Space Weather

RESEARCH ARTICLE
10.1029/2018SW001987

An Ionosphere Specification Technique Based on Data Ingestion Algorithm and Empirical Orthogonal Function Analysis Method

Ercha Aa[1,2], Aaron Ridley[2], Wengeng Huang[1], Shasha Zou[2], Siqing Liu[1,3], Anthea J. Coster[4], and Shunrong Zhang[4]

[1]National Space Science Center, Chinese Academy of Sciences, Beijing, China, [2]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA, [3]College of Earth and Planetary Science, University of Chinese Academy of Sciences, Beijing, China, [4]Haystack Observatory, Massachusetts Institute of Technology, Cambridge, MA, USA

Key Points:
- The Madrigal TEC data are ingested into the NeQuick 2 model through deriving an effective ionization parameter (Az)
- The Empirical Orthogonal Function (EOF) analysis technique is used to construct a parameterized time-varying Az model to make a prediction
- The TEC data ingestion and EOF modeling are effective in bringing certain systematic improvement of ionosphere now-cast/forecast

Correspondence to:
E. Aa,
aercha@nssc.ac.cn

Citation:
Aa, E., Ridley, A. J., Huang, W., Zou, S., Liu, S., Coster, A. J., & Zhang, S. (2018). An ionosphere specification technique based on data ingestion algorithm and empirical orthogonal function analysis method. *Space Weather*, *16*, 1410–1423. https://doi.org/10.1029/2018SW001987

Received 22 JUN 2018
Accepted 11 SEP 2018
Accepted article online 17 SEP 2018
Published online 25 SEP 2018

Abstract A data ingestion method in reproducing ionospheric electron density and total electron content (TEC) was developed to incorporate TEC products from the Madrigal Database into the NeQuick 2 model. The method is based on retrieving an appropriate global distribution of effective ionization parameter (*Az*) to drive the NeQuick 2 model, which can be implemented through minimizing the difference between the measured and modeled TEC at each grid in the local time-modified dip latitude coordinates. The performance of this Madrigal TEC-driven-NeQuick 2 result is validated through the comparison with various International Global Navigation Satellite Systems Services global ionospheric maps and ionosonde data. The validation results show that a general accuracy improvement of 30–50% can be achieved after data ingestion. In addition, the empirical orthogonal function (EOF) analysis technique is used to construct a parameterized time-varying global *Az* model. The quick convergence of EOF decomposition makes it possible to use the first six EOF series to represent over 90% of the total variances. The intrinsic diurnal variation and spatial distribution in the original data set can be well reflected by the constructed EOF base functions. The associated EOF coefficients can be expressed as a set of linear functions of $F_{10.7}$ and Ap indices, combined with a series of trigonometric functions with annual/seasonal variation components. The NeQuick TEC driven by EOF-modeled *Az* shows 10–15% improvement in accuracy over the standard ionosphere correction algorithm in the Galileo navigation system. These preliminary results demonstrate the effectiveness of the combined data ingestion and EOF modeling technique in improving the specifications of ionospheric density variations.

1. Introduction

The Earth's ionosphere is a highly variable region of space that exhibits both climatological variations and weather disturbances. In order to better mitigate the detrimental effects of the ionosphere on radio propagation and satellite navigation, it is of great importance to provide timely and reliable ionospheric specification and prediction through utilizing various ionospheric empirical and/or theoretical models. Ionospheric empirical models, such as International Reference Ionosphere (IRI; ; Bilitza, 2001; Bilitza & Reinisch, 2008) and NeQuick (Giovanni & Radicella, 1990; Nava et al., 2008; Radicella & Leitinger, 2001), are mainly built on the basis of statistical analysis of large data sets. Empirical models have the merits of simplicity and accuracy in reproducing the climatological characteristics of the ionosphere, yet are limited to the way the suitable function was chosen and the quality of the data that were used. Ionospheric theoretical models are constructed on the basis of fundamental physical laws (mass balance, energy balance, heat transfer relations, etc.), and can be run under a much wider set of conditions to test the theories, yet are limited by a lack of accurate estimation of the external drivers and initial conditions. With the continuous increase of ionospheric measurements from diverse sources, such as the total electron content (TEC) data from ground-based Global Navigation Satellite Systems (GNSS) networks, radio occultation data from low-Earth orbit satellites, global digisonde profiles, in situ *Ne* measurements, and ultraviolet airglow data, it has been realized that the dynamic processes and subtle variations in the ionosphere could be better specified and predicted through data assimilation/ingestion techniques to incorporate ionospheric observations into background models (e.g., Nava et al., 2011; Schunk et al., 2014; Yue et al., 2012).

Data assimilation and ingestion techniques are usually associated with each other yet not clearly distinguished. For data assimilation, the observations are projected by certain optimization algorithm (e.g., Kalman filter, 3-D/4-D variational method) into proper global or regional scales to get the best estimation of the external drivers and initial/boundary conditions of the first-principle ionospheric models. For example, Utah State University constructed a Global Assimilation of Ionospheric Measurements, which uses a physics-based Ionosphere Forecast Model and a Kalman filter as a basis for assimilating a diverse set of near real-time measurements (Scherliess et al., 2004, 2006; Schunk et al., 2004, 2005). The Jet Propulsion Laboratory and University of Southern California have cooperatively constructed another Global Assimilation Ionospheric Model (JPL/USC Global Assimilation Ionospheric Model), which uses a traditional Kalman filter method to estimate the three-dimensional density state and a four-dimensional variational approach to estimate ionospheric drivers such as neutral winds and the equatorial E×B drift (Mandrake et al., 2005; Pi et al., 2003; Wang et al., 2004). Some studies use sophisticated empirical models to define the a priori state in order to implement data assimilation, such as Ionospheric Data Assimilation Three-Dimensional (Bust et al., 2004, 2007), Electron Density Assimilative Model (Angling & Cannon, 2004; Angling & Khattatov, 2006), North American/United States TEC (Fuller-Rowell et al., 2006), and China assimilation TEC Model (Aa et al., 2015, 2016). Moreover, there are extensive studies that described the development of ionospheric and thermospheric data assimilation models/procedures (e.g., Komjathy et al., 2010; Lee et al., 2012; Matsuo et al., 2012; Pi et al., 2009; Schunk et al., 2014; Yue et al., 2011, 2012; Zhu et al., 2012).

However, developing an ionospheric data assimilation model is very complicated with many trade-offs and approximations. The computational ease and simplicity makes data ingestion techniques readily accessible to the wide audience of space weather research and application communities. Generally, data ingestion differs from data assimilation in the following two aspects: first, instead of using complex physics-based models, data ingestion usually uses simplified and parameterized models in terms of a given set of *effective*-driven factors; second, data ingestion usually drives the background model toward experimental data sets by using a simple optimization algorithm such as least squares estimation to minimize the deviations between experimental and model values, which has the merits of computation efficiency in contrast to time-consuming calculation in an assimilation process that involves complicated error covariance matrices. There are some studies that ingested global ionosonde measurements into the IRI empirical model. For example, the IRI real-time assimilative mapping incorporate data from the Global Ionospheric Radio Observatory to adjust the Consultative Committee of International Radio coefficients (Galkin et al., 2012). Moreover, some studies adapted the NeQuick empirical model by ingesting GNSS-derived slant TEC (Nava et al., 2006), global ionospheric maps (GIMs) of vertical TEC (Nava et al., 2005), and COSMIC-derived radio occultation TEC/Ne (Brunini et al., 2011; Nava et al., 2011). Although these NeQuick data ingestion attempts are able to improve the model capability in specifying three-dimensional ionospheric *Ne* by updating the ionization level parameters *Az*, there are a few things that worth noting: (1) Since preliminary fitting and approximations are needed to generate vertical TEC GIMs, ingesting these *secondary* data could get complete global *Ne* profiles, but with compromises in accuracy and resolution; (2) Ingesting GNSS-derived slant TEC can improve the accuracy of the reconstructed global ionospheric *Ne* with acceptable computation, though it is more suitable at a single station or over regional grids. Therefore, in the current study, we will use a modified data ingestion technique to adapt the NeQuick 2 model by ingesting TEC products derived from the Madrigal Database of the Massachusetts Institute of Technology Haystack Observatory, then an empirical orthogonal function (EOF) analysis method will be used to give time-dependent specifications of the three-dimensional electron density of the ionosphere. We aim to make this product applicable to precisely reproducing the global ionospheric morphology for scientific study and to providing an alternative ionospheric correction algorithm for GNSS single-frequency users.

