Global prediction of the vertical total electron content of the ionosphere based on GPS data

A. García-Rigo,[1,2] E. Monte,[3] M. Hernández-Pajares,[1] J. M. Juan,[1] J. Sanz,[1] A. Aragón-Angel,[1,2] and D. Salazar[1]

Received 4 January 2011; revised 21 September 2011; accepted 12 October 2011; published 29 December 2011.

[1] Although vertical total electron content (VTEC) forecasting is still an open subject of research, the use of predictions of the ionospheric state at a scale of several days is an area of increased interest. A global VTEC forecast product for two days ahead, which is based exclusively on actual Global Positioning System (GPS) data, has been developed in the frame of the International Global Navigation Satellite Systems (GNSS) Service (IGS) Ionospheric Working Group (IGS Iono-WG). The UPC ionospheric VTEC prediction model is based on the Discrete Cosine Transform (DCT), which is widely used in image compression (for instance, in JPEG format). Additionally, a linear regression module is used to forecast the time evolution of each of the DCT coefficients. The use of the DCT coefficients is justified because they represent global features of the whole two-dimensional VTEC map/image. Also, one can therefore introduce prior information affecting the VTEC, for instance, smoothness or the distribution of relevant features in different directions. For this purpose, the use of a long time series of final/rapid UPC VTEC maps is required. Currently, the UPC Predicted product is being automatically generated in test mode and is made available through the main IGS server for public access. This product is also used to generate two days ahead preliminary combined IGS Predicted product. Finally, the results presented in this work suggest that the two days ahead UPC Predicted product could become an official IGS product in the near future.

Citation: García-Rigo, A., E. Monte, M. Hernández-Pajares, J. M. Juan, J. Sanz, A. Aragón-Angel, and D. Salazar (2011), Global prediction of the vertical total electron content of the ionosphere based on GPS data, *Radio Sci.*, *46*, RS0D25, doi:10.1029/2010RS004643.

1. Introduction

[2] Few decades ago, the satellite navigation dual-frequency Global Positioning System (GPS) became very popular for ionospheric sounding. This system offered an unprecedented combination of accuracy, temporal and spatial resolution, and availability (low cost). Among other achievements, this led to important progress in fields such as ionospheric tomography (see review by *Bust and Mitchell* [2008]), which has enabled to enhance precise Global Navigation Satellite Systems (GNSS) navigation [*Hernández-Pajares et al.*, 2000] and generate reliable global vertical total electron content (VTEC) maps in a continuous operational way in the context of the International Global Navigation Satellite Systems (GNSS) Service (IGS [*Dow et al.*, 2009; *Hernández-Pajares et al.*, 2009]). This manuscript focuses on the improvement of the GNSS application for operational monitoring of the ionospheric electron content. A summary of a new approach used to predict global VTEC maps from GPS data in a very demanding operational scenario is presented. Developing short-term ionospheric predictions can be of great importance for certain scientific and technological applications. For instance, there is a great interest from the European Space Agency (ESA) in the Soil Moisture and Ocean Salinity (SMOS [*Sivestrin et al.*, 2001]) mission to retrieve ionospheric products provided by the IGS Iono-WG, including global VTEC predictions for two days ahead [*Krankowski and Hernández-Pajares*, 2008]. Currently, predictions are being used by the authors as background model to generate preliminary real-time global VTEC maps (see details in the work of *Orús et al.* [2010]). Among many other applications, this could be used to achieve sub-meter accuracies for mass-market single-frequency receivers. This would improve the quality of current GNSS applications, such as in automobiles, road mapping, location-based advertising, personal navigation or logistics.

[3] In the past and recent years, methods for ionospheric forecasting have been developed to predict ionospheric parameters such as the critical frequency of the F2 layer (f_0F2), the maximum usable frequency at a distance of 3000 km of the F2-layer ($M(3000)F2$) or the TEC itself.

[1]Research Group of Astronomy and Geomatics, Technical University of Catalonia, Barcelona, Spain.
[2]gAGE-NAV, S.L., Barcelona, Spain.
[3]Department of Signal Theory and Communications, Technical University of Catalonia, Barcelona, Spain.

0048-6604/11/2010RS004643

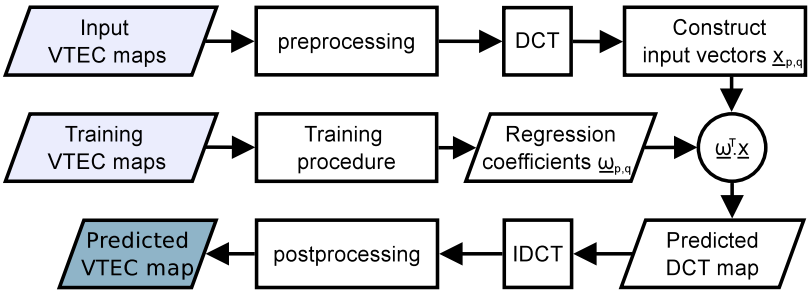

Figure 1. Diagram showing the main steps of the UPC prediction approach. DCT stands for Discrete Cosine Transform, IDCT for Inverse DCT and superindex *T* denotes transpose.

These methods are based on neural networks [*Cander et al.*, 1998; *Francis et al.*, 2000; *Tulunay et al.*, 2006], auto-correlation and auto-covariance procedures [*Muhtarov and Kutiev*, 1999; *Dick et al.*, 1999; *Stanislawska and Zbyszynski*, 2001], linear regression [*Muhtarov et al.*, 2001; *Krankowski et al.*, 2005], among others.

[4] In terms of global VTEC forecasting using GNSS, the first IGS Associate Analysis Center (IAAC) to generate its own ionospheric predicted product was the Center for Orbit Determination in Europe (CODE). Their model, which is based on the extrapolation of the Spherical Harmonics coefficients, is explained in detail by *Schaer* [1999]. Their one and two days ahead VTEC maps (named *C1PG* and *C2PG*, respectively) have been released for public access for years via the FTP server of the Astronomical Institute of the University of Bern (AIUB; http://www.aiub.unibe.ch). More recently, the ESA's Space Operations Centre (ESA-ESOC) and the Technical University of Catalonia (UPC) have created their own two days ahead forecast product in the frame of the IGS Iono-WG (named *E2PG* and *U2PG*, respectively) to fulfill the requests of the SMOS mission. These products have been released for public access through the FTP site of the Crustal Dynamics Data Information System (CDDIS; ftp://cddis.gsfc.nasa.gov/) since September, 2009. Furthermore, preliminary combined IGS Predicted products for one day and two days ahead have also been generated by ESA-ESOC since the end of December, 2009. These products are produced by combining up to the three individual preliminary VTEC predicted products and are designated as *I1PG* and *I2PG* for one day and two days ahead forecast, respectively.

[5] The first part of the article is devoted to explain the UPC prediction approach in more detail. Afterwards, a section is included on how the performance analysis is carried out. And finally, before the conclusions, the results for two days ahead forecast are presented for three periods (in 2004, 2006 and 2011) and validated against UPC Final IGS products (designated as *UPCG*) and JASON external data. A first performance comparison is also included between UPC and the other preliminary IGS forecast products.

2. Technical Description

[6] A model for short-term VTEC predictions was developed by the authors to generate a two days ahead forecast product in the frame of the IGS Iono-WG and to fulfill the demanding SMOS mission requirements. In brief, the forecast model is based on applying linear regression to a temporal window of VTEC maps in the Discrete Cosine Transform (DCT [*Ahmed et al.*, 1974; *Oppenheim and Schafer*, 2010]) domain. Figure 1 shows a simplified diagram with the main steps that are required to compute the UPC Predicted product.

[7] The UPC Final/Rapid IGS products are the input data for the prediction model. These products are encapsulated in IONosphere map EXchange format (IONEX [see *Schaer et al.*, 1998]) and contain 13 global VTEC maps at a temporal resolution of two hours, from 00 UT of the corresponding day to 00 UT of the following day. Each VTEC map is a two-dimensional map with spatial grid points every 2.5°/5° in the latitude/longitude range assuming a thin shell layer model of the ionosphere at a height of 450 km. Thus, the map contains $MxN = 5112$ independent VTEC samples, where $M = 71$, the number of grid points in the latitudinal/vertical direction (the number of rows), and $N = 72$, the number of grid points in the local-time/horizontal direction (the number of columns). From now on, the VTEC value at row *m* and column *n* will be referred to as $V_{m,n}$. Since the DCT is used as an image processing tool, the standard nomenclature in image processing has been followed. Therefore, the latitude/ordinate direction is also denoted as vertical direction and the local-time/abscissa direction as horizontal direction.

