*fo*F1 and *fo*E) and to the vertical asymptote for the F2 layer (*fo*F2 and *fx*I), possibly present in the ionogram. At each critical frequency is associated a maximum in the electronic density through the relation (1.9), absolute for the F2 layer (*Nm*F2), relative for the F1 and E layers (*Nm*F1 and *Nm*E, respectively).

In the right panel of Fig. 1.6 an ionogram recorded by a DPS4 Digisonde and autoscaled by ARTIST software is represented. The ionospheric characteristics automatically scaled by ARTIST software are shown to the left of the ionogram. Besides the ordinary and extraordinary traces (red and green, respectively), also the estimated vertical electron density profile (thin black line, solid for the bottomside, dashed for the topside) is reported on the ionogram. The estimation of the vertical electron density profile is a mathematical inversion problem in which the real heights have to be deduced from the virtual ones (Budden 1961). In particular, the real heights (*hm*F2, *hm*F1, and *hm*E) associated to the maxima of the profile, as well as the real height for each reflected frequency can be obtained, allowing the specification of the whole bottomside electron density profile. For the topside part, being hidden to the ground-based ionosondes, only an estimate based on empirical models can be provided.

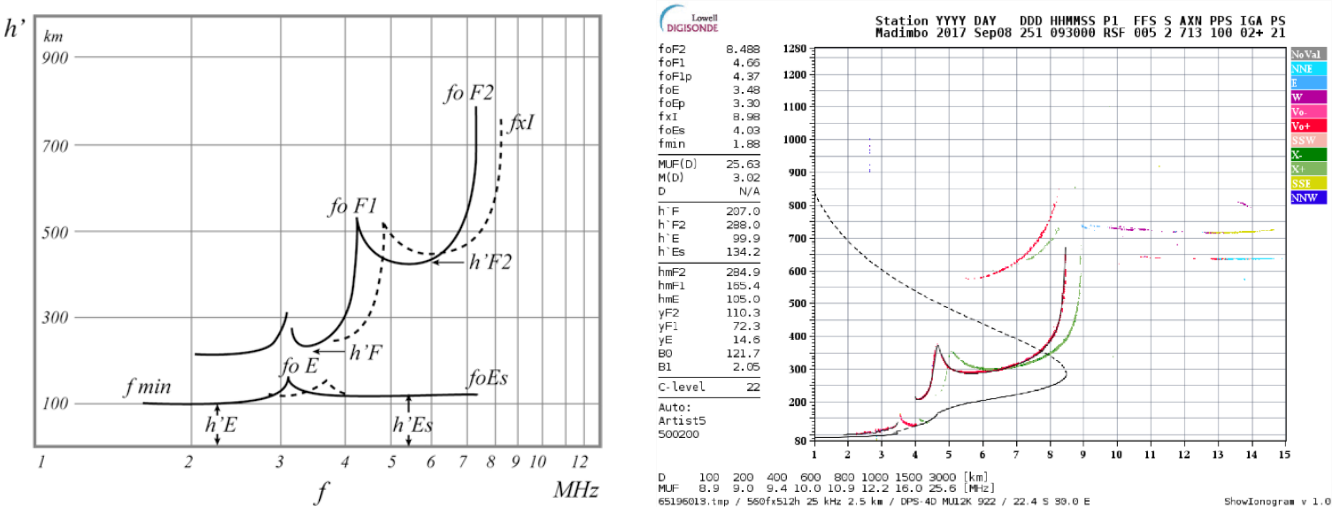


**Figure 1.6:** (Left panel) Schematic representation of a daytime ionogram with routinely scaled characteristics (from Zolesi and Cander 2014). (Right panel) Ionogram recorded by a DPS4 Digisonde and autoscaled by the ARTIST software.

### The oblique incidence ionospheric sounding

The main purpose of the radio communication is linking a transmitter to a receiver, located in different points of the Earth's surface, by means of a radio signal. Hence, an oblique incidence ionospheric sounding is needed to get information on the ionosphere for this particular geometry, where the two points can be thousands of kilometres apart and the link can be obtained with multiple reflections.

The experimental instrumentation employed to accomplish this task is again the ionosonde emitting radio impulses of increasing frequency. In this case, two ionosondes placed very distant from each other (order of thousands of kilometres), and perfectly synchronized, are needed to make an oblique ionospheric sounding. The output is an oblique ionogram.

Fig. 1.7 shows a simple scheme of a radio link between two distant points on the Earth's surface, depending on the angle of elevation $\beta$ (measured counter clockwise from the ground) at which the radio signal is emitted, considering only one reflection in the ionosphere. The experimental evidence shows that radio signals emitted

obliquely with a small angle of incidence $\varphi$ to the ionosphere (rays *a* and *b* in Fig. 1.7), escape into space; radio signals emitted with an angle of incidence greater or equal than $\varphi_o$ (rays *c, d, e,* and *f* in Fig. 1.7), are reflected by the ionosphere. The critical angle of incidence being (Zolesi and Cander 2014)

$$\varphi_0 = \arccos\left(\frac{fo}{f}\right), \tag{1.11}$$

where *fo* is the critical frequency that would be measured by a vertical sounding performed in the middle of the link (it is *fo*F2 if the reflection occurs in the F2 region), and *f* the frequency of the obliquely emitted ray.

Rays reflected by the ionosphere will arrive at greater distances as the elevation angle $\beta$ is decreased till to be tangential to the ground. The *skip distance* is defined as the minimum distance for which it is possible to link two points at a definite frequency by means of an ionospheric reflection. Within this distance, also known as the *silent distance*, only the ground wave propagation is possible.

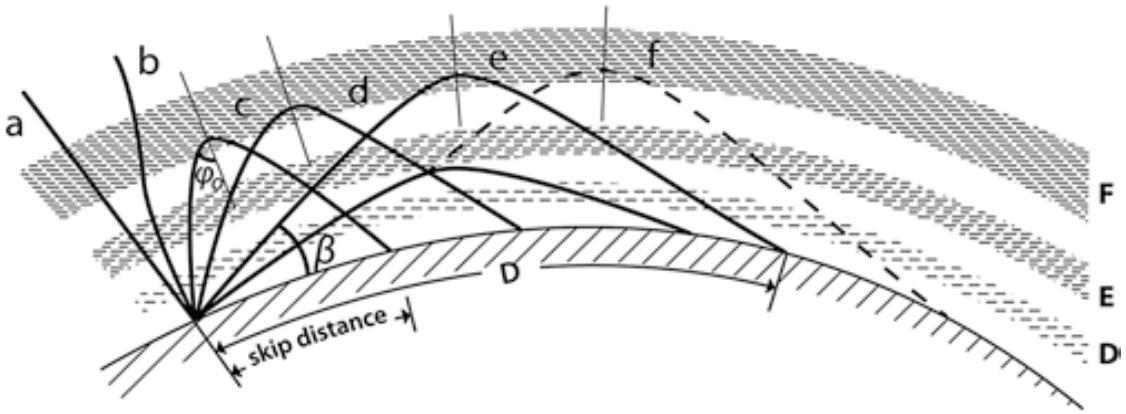


**Figure 1.7:** Simple scheme of an oblique radio link (from Zolesi and Cander 2014).

For radio communication purposes is important to know which is the best frequency and at which angle to emit the radio signal for obtaining a reliable and stable linkage. This depends on the geometry of the linkage and on the properties of the ionosphere and can be deduced from theoretical considerations.

In a simplified model of a flat ionosphere over a flat Earth, as illustrated in the left panel of Fig. 1.8, the reflection of a radio signal emitted with an angle of incidence $\varphi_o$ to the vertical, starts at the point $\delta$, at the beginning of the ionosphere, changes to downwards at point B, and achieves a complete reflection at point $\varepsilon$ when the radio signal comes out from the ionosphere. According to the *Theorem of Breit and Tuve* (Breit and Tuve 1926) the time of propagation of the ray between the points TBR in Fig. 1.8 (left panel) is equal to the time taken by a ray propagating at the speed light *c* to cover the path TCR. Furthermore, *Martyn's Theorem* (Martyn 1935) says that the virtual height of an oblique reflection at a given radio frequency $f_{obl}$ is equal to the virtual height corresponding to a vertical reflection of a frequency $f_v$. Combining this two theorems, the *Secant law* is obtained

$$f_{obl} = f_v \cdot \sec\varphi_0 \quad . \tag{1.12}$$

In the simplified geometry shown in the left panel of Fig. 1.8, from simple geometrical considerations one gets $\sec\varphi_0 = \sqrt{1+\left(d/2h_v'\right)^2}$, where $d$=TR and $h_v'$=AC. Considering a spherical ionosphere over a spherical Earth's surface (right panel of Fig. 1.8), from geometrical considerations one obtains

$$\varphi_0 = \arctan\left[\frac{\sin\left(\frac{\theta}{2}\right)}{1+\frac{h_v'}{R_E}-\cos\left(\frac{\theta}{2}\right)}\right], \tag{1.13}$$

where $R_E$ is the Earth's radius, and $\theta = D/R_E$ (see right panel of Fig. 1.8), *i.e.* the angle underlying T and R at the center of the Earth.

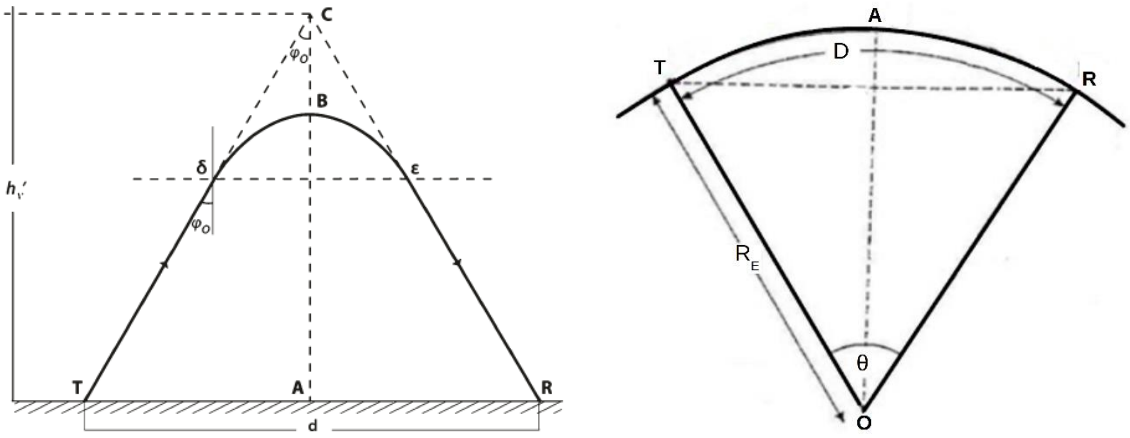


**Figure 1.8:** (Left panel) Simplified model of a flat ionosphere over a flat Earth, for an oblique incidence ionospheric sounding (from Zolesi and Cander 2014). (Right panel) Angle at the center of the Earth $\theta$ underlying the transmitter T and the receiver R, in a spherical Earth centered coordinate system (adapted from Zolesi and Cander 2014).

Considering an oblique radio linkage occurring with only one reflection in the ionosphere (corresponding to $f_v$ measured at the mid-point A between T and R), the *Maximum Usable Frequency* (*MUF*) to establish such linkage can be defined applying the Eq. (1.12) as

$$MUF(d)\text{F2} = \max\left(f_v \cdot \sec\varphi_0\right) \ . \tag{1.14}$$

Because the term $\sec\varphi_0$ depends on the geometry of the link and consequently on the distance *d* between the transmitter and the receiver, different values of the *MUF* are obtained for different *d* values.

The *MUF* for a given distance *d* is dependent only on the vertical frequency reflected at the mid-point, $f_v$, and on the geometry of the radio link through $\sec\varphi_0$. Thanks to that, the *MUF*(*d*)F2 can be calculated directly from an ionogram obtained after performing a vertical incidence sounding once the distance *d* is chosen. In particular, the *MUF* for a distance *d*=3000 km is the ionospheric characteristic scaled from an ionogram, which is then called *MUF*(3000)F2. Specifically, from Eqs. (1.12) and (1.13) for a given oblique frequency $f_{obl}$ it is possible to obtain a curve as a function of $h'$ and $f_v$ (for *d*=3000km) known as the *transmission curve* (Smith 1939). The intersections of the transmission curve with the curve $h' = f(f_v)$, i.e. the ordinary trace on the ionogram, give graphical solutions. The frequency $f_{obl}$ for which the transmission curve is tangential to the part of the ionogram trace referring to the F2 layer, is the *MUF*(3000)F2 defined by Eq. (1.14). For this point only one propagation path is possible, the one occurring when the product between $f_v$ and $\sec\varphi_0$ is maximum, with only one reflection in the ionosphere.

The *MUF*(3000)F2 is a characteristic routinely scaled from an ionogram, from which the *M*(3000)F2 characteristic is obtained as

$$M(3000)\mathrm{F2} = \frac{MUF(3000)\mathrm{F2}}{fo\mathrm{F2}} \quad . \tag{1.15}$$

### 1.3.2 Ground-based GNSS receivers

Ground-based Global Navigational Satellite Systems (GNSS) receivers are instruments able to detect signals emitted from GPS, GLONASS, Galileo, and COMPASS satellites orbiting the Earth at about 20,000 km of height. GPS satellites emit two L-band frequencies, L1=1.575 GHz and L2=1.228 GHz, which propagating through the ionosphere suffer a delay $d_\mathrm{I}$ in the signal carrier phase which is inversely proportional to the square of the frequency $f$ of the emitted signal and directly proportional to the integral of the electron density $N_\mathrm{e}$ along the ray path from the GPS satellite to the ground-based GNSS receiver

$$d_\mathrm{I} = \frac{K}{f^2} \int_{rec}^{sat} N_\mathrm{e} ds \ , \tag{1.16}$$

where $K$=40.3 $\mathrm{m^3 s^{-2}}$ (Davies 1990).

The integral of the electron density along the ray path is called *slant total electron content* (sTEC)

$$\mathrm{sTEC} = \int_{rec}^{sat} N_\mathrm{e} ds \ . \tag{1.17}$$

The adjective *slant* is due to the fact that, usually, the ray path signal between the transmitter on the GPS satellite and the ground-based GNSS receiver is slanted rather than vertical (see Fig. 1.9).
The sTEC can be converted to the equivalent *vertical total electron content* (vTEC) by using different model assumptions about the horizontal structure of the ionosphere. Usually, the vTEC at the *Ionospheric Pierce Point* (IPP) altitude of 350-400 km is calculated assuming the thin spherical shell model (Ciraolo and Spalla 1997). In this assumption the ionosphere is considered as compressed in a thin spherical shell located at a fixed height (more or less coincident with the median *hm*F2 value), so that the whole phase delay of the GPS signal is experimented in this small region. Hence, looking at Fig. 1.9, one can assume $\mathrm{d}s = \sec\chi \mathrm{d}h$ (with $h$ being the vertical coordinate) and

$$\mathrm{vTEC} = \sec\chi \cdot \mathrm{sTEC} \ . \tag{1.18}$$

The sTEC represents the number of free electrons in a column of unit cross section taken along the propagation path from the receiver to the satellite; for vTEC the column is vertical and placed at the IPP coordinate location. For convenience, the units used for both quantities is $10^{16}$ electrons/$\mathrm{m^2}$, called total electron content unit (TECU).

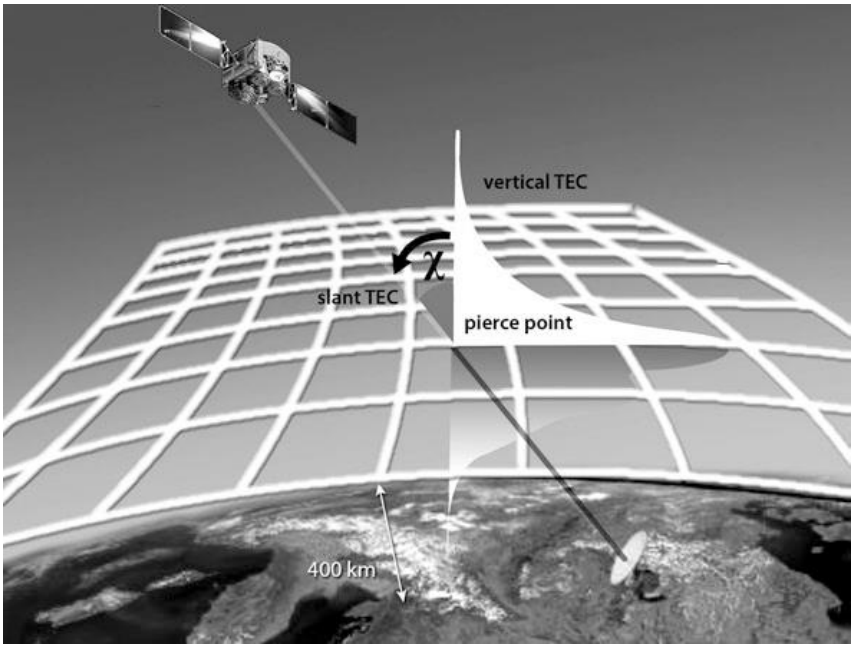


**Figure 1.9:** Geometry of the signal ray path between the GPS satellite and the ground-based GNSS receiver (from Zolesi and Cander 2014).

The vTEC provides an overall description of the ionization in the ionosphere-plasmasphere system. Therefore, it is one of the most important ionospheric parameters not only for trans-ionospheric communications, navigation, and solar-terrestrial physics, but also for a number of modern technological applications.

L1 and L2 code and carrier phase data recorded by GNSS receivers are stored in daily Receiver INdependent EXchange (RINEX) formatted files. The recording rate is usually set to 30 seconds.

A reliable estimation of vTEC is of particular importance because the ionosphere propagation delay is the largest source of error for the single GNSS frequency, and the second for differential GPS applications. The range error caused by 1 TECU at L1 frequency is about 0.162 m. With tens of TEC units along a GNSS signal path through the ionosphere during different solar-terrestrial conditions, ionosphere delays can account for positioning errors of the order of tens of meters.

### 1.3.3 Langmuir probes on-board satellites

Langmuir probes are instruments widely used on board satellites or rockets, which are able to get information on electronic and ionic densities and temperatures of the medium in which they are immersed. Being the ionosphere a conductive medium, the Langmuir probe draws from it an electric current of electrons or ions depending on the sign and magnitude of the potential applied to it.

The current-voltage relation for a plane Langmuir probe (i.e., a Langmuir probe where the potential is applied to a conducting plate) immersed in a plasma, is a characteristic curve with a shape like the one shown in Fig. 1.10.

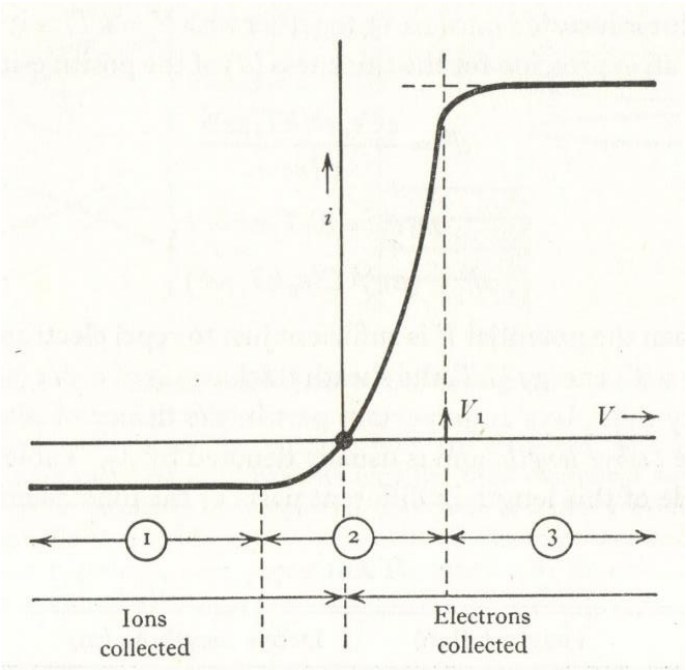


**Figure 1.10:** Current-voltage characteristic curve for a planar Langmuir probe (from Ratcliffe 1972).

In the current-voltage curve shown in Fig. 1.10, $i$ is the current flowing from the probe immersed in the plasma, which depends on the potential $V$ applied to it. $V_1$ corresponds to the *space potential* at which electrons and ions of all velocities reach the probe.

In the Langmuir probe current-voltage curve three different situations can occur:

1) The part marked as (1) in Fig. 1.10 corresponds to a situation where all incident positive ions reach the probe but all the electrons are repelled, thus only ions are collected;
2) The part marked as (2) in Fig. 1.10 corresponds to a situation where the electron current is carried only by the more energetic electrons in the Maxwellian distribution. Because the total electron current is much greater than the total ion current, the shape of this part is largely determined by the electron current;
3) The part marked as (3) in Fig. 1.10 is the situation complementary to (1), where no ions, but all the incident electrons, are collected.

Considering electrons following the Maxwellian distribution, if $u$ is the component of the velocity perpendicular to the plate of the Langmuir probe, it can be demonstrated (Ratcliffe 1972) that the number of electrons with velocities between $u$ and $u+du$ impacting the plate is proportional to $\exp\left(-mu^2 / 2kT\right)\mathrm{d}u$, where $m$ is the electron mass, $k$ the Boltzmann's constant, and $T$ the electron temperature. If a retarding potential $V$ is applied to the probe, each electron with $u > \sqrt{2Ve/m}$ can reach it, so that the number of electrons arriving in unit time, or the current $i$, is $i = i_0 \exp\left(-eV / kT\right)$, where $i_0$ is the current when the retarding potential is zero, i.e. when the plate is at the space potential $V_1$ (this is the part of the curve marked as (2) in Fig. 1.10).

From the shape of the current-voltage characteristic curve it is possible to determine the electron temperature $T$ and, when this is known, the $i_0$ value can be calculated. Knowledge of the probe's plate area $A$ then allows to calculate the density current $j_0 = \frac{i_0}{A} = \frac{1}{4} N_\mathrm{e} e u = \frac{1}{4} N_\mathrm{e} e \sqrt{\frac{8kT}{\pi m}}$, from which the electron density $N_\mathrm{e}$ can be obtained.

### 1.3.4 Radio Occultation technique

The Radio Occultation (RO) technique is a very powerful tool for remote sensing of the Earth's neutral atmosphere and ionosphere based on the GPS limb sounding technique (Kursinski et al. 1997; Hajj and Romans 1998; Jakowski et al. 2002). Low Earth Orbit (LEO) satellites which carry a dual frequency GPS receiver onboard offer the possibility to retrieve the vertical refractivity profile from the LEO satellite orbit height down to the Earth's surface. Since the index of refraction of the ionosphere depends mainly on the number of free electrons, the inversion of measured signals can provide the vertical electron density profile.

When a signal transmitted by a GPS satellite and received by a LEO satellite passes through the Earth's atmosphere in a limb sounding geometry (see Fig. 1.11), its phase and amplitude are affected in ways that are characteristic of the index of refraction of the medium along the path of propagation. Notably, it is the phase change of the signal transmitted by a GPS satellite, passing through the ionospheric medium, the basic observable in the RO technique. In fact, a GPS signal traveling in the atmospheric medium (where the refraction index can be different from the unity, particularly in the ionosphere) undergoes a phase delay which can be calculated (see Eq. (1.16)).

Thanks to the frequency dispersion experimented by L1 and L2 signals, calibration techniques allow to separate the contribution to the GPS signal phase delay due to the atmosphere from the ones caused for instance by clocks, receiver and transmitter biases. Thus, the differential GPS phases provide the TEC along the ray path through a spherically layered voxel structure (Fig. 1.11). The measured line integral TEC is the sum of the product $N_e ds_i$ where $N_e$ is the mean electron density in the voxel $i$ and $ds_i$ corresponds to the ray path length in the voxel $i$ at measurement $j$. Hence, the ray path elements can be computed easily according to the satellite geometry defined by the positions of the transmitting GPS satellite and the LEO satellite where the signal is received. The electron density of different shells can successively be derived from a series of 1s sampled measurements $j$ when the tangential point of occultation rays comes closer and closer to the Earth down to the bottom of the ionosphere (see Fig. 1.11). These are the main steps of the usually called *Abel inversion technique*.

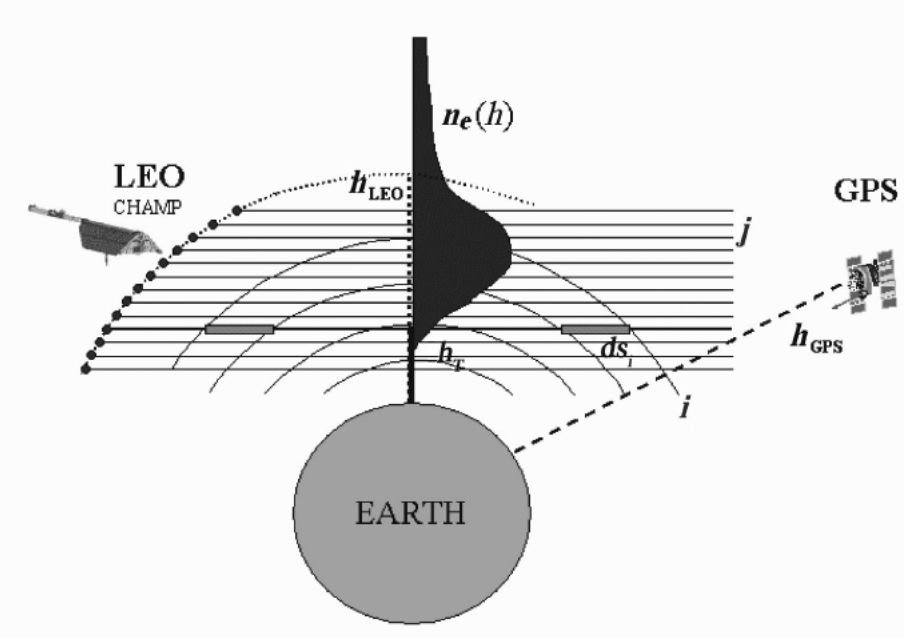


**Figure 1.11:** Illustration of the retrieval technique used in the RO technique (from Jakowski et al. 2004).

The Abel inversion makes use of the spherical geometry assumption (i.e., the signal velocity depends only on the distance $r$ from the centre of the Earth). Making this assumption, one neglects any effect caused by horizontal gradients characterizing an ionospheric shell at a certain height. This assumption can be very weak in certain regions like the low-latitudes (where the equatorial ionization anomaly sets up very large horizontal electron density gradients), for certain hours of the day like dawn and dusk ones (when the solar terminator path produces

localized horizontal electron density gradients), and during magnetically disturbed periods. Furthermore, what is calculated is a horizontally averaged representation of the electron density profile. Recently, some attempts have been made to include the horizontal gradients information in the RO inversion method (for example, see Garcia Fernandez et al. 2003) to improve the retrieval of the corresponding electron density profile.

## 2. Modeling the ionosphere: the IRI empirical model

A comprehensive knowledge of the ionosphere structure and dynamics is necessary to guarantee the effective operation, planning, and management of numerous radio-communication, telecommunication, and navigation systems, as well as for the improvement of scientific knowledge.

A three-dimensional modeling of the ionospheric electron density over the whole ionospheric altitude range for all geographic positions, time spans, and solar-terrestrial conditions is an essential goal of ionospheric physics and modeling.

This is why, many models of the ionosphere have been developed over the years, which can be divided into four main groups:

1. Theoretical or physical models. Such models give a description of the main ionospheric parameters and the electron density distribution starting from first principles arising from physical and chemical processes that control the behavior of the ionospheric plasma. They attempt to solve numerically the differential equations (continuity, momentum, and energy equations) for the ionized species of the ionospheric plasma. These differential equations are time-dependent, coupled, and nonlinear, making their solution a difficult numerical task. These models require magnetospheric and atmospheric input parameters and their accuracy depends on the quality of the input data.

   Main theoretical models are:

   - Time Dependent Ionospheric Model (TDIM, Schunk 1988) developed at Utah State University, USA;
   - Thermosphere-Ionosphere-Electrodynamics General Circulation Model (TIEGCM, Roble et al. 1988) developed at the National Center for Atmospheric Research, USA;
   - Coupled Thermosphere-Ionosphere Model (CTIM, Fuller-Rowell and Rees 1980) developed at University College London and Sheffield University, UK;
   - Global Theoretical Ionospheric Model (GTIM, Anderson et al. 1996) developed at Phillips Laboratory, USA.

2. Parameterised or semi-empirical models. Such models are based on a parameterization of theoretical models by expressing them in terms of solar-terrestrial-ionospheric parameters, giving a realistic representation of the ionospheric spatial and temporal structure using a limited number of numerical coefficients.

   At present, there are a few parameterized versions of theoretical models available for different ionospheric regions:

   - Parameterised Ionospheric Model (PIM, Daniell et al. 1995) obtained by combining GTIM and TDIM theoretical models;
   - Semi-Empirical Low Latitude Ionospheric Model (SLIM, Anderson et al., 1987).

3. Empirical models. These models are established by statistical analysis of long records of measured data. They are climatological models, representing monthly or seasonal average conditions; hence, ionospheric parameters measured at a specific location and time can deviate considerably from values given as output by these models. Empirical models have the advantage of representing the ionosphere based on actual measurements, not depending on the assumptions of an evolving theory. Their major disadvantage is that they are biased towards times and locations that are well represented in the underlying data record.

   The main exponents of this family are:

- International Reference Ionosphere model (IRI, Bilitza et al. 2017), which will be extensively described in this chapter;
- NeQuick model (Nava et al. 2008) developed at the Aeronomy and Radiopropagation Laboratory of the Abdus Salam International Centre for Theoretical Physics (ICTP), Italy.

4. Assimilation models. These are data-driven models that apply data assimilation techniques to improve the ionospheric plasma representation made by an underlying background model, which can be either theoretical or empirical. Currently, the most successful assimilation model based on the IRI empirical model is the IRI Real-Time Assimilative Model (IRTAM, Galkin et al. 2012). Some others assimilative techniques employed in the past for updating the empirical IRI model, used as background, are described in section 3.1.2, being this subject the main scope of the present work.

   For what concerns assimilation models based on theoretical models as background, noteworthy are:

   - Global Assimilation of Ionospheric Measurements model (GAIM, Schunk et al. 2004) developed by the Utah State University, USA;
   - Electron Density Assimilative Model (EDAM, Angling and Cannon 2004) developed by QinetiQ in the UK.

Ionospheric models listed above are only a selection of the most popular ones.

Among these models, the IRI is the one recognized by the International Standardization Organization (ISO) as the recommended for technical specification of ionospheric parameters (ISO 2009). For this reason, and for other reasons linked to the free availability of the IRI Fortran code and to the numerical and mathematical description of the ionosphere made by it, the IRI model is particularly suited to be employed as background model in a data assimilation scheme.

Therefore, the IRI model is extensively described in this chapter for what concerns the electron density profile modeling (the modeling of the ionic constituents and of electronic and ionic temperatures will be not considered in the present work). The attention is focused on those IRI features which are fundamental for the development of the IRI UP method (chapters 3, 4, and 5) and of the topside modeling of the ionosphere (chapter 6) which are proposed in this thesis work.

## 2.1 The IRI model: history and perspectives

The International Reference Ionosphere (IRI) is a joint project of the COmmittee on SPAce Research (COSPAR) and the International Union of Radio Science (URSI) initiated in 1968 with the goal of developing and improving an international standard for the specification of plasma parameters in the Earth's ionosphere.

Its goals are well described in its Terms of Reference as follows:

"*The Task Group was established to develop and improve a standard model of the ionospheric plasma parameters. The model should be primarily based on experimental evidence using all available ground and space data sources; theoretical considerations can be helpful in bridging data gaps and for internal consistency checks. Where discrepancies exist between different data sources the IRI team should promote critical discussion to establish the reliability of the different databases. IRI should be updated as new data become available and as old data sources are fully evaluated and exploited. IRI is a joint working group of COSPAR and URSI. COSPAR's prime interest is in a general description of the ionosphere as part of the terrestrial environment for the evaluation of environmental effects on spacecraft and experiments in space. URSI's prime interest is in the electron density part of IRI for defining the background ionosphere for radiowave propagation studies and applications.*"

As requested by COSPAR and URSI, IRI was built as an empirical model representing the synthesis of most of the available ground and space measurements of ionospheric characteristics. IRI represents monthly averages of electron and ion densities and temperatures in the altitude range of 50 km–2000 km. It also provides the vertical total electron content (vTEC) from the lower boundary (60–80 km) to a user-specified upper boundary (within 2000 km of height). As an empirical model, IRI has the advantage that it does not depend on the evolving theoretical understanding of the processes that shape the ionospheric plasma.

## 2.2 IRI modeling of the bottomside ionosphere

The IRI models the ionospheric vertical electron density profile dividing it into six subregions as shown in Fig. 2.1, including the topside (the region highlighted by the number (1) in Fig. 2.1), the F2 bottomside (2), the F1-layer (3), the intermediate region (4), the E-valley (5), and the E-bottomside and D-region (6). The boundaries of these subsections are marked by several characteristics profile points including the F2, F1, and E-layer peaks. Among these, the F2- and E-layer peaks (with their associated electron density values, *Nm*F2 and *Nm*E) are of particular importance because they are anchor points for the whole vertical electron density profile; thus, values of electron density are normalized to *Nm*F2 and *Nm*E peak values. The F1-layer peak point, although it is not an anchor point, is however, when present, an important point to model the F1-ledge. These three points (*Nm*F2, *Nm*F1, and *Nm*E) are marked with red circles in Fig. 2.1, at the corresponding real height *hm*F2, *hm*F1, and *hm*E, respectively.

The full vertical electron density profile is described by a set of mathematical expressions, each valid in a certain range (see Fig. 2.1), which allows a correct representation of the most important ionospheric characteristics, like the peak density and the height of the main ionospheric layers. The mathematical formulation of each sub-region is described in detail in the following sections.

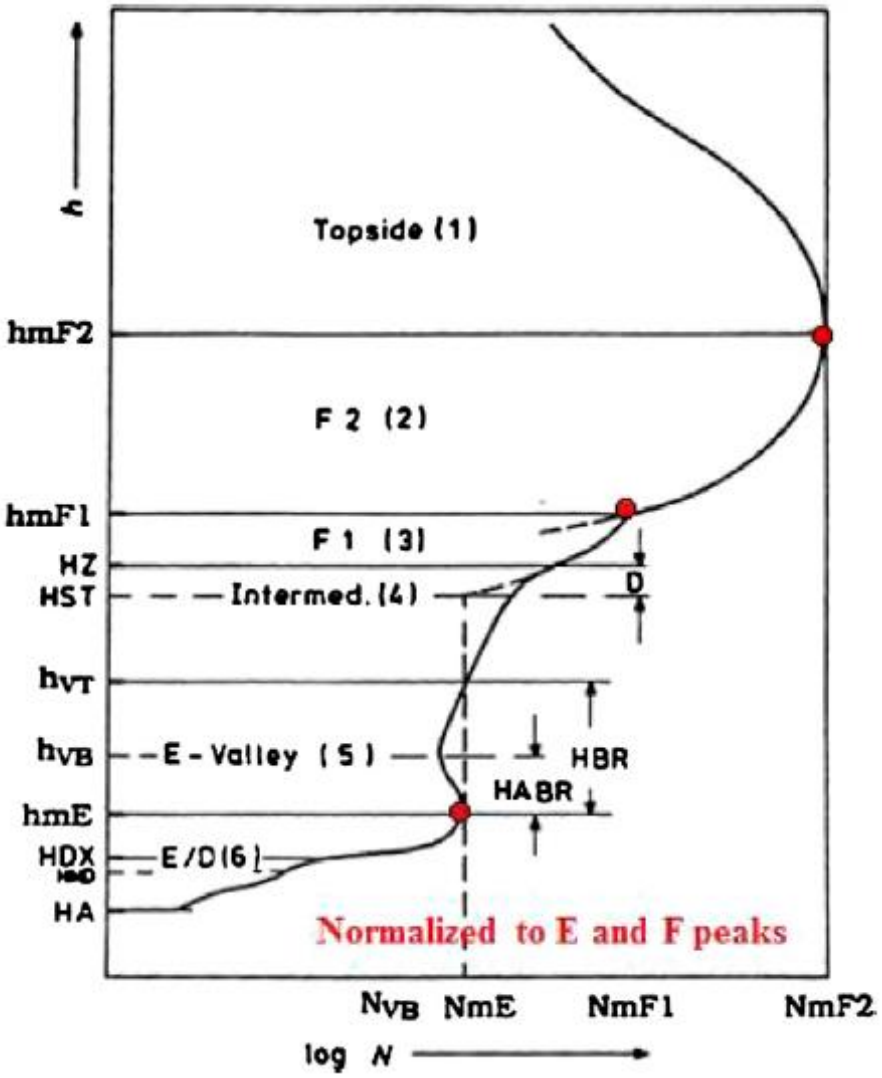


**Figure 2.1:** Schematic representation of the whole IRI vertical electron density profile and its different sub-regions (from Bilitza et al. 2014).

### 2.2.1 IRI's *fo*F2 and *M*(3000)F2 description

IRI modeling of F2 layer characteristics, *fo*F2 (hence *Nm*F2) and *M*(3000)F2, relies on CCIR and URSI long-term global maps; users can choose between these two options, or give their own values of *fo*F2 or *M*(3000)F2. CCIR maps are based on a procedure of numerical mapping pioneered by Jones and Gallet (1962) and Jones et al. (1969). URSI maps make use of the same CCIR mapping procedure, including a number of synthesized data points based on the thermospheric wind theory, as explained by Rush et al. (1989). URSI maps try to overcome the limitations of CCIR maps due to the geographical coverage of assimilated measurements, mainly located in the Northern hemisphere, and without coverage of ocean areas.

In either mapping procedures diurnal/spherical harmonic expansions are used to represent the diurnal cycle of *fo*F2 and *M*(3000)F2 and the corresponding spatial variation. The diurnal variation is based on a Fourier analysis of the hourly monthly median *fo*F2 and *M*(3000)F2 values as observed by a worldwide network of ionosondes used as database (for solar minimum years 1954, 1955 and 1964, and solar maximum years 1956-1958).

The general form of the numerical map function $\Omega(\lambda,\theta,T)$ ($\Omega$ being either *fo*F2 or *M*(3000)F2, $\lambda \in [-90,90]$ is the geographic latitude, $\theta \in [0,360)$ is the East geographic longitude, and $T \in [-180,180)$ is the Universal Time (UT) expressed as an angle) is (ITU-R 2009):

$$\Omega(\lambda,\theta,T) = a_0(\lambda,\theta) + \sum_{j=1}^{H}\left[a_j(\lambda,\theta)\cos jT + b_j(\lambda,\theta)\sin jT\right] \ , \tag{2.1}$$

where $H$ is the maximum number of harmonics used to represent the diurnal variation. In particular, $H = 6$ for *fo*F2 (thus 13 coefficients are considered in the expansion (2.1) for *fo*F2), and $H = 4$ for *M*(3000)F2 (thus 9 coefficients are considered in the expansion (2.1) for *M*(3000)F2).
The diurnal expansion Fourier coefficients $a_0, a_j, b_j$ vary with geographic coordinates and this variation is captured in the spatial spherical expansion over the geographic latitude and longitude:

$$a_0(\lambda,\theta) = \sum_{k=0}^{K} U_{0,k} G_k(\lambda,\theta) \ , \tag{2.2a}$$

$$a_j(\lambda,\theta) = \sum_{k=0}^{K} U_{2j,k} G_k(\lambda,\theta) \qquad j = 1,2,\ldots,H, \tag{2.2b}$$

$$b_j(\lambda,\theta) = \sum_{k=0}^{K} U_{2j-1,k} G_k(\lambda,\theta) \qquad j = 1,2,\ldots,H, \tag{2.2c}$$

where $K$ is the highest harmonic order used in the spherical expansion that determines the number of coefficients used in longitude ($K$=76 for *fo*F2, while $K$=49 for *M*(3000)F2).
The $G_k(\lambda,\theta)$ terms contain the 2-D "geographical" expansion basis functions, optimized for the representation of *fo*F2 by using power sine functions of modip (modified dip latitude), power cosine of latitude, and longitudinal harmonics (for more information see ITU-R 2009). In fact, $G_k(\lambda,\theta)$ terms are only similar in their formulation to the classic surface spherical harmonics that arise naturally in physical problems with spherical symmetry. In particular, the zonal dimension of $G_k(\lambda,\theta)$ basis functions is actually expressed as a conventional harmonic expansion over geographic longitude $\theta$; instead, the meridional dimension of the spherical expansion does not use the traditional Legendre definition, but rather a custom power function that captures, in addition to the solar-controlled geographic latitude dependence on $\lambda$, the meridional dependence of *fo*F2 on the geometry of the Earth's magnetic field:

$$P_{mn}(\lambda,\theta) = \left[\sin\chi(\lambda,\theta)\right]^m (\cos\lambda)^n \ ; \tag{2.3}$$

$\chi(\lambda,\theta)$ is the modified dip latitude introduced by Rawer for an optimal global F peak mapping:

$$\chi(\lambda,\theta) = \arctan\frac{I(\lambda,\theta)}{\sqrt{\cos\theta}} \ , \tag{2.4}$$

where $I(\lambda,\theta)$ is the true magnetic dip angle at a height of 350 km over location $(\lambda,\theta)$.

The $\left[\sin\chi(\lambda,\theta)\right]^m$ term in Eq. (2.3) is frequently called the "geomagnetic correction", and is responsible for placing the Appleton anomaly of ionization around the geomagnetic equator, rather than the geographic equator (Galkin et al. 2012). The full mathematical description of $G_k(\lambda,\theta)$ spatial functions can be found in ITU-R (2009).

The maximum order $K$ (in Eqs. (2.2)) is selected to resolve the smallest spatial scale size of the ionospheric climate variability, which detail will be at times insufficient for representing the real-time ionosphere. Increasing the order $K$ for nowcasting applications as to provide a finer detail of the spatial resolution, however, is limited by several practical considerations, including instrument data noise. As Bradley et al. (2009) argued, no advantage would have been gained by considering a higher order leading to a smaller scale size because the order of the expansion is strictly linked to the spatial instrument coverage. CCIR maps are recommended for mainly land regions, whilst URSI maps when the mapping area includes large ocean areas. In our studies CCIR maps are considered.

Thus IRI needs 13 diurnal × 76 spherical = 988 coefficients to globally represent the *fo*F2 (9 diurnal × 49 spherical = 441 coefficients for *M*(3000)F2) diurnal variation for a definite month. The seasonal variation of F2 characteristics is taken into account calculating these coefficients for every month of the year.

The solar activity dependence is modeled exploiting the close correlation between F2 characteristics and solar/ionospheric activity indices. In particular, *fo*F2 is modeled as a function of the $IG_{12}$ ionospheric activity index, while *M*(3000)F2 as a function of the $R_{12}$ solar activity index. $R_{12}$ is the twelve-months running mean of the Zurich sunspot number *R*. Like other solar activity indices, *R* index is based on solar observations of phenomena occurring at the solar surface or atmosphere, and is used simply as proxy of ionospheric phenomena. $IG_{12}$ is the twelve-months running mean of the ionospheric activity index $IG$ (Liu et al. 1983). It is calculated from linear regression equations based on CCIR modeled monthly median *fo*F2 values (dependent on $R_{12}$) and monthly medians *fo*F2 values recorded at noon at 13 ionospheric stations (Liu et al. 1983). The number of ionospheric stations used in the $IG_{12}$ index calculation has changed over time. Currently, only four ionospheric stations are used for this purpose (Brown et al. 2018). The resulting index is the median of $IG$ index values calculated in those selected stations, and the corresponding $IG_{12}$ is used in CCIR and URSI to model *fo*F2. Indices like $IG_{12}$, despite being calculated by means of a solar activity proxy, include also the variability caused by other phenomena affecting the ionosphere, like the geomagnetic activity, the neutral composition change or dynamical disturbances, because *fo*F2 values, both measured and modeled, have these footprints. Thus, an index calculated merging available information deduced from measured and modeled ionospheric characteristics can tell us much more on ionospheric conditions than a direct solar activity index.

Coefficients of the above exposed spatial and temporal expansions (Eqs. (2.2)) were calculated for two selected levels of solar activity ($IG_{12} = 0$ and $IG_{12} = 100$ for *fo*F2, $R_{12} = 0$ and $R_{12} = 100$ for *M*(3000)F2); thus, the total number of stored coefficients that are used by the IRI model to obtain *fo*F2 for any given time and location is 988×12 months × 2 levels of solar activity = 23,712 coefficients (441×12 months × 2 levels of solar activity = 10,584 coefficients for *M*(3000)F2).

IRI model assumes that the coefficients of the spatial and temporal expansions (Eqs. (2.2)) are linearly correlated with the solar activity proxies $IG_{12}$ (for *fo*F2) and $R_{12}$ (for *M*(3000)F2). This leads also to a linear dependence between the solar activity proxies and the modeled ionospheric characteristics; in particular, direct proportionality between $IG_{12}$ and *fo*F2, and reverse proportionality between $R_{12}$ and *M*(3000)F2), until a threshold value set to 150 (for both indices) beyond which both *fo*F2 and *M*(3000)F2 are kept constants, to account for the well-known ionospheric saturation effect. These coefficients are described by the following formulas (Komjathy and Bradley 1996):

$$U_{2j,k} = \left[ U^{\text{low}}_{2j,k} \left( 1 - \frac{IG_{12}}{100} \right) + U^{\text{high}}_{2j,k} \frac{IG_{12}}{100} \right], \tag{2.5a}$$

$$U_{2j-1,k} = \left[ U^{\text{low}}_{2j-1,k} \left( 1 - \frac{IG_{12}}{100} \right) + U^{\text{high}}_{2j-1,k} \frac{IG_{12}}{100} \right], \tag{2.5b}$$

for *fo*F2, and

$$U_{2j,k} = \left[ U^{\text{low}}_{2j,k} \left( 1 - \frac{R_{12}}{100} \right) + U^{\text{high}}_{2j,k} \frac{R_{12}}{100} \right], \tag{2.6a}$$

$$U_{2j-1,k} = \left[ U^{\text{low}}_{2j-1,k} \left( 1 - \frac{R_{12}}{100} \right) + U^{\text{high}}_{2j-1,k} \frac{R_{12}}{100} \right], \tag{2.6b}$$

for *M*(3000)F2, with the specification that $U_{2j,k}(IG_{12} \text{ or } R_{12} > 150) = U_{2j,k}(IG_{12} \text{ or } R_{12} = 150)$ and $U_{2j-1,k}(IG_{12} \text{ or } R_{12} > 150) = U_{2j-1,k}(IG_{12} \text{ or } R_{12} = 150)$.

In Eqs. (2.5)-(2.6) $U_{2j,k}^{\text{low}}$ and $U_{2j,k}^{\text{high}}$ (similarly for $j$-1 terms) are the coefficient values calculated for $IG_{12}$ or $R_{12}$ equal to 0 (low solar activity) and  equal to 100 (high solar activity), respectively.

### 2.2.2 IRI's *hm*F2 description

Since the 2016 version, IRI includes three different options to calculate the height of the F2 layer peak. The first option was developed by Bilitza et al. (1979) based on the anti-correlation between *hm*F2 and the propagation factor *M*(3000)F2 (Shimazaki 1955), and a correction factor accounting for the E layer presence (Bradley and Dudeney 1973). The second and the third options, which have been recently introduced in IRI based on the works of Altadill et al. (2013) and Shubin et al. (2015), are both based on the spherical harmonic expansion of measured *hm*F2 values.

**Bilitza et al. (1979) *hm*F2 formulation**

In this formulation, IRI calculates *hm*F2 through the empirical formula (Bilitza et al. 1979):

$$hm\text{F2}(\text{km}) = \frac{1490}{\left[M(3000)\text{F2} + DM\right] - 176} \ , \tag{2.7}$$

where the correction factor *DM* is:

$$DM = \frac{f_1 f_2}{\left(\frac{fo\text{F2}}{fo\text{E}} - f_3\right) + f_4} \ , \tag{2.8}$$

and the solar activity functions are:

$$f_1 = 0.00232 R_{12} + 0.222 \ , \tag{2.9a}$$

$$f_2 = 1 - \frac{R_{12}}{150 \exp\left[-\left(\frac{\psi}{40}\right)^2\right]} \ , \tag{2.9b}$$

$$f_3 = 1.2 - 0.0116 \exp\left(\frac{R_{12}}{41.84}\right) \ , \tag{2.9c}$$

$$f_4 = 0.096 \frac{R_{12} - 25}{150} \ , \tag{2.9d}$$

where $\psi$ is the magnetic dip latitude.

*hm*F2 is then explicitly dependent on $R_{12}$, through $f_1$, $f_2$, $f_3$ and $f_4$ (Eqs. (2.9)), but, at the same time, is also implicitly dependent on $R_{12}$ through *M*(3000)F2 (Eq. (2.7)), as described in the previous section, and *fo*E (Eq. (2.8)) being *fo*E dependent on $R_{12}$ itself (Eqs. (2.23)-(2.25)). The presence of *fo*F2 in the *DM* factor (Eq. (2.8)) makes *hm*F2 implicitly dependent also on $IG_{12}$.

The Bilitza et al. (1979) equation to convert *M*(3000)F2 to *hm*F2 is true under the assumption that the ionospheric electron density height dependency is parabolic in the F2 layer region, though this is not always a good approximation (Shubin 2015).

Though this formulation models reasonably well the main trends of the quiet *hm*F2, it cannot capture the features of small spatial scales and of relative short time life (Adeniyi et al. 2003; Oyeyemi et al. 2007).

**Altadill et al. (2013) *hm*F2 formulation**

Altadill et al. (2013) developed a model of the F2 layer peak height *hm*F2 quiet time behavior based on a spherical harmonic analysis of *hm*F2 values recorded by 26 digisonde stations globally spread encompassing the time interval 1998-2006. In the spherical harmonic analysis the modip latitude (Eq. (2.4)) was used as the coordinate of the reference system (as in IRI's *fo*F2 and *M*(3000)F2 formulation). Model coefficients are bounded to the solar

activity (by using $R_{12}$ index), and the temporal and seasonal variations are considered by a Fourier expansion of the coefficients.
The spherical harmonic mathematical formulation employed by Altadill et al. (2013) is:

$$hm\mathrm{F2}(\vartheta,\varphi,t)=\sum_{n=0}^{N}\sum_{m=0}^{n}P_n^m(\cos\vartheta)\left\{g_n^m(t)\cos\left(m\varphi\right)+h_n^m(t)\sin\left(m\varphi\right)\right\}, \tag{2.10}$$

$$\begin{Bmatrix} g_n^m(t) \\ h_n^m(t) \end{Bmatrix}=\sum_{q=0}^{2}\begin{Bmatrix} ga_{n,q}^m \\ ha_{n,q}^m \end{Bmatrix}\cos\left(\frac{2\pi qt}{12}\right)+\begin{Bmatrix} gb_{n,q}^m \\ hb_{n,q}^m \end{Bmatrix}\sin\left(\frac{2\pi qt}{12}\right), \tag{2.11}$$

where $\vartheta$ and $\varphi$ are the geocentric spherical coordinates colatitude and longitude, respectively, and $t$ is the time; $P_n^m$ are the Legendre polynomials of degree $n$ (with a maximum degree $N=8$) and order $m$, and $g_n^m(t)$ and $h_n^m(t)$ are the Gauss coefficients which are functions of time through the coefficients $ga_{n,q}^m$, $gb_{n,q}^m$, $ha_{n,q}^m$, and $hb_{n,q}^m$.
The Altadill et al. (2013) *hm*F2 model was tested against the IRI prediction, made by Bilitza et al. (1979) *hm*F2 formulation, by comparing them with the experimental observations. This model of *hm*F2 improves the IRI prediction by 25% in average at low latitudes and by 10% in average at high latitudes in terms of RMSE (Root Mean Square Error), while no significant improvement is observed for mid-latitudes, resulting an average improvement of about 10% at global scale.

**Shubin et al. (2015) *hm*F2 formulation**
Similarly to the Altadill et al. (2013) *hm*F2 formulation, the Shubin et al. (2015) one employs the spherical harmonic expansion directly applied on measured *hm*F2 values. Differently from Altadill et al. (2013), Shubin et al. (2015), besides considering 62 digisonde stations (1987-2012), makes use also of Radio Occultation data from CHAMP (2001-2008), GRACE (2007-2011), and COSMIC (2006-2012) satellite missions.
As input parameters this model uses the year, month and UT (Universal Time) as well as the geographic coordinates and the $F_{10.7}$ solar index averaged over 3 Sun rotations ($F_{10.7A}$). The spatial decomposition (like Eqs. (2.10)-(2.11)) is based on 12 harmonics for the longitude and 8 ones for the modip latitude. For the diurnal variation three Fourier harmonics are used. The dependency of *hm*F2 on $F_{10.7A}$ index is assumed logarithmic.
Comparing his results to the Bilitza et al. (1979) IRI ones over 10 digisonde stations, Shubin et al. (2015) found that his model produces NRMSE (Normalized Root Mean Square Error) values that do not exceed 8%, while IRI ones are systematically higher (up to 12% in some cases). Shubin et al. (2015) model is able to reconstruct the original *hm*F2 dataset with RMSE=14.3 km and NRMSE= 3.7 % for low solar activity, and RMSE=23.4 km and NRMSE=5.4% for high solar activity.

### 2.2.3 The F2 bottomside region in IRI

The F2 bottomside region (region (2) in Fig. 2.1) is the part of the electron density profile connecting the F2 layer (*hm*F2) to the F1 layer (*hm*F1), when the latter is present. When the F1 layer is not present, it extends to the *HZ* height (Eq. (2.18)).

It was described by Ramakrishnan and Rawer (1972) as:

$$N(h)=Nm\mathrm{F2}\exp\left(-\frac{x^{B_1}}{\cosh x}\right), \tag{2.12}$$

with

$$x=\frac{hm\mathrm{F2}-h}{B_0}, \tag{2.13}$$

where the parameter $B_0$ represents the bottomside thickness parameter, and $B_1$ the bottomside shape parameter.

The current IRI 2016 version offers three different options for both $B_0$ and $B_1$.

In its first formulation (Ramakrishnan and Rawer 1972), the $B_1$ was kept fixed to 3, while the appropriate $B_0$ parameter was selected from a table (Bilitza 1990) where $B_0$ values were calculated for different seasons, local time hours (0 and 12 LT), modip latitudes (18 and 45 degrees), and solar activity levels ($R_{12}$=0 and 100), based on a large database of electron density profiles recorded by ionosondes. $B_0$ values for different local time hours, modip latitudes, and solar activity levels were obtained by interpolating betweem $B_0$ table's values, as described in Bilitza (1990). Such $B_0$ table values were updated by Bilitza et al. (2000) based on the analysis of a large volume of ionosonde data, and $B_0$ values calculated for the magnetic equator (modip=0°) were added. Furthermore, because a clear difference in shape between nighttime and day-time profiles in the bottomside F2 region emerged, two different $B_1$ values were introduced: $B_1 = 1.9$ for nighttime; $B_1 = 2.6$ for daytime. This reflects the difference in shape from the typical nighttime profile with a sharp drop into a deep nighttime valley compared to the gradual decrease during daytime down to first the F1 layer and then a shallow E-valley (Bilitza 2003).

A second option for $B_0$ and $B_1$ selection makes use of the Gulyaeva (1987) model for the half-density height, i.e., the height at which the F2 bottomside profile assumes half the value of the F2 maximum electron density (*Nm*F2). Adaption of the Gulyaeva (1987) model to the IRI model F2 bottomside description was accomplished by Bilitza and Rawer (1990). In such formulation the $B_1$ parameter can take the values (3, 3.5, 4, 4.5, 5), and different $B_0$ values were obtained for different $B_1$values as a function of the season and hour of the day.

The third option for $B_0$ and $B_1$ selection (the latest and the recommended one) is based on the work made by Altadill et al. (2009). In this model the spherical harmonics analysis was applied on $B_0$ and $B_1$ parameters deduced from ionograms measured by 27 globally distributed ionosonde stations for the years 1998-2006. The variations with modip latitudes, local time hours, month, and sunspot number were included in the model. With their model Altadill et al. (2009) found an improvement in the $B_0$ parameter modeling of up to 32% over the Bilitza et al. (2000) model (first option), and up to 40% over the Gulyaeva (1987) model (second option). An improvement of up to 20%, compared to the other options, was also found for the $B_1$ parameter.

### 2.2.4 The F1 layer region

The F1-peak critical frequency *fo*F1 in IRI is described by the Ducharme et al. (1971, 1973) model. Using a large amount of ionosonde data, *fo*F1was modeled as a function of the solar zenith angle ($\chi$), solar activity index $R_{12}$, and magnetic dip latitude ($\psi$):

$$foF1 = f_s \cos^n \chi \ , \tag{2.14}$$

where

$$f_s = \frac{f_0 \left( f_{100} - f_0 \right) R_{12}}{100} \ , \tag{2.15a}$$

$$f_0 = 4.35 + 0.058|\psi| - 0.00012\psi^2 \ , \tag{2.15b}$$

$$f_{100} = 5.348 + 0.011|\psi| - 0.00023\psi^2 , \tag{2.15c}$$

$$n = 0.093 + 0.0046|\psi| - 0.000054\psi^2 + 0.0003R_{12} . \tag{2.16c}$$

In this model an F1 layer is present only when the solar zenith angle is smaller than the critical value $\chi_s$:

$$\chi_s = \frac{\chi_0 \left( \chi_{100} - \chi_0 \right) R_{12}}{100} \ , \tag{2.17a}$$

where

$$\chi_0 = 49.85 + 0.35|\chi| \ , \tag{2.17b}$$

$$\chi_{100} = 38.96 + 0.51|\chi| . \tag{2.17c}$$

In accordance with the experimental evidence, IRI omits the F1 layer at night and during winter. Later, Scotto et al. (1997, 1998) introduced a statistical description of the occurrence statistics of the F1 layer in terms of the solar zenith angle, solar activity, and geomagnetic latitude. They found that the IRI model describes the measured *fo*F1 values quite well but underestimates the time span (diurnal and seasonal) for which the F1 layer is observed. Describing the F1 appareance with actual probabilities (percentage of days for which an F1 layer is expected) rather than with a cutoff criterion (like in Ducharme model, Eqs. (2.17)), Scotto et al. (1997, 1998) found a much better match with the data.

The F1 layer peak height *hm*F1 is found as the height at which the F2 bottomside profile (Eq. (2.12)) reaches the F1 peak density (*Nm*F1, obtained from (2.14)). Thus, it depends on the choice of the bottomside thickness parameter $B_0$.

The F1 region electron density profile (region (3) in Fig. 2.1) extends from *hm*F1 to *HZ*, where *HZ* is the upper boundary of the intermediate region defined by:

$$HZ = \frac{(hm\mathrm{F1} + HST)}{2} \ . \tag{2.18}$$

The height *HST* is found where the F1 region profile function reaches the *Nm*E value (derived from Eq. (2.23)).

The F1 region electron density profile is described by the F2 bottomside formulation (Eq. (2.12)) with a modified height named $h^*$:

$$h^* = hm\mathrm{F1} \cdot \left[ 1 - \left( \frac{hm\mathrm{F1} - h}{hm\mathrm{F1}} \right)^{1+D_1} \right]. \tag{2.19}$$

$D^1$ is the F1 layer shape factor and varies with the modified dip latitude and local time (see Bilitza et al. 2000).

When the F1 layer is not present $D^1$=0 hence $h^*$=*h* and the (2.12) profile extends also in the F1 region.

### 2.2.5 The intermediate region

The intermediate region (region (4) in Fig. 2.1) is the region merging the upper electron density profile (normalized to *Nm*F2) to the lower one (normalized to *Nm*E), extending from *HZ* to the E-valley top height $h_{\mathrm{VT}}$, which is the height for which the intermediate region profile reaches the *Nm*E value (derived from Eq. (2.23)).

The intermediate region profile is described by the F2 bottomside formulation (Eq. (2.12)) with a modified height named $h^{**}$:

$$h^{**} = hm\mathrm{F1} \cdot \left[ 1 - \left( \frac{hm\mathrm{F1} - \eta}{hm\mathrm{F1}} \right)^{1+D_1} \right], \tag{2.20}$$

with

$$\begin{cases} \eta = h \qquad \text{if} \quad HST = h_{\mathrm{VT}} \\ \eta = HZ + \dfrac{T}{2} - \sqrt{T\left[\dfrac{T}{4} - \left(h - HZ\right)\right]} \quad \text{if} \quad HST > h_{\mathrm{VT}} \\ \eta = HZ + \dfrac{T}{2} + \sqrt{T\left[\dfrac{T}{4} - \left(h - HZ\right)\right]} \quad \text{if} \quad HST < h_{\mathrm{VT}} \end{cases}, \tag{2.21}$$

and

$$T = \frac{\left(HZ - HST\right)^2}{HST - h_{\mathrm{VT}}}. \tag{2.22}$$

With this formulation, the same function Eq. (2.12) is used by IRI for modeling the F2, the F1, and the intermediate regions by using different height definitions (from $h$ in the F2 layer to $h^*$ in the F1 layer to $h^{**}$ in the intermediate layer). This approach has been developed to avoid discontinuities or artificial valleys when merging different regions.

### 2.2.6 The E layer region

The E-peak critical frequency *fo*E in IRI is described by the model developed by Kouris and Muggleton (1973a,b) according to the following relationship:

$$fo\mathrm{E}^4 = A \cdot B \cdot C \cdot D. \tag{2.23}$$

$$A = 1 + 0.0094\left(COV_{12} - 66\right), \tag{2.24a}$$

$$B = \cos^m \chi_{\mathrm{noon}}, \tag{2.24b}$$

$$m = \begin{cases} -1.93 + 1.92\cos\phi & \text{for } |\phi| < 32° \\ 0.11 - 0.49\cos\phi & \text{for } |\phi| \geq 32° \end{cases}, \tag{2.24c}$$

$$C = \begin{cases} 23 + 116\cos\phi & \text{for } |\phi| < 32° \\ 92 + 35\cos\phi & \text{for } |\phi| \geq 32° \end{cases}, \tag{2.24d}$$

$$D = \cos^n \chi_a, \tag{2.24e}$$

$$n = \begin{cases} 1.2 & \text{for } |\phi| > 12° \\ 1.31 & \text{for } |\phi| \leq 12° \end{cases}. \tag{2.24f}$$

The four factors $A, B, C, D$ in Eq. (2.23) depend on the solar activity through the 12-months running mean of the $F_{10.7}$ solar flux ($COV_{12}$), on the season through the solar zenith angle at noon ($\chi_{\mathrm{noon}}$), on the geodetic latitude ($\phi$), and on the modified solar zenith angle ( $\chi_a$, see Bilitza 1990). The use of $\chi_a$ instead of $\chi$ improves the description of the *fo*E nighttime variation.
IRI estimates $COV_{12}$ from $R_{12}$ values through the formula:

$$COV_{12} = 63.75 + R_{12}\left(0.728 + 0.00089R_{12}\right). \tag{2.25}$$

Thus, *fo*E values depend on $R_{12}$ values.
IRI assumes a constant E layer peak height $hm\mathrm{E} = 105\ km$ for all conditions.

The E valley, the region ((5) in Fig. 2.1) from *hm*E to $h_{VT}$, is described by a fifth order power function:

$$N(h) = NmE\left(1 + E_1 x^2 + E_2 x^3 + E_3 x^4 + E_4 x^5\right) , \tag{2.26}$$

with $x = h - hmE$ .
The parameters $E_1, E_2, E_3,$ and $E_4$ of Eq. (2.26) are obtained by fitting the above function to electron density profiles of the E-valley region derived from incoherent scatter radar measurements (Bilitza (1990)).

### 2.2.7 The D layer region

Most electron density profiles in the D region exhibit a characteristic inflection point at a height *hm*D and density *Nm*D. Based on the analisys of profiles derived from rocket measurements, Mechtley and Bilitza (1974) found the following empirical relationship:

$$NmD\ [10^8 m^{-3}] = \max\left[4 \cdot 10^8, \left(6.05 + 0.088 R_{12}\right) \exp\left(-\frac{0.1}{\cos^{2.7}\chi}\right)\right] . \tag{2.27}$$

It is seen that the *Nm*D value is greater or equal to a minimum (nighttime) value of $4 \cdot 10^8 el/m^3$.
The density profile starting at height *HA* is represented by a third order polynomial:

$$N(h) = NmD \exp\left(F_1 x + F_2 x^2 + F_3 x^3\right) , \tag{2.28}$$

with $x = h - hmD$ , and *F* parameter values specified in Bilitza (1990).

The E region merges with the underlying D region by means of the formula (region (6) in Fig. 2.1):

$$N(h) = NmE \cdot \exp\left[-D_1 \left(hmE - h\right)^K\right] , \tag{2.29a}$$

where the parameters $D_1$ and *K* are determined such that the function agrees with the D-region profile function (Eq. (2.28)) and its first derivative at the height *HDX* of Fig. 2.1:

$$K = -\frac{DN(hmE - HDX)}{NDX \ln\left(\frac{NDX}{NmE}\right)} , \tag{2.29b}$$

$$D_1 = \frac{DN}{NDX \cdot K\left(hmE - HDX\right)^{k-1}} , \tag{2.29c}$$

where *NDX* is the electron density at *HDX* and *DN* is the electron density derivative at that height.

## 2.3 IRI modeling of the topside ionosphere

The IRI 2016 version offers three options to model the electron density in the topside ionosphere (region (1) in Fig. 2.1). These options are called: IRI-2001, IRI-2001 corrected, and IRI-NeQuick.

### 2.3.1 IRI-2001 topside option

For the IRI 1978 version Rawer (1984) developed an analytical description of the information contained in the Bent's parameter plots (Bent 1972). When examining the Bent model, they found that the deduced topside scale height exhibits a highly irregular behavior across Bent's class boundaries. Their analytical representation avoids these shortcomings and produces a more reasonable smoothly varying scale height.

When comparing IRI with a large database of TEC measurements McNamara (1984) found good agreement at mid-latitudes, but large discrepancies close to the magnetic equator. IRI predictions were almost a factor of two below the measurements during high solar activity. This was found to be partially a result of the limited latitudinal resolution of the original sampling by Bent (1972).

The IRI-2001 topside formulation is based on the approach to model the profile by using different sections with nearly constant gradients. The analytical representation of the electron density profile $N(h)$ as a function of the height (Bilitza 1990, 2001) is an exponential layer modulated by the Booker function $B(x)$ (Eq. (2.31a)). The Booker (1977) approach divides the profile into segments with a constant gradient and uses Epstein step-functions to represent the whole gradient profile analytically. Integrating the gradient profile the Booker function $B$ is obtained.

The electron density $N(h)$ related to the F2 peak density $Nm$F2 is then given by:

$$N(h) = Nm\mathrm{F2}\exp\left[-\frac{B(x)}{\alpha}\right], \tag{2.30}$$

where $Nm$F2 is the electron density of the F2 maximum and $B(x)$ is expressed by:

$$B(x) = \beta\eta E_1(x) + \xi\left[100E_2(x) - \alpha\left(h - hm\mathrm{F2}\right)\right], \tag{2.31a}$$

with

$$E_1(x) = \ln\left[\frac{1+\exp\left(\frac{x-394.5}{\beta}\right)}{1+\exp\left(\frac{x_0-394.5}{\beta}\right)}\right], \tag{2.31b}$$

$$E_2(x) = \ln\left[\frac{1+\exp\left(\frac{x-300}{100}\right)}{1+\exp\left(\frac{x_0-300}{100}\right)}\right], \tag{2.31c}$$

where $hm$F2 is the height of the F2 maximum. $\eta$ and $\xi$ are the gradients and $\beta$ the thickness of the transition region involved in the Booker's approach. $\beta$, $\eta$, and $\xi$ are computed through empirical relations involving the geomagnetic latitude, the F2 layer critical frequency $fo$F2 and the solar flux $F_{10.7}$.

$x$ is a modified altitude variable transforming the F2-peak near 300 km:

$$x - x_0 = \alpha\left(h - hm\mathrm{F2}\right), \tag{2.31d}$$

with the transformation factor $\alpha$

$$\alpha = \frac{700}{1000 - hm\mathrm{F2}} \ , \tag{2.31e}$$

and the new peak height $x_0$

$$x_0 = 300 - \delta \ . \tag{2.31f}$$

The corrective term $\delta$ is computed using $\beta$, $\eta$, and $\xi$.

The $\beta$, $\eta$, and $\xi$ parameters are function of the geomagnetic latitude $\phi$, the F2 layer critical frequency *fo*F2, and the monthly solar radio flux $F_{10.7}$:

$$\xi, \beta = t_0 + t_1 T(\phi) + t_2 R(F10.7) + t_3 T(\phi) R(F10.7) + t_4 \, fo\mathrm{F2} + t_5 \, fo\mathrm{F2} T(\phi) + t_6 \left( fo\mathrm{F2} \right)^2 \ , \tag{2.31g}$$

$$\eta = t_0 + t_1 T_{\mathrm{m}}(\phi) + t_2 R(F10.7) + t_3 T(\phi) R(F10.7) + t_4 \, fo\mathrm{F2} + t_5 \, fo\mathrm{F2} T(\phi) + t_6 \left( fo\mathrm{F2} \right)^2 , \tag{2.31h}$$

where,

$$T(\phi) = \cos^2 \phi \ , \tag{2.31i}$$

$$T_{\mathrm{m}}(\phi) = \frac{1}{\left[ 1 + \exp\left( -\frac{\phi}{15} \right) \right]^2} \ , \tag{2.31j}$$

$$R(F10.7) = \begin{cases} \dfrac{F10.7 - 40}{30} & \text{for} \quad \mathrm{F10.7<193} \\ 5.1 & \text{for} \quad \mathrm{F10.7} \geq 193 \end{cases} \ . \tag{2.31k}$$

The parameters $t_i$ of Eq. (2.31g-h) are obtained from a non-linear least-square fitting procedure approximating the Bent (1972) topside profiles and are listed in Table 2.1. This analytical representation helps to smooth out some of the unreasonable sharp transitions between different classes in the Bent model.

| | $\eta$ | $\xi$ | $\beta$ |
|---|---|---|---|
| $t_0$ | 0.058798 | 0.078922 | -128.03 |
| $t_1$ | -0.08 | -0.0046702 | 20.253 |
| $t_2$ | -0.014065 | -0.019132 | -8.0755 |
| $t_3$ | 0.0069724 | 0.0076545 | -0.65896 |
| $t_4$ | 0.0024287 | 0.0032513 | 0.44041 |
| $t_5$ | 0.004281 | 0.006029 | 0.71458 |
| $t_6$ | -0.0001528 | -0.00020872 | -0.042966 |

**Table 2.1:** IRI-2001 topside profile parameters.

Minimizing for the corrective term $\delta$ one finds:

$$\delta = \frac{\dfrac{\eta}{1+Z} - \dfrac{\xi}{2}}{\dfrac{\eta}{\beta} \dfrac{Z}{\left(1+Z\right)^2} + \dfrac{\xi}{400}} \ , \tag{2.31l}$$

$$Z = \exp\left(\frac{94.45}{\beta}\right) . \tag{2.31m}$$

### 2.3.2 IRI-2001 corrected topside option

Some discrepancies between the IRI-2001 topside model and measurements taken in this altitude region were found in the past (Ezquer et al. 1998; Bilitza and Williamson 2000; Iwamoto et al. 2002; Triskova et al. 2002). In particular, it was found that the IRI-2001 topside model overestimates the electron density in the upper topside and unrealistically steep (almost vertical) profiles could result during high solar activity at high latitudes. Additionally, IRI-2001 predictions of the TEC turned out to be consistently lower than the observations at low latitudes especially during high solar activity, pointing to an underestimation of the electron density in the lower topside.
To overcome some of these problems Bilitza (2004) using a large database of about 200,000 topside profiles from Alouette 1, 2, and ISIS 1, 2, derived a correction term for the IRI-2001 topside model. The model thus obtained is named IRI-2001 corrected and it was put as an option for modeling the topside profile shape in the IRI-2007 version.

The IRI-2001 topside model is normalized to the F2 peak density and peak height (Eq. (2.30)), thus discrepancies between data and model can have two reasons: (1) misrepresentation of the F2 peak density and height by IRI, and/or (2) misrepresentation of the topside profile shape. To eliminate the influence of uncertainties in the F2 peak characteristics modeling, Bilitza (2004) normalized the measured profiles to the measured peak parameters. Then the IRI profile was computed for each normalized topside sounder profile and determined the ratio between data and model values. It was found that IRI-2001 topside model underestimates the topside sounder data at altitudes just above the F2 peak and overestimates the sounder data in the upper topside. Furthermore, a strong dependence on the modified dip latitude was found, especially below 1000 km, with smaller ratios at low and equatorial modified dip latitude values.
To bring the model in agreement with the data, IRI has to be multiplied by a factor varying with the height and the modified dip latitude. In addition, Bilitza (2004) found also differences between noontime and midnight (the only two times studied). Thus, Bilitza (2004) calculated the correction factors for different heights, modified dip latitudes, and for noon and midnight. Using these correction term values Bilitza (2004) established a simple, first-order correction model for IRI-2001. The correction factor varies with altitude from a value slightly greater than 1 just above the F peak to about 0.15 at 1500 km. The lowest values were found near the modip equator and the best match with IRI-2001 was found (as expected) at middle latitudes. This study did not reveal a dependence on the solar activity for the low and medium solar activities covered by the database.

### 2.3.3 IRI-NeQuick topside option

The NeQuick topside was included in the IRI-2007 version (Coisson et al. 2009). Its mathematical representation is an Epstein layer, with a height-dependent thickness parameter *H*:

$$N(h) = 4Nm\mathrm{F2}\frac{\exp(z)}{\left[1+\exp(z)\right]^2} , \tag{2.32a}$$

$$z = \frac{h - hm\mathrm{F2}}{H} , \tag{2.32b}$$

$$H = H_0\left[1+\frac{rg\left(h-hm\mathrm{F2}\right)}{rH_0+g\left(h-hm\mathrm{F2}\right)}\right] , \tag{2.32c}$$

where $NmF2$ [$el/m^3$] is the maximum electron density of the F2 layer, $hmF2$ [$km$] is the height of this maximum, $r$ = 100 and $g$ = 0.25 are constant factors.

The thickness at the F2 peak height, $H_0$, is linked to the thickness of NeQuick bottomside:

$$H_0 = kB2_{\text{bot}} \ , \tag{2.32d}$$

where

$$k = 3.22 - 0.0538\, foF2 - 0.00664 hmF2 + 0.113\frac{hmF2}{B2_{\text{bot}}} + 0.00257 R_{12} \ , \tag{2.32e}$$

in which $foF2$ [$MHz$] is the critical frequency of the F2 layer, $B2_{\text{bot}}$ [$km$] the NeQuick bottomside thickness parameter, and $R_{12}$ the twelve-month running-mean of the relative sunspot number.

The inflection point of the Epstein layer representing the NeQuick bottomside defines $B2_{\text{bot}}$ :

$$B2_{\text{bot}} = \frac{0.385 NmF2}{\left(\frac{\mathrm{d}N}{\mathrm{d}h}\right)_{\max}} \ , \tag{2.32f}$$

for which

$$\ln\left(\frac{\mathrm{d}N}{\mathrm{d}h}\right)_{\max} = -3.467 + 0.857 \ln\left(foF2\right)^2 + 2\ln\left[M(3000)F2\right]. \tag{2.32g}$$

### 2.3.4 An evaluation of the IRI model options for the topside electron density

Bilitza (2009) using a large volume of Alouette 1, 2 and ISIS 1, 2 topside sounder data evaluated the three IRI topside model options. He also studied the accurate representation of the different profiles in the equatorial anomaly region where the profile function has to accommodate two latitudinal maxima (crests) at lower altitudes but only a single maximum (at the equator) higher up.

To get an overall estimation of the data-model discrepancies, Bilitza computed the mean and the standard deviation of the Relative Deviation between data and models, $RD = \frac{N_{\text{data}}(h) - N_{\text{model}}(h)}{N_{\text{data}}(h)}$ . The results are listed in Table 2.2.

Considering all cases joined, he got a negative $RD$ indicating that, on average, all three topside options overestimate the sounder data. As expected, the largest discrepancies were found with the IRI-2001 model in the upper topside. The other two options produce significant improvements over the older model reducing the data-model difference by a factor of up to 10. Best results were obtained with the IRI-NeQuick option. Taking all Alouette and ISIS data together it was found (last column of Table 2.2) that the 165% overestimation of the data with IRI-2001 (mostly due to the error at high altitudes) is reduced to 46% with IRI-2001 corrected and to 24% with IRI-NeQuick. These two topside models also significantly reduce the scatter of difference values as shown by the decrease of the standard deviation by a factor of 2-5.

| | Alouette-1 | | Alouette-2 | | ISIS-1 | | ISIS-2 | | ALL |
|---|---|---|---|---|---|---|---|---|---|
| | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** | **Mean** |
| **IRI-2001** | -0.50 | 4.50 | -2.23 | 9.90 | -1.79 | 7.44 | -2.12 | 10.39 | -1.65 |
| **IRI-2001 corrected** | -0.22 | 2.44 | -0.38 | 2.04 | -0.43 | 2.72 | -0.61 | 2.97 | -0.46 |
| **IRI-NeQuick** | -0.11 | 1.96 | -0.04 | 0.98 | -0.21 | 1.74 | -0.31 | 1.92 | -0.24 |
| **Number of data** | 19,434 | | 5,166 | | 20,105 | | 25,214 | | 120,059 |

**Table 2.2:** Mean and standard deviation of the relative deviation between different topside sounder datasets and IRI topside models; and considering all datasets joined (last column).

Furthermore, the typical altitude-latitude structure of the Equatorial Anomaly region is better represented by IRI-2001 corrected. This model option reproduces the merging of the F-region crests into a single crest at the magnetic equator higher up, while IRI-NeQuick shows the F-region crests continuing to higher altitudes. Best results were obtained at mid-latitudes while worse results characterize high and low latitudes.

## 2.4 STORM option in IRI

Since the IRI 2001 version, also a storm option, as a correction factor for magnetically disturbed conditions, is included in IRI (Fuller-Rowell et al. 2000; Bilitza 2001; Araujo-Pradere et al. 2002). This option consists of an empirical ionospheric storm-time correction model that scales the quiet time F region critical frequency (*fo*F2) to account for storm-time changes in the ionosphere. IRI uses the 3-hourly $a_p$ index for the description of magnetic storm effects, and the storm model option is driven by a new index based on the integral of the $a_p$ index over the previous 33 hours, weighted by a filter obtained by the method of the singular value decomposition. The storm option gives reliable results at mid-latitudes during summer and equinox, but during winter and near the equator, the model does not improve significantly the IRI representation.

It is worth highlighting that the IRI storm model option was implemented mostly to represent the mid-latitude F2 peak density variations for disturbed conditions. Anyhow, setting on this option clearly influences the whole electron density profile over the entire terrestrial globe.

# 3. IRI UP: an assimilative method for updating the IRI model

Prediction of solar activity, geomagnetic activity, and neutral atmosphere changes are essential for forecasting actual ionospheric conditions. The complex interactions occurring between the atmosphere, the ionosphere, and the plasmasphere, make the spatial and temporal variations of the ionospheric plasma, for both quiet and particularly disturbed conditions, difficult to describe. Ionospheric empirical climatological models, like the International Reference Ionosphere (IRI) model (Bilitza et al. 2017), are not able to predict the complete ionospheric variability, specifically under disturbed magnetic conditions (Mirò Amarante et al. 2007). Therefore, as established by the IRI community (Bilitza et al. 2011, 2017), real-time space-sparse ionospheric data should be used in conjunction with climatological models in order to get a picture of the ionospheric plasma variability as near as possible to the real conditions.

Assuming there are several spatially-sparse ionospheric measurements, data assimilation is a process of merging measurements with a model to improve the estimate of the ionospheric conditions over the area covered by the model, also where direct measurements are not available. Thus, by means of data assimilation, it is possible to spatially expand the effectiveness of limited measurements using the model and, at the same time, increase the accuracy of model estimates.

For this reason, a new assimilation method to update the IRI model has been developed. International Reference Ionosphere UPdate (IRI UP) is an assimilative method aiming to produce an updated representation of the ionospheric plasma distribution by assimilating *fo*F2 and *M*(3000)F2 values in the climatological IRI model. Assimilated measurements are used to calculate ionospheric effective indices, $IG_{12\mathrm{eff}}$ and $R_{12\mathrm{eff}}$, at the assimilation stations which, in turn, are used to produce effective indices maps through the Universal Kriging method (Kitanidis 1997; Pignalberi et al. 2018). By ingesting effective indices maps, the IRI model is able to produce an updated representation of the F2 layer region by far better than the climatological one for both quiet and disturbed conditions.

This chapter introduces the main ideas behind the IRI UP method development. In section 3.1, the importance of using effective ionospheric indices in the ionospheric modeling is introduced, along with the need to spatially characterize their behaviour. In section 3.2, the algorithm used to calculate $IG_{12\mathrm{eff}}$ and $R_{12\mathrm{eff}}$ indices is described. In section 3.3, the Universal Kriging geostatistical interpolation method, by means of which effective indices maps are calculated, is theoretically explained.

Topics covered in this chapter are published in:

*Pignalberi, A., Pezzopane, M., Rizzi, R., Galkin, I. (2018) Effective solar indices for ionospheric modeling: A review and a proposal for a real-time regional IRI. Survey Geophys. 39(1), 125-167. doi:10.1007/s10712-017-9438-y.*

and:

*Pignalberi, A., Pietrella, M., Pezzopane, M., Rizzi, R. (2018) Improvements and validation of the IRI UP method under moderate, strong, and severe geomagnetic storms. Earth, Planets and Space, 70(180). https://doi.org/10.1186/s40623-018-0952-z*

## 3.1 Updating the IRI model: background and perspectives

Empirical ionospheric models of the standard vertical incidence ionospheric characteristics are based on analytical spatial and temporal functions that oversimplify a number of processes that are instead important for radio communication applications. In these models, ionospheric characteristics such as *fo*F2 and *M*(3000)F2, the vertical electron density profile, and other ionospheric parameters as vTEC (vertical total electron content), are usually represented as monthly median values varying as a function of geographic location, local time and solar activity. Models of this kind are then climatological since they are not designed to predict the day-to-day variability characterizing the ionospheric plasma, particularly during disturbed conditions. The major exponent of this category is the IRI model described in chapter 2 (Bilitza et al. 2014, 2017). IRI model provides some options describing the behavior of some ionospheric characteristics under magnetically disturbed periods (STORM option described in section 2.4); nevertheless, these options are not able to catch very large deviations from climatology occurring under typical geomagnetic storms (Bradley et al. 2009).

### 3.1.1 Effective solar and ionospheric activity indices for modeling the ionosphere

To catch the ionospheric variability induced by solar activity, different solar and ionospheric activity indices have been used over the years. Solar activity indices are based on solar observations of phenomena occurring at the solar surface or atmosphere regardless of the effects that these phenomena can have on the ionospheric characteristics, and they are used simply as proxies.

The most popular solar index is based on the long-term sunspot number that is used to define the daily Zurich sunspot number $R$ and the twelve-months Zurich sunspot number running mean $R_{12}$, which is used in CCIR and URSI models encapsulated in the IRI model to model *M*(3000)F2 (section 2.2.1, Bilitza 1990). Other solar indices used in some ionospheric models are the $F_{10.7}$ radio solar flux and the EUV solar emission flux.

One of the most used ionospheric activity index is the $IG$ index (Liu et al. 1983) and its twelve-months running mean $IG_{12}$. It is calculated from linear regression equations based on CCIR modeled monthly median *fo*F2 values, at that time dependent on $R_{12}$, and monthly medians *fo*F2 values recorded at noon at 13 ionospheric stations globally spread (Liu et al. 1983). As stated in chapter 2, currently, only four ionospheric stations are used in the $IG_{12}$ index calculation (Brown et al. 2018). The resulting index is the median of $IG$ index values calculated in those selected locations, and the corresponding $IG_{12}$ is used in CCIR and URSI to model *fo*F2 (section 2.2.1). Indices like $IG$, despite being calculated by means of a solar activity proxy, include also the variability caused by other phenomena affecting the ionosphere, like the geomagnetic activity, the neutral composition change or dynamical disturbances, because *fo*F2 values, both measured and modeled, have these footprints. Thus, an index calculated merging available information deduced from measured and modeled ionospheric characteristics can tell us much more on ionospheric conditions than a direct solar activity index. This is the idea behind the development of effective indices made by several authors.

Secan and Wilkinson (1997) developed an effective sunspot number $SSN_{\mathrm{e}}$ by fitting a model of *fo*F2 to observed *fo*F2 values. Hence, $SSN_{\mathrm{e}}$ shares with $IG$ index the concept of fitting a global model for *fo*F2 to observations. The $SSN_{\mathrm{e}}$ index, deduced from a set of *fo*F2 observations globally spread, is defined as the value that, when given as input to the URSI *fo*F2 model, produces a weighted zero-mean difference between the observed and modeled *fo*F2 values. In this formulation, every measured *fo*F2 value is weighted according to the absolute value of the geomagnetic latitude and the local solar time. $SSN_{\mathrm{e}}$ values were demonstrated to catch more ionospheric variability compared to long-term solar activity indices, like $R_{12}$ or $IG_{12}$, for both quiet and disturbed geomagnetic conditions. Thus, continuously updated indices have the potential to represent fast-time ionospheric variations, even though the global character of these indices does not allow to represent small-scale ionospheric variations.