The rest of the paper is organized as follows: the NeQuick 2 model and Madrigal TEC data will be briefly introduced in section 2. The data ingestion technique and its validation will be given in section 3. The EOF modeling method and its verification will be presented in section 4, and the conclusions in section 5.

2. Description of the Model and Data

The NeQuick 2 (Nava et al., 2008) is used here as a background model to describe the global distribution of electron density. The NeQuick 2 model is developed at the Aeronomy and Radio propagation Laboratory of the Abdus Salam International Centre for Theoretical Physics, Trieste, Italy, and at the Institute for Geophysics, Astrophysics and Meteorology of the University of Graz, Austria. NeQuick 2 describes the vertical profile of bottomside *Ne* in terms of a modified DGR profile formulation (Giovanni & Radicella, 1990; Radicella &

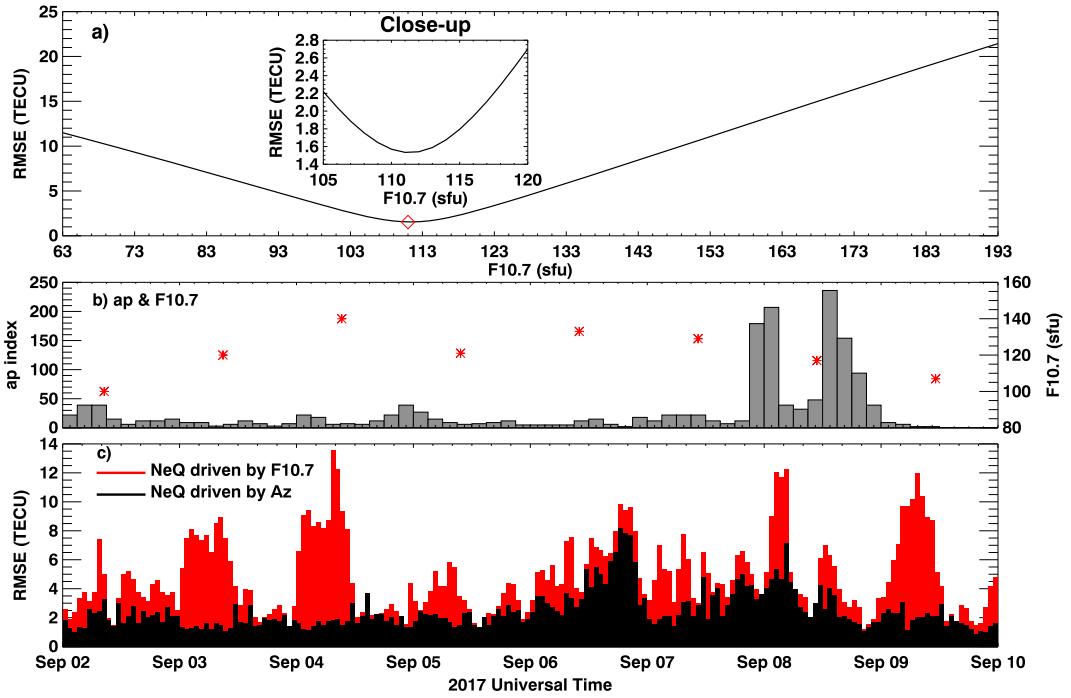

Figure 1. (a) A schematic diagram of RMSE variation with respect to $F_{10.7}$. The minimum error that corresponds to *Az* is marked with a diamond. (b) Temporal variation of Ap index and $F_{10.7}$ during 02–10 September 2017. (c) An example of RMSE comparison between NeQuick 2 driven by $F_{10.7}$ (red) and driven by *Az* (black) at BJFS station (39.4°N, 115.9°E) during this interval. RMSE = Root Mean Square Error; TECU = total electron content unit, 1 TECU = 10^{16}el/m^2.

Leitinger, 2001), which includes five semi-Epstein functions with modeled thick parameters to represent the lower and upper parts of the E and F1 layers, as well as the lower part of the F2 layer. The functions are anchored to the N_m (electron density) and the h_m (height) of the E, F1, and F2 layer peaks, which can be either experimentally derived from ionosonde measurement or modeled as indicated by Leitinger et al. (2005). The topside ionosphere is represented by a sixth semi-Epstein function with a height-dependent thickness parameter that can be empirically determined as described by Coïsson et al. (2006). The basic inputs of the NeQuick 2 model are position, time, and solar flux (or sunspot number); the outputs are the electron density along the ray-path and the numerically integrated TEC. For more details about NeQuick 2, readers may refer to Nava et al. (2008) and the references therein.

The TEC products derived from the Madrigal database are used here for data ingestion, which are developed at the Massachusetts Institute of Technology Haystack Observatory by using dense networks of worldwide GNSS receivers (Rideout & Coster, 2006; Vierinen et al., 2016). The gridded TEC cover locations where GNSS data are available and have a resolution of 1° (latitude) × 1° (longitude) × 5 min. Madrigal gridded TEC is strictly data driven with no postprocess interpolation or fitting that might smooth out real gradients, which can thus be considered as a suitable TEC source for data ingestion.

3. TEC Ingestion Technique and Validation

For a given time and location, the TEC value derived from the integration of NeQuick 2 electron density profile varies monotonically as a function of the 10.7 cm solar radio flux. For the technique presented here, the optimum solar flux that produces the best TEC value from NeQuick 2 is usually termed as *Az*, which is an effective parameter to represent local ionization level and is calculated by minimizing the Root Mean Square Error (RMSE) between the modeled and observational TEC (Nava et al., 2005, 2006):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(TEC_{mod}(Az) - TEC_{obs})^2}. \quad (1)$$

Here *N* is the number of individual observations during the current interval. Figure 1a illustrates the process in calculating the minimum RMSE to derive *Az*. As an example to quantitatively illustrate the effect of data

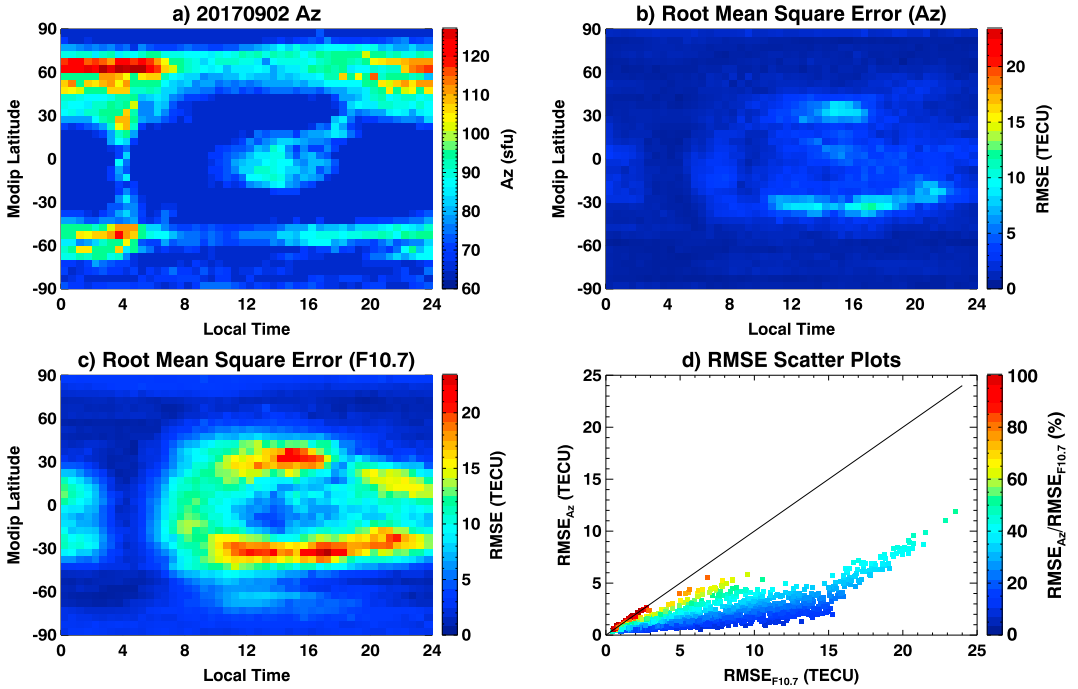

Figure 2. (a) An example of Modip-LT distribution of *Az* on 02 September 2017 after ingesting Madrigal total electron content into NeQuick 2 model. (b) Distribution of RMSE after data ingestion. (c) Distribution of RMSE before data ingestion. (d) Scatter plot of $RMSE_{Az}$ versus $RMSE_{F10.7}$, while the color represent the ratio of RMSE after and before data ingestion. Modip = modified dip; RMSE = Root Mean Square Error; TECU = total electron content unit, 1 TECU = 10^{16}el/m^2.

ingestion, Figure 1c shows the RMSE comparison between NeQuick 2 driven by $F_{10.7}$ (red) and driven by *Az* (black) at a GNSS station: BJFS (39.4°N, 115.9°E) during 02–10 September 2017. It can be seen that the RMSE calculated via using a modified *Az* to drive NeQuick 2 is generally 30–50% smaller that the model driven by the observed $F_{10.7}$. This partly demonstrates the systematic improvements when TEC data are ingested into the NeQuick 2 model.

