ABSTRACT. Spatial and temporal variation of the electron density in the atmosphere makes the ionosphere a difficult region to model. A major difficulty arises due to the incomplete geometrical coverage of the Global Navigation Satellite System (GNSS) for tomographic applications, making the ionospheric tomographic system an ill-conditioned problem. Although the tomographic system is ill-conditioned, several techniques have been developed to partially overcome the problem. There is great interest in using tomographic techniques for ionospheric imaging, because it allows describing the ionosphere in terms of electron density, which is an important parameter for studying the behavior of the physical processes that occur in the upper atmosphere. In Brazil, there are additional interests in the tomographic techniques, due to the peculiar characteristics of the ionosphere and of the geomagnetic field over the region. In this direction, methods of ionospheric tomographic reconstruction are presented and discussed in this review. Particular emphasis is given to the mathematical formulation from grid-based and function-based methods and are presented some of their main advantages and limitations.

Keywords: TEC, Inverse Problem, Grid-Based Tomography, Function-Based Tomography, Ionospheric Imaging.

RESUMO. A variação espaço-temporal da densidade eletrônica na atmosfera terrestre torna a ionosfera uma região de difícil modelagem. A principal dificuldade no imageamento da ionosfera com o GNSS (Global Navigation Satellite System) é devido à geometria dos satélites, pois torna o sistema tomográfico mal condicionado. Muitos embora a geometria não permita solução direta do sistema, diversas técnicas foram desenvolvidas para parcialmente superar tal problema. Há grande interesse no uso de técnicas de tomografia para o imageamento da ionosfera, pois permitem descrever a ionosfera ao nível da densidade eletrônica, sendo este um importante parâmetro para compreender os processos físicos que ocorrem na alta atmosfera. No Brasil, existem interesses adicionais no imageamento ionosférico por meio de técnicas tomográficas devido às características peculiares da ionosfera e do campo geomagnético sobre a região. Neste sentido, métodos utilizados para a reconstrução tomográfica da ionosfera são apresentados e discutidos nesta revisão. Uma ênfase especial é dada para a formulação matemática dos métodos baseados em células e em funções, além da apresentação de algumas de suas principais vantagens e limitações.

Palavras-chave: TEC, Problema Inverso, Tomografia Baseada em Células, Tomografia Baseada em Funções, Imageamento Ionosférico.
INTRODUCTION

Ionospheric imaging is an important technique for understanding the space weather behavior. The resulting images are interesting from a scientific point of view, since the ionosphere is a region where many studies related to physical and chemical processes in the upper atmosphere are performed (Kelley, 2009). In the technological scenario, the density of ions and free electrons is enough to affect the electromagnetic waves propagation. Therefore, the ionosphere plays an important role in telecommunication systems because it acts as a reflective surface for Very Low Frequencies (VLF) to High Frequencies (HF). In the case of higher frequencies, such as VHF (Very High Frequency) and UHF (Ultra High Frequency), the ionosphere causes refraction and/or diffraction of the signals, which can introduce high levels of noise and telecommunication failures. More details about introductory texts about the ionosphere are given by Davies (1990).

Terrestrial and space-borne sounders that emit Radio Frequency (RF) pulses in direction to the atmosphere are the main equipment for describing the ionosphere. Usually, the ionospheric variations are measured by equipments specifically constructed to observe it, such as the ionosonde and incoherent scatter radars. But in recent years, the Global Navigation Satellite System (GNSS) has been widely used as an effective technology to represent the ionosphere because it is possible to combine the ionospheric refraction in the GNSS signals and integrate the electron density present in the signal path, in other words, the Total Electron Content (TEC).

Many models were developed using GNSS observations to represent the ionosphere in maps of Vertical TEC (VTEC) values (Camargo, 2009; Hernández-Pajares et al., 2009; Mitch et al., 2013). In those maps, errors occur because the ionosphere is considered as a thin shell with constant height (Brunini et al., 2011). However, this assumption is only an approximation to the reality and restricts the GNSS data for applications that are applied to observe it, such as the Ionospheric Tomography System (ISM) identifying of the ionospheric variability causing scintillations. In addition, Pokhrel et al. (2011) indicate that the GPS tomography correctly predicted the time of arrival and passage of an ionospheric tongue of ionization over the magnetic pole area.

In Brazil, few studies have been using ionospheric tomographic techniques (Muella et al., 2011). Therefore, more research may contribute with new perspectives to science and technology. Also, Brazil is under some peculiar ionospheric characteristics, as it is affected by ionospheric plasma bubbles, and by the Equatorial Ionization Anomaly (EIA), and because it is close to the South America Magnetic Anomaly (SAMA) center. For these reasons, methods used for ionospheric tomography are presented in this review, providing an overview about the mathematical formulation of grid-based and function-based algorithms and some of their main advantages and limitations.

IONOSPHERIC TOMOGRAPHY USING GNSS

Extracting Ionospheric Information from GNSS

Ionospheric information may be retrieved from code or phase measurements of GNSS. The observed distance from the code (P), is the required time to align the code generated in the receiver with the code received from the satellite, multiplied by the speed of light. The non-synchronization between receiver and satellite time systems introduces a systematic error (clock error) in the distance and, therefore, it can be identified as a pseudo distance (pseudo-range). On the other hand, the observed phase (ϕ) is equal to the sum of its fractional part and the integer number of cycles counted from the beginning of the tracking. The integer number of cycles between antennae of the receiver and the satellite is unknown at the beginning of the tracking and is called ambiguity, making the phase an ambiguous term. The main systematic effects (in meters) of these two observables are presented by Eqs. (1) and (2) (Seeber, 2003; Monico, 2008).