### 3.1.2 Ionospheric modeling based on assimilated data in IRI

It was demonstrated (Mirò Amarante et al. 2007) that the ionosphere shows small-scale variations that global or regional empirical model have difficulties in describing, especially for very disturbed periods. Therefore, ionospheric activity indices perform to their full potential if they are able to represent fast-time and small-scale ionospheric variations (assuming that slow-time and large-scale variations are sufficiently described by a global or regional model). In order to accomplish this task, such ionospheric indices should be quickly updated in time (one hour or less) and have a fine spatial structure. For this purpose, some attempts were made by several authors using IRI as first-guess model, using different ionospheric assimilated data (vertical ionospheric soundings made by ionosondes or GNSS derived vTEC values), and spatial interpolation methods (Kriging, spline, inverse weighted functions, natural neighbor, etc.).

A first approach to update the IRI model through updated indices was made by Komjathy and Langley (1996), who used hourly vTEC maps with a 1°x1° spatial resolution. They were able to improve IRI-95 vTEC outputs with an error reduction of about 32.5%. A similar approach was pursued by Hernandez-Pajares et al. (2002), who used 2-hourly vTEC global maps, with a resolution of 5°x2.5° in longitude and latitude. They performed a local fit, over each IGS (International GPS Service) station (about 500 at that time), between IRI values and those of the vTEC map by tuning the Sun spot number.

In order to improve the IRI predictions, Bilitza et al. (1997) used *fo*F2 measurements from more than 70 ionosondes worldwide, for the period 1986-1989, to calculate $IG_{12}$ updated values. Updated indices were calculated matching measured *fo*F2 values with those calculated by URSI model (Rush et al. 1989), for each station, and then were globally or zonally averaged. They found that zonally averaged updated indices bring to better results than globally ones, thus highlighting the importance of their spatial characterization.

An attempt to update the IRI model by means of measured GNSS vTEC data collected in the African region has been recently pursued by Habarulema and Ssessanga (2016). They implemented an iterative procedure to minimize the difference between measured and modeled (by IRI 2012 model) vTEC values modifying iteratively $R_{12}$ and $IG_{12}$ solar activity indices used by IRI to model the vertical electron density profile. This procedure is repeated for every GNSS station assimilated and then adjusted solar activity indices are averaged to obtain two single $R_{12}$ and $IG_{12}$ values which are used to update the IRI model. They evaluated their method for a disturbed (March 9 2012) and a quiet (March 25 2012) day comparing the corresponding output with both ionosonde and Radio Occultation COSMIC derived data. By averaging results over the whole African region, the modified IRI 2012 gave an improvement of about 18% in estimating vTEC during the stormy day, and about 28% during the quiet day. Anyway, the authors highlighted that the use of single averaged solar activity indices strongly smooth the spatial information embedded in them, above all considering a large region as Africa.

vTEC-based methods used to update the IRI model suffer the limitation related to the calculation of vTEC maps and measurements and to the intrinsic limit in height of the model (2000 km of height). vTEC maps and measurements have not yet reached an appropriate accuracy owing to noise, multipath propagation, phase jumps and subjectivity of the ionospheric model choice (being the calculation of vertical vTEC values, derived from slant ones, dependent on the choice of the height of the ionospheric shell). Furthermore, the IRI model does not consider the plasmaspheric contribution, being the vTEC calculation limited in height (2000 km), whilst GPS-based vTEC measurements contain it (Klimenko et al. (2014) suggested that the plasmaspheric contribution to vTEC values can potentially be up to 20% during the day and up to 50% at night), thus causing a mismatch which can degrade the accuracy of results regardless the employed updating method. Another important difficulty in assimilating vTEC values in IRI is due to the fact that vTEC values, in IRI, are calculated a posteriori from the vertical electron density profile, while other characteristics as *fo*F2 and *M*(3000)F2 are directly parametrized by the IRI model. Thus, an erroneous shape of the electron density profile (above all in the topside part, as recently demonstrated by Pignalberi et al. (2016)), can lead to an incorrect vTEC estimate, even though the modeled *fo*F2 value is improved through an updating procedure.

### 3.1.3 Describing the spatial variations of the effective activity indices

The methods above reported focused principally on the choice of the best ionospheric parameters to constrain an activity index (solar or ionospheric), but considering the spatial description of lesser importance. Some of these effective indices, like $SSN_e$ in Secan et al. (1997), are characterized by a single value which is considered valid for the whole area under study; others are estimated in all grid points of the assimilated ionospheric parameter map, like in Komjathy and Langley (1996). In both cases, the spatial behavior of indices is overshadowed; the spatial description is delegated to the underlying climatological empirical model used or to the method used to obtain the assimilated map.

In the case of a single index valid for the whole region, errors can arise due to an insufficiently accurate spatial description of the ionospheric parameters in the underlying empirical model. However, even assuming a perfect spatial description by the underlying model, errors arise because the effective indices contain inside them the imprint of several geophysical phenomena affecting the ionospheric parameter used in their computation, which are, in most cases, heavily spatially varying. Hence, it is clear the need to take into account the spatial behavior of an effective index.

In the second case, that is the one of multiple indices calculated in each point of the grid region, errors can arise because the assimilated ionospheric parameter grid is itself obtained by means of interpolation of spatial scattered measures, thus introducing additional errors that are difficult to estimate without a thorough a priori knowledge of how the assimilated grid was obtained.

Hence, it is clear the need to take the full control on the spatial description of the effective indices through two main steps:

1. to obtain them from reliable measures taken in geographically scattered ionospheric stations;

2. to spatially interpolate the values obtained in 1) with a fully manageable spatial interpolation method, whose benefits and shortcomings can be clearly estimated and that can be adapted to the different geophysical conditions that can occur.

In an attempt to pursue both objectives, several spatial interpolation methods were developed to obtain an instantaneous mapping of several ionospheric parameters. Samardjiev et al. (1993) compared three spatial interpolation methods for an instantaneous ionospheric mapping of *fo*F2 and *M*(3000)F2 characteristics over the European region, that is: the Inverse-distance power law method, the Minimum-curvature analysis and the geostatistical procedure known as Kriging method. This comparison showed that Kriging method gives the lowest mean square deviation for both small and large distances from the scattered assimilated measures, for both *fo*F2 and *M*(3000)F2.

A Kriging method for an instantaneous mapping of *fo*F2, was used in particular by Stanislawska et al. (1996). They applied the ordinary Kriging method (Matheron 1963; Oliver and Webster 1990; Kitanidis 1997) to obtain *fo*F2 maps in the European region after assimilating *fo*F2 values measured by a network of 10 ionosondes. They described the spatial variation of the sampled data through a linear variogram model. To build the experimental variogram, and in order to have a consistent statistical sample, regardless of the hour of the day, they used *fo*F2 values recorded for consecutive days, since semivariance values for a definite time moment were very randomly distributed. It is clear that this way to proceed neglects any spatial correlation property linked to the ionospheric daily variability. Unfortunately, the proposed variogram's construction method turns out to be inapplicable for real-time applications in which only measures collected in a precise moment are available. Furthermore, several variogram models, i.e. several mathematical functions fitting the semivariance values in the experimental semivariogram, can be employed, each of them having different small-scale and large-scale properties (McBratney and Webster 1986). Limiting the analysis to only the linear variogram model is quite restrictive.

Regional ionosphere vTEC mapping using the Kriging method was investigated by Grynyshyna-Poliuga et al. (2014), who used IGS vTEC values recorded by 15 stations located in the European region to reconstruct vTEC maps by means of the ordinary Kriging method. They probed the opportunity to use different stationary variogram

models, in particular spherical, exponential and Gaussian. They considered 15-minutes vTEC values during three days including a geomagnetic storm, obtaining maps in a region located between 30°N-60°N in latitude and 40°W-45°E in longitude on a 2.5°x2.5° grid. These maps could properly represent significant features of the ionosphere response, as latitudinal variations of the ionospheric trough for disturbed conditions.

The aforementioned considerations about the application of the ordinary Kriging method to generate ionospheric parameter maps led us to conclude that it might be a good spatial interpolation method. However, it is characterized by several user arbitrary choices (the variogram to be used, which and how many values to be used to build it, the use of all values or only a selection of it, etc.) that make its use "dangerous", in the sense that certain choices provide worse results than other, more user-friendly, interpolation methods. In the IRI UP method the Universal Kriging interpolation method is used, because it is considered more suitable than the ordinary one for ionospheric issues. Several stationary and non-stationary variogram models are investigated, in order to exploit the potentiality of the Kriging method.

### 3.1.4 Five steps to make a good ionospheric model update based on effective activity indices

The critical review of previous subsections highlights five points to be followed in the development of a method to generate reliable ionospheric maps:

1. An empirical climatological ionospheric model is adopted with improved temporal and spatial description of the ionosphere, and whose ionospheric characteristics depend on some solar/ionospheric activity indices that can be chosen by the user;
2. The most reliable ionospheric characteristics must be assimilated in order to have the highest confidence. Specifically, those characteristics that are directly parametrized by the underlying model, preferably those depending on only one solar/ionospheric activity index, should be considered;
3. The assimilated ionospheric characteristic is used to update the solar/ionospheric activity index by which it depends, for several points on a chosen region;
4. A spatial interpolation method is used to obtain maps of updated solar/ionospheric activity index/indices on a region whose extension depends on the geographical position of stations from which data are assimilated;
5. The underlying model is fed with the updated solar/ionospheric activity index/indices to obtain updated values of the selected ionospheric characteristics.

The choices made to build the IRI UP method, strictly linked to the aforementioned five points, are also guided by the considerations made in the previous subsections about potentialities and shortcomings of each developed method.

## 3.2 $IG_{12\mathrm{eff}}$ and $R_{12\mathrm{eff}}$ effective indices calculation

IRI gives the user the opportunity to input a value for both $R_{12}$ and $IG_{12}$ indices. In this way, using real-time updated $IG_{12}$ and $R_{12}$ values, improved *fo*F2 and *M*(3000)F2 values can be obtained. This is because the IRI model for the current time uses forecasted values of $IG_{12}$ and $R_{12}$. Moreover, $IG_{12}$ and $R_{12}$, being 12-months running means, are strongly smoothed proxies that are good to model long-term variations of the ionosphere but cannot catch short-term phenomena, like those caused by geomagnetic storms. IRI, of course, is a climatological model, so one does not expect that phenomena occurring at shorter time scale, like for instance those characterizing geomagnetic disturbed conditions, are reliably predicted by the model. At the same time, however, because the ionospheric characteristics associated to the two anchor points of the IRI vertical electron density profile (the E and F2 peaks) are modeled through the use of these two indices, one would expect that feeding the model with

updated $IG_{12}$ and $R_{12}$ indices could make the model sensitive also to the short-term variability. In addition, $IG_{12}$ and $R_{12}$ indices are global, thus they do not catch geographical variations that are commonly observed in several ionospheric phenomena, leaving the description of these to the spherical harmonic analysis or to other functions (as described in chapter 2).

In the IRI UP method, values of *fo*F2 (for $IG_{12} = 0$ and 100) and *M*(3000)F2 (for $R_{12} = 0$ and 100) are calculated by IRI at each of the assimilated ionospheric station for the epoch at which the data assimilation is executed (i.e. for a fixed year, month, day, and hour UT). The two values are used to linearly interpolate *fo*F2 vs $IG_{12}$ and *M*(3000)F2 vs $R_{12}$ over a large range of values. An example is shown in the upper panel of Figs. 3.1 and 3.2 for the Athens station on March 9 2015 at 17 UT, for $IG_{12}$ and $R_{12}$, respectively. In both Figs. 3.1 and 3.2 the x-axis is limited in the range [-100; 200] only for a graphical representation convenience.

The square difference between observed and modeled values of *fo*F2 and *M*(3000)F2 is calculated, for each station, according to following formulas:

$$\Delta_{IG_{12}} = \left[ foF2_{\text{obs}} - foF2_{\text{IRI}}\left(IG_{12}\right)\right]^2, \tag{3.1}$$

$$\Delta_{R_{12}} = \left[ M\left(3000\right)F2_{\text{obs}} - M\left(3000\right)F2_{\text{IRI}}\left(R_{12}\right)\right]^2 ; \tag{3.2}$$

$foF2_{\text{obs}}$ and $M(3000)F2_{\text{obs}}$ are the observed values at the station, and $foF2_{\text{IRI}}(IG_{12})$ and $M(3000)F2_{\text{IRI}}(R_{12})$ are the corresponding modeled values, represented by the linear fits shown in the upper panel of Figs. 3.1 and 3.2. An example is shown in the lower panel of Figs. 3.1 and 3.2. The square difference functions expressed by Eqs. (3.1)-(3.2) are parabolas for which it is simple to find the minimum corresponding to the effective values of $IG_{12}$ and $R_{12}$. In the example shown in Figs. 3.1 and 3.2, for the Athens station on March 9 2015 at 17 UT, $IG_{12\text{eff}} = 33$ and $R_{12\text{eff}} = -27$ are found.

Thus, $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ values represent the indices by which the IRI model reproduces exactly the *fo*F2 and *M*(3000)F2 values recorded at the considered station.

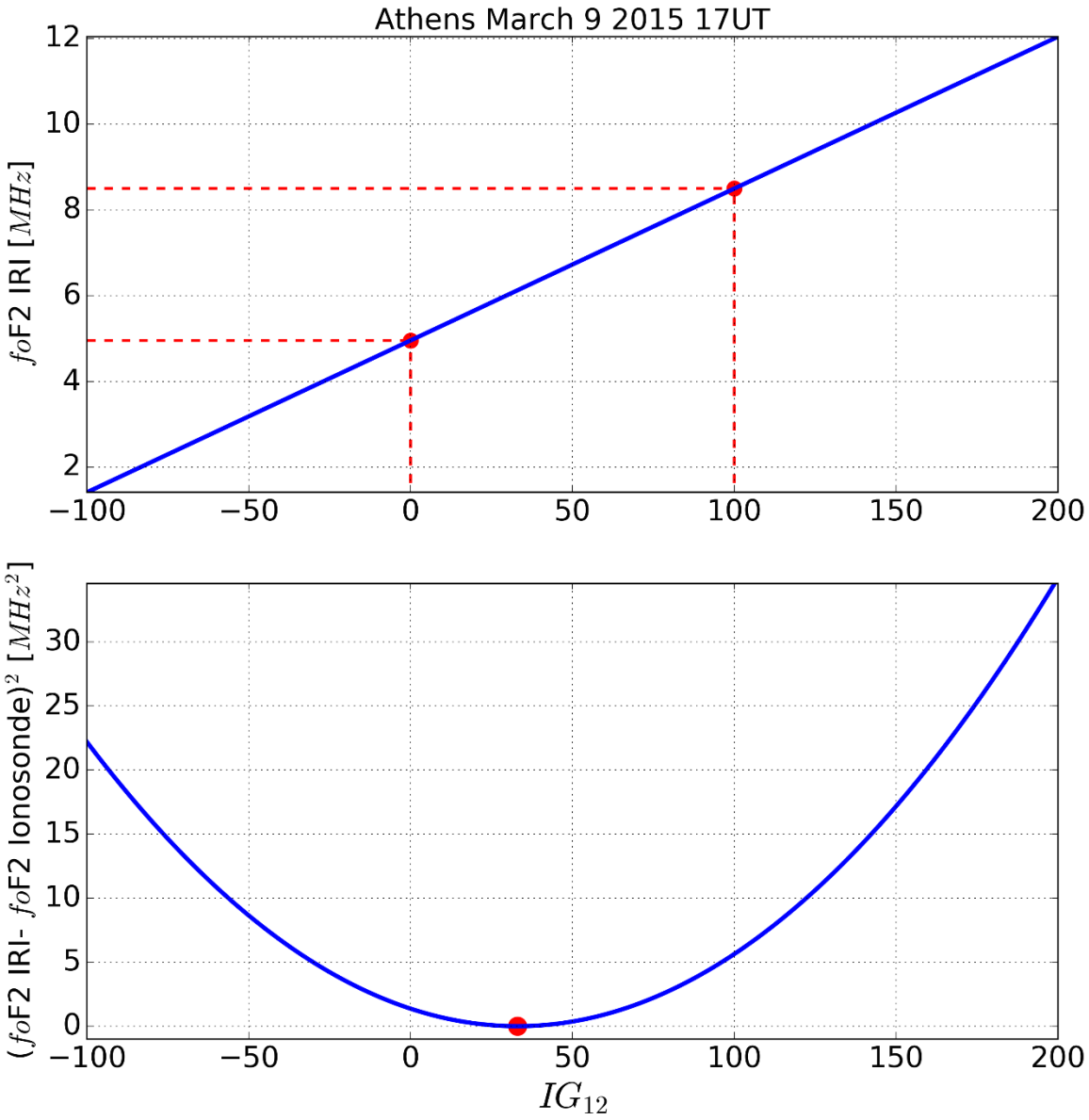


**Figure 3.1:** Linear fit between *fo*F2 and $IG_{12}$ values (upper panel). Two red points represent the IRI simulated *fo*F2 values for $IG_{12} = 0$ and $IG_{12} = 100$. Plot of the square difference between observed and modeled values of *fo*F2 vs $IG_{12}$ (lower panel), as described by Eq. (3.1). Red point in the lower panel is the calculated minimum which corresponds to the effective value of $IG_{12}$.

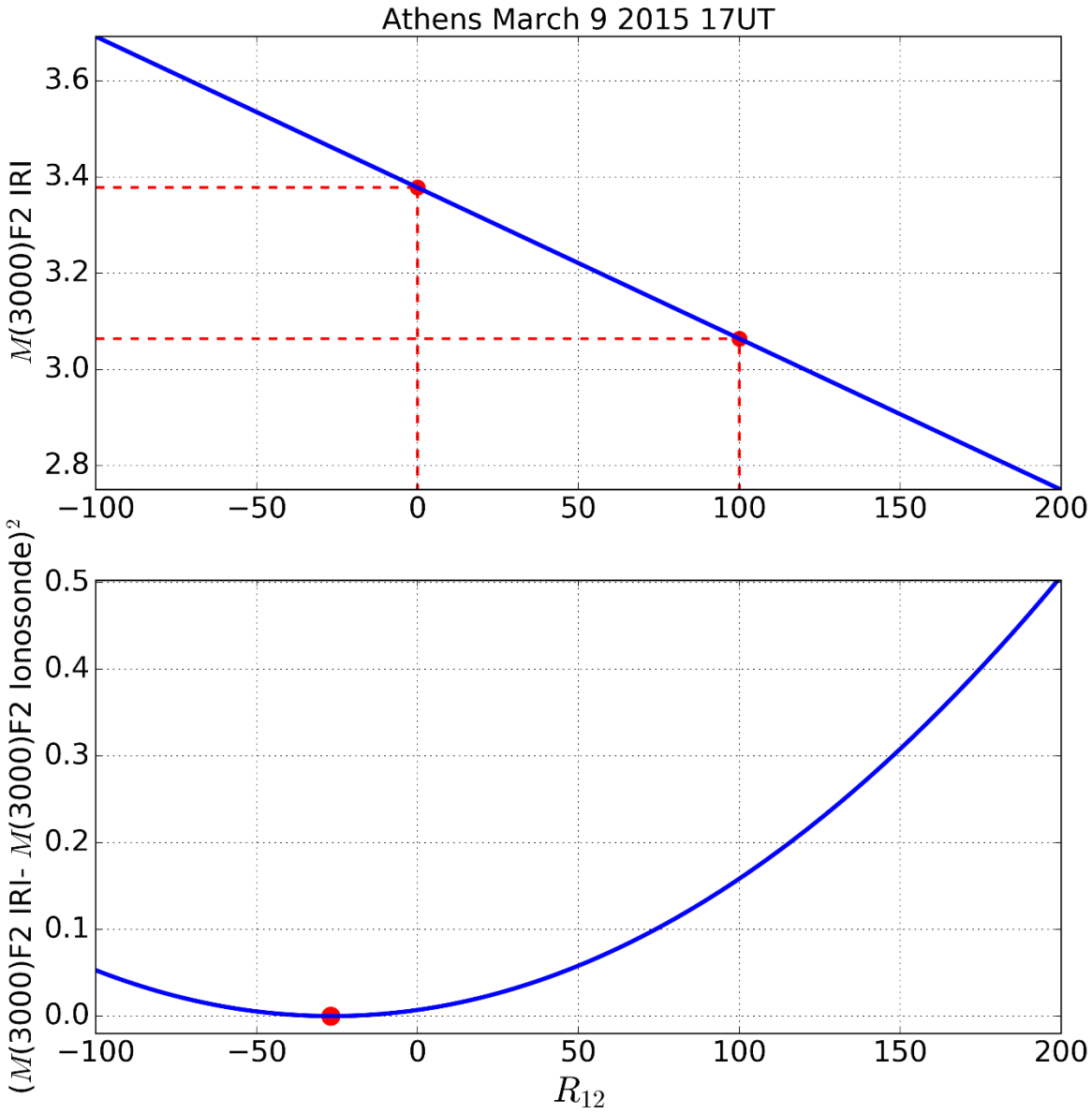


**Figure 3.2:** Linear fit between *M*(3000)F2 and $R_{12}$ values (upper panel). Two red points represent the IRI simulated *M*(3000)F2 values for $R_{12} = 0$ and $R_{12} = 100$. Plot of the square difference between observed and modeled values of *M*(3000)F2 vs $R_{12}$ (lower panel), as described by Eq. (3.2). Red point in the lower panel is the calculated minimum which corresponds to the effective value of $R_{12}$.

Since values of $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ are calculated for each assimilated station, an interpolation of these through the chosen grid is needed. To accomplish this task, the Universal Kriging method is considered to obtain maps of $IG_{12\text{eff}}$ and $R_{12\text{eff}}$.

## 3.3 Kriging interpolation method

The Kriging method is an advanced geostatistical procedure that, used as an interpolator, generates an estimated surface from a scattered set of punctual geophysical measurements (Matheron 1963; Oliver and Webster 1990; Kitanidis 1997).

It belongs to statistical *BLUE* (Best Linear Unbiasedness Estimation) methods, for which the first two statistical moments of the random field, which are the *mean function* (first moment) and the *covariance function* (second moment), are used to select the best estimation method. Thus, the model of the spatial structure consists of mathematical expressions chosen to describe the mean function and the covariance function.

Given $n$ measurements of the variable $z$ at locations with spatial coordinates $\mathbf{x}_1$, $\mathbf{x}_2$, . . . , $\mathbf{x}_n$, Kriging estimates the value of $z$ at point $\mathbf{x}_0$, making a linear combination of measurements (Oliver and Webster 1990):

$$\hat{z}_0 = \sum_{i=1}^{n} \lambda_i z(\mathbf{x}_i) \ . \tag{3.3}$$

Thus, the problem is reduced to select a set of coefficients $\lambda_1$ , . . . , $\lambda_n$ that match the condition of *unbiasedness*:

$$E\left[\hat{z}_0 - z(\mathbf{x}_0)\right] = 0 \ , \tag{3.4}$$

i.e., the expected value of the difference between the Kriging estimate $\hat{z}_0$ and the actual value $z(\mathbf{x}_0)$, the *estimation error*, must be zero; and of *minimum variance*:

$$\text{Minimum } \left\{E\left[\left(\hat{z}_0 - z(\mathbf{x}_0)\right)^2\right]\right\} \ , \tag{3.5}$$

i.e., the mean squared estimation error must be minimum.

Eq. (3.3) highlights how Kriging is a global interpolation method. Global interpolation methods use a weighted sum of all data to compute output values, thus the information contained in each measured point is shared with all other points, regardless the distance between two points, taking into account that the spatial correlation between points is embedded in the measured data itself.

### 3.3.1 The experimental variogram

The building of the *experimental variogram* (or semivariogram) is the core of the Kriging method, because it allows to investigate the spatial correlation between measurements, and is the first step of the structural analysis (McBratney and Webster 1986; Kitanidis 1997).

Considering *n* measurements $z(\mathbf{x}_1)$, $z(\mathbf{x}_2)$,..., $z(\mathbf{x}_n)$ taken at different locations, the *semivariance* $\gamma$ is defined as:

$$\gamma(h_i) = \frac{1}{2}\left[z(\mathbf{x}_i) - z(\mathbf{x}_i')\right]^2 \ , \tag{3.6}$$

i.e., it is the squared difference between a pair of measurements $z(\mathbf{x}_i)$ and $z(\mathbf{x}_i')$ whose distance is $h_i = |\mathbf{x}_i - \mathbf{x}_i'|$ (where | | means the length of a vector). The index *i* refers to each pair of measurements.

For *n* measurements there are $n(n-1)/2$ pairs for which it is possible to calculate semivariance and distance values. These $i$ points $(h_i, \gamma(h_i))$ form a scatter plot which is the experimental variogram.

For cases where a large number *n* of measurements is considered, it would be better to arrange the $n(n-1)/2$ pairs of measurements in *K* bins having all the same width $W = \frac{h_{\max} - h_{\min}}{K}$, being $h_{\min}$ and $h_{\max}$the absolute minimum and maximum distance among the $n(n-1)/2$ pairs of measurements, respectively.

The lower and upper limit of each bin is defined through an iterative procedure which starts from the value of $h_{\min}$ (lower limit of the first bin) and ends with $h_{\max}$ (upper limit of the last bin). For example, the first, second, third…and last bin corresponds to the intervals [$h_{\min}$, $h_{\min}$+*W*), [$h_{\min}$+*W*, $h_{\min}$+2*W*), [$h_{\min}$+2*W*, $h_{\min}$+3*W*],…, ($h_{\min}$+ (*K*-1)*W*, $h_{\max}$], respectively.

Since each bin can contain a different number, $N_K$, of pairs of measurements, the distance pertinent to each bin is then defined as the average value, $\bar{h}_K$ of the distances $h_k$ falling inside the bin:

$$\bar{h}_K = \frac{\sum_{k=1}^{N_K} h_k}{N_K} \ . \tag{3.7}$$

Therefore, the semivariance associated to a given bin is the average value of the semivariances $\gamma(h_K)$ falling inside that bin:

$$\bar{\gamma}_K = \frac{\sum_{k=1}^{N_K} \gamma(h_k)}{N_K} = \frac{1}{2(N_K)} \sum_{k=1}^{N_K} [z(\mathbf{x_k}) - z(\mathbf{x'_k})]^2 \ ; \tag{3.8}$$

the plot obtained with the *K* points (one for each bin) of coordinates ($\bar{h}_K$, $\bar{\gamma}_K$), as before, constitutes the experimental variogram.

It can be demonstrated that the semivariance is equal to minus the *covariance* function $R(h)$ plus a constant, where the constant is the *variance* $R(h=0)$(Kitanidis 1997):

$$\gamma(h) = -R(h) + R(h=0) \ . \tag{3.9}$$

The experimental variogram contains information about the *scale of fluctuations* of the variable: the behavior of the experimental variogram near the origin represents the small-scale variability while the behavior of the experimental variogram at large distances is indicative of the large-scale behavior (Kitanidis 1997).

### 3.3.2 Variogram models

The second step of the structural analysis is to fit a mathematical expression, called *variogram model*, to the points of the experimental variogram. Variogram models are linear combinations of simple mathematical functions with the restriction that these should be positive-definite.

Considering their large-scale behavior, variogram models can be grouped in two distinct families: *stationary* and *non-stationary* models. Stationary models manifest small-scale fluctuations, compared to the size of the domain, around some well-defined mean values; the corresponding experimental variogram should stabilize around a value, called *sill*, and the length scale at which the sill is obtained, called *range* or *correlation length*, describes the scale at which two measurements of the variable become practically uncorrelated. On the contrary, variograms of non-stationary models increase even at distances comparable to the maximum separation of interest.

The behavior of the variogram for short distances determines whether the spatial function appears continuous and smooth. Three different small-scale behaviors are identifiable: *discontinous*, *parabolic*, and *linear*. A variogram is discontinuous if the average sampling interval (i.e., the distance between measurement locations) is several

times larger than the scale of fluctuations of the variable; in this case, two adjacent values are about as different as two distant values. At the scale of the sampling interval, the variable $z$ is discontinuous because it changes abruptly from one sampling point to the next. In this case, the experimental variogram is approximately a straight horizontal line, and since it does not converge to zero as the separation decreases, there is a discontinuity of the experimental variogram at the origin, or a *nugget effect* (a condition called *microvariability*). Instead, if all the variability is at a scale much larger than that of the sampling interval, changes of values are so gradual that both $z$ and its slope vary continuously. In this case, the experimental variogram has a parabolic behavior near the origin, that is, it is proportional to $h^2$ for small values of $h$. An intermediate case is when the variable presents most of its variability at a scale larger than the average sampling interval, but also some variability at a scale comparable to that of the measurement spacing; in this case, changes of $z$ between adjacent sampling points are gradual so that $z$ is practically continuous at the scale of measurements, and the experimental variogram has approximately a linear behavior near the origin, that is, the variogram is proportional to $h$ for small values of $h$.

Five different variogram models are used in the IRI UP method (Kitanidis 1997; Pignalberi et al. 2018): *Gaussian*, *spherical*, and *exponential*, that are stationary models; *power* and *linear*, that are non-stationary models.

- Gaussian:

$$\gamma(h) = c_0 + (\sigma^2 - c_0)\left(1 - \exp\left(-\frac{7h}{4\alpha}\right)^2\right), \tag{3.10}$$

  where $c_0$ is a constant, called *nugget value* (is the intercept of the variogram model), $\sigma^2$ is the *sill*, and $\alpha$ is the *range*. The Gaussian model has a parabolic behavior at the origin.

- Spherical:

$$\gamma(h) = \begin{cases} c_0 + (\sigma^2 - c_0)\left(\frac{3}{2}\frac{h}{\alpha} + \frac{1}{2}\left(\frac{h}{\alpha}\right)^3\right) & for \quad 0 \le h \le \alpha, \\ \sigma^2 & for \quad h > \alpha. \end{cases} \tag{3.11}$$

  The spherical model exhibits a linear behavior at the origin.

- Exponential:

$$\gamma(h) = c_0 + \left(\sigma^2 - c_0\right)\left(1 - \exp\left(-\frac{3h}{\alpha}\right)\right). \tag{3.12}$$

  The exponential model exhibits a linear behavior at the origin.

- Power:

$$\gamma(h) = c_0 + \theta h^s, \tag{3.13}$$

  where $\theta > 0$ is the *scale* and $0 < s < 2$ is the *exponent*.

- Linear:

$$\gamma(h) = c_0 + \theta h, \tag{3.14}$$

  where $\theta > 0$ is the *slope*.

Examples of variogram models defined by Eqs. (3.10)-(3.14) are shown in Fig. 3.3. Each variogram of Fig. 3.3 has the same nugget value (this is testified by the same intercept of each curve), and stationary models are considered with the same sill and range values. Fig. 3.3 shows clearly the different large-scale behavior between non-stationary and stationary models. It is worth noting that stationary models present a different small-scale behavior, also considering same values of both range and sill. In particular, the Gaussian model is the smoothest one and reaches the sill value well beyond the range value; instead, exponential and spherical models are steeper near the origin and reach the sill value near the range value.

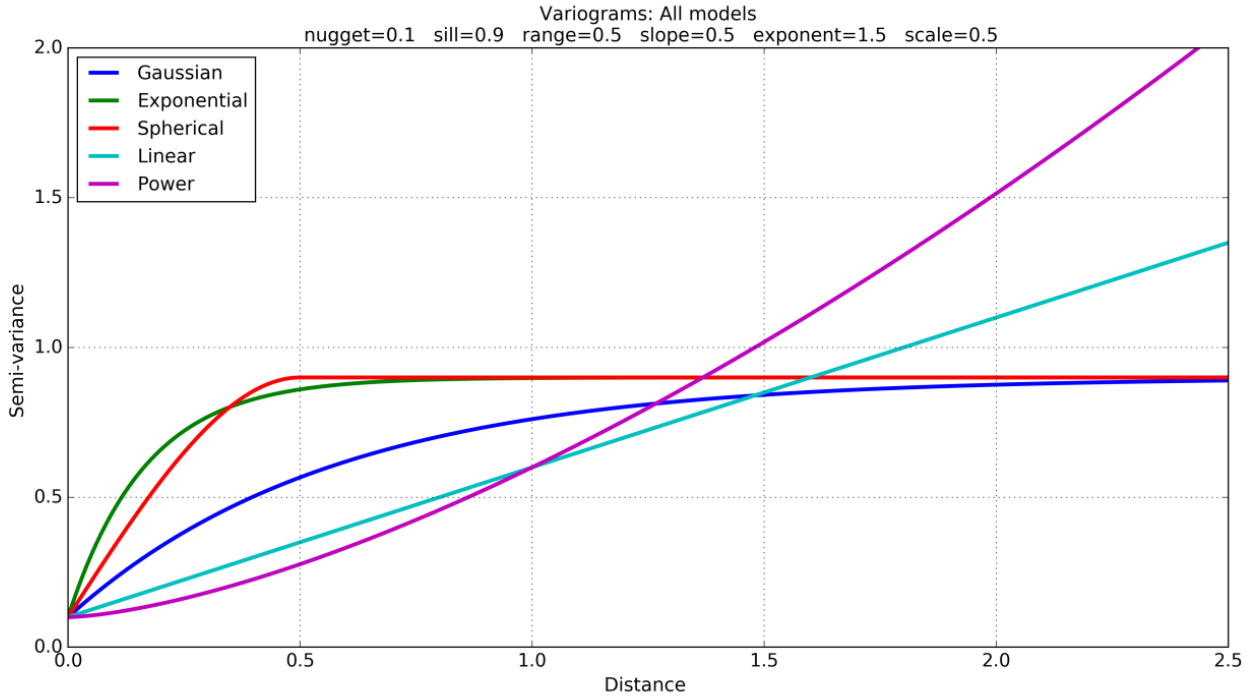


**Figure 3.3:** Examples of variogram models: Gaussian (blue), Exponential (green), Spherical (red), Linear (cyan) and Power (magenta). Used parameters are shown at the top of the figure. The horizontal axis represents the independent variable $h$ and the vertical axis the $\gamma(h)$ value used in Eqs. (3.10)-(3.14).

### 3.3.3 On the choice of the best variogram model in the Universal Kriging procedure

The Universal Kriging method performance strongly relies on the choice of the variogram model which, in principle, should be the one that best fits the experimental semivariance values. Therefore, the problem of how to choose among the possible variogram models described before is of crucial importance, because this choice greatly affects the goodness of the maps of the interpolated effective index.

Some statistical quantities can be calculated by looking at residuals between experimental and modeled (by variogram model) semivariance values in order to select the more reliable variogram model for any particular situation.

#### $Q_1$ statistics

It is possible to define the statistical variable $Q_1$:

$$Q_1 = \frac{1}{n-1}\sum_{k=2}^{n}\varepsilon_k \,, \tag{3.15}$$

where *n* is the number of observations and $\varepsilon_k = \frac{\delta_k}{S}$ are the normalized residuals, being $\delta_K$ the common residuals and *S* the variance of their distribution. It can be proven (see Kitanidis 1997) that $Q_1$ is a statistical variable which follows the normal distribution with a *probability density function* (*pdf*) given by:

$$f(Q_1) = \frac{1}{\sqrt{\frac{2\pi}{n-1}}} \exp\left( -\frac{Q_1^2}{\frac{2}{n-1}} \right) , \tag{3.16}$$

with a mean value *m* = 0 and a variance $\sigma^2 = \frac{1}{n-1}$.

Therefore, with a probability of about 95%, $Q_1$ is ranged in the interval:

$$-2\sigma < Q_1 < +2\sigma \text{ i.e. } |Q_1| < \frac{2}{\sqrt{n-1}} , \tag{3.17}$$

and with a probability of about 68%, $Q_1$ is ranged in the interval:

$$-\sigma < Q_1 < +\sigma \text{ i.e. } |Q_1| < \frac{1}{\sqrt{n-1}} . \tag{3.18}$$

An example of the *pdf* and the *cumulative probability function* (*cpf*) of the variable $Q_1$, obtained for *n* = 12, is shown in Fig. 3.4.

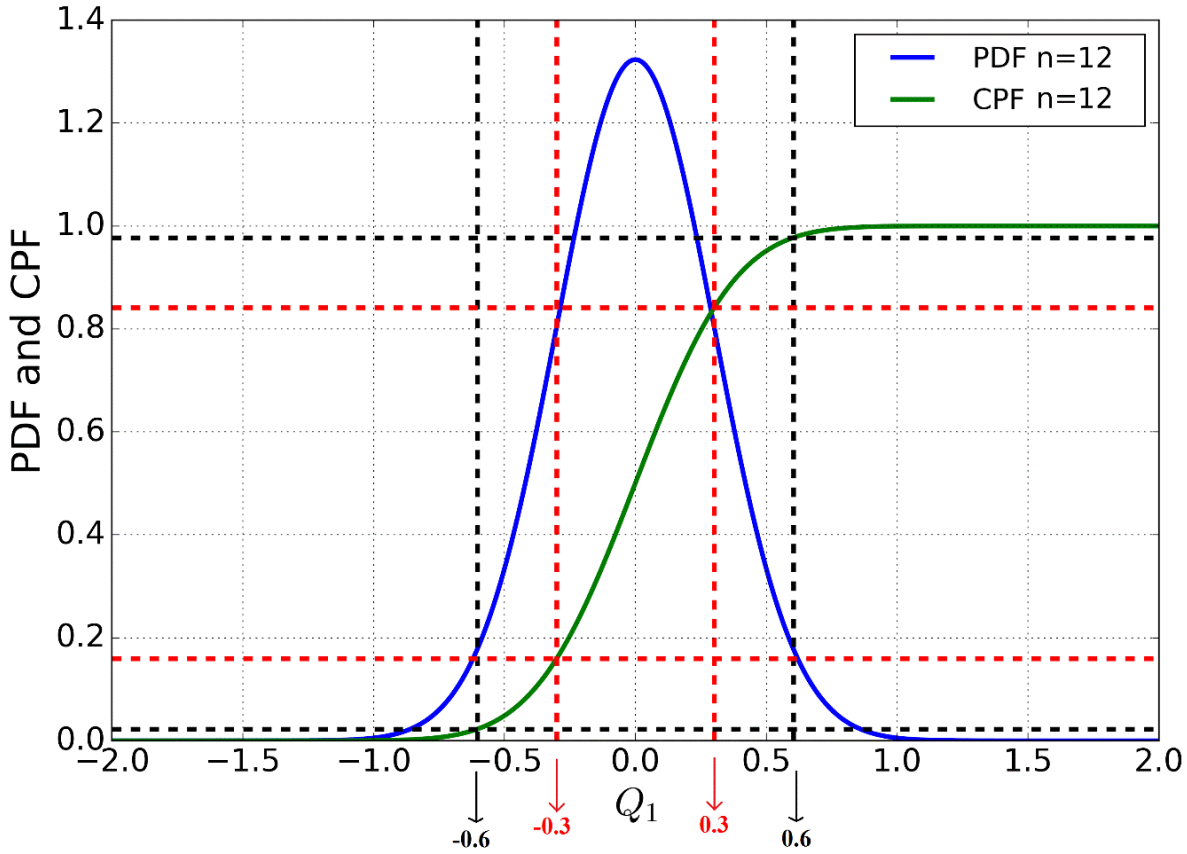


**Figure 3.4:** Example of *pdf* (solid blue line) and *cpf* (solid green line) for $Q_1$, for $n$ = 12. Red and black arrows (corresponding to red and black dashed vertical lines) indicate respectively the thresholds ± 1$\sigma$ and ± 2$\sigma$. Black and red dashed horizontal lines represent the values referred to the *cpf* corresponding to the conditions (3.17) and (3.18) which, for the considered case, provide respectively the numerical solutions -0.60 < $Q_1$ < +0.60 and -0.30 < $Q_1$ < +0.30.

When the condition (3.17) is fulfilled, the variogram model under examination is accepted and it is significant, from the statistical point of view, with a probability of about 95%; this implies that there exists a 5% of probability to accept an incorrect variogram model. However, a 5% cut off is customary in statistics. Therefore, when the condition (3.17) is met, it is assumed that the variogram model has passed the $Q_1$ statistical test.

### $Q_2$ statistics

Another test that can be carried out to test the goodness of the variogram model is that relying on the statistical variable $Q_2$ defined as:

$$Q_2 = \frac{1}{n-1}\sum_{k=2}^{n}\varepsilon_k^{\,2} \,. \tag{3.19}$$

It can be proven (see Kitanidis 1997) that $Q_2$ is a statistical variable whose *pdf* is:

$$f(Q_2) = \frac{(n-1)^{\frac{n-1}{2}} Q_2^{\frac{n-3}{2}} \exp\left(-\frac{(n-1)Q_2}{2}\right)}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)}, \tag{3.20}$$

with mean value $m \to 1$, for $n \to \infty$, and variance $\sigma^2 = \frac{2}{n-1}$.

An example of the *pdf* and *cpf* of the variable $Q_2$, obtained for $n$ = 12, is shown in Fig. 3.5.

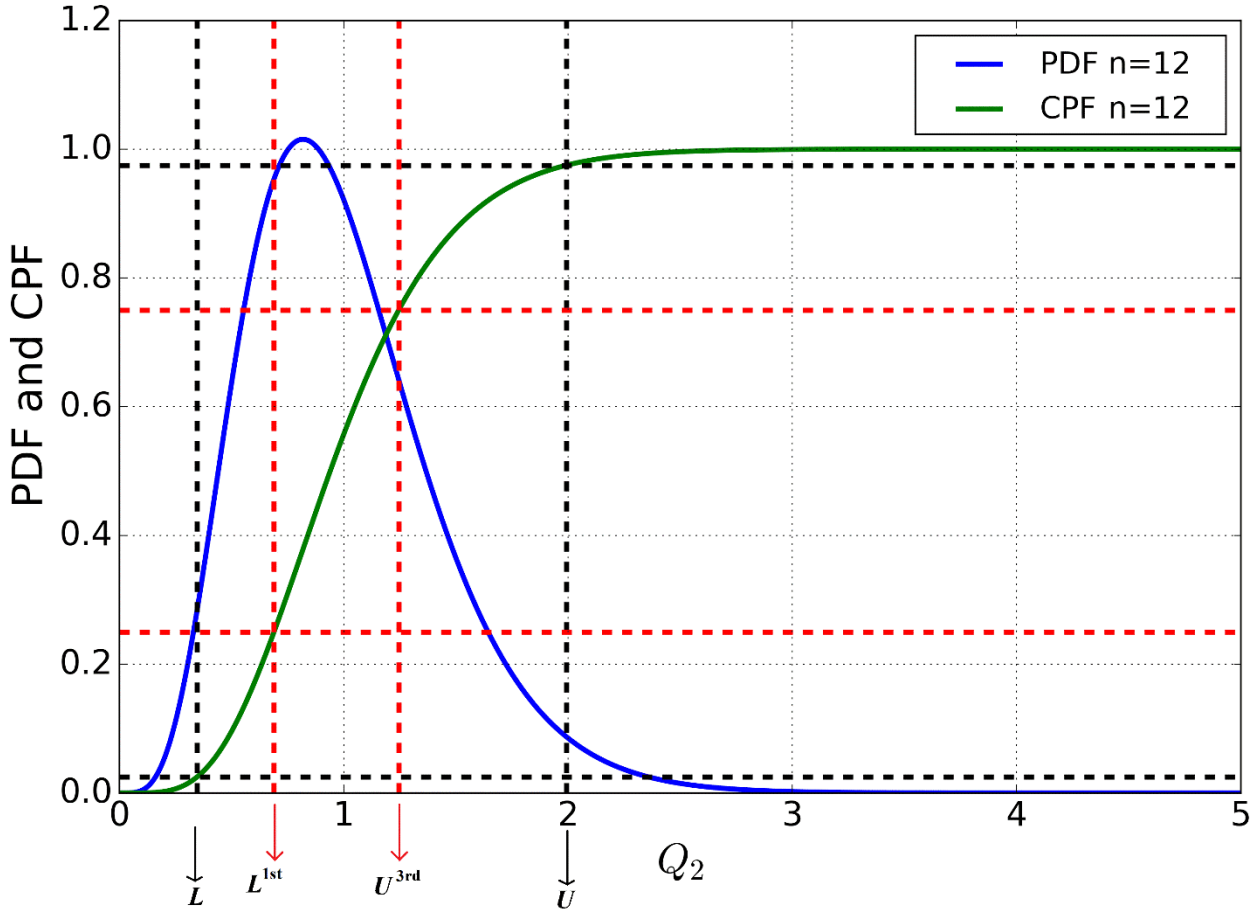


**Figure 3.5:** Example of *pdf* and *cpf* for $Q_2$ for $n$ = 12. The black dashed horizontal lines intersect the *cpf* at the two points of coordinates ($L$, 0.025) and ($U$, 0.975), while the red dashed horizontal lines intersect the *cpf* at the two points of coordinates ($L^{1st}$, 0.25) and ($U^{3rd}$, 0.75), where $L^{1st}$ and $U^{3rd}$ identify the first and third quartile, respectively. Therefore, the black and red arrows mark the intervals of the variable $Q_2$ corresponding to a statistical significance of 95% and 50%, respectively.

Therefore, according to Fig. 3.5, if the value of $Q_2$ given by Eq. (3.19) is included in the interval:

$$L < Q_2 < U, \tag{3.21}$$

the variogram model under examination is significant, from the statistical point of view, with a probability of 95%. When this condition is met, one could say that the variogram model has passed the $Q_2$ statistical test to the usual

confidence level of 5%; in the sense that there exists however a 5% probability that an incorrect variogram model is accepted.

It is worth noting that the form of *pdf* and *cpf* depends on the number *n* of assimilated ionosonde data and, consequently, the same stands for the values of the two thresholds (*L*,*U*) and ($L^{1st}$, $U^{3rd}$).

**The *cR* criterion**

The residuals are particularly important in evaluating how closely the variogram model fits the data, since smaller residuals imply a better fit. To construct stable (i.e., less affected by random error) criteria for the choice of the best variogram model, it is used the stabilized geometric mean of the square residuals (see Kitanidis 1997), i.e., the parameter:

$$cR = \exp\left[\frac{1}{n-1}\sum_{k=2}^{n}\ln\left(S_k^{\,2}\right)\right]. \tag{3.22}$$

The condition:

$$\text{Minimum}\left(cR\right), \tag{3.23}$$

allows to choose among the various variogram models.

### 3.3.4 The Universal Kriging

The geostatistics method known as *Universal Kriging* (or *Kriging with a drift*) (Olea 1974; Kitanidis 1997) is based on the assumption that the measured geophysical field displays a large-scale spatial variability that can be represented as a *deterministic* function, and a small-scale spatial variability, the *stochastic* part, involving spatial correlations inferable from the variogram. Then:

$$z(\mathbf{x}) = m(\mathbf{x}) + \varepsilon(\mathbf{x}), \tag{3.24}$$

where $m(\mathbf{x})$, called the *drift*, represents the deterministic part of $z(\mathbf{x})$, while the function $\varepsilon(\mathbf{x})$ represents the stochastic part of $z(\mathbf{x})$, with zero mean ($\mathrm{E}[\varepsilon(\mathbf{x})] = 0$) and a definite correlation structure described by its covariance *R* (linked to the semivariance by Eq. (3.9)). The ordinary Kriging method is based on the assumption that $m(\mathbf{x})$ is a constant without any spatial characterization, and the spatial description is left only to the stochastic part $\varepsilon(\boldsymbol{x})$. The Universal Kriging method, however, requires an a priori knowledge of the geophysical quantity to be described, which is used in the characterization of the deterministic part. For example, *fo*F2, on average, is characterized by spatial gradients that are bigger at mid-low latitudes than at mid-high latitudes, and the same spatial gradient features will be embedded in the $IG_{12}$ activity index. Using the Universal Kriging method, this large-scale deterministic behavior can be included in the drift part, which allows to characterize also regions distant from assimilated points, without the need to employ non-stationary variogram models.

The drift part can be represented by (Kitanidis 1997):

$$m(\mathbf{x}) = \sum_{k=1}^{p} f_k(\mathbf{x})\beta_k, \tag{3.25}$$

where $f_1(\mathbf{x}), \ldots, f_p(\mathbf{x})$ are known functions of spatial coordinates $\mathbf{x}$ called *base functions*, and $\beta_1, \ldots, \beta_p$ are unknown deterministic coefficients referred to as *drift coefficients*.

In this study:

$$m(x,y) = A + Bx + Cy, \tag{3.26}$$

namely, the deterministic part is a two-dimensional *regional linear mean* in the $x$ (longitude) and $y$ (latitude) variables, with coefficients *A*, *B*, and *C* to be determined.

Hence, the spatial function is given by:

$$z(\mathbf{x}) = \sum_{k=1}^{p} f_k(\mathbf{x})\beta_k + \varepsilon(\mathbf{x}), \tag{3.27}$$

Thus, the goal is to use Eq. (3.3) to derive an estimate of $z$ at $\mathbf{x_0}$, from measurements $z(\mathbf{x_1}), \ldots, z(\mathbf{x_n})$, with the assumption that Eq. (3.27) follows the unbiasedness and minimum variance conditions Eqs. (3.4)-(3.5).

The unbiasedness condition, using Eq. (3.3), may be written:

$$\mathrm{E}\left[\hat{z}(\mathbf{x}_0) - z(\mathbf{x}_0)\right] = \mathrm{E}\left[\sum_{i=1}^{n} \lambda_i z(\mathbf{x}_i) - z(\mathbf{x}_0) = 0\right]; \tag{3.28}$$

making use of Eqs. (3.25) and (3.27) (remembering that $\mathrm{E}[\varepsilon(\mathbf{x})] = 0$) (Kitanidis 1997):

$$\sum_{k=1}^{p}\left[\sum_{i=1}^{n} \lambda_i f_k(\mathbf{x_i}) - f_k(\mathbf{x}_0)\right]\beta_k = 0. \tag{3.29}$$

For this condition to hold for any values of $\beta_k$ it is required that:

$$\sum_{i=1}^{n} \lambda_i f_k(\mathbf{x_i}) = f_k(\mathbf{x_0}) \qquad k = 1, \ldots\ldots, p. \tag{3.30}$$

This set of $p$ constraints must be satisfied by the $\lambda$ coefficients, and each unbiasedness condition has the effect of eliminating an unknown drift coefficient.

The variance of the estimation error, making use of the unbiasedness condition, involves only the stochastic part $\varepsilon(\boldsymbol{x})$, which is described by its covariance $R(\boldsymbol{x}, \boldsymbol{x}')$, leading to (Kitanidis 1997):

$$\mathrm{E}\left[\left(\hat{z}(\mathbf{x_0}) - z(\mathbf{x_0})\right)^2\right] = \mathrm{E}\left[\left(\sum_{i=1}^{n} \lambda_i \varepsilon(\mathbf{x_i}) - \varepsilon(\mathbf{x_0})\right)^2\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j R(\mathbf{x_i}, \mathbf{x_j}) - 2\sum_{i=1}^{n} \lambda_i R(\mathbf{x_i}, \mathbf{x_0}) + R(\mathbf{x_0}, \mathbf{x_0}). \tag{3.31}$$

Thus, coefficients $\lambda_1, \ldots, \lambda_n$ are estimated by minimizing the expression in Eq. (3.31), which in turn depends on the $p$ linear constraints of the unbiasedness condition Eq. (3.30).

The minimum variance condition makes the Kriging method an optimization problem, with assigned constraints, that can be resolved by means of *Lagrangian multiplier method*, which has the general form:

$$\min f(\lambda_1, \ldots, \lambda_n),$$

$$g_k(\lambda_1, \ldots, \lambda_n) = b_k \quad k = 1, \ldots, p. \tag{3.32}$$

In this case:

$$f(\lambda_1, \ldots, \lambda_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j R(\mathbf{x_i}, \mathbf{x_j}) - 2\sum_{i=1}^{n} \lambda_i R(\mathbf{x_i}, \mathbf{x_0}) + R(\mathbf{x_0}, \mathbf{x_0}), \tag{3.33a}$$

$$g_k(\lambda_1,\ldots,\lambda_n)=\sum_{i=1}^{n}\lambda_i f_k(\mathbf{x_i}), \tag{3.33b}$$

$$b_k = f_k(\mathbf{x_0}). \tag{3.33c}$$

The *Lagrangian* (*L*) is:

$$L(\lambda_1,\ldots,\lambda_n,\nu_1,\ldots,\nu_p)= f(\lambda_1,\ldots,\lambda_n)+2\nu_1(g_1(\lambda_1,\ldots,\lambda_n)-b_1)+\ldots+2\nu_p(g_p(\lambda_1,\ldots,\lambda_n)-b_p), \tag{3.34a}$$

where $\nu_1,\ldots,\nu_p$ are parameters called *Lagrange multipliers*, and number 2 is used only for convenience.

By taking derivatives of *L* with respect to $\lambda_1,\ldots,\lambda_n$ and setting them to zero:

$$\frac{\partial f}{\partial\lambda_i}+2\nu_i\frac{\partial g_i}{\partial\lambda_i}+\ldots+2\nu_p\frac{\partial g_p}{\partial\lambda_i}=0 \quad i=1,\ldots,n, \tag{3.34b}$$

the following set of *n* equations is obtained:

$$\sum_{j=1}^{n}R(\mathbf{x_i},\mathbf{x_j})\lambda_j+\sum_{k=1}^{p}f_k(\mathbf{x_i})\nu_k=R(\mathbf{x_i},\mathbf{x_0}) \quad i=1,\ldots,n. \tag{3.35}$$

Eqs. (3.30) and (3.35) form the Kriging system which consists of $n + p$ linear equations with $n + p$ unknowns which can be easily resolved.

The Kriging system can be put also in a matrix notation. Let **x** be the vector of unknowns

$$\mathbf{x}=\begin{pmatrix}\lambda_1\\ \ldots\\ \lambda_n\\ \nu_1\\ \ldots\\ \nu_p\end{pmatrix}, \tag{3.36a}$$

**b** the right-hand side vector

$$\mathbf{b}=\begin{pmatrix}R(\mathbf{x_1},\mathbf{x_0})\\ \ldots\\ R(\mathbf{x_n},\mathbf{x_0})\end{pmatrix}, \tag{3.36b}$$

and **A** the matrix of coefficients

$$\mathbf{A}=\begin{pmatrix}R(\mathbf{x_1},\mathbf{x_1}) & R(\mathbf{x_1},\mathbf{x_2}) & \ldots & R(\mathbf{x_1},\mathbf{x_n}) & f_1(\mathbf{x_1}) & f_2(\mathbf{x_1}) & \ldots & f_p(\mathbf{x_1})\\ R(\mathbf{x_2},\mathbf{x_1}) & R(\mathbf{x_2},\mathbf{x_2}) & \ldots & R(\mathbf{x_2},\mathbf{x_n}) & f_1(\mathbf{x_2}) & f_2(\mathbf{x_2}) & \ldots & f_p(\mathbf{x_2})\\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots\\ R(\mathbf{x_n},\mathbf{x_1}) & R(\mathbf{x_n},\mathbf{x_2}) & \ldots & R(\mathbf{x_n},\mathbf{x_n}) & f_1(\mathbf{x_n}) & f_2(\mathbf{x_n}) & \ldots & f_p(\mathbf{x_n})\end{pmatrix}. \tag{3.36c}$$

Hence, the Kriging system can be written as:

$$\mathbf{Ax}=\mathbf{b}. \tag{3.36d}$$

## Summary

Topics covered by this chapter are the core of the IRI UP method, of which two different implementations are described in chapters 4 and 5.

In particular, the importance of using effective ionospheric indices in the ionospheric modeling along with the need of spatially characterize their behaviour have been addressed. From this critical review emerges how important is to use ionospheric activity indices, instead of the commonly used solar activity indices, to describe very fast and small-scale variations occurring in the ionosphere. The introduction of the ionospheric activity indices $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ responds to this need. Because they are calculated by means of actual assimilated ionospheric data from spatially scattered ionospheric stations, they are able to catch such very fast and small-scale ionospheric variations, much better than global solar activity indices usually used by empirical climatological ionospheric models.

# 4. On the assimilation of ionosonde derived F2 layer ionospheric characteristics with the IRI UP method

In this chapter, a first implementation of the IRI UP method (described in chapter 3) is proposed and statistically evaluated. In particular, *fo*F2 and *M*(3000)F2 values collected by a European ionosonde network, are assimilated to calculate $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ indices at each assimilated station (as described in section 3.2). From these point values of $IG_{12\text{eff}}$ and $R_{12\text{eff}}$, corresponding maps are generated by means of the Universal Kriging method (as described in section 3.3); these maps are then used as input to the IRI model, thus obtaining updated values of *fo*F2 and *M*(3000)F2, hence of *hm*F2.

In section 4.1, the assimilation procedure of ionosonde derived F2 layer characteristics is carefully described. Each step, from the collection and filtering of ionosonde measurements (section 4.1.1), passing through the effective indices calculation and their mapping by means of the Universal Kriging method (section 4.1.2), the choice of the best variogram (section 4.1.3), and, finally, the ingestion of effective indices maps in IRI to produce updated F2 layer characteristics (section 4.1.4), is described and graphically explained for a definite period taken as example. Then, the procedure described in section 4.1 is tested on 30 periods embedding geomagnetic storms occurred between 2004 and 2016. Firstly (section 4.2), the IRI UP performance for the very disturbed period including the St. Patrick geomagnetic storm occurred in 2015 (the most intense storm occurred in the current solar cycle) is evaluated by comparing its output with that given by both the IRI model, with the STORM option on, and the IRTAM (IRI Real Time Assimilative Model, Galkin et al. 2012) model. Secondly (section 4.3), this validation is extended on all the 30 selected geomagnetic storms to highlight IRI UP performance for different conditions. The ability of the variogram selection procedure in picking the best variogram model used to produce the effective indices maps, is also evaluated in section 4.3.

Results described in this chapter are published in:

*Pignalberi, A., Pezzopane, M., Rizzi, R., Galkin, I. (2018) Effective solar indices for ionospheric modeling: A review and a proposal for a real-time regional IRI. Survey Geophys. 39(1), 125-167. doi:10.1007/s10712-017-9438-y.*

and :

*Pignalberi, A., Pietrella, M., Pezzopane, M., Rizzi, R. (2018) Improvements and validation of the IRI UP method under moderate, strong, and severe geomagnetic storms. Earth, Planets and Space, 70(180). https://doi.org/10.1186/s40623-018-0952-z*

## 4.1 Assimilation procedure of ionosonde derived F2 layer ionospheric characteristics in the IRI model: the IRI UP method

In order to calculate the $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ indices, by means of the procedure described in chapter 3.2, values of *fo*F2 and *M*(3000)F2 measured by a network of ionosondes are used. These data are downloaded from the DIDBASE (Digital Ionogram DataBASE) repository (Reinisch and Galkin 2011) by means of the SAO Explorer software, developed by the University of Massachusetts, Lowell. Fourteen ionospheric stations in the European sector, from 15°E to 45°W in longitude and from 30°N to 60°N in latitude, are considered (see Fig. 4.1). Table 4.1 provides an overview of the ionospheric stations considered, highlighting the ionosonde type, the autoscaling software, and the state of the station. The state "*assimilated*" means that *fo*F2 and *M*(3000)F2 hourly values are assimilated to update the background model, while the state "*used as test site*" indicates that *fo*F2 and *M*(3000)F2 hourly values are used only to test the IRI UP performance, which means that they are not assimilated. Data from Rome and Gibilmanna are autoscaled from the ionograms recorded by an AIS-INGV ionosonde (Zuccheretti et al. 2003), and those from Warsaw are autoscaled from the ionograms recorded by a VISRC2 ionosonde (Pezzopane et al. 2009); in both cases it is applied the Adaptive Ionospheric Profiler (AIP) (Scotto 2009; Scotto et al. 2012), an integral part of Autoscala (Pezzopane and Scotto 2005, 2007). The Automatic Real-Time Ionogram Scaler with True height (ARTIST, Galkin et al. 2008) system is instead applied to autoscale ionospheric characteristics from the ionograms recorded by digisondes (Bibl and Reinisch 1978) installed in the remaining ionospheric stations.

Two of these stations, specifically Fairford and San Vito, are used only as test sites to verify the model performance. The two test stations are well spaced in latitude, to test the model performance at different latitudes. Furthermore, the Fairford station is located near Chilton station, while San Vito station is relatively isolated and surrounded by Rome, Gibilmanna and Athens stations; therefore, they experiment different boundary conditions.

| **Ionospheric stations** | **Lat (degrees)** | **Lon (degrees)** | **Ionosonde type** | **Autoscaling software** | **State** |
|---|---|---|---|---|---|
| Athens (Ath) | 38.0° N | 23.5° E | Digisonde DPS-4D | ARTIST 5 | assimilated |
| Chilton (Chi) | 51.5° N | 0.6° W | Digisonde DPS-1 | ARTIST 4 | assimilated |
| Dourbes (Dou) | 50.1° N | 4.6° E | Digisonde DPS-4D | ARTIST 5 | assimilated |
| El Arenosillo (ElA) | 37.1° N | 6.7° W | Digisonde DPS-4D | ARTIST 5 | assimilated |
| **Fairford (Fai)** | 51.7° N | 1.5° W | Digisonde DPS-4D | ARTIST 5 | **used as test site** |
| Gibilmanna (Gib) | 37.9° N | 14.0° E | AIS-INGV | Autoscala 4.1 | assimilated |
| Juliusruh (Jul) | 54.6° N | 13.4° E | Digisonde DPS-4D | ARTIST 5 | assimilated |
| Moscow (Mos) | 55.5° N | 37.3° E | Digisonde DPS-4 | ARTIST 5 | assimilated |
| Nicosia (Nic) | 35.0° N | 33.2° E | Digisonde DPS-4D | ARTIST 5 | assimilated |
| Pruhonice (Pru) | 50.0° N | 14.6° E | Digisonde DPS-4D | ARTIST 5 | assimilated |
| Rome (Rom) | 41.8° N | 12.5° E | AIS-INGV | Autoscala 4.1 | assimilated |
| Roquetes (Roq) | 40.8° N | 0.5° E | Digisonde DPS-4D | ARTIST 5 | assimilated |
| **San Vito (SaV)** | 40.6° N | 17.8° E | Digisonde DPS-4D | ARTIST 5 | **used as test site** |
| Warsaw (War) | 52.2° N | 21.1° E | VISRC2 | Autoscala 4.1 | assimilated |

**Table 4.1:** List of European ionospheric stations and their geographical coordinates. The ionosonde type, the autoscaling software, and the state of the station are also highlighted. In bold the stations used as test site.

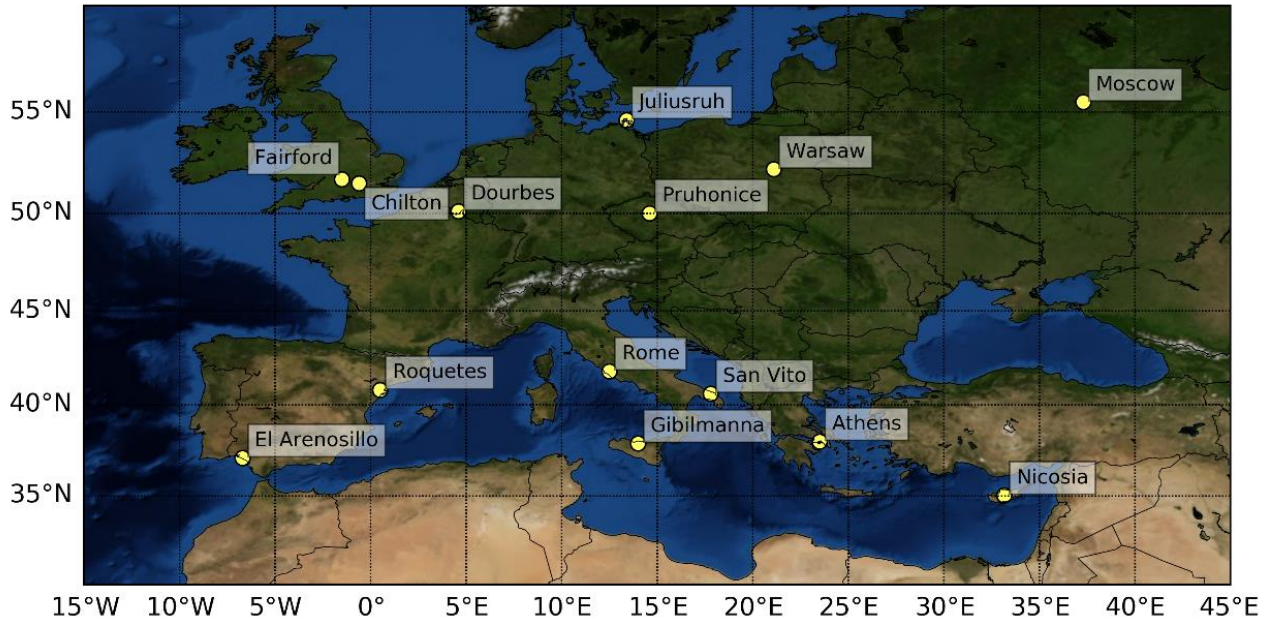


**Figure 4.1:** Geographical distribution of the ionospheric stations listed in Table 4.1.

### 4.1.1 Filtering out wrongly auto-scaled F2 layer characteristics

When assimilating measured data a particular attention should be given to avoid the ingestion of bad data (outliers, spikes, data affected by excessive calibration errors, etc.), which could affect the overall performance of the method. This is particularly true for the IRI UP method, being based on the Universal Kriging geostatistical interpolation method, which performs a global weighted mean of the assimilated quantities, as highlighted by Eq. (3.3).

This is why IRI UP implements a filter (called *F* algorithm) able to automatically discard those ionosonde data clearly wrong (spikes) which, once assimilated, would affect negatively the ionospheric modeling. The *F* algorithm has a dual objective:

1) to improve the level of reliability of assimilated data;
2) to speed up the selection process of data which are going to be assimilated.

The way the *F* algorithm works is as follows. In order to understand whether at a given hour *h* a value of *fo*F2 (*fo*F2$_h$) or *M*(3000)F2 (*M*(3000)F2$_h$) has been correctly autoscaled in a definite ionosonde station, *fo*F2 and *M*(3000)F2 data autoscaled in the previous 15 days, at the same considered hour, are taken into account to calculate the mean values $\bar{m}_{foF2,h,15prevdays}$ and $\bar{m}_{M(3000)F2,h,15prevdays}$, and the corresponding standard deviations $sd_{foF2,h,15prevdays}$ and $sd_{M(3000)F2,h,15prevdays}$. Specifically, standard deviations are calculated when time series have a number *N* of available data greater than 5, otherwise they are fixed to 0.5 MHz and 0.15 for *fo*F2 and *M*(3000)F2, respectively.

Subsequently, if $sd_{foF2,h,15prevdays} \geq 0.5$ MHz, *fo*F2$_h$ is considered reliable if the following inequalities are fulfilled:

$$\bar{m}_{foF2,h,15prevdays} - 5\, sd_{foF2,h,15prevdays} \leq foF2_h \leq sd_{foF2,h,15prevdays} + 5\, sd_{foF2,h,15prevdays}; \quad (4.1)$$

otherwise, if $sd_{foF2,h,15prevdays}$ < 0.5 MHz, *fo*F2$_h$ is considered reliable if the following inequalities are fulfilled:

$$\bar{m}_{foF2,h,15\,prevdays} - 5\cdot(0.5) \leq foF2_h \leq \bar{m}_{foF2,h,15\,prevdays} + 5\cdot(0.5)\,. \quad (4.2)$$

The inequalities (4.2) are considered because it may happen that when very quiet days occur, the previous 15 days are characterized by *fo*F2 values very close to each other; in these cases, the value of $sd_{foF2,h,15prevdays}$ would be too small and, consequently, the inequalities (4.1) would be too selective.

The values of *fo*F2$_h$ considered reliable are then used to calculate $IG_{12eff}$ values at the hour *h* in the corresponding assimilated ionosonde station (as described in chapter 3).

Likewise, if $sd_{h,M(3000)F2,15prevdays} \geq 0.15$, *M*(3000)F2$_h$ is considered reliable if the following inequalities are fulfilled:

$$\bar{m}_{M(3000)F2,h,15prevdays} - 5\,sd_{M(3000)F2,h,15prevdays} \leq M(3000)F2_h \leq \bar{m}_{M(3000)F2,h,15prevdays} + 5\,sd_{M(3000)F2,h,15prevdays}\,; \quad (4.3)$$

otherwise, if $sd_{h,M(3000)F2,15prevdays} < 0.15$, *M*(3000)F2$_h$ is considered reliable if the following inequalities are fulfilled:

$$\bar{m}_{M(3000)F2,h,15prevdays} - 5\cdot(0.15) \leq foF2_h \leq \bar{m}_{M(3000)F2,h,15\,prevdays} + 5\cdot(0.15)\,. \quad (4.4)$$

The values of *M*(3000)F2$_h$ considered reliable are then used to calculate $R_{12eff}$ in the corresponding assimilated ionosonde station (as described in chapter 3).

The aim of the proposed *F* algorithm is to remove spikes, that is values of *fo*F2 and *M*(3000)F2 which are evidently wrongly autoscaled. Fig. 4.2 shows an example of the effectiveness of the "filter" applied on *fo*F2 and *M*(3000)F2 data recorded at Fairford over the period 24-29 October 2016.

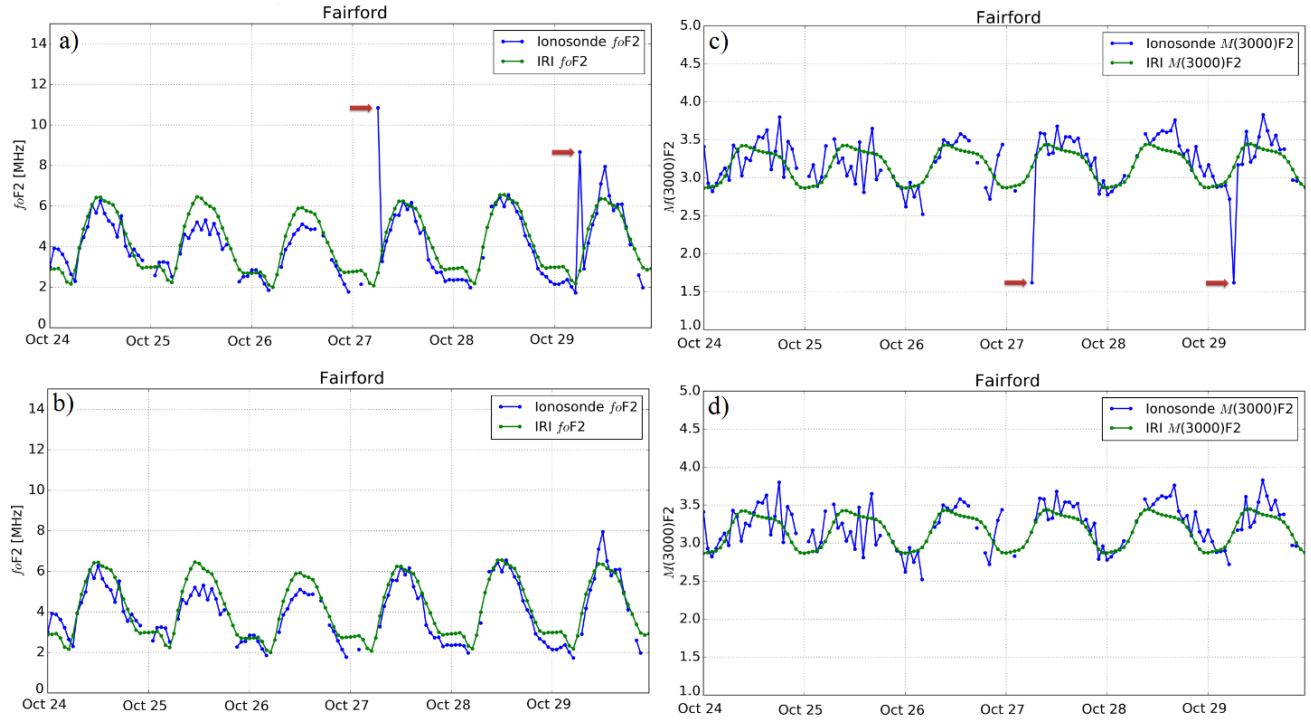


**Figure 4.2:** *fo*F2 and *M*(3000)F2 time series for the period 24 - 29 October 2016, before (panel a) and c), respectively) and after (panel b) and d), respectively) applying the *F* algorithm. Red arrows indicate spikes which are discarded by the *F* algorithm.

The *F* algorithm described by Eqs. (4.1)-(4.4) has proven to be very effective in discarding ionosonde data which, once assimilated, would affect badly the modeling of $IG_{12eff}$ and $R_{12eff}$, leading to unrealistic *fo*F2 and *M*(3000)F2 maps and, consequently, to unlikely *hm*F2 maps. It must be noted that considering five times the standard deviation and thresholds values equal to 0.5 MHz, for *fo*F2, and 0.15, for *M*(3000)F2, are subjective choices, aiming to remove especially those measurements which are clearly out of range (spikes), as shown in Fig. 4.2.

### 4.1.2 Effective indices maps calculation

For a definite moment of time, all the available *fo*F2 and *M*(3000)F2 values from assimilated ionosonde stations listed in Table 4.1 are used to calculate $IG_{12eff}$ and $R_{12eff}$ indices by using the procedure described in chapter 3.2. For example, for the 17 March 2015 (day 76) at 11 Universal Time (UT) (the main phase day of the St. Patrick geomagnetic storm occurred in 2015), 10 ionosonde stations were available to be assimilated (for that particular moment, Athens and Nicosia stations were not available). Assimilated and test *fo*F2, *M*(3000)F2, and *hm*F2 values are listed in Table 4.2.

| **Ionospheric stations** | ***fo*F2 [MHz]** | ***M*(3000)F2** | ***hm*F2 [km]** | **State** |
|---|---|---|---|---|
| Athens | Values not available | | | assimilated |
| Chilton | 9.575 | 2.623 | 333.0 | assimilated |
| Dourbes | 10.100 | 2.592 | 350.9 | assimilated |
| El Arenosillo | 10.688 | 2.703 | 330.4 | assimilated |
| **Fairford** | **9.700** | **2.570** | **353.3** | **used as test site** |
| Gibilmanna | 11.100 | 2.597 | 341.0 | assimilated |
| Juliusruh | 9.938 | 2.636 | 333.6 | assimilated |
| Moscow | 11.625 | 2.915 | 302.6 | assimilated |
| Nicosia | Values not available | | | assimilated |
| Pruhonice | 10.775 | 2.646 | 352.1 | assimilated |
| Rome | 10.800 | 2.577 | 344.0 | assimilated |
| Roquetes | 10.725 | 2.535 | 356.4 | assimilated |
| **San Vito** | **11.075** | **2.625** | **338.3** | **used as test site** |
| Warsaw | 10.600 | 2.637 | 369.0 | assimilated |

**Table 4.2:** List of ionospheric stations used for the 17 March 2015 at 11 UT example. For each station, measured values of *fo*F2, *M*(3000)F2, and *hm*F2 (where available) are reported. In bold the stations used as test site.

Fig. 4.3 shows maps of *fo*F2 (top panel) and *M*(3000)F2 (bottom panel) assimilated values, with a $0.1° \times 0.1°$ of spatial resolution, obtained applying the Universal Kriging method directly on assimilated *fo*F2 and *M*(3000)F2 values, for the 17 March 2015 at 11 UT. The spherical variogram model for *fo*F2 and the linear one for *M*(3000)F2 have been used. Black circles in Fig. 4.3 highlight the assimilated stations; the color inside them represent the actual assimilated value, as reported in Table 4.2.

In Fig.4.3, both *fo*F2 and *M*(3000)F2 exhibit large-scale gradients. In particular, *fo*F2 manifests both longitudinal and latitudinal gradients, with highest values on the south-east sector, the direction from where the solar terminator comes from at 11 UT in March. *M*(3000)F2 manifests mostly a longitudinal gradient with highest values in the east sector. Both *fo*F2 and *M*(3000)F2 assimilated values at ionosonde stations (black circles) are well reproduced.

These maps are shown here only to highlight the spatial variability embedded in assimilated *fo*F2 and *M*(3000)F2 values. In fact, $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ indices are calculated only on assimilated stations by using measured *fo*F2 and *M*(3000)F2 values (as described in section 3.2). These scattered effective indices are then interpolated by means of the Universal Kriging method (as described in section 3.3) to obtain effective indices maps.

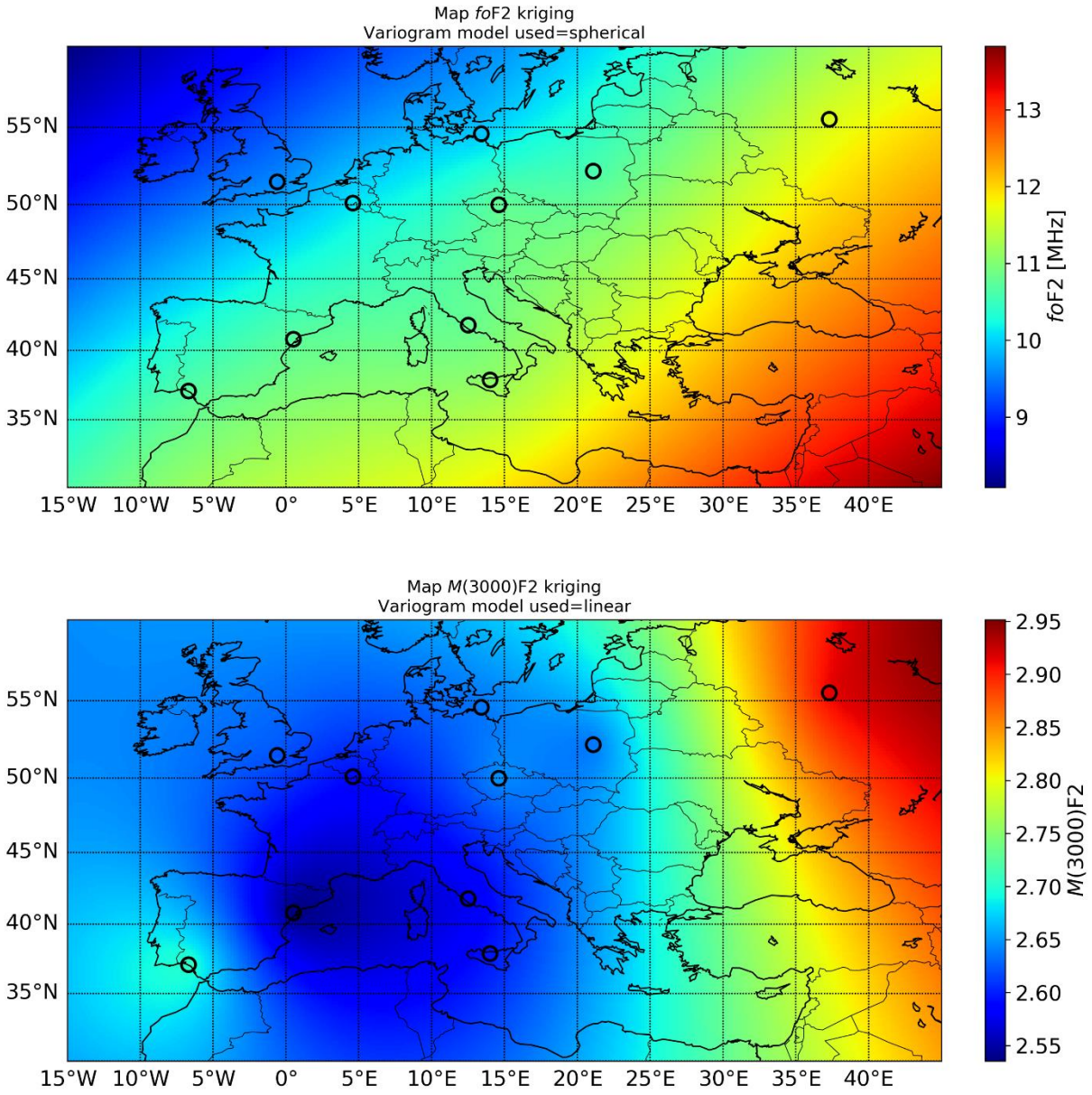


**Figure 4.3:** (top panel) *fo*F2 and (bottom panel) *M*(3000)F2 maps for the 17 March 2015 at 11 UT. They were generated by using the spherical variogram model for *fo*F2 and the linear variogram model for *M*(3000)F2 . Black circles highlight the assimilated stations, and the color inside them represent the actual assimilated value.

| **Ionospheric stations** | $IG_{12\text{eff}}$ | $R_{12\text{eff}}$ |
|---|---|---|
| Athens | Values not available | |
| Chilton | 113 | 209 |
| Dourbes | 117 | 212 |
| El Arenosillo | 101 | 216 |
| Gibilmanna | 101 | 201 |
| Juliusruh | 123 | 187 |
| Moscow | 155 | 103 |
| Nicosia | Values not available | |
| Pruhonice | 129 | 182 |
| Rome | 105 | 208 |
| Roquetes | 106 | 249 |
| Warsaw | 126 | 180 |

**Table 4.3:** $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ indices calculated by means of measured *fo*F2 and *M*(3000)F2 values of Table 4.2 for the 17 March 2015 at 11 UT.

$IG_{12\text{eff}}$ and $R_{12\text{eff}}$ indices are calculated at each assimilated station making use of *fo*F2 and *M*(3000)F2 values listed in Table 4.2, by means of the procedure described in section 3.2. They are listed in Table 4.3. Once $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ indices have been calculated, the variogram plot can be built by calculating the experimental semivariance values (Eqs. (3.6)-(3.8)) for each couple of assimilated stations. Red points in the variogram plots of Fig. 4.4 represent the calculated semivariance values. Solid blue lines in Fig. 4.4 represent the fitted variogram model, spherical for $IG_{12\text{eff}}$ (top panel), linear for $R_{12\text{eff}}$ (bottom panel).

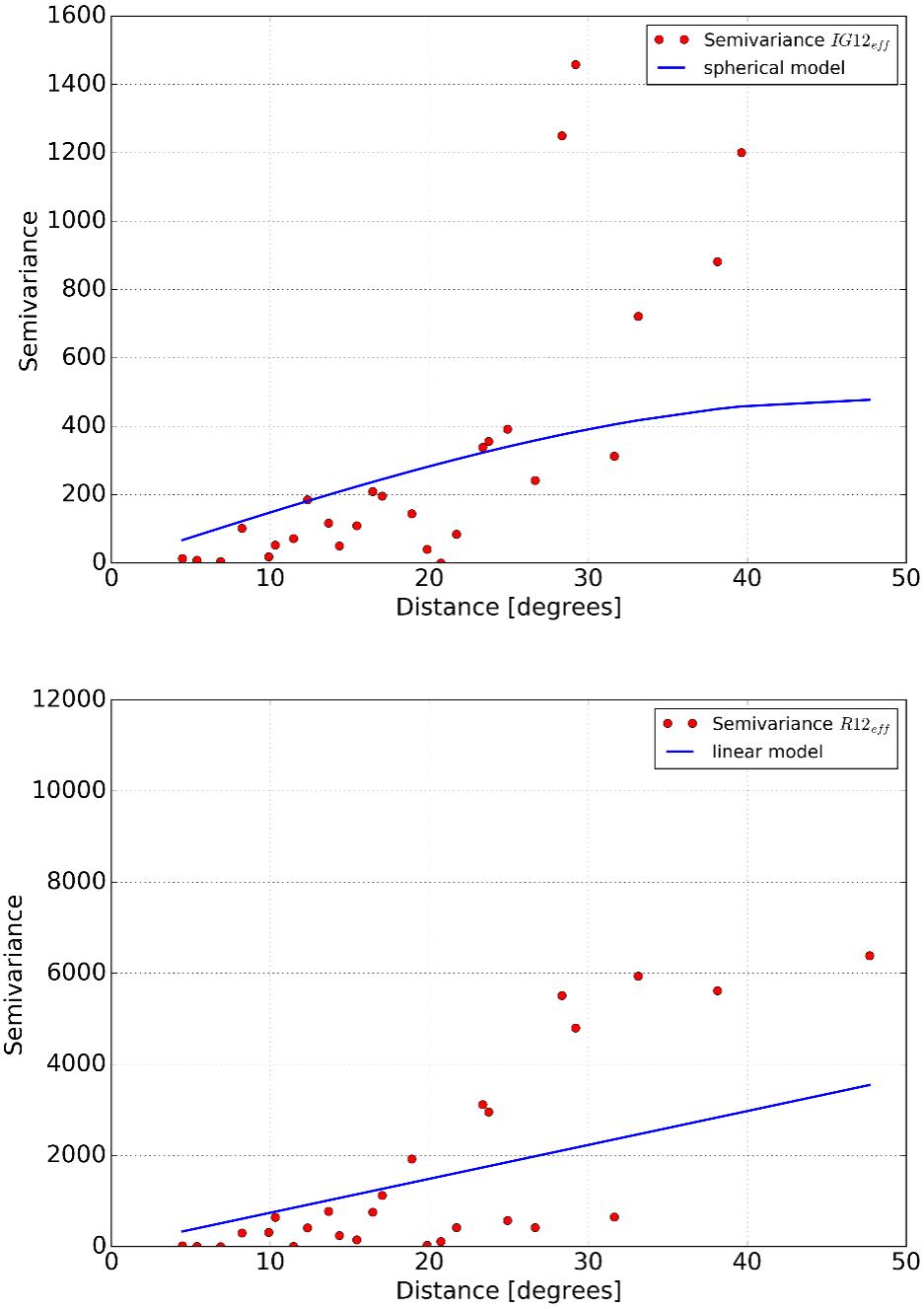


**Figure 4.4:** (top panel) $IG_{12\text{eff}}$ and (bottom panel) $R_{12\text{eff}}$ variogram plot for the 17 March 2015 at 11 UT. Red points depict semivariance values. The spherical variogram model is used for $IG_{12\text{eff}}$, and the linear variogram model for $R_{12\text{eff}}$ (solid blue curves).