In the traditional TEC ingestion method, the above-mentioned technique is either applied to all grids of vertical TEC GIMs (normally 2.5° in latitude and 5° in longitude) to get a global distribution of *Az* maps (e.g., Nava et al., 2005; Yu et al., 2015), or applied to slant TEC at a single or multiple GNSS stations to get a scattered *Az* distribution that might be further interpolated or fitted into regular size (e.g., Nava et al., 2006; Nigussie et al., 2016). Both of these techniques focus on obtaining fixed geographical distribution of *Az* to drive NeQuick to get 3-D specification of the ionospheric electron density. However, the extent of photoionization and ionospheric dynamics are also strongly dependent on the geomagnetic field and local time. Rawer (1963) proposed a parameter called the modified dip latitude μ (Modip), which combines the geomagnetic dip I and the geographic latitude ϕ,

$$Modip = arctan(\frac{I}{\sqrt{cos\phi}}). \quad (2)$$

Thus, in order to consider both the geomagnetic field and the spinning of Earth, the current study will use local time-Modip coordinate system to represent the variation of *Az* instead of using simple geographic or geomagnetic coordinates. The grid points are spaced 5° in Modip latitude by 0.5 hr in local time. For a certain day, the Madrigal TEC within each grid will be used to derive the *Az* on the basis of the above-mentioned data ingestion technique, then global LT-Modip maps of *Az* and associated RMSE distribution can be generated accordingly. Figure 2a illustrates an example of the *Az* distribution on 02 September 2017 ($F_{10.7}$ = 120). Figures 2b and 2c show the associated RMSE after and before TEC ingestion, respectively. It can be seen from Figure 2a that *Az* has relatively large values around the auroral zone and equatorial ionization anomaly (EIA) regions. This might be due to the fact that NeQuick 2 is only using solar radio flux as the effective driver, as well as that the NeQuick 2 model does not include the effect of ionization enhancement around 125 km due

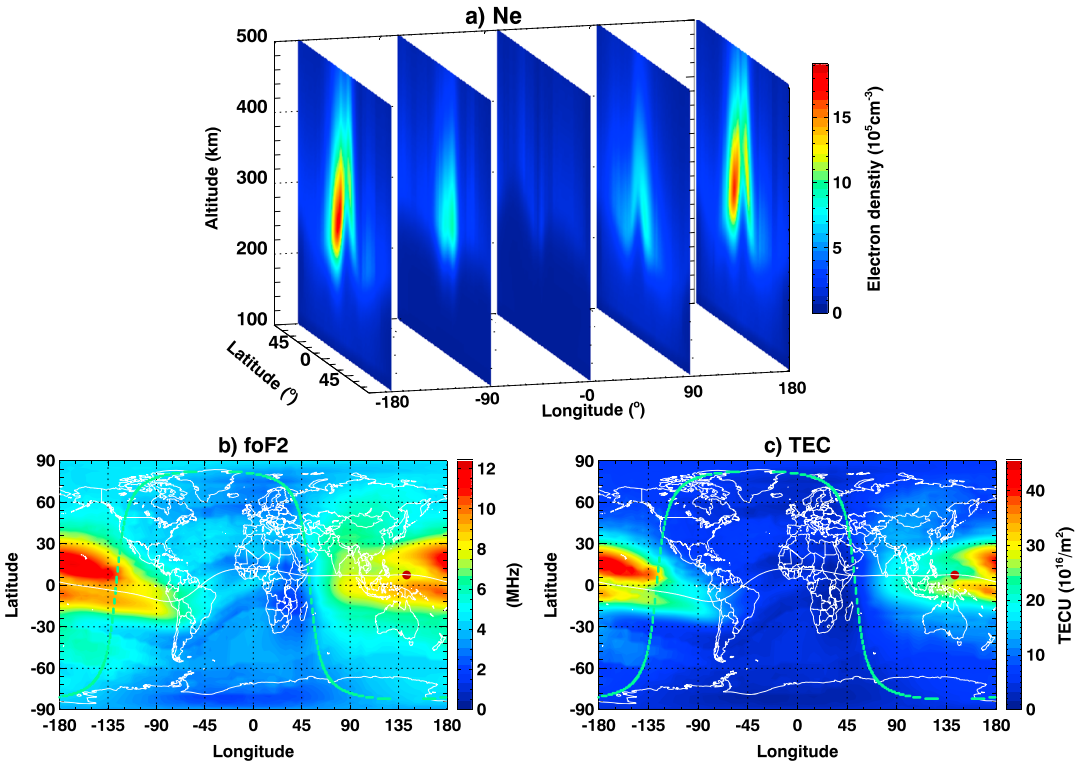

Figure 3. An example of data ingestion results at 0230 UT on 02 September 2017: (a) global 3-D electron density distribution; (b) foF2; (c) vertical TEC map. The terminator, subsolar point, and geomagnetic equator are also marked in the maps. TEC = total electron content; TECU = TEC unit, 1 TECU = 10^{16} electron per m^2; foF2 = F2 layer critical frequency.

to particle precipitation. So in this case, the model may be expected to underestimate the electron density around the auroral and EIA regions, which need to be compensated via the enhancement of *Az*.

Furthermore, the following characteristics on Modip latitude distribution can also be seen from the figure. First, for the equatorial and low latitude regions (~30°S–30°N), the *Az* values are generally lower than $F_{10.7}$. Similar to the EIA, there are double-peak structures near the EIA crest both for $RMSE_{Az}$ (~10 TECU, total electron content unit, 1 TECU = 1,016 electron per m^2) and $RMSE_{F10.7}$ (~20 TECU), while the RMSE values around the equator are much lower. Second, for midlatitude to high-latitude bands, there is a hemispheric asymmetry in the *Az* distribution especially around the auroral zones, with the northern hemisphere exhibiting larger *Az* values (close to real $F_{10.7}$) than those of the southern hemisphere. This might be ascribed to higher numbers of GNSS observations in the northern hemisphere both in the Madrigal database and used to construct the NeQuick 2 model. It could also be that the performance difference of the model was due to the effect of seasonal and hemispheric variation of precipitation in the auroral zone. Moreover, the RMSE have lower values for midlatitude to high-latitude bands, while the polar regions have lower *Az* and RMSE values.

Figure 2d shows a scatter plot of the comparison between $RMSE_{Az}$ and $RMSE_{F10.7}$, while the color represents the percentages ratio of $RMSE_{Az}$ to $RMSE_{F10.7}$. Through adjusting *Az* values, the data ingestion method generally reduced the errors around 50% since most of the points have green-to-blue colors. Another important thing worth noting is that the local time variation of *Az* is usually ignored in the past ingestion method, where *Az* at a fixed location will be updated by using 24 hr of data and then expressed as a sole function of Modip latitude. Thus, the diurnal variation of errors was smoothed out to a great extent in this way. However, it can be seen from Figure 2a that the optimized *Az* has an obvious local time variation pattern. For equatorial and low latitude regions, *Az* has relatively large values around the local noon sectors, which could be attributed to the EIA enhancement. For midlatitude to high-latitude regions, *Az* has maximum values around local night hours, which might indicate that NeQuick 2 tends to underestimate (overestimate) ionosphere electron density during nighttime (daytime) around these latitudinal regions, since NeQuick 2 did not include the effects of ionization enhancement around 125 km due to particle precipitation.