\[ P_L = \rho + c(\tau_L - \tau_s) + T - I_{L_s} + dm_{L_s} + \varepsilon, \quad [m] \] (1)

\[ \phi_L = \rho + c(\tau_s - \tau_s) + T - I_{L_s} + dm_{L_s} + \lambda_{L_s}N_{L_s} + \varepsilon, \quad [m] \] (2)
The frequency terms of L-band from GNSS are identified by the subscript \((L_i)\), that represents the carrier number, \(p\) is the geometric distance between the satellite and the receiver in meters (m); \(c\) is the speed of the light (m/s); \(\lambda\) is the wavelength (m); \(\tau_r\) and \(\tau_s\) are the receiver \((r)\) and satellite \((s)\) clock errors with respect to the GNSS system time in the respectively instant in seconds (s); \(T\) is the tropospheric delay (m); \(I\) is the ionospheric delay (m); \(dm\) is the multipath (m); \(N\) is the ambiguity (in cycles), which represents the integer number of cycles between the receiver and the satellite in the beginning of the tracking; and \(e\) (m) indicates the thermal and instrumental noise of the equipment. More details about noise in GPS receivers are presented by Langley (1997).

The ionospheric delay is positive for the code, being related to a group delay, and the negative sign in the carrier-phase represents a phase advance. Thus, the ionospheric refraction effect to a group delay, and the negative sign in the carrier-phase representation causes delays in the order of a few millimeters, as can be seen in the experiments from Marques et al. (2011).

Due to some approximations in the refraction index to obtain Eq. (3), described by Hartree-Appleton, the ionospheric delay, used in most of the ionospheric models, considers only the first order effect. The first order effect accounts for the electron density within the ionosphere, while the effect of the geomagnetic field and its interactions with the ionosphere are considered in the higher order terms. The higher orders terms are commonly neglected in ionospheric modeling using GNSS because the second order errors are in the range of a few centimeters, and the third order causes delays in the order of a few millimeters, as can be seen in the experiments from Marques et al. (2011).

Since the ionospheric delay is inversely proportional to the squared frequency, the ionosphere impact is not eliminated when a geometry-free linear combination is employed. In this way, such combination can be used to estimate the ionospheric delay. For GPS, the geometry-free combination between \(L_1\) (1,5754 GHz) and \(L_2\) (1,2276 GHz) from code is represented as:

\[
(P_{L_2} - P_{L_1}) = (I_{L_2} - I_{L_1}) + c(\Delta b^\prime + \Delta b_\rho) + \Delta e. \quad [m]
\]

In this linear combination, additional terms from the hardware delay, named Differential Code Bias (DCB), affect the satellite \((\Delta b^\prime)\) and the receiver \((\Delta b_\rho)\), as they are also dependent on the signal frequency. In Eq. (5), DCBs are expressed in seconds and by replacing Eq. (3) in Eq. (5), it is possible to obtain the relationship between the GNSS observations and TEC:

\[
F(P_{L_2} - P_{L_1}) = \mathrm{TEC} + F[c\Delta b_\rho + c\Delta b^\prime + \Delta e], \quad [\frac{cm}{m}]
\]

Equation (6) is the main expression for TEC estimation with code observations. Beyond the direct use of raw measurement, it is possible either to minimize the code noise by using filters, such as the Hatch filter (Hatch, 1983), which uses the carrier phase to smooth the pseudorange, or to use the phase observations by applying some method to fix the ambiguities, such as the leveled carrier-phase technique (Ciraolo et al., 2007). Furthermore, bearing in mind that the code can be demodulated by using different techniques, a bias correction is normally applied. For example, IGS (International GNSS Service) recommends applying the C1-P1 correction (Nacho, 2008) to make the civil code (C/A – Coarse/Acquisition) consistent with the precise code (P).

DCB may be estimated in the TEC adjustment process, but is necessary to impose constraints in the equation system. Camargo (2009), for example, imposed constraints to the DCB with respect to one network receiver or one satellite. IGS usually imposes a zero-mean condition, where the mean of all satellites’ DCB is equal to zero. Also, DCB may be estimated by taking ionospheric maps and computing average values over several hours (Prol & Camargo, 2014).

**Ionospheric Tomographic Inverse Problem**

Computerized Tomography (CT) is a classical example of a Digital Signal Process (DSP). CT is based on signals from many directions that pass through an object. Instead of simply forming an
image, the signals are converted into digital data and stored in a computer. Each signal trajectory is projected on a set of line integrals that pass through the object. The projections are then used to calculate tomographic images (virtual slices) by solving an inverse problem. Figure 1 illustrates an example of an acquired set of views to reconstruct the corresponding image, where the reconstructed image depends on how many views passed through the object (Smith, 1999).

Figure 1 – Set of views acquired by CT and reconstructed images (adapted from Smith, 1999).

Ionospheric tomography using TEC was first proposed by researchers at the University of Illinois (United States). From these initial studies, Austen et al. (1988) presented simulated results and identified the technique as Computerized Ionospheric Tomography (CIT). In the CIT algorithm, electron density is the parameter to be reconstructed and TEC corresponds to the line that slices the ionosphere into projections. CIT has the advantage that GNSS signals can be approximated as straight-line integrals. On the other hand, the reconstruction is compromised by incomplete geometrical coverage (Bust & Mitchell, 2008). Figure 2 illustrates the CIT geometry used with receivers located on the Earth's surface.

Figure 2 – CIT geometry used with receivers located on the Earth's surface.

In comparison to Figure 1, a reduced number of different inclinations is observed, mainly due to the absence of signals in the horizontal direction. Radio Occultation can be used to increase the number of inclined signals (Jakowski et al., 2004) but special attention is given to CIT using the receivers located on the Earth's surface in this study.