Fig. 4.5 shows an example of $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ maps (with a $0.1° \times 0.1°$ of spatial resolution) for the 17 March 2015 at 11 UT, generated by using the spherical variogram model for $IG_{12\text{eff}}$ and the linear variogram model for $R_{12\text{eff}}$, both shown in Fig. 4.4. Calculated $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ indices, as in Table 4.3, for each available assimilated station are shown inside the black circles of Fig. 4.5.

Both maps of Fig. 4.5 show important spatial gradients with very extended magnitude scales which reflect the assimilated spatial variability. It is worth noting that, in Fig. 4.5, maxima and minima of $IG_{12\mathrm{eff}}$ and $R_{12\mathrm{eff}}$ are opposite; this is due to the fact that, for that moment, the highest values of *fo*F2 and *M*(3000)F2 are recorded at Moscow and Warsaw, and the dependence between *fo*F2 and $IG_{12}$ is direct, while that between *M*(3000)F2 and $R_{12}$ is inverse.

Of course, to catch small-scale spatial structures, the ionospheric assimilated stations distribution is fundamental for two different reasons:

1) having many stations at very small distance allows to populate the part of the experimental variogram near the origin, which is essential to describe the small-scale spatial behavior of the represented characteristic;
2) to increase the goodness of the fitted variogram model. The large-scale spatial distribution of assimilated stations is particularly important to spread over the map the information embedded in the variogram (even though this need is highly decreased using the Universal Kriging, whatever variogram model is used).

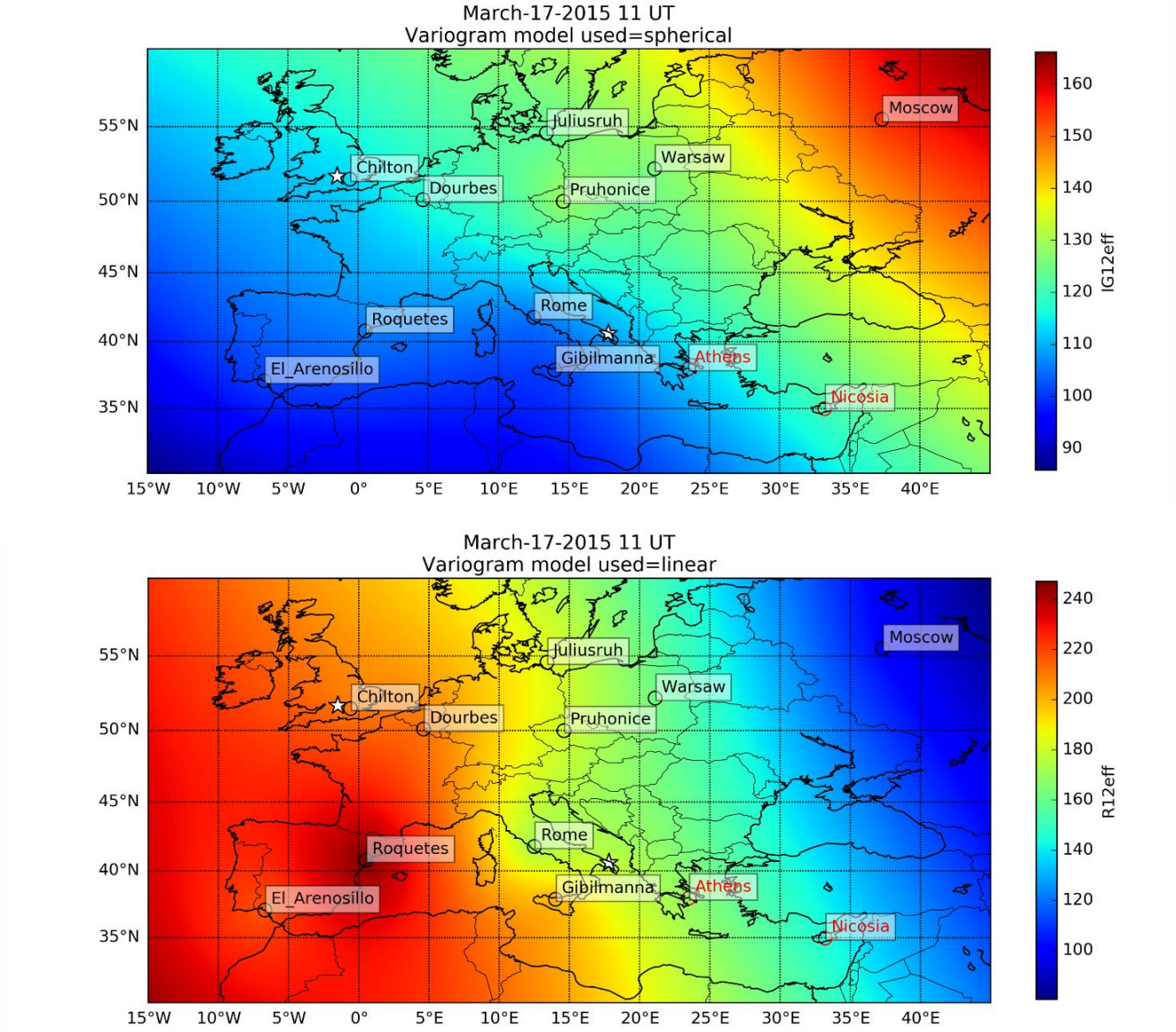


**Figure 4.5** (top panel) $IG_{12\text{eff}}$ and (bottom panel) $R_{12\text{eff}}$ maps for the 17 March 2015 at 11 UT. They were calculated on the basis of assimilated *fo*F2 and *M*(3000)F2 values of Table 4.2, and variogram plots of Fig.4.4. Black circles highlight the assimilated stations. Red circles highlight stations, included in the assimilated stations list, for which values were not available. The two white stars mark the position of the two test stations, Fairford in the upper left part and San Vito in the middle lower part of both panels.

### 4.1.3 On the choice of the best variogram model

Reliable *fo*F2 and *hm*F2 maps can be obtained when the IRI model is updated with a realistic $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ representation. Such representation is obtained on the base of a discrete set of $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ values, obtained at locations of selected assimilation stations, applying the Universal Kriging method. It must be pointed out that the Universal Kriging method performance strongly relies on the choice of the variogram model which, in principle,

should be the one that best fits the experimental variogram (i.e., the variogram built on the discrete set of $IG_{12eff}$ and $R_{12eff}$ values).

Therefore, the problem of how to choose among the possible variogram models, as described in section 3.3, is of crucial importance, because this choice greatly affects the goodness of the $IG_{12eff}$ and $R_{12eff}$ maps and, consequently, the capability of delivering an accurate and trustworthy *fo*F2 and *hm*F2 modeling.

Depending on the availability of stations and on the quality of assimilated data, some cases may occur where the variogram model is an approximately straight horizontal line. In this case, the semivariance does not vary with the distance, as it should be expected; therefore, these cases correspond to unrealistic situations which would lead to unrealistic maps of $IG_{12eff}$ and $R_{12eff}$. Examples of variogram models and related maps of $IG_{12eff}$ and $R_{12eff}$ which should be discarded by the IRI UP method are shown in Fig. 4.6.

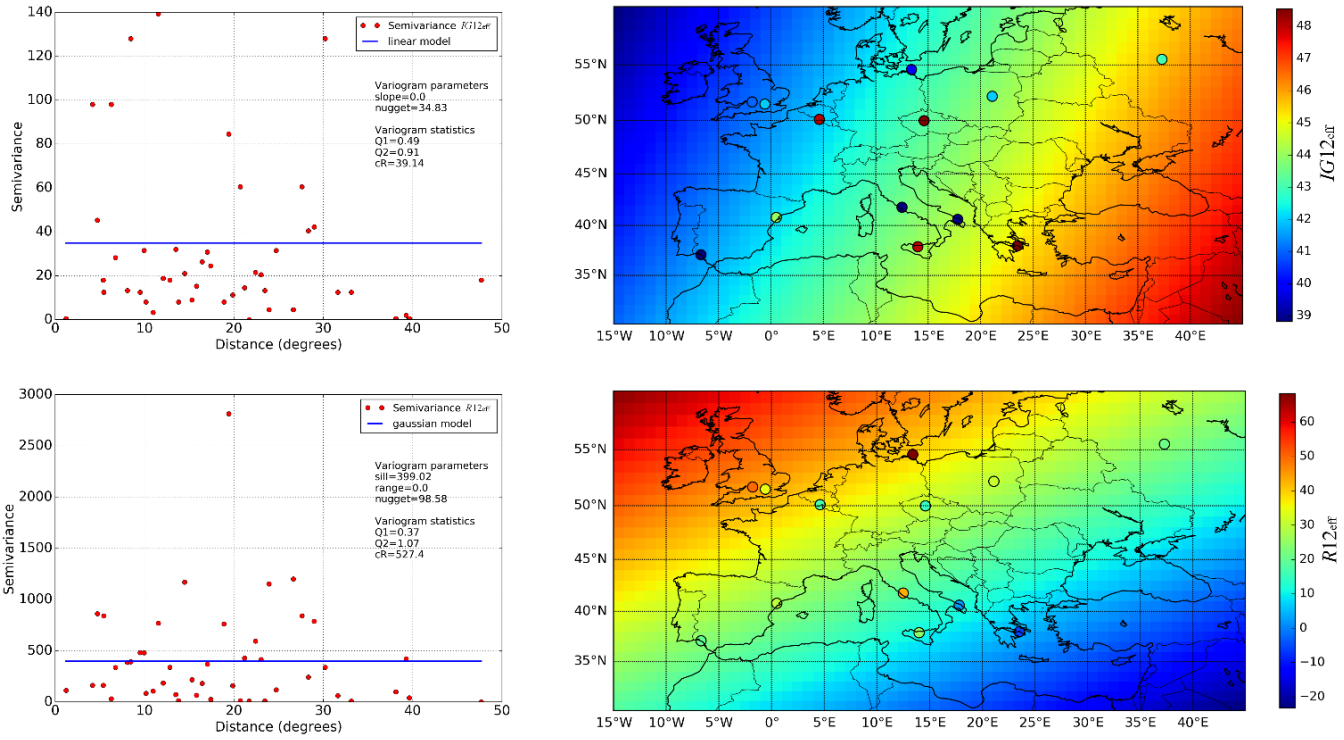


**Figure 4.6:** Examples of discarded variogram models (linear (top left), Gaussian (bottom left)), with corresponding maps of $IG_{12eff}$ (top right) and $R_{12eff}$ (bottom right), for the 20 March 2015 at 16UT.

A statistical procedure, based on statistical tests carried out using the statistical variables $Q_1$, $Q_2$, and $cR$ (Eqs. (3.15), (3.19), and (3.22)), built on the residuals between the observed (experimental variogram) and the modeled (variogram model) semivariance values, is implemented in the IRI UP method. This procedure allows to select objectively, and with a high acceptable degree of confidence, the "best" variogram model to be used.

The statistical criteria (3.17), (3.21), and (3.23), constitute the statistical procedure of the variogram model selection implemented in the IRI UP method.

For the selection of the "best" variogram model, one could be tempted to consider only the $cR$ criterion (3.23) leaving out the $Q_1$ (3.17) and $Q_2$ (3.21) criteria. Nevertheless, from a preliminary investigation conducted over a

large number of variogram models, it emerges that if only the *cR* criterion were applied, several variogram models which do not satisfy the $Q_1$ and/or the $Q_2$ criteria would be accepted.

For these reasons, the variogram model is selected through the iterative procedure depicted by the flowchart of Fig. 4.7. The figure shows the various steps carried out to select that variogram model which, to an acceptable degree of confidence, fits the data more reliably than the other ones.

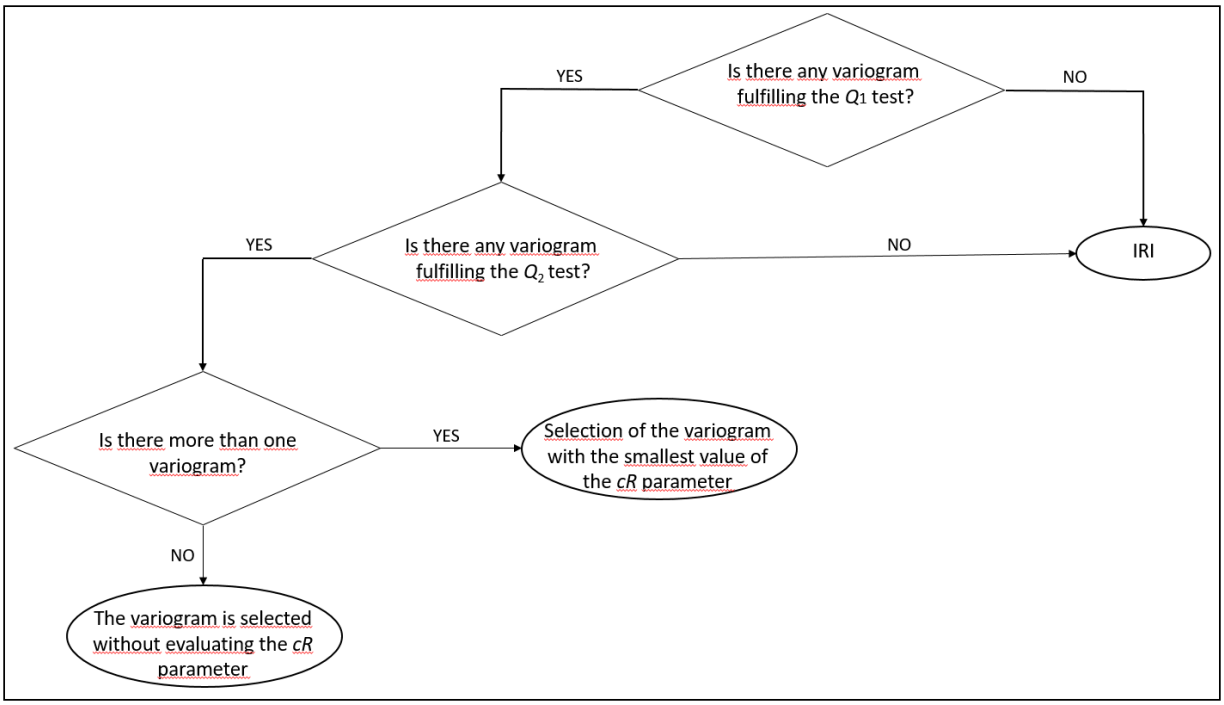


**Figure 4.7:** Flow chart representing the main steps of the statistical procedure for selecting the best variogram model.

When the distribution of semivariance values in the variogram plot is such that the experimental variogram cannot be adequately represented by any of the variogram models (3.10)-(3.14), one is forced to discard the variogram modeling because it would lead to unrealistic maps. In this case, the ionospheric representation made by the IRI model is kept.

Fig. 4.6 shows examples of discarded variogram models for $IG_{12eff}$ and $R_{12eff}$; in these cases the fitting function is an approximately straight horizontal line whatever is the considered variogram model, and the corresponding maps are not able to reproduce $IG_{12eff}$ and $R_{12eff}$ values over the assimilated stations. This is due to the very large nugget values ($c_0 = 34.8$ for $IG_{12eff}$ and $c_0 = 98.6$ for $R_{12eff}$) describing the micro-scale variability. This implies that coloured circles, marking the assimilated stations, are in contrast with the colours characterizing the regions around the assimilated stations, which means that the maps are not realistic since do not match the measured data.

On the other hand, looking at the statistical parameters related to these two experimental variograms (in Fig. 4.6), the associated values of $Q_1$ (0.49 for $IG_{12eff}$, 0.37 for $R_{12eff}$) and $Q_2$ (0.91 for $IG_{12eff}$, 1.07 for $R_{12eff}$) are such that the two experimental variograms cannot pass the $Q_1$ and $Q_2$ statistical test (Eqs. (3.17) and (3.21)), and hence they must be rejected along with their corresponding maps.

On the contrary, when the nugget effect is not so relevant and the distribution of semivariance values can be well described by most variogram models, the experimental variogram plot likely produce realistic maps. This is what

happens for the case shown in Fig. 4.8 where the null nugget effect ($c_0$ = 0 for $IG_{12eff}$ and $R_{12eff}$), and the associated values of $Q_1$ (0.22 for $IG_{12eff}$, 0.49 for $R_{12eff}$) and $Q_2$ (0.98 for $IG_{12eff}$, 0.76 for $R_{12eff}$) are such that the two experimental variogram plots pass both the $Q_1$ and the $Q_2$ statistical tests. In this case, the corresponding maps show $IG_{12eff}$ and $R_{12eff}$ values, over the stations, compatible with those of the regions close to the stations.

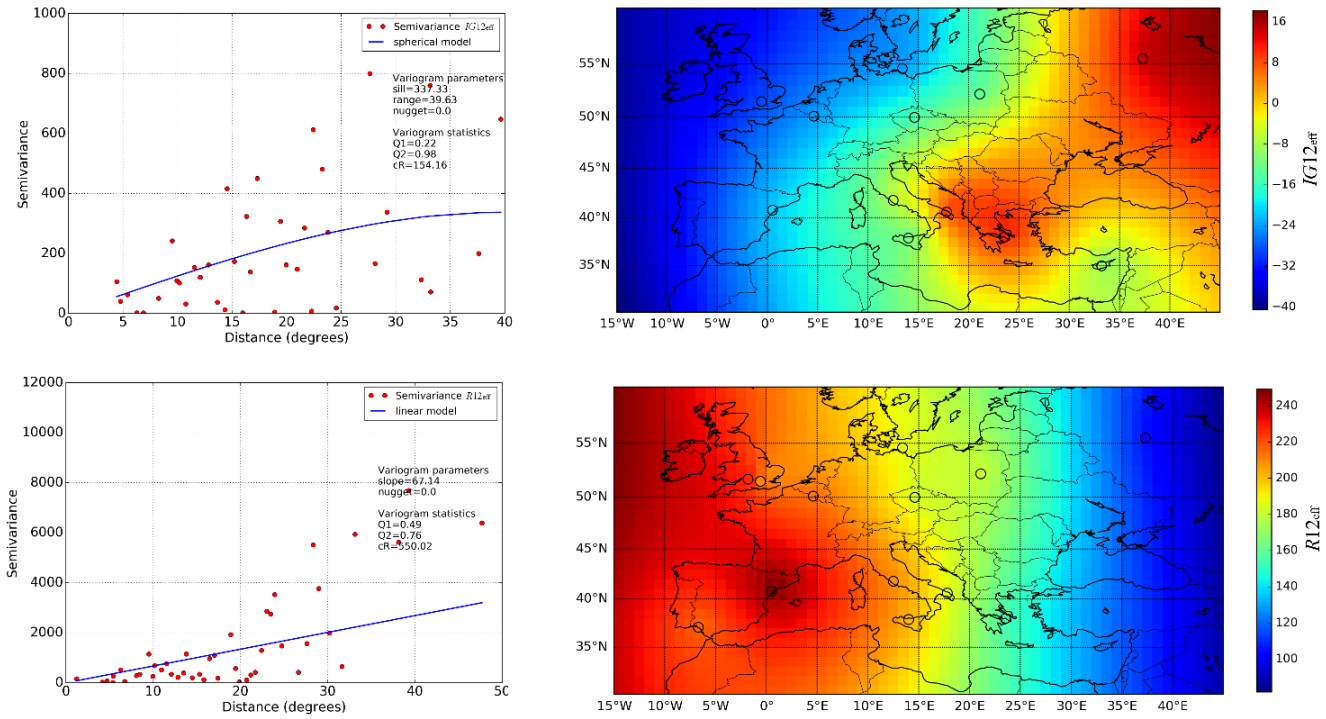


**Figure 4.8:** (top left) Example of spherical variogram model which has passed the statistical procedure for selecting the best variogram model, and (top right) corresponding map of $IG_{12eff}$ for the epoch 16 October 2016 at 12 UT. (bottom left) Example of linear variogram model which has passed the statistical procedure for selecting the best variogram model, and (bottom right) corresponding map of $R_{12eff}$ for the epoch 17 March 2015 at 11 UT.

### 4.1.4 Ingestion of Effective indices maps in the IRI model

Maps of effective indices as those depicted in Fig. 4.5 are ingested by the IRI model in order to get an updated representation of *fo*F2 (directly related to $IG_{12eff}$, Eq. (2.1)) and *M*(3000)F2 (directly related to $R_{12eff}$, Eq. (2.1)), and hence also *hm*F2 (Eq. (2.7)). Thus, for each grid point of the chosen map (whose limits are, in this case, from 15°E to 45°W in longitude and from 30°N to 60°N in latitude, with a 0.1°x 0.1° of spatial resolution), calculated $IG_{12eff}$ and $R_{12eff}$ values of Fig. 4.5 maps are ingested by the IRI model producing updated *fo*F2 and *hm*F2 values. Because the F2 layer peak is an anchor point for the IRI model, changing *fo*F2 and *hm*F2 has an effect on the whole vertical electron density profile. In particular, the F2 layer region (with its bottomside and topside region) is the part of the vertical electron density profile that undergoes the major changes, as evident in Fig. 4.11. This operation, being repeated for each grid point of the map, produces updated *fo*F2 (bottom panel of Fig. 4.9) and *hm*F2 (bottom panel of Fig. 4.10) maps, which are one of the IRI UP method output. In fact, also a three-dimensional matrix of updated electron density values is produced.

Fig. 4.9 shows *fo*F2 maps, for the 17 March 2015 at 11 UT, produced by the IRI model (namely by using $IG_{12}$ values), and produced by the IRI UP method after ingesting $IG_{12eff}$ map of Fig. 4.5. The figure shows that the *fo*F2

spatial representation made by IRI and IRI UP is quite different both in magnitude and concerning the direction of gradients. In fact, while IRI models an *fo*F2 pattern characterized by only a latitudinal gradient (that is common at 11 UT for this longitudes and latitudes), IRI UP exhibits also an important longitudinal gradient, as already evident from assimilated *fo*F2 values of Fig. 4.3, and from $IG_{12eff}$ values of Fig. 4.5. Measured *fo*F2 values at test stations turn out to be very similar to those modeled by the IRI UP method.

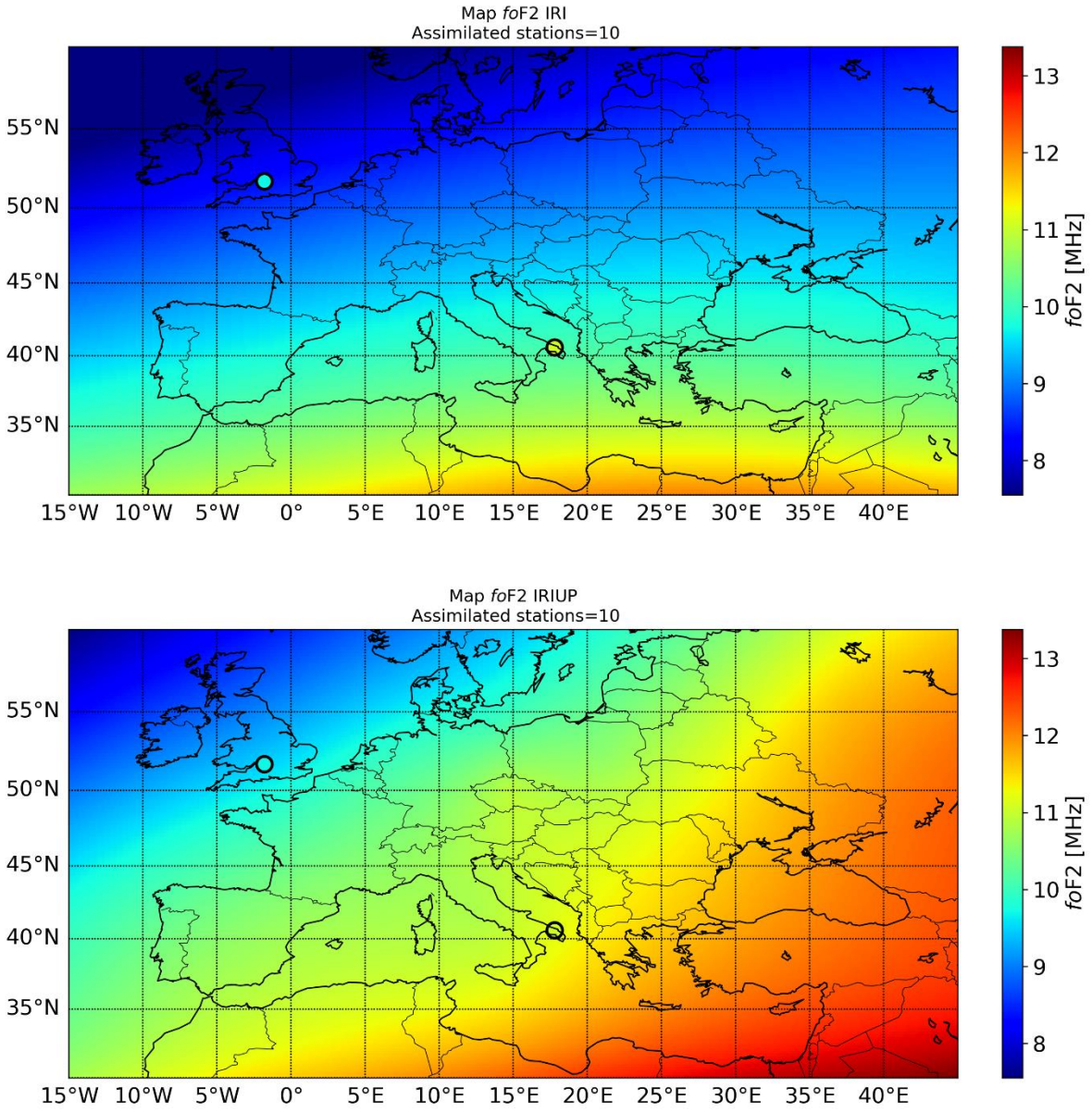


**Figure 4.9** (top panel) IRI and (bottom panel) IRI UP *fo*F2 maps for the 17 March 2015 at 11 UT. Black circles highlight the two test stations; the color inside them represent the actual measured value.

Fig. 4.10 shows *hm*F2 maps, for the 17 March 2015 at 11 UT, produced by the IRI model (namely by using $IG_{12}$ and $R_{12}$ values), and produced by the IRI UP method after ingesting $IG_{12eff}$ and $R_{12eff}$ maps of Fig. 4.5. The figure shows that, beyond the great difference in the *hm*F2 magnitude as modeled by IRI and IRI UP, very important

differences are evident in the *hm*F2 spatial characterization made by the two models. Longitudinal gradients are indeed opposite. Measured *hm*F2 values at test stations turn out to be more similar to those modeled by the IRI UP method than those modeled by the IRI model.

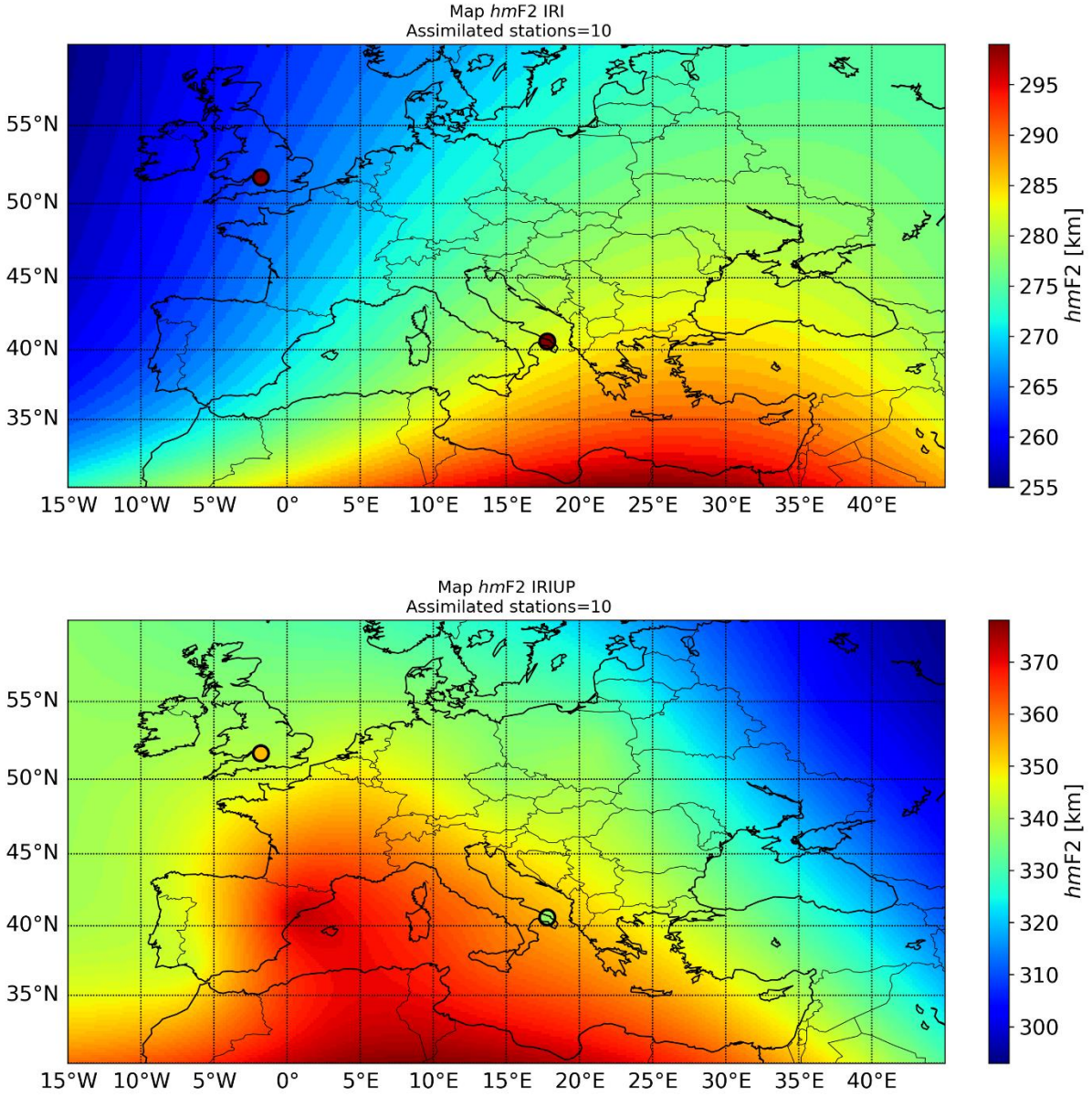


**Figure 4.10** (top panel) IRI and (bottom panel) IRI UP *hm*F2 maps for the 17 March 2015 at 11 UT. Black circles highlight the two test stations; the color inside them represent the actual measured value. The scale of the colorbars is different for the two maps because of the large differences between *hm*F2 values.

In order to better understand the differences between the IRI and IRI UP modeling of the ionosphere, the whole electron density profile is considered (Fig. 4.11) for the two test stations. In Fig. 4.11, *fo*F2 and *hm*F2 values for the test ionosonde stations are those shown by the color in the small circles of Figs. 4.9 and 4.10, while IRI and IRI UP *fo*F2 and *hm*F2 values are those extrapolated from Figs. 4.9 and 4.10 at test ionosonde locations. As already

said before, the ingestion of the effective indices in IRI produces a better representation of the F2 layer region, as it is evident looking at IRI UP profiles of Fig. 4.11. In fact, the ingestion of $IG_{12eff}$ updates the *fo*F2 value, while the ingestion of both $IG_{12eff}$ and $R_{12eff}$ updates the *hm*F2 value. In this way, the F2 layer peak (*fo*F2, *hm*F2), and the F2 bottomside and topside regions are more similar to those measured by ionosondes located at Fairford and San Vito.

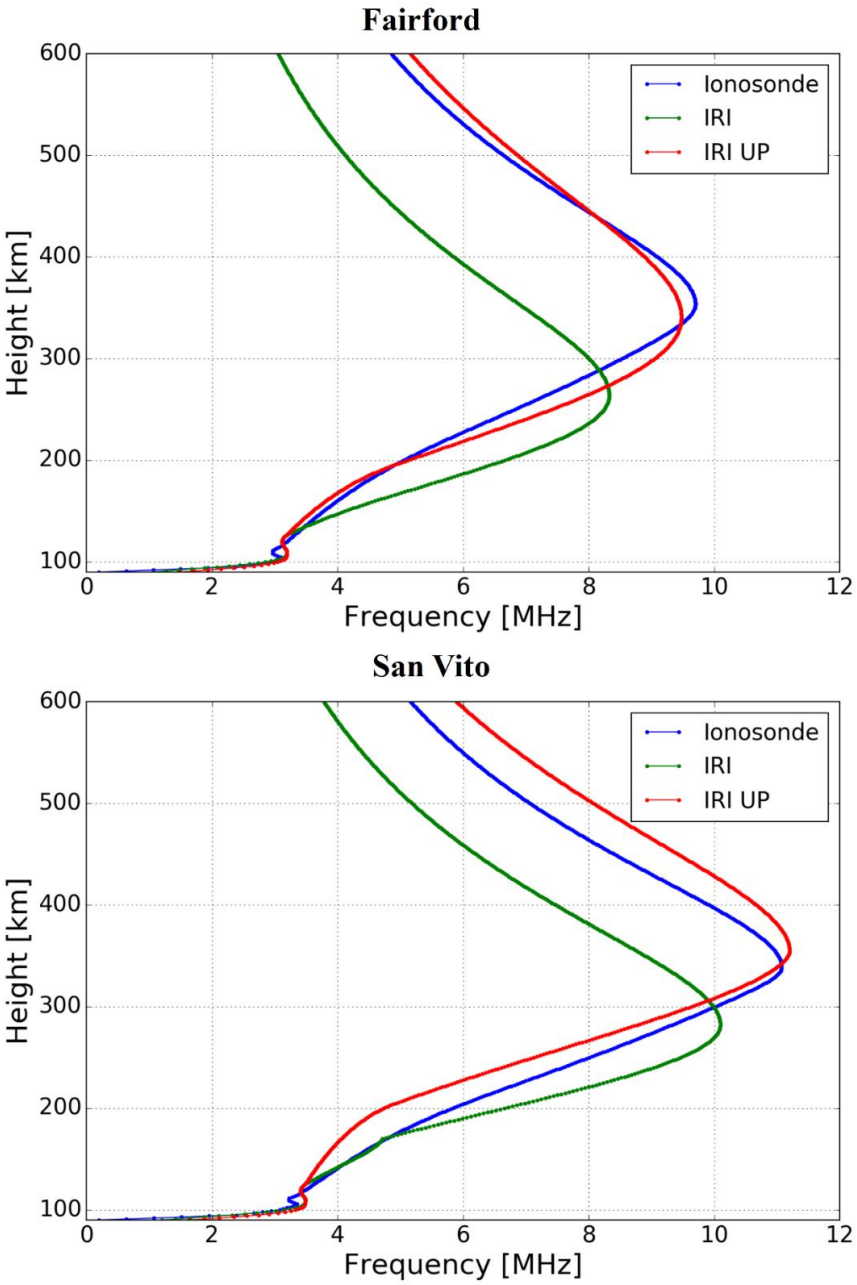


**Figure 4.11** (top panel) Fairford and (bottom panel) San Vito vertical electron density profiles for the 17 March 2015 at 11 UT. Blue lines refers to measured values, green lines to IRI modeled values, and red lines to IRI UP modeled values after ingesting effective indices maps.

## 4.2 A case study: the geomagnetic storm occurred on March 17th 2015

IRI UP method has been first validated during the geomagnetic storm occurred on March 17th 2015, comparing it with IRI and IRTAM (Galkin et al. 2012) models.

### 4.2.1 Data used and period under study

The considered period is the one between the March 9 2015 at 0 UT and the March 25 2015 at 23 UT, a time period including the so-called St. Patrick geomagnetic storm, the most intense geomagnetic storm observed during solar cycle 24, classified as severe and for which the $K_p$ index reached the maximum value of 8. The sudden storm commencement was observed from ground observatories on March 17 2015 at around 04:45 UT with the arrival at the Earth of a Coronal Mass Ejection. The storm reached its maximum intensity on March 17 at around 23:00 UT and it was characterized by a minimum value of the $D_{st}$ index of −223 nT. The considered time period was characterized initially by a magnetically quiet period, as it is shown by the $D_{st}$ and AE values close to zero (see Fig. 4.12), lasting until the sudden storm commencement occurred on March 17 2015 at about 05 UT. Then, a strongly magnetically disturbed period characterized the main phase of the storm, from the sudden storm commencement until the end of March 18 2015, when the $D_{st}$ index becomes deeply negative and the AE index increases significantly its value. Hence, a long recovery phase characterized the days until the end of the studied period, for which the $D_{st}$ index slowly increases to zero, while the AE index is still characterized by several perturbations due to the occurrence of sub-storms.

Hourly values of *fo*F2, *M*(3000)F2, and *hm*F2 are used in this validation. These data are downloaded from the DIDBASE (Digital Ionogram DataBASE) (Reinisch and Galkin 2011) by means of the SAO Explorer software, developed by the University of Massachusetts, Lowell. Fourteen ionospheric stations in the European sector, from 15°E to 45°W in longitude and from 30°N to 60°N in latitude, are considered (see Table 4.1).

### 4.2.2 Variogram statistics

For every hour of the studied period (March 9 2015 at 0 UT – March 25 2015 at 23 UT), *fo*F2 and *M*(3000)F2 values are assimilated from stations listed in Table 4.1 to calculate punctual $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ values. Then, the Universal Kriging method is used to interpolate these values over the chosen grid, with a 0.1°x0.1° resolution, using any of the five variogram models (3.10)-(3.14), in order to evaluate which of these gives the best results against *fo*F2, *M*(3000)F2, and *hm*F2 values recorded at the test stations. Modeled F2 layer characteristics values, calculated as described in section 4.1, are then compared with those measured at the test stations through the following statistical quantities (Root Mean Square Error (RMSE), Normalized Root Mean Square Error (NRMSE), Pearson correlation coefficient $\rho$, Mean delta, Standard Deviation delta):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(X_{\text{calculated},i} - X_{\text{ionosonde},i}\right)^2}{N}}, \tag{4.5}$$

where $X$ stands for *fo*F2, *M*(3000)F2, or *hm*F2; the subscript *calculated* refers to modeled values while the subscript *ionosonde* refers to values recorded by the ionosonde (the index $i$ runs on the $N$ hourly values of the time series);

$$\text{NRMSE} = \frac{\text{RMSE}\left(X_{\text{calculated}}, X_{\text{ionosonde}}\right)}{\overline{X_{\text{ionosonde}}}} \cdot 100, \tag{4.6}$$

where $\overline{X_{\text{ionosonde}}}$ is the arithmetic mean over time of $X_{\text{ionosonde}}$ ;

$$\rho = \frac{\text{cov}\left(X_{\text{calculated}}, X_{\text{ionosonde}}\right)}{\sigma_{X_{\text{calculated}}}\sigma_{X_{\text{ionosonde}}}} \in [-1,1], \tag{4.7}$$

where cov is the covariance and $\sigma$ the standard deviation;

$$\text{Mean delta} = \frac{\sum_{i=1}^{N}\left(X_{\text{calculated},i} - X_{\text{ionosonde},i}\right)}{N}\ , \tag{4.8}$$

$$\text{Standard Deviation delta} = \sqrt{\frac{\sum_{i=1}^{N}\left[\left(X_{\text{calculated},i} - X_{\text{ionosonde},i}\right) - \text{Mean delta}\left(X_{\text{calculated}} - X_{\text{ionosonde}}\right)\right]^2}{N-1}}\ . \tag{4.9}$$

In addition, also the following percentage of discarded maps is calculated:

$$\%\text{discarded} = \frac{\#\text{total maps} - \#\text{good maps}}{\#\text{total maps}}, \tag{4.10}$$

where #good maps is the cumulative number of $IG_{12\text{eff}}$ or $R_{12\text{eff}}$ maps which have passed both the $Q_1$ and $Q_2$ tests, and #total maps is the total potential number of maps. Thus, %discarded gives the percentage of maps not produced by the method, since the number of assimilated stations was lower than three (at least three measurement stations are needed to build the experimental variogram plot), or because discarded in accordance with the selection procedure based on $Q_1$ and $Q_2$ statistical parameters.

In Table 4.4 and 4.5 statistical parameters (4.5)-(4.10) are calculated for Fairford and San Vito ionospheric stations, for each of five variogram methods used, for *fo*F2 and *M*(3000)F2 respectively. In Table 4.6 the same statistical parameters are calculated for *hm*F2 using the spherical model for $IG_{12\text{eff}}$ and the linear model for $R_{12\text{eff}}$. Soon it will be clear why these two variogram models for $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ have been chosen in the *hm*F2 calculation.

| Station | Ionospheric characteristic | Variogram method | RMSE [MHz] | NRMSE [%] | Correlation Coefficient | Mean delta [MHz] | Standard Deviation delta [MHz] | discarded % |
|---|---|---|---|---|---|---|---|---|
| Fairford | *fo*F2 | linear | 0.26 | 3.97 | 0.99 | 0.04 | 0.27 | 6.37 |
| | | power | 0.26 | 3.92 | 0.99 | 0.03 | 0.27 | 6.37 |
| | | Gaussian | 0.27 | 4.12 | 0.99 | 0.02 | 0.28 | 6.37 |
| | | spherical | 0.26 | 3.90 | 0.99 | 0.02 | 0.27 | 6.37 |
| | | exponential | 0.26 | 3.93 | 0.99 | 0.02 | 0.27 | 6.37 |
| San Vito | *fo*F2 | linear | 0.38 | 5.17 | 0.99 | -0.08 | 0.39 | 6.37 |
| | | power | 0.39 | 5.27 | 0.99 | -0.09 | 0.40 | 6.37 |
| | | Gaussian | 0.37 | 5.00 | 0.99 | -0.07 | 0.38 | 6.37 |
| | | spherical | 0.37 | 5.02 | 0.99 | -0.08 | 0.38 | 6.37 |
| | | exponential | 0.37 | 5.03 | 0.99 | -0.08 | 0.38 | 6.37 |

**Table 4.4:** Statistical results, for both test stations, Fairford and San Vito, for *fo*F2 using each of five variogram models. The number of data involved in the stastics is *N*=408 (hourly values from March 9 2015 at 0 UT to March 25 2015 at 23 UT).

| Station | Ionospheric characteristic | Variogram method | RMSE | NRMSE % | Correlation Coefficient | Mean delta | Standard Deviation delta | discarded % |
|---|---|---|---|---|---|---|---|---|
| Fairford | *M*(3000)F2 | linear | 0.081 | 2.64 | 0.92 | 0.006 | 0.093 | 23.77 |
| | | power | 0.082 | 2.68 | 0.93 | 0.007 | 0.095 | 23.77 |
| | | Gaussian | 0.086 | 2.81 | 0.92 | 0.010 | 0.099 | 23.77 |
| | | spherical | 0.084 | 2.75 | 0.93 | 0.012 | 0.097 | 23.77 |
| | | exponential | 0.083 | 2.72 | 0.93 | 0.011 | 0.095 | 23.77 |
| San Vito | *M*(3000)F2 | linear | 0.098 | 3.18 | 0.89 | 0.051 | 0.102 | 23.77 |
| | | power | 0.101 | 3.28 | 0.88 | 0.054 | 0.104 | 23.77 |
| | | Gaussian | 0.108 | 3.48 | 0.87 | 0.058 | 0.110 | 23.77 |
| | | spherical | 0.103 | 3.35 | 0.88 | 0.056 | 0.106 | 23.77 |
| | | exponential | 0.102 | 3.30 | 0.88 | 0.054 | 0.104 | 23.77 |

**Table 4.5:** Same as Table 4.5, for *M*(3000)F2.

| Station | Ionospheric characteristic | Variogram method | RMSE [km] | NRMSE % | Correlation Coefficient | Mean delta [km] | Standard Deviation delta [km] | discarded % |
|---|---|---|---|---|---|---|---|---|
| Fairford | *hm*F2 | IG12eff = spherical<br>R12eff = linear | 11.57 | 3.96 | 0.94 | -0.44 | 13.78 | 28.68 |
| San Vito | *hm*F2 | IG12eff = spherical<br>R12eff = linear | 13.57 | 4.63 | 0.92 | -5.34 | 15.44 | 28.68 |

**Table 4.6:** Same as Table 4.5, for *hm*F2, using the spherical model for $IG_{12\text{eff}}$ and the linear model for $R_{12\text{eff}}$.

Tables 4.4 and 4.5 highlights that, for *fo*F2 and *M*(3000)F2, all five methods show little differences. A close inspection of results of Tables 4.4 and 4.5 leads, however, to the choice of the spherical model for *fo*F2 and the linear model for *M*(3000)F2 as the best variogram model. Thus, using the spherical method to obtain $IG_{12\text{eff}}$ maps, with which updated *fo*F2 values are obtained, and the linear method to obtain $R_{12\text{eff}}$ maps, with which updated *M*(3000)F2 values are obtained, updated *hm*F2 values are obtained. $R_{12\text{eff}}$ values are used also to update *fo*E values, that are used in the *hm*F2 calculation. As explained in section 3.3, each variogram model displays its own small-scale and large-scale behavior to which are linked some mathematical implications, which imply the choice of the proper experimental variogram fitting procedure. However, a thorough knowledge of the geophysical quantities that have to be interpolated can help in the variogram model choice. For example, the electron density spatial distribution at mid latitudes shows, on average, spatial gradients with a very large scale, ranging from higher values at mid-low latitudes to lower values at mid-high latitudes. With the use of the Universal Kriging method, this large-scale spatial component can be represented by the regional linear mean described by the drift coefficients of Eq. (3.25). This is why for *fo*F2 the spherical variogram model is chosen. For a more spatially

variable characteristic like *M*(3000)F2, things become harder. In this case, the choice of the best variogram model is guided by the overall statistical analysis of Table 4.5, without any a priori hypothesis. Hence, for *M*(3000)F2, by choosing the linear variogram model, the stochastic part contributes to the description of the *M*(3000)F2 large-scale spatial behavior.

The last column of Tables 4.4-4.6 shows the percentage of discarded maps and then the percentage of missing updated values. This value is small for *fo*F2 (6.37%) and becomes larger for *M*(3000)F2 (23.77%) and *hm*F2 (28.68%). The large value for *hm*F2 is simply due to the fact that it depends on both *fo*F2 and *M*(3000)F2 values. The difference between *fo*F2 and *M*(3000)F2 percentages is instead more difficult to explain; the reason is likely due to the intrinsic different spatial behavior between these two ionospheric characteristics, which makes the spatial description of *M*(3000)F2 more challenging.

### 4.2.3 $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ indices

Fig. 4.12 shows time series of $IG_{12\text{eff}}$, $IG_{12}$, $R_{12\text{eff}}$ and $R_{12}$, for Fairford and San Vito, respectively, along with those of geomagnetic indices $D_{st}$ and AE, which highlight the effects of the St. Patrick storm. The figure clearly shows that effective indices $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ describe short-temporal variations, for both quiet and disturbed conditions. The power of the method is however mainly evident in the main and recovery phases of the storm. In particular, the $IG_{12\text{eff}}$ time series correlates well with the $D_{st}$ index, a sign that *fo*F2 values, from which it derives, are significantly affected by storm induced effects; also oscillations in the recovery phase, seen especially at Fairford, tell us that also sub-storm induced effects are well described by this effective index. The pattern of variability shown by the $R_{12\text{eff}}$ time series is more irregular than that of $IG_{12\text{eff}}$, highlighting however a more definite variability during the main phase of the storm.

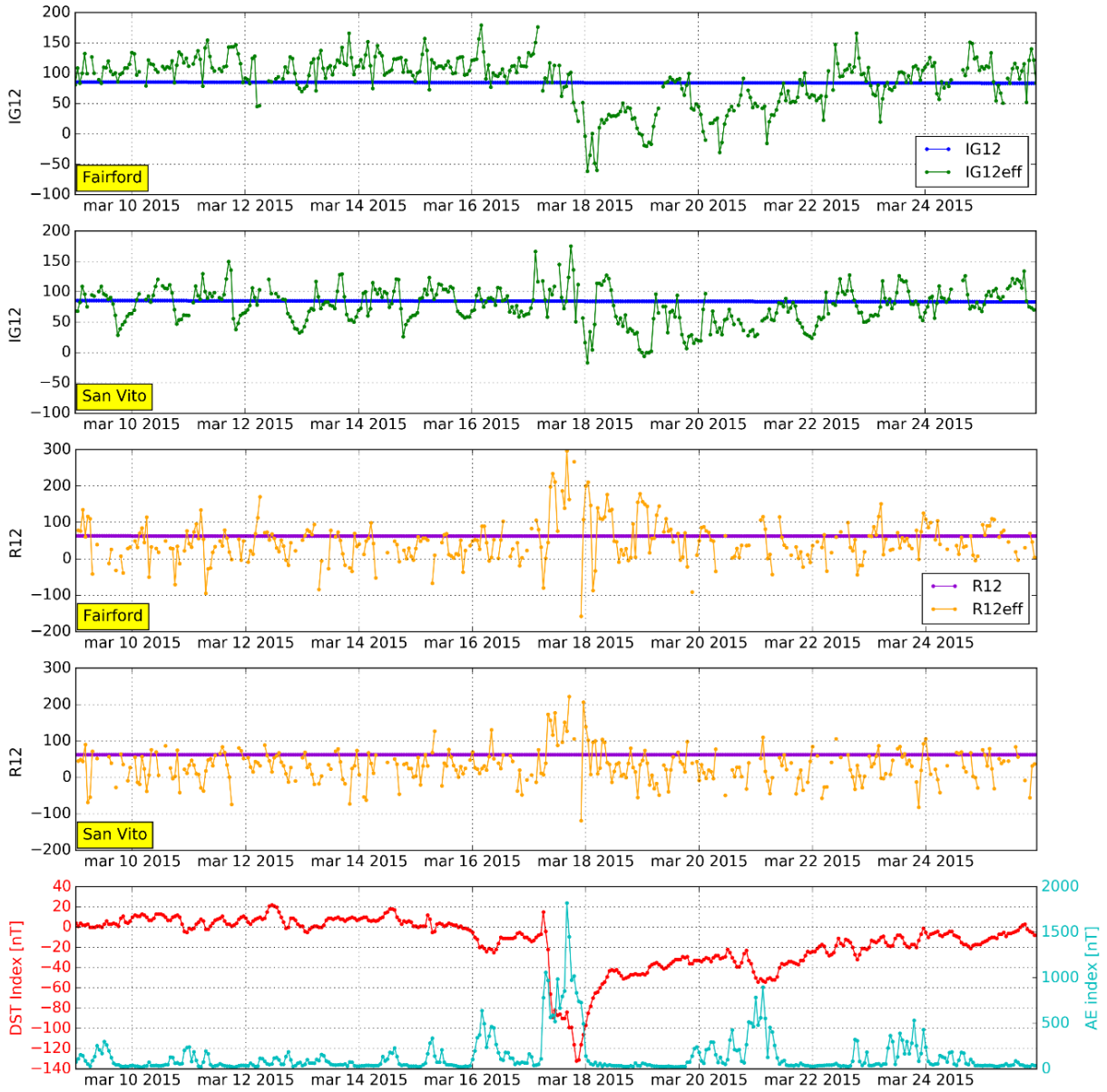


**Figure 4.12:** From top to bottom, for the studied period, time series of: $IG_{12}$ (blue) and $IG_{12\text{eff}}$ (green), for Fairford and San Vito; $R_{12}$ (violet) and $R_{12\text{eff}}$ (orange), for Fairford and San Vito; $D_{\text{st}}$ (red) and AE (cyan) magnetic indices.

### 4.2.4 Comparison with IRI and IRTAM models

Each method that, through data assimilation processes, has the purpose of improving an underlying model, needs to be compared, first of all, with the underlying model itself, and, when possible, with a model characterized by the same goal but based on different mathematical techniques. Hence, IRI UP is compared first with IRI, while the second task is accomplished by considering the IRTAM model (http://giro.uml.edu/IRTAM/, publicly available at http://giro.uml.edu/GAMBIT/) developed by Galkin et al. (2012). IRTAM assimilates data from the Global Ionospheric Radio Observatory (GIRO) (Reinisch and Galkin 2011) to update the climatological IRI

background in terms of *fo*F2 and *hm*F2. The IRTAM model is global by design, i.e., it gives preference to a smoothed rather than a detailed spatial representation, but, given that in the European sector it assimilates data from the same ionospheric stations considered by IRI UP, it seems appropriate to perform a comparison between the two. Notably, the publicly available version of IRTAM does not treat Fairford and San Vito as test sites during its assimilation, but instead excludes nearby Chilton (66 km away) and Rome (460 km) locations where older digisondes operate. This means that IRI UP is disadvantaged in this comparison as it does not assimilate the test sites.

### *fo*F2 statistical comparison

Fig. 4.13a, for the quiet period from March 9 2015 to March 16 2015, and Fig. 4.13b, for the disturbed period from March 17 2015 to March 25 2015, show a comparison between *fo*F2 values autoscaled at Fairford and San Vito, and calculated by IRI (with the STORM option on), by IRI UP, and by IRTAM. Same figures show also the time series of the difference between values recorded by the ionosonde and those calculated by models. Statistical quantities (4.5)-(4.9) related to these time series are shown in Table 4.7. Fig. 4.14 shows the histograms of differences between *fo*F2 values measured at Fairford and San Vito, and calculated by IRI, IRI UP, and IRTAM.

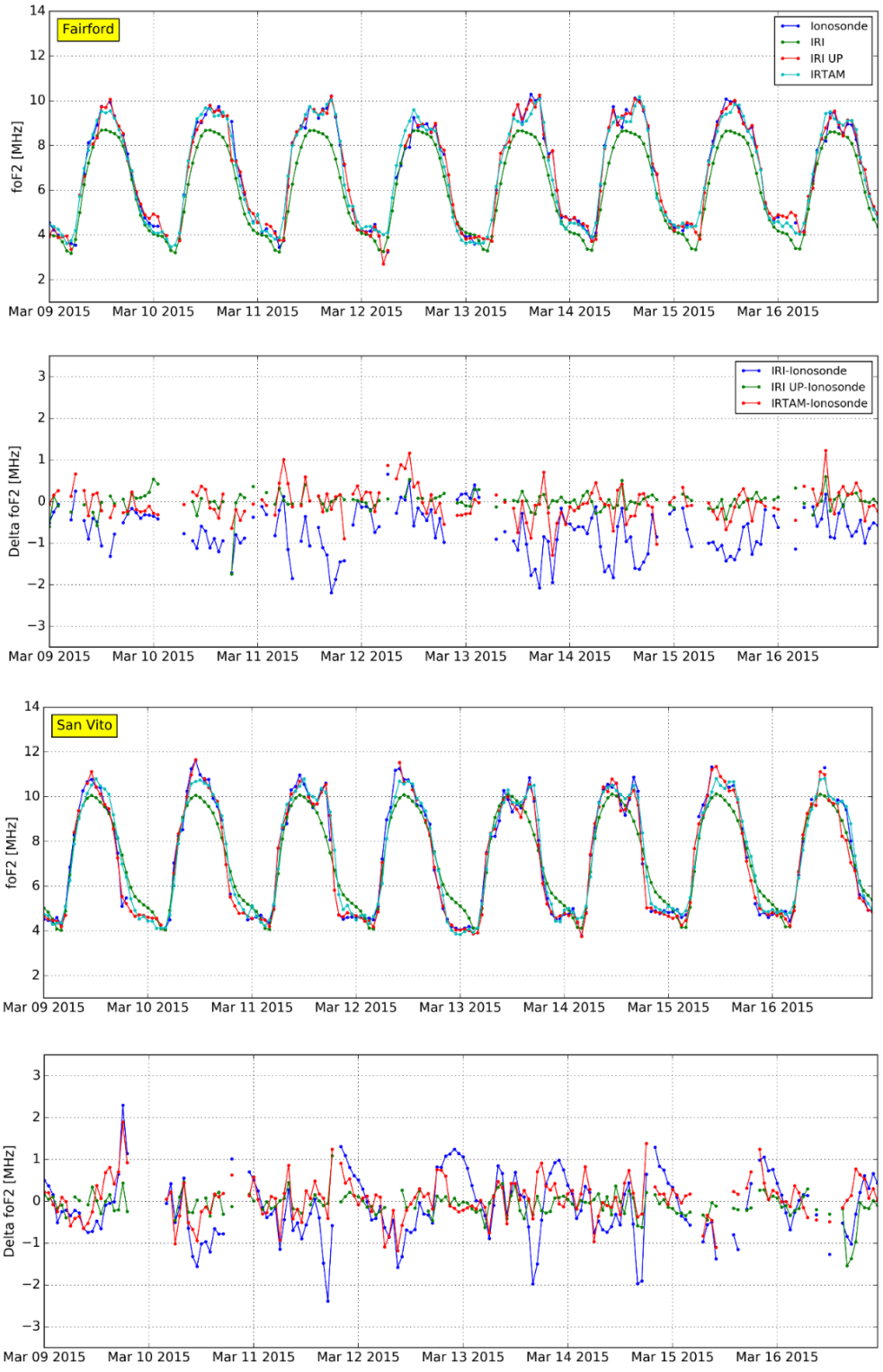


**Figure 4.13a:** Comparison between *fo*F2 values as measured by the ionosonde (blue), and calculated by IRI (green), IRI UP (red), and IRTAM (cyan), for (first panel from the top) Fairford and (third panel from the top) San Vito, for the period from March 9 2015 to March 16 2015. Corresponding differences between measured and modeled values are also reported for Fairford (second panel from the top) and San Vito (fourth panel from the top).

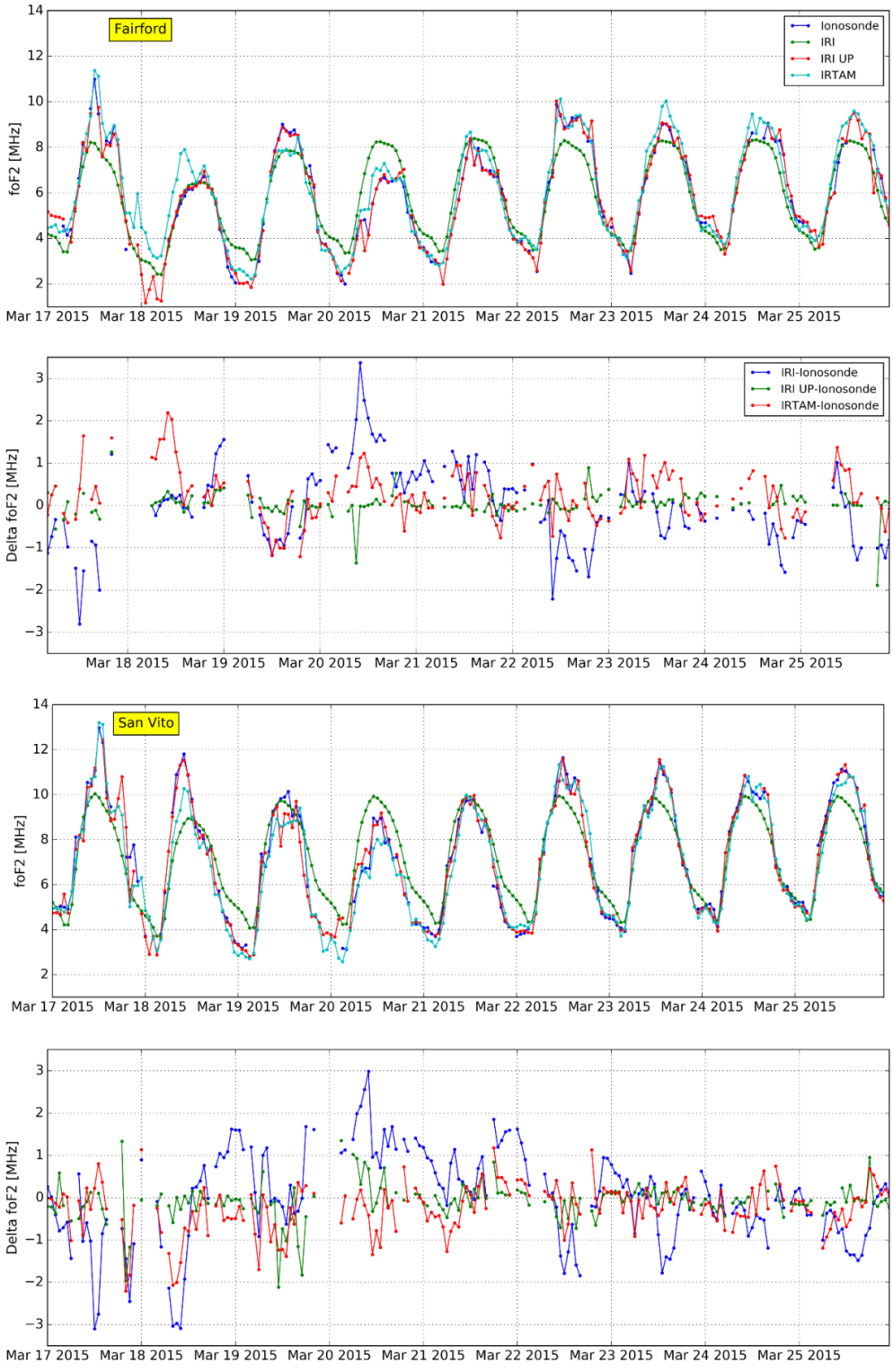


**Figure 4.13b:** Same as Fig. 4.13a but from March 17 2015 to March 25 2015.

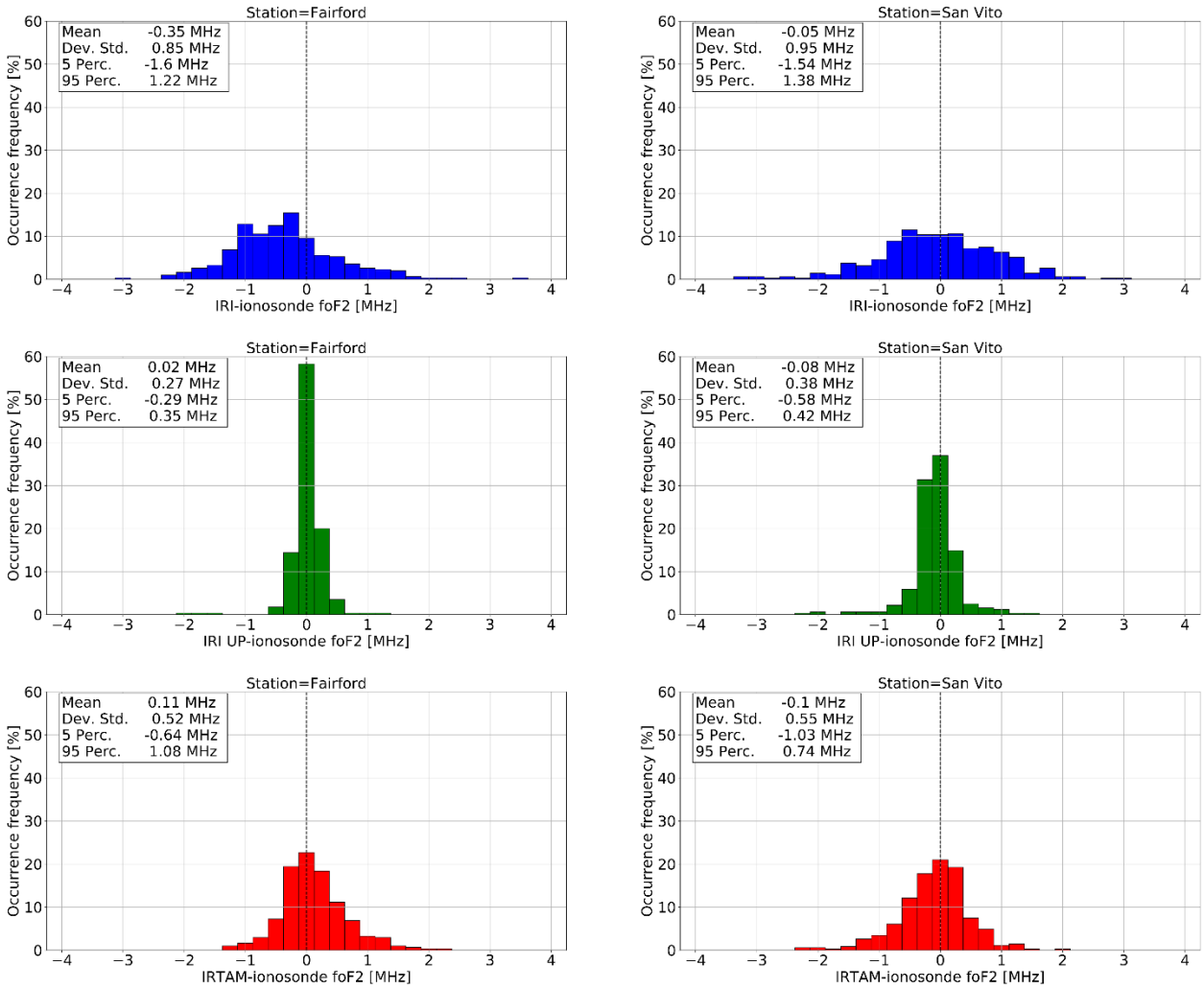


**Figure 4.14:** Histograms representing the occurrence frequency of differences between *fo*F2 measured by the ionosonde and that calculated by IRI (blue), IRI UP (green), and IRTAM (red), for Fairford (left side) and San Vito (right side), for the time period under study (from March 9 2015 at 0 UT to March 25 2015 at 23 UT). In the upper left corner of each histogram the distribution mean, the standard deviation, the 5$^{th}$ and 95$^{th}$ percentiles, are reported. The number of points in IRI and IRTAM histograms are the same (408), while those for IRI UP are slightly lower (382), taking into account the discarded maps (whose percentage is highlighted by the rightmost columns of Table 4.4).

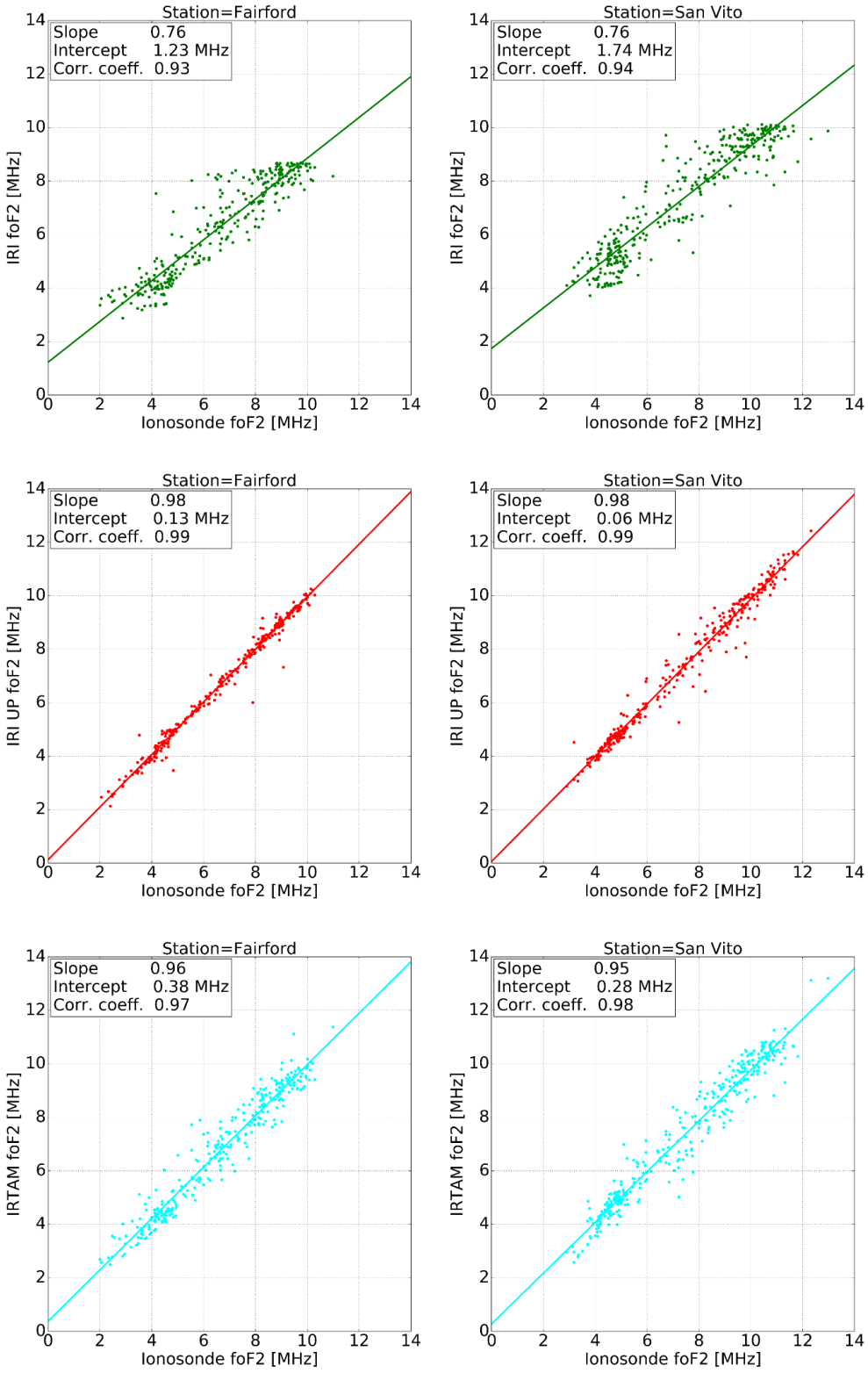


**Figure 4.15:** Scatter plots of *fo*F2 values as measured by the ionosonde against ones modeled by IRI (green), IRI UP (red), and IRTAM (cyan), for Fairford (left side) and San Vito (right side). In each plot the linear fit is drawn, with the corresponding slope and intercept shown in the upper left corner, where also the value of the Pearson's correlation coefficient is reported.

Figs. 4.13a,b show that IRI performs:

a) for quiet days, before the storm commencement and in the last part of the time window, a general underestimation of measured values at Fairford, while a more wavy behavior holds for San Vito;
b) a general overestimation of measured values during the recovery phase of the storm;
c) an important underestimation of measured values during the main phase of the storm (March 17-18).

The relevant feature is however that the IRI STORM option, for this particularly severe event, cannot properly describe the ionospheric plasma. On the other hand, IRI UP and IRTAM show, as expected, a better agreement with measured *fo*F2 values. Specifically, the IRTAM model, although it is very efficient for quiet days, displays some deficiencies in the main phase of the storm. For example, at Fairford on March 18, IRTAM shows a general overestimation, while this does not happen for IRI UP. For the same period, IRTAM shows a general underestimation at San Vito, like IRI, but in a minor extent. This particular day (March 18), under severely disturbed conditions, demonstrates that, during the main phase of the storm, the ionosphere is really spatially variable, with very different *fo*F2 values at Fairford and San Vito. This spatial variability is challenging for every model except IRI UP, whose behavior is good for both quiet and, more importantly, for disturbed conditions. This means that the spatial description of effective indices, made by the application of the Universal Kriging spatial interpolation method, turns out to be very powerful for *fo*F2, allowing to catch the different effects that the storm has at different latitudes. The latter statement is statistically supported by results shown in Table 4.7. Specifically, histograms shown in Fig. 4.14 point out that IRI UP performs better than IRI and IRTAM, with distributions that are more picked around the zero difference and with smaller dispersion values. Fig. 4.14 shows that IRTAM has a good accuracy and it is less precise than IRI UP. The precision of each model can be verified visually through the scatter plots of Fig. 4.15, which show that IRI and IRTAM values are more scattered than IRI UP ones. In conclusion, Figs. 4.13-4.15 and Table 4.7 tell us that, for *fo*F2, the IRI UP model leads to a net improvement with respect to IRI and, to a minor extent, IRTAM.

### *hm*F2 statistical comparison

An analogous comparison is made between measured *hm*F2 values at Fairford and San Vito and those calculated by IRI, IRI UP, and IRTAM. Statistical quantities (4.5)-(4.9) related to these time series are shown in Table 4.8. Fig. 4.16 shows histograms of differences between *hm*F2 values measured at Fairford and San Vito, and those calculated by IRI, IRI UP, and IRTAM.

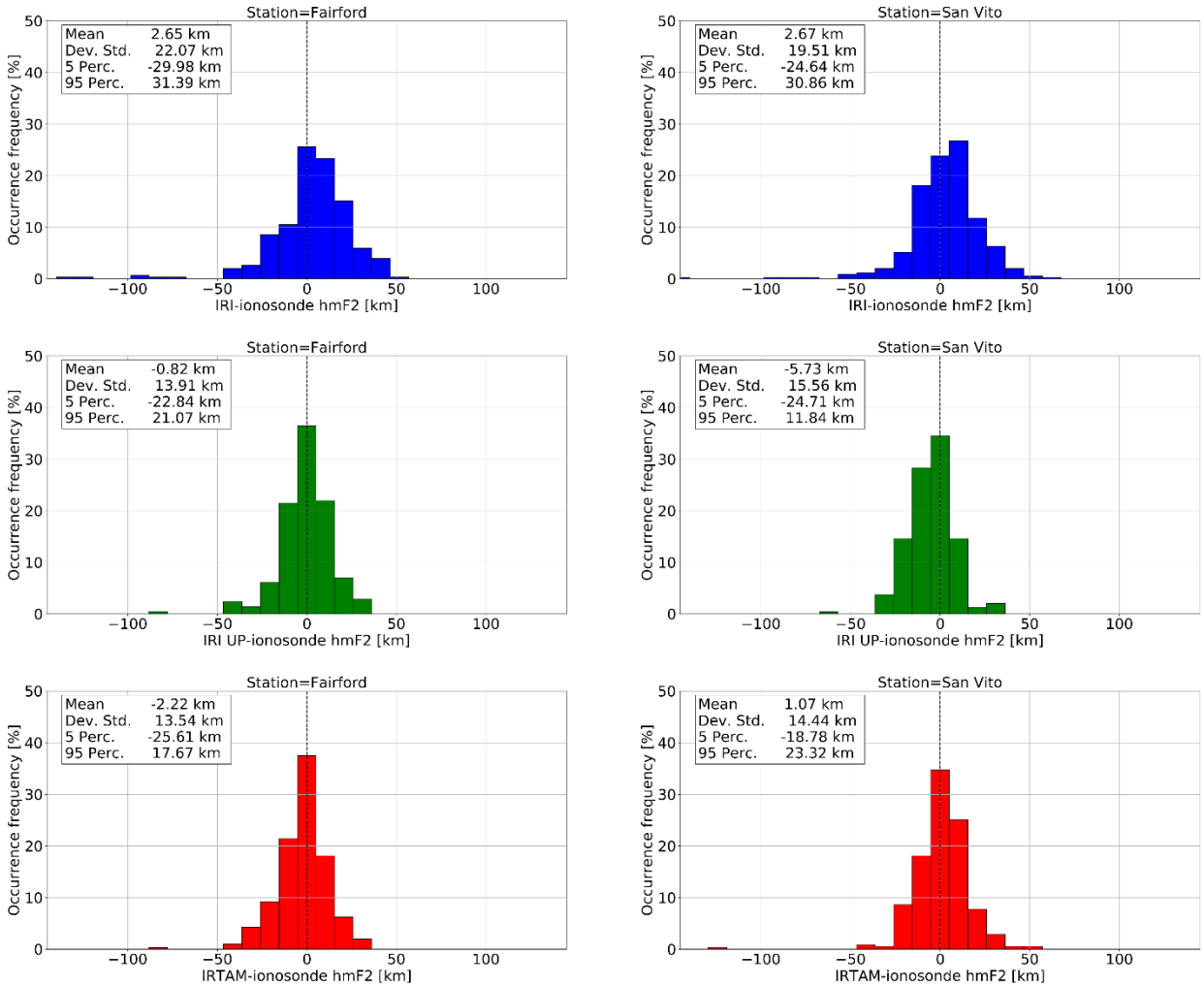


**Figure 4.16:** Same as Fig. 4.14 but for *hm*F2. The number of points in IRI and IRTAM histograms are the same (408), while those for IRI UP are lower (291), taking into account the discarded maps (whose percentage is highlighted by the rightmost columns of Table 4.6).

Statistical quantities of Table 4.8 show that IRI UP and IRTAM both improve the IRI output, but in this case values given as output by IRTAM are slightly better than IRI UP ones. These outcomes are supported also by statistical distributions of Fig. 4.16. However, in general, assimilation of measured ionospheric characteristics seems to be more effective to update *fo*F2 than *hm*F2. The reason of this is that *hm*F2 values are not directly measured or inferred by the ionogram, but calculated through an analytical formula (Eq. (2.7)). This mathematical passage, unavoidably, lowers the degree of precision of this characteristic. Furthermore, *hm*F2 is not directly modeled by means of a single solar/ionospheric index, but it depends on both $IG_{12}$ (through *fo*F2) and $R_{12}$ (through *M*(3000)F2 and *fo*E); thus, errors related to the calculation of these indices can propagate and, normally, increase.

| Station | Ionospheric characteristic | Comparison | RMSE [MHz] | NRMSE [%] | Correlation Coefficient | Mean delta [MHz] | Standard Deviation Delta [MHz] |
|---|---|---|---|---|---|---|---|
| Fairford | *fo*F2 | Ionosonde vs IRI | 0.92 | 13.81 | 0.93 | -0.35 | 0.85 |
| | | Ionosonde vs IRI UP | 0.26 | 3.90 | 0.99 | 0.02 | 0.27 |
| | | Ionosonde vs IRTAM | 0.53 | 8.07 | 0.97 | 0.11 | 0.52 |
| San Vito | *fo*F2 | Ionosonde vs IRI | 0.95 | 12.82 | 0.94 | -0.05 | 0.95 |
| | | Ionosonde vs IRI UP | 0.37 | 5.02 | 0.99 | -0.08 | 0.38 |
| | | Ionosonde vs IRTAM | 0.56 | 7.55 | 0.98 | -0.10 | 0.55 |

**Table 4.7:** Statistical results of the comparison between measured *fo*F2 and calculated by IRI, IRI UP, and IRTAM for both test stations.

| Station | Ionospheric characteristic | Comparison | RMSE [km] | NRMSE [%] | Correlation Coefficient | Mean delta [km] | Standard Deviation Delta [km] |
|---|---|---|---|---|---|---|---|
| Fairford | *hm*F2 | Ionosonde vs IRI | 22.23 | 7.61 | 0.84 | 2.65 | 22.07 |
| | | Ionosonde vs IRI UP | 11.57 | 3.96 | 0.94 | -0.44 | 13.78 |
| | | Ionosonde vs IRTAM | 13.72 | 4.70 | 0.94 | -2.22 | 13.54 |
| San Vito | *hm*F2 | Ionosonde vs IRI | 19.69 | 6.72 | 0.86 | 2.67 | 19.51 |
| | | Ionosonde vs IRI UP | 13.57 | 4.63 | 0.92 | -5.34 | 15.44 |
| | | Ionosonde vs IRTAM | 14.48 | 4.94 | 0.93 | 1.07 | 14.44 |

**Table 4.8:** Same as Table 4.7, for *hm*F2.

### *fo*F2 and *hm*F2 example maps

Figs. 4.17-4.19 are examples of *fo*F2 and *hm*F2 maps, for the chosen European grid, generated by IRI with the STORM option off, IRI with the STORM option on, and IRI UP, for three selected moments. Fig. 4.17 represents a quiet moment on March 9 2015 (calendar day 68) at 15 UT, Figs. 4.18 and 4.19 disturbed moments on March 17 2015 (calendar day 76) at 11 UT and on March 18 2015 (calendar day 77) at 12 UT.

Fig. 4.17 represents an afternoon snapshot of F2-layer characteristics, for quiet conditions. Quiet conditions represented by this figure are testified by the fact that IRI and IRI-STORM maps are identical, for both *fo*F2 and *hm*F2; those associated to IRI UP show instead slight differences for *fo*F2 and significant differences for *hm*F2. For *fo*F2, IRI UP preserves more or less the same gradient direction shown by IRI but with a sensible intensification in magnitude, clear in the upper right and lower left corners; furthermore, a slight increase of *fo*F2 values over Rome is evident, which is caused by an assimilation of an *fo*F2 value at the ionospheric station of Rome well above the monthly median value. On the contrary, the *hm*F2 IRI UP map shows an important decrease of values, well below monthly median ones, in the eastern part of the map, preserving however the gradient direction in the other parts of the map.

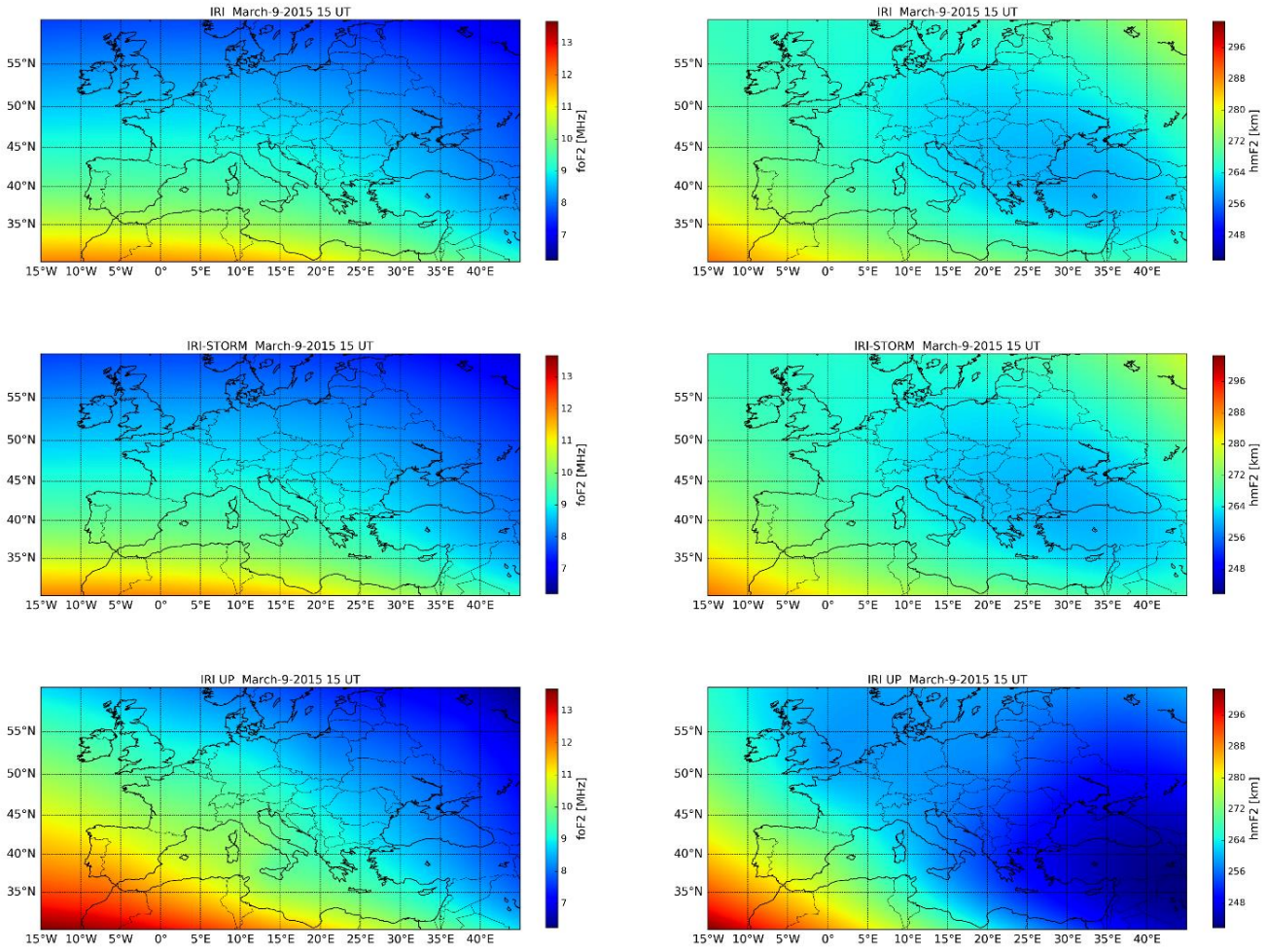


**Figure 4.17:** Examples of maps for (left side) *fo*F2 and (right side) *hm*F2, generated by (top) IRI with the STORM option off, (middle) IRI with the STORM option on, and (bottom) IRI UP. They refer to the quiet day of March 9 2015 (calendar day 68) at 15 UT.

Fig. 4.18 represents a near-noon snapshot of F2-layer characteristics, for disturbed conditions. Despite being a disturbed moment, IRI and IRI-STORM maps are identical, both for *fo*F2 and *hm*F2, since the STORM option uses a weighted mean of $a_p$ values for the previous 33 hours. Maps of Fig. 4.18 are calculated for a moment for which the storm has started only for a few hours (~6 h), and has not reached its maximum development ($D_{st}$ index achieves its minimum on March 17 at 23 UT). The IRI STORM option in the early phase of the storm cannot be effective in catching storm effects on the ionosphere. Instead, IRI UP maps turn out to be very effective, and their

goodness is statistically supported by the analysis described before. These maps, obtained by means of updated effective indices, can properly represent both very fast ionospheric changes, from an hour to the next (and also every 15 minutes, it depends on the ionosonde sounding repetition rate), and also small spatial changes thanks to the Universal Kriging interpolation method.