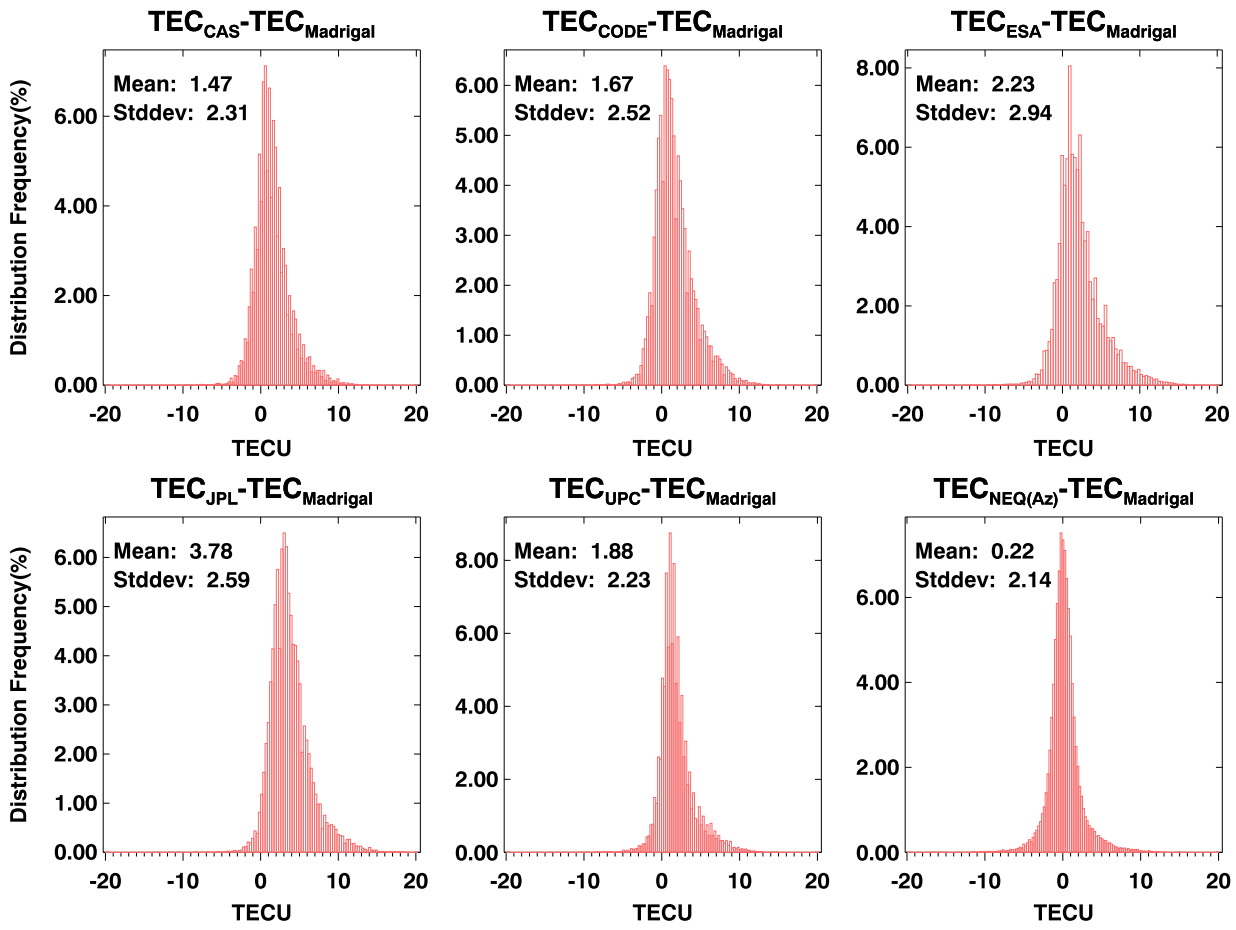

Figure 4. Histogram comparison of data ingestion results and International Global Navigation Satellite Systems Services global ionospheric maps with respect to Madrigal TEC on 02 September 2017. TEC = total electron content; CAS = Chinese Academy of Sciences; CODE = Center for Orbit Determination in Europe; ESA = European Space Agency; TECU = total electron content unit, 1 TECU = 1,016 electron per m^2; JPL = Jet Propulsion Laboratory; UPC = Polytechnical University of Catalonia.

With the availability of the optimized *Az* distribution map, the local ionization level for different locations and times can be derived correspondingly, and the Madrigal-driven NeQuick 2 results can then be generated by using derived *Az* as model inputs. Figure 3 gives an example of the reproduced global 3-D electron density in latitude/height slices, F2 layer critical frequency (foF2), and the vertical TEC map at 0230 UT on 02 September 2017. It can be seen that the large-scale features of ionosphere such as the EIA and hemispheric asymmetry could be reasonably reproduced. Since TEC measurements were used for data ingestion, it is expected that comparisons with TEC would be better than comparisons with electron density profiles.

3.1. Comparison with IGS GIMs

In order to verify the validity of the data ingestion technique, the TEC GIMs provided by International GNSS Services (IGS) are used here to make a comparison. Currently, there are five IGS ionospheric analysis centers routinely providing TEC GIMs by using ever-growing measurements from dense GNSS receivers. These centers include the Center for Orbit Determination in Europe (CODE), European Space Agency (ESA), JPL, Polytechnical University of Catalonia (UPC), and Chinese Academy of Sciences (CAS)/Wuhan University. The TEC GIMs of CODE and ESA are modeled by using a series of Spherical Harmonic functions up to degree and order of 15 (Feltens, 2007; Feltens & Schaer, 1998; Schaer, 1999). JPL adopted a grid-based modeling method to represent the TEC by using a linear composition of bicubic splines with 1,280 spherical triangles (Komjathy et al., 2005; Mannucci et al., 1998). The approaches used by UPC are similar to those of JPL, while UPC modeled the ionospheric TEC variation over each station separately by using a rectangular grid of two layers (Hernández-Pajares et al., 1999, 2009). CAS GIMs are generated by using a function-based plus grid-based approach that combine the Spherical Harmonic functions and the generalized Trigonometric Series functions (Li et al., 2015).

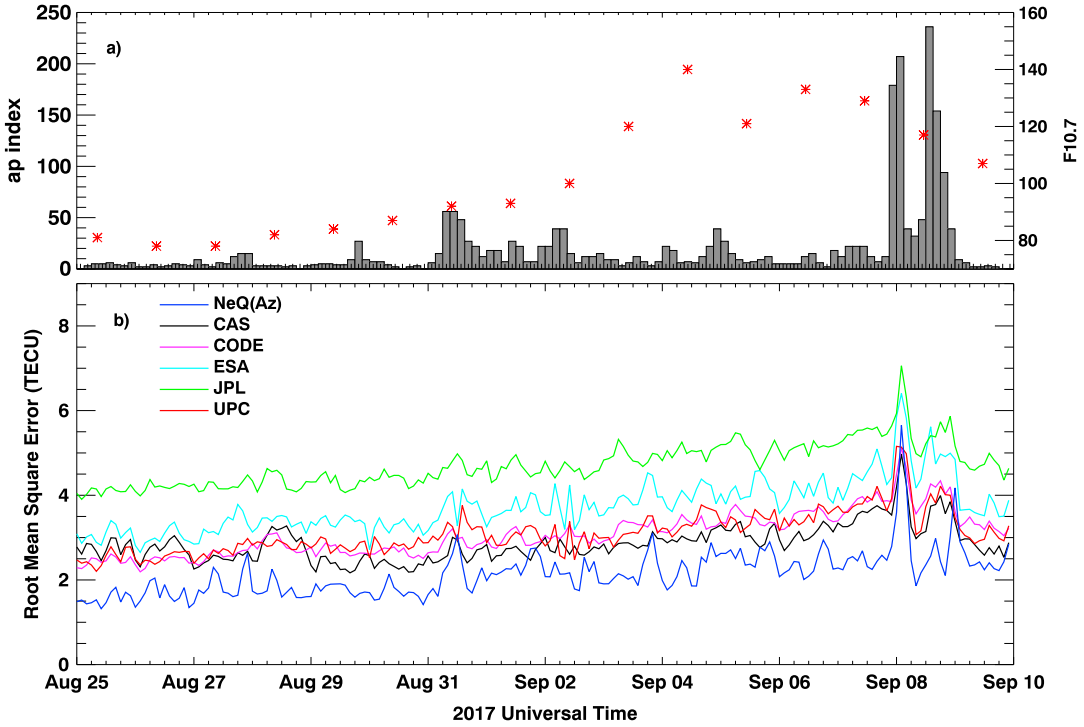

Figure 5. (a) Temporal variation of Ap index and $F_{10.7}$ from 25 August to 10 September 2017. (b) Root Mean Square Error variation for different TEC global ionospheric maps and NeQuick 2 driven by *Az*. TECU = total electron content unit, 1 TECU =10^{16}el/m^2 CAS = Chinese Academy of Sciences; CODE = Center for Orbit Determination in Europe; ESA = European Space Agency; JPL = Jet Propulsion Laboratory; UPC = Polytechnical University of Catalonia.

