Neural network modeling of the ionospheric electron content at global scale using GPS data

M. Hernández-Pajares, J. M. Juan, and J. Sanz
Research Group of Astronomy and Space Geodesy, Universitat Politècnica de Catalunya
Barcelona, Spain

Abstract. The adaptative classification of the rays received from a constellation of geodetic satellites (the Global Positioning System (GPS)) by a set of ground receivers is performed using neural networks. This strategy allows us to improve the reliability of reconstructing the ionospheric electron distribution from GPS data. As an example, we present the evolution at global scale of the radially integrated electron density (total electron content (TEC)) for October 18, 1995, coinciding with an important geomagnetic storm.

Introduction

Problem

As is well known, the ionosphere is the part of the Earth atmosphere that contains free ions, reaching a maximum density at typical heights of 300-400 km. It causes a frequency-dependent delay in the propagated electromagnetic signals, being proportional to the integrated density of electrons along the path \( I \), also called slant total electron content (STEC) [see, for example, Davies, 1990, page 73]. This is a distorting physical effect for space geodesy and satellite telecommunication activities that can be used in positive to estimate the global distribution of the free electrons in the atmosphere, from dual-frequency delay observations.

Data

To achieve this objective, we need during a certain time interval a high sampling rate of the atmosphere, with as many rays in as many orientations as possible. Nowadays, the unique system that provides so many observations, continuously and on a planetary scale, is the Global Positioning System (GPS). Its space segment contains a constellation with more than 24 satellites continuously emitting carrier and code phases in two frequencies: L1 (\( \approx1.6 \) GHz) and L2 (\( \approx1.2 \) GHz) [see, for example, Seeber, 1993, p. 209-349]. In the GPS user segment, it is possible to get a few hours later the public domain data gathered from a global network of permanent receivers, such as the International GPS Service for Geodynamics (IGS), [Zumberge et al., 1994], with more than 100 stations distributed worldwide, mainly concentrated in the northern hemisphere, in North America, and in Europe.

Model and Goals

The amount of data implied, however (more than one million delays collected each day in the IGS network), jointly with its inhomogeneous distribution, makes it difficult to solve the problem, so that new algorithms and strategies must be considered in order to perform the tomography of the ionosphere. Several approaches has been developed recently by some authors.

Sardon et al., [1994] estimates the total electron content (TEC) for each GPS receiver, in a local reference frame, using the Kalman filter to update the TEC. This strategy is adapted to the data, i.e., it provides more TEC estimations where more receivers are available. However, this method is not adequate for global TEC modeling using a high number of stations (50 or more) due to the very high computation load that introduces the use of Kalman filtering for each update epoch (each 2 min in that work).

Wilson et al., [1995] and Juan et al., [1997] avoid the high update frequency in the Kalman filtering by choosing an inertial reference frame. However, in
both works the orthogonal basis in which the TEC distribution is explained (spherical harmonics and regular cells (boxels), respectively) is not adapted to the data.

In this paper we mainly discuss the different data analysis problems encountered in the estimation of the ionospheric electron distribution using ground data from the IGS network and the solutions adopted, emphasizing the adaptative clustering of the rays using the Kohonen neural network algorithm. This technique will be applied for bidimensional modeling of the overall ionosphere, i.e., at global scale, in an inertial reference frame generating cells adapted in size to the variable sparsity of the data.

Model
Scenario
We have the following situation.

From each ground station we simultaneously measure with a certain sampling rate (i.e., one time per 30 s) the ionospheric delays experienced by the rays received from the visible satellites (i.e., 4-8). These rays cross different parts of the nearby ionosphere to the respective station. Between observation epochs the Earth rotates, and the part of the sounded ionosphere has changed. We assume that the ionosphere is stationary in a Sun-fixed reference system. (This is not true if we consider the magnetic field effects and the variability of geomagnetic conditions [see, for example, Sanz et al. 1996].)