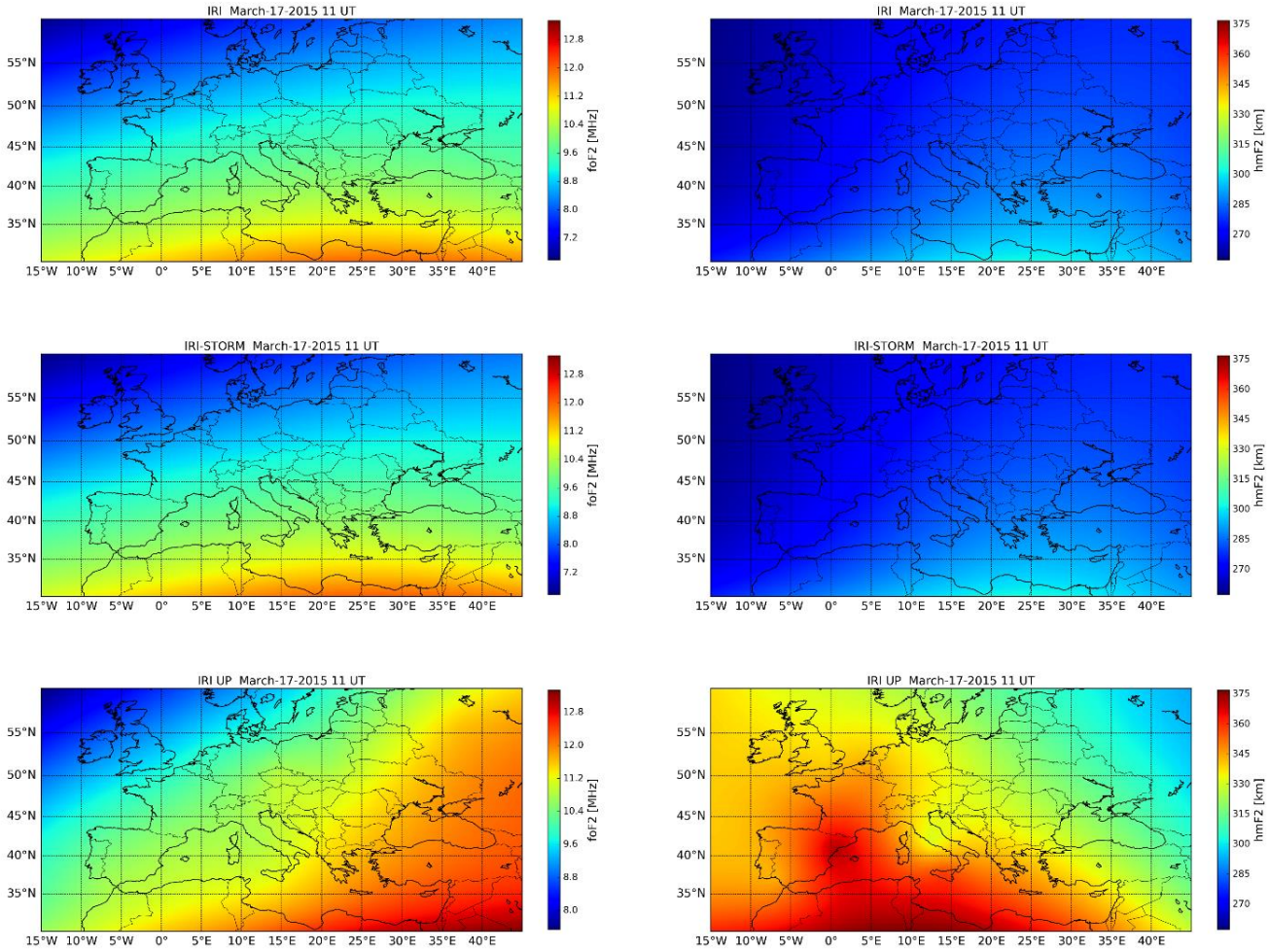


**Figure 4.18:** Same as Fig.4.17 but for the disturbed day March 17 2015 (calendar day 76) at 11 UT.

Fig. 4.19 represents a noon snapshot of F2-layer characteristics, for very disturbed conditions. In this case, being the storm well developed, IRI and IRI-STORM maps show large differences in magnitude, for both *fo*F2 and *hm*F2, even though the gradient direction is similar. IRI UP map for *fo*F2 shows a gradient which is similar to that of IRI and IRI-STORM, although it manifests a wavier behavior than the more linear IRI ones. For *hm*F2, differences between IRI and IRI UP are again very important, both in gradient direction and magnitude.

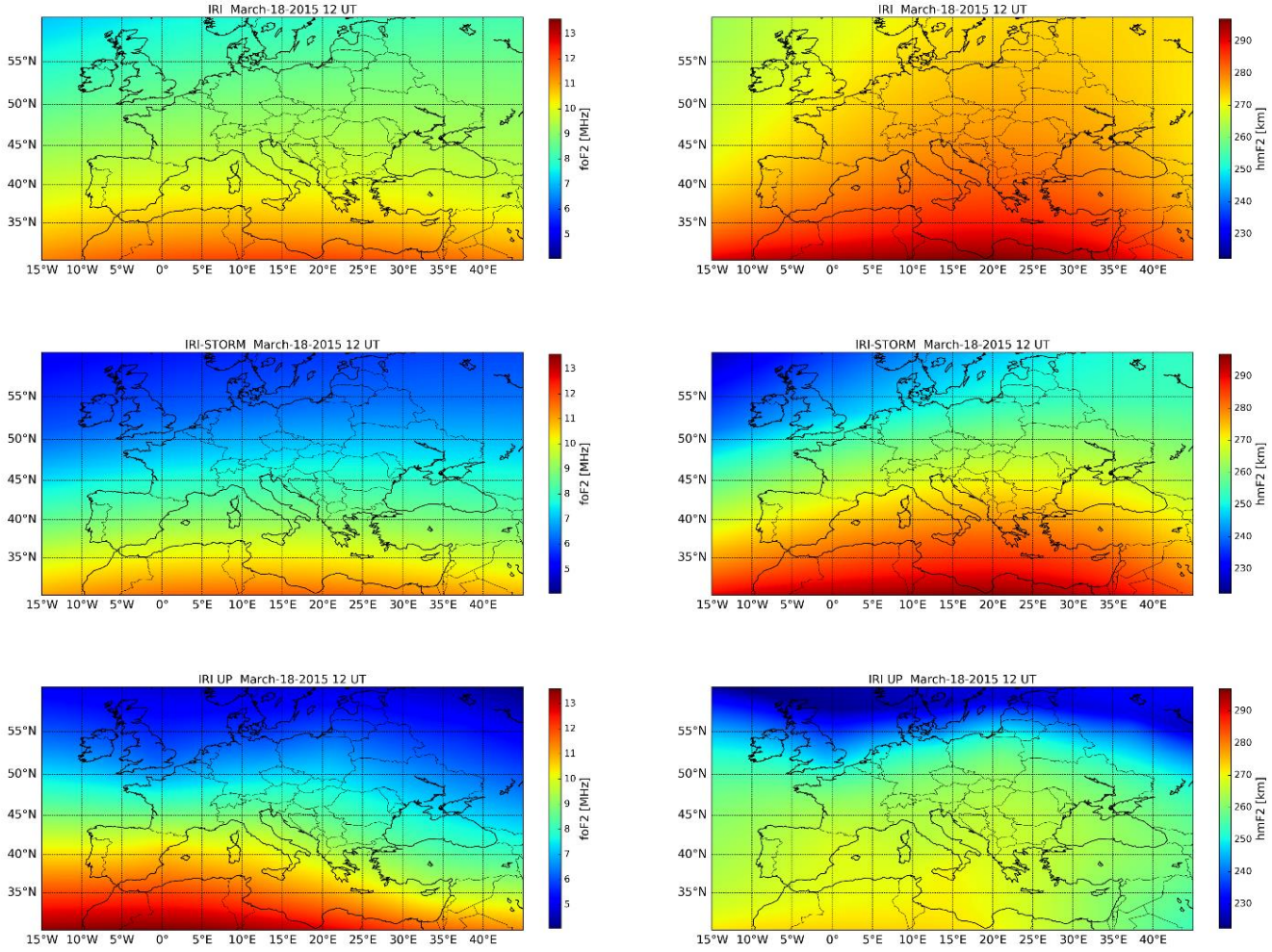


**Figure 4.19:** Same as Fig. 4.17 but for the disturbed day March 18 2015 (calendar day 77) at 12 UT.

Visual analysis of *fo*F2 and *hm*F2 maps confirms the goodness of the IRI UP method, highlighting its ability to catch rapid and small-scale electron density changes occurring especially under disturbed conditions.

## 4.3 Validation of IRI UP method on 30 moderate, strong, and severe geomagnetic storms occurred during 2004-2016 years

In order to explore more thoroughly the IRI UP performance under geomagnetic storm conditions, the IRI UP method is further validated applying it on 30 selected time intervals characterized by moderate, strong, and severe geomagnetic storms occurred from January 1st 2004 to December 31st 2016. Such period embraces part of the past 23rd solar cycle (August 1996 - December 2008) and the current one.

The validation of the IRI UP method is carried out comparing the IRI UP performance with the one of the IRI model with the STORM option on, at the two truth-sites of Fairford (51.7° N, 1.5° W) and San Vito (40.6° N, 17.8° E).

### 4.3.1 Data used and periods under study

Magnetically disturbed conditions have been selected based on $A_p$ geomagnetic index data, and $K_p$ geomagnetic index data. $A_p$ daily mean values > 50 and $K_p$ maximum values > $5^+$, as recorded during the main phase day of the storm, are the thresholds adopted to discard minor storms, thus selecting only the moderate, strong, and severe ones.

The ionospheric stations used for this validation are those listed in Table 4.1 and Fig. 4.1. Availability of the assimilated ionosondes for each year is shown in Table 4.9.

Specifically, each time interval is selected considering the day before the main phase of the storm (which, in the selected cases, is always a quiet day), the day of the main phase, and the next four days of the recovery phase, for a total number of six days for each time interval. Nevertheless, when the storm is characterized by substorms with $A_p$ > 50, then from the last substorm occurred, four days of the recovery phase are further considered; this means that, in these cases, the period under study can be characterized by more than six days. Some info about the geomagnetic storms considered in this study are summarized in Table 4.10.

| **Year** | **Number of available ionospheric stations for assimilation** | **Available ionospheric stations for assimilation** |
|---|---|---|
| 2004 | 8 | Ath, Chi, Dou, ElA, Jul, Pru, Rom, Roq |
| 2005 | 8 | Ath, Chi, Dou, ElA, Jul, Pru, Rom, Roq |
| 2006 | 7 | Ath, Chi, Dou, ElA, Jul, Pru, Rom |
| 2007 | 8 | Ath, Chi, Dou, ElA, Jul, Pru, Rom, Roq |
| 2008 | 10 | Ath, Chi, Dou, ElA, Jul, Mos, Nic, Pru, Rom, Roq |
| 2009 | 11 | Ath, Chi, Dou, ElA, Gib, Jul, Mos, Nic, Pru, Rom, Roq |
| 2010 | 10 | Ath, Chi, Dou, ElA, Jul, Mos, Nic, Pru, Rom, Roq |
| 2011 | 10 | Ath, Chi, Dou, ElA, Jul, Mos, Nic, Pru, Rom, Roq |
| 2012 | 12 | Ath, Chi, Dou, ElA, Gib, Jul, Mos, Nic, Pru, Rom, Roq, War |
| 2013 | 12 | Ath, Chi, Dou, ElA, Gib, Jul, Mos, Nic, Pru, Rom, Roq, War |
| 2014 | 12 | Ath, Chi, Dou, ElA, Gib, Jul, Mos, Nic, Pru, Rom, Roq, War |
| 2015 | 12 | Ath, Chi, Dou, ElA, Gib, Jul, Mos, Nic, Pru, Rom, Roq, War |
| 2016 | 12 | Ath, Chi, Dou, ElA, Gib, Jul, Mos, Nic, Pru, Rom, Roq, War |

**Table 4.9:** Ionospheric stations available for data assimilation for each year, from 2004 to 2016.

| Storm Number | Main Phase Day [dd mm yyyy] | $A_p$ | $K_p$ max | Analyzed period [dd-dd mm-mm yyyy] | $R_{12}$ |
|---|---|---|---|---|---|
| Solar cycle 23 | | | | | |
| 1 | 22 01 2004 | 64 | 7 | 21-26 01 2004 (Str) | 80.1 |
| 2 | 27 07 2004 | 186 | 9- | **22-31 07 2004 (Sev)** | 64.8 |
| 3 | 10 11 2004 | 161 | 9- | **06-14 11 2004 (Sev)** | 56.6 |
| 4 | 18 01 2005 | 84 | 8- | **16-25 01 2005 (Sev)** | 54.5 |
| 5 | 05 04 2005 | 50 | 7 | 04-09 04 2005 (Str) | 49.3 |
| 6 | 08 05 2005 | 91 | 8+ | 07-12 05 2005 (Sev) | 45.0 |
| 7 | 15 05 2005 | 87 | 8+ | 14-19 05 2005 (Sev) | 45.0 |
| 8 | 30 05 2005 | 90 | 8- | **29-4 05-06 2005 (Sev)** | 45.0 |
| 9 | 12 06 2005 | 54 | 7+ | 11-16 06 2005 (Str) | 44.5 |
| 10 | 23 06 2005 | 50 | 7 | 22-27 06 2005 (Str) | 44.5 |
| 11 | 10 07 2005 | 57 | 6+ | 09-14 07 2005 (Mod) | 44.6 |
| 12 | 24 08 2005 | 102 | 9- | 23-28 08 2005 (Sev) | 41.9 |
| 13 | 11 09 2005 | 101 | 8- | **10-16 09 2005 (Sev)** | 39.4 |
| 14 | 14 04 2006 | 65 | 7 | 13-18 04 2006 (Str) | 27.0 |
| 15 | 15 12 2006 | 94 | 8+ | 14-19 12 2006 (Sev) | 20.2 |
| Solar cycle 24 | | | | | |
| 16 | 05 04 2010 | 55 | 8- | 04-09 04 2010 (Sev) | 20.8 |
| 17 | 09 03 2012 | 87 | 8 | 08-13 03 2012 (Sev) | 98.3 |
| 18 | 15 07 2012 | 78 | 7 | 14-19 07 2012 (Str) | 84.5 |
| 19 | 17 03 2013 | 72 | 7- | 16-21 03 2013 (Str) | 84.4 |
| 20 | 01 06 2013 | 58 | 7 | **30-05 05-06 2013 (Str)** | 90.6 |
| 21 | 29 06 2013 | 50 | 6+ | 28-02 06-07 2013 (Mod) | 90.6 |
| 22 | 02 10 2013 | 58 | 8- | 01-06 10 2013 (Sev) | 107.0 |
| 23 | 17 03 2015 | 108 | 8- | 16-21 03 2015 (Sev) | 82.2 |
| 24 | 22 06 2015 | 57 | 8- | 21-26 06 2015 (Sev) | 72.1 |
| 25 | 27 08 2015 | 52 | 6+ | 26-31 08 2015 (Mod) | 66.4 |
| 26 | 09 09 2015 | 60 | 7 | 08-13 09 2015 (Str) | 65.9 |
| 27 | 07 10 2015 | 74 | 7+ | 06-11 10 2015 (Str) | 64.3 |
| 28 | 20 12 2015 | 70 | 7- | 19-24 12 2015 (Str) | 57.8 |
| 29 | 08 05 2016 | 70 | 6+ | 07-12 05 2016 (Mod) | 36.0 |
| 30 | 25 10 2016 | 57 | 6+ | 24-29 10 2016 (Mod) | 31.5 |

**Table 4.10:** The Main Phase Day (MPD), $A_p$ and $K_p$ max geomagnetic indices values as recorded during the MPD, the analyzed periods, and the $R_{12}$ solar activity index are shown for each moderate (Mod), strong (Str), and severe (Sev) geomagnetic storm. Periods characterized by significant substorms are highlighted in bold.

### 4.3.2 $Q_1$, $Q_2$, and cR statistical test: some results

Some examples of $Q_1$, $Q_2$, and $cR$ time series, for linear and spherical variogram models used to obtain $IG_{12eff}$ and $R_{12eff}$ maps, are shown in Figs. 4.20 and 4.21, respectively. For the $Q_1$ statistical test, each case exceeding the threshold (3.17) is rejected; as well as for the $Q_2$ statistical test each case not included in the interval (3.20) is

rejected as well. Note that the epochs characterized by the greatest values of *cR* are those following the main phase of the storm.

Blue, red and green dots represent the values of $Q_1$, $Q_2$, and *cR*, respectively, computed by Eqs. (3.15), (3.19), and (3.22). For the $Q_1$ and $Q_2$ statistics the thresholds depend on the number *n* of *fo*F2 and *M*(3000)F2 values available in the assimilation stations which, at a given hour, are considered. Obviously, the value of *n* can be different from epoch to epoch because, at a given moment, it is possible that a station is not working and/or the ionospheric characteristics of a station are wrongly autoscaled. This explains why the continuous black lines representing the $Q_1$ and $Q_2$ thresholds in Figs. 4.20 and 4.21 are not flat.

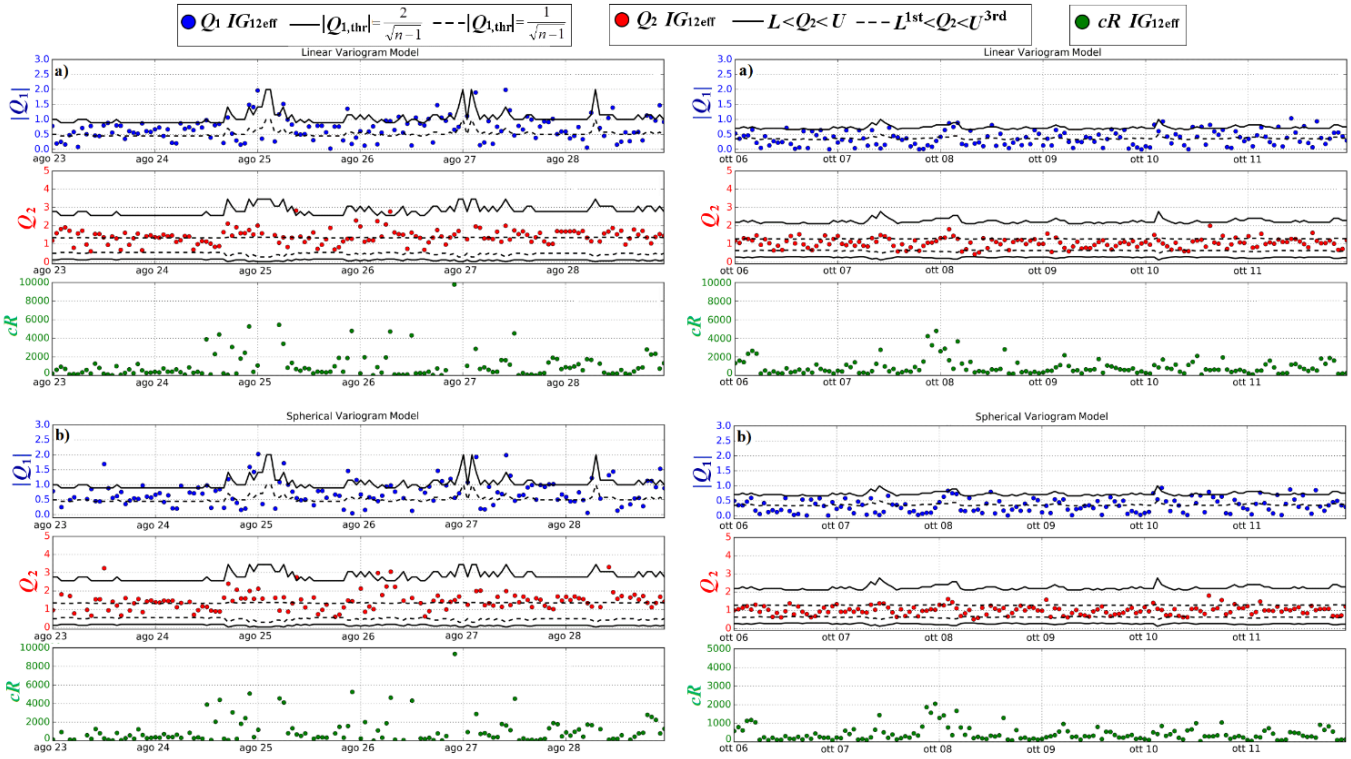


**Figure 4.20:** $Q_1$, $Q_2$, and *cR* time series for a) linear and b) spherical variogram models for $IG_{12eff}$ corresponding to the severe geomagnetic storms listed in Table 4.10 as number 12 (left) and 27 (right). The continuous and dashed black lines in $Q_1$ and $Q_2$ plots represent the threshold values as highlighted at the top of the figure.

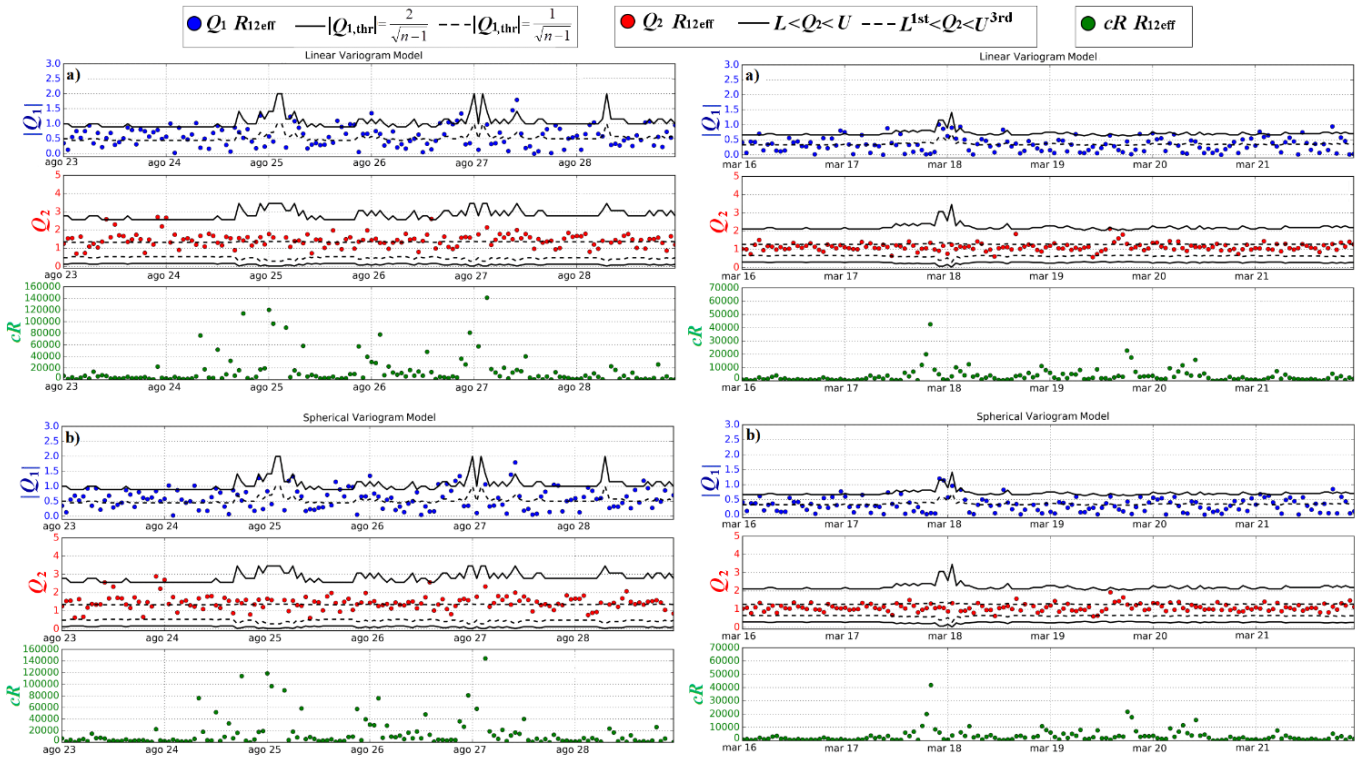


**Figure 4.21:** Same as Fig. 4.20, but for $R_{12eff}$.

Looking at Figs. 4.20 and 4.21 clearly emerges that, in the case of the storm number 12 (23-28 August 2005), the number of rejected variogram models is relatively large when the threshold is lowered from $|2\sigma|$ to $|\sigma|$ (for the statistical test $Q_1$) and from [*L*-*U*] to [$L^{1st}$-$U^{3rd}$] (for the statistical test $Q_2$). The threshold effect is however much less evident in the case of the storm number 27 (6-11 October 2015), for which a limited number of variograms is discarded when reducing the threshold.

This is probably due to the different number of available assimilated stations used, which for the period 6-11 October 2015 ($n$ = 12) is larger than that for the period 23-28 August 2005 ($n$ = 8), thus allowing a better representation of the spatial gradients over the area under study.

As it is easy to realize looking at Figs. 4.20 and 4.21, the number of variogram models discarded by the variogram model selection procedure depends on the established thresholds values. As a general rule, if, at a given epoch, the number of reliable ionosonde data to be assimilated is relatively large, one can choose a more selective threshold, thus providing ionospheric characteristics maps with a high confidence level. In the event that the number of reliable data is lower, one has to increase the threshold and, in this case, a map with a lower confidence level is provided. It is clear that, when there are very few reliable data to be assimilated, one cannot provide a statistically significant updated map. In this case, the IRI UP method is not applicable and one relies on the IRI background map.

### 4.3.3. Validation of IRI UP method: some results

The IRI UP method is systematically tested over the 30 disturbed time intervals listed in Table 4.10, in order to investigate its performance during moderate, strong, and severe geomagnetic storms.

For each epoch, the testing procedure follows 4 steps:

1) the variogram models which have passed the $Q_1$ and $Q_2$ statistical tests, and the $cR$ criterion, are considered;

2) applying the Universal Kriging method for each selected variogram model, $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ maps are calculated over the European area depicted in Fig. 4.1;

3) *fo*F2 and *M*(3000)F2 maps are calculated giving as input to the IRI model the $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ maps calculated in 2); then, applying the empirical formula (2.7) which relates *hm*F2 to *fo*F2 and *M*(3000)F2 (Bilitza et al. 1979), also *hm*F2 maps are obtained;

4) from *fo*F2 and *hm*F2 maps, values at the truth-sites of Fairford and San Vito are extracted and compared with corresponding measurements.

For each storm the statistical parameters RMSE, NRMSE, Pearson correlation coefficient $\rho$, Mean delta MD, and the percentage of discarded maps (Eqs. (4.5)-(4.10)), are calculated.

Figs. 4.22 and 4.23 show some examples of comparison between IRI and IRI UP, in terms of the aforementioned statistical quantities carried out at Fairford and San Vito for each storm listed in Table 4.10, for both *fo*F2 and *hm*F2. Fig. 4.24 shows instead, for each storm individually and for the whole group of storms, the percentage of discarded variogram models calculated by the Eq. (4.10).

Finally, the mean delta quantities provided by Eq. (4.8), taken as absolute values (to avoid that positive and negative values around zero could cancel each other), are used to calculate the following mean value:

$$\text{MMDAV (Mean of Mean Delta Absolute Value)} = \sum_{j=1}^{30} \frac{|MD_j|}{30}. \tag{4.11}$$

Tables 4.11 and 4.12 summarize the statistical results at the truth-sites of Fairford and San Vito for *fo*F2 and *hm*F2 respectively, in the following three cases:

a) IRI UP method running with a fixed variogram model, considering only those cases passing the first two steps of the variogram selection procedure, namely $Q_1$ and $Q_2$ tests;
b) IRI background model;
c) IRI UP method embedding the complete variogram selection procedure.

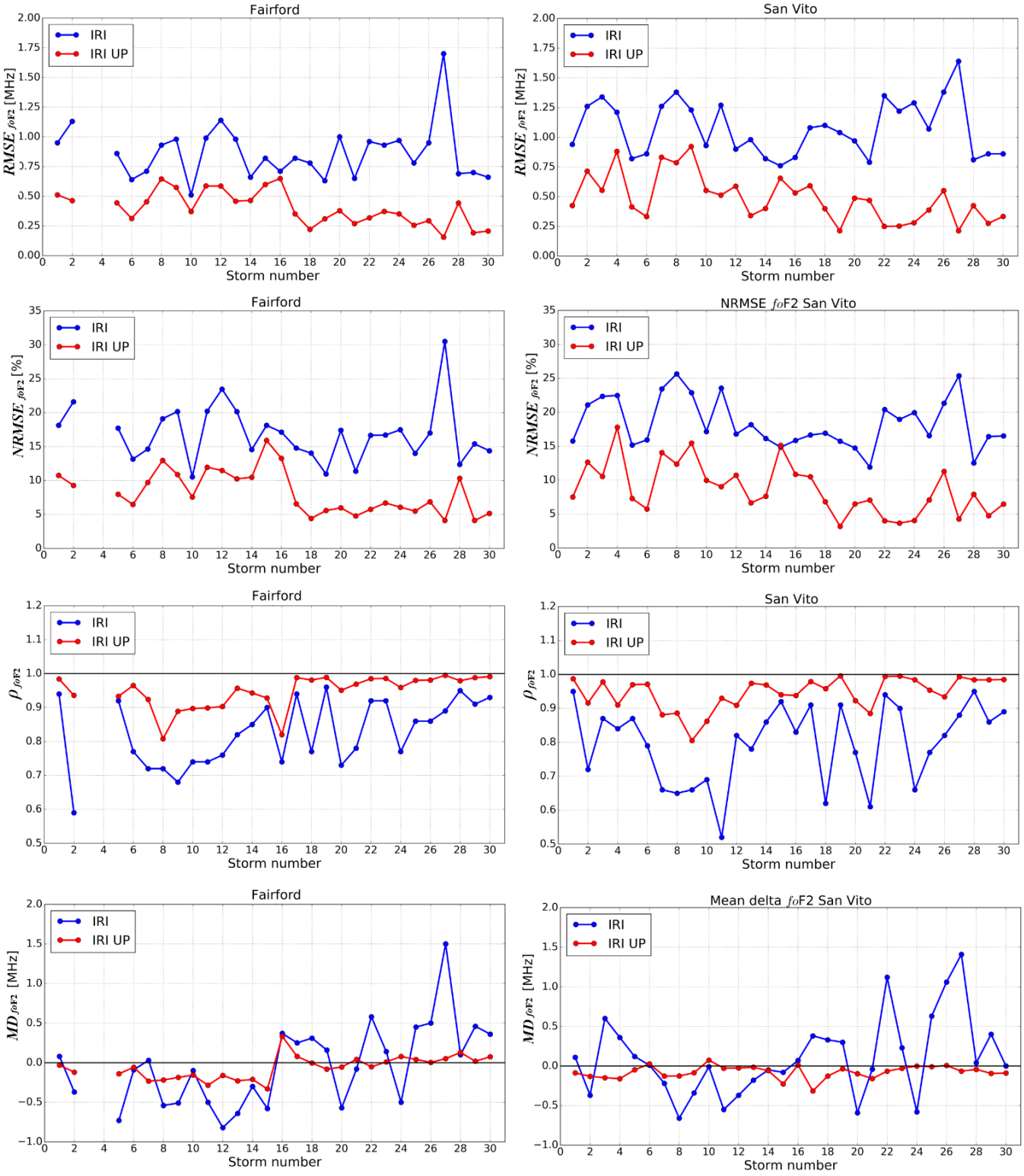


**Figure 4.22:** Trend of statistical parameters defined by Eqs. (4.5)-(4.8) related to *fo*F2 for Fairford (left) and San Vito (right). The gap between storms number 2 and 5 at Fairford is due to the lack of ionosonde data.

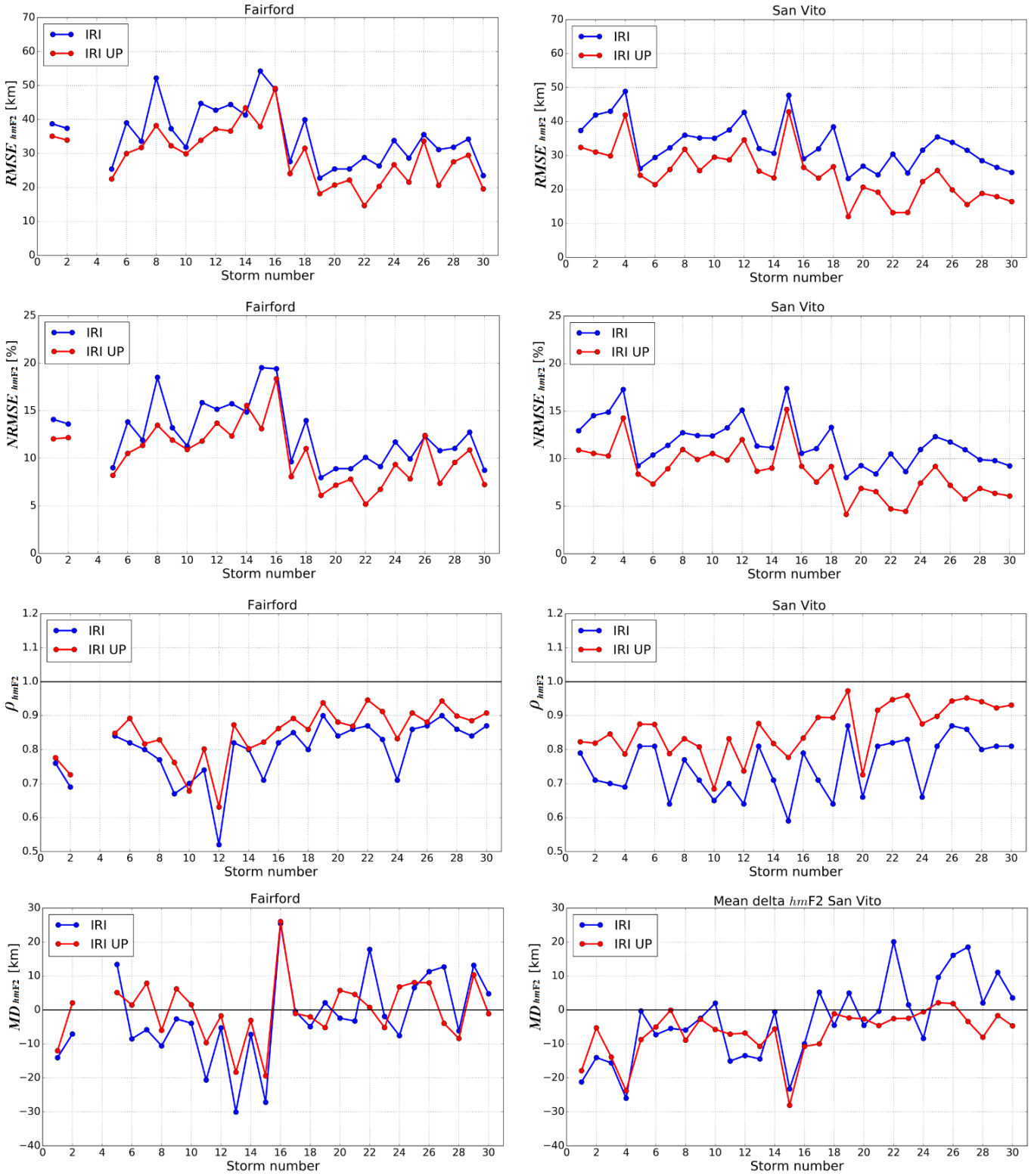


**Figure 4.23:** Same as Fig. 4.22, but for *hm*F2.

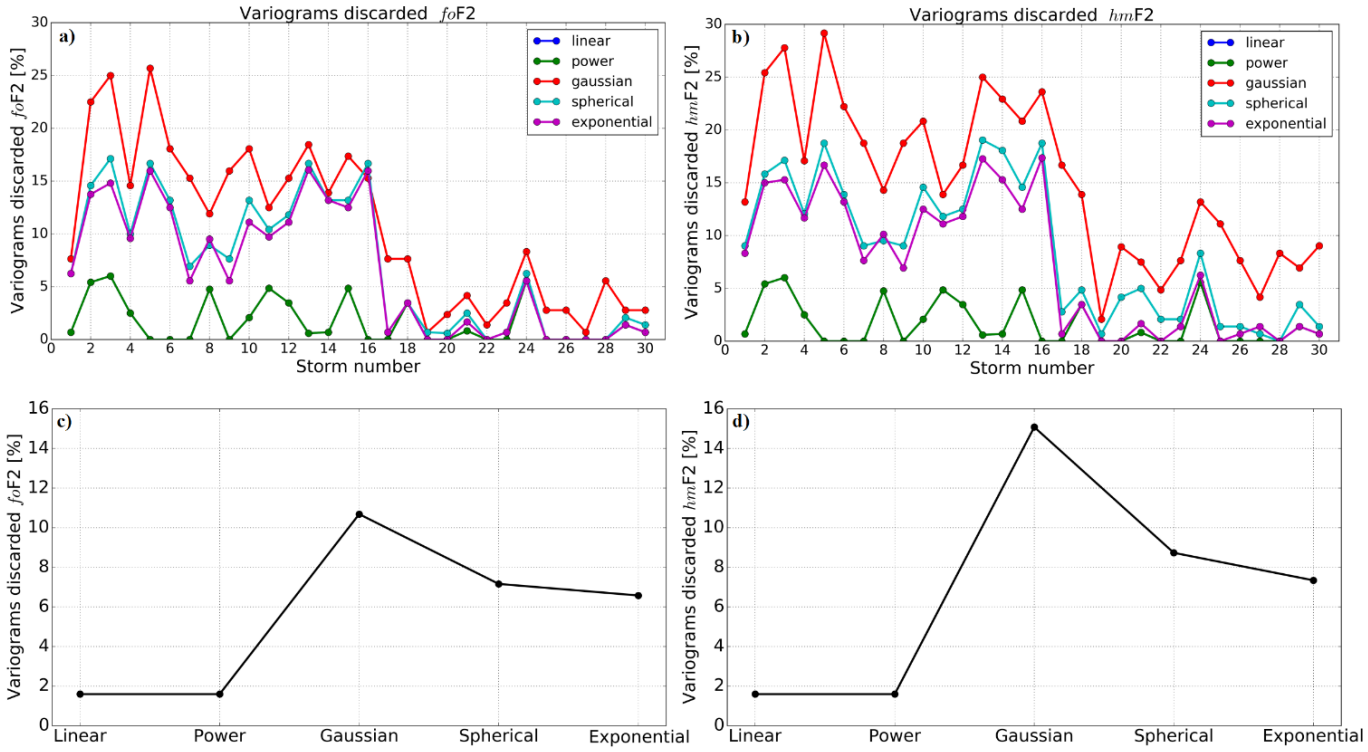


**Figure 4.24:** Percentage of discarded variogram models for (top) each storm listed in Table 4.10 and (bottom) all storms as a whole, for (left) *fo*F2 and (right) *hm*F2. On the top panels, the percentage of the linear model is not visible because hidden by the power model.

| Station | Ionospheric characteristic | Model | RMSE [MHz] | NRMSE [%] | ρ | MMDAV [MHz] |
|---|---|---|---|---|---|---|
| Fairford | *fo*F2 | IRI UP - *linear* | 0.474 | 9.40 | 0.927 | 0.137 |
| | | IRI UP - *power* | 0.471 | 9.37 | 0.929 | 0.139 |
| | | IRI UP - *Gaussian* | 0.467 | 9.28 | 0.920 | 0.152 |
| | | IRI UP - *spherical* | 0.464 | 9.23 | 0.926 | 0.146 |
| | | IRI UP - *exponential* | 0.461 | 9.14 | 0.927 | 0.144 |
| | | IRI | 0.865 | 16.85 | 0.824 | 0.415 |
| | | **IRI UP** | 0.401 | 8.25 | 0.947 | 0.122 |
| | | | | | | |
| San Vito | *fo*F2 | IRI UP - *linear* | 0.558 | 9.74 | 0.927 | 0.088 |
| | | IRI UP - *power* | 0.555 | 9.69 | 0.927 | 0.090 |
| | | IRI UP - *Gaussian* | 0.538 | 9.38 | 0.922 | 0.083 |
| | | IRI UP - *spherical* | 0.539 | 9.39 | 0.927 | 0.088 |
| | | IRI UP - *exponential* | 0.538 | 9.36 | 0.927 | 0.090 |
| | | IRI | 1.075 | 18.37 | 0.797 | 0.374 |
| | | **IRI UP** | 0.485 | 8.70 | 0.946 | 0.084 |

**Table 4.11:** Statistical results for *fo*F2 obtained at the two truth-sites for: IRI UP running with a fixed variogram model (IRI UP – *variogram model* chosen, in the table*)*, IRI, and IRI UP method applying the variogram selection procedure (boldface).

| Station | Ionospheric characteristic | Variogram model | RMSE [km] | NRMSE [%] | ρ | *MMDAV* [km] |
|---|---|---|---|---|---|---|
| Fairford | *hm*F2 | IRI UP - *linear* | 31.567 | 11.31 | 0.817 | 7.361 |
| | | IRI UP - *power* | 31.552 | 11.27 | 0.819 | 7.376 |
| | | IRI UP - *Gaussian* | 30.334 | 10.83 | 0.796 | 7.457 |
| | | IRI UP - *spherical* | 30.744 | 10.97 | 0.811 | 7.388 |
| | | IRI UP - *exponential* | 30.844 | 11.00 | 0.811 | 7.479 |
| | | IRI | 35.241 | 12.57 | 0.797 | 9.879 |
| | | **IRI UP** | 29.369 | 10.44 | 0.845 | 6.844 |
| | | | | | | |
| San Vito | *hm*F2 | IRI UP - *linear* | 30.132 | 10.62 | 0.798 | 7.469 |
| | | IRI UP - *power* | 30.197 | 10.64 | 0.799 | 7.674 |
| | | IRI UP - *Gaussian* | 28.193 | 10.19 | 0.780 | 8.162 |
| | | IRI UP - *spherical* | 29.805 | 10.50 | 0.783 | 8.053 |
| | | IRI UP - *exponential* | 29.685 | 10.46 | 0.792 | 7.901 |
| | | IRI | 33.252 | 11.70 | 0.749 | 9.577 |
| | | **IRI UP** | 24.671 | 8.62 | 0.859 | 6.959 |

**Table 4.12:** Same as Table 4.11, but for *hm*F2.

The winning percentage of each variogram model is computed for each single storm and for the complete storm set, evaluating the following parameters:

$$\%_{\text{single\_storm},i} = \frac{n_{ss,i}}{N_{ss}}, \tag{4.12}$$

$$\%_{\text{all\_storms},i} = \frac{\sum_{ss=1}^{30} n_{ss,i}}{\sum_{ss=1}^{30} N_{ss}}, \tag{4.13}$$

where $i$ is the index running on the 5 possible variogram models, $ss$ is the index running on the considered storms, $n_{ss,i}$ is the number of times the $i$ variogram model is declared as the "winner" by the variogram selection procedure, and $N_{ss}$ is the total number of epochs included in each single storm.

Figs. 4.25a and 4.26a show the winning percentage of each variogram model for each single storm, for $IG_{12\text{eff}}$ and $R_{12\text{eff}}$, respectively; Figs. 4.25b and 4.26b show the same percentage for the complete storm set.

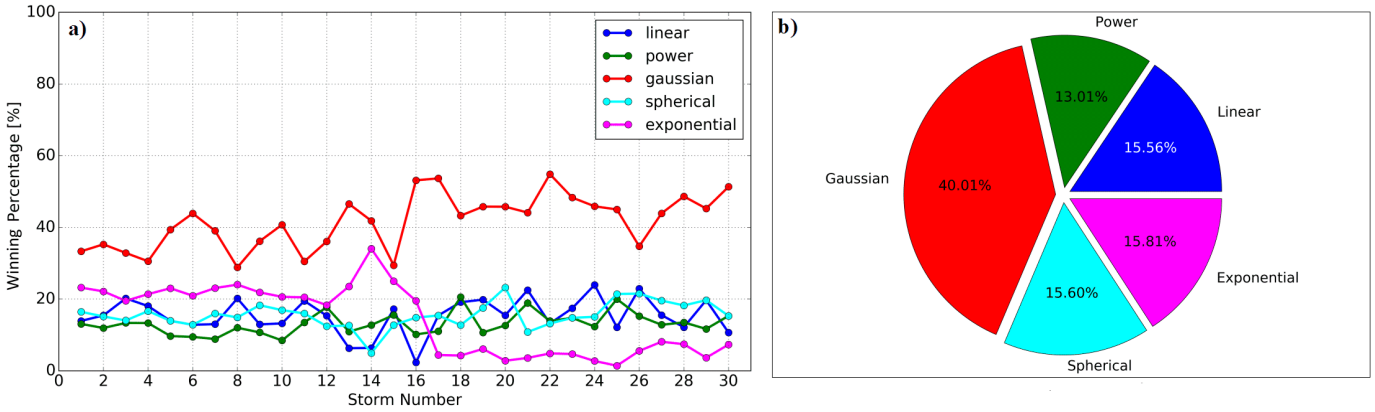


**Figure 4.25:** a) Winning percentages related to each $IG_{12\text{eff}}$ variogram model, for each single storm, after applying the variogram selection procedure; b) same as a) but considering all storms listed in Table 4.10 as a whole.

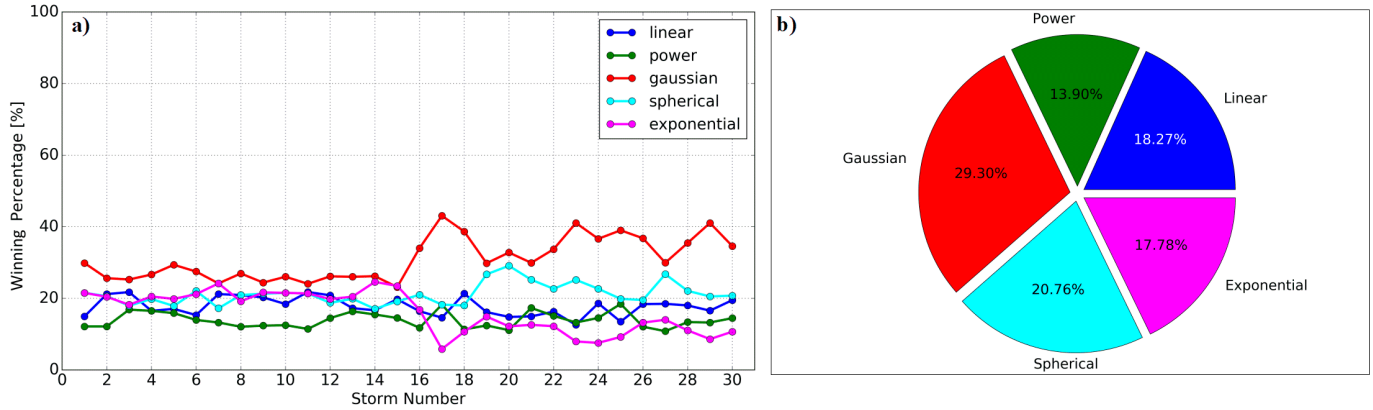


**Figure 4.26:** same as Fig. 4.25, but for $R_{12\text{eff}}$.

Statistical results shown in Figs. 4.22 and 4.23, as well as in Tables 4.11 and 4.12, clearly indicate that IRI UP performs significantly better than IRI, for each of the 30 considered cases. Moreover, results of Tables 4.11 and 4.12, show that when IRI UP is applied deciding a priori the variogram model, a clear difference among the various variogram models does not emerge. In fact, NRMSE values for *fo*F2 (*hm*F2) range between 9.14 % and 9.40 % (10.83 % and 11.31%) and 9.36 % and 9.74 % (10.19 % and 10.64%) at Fairford and San Vito, respectively, while MMDAV values for *fo*F2 (*hm*F2) range between 0.137 MHz and 0.152 MHz (7.361 km and 7.479 km) and 0.083 MHz and 0.090 MHz (7.469 km and 8.162 km) at Fairford and San Vito, respectively.

When the IRI UP method runs with the complete variogram selection procedure, its performance shows a noticeable improvement at the considered truth-sites, for both *fo*F2 and *hm*F2. In fact, in this case, NRMSE values for *fo*F2 (*hm*F2) are 8.25% (10.44%) and 8.70% (8.62%) at Fairford and San Vito, respectively, while MMDAV values for *fo*F2 (*hm*F2) are 0.122 MHz (6.844 km) and 0.084 MHz (6.959 km) at Fairford and San Vito, respectively.

NRMSE and MMDAV values for IRI UP are systematically smaller than ones obtained for a fixed variogram model and IRI; that proves the effectiveness of the variogram selection procedure in providing more precise and accurate results.

In general, from Figs. 4.22 and 4.23 it emerges that IRI UP performs slightly better for *fo*F2 than for *hm*F2. This happens for different reasons:

1) *fo*F2 assimilated data are generally more reliable than the corresponding *M*(3000)F2 data. This is mostly due to the fact that ionograms can be characterized by multiple reflections of the F2 layer, and when this happens an autoscaling program can be misled and the second-order reflection can be identified as the real trace; in this case the *fo*F2 value is usually not affected by a significant error, while the *M*(3000)F2 is significantly wrongly scaled (Scotto and Pezzopane 2008);

2) the spherical harmonic expansion (chapter 2) used to describe the *fo*F2 and *M*(3000)F2 spatial behavior stops when the maximum order of the harmonics is $K = 76$, for *fo*F2, and $K = 49$, for *M*(3000)F2. This means that *fo*F2 maps present a higher spatial resolution than *M*(3000)F2 ones;

3) *hm*F2 is calculated applying the empirical formula (2.7) of Bilitza et al. (1979), and this implies that the error characterizing *hm*F2 depends on the errors relative to *fo*F2, *M*(3000)F2, *fo*E, and $R_{12eff}$; therefore, the error propagation leads to an *hm*F2 prediction error which is intrinsically larger than that of *fo*F2;

4) last but not least, *fo*F2 predictions are based on the $IG_{12eff}$ index, which is an ionospheric index because it is "built" just starting from *fo*F2 values recorded at several ionospheric stations (Liu et al. 1983), while *M*(3000)F2 predictions, which come into play to calculate *hm*F2, are not based on an ionospheric index, but on $R_{12eff}$.

Another positive aspect of the IRI UP method is that the experimental variogram has a higher spatial variability the larger is the number *n* of data assimilated, so that corresponding maps are statistically more reliable and are not discarded. This situation is clear from the results of Fig. 4.24 where, for each kind of variogram model, a decreasing trend of the number of discarded maps is observed starting from the storm number 17, corresponding to 2012, i.e., the year from which the number of available ionospheric stations maximize and are more uniformely distributed over the European region (see Table 4.9 and Fig. 4.1). It is also to be noted that stationary variogram models (Gaussian, spherical, and exponential) are more sensitive than non-stationary ones to the number *n* of assimilated data; this is probably due to their more complex mathematical formulation which requires a greater value of *n* to represent more adequately the spatial correlations between measurements, for every spatial scale. In fact, considering all storms as a whole, it results that percentages of rejected stationary variograms for *fo*F2 (Fig. 4.24c) and *hm*F2 (Fig. 4.24d) are greater than those related to non-stationary variograms (linear and power). This result is probably due to the cumulative effect of the first 16 storms listed in Table 4.10, which are relative to years characterized by a low value of *n*. Nevertheless, in Figs. 4.24a and 4.24b, the trend observed starting from the storm number 17, corresponding to years for which *n* is increased ($n = 12$), suggests that the percentages of stationary and non-stationary discarded variogram models may converge to similar values with increasing number of stations.

The winning percentages shown in Figs. 4.25b and 4.26b indicate that among the linear, power, spherical, and exponential variogram models there is not a clear predominance of one model over another, and that the Gaussian variogram model shows the higher percentages, $\approx 40\%$ and 29% for $IG_{12eff}$ and $R_{12eff}$, respectively. This fact is surprising, because considering Fig. 4.24 the Gaussian model is the most rejected one. This means that for the Gaussian variogram model is more difficult to pass the $Q_1$ and $Q_2$ statistical tests, but when this happens, it is more

likely that the Gaussian variogram will be the best one according to the variogram selection procedure. A fact that clearly emerges also when each single storm is considered (Figs. 4.25a and 4.26a).

## Summary

In this chapter, the application of the IRI UP method, by assimilating *fo*F2 and *M*(3000)F2 values measured by ionosondes over the European region, has been described and statistically evaluated under disturbed geomagnetic conditions.

The IRI UP method relies on the assimilation of *fo*F2 and *M*(3000)F2 data routinely recorded by a network of ionosonde stations, to calculate updated effective indices $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ at each station of the network; then the Universal Kriging method is applied to generate maps of these indices. Maps of these updated indices are then used to update the IRI output in terms of *fo*F2 (directly through $IG_{12\text{eff}}$) and *hm*F2 (indirectly through *M*(3000)F2 and *fo*E values, dependent on $R_{12\text{eff}}$, and *fo*F2 values, dependent on $IG_{12\text{eff}}$).

The IRI UP method has been tested by comparing its output with IRI (STORM option on), and IRTAM (only for what concerns the St. Patrick storm occurred in 2015), at Fairford and San Vito testing stations, for 30 periods embedding geomagnetic storms occurred between 2004 and 2016. The statistical analysis shows that IRI UP for *fo*F2 presents the best precision and accuracy. Concerning *hm*F2, the output of IRI UP is slightly better than IRI and similar to that of IRTAM. Analysis of statistical distributions and scatter plots pointed out that IRI UP, both for *fo*F2 and *hm*F2, is precise and accurate. Visual analysis of *fo*F2 and *hm*F2 maps confirms the goodness of the IRI UP method, highlighting its ability to catch rapid and small-scale electron density changes occurring especially under disturbed conditions.

The results achieved in this chapter prove that reliable and trustworthy updated maps of the main ionospheric characteristics can be provided with a satisfactory degree of confidence, especially under moderate, strong and severe geomagnetic storm conditions. This means that IRI UP method represents an interesting approach for Space Weather purposes in the ionospheric domain, for any region characterized by an adequately distributed network of ionosondes. One of the advantages of the IRI UP method is the short computational time (from 1 to 2 minutes, depending on the spatial resolution needed, on a ordinary notebook) required to run it, an essential feature for a nowcasting approach.

# 5. On the assimilation of ground-based GNSS receiver derived vTEC measurements in the IRI UP method

In this chapter a different implementation of IRI UP is presented to assimilate vertical total electron content (vTEC) measurements from a Global Navigational Satellite Systems (GNSS) receiver network rather than ionosonde derived F2 layer ionospheric characteristics (as described in chapter 4).

To accomplish this task, a mathematical procedure for obtaining *fo*F2 and *M*(3000)F2 values from assimilated vTEC measurements has been implemented. Mathematical relationships between F2 layer characteristics and vTEC values have been derived using South African co-located ionosondes and GNSS stations. The same procedure can be applied with profit in each region where such data are available. This allows for the calculation of effective indices $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ which are used to update the IRI model.

Being a data assimilation procedure, the IRI UP method strongly relies on both the availability and the spatial distribution of assimilated observations. In regions well covered by ionosonde stations, like Europe, the IRI UP method can be successfully applied (as demonstrated in chapter 4). This situation however represents an exception rather than normality. Many terrestrial regions are poorly covered by ionosondes but are fairly well covered by GNSS receiver stations; this is the case of the South African region where 4 ionosondes are available compared to 68 GNSS stations. This makes very attractive the prospective of using the huge mass of vTEC measurements collected by GNSS receiver networks.

The goodness of the proposed IRI UP method, based on assimilated vTEC values, has been tested for several quiet and disturbed days in 2017, comparing *fo*F2 and *hm*F2 values modeled by both IRI and IRI UP with those measured at ionosonde testing stations.

In section 5.1, South African ionosonde and GNSS receiver networks are presented. Section 5.2 describes the mathematical procedure developed to model *fo*F2 and *MUF*(3000)F2 values from vTEC measured data; it also contains a test of the proposed procedure. Section 5.3 describes the assimilation procedure and the IRI UP steps for obtaining *fo*F2 and *hm*F2 updated values. Finally, section 5.4 contains a detailed validation of the proposed method, for different seasons and geomagnetic conditions.

Results described in this chapter are the subject of a paper currently under review in the *Space Weather* journal:

*Pignalberi, A., Habarulema, J.B., Pezzopane, M., Rizzi, R., On the development of a method for updating an empirical climatological ionospheric model by means of assimilated vTEC measurements from a GNSS receiver network*

## 5.1 vTEC and F2 layer ionospheric characteristics datasets preparation

### 5.1.1 Ionosonde derived F2 layer ionospheric characteristics data

Ionosonde data are downloaded from the Digital Ionogram DataBASE (Reinisch and Galkin 2011) by means of the SAO Explorer software developed by the University of Massachusetts, Lowell (http://ulcar.uml.edu/SAO-X/SAO-X.html). Ionograms recorded at the ionospheric stations listed in Table 5.1 (see also Fig. 5.1) are recorded by DPS digisondes (Bibl and Reinisch 1978), and are autoscaled by the Automatic Real-Time Ionogram Scaler with True height analysis (ARTIST) software (Reinisch and Huang 1983; Galkin and Reinisch 2008). The reliability of ionosonde's autoscaled data is based on the Confidence Score (C-Score) parameter (see http://www.ursi.org/files/CommissionWebsites/INAG/web-73/confidence_score.pdf) accompanying the autoscaling performed by the ARTIST software.

Ionosonde derived dataset consists of autoscaled F2 layer ionospheric characteristics *fo*F2, *MUF*(3000)F2, and *hm*F2. Those ionospheric characteristics have been routinely recorded by the ionosondes listed in Table 5.1 with a 15 minutes sounding repetition rate for several years. The data period ranges of 2006-2017, 2008-2017, 2004-2017 and 2003-2017 are used for Grahamstown, Hermanus, Louisvale and Madimbo, respectively.

| **Ionosonde station** | **Lat [°]** | **Lon [°]** | **GNSS stations** | **Lat [°]** | **Lon [°]** | **Distance [km]** |
|---|---|---|---|---|---|---|
| Grahamstown | 33.30° S | 26.50° E | Grahamstown | 33.30° S | 26.50° E | 0 |
| Hermanus | 34.42° S | 19.22° E | Hermanus | 34.42° S | 19.22° E | 0 |
| Louisvale | 28.50° S | 21.20°E | Upington | 28.40°S | 21.25°E | 12 |
| Madimbo | 22.39° S | 30.88°E | Thohoyandou | 23.08° S | 30.38° E | 92 |

**Table 5.1:** South African ionosondes and co-located GNSS stations used to establish the relationship between *fo*F2 and vTEC. Their geographical position is shown along with the distance between the ionosonde and GNSS receiver stations, in the last column.

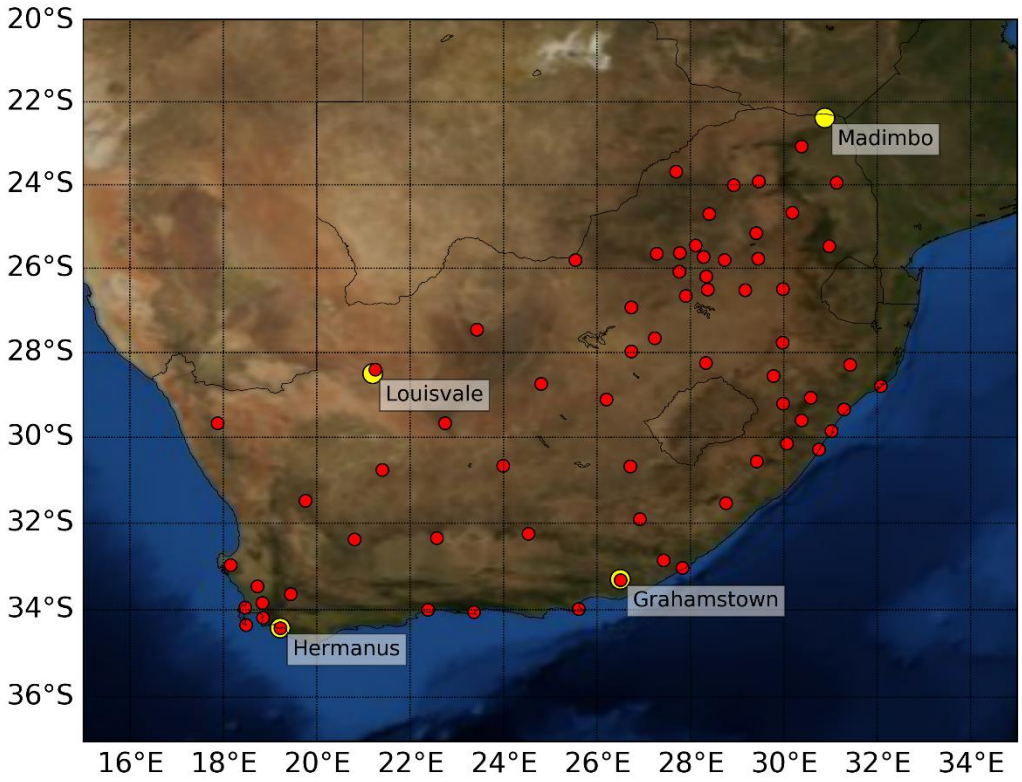


**Figure 5.1:** Geographical distribution of the South African TrigNet GNSS receivers network (red circles), and that of the South African ionosonde network (yellow circles). Ionosonde stations are also labelled with their location names.

### 5.1.2 Deriving vTEC calibrated values from TrigNet GNSS receivers' network

TrigNet is a network of continuously operating GNSS ground-based stations that are located throughout South Africa at an interstation spacing of between 80 km and 300 km (http://www.trignet.co.za/). Currently, TrigNet network consists of 68 GNSS ground-based stations (red circles in Fig. 5.1) equipped with a GNSS receiver and a geodetic choke ring antenna that are used to record GNSS observables. Data are continuously streamed from each TrigNet base station to the TrigNet control centre located at the National Geospatial Information (NGI) in Cape Town. The TrigNet control centre processes the data and publishes the post processing data files on the website http://www.trignet.co.za/.

In particular, daily Receiver INdependent EXchange (RINEX) formatted files containing L1 and L2 code and carrier phase data at a 30 second epoch rate, are used. RINEX files are downloaded by means of TrigNet File Transfer Protocol (FTP), freely available at ftp://ftp.trignet.co.za. after registration.

Because of the measurement geometry between GPS satellite and ground-based receiver, TEC is determined at a slant column dependent on the satellite elevation angle. Thus, what is calculated is the slant TEC (sTEC) value, which is then mapped to the local vertical direction to provide the vertical TEC (vTEC).

Calibration of vTEC values is performed by means of the GPS-TEC software (v. 2.9.6) developed by the Boston College (Seemala and Valladares 2011). The software uses both L1 and L2 signals to calculate the relative sTEC by removing errors due to clock biases and the tropospheric water vapour effect. To obtain absolute sTEC values,

the differential satellite biases (published by the University of Bern) are included along with the receiver bias values that are calculated by minimizing the sTEC variability between 2 and 6 LT (Local Time) (Doherty et al. 2004; Ssessanga et al. 2014). The equivalent vTEC at the Ionospheric Pierce Point (IPP) altitude of 350 km is calculated assuming the thin spherical shell model (Ciraolo and Spalla 1997). Possible outliers in vTEC values are filtered out applying the procedure described in section 5.1.3. For each GNSS station, at a particular time, a single vTEC value is obtained averaging vTEC values calculated for each satellite in view, whose number depends on the chosen elevation angle (Ssessanga et al. 2014). In this study, an elevation angle of 50° has been chosen. Making this choice, GPS satellites whose IPP is located inside a circle with a radius of about 300 km, centered on the ground-based GNSS receiver, are used in the vTEC averaging process.

### 5.1.3 Filtering out outliers in GNSS derived vTEC and ionosonde derived F2 layer ionospheric characteristics datasets

The calibration process to obtain vTEC values, starting from information embedded in GPS signals, is subject to many error sources due mainly to transmitter and receiver biases, and to the geometric approximation made to calculate vTEC values from sTEC ones. These, and other minor error sources, lead to an uncertainty in the absolute vTEC value which can be of the order of some TECU (TEC Units whose value is 1 TECU=$10^{16}$ el/m$^2$) also for quiet and stationary conditions (Brunner and Gu 1991; Bassiri and Hajj 1992; Langley 1996; Ciraolo et al. 2007). The uncertainty can dramatically increase during storm disturbed conditions (Brunini et al. 2003, 2005), during hours when important horizontal gradients in the electron density can arise (dawn and dusk, around the solar terminator path) (Somsikov 2011; Makarevich and Nicolls 2013), or for ionospheric regions where important horizontal gradients are already present (like in the Equatorial Ionization Anomaly region, (Brunini and Azpilicueta 2010; Sunehra et al. 2014).

Because it is difficult to estimate such errors, the Median Absolute Deviation (MAD) method (Huber and Ronchetti 2009) is applied to obtain a robust statistic of the calibrated vTEC values. MAD is a robust measure of the variability of a univariate sample of quantitative data which can be successfully applied in the process of removing bad sample data (outliers and spikes) from a time series. MAD is a measure of the statistical dispersion more robust than that given by the standard deviation, being less dependent on outliers in the tails of the distribution. In fact, in the standard deviation, the differences from the mean are squared, so that large deviations are weighted heavily and outliers can heavily influence it. Concerning MAD, the deviations of a small number of outliers are irrelevant being based on calculated medians. For symmetric distributions MAD is asymptotically equivalent to one-half of the interquartile distance.

MAD filter is applied to calibrated vTEC values, according to the following procedure:

1. Calibrated vTEC values for each GPS satellite in view are calculated for the entire day to which the RINEX file refers, with a 30 seconds time resolution;
2. Daily calibrated vTEC values are organized in 24 bins (one for each hour of the day) depending on the Universal Time (UT) at which they are measured;
3. The median of vTEC values falling in each bin is calculated:

$$M_h = \text{median}\left(\text{vTEC}_{h,i}\right) , \tag{5.1}$$

   where the subscript $h \in [0,1,\ldots,22,23]$ refers to the UT hour of the day, and the subscript $i$ runs on all the vTEC values falling in the bin $h$;
4. For each bin, the median of the absolute differences between measured vTEC values and $M_h$ (Eq. (5.1)) is calculated. Such median is what is called MAD:

$$MAD_h = \text{median}\left(\left|\text{vTEC}_{h,i} - M_h\right|\right); \tag{5.2}$$

5. To make MAD a consistent estimator of the dispersion associated to the particular distribution shown by data, it is necessary to multiply it by a scale factor *k* dependent on the distribution. In the hypothesis that absolute differences between measured vTEC values and their median are normally distributed, it can be demonstrated that $k \cong 1.4826$ is the right scale factor (Huber and Ronchetti 2009):

$$MAD_h = k \cdot MAD_h \cong 1.4826 \cdot MAD_h \ ; \tag{5.3}$$

6. For each of the 24 hourly bins, vTEC values which do not comply with the following conditions are discarded:

$$M_h - s \cdot MAD_h \leq \mathrm{vTEC}_{h,i} \leq M_h + s \cdot MAD_h \ , \tag{5.4}$$

   where the multiplicative factor *s* is subjectively chosen based on how much selective one wants the MAD filter to be. The choice of the *s* factor is described in the next section;

7. Steps described by points 3-6 are repeated for each of the 24 hourly bin;

8. vTEC values which have passed the MAD filter procedure are used to calculate the average vTEC values, as described in section 5.1.2.

When applying the MAD filter to the vTEC dataset, and evaluating the reliability of the autoscaled *fo*F2 and *MUF*(3000)F2 dataset through the C-Score parameter, one must choose the values of the parameter *s* of Eq. (5.4) and of the C-Score. Different choices have been tested by calculating the RMSE (Eq. (4.5)), the NRMSE (Eq. (4.6)), and the Pearson correlation coefficient $\rho$ (Eq. (4.7)), between modeled (following the procedure described in section 5.3) and measured (by the Hermanus ionosonde, for years 2011-2017) *fo*F2 and *MUF*(3000)F2 values.

For this analysis Hermanus ionosonde derived data and vTEC data from a co-located GNSS receiver have been used. C-Score threshold values equal to $[0,50,75,90]$ have been tested, discarding those ionograms for which C-Score $\leq$ threshold. Obviously, the case for which the C-Score threshold value is set to 0 means that no filtering procedure has been applied on ionosonde data. The filter will be more and more selective by increasing the C-Score threshold value.

In the case of MAD filter, three options have been tested:

- NO MAD filter, which means that no filtering procedure has been applied on derived vTEC data;
- 1 MAD (*s*=1), where the *s* parameter has been set to 1;
- 2 MAD (*s*=2), where the *s* parameter has been set to 2.

Obviously, the 1 MAD option is more selective than the 2 MAD one.

Results for each possible permutation of both thresholds are listed in Table 5.2.

| C-Score threshold | MAD threshold | # Data | Ionospheric characteristics | RMSE [MHz] | NRMSE [%] | $\rho$ |
|---|---|---|---|---|---|---|
| 0 | NO MAD filter | 136417 | *fo*F2 | 0.812 | 13.838 | 0.948 |
| | | | *MUF*(3000)F2 | 3.333 | 18.025 | 0.904 |
| 50 | NO MAD filter | 129899 | *fo*F2 | 0.801 | 13.531 | 0.951 |
| | | | *MUF*(3000)F2 | 3.300 | 17.731 | 0.902 |
| 75 | NO MAD filter | 112941 | *fo*F2 | 0.787 | 12.948 | 0.952 |
| | | | *MUF*(3000)F2 | 3.295 | 17.284 | 0.906 |
| 90 | NO MAD filter | 86983 | *fo*F2 | 0.749 | 12.344 | 0.957 |
| | | | *MUF*(3000)F2 | 3.176 | 16.771 | 0.907 |
| 0 | 1 MAD ($s$=1) | 129428 | *fo*F2 | 0.809 | 13.825 | 0.949 |
| | | | *MUF*(3000)F2 | 3.301 | 17.921 | 0.905 |
| 50 | 1 MAD ($s$=1) | 123397 | *fo*F2 | 0.797 | 13.515 | 0.951 |
| | | | *MUF*(3000)F2 | 3.269 | 17.632 | 0.908 |
| 75 | 1 MAD ($s$=1) | 107262 | *fo*F2 | 0.782 | 12.912 | 0.953 |
| | | | *MUF*(3000)F2 | 3.262 | 17.177 | 0.908 |
| 90 | 1 MAD ($s$=1) | 82701 | *fo*F2 | 0.742 | 12.271 | 0.958 |
| | | | *MUF*(3000)F2 | 3.140 | 16.641 | 0.913 |
| 0 | 2 MAD ($s$=2) | 135020 | *fo*F2 | 0.799 | 13.612 | 0.950 |
| | | | *MUF*(3000)F2 | 3.298 | 17.836 | 0.906 |
| 50 | 2 MAD ($s$=2) | 128592 | *fo*F2 | 0.787 | 13.297 | 0.952 |
| | | | *MUF*(3000)F2 | 3.264 | 17.540 | 0.909 |
| 75 | 2 MAD ($s$=2) | 111799 | *fo*F2 | 0.772 | 12.713 | 0.954 |
| | | | *MUF*(3000)F2 | 3.260 | 17.101 | 0.909 |
| 90 | 2 MAD ($s$=2) | 86164 | *fo*F2 | 0.732 | 12.075 | 0.959 |
| | | | *MUF*(3000)F2 | 3.137 | 16.571 | 0.909 |

**Table 5.2:** Number of couples of values (vTEC/*fo*F2 and vTEC/*MUF*(3000)F2) which passed both C-Score and MAD filters and statistical results for each possible permutation of C-Score and MAD threshold values, for both *fo*F2 and *MUF*(3000)F2. Data collected by Hermanus co-located GNSS and ionosonde stations for years 2011-2017 have been used.

Some important considerations can be drawn from Table 5.2:

- application of the C-Score filter when set at 50, 75 and 90 reduces the amount of data. A sharp decrease is observed for the C-Score threshold value of 90;
- from a statistical point of view, considering data with a higher C-Score improves the performance of the method;
- the 2 MAD threshold filter gives the best performance despite being less selective than the 1 MAD option; furthermore, the 2 MAD option discards much less data than the 1 MAD one.

Results shown by Table 5.2 should point towards the choice of a C-Score threshold equal to 90 and 2 MAD. But, because of the huge loss of data caused by a C-Score equal to 90, it has been decided to set the C-Score threshold to 75, since, as it will be clearer in the next section, the number of available data is of fundamental importance for deriving reliable relationships between vTEC and F2 layer characteristics.

### 5.1.4 vTEC dataset building

Calibration and filtering procedures of vTEC values, described in sections 5.1.2 and 5.1.3, are applied to both assimilated and not-assimilated vTEC values. The *assimilated* are the vTEC values which are used to update the background climatological ionospheric model. Instead, the *not-assimilated* vTEC values are the values of the four GNSS stations listed in Table 5.1, used to derive the relationships between vTEC and F2 layer ionospheric characteristics.

The data period ranges of 2006-2017, 2008-2017, 2004-2017 and 2003-2017 are used for Grahamstown, Hermanus, Upington and Thohoyandou GNSS receiver stations, respectively. These periods are those for which data from co-located ionosondes are simultaneously available.

Because of the 15 minutes time resolution of F2 layer characteristics time series measured by ionosondes, a one to one comparison between ionosonde and GNSS derived data would result in a dramatic decimation of GNSS dataset. This decimation, besides causing a loss of a huge quantity of vTEC data, might be dangerous in the case of isolated vTEC outliers recorded exactly at the same time of ionosonde data. To avoid this waste of information and the occurrence of isolated outliers, the whole vTEC time series is used applying a sliding window centered at minutes 0, 15, 30, and 45 of each hour (according to the sounding repetition rate of the ionosonde), and only vTEC values falling in the window are used to calculate the average vTEC value. Specifically, vTEC values falling in the window (at most 31 in this case) are multiplied by a Gaussian weight $W_i$:

$$W_i = \exp\left[-\frac{(t_i - t_0)^2}{2\sigma^2}\right], \tag{5.5}$$

where $t_0 \in (0,15,30,45)$ is the time of one of the four possible points at the center of the window, $t_i$ is the time of each point falling in the window, and $\sigma$ is a parameter acting as the variance in the Gaussian distribution. In this case, $\sigma = 150$ seconds (i.e. two and a half minutes); the two points at the edge of the window, distant $3\sigma$ from the center, have a very low weight, while the weight due to points very close to the center of the window is by far higher.

Once each vTEC value falling in a definite window has been multiplied by the appropriate weight, the weighted mean is calculated and assigned to the center of the window:

$$\overline{(\mathrm{vTEC}_0)} = \frac{\sum_{i=1}^{31} \mathrm{vTEC}_i \cdot W_i}{\sum_{i=1}^{31} W_i} \ . \tag{5.6}$$

## 5.2 Deriving mathematical relationships between vTEC and *fo*F2/*MUF*(3000)F2

It has been demonstrated by many authors that a strong correlation exists between vTEC and the F2 layer maximum of electron density *Nm*F2 (Spalla and Ciraolo 1994; Krankowski et al. 2007). From a mathematical point of view this is understandable because there should exist some relation between the integral of a function calculated over its domain (the vTEC value in our case, where the function is the electron density and the domain goes from the base of the ionosphere to the GPS satellite height) and the absolute maximum of the function, if the function is continuous and regular. If the continuity of the electron density as a function of the height is a hypothesis well verified in many cases, the hypothesis of regularity is a bit trickier. In fact, as it is known, the ionosphere is characterized by several layers whose properties heavily change with the hour of the day, the season, the solar activity level, and the geomagnetic activity (see chapter 1). This means that the shape of the vertical electron density profile, thus the shape of the function, changes with time. Thus, when finding a relationship between a function whose shape can rapidly change and its integral, considering only the absolute maximum of the function represents certainly an important simplification. Aware of these limitations, a strong correlation between vTEC and *Nm*F2 was found in the past (Kouris et al. 2004; Ssessanga et al. 2014), due to the very important contribution that the F2 region gives to the vTEC value, above all during daytime when the most important contributor to the ionospheric electron density is the solar radiation. During night time, things become more difficult because the contribution of the plasmasphere to the vTEC is not negligible compared to that provided by the ionosphere (Klimenko et al. 2015). Furthermore, difficulties arise when the ionosphere and the plasmasphere undergo very rapid changes in both time and space as during dawn and dusk (when the solar terminator sets very important horizontal gradients in the electron density distribution), or under very disturbed geomagnetic conditions. A complicated relation between *hm*F2 and its integral is expected. This is why one does not try to find a relationship between *hm*F2 and vTEC. Instead, a meaningful relation can be found between *MUF*(3000)F2 and vTEC, being *MUF*(3000)F2 related to the intensity and the shape of the F2 region around its maximum; moreover, *MUF*(3000)F2 can be used to calculate *M*(3000)F2 values needed to calculate $R_{12\mathrm{eff}}$ indices through the IRI UP method (chapter 3).

Therefore, *fo*F2 and *MUF*(3000)F2 are modeled as a function of vTEC. For this purpose, vTEC time series from GNSS stations and *fo*F2 and *MUF*(3000)F2 values from ionosondes are used, for the four co-located couples of stations listed in Table 5.1.

The data for the four co-located stations listed in Table 5.1 are joined to obtain relationships valid for the entire South African region. This choice is justified by the homogeneous coverage of the South African region provided by these stations (Fig. 5.1), and by the limited spatial extension of the region. Furthermore, the spatial characterization of the physical quantities under investigation is provided by the network of GNSS stations (Fig. 5.1) from which vTEC data are assimilated.

From a preliminary comparison between vTEC values expressed in TECU and *fo*F2 (*MUF*(3000)F2) values expressed in MHz, collected by co-located ionosonde and GNSS receiver stations, emerged that non-linear functions are more suitable to describe their mathematical dependence than the linear one.

Instead of using vTEC values expressed in TECU, one use directly el/m$^2$. Similarly, instead of *fo*F2 in MHz, one consider *Nm*F2, expressed in el/m$^3$ (*fo*F2 and *Nm*F2 are linked through the relationship *fo*F2[MHz]=(*Nm*F2[el/m$^3$]/1.24·10$^{10}$)$^{1/2}$) and, instead of *MUF*(3000)F2 in MHz, starting from the modified secant law (Davies 1990), one uses *MUF*(3000)F2 expressed in el/m$^3$ according to the following expression:

$$MUF(3000)\mathrm{F2}_{\mathrm{el/m^3}} = \frac{\cos\left[\arctan\left(\dfrac{\sin\dfrac{\theta}{2}}{1+\dfrac{h_0}{R_E}-\cos\dfrac{\theta}{2}}\right)\right]}{k}\cdot 1.24\cdot 10^{10}\cdot\left(MUF(3000)\mathrm{F2}_{\mathrm{MHz}}\right)^2 , \qquad (5.7)$$

where $k$ is a correction factor, varying from 1.000 and 1.200, to account for the ionospheric curvature and that is equal to 1.114 in the calculation of the *MUF*(3000)F2; $h_0$ is the height of the reflecting layer, which one approximates to *hm*F2; $\theta$ is the angle at the center of the Earth subtended by the path (which for 3000 km is equal to 0.471 radians), and $R_E$ is the Earth's radius.

The reasons for these choices are that linear relationships, achievable working with vTEC in el/m$^2$, and *Nm*F2 and *MUF*(3000)F2 in el/m$^3$, give better performances (not shown here) than non-linear relationships achievable working with vTEC in TECU, and *fo*F2 and *MUF*(3000)F2 in MHz. Furthermore, the base-10 logarithm of *Nm*F2, *MUF*(3000)F2, and vTEC is considered.

Hence, the linear relationships that are used in the fitting procedure are the following:

$$\log_{10} NmF2 = a_{NmF2} \cdot \log_{10} \text{vTEC} + b_{NmF2}\,, \tag{5.8}$$

$$\log_{10} MUF(3000)F2 = a_{MUF(3000)F2} \cdot \log_{10} \text{vTEC} + b_{MUF(3000)F2}\,, \tag{5.9}$$

where $a_{NmF2}$ and $a_{MUF(3000)F2}$ are the slopes, while $b_{NmF2}$ and $b_{MUF(3000)F2}$ are the intercepts, of the linear relationships used, for both *Nm*F2 and *MUF*(3000)F2.

In Fig. 5.2, two-dimensional density plots are shown of $\log_{10} NmF2$ vs $\log_{10}$ vTEC, and $\log_{10} MUF(3000)F2$ vs $\log_{10}$ vTEC, together with histograms of the residuals expressed in terms of *fo*F2 and *MUF*(3000)F2 (in MHz), obtained by using all data recorded from 2003 to 2017 by the four co-located couples of South African stations of Table 5.1. As one can see, using the approach described above, it is possible to linearly relate *Nm*F2 (*MUF*(3000)F2) and vTEC values, with a good degree of confidence, using a single formula, although values measured for different hours of the day, for different seasons, and for different years (hence, for different solar activity levels) have been used.

Some interesting features emerge from a careful inspection of Fig. 5.2:

- the linear fitting procedure turns out to be very accurate for *fo*F2 with a residuals mean of 0.062 MHz, and less accurate for *MUF*(3000)F2 that shows a residuals mean of 0.240 MHz;
- both *fo*F2 and *MUF*(3000)F2 show important dispersion around the fitting linear function, 0.754 MHz for *fo*F2 and 2.782 MHz for *MUF*(3000)F2. In terms of NRMSE, a value of 13.5 % is obtained for *fo*F2 and 15.5 % for *MUF*(3000)F2;
- the dispersion is particularly accentuated during nighttime hours, when lower *Nm*F2 (*MUF*(3000)F2) and vTEC values are reached, compared to daytime hours where higher values are recorded;
- good correlation coefficients of 0.948 and 0.921 are obtained for *fo*F2 and *MUF*(3000)F2, respectively.

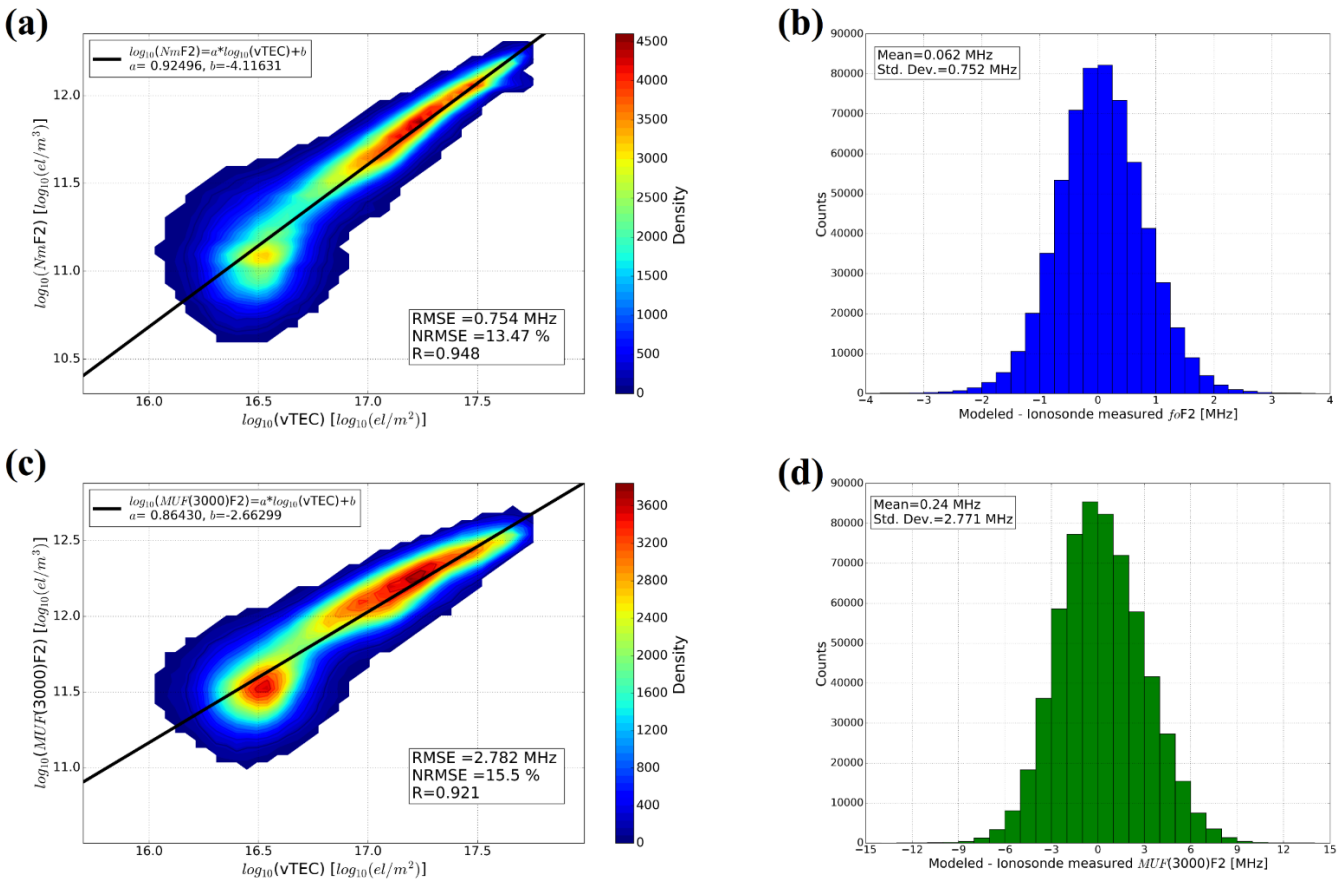


**Figure 5.2:** Two-dimensional density plots of (a) $\log_{10} NmF2$ vs $\log_{10}$ vTEC, and (c) $\log_{10} MUF(3000)F2$ vs $\log_{10}$ vTEC, with corresponding fitting linear functions (solid black lines), obtained by using all data recorded from 2003 to 2017 by the four co-located couples of South African stations of Table 5.1. Histograms of the residuals between modeled and measured *fo*F2 and *MUF*(3000)F2 values, are respectively shown in panels (b) and (d).