As a preliminary verification of the data ingestion method, around 10% of the Madrigal TEC data set was selected as a control group (i.e., not used for data ingestion). Then the IGS GIMs and data ingestion TEC results were compared with the control group, respectively. Figure 4 shows the histogram statistics of the comparison on 02 September 2017. The data ingestion errors ($TEC_{NEQ(Az)}$-$TEC_{Madrigal}$) exhibit a nearly unbiased Gaussian distribution with relatively low mean value (0.22 TECU) and standard deviation (2.14 TECU). The different IGS errors are generally more skewed and dispersed with larger standard deviation values.

In order to get a more comprehensive comparison under different solar and geomagnetic activities, Figure 5 displays the RMSE results of IGS GIMs and data ingestion TEC with respect to the Madrigal control group during the time interval from 25 August to 10 September 2017. The temporal variation of $F_{10.7}$ and 3-hr Ap index for this interval are shown in Figure 5a. The $F_{10.7}$ gradually increased from ~80 to 140, then decreased to 107. Meanwhile, the Ap index has two peaks of 207 and 236, which correspond to a double main phase of an intense geomagnetic storm on 07–08 September 2017. Thus, this time interval covers varied levels both for solar and geomagnetic activities, which is a suitable period to test the effectiveness of the data ingestion technique. Figure 5b shows that the RMSE of the data ingestion technique is generally smaller (around 1 TECU) than those of IGS GIMs, while all products have relatively large errors during the storm time. One thing worth noting is that the differences between data ingestion results and IGS GIMs could be caused by systematic errors that are generated by different processing algorithms of biases correction. The GIMs of JPL and UPC have larger mean bias: ~3 and ~2.5 TECU, respectively, while those of ESA (~2.2 TECU), CODE (~1.8 TECU), and CAS (~1.5 TECU) are smaller. For more details about the bias of different IGS GIMs, readers may refer to Li et al. (2015) and Hernández-Pajares et al. (2009). Also, the control group might not be strictly *independent* due to possible interference from surrounding measurements. Therefore, this initial comparison might indicate the effectiveness of data ingestion technique, though further comparisons are still needed.

3.2. Comparison with Ionosonde Data

Ionospheric foF2 measured by ground-based ionosonde can be considered as a suitable reference to verify the data ingestion results. Six ionosonde stations at different latitude/longitude locations were used to make the comparison. Figure 6 shows the comparisons of ionosonde foF2 measurements with those calculated via NeQuick 2 driven by $F_{10.7}$ (red) and *Az* (blue) during 25 August to 10 September 2017. The RMSE, correlation

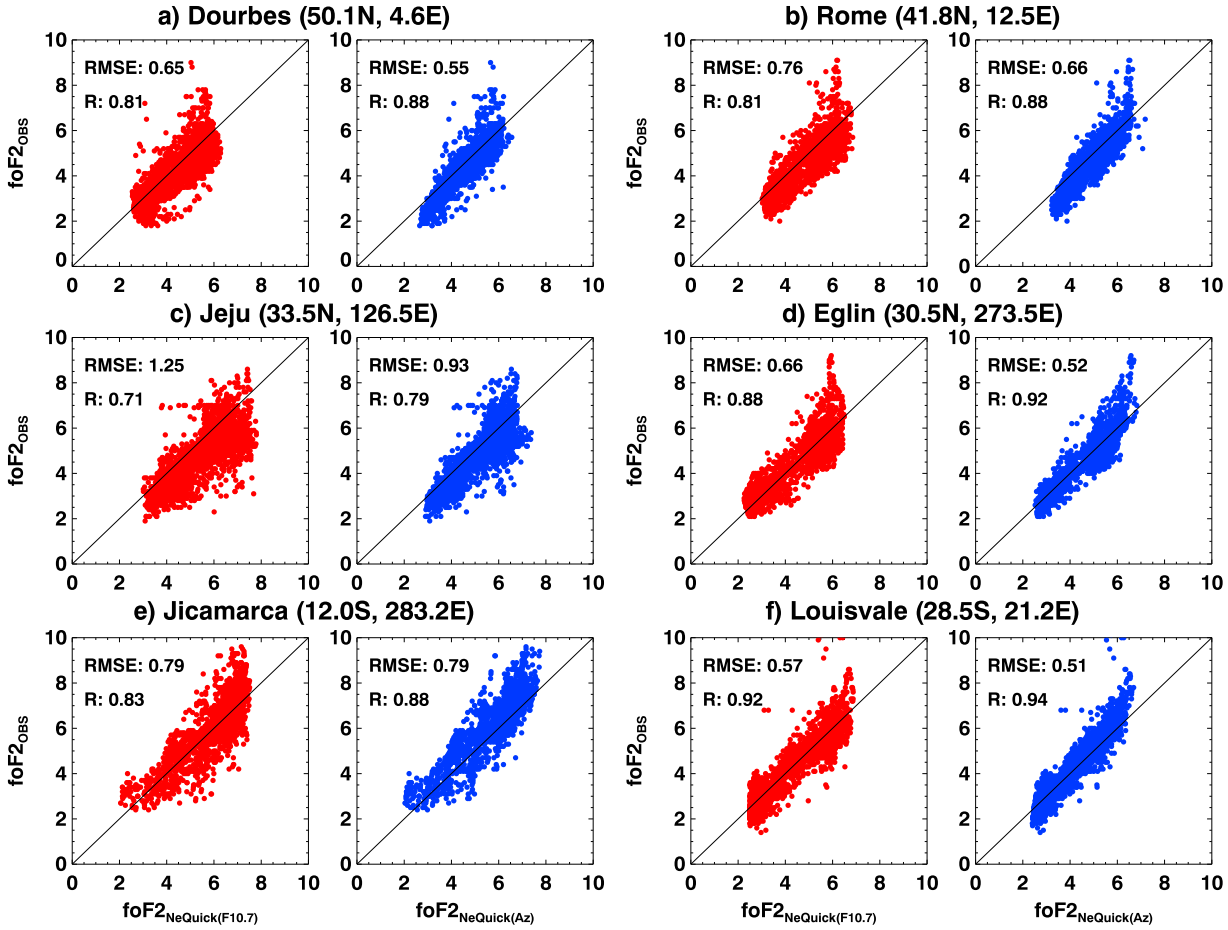

Figure 6. Scatter points comparison of ionosonde foF2 with NeQuick 2 driven by $F_{10.7}$ (red) and *Az* (blue) at six ionosonde stations. RMSE = Root Mean Square Error; foF2 = F2 layer critical frequency.

coefficient, as well as the geographic coordinates are marked in the figure. Generally, the data ingestion results had lower RMSE values and higher correlation coefficients, which illustrate that the ability of NeQuick 2 in reconstructing the foF2 is also improved after the Madrigal TEC are ingested into the model.

4. EOF Modeling of Az

The data ingestion is basically a now-casting method of measurement update. Considering the forecasting needs of ionospheric correction for navigation and communication customers, it is of great importance to construct a time-dependent model of *Az* so that the spatiotemporal variability after previous TEC ingestion could be extracted and parameterized to make a prediction for future use. In this study, the EOF analysis technique, also known as Principal Component Analysis method, was used to build this time-dependent *Az*

Table 1
Variances of the First Six EOF Series

EOF series	Variances (%)	Cumulative variances (%)
EOF 1	76.11	76.11
EOF 2	6.18	82.29
EOF 3	3.75	86.04
EOF 4	2.74	88.78
EOF 5	1.70	90.48
EOF 6	1.29	91.77