[8] In daily computations, the last seven consecutive UPC Final IGS products (*UPCG*) are used as the input dataset for the model (first row in Figure 1). If any of these products is not available or is not generated, the corresponding UPC Rapid IGS product (*UPRG*) is retrieved instead. Note that UPC Rapid IGS products are generated with a one-day latency while the UPC Final IGS products are internally generated with a five- to six-day latency. This is an acceptable procedure because discrepancies between both the UPC Final and Rapid IGS products are approximately 5–7%, as stated by *Hernández-Pajares et al.* [2009]. For the prediction on day *T*, the input dataset includes the UPC products from day $T - 8$ to $T - 2$ (being $T - 1$ the day before day *T*).

[9] Regardless, because 00 UT VTEC maps are calculated twice for adjacent products, a total of $U = 85$ input VTEC maps are considered (one map every two hours for seven days, taking into account the fact that 00 UT maps are not accounted for twice; $U = 12*7 + 1$). The U length was tested for values between 13 and 481 (1 to 40 daily UPC IGS products as input dataset) for three independent weeks (in 2004, 2010 and 2011 with or without disturbed ionospheric conditions, respectively) giving a consistent optimum value of

seven. As the performance curve is convex, the cutoff at 481 is justified because of the monotonous decrease in performance and because the degradation is so severe when U tends to 481 that higher values were not considered.

[10] The preprocessing step in Figure 1 includes the transformation of the input data to a sun-fixed reference frame (local time and latitude). Afterwards, each VTEC map is represented by a set of parameters of a generative model in the frequency domain, in this case the two-dimensional DCT. The basis functions of the DCT represent the latitude, longitude and crossed latitude/longitude spatial frequency components of the original VTEC map. In other words, the DCT provides information on the vertical, horizontal and diagonal patterns of the input VTEC grid image. The output values of the DCT of a given map are referred to as DCT coefficients, following the notation $C_{p,q}$, where p and q correspond to the vertical and horizontal discrete frequencies, respectively. Thus, the DCT coefficient $C_{0,0}$ is related to the continuous component (i.e., the mean VTEC value of the map), and the DCT coefficients correspond to higher frequency components as the p and q indices increase.

[11] Note that the DCT allows physical properties of the VTEC to be modeled. Specifically, this includes the smooth spatial evolution of the VTEC (i.e., the fact that the absolute value of high frequency components is smaller than that of low frequency components) and the fact that energy is more concentrated along each frequency axis. Therefore, the original VTEC map in the spatial domain can be generated with a subset of the DCT coefficients without significantly degrading the quality. In this work, a triangular subset considering 2556 DCT basis functions was selected to reduce the noise produced by high frequency coefficients and to smooth the spatial variation of the VTEC. Thus, the index of the vertical discrete frequencies p ranges between 0 and $P - 1 = 70$, and the index of the horizontal discrete frequencies q ranges between 0 and $Q - p = 71 - p$. In equation (1), the mathematical means of computing the DCT coefficients of a given map represented by the VTEC values $V_{m,n}$ is shown.

$$C_{p,q} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} V_{m,n} \cos\frac{\pi(2m+1)p}{2M} \cos\frac{\pi(2n+1)q}{2N}, \quad 0 \le p \le P-1, 0 \le q \le Q-p \tag{1}$$

[12] It is important to note that the future values of each DCT coefficient, rather than the raw spatial VTEC values, are predicted. The predicted DCT coefficients can then be used to reconstruct the predicted VTEC map by applying the Inverse Discrete Cosine Transform (IDCT; see diagram in Figure 1) and a post-processing, where the change in the kind of coordinates is applied. This approach is more robust, as the evolution of the VTEC for a specific grid point may not be correlated with its previous values, but may be highly correlated with the time and spatial evolution of the physical processes affecting the whole VTEC map. In fact, the transformation implies that the spatial information of the whole map/image is implicit in each DCT coefficient.

[13] The relationship between past and future values of each DCT coefficient is obtained by linear regression. The input vectors $\underline{x}_{p,q}$ in the diagram in Figure 1 are built with the last $U = 85$ values of each DCT coefficient and include a constant term to model the offset. Additional terms could have been included with information such as the 10.7 cm Solar Flux, the Sunspot number, the *Kp* geomagnetic index or the ionospheric parameter Global Electron Content (GEC [*Afraimovich et al.*, 2006]). However, the impact of these parameters on prediction performance is currently under investigation. On the other hand, the corresponding linear regression coefficients are computed via a training process and are named $\underline{w}_{p,q}$. The model yields a high number of parameters and the inputs could be collinear. As a result, ridge regression is used, which is a modification of the linear regression where a regularization term $\lambda \cdot \underline{\underline{I}}$ has been added to the covariance matrix [*Hastie et al.*, 2001], where λ is a scalar below one and $\underline{\underline{I}}$ is the identity matrix.

[14] In this work, the following training process is executed each day (second row in Figure 1) taking approximately 45 min of computer processing in a standard PC under Linux. First, the training data obtained from 366 days of UPC Final IGS products are transformed to a sun-fixed reference frame. For the prediction for day T, the training dataset includes the UPC IGS products from day $T - 9 - 365$ to $T - 9$. Secondly, all VTEC maps are converted into the DCT frequency domain and one-dimensional one-year time series are generated for each DCT coefficient. A sliding window moving one VTEC map per step determines the number of training subsets of length $U = 85$ VTEC maps. Afterwards, the coefficients $\underline{w}_{p,q}$ are computed as shown in Algorithm 1. These regression coefficients model the relationship between the training subsets and their corresponding reference values $\underline{r}_{p,q}$ (known "predictions"). These reference values are obtained from the VTEC map that is Δt maps ahead from the last input VTEC map. Thus, t is the integer temporal index for each input VTEC map. Note that the linear regression coefficients are calculated independently for each DCT coefficient and for each of the VTEC maps to be predicted.

[15] The training process has to be configured in 13 independent ways because the predicted product includes 13 VTEC maps as it is encapsulated in IONEX format. Thus, the regression coefficients are calculated considering that the distance Δt between the last VTEC map in each training subset and the reference can be 12 to 24 maps ahead. Note that for the two days ahead forecast on day T, the prediction model is executed using input products until day $T - 2$ (see the third paragraph in this section 2). Nevertheless, the last input VTEC map is the one at 00 UT on day $T - 1$. Therefore, the prediction for 00 UT on day T is 12 maps ahead and the prediction for 00 UT on day $T + 1$ is 24 maps ahead.

[16] Algorithm 1: Pseudocode used to compute the regression coefficients $\underline{w}_{p,q}$ for $i \leftarrow 1$, *nTrnSubsets* do

$$X_{p,q}(1,i) \leftarrow 1(\text{Offset})$$

for k $\leftarrow$ *iFirstMap* to *iLastMap* do

$$X_{p,q}(1+k,i) \leftarrow C_{p,q}[k]\ (\text{Values assignment})$$

end for

$$r_{p,q}(i) = C_{p,q}[iLastMap + \Delta t](\text{Reference is } \Delta t \text{ maps ahead})$$

end for

$$\underline{w}_{p,q} \leftarrow \left(\underline{\underline{X}}_{p,q} \cdot \underline{\underline{X}}_{p,q}^{T} + \lambda \cdot \underline{\underline{I}}\right)^{-1} \cdot \underline{\underline{X}}_{p,q} \cdot \underline{r}_{p,q} \text{ (Pseudo inverse to get the regression coefficients)}$$

[17] Finally, each predicted value is computed by performing a dot product between the coefficients $\underline{w}_{p,q}$ and the sequence of DCT coefficients (input vector $\underline{x}_{p,q}$), as can be seen from equation (2) and in the diagram in Figure 1.

$$\widehat{C_{p,q}}[t+\Delta t] = \omega_{p,q}[0] + \sum_{u=1}^{U} \omega_{p,q}[u] \cdot C_{p,q}[t-u+1] \qquad U = 85; 12 \le \Delta t \le 24 \tag{2}$$

[18] The resulting UPC Predicted product is automatically generated on a daily basis and is designated as *U2PG*, which is compatible with the IONEX naming convention. Note that the predicted product for day T uses information until day $T-2$ and it is computed and made publicly available on day $T-1$ before midday (following SMOS mission request).

3. Results

3.1. Test Data

[19] As mentioned before, the final, rapid and preliminary predicted IGS products for the different IAACs and their combined products were used as test data. And an alternative approach, termed the "frozen" or time-invariant ionosphere method in which no prediction model is applied, was also considered for comparative purposes. In this last case, the prediction relies directly on the raw VTEC values provided by the Final/Rapid IGS product produced two days before. The Rapid product for $T-2$, which is made available at CDDIS on day $T-1$, is directly considered as the Time-invariant ionosphere product for day T (no prediction being applied; ionosphere is considered frozen for two days). Note that this internal product is designated as *UPR*2. Also, Time-invariant products for CODE, ESA, JPL and IGS are generated internally (named *COR*2, *ESR*2, *JPR*2 and *IGR*2, respectively). In cases where Rapid IGS products are not generated, the corresponding Final IGS products are used instead (for instance, the Time-invariant product for UPC would be designated *UPC*2).