In order to derive the electron density, CIT approximates the ionosphere as a set of pixels or voxels. Taking \( n_{ij} \) as the electron density for a voxel \( (j) \) and \( d_{ij} \) the path length by the GNSS signal \( (i) \) inside the boundaries that intersect the voxel \( (j) \), an example of a three-dimensional grid of voxels is presented in Figure 3. When a voxel is sectioned by a signal, it is called illuminated (gray voxels) but, in general, many voxels are not illuminated. Also, to find an optimum voxel resolution which supports a better reconstruction of the electron density, experimental analysis may be required performing the tomographic reconstruction and checking the best computational performance for distinct voxel sizes. Das & Shukla (2011), for example, found from experimental analysis that the optimum pixel size is \( 5^\circ \times 50 \) km (latitude × altitude) over the Indian region using grid-based algorithms.

Figure 3 – Ionosphere stratified into a grid of voxels.

In this example, TEC may be approximated to a finite sum of voxels:

\[
\text{TEC}_i = d_{i1}n_{1} + d_{i2}n_{2} + d_{i3}n_{3} + d_{i4}n_{4} + d_{i5}n_{5} + d_{i6}n_{6} + d_{i7}n_{7} + d_{i8}n_{8} + d_{i9}n_{9}, \quad [\text{el/m}^2]
\]

or:

\[
\text{TEC}_i = \sum_{j=1}^{J} n_{ij}d_{ij}, \quad [\text{el/m}^2]
\]

where \( j \) ranges from 1 to \( J \) (number of voxels into the grid) and \( d_{ij} = 0 \) if the signal does not intercept the corresponding voxel.

Distance \( (d_{ij}) \) may be calculated from the projection of an IPP (Ionospheric Pierce Point) at each point where the signal transverses a grid cell (Shukla et al., 2010). Figure 4 illustrates the IPP's projection into a model that considers the ionosphere represented by voxels.

Once the receiver position \((\phi_r, \lambda_r)\), the azimuth \((\alpha)\) and elevation \((\epsilon)\) satellite angle are known, the geographic IPP
coordinates \((\varphi_{gi}, \lambda_{gi})\) depend on the IPP height \((h_{ip})\), as can be seen in the following formulation (Skone, 2002):

\[
\varphi_{gi} = \varphi + \psi \cos(az), \quad \text{[rad]} \tag{10}
\]

and

\[
\lambda_{gi} = \lambda + \psi \sin(az) \cos(\varphi_{gi}), \quad \text{[rad]} \tag{11}
\]

with:

\[
\psi = \cos^{-1} \left( \frac{R_m}{R_m + h_{ip}} \cos(cl) \right) - cl, \quad \text{[rad]} \tag{12}
\]

where the IPP heights can be estimated since they coincide with the known limits from the voxels and \(R_m\) in meters is the average radius of the Earth. The geometric distances between successive IPPs are used then as a voxel distance \((d_{ij})\) approximation in meters.

\[
\begin{bmatrix}
    \text{TEC}_1 \\
    \text{TEC}_2 \\
    \text{TEC}_i
\end{bmatrix}_{\text{m}} =
\begin{bmatrix}
    d_{11} & d_{12} & \cdots & d_{1j} \\
    d_{21} & d_{22} & \cdots & d_{2j} \\
    \vdots & \vdots & \ddots & \vdots \\
    d_{11} & d_{12} & \cdots & d_{ij}
\end{bmatrix}_{m \times j}
\begin{bmatrix}
    n_{e_1} \\
    n_{e_2} \\
    \vdots \\
    n_{e_j}
\end{bmatrix}_{j}

+ \begin{bmatrix}
    v_1 \\
    v_2 \\
    \vdots \\
    v_{ij}
\end{bmatrix}_{j} \tag{14}
\]

where \(v\) is a residual value of \(V\).

The parameter \(X\) can be solved by an estimation process, but the reduced numbers of GNSS inclinations often make this inverse problem ill-conditioned. In general, this problem also involves more parameters than observations because there are some non-illuminated voxels or voxels illuminated by only one signal, or by very few signals. However, the model allows the imposing of electron density information directly on the adjustment. Wen et al. (2012), for example, imposed a smoothing constraint to take advantage of the fact that the electron densities of neighboring voxels vary smoothly across the ionospheric tomography system. The number of parameters is directly proportional to the number of voxels and considering the ionospheric dimensions and its time dependence, in general, this is a method with intense computational effort, typically involving several parameters.

Many techniques have been developed to overcome these difficulties. For practical convenience, this large number of different techniques can be organized into two categories: grid-based and function-based. Some of their main characteristics may be then outlined.

As a result, a functional model that relates TEC and \(n_e\) is created to represent the inverse problem as a linear system:

\[
L = AX + V, \tag{13}
\]

where \(L\) corresponds to the observation vector having scalar values of TEC, \(X\) vector stands for the electron density parameters in each voxel, \(A\) is the design matrix composed by \(d_{ij}\) elements and \(V\) is the residual vector. In this way:

GRID-BASED IONOSPHERIC TOMOGRAPHY

In the grid-based technique, the information missing due to the GNSS geometry is overcome using a starting point known as a background ionosphere. Grid-based uses algebraic techniques to distribute the difference between TEC calculated from back-ground and TEC from GNSS onto the resulting image. So the CIT grid-based is, usually, a two-stage method, involving the selection of the initial condition followed by an algebraic algorithm.

Background Ionosphere

The background is usually obtained from empirical models of the ionosphere, such as the International Reference Ionosphere (IRI) model, to fulfill each cell from the CIT grid. This process may have many parameters, especially when considering that many backgrounds are constructed in a daily period or also in every instant that the GNSS measurements are recorded. Therefore, some methods are applied to reduce the number of parameters and hence the computational efforts.