Results of Fig. 5.2 show that the linear fitting procedure leans on solid foundations. Additionally, to take into account the daily and seasonal variability shown by both vTEC and F2 layer characteristics values, the dataset has been divided into sub-datasets binned as a function of the UT hour of the day (LT=UT+2 for the South African region) and of the month of the year. So, for each of the 288 obtained sub-datasets (24 hours of the day for each of the 12 months of the year) the fitting procedure has been applied to calculate the coefficients of each sub-dataset. Corresponding results are shown in Fig. 5.3.

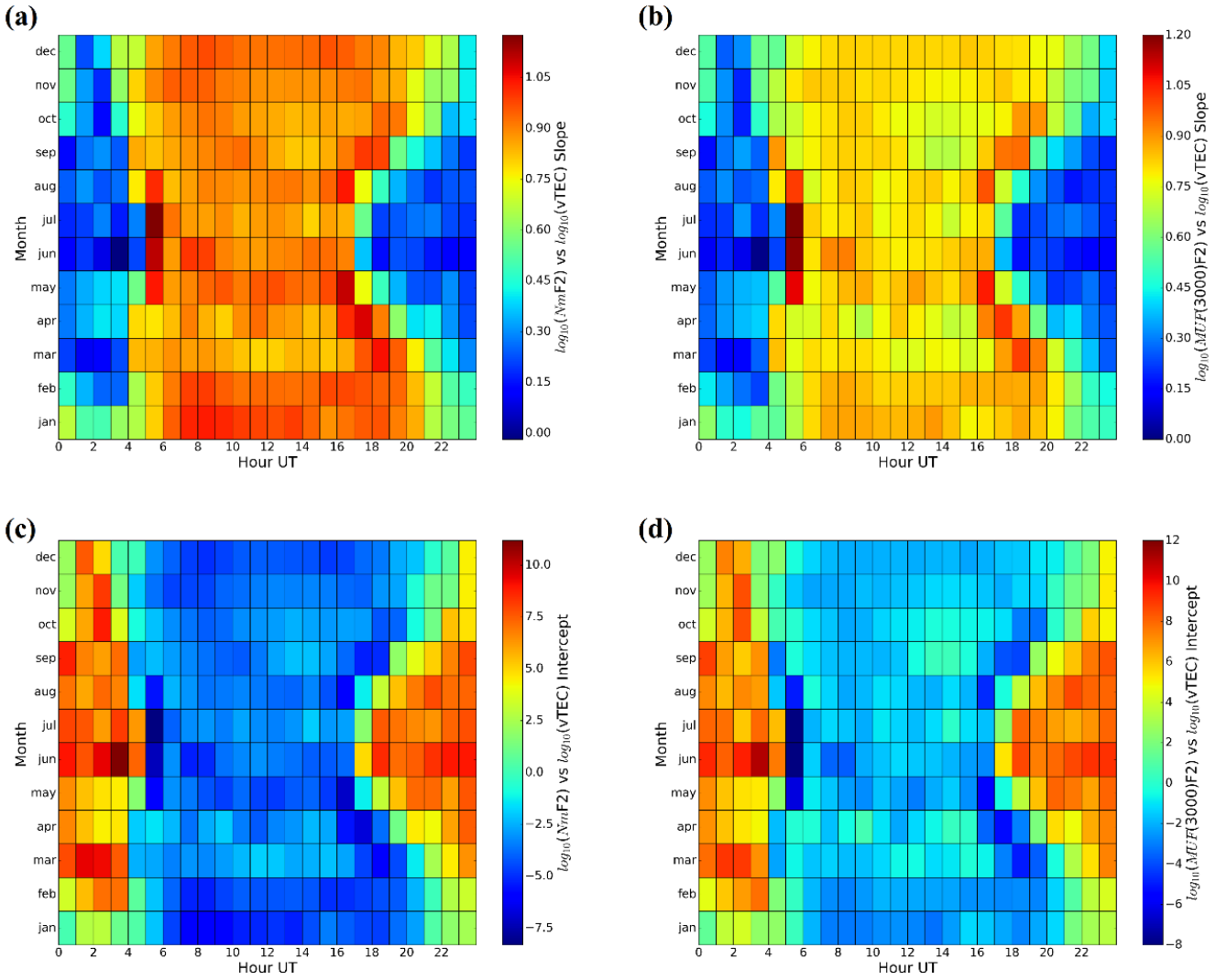


**Figure 5.3:** Matrices of the linear fitting coefficients calculated using Eqs. (5.8) and (5.9), for measurements taken at a specific UT hour (x axis), for a specific month (y axis). Panels (a) and (b) show slope coefficients for *Nm*F2 and *MUF*(3000)F2. Panel (c) and (d) show intercept coefficients for *Nm*F2 and *MUF*(3000)F2, respectively.

Matrices of Fig. 5.3 highlight some interesting features:

- hourly and seasonal variations in fitting coefficients (slope and intercept) are clearly visible and of great magnitude;
- fitting coefficients show a very different behaviour between nighttime (uniforma anche questa parola) and daytime hours. Slopes are higher during day, while intercepts are higher during night;
- the seasonal variability is quite evident considering the different daily behaviour across different months, and the solar terminator variability is well reproduced.

Results shown in Fig. 5.3 strongly support the choice made to consider the hourly and seasonal variations in describing the relation between vTEC and F2 layer characteristics.

### 5.2.1 Calibration of the proposed vTEC vs F2 layer characteristics modeling procedure

Using the whole dataset coming from the four couples of co-located GNSS and ionosonde stations listed in Table 5.1, *Nm*F2 and *MUF*(3000)F2 values are modeled through the linear relationships of Eqs. (5.8) and (5.9), selecting the appropriate fitting coefficients from matrices of Fig. 5.3, according to the UT hour and month at which measured vTEC values are available. Then, these *Nm*F2 and *MUF*(3000)F2 modeled values are compared to measured ones collected by ionosonde stations. Statistical quantities RMSE, NRMSE, and $\rho$ (Eqs. (4.5)-(4.7)) are then calculated, for each of the 288 considered bins. Fig. 5.4 shows the results obtained from this calibration procedure and highlights the following important features:

- Clearly different statistical differences are visible for nighttime and daytime hours. Lower NRMSE values characterize daytime hours (6-8 % for *fo*F2, 7-9 % for *MUF*(3000)F2), while higher NRMSE values characterize nighttime hours (10-20 % for *fo*F2, 12-20 % for *MUF*(3000)F2);

- Particularly high NRMSE values characterize solar terminator hours in winter season (May-August) and during late spring-early summer for both *fo*F2 and *MUF*(3000)F2;

- For both characteristics very high correlation coefficients are visible for daytime hours, while low correlation coefficients characterize nighttime hours.

Results of Fig. 5.4 highlight the potential and the limitation of the proposed procedure in modeling F2 layer characteristics from GNSS derived vTEC values:

- As expected, the fitting procedure is very accurate during daytime hours when the F2 region gives the major contribution to vTEC (Ssessanga et al., 2014). On the contrary, during nighttime, when the plasmaspheric contribution to vTEC becomes important, the correlation between F2 region peak parameters and vTEC decreases (Klimenko et al., 2015), and the performance of the fitting procedure degrades;

- Solar terminator hours are very difficult to model because of their intrinsic ionospheric and plasmaspheric variability. Solar terminator sets up significant horizontal electron density gradients at a small spatial scale. Thus, joining measurements derived from the four co-located South-African stations can produce a significant dispersion in hours characterized by rapidly changing conditions. This could be the reason behind the high *NRMSE* values characterizing solar terminator hours;

- Mean values of the elements of matrices in panels (a), (b), (c), and (d) of Fig. 5.4 are:
  $\overline{\text{RMSE } fo\text{F2}} = 0.553\,\text{MHz}$, $\overline{\text{RMSE } MUF(3000)\text{F2}} = 1.897\,\text{MHz}$, $\overline{\text{NRMSE } fo\text{F2}} = 11.662\,\%$,
  and $\overline{\text{NRMSE } MUF(3000)\text{F2}} = 12.105\,\%$.
  The comparison of these values with the ones calculated using the relationships of Fig. 5.2
  (RMSE $fo\text{F2} = 0.754$ MHz, RMSE $MUF(3000)\text{F2} = 2.782$ MHz, NRMSE $fo\text{F2} = 13.470$ %,
  NRMSE $MUF(3000)\text{F2} = 15.500$ %), shows that a significant improvement has been achieved by considering the daily and seasonal dependence.

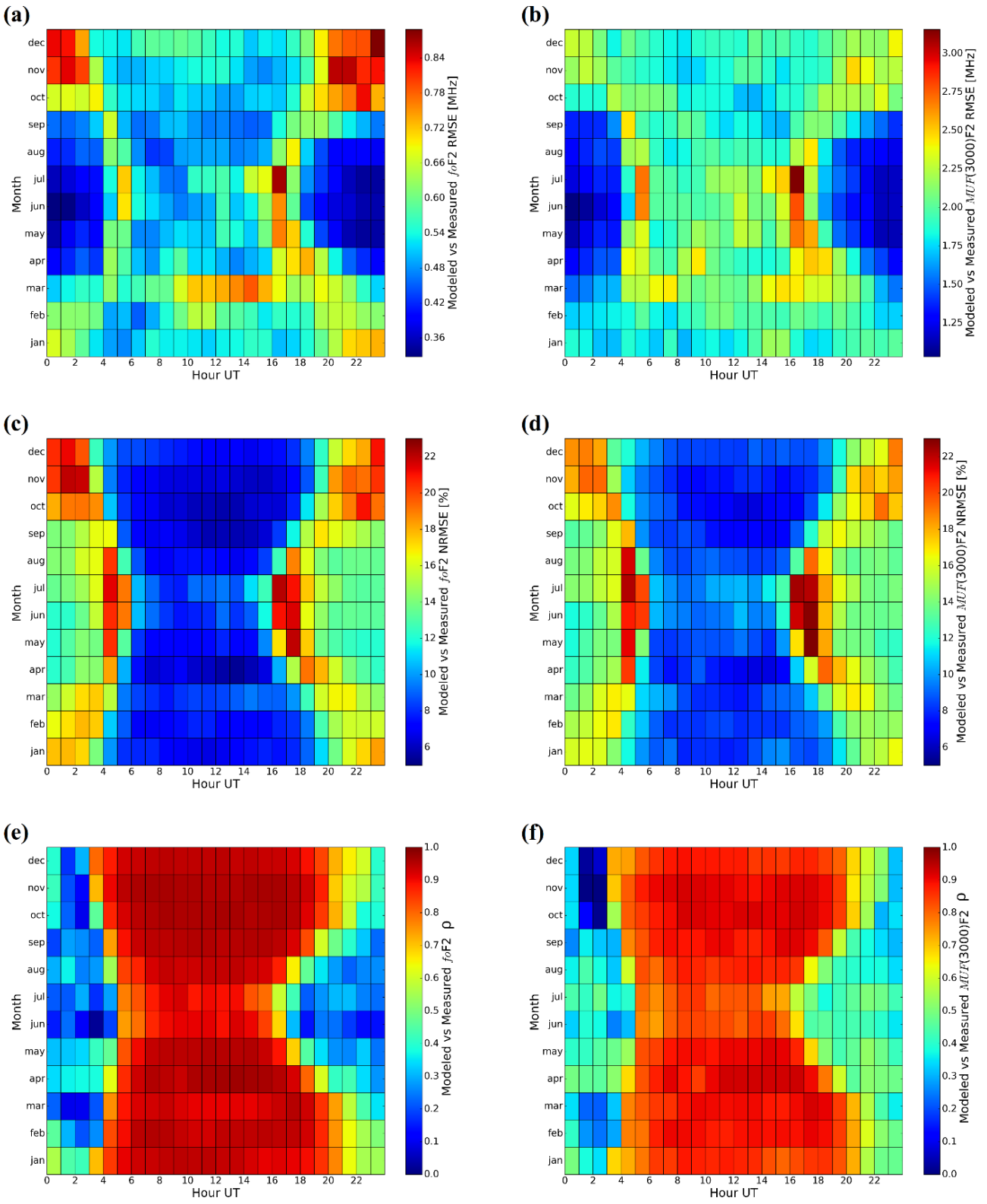


**Figure 5.4:** Matrices, depending on the UT hour (x axis) and month (y axis), of the statistical quantities RMSE, NRMSE, and $\rho$, of the residuals between modeled and measured *fo*F2 (panels (a), (c), and (e), respectively) and *MUF*(3000)F2 (panels (b), (d), and (f), respectively) values.

## 5.3 Ground-based GNSS derived vTEC measurements assimilation procedure

The IRI UP method described in chapter 4 exploits the *fo*F2 and *M*(3000)F2 ionospheric characteristics assimilated by European ionosondes to update the F2 layer description made by IRI. As explained in chapter 3, this method uses the Universal Kriging method to produce effective indices maps by interpolating effective indices values calculated at each ionosonde station. The Universal Kriging method needs at least three assimilating stations in order to build the experimental variogram, from which information about the spatial correlation of the assimilated measurements is extracted. Furthermore, the number and dislocation of stations is of utmost importance. The application of the same procedure over the South Africa region, where only four ionosondes are available, results in minor performances than those obtained for the European region. This is the main reason why one is looking for implementing an assimilation procedure based on GNSS derived vTEC measurements. The TrigNet GNSS receiver network consists of 68 GNSS stations spread all over the South African region with a mean interspacing of 100-200 km. From this point of view, GNSS derived measurements constitute a very attractive source of data to be assimilated.

The fitting procedure described in section 5.2 has been developed to exploit the information embedded in the GNSS derived vTEC measurements, without modifying the IRI UP method framework. In this way, in fact, assimilated vTEC measurements are converted to *fo*F2 and *MUF*(3000)F2 (and then *M*(3000)F2) measurements, as if virtual ionosondes co-located with GNSS receivers were available. In this way, except for the procedure used to obtain values of *fo*F2 and *M*(3000)F2 to be assimilated, the IRI UP method is exactly the same described in chapters 3 and 4.

### 5.3.1 GNSS derived vTEC assimilation procedure

Before calculating *fo*F2 and *MUF*(3000)F2 values from vTEC values, through the procedure described in section 5.2, a spatial filtering is applied on vTEC measurements since the vTEC calibration procedure is inherently affected by some errors. Furthermore, vTEC from some GNSS receiver stations could be noisier than others. In assimilation procedures, the spatial correlation between assimilated data plays a very crucial role. One expects that very close stations provide similar values, and that they are more correlated than distant ones. Because it is difficult to evaluate errors associated to calibrated vTEC values (which could be used to weight values derived by GNSS receivers characterized by different errors), a spatial filter, which has the purpose to detect erroneously calibrated vTEC values, has been developed. According to this, GNSS stations of the considered South African region are grouped in four spatial sectors:

- sector 1: $lat \in [30°S, 20°S] \cup lon \in [15°E, 25°E]$;
- sector 2: $lat \in [40°S, 30°S] \cup lon \in [15°E, 25°E]$;
- sector 3: $lat \in [40°S, 27.5°S] \cup lon \in [25°E, 35°E]$;
- sector 4: $lat \in [27.5°S, 20°S] \cup lon \in [25°E, 35°E]$.

The MAD filter, described in section 5.1.3, is then applied, at each considered UT time, on vTEC values recorded by each GNSS station included in each sector; in this way, a different MAD value characterizes each sector, allowing a spatial description of the dispersion of vTEC values in each sector. The MAD threshold *s* is set equal to 1 in this case. Thus, each GNSS station whose vTEC value does not fulfil the conditions of Eq. (5.4) is discarded from the assimilation procedure. The number of GNSS stations available for the assimilation process, for a definite moment, depends critically on:

a) how many GNSS stations are operating in that moment;
b) the quality of the calibrated vTEC value at each station;
c) the geomagnetic conditions of the ionosphere.

Actually, points b) and c) are strictly interrelated because the quality of calibrated vTEC values decreases during disturbed periods. In fact, for the periods analyzed in the validation process described in section 5.4, it is found

that for quiet conditions the number of GNSS stations passing the MAD filter on average is from 30 to 50 (out of 68); while, under disturbed conditions, this number can lower to 15. Hence, the MAD filter is particularly effective under disturbed geomagnetic conditions when calibrated vTEC values are noisier.

Fig. 5.5 shows a map of vTEC values, with a spatial resolution of $0.25° \times 0.25°$, obtained applying the Universal Kriging method directly on assimilated vTEC values, according to the aforementioned spatial division, for the 8 September 2017 at 9:30 UT, the day of the main phase of the most severe geomagnetic storm which occurred in 2017. After applying the MAD filter, 42 GNSS stations are assimilated, those represented by small circles in Fig. 5.5b. This map is shown only to highlight the spatial variability embedded in vTEC data and the ability of the MAD filter in discarding outliers. In fact, as it is evident from the experimental variogram (panel (a) of Fig. 5.5), and directly from the interpolated map (panel (b) of Fig. 5.5), the spatial variability of vTEC values is well described by the power variogram model which fits very well measured semivariance values. This is reflected by the good agreement between assimilated vTEC values and the background interpolated values. Semivariance values in the variogram plot of Fig. 5.5a (red points) are not calculated for each couple of assimilated stations; in fact, these are previously binned in 16 bins (whose width is one degree), for each of which the mean value of the distance and semivariance is calculated (Eqs. (3.7)-(3.8)). The binning procedure improves the quality of the fitted variogram model (solid blue line in Fig. 5.5a), as can be demonstrated calculating the *cR* statistical parameter (eq. (3.22)).

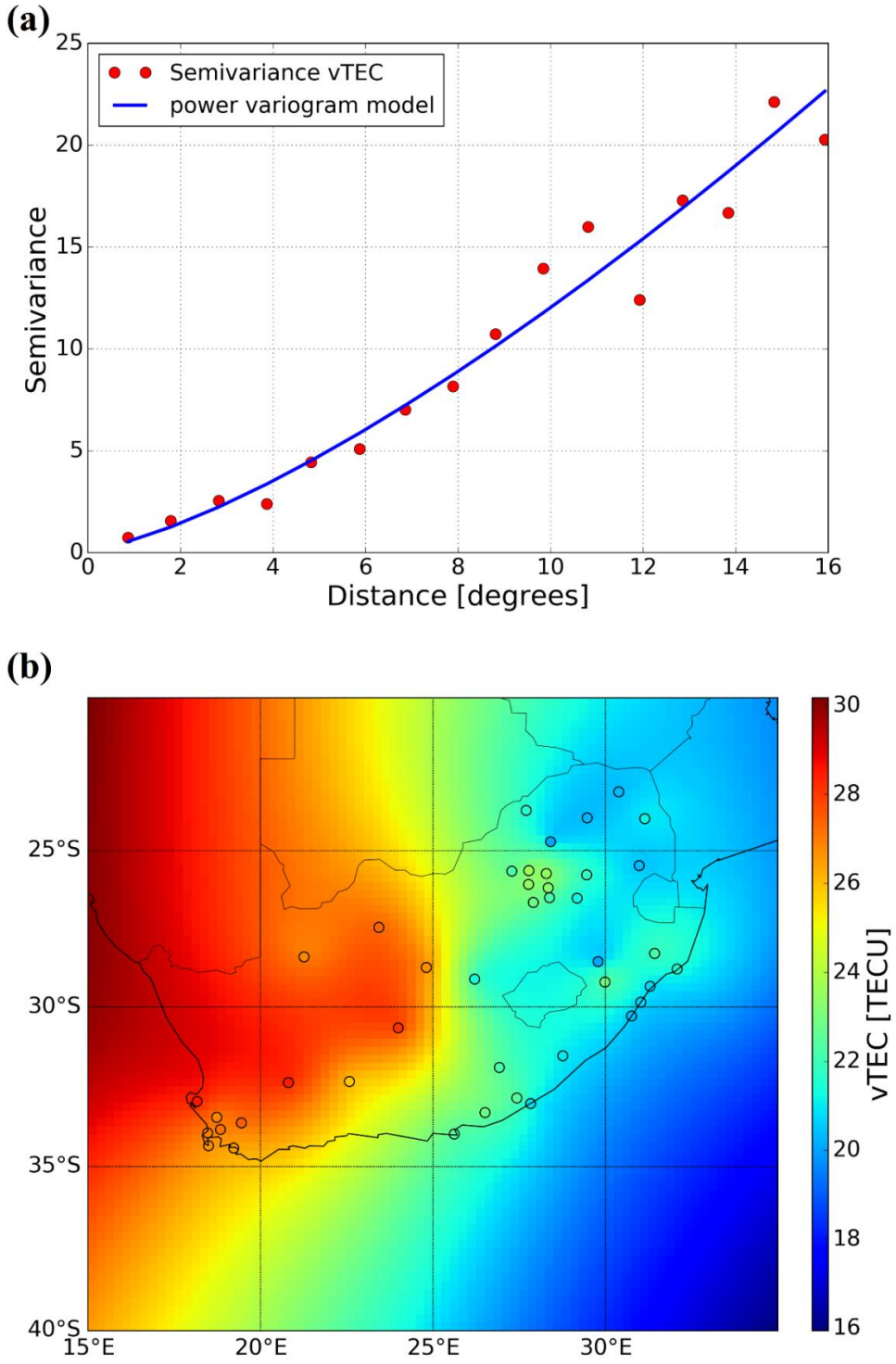


**Figure 5.5:** Panel (a), variogram plot of vTEC assimilated values. Red points are the binned experimental semivariance vTEC values. The blue line is the power variogram model fitting red points. Panel (b), vTEC map obtained applying the Universal Kriging method on vTEC assimilated values starting from the variogram plot of panel (a) (small circles represent GNSS stations which passed the MAD filter and the colour inside them corresponds with the vTEC value assimilated).

### 5.3.2 Effective indices' maps calculation and ingestion in IRI

The *N* assimilated vTEC values, which passed MAD filter, are used to calculate *fo*F2 and *MUF*(3000)F2, using the linear relationships (5.8) and (5.9) and fitting coefficients of Fig. 5.3. The *fo*F2 values are directly used to calculate $IG_{12\text{eff}}$ effective indices, following the procedure described in section 3.2. *MUF*(3000)F2 and *fo*F2 are used to calculate *M*(3000)F2=*MUF*(3000)F2/*fo*F2 for each station and then $R_{12\text{eff}}$ effective indices (section 3.2). Maps of these effective indices are calculated by applying the Universal Kriging method (section 3.3). An example is shown in Fig. 5.6, where vTEC values of Fig. 5.5 are used to obtain effective indices for each GNSS station. In this case, the power variogram model has been used for both indices. In both cases, binned experimental semivariance mean values are well fitted by the variogram model function, and this is reflected in the good description of effective indices spatial variability in the interpolated maps. Specifically, in the $IG_{12\text{eff}}$ map (panel (c) of Fig. 5.6), the calculated $IG_{12\text{eff}}$ indices (small circles) are slightly different from those interpolated, and this is because the $IG_{12\text{eff}}$ semivariogram model manifests a nugget value (the zero-intercept of the variogram model function) slightly different from zero. On the contrary, in the $R_{12\text{eff}}$ case (panel (d) of Fig. 5.6), calculated effective indices are better reproduced, being the nugget null.

As described in section 3.3, semivariance values in the variogram plot describe the spatial correlation between couples of assimilated stations; thus, they characterize the spatial structure of the interpolated effective index. Looking at variogram plots of Fig. 5.6a,b, a very similar spatial behaviour for $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ is evident; this is partly due to the fact that *fo*F2 and *MUF*(3000)F2, used to derive the effective indices, are both calculated from vTEC values. Hence, they manifest the same spatial structure embedded in the assimilated vTEC values (as evident comparing variogram plots of Figs. 5.5 and 5.6). Instead, interpolated $IG_{12\text{eff}}$ and $R_{12\text{eff}}$ maps (Fig. 5.6c,d) can be different because *M*(3000)F2 values are used to derive $R_{12\text{eff}}$ values, and not directly *MUF*(3000)F2 ones.

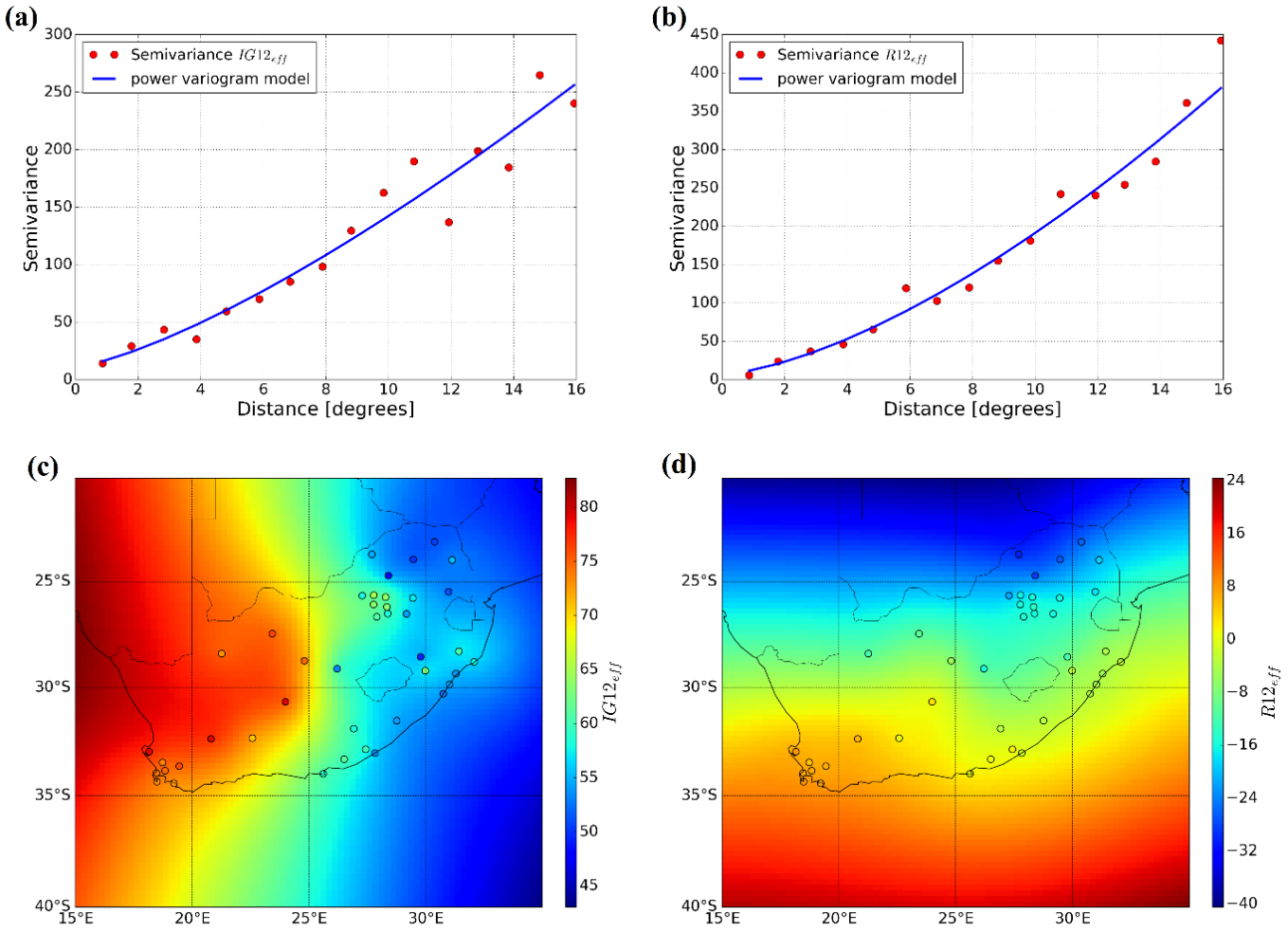


**Figure 5.6:** Left side, variogram plot (panel (a)) and Universal Kriging map (panel (c)) for $IG_{12\text{eff}}$ index. Right side, variogram plot (panel (b)) and Universal Kriging map (panel (d)) for $R_{12\text{eff}}$ index. Both maps refer to the 8 September 2017 at 9:30 UT.

Finally, effective indices maps of Fig. 5.6 are ingested by IRI: the ingestion of $IG_{12\text{eff}}$ map produces updated *fo*F2 values (section 3.2), while the ingestion of the $R_{12\text{eff}}$ map produces updated *hm*F2 values (section 3.2). This operation is done for each grid point of the map.

Fig. 5.7 compares the result of the ingestion with the initial background IRI (using Shubin et al. (2015) option for modeling *hm*F2, chapter 2).

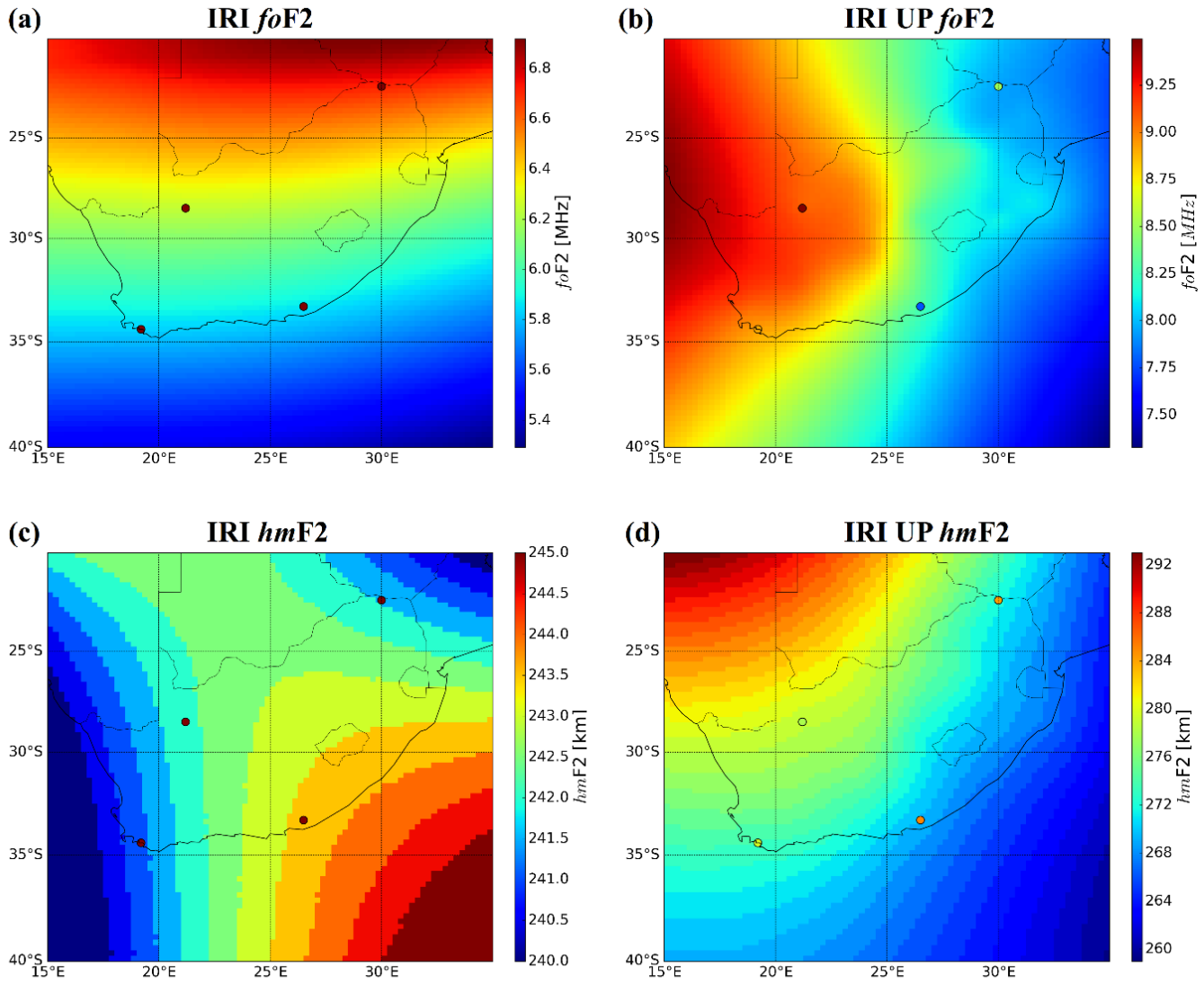


**Figure 5.7:** Left side, maps of *fo*F2 (panel (a)) and *hm*F2 (panel (c)) as modeled by the climatological background IRI model, for the September 8 2017 at 9:30 UT. Right side, panels (b) and (d), same as maps on the left side but for the IRI UP model. Colour in the circles represents the value measured at the corresponding ionosonde testing station. The reader has to pay attention to the fact that colorbar scales are different, this choice being due to the very different values characterizing IRI and IRI UP output.

Fig. 5.8 shows vertical electron density profiles modeled by IRI and IRI UP at ionosonde testing stations along with those measured. The IRI UP method produces *fo*F2 and *hm*F2 values clearly better than those modeled by IRI as the current ionospheric situation is described much better. Results of Figs. 5.7 and 5.8 refer to a disturbed moment during the main phase of a geomagnetic storm.

In Fig. 5.9 ionograms recorded by testing ionosondes and autoscaled by ARTIST are shown in order to highlight the goodness of autoscaled values.

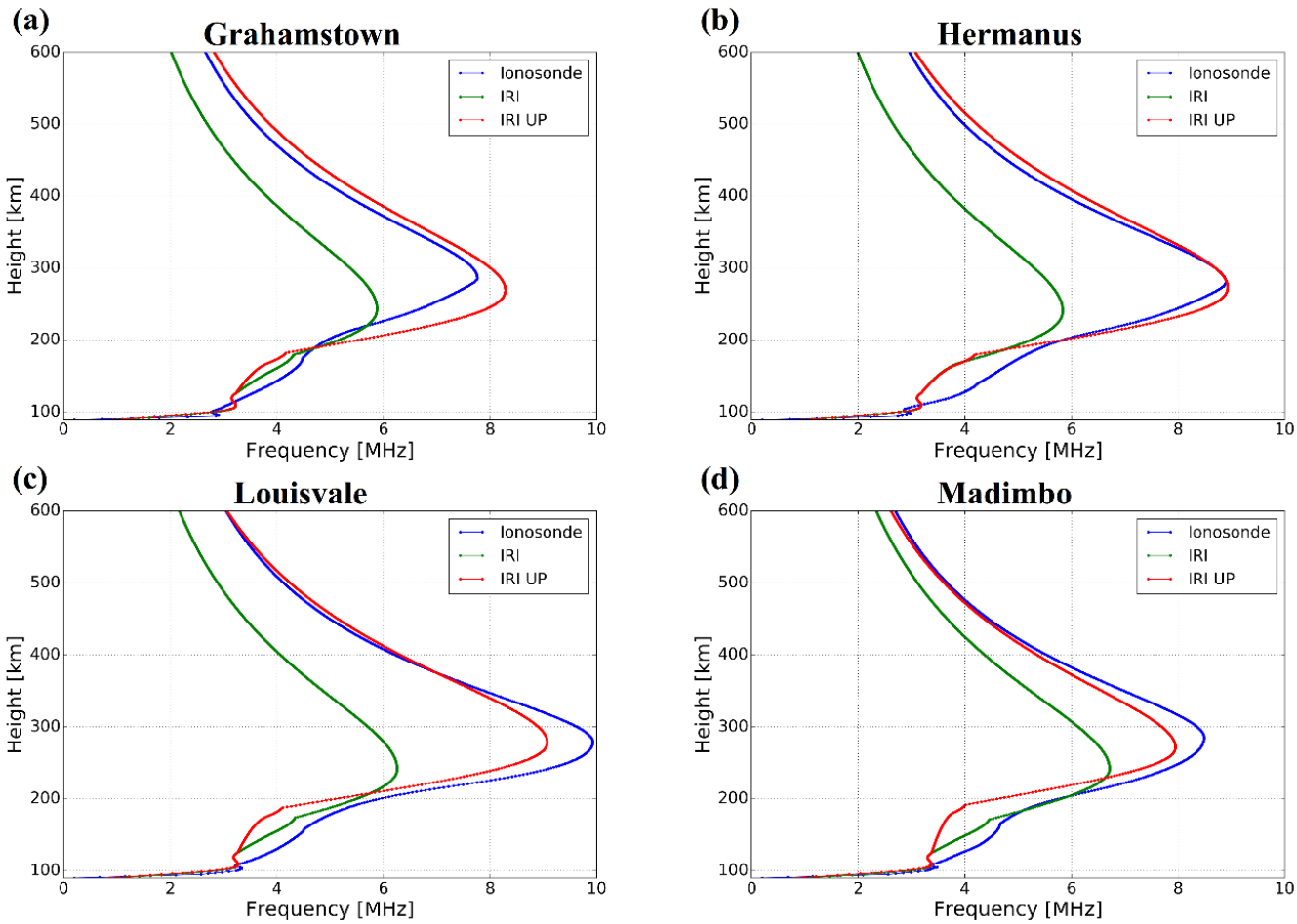


**Figure 5.8:** Vertical electron density profiles at ionosonde testing locations, for the September 8 2017 at 9:30 UT. Blue, green, and red curves refer to profiles measured by ionosondes, modeled by IRI, and modeled by IRI UP, respectively, for (a) Grahamstown, (b) Hermanus, (c) Louisvale, and (d) Madimbo.

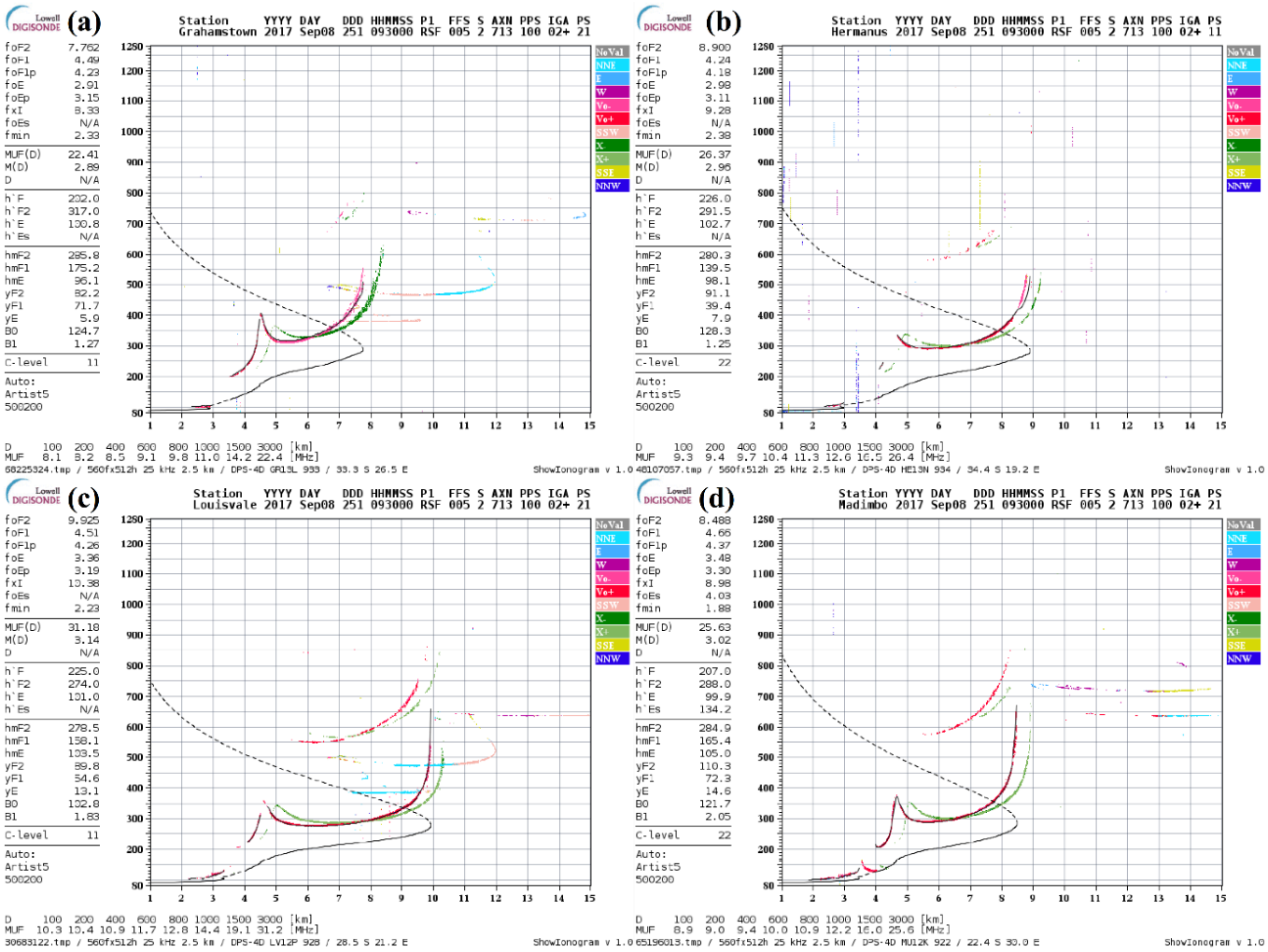


**Figure 5.9:** Ionograms recorded and autoscaled by the ARTIST software, for the September 8 2017 at 9:30 UT, for (a) Grahamstown, (b) Hermanus, (c) Louisvale, and (d) Madimbo. Calculated vertical electron density profiles are shown as black curves, solid for the bottomside ionosphere, dotted for the topside one.

## 5.4 Validation of the IRI UP method for quiet and disturbed periods

Different quiet and disturbed periods for the year 2017 have been chosen to validate the model. For each considered day the time resolution is set to 15 minutes (specifically, minutes 0, 15, 30, and 45 of each hour are considered) to be synchronized with the sounding repetition rate of ionosondes used as comparison in the validation procedure. All available GNSS data from TrigNet stations are used in the assimilation procedure, while ionosonde data from the four South African ionosondes of Table 5.1 are used only as testing values. Validation is performed on *fo*F2 and *hm*F2 characteristics, comparing values modeled by IRI UP and those modeled by IRI (IRI 2016 version, with the STORM option on) with those measured over the ionosonde testing stations.

### 5.4.1 Validation of IRI UP method on 6 geomagnetic storms occurred on 2017

The IRI UP method has been validated during six disturbed periods including the most intense geomagnetic storms that occurred in 2017. Each period consists of some quiet days before the storm commencement, full storm duration (initial, main and recovery phases) and relatively quiet days after the recovery phase. Table 5.3 summarises the considered days for each considered storm period. The different geomagnetic conditions are highlighted by $D_{st}$ and $K_p$ geomagnetic indices downloaded from the OMNIWeb Data Explorer - NASA site at https://omniweb.gsfc.nasa.gov/form/dx1.html. Specifically, minimum $D_{st}$ and maximum $K_p$ values reached during such periods are reported in the last two columns of Table 5.3.

| **STORM NAME** | **Main Phase day** | **Initial analyzed day** | **Final analyzed day** | $\boldsymbol{Dst_{min}}$ **[nT]** | $\boldsymbol{Kp_{max}}$ |
|---|---|---|---|---|---|
| MARCH | 27/03/2017 | 24/03/2017 | 02/04/2017 | -74 | $6^+$ |
| MAY | 28/05/2017 | 25/05/2017 | 05/06/2017 | -125 | 7 |
| JULY | 16/07/2017 | 12/07/2017 | 22/07/2017 | -72 | 6 |
| SEPTEMBER 1 | 08/09/2017 | 01/09/2017 | 19/09/2017 | -142 | $8^+$ |
| SEPTEMBER 2 | 28/09/2017 | 23/09/2017 | 07/10/2017 | -76 | $7^-$ |
| NOVEMBER | 08/11/2017 | 02/11/2017 | 14/11/2017 | -72 | $6^+$ |

**Table 5.3:** Disturbed periods used for validating the IRI UP method. Minimum $D_{st}$ and maximum $K_p$ values occurred during such periods are reported in the last two columns.

In Fig. 5.10 time series of *fo*F2 values from Hermanus ionosonde, modeled by IRI, and IRI UP, for the storm named SEPTEMBER 1 in Table 5.3 (the most intense storm in 2017), are shown with corresponding residuals. For the comparison, only ionograms with C-score $\geq$ 75 have been employed. Corresponding histograms of the residuals and scatter plots are shown in Fig. 5.11. Statistical quantities RMSE, NRMSE, and $\rho$ (Eqs. (4.5)-(4.7)), and the mean of the residuals (Eq. (4.8)), have been calculated for both IRI and IRI UP separating quiet, very disturbed, and moderately disturbed days (Table 5.4), and for the whole time series, for each ionosonde testing station (Table 5.5).

Several conclusions can be drawn from the results obtained for Hermanus for the SEPTEMBER 1 storm:

- for *fo*F2 IRI UP performs better than IRI for the whole time period (1-19 September 2017), while for *hm*F2 results are equivalent (Table 5.5);
- the assimilation of vTEC by IRI UP leads to an *fo*F2 RMSE=0.474 MHz, which is by far better than the IRI RMSE=0.823 MHz. Thus, also NRMSE is reduced from 16.376 % for IRI to 9.430 % for IRI UP. This involves an improvement of the precision of the *fo*F2 modeling, as it is clearly visible from time series of Fig. 5.10 and from histograms of Fig. 5.11. The correlation coefficient also manifests an improvement for *fo*F2 from 0.930 for IRI to 0.970 for IRI UP, as it is clearly visible from Fig. 5.11. This is reflected in a smaller dispersion of modeled vs measured *fo*F2 values around the linear fitting function. The mean of the residuals between modeled and measured *fo*F2 values is also improved from -0.430 MHz for IRI to 0.006 MHz for IRI UP, which testifies an improvement of the accuracy of the modeling;
- looking at time series of panel (a) of Fig. 5.10, it is clear that IRI UP describes the ionospheric variability depicted by measured *fo*F2 values better than the IRI model, for both quiet, moderately disturbed and very

disturbed periods. IRI tends to underestimate *fo*F2 values during dayside for quiet days (244-250 days of the year 2017), before the storm commencement, and it is not able to follow the positive phase of the day 251 (September 8 2017, main phase day) and the successive negative phase of the day 252. On the contrary, IRI UP describes successfully the above-mentioned variability, particularly during daytime. Some problems arise during nighttime for the main phase days and for the successive recovery phase days (251-255), where IRI UP tends to overestimate *fo*F2 at night. This might be ascribed to depletion/refilling processes characterizing the ionosphere-plasmasphere system especially for disturbed conditions (e.g., Chi et al. 2000; Förster and Jakowski 2000);

- results obtained separating quiet, moderately disturbed, and very disturbed days (Table 5.4), highlight IRI's difficulties in modeling the ionosphere response to geomagnetic storms, at these latitudes, even with the STORM option on;
- *hm*F2 modeling made by IRI UP produces results similar to those modeled by IRI for quiet and slightly disturbed conditions, while an important improvement is obtained under very disturbed conditions (Table 5.4);
- similar considerations made for Hermanus hold also for the other ionosonde stations (Table 5.5), even though they are characterized by slightly lower performances, especially Madimbo.

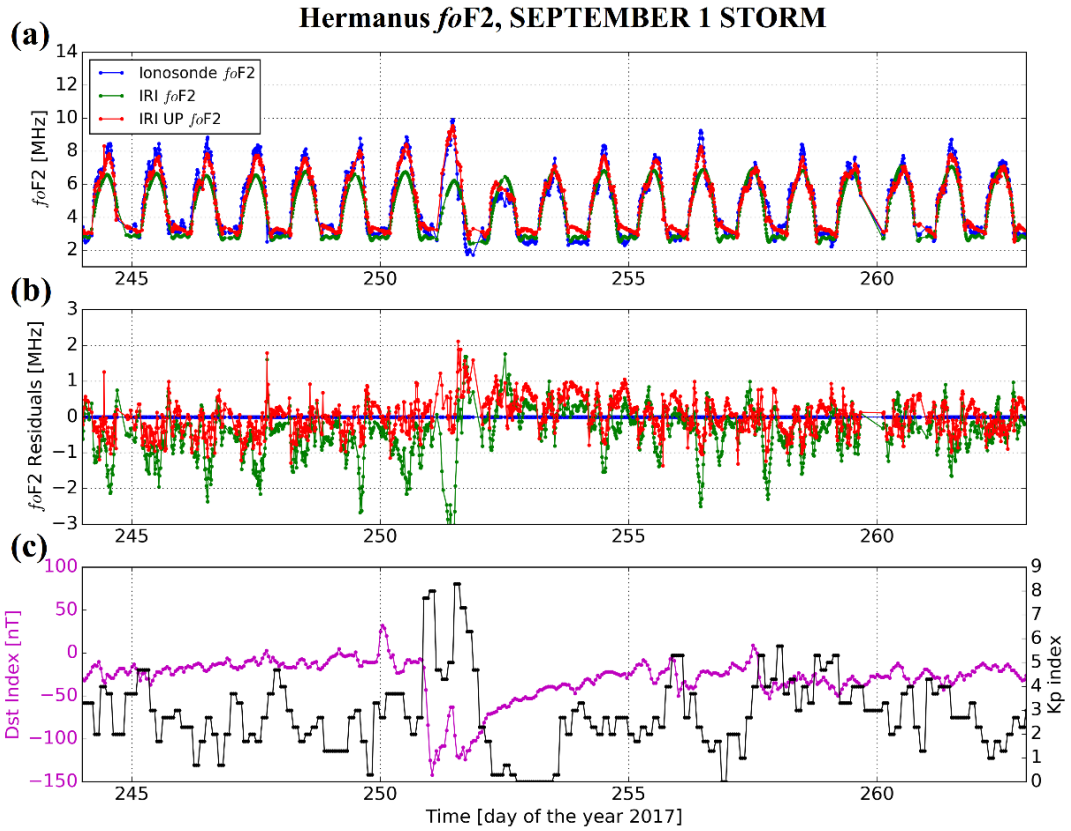


**Figure 5.10:** (a) Time series of *fo*F2 values at Hermanus measured by the ionosonde (blue line), modeled by IRI (green line), and modeled by IRI UP (red line), for the time period 1-19 September 2017 (SEPTEMBER 1 storm in Table 5.3). (b) Time series of corresponding residuals between modeled and measured *fo*F2 values, color coded as in panel (a). (c) $D_{st}$ and $K_p$ geomagnetic indices time series, in magenta and black, respectively.

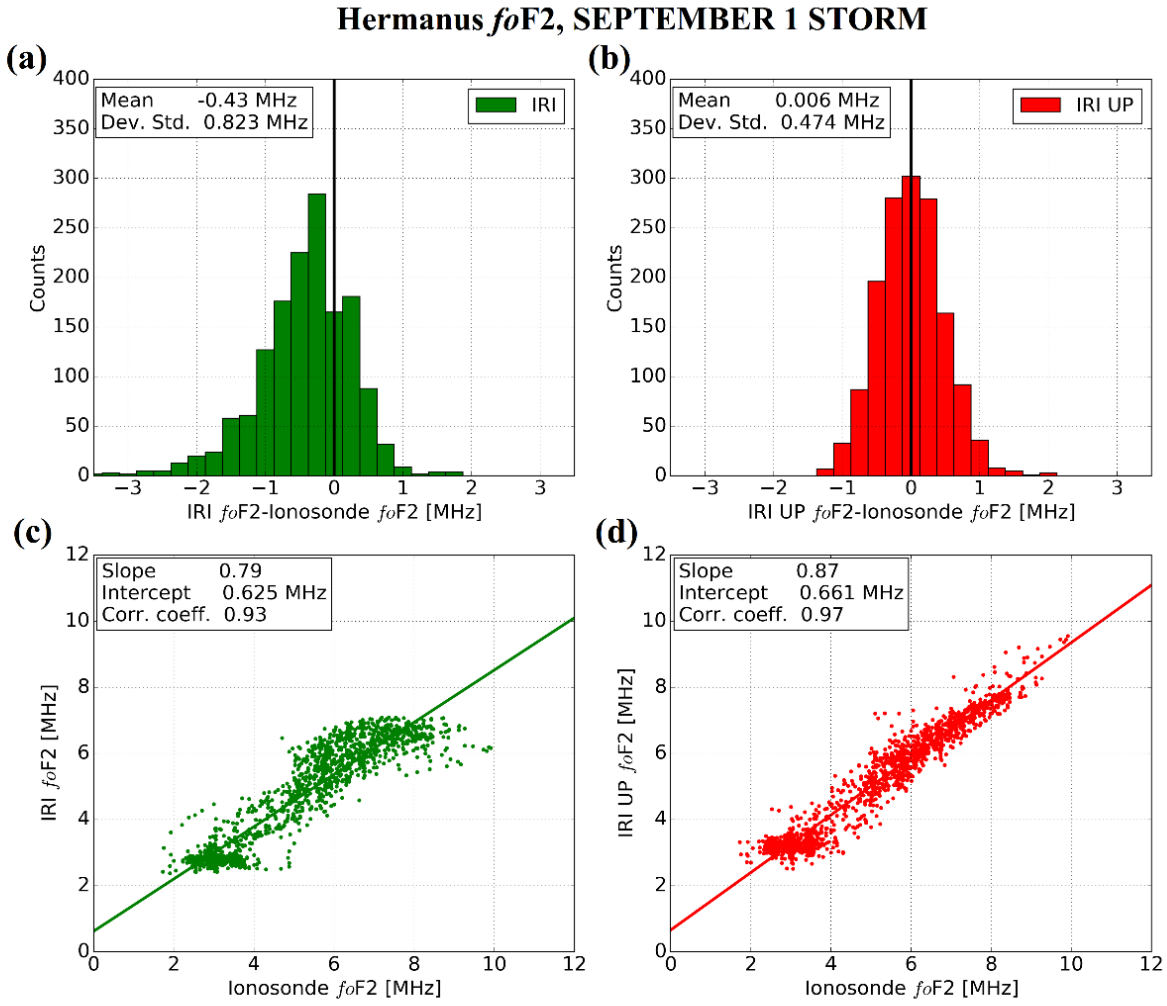


**Figure 5.11:** Panels (a) and (b), histograms derived from time series of residuals between modeled and measured *fo*F2 values, shown in the panel (b) of Fig. 5.10, for the SEPTEMBER 1 storm of Table 5.3. Green color refers to IRI, red to IRI UP. The mean and the standard deviation of the residuals are shown in the upper left box of each histogram. Panels (c) and (d) show, respectively, the scatter plots of modeled and measured *fo*F2 values for IRI and for IRI UP; straight solid lines are the best fitting linear functions whose slope and intercept are shown in the upper left box of each scatter plot, along with the Pearson correlation coefficient value.

| STATION | Model & Characteristic | RMSE [MHz for *fo*F2, km for *hm*F2] | NRMSE [%] | $\rho$ | Residuals mean [MHz for *fo*F2, km for *hm*F2] |
|---|---|---|---|---|---|
| **SEPTEMBER 1 STORM, Quiet days 1-6 September 2017** | | | | | |
| Hermanus | IRI *fo*F2 | 0.926 | 17.586 | 0.962 | -0.694 |
| | IRI UP *fo*F2 | 0.457 | 8.677 | 0.979 | -0.186 |
| | IRI *hm*F2 | 18.291 | 7.271 | 0.718 | -5.286 |
| | IRI UP *hm*F2 | 21.423 | 8.517 | 0.781 | 14.416 |
| **SEPTEMBER 1 STORM, Main Phase days 7-8 September 2017** | | | | | |
| Hermanus | IRI *fo*F2 | 1.406 | 26.486 | 0.910 | -0.844 |
| | IRI UP *fo*F2 | 0.627 | 11.817 | 0.966 | 0.188 |
| | IRI *hm*F2 | 32.642 | 11.995 | 0.505 | -22.790 |
| | IRI UP *hm*F2 | 23.415 | 8.604 | 0.507 | -2.445 |
| **SEPTEMBER 1 STORM, Recovery Phase days 9-15 September 2017** | | | | | |
| Hermanus | IRI *fo*F2 | 0.645 | 13.815 | 0.936 | -0.203 |
| | IRI UP *fo*F2 | 0.485 | 10.387 | 0.966 | 0.132 |
| | IRI *hm*F2 | 19.244 | 7.472 | 0.792 | -5.422 |
| | IRI UP *hm*F2 | 20.650 | 8.018 | 0.787 | 10.040 |

**Table 5.4:** RMSE, NRMSE, $\rho$ correlation coefficient, and mean of the residuals between modeled and measured values are shown for *fo*F2 and *hm*F2, for Hermanus, for the quiet days before the storm commencement (1-6 September 2017), for the main phase days (7-8 September 2017), and for the recovery phase days (9-15 September 2017), for the SEPTEMBER 1 storm (Table 5.3). Best statistical results, between IRI and IRI UP, are highlighted in red for each ionospheric characteristic.

| STATION | Model & Characteristic | RMSE [MHz for *fo*F2, km for *hm*F2] | NRMSE [%] | $\rho$ | Residuals mean [MHz for *fo*F2, km for *hm*F2] |
|---|---|---|---|---|---|
| **SEPTEMBER 1 STORM, 1-19 September 2017** | | | | | |
| Grahamstown | IRI *fo*F2 | 0.833 | 16.710 | 0.922 | -0.439 |
| | IRI UP *fo*F2 | 0.514 | 10.311 | 0.962 | 0.033 |
| | IRI *hm*F2 | 25.310 | 9.699 | 0.697 | -10.092 |
| | IRI UP *hm*F2 | 21.653 | 8.298 | 0.762 | 3.828 |
| Hermanus | IRI *fo*F2 | 0.823 | 16.376 | 0.930 | -0.430 |
| | IRI UP *fo*F2 | 0.474 | 9.430 | 0.970 | 0.006 |
| | IRI *hm*F2 | 20.049 | 7.850 | 0.708 | -5.261 |
| | IRI UP *hm*F2 | 21.620 | 8.465 | 0.747 | 11.671 |
| Louisvale | IRI *fo*F2 | 0.862 | 17.573 | 0.926 | -0.495 |
| | IRI UP *fo*F2 | 0.505 | 10.302 | 0.971 | -0.167 |
| | IRI *hm*F2 | 26.230 | 9.837 | 0.632 | -12.536 |
| | IRI UP *hm*F2 | 22.870 | 8.577 | 0.671 | 5.796 |
| Madimbo | IRI *fo*F2 | 0.796 | 15.513 | 0.925 | -0.380 |
| | IRI UP *fo*F2 | 0.544 | 10.610 | 0.960 | -0.067 |
| | IRI *hm*F2 | 26.024 | 9.918 | 0.667 | -9.176 |
| | IRI UP *hm*F2 | 24.810 | 9.456 | 0.697 | 7.255 |

**Table 5.5:** Same as Table 5.4 but for each ionosonde testing station, for the time period from 1 to 19 September 2017 (SEPTEMBER 1 storm in Table 5.3). Best statistical results, between IRI and IRI UP, are highlighted in red for each ionospheric characteristic.

The geographical dislocation of the stations is at the base of the different IRI UP performances for different testing ionosonde stations. In fact, looking at Fig. 5.1, it is evident that Hermanus station is surrounded by several very close GNSS stations, which is also the case for Grahamstown station; on the contrary, Louisvale and Madimbo ionosonde stations are surrounded by few GNSS stations. Louisvale station, apart from the quasi co-located Upington GNSS station, is quite isolated and is at the northern edge of the TrigNet network. Madimbo station experiences an even worse situation: it is at the eastern edge of the TrigNet network and the closest GNSS station is Thohoyandou, at a distance of 92 km. Relying on assimilated vTEC data, improved IRI UP performances are expected in regions well covered by GNSS stations. The performance is degraded by moving away from assimilated stations and particularly at the border of the map. Results of Table 5.5 agree pretty well with these considerations. Nevertheless, obtained results represent all the same an improvement of the IRI model, for each considered station.

The different IRI UP performances obtained for *fo*F2 and *hm*F2 are also somewhat expected. This happens mainly for the following reasons:

- the spherical harmonic expansion (chapter 2) used to describe the *fo*F2 and *M*(3000)F2 spatial behavior stops when the maximum order of the harmonics is $K = 76$, for *fo*F2, and $K = 49$, for *M*(3000)F2. This means that *fo*F2 maps present a higher spatial resolution than *M*(3000)F2 ones;

- *fo*F2 predictions made by IRI are based on $IG_{12}$, which is an ionospheric index (Liu et al. 1983). Instead *M*(3000)F2 predictions, which come into play to calculate *hm*F2, are not based on an ionospheric index, but on the solar index $R_{12}$;

- in the IRI UP procedure, *hm*F2 is calculated applying the empirical formula of Bilitza et al. (1979) (Eq. (2.7)). This implies that the error characterizing *hm*F2 depends on the errors relative to *fo*F2, *M*(3000)F2, *fo*E, and $R_{12\text{eff}}$; therefore, the error propagation leads to an error associated to the *hm*F2 prediction which is intrinsically larger than that of *fo*F2;

- Furthermore, there is also another very important difference in the process of modeling *fo*F2 and *hm*F2. In fact, unlike the IRI UP procedure described in chapter 4, for which *fo*F2 and *M*(3000)F2 measured by ionosondes are assimilated, now *fo*F2 and *M*(3000)F2 values are derived from assimilated vTEC measurements. With regard to this, while the relationship between *fo*F2 and vTEC values can be easily understood and justified from a physical point of view, the one between *MUF*(3000)F2 and vTEC values is more empirical and speculative. As Fig. 5.4 demonstrates, better performances in terms of NRMSE are achieved for *fo*F2 than for *MUF*(3000)F2, when modeled as a function of vTEC.

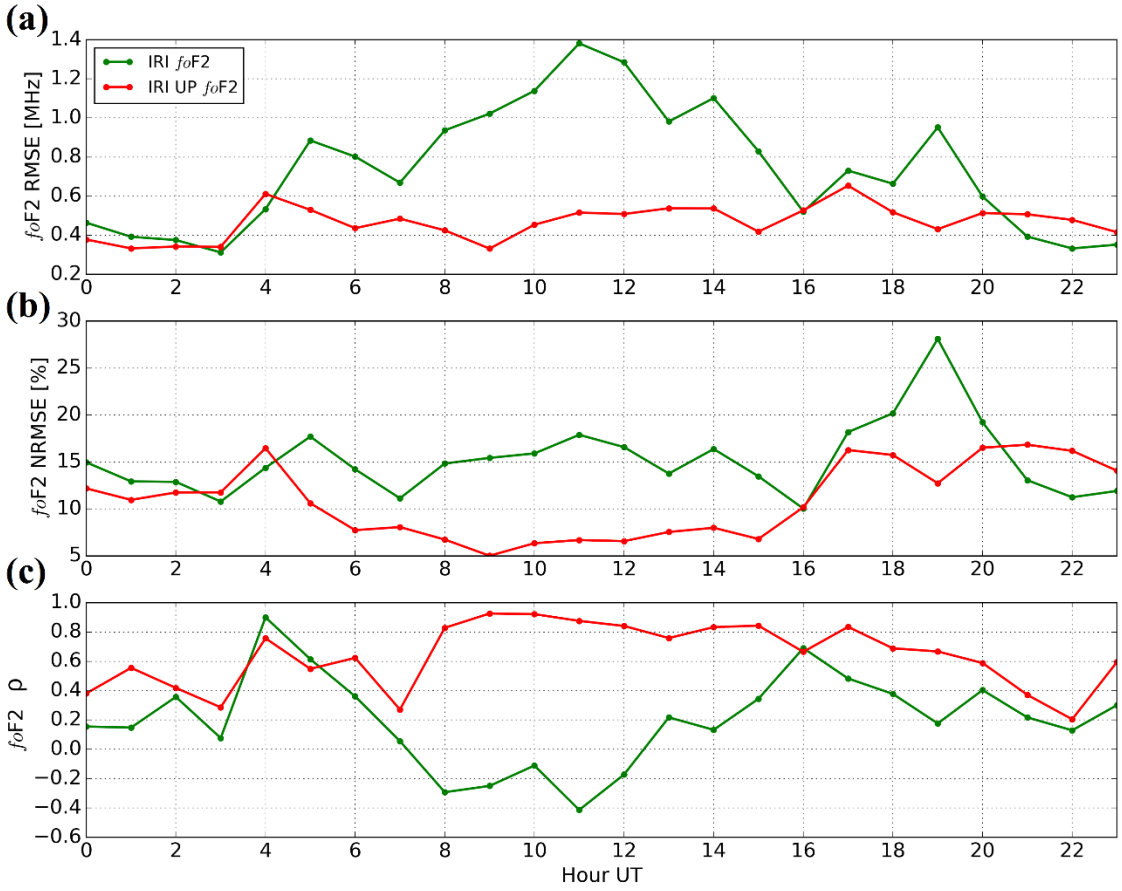


**Figure 5.12:** Panel (a), time series of RMSE between measured and modeled *fo*F2 values, for the Hermanus testing station, for the period 1-19 September 2017 (SEPTEMBER 1 storm), as a function of UT hour. Green line refers to IRI's RMSE values, red line to IRI UP ones. Panel (b), same as panel (a) but for NRMSE. Panel (c), same as panel (a) but for the $\rho$ correlation coefficient.

Because of differences in the *fo*F2 and *MUF*(3000)F2 derivation from vTEC values for different hours of the day (see Figs. 5.3 and 5.4), one would expect corresponding differences also in the IRI UP performance. The diurnal variability has been studied for the time period 1-19 September 2017 (SEPTEMBER 1 storm in Table 5.3) for Hermanus, calculating statistical quantities RMSE, NRMSE, and $\rho$, for different hours, for both IRI and IRI UP. Corresponding results are shown in Fig. 5.12, where IRI UP exhibits better performances than IRI, except for hours 3, 4, 21, 22, and 23 UT; these are dawn and nighttime hours for which, as explained in section 5.2, modeling of *fo*F2 and *MUF*(3000)F2 as a function of vTEC is difficult. IRI UP performs much better during daytime and at dusk, compared to IRI, and these results are perfectly compatible with ones shown in Fig. 5.4. Exceptionally low IRI correlation coefficient values in panel (c) of Fig. 5.12 are due to the fact that the large day-to-day variability is not considered by the IRI model (Bilitza 1990), even with the STORM option on (Fuller-Rowell et al. 1998; Araujo-Pradere et al. 2002).

Statistical results for each ionosonde testing station, for the rest of disturbed periods listed in Table 5.3, are collected in Table 5.6. In general, for Grahamstown and Hermanus, IRI UP results are better than IRI ones, for both *fo*F2 and *hm*F2. Louisvale and Madimbo results, also because of some missing ionosonde data, are instead

more equilibrated. On the other hand, the storm occurred in May deserves a more careful inspection, as it is the only case for which IRI UP does not perform properly.

| STATION | Model & Characteristic | RMSE [MHz for *fo*F2, km for *hm*F2] | NRMSE [%] | $\rho$ | Residuals mean [MHz for *fo*F2, km for *hm*F2] |
|---|---|---|---|---|---|
| **STORM MARCH, 24 March-2 April 2017** | | | | | |
| Grahamstown | IRI *fo*F2 | 0.760 | 15.513 | 0.916 | -0.180 |
| | IRI UP *fo*F2 | 0.647 | 13.206 | 0.963 | 0.398 |
| | IRI *hm*F2 | 23.304 | 8.923 | 0.738 | -11.305 |
| | IRI UP *hm*F2 | 20.197 | 7.733 | 0.766 | 4.937 |
| Hermanus | IRI *fo*F2 | 0.751 | 15.205 | 0.926 | -0.182 |
| | IRI UP *fo*F2 | 0.544 | 11.022 | 0.973 | 0.261 |
| | IRI *hm*F2 | 20.005 | 7.788 | 0.680 | -6.629 |
| | IRI UP *hm*F2 | 20.324 | 7.912 | 0.736 | 10.460 |
| Louisvale | NO IONOSONDE DATA | | | | |
| Madimbo | IRI *fo*F2 | 0.686 | 13.126 | 0.948 | -0.194 |
| | IRI UP *fo*F2 | 0.695 | 13.321 | 0.961 | 0.298 |
| | IRI *hm*F2 | 26.496 | 10.045 | 0.615 | -12.191 |
| | IRI UP *hm*F2 | 25.859 | 9.803 | 0.563 | 7.843 |
| **STORM MAY, 25 May-5 June 2017** | | | | | |
| Grahamstown | IRI *fo*F2 | 0.478 | 12.222 | 0.944 | 0.085 |
| | IRI UP *fo*F2 | 0.553 | 14.139 | 0.932 | 0.204 |
| | IRI *hm*F2 | 24.358 | 10.134 | 0.793 | -6.836 |
| | IRI UP *hm*F2 | 23.563 | 9.803 | 0.795 | 6.028 |
| Hermanus | IRI *fo*F2 | 0.522 | 13.729 | 0.948 | 0.251 |
| | IRI UP *fo*F2 | 0.537 | 14.124 | 0.942 | 0.246 |
| | IRI *hm*F2 | 22.742 | 9.576 | 0.776 | -3.535 |
| | IRI UP *hm*F2 | 25.250 | 10.632 | 0.777 | 11.835 |
| Louisvale | NO IONOSONDE DATA | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Madimbo | IRI *fo*F2 | 0.470 | 10.246 | 0.944 | -0.007 |
| | IRI UP *fo*F2 | 0.617 | 13.453 | 0.910 | 0.011 |
| | IRI *hm*F2 | 26.154 | 11.204 | 0.718 | -6.339 |
| | IRI UP *hm*F2 | 27.188 | 11.647 | 0.706 | 9.115 |
| **STORM JULY, 12-22 July 2017** | | | | | |
| Grahamstown | IRI *fo*F2 | 0.484 | 12.802 | 0.934 | -0.068 |
| | IRI UP *fo*F2 | 0.402 | 10.634 | 0.956 | 0.062 |
| | IRI *hm*F2 | 25.855 | 10.788 | 0.749 | -5.736 |
| | IRI UP *hm*F2 | 34.371 | 14.341 | 0.607 | 16.935 |
| Hermanus | IRI *fo*F2 | 0.470 | 12.506 | 0.945 | 0.052 |
| | IRI UP *fo*F2 | 0.425 | 11.318 | 0.959 | 0.163 |
| | IRI *hm*F2 | 24.780 | 10.556 | 0.730 | -1.304 |
| | IRI UP *hm*F2 | 42.832 | 18.246 | 0.350 | 27.508 |
| Louisvale | IRI *fo*F2 | 0.397 | 10.983 | 0.956 | 0.003 |
| | IRI UP *fo*F2 | 0.401 | 11.110 | 0.955 | 0.058 |
| | IRI *hm*F2 | 28.913 | 11.691 | 0.722 | -10.146 |
| | IRI UP *hm*F2 | 43.470 | 17.577 | 0.165 | 20.862 |
| Madimbo | IRI *fo*F2 | 0.464 | 11.925 | 0.949 | -0.077 |
| | IRI UP *fo*F2 | 0.576 | 14.811 | 0.917 | 0.007 |
| | IRI *hm*F2 | 29.071 | 12.020 | 0.731 | -8.541 |
| | IRI UP *hm*F2 | 41.992 | 17.362 | 0.348 | 20.604 |
| **STORM SEPTEMBER 2, 23 September-7 October 2017** | | | | | |
| Grahamstown | IRI *fo*F2 | 0.894 | 16.278 | 0.942 | -0.634 |
| | IRI UP *fo*F2 | 0.453 | 8.243 | 0.972 | -0.086 |
| | IRI *hm*F2 | 21.423 | 7.959 | 0.758 | -9.498 |
| | IRI UP *hm*F2 | 21.907 | 8.139 | 0.748 | 8.796 |
| Hermanus | IRI *fo*F2 | 0.918 | 16.621 | 0.941 | -0.654 |
| | IRI UP *fo*F2 | 0.471 | 8.532 | 0.975 | -0.170 |
| | IRI *hm*F2 | 18.159 | 6.908 | 0.745 | -4.183 |
| | IRI UP *hm*F2 | 24.802 | 9.435 | 0.703 | 16.084 |

| | | | | | |
|---|---|---|---|---|---|
| Louisvale | IRI *fo*F2 | 0.904 | 17.158 | 0.942 | -0.616 |
| | IRI UP *fo*F2 | 0.556 | 10.564 | 0.968 | -0.221 |
| | IRI *hm*F2 | 23.462 | 8.565 | 0.696 | -9.699 |
| | IRI UP *hm*F2 | 25.820 | 9.425 | 0.617 | 11.090 |
| Madimbo | IRI *fo*F2 | 0.834 | 14.726 | 0.933 | -0.431 |
| | IRI UP *fo*F2 | 0.589 | 10.406 | 0.960 | -0.142 |
| | IRI *hm*F2 | 24.837 | 9.176 | 0.647 | -9.205 |
| | IRI UP *hm*F2 | 27.464 | 10.146 | 0.590 | 12.210 |
| **STORM NOVEMBER, 2-14 November 2017** | | | | | |
| Grahamstown | IRI *fo*F2 | 0.879 | 16.762 | 0.876 | 0.263 |
| | IRI UP *fo*F2 | 0.591 | 11.300 | 0.948 | 0.141 |
| | IRI *hm*F2 | 21.633 | 8.020 | 0.731 | -5.758 |
| | IRI UP *hm*F2 | 30.564 | 11.331 | 0.561 | 18.086 |
| Hermanus | IRI *fo*F2 | 0.806 | 15.254 | 0.884 | 0.199 |
| | IRI UP *fo*F2 | 0.523 | 9.931 | 0.957 | 0.108 |
| | IRI *hm*F2 | 19.128 | 7.172 | 0.730 | -3.214 |
| | IRI UP *hm*F2 | 32.538 | 12.202 | 0.572 | 23.452 |
| Louisvale | NO IONOSONDE DATA | | | | |
| Madimbo | NO IONOSONDE DATA | | | | |

**Table 5.6:** Same as Table 5.5 but for the other five time periods listed in Table 5.3. Best statistical results, between IRI and IRI UP, are highlighted in red for each ionospheric characteristic.

In fact, as it is evident from Table 5.6, results obtained for the MAY storm are very different from those obtained for all the other analyzed storms. This is the only case for which, concerning *fo*F2, IRI performs better than IRI UP for each testing station. Values of measured and modeled *fo*F2 values at each ionosonde station in May 2017 (Grahamstown, Hermanus, and Madimbo), and assimilated vTEC values recorded by the closest GNSS stations (Grahamstown, Hermanus, and Thohoyandou, respectively), are plotted in Fig. 5.13 for days 27-29 May 2017, around the main phase of the storm. Looking at vTEC time series, one can see that vTEC values collected at Hermanus and Grahamstown remain pretty similar during quiet and disturbed periods. On the contrary, vTEC values collected at Thohoyandou manifest a very important increase in the main phase day compared to the other days, probably to be ascribed to a southern expansion of the equatorial ionization anomaly. IRI UP modeled *fo*F2 values reflect the daily variation recorded by the closest GNSS stations: for Hermanus and Grahamstown, *fo*F2 values have a similar trend for all the three analyzed days, while at Madimbo *fo*F2 values during the main phase day are really higher than those of the adjacent days. Unfortunately, *fo*F2 values recorded at Hermanus and Grahamstown present a really different scenario, with a negative phase characterizing the May 28 2017 which is not reflected in vTEC data. This is likely due to the fact that, for disturbed days, at Hermanus and Grahamstown, the bottomside and topside ionosphere are affected by different physical processes. This highlights a very

important issue, that results are as good as the assimilated data: if the assimilated data reproduces the physical property, then the outcome reflects the same.

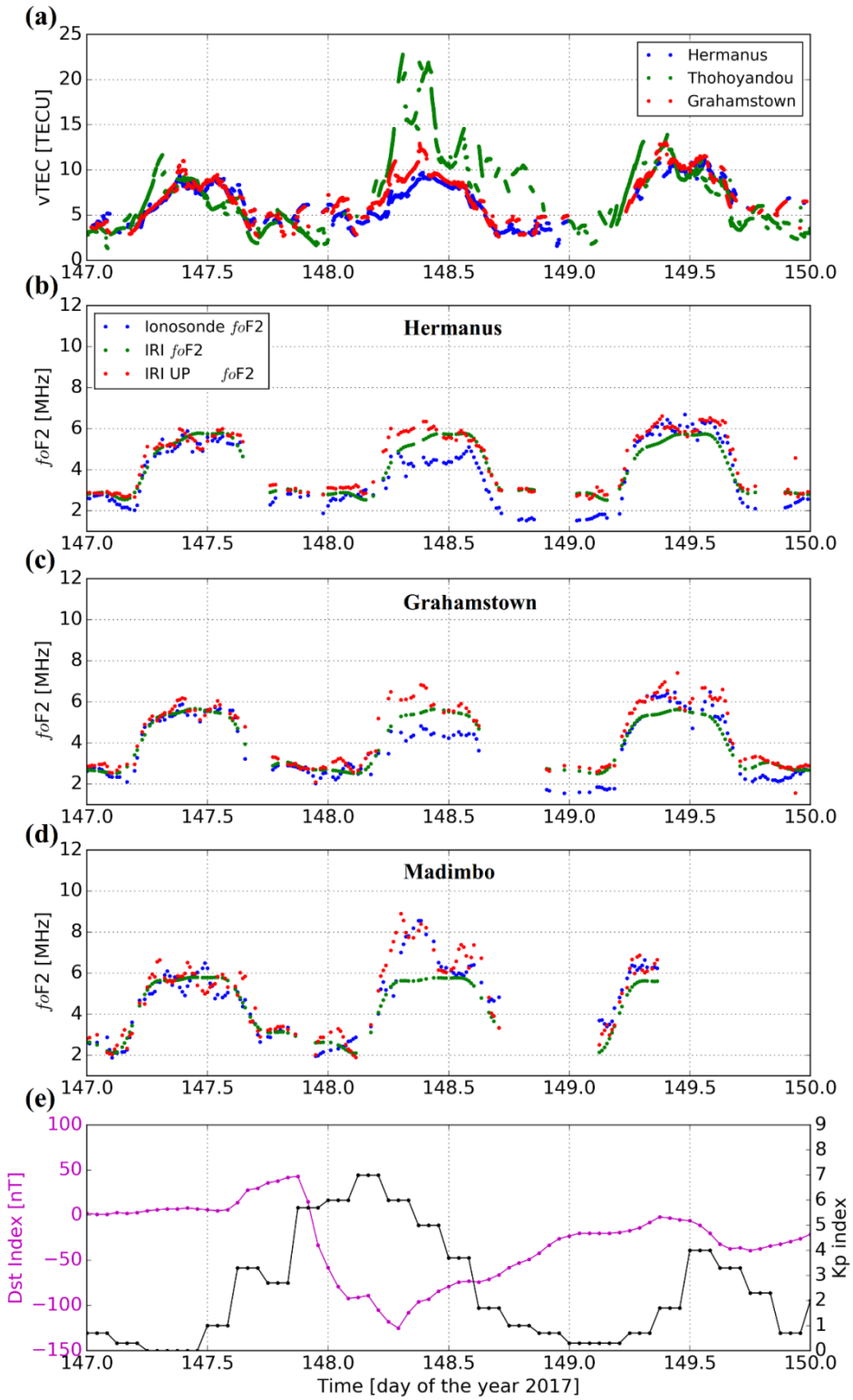


**Figure 5.13:** Panel (a), time series of vTEC values measured by Hermanus (blue), Thohoyandou (green), and Grahamstown (red) for the days 27-29 May 2017. Panels (b)-(d), time series of *fo*F2 values measured by the ionosonde (blue), modeled by IRI (green), and modeled by IRI UP (red), for the same time period of panel (a), for Hermanus, Grahamstown, and Madimbo, respectively. Panel (e), $D_{st}$ and $K_p$ geomagnetic indices time series, in magenta and black, respectively.

### 5.4.2 Validating the seasonal behavior of the IRI UP method

A validation of the IRI UP model for different seasons has been performed. In fact, linear fit coefficients shown in Fig. 5.3 exhibit a different behavior for different months, so one would expect that different performances can be achieved for different seasons. This is why, for the year 2017, four time periods representing different seasons have been selected. All selected days are quiet ($K_p$<4).

Specifically, selected days are:

- SUMMER, 1-4 January 2017;
- AUTUMN, 22-25 March 2017;
- WINTER, 22-25 June 2017;
- SPRING, 23-26 September 2017.

Each period is composed of four consecutive days around either the equinox or the solstice, except in summer due to an unavailability of data at the ionosonde testing stations.

Time series of *fo*F2 values, for the four selected periods, measured at Hermanus, modeled by IRI and IRI UP are shown in Fig. 5.14. Statistical results for each ionosonde testing station are collected in Table 5.7.

Some conclusions can be drawn from this analysis:

- focusing on Hermanus results for *fo*F2, IRI shows better performances in autumn (NRMSE=10.951 %) and winter (NRMSE=10.753 %) than in summer (NRMSE=15.407 %) and spring (NRMSE=13.736 %). On the contrary, IRI UP exhibits small differences for different seasons, NRMSE values are slightly lower in spring (NRMSE=8.892 %) than in summer (NRMSE=9.698 %), autumn (NRMSE=9.959 %), and winter (NRMSE=10.169 %). Hence, IRI and IRI UP performances are quite similar in autumn and winter, while in summer and spring IRI UP strongly improves the *fo*F2 modeling. A similar seasonal behaviour characterizes also the other testing stations;

- focusing on Hermanus results for *hm*F2, IRI shows better performances in autumn (NRMSE=6.302 %) and spring (NRMSE=5.889 %) than in summer (NRMSE=7.870 %) and winter (NRMSE=8.103 %). Differently, IRI UP shows slightly better performances in autumn (NRMSE=9.507 %) than in summer (NRMSE=11.187 %), winter (NRMSE=10.085 %), and spring (NRMSE=10.388 %). From these results, one can deduce how the output of the Shubin et al. (2015) IRI option for *hm*F2 modeling is pretty reliable for quiet periods; instead, the Bilitza et al. (1979) one, used by IRI UP, in spite of the use of the effective value of $R_{12}$, shows worse results, although by far better than those obtained with $R_{12}$ (not shown here). A similar seasonal behaviour characterizes also the other testing stations;

- as pointed out in section 5.4.1, some differences arise for different ionosonde testing stations because of their different geographical location with respect to the closest GNSS stations. Usually, results for Madimbo are the worst;

- with very few exceptions, IRI UP turns out to be more accurate than IRI, for *fo*F2 , as it is perceivable by looking at the residuals mean column in Table 5.7. Considering *hm*F2, in most cases IRI is more accurate than IRI UP;

- looking at panels in Fig. 5.14, it is clear how IRI UP is able to follow the diurnal ionospheric variability shown by *fo*F2 better than IRI, above all during daytime. Some evident deviations from the climatological pattern, especially for summer and spring seasons, are well reproduced by IRI UP (as testified by the corresponding correlation coefficients which are by far better than the IRI ones).