3.2. Reference Data

[20] As mentioned before, the UPC post-processed maps corresponding to the UPC Final IGS product (*UPCG*) are used as reference data. Additionally, the combined Final and Rapid IGS products (*IGSG* and *IGRG*, respectively), are taken into account since they provide the most accurate ionospheric VTEC values at a global scale when using GNSS data. The suitability and reliability of the above mentioned products in representing ionospheric VTEC have been demonstrated in previous years [*Hernández-Pajares et al.*, 2009]. In this way, IGS products are widely used in the scientific community. Nevertheless, it is important to note that combined Final and Rapid IGS products are not totally independent because UPC Final and Rapid products are used to generate these products. Note that the combined VTEC maps have a better performance than the individual products because systematic errors are partially filtered thanks to the combination process [see *Orús et al.*, 2007].

[21] Additionally, an external source of global VTEC measurements is recorded with the dual-frequency altimeter instrument on board the JASON spacecraft, which was launched on December 7, 2001. The VTEC observables provided by these altimetry data can fill data coverage gaps over the oceans between latitudes of 66°N and 66°S (this restriction is given through the inclination of the JASON orbit), where no permanent GNSS receivers can be placed. In this way, JASON provides independent reference data that can be used to evaluate the performance of GNSS-derived VTEC maps over the oceans except in the polar regions.

[22] JASON data have been used to validate Final IGS products [*Hernández-Pajares et al.*, 2009]. In this respect, two considerations have to be taken into account. First, JASON VTEC measurements are very accurate but are affected by an offset with respect to the IGS products [*Azpilicueta and Brunini*, 2008; *Hernández-Pajares*, 2004]. In this regard, it should be noted that the GPS constellation orbits the earth at an altitude of $\cong$ 20200 km. Therefore, GPS VTEC observations for receiver-satellite pairs include the plasmaspheric contribution, in contrast to the case of JASON that orbits at an altitude of $\cong$ 1300 km. Second, this is a pessimistic scenario because JASON direct measurements are compared with interpolated values at the same exact location derived from the nearby grid points of the VTEC map being validated. And it is likely that even the VTEC at these grid points had to be interpolated from insufficient real data due to the lack of permanent GNSS stations over the oceans.

[23] As a final remark, JASON VTEC data are made publicly available from year 2003 through the FTP site of the National Aeronautics and Space Administration-Jet Propulsion Laboratory (NASA-JPL) Physical Oceanography Distributed Active Archive Center (PO.DAAC; http://podaac.jpl.nasa.gov).

3.3. Performance Tests

[24] As a first validation of the results, the UPC Predicted product is compared to the UPC Final IGS product, which includes the most accurate post-processed VTEC data that UPC provides. As mentioned before, a long time series of UPC IGS products is used as input dataset for the training process. Therefore, the model is specifically designed to predict the UPC Final VTEC values as accurately as possible (Figure 2).

[25] Table 1 shows the bias, standard deviation, root mean square (RMS) and the minimum and maximum values of the differences between the predicted UPC VTEC maps and the UPC Final IGS VTEC maps for the periods being investigated. Its main purpose is to determine whether there is an offset or a deviation that is systematically affecting all of the values of the VTEC maps. Histograms on a semi-logarithmic scale are also provided (Figures 4 and 5) as they reflect the behavior of the model at the tails of the data distribution, which correspond to the areas with the worst prediction performance (i.e., details of the extreme values are emphasized and an idea of how likely they are is given).

[26] As a second validation, a test against the VTEC records provided by the JASON altimeter instrument is used to verify that the predicted product is consistent with an external

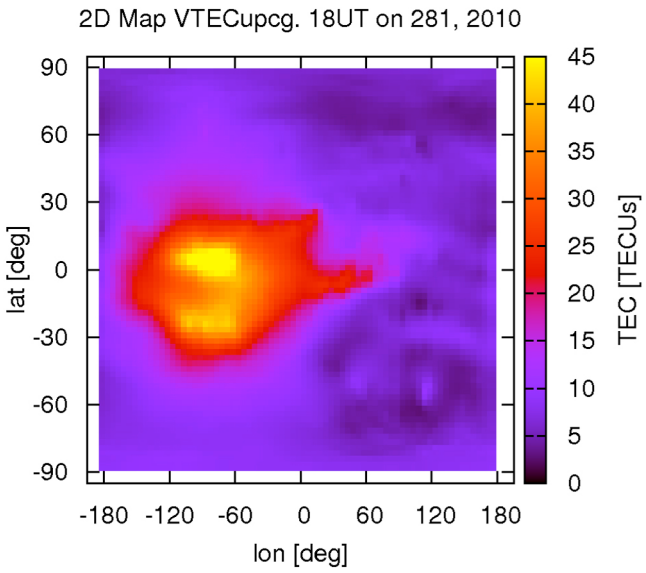

Figure 2. Representation of the UPC Final VTEC map in geomagnetic longitude/latitude range at 18 UT on day 281, 2010.

VTEC data source. The purpose of this test is to show that the prediction performs well in modeling the real VTEC and not only the VTEC of the corresponding UPC Final IGS product. This test is also sensitive to the quality of the input data used in the prediction model.

[27] In this case, Table 2 shows the bias, standard deviation, RMS, the minimum and maximum values, and the number of comparisons of the differences between the UPC Final/Predicted/Time-invariant products and the JASON VTEC determinations. Furthermore, plots of the bias and standard deviation in terms of the geomagnetic latitude are provided (Figures 6 and 7) to quantify and show where the prediction systematically over- or underestimates VTEC. When multiple test products are considered, these plots allow for easier comparisons than with two-dimensional bias or standard deviation maps (see Figure 3). In this case, a whole map would have to be included for each product rather than including only one graph for the maps to compare. Last but not least, these plots can be useful for determining whether the peaks of the ionospheric anomaly have been correctly determined. For instance, this can be of great interest to identify whether the prediction model is providing excessively smoothed predicted VTEC values.

[28] Finally, a comparison between the preliminary IGS predicted products and the JASON VTEC values is performed for the different IAACs. This test is of interest to performance comparisons between the UPC Predicted product and the predicted products produced by the other IAACs (not to comparisons between prediction methods). It is also useful to first evaluate the two days ahead combined IGS Predicted product (*I2PG*).

[29] For this purpose, boxplots of the differences between the predicted products and JASON VTEC values are depicted (Figure 8). The boxplots allow the distribution of the results to be summarized for the different prediction methods in a simplified and compact way by observing the quartiles (see labels on the rightmost boxplot in Figure 8) as well as the smallest and largest daily standard deviation observations. Remember that the lower quartile (designated as $Q1$) is the threshold for the lowest 25% values in the range of results, $Q2$ corresponds to the median and $Q3$ is used to indicate the highest 25% of data. Note that, in this work, the boxplots were computed following the Tukey method [*Hoaglin et al.*, 1983].

3.4. Selected Periods

[30] One month of data has been analyzed for each of the years from 2004 to 2006, as well as 115 days in 2010. In this way, representative datasets have been selected to cover conditions of both low and high solar and geomagnetic activity. Although data from the maximum activity period of the 23rd Solar Cycle has not been considered, the effects derived from certain solar flares and geomagnetic storms are present during all the three selected periods. Nevertheless, it should be stated that the correct predictability of impulsive events is not the current concern of the UPC prediction model and can be the subject of further research.

[31] In 2004, the month of August was analyzed, i.e., from day of year 214 to 244. This period corresponds to medium solar cycle conditions. Nevertheless, the geomagnetic activity was strong at the end of the month due to a solar storm and two X-class flares were recorded by the Geostationary Operational Environmental Satellite (GOES [*Hill et al.*, 2005]) on days 226 and 231. In fact, the *Kp* index ranged from 0 to 7 with a mean value of approximately 1.76 and its maximum value at 00–03 UT on day 244. At that time, the IGS did not provide Rapid products so the Time-invariant product has been generated using the corresponding UPC Final IGS products (*UPC2*).

[32] A second period is selected (December, 2006) for the minimum solar cycle conditions. In this case, the *Kp* index ranged from 0 to 8.3 with a mean value of 2.25 and its maximum value at 00–03 UT at day 349. In addition, a geomagnetic storm occurred on day 348, and three X-class flares were recorded by GOES on days 340, 347 and 348.