One way to reduce this set of parameters is by defining some discrete backgrounds for a time interval and imposing a tempo-
temporal variations are calculated using only the cells where the GNSS signals were projected in the CIT grid.

In this way, the background construction is generally based on empirical models, representing the ionosphere by assumption. However, the grid-based technique is highly dependent on the background. This mainly occurs because the algebraic technique does not couple the changes applied to the voxels that may have no intersections with signal projections. So the background ionosphere must be defined with special concern. An example of high attention is given by the method developed by Mitchell et al. (1997), where a great number of ionospheric backgrounds are created at a given instant, and the vertical information of the ionosphere is represented by many possible models. The algebraic technique is then processed for all these backgrounds and a numerical selection criterion is employed to define the image that best fits the measurements.

The algebraic technique application is required because some ionospheric structures may not be represented in the background. Even when the background is constructed by the incorporation of ionosonde data, the information is restricted to the ionosonde geographic location. Thus, the algebraic algorithms are employed to include GNSS information in the tomographic system and to produce ionospheric images using the TEC measurements directly on the grid.

It is important to mention that the GNSS bias (related to the DCB) may be calculated before the background construction, estimated together with a background modeling (by imposing constraints on the receiver or satellite), or even with the iterative process. Hobiger et al. (2008) mentions that, to estimate the DCB during the iterations, they are treated as “virtual” cells and each projection length is imposed equal to one. Therefore, in the example of the zero-mean condition, the mean of the artificial observations of the satellite virtual grid is considered equal to zero.

It must also be considered that TEC observations from GNSS are related to the electron content along the path between the receiver and satellite. By contrast, the background ionosphere is related to some portion of atmospheric heights. So TEC needs to be corrected by a value that corresponds to the electron density over the upper or bottom grid limit. The height limits of the ionosphere are not well defined, however, the atmosphere is subject to ionization mainly in heights between 60 km and 1000 km above the Earth’s surface. In general, vertical functions are used to estimate this electron density from the outside of these heights. For example, Materassi & Mitchell (2005) applied a Chapman function to correct TEC from heights above 1000 km.

### Algebraic Techniques

The Algebraic Reconstruction Technique (ART), first technique used in CT, was simultaneously proposed by Gordon et al. (1970) and Hounsfield (1972) as an iterative algorithm derived from the Kaczmarz (1937) method. ART assumes initial conditions of the image to be reconstructed, where a set of lines are projected in this initial image. The integral values of these projections are known and are compared with the initial image. The differences between the measured value and the initial values from the projections are distributed into the grid, with a known weight value. This process is repeated for a specified number of iterations, that successively corrects the initial image.

To apply ART for CIT, an initial image is defined by ionospheric background concepts. In the cell grid, two projections are constructed for each GNSS signal (\( i \)). The first is the measured TEC from GNSS and the second represents TEC calculated from the background. The electron density in each cell (j) is then calculated with the following expression:

\[
n_{ij}^{K+1} = n_{ij}^K + \frac{\bigg(T_{EC} - \sum_{j=1}^J d_{ij} n_{ij}^K\bigg)}{\sum_{j=1}^J d_{ij}^2} d_{ij} \quad \text{[eq. 15]}
\]

where \( n_{ij}^{K+1} \) is the electron density value obtained from iteration \((K + 1)\) and \( w \) is a weighting parameter used to control the convergence of the algorithm with \( 0 < w < 1 \) (Bust & Mitchell, 2008). The weighting parameter is empirically determined considering the relative precision of GNSS TEC observations, which may be determined by statistical analysis, such as in the Improved ART (IART) algorithm described by Wen et al. (2007). The \( \sum_{j=1}^J d_{ij} n_{ij}^K \) term is equivalent to a scan through each cell from the ionospheric grid that calculates the TEC values of the correspondent background projection.

In ionospheric imaging, different versions from ART are commonly applied, such as the Simultaneous Algebraic Reconstruction Technique (SART) and the Multiplicative Algebraic Reconstruction Technique (MART). While the original algorithm (ART) is based on an analysis of each projection, SART performs the iteration after all the differences between the projections (initial and measured) are computed (Anderson & Kak, 1984). By contrast, the multiplicative version (MART) differs in how the differ-
ence errors are distributed in the image. The MART algorithm is based on the following nonlinear iteration (Pryse et al., 1998):

\[
r_{e}^{K+1} = n_{e}^{K} \left( \sum_{i=1}^{J} d_{i} n_{e}^{K} \right)^{-1/2} \left( \sum_{j=1}^{N} \frac{TEC_{j}}{d_{i} n_{e}^{K}} \right)^{-1} \frac{d_{i}}{d_{\text{max}}} \quad \text{[el/m}^{2}]\]

(16)

where \(d_{\text{max}}\) is the largest path length of the respective projection. Even when \(\sum_{j=1}^{N} d_{i} n_{e}^{K} > TEC_{i}\), MART produces non-negative values, so an advantage to using MART, instead of ART or SART, is the guarantee of non-negative values for the reconstructed image. An interesting experimental comparison between ART and MART algorithms is presented by Das & Shukla (2011), showing that MART presented better results for the Indian region.