Then, we have chosen an Earth-centered pseudoinertial reference system (ECI), where the X axis points toward the vernal equinox and the Z axis points toward the geographic north pole; the XY plane is the celestial equator. In the ECI the Sun is only moving 1° per day. The associated spherical coordinates are the right ascension \( \alpha \) (azimuthal angle) and the declination \( \delta \) (angle referred to the equator). For practical purposes, i.e., to know which is the TEC over a given observer, it can be interesting to know the ionospheric electron distribution in an earth-fixed reference frame, in terms of longitude and latitude. As the main difference is just a rotation over the Earth pole, we will represent also in our final results the projection of the Greenwich meridian.

Getting the Ionospheric Delays (or the Modified STEC)

The slant total electron content (STEC) for each ray “GPS satellite station” can be obtained from the same GPS measurements, due to the effects of the ionosphere (a dispersive medium) on the phase and group delay of the electromagnetic (EM) waves crossing this medium, \( \Delta s_g, \Delta s_p \). As is well known [see, for example, Davies, 1990, page 73], the delays (in meters) are related to the STEC \( I \) in electrons per square meter and the frequency of the signal \( f \) in hertz:

\[
\Delta s_g = -\Delta s_p = \frac{40.3 I}{f^2}
\]

(1)

For the Global Positioning System, in particular, as was commented above, the satellites are orbiting \( \sim 20,000 \) km above the Earth surface and transmit codes within two carriers: L1 (\( f_1 \sim 1.6 \) GHz) and L2 (\( f_2 \sim 1.2 \) GHz). Then, each observation of GPS code and phase between a given receiver \( i \) and a satellite \( j \) (\( \rho_{mi}^j \) and \( \varphi_{mi}^j \), respectively, in the frequency \( f_m \)) is affected by an ionospheric delay (equation (1)), which depends on the frequency. This dependence allows us to eliminate these delays using the ionospheric-free combination of the corresponding observations in L1 and L2. Nevertheless, if we are interested on modeling the radially integrated electron density of the ionosphere \( T \) (also called total electron content (TEC)), we can estimate it following, for instance, the strategy of Sardon et al. [1994]. Indeed, we have the following relationships for the observations:

\[
\begin{align*}
\rho_{mi}^j &= \rho_i^j + D_i + D_j^j + \alpha_m \varphi_{mi}^j + \lambda_m \varphi_{mi}^j, \quad m = 1, 2 \\
\varphi_{mi}^j &= \varphi_i^j + \alpha_m \rho_{mi}^j
\end{align*}
\]

(2)

where \( \rho_i^j = \rho_i^j + D_i + D_j^j \) is the modified STEC that includes the L1-L2 differential instrumental delays for receiver \( i \) and satellite \( j \), \( D_i \) and \( D_j \); \( \alpha_m \) is a modified phase bias; \( \rho_i^j \) is the geometrical delay between satellite and receiver and \( \alpha_m = 40.3/f_m^2 \) in units of meters x Hz^2/(electrons per square meter).

After detecting and repairing the cycle-slips of the phases and estimating the modified bias \( \rho_{mi}^j \) in equation (2), we are able to estimate the modified STEC (including the instrumental delays) using the ionospheric combination

\[
\rho_i^j = \rho_i^j + D_i + D_j^j = \frac{I_i^j - I_{2i}^j}{\alpha_2 - \alpha_1}
\]

(3)

General Model

In the general tomographic approach, and not taking into account the bending effect associated with the refraction of the ray, we can relate the slant TEC with the electron density \( N \) as \( I_i^j = \int_{\vec{r}_i} N(\vec{r}, t) \) ds.
In this manner, equation (3) can be rewritten as

\[ \overline{T_i^j} = \int_{r_i}^{r_j} N(\vec{r}, t) \, ds + D_i + D^j \]  

where \( N(\vec{r}, t) \) is the electron density at epoch \( t \) and position \( \vec{r} \), which is a point that belongs to the ray between station \( i \) and satellite \( j \) at distance \( s \) of the station; \( \overline{T_i^j} \) is the ionospheric combination in units of integrated electron density (i.e., the modified STEC) corresponding to the ray from the satellite \( j \) to the station \( i \) at time \( t \); \( \vec{r}_i, \vec{r}_j \) are the position vectors of the station \( i \) and satellite \( j \) at the observation time; and \( D_i \) and \( D^j \) are the instrumental delays associated with the station \( i \) and satellite \( j \). The integral path is assumed to be extended over the linear ray.