Table 1. Statistics of the Differences Between the UPC Predicted/Time-Invariant VTEC Data and the UPC Final IGS Reference Data, Including Bias, Standard Deviation, RMS, Minimum and Maximum Differences for the Three Selected Periods in 2004, 2006 and 2010 in TECUs

Period	Product	BIAS	Standard Deviation	RMS	Minimum	Maximum	Number of Comparisons
214–244, 2004	Predicted	0.23	2.37	2.38	−23.10	18.20	2,060,136
	Frozen	−0.25	2.63	2.64	−25.50	27.00	2,060,136
335–365, 2006	Predicted	0.07	2.27	2.27	−46.90	24.20	2,060,136
	Frozen	0.11	2.69	2.69	−50.40	53.80	2,060,136
184–355, 2010	Predicted	−0.27	2.49	2.50	−34.40	21.30	7,642,440
	Frozen	−0.28	2.89	2.90	−35.60	31.90	7,642,440

Table 2. Statistics of the Differences Between the UPC Predicted/Time-Invariant VTEC Data and the JASON Reference Data, Including Bias, Standard Deviation and RMS for the Three Selected Periods in 2004, 2006 and 2010 in TECUs

Period	Product	BIAS	Standard Deviation	RMS	Number of Comparisons
214–244, 2004	IGS Final	−0.43	3.22	3.25	1,167,252
	Final	−1.29	3.14	3.40	1,167,252
	Predicted	−0.97	4.00	4.12	1,167,252
	Frozen	−1.54	4.16	4.43	1,167,252
335–365, 2006	IGS Final	−1.65	2.83	3.28	1,249,145
	Final	−2.02	2.78	3.44	1,249,145
	Predicted	−1.97	3.70	4.19	1,249,145
	Frozen	−1.87	4.16	4.56	1,249,145
184–355, 2010	IGS Final	−0.57	2.77	2.83	4,341,359
	Final	−1.30	2.61	2.92	4,341,325
	Predicted	−1.83	3.34	3.81	4,341,359
	Frozen	−1.83	3.61	4.05	4,341,359

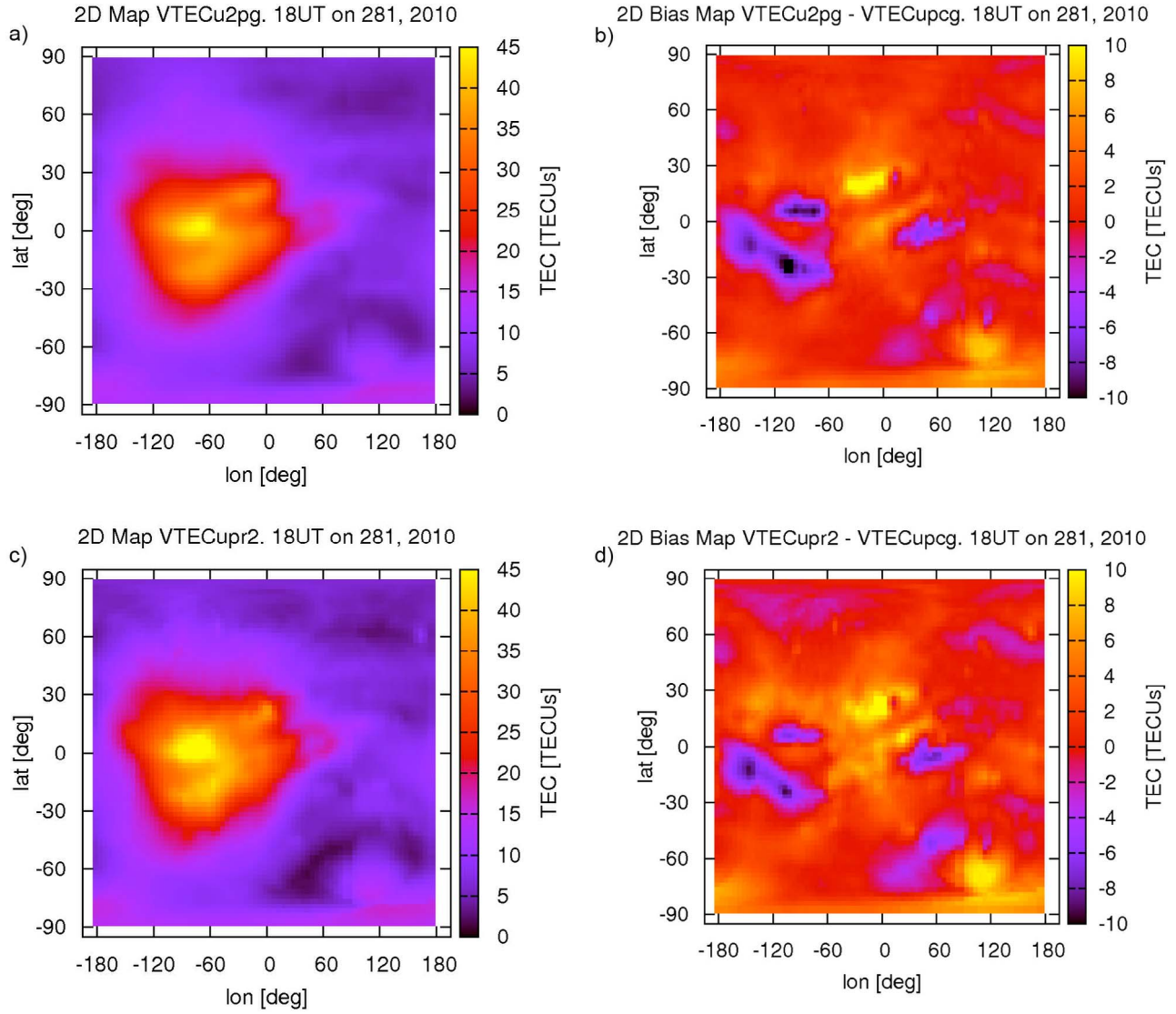

Figure 3. (a, c) Representation of the UPC Predicted/Time-invariant VTEC map in geomagnetic longitude/latitude range at 18 UT on day 281, 2010, and (b, d) the differences with respect to the corresponding UPC Final VTEC map. Note that the UPC Final VTEC map for this day is depicted in Figure 2. The same range of colors are used but for different scales: 0 to 45 for Figures 3a and 3c and −10 to 10 for Figures 3b and 3d.

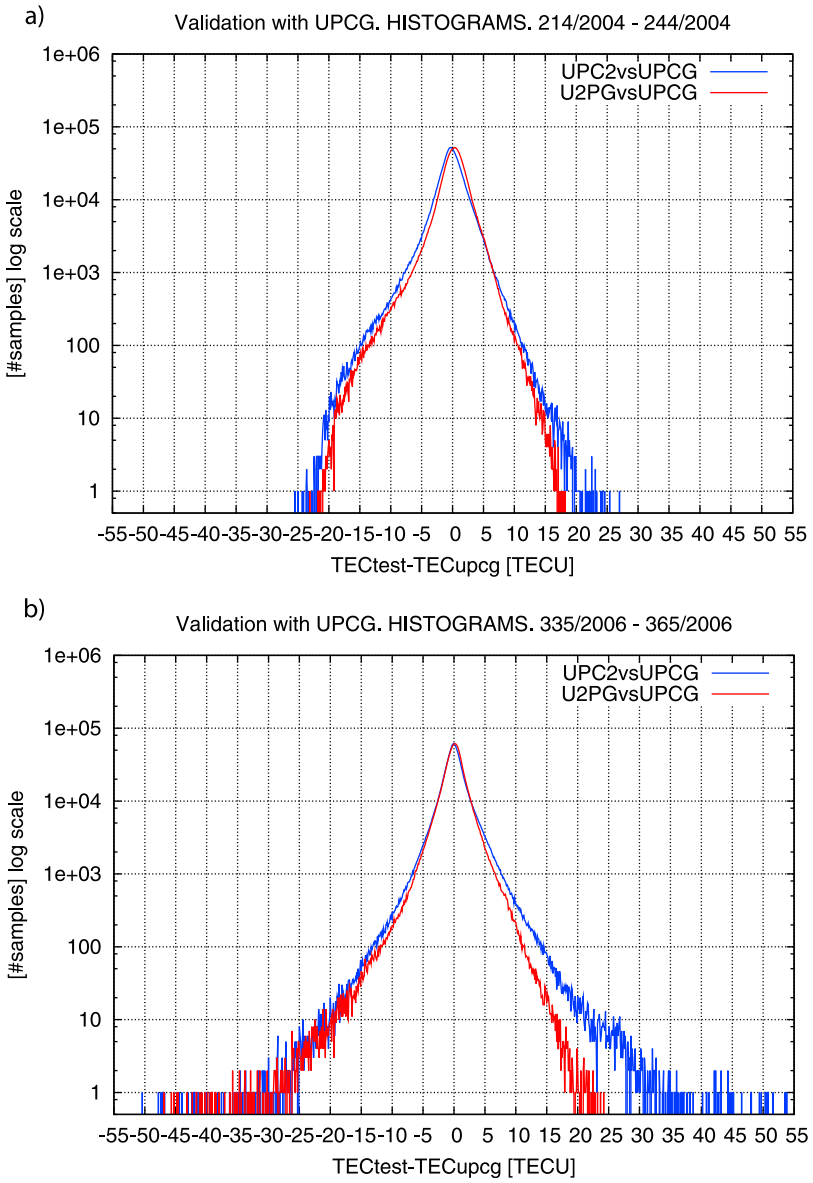

Figure 4. Histogram on a semi-logarithmic scale of the differences between UPC predicted/Time-invariant VTEC values and the UPC Final IGS VTEC values (*U2PG* versus *UPCG* and *UPC2* versus *UPCG*, respectively) for the periods from days (a) 214–244, 2004 and (b) 335–365, 2006.