Many iterative algorithms are being constantly developed to improve the algebraic techniques. However, most of them are based on the same principles of ART, SART or MART (Hobiger et al., 2008; Wen & Liu, 2010; Wen et al., 2012). Iterations of the algebraic techniques can be performed until the instant when the iteration error is below a certain threshold. The error of the iteration (\(\epsilon^{K}\)) is given by:

\[
\epsilon^{K} = \frac{\sum_{i=1}^{n_{i}} \left( TEC_{i} - \sum_{j=1}^{J} d_{i} n_{e}^{K} \right)^{2}}{\sum_{i=1}^{n_{i}} TEC_{i}}^{1/2} \quad \text{[el/m}^{2}]\]

(17)

where for the first iteration, an initial electron density for each cell is necessary and such information is obtained from the ionospheric background.

FUNCTION-BASED IONOSPHERIC TOMOGRAPHY

In function-based techniques, the electron density is represented by coefficients that define the horizontal, vertical or three-dimensional surface. When using only horizontal functions, the ionosphere is represented by conventional TEC maps. In the case of vertical functions, empirical (or semi-empirical) models are used to define the ionospheric profile for a given position. Combining both these characteristics, the TEC measurements are directly used to reconstruct a three-dimensional ionospheric surface and produce tomographic images. The temporal dimension is usually described by linear imposition or also using specific constraints, such as Van de Kamp (2013) that imposed conditions by using the differential TEC.

Horizontal Functions

The horizontal function considers that the ionosphere is represented by a projection of TEC into a vertical point of the ionosphere (IPP). This point is formed by the intersection of the receiver/satellite vector and a thin-shell layer that is generally located between 300 km and 450 km above the Earth’s surface. The VTEC projection is commonly related to mapping functions and the standard geometric mapping function is expressed as (Hofmann-Wellenhof et al., 2001):

\[
VTEC = \cos (z^{\text{ip}}) TEC \quad \text{[el/m}^{2}]\]

(18)

and

\[
\sin (z^{\text{ip}}) = \frac{R_{m}}{R_{m} + h_{m} \sin (z_{r})}, \quad \text{(19)}
\]

where \(z_{r}\) is the zenithal angle of the path between the receiver and the satellite, \(z^{\text{ip}}\) is the zenithal angle between the IPP and the satellite and \(R_{m}\) is the average radius of the Earth in meters. Figure 5 illustrates the IPP projection geometry.

**Figure 5** – Standard geometric mapping function geometry to project an IPP (adapted from Hofmann-Wellenhof et al., 2001).

Substituting Eq. (18) in (6), the main observation equation for modeling the ionosphere with VTEC and code measurements is obtained:

\[
F \left( \frac{P_{T,1} - P_{B,1}}{P_{B,1}} \right) = VTEC \cos (z^{\text{ip}}) + F \left[ c \Delta b_{e} + c \Delta b_{e} + \Delta \varepsilon \right] \quad \text{[el/m}^{2}]\]

(20)

Due to the periodic nature of ionospheric refraction, the VTEC is represented by a mathematical surface function, such as series, polynomials and spherical harmonics. The surface function choice must consider the coverage area, typically, classifying the VTEC models as global, regional or local.
For global modeling, data from worldwide GNSS network are used to estimate VTEC values and produce Global Ionospheric Maps (GIM). GIMs are daily produced by the IGS Ionosphere Working Group (IonoWG) analysis centers, where the IGS IonoWG is composed of four centers: CODE (Centre for Orbit Determination in Europe), JPL (Jet Propulsion Laboratory), UPC (Universitat Politècnica de Catalunya) and ESA (European Space Agency) (Hernández-Pajares et al., 2009). Among these centers, CODE developed the method available in the Bernese software, in which VTEC is modeled by the following spherical harmonic expansion (Schaer, 1999):

\[
VTEC = \sum_{n=0}^{n_{\text{max}}} \sum_{m=-n}^{n} \tilde{P}_m(\sin(\varphi_{ip})) \left[ a_{nm} \cos(m \lambda_{ip}^\beta) + b_{nm} \sin(m \lambda_{ip}^\beta) \right] \left[ \text{e} \text{l/m}^3 \right]
\]

with:

\[
\lambda_{ip}^\beta = \lambda_{ip}^\alpha - \lambda_s \quad \text{[rad]}
\]

where \(\varphi_{ip}\) is the IPP geomagnetic latitude, \(\lambda_{ip}\) is the IPP geographic longitude, \(\lambda_s\) is the longitude of the Sun \((\lambda_s = \pi - \pi \frac{224}{s_e} t \text{[UT]}\), \(n_{\text{max}}\) is the maximum degree of the expansion, \(a\) and \(b\) are the spherical harmonic coefficients to be estimated in units of \(\text{e} \text{l/m}^3\) and \(\tilde{P}_m(x)\) is the normalized associated Legendre function of degree \(n\) and order \(m\). The number of parameters is calculated by \((n_{\text{max}} + 1)^2\) and the series is expanded to 15 degrees to do the global modeling. In addition, the temporal evolution of the VTEC is modeled by imposing linear conditions in the spherical harmonics coefficients. Otherwise, some different methods can be applied to produce GIMs. UPC uses tomographic techniques (Hernández-Pajares et al., 1999), ESA performs a three-dimensional modeling based on the Chapman functions (Feltens, 1998), and JPL represents the ionosphere as a spherical grid from bilinear spline functions (Mannucci et al., 1998). Finally, the IonoWG combines the results of the four agencies, and produces the IGS final maps.

On another scale, the regional term can be used when the ionospheric modeling is held in a specific coverage area (e.g., South America), and aims to detect regional VTEC structures. Several strategies are adopted to represent VTEC regionally, like interpolation kriging methods (Hernandez-Pajares et al., 1999), trigonometric series (Brunini et al., 2004) and Fourier series (Camargo, 2009). In local cases, the same regional concepts are applicable but within a small area of one or more GNSS stations. More details about local models are presented by Mitch et al. (2013).