In order to estimate the density \( N \) from equation (4), we can expand it within certain basis functions

\[ N(\vec{r}, t) = \sum_i \lambda_i g_i(\vec{r}) \]  

in such a way that

\[ \overline{T_i^j} = \sum_i \lambda_i \int_{r_i}^{r_j} g_i(\vec{r}) \, ds + D_i + D^j \]  

Our final purpose is to get an estimation of \( N \), from the data \( I \), knowing \( \lambda_i \). The instrumental delays \( D_i, D^j \) are also unknowns. Then, the general model proposed consists of a certain number of geocentric spherical shells covering the ionosphere sampled by the GPS rays. These shells define the floor and ceiling of each layer, which is going to be partitioned within pixels.

For each layer \( k \) (see Figure 1) we define the pixels or cells \( \{P_{k,l}\}_{l \in I} \) as those given by a set of centers \( \{w_{k,l}\} \) with the minimum distance criterion:

\[ P_{k,l}(\vec{r}) = \begin{cases} 1 & \text{if } ||\vec{r} - \vec{w}_{k,l}|| \leq ||\vec{r} - \vec{w}_{k,l'}|| \forall l' \\ 0 & \text{otherwise} \end{cases} \]  

Then, we can write applying to equation (6):

\[ \overline{T_i^j} = \sum_k \sum_l N_{k,l} \int_{r_i}^{r_j} P_{k,l}(\vec{r}) \, ds + D_i + D^j \]  

where the unknowns \( \lambda \) have been reinterpreted as the mean electronic density in the cell \( N_{k,l} \). From the last two equations, we get

\[ \overline{T_i^j} = \sum_k \sum_l N_{k,l} \Delta s_{k,l} + D_i + D^j \]  

where \( N_{k,l}, \Delta s_{k,l} \) are the mean density and fraction of the ray length, respectively, corresponding to the cell \( l \) of layer \( k \).

Finally, we have in equation (9) a linear overdetermined system that can be solved by the least squares method and by taking into account the difficulties in reconstructing the vertical structure pointed out by Hajj et al. [1994]. We focus our attention on a global description of the electron content considering only one layer \( (k = 1 \text{ in equation 9}) \), which would correspond with the simple physical model of assuming all the electrons distributed in one spherical layer without radial variation, centered at the maximum density height.

So the considered model can be written as

\[ \overline{T_i^j} = \sum_l N_{1,l} \Delta s_{1,l} + D_i + D^j = \sum_l T_l M_l + D_i + D^j \]  

where \( T_l = N_{1,l} \times h \) is the TEC of the cell \( l \), \( h \) being the thickness of the layer, and \( M_l = \Delta s_{1,l} / h \) is the corresponding mapping function.

**Adaptative Pixels Versus Regular Pixels**

If we choose a pixel basis function to expand the electron density, two approaches, among others, are possible to define the pixels or cells: (1) to consider a regular grid and (2) to get adaptative cells in such a way that an approximately equal number of rays per cell can be assured. Both models present some advantages and disadvantages identified from our own experience [Juan et al. 1997].

1. While the regular grid provides a faster algorithm to solve the model, compared with the same number of parameters, the adaptative approach does
not provide poorly populated cells, avoiding the associated ill-conditioning problems when we solve the model.

2. When an important density of data is available in certain regions (ionosphere above North America and Europe), the adaptive gridding provides higher spatial resolution for the TEC description in these zones, with the same number of parameters. To fulfill such resolution with a regular grid, the number of parameters would be several times greater, with higher computation time and ill-conditioning problems.

3. Another point is the discretization error, which is comparable between the two models when equivalent constraints are imposed (see the end of Appendix B). In the adaptive approach, the error introduced in the description by large cells affects few rays, so both effects tends to cancel.