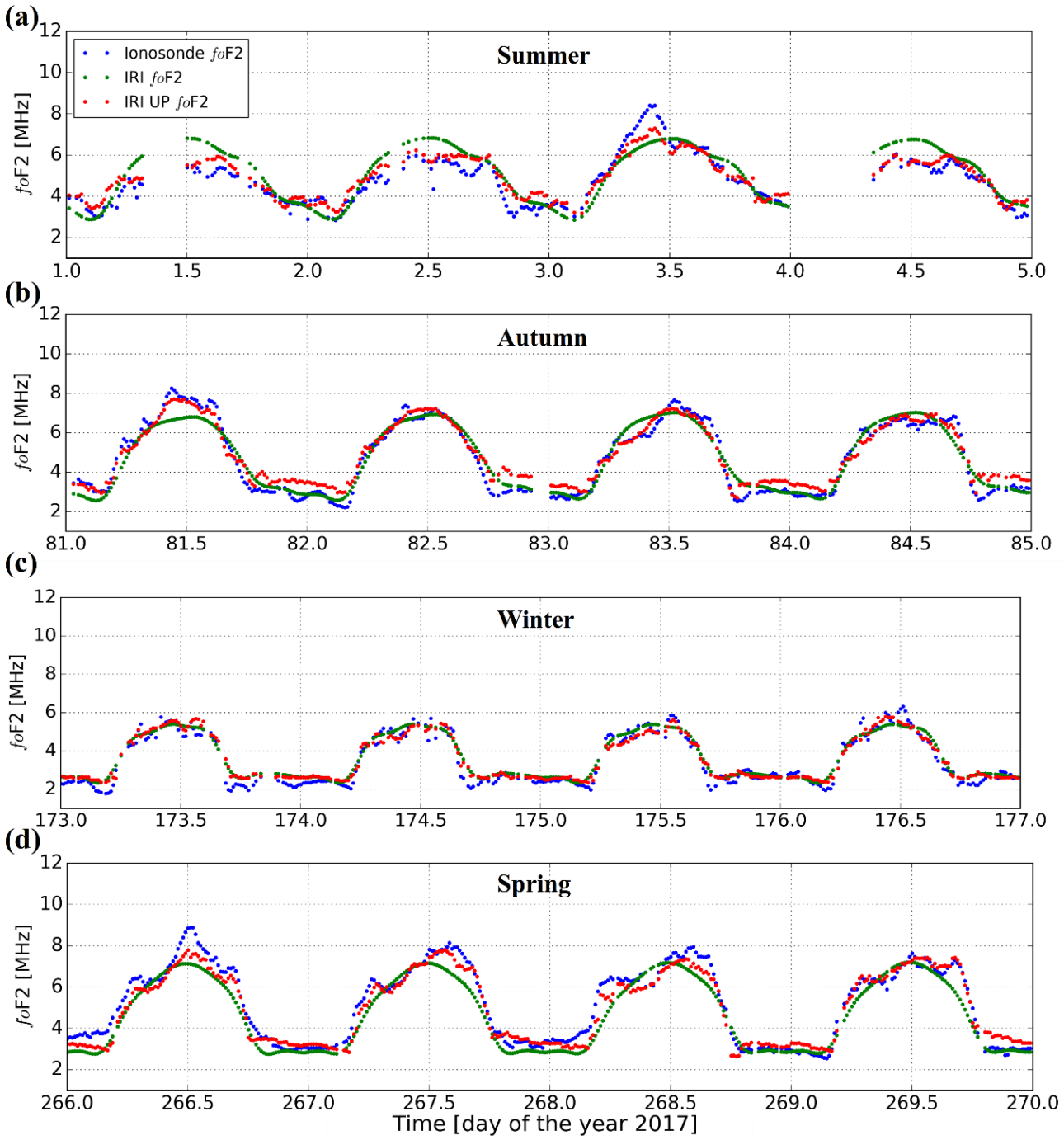


**Figure 5.14:** Time series of *fo*F2 values measured at Hermanus (blue), modeled by the IRI model (green), and modeled by the IRI UP method (red), in (a) summer (1-4 January 2017), (b) autumn (22-25 March 2017), (c) winter (22-25 June 2017), and (d) spring (23-26 September 2017).

| STATION | Model & Characteristic | RMSE [MHz for *fo*F2, km for *hm*F2] | NRMSE [%] | $\rho$ | Residuals mean [MHz for *fo*F2, km for *hm*F2] |
|---|---|---|---|---|---|
| **SUMMER, 1-4 January 2017** | | | | | |
| Grahamstown | IRI *fo*F2 | 0.764 | 15.407 | 0.865 | 0.267 |
| | IRI UP *fo*F2 | 0.400 | 8.063 | 0.960 | 0.117 |
| | IRI *hm*F2 | 21.406 | 7.971 | 0.754 | -4.349 |
| | IRI UP *hm*F2 | 26.788 | 9.975 | 0.607 | 13.586 |
| Hermanus | IRI *fo*F2 | 0.752 | 15.686 | 0.848 | 0.282 |
| | IRI UP *fo*F2 | 0.465 | 9.698 | 0.940 | 0.168 |
| | IRI *hm*F2 | 20.849 | 7.870 | 0.714 | -0.764 |
| | IRI UP *hm*F2 | 27.459 | 10.365 | 0.644 | 16.542 |
| Louisvale | IRI *fo*F2 | 0.762 | 14.983 | 0.871 | 0.189 |
| | IRI UP *fo*F2 | 0.456 | 8.964 | 0.948 | -0.052 |
| | IRI *hm*F2 | 23.368 | 8.550 | 0.660 | -7.353 |
| | IRI UP *hm*F2 | 32.451 | 11.873 | 0.286 | 15.655 |
| Madimbo | IRI *fo*F2 | 0.886 | 15.914 | 0.868 | 0.217 |
| | IRI UP *fo*F2 | 0.720 | 12.940 | 0.935 | -0.236 |
| | IRI *hm*F2 | 24.909 | 8.836 | 0.601 | -13.721 |
| | IRI UP *hm*F2 | 28.884 | 10.246 | 0.243 | 10.802 |
| **AUTUMN, 22-25 March 2017** | | | | | |
| Grahamstown | IRI *fo*F2 | 0.515 | 10.702 | 0.951 | -0.018 |
| | IRI UP *fo*F2 | 0.539 | 11.212 | 0.970 | 0.322 |
| | IRI *hm*F2 | 18.928 | 7.296 | 0.819 | -7.879 |
| | IRI UP *hm*F2 | 18.992 | 7.321 | 0.809 | 7.737 |
| Hermanus | IRI *fo*F2 | 0.527 | 10.951 | 0.956 | -0.007 |
| | IRI UP *fo*F2 | 0.479 | 9.959 | 0.976 | 0.165 |
| | IRI *hm*F2 | 16.032 | 6.302 | 0.782 | -3.164 |
| | IRI UP *hm*F2 | 19.977 | 7.852 | 0.799 | 13.084 |

| | | | | | |
|---|---|---|---|---|---|
| Louisvale | IRI *fo*F2 | 0.610 | 11.652 | 0.942 | -0.025 |
| | IRI UP *fo*F2 | 0.523 | 9.986 | 0.966 | 0.108 |
| | IRI *hm*F2 | 20.259 | 7.701 | 0.782 | -8.480 |
| | IRI UP *hm*F2 | 22.190 | 8.435 | 0.719 | 10.666 |
| Madimbo | IRI *fo*F2 | 0.542 | 10.768 | 0.958 | -0.036 |
| | IRI UP *fo*F2 | 0.647 | 12.869 | 0.960 | 0.284 |
| | IRI *hm*F2 | 21.327 | 8.158 | 0.680 | -7.487 |
| | IRI UP *hm*F2 | 26.978 | 10.319 | 0.506 | 13.187 |
| **WINTER, 22-25 June 2017** | | | | | |
| Grahamstown | IRI *fo*F2 | 0.383 | 10.753 | 0.949 | -0.044 |
| | IRI UP *fo*F2 | 0.380 | 10.644 | 0.951 | 0.041 |
| | IRI *hm*F2 | 22.923 | 9.300 | 0.837 | -10.285 |
| | IRI UP *hm*F2 | 20.903 | 8.481 | 0.830 | 3.193 |
| Hermanus | IRI *fo*F2 | 0.375 | 10.729 | 0.968 | 0.152 |
| | IRI UP *fo*F2 | 0.356 | 10.169 | 0.968 | 0.112 |
| | IRI *hm*F2 | 19.429 | 8.103 | 0.841 | -4.441 |
| | IRI UP *hm*F2 | 23.884 | 9.961 | 0.819 | 12.836 |
| Louisvale | IRI *fo*F2 | 0.378 | 10.680 | 0.954 | -0.031 |
| | IRI UP *fo*F2 | 0.453 | 12.763 | 0.932 | -0.001 |
| | IRI *hm*F2 | 22.607 | 9.088 | 0.830 | -9.988 |
| | IRI UP *hm*F2 | 23.692 | 9.524 | 0.801 | 8.270 |
| Madimbo | IRI *fo*F2 | 0.357 | 9.313 | 0.969 | -0.055 |
| | IRI UP *fo*F2 | 0.564 | 14.712 | 0.919 | 0.062 |
| | IRI *hm*F2 | 21.389 | 8.976 | 0.803 | -6.209 |
| | IRI UP *hm*F2 | 25.875 | 10.859 | 0.749 | 12.201 |
| **SPRING, 23-26 September 2017** | | | | | |
| Grahamstown | IRI *fo*F2 | 0.731 | 14.064 | 0.955 | -0.501 |
| | IRI UP *fo*F2 | 0.429 | 8.259 | 0.972 | -0.143 |
| | IRI *hm*F2 | 18.541 | 7.067 | 0.824 | 0.190 |
| | IRI UP *hm*F2 | 22.169 | 8.449 | 0.823 | 11.461 |

| | | | | | |
|---|---|---|---|---|---|
| Hermanus | IRI *fo*F2 | 0.715 | 13.736 | 0.958 | -0.484 |
| | IRI UP *fo*F2 | 0.463 | 8.892 | 0.973 | -0.173 |
| | IRI *hm*F2 | 15.144 | 5.889 | 0.866 | 4.656 |
| | IRI UP *hm*F2 | 24.523 | 9.537 | 0.812 | 18.396 |
| Louisvale | IRI *fo*F2 | 0.731 | 14.398 | 0.950 | -0.445 |
| | IRI UP *fo*F2 | 0.548 | 10.800 | 0.962 | -0.271 |
| | IRI *hm*F2 | 18.694 | 7.082 | 0.747 | 0.999 |
| | IRI UP *hm*F2 | 26.795 | 10.151 | 0.662 | 17.059 |
| Madimbo | IRI *fo*F2 | 0.670 | 12.450 | 0.951 | -0.255 |
| | IRI UP *fo*F2 | 0.470 | 8.725 | 0.962 | -0.127 |
| | IRI *hm*F2 | 23.452 | 8.938 | 0.700 | 1.655 |
| | IRI UP *hm*F2 | 29.755 | 11.340 | 0.634 | 16.003 |

**Table 5.7:** Same as Table 5.5 but for the four time periods representing different seasons. Best statistical results, between IRI and IRI UP, are highlighted in red for each ionospheric characteristic.

## Summary

In this chapter an assimilative procedure to update the ionospheric empirical climatological IRI model by assimilating GNSS derived vTEC measurements, developed in the framework of the IRI UP method, has been described. Differently from the IRI UP procedure described in chapter 4, *fo*F2 and *M*(3000)F2 values required to calculate the ionospheric effective indices $IG_{12\text{eff}}$ and $R_{12\text{eff}}$, are now derived from assimilated vTEC values recorded by ground-based GNSS stations and not from a network of ionosondes. The mathematical procedure for deriving *fo*F2 and *MUF*(3000)F2 values from vTEC values has been described and tested. Mathematical relationships between *fo*F2 (*MUF*(3000)F2) and vTEC are derived using South African co-located ionosondes and GNSS stations, thus they are valid only for the South African region. Nevertheless, the same procedure can be applied with profit in each region where such data are available. The goodness of the proposed IRI UP method based on assimilated vTEC values has been tested for several quiet and disturbed days in 2017, comparing *fo*F2 and *hm*F2 values modeled by IRI and IRI UP with those measured by ionosondes. IRI UP exhibits better performances than IRI for *fo*F2, for most of the analyzed cases. Concerning *hm*F2, evident improvements are achieved by IRI UP for very disturbed conditions, similar results between IRI UP and IRI characterize instead moderately disturbed periods, while for quiet conditions the IRI UP output seems to be worse than the IRI one.

Compared to the performance achieved by the IRI UP method when assimilating *fo*F2 and *M*(3000)F2 ionosonde data (chapter 4), the IRI UP procedure here described, based on the assimilation of vTEC data, produces slightly worse results. Nevertheless, it produces better results than the IRI model, particularly during disturbed periods. The proposed method has however the great ability to be applied also in very remote regions, due to the very good coverage of the terrestrial surface that GNSS receivers have been achieved in recent years. From this point of view, the proposed method, while being less accurate than that relying on ionosonde assimilated data, has a field of applicability wider than the previous one and can be a good implementation of the IRI model for nowcasting purposes.

# 6. On the use of SWARM's satellites data for modeling the topside ionosphere

The topside ionosphere extends from the F2-layer peak height (*hm*F2) to the Upper Transition Height (UTH, the height between the oxygen ion and hydrogen ion dominated plasma regions). The plasma density distribution of this region is largely determined by field-aligned plasma flows and plasma transport processes (chapter 1) and, because of the large fraction of the Total Electron Content (TEC) it contains, its modeling is extremely important for telecommunication's purposes.

Knowledge of the physical and chemical state of the plasma in this region is very problematic because equipment commonly used to sound the ionosphere are not able to probe it (chapter 1). In fact, ground-based ionosondes can only measure the bottomside part of the vertical electron density profile, up to the height (*hm*F2) of the F2-layer electron density peak (*Nm*F2). This task requires the use of more sophisticated and expensive techniques and equipment: topside sounders, Radio Occultation (RO), Incoherent Scatter Radars (ISR), and Langmuir probes on board Low Earth Orbit (LEO) satellites.

Difficulties in modeling the topside part of the ionosphere are testified by the fact that often the IRI model (Bilitza et al. 2014, 2017) does not represent properly the real features of this part of the ionosphere (Bilitza et al. 2006), as described in chapter 2. Over the years, different topside formulations were introduced in IRI but, despite these huge efforts, the topside modeling is still a challenge for the IRI model. This is why, after having reached good and reliable results for the bottomside region with the IRI UP method, the attention is now addressed on the topside region modeling.

In section 6.1, some results of a comparison between IRI modeled values and preliminary Swarm Langmuir probe measurements of the electron density in the topside ionosphere, during the St. Patrick storm period (March 2015), are described. Overall results are published in:

*Pignalberi, A., Pezzopane, M., Tozzi, R., De Michelis, P., Coco, I. (2016) Comparison between IRI and preliminary Swarm Langmuir probe measurements during the St. Patrick storm period. Earth, Planets and Space, 68(93). doi:10.1186/s40623-016-0466-5*

In section 6.2, a procedure aiming to model the topside part of the vertical electron density profile of the ionosphere by using Swarm Langmuir probe measurements and IRI UP method, is described. Overall results are published in:

*Pignalberi, A., Pezzopane, M., Rizzi, R. (2018) Modeling the lower part of the topside ionospheric vertical electron density profile over the European region by means of Swarm satellites data and IRI UP method. Space Weather, 16, 304–320. https://doi.org/10.1002/2017SW001790*

## 6.1 Comparison between IRI and preliminary Swarm Langmuir probe measurements during the St. Patrick storm period (March 2015)

To evaluate the IRI performance in modeling the topside ionosphere, under quiet and disturbed conditions, preliminary Swarm electron density ($N_e$) measurements recorded during six time periods of March 2015, including the St. Patrick geomagnetic storm occurred on March 17 2015, are considered and compared with the corresponding output given by the IRI model. The six time intervals considered to carry out this comparison are chosen to have two quiet periods before the onset of the storm, two periods including the main phase of the storm, and two periods during the recovery phase of the storm.

The Swarm $N_e$ measurements, although preliminary and under validation, are considered nowadays reliable by the reference community. For example, Pedatella et al. (2015) showed a comparison between Swarm densities and those inferred from COSMIC Radio Occultation measurements and found a very good agreement.

Nevertheless, since the Swarm data validation is still on-going, a comparison with IRI in terms of absolute values would have not been suitable. So, a correlation analysis is considered to evaluate the trends embedded in the Swarm and IRI time series.

### 6.1.1 Swarm satellites

Swarm is a satellites constellation launched at the end of 2013 by the European Space Agency (ESA) (Friis-Christensen et al. 2006). It is constituted by three LEO satellites in a circular near-polar orbit. Two of them (called Alpha (A) and Charlie (C)) are orbiting the Earth side-by-side at the same altitude of about 460 km (with an inclination of 87.4°, an east-west separation of 1-1.5° in longitude, and a maximal differential delay in orbit of approximately 10 seconds), while the third (Bravo (B)) is flying about 60 km above (with an inclination of 88°) in an orbital plane which will gradually get farther from those of the other two satellites during the mission lifetime (9 hours in local time after 4 years), as depicted in Fig. 6.1.

They are all equipped with identical instruments consisting of high-resolution sensors for measuring both geomagnetic and electric fields, as well as plasma density. In particular, the electron density measurements at 2 Hz rate made by the Langmuir probes carried by each satellite (Lomidze et al. 2017) have been used for this work.

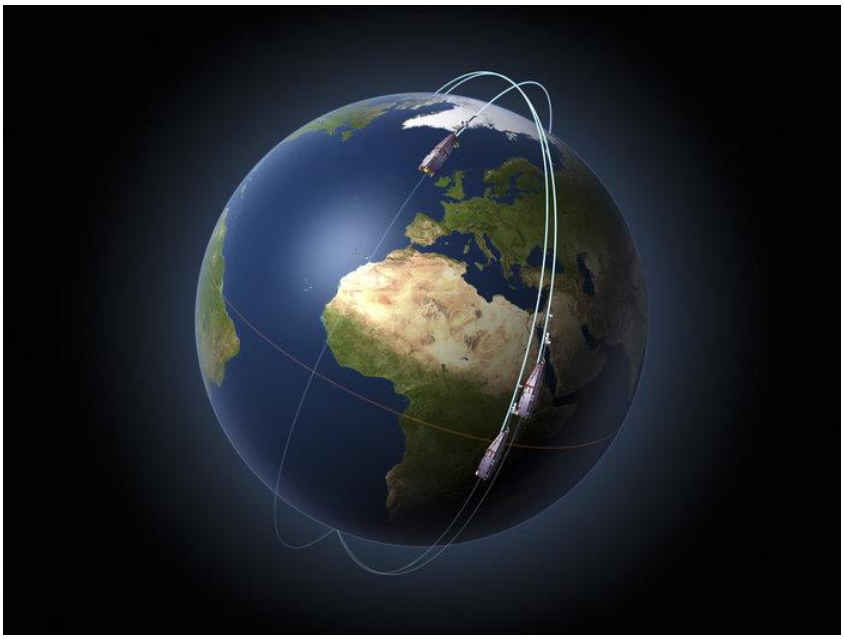

**Figure 6.1:** The Swarm satellites constellation geometry.

### 6.1.2 Data used and period under study

Preliminary measurements of electron density recorded by the Swarm constellation during March 2015, a period of time including the so-called "St. Patrick storm", are analysed. This storm was the most intense storm observed during the Solar Cycle 24, that is classified as severe and for which the $K_p$ index reached the maximum value of 8. At ground observatories the sudden storm commencement was observed at around 04:45 UT of 17 March 2015 with the arrival at the Earth of a coronal mass ejection. Fig. 6.2 shows the temporal trend (from 9 to 25 March 2015) of the South component of the interplanetary magnetic field ($B_z$ in Fig. 6.2), and of some geomagnetic indices ($D_{st}$, *AE* and 3-hourly $a_p$) describing the global level of the Earth's magnetic disturbance. The maximum intensity of the storm was reached at around 23:00 UT of 17 March 2015 and was characterized by the minimum value of $D_{st}$ index of -223 nT.

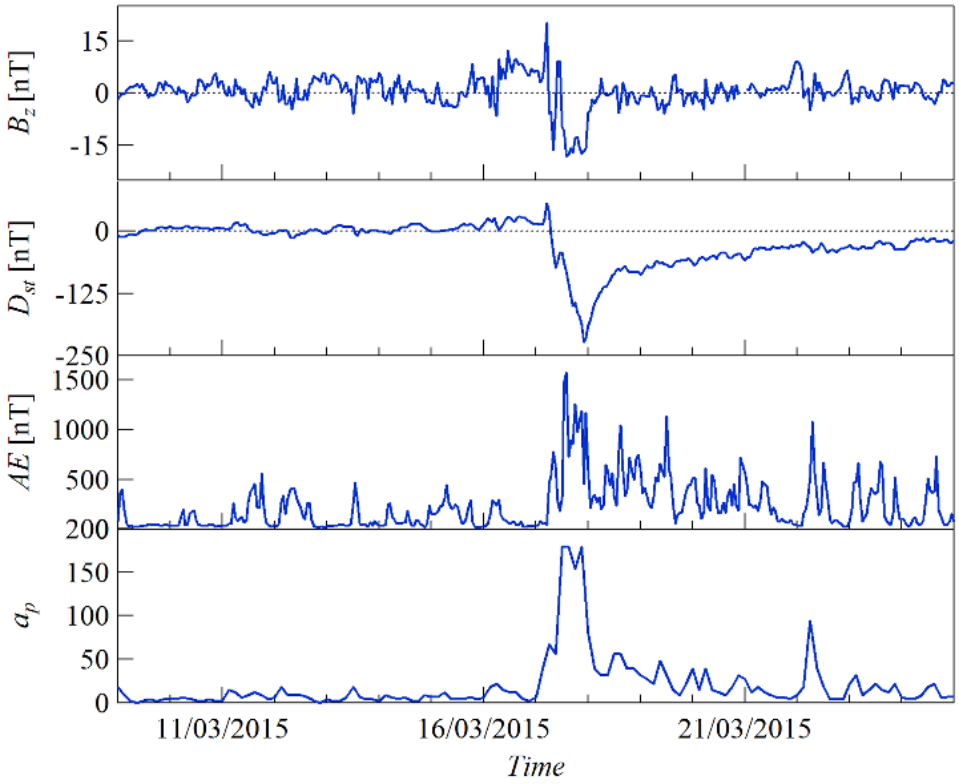


**Figure 6.2:** The St. Patrick magnetic storm as seen by (from top to bottom): Interplanetary Magnetic Field $B_z$ component and $D_{st}$, *AE*, and 3-hourly $a_p$ magnetic indices.

Data used consist of electron density measurements made on board the three satellites of Swarm constellation during six selected time periods between 9 and 25 March 2015. During this time window, the Swarm's available data are those from the Preliminary Plasma Dataset prepared by the Swedish Institute for Space Science (IRF) at Uppsala (Knudsen et al. 2015). They are not calibrated. Specifically, only data with a quality flag value lower than 256 are considered (Knudsen et al. 2015).

The purpose of this investigation is to compare Swarm and IRI electron density representations for both disturbed and quiet magnetic conditions. For this reason, are chosen:

a) two quiet periods before the onset of the St. Patrick storm, named as pre-storm time intervals (P1 and P2), characterised by a significant low value of the $D_{st}$ index;
b) two disturbed periods including the main phase of the storm, named as main phase periods (M1 and M2);

c) two moderately disturbed periods during the recovery phase of the storm, named as recovery phase periods (R1 and R2).

Detailed information on the bounds of the selected periods are summarised in Table 6.1, together with the range of variability and average values of $D_{st}$ and *AE* indices in each period.

| **Period code** | **Day - Start time [UT]** | **Day - End time [UT]** | $D_{st}$ **[nT]** | $\langle D_{st} \rangle$ **[nT]** | ***AE*** **[nT]** | $\langle AE \rangle$ **[nT]** |
|---|---|---|---|---|---|---|
| P1 | 09 – 03 04:30:00 | 11 – 03 04:30:00 | [-9,15] | 4 | [22,261] | 57 |
| P2 | 12 – 03 17:30:00 | 13 – 03 11:30:00 | [-4,12] | 5 | [-17,259] | 48 |
| M1 | 17 – 03 00:00:00 | 17 – 03 23:59:59 | [-223,56] | -72 | [37,1570] | 633 |
| M2 | 18 – 03 00:00:00 | 18 – 03 23:59:59 | [-189,-70] | -105 | [200,1043] | 488 |
| R1 (A) | 19 – 03 00:00:00 | 20 – 03 11:57:33 | [-88,-48] | -63 | [61,1134] | 381 |
| R1 (B) | 19 – 03 00:00:00 | 20 – 03 03:42:33 | [-88,-53] | -65 | [61,1134] | 396 |
| R1 (C) | 19 – 03 00:00:00 | 20 – 03 23:59:59 | [-88,-44] | -59 | [61,1134] | 373 |
| R2 | 24 – 03 00:00:00 | 25 – 03 23:59:59 | [-36,9] | -16 | [28,729] | 165 |

**Table 6.1:** Details on the time periods selected and on the corresponding level of magnetic activity expressed by means of $D_{st}$ and *AE* magnetic indices. Due to gaps in the 20 March Swarm A and B data, R1 time periods differ from satellite to satellite.

The pre-storm periods P1 and P2 are chosen according to simultaneously low values of both $D_{st}$ and *AE* indices, in order to be quite confident that the magnetic activity was low at all latitudes. In fact, the well-known $D_{st}$ index is able to represent the disturbance observed on the ground at low- and mid-latitudes produced by the ring current, the partial ring current, and by magnetopause and magnetotail currents during magnetic storms. Differently, *AE* index indicates the total intensity of the auroral electrojets and is used to represent the disturbance observed at high latitudes due to geomagnetic substorms. Consequently, when $D_{st}$ is low, *AE* is not necessarily low as well. The main phase periods M1 and M2 correspond to the growth of the ring current till its maximum intensity and up to its initial decay, respectively. With regard to the recovery phase periods, R1 is selected midway, in terms of $D_{st}$ index, between the main phase and quiet conditions, while during R2 quiet conditions are almost achieved.

In order to compare the 2 Hz Swarm $N_e$ measurements with the $N_e$ values provided by the IRI model at the same time and location, Swarm data are resampled, actually decimated, taking 1 measurement every 9, which corresponds to a sampling of 4.5 s. This value descends by the fact that the IRI temporal step is expressed as tenths of hour, and at the same time has to be a multiple of 0.5 s, that is the Swarm sampling. The smallest value matching these two constraints is the decimal temporal increment of 0.00125 hours, which corresponds to 4.5 s. The other temporal steps multiple of 0.5 s, and lower than 4.5 s, would give rise to circulating decimal temporal increments, which would result in an inaccurate analysis.

Data from each period shown in Table 6.1 are grouped according to Magnetic Local Time (MLT) sectors and magnetic latitude bands. Partition into MLT is made to consider that, for each Swarm orbit, half measurements are taken in the morning sector (descending phase of satellite orbit) and half in the evening sector (ascending phase of satellite orbit). So, dividing data in this way one distinguishes among the different dynamics characterizing

morning and evening ionospheric sectors, especially at low and equatorial latitudes that are characterized by the fountain effect (Davies 1990; Kelley 2009). Since Swarm satellites move along near-polar orbits, MLTs are clustered around morning and evening sectors and partially spread over the entire 24 h MLT range at the poles. So, to consider disjoint set of measurements, the MLT ranges considered for the descending and ascending phases are 04-12 h and 16-24 h. Within these time intervals, over 99% of measurements taken between magnetic latitudes of 60° S and 60° N fall in the range 06-09 h for Swarm A and C and in the range 08-11 h for Swarm B, for the morning sector, and in the range 18-21 h for Swarm A and C and in the range 20-23 h for Swarm B, for the evening sector. Differently, at latitudes higher than 60° the percentage of measurements taken in the morning and in the evening sectors decreases to around 60%, being the orbits not really polar. Fig. 6.3 shows the overall distribution of measurements in MLT without distinguishing between high and low/mid-latitudes.

The reason for the splitting into magnetic latitude bands is similar to that behind the partition in MLT. In fact, most of processes occurring in the ionosphere have a marked magnetic latitudinal dependence (Davies 1990; Kelley 2009). So, geographical coordinates are converted into quasi-dipole coordinates (Emmert et al. 2010), and the following magnetic latitude bands are defined: between -90° S and -60° S (SP, South Pole), between -60° S and -30° S (SM, South Mid), between -30° S and 30° N (EQ, Equator), between 30° N and 60° N (NM, North Mid), between 60° N and 90° N (NP, North Pole). The limits of these bands are chosen also on the base of the distribution in magnetic latitude of Swarm $N_e$ measurements. Two examples are shown in Fig. 6.4 for Swarm A, during the quiet period P1 for measurements recorded in the morning sector, and during the perturbed period M2 for measurements in the evening sector.

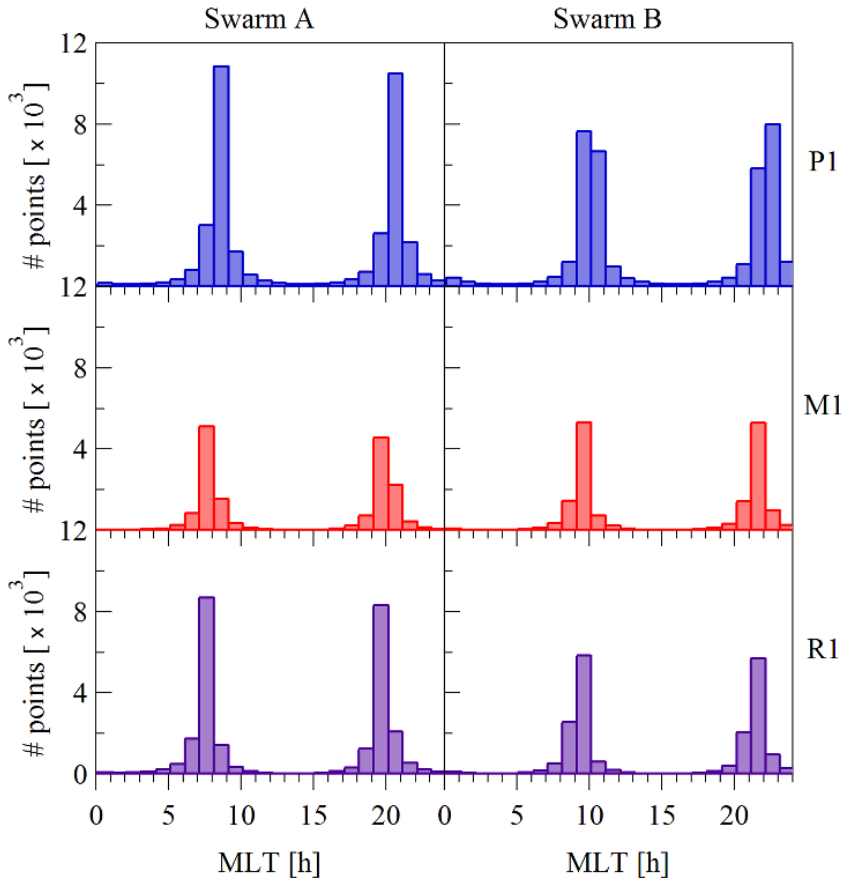


**Figure 6.3:** Histograms of available electron density values as a function of MLT, for P1, M1, and R1, for Swarm A and B. Due to the geometry of Swarm constellation the MLT distribution of Swarm C is identical to that of Swarm A.

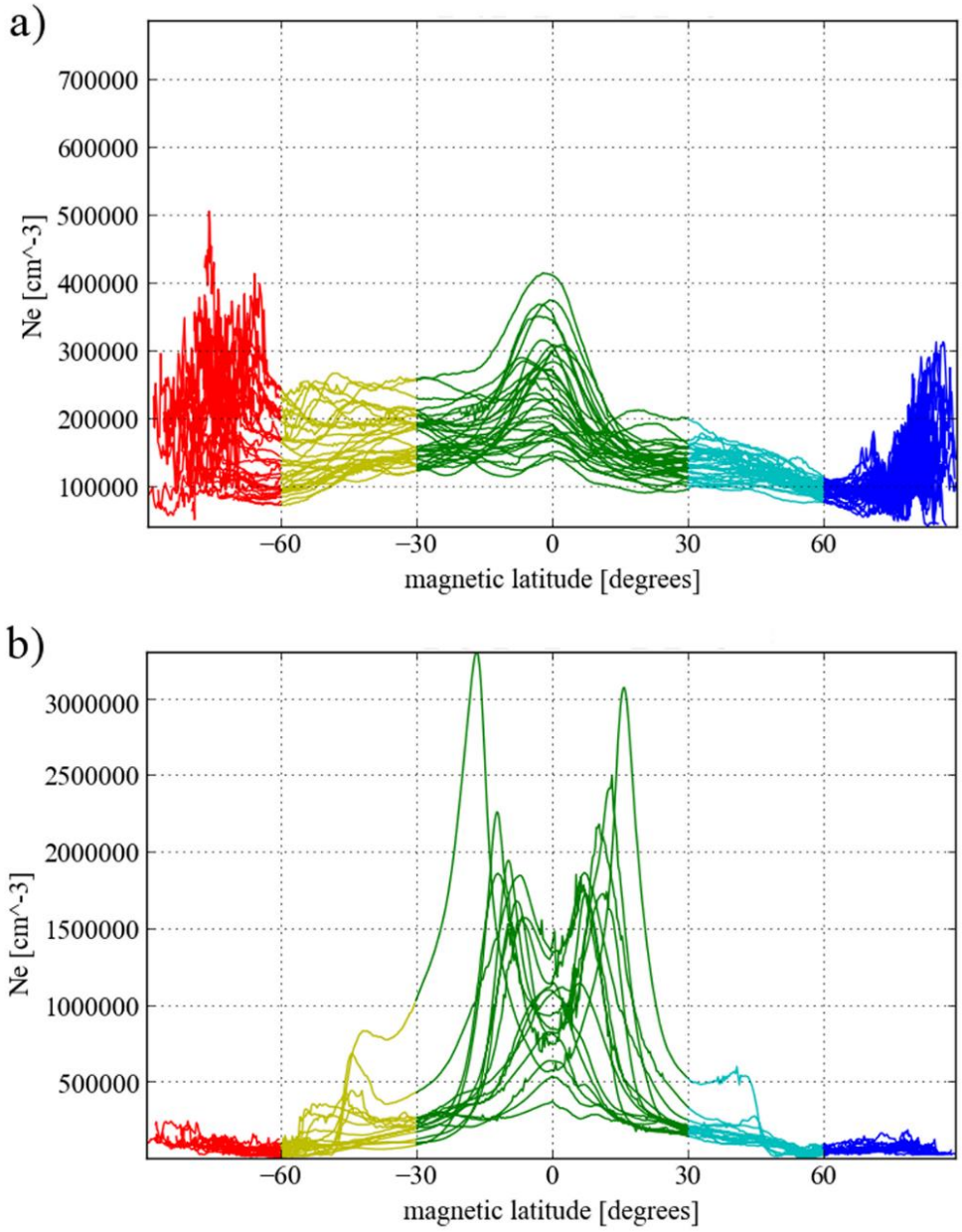


**Figure 6.4:** Electron density as measured by Swarm A during a) the pre-storm period P1 in the morning sector (04-12 MLT), and b) during the main-phase period M2 in the evening sector (16-24 MLT).

### 6.1.3 Results of the comparison

The IRI model is used to estimate $N_e$ at the same time and location (geographical latitude and longitude, altitude) of Swarm measurements falling in the six selected periods listed in Table 6.1. In detail, the URSI coefficients (Rush et al. 1989) are used (URSI coefficients are recommended when considering areas mainly covered by oceans), according to the three IRI topside options (IRI-2001, IRI-2001corr, IRI-NeQuick) described in chapter 2, and with the STORM option on (StON).

In order to assess quantitatively the behavior of the different IRI topside models, for each selected period, and for each model, the Pearson correlation coefficient between Swarm and IRI time series is calculated, according to the following formula:

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{E}\big(\big(X - \mathrm{E}(X)\big)\big(Y - \mathrm{E}(Y)\big)\big)}{\sigma_X \sigma_Y}, \tag{6.1}$$

where cov() is the covariance between the variables $X$ and $Y$, $\sigma_X$ e $\sigma_Y$ are the corresponding standard deviations, and E() represents the expected value.

This approach is chosen because the Swarm Langmuir Probes data are still undergoing a thorough validation, hence a comparison in terms of absolute values is not appropriate. On the contrary, the value of the Pearson coefficient can give an idea about the similarity of trends embedded in the IRI and Swarm time series.

Figs. 6.5-6.8 show electron density values measured by Swarm A with corresponding ones calculated by IRI-NeQuick(StON), and electron densities measured by Swarm B with corresponding ones calculated by IRI-2001corr(StON), for some of the studied periods, for morning and evening sectors. Overall results, in terms of Pearson correlation coefficient, are shown in Figs. 6.9-6.11. Figs. 6.9-6.10 show the average of Pearson coefficients calculated for Swarm A and C, and Pearson coefficients calculated for Swarm B, by considering all the three IRI topside options (IRI-2001, IRI-2001corr, IRI-NeQuick), respectively for morning and evening sectors, and for each magnetic latitude band. Correlations for Swarm A and C are averaged since obtained results are practically identical. Fig. 6.11 displays instead the magnetic latitude dependence of Pearson coefficients shown in Figs. 6.9-6.10. At Swarm altitudes, the three IRI topside options give a very similar ionospheric representation, and that is why different IRI topside options for Swarm A and B have been represented.

Looking at Figs. 6.5-6.8 several interesting features measured by Swarm satellites, and differences between these and IRI, come out.

Concerning the period P1 (the same is for period P2), for the morning sector (Fig. 6.5), the equatorial band shows, for both IRI and Swarm, the same usual pattern characterized by a maximum of electron density along the magnetic equator (e.g., Balan and Bailey 1995). Anyway, some differences about the electron density intensity appear: Swarm A measures $N_e$ values lower than those calculated by IRI, while the contrary holds for Swarm B. This dissimilarity could be related to the local time shift characterizing the two satellites (see Figs. 6.1 and 6.3) but, more likely, is due to their different orbital altitudes, a fact that, when having accurate measurements, will turn out to be really useful to obtain new insights about the topside plasma scale height, which is so important to reliably model the topside profile. Fig. 6.5 shows also, for Swarm A, a general underestimation made by IRI in the southern part of Atlantic Ocean. Concerning the polar regions, the values measured by Swarm satellites are pretty different from those given by IRI and, with regard to this, the most striking feature is the very low values of correlation coefficients characterizing the northern polar region for Swarm B (Fig. 6.10).

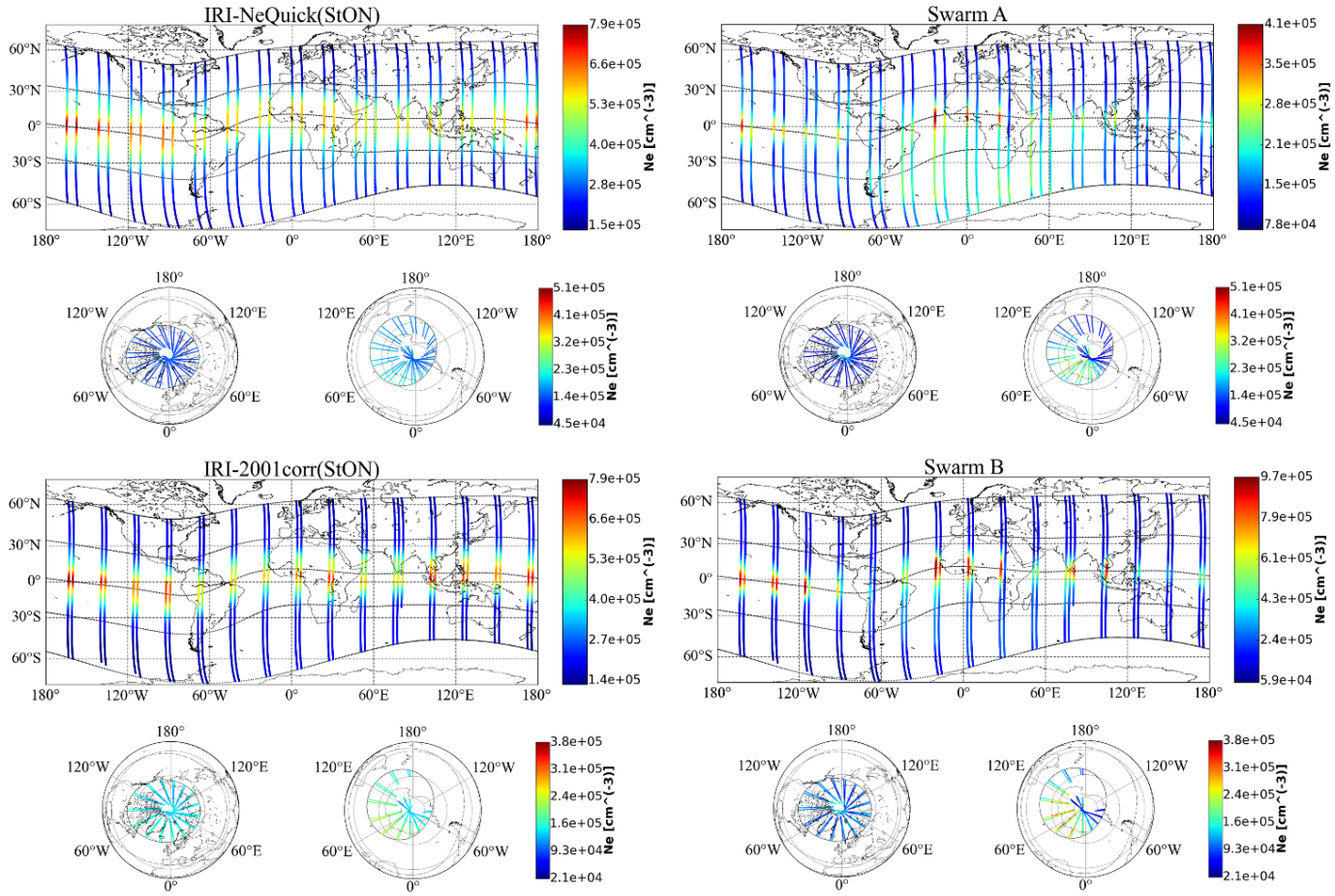


**Figure 6.5:** Electron densities measured by Swarm A (top-right panels) and Swarm B (bottom-right panels), and the corresponding ones calculated by IRI-NeQuick(StON) (top-left panels) and IRI-2001corr(StON) (bottom-left panels), for the period P1, for the morning sector. Magnetic latitude bands between -60° and 60° are plotted in a Gall stereographic projection, while the high latitude bands are plotted in an orthographic projection (on the left the north pole, on the right the south pole). Coordinates are geographic and bold lines in both Gall stereographic projections and polar orthographic projections represent magnetic parallels drawn with a 30° step. Due to the large difference between $N_e$ values measured by Swarm A and B and those estimated by IRI, it is not possible to draw the values into the Gall stereographic projections with the same colour scale.

About period M1 (the same is for the period M2) for the morning sector (Fig. 6.6), IRI still models an equatorial pattern characterized by a maximum centered on the magnetic equator, while Swarm measures a double-crest pattern in the west longitude sector of the globe, that is unusual for these local times. In fact, at these local times, the zonal electric field is westward and gives rise to a reverse fountain causing an increase of electron density around the magnetic equator, according to the mechanism proposed by Balan and Bailey (1995). The double-crest pattern measured by Swarm can be ascribed to an early fountain effect caused by ionospheric electric fields and currents that at low- and mid-latitudes, during geomagnetic disturbed periods, can significantly differ from their quiet-day patterns. In addition, the polar patterns given by IRI are different from those measured by Swarm, showing a general overestimation of $N_e$. This feature is confirmed by the low values of correlation coefficients characterizing the polar bands (Figs.6.9-6.10), especially for Swarm B.

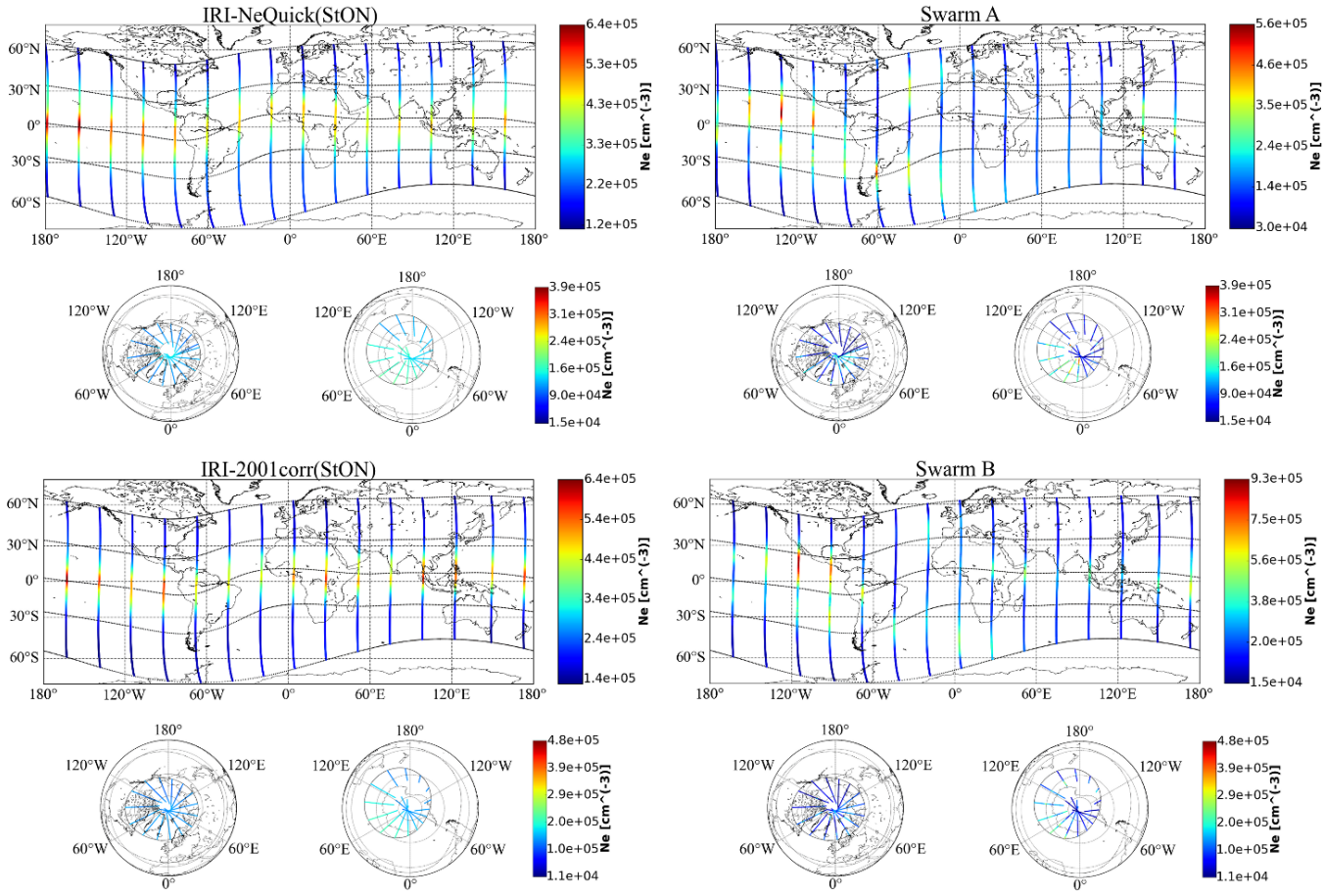


**Figure 6.6:** Same as Fig. 6.5, but for M1 morning sector.

Regarding the period R1 (the same is for R2), for the morning sector, Swarm comes back to the usual pattern characterized by a maximum centered on the magnetic equator, as on the other hand is modeled by IRI; anyway, along the Pacific Ocean sector, Swarm measures $N_e$ values higher than IRI ones, while the rest of values are lower than those modeled by IRI. Again, the correlation coefficients of polar regions are the lowest ones, confirming that also for the R1 period Swarm and IRI patterns are different.

Concerning the period P1 (the same is for the period P2), for the evening sector (Fig. 6.7), the equatorial band shows for both IRI and Swarm the same usual electron density double-crest pattern around the magnetic equator. The values measured by Swarm, that show a maximum over the south American sector, are however higher than those given as output by IRI. Moreover, the IRI values are significantly asymmetric with those of the southern crest that are higher than those of the northern crest. On the contrary, Swarm satellites measure two crests that are very similar, there is only a slight difference over the South America, where Swarm presents the maximum of $N_e$, for which the northern crest is more intense than the southern one. As for the morning sector, also for the evening sector, the northern polar region is characterized by the lowest values of correlation coefficients (Figs. 6.9-6.10).

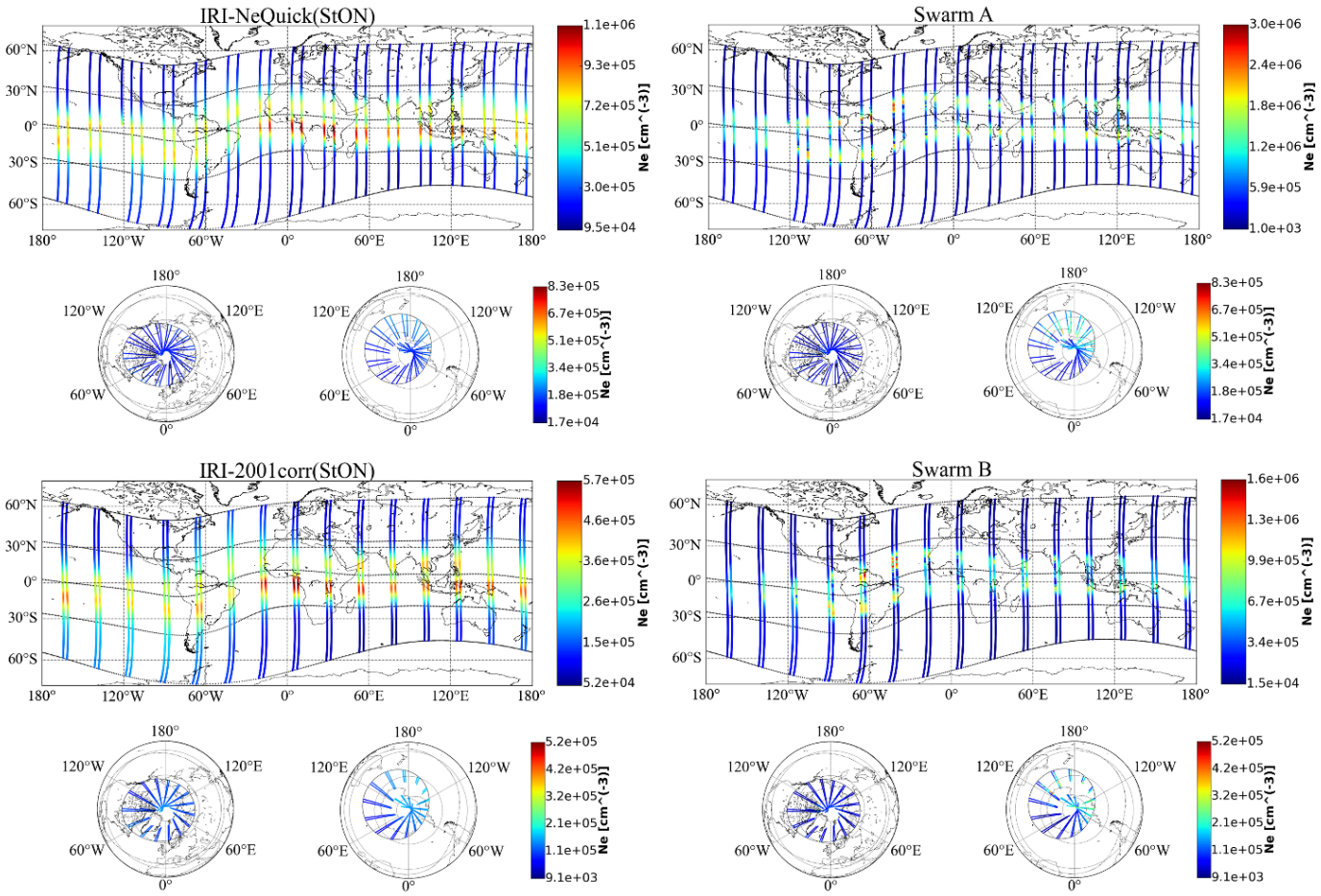


**Figure 6.7:** Same as Fig. 6.5, but for P1 evening sector.

With reference to period M1 (the same is for the period M2), for the evening sector (Fig. 6.8), the usual electron density double-crest pattern around the magnetic equator is still shown by both IRI and Swarm, even though the crests measured by Swarm are noticeably narrower than those modeled by IRI. Anyway, also during the main phase of the storm, the values measured by Swarm, which present again a maximum over the South-American sector, are higher than those modeled by IRI. Moreover, as for the period P1, IRI still models electron density crests that are significantly asymmetric, with the southern crest which is notably more intense than the northern one. It is not the same for the electron density crests measured by Swarm, which appear quite symmetric. During the main phase of the storm, the polar patterns characterizing IRI and Swarm are again sensibly different, especially for the period M2; the most striking feature is the very low values of correlation coefficient associated to the northern polar region.

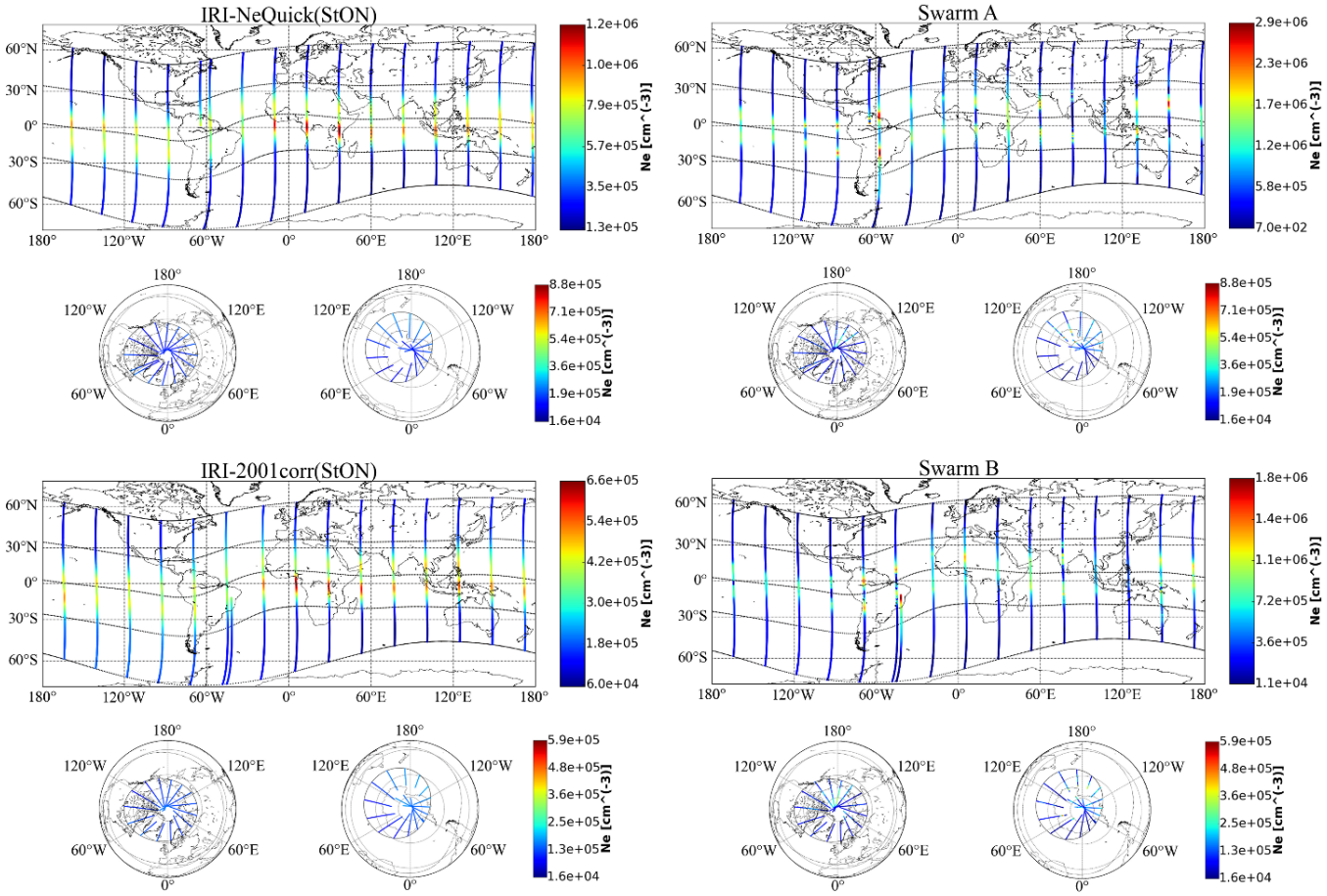


**Figure 6.8:** Same as Fig. 6.5, but for M1 evening sector.

Concerning the period R1 (the same is for the period R2), for the evening sector, the morphology of all latitude bands is very similar to that of periods M1/M2. The only difference is that the maximum values of $N_e$ measured by Swarm are now spread on a wider longitude sector including also the Atlantic Ocean.

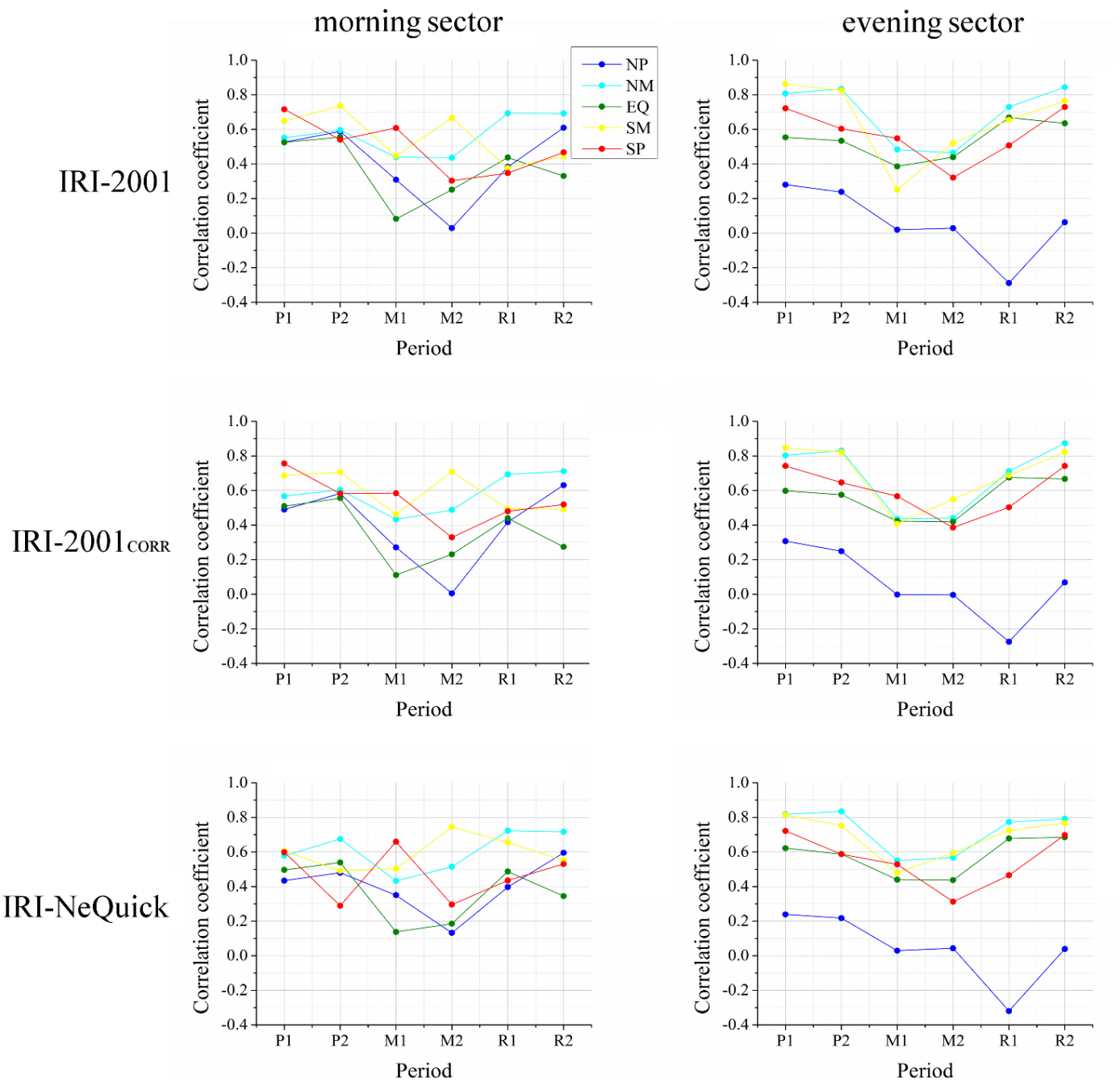


**Figure 6.9:** Average of Pearson correlation coefficients calculated for Swarm A and C, for morning (left panels) and evening (right panels) sectors, for each magnetic latitude band (NP, NM, EQ, SM, SP), by considering all the three IRI topside models: IRI-2001 (top panels), IRI-2001corr (middle panels), and IRI-NeQuick (bottom panels).

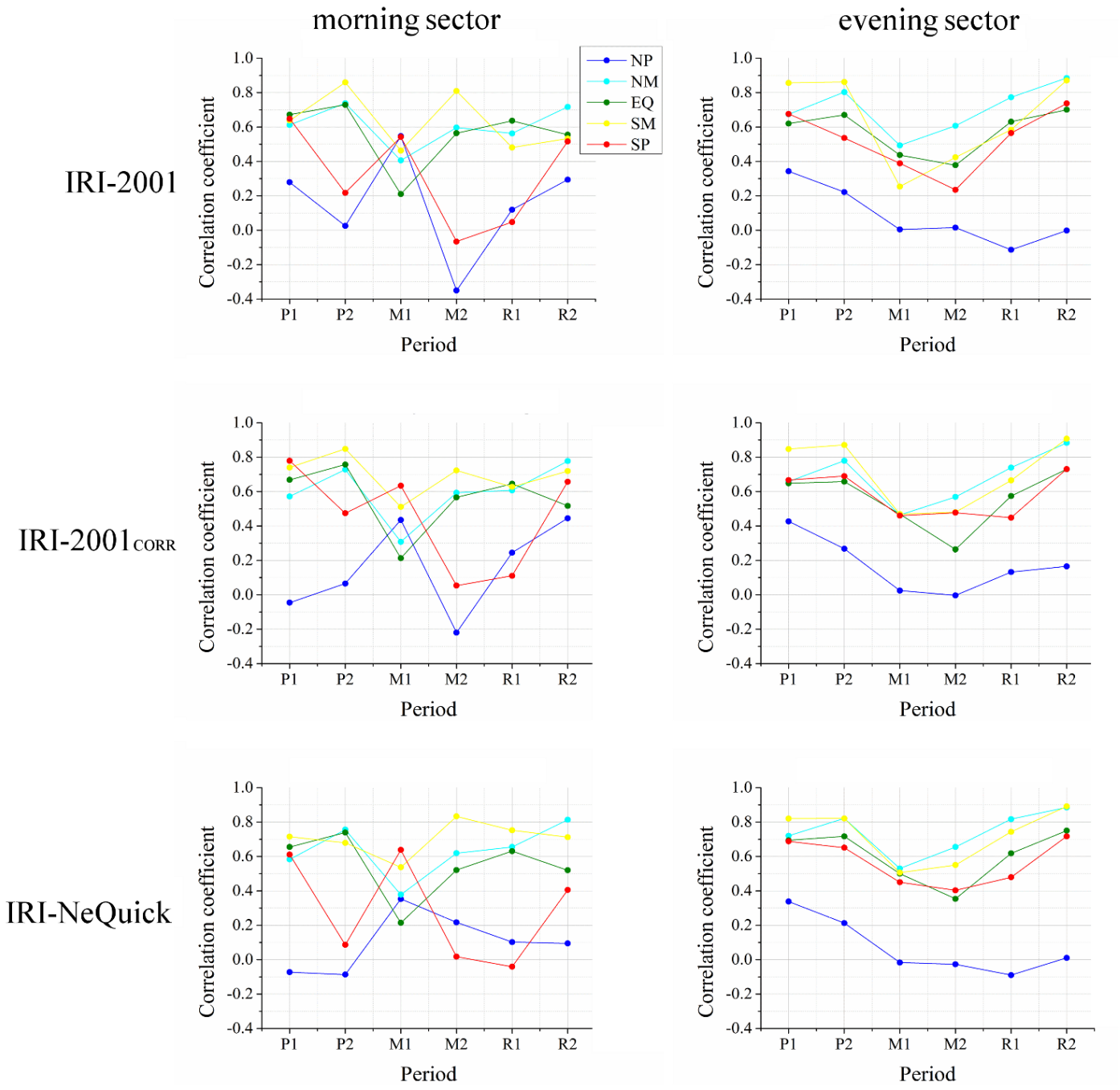


**Figure 6.10:** Same as Fig. 6.9 but for Swarm B.

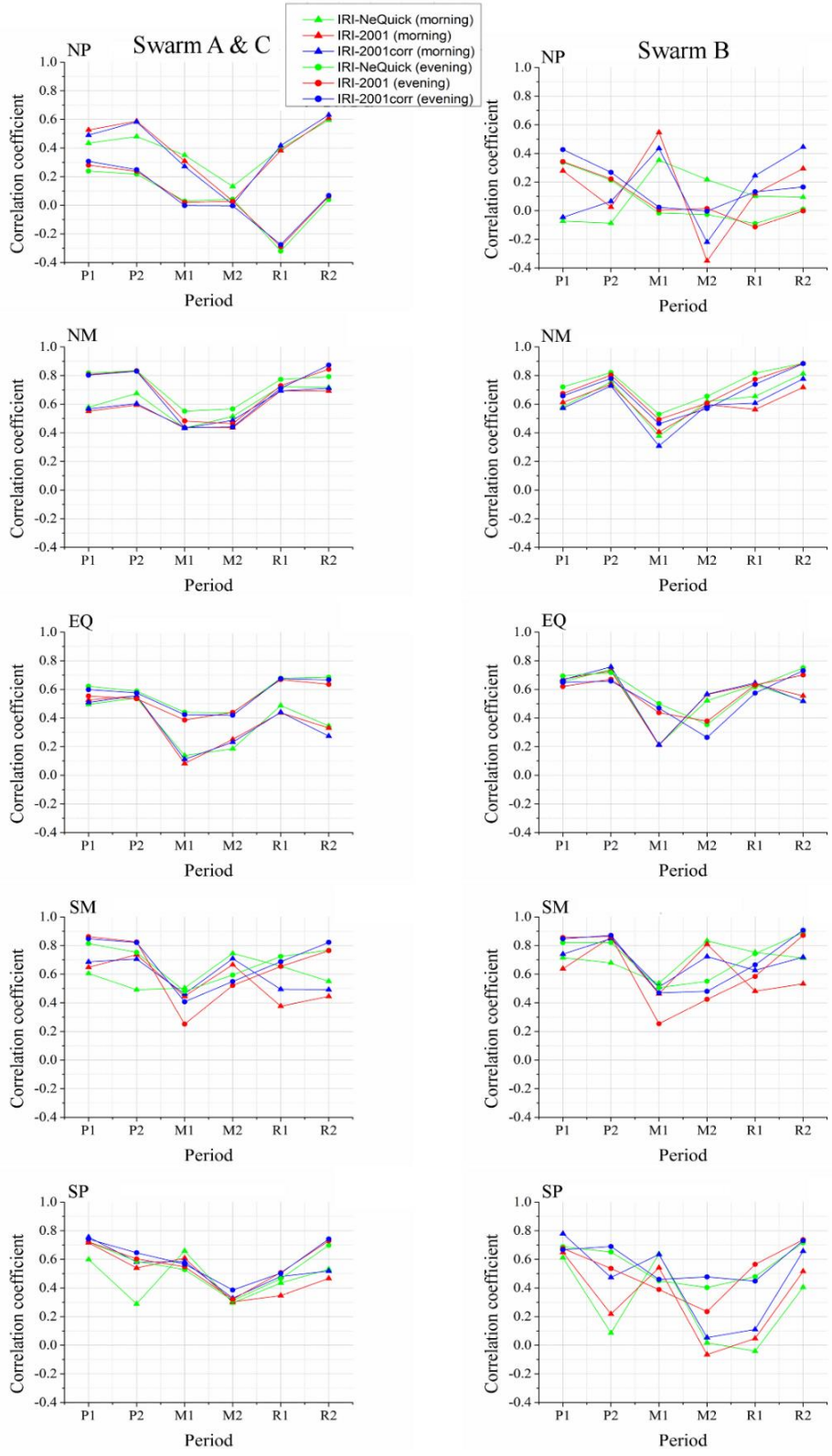


**Figure 6.11:** Average of Pearson correlation coefficients calculated for Swarm A and C (left panels), and Pearson correlation coefficients calculated for Swarm B (right panels), for each magnetic latitude band (from top to bottom: NP, NM, EQ, SM, and SP), by considering all the three IRI topside models (IRI-2001, IRI-2001corr, IRI-NeQuick), for morning and evening sectors.

In summary, Figs. 6.5-6.11 show that, from a morphological point of view, the electron density patterns measured by Swarm and those modeled by IRI are different, especially during the main phase of the storm, for the morning sector, when Swarm highlights an unusual double-crest pattern. As a consequence, the correlation coefficients between IRI and Swarm of all magnetic latitude bands are somewhat low, mainly in the periods M1, M2, and R1. In general, the correlation coefficients of mid latitude regions are higher than those of equatorial and polar regions, confirming a well-known feature of IRI, that is IRI predictions are less accurate at equatorial and auroral latitudes (Bilitza and Reinisch 2008).

The correlation coefficients of the northern polar region deserve a special mention, because they are often very low (even negative at times), for both the morning and evening sector, and, although less evident, the same happens for the southern polar region. This result is most likely caused by the high latitude current systems which are activated during disturbed magnetic periods. In fact, the lowest values of correlation are found mainly in the periods M2 and R1. During these periods the $a_p$ index, which is the magnetic index used by IRI, is not high ($a_p < 50$), especially when compared with that relative to the M1 period ($a_p > 150$). Nevertheless, the values of the *AE* index clearly show an intense global electrojet activity in the auroral zone during both the main and the recovery phase of the storm; this means that from 18 to 22 March 2015 the auroral regions are characterized by an intense substorm activity, with a consequent enhancement of the auroral electrojet systems. This may explain the considerable difference obtained between Swarm measurements and IRI modeled values not only at polar regions, but also between morning and evening sectors. Indeed, the spatial distribution of the polar ionospheric convection and current systems is not uniform at high latitude, showing a greater intensity in the evening sector than in the morning one. Moreover, during this particular geomagnetic event, also a difference of the current intensities could have characterized the two hemispheres as reported by Cherniak et al. (2015).

Results just shown represents an evidence that the topside ionosphere modeling, especially during magnetically disturbed periods, is still a challenge. In fact, Swarm electron density data considered in this study showed patterns that are, at the moment, difficult to model. Specifically, the analysis done, based on the Pearson correlation coefficient, showed that, independently of the chosen topside option (IRI-2001, IRI-2001corr, IRI-NeQuick), the trend embedded in the Swarm and IRI time series are fairly different. In particular, the analysis did not show an IRI topside option that behaves definitely better than the others.

With regard to the topside modeling, *in situ* measurements of the thin electron plasma density around the Earth carried out by the Swarm constellation can be extremely valuable. In fact, when having accurate and calibrated measurements, the peculiar configuration of Swarm satellites will allow to achieve new insights about the topside plasma scale height, a parameter of crucial significance to reliably model the topside profile.

## 6.2 Modeling the lower part of the topside ionospheric vertical electron density profile over the European region by means of Swarm satellites data and IRI UP method

Results described in the previous section highlight how difficult it is to model the topside ionosphere for both quiet and disturbed magnetic conditions. Anyway, electron density data collected by Langmuir probes on board Swarm satellites provide a valuable source of information in order to get a better characterization of the topside ionosphere.

To this regard, an empirical method to model, over the European region, the lower part of the ionospheric topside region, from the F2-layer peak height to about 500-600 km of altitude, has been developed. The method is based on calibrated electron density values recorded from December 2013 to June 2016 by Langmuir Probes on board Swarm satellites, and on *fo*F2 and *hm*F2 values provided by the IRI UP method. Differently from section 6.1, for this work the "Extended Set of Swarm Langmuir Probe Data" dataset (SW-RN-IRF-GS-005, Rev: 1, 2016-06-23), released by S. Buchert (Swedish Institute of Space Physics) on 23 June 2016, has been used. This dataset comprises calibrated plasma density measurements collected by the Swarm satellite constellation from 12 December 2013 to 11 June 2016. Topside effective scale heights are calculated by fitting some definite analytical functions (α-

Chapman, β-Chapman, Epstein and Exponential) through the values recorded by Swarm and the ones output by IRI UP, with the assumption that the effective scale height is constant in the altitude range considered. Calculated effective scale heights are then modeled as a function of *fo*F2 and *hm*F2, in order to be operationally applicable to both ionosonde measurements and ionospheric models, like IRI. The method produces two-dimensional maps of the median effective scale height binned as a function of *fo*F2 and *hm*F2, for each of the considered topside profiles. A statistical comparison with COSMIC/FORMOSAT-3 collected Radio Occultation profiles have been carried out to assess the validity of the proposed method, and to investigate which of the considered topside profiles is the best one.

### 6.2.1 The effective scale height of the topside ionosphere

The need to mathematically model the topside ionosphere led many scientists to apply several analytical functions to fit the sparse information in this region coming from topside sounders, ISRs, Langmuir probes in-situ data, and Radio Occultation data. The most used analytical functions are Chapman's (Chapman 1931), Epstein (Rawer 1988), Exponential, and Parabolic functions, or a linear combination of these (Kutiev and Marinov 2007).

All these formulations strongly rely on a parameter called *scale height* whose definition and calculation is the most difficult task when looking for the best topside formulation. In all the aforementioned functions, the scale height controls the shape of the topside profile, thus the vertical distribution of the electron density in the topside ionosphere; this is why the scale height is strongly connected with the ionospheric dynamics, plasma thermal structure, and composition (Liu et al. 2007a, 2007b; Tulasi Ram et al. 2009).

The definition of the scale height parameter depends on the analytical formulation used and on the type of data on which it is built. From a theoretical point of view, the *plasma scale height* $H_\mathrm{p}$ is defined as (Hargreaves 1992):

$$H_\mathrm{p} = \frac{k_\mathrm{b} T_\mathrm{p}}{m_\mathrm{i} g} \ , \tag{6.2}$$

where $k_\mathrm{b}$ is the Boltzmann constant, $T_\mathrm{p} = T_\mathrm{e} + T_\mathrm{i}$ is the plasma temperature ($T_\mathrm{e}$ and $T_\mathrm{i}$ are the electron and ion temperature, respectively), $m_\mathrm{i}$ is the ion mean molar mass, and $g$ is the acceleration due to gravity.

However, to apply this definition one should know the vertical distribution of plasma temperature and that of each ion constituent, which is difficult. This is why more practical approaches based directly on electron density measurements are instead used. With regard to this, the *effective scale height*, frequently called $H_\mathrm{m}$ in the literature (Liu et al. 2007a, 2007b), is the parameter that can be inferred by fitting some analytical functions to electron density values. Then, the effective scale height is a mere empirical parameter which is used to fit measured data with analytical functions, in order to obtain the most reliable representation of the topside vertical electron density distribution.

The effective scale height can be calculated in various ways, depending on the type of available ionospheric data. Bottomside electron density values retrieved by ionosondes can be used to estimate the topside ionospheric profile approximating the F2-layer region, and the overlying region, by an α-Chapman function with an effective scale height determined around *hm*F2 (Huang and Reinisch 2001; Reinisch et al. 2004).

Indeed, several empirical methods exploit topside ionosphere retrieved data. Some of these are based on topside sounders flown from sixties to eighties, such as Alouette-1 & -2, ISIS-1 & -2 and Intercosmos 19, that have provided sets of topside data, but with a limited temporal and spatial coverage (Kutiev et al. 2006; Kutiev and Marinov 2007; Nsumei et al. 2012; Wang et al. 2015b; Zhu et al. 2015). Others are based on Radio Occultation measurements made by LEO satellites such as CHAMP and COSMIC, now readily available with a good spatial and temporal global coverage (Stankov and Jakowski 2006; Wu et al. 2016; Olivares Pulido et al. 2016; Xu et al. 2013; Wang et al. 2015a). Effective scale heights are calculated also by using vTEC data in conjunction with bottomside ionosonde derived data to constrain the topside profile to meet some requirements on topside vTEC

values (Stankov et al. 2003; Belehaki et al. 2012). In-situ electron density measurements collected by Langmuir probes on board LEO satellites can also be used to model the topside profile constraining some analytical functions to pass through the electron density value measured by the satellite and that measured by an underlying ionosonde (Venkatesh et al. 2011; Tulasi Ram et al. 2009; Liu et al. 2014), or using electron density values measured by satellites flying at different heights (Triskova et al. 2006).

Liu et al. (2007a, 2007b), using Arecibo (66.7°W, 18.4°N) and Millstone Hill (71.5°W, 42.6°N) ISR measurements made a careful morphological and theoretical examination of differences between $H_p$ and $H_m$. From a theoretical point of view, Liu et al. (2007a, 2007b) pointed out that $H_p$ is expected to be twice $H_m$ only in a topside ionosphere mainly driven by a diffusive equilibrium and neglecting the altitude gradient of the thermospheric temperature. Analyzing Arecibo ISR data, they found that median ratios of $H_p$ and $H_m$ are between 3.2 and 3.6 by daytime and equal to 2.8 at night. Similar results were found also analyzing Millstone Hill ISR data. Liu et al. (2007a, 2007b) studies led to the important conclusions that: a) the dynamics of the topside ionosphere is dominated by the plasma diffusion; b) motions caused by the vertical thermal structure of this region, field-aligned fluxes, the ion-neutral drag, and neutral winds cannot, however, be neglected. Besides differences in the absolute values, plasma and effective scale heights exhibit different diurnal and seasonal behavior as well. It is thus important to carefully specify that the scale height considered in this study is the *effective scale height*.

### 6.2.2 Topside analytical formulations

Four different analytical functions are commonly used to model the topside ionospheric electron density profile as a function of both the effective scale height ($H_m$) and the F2-layer peak characteristics ($NmF2$ and $hmF2$). These are:

- $\alpha$-Chapman:

$$N(h) = NmF2 \cdot \exp\left\{\frac{1}{2}\left[1 - \frac{h - hmF2}{H_m} - \exp\left(-\frac{h - hmF2}{H_m}\right)\right]\right\} \; ; \tag{6.3}$$

- $\beta$-Chapman:

$$N(h) = NmF2 \cdot \exp\left\{\left[1 - \frac{h - hmF2}{H_m} - \exp\left(-\frac{h - hmF2}{H_m}\right)\right]\right\} \; ; \tag{6.4}$$

- Epstein:

$$N(h) = 4 \cdot NmF2 \frac{\exp\left(\frac{h - hmF2}{H_m}\right)}{\left[1 + \exp\left(\frac{h - hmF2}{H_m}\right)\right]^2} \; ; \tag{6.5}$$

- Exponential:

$$N(h) = NmF2 \cdot \exp\left(-\frac{h - hmF2}{H_m}\right). \tag{6.6}$$

Among these, the most used and studied is the α-Chapman with a fixed (Chapman 1931; Wright 1960) or variable (Vary-Chap) effective scale height (Rishbeth and Garriott 1969; Reinisch et al. 2007). The use of Chapman's functions to model the topside ionosphere was carefully investigated by Luan et al. (2006) studying which

Chapman's shape factor (keeping constant the effective scale height) could better fit ISR derived profiles at Arecibo (18.4°N, 66.7°W) and Millstone Hill (42.6°N, 71.5°W). They found that a value of the Chapman shape factor of about 0.5 (as it is in the α-Chapman function) gives a reasonable description of the topside ionosphere, but with some departures in the morning (where values of about 0.35-0.45 were identified) and during daytime (where values of about 0.55-0.75 were identified), showing also a seasonal variation. The Chapman's shape factor exhibits a high correlation with the F2-layer peak electron density, and shows a strong solar cycle dependence during the late morning hours. Zhang et al. (2002) tried to model the topside ionosphere by fitting the α-Chapman and Epstein functions, with a linearly variable effective scale height, to ISR profiles collected in Malvern (52.1°N, 2.3°W). They found that topside profiles can be fitted very well (about 94% of profiles met the criterion $\varepsilon<10\%$, ε being the percentage difference between modeled and measured topside electron density values) to a height of about 400 km above the F2-peak using either functions.

Despite being widely used, Chapman's functions are not based on any theoretical consideration, because they are derived according to the simplifying hypotheses of monochromatic solar irradiance, single ion component, and, more importantly, absence of any dynamics (Chapman 1931). Such hypotheses do not hold in the F2 region, where the dynamics of the ions is deeply influenced by both zonal and meridional neutral winds, and by the effect of the diffusion along the geomagnetic field lines, and do not hold even more so in the topside ionosphere. For these reasons, other analytical functions were used to model the topside ionosphere. Among these, Epstein and Exponential functions meet the constraints to pass through the F2-layer peak and to monotonically decrease in the topside ionosphere. All the analytical functions (6.3)-(6.6) are purely empirical, thus the need to find which of these can better describe the topside ionosphere.

Fonda et al. (2005) compared α-Chapman, β-Chapman, Epstein and modified Epstein (as used in the NeQuick topside model by Radicella and Leitinger (2001) and Nava et al. (2008)) profiles to those measured by ISIS-2 and IK19 topside sounders, finding that α-Chapman gives the best results.

Verhulst and Stankov (2014, 2015) compared α-Chapman, β-Chapman, Epstein, and Exponential topside profiles (calculating for each of these a different effective scale height) to topside sounders data, finding that in almost 75% of the analyzed cases the best fit was provided by the Exponential profile, followed by the α-Chapman profile. A local time, seasonal, latitudinal, longitudinal, solar and geomagnetic activity influence on the topside profile shape was also identified, but highlighting, at the same time, an objective difficulty in modeling the effective scale height dependence on these parameters. More obvious relations were instead identified between the profile shape and the characteristics associated to the F2-layer peak (basically *hm*F2 and *Nm*F2), whose variations strongly influence the topside profile.

### 6.2.3 Effective scale height calculation

By assuming a constant effective scale height $H_m$ for the first hundreds of kilometers from the F2-layer peak, one can model the topside above some selected ionosonde stations using an analytical formula joining the F2-layer peak characteristics *Nm*F2 and *hm*F2, measured by the ionosonde, to the in-situ electron density value ($N(h_{sat})$, $h_{sat}$) measured by Langmuir Probes on board Swarm. Effective scale heights are calculated by forcing the passage of the considered analytical functions through the two anchor points, one at the height of the F2-layer peak, and the other one at the height of the satellite.

Tulasi Ram et al. (2009), and Venkatesh et al. (2011) pursued this approach using simultaneous observations of ROCSAT-1 electron density values (collected around 600 km of altitude) and ionosonde or ISR deduced F2-layer characteristics. However, selecting only satellite's passages on a definite point (the one where a ground-based measuring station is installed) heavily reduces the available data, and consequently the possibility to do any spatial study on the topside effective scale height. Furthermore, ionosondes (or ISRs) positioning and the corresponding sounding repetition rate further reduces available data, thus limiting the spatial and temporal description of any modeled parameter.

Venkatesh et al. (2014) carried out a study to identify the altitude, above *hm*F2, at which to calculate the effective scale height giving the most reliable representation of the topside profile, when compared to that recorded by ISRs at Arecibo (66.7°W, 18.4°N) and Jicamarca (77.3°W, 11.9°S). To accomplish this task, they used an α-Chapman profile joining the F2-layer peak point with that above with a constant effective scale height. After calculating effective scale heights for different local times and seasons, they found that $H_m$ generally exhibits an increase with altitude for the first 50-100 km above the F2-layer peak height, and then settles on a fairly constant value for the following 200-250 km. Comparing calculated α-Chapman topside vertical electron density profiles (using effective scale heights calculated between the F2-layer peak height and a varying point above) and ISR profiles, they got the best results for $H_m$ values obtained using electron density values around 550 km for the Jicamarca equatorial station, and around 500 km for the Arecibo low-latitude station.

An ionospheric model able to spatially describe the ionospheric plasma, more specifically the F2-layer characteristics, would maximize the use of the data recorded by the Swarm constellation. This task can be accomplished, for the European region, by the IRI UP method, which can generate European maps of updated values of *fo*F2 and *hm*F2 (as described in chapter 4). In this way, the effective scale height can be calculated for each satellite's passage over Europe, averaging the satellite's measurements falling in each grid point of the map, with a 1°x1° spatial resolution. This procedure allows to take advantage of every satellite track over the European region and not only the ones co-located with a ground-based station. Carrying out this analysis for every transit of each satellite of the Swarm constellation over each grid point of the European region, a huge amount of effective scale height values can be obtained. This gives the possibility to perform a robust statistical and spatial characterization of this parameter.

Electron density data in the topside are those collected by the Swarm satellites constellation whose features have been already described in section 6.1.1. In particular, to calculate these effective scale height values, the "Extended Set of Swarm Langmuir Probe Data" dataset (SW-RN-IRF-GS-005, Rev: 1, 2016-06-23), released by S. Buchert (Swedish Institute of Space Physics) on 23 June 2016 (freely downloadable at https://earth.esa.int/web/guest/swarm/data-access after registration), is used. This dataset comprises calibrated plasma density measurements collected by the Swarm satellites constellation from 12 December 2013 to 11 June 2016. Specifically, Swarm's measurements collected over the European region (from 15°W to 45°E in longitude and from 30°N to 60°N in latitude) have been considered. In particular, 4259984 electron density measurements for Swarm A, 4229607 for Swarm B, and 4251095 for Swarm C have been used. From this dataset, unusable data were eliminated according to a flag embedded in the downloaded files, as recommended in "Extended Set of Swarm Langmuir Probe Data" dataset (SW-RN-IRF-GS-005, Rev: 1, 2016-06-23).

The near-polar orbit of Swarm satellites, the particular geometry of the constellation, and the height at which satellites fly, are particularly appropriate to study the topside ionosphere. As it is evident from Fig. 6.12, Swarm satellites provide a good spatial (both in longitude and latitude) and temporal (local time and seasonal) coverage of the European region.

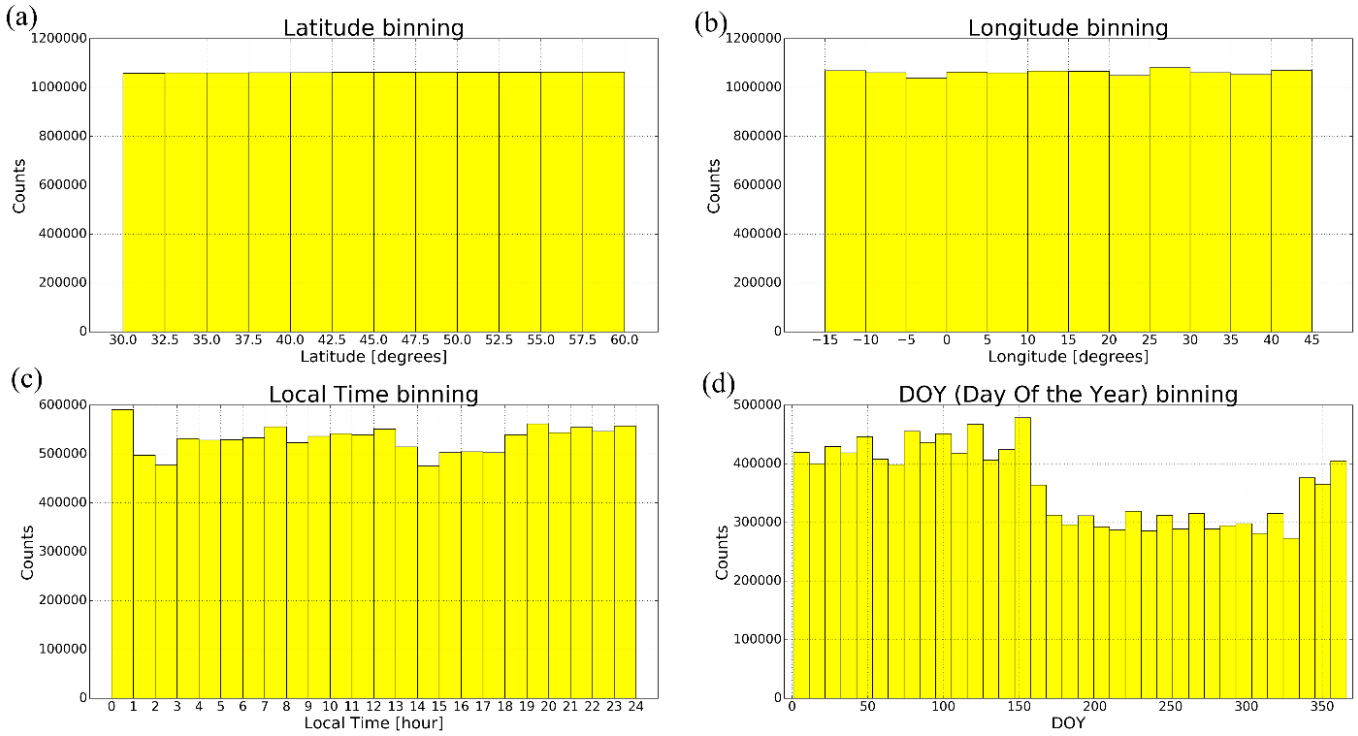


**Figure 6.12:** Swarm satellites' data distribution over the European region from 12 December 2013 to 11 June 2016 in terms of: (a) latitude, (b) longitude, (c) Local Time, and (d) Day Of the Year.