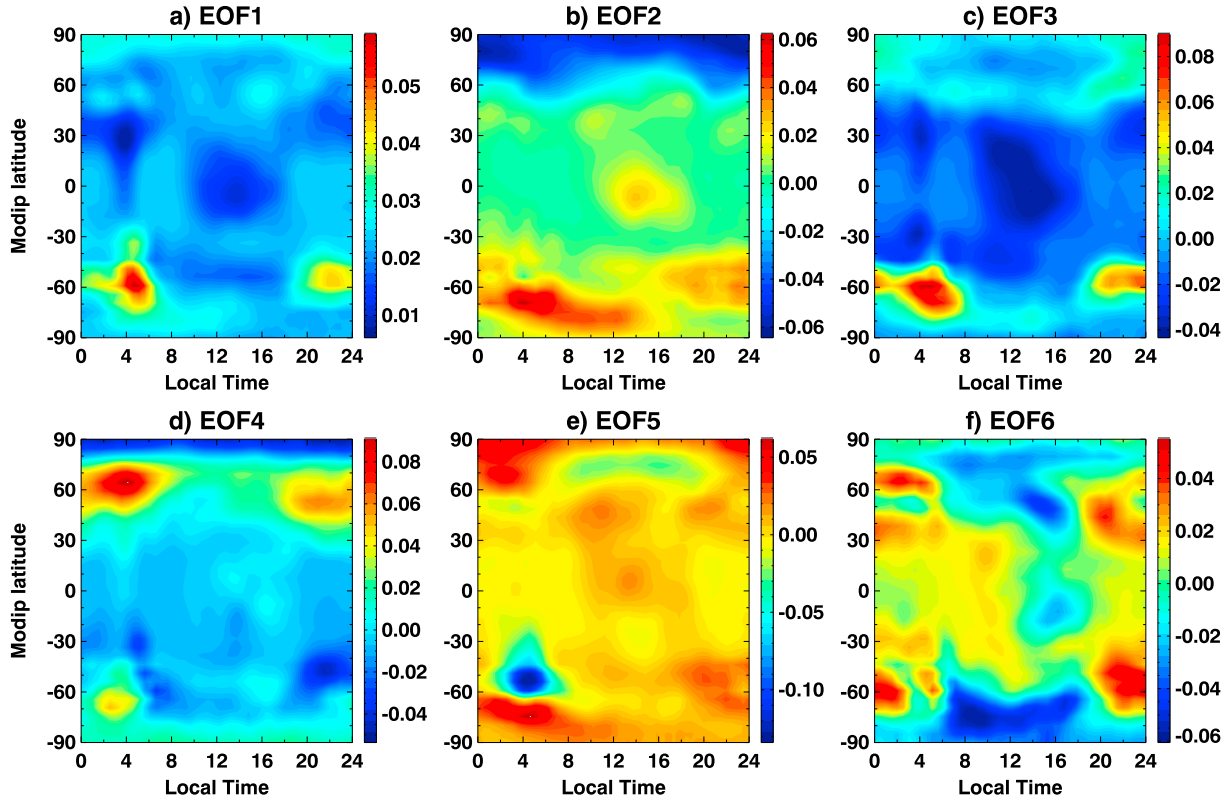

Figure 7. The Modip latitude and local time distribution of the first six EOF base functions through decomposition of $Az/F_{10.7}$. EOF = empirical orthogonal function; Modip = modified dip.

model. The EOF technique is a statistical procedure, capturing the most significant components of the variability in the original data set, which is implemented by using an orthogonal transformation to decompose the original data set into a series of uncorrelated base functions, with each succeeding base function accounting for as much residual variance as possible (Jolliffe, 1990). The merit of the EOF technique is that it converges quickly, which makes it possible to succinctly represent the majority of the original variances by using only a few base functions and associated coefficients. For more details about the mathematical explanation of the EOF analysis method, readers may refer to Dvinskikh (1988) and Singer and Dvinskikh (1991) and the references therein.

In order to get a balance between capturing the variances as much as possible and making ionospheric *weather* prediction, the *Az* data set needed an appropriate time length. Using multiple years of *Az* data to do EOF decomposition would generate the maximum variances, yet making this model a *climatological* one. On the other hand, if the time length of *Az* data was too short, then it would be unlikely to extract effective base functions with enough variances to make the forecast. Thus, in the current study, after above-mentioned Madrigal TEC ingestion, a moving data set of *Az* ratio (i.e., $Az/F_{10.7}$) with a time length of 81 days was reorganized into the following matrix Az_{ratio}(LT, Modip, d), in which LT, Modip, and *d* stand for the local time (48 grids), modified dip latitude (36 grids), and day of year (81 days), respectively. This Az_{ratio} data set could then be decomposed into EOF base functions and associated coefficients:

$$Az_{ratio}(LT, Modip, d) = \sum_{i=1}^{N} EOF_i(LT, Modip) \times Coef_i(d), \tag{3}$$

where EOF_i(LT,Modip) is the i_{th} EOF base functions that vary with local time and modified dip latitude, which represent the diurnal fluctuation and spatial distribution of original data set. $Coef_i(d)$ is the associated i_{th} coefficient that indicates the temporal variation of original data set. *N* is the total number of EOF decomposition series. The order of the EOF series is ranked according to their variance, and the variances contributed by the first six EOF series are listed in Table 1. Since 91.77% of the total variance in the original data set can be

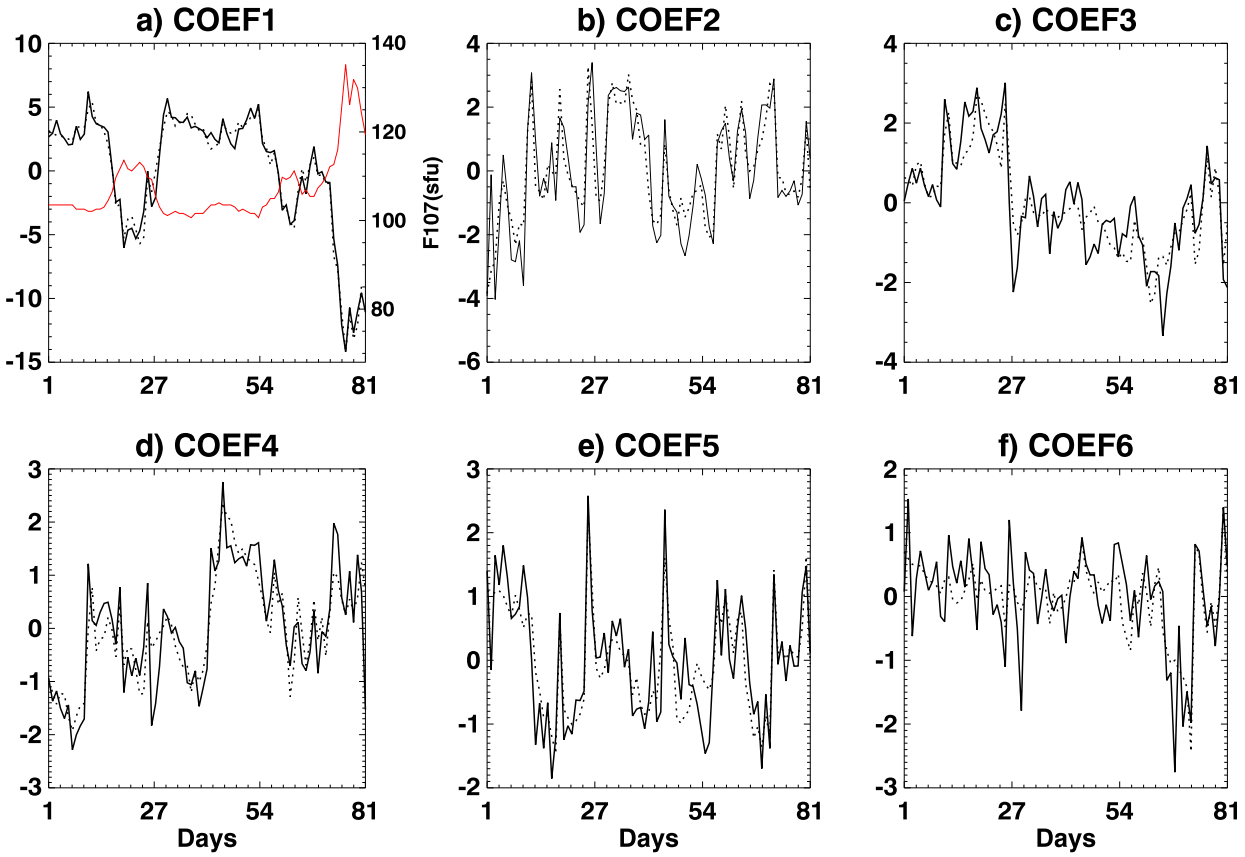

Figure 8. Temporal variation of the first six empirical orthogonal function coefficients through decomposition of $Az/F_{10.7}$. The observational $F_{10.7}$ (red) and modeled coefficients (dashed) are also marked.

reproduced via the first six EOF series, so it is an effective and efficient modeling method to use only six EOF base functions and coefficients to represent most of the variation in the original data set.