These features are the reasons to propose a new adaptive gridding, where the centers are obtained by applying the unsupervised classifier known as the Kohonen neural network, or self organizing map [Kohonen, 1990], to the crossing points of each ray with the mean shell of the given layer. The resulting cells are adapted in size, guaranteeing a minimum number of crossing points per cell (see Appendix A for a detailed explanation of the algorithm and Murtagh [1995] for an assessment).

Then, we propose to estimate the electron density in equation (9) as a two-dimensional (2-D) global model of the overall ionosphere (using all the IGS data). To do so, we solve equation (9) with only one layer [equation 10], avoiding in this manner the high correlations between the estimations of electron density at different heights, as is commented on by Hajj et al. [1994] and by using the adaptive basis of functions to avoid the problem of poorly populated cells (mainly in the southern hemisphere due to the lack of stations). (To avoid the existence of empty or sparsely populated cells in our model, we take the initial values for the cell centers randomly from the input subionospheric points. Also, if any sparsely populated cells would appear, our software automat-

<table>
<thead>
<tr>
<th>Table 1. Description of the Data Set Considered in the Computations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Data set</td>
</tr>
<tr>
<td>GPS receivers</td>
</tr>
<tr>
<td>Time Interval</td>
</tr>
<tr>
<td>Number of subsets</td>
</tr>
<tr>
<td>Right ascension range of the rays at 350 km height</td>
</tr>
<tr>
<td>Declination range of the rays at 350 km height</td>
</tr>
<tr>
<td>Elevation mask</td>
</tr>
<tr>
<td>Number of rays</td>
</tr>
</tbody>
</table>

These data correspond to October 18, 1995, with a geomagnetic storm and P code not encrypted (antispoofing off).

Figure 2. International GPS Service for Geodynamics permanent stations used in the computations.
Computation and Results

Data

The main features of the data sets used (provided by the IGS via anonymous ftp) are summarized in Table 1. Also, the distribution of stations is plotted in Figure 2. In order to maximize the amount of data, we do not reject observations taken at low elevations, due to the fact that the multipath practically does not affect the L1-L2 phase data.

Estimated Model

We have computed the 2-D global ionosphere using the least squares method. We use the self-organizing map algorithm to generate the adaptative cells in order to compute the TEC for the data set corresponding to October 18, 1995 (subsets 0200-0700, 0700-1200, 1200-1700, and 1700-2200 UT; see Figure 3). (During this period one geomagnetic storm occurred, with a high variability of the ionospheric electron distribution (see, for instance, the document http://bolero.gsfc.nasa.gov/~solar/cloud/cloud.html).) We have solved equation (9) for one unique layer between 300 and 400 km, taking 400 cells (self-organized along a 20x20 Kohonen map) and a subray length of 5 km (see appendices for the details). The cells present sizes ranging from few squared degrees (above the cluster of GPS stations in California) to 10-100 times greater or more (southern oceans), depending on the small or large sparsity of the data. In Figure 3 an interesting result appears: the detection of the TEC increase coincident with the start of the geomagnetic storm in the last time interval (from 1700 to 2200 UT). Indeed, the geomagnetic Ap index provided each 3 hours by NOAA (ftp://140.172.172.193/stp/geomagnetic_data/indices/kp_ap /1995) reaches the value of 111 (severe storm) between 2100 and 2400 UT, October 18, 1995. Other remarkable results are the following:

1. The instrumental delays are similar to the one obtained by the German Aerospace Research Establishment (the initials in german are DLR) following the strategy of Sardon et al. [1994] (ftp://ftp.nzl.dlr.de/pub/navigation/bias/). The rms of the differences is \( \approx 10 \) cm (\( \approx 0.3 \) ns) in front of DLR internal rms for each instrumental delay of the order of 3 cm.

2. The mean residual of the slant TEC is \( \approx 2.10^{16} \) el/m² (\( \approx 20 \) cm in delay units), comparable with that obtained by Wilson et al. [1995].

Conclusions

We present in this paper the estimation of the global 2-D ionospheric electron distribution from GPS ground data using an adaptative neural network algorithm. The use of the neural network in this work makes it possible to automatically overcome the problem of the nonhomogeneous sampling, dividing each layer into a partition of cells or clusters with a similar number of rays. At the same time, solving for the same number of parameters as in a regular model, the resolution obtained in the regions of the ionosphere sounded by more rays is greater and the model is not ill-conditioned. From this point of view, this method is not more expensive in computing time than the regular models.