Also in this case, Time-invariant VTEC maps are generated using UPC Final IGS products.

[33] The period in 2010 covers 115 days from day of year 184 to 355. Here, performance comparisons could be made with the other IAACs predicted products for the exact same period. During this time interval, a geomagnetic storm happened on days 215–216, but no powerful X-class solar flares were recorded by GOES. The *Kp* index ranged from 0 to 6.7, with a mean value of 1.24 and its maximum at 00–03 UT on day 216.

[34] Note that the selected days in 2010 are not consecutive, as days for which rapid/predicted/time-invariant products of each IAAC center and the combined IGS products were not available were discarded. Additionally, days where there were a low number of JASON observations (in this work, less than 10,000 observations) were also discarded. Furthermore, it is also required that the combined IGS Predicted product for two days ahead (*I2PG*) was generated using the three individual predicted products from CODE, ESA and UPC. In this regard, the first day in which all three products were combined was day of year 184 in 2010. Taking these factors into account, a total of 115 days were selected between days 184 and 355 in 2010.

3.5. Test Against UPC Final IGS Product

[35] In Table 1, the differences between the UPC Predicted/Time-invariant products (the test data) and the UPC Final IGS product (the reference data) are compared for the three selected periods. The bias, standard deviation, RMS and minimum and maximum values are included expressed in TEC units or TECUs (1 TECU is equivalent to 10^{16} el/m^2), as well as the number of comparisons.

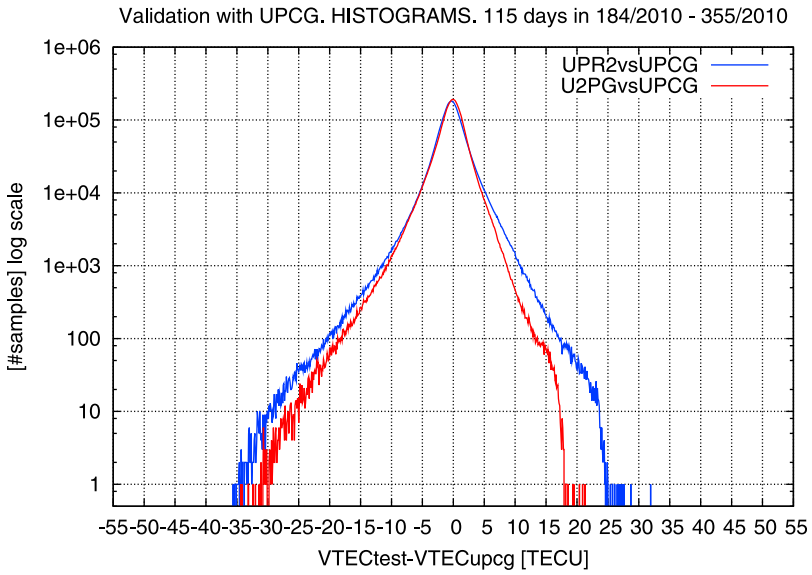

Figure 5. Histogram of the differences between the UPC predicted/Time-invariant VTEC values and the UPC Final IGS VTEC values (*U2PG* versus *UPCG* and *UPR2* versus *UPCG*, respectively) on a semi-logarithmic scale, shown for the 115 day period from days 184–355, 2010.

[36] In this table, the UPC prediction model systematically provides better results than the Time-invariant product in a solar-fixed reference frame. In fact, the RMS of the differences between the UPC Time-invariant product and the UPC Final IGS product (*UPR2*/*UPC2* versus *UPCG*) is 10.92%, 18.50% and 16% worse than the RMS of the differences for UPC Predicted product (for 2004, 2006 and 2010 periods, respectively). Regarding the bias over all grid points in each interval, a systematic under- or overestimation cannot be confirmed. The UPC prediction model tends to overestimate VTEC in the 2004 interval while it underestimates VTEC in the 2010 interval.

[37] It is also interesting to look at the histograms of the above mentioned comparisons. In Figures 4 and 5, the histograms of the differences between the UPC Predicted/Time-invariant products and the UPC Final IGS products (*U2PG* versus *UPCG* and *UPR2*/*UPC2* versus *UPCG*, respectively) are plotted. For the selected periods, the results for the UPC Predicted product show lower residual values than for the UPC Time-invariant ionosphere. Both the negative and positive tails of the distributions finish at lower values for the UPC Predicted product (see also Table 1). A significant improvement is obtained looking at the relative differences between the maximum values of the UPC Predicted product and the Time-invariant product. The Time-invariant maximum values are 48.35 %, 122.31 % and 49.77 % higher than the Predicted ones (in 2004, 2006 and 2010 periods, respectively), while the minimum values also experience higher values but in a lower proportion.

[38] These results are caused by the disturbed conditions affecting the three selected periods. Although such effects cannot be predicted, the UPC Predicted product is more robust against them than the Time-invariant product. This is because the disturbed VTEC values are directly used to generate the Time-invariant product two days afterwards (for $T + 2$). In contrast, the UPC Predicted product is generated taking into account a set of seven VTEC files as input dataset so that the impact two days afterwards is lower (see section 2).

[39] The histograms in 2004 interval (Figure 4a) show an important asymmetry affecting the negative values (underestimation of VTEC). This is due to the increase in VTEC related to the storm that occurred at the end of August, 2004, which is not predicted. The histograms in 2006 interval (Figure 4b) show that the UPC Predicted product histogram (*U2PG* versus *UPCG*) is asymmetrical with a higher number of samples on the negative side due to the underestimation of VTEC for the disturbed period on day 348. On the other hand, the Time-invariant histogram (*UPC2* versus *UPCG*) is symmetrical because the underestimation on day 348 is compensated by the corresponding overestimation on day 350 (as the UPC Time-invariant product is generated from the disturbed UPC Final IGS product). Finally, for the 115 days in 2010 (Figure 5), the histograms show a sudden decrease in the positive tail related to the days not taken into account (remember that the selected days may not be consecutive) that are close to the most disturbed days in the period being considered.

3.6. Test Against JASON Data

[40] For the external validation with JASON, bias, standard deviation, RMS and the number of comparisons are shown in Table 2. In this table, the combined Final IGS product and the UPC Final IGS product against the JASON VTEC values are shown as reference (*IGSG* versus JASON and *UPCG* versus JASON, respectively). The standard deviation, which is not affected by the JASON offset (see section 3.2), and RMS results obtained in cases where the UPC prediction model is applied are systematically better than for the time-invariant ionosphere approach.