**Vertical Functions**

Vertical functions are used to describe the electron density profile distribution at any given position and, usually, the functions are represented by mathematical functions obtained from measured data. Among various existing vertical functions, Empirical Orthogonal Functions (EOF), IRI and the Chapman model are widely applied in ionospheric tomography and particular attention is given to them.

**Chapman Model**

The Chapman model is characterized by a balance between the ionization and recombination process of gases in the ionosphere. The method was deduced from the following relation: while neutral gases have exponential decreasing with altitude, ionized gases have exponential increase (Chapman, 1931).

According to the model, Chapman-\(\alpha\) function (quadratic loss) is adapted to describe low altitudes in the ionosphere, while the Chapman-\(\beta\) function (linear loss) characterizes higher altitudes in the ionosphere. The electron density at an altitude \(h\) is represented by Chapman-\(\alpha\) with \(\gamma = 0.5\) and Chapman-\(\beta\) with \(\gamma = 1\), as follows:

\[
n_e = n_m \text{e}^{\gamma \left[1 - z - z^{-\gamma}\right]} \quad [\text{e} \text{l/m}^3]
\]

with:

\[
z = \frac{(h - h_m)}{H_s}
\]

where \(n_m\) is the critical electron density, \(h_m\) (m) is the altitude where \(n_m\) occurs and \(H_s\) (m) is the scale height, which is an essential parameter in the model because it describes the profile shape. Despite the two Chapman functions, some observations of vertical profiles showed that the Chapman-\(\alpha\) is sufficient to describe the electron density at all the altitudes of the ionosphere (Mei & Wan, 2008). Besides that, a variation of Chapman-\(\alpha\) function has been evaluated by IRI as a new representative model of the topside electron density distribution. Rishbeth & Garrriott (1969) first proposed this function and it describes the ionosphere topside using a Chapman function with continuously varying scale height. It is, therefore, called the Vary–Chap function (Bilitza et al., 2011).

**International Reference Ionosphere**

The IRI model was developed and continues to be improved by the Committee on Space Research (COSPAR), by the International Union of Radio Science (URSI) and with cooperation from members of various nations. The main intention is to obtain components that describe the ionosphere, such as electron density,
electron temperature, ion temperature, ion composition (O+, H+, He+, NO+, O2+), ion drift, Spread-F probability and TEC. IRI is an empirical model based on several terrestrial and spatial data sources and the electron density is represented by analytic functions, whose parameters are organized in tables (Bilitza et al., 2011).

The IRI online platform is available through the web site: http://iri.gsfc.nasa.gov, where the algorithm source code can be found. Solar parameters such as the 12-month-running mean sunspot number (R12), the solar radiation flux per unit frequency found. Solar parameters such as the 12-month-running mean sunspot number (R12) and magnetic indices (Ap and Kp) are input parameters. In addition, global ionospheric index based on ionosonde data (IG12) and the Solar Radiation Flux per unit frequency (foF2) and the ratio of the Maximum Usable Frequency (MUF) at a distance of 3000 km to foF2 (M(3000) F2).

Currently, IRI2012 is the latest version and represents the global behavior of the ionosphere, such as the F2 layer critical frequency (foF2) and the ratio of the Maximum Usable Frequency (MUF) at a distance of 3000 km to foF2 (M(3000) F2).

Empirical Orthogonal Functions

Principal Components Analysis is a statistical technique widely used in Meteorology. Lorenz (1956) made PCA popularized for atmospheric analysis and renamed it EOF analysis. His approach was to capture the main spatial patterns variability from data and expand the data in time-dependent functions. In fact, the EOF analysis transforms a large number of data sets into a smaller number of coefficients to find a set of orthogonal spatial patterns along with a set of associated uncorrelated time series (Hannachi, 2004).

Assuming a sample data set that represents the electron density profiles obtained at different times (t₁, t₂, ..., t₂₆) and heights (h₁, h₂, ..., h₂₆), a data matrix can be formulated as (Amerian et al., 2010):

\[ \mathbf{Ne} = \mathbf{EQ} \]

The \( \mathbf{E} \) matrix describes the electron density altitude variation and \( \mathbf{Q} \) is the time dependent matrix. The original data matrix (\( \mathbf{Ne} \)) may be constructed by other vertical functions (e.g., IRI) or by equipment that provides observation of the ionospheric profile, such as an ionosonde. From these observations, it is necessary to determine the basis functions (\( \mathbf{E} \)) and the temporal coefficients (\( \mathbf{Q} \)).

In the EOF estimation, a datum pre-processing is performed to remove the less important trends. This can be done by calculating the mean value of each \( \mathbf{Ne} \) column and subtracting each average from the original data values. The average value of each column is obtained by:

\[ \overline{\mathbf{Ne}}(h) = \frac{1}{Q} \sum_{q=1}^{Q} n_e(t_q, h), \quad [\text{el/m}^3] \]

and EOF can be expressed in the form (Lorenz, 1956):

\[ \mathbf{Ne} = \mathbf{EQ} \]

The \( \mathbf{E} \) matrix describes the electron density altitude variation and \( \mathbf{Q} \) is the time dependent matrix. The original data matrix (\( \mathbf{Ne} \)) may be constructed by other vertical functions (e.g., IRI) or by equipment that provides observation of the ionospheric profile, such as an ionosonde. From these observations, it is necessary to determine the basis functions (\( \mathbf{E} \)) and the temporal coefficients (\( \mathbf{Q} \)).