Figure 7 shows the Modip latitude and local time distribution of the first six EOF base functions. Take EOF 1 to EOF 3 as example, EOF 1 appears to represent the most dominant feature of global spatial and temporal variation of the original data set, which is day-to-night variability as well as high-latitude-to-low-latitude difference due to solar ionization. EOF 2 mainly displays hemispheric asymmetries, which can be attributed to the summer-to-winter annual variation induced by the uneven solar EUV illumination. EOF 3 captures mostly auroral and high-latitude variations, which can be ascribed to Joule heating and auroral precipitation under the influence of geomagnetic activity. EOF 4–EOF 6 have similar distribution features but with more small-scale variations. The physical meaning of these EOF components are not always apparent, particularly for high order ones whose contributions to the overall variance in the data are often very small. Figure 8 shows the temporal variation of the first six EOF coefficients. The corresponding $F_{10.7}$ is also plotted in Figure 8a for comparison. The $Coef_1$ and $F_{10.7}$ are roughly anticorrelated, which indicate that there are certain variations in the original data set that are dependent on solar activity. In order to further investigate the intrinsic dependence of Az_{ratio} on solar and geomagnetic activity, Table 2 gives the correlation values of the first six EOF coefficients with respect to $F_{10.7}$ and daily Ap index. The correlation between $Coef_1$ and the $F_{10.7}$ index is −0.96, while all coefficients more or less have some correlation with Ap index. This shows that the local ionization parameter Az changes mainly as a function of solar activity, while geomagnetic activity also plays a nonnegligible role in affecting it.

Therefore, the first six EOF coefficients could be parameterized and modeled as follows:

$$Coef_i(d) = F^i_{SG} \times F^i_{time}, \tag{4}$$

Table 2
Correlation of the First Six Empirical Orthogonal Function Coefficients With Respect to $F_{10.7}$ and Ap Index

Correlation	$F_{10.7}$	Ap
Coef 1	−0.96	−0.31
Coef 2	0.01	0.36
Coef 3	0.11	−0.08
Coef 4	0.10	0.27
Coef 5	0.08	0.49
Coef 6	−0.06	0.16

where F^i_{SG} represents the effects of solar and geomagnetic activity and F^i_{time} refers to annual/seasonal variations. These two parameters can be expressed as follows:

$$F^i_{SG} = a_i + b_i F_{10.7}(d) + c_i Ap(d), \tag{5}$$

$$F^i_{time} = d_i + e_i cos(\frac{2\pi d}{365.25}) + f_i sin(\frac{2\pi d}{365.25}) + g_i cos(\frac{2\pi d}{81}) + h_i sin(\frac{2\pi d}{81}), \tag{6}$$

where a–h are amplitudes of various terms in the above equations and can be calculated via a multiple linear regression analysis method. Thus, the EOF coefficients can be expressed as a parameterized function of $F_{10.7}$ and Ap index, which can be reconstructed with observed or predicted $F_{10.7}$ and Ap index. Figure 8 also shows an example of reconstructed EOF coefficients as dashed lines, which agrees well with the original solid lines. In this way, for the time-window of 81 days, the naturally decomposed EOF base functions and artificially fitted coefficients can be combined with each other to generate modeled *Az* values, which can then be used to drive the NeQuick 2 model to make a short-term prediction of *Ne* and TEC for the next day (or even longer). This whole procedure of data ingestion and EOF modeling can be rolled over with a moving time-window.

In order to verify the effectiveness of the EOF modeling technique, the *Az* modeling method of the Galileo navigation system is introduced here to make a comparison. In the Galileo ionospheric correction algorithm, the *Az* values of the previous day are expressed as a 2nd order polynomial function of

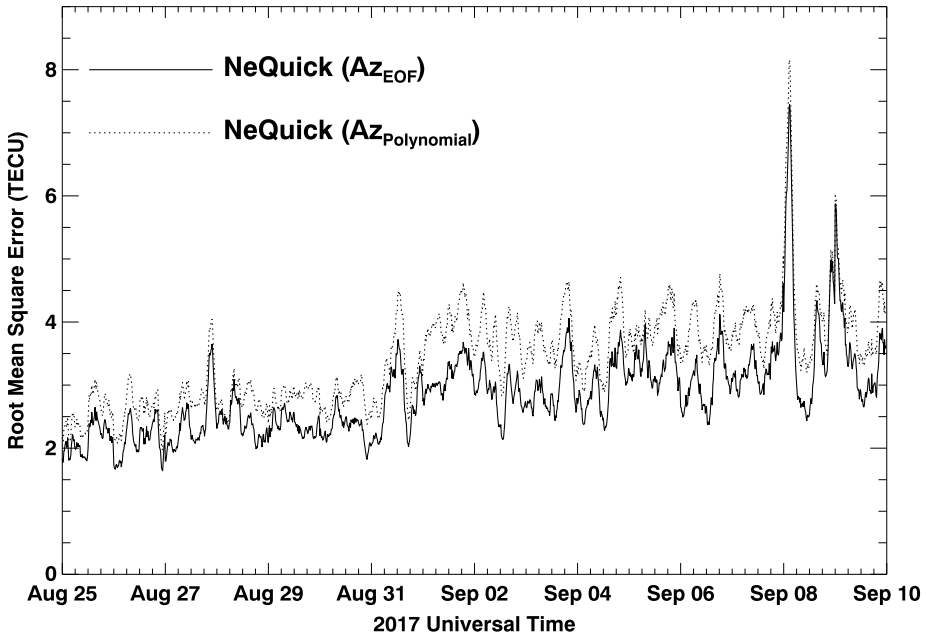

Figure 9. Root Mean Square Error variation of NeQuick total electron content driven by EOF-modeled *Az* (solid) and polynomial fitted *Az* (dashed) during the time period from 25 August to 10 September 2017. EOF = empirical orthogonal function; TECU = total electron content unit, 1 TECU = 10^{16}el/m^2.

Modip: $Az = a_0 + a_1 \times Modip + a_2 \times Modip^2$, then the set of three coefficients are calculated and broadcasted in the navigation file so that the ionospheric delay at a specific frequency for the next day can be corrected using NeQuick driven by reconstructed *Az* parameters (Bidaine & Warnant, 2011). Figure 9 shows the temporal variation of RMSE comparison between NeQuick TEC driven by EOF-modeled *Az* (solid line) and that driven by polynomial fitted *Az* (dashed line) during the time period of 25 August to 10 September 2017. The NeQuick RMSE driven by the EOF-modeled *Az* are generally smaller than those of the polynomial fitted *Az* with an average improvement of ∼10–15%. Both of which have larger errors around the geomagnetic storm time, which is consistent with those indicated in Figure 5. Moreover, considering the polynomial method needs fewer parameters transmitted to the users, it is still more functional in real application.

5. Conclusion

In this paper, a data ingestion technique is described to incorporate the Madrigal TEC data into the NeQuick 2 model. The global LT-Modip distribution map of the effective ionization parameter (*Az*) was estimated accordingly through this ingestion procedure, then the NeQuick 2 model could be driven by an *Az* map to reproduce ionospheric parameters, such as *Ne*, TEC, NmF2, hmF2, and so forth. In general, the performance of the Madrigal TEC-driven-NeQuick 2 can reduce the errors around 30–50% compared with those before data ingestion and it can capture more subtle ionospheric features. The accuracy of the ingestion results are further validated through comparison with various IGS GIMs, and the statistical analysis demonstrates that the data ingestion results have slightly lower RMSE (∼1 TECU) and bias than those of IGS GIMs. A further comparison with ionosonde data shows that the ability of NeQuick 2 to reproduce the foF2 is also improved after data ingestion. Moreover, the EOF technique is used to construct a time-dependent model of *Az*. The intrinsic diurnal variation and spatial distribution of the original data set can be well represented by EOF base functions, and 90% of the total variances can be well captured by using the first six EOF series. The associated EOF coefficients can be expressed as a combination of (1) linear functions of $F_{10.7}$ and Ap index to show the dependence on solar/geomagnetic activity and (2) a series of trigonometric functions with different periods to represent annual/seasonal variation components. In comparison with the Galileo ionospheric correction algorithm, the accuracy of TEC prediction by using the EOF-modeled *Az* is improved to some extent (∼10–15%) though both results have large deviations for a short period during the storm recovery phase. These preliminary results indicate the effectiveness of this data ingestion and EOF modeling technique in bringing certain systematic improvement of ionosphere now-cast/forecast, while further modification could still be needed in the future to make this product more robust for both scientific study and space weather applications.