[41] On the one hand, the standard deviation results of the differences between UPC Predicted products and the

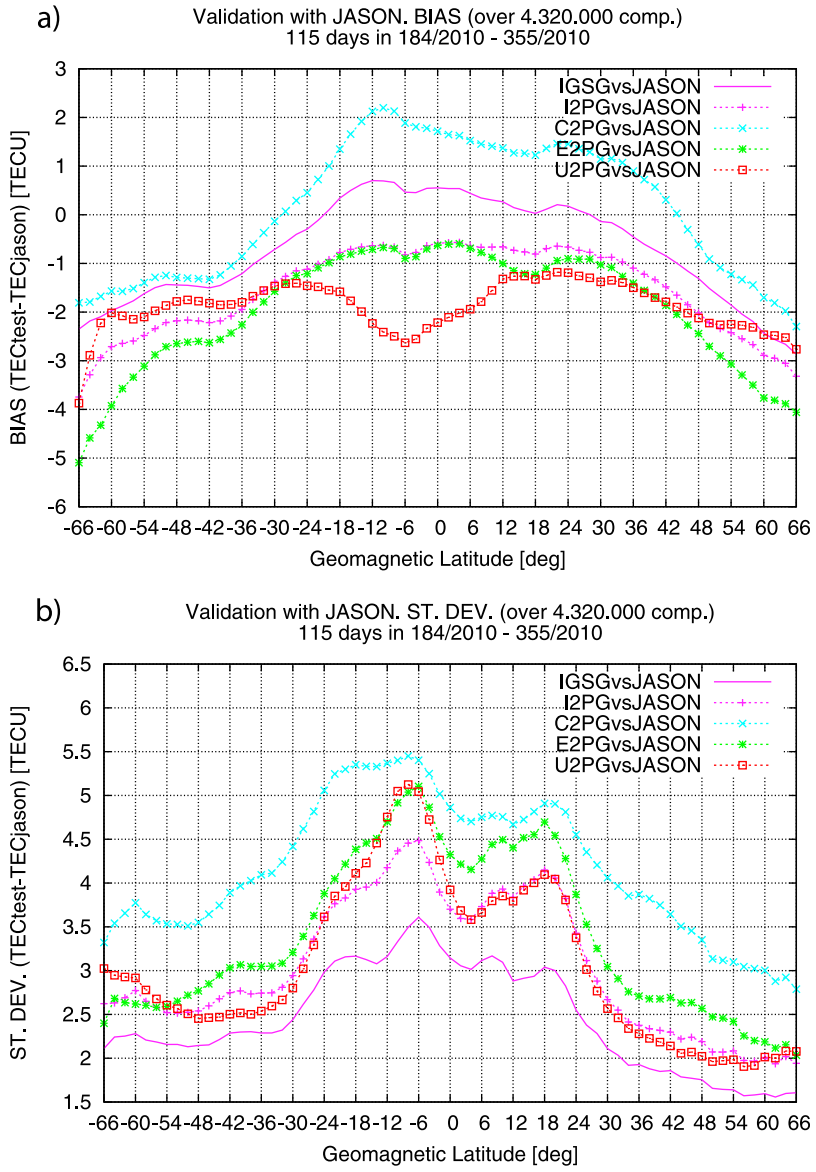

Figure 6. (a) Bias) and (b) standard deviation of the differences between the IGS, CODE, ESA and UPC two days ahead predicted products and the JASON data (*I2PG* versus JASON, *C2PG* versus JASON, *E2PG* versus JASON and *U2PG* versus JASON; dotted lines). The assessed values were in two degrees geomagnetic latitude bins for the 115 day period from days 184–355, 2010. The differences between the combined Final IGS product and JASON data are also included as reference (*IGSG* versus JASON; continuous line).

JASON reference data (*U2PG* versus JASON) are 27.39%, 33.09% and 27.97% worse than that of UPC Final IGS product (*UPCG* versus JASON) for the 2004, 2006 and 2010 periods, respectively. On the other hand, the standard deviation results for the UPC Time-invariant products (*UPR2*/*UPC2* versus JASON) are 32.48%, 49.64% and 38.31% worse than that of the UPC Final IGS product (*UPCG* versus JASON) for the three periods. In this way, the percentage of improvement, which is the decrease in the variability with respect to JASON, is 5.09 %, 16.55 % and 10.34 % for the three periods. Apart from that, the bias results show that all IGS VTEC values are underestimated with respect to the JASON ones. This is compatible with the offset affecting JASON (see section 3.2).

3.6.1. Latitudinal Behavior

[42] The dependence of model performance on the geomagnetic latitude is studied for the different IAACs as well as for the IGS combination for the period in 2010. The latitudinal behavior of the VTEC predictions can be analyzed; for example, this can be accomplished by plotting the bias and standard deviation values in multiple geomagnetic latitude bins (of two degrees in this work) when comparing these values with external JASON data.

[43] In Figure 6a, the bias of the differences between the predicted products and the JASON data is depicted. In it, the plasmaspheric component of the VTEC clearly affects GNSS measures at low latitudes and thus produces the typical (inverted) *U* shape in terms of the latitude [*Aragón-Angel*,

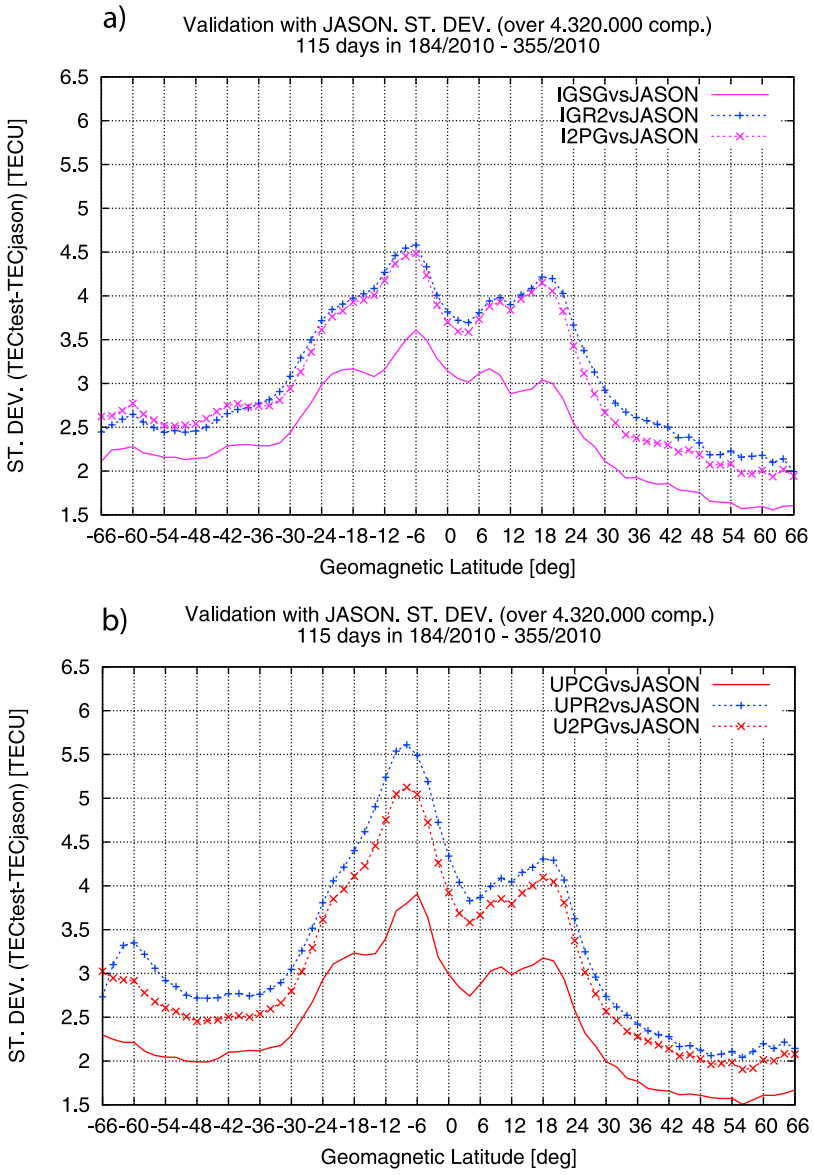

Figure 7. Standard deviation of the differences between the UPC and IGS Final/Predicted/Time-invariant products and the JASON data. (a) *IGSG*/*I2PG*/*IGR*2 versus JASON. (b) *UPCG*/*U2PG*/*UPR*2 versus JASON. The assessed values were in two degrees geomagnetic latitude bins for the 115 day period from days 184–355, 2010.

2010]. Note that the bias of the differences between UPC Predicted product and JASON data (*U2PG* versus JASON) experiences a decrease in the equatorial region centered at −6 degrees. This VTEC underestimation is also present in the bias of the differences obtained from the UPC Rapid IGS products (*UPRG* versus JASON; not plotted). In this regard, note that UPC Rapid IGS products are used as the input dataset for the UPC prediction model to generate the UPC predicted VTEC maps (see section 2). This issue will have to be further investigated in the future.

[44] In Figure 6b, the standard deviation of the differences are plotted showing a clear correlation with the effect of the Appleton-Hartree equatorial anomaly. In particular, higher standard deviation values are obtained when modeling the anomaly in the Southern Hemisphere.

[45] First of all, the behavior of the differences between the combined IGS Predicted product and the JASON data (*I2PG* versus JASON) is quite similar to the differences between the combined IGS Final product and the JASON data (*IGSG* versus JASON; used as a point of reference). This shows the importance of the combination process as the performance of the combined IGS Predicted product is better than the individual predicted products. Additionally, the UPC Predicted product shows a good overall standard deviation performance compared to the results obtained from CODE and ESA Predicted products. Nevertheless, VTEC is still underestimated when modeling the Southern Hemisphere peak. This feature seems to be related to the decrease in the bias that can be seen in Figure 6a.