In the EOF estimation, a datum pre-processing is performed to remove the less important trends. This can be done by calculating the mean value of each \( \mathbf{Ne} \) column and subtracting each average from the original data values. The average value of each column is obtained by:

\[ \overline{\mathbf{Ne}}(h) = \frac{1}{Q} \sum_{q=1}^{Q} n_e(t_q, h), \quad [\text{el/m}^3] \]

and the anomaly field is then calculated from this mean vector:

\[ \overline{\mathbf{Ne}} = \begin{bmatrix} n_e(t_1, h_1) - \overline{n_e}(h_1) & \cdots & n_e(t_1, h_H) - \overline{n_e}(h_H) \\ \vdots & \ddots & \vdots \\ n_e(t_Q, h_1) - \overline{n_e}(h_1) & \cdots & n_e(t_Q, h_H) - \overline{n_e}(h_H) \end{bmatrix}, \quad [\text{el/m}^3] \]

EOF estimation is based on the covariance matrix (\( \mathbf{\Sigma_{Ne}} \)) of the anomaly matrix (\( \overline{\mathbf{Ne}} \)), that can be computed by the multiplication of \( \overline{\mathbf{Ne}}^T \) for its transpose (Björnsson & Venegas, 1997):

\[ \mathbf{\Sigma_{Ne}} = \overline{\mathbf{Ne}}^T \overline{\mathbf{Ne}}. \]
After the covariance matrix computation, it is necessary to solve the eigenvalues problem:

$$\Sigma_{N_e} E = E \Lambda,$$

(30)

where the eigenvectors ($E$) represents the EOF basis and have the following property:

$$E^T E = I,$$

(31)

which turns the EOF orthogonal in space, because $I$ denotes the identity matrix. Eigenvalues ($\Lambda$) are represented in a matrix where the non-diagonal elements tend to zero, making EOF orthogonal in time. Also, the projection of each eigenvector in the original field ($Ne$) is used to compute the temporal coefficients ($Q$).

It is a usual procedure to sort the eigenvectors in ascending order of importance. The first EOF basis ($E_1$) is the most important (larger eigenvalue), the second EOF ($E_2$) is the second larger and so on. An important concern is the definition of how many eigenvalues must be retained in order to separate the significant part. Therefore, the real dimension of EOF is given by a matrix $H x k_{max}$ where the number of used eigenvalues is $k_{max}$. Usually, the ionospheric tomographic reconstruction uses up to four modes (Materassi & Mitchell, 2005). So the electron density is represented as:

$$n_e(h, t) = \sum_{k=1}^{k_{max}} E_k(h) Q_k(t), \quad [\text{el/m}^3]$$

(32)

For the system equations construction, the electron density estimation may be parameterized in terms of spherical harmonics and vertical functions by the following expression (Bruini et al., 2004):

$$n_e = \sum_{n=0}^{n_{max}} \sum_{m=0}^{n} f(h) \tilde{P}_{nm}(\sin(\psi^h_m) \left[ a_{nm} \cos(m \lambda^h_m) + b_{nm} \sin(m \lambda^h_m) \right] \quad [\text{el/m}^3]$$

(35)

where $f(h)$ is the vertical function that describes the ionospheric profile without the EOF concepts.

Since the relationship between the electron density and the GNSS observations is performed by TEC, Eq. (35) needs to be integrated. This integration is done approximating TEC to a finite sum. Therefore, TEC is given by:

$$\text{TEC} = \sum_{n=0}^{n_{max}} \sum_{m=0}^{n} \{ f(h) \tilde{P}_{nm}(\sin(\psi^h_m)) \cos(m \lambda^h_m) \} \quad [\text{el/m}^2]$$

(36)

or

$$\text{TEC} = \sum_{n=0}^{n_{max}} \sum_{m=0}^{n} a_{nm} \int \{ f(h) \tilde{P}_{nm}(\sin(\psi^h_m)) \cos(m \lambda^h_m) \} \quad [\text{el/m}^2]$$

$$+ b_{nm} \int \{ f(h) \tilde{P}_{nm}(\sin(\psi^h_m)) \sin(m \lambda^h_m) \} \quad [\text{el/m}^2]$$

(37)

In this way, the matrix $A$ is defined through the path lengths within each voxel from the grid. $X$ is the derivative of the equation in relation to the coefficients and $W$ stands for the $a$ and $b$ coefficients. If a method is previously defined to calculate the vertical function
parameters, the vertical information can be directly used on the model given by Eq. (37) and the number of parameters is equal to the horizontal function modeling \((n_{\text{max}} + 1)^2\). For this application, a normalization process can be applied to the ionospheric profile, where the critical density at the correspondent position is set equal to one. In this normalization, the estimated coefficients will be related to the critical density and the temporal variation can be imposed linearly, as in the horizontal functions.

Otherwise, by using EOF, the three-dimensional function becomes (Howe et al., 1998):

\[
\text{TEC} = \sum_{k=1}^{n_{\text{max}}} \sum_{n=0}^{n_{\text{max}}} \sum_{m=0}^{n_{\text{max}}} a_{knm} \int_{r} \{ E_k(h) \hat{P}_{nm}(\sin(\phi_{\text{ip}})) \cos(m\lambda_{\text{ip}}) \} ds + b_{knm} \int_{r} \{ E_k(h) \hat{P}_{nm}(\sin(\phi_{\text{ip}})) \sin(m\lambda_{\text{ip}}) \} ds \quad [\text{el/m}^2]
\]

If EOF is previously estimated, the vertical information from EOF can be directly used in the model, imposing constraints related to the eigenvalues. The equation system has \([(n_{\text{max}} + 1)^2] \times k_{\text{max}}\) coefficients and when \(k_{\text{max}} > 1\) the number of parameters will be higher than the model presented in Eq. (37). Now the coefficients are related to the critical density for each EOF mode and the temporal dependencies may also be imposed as linear conditions (or by other gradient constraint).