In order to do this, we have considered a new basis of adaptative pixels (Kohonen adaptative pixel basis) that are defined from a certain number of centers, obtained with the Kohonen artificial neural network. The existence of the topological relationships between the neighbors centers in the Kohonen map, simultaneously generated during the computation of the cells, is also taken into account to dramatically reduce the computation load. (This important feature, from a practical point of view, is not directly provided by other “adaptative”/clustering techniques such as k-means [see, for example, Murtagh and Hernández-Pajares, 1995]). This new strategy is supported by the results obtained for October 18, 1995, with the detection of the increase of electron content during the starting of the geomagnetic storm.

Future research, taking this model as starting point, can be done by studying the electron vertical distribution of the ionosphere, using data with more vertical information like the GPS/MET low earth orbiter data (see http://pocc.gpsmet.ucar.edu/). Then, we could assume several layers with the general model (equation (9)) or we could use three-dimensional (3-D) cells computed in the same adaptative way (model without layers), without increasing the correlations.

Appendix A: The Self-Organizing Feature Map Algorithm

One of the competitive learning neural network algorithms is the self-organizing map (SOMA), also called the Kohonen network. It has the special prop-
Figure 3. Global model of the total electron content (TEC) for October 18, 1995, for the data subsets (0200-0700, 0700-1200, 1200-1700, and 1700-2200 UT) in Earth-centered, Earth-fixed coordinates (right ascension in X axis and declination in Y axis). The projection of the Greenwich meridian is also plotted.
property of creating spatially organized representatives of the centroids (weights of the output neurons) found in the input vectors. The resulting maps resemble real neural structures that appear in the cortices of developed animal brains. The SOMA has also been successful in various pattern recognition tasks involving noisy signals such as speech recognition (see a summarized review by Kohonen [1990]).

The basic aim of this neural network is to find a smaller set \( \{\bar{w}_1, \ldots, \bar{w}_c\} \) of \( c \) centroids that provides a good approximation of the original set \( S \) of \( n \) objects (input space), with \( m \) attributes, encoded as vectors \( \bar{x} \in S \). Intuitively, this should mean that for each \( \bar{x} \in S \) the distance \( \| \bar{x} - \bar{w}_{i(j)} \| \) between \( \bar{x} \) and the closest centroid \( \bar{w}_{i(j)} \) should be small. Simultaneously, the algorithm arranges the centroids so that the associated map \( f \) from \( S \) to \( A \),

\[
f : S \subset \mathbb{R}^n \rightarrow A \subset \mathbb{N}^2
\]

reflects the topology of the set \( S \) in a least distorting way, where \( A \) is the representation space, a two-dimensional set of indices known as the self-organizing feature map, or Kohonen map. Proximity in \( A \) means similarity between the global properties of the associated groups of objects.

From a detailed point of view, the Kohonen network is composed of a set of \( c \) nodes or neurons. The algorithm scheme is as follows:

1. We initialize at random the weights of the \( c \) nodes of the grid with small values: \( \{\bar{w}_1(0), \ldots, \bar{w}_c(0)\} \). After training, every neuron \( l \in \{1, \ldots, c\} \) will represent a group of objects with similar features, and the weight vector \( \bar{w}_l \) will approximate the centroid of these associated objects.

2. For each of the \( n \) training vectors of the overall database, \( \bar{x}^p \), we do the following: (1) We find the node \( k \) whose weight \( \bar{w}_k \) best approaches \( \bar{x}^p \) (\( d \) can represent the Euclidean distance): \( d(\bar{w}_k, \bar{x}^p) \leq d(\bar{w}_l, \bar{x}^p), \forall l \in \{1, \ldots, c\} \). (2) We update the weight of the winning node \( k \) and its neighbors, \( N_k(t) \), approaching the training vector as closely as possible:

\[
\bar{w}_l(t) = \begin{cases} 
\bar{w}_l(t-1) + \alpha(t) H(t) (\bar{x}^p - \bar{w}_l(t-1)) & l \in N_k(t) \\
\bar{w}_l(t-1) & l \notin N_k(t)
\end{cases}
\]