[46] In Figure 7a, the standard deviation in terms of the latitude band is shown for the differences between UPC

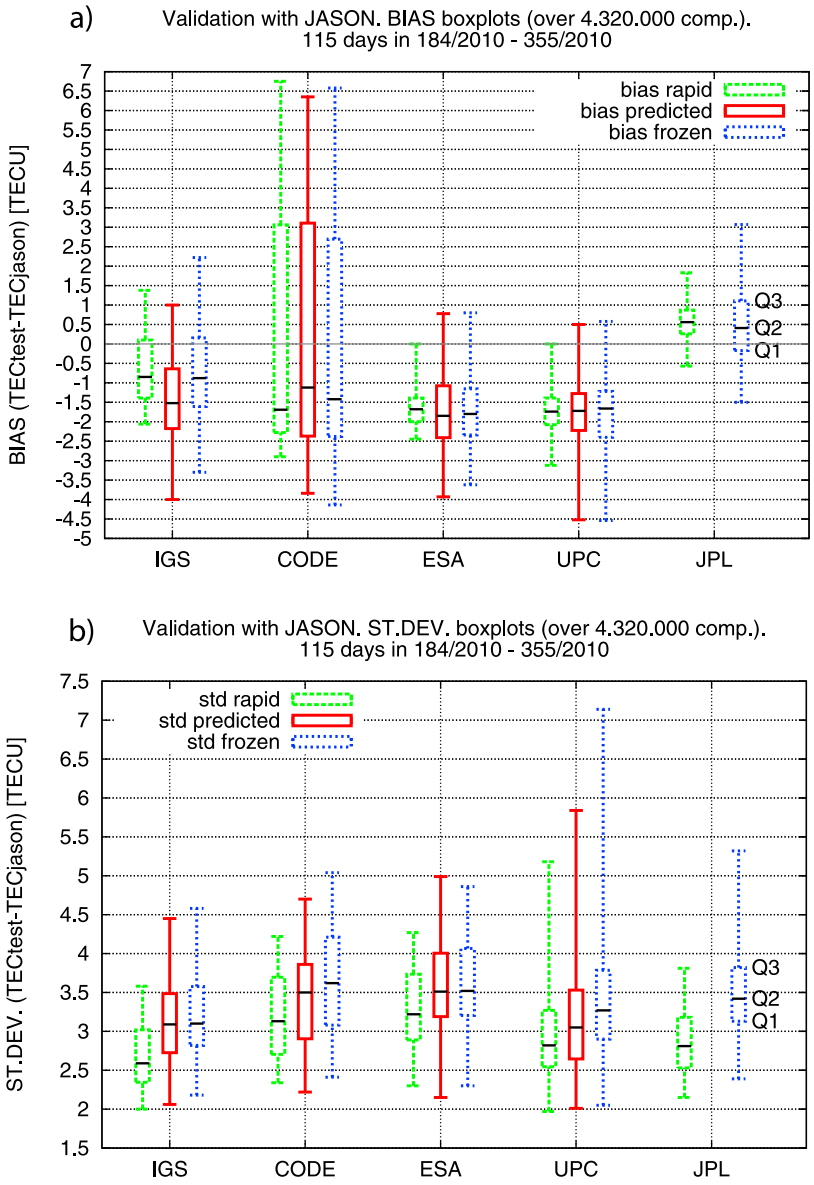

Figure 8. (a) Bias and (b) standard deviation boxplots of the differences between the rapid/predicted/time-invariant test data and the JASON reference data for the 115 day period from days 184–355, 2010. The results for the rapid/predicted/time-invariant test data are plotted in the leftmost/center/rightmost boxplots for each test source on the x-axis. Test data include CODE, ESA, UPC, JPL and combined IGS rapid products, preliminary predicted products, and time-invariant products (generated internally), if available.

Final/Predicted/Time-invariant product and the JASON data (*UPCG*/*U2PG*/*UPR2* versus JASON). The UPC Predicted product systematically yields better results than for the UPC Time-invariant ionosphere. In Figure 7b, the combined IGS Predicted product (*I2PG* versus JASON) behavior is shown and yields to slightly better results than for the IGS Time-invariant ionosphere.

3.6.2. Boxplots

[47] The boxplot is a parametric representation that summarizes graphically the important statistics of a sample. This allows to grasp the important features of the distribution (for a more detailed description of this tool, see the last paragraph in section 3.3). Figure 8 shows the boxplots for the GNSS VTEC map sources under consideration, i.e., IGS as well as the IAACs providing predicted products, against JASON data. For each of these products, three types of data are plotted for comparison: the IGS and IAACs official Rapid products (on the left hand side for each case; indicated with a dotted green line), the preliminary predicted products (in the center; a continuous red line) and the Time-invariant products (on the right; a dotted blue line). Note that the rapid products are being used rather than the final ones to show the variability of the most recent VTEC maps used to generate the predictions (at least in the UPC case).

[48] The results show that there is a strong dependence on the performance of the corresponding rapid products.

Applying a prediction model leads to better results than does the use of the time-invariant ionosphere for all of the individual products that are taken into account. In addition, applying a prediction model generates an increase in the data dispersion which is, in general, lower than the dispersion observed in the time-invariant ionosphere VTEC maps.

[49] In the case of the differences between the UPC Predicted product and JASON reference data (*U2PG* versus JASON), the obtained standard deviation is lower than that observed for the other IAACs for the analyzed dataset of 115 days in 2010. Note that the standard deviation values for the differences between UPC Predicted product and JASON (*U2PG* versus JASON) and the differences obtained from a CODE Predicted product for two days ahead forecast (*C2PG* versus JASON) are clearly better than that observed for the corresponding time-invariant VTEC values (i.e., *UPR*2 versus JASON and *COR*2 versus JASON). It should also be noted that the standard deviation boxplot displays a slightly lower median (*Q*2 quartile) for the UPC Predicted product than for the combined IGS Predicted product. Regarding the bias boxplots, which are depicted in Figure 8a, the differences for the UPC Predicted product (U2PG versus JASON) and the ESA Predicted product (E2PG versus JASON) indicate a similar performance compared to the corresponding Time-invariant products (*UPR*2 versus JASON and *ESR*2 versus JASON). In the case of the CODE Predicted products, the bias results show a larger dispersion but a lower median.

[50] Regarding the combined forecast product, the differences between the combined IGS Predicted product and the JASON data (*I2PG* versus JASON) indicate a slightly better performance compared to the corresponding Time-invariant product (*IGR*2 versus JASON) based on the standard deviation values. However, the bias results are worse, as can be seen in Figure 8a. This is of great importance because it indicates that using IGS Time-invariant prediction (*IGR*2) may lead to better results than considering the combined IGS Predicted product. This could be related to the weighting scheme that has been applied during the combination process [see *Orús et al.*, 2007]. For the period under consideration, this may be related to the bias performance of the JPL products as there is no predicted product provided by JPL. Note that the JPL Rapid IGS product (*JPRG*) and the Time-invariant product (*JPR*2) have also been added for reference. In this regard, it seems that the future availability of a JPL predicted product could further improve the combined product.

4. Conclusions and Further Steps

[51] The use of representative datasets has demonstrated that the UPC prediction model works well, particularly when results are compared with those obtained by other IAACs. This study shows that applying a prediction model leads to better results than the use of time-invariant ionosphere for two days ahead predictions. These conclusions can be extended to the prediction models of CODE and ESA as well. Nevertheless, the predicted products still have the potential for further improvements to achieve better results. The results obtained here suggest that the inclusion of a potential future predicted product from JPL could be of great importance. In this way, the accuracy of the combined IGS predicted product could be increased significantly. In addition, it might be possible to apply a prediction model directly to the IGS final/rapid products, although this may deviate from the philosophy of IGS of combining several independent products.

[52] We would like to emphasize the fact that this first UPC Predicted product has been automatically generated and uploaded to the IGS main server since September 2009 without major problems. In this context, another improvement for the near future is to make the generation of the predicted UPC VTEC maps more robust against corrupted files in the input dataset of the model. Such files would penalize the performance of the predicted VTEC products if undetected.

[53] Currently, the prediction of TEC disturbances induced by geomagnetic storms, solar flares and other impulsive events is not yet accounted for in the UPC prediction model design. Nevertheless, the handling of these phenomena should be a key priority in future. In this regard, tests feeding the UPC prediction model with physical information including indexes of solar/geomagnetic activity, such as *Kp* and Solar Flux, are currently being conducted. Also in this context, a detector of solar flares facing the earth that may be used to input additional physical information into the prediction model has been developed by the authors [see *García-Rigo et al.*, 2008, 2006]. Apart from that, further improvement could be obtained by using a geomagnetic grid rather than a geographic one or by using a subset of the regression coefficients (for instance, they could be determined by feature selection [*Hastie et al.*, 2001]).