Maps of updated *fo*F2 and *hm*F2 values, over the European region, for each Swarm satellites' track, are produced by the IRI UP method by assimilating values of *fo*F2 and *M*(3000)F2 recorded at minutes 00 and 30 of each hour by the European stations listed in Table 4.1. In this way, the time difference between the satellite's passage and the assimilation of the ionosondes' measured characteristics is 15 minutes at most.

A visual example of the procedure used to calculate the effective scale height, for a particular moment for each topside profile used, is displayed in Fig. 6.13. Each topside profile is constrained to pass through two anchor points; the one modeled by the IRI UP method representing the F2 layer peak (which coordinates are (*fo*F2, *hm*F2), the red point in Fig. 6.13), and the one measured by a Swarm satellite in the topside (which coordinates are $(f(h_{\mathrm{sat}}), h_{\mathrm{sat}})$, the green point in Fig. 6.13). Fixing these two anchor points, the only free parameter in the topside profiles described by Eqs. (6.3)-(6.6) remain the effective scale height $H_{\mathrm{m}}$, which can be numerically deduced. It is worth noting that, despite having the same two anchor points, the topside profiles are different, especially in the region right above the F2-layer peak, which is the one the focus is on. This means that calculated effective scale heights are different. Most of the work is then devoted to the search of the function able to better represent the shape of the topside. With regard to this, Chapman and Epstein functions were introduced to better describe the shape of the F2-layer. The Exponential one, which is not able to describe the characteristic curvature of this layer, has instead the potential to better describe the upper part of the topside, where the ionosphere transitions to the plasmasphere (in which the vertical profile is essentially exponential with a scale height dependent on the $H^+$ions vertical distribution). In light of these considerations, the choice of which function has to be used to model the topside ionosphere is not trivial and deserves a close inspection.

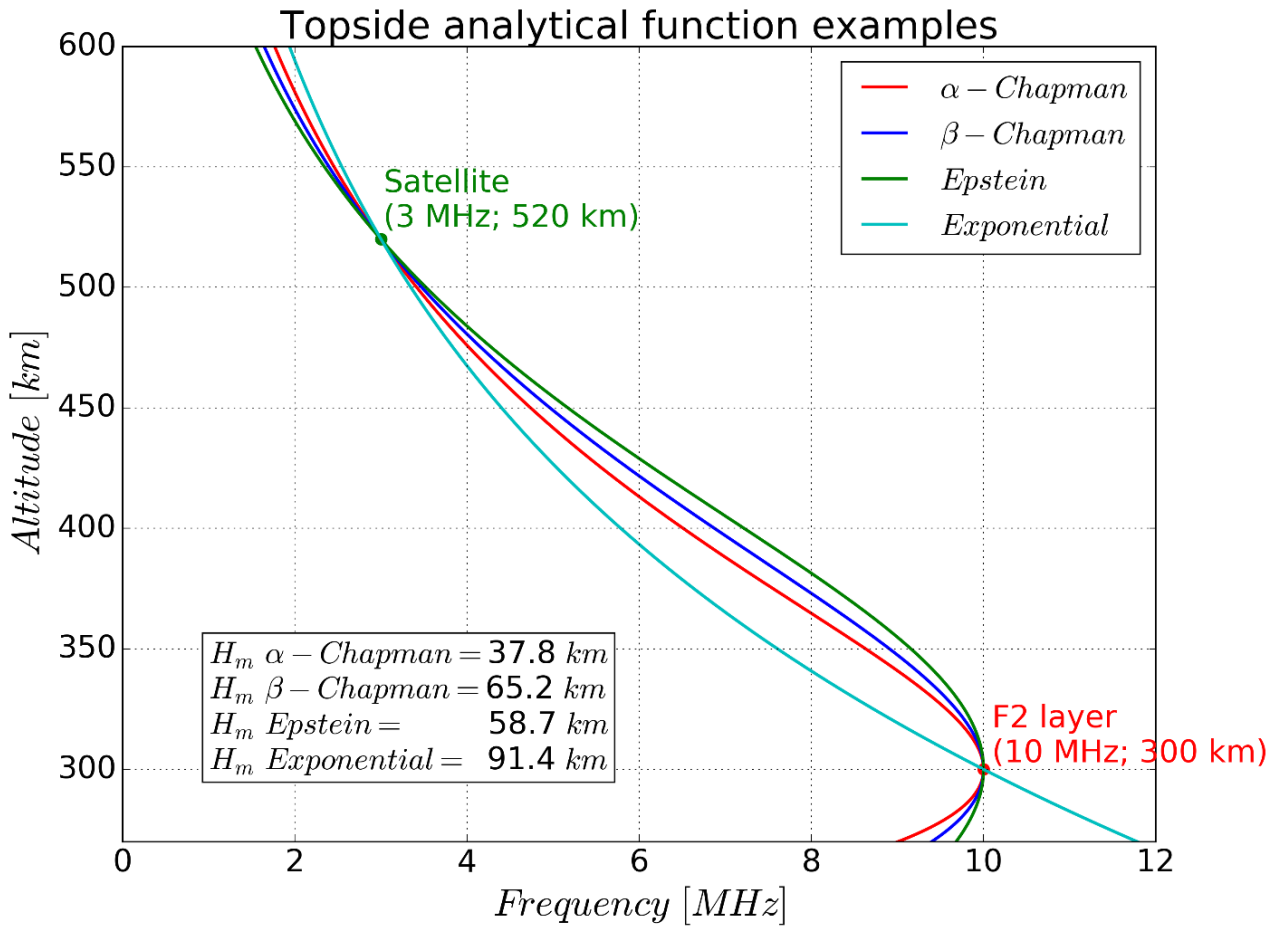


**Figure 6.13:** Examples of topside analytical functions (6.3)-(6.6) obtained after forcing them to meet the constrain to join the F2-layer peak point $(foF2, hmF2)$ to the satellite point $(f(h_{sat}), h_{sat})$. The corresponding calculated effective scale heights are all different.

### 6.2.4 Effective scale height modeling

Effective scale height values, calculated using the procedure described in section 6.2.3, need to be modeled as a function of some measured or modeled parameters. Many attempts were made in the past to model effective scale height values as a function of spatial (geocentric or magnetic longitude and latitude), temporal (local time, day of the year, season), and solar and magnetic activity parameters. All these approaches led to different results, critically depending on the chosen modeling parameters, and on the chosen modeling methods. Verhulst and Stankov (2014), studying which of the (6.3)-(6.6) topside profiles could better match Alouette and ISIS topside sounder profiles, using the Kutiev and Marinov (2007) topside scale height calculation method, came to the conclusion that:

- Local time, season, magnetic longitude/latitude, solar activity, and geomagnetic activity have a crucial influence on the topside profile shape. However, they say: "commonly used solar/magnetic activity indices do not provide much useful information and criteria to select a particular topside profile. The same holds true for local time, season, and magnetic latitude/longitude";

- Difficulties in modeling the effective scale height as a function of commonly used indices or geophysical parameters arise from the simultaneous influence that these drivers have on the topside ionosphere. Thus, it is difficult to discriminate the influence of each driver;
- More obvious relations exist between the topside profile shape and the F2-layer peak characteristics, especially between *hm*F2 and the effective scale height.

As it is easily understandable looking at topside analytical profiles of Eqs. (6.3)-(6.6), and taking into account the procedure used to derive effective scale height values described in section 6.2.3, calculated effective scale height values are function of four variables:

$$H_{\mathrm{m}} = H_{\mathrm{m}}\left(hm\mathrm{F2}, Nm\mathrm{F2}, h_{\mathrm{sat}}, N\left(h_{\mathrm{sat}}\right)\right) \ . \tag{6.7}$$

From an operational point of view, it would be better to describe $H_{\mathrm{m}}$ as a function of only *Nm*F2 and *hm*F2, which are measured by ionosondes or easily modeled by several ionospheric models, while satellite's related parameters are not routinely available. Therefore, following Verhulst and Stankov (2014), $H_{\mathrm{m}}$ values are modeled as a function of the sole F2-layer peak characteristics, *hm*F2 and *fo*F2, applying a two-dimensional binning of them with a bin width of 5 km and 0.25 MHz, for *hm*F2 and *fo*F2, respectively.

In this way any dependence on $h_{\mathrm{sat}}$ and $N(h_{\mathrm{sat}})$ is neglected:

$$H_{\mathrm{m}} = H_{\mathrm{m}}\left(hm\mathrm{F2}, fo\mathrm{F2}\right) \ . \tag{6.8}$$

The use of *fo*F2, directly related to *Nm*F2 by means of the relation:

$$fo\mathrm{F2}\left(\mathrm{MHz}\right) = \sqrt{Nm\mathrm{F2}\left(\mathrm{el}/\mathrm{cm}^3\right)/\left(1.24\cdot 10^4\right)} \ , \tag{6.9}$$

is done here only for convenience.

In this way, for each of the four proposed topside profiles, and for each Swarm's satellite derived dataset, a two-dimensional binning procedure is carried out, by selecting calculated $H_{\mathrm{m}}$ values derived from a defined pair (*fo*F2, *hm*F2) of binning indices. In order to obtain a two-dimensional map of $H_{\mathrm{m}}$, function of *fo*F2 and *hm*F2, the median of the $H_{\mathrm{m}}$ values falling in each bin is calculated.

Fig. 6.14 shows calculated $H_{\mathrm{m}}$ two-dimensional binning maps, after joining Swarm A and C datasets (this choice to join Swarm A and C datasets will be justified later), using each of the (6.3)-(6.6) topside profiles. The distribution of the number of values falling in each bin is also shown in Fig. 6.14a; for statistical robustness, a value of $H_{\mathrm{m}}$ is calculated only when this number is greater than 10. The minimum number of values in a bin is 0, represented by the white color, while the maximum number of values in a bin is 735, corresponding to the dark red color. Overall, there are 204086 values to be binned in 60x64=3840 bins. Fig. 6.14a shows also that binning indices ranges are $fo\mathrm{F2} \in [0{,}16]$ MHz and $hm\mathrm{F2} \in [150{,}450]$ km, and that the most filled bins (the orange/red colored ones) are in the range $[2{,}11]$ MHz for *fo*F2, and $[225{,}350]$ km for *hm*F2. Obviously, these are the bins for which the highest confidence level, in a statistical sense, is achieved. Fig. 6.14a shows also the presence of two well distinguished patterns highlighted by the dark red colored sectors, one in the sector between 2 and 4 MHz for *fo*F2 and between 300 and 350 km for *hm*F2, the other one in the sector between 5 and 9 MHz for *fo*F2 and between 250 and 300 km for *hm*F2. These two patterns describe, approximately, the nighttime and daytime behavior of the pair (*fo*F2, *hm*F2), respectively. This means that, for most of local time hours, daytime and nighttime behaviors do not overlap each other, allowing a reliable description of the corresponding daytime and nighttime behavior of the topside ionosphere.

Figs. 6.14b,c,d,e, depict the two-dimensional binning maps of the effective scale height median values, for each of the considered topside analytical function. The bin distribution reflects what is shown in Fig. 6.14a, but the bins with a number of $H_{\mathrm{m}}$ values lower than 10 are not displayed. For each topside profile, highest $H_{\mathrm{m}}$ median values are obtained for low *fo*F2 values, and then they decrease for higher *fo*F2 values. Extremely high $H_{\mathrm{m}}$ median values,

obtained mainly for very low *fo*F2 values (dark red bins), could be due to an unsatisfactory statistical description in those bins. Generally speaking, α-Chapman derived $H_m$ median values are characterized by the lowest values. β-Chapman and Epstein calculated $H_m$ median values show slightly higher values compared to α-Chapman ones. Note that the scale for the Exponential $H_m$ median values is double compared to the other three functions. This significant difference between the Exponential profile and the other ones can be ascribed to the very different behavior of the Exponential profile immediately above the F2-layer peak.

Once maps of the median effective scale height as a function of *fo*F2 and *hm*F2 are obtained (like those shown in Fig. 6.14), one can use them to model the topside profile. In fact, once a pair of values (*fo*F2, *hm*F2) is available, either measured or modeled, a value of the corresponding effective scale height can be extrapolated from the map of each topside function.

In addition to maps of effective scale height median values, also maps of effective scale height first and third quartile values (not shown here) have been calculated in order to characterize the variability of this parameter, and to provide an estimation of the error associated to a defined effective scale height value. In this way, topside vertical electron density profiles derived using effective scale height median values can be complemented with lower and upper profiles calculated by using effective scale height first and third quartile values.

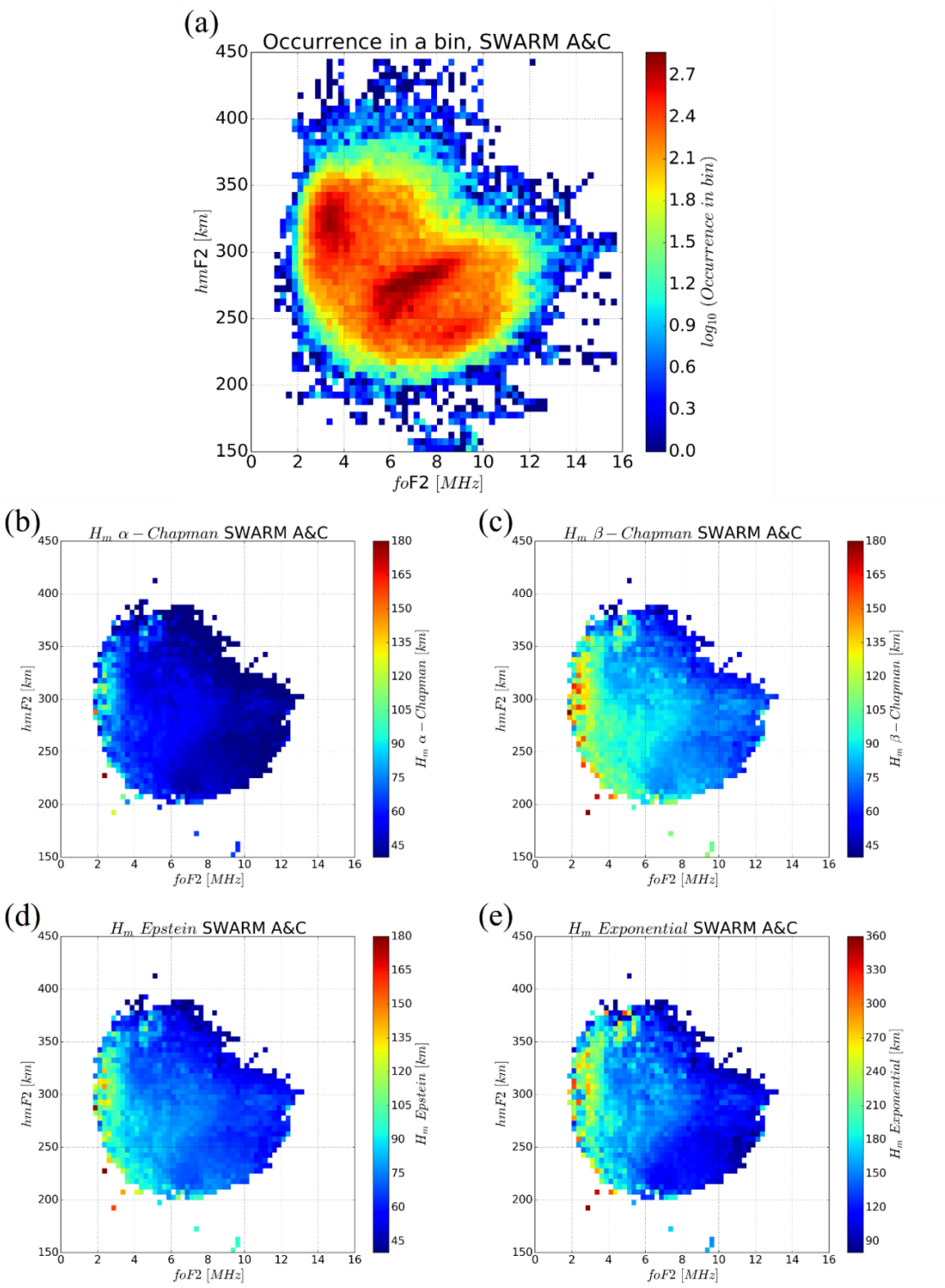


**Figure 6.14:** (a) Effective scale height values occurrence in a bin (logarithmic scale). Median values of the effective scale height are shown for (b) the α-Chapman topside profile, (c) the β-Chapman topside profile, (d) the Epstein topside profile, and (e) the Exponential topside profile. It is worth noting that the scale of the Exponential effective scale height is doubled. For each plot the joint dataset Swarm A&C is considered. In panels (b), (c), (d), and (e), bins including a number of $H_m$ values lower than 10 (the blue/dark blue colored ones in the panel (a)) have been discarded.

### 6.2.5 Effective scale height dependence on the satellite altitude

Neglecting the dependence of the effective scale height on satellite's parameters ($h_{\mathrm{sat}}, N(h_{\mathrm{sat}})$) implies an approximation; this is because one is trying to describe a four-dimensional mathematical function while utilizing only two of its four independent variables. Since the three Swarm satellites fly at different height (around 460 km for Swarm A and C, and around 520 km for Swarm B), a careful analysis of the effective scale heights calculated using electron density data derived from different satellites have to be carried out. For this reason, effective scale height median values are calculated using, separately, each of the measurement datasets provided by Swarm satellites, and differences, bin by bin, are calculated for each of the possible permutations.

An example of this analysis is visible in Fig. 6.15, for the α-Chapman topside profile. This figure shows that, except for some bins in sectors statistically not so significant, differences between Swarm A and C are negligible (Fig. 6.15f). This is the reason why Swarm A and C datasets are merged to derive the effective scale heights shown in Fig. 6.14. The same does not hold true for Swarm B derived effective scale heights, which are different from those derived from Swarm A and C (Fig. 6.15d,e). Differences of the order of about 10-20 km, for *fo*F2 values lower than 6 MHz, and of about 0-10 km, for *fo*F2 values higher than 6 MHz, are appreciable. Thus, effective scale height values derived from Swarm B cannot be merged with those calculated by using Swarm A and C.

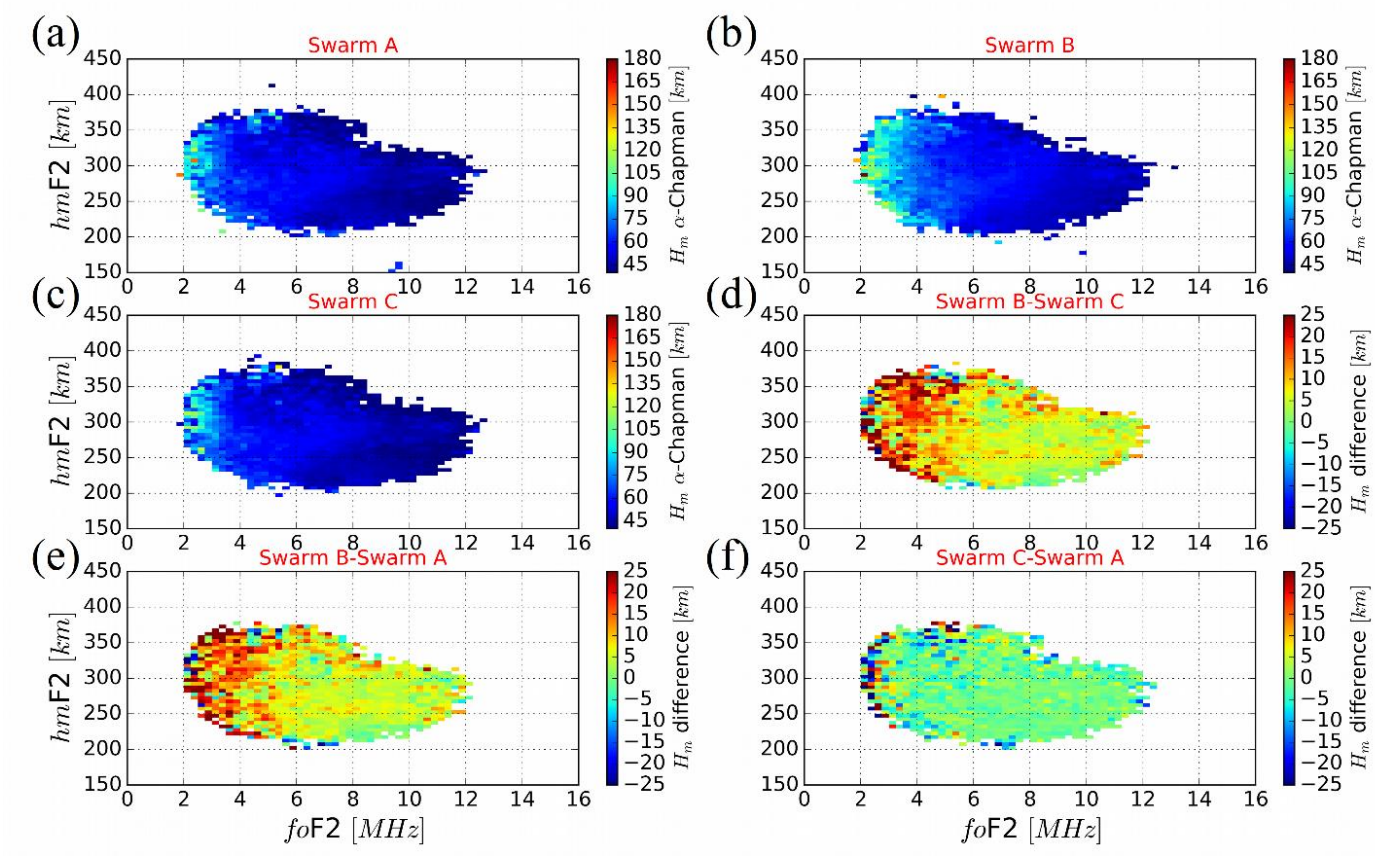


**Figure 6.15:** Median effective scale height values for the α-Chapman topside profile, for (a) the Swarm A dataset, (b) the Swarm B dataset, (c) the Swarm C dataset. Corresponding differences are shown in the rest of panels (d, e, and f).

Results shown in Fig. 6.15 seem to contradict the work hypothesis behind the whole analysis, namely, the constancy of the topside effective scale height for the lower topside ionospheric region. In reality, this should not

be a surprise because it is only a consequence of the fact that the dependence of the effective scale height on satellite's parameters has been deliberately neglected.

The work hypothesis can be verified by using normalized variables in the description of the effective scale height, that is:

$$n = \frac{N\left(h_{\text{sat}}\right)}{N\text{m}\text{F2}}, \qquad z = h_{\text{sat}} - h\text{m}\text{F2}\ . \tag{6.10}$$

Using (6.10) the topside analytical formulas become:

- Normalized $\alpha$-Chapman:

$$n(z) = \exp\left\{\frac{1}{2}\left[1 - \frac{z}{H_{\text{m}}} - \exp\left(-\frac{z}{H_{\text{m}}}\right)\right]\right\}\ ; \tag{6.11}$$

- Normalized $\beta$-Chapman:

$$n(z) = \exp\left\{\left[1 - \frac{z}{H_{\text{m}}} - \exp\left(-\frac{z}{H_{\text{m}}}\right)\right]\right\}\ ; \tag{6.12}$$

- Normalized Epstein:

$$n(z) = 4\frac{\exp\left(\dfrac{z}{H_{\text{m}}}\right)}{\left[1 + \exp\left(\dfrac{z}{H_{\text{m}}}\right)\right]^2}\ ; \tag{6.13}$$

- Normalized Exponential:

$$n(z) = \exp\left(-\frac{z}{H_{\text{m}}}\right)\ . \tag{6.14}$$

In this way, calculated effective scale heights depend only on two normalized variables:

$$H_{\text{m}} = H_{\text{m}}\left(n, z\right)\ . \tag{6.15}$$

Using (6.15) to make the same calculations done to obtain Fig. 6.14, results shown in Fig. 6.16 are obtained.

This figure shows that, independently of which Swarm satellite is used, the difference between effective scale height median values is close to zero (the scale of the last three panels in Fig. 6.16 is reduced by one fifth compared to Fig. 6.14), which confirms the validity of the effective scale height constancy hypothesis, for the studied ionospheric region.

The use of normalized variables, even though it looks attractive, is not feasible from an operational point of view because of the lack of available real-time electron density values recorded by LEO satellites. Thus, the need to use the approach described before, relying on only the F2-layer peak characteristics, separating the dataset formed by Swarm A&C data from the one related to Swarm B.

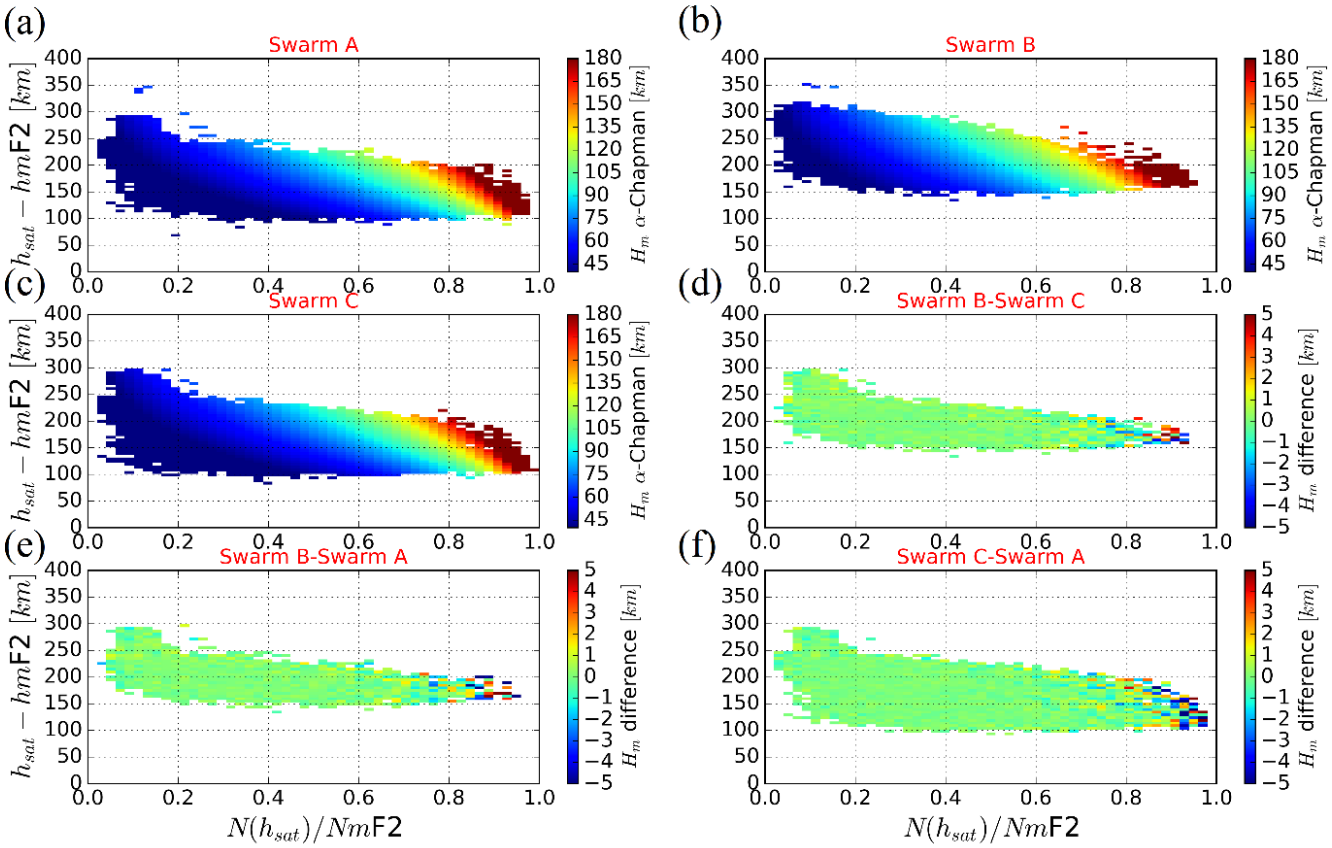


**Figure 6.16:** Same as Fig. 6.14 but using normalized variables. Compared to Fig. 6.14, the scale of the last three panels is reduced by one fifth.

### 6.2.6 Topside analytical formulation statistical assessment

With the procedure described before, two-dimensional binning maps of median effective scale heights for each of the four proposed topside profile, and for each of the two satellite's derived datasets related to Swarm A&C and Swarm B, respectively, are calculated. The number of calculated binned maps is then eight.

To investigate the performance of the proposed method in modeling the lower ionospheric topside profile, and to asses which of the four proposed topside profiles better represents this region, a careful statistical analysis is carried out using an independent dataset of topside profiles. Specifically, Radio Occultation derived electron density profiles collected by Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC), also known as FORMOSAT-3 in Taiwan.

COSMIC is a constellation made by six microsatellites launched on 15 April 2006 into a circular orbit (with 72° of inclination) at about 800 km of height (gradually reached 17 months after the launch) and a separation angle of 30° in longitude between neighboring satellites (Anthes et al. 2008). The mission is a collaborative project between the National SPace Organization (NSPO) in Taiwan and the University Corporation for Atmospheric Research (UCAR) in the United States. Each satellite carries a GPS radio occultation receiver able to measure the phase delay of radio waves from GPS satellites as they are occulted by the Earth's atmosphere, allowing an accurate determination of the ionospheric vertical electron density profile. It is worth noting that the accuracy and the precision of the Radio Occultation data critically depend on the algorithm used in the inversion procedure to obtain a vertical electron density profile. In particular, COSMIC Radio Occultation profiles are obtained by means of the Abel inversion technique (Haji and Romans 1998; Schreiner et al. 1999). Because of the assumed spherical

symmetry in the Abel inversion procedure, some errors arise above all in the Equatorial Ionization Anomaly region, during dawn and dusk hours, and during intense magnetically disturbed events (Garcia-Fernandez et al. 2003; Yue et al. 2010). This is due to the strong horizontal electron density gradients characterizing these ionospheric states, a condition that the spherical symmetry assumption, embedded in the Abel inversion technique, cannot take into account.

COSMIC Radio Occultation data are used as truth reference because of their reliability in the description of the topside ionosphere and the underlying F-region (Krankowski et al. 2011; Yue et al. 2010; Lei et al. 2007; Wu et al. 2009; Habarulema et al. 2014; Chu et al. 2010). Comparing COSMIC retrieved *Nm*F2 and *hm*F2 F2-layer peak characteristics with those measured by co-located mid-latitude ionosondes in the European region for the year 2008, Krankowski et al. (2011) found a good agreement in terms of both absolute value and degree of correlation, thus highlighting the good reliability of COSMIC data over the European region.

Radio occultation derived electron density profiles collected by COSMIC satellites, for the same period covered by Swarm's data (from 12 December 2013 to 11 June 2016), and whose tangent points of signal ray path is over the European region (from 15°W to 45°E in longitude, and from 30°N to 60°N in latitude), are used to validate the method and to decide which of the four proposed topside analytical formulations performs better. The dataset consists of 17872 "ionprf" Radio Occultatio profiles, downloaded through the COSMIC Data Analysis and Archive Center (CDAAC, http://cdaac-www.cosmic.ucar.edu/cdaac/products.html). Radio Occultation profiles showing *fo*F2 and *hm*F2 values outside the range [1,16] MHz and [150,450] km, respectively, are discarded, as well as profiles with excessive and unrealistic fluctuations in the electron density. After applying this filtering procedure, the remaining profiles are 9672 (about 54% of the total).

For each of the considered 9672 COSMIC profiles, the Root Mean Square Error Eq. (6.16) and the Normalized Root Mean Square Error Eq. (6.17), between modeled topside electron density values and those measured by COSMIC (both expressed as plasma frequency, $f_p$, obtained replacing *Nm*F2 with $N_e$ in Eq. (6.9)), are calculated for the height range from *hm*F2 to the height of Swarm's satellites:

$$\mathrm{RMSE}\left[\mathrm{MHz}\right]=\sqrt{\frac{\sum_{i=1}^{N}\left(f_{\mathrm{p}_{\mathrm{model},i}}-f_{\mathrm{p}_{\mathrm{COSMIC},i}}\right)^{2}}{N}}\,, \tag{6.16}$$

$$\mathrm{NRMSE}\left[\%\right]=\frac{\mathrm{RMSE}}{f_{\mathrm{p}_{\mathrm{COSMIC}}}}\cdot 100\,. \tag{6.17}$$

Modeled topside profiles are calculated using the four profiles (6.3)-(6.6) and considering: a) *hm*F2 and *fo*F2 values measured by COSMIC; b) the effective scale height value, correspondent to values in a), given by the two-dimensional binning map (as those shown in Fig. 6.14) related to the definite topside profile. An example of the carried out analysis, for one of the analysed COSMIC RO profile, is shown in Fig. 6.17.

The use of RO topside profiles leads, unavoidably, to some approximations because these are slanted than vertical. *hm*F2 and *fo*F2 values measured by RO method have a different spatial location compared to other points of the profile, while the effective scale height calculated with the described method refers to the whole topside vertical electron density profile. To get an idea of the co-location errors affecting this comparison, the geographical difference (in degrees) between the F2-layer peak and the topside point at 460 km (the height of Swarm A and C), and 520 km (the height of Swarm B), are calculated for each of the considered 9672 RO profiles. The corresponding mean and standard deviation are:

- Mean co-location errors, Swarm A and C = 2.4°;
- Mean co-location errors, Swarm B = 3.3°;
- Standard Deviation co-location errors, Swarm A and C = 1.6°;
- Standard Deviation co-location errors, Swarm B = 2.2°.

The mean co-location error is of the order of 2°-3°, with a dispersion of about 2°. Because of the known correlation distance of the ionosphere at mid latitudes, which is equal to 1000-1500 km for quiet geomagnetic conditions (Klobuchar and Kunches 2000; Yue et al. 2007), the co-location error is not considered so important as to invalidate the analysis, but a reduced performance of the proposed method can be expected when using RO profiles as truth reference.

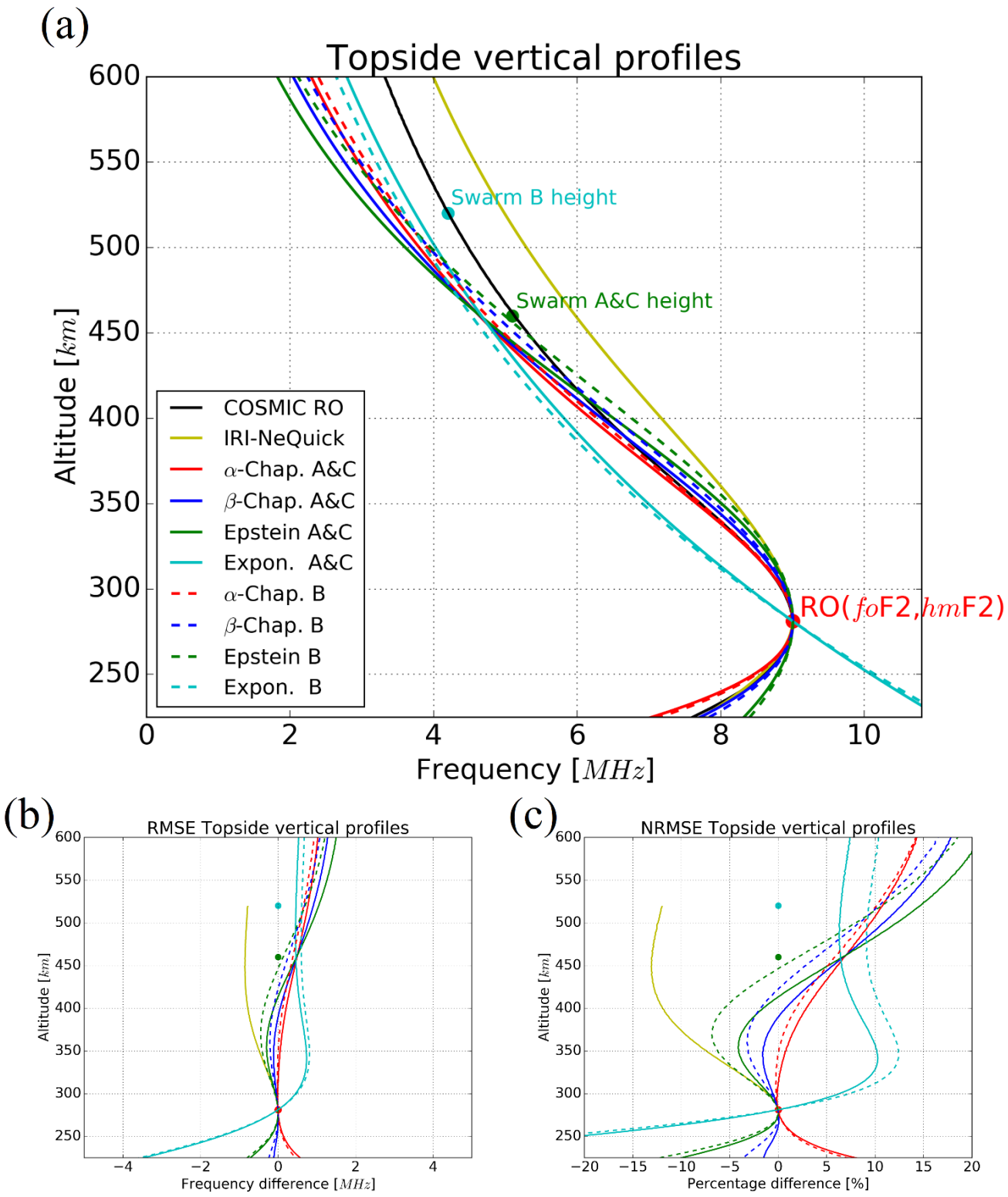


**Figure 6.17: (**a) Topside electron density profile measured by COSMIC RO (solid line in black), and those modeled by using the Swarm A&C dataset (red, blue, green, and light-blue solid lines), and by using the Swarm B dataset (red, blue, green, and light-blue dashed lines). The solid dark yellow line depicts the topside profile as modeled by the IRI model (using the NeQuick topside option), after forcing the model to pass through the measured RO derived F2-layer peak point. (b) RMSE values. (c) NRMSE values.

A statistical summary of the analysis is shown in Table 6.2, where the mean and the standard deviation of both RMSE and NRMSE of all the 9672 analyzed topside profiles, are reported for each of the four studied topside profiles, and also for the NeQuick topside option of the IRI model (Coisson et al. 2006), for both considered Swarm datasets. The NeQuick topside option of IRI has been chosen because, when compared to the other two possible options proposed by IRI, it turns out to be the best one (Bilitza 2009). Moreover, as demonstrated in section 6.1, the IRI topside description performs better for mid latitudes.

Table 6.2 points out that the α-Chapman topside profile is the best one compared to both the other studied topside profiles and the IRI-NeQuick topside model, by using both Swarm A&C and Swarm B derived effective scale heights. With regard to the Swarm A&C dataset, all the topside profiles (6.3)-(6.6), with the exception of the Exponential one, provide a better accuracy than IRI-NeQuick. Instead, for the Swarm B dataset, this holds true only for α-Chapman and Exponential profiles. Moreover, all the topside profiles, with the exception of the Exponential one, present worse performances for the Swarm B dataset than for the Swarm A&C one.

| | **SWARM A&C** | | | | | **SWARM B** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **α-Chapman** | **β-Chapman** | **Epstein** | **Exponential** | **IRI-NeQuick** | **α-Chapman** | **β-Chapman** | **Epstein** | **Exponential** | **IRI-NeQuick** |
| | **RMSE [MHz] Dataset (2014/121 – 2016/163)** | | | | | | | | | |
| **Mean** | 0.30 | 0.31 | 0.33 | 0.42 | 0.34 | 0.34 | 0.38 | 0.44 | 0.37 | 0.37 |
| **Standard Deviation** | 0.22 | 0.22 | 0.23 | 0.31 | 0.35 | 0.20 | 0.22 | 0.25 | 0.23 | 0.34 |
| | **NRMSE [%] Dataset (2014/121 – 2016/163)** | | | | | | | | | |
| **Mean** | 6.14 | 6.44 | 7.14 | 7.94 | 7.00 | 8.69 | 9.59 | 11.04 | 8.39 | 8.31 |
| **Standard Deviation** | 4.32 | 4.57 | 5.07 | 4.08 | 6.67 | 6.29 | 6.57 | 7.10 | 4.31 | 7.12 |

**Table 6.2:** Statistical summary of the analysis made by comparing modeled topside profiles with those measured by COSMIC satellites.

These results highlight that:

a) α-Chapman, β-Chapman, and Epstein profiles can properly model the topside region immediately above the F2-layer peak, allowing a reliable description of the F2-layer shape;
b) in the upper topside region, by using a constant effective scale height for the entire topside region, α-Chapman, β-Chapman, and Epstein formulations no longer represent reliably its particular shape;
c) the Exponential profile is the best one to model the upper part of the ionospheric topside region.

Table 6.2 shows also that each of the proposed topside profiles is characterized by a standard deviation lower than that associated to IRI-NeQuick, thus highlighting a higher precision of the proposed topside modeling method.

## Summary

In this chapter, after having verified the IRI's difficulty in modeling the topside part of the vertical electron density profile (section 6.1), an empirical method to model the lower part of the ionospheric topside region from the F2-layer peak height to about 500-600 km of altitude over the European region has been proposed (section 6.2).

The method is based on electron density values recorded by Swarm satellites, and on *fo*F2 and *hm*F2 values provided by IRI UP, relying on the assimilation of *fo*F2 and *M*(3000)F2 data routinely recorded by a network of European ionosonde stations (as described in chapter 4). Topside effective scale heights have been calculated by fitting some definite analytical functions (α-Chapman, β-Chapman, Epstein, and Exponential) through values recorded by Swarm satellites and the ones output by IRI UP, with the assumption that the effective scale height is constant in the considered range of altitude. Calculated effective scale heights have been then modeled as a function of *fo*F2 and *hm*F2. The method produces two-dimensional maps of the median effective scale height binned as a function of *fo*F2 and *hm*F2, for each of the considered topside profiles.

From the statistical comparison with COSMIC/FORMOSAT-3 collected Radio Occultation profiles emerges that: the α-Chapman topside function displays the best performance compared to the others, and also when compared to the NeQuick topside option of IRI. Furthermore, all the proposed topside profiles are characterized by a standard deviation lower than that associated to IRI-NeQuick. This highlights a good precision of the proposed topside modeling method which then represents a valid Space Weather operational tool to reliably model the topside over the European region once the F2-layer characteristics, namely *fo*F2 and *hm*F2, are known, either measured or modeled.

Swarm electron density measurements have the potentiality to be used to model the topside ionospheric region because of the particular geometry of their orbit. In addition, the IRI UP method turns out to be particularly suited to this task, being able to spatially describe the F2-layer peak characteristics, in regions where a good number of ionosonde stations are available. Thus, the proposed method can be used, with profit, in other regions besides Europe.

# 7. IonoPy: a Python package for modeling the ionosphere

In this chapter the IonoPy model is presented. The name IonoPy derives from the contraction of the words *Ionosphere* and *Python* and highlights the main purpose of the model, namely the modeling of the ionosphere, and the programming language used for its implementation, namely Python (https://www.python.org/).

The IonoPy model as a first step embeds the IRI UP method (the one assimilating directly ionosonde data, described in chapters 3 and 4) for updating the climatological background IRI model; then, as a second step, it updates the topside ionosphere description by means of the topside model described in chapter 6.

However, the ionosondes are able to sound the whole bottomside ionosphere (chapter 1) and provide the bottomside vertical electron density profile by inverting the ordinary mode of propagation recorded in the ionogram. It is to exploit such information that a procedure for assimilating the whole bottomside electron density profile recorded by ionosondes has been implemented. From now on, one assumes that before applying such procedure, the IRI UP method and the topside modeling have already been applied, therefore one simply refers to it as the IonoPy model.

In section 7.1, the IonoPy procedure for assimilating the whole bottomside electron density profile measured by ionosondes is first mathematically described and then its application for a very disturbed period is shown.

In section 7.2, a comparison between the IonoPy model and the ISP model, another model which assimilates the whole bottomside vertical electron density profile developed by Pezzopane et al. (2011, 2013), for the Mediterranean region, during some intense and severe geomagnetic storms, is provided.

Results described in section 7.1 and 7.2 are published in:

*Pietrella, M., Pignalberi, A., Pezzopane, M., Pignatelli, A., Azzarone, A., Rizzi, R. (2018) A comparative study of ionospheric IRIEup and ISP assimilative models during some intense and severe geomagnetic storms. Advances in Space Research, 61(10), 2569-2584. https://doi.org/10.1016/j.asr.2018.02.026.*

## 7.1 The IonoPy algorithm: on the assimilation of the whole bottomside electron density profile from ionosondes

### 7.1.1 Description of the mathematical procedure to assimilate the whole bottomside profile recorded by a ionosonde

The IonoPy model as a first step is characterized by the application of the IRI UP method, through which measurements of *fo*F2 and *M*(3000)F2 carried out at each assimilated station are assimilated, to calculate effective values of $IG_{12}$, and $R_{12}$, namely $IG_{12eff}$ and $R_{12eff}$ (as described in chapter 3).

From the scattered values of $IG_{12eff}$ and $R_{12eff}$, corresponding maps are generated by means of the Universal Kriging geostatistical interpolation method (chapter 3). Then, these maps are given as input to the IRI model (chapter 4), obtaining a three-dimensional updated representation of the electron density identified as $B_{ijk}$, where: $i = 1,\ldots,nlon$ is the index of longitudes, being $nlon$ the grid point number running on the $x$-axis; $j = 1,\ldots,nlat$ is the index of latitudes, being $nlat$ the grid point number running on the $y$-axis; $k = 1,\ldots,nheight$ is the index of heights, being $nheight$ the grid point number running on the $z$-axis.

The matrix $B_{ijk}$ consists of $nlon \cdot nlat$, that is $i \cdot j$, one-dimensional updated IRI vertical electron density profiles $b_k$ .

To better describe the assimilation made by the IonoPy algorithm of the whole bottomside electron density profile from ionosondes, the following mathematical notation will be used: the geometric dimension of each mathematical entity used in the algorithm is described by the number of subscript indices to the left of the comma, while the mathematical product of the number of elements to the right of the comma represents the number of these entities. For example, with this notation, the $nlon \cdot nlat$ one-dimensional vertical electron density profiles $b_k$ can be expressed as $b_{k,ij}$ ; the single index $k$ to the left of the comma means that the entity $b$ is a one-dimensional vector whose values depend on the height, and the product of the number of elements of the two indices to the right of the comma $i \cdot j$ means that there are $nlon \cdot nlat$ vectors $b_k$ .

$B_{ijk}$ is instead a three-dimensional matrix function of longitude, latitude and height. $B_{ijk}$ can be also considered as $k$ bi-dimensional matrices $B_{ij}$, thus, with the aforementioned notation, it can be written as $B_{ij,k}$. The $B_{ij,k}$ matrix is the background reference state.

Let us assume to have *nstation* vertical radio-sounding ionospheric stations, described by the index $l = 1,\ldots,nstation$, whose longitude and latitude are respectively $X_{,l}$ and $Y_{,l}$ (there are *nstation* scalar values of longitude and latitude, and consequently there is no index to the left of comma). Vertical electron density profiles provided by the assimilation ionospheric stations, that will be used in the assimilation process, are identified as $I_{k,l}$.

The idea behind the assimilation algorithm is to spread the information of each assimilated vertical profile on every grid point $(i, j, k)$, with a weight inversely proportional to the distance between the assimilated profile location and the grid point. To accomplish this task, several *Radial Basis Functions* can be used as weighting function. The choice fell on a bi-dimensional height independent spherical Gaussian function.

It is worth noting that in a spherical surface the length of one degree in latitude has the same value all over the surface, while the length of one degree in longitude depends on the considered latitude. As a consequence, to take into account the Earth's spherical symmetry, one would need to calculate the length of the maximum circle between every grid point and the assimilated station. This calculation is however time-consuming and it should be done for every height (the maximum circle length over a sphere depends on the sphere's radius); this is why it has

been chosen to redefine the sigma parameter of the spherical Gaussian function, by including in it the latitude dependence.

According to these considerations, the considered Gaussian radial basis weighting function is:

$$g_{ij,l} = \exp\left\{ -\frac{\sigma^2 \left(x_i - X_{,l}\right)^2 + \left[\frac{\sigma}{\cos\left(Y_{,l}\right)}\right]^2 \left(y_j - Y_{,l}\right)^2}{2\sigma^2 \left[\frac{\sigma}{\cos\left(Y_{,l}\right)}\right]^2} \right\}, \tag{7.1}$$

where:

- $x_i \in \left[lonini, lonfin\right]$ is the longitude array indexed by $i$, with bounds $lonini$ (the initial longitude) and $lonfin$ (the final longitude), and length $nlon$, dependent on the chosen size step;
- $y_j \in \left[latini, latfin\right]$ is the latitude array indexed by $j$, with bounds $latini$ (the initial latitude) and $latfin$ (the final latitude), and length $nlat$, dependent on the chosen size step;
- $\sigma$ is the parameter of the Gaussian function.

With this formulation, a two-dimensional map (on the chosen grid) of the Gaussian function (the same for every height), is obtained for each assimilated station. Each Gaussian function has a maximum equal to 1 at the station point $\left(X_{,l}, Y_{,l}\right)$, and it decreases getting away from it, according to the value of $\sigma$.

Since all the assimilated profiles must be simultaneously taken into account in the assimilation procedure, while maintaining over each station electron density values as similar as possible to the assimilated ones, the weights provided by each Gaussian function defined in Eq. (7.1) are properly normalized according to the following expression:

$$G_{ij} = \sum_{l=1}^{nstation} \left( \frac{g_{ij,l}}{\sum_{l=1}^{nstation} g_{ij,l}} \right). \tag{7.2}$$

After that, a weighted mean ( $A_{ij,k}$ ) of the assimilated electron density values ( $I_{k,l}$ ) is performed multiplying them by the associated Gaussian map defined in Eq. (7.2), for each grid point $\left(i, j\right)$ and for every height *k*, according to the following expression:

$$A_{ij,k} = \sum_{l=1}^{nstation} \left( \frac{g_{ij,l}}{\sum_{l=1}^{nstation} g_{ij,l}} \right) \cdot I_{l,k} = G_{ij} \cdot I_{k,l} \ . \tag{7.3}$$

In Eq. (7.3) it has been written $I_{l,k}$ (instead of $I_{k,l}$) because the summation is carried out on the station index $l$ for every *k* height. Being this operation made for every height, a $k$ index appears to the right of comma in $A$.

Concurrently, also the weight $C_{ij,k}$ of the background is calculated as follows:

$$C_{ij,k} = \sum_{l=1}^{nstation} \left( \mathbf{1}_{ij} - \frac{g_{ij,l}}{\sum_{l=1}^{nstation} g_{ij,l}} \right) \cdot b_{l,k} = \left( \mathbf{1}_{ij} - G_{ij} \right) \cdot b_{k,l} \,, \tag{7.4}$$

where $\mathbf{1}_{ij}$ is a two-dimensional matrix made up of one, and $b_{l,k}$ are $k$ (one for each height) one-dimensional vectors made up of background electron density values at the assimilated station grid points.

In this way, both the assimilated profiles ($I_{k,l}$) and background values ($b_{k,l}$) are weighted by the same function ($G_{ij}$) defined in Eq. (7.2). Finally, to obtain $k$ modeled two-dimensional electron density matrices ($M_{ij,k}$), the information obtained after assimilating vertical electron density profiles (now spread on all grid points) is merged with that given by the background reference state, as follows:

$$M_{ij,k} = A_{ij,k} + C_{ij,k} = G_{ij} \cdot I_{k,l} + \left( \mathbf{1}_{ij} - G_{ij} \right) \cdot b_{k,l} = B_{ij,k} + G_{ij} \cdot \left( I_{k,l} - b_{k,l} \right) , \tag{7.5}$$

where $B_{ij,k} = \sum_{l=1}^{nstation} \mathbf{1}_{ij} \cdot b_{l,k}$ .

Eq. (7.5) tells us that the IonoPy model output matrix $M_{ij,k}$ consists of the summation of two parts: the background one ($B_{ij,k}$) and the difference between assimilated ($I_{k,l}$) and background ($b_{k,l}$) values (at assimilated station locations), known as *Innovation*, weighted and spatially spread by the so called *Gain Matrix* $G_{ij}$. In this way, where $G_{ij}$ tends to zero (far enough from the assimilated stations, depending on the $\sigma$ value), the output smoothly tends to the background one, while near the assimilated stations the background is corrected, depending on the degree of innovation brought by the assimilated values.

### 7.1.2 An example of application of the bottomside profile assimilation procedure

To better understand how the IonoPy assimilation procedure works, the algorithm mathematically described in section 7.1.1 is here applied and graphically illustrated for the 17th March 2015 at 08 UT, the main phase day of the St. Patrick storm occurred in 2015, so a very disturbed moment of time. The vertical resolution of the model has been fixed to 1 km, while the horizontal resolution to 1°x1°.

Fig. 7.1 represents schematically Eq. (7.5) highlighting with different colors the different terms at the base of the IonoPy algorithm.

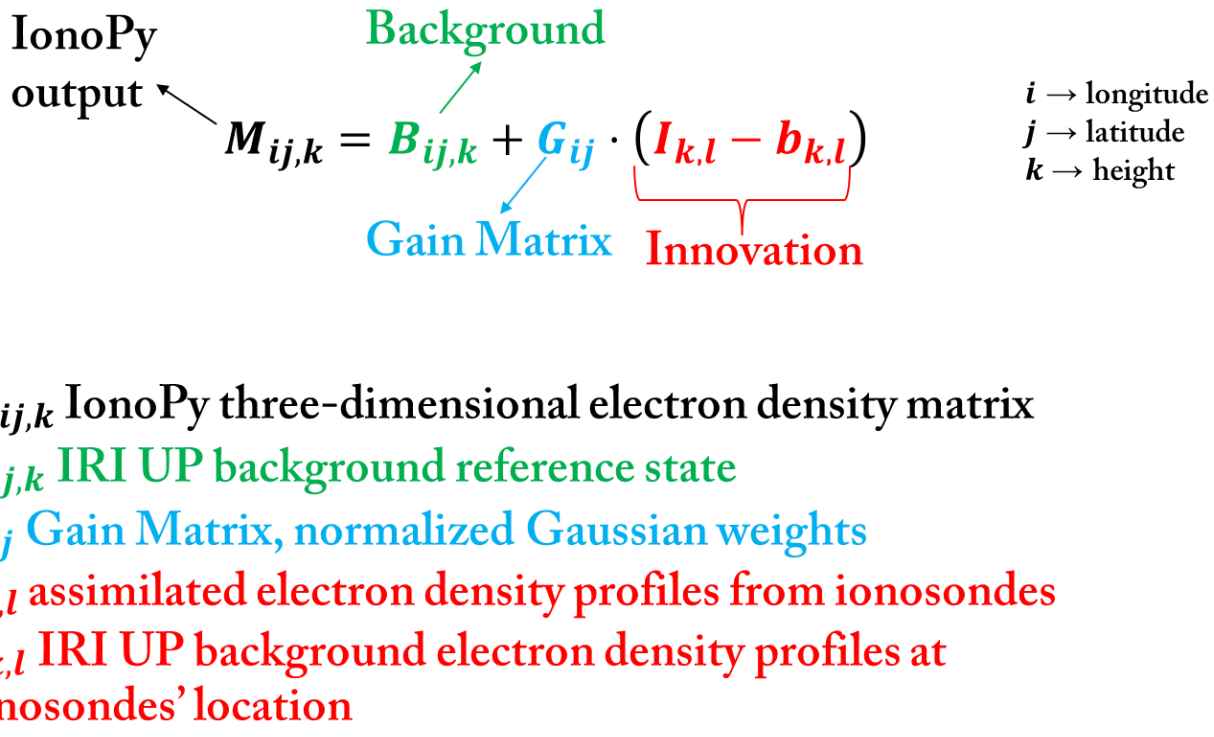


**Figure 7.1:** Schematic representation of Eq. (7.5). The different terms of the IonoPy assimilation algorithm are coloured differently.

For what concerns the background reference state matrix $B_{ij,k}$, this is provided by the IRI UP method that has been extensively described in chapters 3 and 4. Consequently, IRI UP method provides also the background electron density profiles at assimilated ionosondes' location $b_{k,l}$, in the Innovation term.

The electron density profiles $I_{k,l}$ measured by ionosondes at the available assimilated ionospheric stations are downloaded from the DIDBASE (Digital Ionogram DataBASE) (Reinisch and Galkin 2011) through the SAO Explorer software, developed by the University of Massachusetts, Lowell. Assimilated stations and their properties have been already described in section 4.1. From the downloaded XML files (one for each station available at the moment of the assimilation) the vertical electron density profile is extracted. The inversion of the vertical electron density profile is performed by either ARTIST or Autoscala software, depending on the assimilation station considered (see Table 4.1 for reference). Only the bottomside part of the profile is considered in this procedure, the one directly inverted from the ordinary mode of propagation shown in the ionogram. In particular, values of the electron density from 90 km of height (the bottom limit of IonoPy model in this case) to *hm*F2 are considered. An example of assimilated vertical electron density profile recorded at Rome (17$^{th}$ March 2015 at 12 UT) is represented in Fig. 7.2. For each of the *l* available ionospheric station, for a definite moment of time, the inverted vertical electron density profile (as that shown in Fig. 7.2) is considered in the assimilation process.

The difference between the vertical electron density profile $I_{k,l}$ measured by the ionosonde and that $b_{k,l}$ modeled by the background model (IRI UP in this case), at each of the *l* assimilated ionosonde locations, forms the Innovation term of Fig. 7.1.

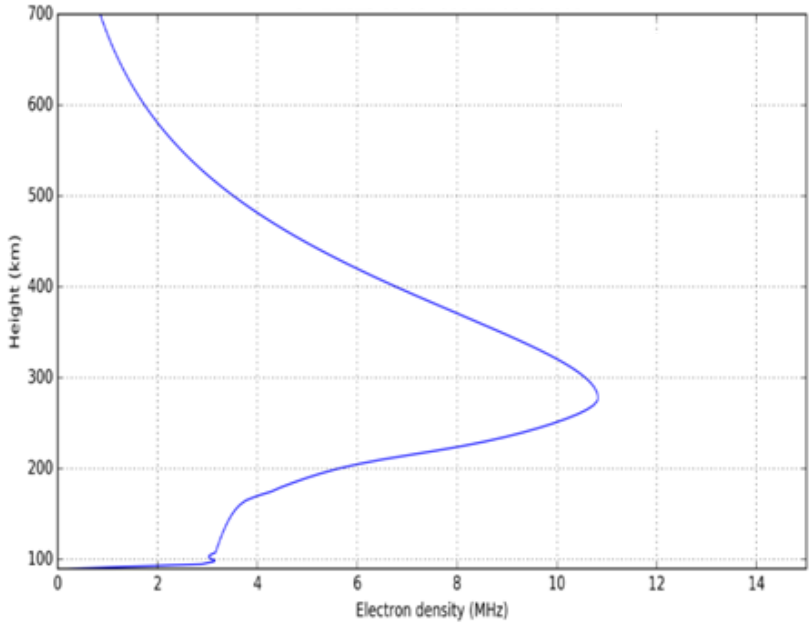


**Figure 7.2**: Example of assimilated vertical electron density profile $I_{k,l}$ measured from the ionosonde at Rome (one of the *l* available ionosonde). It is worth noting that only bottomside electron density values are considered in the assimilation process.

The Innovation term of Eq. (7.5) and Fig. 7.1 contains the information about how much the assimilated measurements update the description of the ionospheric plasma made by the background reference model, at each of the assimilated stations. The information carried by the Innovation term has to be spread among each grid point of a slice *k* of the three-dimensional matrix (in fact, as described in section 7.1.1, each three-dimensional matrix is considered as *k* two-dimensional slices). This operation is carried out by the Gain Matrix $G_{ij}$ which is a normalized spatial mean (Eq. (7.2)) of the $g_{ij,l}$ Gaussian radial basis weighting functions calculated over each assimilated station (Eq. (7.1)). In Fig. 7.3 the $g_{ij,l}$ function is three-dimensionally represented for the Rome station (lat = 41.8° N = $Y_{,l}$ , lon = 12.5° E = $X_{,l}$ ), with $\sigma = 3°$ , for each of the $(x_i, y_j)$ grid point of the plane *k* (because the $g_{ij,l}$ function is height independent, this is the same for each of the *k* heights). As it is clear by looking at Fig. 7.3, the Gaussian weighting function is centered on Rome location ( $X_{,l}$ = 12.5° E, $Y_{,l}$ = 41.8° N) where its value is 1 and it decreases getting away from it according to the value of $\sigma$. With $\sigma = 3°$ , the grid points located within about $2\sigma = 6°$ are the most affected by the Gaussian function, while $g_{ij,l} \simeq 0$ for points at a distance greater than $3\sigma = 9°$ . Obviously, the distance from the assimilated station at which $g_{ij,l} \simeq 0$ depends on the choice of $\sigma$. The $g_{ij,l}$ function, for different choices of the $\sigma$ parameter, is represented in Fig. 7.4. Looking at Fig.7.4 it is clear how the choice of $\sigma$ is of paramount importance to describe the spatial spreading of the information embedded in the Innovation term and assigne to it the right weight for each grid point. The greater is $\sigma$, the higher is the weight assigned to the Innovation term (compared to the weight of the background) at grid points surrounding the assimilated station, and wider is the influence of the assimilated measurements.

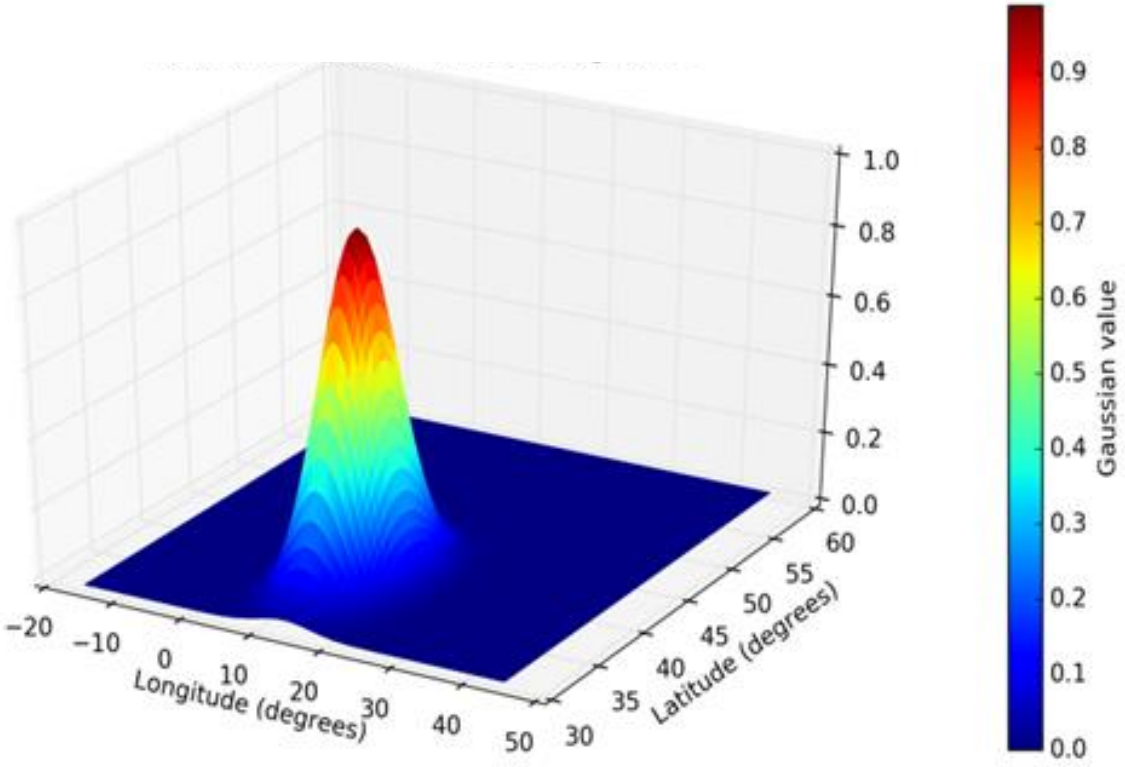


**Figure 7.3**: Three-dimensional representation of the Gaussian Radial Basis Function $g_{ij,l}$, as described by Eq. (7.1), with $\sigma = 3°$, centered on Rome location.

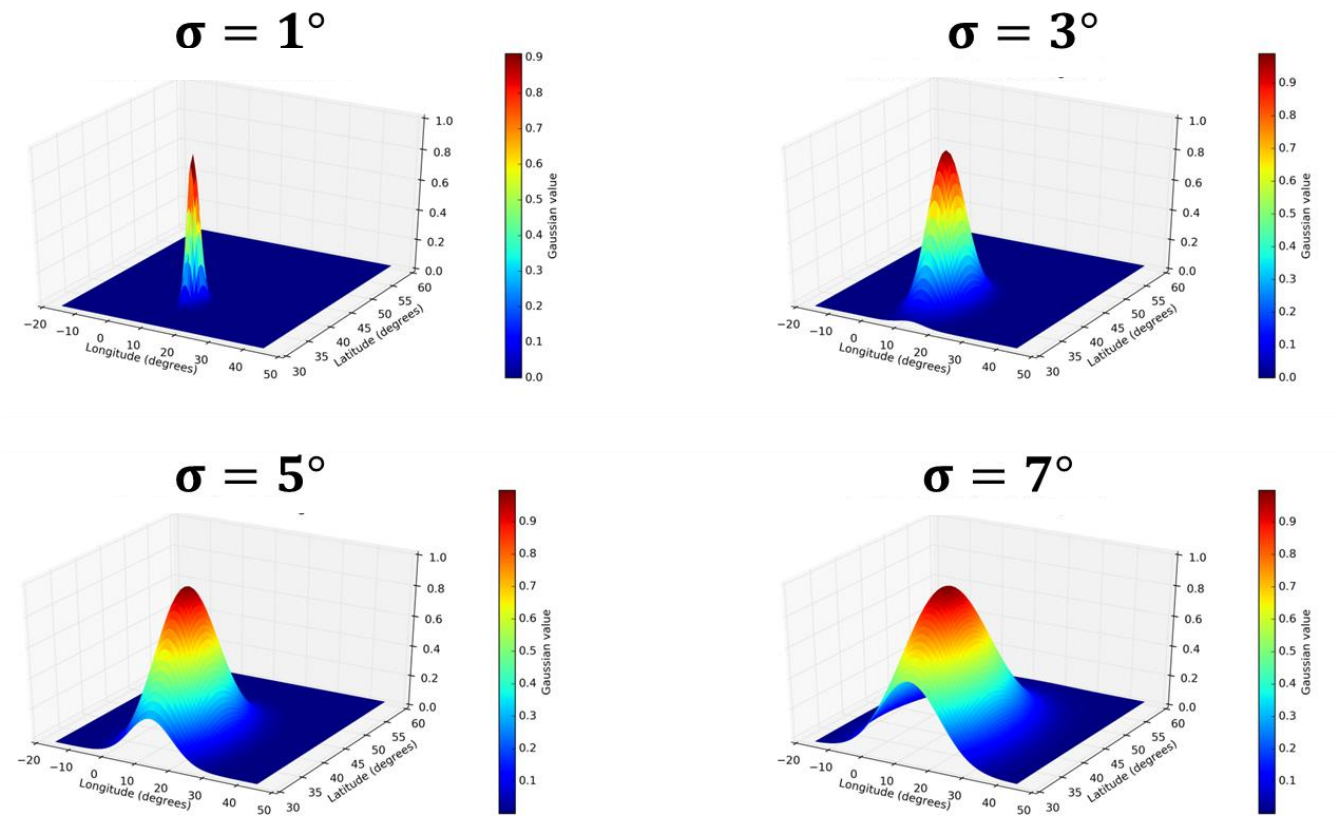


**Figure 7.4**: Three-dimensional representation of the Gaussian Radial Basis Function $g_{ij,l}$, centered on Rome location, for different values of the $\sigma$ parameter.

For each of the $l$ available assimilated station, the $g_{ij,l}$ function is calculated and then these are normalized to obtain the Gain Matrix $G_{ij}$ (Eq. (7.2)). This operation is necessary to assign at each grid point the right weight due to every $g_{ij,l}$. In this way, each assimilated station has some influence on the grid point depending on the distance between the stations, and on the $\sigma$ parameter choice. Then, such weights are normalized as shown by Eq. (7.2), so that their sum at the $(i,j)$ grid point is equal to the unity. In Fig. 7.5 the term $\frac{g_{ij,l}}{\sum_{l=1}^{nstation} g_{ij,l}}$ of Eq. (7.2) is three-dimensionally represented, with $\sigma = 3°$. Obviously, with $\sigma = 3°$ and with assimilated stations not too close, for a grid point located exactly at the assimilated station location the weight deriving from the co-located assimilation station is very close to 1 while the influence of the other assimilated stations is very small. Instead, a grid point located midway between more assimilated stations experiences the influence of more stations. In every case, the sum of the weights due to each $g_{ij,l}$, at a particular grid point, is always equal to the unity.

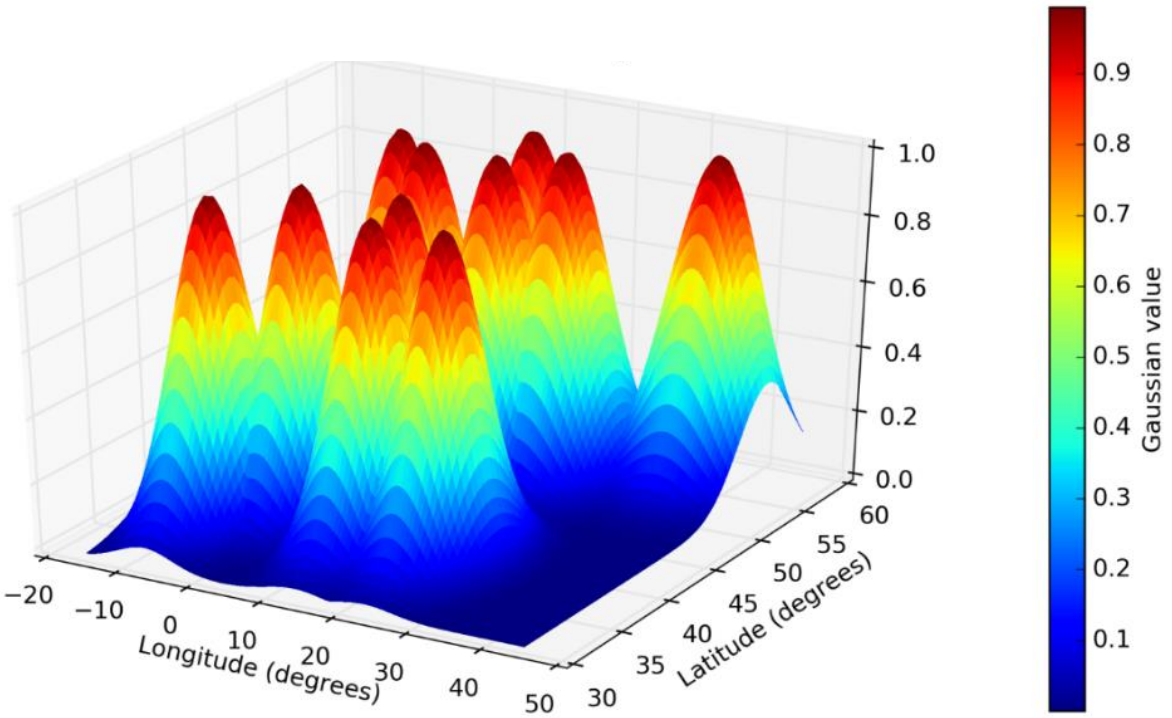


**Figure 7.5**: Three-dimensional representation of the term $\frac{g_{ij,l}}{\sum_{l=1}^{nstation} g_{ij,l}}$ of Eq. (7.2), with $\sigma = 3°$, by considering a number of assimilated stations equal to 11.

After describing each of the terms of Eq. (7.5), the IonoPy model is applied for the 17th March 2015 at 08 UT, the main phase day of the St. Patrick storm occurred in 2015. In particular, the electron density values at 250 km of height (then, for a particular value of the *k* index, in the bottomside ionosphere) are shown for each of the terms of Fig. 7.1. The same procedure holds true for different values of the height (so, for different *k* index values), building as a consequence the $M_{ij,k}$ three-dimensional IonoPy model output matrix.

In Fig. 7.6 the electron density at 250 km of height as modeled by the IRI UP method, through the assimilation of available *fo*F2 and *M*(3000)F2 measured values, and used as background $B_{ij,k}$, is shown. This represents already a very important update of the underlying IRI climatological model, as demonstrated in chapter 4.

Figs. 7.7 and 7.8 show separately the effect of the Innovation term weighted by the Gain Matrix. In particular, Fig.7.7 shows the effect of the term $B_{ij,k} - G_{ij,k} \cdot b_{k,l}$, while Fig. 7.8 shows the effect of the term $G_{ij,k} \cdot I_{k,l}$, of the Eq. (7.5). Looking at Fig. 7.8, it appears clear how the term $-G_{ij,k} \cdot b_{k,l}$ has the effect to subtract from the background $B_{ij,k}$ the contribution of the background at the assimilated station weighted by the Gain Matrix $G_{ij,k}$. This operation has the effect to produce "holes" in the electron density above the assimilated stations, well defined for very isolated stations like Moscow. Such holes tend to merge each other for relatively close stations like for example those of Rome, Gibilmanna, and San Vito in Italy. The holes shown in Fig. 7.7 are "filled" by the assimilated electron density values $I_{k,l}$ weighted by the Gain Matrix $G_{ij,k}$, as shown in Fig. 7.8.

Finally, the sum of Fig. 7.7 and 7.8 gives the IonoPy modeled electron density values depicted in Fig. 7.9, where the effect of the assimilated electron density values is particularly visible around Moscow, Warsaw, and Athens stations.

$$M_{ij,k} = B_{ij,k} + G_{ij} \cdot (I_{k,l} - b_{k,l})$$

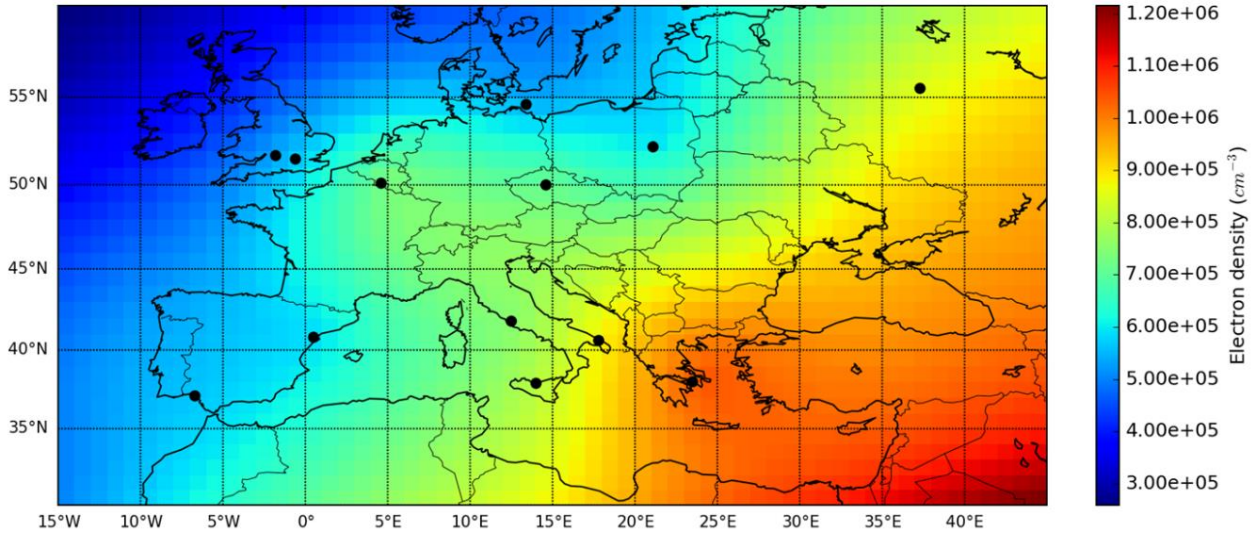


**Figure 7.6**: Two-dimensional representation (at a height of 250 km) of the electron density as modeled by the IRI UP method for the 17th March 2015 at 08 UT, used as background in the IonoPy model. Black dots represent the available assimilated stations. Above the map is reported Eq. (7.5) with highlighted in red the term to which the figure refers.

$$M_{ij,k} = B_{ij,k} + G_{ij} \cdot (I_{k,l} - b_{k,l})$$

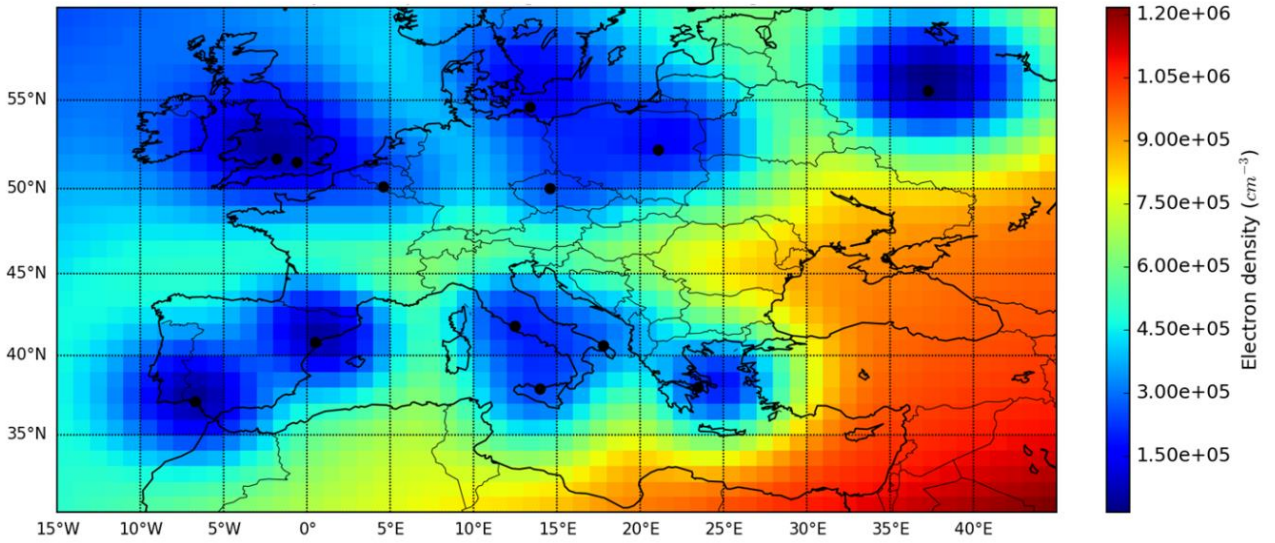


**Figure 7.7**: Same as Fig. 7.6 but for terms highlighted in red in the equation right above the map.

$$M_{ij,k} = B_{ij,k} + G_{ij} \cdot (I_{k,l} - b_{k,l})$$

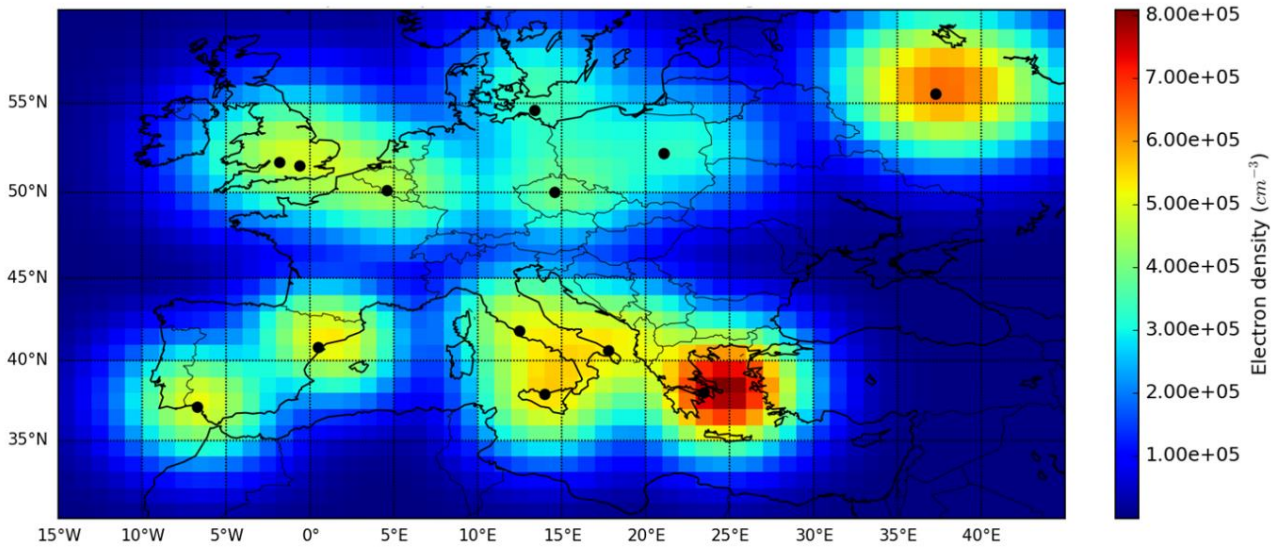


**Figure 7.8**: Same as Fig. 7.6 but for terms highlighted in red in the equation right above the map.

$$M_{ij,k} = B_{ij,k} + G_{ij} \cdot (I_{k,l} - b_{k,l})$$

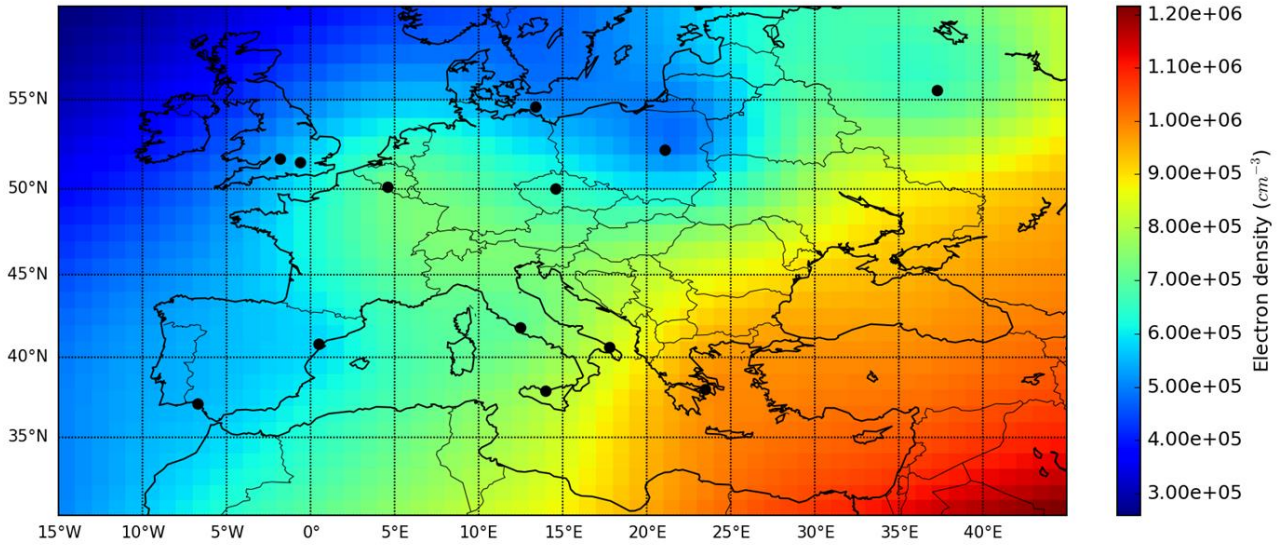


**Figure 7.9**: The sum of maps represented in Figs. 7.7 and 7.8, that is the matrix $M_{ij,k}$ highlighted in red in the equation right above the map. This map represents the electron density at a 250 km of height for the 17th March 2015 at 08UT produced by the IonoPy model.

## 7.2 Validation of the IonoPy model against the ISP model for the Mediterranean region during some intense and severe geomagnetic storms

In this section the behavior of the IonoPy model is compared to that of IRI and ISP models for some very disturbed epochs included in the geomagnetic storms occurred in March and December 2015, and October 2016.