However, instead of using vertical information directly on any of these models, the required vertical parameters can be estimated by imposing vertical constraints in the adjustment process. A detailed resolution of the systems equation is given by Mitchell & Spencer (2003), where Singular Value Decomposition is applied and constraints are imposed to EOF be assumed proportional to some other vertical function, as the Chapman model.

**EXPERIMENTAL RESULTS**

This section presents a real case example of ionospheric representations that can be obtained using tomographic reconstruction techniques in Brazil. Figures 6 and 7 show VTEC maps and ionospheric profiles constructed using MART algorithm developed in C/C++ language programming. These figures were obtained using the same algorithm and data configurations of Prol & Camargo (2015), but now representing the date October 16, 2014. In general, TEC was estimated using the leveled carrier-phase ionospheric observable (Ciracolo et al., 2007) and OCB was previously estimated using VTEC maps from IGS (Prol & Camargo, 2014). The background and the plasmapause correction were calculated using the IRI2012 model (Bilitza et al., 2011) and MART was applied on a voxel-based grid with a resolution of 4° × 4° × 10 km in latitude, longitude and height.

Although the representation of peak height and vertical drift is inherent to the background, we can see in Figures 6 and 7 that EIA intensification in the evening pre-reverse period becomes greater around 22 hours UT (Universal Time), which represents 19 hours LT (Local Time). In this period, VTEC values start to get lower on the magnetic equator while a major concentration of VTEC appears around 20°S in geographic latitude (15°W on magnetic latitude). At the same time, we see in Figure 8 (longitudinal sector of 50°W) the elevation of the ionosphere for between 22 and 00 hours UT, representing the vertical drift that occurs in the evening pre-reverse period. This vertical drift was represented mainly due to the IRI model (used as background), but MART played an important role in estimating the electron density by incorporating TEC information, which better defined the EIA southern crest. Furthermore, it is also possible to check that the E-layer and F1-layer disappear in the evening pre-reverse period. In this way, MART linked with the IRI model enables the analysis, not only of the EIA and its intensification in the evening pre-reverse peak, but also of the variations in the vertical gradients of electron density, which may be interesting to analyze in future works.

**CONCLUSIONS**

This study presented some methods used in ionospheric tomography, contextualizing a discussion about grid-based and function-based techniques. There are two main steps in the grid-based technique: the background ionosphere definition and the algebraic technique application to solve the problem of inverting TEC in electron density. Since the number of parameters is equal to the number of cells, the grid-based technique involves a large number of unknowns. However, this method enables the adding of information related to electron density directly in the reconstruction, which makes it appropriate for data assimilation from different technologies. In addition, measured TEC from GNSS is used directly in the grid image, making this method applicable for analyzing regional structures in the ionosphere. But some difficulties arise because the grid-based technique does not couple the changes applied to the voxels which may have no intersections with signal projections.
In function-based techniques, it was observed that horizontal and vertical functions reduce the ionospheric variations to a set of coefficients. The horizontal component can only be well represented using GNSS observations, as could be seen on the formulation to produce GIM. However, external information is necessary to represent the vertical profiles. It is possible to obtain vertical information directly from ionospheric models (such as IRI) or by reducing the vertical information into EOF. A more mathematical operation is required and larger numbers of parameters are used in the EOF estimation but it allows imposing vertical information from measured data (like from ionosonde).

It was noticed that several versions of tomographic ionospheric reconstruction could be formulated. For example, different techniques can be used to construct the background, also, many algebraic techniques can be employed to perform the iterations. Moreover, a wide variety of surface functions can be used to represent the ionosphere and different constraints can be imposed. But, in general, the reconstruction method mainly depends on vertical functions or on the background model because both techniques (function-based and grid-based) share a common limitation: a deficiency in representing ionospheric vertical variations using only GNSS.
It is possible to combine these two characteristics since the function-based technique can be used to define the whole background of the ionosphere, whereas the grid-based technique is used to improve the image locally. In a different way, Yao et al. (2013) used the grid-based technique to construct initial ionospheric conditions and imposed constraints on the EOF. Therefore, the problems due to the missing information in some voxels from grid-based techniques are overcome by the function-based techniques. In this way, based on this principles, a quantitative comparison between the mentioned algorithms in this work is a research topic and this application is possible for Brazil by considering the IGS network and the RBMC (Rede Brasileira de Monitoramento Continuo dos Sistemas GNSS) geometry.

ACKNOWLEDGEMENTS

This work was jointly founded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Grant: 2015/15027-7).

REFERENCES


BJÖRNSSON H & VENEGAS SA. 1997. A manual for EOF and SVD analysis of climatic data. Department of Atmospheric and Oceanic Sciences, Centre for Climate and Global Change Research, McGill University, Canada, 52 pp.


Prol FS & Camargo PO. 2015. Ionospheric tomography using GNSS: multiplicative algebraic reconstruction technique applied to the area of Brazil. GPS Solutions, Published First Online: 22 September 2015, doi:10.1007/s10291-015-0490-0.


NOTES ABOUT THE AUTHORS

Fabricio dos Santos Prol. Graduated (2011) in Cartographic Engineering from Universidade Estadual Paulista (UNESP) and received M.Sc. degree (2015) in Cartographic Sciences from UNESP. Ph.D. candidate at UNESP, and the focus of the current research lies in ionospheric modeling using data assimilation.

Paulo de Oliveira Camargo. Graduated (1985) in Cartographic Engineering from Universidade Estadual Paulista (UNESP) and received Master (1992) and Doctor (1999) degrees in Geodetic Sciences from the Universidade Federal do Paraná (UFPR). Professor of the graduate and postgraduate program at UNESP focusing on the following topics: GNSS, ionosphere, quality control, and least-squares adjustment.