[54] Another key point in assessing UPC prediction model performance should be its comparison with Klobuchar, NeQuick and the International Reference Ionosphere (IRI), among other models [see, e.g., *Klobuchar*, 1986; *Nava et al.*, 2008; *Bilitza*, 2001].

[55] Last but not least, the future multifrequency/multiconstellation GNSS scenario, as well as the deployment of more permanent GNSS stations distributed worldwide within the IGS framework, will enable the generation of improved UPC Final/Rapid VTEC maps. Consequently, better predicted products could be released in case these products are used as input data for the prediction models, as is the case of the UPC prediction model.

[56] **Acknowledgments.** The authors are grateful to IGS for providing ionosphere data products and to NASA/CNES for JASON data. Thanks to Carey Noll as well for her support in the broadcasting of the predicted products. Also, the first author wishes to express his gratitude to Andrzej Krankowski and Pawel Wielgosz for their support during his visit to the University of Warmia and Mazury in Olsztyn, Poland. This study has been partially funded by the Spanish Ministry of Science, Technology and Innovation under the PTQ-10-03279, PTQ-10-03294, CTM2010-21312-C03-02 and ESP2007-62676 projects, as well as by the BUCEADOR project TEC2009-14094-C04-01.

References

Afraimovich, E. L., E. I. Astafyeva, and I. V. Zhivetiev (2006), Solar activity and global electron content, *Dokl. Earth Sci.*, *409A*(N6), 921–924.

Ahmed, N., T. Natarajan, and K. R. Rao (1974), Discrete Cosine Transform, *IEEE Trans. Comput.*, *C-23*, 90–93.

Aragón-Angel, A. (2010), Contributions to ionospheric electron density retrieval, dissertation, Doctorate in Aerospace Science and Technology, Tech. Univ. of Catalonia, Barcelona, Spain.

Azpilicueta, F., and C. Brunini (2008), Analysis of the bias between TOPEX and GPS vTEC determinations, *J. Geod.*, *83*(2), 121–127.

Bilitza, D. (2001), International Reference Ionosphere 2000, *Radio Sci.*, *36*, 261–275, doi:10.1029/2000RS002432.

Bust, G. S., and C. N. Mitchell (2008), History, current state, and future directions of ionospheric imaging, *Rev. Geophys.*, *46*, RG1003, doi:10.1029/2006RG000212.

Cander, L. R., M. M. Milosavljevic, S. S. Stankovic, and S. Tomasevic (1998), Ionospheric forecasting technique by artificial neural network, *Electron. Lett.*, *34*(6), 1573–l574.

Dick, M. I., M. F. Levy, L. R. Cander, I. Kutiev, and P. Muhtarov (1999), Short-term ionospheric forecasting over Europe, *IEE Publ.*, *461*, 105.

Dow, J. M., R. E. Neilan, and C. Rizos (2009), The International GNSS Service in a changing landscape of Global Navigation Satellite Systems, *J. Geod.*, *83*, 191–198.

Francis, N. M., P. S. Cannon, A. G. Brown, and D. S. Broomhead (2000), Nonlinear prediction of the ionospheric parameter on hourly, daily, and monthly timescales, *J. Geophys. Res.*, *105*(A6), 12,839–12,849.

García-Rigo, A., M. Hernández-Pajares, J. M. Juan, and J. Sanz (2006), Solar flare detection system based on Global Positioning System data: First results, *Adv. Space Res.*, *39*(5), 889–895.

García-Rigo, A., M. Hernández-Pajares, J. M. Juan Zornoza, and J. Sanz (2008), Real time ionospheric TEC monitoring method applied to detect solar flares, paper presented at General Assembly, Eur. Geosci. Union, Vienna.

Hastie, T., R. Tibshirani, and J. Friedman (2001), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York.

Hernández-Pajares, M. (2004), IGS Ionosphere WG status report: Performance of IGS Ionosphere TEC Maps, paper presented at IGS Technical Meeting, Int. GNSS Serv., Bern.

Hernández-Pajares, M., J. M. Juan, J. Sanz, and O. L. Colombo (2000), Application of ionospheric tomography to real-time GPS carrier-phase ambiguities resolution, at scales of 400–1000 km and with high geomagnetic activity, *Geophys. Res. Lett.*, *27*(13), 2009–2012.

Hernández-Pajares, M., J. M. Juan, J. Sanz, A. Garcia-Rigo, J. Feltens, A. Komjathy, S. C. Schaer, and A. Krankowski (2009), The IGS VTEC maps: A reliable source of ionospheic information since 1998, *J. Geod.*, *83*(3–4), 263–275.

Hill, S. M., et al. (2005), The NOAA Goes-12 Solar X-Ray Imager (SXI): 1. Instrument, operations, and data, *Sol. Phys.*, *226*, 255–281.

Hoaglin, D., F. Mosteller, and J. Tukey (1983), *Understanding Robust and Exploratory Data Analysis*, Wiley, New York.

Klobuchar, J. A. (1986), Design and characteristics of the GPS ionospheric time delay algorithm for single frequency users, paper presented at PLANS 1986 Position, Location and Navigation Symposium, Inst. of Electr. and Electron. Eng., Las Vegas, Nev.

Krankowski, A., and M. Hernández-Pajares (2008), Status of the IGS ionosphere products and future developments, paper presented at IGS Analysis Center Workshop, Natl. Geod. Surv., Miami Beach, Fla.

Krankowski, A., W. Kosek, L. W. Baran, and W. Popiński (2005), Wavelet analysis and forecasting of VTEC obtained with GPS observations over European latitudes, *J. Atmos. Sol. Terr. Phys.*, *67*, 1147–1156.

Muhtarov, P., and I. Kutiev (1999), Autocorrelation method for temporal interpolation and short-term prediction of ionospheric data, *Radio Sci.*, *34*(2), 459–464, doi:10.1029/1998RS900020.

Muhtarov, P., I. Kutiev, Lj. R. Cander, B. Zolesi, G. de Franceschi, M. Levy, and M. Dick (2001), European ionospheric forecast and mapping, *Phys. Chem. Earth, Part C, Sol. Terr. Planet. Sci.*, *26*(5), 347–351.

Nava, B., P. Coisson, and S. M. Radicella (2008), A new version of the NeQuick ionosphere electron density model, *J. Atmos. Sol. Terr. Phys.*, 1856–1862, doi:10.1016/j.jastp.2008.01.015.

Oppenheim, A. V., and R. W. Schafer (2010), *Discrete-Time Signal Processing*, 3rd ed., Prentice Hall, Upper Saddle River, N. J.

Orús, R., L. R. Cander, and M. Hernández-Pajares (2007), Testing regional vertical total electron content maps over Europe during the 17–21 January 2005 sudden space weather event, *Radio Sci.*, *42*, RS3004, doi:10.1029/2006RS003515.

Orús, R., M. Hernández-Pajares, J. M. Juan, J. Sanz, A. Aragón-Angel, and A. García-Rigo (2010), Real time application of TOMION model, paper presented at Beacon Satellite Symposium, Int. Union of Radio Sci., Barcelona, Spain.

Schaer, S. (1999), Mapping and predicting the Earth's ionosphere using the Global Positioning System, dissertation, Astron. Inst., Univ. of Bern, Bern.

Schaer, S., W. Gurtner, and J. Feltens (1998), IONEX: The IONosphere Map EXchange Format Version 1, February 25, 1998, paper presented at IGS Analysis Centers Workshop, ESA's Space Oper. Cent., Darmstadt, Germany.

Sivestrin, P., M. Berger, Y. Kerr, and J. Font (2001), ESA's second earth explorer opportunity mission: The soil moisture and ocean salinity mission SMOS, *IEEE Geosci. Remote Sens. Newsl.*, *118*, 11–14.

Stanislawska, I., and Z. Zbyszynski (2001), Forecasting of the ionospheric quiet and disturbed f_0F2 VALues at a single location, *Radio Sci.*, *36*, 1065–1071, doi:10.1029/1999RS002242.

Tulunay, E., E. T. Senalp, S. M. Radicella, and Y. Tulunay (2006), Forecasting total electron content maps by neural network technique, *Radio Sci.*, *41*, RS4016, doi:10.1029/2005RS003285.

A. Aragón-Angel, A. García-Rigo, M. Hernández-Pajares, J. M. Juan, D. Salazar, and J. Sanz, Research Group of Astronomy and Geomatics, Technical University of Catalonia, E-08034 Barcelona, Spain. (agarcia@ma4.upc.edu)

E. Monte, Department of Signal Theory and Communications, Technical University of Catalonia, E-08034 Barcelona, Spain.