The ISP (IRI-SIRMUP-P, Pezzopane et al. 2011, 2013) model is a regional model, developed at the Istituto Nazionale di Geofisica e Vulcanologia (INGV), able to provide a 3-D real time imaging of the ionosphere over the Mediterranean area, extending in latitude from 30°N to 44°N and in longitude from 7°W to 39°E. It is based on the ingestion of *fo*F2 and *M*(3000)F2 values, and the assimilation of the observed electron density profiles from some reference stations. At first, after assimilating *fo*F2 and *M*(3000)F2 values, it outputs a three-dimensional updated matrix of the electron density generated combining the IRI model with the Simplified Ionospheric Regional Model Updated (SIRMUP) (Zolesi et al. 2004; Tsagouri et al. 2005) and, subsequently, an assimilation process of measured electron density profiles from the reference stations further updates the IRI-SIRMUP electron density profiles, thus generating an ISP three-dimensional matrix (Pezzopane et al. 2011, 2013).

In principle, as the comparison between IonoPy and ISP models must be coherent as far as possible, they must operate under the same conditions. This means that they have to assimilate validated measured electron density profiles from the same stations. For these reasons, IonoPy is applied over the same ISP model validity area, i.e., over a limited area which, in this case, is the Mediterranean one; this operational mode is referred as IonoPy_re(V) ("re" stands for "restricted area", and (V) means that before the assimilation the autoscaled ionograms were validated). Nevertheless, the behaviour of the IonoPy model is also investigated under two other different operative

modes, namely assimilating electron density profiles from a larger region including several stations spread across Europe (as described in chapter 4, see Table 4.1): (a) without taking care of validating the autoscaled data before the assimilation process; (b) validating carefully the autoscaled data before their assimilation. These two additional operational modes are referred as IonoPy(NV) ((NV) stands for "not validated") and IonoPy(V) ((V) stands for "validated"), respectively. Moreover, the IRI model, 2016 version, is also considered, to investigate whether the IonoPy and ISP assimilation techniques can output better ionospheric predictions than IRI ones.

Each of these three models output a numerical grid representing the three-dimensional electron density distribution of the ionosphere that can be used to calculate synthesized ionograms by a ray tracing program. In this case, the three-dimensional IRI, IonoPy_re(V), ISP, IonoPy(NV), and IonoPy(V) electron density distributions are ingested by the software tool named IONORT (IONOspheric Ray Tracing), developed at the INGV (Azzarone et al. 2012), to synthesize quasi-vertical ionograms at the test sites. The climatological IRI model is run with the STORM option "ON" in order to model at best the intense and severe geomagnetic-ionospheric storms considered in this investigation.

The IRI, IonoPy_re(V), ISP, IonoPy(NV), and IonoPy(V) *fo*F2 synthesized values are then compared with the corresponding *fo*F2 measurements autoscaled at the test sites, by the ARTIST system, to evaluate the goodness of models.

### 7.2.1 Data and periods used in this study

The geographic area considered in this study is essentially the Mediterranean one (extending in latitude from 30°N to 44°N and in longitude from 7°W to 39°E), that is the validity area of the ISP model. The ionospheric working stations included in this region are: El Arenosillo (37.10° N, 6.70°W, Spain), Roquetes (40.80°N, 0.50°E, Spain), Rome (41.80°N, 12.50°E, Italy), Gibilmanna (37.90°N, 14.00°E, Italy), San Vito (40.60°N, 17.80°E, Italy), Athens (38.00°N, 23.50°E, Greece), and Nicosia (35.03°N, 33.16°E, Cyprus).

To investigate the reliability of IRI, IonoPy_re(V), ISP, IonoPy(NV), and IonoPy(V) models during disturbed ionospheric conditions, some recent geomagnetic storms occurred in March and December 2015, and October 2016, have been selected. Specifically, the periods 17-22 March 2015, 19-22 December 2015, and 13-14 October 2016, have been selected to be focused on periods for which the $D_{st}$ geomagnetic index is below -30 nT. As it can be inferred from Fig. 7.10, these periods basically include the main phase of the geomagnetic storm and the first part of its recovery phase. The geomagnetic storm of 17-22 March 2015 is classified "severe", being the minimum value of $D_{st}$ equal to -223 nT, while geomagnetic storms of 19-22 December 2015, and 13-14 October 2016, are classified "intense", since corresponding minimum values of $D_{st}$ are equal to -155 and -104 nT, respectively.

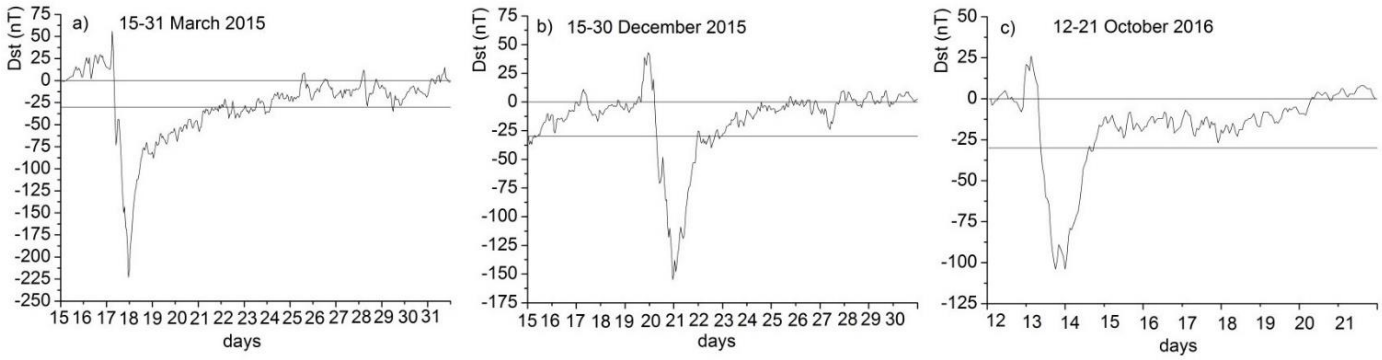


**Figure 7.10**: $D_{st}$ index recorded in the time intervals: a) 15 - 31 March 2015; b) 15-30 December 2015; c) 12-21 October 2016. The horizontal lines mark the levels 0 and -30 nT.

To ensure that chosen periods are representative of a strongly disturbed ionosphere, *fo*F2 hourly monthly median values ($foF2_{med}$) calculated by the SIRM model (Zolesi et al. 1993), are compared to the *fo*F2 hourly autoscaled measurements given by ARTIST ($foF2_{meas}$) for each station included in the Mediterranean area. The reliability of *fo*F2 hourly autoscaled values was checked by the visual inspection of the corresponding ionograms (downloadable at: http://ulcar.uml.edu/DIDBase/). As an example, Fig. 7.11 shows the *fo*F2 monthly median trend predicted by SIRM and the autoscaled one for the period 17-22 March 2015 at Roquetes, San Vito, Athens, and Nicosia.

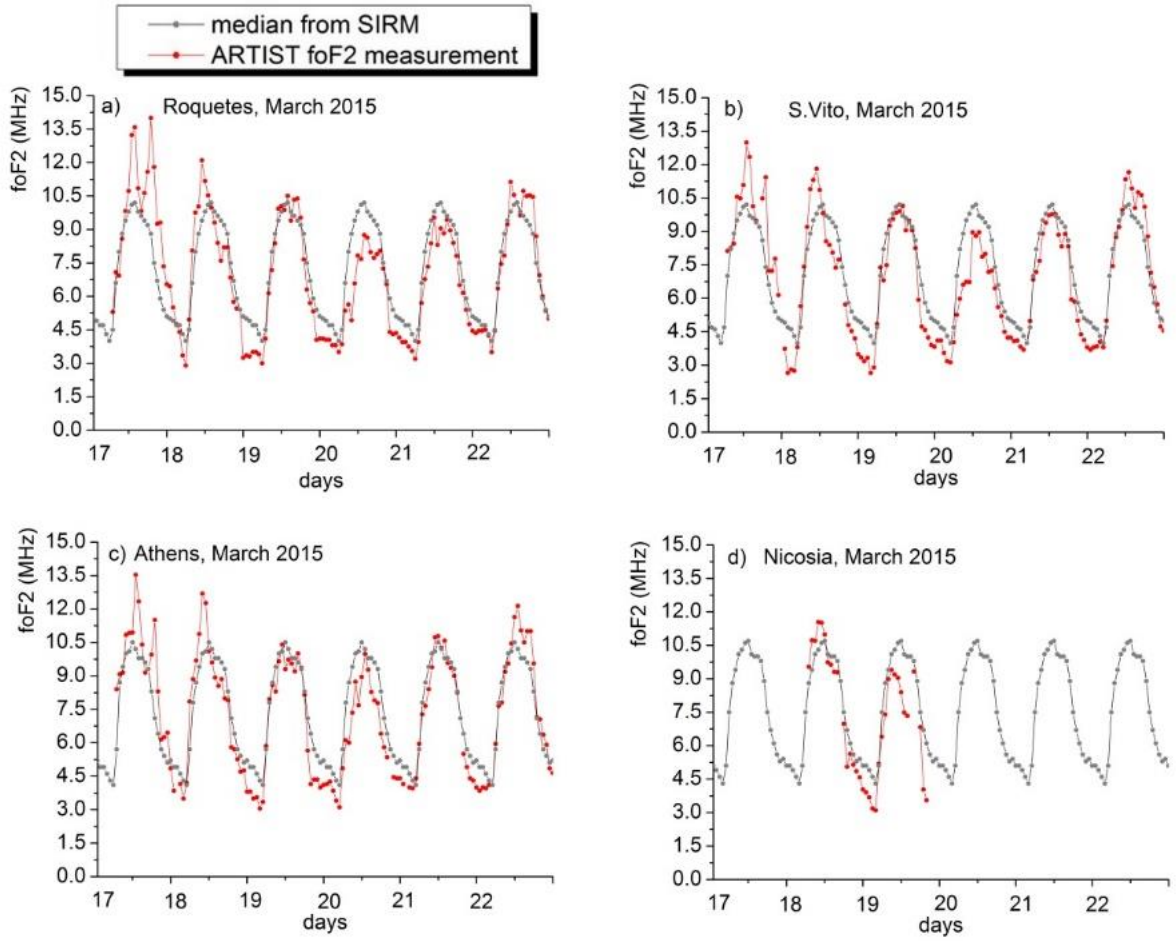


**Figure 7.11:** Hourly *fo*F2 measurements (red) and hourly monthly median *fo*F2 values (grey) for the period 17-22 March 2015 at: a) Roquetes, b) San Vito, c) Athens, d) Nicosia.

Cases showing differences ($foF2_{meas}$ - $foF2_{med}$) ≥ 1 MHz (positive phases) and ($foF2_{meas}$ - $foF2_{med}$) ≤ - 1 MHz (negative phases) are considered significantly disturbed. This criterion allows to select, for each station, only those epochs characterized by very disturbed ionospheric conditions. The "status" of very disturbed ionospheric conditions, simultaneously occurring at least in six out of the seven Mediterranean stations considered in this study, is a further requirement aimed to guarantee that the epochs considered to test all models, refer actually to magneto-ionospheric storm conditions involving all the Mediterranean region. Moreover, ionograms autoscaled by ARTIST and Autoscala have been visually inspected and validated discarding those showing a misleading

autoscaling, so that only consistent electron density profiles have been assimilated by IonoPy and ISP models. Following these criteria, several cases were rejected and only a definite number of epochs were selected. A summary picture, showing the selected epochs and different combinations of reference and test sites is given in Fig. 7.12.

**A) Epochs with 7 measurements simultaneously available**

| Epochs | El Are | Roq | Rom | Gib | S. Vito | Ath | Nic |
|---|---|---|---|---|---|---|---|
| 19/03/2015-19:00 | ART | ART | AUT | AUT | ART | ART | ART |
| 20/12/2015-22:00 | ART | ART | AUT | AUT | ART | ART | ART |
| 20/12/2015-23:00 | ART | ART | AUT | AUT | ART | ART | ART |

**Reference stations (5): El Arenosillo, Rome, Gibilmanna, S. Vito, Nicosia**
**Test sites (2): Roquetes, Athens**

**B) Epochs with 6 measurements simultaneously available**

| Epochs | El Are | Roq | Rom | Gib | S. Vito | Ath | Nic |
|---|---|---|---|---|---|---|---|
| 18/03/2015-09:00* | DNA | ART | AUT | AUT | ART | ART | ART |
| 18/03/2015-23:00 | DNA | ART | AUT | AUT | ART | ART | ART |
| 19/03/2015-00:00 | DNA | ART | AUT | AUT | ART | ART | ART |
| 19/03/2015-01:00 | DNA | ART | AUT | AUT | ART | ART | ART |
| 19/03/2015-02:00 | DNA | ART | AUT | AUT | ART | ART | ART |
| 19/03/2015-03:00 | DNA | ART | AUT | AUT | ART | ART | ART |
| 14/10/2016-07:00 | DNA | ART | AUT | AUT | ART | ART | ART |
| 14/10/2016-11:00 | DNA | ART | AUT | AUT | ART | ART | ART |
| 14/10/2016-12:00 | DNA | ART | AUT | AUT | ART | ART | ART |

**a) Reference stations (5): Roquetes, Rome, Gibilmanna, Athens, Nicosia**
**Test sites (1): S. Vito * case not considered**
**b) Reference stations (5): Roquetes, Rome, Gibilmanna, Athens, S. Vito**
**Test sites (1): Nicosia**

**C) Epochs with 6 measurements simultaneously available**

| Epochs | El Are | Roq | Rom | Gib | S. Vito | Ath | Nic |
|---|---|---|---|---|---|---|---|
| 17/03/2015-13:00 | ART | ART | AUT | AUT | ART | ART | DNA |
| 20/03/2015-08:00 | ART | ART | AUT | AUT | ART | ART | DNA |
| 20/03/2015-09:00 | ART | ART | AUT | AUT | ART | ART | DNA |
| 20/03/2015-10:00 | ART | ART | AUT | AUT | ART | ART | DNA |
| 20/03/2015-14:00 | ART | ART | AUT | AUT | ART | ART | DNA |
| 20/03/2015-15:00 | ART | ART | AUT | AUT | ART | ART | DNA |
| 20/03/2015-16:00 | ART | ART | AUT | AUT | ART | ART | DNA |
| 20/12/2015-20:00 | ART | ART | AUT | AUT | ART | ART | DNA |
| 20/12/2015-21:00 | ART | ART | AUT | AUT | ART | ART | DNA |

**a) Reference stations (5): El Arenosillo, Roquetes, Rome, Gibilmanna, Athens**
**Test sites (1): S. Vito**
**b) Reference stations (5): El Arenosillo, Rome, Gibilmanna, Athens. S. Vito**
**Test sites (1): Roquetes**

**Figure 7.12:** Epochs considered in this study. ART and AUT indicate the cases where the *fo*F2 value and the electron density profiles are those autoscaled by ARTIST and Autoscala, respectively. DNA marks the cases for which *fo*F2 autoscaled data are not available.

The same epochs are also considered in the IonoPy(NV) operational mode to evaluate how a "bad" ingestion of electron density profiles can affect the predictive capabilities of IonoPy model.

### 7.2.2 Analysis and results

Three-dimensional electron density distributions of the ionosphere, calculated by the IRI climatological model and assimilative IonoPy_re(V), ISP, IonoPy(NV), and IonoPy(V) models, are taken as input by IONORT to compute quasi-vertical ionograms at the test sites of Roquetes, San Vito, Athens, and Nicosia, simulating some very short radio links between the test site, considered as the transmitting station (Tx), and a location very close to Tx, considered as the receiving station (Rx). Table 7.1 shows the radio links for which synthesized quasi-vertical ionograms are computed by IONORT. Some examples of synthetic ionograms calculated by IONORT, for some of the epochs listed in Fig. 7.12, and corresponding recorded ionograms, are presented in Figs. 7.13 and 7.14. Corresponding electron density profiles provided by each model are depicted in Fig. 7.15.

| **Tx**-Rx | Azimuth (°) | Distance (km) |
|---|---|---|
| **Athens** (38.0° N, 23.5° E) - Eleusi (38.0° N, 23.3° E) | 270.0 | ≈ 18 |
| **Nicosia** (35.03° N, 33.16° E) - Latsia (35.0° N, 33.2° E) | 89.9 | ≈ 20 |
| **Roquetes** (40.8° N, 0.5° E) - Bitem (40.5° N, 0.3° E) | 206.8 | ≈ 37 |
| **San Vito** (40.6° N, 17.8° E) - Brindisi (40.4° N, 17.6° E ) | 217.3 | ≈ 28 |

**Table 7.1:** Radio links established to calculate synthesized quasi-vertical ionograms for each epoch listed in Fig. 7.12. The test sites for which *fo*F2 autoscaled measurements are available for testing are indicated in bold.

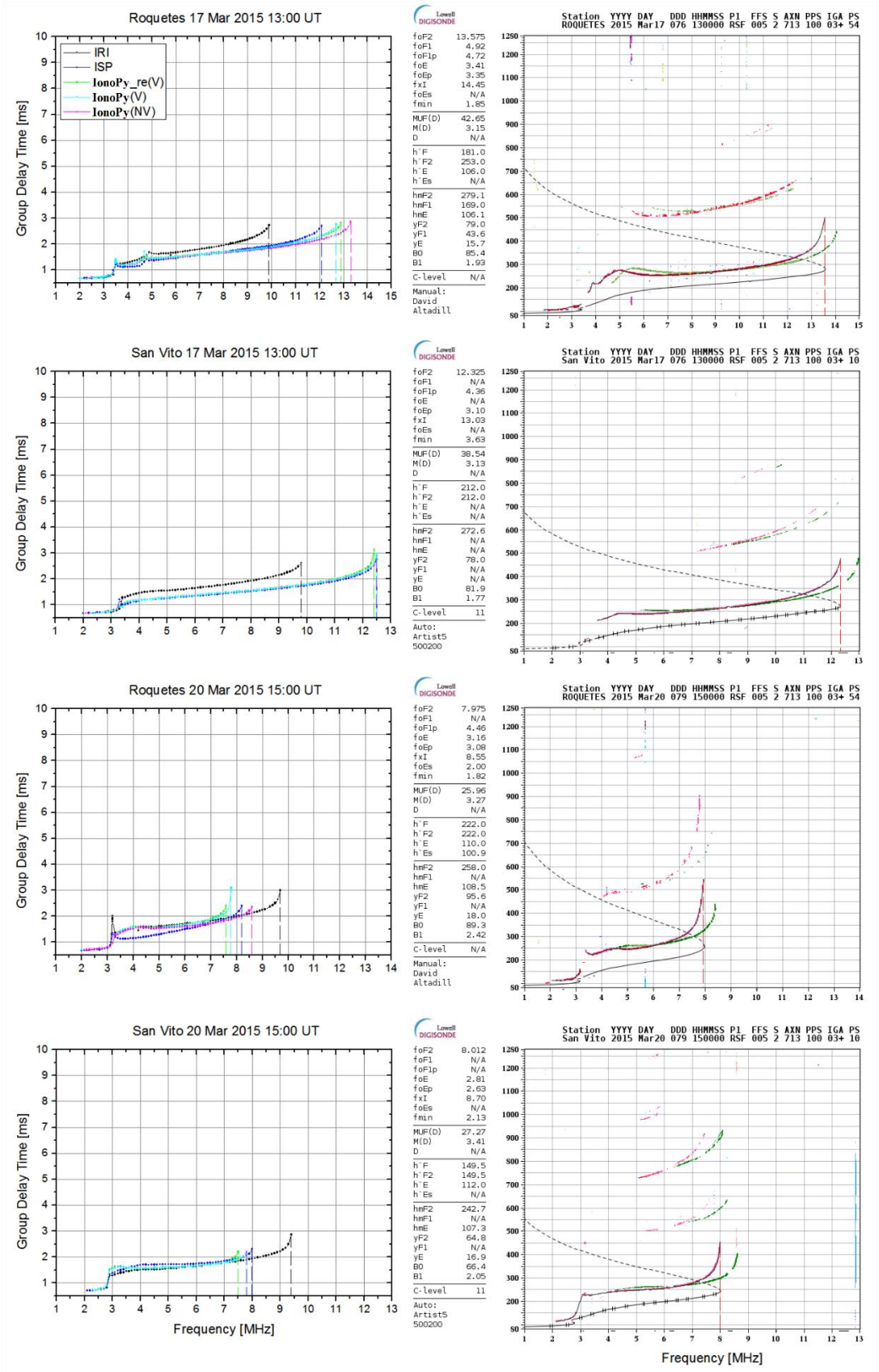


**Figure 7.13:** Synthetic (left) and measured ionograms (right) for some of the epochs listed in Fig. 7.12.

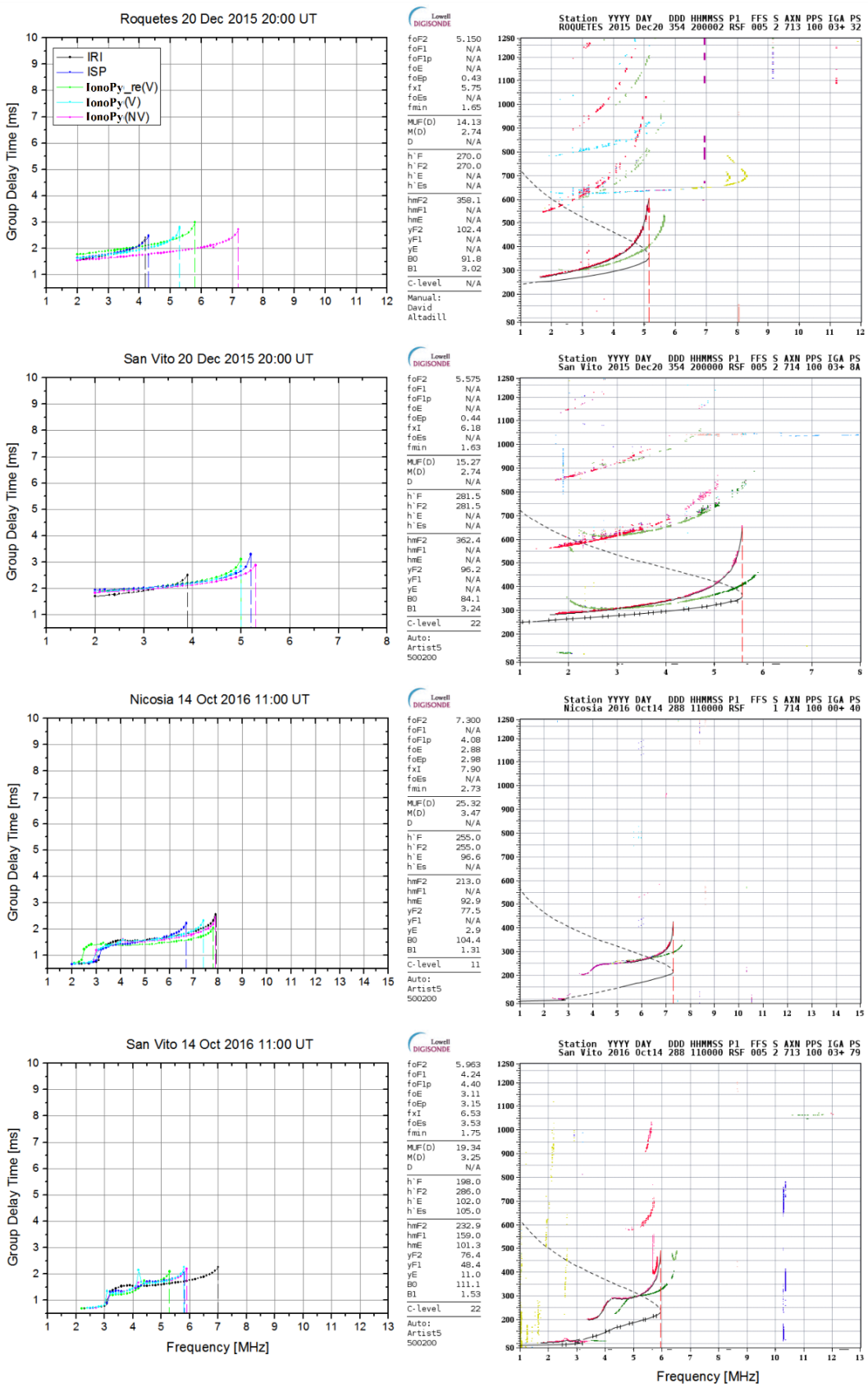


**Figure 7.14:** Same as Fig. 7.13.

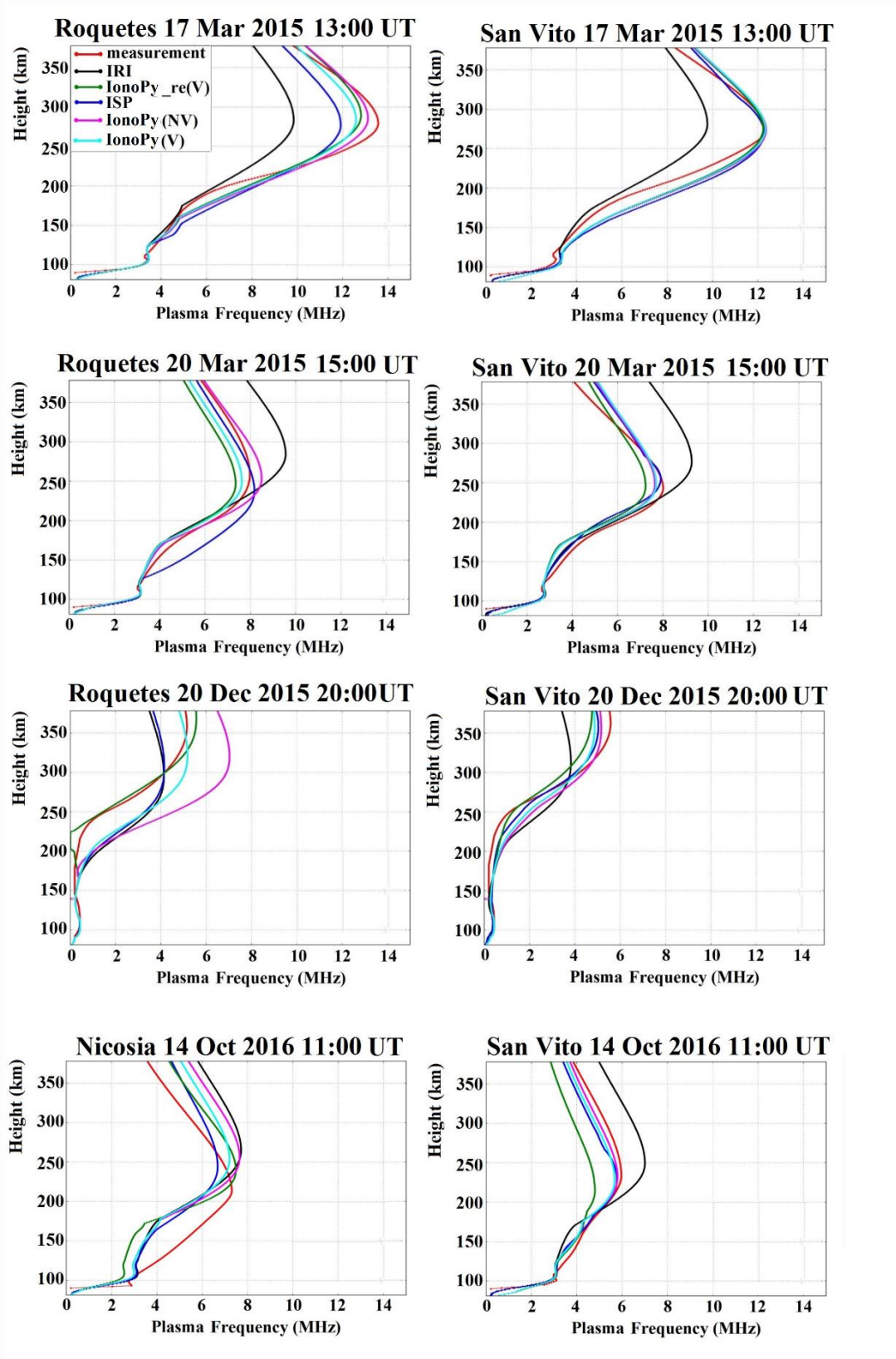


**Figure 7.15:** Electron density profiles for the same epochs considered in Figs. 7.13 and 7.14.

Values of $foF2_{meas}$ available at test sites have been carefully inspected to check their reliability, and then compared to the ordinary mode (O-mode) of propagation Maximum Usable Frequencies (MUFs) obtained from the synthetic quasi-vertical ionograms. This comparison is relevant because, given the relatively small distances between Tx and Rx (see Table 7.1), the synthetic MUF value can be considered basically as the synthetic critical frequency ($foF2_{Syn}$) of the O-mode of propagation.

To have an idea about models accuracy, the frequency distribution of the difference $(foF2_{meas} - foF2_{Syn})$ using the intervals $[(J+1)\cdot(-0.3), J\cdot(-0.3)) \cup (J\cdot 0.3, (J+1)\cdot 0.3]$ MHz for $J$= 0,1,…,12 as bins, have been calculated. Corresponding results are presented in the form of hystograms in Fig. 7.16.

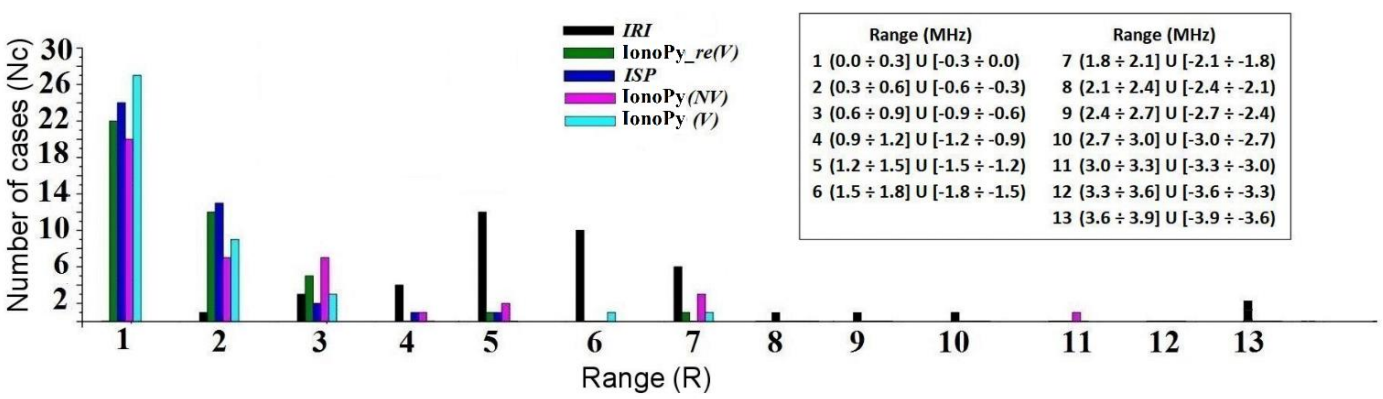


**Figure 7.16:** Comparison between IRI, IonoPy_re(V), ISP, IonoPy(NV), and IonoPy(V) hystograms of $(foF2_{meas} - foF2_{Syn})$. The meaning of the abscissa *R* is reported in the upper right part of the panel.

The goodness of each model used is estimated by calculating the RMSE of each model for the single test site, and then taking into account all the test sites as a whole to calculate the following global RMSE:

$$\mathrm{RMSE}_{\text{meas-Syn}} = \sqrt{\frac{\sum_{i=1}^{N} (foF2_{\text{meas},i} - foF2_{\text{Syn},i})^2}{N}}, \tag{7.6}$$

where *N* represents the number of cases for which the RMSE is calculated.

The results obtained in terms of RMSE are shown in Table 7.2.

| Test Sites | IRI | IonoPy_re(V) | ISP | IonoPy (NV) | IonoPy (V) | N |
|---|---|---|---|---|---|---|
| Athens | 1.97 | 0.39 | 0.27 | 1.30 | 0.05 | 3 |
| Nicosia | 1.45 | 0.83 | 0.36 | 1.36 | 0.85 | 9 |
| Roquetes | 2.03 | 0.43 | 0.68 | 0.84 | 0.44 | 12 |
| San Vito | 1.64 | 0.35 | 0.23 | 0.33 | 0.29 | 17 |
| Global | 1.75 | 0.52 | 0.44 | 0.88 | 0.50 | 41 |

**Table 7.2:** Models performance in terms of RMSE in MHz.

### 7.2.3 Discussion of results

The statistical analysis of ($foF2_{meas}$ - $foF2_{Syn}$) frequency distributions (see Fig. 7.16) provides the following results:

a) a decreasing trend of $N$, for $R$ increasing from 1 to 3, values of $N \leq 1$ for $4 \leq R \leq 7$, and $N = 0$ for $R \geq 8$, are the peculiarities of IonoPy_re(V), ISP and IonoPy(V);
b) an increasing trend of $N$ for $R$ increasing from 1 to 5, values of $N$ remaining relatively high for $6 \leq R \leq 7$, $N = 1$ for $8 \leq R \leq 10$, and $N = 2$ for $R = 13$, are the characteristics observed for the IRI model;
c) a decreasing trend of $N$ for $R$ increasing from 1 to 2, a value of $N$ relatively high for $R = 3$, values of $N \leq 3$ for $4 \leq R \leq 7$, and $N = 1$ for $R = 11$, characterize the IonoPy(NV) model.

Given that $R$ defines ranges for which the difference ($foF2_{meas}$ - $foF2_{Syn}$) becomes significantly larger as $R$ increases, the results listed above suggest that $foF2_{Syn}$ values calculated using IonoPy_re(V), ISP and IonoPy(V) three-dimensional electron density matrices, are clearly more accurate than those obtained using IRI and IonoPy(NV) ones.

This issue is also confirmed by the results (see Table 7.2) showing the global accuracy expressed in terms of RMSE provided by each model: 1.75, 0.88, 0.52, 0.50, and 0.44 MHz for IRI, IonoPy(NV), IonoPy_re(V), IonoPy(V), and ISP, respectively. This indicates that the global performance provided by the IRI model is by far the worst, followed by the IonoPy(NV) model, while the IonoPy_re(V) and IonoPy(V) performance can be considered equivalent and just a little below the ISP one.

A more detailed analysis of the results about the accuracy provided by the different models in each single test site, suggests other considerations:

a) when comparing IonoPy_re(V) and ISP models, assimilating validated measured electron density profiles from the same stations, one finds that:

   (1) at Roquetes, IonoPy_re(V) performs better than ISP;
   (2) at Athens and San Vito, ISP performs slightly better than IonoPy_re(V);
   (3) at Nicosia, ISP performs better than IonoPy_re(V).

   It is worth noting that in the ISP procedure (Pezzopane et al. 2011, 2013) is embedded the SIRMUP regional model whose validity area, extending approximately in latitude from 38°N to 60°N and in longitude from 3.5°W to 21°E, covers pretty well the Mediterranean sector; the operational mode IonoPy(V) is instead based only on the IRI global model.

   These considerations could explain why the ISP performance is overall better than the IonoPy_re(V) one; in fact, due to past international efforts devoted to the development and improvement of regional ionospheric modeling (Bradley 1999; Hanbaba 1999), it is currently recognized by the scientific

community that regional models can provide, over limited areas, a representation of the ionosphere more accurate than the one that can be achieved by global models.

IonoPy_re(V) results obtained at Nicosia deserve a particular attention since it is a somewhat isolated test-site, pretty far from assimilated reference stations (≈10° in longitude away from Athens). For these conditions, the weight of assimilated vertical electron density profiles, which depends on the value of the $\sigma$ parameter, can be lower than the weight due to the already updated background (calculated by IRI-SIRMUP and IRI UP procedures, respectively for ISP and IonoPy). In virtue of these considerations, the worsening of IonoPy_re(V) at Nicosia can be ascribed to the weakness of the IRI UP method in describing the effective indices distribution for both points pretty far from assimilated reference stations and at the outer edge of the considered Mediterranean grid, where there are few available stations, making it difficult to obtain a good description of the spatial correlation of the modeled quantity through the Kriging method.

b) it must be pointed out that, owing to the limited number of stations used in the operational mode IonoPy_re(V), the IRI UP in the IonoPy model does not exploit fully its potentiality in providing a three-dimensional ionospheric mapping as much as possible close to the real ionospheric conditions. For this reason, the IonoPy model have also been applied considering the operational modes IonoPy(NV) and IonoPy(V) including additional European stations for which ARTIST autoscaled electron density profiles were available. In this way, the use of a larger number of available stations, over an extended geographical area, allows to obtain an improved empirical variogram, which is necessary to properly describe the spatial correlation of effective indices through the Universal Kriging method (chapter 3).

   Comparing IonoPy_re(V), IonoPy(NV), and IonoPy(V) models, the last two assimilating not validated and validated measured electron density profiles from all reference stations listed in Table 7.1, one finds that:

   (1) at Athens, Nicosia, Roquetes, and San Vito, IonoPy(NV) performs much worse than IonoPy_re(V) and IonoPy(V); these results demonstrate how the assimilation of correct electron density profiles is of utmost importance for obtaining a reliable ionospheric modeling;

   (2) while at Nicosia and Roquetes IonoPy_re(V) and IonoPy(V) performance can be considered practically equivalent, at San Vito IonoPy(V) performs slightly better than IonoPy_re(V), while a significant improvement is obtained at Athens.

   These outcomes highlights that the larger is the number of assimilated stations, from which the electron density profiles are assimilated, the more the IonoPy algorithm works at best of its possibilities.

c) when comparing ISP with IonoPy(V), one finds that:

   (1) at Roquetes, IonoPy(V) performs better than ISP, but a significant improvement compared to IonoPy_re(V) is not observed;
   (2) at Athens, IonoPy(V) performs better than ISP and a significant improvement is also observed compared to IonoPy_re(V);
   (3) at San Vito, although IonoPy(V) shows slightly better results than IonoPy_re(V), ISP continues to perform better;
   (4) at Nicosia, ISP performances continue to be better because IonoPy(V) does not show any improvement compared to IonoPy_re(V).

   It must be pointed out that these results come out from a non-peer to peer comparison between ISP and IonoPy(V), since the autoscaled data assimilated do not come from exactly the same assimilation stations. Notwithstanding the overall improvement of IonoPy(V) compared to IonoPy_re(V), mainly due to its capabilities to produce a more accurate ionospheric representation over Athens, ISP performances globally continue to be better than IonoPy(V) ones. The greater precision of ISP is much probably due to the regional model SIRMUP underlying the ISP algorithm which, in the Mediterranean area considered in this

study, offers a more realistic representation of the ionosphere than the one provided by the IRI global model underlying the IonoPy model.

d) from the comparison between the IRI background model and nowcasting models one finds that:

   (1) at Athens, Nicosia, Roquetes, and San Vito, the IRI performance is considerably the worst; in fact, values of RMSE calculated for the IRI model at Athens, Nicosia, Roquetes, and San Vito are always greater than 1.4 MHz, while those referring to IonoPy_re(V), ISP, and IonoPy(V) nowcasting models are at most around 0.8 MHz;

   (2) it is worth noting that also IonoPy(NV), although it works assimilating not validated electron density profiles, can perform better than IRI at all test sites. This is also testified from the comparison between the magenta (IonoPy(NV)) and black (IRI) electron density profiles, and the measured electron density profiles shown in red in Fig. 7.15.

   These outcomes show unequivocally that the IRI climatological model does not adequately "capture" the geomagnetic-ionospheric storm effects occurring over a relatively large area. On the contrary, the availability of autoscaled data, allows nowcasting models to "reproduce" the storm effects impacting the ionosphere and hence to get a three-dimensional ionospheric image much more accurate than that provided by the IRI model. This conclusion is most noticeable by comparing the electron density profiles shown in Fig. 7.15.

Relying on the SIRMUP regional model, the ISP algorithm is applicable only on the Mediterranean region. On the contrary, the IonoPy model can be potentially applied to any region characterized by a network of ionosondes. This greatly compensates its slightly lower performance when compared to ISP. Another benefit of the IonoPy model is a computer run time much lower than that of the ISP one, a very important feature for nowcasting models.

## Summary

In this chapter, the algorithm to update the bottomside part of the vertical ionospheric electron density profile, by assimilating the whole bottomside profile recorded by an ionosonde, has been first mathematically described and then compared to the IRI and ISP models over the Mediterranean region for very disturbed periods.

It has been demonstrated that this algorithm improves the bottomside ionosphere modeling if the electron density measurements, at a particular height, are carefully merged with those values provided by the underlying background model. It is worth noting how the background model of IonoPy is the already updated description of the ionosphere given by the IRI UP procedure; thus, the IonoPy model is expected to provide a greater improvement above all in the lower E and F1 regions, being the F2 region already pretty well improved by the IRI UP method.

Furthermore, the application of the topside model described in chapter 6 allows to update also the lower topside part of the ionosphere; thus, the IonoPy model is able to provide an update of the whole ionospheric vertical electron density profile from the base of the ionosphere (≅90 km) to an altitude of about 600 km.

# Conclusions

In this thesis a three-dimensional regional assimilative model of the ionospheric electron density has been presented. The main ideas which led to the development of this model, its implementation, and the subsequent validation have been described.

The study highlights the importance of the use of effective indices in the ionospheric modeling. In fact, effective indices, calculated through the assimilation of ionospheric measurements, are able to catch small-scale and fast-changing behaviours of the ionosphere. The implemented method, based on the calculation of effective indices, $IG_{12eff}$ and $R_{12eff}$, has proven to be a powerful way to obtain a more reliable and up to date description of the ionospheric plasma distribution over a particular region. Notably, the proposed method has proven to be much more accurate and precise compared to the original IRI background model, for both quiet and, particularly, magnetically disturbed conditions. The data assimilation is based on both ionosonde data and vTEC data, the former being possible only in regions well covered by a dense network of ionosondes. Compared to the performance achieved when assimilating ionosonde data, the assimilation based on vTEC data produces slightly worse results, but can be applied also in very remote regions, due to the very good coverage of the terrestrial surface that GNSS receivers have achieved in recent years. The validation of the proposed method, called IRI UP, has proven that reliable and trustworthy updated maps of the main ionospheric characteristics can be provided with a satisfactory degree of confidence, especially under moderate, strong and severe geomagnetic storm conditions. This means that IRI UP represents an interesting approach for Space Weather purposes in the ionospheric domain.

A model based on the IRI UP method and in-situ electron density values collected by Langmuir probes on-board Swarm satellites has been implemented to improve the description of the topside ionosphere, which is modeled with difficulty by the IRI model, when compared to COSMIC Radio Occultation profiles.

Finally, a procedure to assimilate the whole bottomside part of the vertical ionospheric electron density profile recorded by an ionosonde, has been implemented. This procedure, called IonoPy and embedding the IRI UP method, has demonstrated to significantly improve the bottomside ionosphere modeling if electron density measurements, at a particular height, are carefully merged with those provided by the underlying background model.

To sum up, IonoPy is a model that by assimilating space-sparse ionospheric data measured by different ionospheric instruments is able to reliably update the IRI model, thus obtaining a more consistent description of the plasma distribution in the ionospheric domain. This makes the IonoPy model an interesting tool for the Space Weather community.

## References


Adeniyi, J.O., Bilitza, D., Radicella, S.M., Willoughby, A.A. (2003) Equatorial F2-peak parameters in the IRI model. Adv. Space Res., 31(3), 507-512, https://doi.org/10.1016/S0273-1177(03)00039-5

Altadill, D., Torta, J., Blanch, E. (2009) Proposal of new models of the bottom-side B0 and B1 parameters for IRI. Adv. Space Res., 43, 1825-1834. doi:10.1016/j.asr.2008.08.014

Altadill, D., Magdaleno, S., Torta, J.M., Blanch, E. (2013) Global empirical models of the density peak height and of the equivalent scale height for quiet conditions. Adv. Space Res., 52, 1756–1769. doi:10.1016/j.asr.2012.11.018

Anderson, D.N., Mendillo, M., Herniter, B. (1987) A Semi-Empirical, Low-Latitude Ionospheric Model. Radio Sci., 22, 292-306

Anderson, D.N., Decker, D.T., Valladares, C. E. (1996) Global Theoretical Ionospheric Model (GTIM), in Handbook of Ionospheric Models, edited by R.W. Schunk, Boulder, USA, Solar-Terrestrial Energy Program (STEP), Scientific Committee on Solar Terrestrial Physics (SCOSTEP), Special Report, pp. 133-152

Angling, M.J., Cannon, P.S. (2004) Assimilation of radio occultation measurements into background ionospheric models, Radio Sci., 39, RS1S08, https://doi.org/10.1029/2002RS002819

Anthes, R.A., Ector, D., Hunt, D.C., Kuo, Y., Rocken, C., Schreiner, W.S., Sokolovskiy, S.V., Syndergaard, S., Wee, T., Zeng, Z., Bernhardt, P.A., Dymond, K.F., Chen, Y., Liu, H., Manning, K., Randel, W.J., Trenberth, K.E., Cucurull, L., Healy, S.B., Ho, S., McCormick, C., Meehan, T.K., Thompson, D.C., Yen, N.L. (2008) Bull. Amer. Meteor. Soc., 89, 313–333. doi:10.1175/BAMS-89-3-313

Araujo-Pradere, E.A., Fuller-Rowell, T.J., Codrescu, M.V. (2002) STORM: An empirical storm-time ionospheric correction model 1. Model description. Radio Sci., 37(5), doi:10.1029/2001RS002467

Balan, N., Bailey, G.J. (1995) Equatorial plasma fountain and its effect: Possibility of an additional layer. J. Geophys. Res., 100(21), 421-432. https://doi.org/10.1029/95JA01555

Bassiri, S., Hajj, A. (1992) Higher order ionospheric effects on the Global Positioning System observables and means of modelling them. *Manuscripta Geodaetica*, 18, 280–290

Belehaki, A., Tsagouri, I., Kutiev, I., Marinov, P., Fidanova, S. (2012) Upgrades to the topside sounders model assisted by Digisonde (TaD) and its validation at the topside ionosphere. J. Space Weather Space Clim., 2, A20. doi:10.1051/swsc/2012020

Bent, R.B., Llewellyn, S.K., Schmid, P.E. (1972) Description and evaluation of the Bent ionospheric model. Vol. 1–3, Space and Missile Syst. Org., Rep. SAMSO

Bibl, K,, Reinisch, B,W, (1978) The universal digital ionosonde. Radio Sci., 13, 519-530. http://dx.doi.org/10.1029/RS013i003p00519

Bilitza, D., Sheikh, M., Eyfrig, R. (1979) A global model for the height of the F2-peak using M3000 values from the CCIR numerical map. Telecommunication J., 46, 549-553

Bilitza, D. (1990) International Reference Ionosphere 1990. NSSDC/WDC-A-R&S 90-22.

Bilitza, D., Rawer, K. (1990) New options for IRI electron density in the middle ionosphere. Adv. Space Res., 10, 7-16. doi:10.1016/0273-1177(90)90299-F

Bilitza, D., Bhardwaj, S., Koblinsky, C. (1997) Improved IRI predictions for the GEOSAT time period. Adv. Space Res., 20, 9, 1755-1760, doi:10.1016/S0273-1177(97)00585-1

Bilitza, D., Radicella, S.M., Reinisch, B.W., Adeniyi, J.O., Mosert de Gonzalez, M.E., Zhang, S., Obrou, O. (2000) New $B_0$ and $B_1$ models for IRI. Adv. Space Res., 251, 89–96. https://doi.org/10.1016/S0273-1177(99)00902-3

Bilitza, D., Williamson, R. (2000) Toward a better representation of the IRI topside based on ISIS and Alouette data. Adv. Space Res., 25(1), 149-152. https://doi.org/10.1016/S0273-1177(99)00912-6

Bilitza, D. (2001) International Reference Ionosphere 2000. Radio Sci., 36, 261–275. doi:10.1029/2000RS002432

Bilitza, D. (2003) International Reference Ionosphere 2000: Examples of improvements and new features. Adv. Space Res., 31(3), 757-767. doi:10.1016/S0273-1177(03)00020-6

Bilitza, D. (2004) A correction for the IRI topside electron density model based on Alouette/ISIS topside sounder data. Adv. Space Res., 33, 838–843. https://doi.org/10.1016/j.asr.2003.07.009

Bilitza, D., Reinisch, B.W., Radicella, S.M., Pulinets, S., Gulyaeva, T., Triskova, L. (2006) Improvements of the International Reference Ionosphere model for the topside electron density profile. Radio Sci., 41, RS5S15. doi:10.1029/2005RS003370

Bilitza, D., Reinisch, B.W. (2008) International Reference Ionosphere 2007: improvements and new parameters. Adv. Space Res., 42(4), 599-609. http://dx.doi.org/10.1016/j.asr.2007.07.048

Bilitza, D. (2009) Evaluation of the IRI-2007 model options for the topside electron density, Adv. Space Res., 44, 701-706. doi:10.1016/j.asr.2009.04.036

Bilitza, D., McKinnell, L.A., Reinisch, B., Fuller-Rowell, T. (2011): The International Reference Ionosphere today and in the future. J. Geod., 85, 909-920. doi:10.1007/s00190-010-0427-x

Bilitza, D., Altadill, D., Zhang, Y., Mertens, C., Truhlik, V., Richards, P., McKinnell, L.A., Reinisch, B. (2014) The International Reference Ionosphere 2012 – a model of international collaboration. J. Space Weather Space Clim., 4, A07. doi:10.1051/swsc/2014004

Bilitza, D., Altadill, D., Truhlik, V., Shubin, V., Galkin, I., Reinisch, B., Huang, X. (2017) International Reference Ionosphere 2016: From ionospheric climate to real-time weather predictions. Space Weather, 15, 418–429. doi:10.1002/2016SW001593

Booker, H.G. (1977) Fitting of multi-region ionospheric profiles of electron density by a single analytic function of height. J. Atmos. Terr. Phys., 39, 619–623. https://doi.org/10.1016/0021-9169(77)90072-1

Bradley, P.A, Dudeney, J.R. (1973) A simple model of the vertical distribution of electron concentration in the ionosphere. J. Atm. Terr. Phys., 35, 2131-2146. doi:10.1016/0021-9169(73)90132-3

Bradley, P. (1999) PRIME (Prediction and Retrospective Ionospheric Modelling over Europe), Cost Action 238, Final Report, Rutherford Appleton Laboratory, Chilton Didcot, UK.

Bradley, P.A., Stanislawska, I., Juchnikowski, G. (2009) Options for mapping *fo*F2. Adv. Space Res., 43, 1776-1779. doi:10.1016/j.asr.2008.09.028

Breit, G., Tuve, M.A. (1926) A test of the existence of the conducting layer. Phys. Rev., 28, 554-575

Brown, S., Bilitza, D., Yigit, E. (2018) Ionosonde-based indices for improved representation of solar cycle variation in the International Reference Ionosphere model. J. Atm. Sol. Terr. Phys., 171, 137-146. https://doi.org/10.1016/j.jastp.2017.08.022

Brunini, C., van Zele, M.A., Meza, A., Gende, M. (2003) Quiet and perturbed ionospheric representation according to the electron content from GPS signals. J. Geophys. Res., 108, SIA4-1. doi:10.1029/2002JA009346

Brunini, C., Meza, A., Bosch, W. (2005) Temporal and spatial variability of the bias between TOPEX- and GPS-derived total electron content. J. Geod., 79. doi:10.1007/s00190-005-0448-z

Brunini, C., Azpilicueta, F. (2010) GPS slant total electron content accuracy using the single layer model under different geomagnetic regions and ionospheric conditions. J. Geod., 84, 293-304. doi:10.1007/s00190-010-0367-5

Brunner, F.K., Gu, M. (1991) An improved model for the dual frequency ionospheric correction of GPS observations. Manuscript Geodaetica, 16, 205–214

Budden, K.G. (1961) Radio waves in the ionosphere. Cambridge University Press

Chapman, S. (1931), The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating Earth, Proc. Phys. Soc. Lond., 43 (1), 26-45

Cherniak, I., Zakharenkova, I., Redmon, R.J. (2015) Dynamics of the high-latitude ionospheric irregularities during the 17 March 2015 St. Patrick's Day storm: Ground-based GPS measurements. Space Weather, 13. doi:10.1002/2015SW001237

Chu, Y.H., Su, C.L., Ko, H.T. (2010) A global survey of COSMIC ionospheric peak electron density and its height: A comparison with ground-based ionosonde measurements. Adv. Space Res., 46, 431-439. doi:10.1016/j.asr.2009.10.014

Ciraolo, L., Spalla, P. (1997) Comparison of ionospheric total electron content from the Navy Navigation Satellite System and GPS. Radio Sci., 32, 1071-1080. doi:10.1029/97RS00425

Ciraolo, L., Azpilicueta, F., Brunini, C., Meza, A., Radicella, S.M. (2007) Calibration errors on experimental slant total electron content (TEC) determined with GPS. *J. Geod.*, 81, 111–120. doi:10.1007/s00190-006-0093-1

Coisson, P., Radicella, S.M., Leitinger, R., Nava, B. (2006) Topside electron density in IRI and NeQuick: Features and limitations. Adv. Space Res., 37 (5), 937-942. doi:10.1016/j.asr.2005.09.015

Coisson, P., Nava, B., Radicella, S.M. (2009) On the use of NeQuick topside option in IRI-2007. Adv. Space Res., 43(11), 1688-1693, https://doi.org/10.1016/j.asr.2008.10.035

Daniell, R. E., Brown, L. D., Anderson, D. N., Fox, M.W., Doherty, P. H., Decker, D. T., Sojka, J. J., Schunk, R. W., (1995) Parameterized ionospheric model: A global ionospheric parameterization based on first principles models, Radio Sci., 30, 1499-1510. doi: 10.1029/95RS01826.

Davies, K. (1990) Ionospheric Radio. IEE Electromagnetic Waves Series 31. Peter Peregrinus Ltd., London

Doherty, P., Coster, A.J., Murtagh, W. (2004) Space weather effects of October-November 2003. GPS Solut., 8, 267-271. doi:10.1007/s10291-004-0109-3

DuCharme, E.D., Petrie, L.E., Eyfrig, R. (1971) A Method for Predicting the $F_1$ Layer Critical Frequency. Radio Sci., 6(3), 369–378, doi: 10.1029/RS006i003p00369

DuCharme, E.D., Petrie, L.E., Eyfrig, R. (1973) A method for predicting the $F_1$ layer critical frequency based on the Zurich smoothed sunspot number. Radio Sci., 8(10), 837–839. doi: 10.1029/RS008i010p00837

Emmert, J.T., Richmond, A.D., Drob, D.P. (2010) A computationally compact representation of Magnetic-Apex and Quasi-Dipole coordinates with smooth base vectors. J. Geophys. Res., 115. doi:10.1029/2010JA015326

Ezquer, R.G., Jadur, C.A., Mosert de Gonzalez, M. (1998) IRI-95 TEC predictions for the South American peak of the equatorial anomaly. Adv. Space Res., 22(6), 811-814, https://doi.org/10.1016/S0273-1177(98)00103-3

Friis-Christensen, E., Lühr, H., Hulot, G. (2006) Swarm: a constellation to study the Earth's magnetic field. Earth Planets Space, 58, 351–358. doi:10.1186/BF03351933

Fonda, C., Coisson, P., Nava, B., Radicella, S.M. (2005) Comparison of analytical functions used to describe topside electron density profiles with satellite data. Ann. Geophys., 48(3). doi:10.4401/ag-3213

Fuller-Rowell, T.J., Rees, D. (1980) A three-dimensional Time Dependent Global Model of the Thermosphere. J. Atmos. Sci., 37, 2545-2567

Fuller-Rowell, T.J., Codrescu, M.V., Araujo-Pradere, E., Kutiev, I. (1998) Progress in developing a storm-time ionospheric correction model. Adv. Space Res., 22(6), 821-827. doi:10.1016/S0273-1177(98)00105-7

Fuller-Rowell, T.J., Araujo-Pradere, E., Codrescu, M.V. (2000) An empirical ionospheric storm-time correction model. Adv. Space Res., 25(1), 139–146. https://doi.org/10.1016/S0273-1177(99)00911-4

Galkin, I.A., Khmyrov, G.M., Kozlov, A.V., Reinisch, B.W., Huang, X., Paznukhov, V.V. (2008) The ARTIST 5. Radio Sounding and Plasma Physics. AIP Conf Proc 975, 150-159

Galkin, I.A., Reinisch, B.W. (2008) The new ARTIST 5 for all digisondes. Ionosonde Network Advisory Group Bulletin 69, 8 pp. [Available at: http://www.ips.gov.au/IPSHosted/INAG/web-69/2008/artist5-inag.pdf.]

Galkin, I.A., Reinisch, B.W., Huang, X., Bilitza, D. (2012) Assimilation of GIRO data into a real-time IRI. Radio Sci., 47, RS0L07. doi:10.1029/2011RS004952

García-Fernández, M., Hernández-Pajares, M., Juan, J.M., Sanz, J., Orús, R., Coisson, P., Nava, B., Radicella, S.M. (2003) Combining ionosonde with ground GPS data for electron density estimation. J. Atm. Sol. Terr. Phys., 65(6), 683-691, https://doi.org/10.1016/S1364-6826(03)00085-3.

Grynyshyna-Poliuga, O., Stanislawska, I., Swiatek, A. (2014) Regional ionosphere mapping with kriging and B-spline methods. Chapter 11 in the book "Mitigation of Ionospheric Threats to GNSS: an Appraisal of the Scientific and Technological Outputs of the TRANSMIT Project" edited by Riccardo Notarpietro, Fabio Dovis, Giorgiana De Franceschi and Marcio Aquino, INTECH, 135-147

Gulyaeva, T. (1987) Progress in ionospheric informatics based on electron-density profile analysis of ionograms. Adv. Space Res., 7. 39-48. doi:10.1016/0273-1177(87)90269-9

Habarulema, J.B., Katamzi, Z.T., Yizengaw, E. (2014) A simultaneous study of ionospheric parameters derived from FORMOSAT-3/COSMIC, GRACE, and CHAMP missions over middle, low, and equatorial latitudes: Comparison with ionosonde data. J. Geophys. Res. Space Physics, 119, 7732-7744. doi:10.1002/2014JA020192

Habarulema, J.B., Ssessanga, N. (2016) Adapting a climatology model to improve estimation of ionosphere parameters and subsequent validation with radio occultation and ionosonde data. Space Weather, 15, doi:10.1002/2016SW001549

Hajj, G.A., Romans, L.J. (1998) Ionospheric electron density profiles obtained with the Global Positioning System: Results from the GPS/MET experiment. Radio Sci., 33(1), 175-190. https://doi.org/10.1029/97RS03183

Hanbaba, R. (1999) Improved quality of services ionospheric telecommunication systems planning and operation. Cost Action 251, Final Report, Published by Space Research Centre,Warsaw, Poland

Hargreaves, J.K. (1992) The solar-terrestrial environment, Cambridge Univ. Press, New York

Hernandez-Pajares, M., Juan, J.M., Sanz, J., Bilitza, D. (2002) Combining GPS measurements and IRI model values for space weather specification. Adv. Space Res., 29(6), 949-958. doi:10.1016/S0273-1177(02)00051-0

Huang, X., Reinisch, B.W. (2001) Vertical electron content from ionograms in real time. Radio. Sci., 36(2), 335-342. doi:10.129/1999RS002409

Huber, P.J., Ronchetti, E.M. (2009) Robust statistics. John Wiley & Sons, Hoboken (New Jersey)

Hunsucker, R.D. (1991) Radio Techniques for Probing the Terrestrial Ionosphere, Springer-Verlag, Berlin Heidelberg. doi: 10.1007/978-3-642-76257-4

IEEE Std 211-1997 (1998) IEEE standard definition of terms for radio wave propagation. The Institute of Electrical and Electronics Engineers, New York. doi: 10.1109/IEEESTD.1998.87897

ISO (2009) Space environment (natural and artificial)−Earth's ionosphere model: international reference ionosphere and extension to the plasmasphere. Technical Specification TS16457, International Standardization Organization, Geneva, Switzerland

ITU-R (2009) ITU-R reference ionospheric characteristics. Recommendation ITU-R P.1239-2

Iwamoto, I., Katoh, H., Maruyama, T., Minakoshi, H., Watari, S., Igarashi, K. (2002) Latitudinal variations of solar flux dependence in the topside plasma density: comparison between IRI model and observations. Adv. Space Res., 29(6), 877-882. https://doi.org/10.1016/S0273-1177(02)00054-6

Jakowski, N., Wehrenpfennig, A., Heise, S., Reigber, C., Luhr, H., Grunwaldt, L., Meehan, T. K. (2002) GPS radio occultation measurements of the ionosphere from CHAMP: Early results, Geophys. Res. Lett., 29(10), doi: 10.1029/2001GL014364

Jakowski, N., Leitinger, R., Angling, M. (2004) Radio occultation techniques for probing the ionosphere. Ann. Geophys. 47 (2-3), 1049-1066. doi:10.4401/ag-3285

Jones, W.B., Gallet, R.M. (1962) Representation of diurnal and geographical variations of ionospheric data by numerical methods. Telecommun. J., 29, 129-149

Jones, W.B., Graham, R.P., Leftin, M. (1969) Advances in ionospheric mapping by numerical methods. ESSA Technical Report ERL107-ITS75, US Department of Commerce, Boulder, Colorado, USA

Kelley, M.C. (2009) The Earth's Ionosphere. International Geophysics (Book 96), Academic Press, 2 edition, San Diego, USA

Kitanidis, P.K. (1997) Introduction to geostatistics: application to hydrogeology. Cambridge University Press.

Klimenko, M.V., Klimenko, V.V., Zakharenkova, I.E., Cherniak, I.V. (2015) The global morphology of the plasmaspheric electron content during Northern winter 2009 based on GPS/COSMIC observation and GSM TIP model results. Adv. Space Res., 55, 2077-2085. doi:10.1016/j.asr.2014.06.027

Klobuchar, J.A., Kunches, J.M. (2000) Eye on the ionosphere: The spatial variability of ionospheric range delay. GPS Solut., 3(3), 70-74. doi: /10.1007/PL00012808

Knudsen, D.J., Burchill, J.K., Buchert, S.C., Eriksson, A.I., Gill, R., Wahlund, J.E., Åhlen, L., Smith, M., Moffat, B. (2017) Thermal ion imagers and Langmuir probes in the Swarm electric field instruments. J. Geophys. Res. Space Physics, 122, 2655–2673. doi:10.1002/2016JA022571

Komjathy, A., Langley, R. (1996) Improvement of a Global Ionospheric Model to Provide Ionospheric Range Error Corrections for Single-frequency GPS Users. Proceedings of the 52nd Annual Meeting of The Institute of Navigation, Cambridge, MA, June 1996, 557-566

Kouris, S., Muggleton, M.L. (1973a) Diurnal variation in the E-layer ionization. J. Atm. Solar Terr. Phys., 35, 133-139. doi:10.1016/0021-9169(73)90221-3

Kouris, S., Muggleton, L.M., (1973b) World morphology of the Appleton E-layer seasonal anomaly. J. Atm. Solar Terr. Phys., 35, 141-151. doi:10.1016/0021-9169(73)90222-5

Kouris, S.S., Xenos, T.D., Polimeris, K.V., Stergiou, D. (2004) TEC and *fo*F2 variations: preliminary results. *Ann. Geophys.*, 47 (4). doi:10.4401/ag-3346

Krankowski, A., Shagimuratov, I.I., Baran, L.W. (2007) Mapping of *fo*F2 over Europe based on GPS-derived TEC data. *Adv. Space Res.*, 39(5), 651-660. doi:10.1016/j.asr.2006.09.034

Krankowski, A., Zakharenkova, I., Krypiak-Gregorczyk, A., Shagimuratov, I.I., Wielgosz, P. (2011) Ionospheric electron density observed by FORMOSAT-3/COSMIC over the European region and validated by ionosonde data. J. Geod., 85, 949-964. doi:10.1007/s00190-011-0481-z

Kursinski, E.R., Hajj, G.A., Schofield, J.T., Linfield, R.P., Hardy, K.R. (1997) Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System. J. Geophys. Res., 102, 23429-23465. https://doi.org/10.1029/97JD01569

Kutiev, I.S., Marinov, P.G., Watanabe, S. (2006) Model of topside ionosphere scale height based on topside sounder data, Adv. Space Res., 37(5), 943-950,.doi:10.1016/j.asr.2005.11.021

Kutiev, I.S., Marinov, P.G. (2007) Topside sounder model of scale height and transition height characteristics of the ionosphere, Adv. Space Res., 39(5), 759-766. doi:10.1016/j.asr.2006.06.013

Langley, R. (1996) Propagation of the GPS signals. In: Kleusberg A, Teunissen P (eds) GPS for geodesy. Springer, Berlin Heidelberg New York, pp 103–140

Lei, J. et al. (2007) Comparison of COSMIC ionospheric measurements with ground-based observations and model predictions: Preliminary results. J. Geophys. Res., 112, A07308. doi:10.1029/2006JA012240

Liu, R.Y., Smith, P.A., King, J.W. (1983) A new solar index which leads to improved foF2 predictions using the CCIR Atlas. Telecommun. J., 50, 408-414

Liu, L., Wan, W., Zhang, M.-L., Ning, B., Zhang, S.-R., Holt, J.M. (2007a) Variations of topside ionospheric scale height over Millstone Hill during the 30-day incoherent scatter radar experiment, Ann. Geophys., 25, 2019-2027. doi:10.5194/angeo-25-2019-2007

Liu, L., Le, H., Wan, W., Sulzer, M.P., Lei, J., Zhang, M.-L. (2007b) An analysis of the scale heights in the lower topside ionosphere based on the Arecibo incoherent scatter radar measurements, J. Geophys. Res., 112, A06307. doi:10.1029/2007JA012250

Liu, L., Huang, H., Chen, Y., Le, H., Ning, B., Wan, W., Zhang, H. (2014) Deriving the effective scale height in the topside ionosphere based on ionosonde and satellite in situ observations. J. Geophys. Res. Space Physics, 119, 8472-8482. doi:10.1002/2014JA020505

Lomidze, L., Knudsen, D.J., Burchill, J., Kouznetsov, A., Buchert, S.C. (2018) Calibration and validation of Swarm plasma densities and electron temperatures using ground-based radars and satellite radio occultation measurements. Radio Sci., 53, 15–36. https://doi.org/10.1002/2017RS006415

Luan, X., Liu, L., Wan, W., Lei, J., Zhang, S.-R., Holt, J.M., Sulzer, M.P. (2006) A study of the shape of topside electron density profile derived from incoherent scatter radar measurements over Arecibo and Millstone Hill Radio Sci., 41, RS4006. doi:10.1029/2005RS003367

Makarevich, R.A., Nicolls, M.J. (2013) Statistical comparison of TEC derived from GPS and ISR observations at high latitudes. *Radio Sci.*, 48, 441–452. doi:10.1002/rds.20055

Martyn, D.F. (1935) The propagation of medium radio waves in the ionosphere. Proc. Phys. Soc., 47, 323-339

Matheron, G. (1963) Principles of geostatistics. Economic geology 58, 1246-1266

McBratney, A.B., Webster, R. (1986) Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. J. Soil Sci., 37, 617-639

McNamara, L. F. (1984) Prediction of total electron content using the International Reference Ionosphere. Adv. Space Res., 4, 25–50. https://doi.org/10.1016/0273-1177(84)90470-8

Mechtley, E., Bilitza, D. (1974) Models of *D*-region electron concentration. Rep. IPW-WB1

Mirò Amarante, G., Cueto Santamarìa, M., Alazo, K., Radicella, S.M. (2007) Validation of the STORM model used in IRI with ionosonde data. Adv. Space Res., 39, 681-686. doi:10.1016/j.asr.2007.01.072

Nava, B., Coisson, P., Radicella, S.M. (2008) A new version of the NeQuick ionosphere electron density model. J. Atm. Sol. Terr. Phys., 70, 1856-1862. doi:10.1016/j.jastp.2008.01.015

Nsumei, P., Reinisch, B.W., Huang, X., Bilitza, D. (2012) New Vary-Chap profile of the topside ionosphere electron density distribution for use with the IRI model and the GIRO real time data. Radio Sci., 47, RS0L16. doi:10.129/2012RS004989

Olea, R.A. (1974) Optimal contour mapping using universal kriging. J. Geophys. Res., 79 (5), 695-702

Olivares-Pulido, G., Hernandez-Pajares, M., Aragon-Angel, A., Garcia-Rigo, A. (2016) A linear scale height Chapman model supported by GNSS occultation measurements. J. Geophys. Res. Space Physics, 121. doi:10.1002/2016JA022337

Oliver, M.A., Webster, R. (1990) Kriging: A method of interpolation for geographical information systems. Int. J. Geogr. Inf. Syst., 4(3), 313-332

Oyeyemi, E.O., McKinnell, L.A., Poole, A.W.V. (2007) Neural network-based prediction techniques for global modeling of M(3000)F2 ionospheric parameter. Adv. Space Res., 39(5), 643-650. https://doi.org/10.1016/j.asr.2006.09.038.

Pezzopane, M., Scotto, C. (2005) The INGV software for the automatic scaling of foF2 and MUF(3000)F2 from ionograms: A performance comparison with ARTIST 4.01 from Rome data. J. Atm. Sol. Terr. Phys., 67(12), 1063-1073. doi:10.1016/j.jastp.2005.02.022

Pezzopane, M., Scotto, C. (2007) The Automatic Scaling of Critical Frequency *foF2* and *MUF(3000)F2*: a comparison between Autoscala and ARTIST 4.5 on Rome data. Radio Sci., 42, RS4003, doi:10.1029/2006RS003581.

Pezzopane, M., Scotto, C., Tomasik, Ł., Krasheninnikov, .I (2009) Autoscala: an Aid for Different Ionosondes. Acta Geophys., 58(3), 513-526. doi: 10.2478/s11600-009-0038-1

Pezzopane, M., Pietrella, M., Pignatelli, A., Zolesi, B., Cander, L.R. (2011) Assimilation of autoscaled data and regional and local ionospheric models as input sources for real-time 3-D International Reference Ionosphere modeling. Radio Sci., 46, RS5009. http://dx.doi.org/10.1029/2011RS004697

Pezzopane, M., Pietrella, M., Pignatelli, A., Zolesi, B., Cander, L.R. (2013) Testing the three-dimensional IRI-SIRMUP-P mapping of the ionosphere for disturbed periods. Adv. Space Res., 52(10), 1726-1736. doi: 10.1016/j.asr.2012.11.028

Pietrella, M., Pignalberi, A., Pezzopane, M., Pignatelli, A., Azzarone, A., Rizzi, R. (2018) A comparative study of ionospheric IRIEup and ISP assimilative models during some intense and severe geomagnetic storms. Adv. Space Res., 61(10), 2569-2584. https://doi.org/10.1016/j.asr.2018.02.026

Pignalberi, A., Pezzopane, M., Tozzi, R., De Michelis, P., Coco, I. (2016) Comparison between IRI and preliminary Swarm Langmuir probe measurements during the St. Patrick storm period. Earth, Planets and Space, 68(93). doi:10.1186/s40623-016-0466-5

Pignalberi, A., Pezzopane, M., Rizzi, R., Galkin, I. (2018) Effective solar indices for ionospheric modeling: A review and a proposal for a real-time regional IRI. Survey Geophys. 39(1), 125-167. doi:10.1007/s10712-017-9438-y

Pignalberi, A., Pezzopane, M., Rizzi, R. (2018) Modeling the lower part of the topside ionospheric vertical electron density profile over the European region by means of Swarm satellites data and IRI UP method. Space Weather, 16, 304–320. https://doi.org/10.1002/2017SW001790

Pignalberi, A., Pietrella, M., Pezzopane, M., Rizzi, R. (2018) Improvements and validation of the IRI UP method under moderate, strong, and severe geomagnetic storms. Earth, Planets and Space, 70(180). https://doi.org/10.1186/s40623-018-0952-z

Prolss, G. (2004) Physics of the Earth's space environment, an introduction. Springer-Verlag Berlin Heidelberg. doi: 10.1007/978-3-642-97123-5

Radicella, S.M., Leitinger, R. (2001) The evolution of the DGR approach to model electron density profiles. Adv. Space Res., 27(1), 35-40. doi:10.1016/S0273-1177(00)00138-1

Ramakrishnan, S., Rawer, K. (1972) Model electron density profiles obtained by empirical procedures. **S**pace Res. XII, Vol. 12, 1253-1259

Ratcliffe, J.A. (1962) The magneto-ionic theory and its application to the ionosphere. Cambridge University Press

Ratcliffe, J.A. (1972) An introduction to the ionosphere and magnetosphere. Cambridge University Press

Rawer, K. (1984) Modelling of Neutral and Ionized Atmospheres, Encyclopedia of Physics, Volume XLIX/7, pp. 223-533, Springer-Verlag, Berlin

Rawer, K. (1988) Synthesis of ionospheric electron density profiles with Epstein functions. Adv. Space. Res., 8 (4), 191-199. doi:10.1016/0273-1177(88)90239-6

Reinisch, B.W., Huang, X. (1983) Automatic calculation of electron density profiles from digital ionograms: 3. Processing of bottom side ionograms. Radio Sci., 18(3), 477-492. http://dx.doi.org/10.1029/RS018i003p00477

Reinisch, B.W., Huang, X., Belehaki, A., Shi, J.K., Zhang, M.L., Ilma, R. (2004) Modeling the IRI topside profile using scale heights from ground-based ionosonde measurements. Adv. Space Res., 34, 2026-2031. doi:10.1016/j.asr.2004.06.012

Reinisch, B.W., Nsumei, P., Huang, X., Bilitza, D.K. (2007) Modeling the F2 topside and plasmasphere for IRI using IMAGE/RPI and ISIS data. Adv. Space Res., 39(5), 731-738. doi: 10.1016/j.asr.2006.05.032

Reinisch, B.W., Galkin, I.A. (2011) Global Ionospheric Radio Observatory (GIRO). Earth Planets Space, 63, 377–381. doi:10.5047/eps.2011.03.001

Rishbeth, H., Garriott, O. (1969) Introduction to ionospheric physics. Academic Press, New York, International geophysics series v. 14

Roble, R.G., Ridley, E. C., Richmond, A. D., Dickinson, R. E. (1988) A coupled thermosphere / ionosphere general circulation model. Geophys. Res. Lett., 15, 1325-1328. https://doi.org/10.1029/GL015i012p01325

Rush, C.M., Fox, M., Bilitza, D., Davies, K., McNamara, L., Stewart, F.G., PoKempner, M. (1989) Ionospheric mapping – an update of foF2 coefficients. Telecommun. J., 56, 179-182

Samardjiev, T., Bradley, P.A., Cander, L., Dick, M.I. (1993) Ionospheric mapping by computer contouring techniques. Electron. Lett., 29(20), 1794-1795

Schreiner, W.S., Sokolovskiy, S.V., Rocken, C., Hunt, D.C. (1999) Analysis and validation of GPS/MET radio occultation data in the ionosphere. Radio Sci., 34(4), 949-966. doi: 10.1029/1999RS900034

Schunk, R.W. (1988) A Mathematical Model of the Middle and High Latitude Ionosphere, Pure Appl. Geophys., 127 , 255–303. doi:10.1007/BF00879813

Schunk, R.W., et al. (2004) Global Assimilation of Ionospheric Measurements (GAIM), Radio Sci., 39, RS1S02, doi:10.1029/2002RS002794

Scotto, C., Mosert de Gonzalez, M., Radicella, S.M., Zolesi, B. (1997) On the prediction of F1 ledge and critical frequency. Adv. Space Res., 209, 1773–1775. https://doi.org/10.1016/S0273-1177(97)00589-9

Scotto, C., Radicella, S.M., Zolesi, B. (1998) An improved probability function to predict the F1 layer occurrence and L condition. Radio Sci., 33, 1763–1765. doi: 10.1029/98RS02637

Scotto, C., Pezzopane, M. (2008) Removing multiple reflections from the F2 layer to improve Autoscala performance. J. Atmos. Sol. Terr. Phys, 70(15), 1929–1934. doi:10.1016/j.jastp.2008.05.012

Scotto, C. (2009) Electron Density profile calculation technique for Autoscala ionogram analysis. Adv. Space Res., 44,756-766. doi:10.1016/j.asr.2009.04.037

Scotto, C., Pezzopane, M., Zolesi, B. (2012) Estimating the vertical electron density profile from an ionogram: on the passage from true to virtual heights via the target function method. Radio Sci., 47, RS1007. http://dx.doi.org/10.1029/2011RS004833

Secan, J.A., Wilkinson, P.J. (1997) Statistical studies of an effective sunspot number. Radio. Sci., 32(4), 1717-1724. doi:10.1029/97RS01350

Seemala, G.K., Valladares, C.E. (2011) Statistics of total electron content depletions observed over the South American continent for the year 2008. Radio Sci., 46, RS5019. doi:10.129/2011RS004722

Shimazaki, T. (1955) World daily variability in the height of the maximum electron density of the ionospheric F2-layer. J. Radio Res. Lab. (Japan), 2, 85–97

Smith, N. (1939) The relation of radio sky-wave transmission ionosphere measurements. Preceedings of IRE, 27, 332-347

Shubin, V.N., Karpachev, A.T., Tsybuly, K.G. (2015) Global model of the F2 layer peak height for low solar activity based on GPS radio-occultation data. J. Atmos. Sol. Terr. Phys., 104, 106-115. doi:10.1016/j.jastp.2013.08.024.

Spalla, P., Ciraolo, L. (1994) TEC and *fo*F2 comparison. *Ann. Geophys.*, 37(5), 929-938. doi:10.4401/ag-4182

Somsikov, V.M. (2011) Solar Terminator and Dynamic Phenomena in the Atmosphere: A Review. *Geomagnetism and Aeronomy,* 51(6), 707–719. doi:10.1134/S0016793211060168

Ssessanga, N., Mckinnell, L.A., Habarulema, J.B. (2014) Estimation of *fo*F2 from GPS TEC over the South African region. J. Atm. Sol. Terr. Phys., 112, 20-30. doi:10.1016/j.jastp.2014.02.003

Stanislawska, I., Juchnikowski, G., Cander, L. (1996) The kriging method of ionospheric parameter foF2 instantaneous mapping. Ann. Geophys., 39(4), 845-852. doi:10.4401/ag-4007

Stankov, S.M., Jakowski, N., Heise, S., Muhtarov, P., Kutiev, I., Warnant, R. (2003) A new method for reconstruction of the vertical electron density distribution in the upper ionosphere and plasmasphere. J. Geophys. Res., 108 (A5), 1164. doi:10.1029/2002JA009570

Stankov, S.M., Jakowski, N. (2006) Topside ionospheric scale height analysis and modelling based on radio occultation measurements. J. Atm. Solar-Terr. Phys., 68, 134-162. doi:10.1016/j.jastp.2005.10.003

Sunehra, D., Satyanarayana, K., Viswanadh, C.S., Sarma, A.D. (2014). Estimation of Total Electron Content and Instrumental Biases of Low Latitude Global Positioning System Stations Using Kalman Filter. *IETE Journal of Research,* 56, 235-241. doi:10.4103/0377-2063.72772

Třísková, L., Truhlík, V., Šmilauer, J. (2002) On possible improvements of outer ionosphere ion composition model in IRI. Adv. Space Res., 29(6), 849-858, https://doi.org/10.1016/S0273-1177(02)00041-8

Třísková, L., Truhlík, V., Šmilauer, J. (2006) An empirical topside electron density model for calculation of absolute ion densities in IRI. Adv. Space. Res., 37(5), 928-934. doi:10.1016/j.asr.2005.09.013

Tsagouri, I., Zolesi, B., Belehaki, A., Cander, L. (2005) Evaluation of the performance of the real-time updated simplified ionospheric regional model for the European area. J. Atm. Sol. Terr. Phys., 67, 1137-1146. doi:10.1016/j.jastp.2005

Tulasi Ram, S., Su, S.-Y., Liu, C.H., Reinisch, B.W., McKinnell, L.-A. (2009) Topside ionospheric effective scale height ($H_T$) derived with ROCSAT-1 and ground-based ionosonde observations at equatorial and midlatitude stations. J. Geophys. Res., 114, A10309. doi:10.1029/2009JA014485

Venkatesh, K., Rama Rao, P.V.S., Saranya, P.L., Prasad, D.S.V.V.D., Niranjan, K. (2011) Vertical electron density and topside effective scale height ($H_T$) variations over the Indian equatorial and low latitude stations. Ann. Geophys., 29, 1861-1872. doi:10.5194/angeo-29-1861-2011

Venkatesh, K., Rama Rao, P.V.S., Fagundes, P.R. (2014) The role of altitudinal variation of scale height in determining the topside electron density profile over equatorial and low latitude sectors. J. Atm. Solar Terr. Phys., 121, 72-82. doi:10.1016/j.jastp.2014.10.006

Verhulst, T., Stankov, S.M. (2014) Evaluation of ionospheric profilers using topside sounding data. Radio Sci., 49, 181-195. doi:10.1002/2013RS005263

Verhulst, T., Stankov, S.M. (2015) Ionospheric specification with analytical profilers: Evidences of non-Chapman electron density distribution in the upper ionosphere. Adv. Space Res., 55, 2058-2069. doi:10.1016/j.asr.2014.10.017

Xu, T.L., Jin, H.L., Xu, X., Guo, P., Wang, Y.B., Ping, J.S. (2013) Statistic analysis of the ionospheric topside scale height based on COSMIC RO measurements. J. Atm. Solar Terr. Phys., 104, 29-38. doi: 10.1016/j.jastp.2013.07.012

Yue, X., Wan, W., Liu, L., Mao, T. (2007) Statistical analysis on spatial correlation of ionospheric day-to-day variability by using GPS and incoherent scatter radar observations. Ann. Geophys., 25, 1815-1825. doi:10.5194/angeo-25-1815-2007

Yue, X., Schreiner, W.S., Lei, J., Sokolovskiy, S.V., Rocken, C., Hunt, D.C., Kuo, Y.H. (2010) Error analysis of Abel retrieved electron density profiles from radio occultation measurements. Ann. Geophys., 28, 217-222. doi: 10.5194/angeo-28-217-2010

Wang, S., Huang, S., Fang, H. (2015a) Topside ionospheric Vary-Chap scale height retrieved from the COSMIC/FORMOSAT-3 data at midlatitudes. Adv. Space Res., 56, 893-899. doi:10.1016/j.asr.2015.04.021

Wang, S., Huang, S., Fang, H. (2015b) New method for deriving the topside ionospheric Vary-Chap scale height Radio Sci., 50, 866-875. doi:10.1002/2015RS005724

Wright, J.W. (1960) A model of the F region above $h_{max}F2$. J. Geophys. Res., 65(1), 185-191. doi: 10.1029/JZ065i001p00185

Wu, X., Hu, H., Gong, X., Zhang, X., Wang, X. (2009) Analysis of inversion errors of ionospheric radio occultation. GPS Solut., 13, 231-239. doi: /10.1007/s10291-008-0116-x

Wu, M.J., Guo, P., Fu, N.F., Xu, T.L., Xu, X.S., Jin, H.L., Hu, X.G. (2016) Topside correction of IRI by global modeling of ionospheric scale height using COSMIC radio occultation data. J. Geophys. Res. Space Physics, 121 5675-5692. doi:10.1002/2016JA022785

Zhang, M.L., Radicella, S.M., Kersley, I., Pulinets, S.A. (2002) Results of the modeling of the topside electron density profile using the Chapman and Epstein functions. Adv. Space Res., 29(6), 871-876. doi: 10.1016/S0273-1177(02)00053-4

Zhu, J., Zhao, B., Wan, W., Ning, B., Zhang, S. (2015) A new topside profiler based on Alouette/ISIS topside sounding. Adv. Space Res., 5, 2080-2090. doi:10.1016/j.asr.2015.08.008

Zolesi, B., Cander, L., De Franceschi, G. (1993) Simplified ionospheric regional model for telecommunication applications. Radio Sci., 28(4), 603-612. doi:10.1029/93RS00276

Zolesi, B., Belehaki, A., Tsagouri, I., Cander, L. (2004) Real-time updating of the Simplified Ionospheric Regional Model for operational applications. Radio Sci., 39, RS2011. doi:10.1029/2003RS002936

Zolesi, B., Cander, L. (2014) Ionospheric Prediction and Forecasting. Springer Geophysics

Zuccheretti, E., Tutone, G., Sciacca, U., Bianchi, C., Arokiasamy, B.J. (2003) The new AIS-INGV digital ionosonde. Ann. Geophys., 46(4), 647-659. http://dx.doi.org/10.4401/ag-4377

## Acknowledgements

Many people contributed to the successful drawing up of this PhD thesis that summarises three years of hard but satisfying work. Without the support of these people, my way would have been much more difficult.

My first thank goes to Prof. Rolando Rizzi who believed in me since the first moment. Without his support, this wonderful journey would never have started. He gave me more than I gave him. Thanks to the University of Bologna who economically supported me, and gave me the freedom of conducting my research without any constraint.

Thanks to Michael Pezzopane who everyday guides me in my professional growth. His constant support and help gives me the opportunity of growing as a scientist and as a person. Thanks to the Istituto Nazionale di Geofisica e Vulcanologia (INGV) who hosted and supported me during my PhD. Thanks to yhe INGV colleagues with whom I collaborated, and to all other people I have the pleasure of meeting everyday at INGV.

Thanks to Dr. John Bosco Habarulema and to the South African National Space Agency (SANSA) for giving me the opportunity of spending three wonderful months in Hermanus collaborating with brilliant scientists and, more importantly, fantastic people. That period was really fruitful and stimulating for me. Working with John Bosco has been a pleasure and an honor for me. Thanks to SANSA's staff and students that made me feel at home despite a hemisphere away.

Thanks to my family, to my relatives, and to all my friends who believe in me even without understanding anything of the work I do.

Last but not least, thanks to myself for believing in the success of all of this even in the most difficult moments.