(A2)

where \( \forall l \in \{1, \ldots, c\} \) and \( \alpha(t) \) is the learning rate, a suitable monotonic decreasing sequence of scalar-valued gain coefficients, \( 0 < \alpha(t) < 1 \); the radius \( R_t \) of the activated neighborhood, \( N_k(t) \), is a monotonically decreasing function of the iteration \( t \); and \( H(t) = \exp\left(-d^2/4R_t^2\right) \) is a function that represents in equation (A2) the decay of the activation depending on the distance \( d \) in \( A \) between the winner unit \( k \) and the considered unit \( l \), \( d = \sqrt{(i_l - i_k)^2 + (j_l - j_k)^2} \). Updating neighbor’s weights instead of just that of the winning node assures the ordering of the net in such a way that centroids which are close in the representation space \( A \) (dimension 2) are updated to become similar in the input space \( S \) of dimension \( m \) (see Figure 4).

![Figure 4](image-url)
3. Process 2 is repeated for the overall database until good final training is obtained.

The final point density function of \( \{\bar{w}_1, \ldots, \bar{w}_c\} \) is an approximation of the continuous probability density function of the vectorial input variable \( \vec{g}(\vec{x}) \) [Kohonen, 1990, p. 1466].

Appendix B: Some Practical Aspects of the Problem

In order to implement the adaptative strategy, we have to deal with the problem of the non regular boundaries between the cells. Indeed, to solve equation (10), we have to compute for each ray and for a given layer, the cells crossed and the ray path fraction for each one. One procedure is to digitize the ray within subrays of length \( L \), counting how many of them are contained in the crossed cells (see Figure 5).

This approach implies getting the cell to which each subray belongs; this can be an expensive operation, due to the amount of subrays that we have. One way to avoid such computation load is to use the physical continuity of the ray, knowing the neighboring between the cells. The Kohonen neural network gives a bidimensional topological map of the cells at the same time that it constructs cells with an approximately equal number of elements (see Appendix A).

The strategy adopted is in two steps. The first is to compute the cells. From the crossing points of all the rays with the mean spherical shell of the layer, the centers of the cells are obtained. Simultaneously, we get the cell membership of all the rays and the 2-D Kohonen map of the centers, which maintains their 3-D neighboring relationships. The cells are defined in the usual way by the nearest center criterium. The second step is to compute the crossing fractions of the rays. For a given ray, we know now to which cell the subray belongs to (step 1) and hence where it is placed in the Kohonen map. For a enough small length and taking into account the continuity of the ray, the next upper (lower) subray must belong to the same or to a neighbor cell (see Figure 5); then only a small subset of centers must be explored: neighbors in the Kohonen map to the center of the last subray (again, see figure 5). This approach allows us to
reduce the computation time an order of magnitude regarding the exploration of all the centers to assign each subray.

Another advantage of this approach, of dividing each ray into subrays of a certain length $L$, is that it introduces a fuzzy feature in the borders between cells which is equivalent to an implicit smoothing in the solution, depending on $L$. However, this implicit constraint of continuity could increase the discretization error (which would be like increasing the mean size of the cells).

Acknowledgments. We thank A. Rius of the IEEC/CSIC for some interesting discussions which were useful in the starting of this work. We acknowledge the International GPS Service for Geodynamics for the availability of the GPS data used in this research. This work has been partially supported with funds from Spanish government projects PB94-1205 and PB94-0905 (DGICYT).

References


M. Hernández-Pajares, J. M. Juan, and J. Sanz, Research Group of Astronomy and Space Geodesy, Universitat Politècnica de Catalunya, Campus Nord Mod. C3, c/. Gran Capità s/n., 08034 Barcelona, Spain. (e-mail: manuel@mat.upc.es; miguel@magallanes.upc.es; jaume@mat.upc.es)

(Received October 1, 1996; revised January 22, 1997; accepted February 13, 1997.)