Acknowledgments
This work is sponsored by the National Key R&D Program of China (2016YFB0501503), National Science Foundation of China (41404125, 41674183), AFOSR under DDDAS (Dynamic Data-Driven applications Systems http://www.1dddas.org/) grant FA9550-16-1-0071, and Youth Innovation Promotion Association of Chinese Academy of Sciences. The authors acknowledge S. Radicella and B. Nava from Abdus Salam International Centre for Theoretical Physics (ICTP) in providing NeQuick 2 code. The ionosonde data are downloaded from NOAA Space Weather Prediction Center at ftp.swpc.noaa.gov/pub/. The solar/geomagnetic activity indices data are obtained from the NASA Goddard Space Flight Center Coordinated Data Analysis Web. The IGS GIMs data are acquired from NASA's archive of space geodesy data: Crustal Dynamics Data Information System (CDDIS). GPS TEC data products and access through the Madrigal distributed data system are provided to the community (http://www.openmadrigal.org) by the Massachusetts Institute of Technology (MIT) under support from US National Science Foundation grant AGS-1242204. AJC, and SRZ acknowledge the ONR grant N00014-17-1-2186, and SRZ and AJC acknowledge the AFOSR MURI grant FA9559-16-1-0364. Data for the TEC processing is provided from the following organizations: UNAVCO; Scripps Orbit and Permanent Array Center; Institut Geographique National, France; International GNSS Service; The Crustal Dynamics Data Information System (CDDIS); National Geodetic Survey; Instituto Brasileiro de Geografia e Estatística; RAMSAC CORS of Instituto Geográfico Nacional de la República Argentina; Arecibo Observatory; Low-Latitude Ionospheric Sensor Network (LISN); Topcon Positioning Systems, Inc.; Canadian High Arctic Ionospheric Network; Institute of Geology and Geophysics; Chinese Academy of Sciences; China Meteorology Administration; Centro di Ricerche Sismologiche; Système d'Observation du Niveau des Eaux Littorales (SONEL); RENAG: REseau NAtional GPS permanent; GeoNet—the official source of geological hazard information for New Zealand; GNSS Reference Networks; Finnish Meteorological Institute; and SWEPOS—Sweden.

References

Aa, E., Huang, W., Yu, S., Liu, S., Shi, L., Gong, J., et al. (2015). A regional ionospheric TEC mapping technique over China and adjacent areas on the basis of data assimilation. *Journal of Geophysical Research: Space Physics*, *120*, 5049–5061. https://doi.org/10.1002/2015JA021140

Aa, E., Liu, S., Huang, W., Shi, L., Gong, J., Chen, Y., et al. (2016). Regional 3-D ionospheric electron density specification on the basis of data assimilation of ground-based GNSS and radio occultation data. *Space Weather*, *14*, 433–448. https://doi.org/10.1002/2016SW001363

Angling, M. J., & Cannon, P. S. (2004). Assimilation of radio occultation measurements into background ionospheric models. *Radio Science*, *39*, RS1S08. https://doi.org/10.1029/2002RS002819

Angling, M. J., & Khattatov, B. (2006). Comparative study of two assimilative models of the ionosphere. *Radio Science*, *41*, RS5S20. https://doi.org/10.1029/2005RS003372

Bidaine, B., & Warnant, R. (2011). Ionosphere modelling for Galileo single frequency users: Illustration of the combination of the NeQuick model and GNSS data ingestion. *Advances in Space Research*, *47*, 312–322. https://doi.org/10.1016/j.asr.2010.09.001

Bilitza, D. (2001). International reference ionosphere 2000. *Radio Science*, *36*(2), 261–275. https://doi.org/10.1029/2000RS002432

Bilitza, D., & Reinisch, B. W. (2008). International reference ionosphere 2007: Improvements and new parameters. *Advances in Space Research*, *42*, 599–609. https://doi.org/10.1016/j.asr.2007.07.048

Brunini, C., Azpilicueta, F., Gende, M., Camilion, E., Ángel, A. A., Hernandez-Pajares, M., et al. (2011). Ground- and space-based GPS data ingestion into the NeQuick model. *Journal of Geodesy*, *85*, 931–939. https://doi.org/10.1007/s00190-011-0452-4

Bust, G. S., Crowley, G., Garner, T. W., Gaussiran, T. L., Meggs, R. W., Mitchell, C. N., et al. (2007). Four-dimensional GPS imaging of space weather storms. *Space Weather*, *5*, 02003. https://doi.org/10.1029/2006SW000237

Bust, G. S., Garner, T. W., & Gaussiran, T. L. (2004). Ionospheric Data Assimilation Three-Dimensional (IDA3D): A global, multisensor, electron density specification algorithm. *Journal of Geophysical Research*, *109*, A11312. https://doi.org/10.1029/2003JA010234

Coïsson, P., Radicella, S. M., Leitinger, R., & Nava, B. (2006). Topside electron density in IRI and NeQuick: Features and limitations. *Advances in Space Research*, *37*, 937–942. https://doi.org/10.1016/j.asr.2005.09.015

Dvinskikh, N. I. (1988). Expansion of ionospheric characteristics fields in empirical orthogonal functions. *Advances in Space Research*, *8*, 179–187. https://doi.org/10.1016/0273-1177(88)90238-4

Feltens, J. (2007). Development of a new three-dimensional mathematical ionosphere model at European Space Agency/European Space Operations Centre. *Space Weather*, *5*, S12002. https://doi.org/10.1029/2006SW000294

Feltens, J., & Schaer, S. (1998). *IGS products for the ionosphere, IGS position paper the IGS analysis centers workshop*. Germany: Darmstadt.
Fuller-Rowell, T, Araujo-Pradere, E., Minter, C., Codrescu, M., Spencer, P., Robertson, D., et al. (2006). US-TEC: A new data assimilation product from the Space Environment Center characterizing the ionospheric total electron content using real-time GPS data. *Radio Science*, *41*, RS6003. https://doi.org/10.1029/2005RS003393
Galkin, I. A., Reinisch, B. W., Huang, X., & Bilitza, D. (2012). Assimilation of GIRO data into a real-time IRI. *Radio Science*, *47*, RS0L07. https://doi.org/10.1029/2011RS004952
Giovanni, D. G., & Radicella, S. M. (1990). An analytical model of the electron density profile in the ionosphere. *Advances in Space Research*, *10*, 27–30. https://doi.org/10.1016/0273-1177(90)90301-F
Hernández-Pajares, M., Juan, J. M., & Sanz, J. (1999). New approaches in global ionospheric determination using ground GPS data. *Journal of Atmospheric and Solar-Terrestrial Physics*, *61*, 1237–1247. https://doi.org/10.1016/S1364-6826(99)00054-1
Hernández-Pajares, M., Juan, J. M., Sanz, J., Orus, R., Garcia-Rigo, A., Feltens, J., et al. (2009). The IGS VTEC maps: A reliable source of ionospheric information since 1998. *Journal of Geodesy*, *83*, 263–275. https://doi.org/10.1007/s00190-008-0266-1
Jolliffe, I. T. (1990). Principal component analysis: A beginner's guide - I. Introduction and application. *Weather*, *45*, 375–382. https://doi.org/10.1002/j.1477-8696.1990.tb05558.x
Komjathy, A., Sparks, L., Wilson, B. D., & Mannucci, A. J. (2005). Automated daily processing of more than 1000 ground-based GPS receivers for studying intense ionospheric storms. *Radio Science*, *40*, RS6006. https://doi.org/10.1029/2005RS003279
Komjathy, A., Wilson, B., Pi, X., Akopian, V., Dumett, M., Iijima, B., et al. (2010). JPL/USC GAIM: On the impact of using COSMIC and ground-based GPS measurements to estimate ionospheric parameters. *Journal of Geophysical Research*, *115*, A02307. https://doi.org/10.1029/2009JA014420
Lee, I. T., Matsuo, T., Richmond, A. D., Liu, J. Y., Wang, W., Lin, C. H., et al. (2012). Assimilation of FORMOSAT-3/COSMIC electron density profiles into a coupled thermosphere/ionosphere model using ensemble Kalman filtering. *Journal of Geophysical Research*, *117*, A10318. https://doi.org/10.1029/2012JA017700
Leitinger, R., Zhang, M., & Radicella, S. M. (2005). An improved bottomside for the ionospheric electron density model NeQuick. *Annales de Geophysique*, *48*(3). https://doi.org/10.4401/ag-3217
Li, Z., Yuan, Y., Wang, N., Hernandez-Pajares, M., & Huo, X. (2015). SHPTS: Towards a new method for generating precise global ionospheric TEC map based on spherical harmonic and generalized trigonometric series functions. *Journal of Geodesy*, *89*, 331–345. https://doi.org/10.1007/s00190-014-0778-9
Mandrake, L., Wilson, B., Wang, C., Hajj, G., Mannucci, A., & Pi, X. (2005). A performance evaluation of the operational Jet Propulsion Laboratory/University of Southern California global assimilation ionospheric model (JPL/USC GAIM). *Journal of Geophysical Research*, *110*, A12306. https://doi.org/10.1029/2005JA011170
Mannucci, A. J., Wilson, B. D., Yuan, D. N., Ho, C. H., Lindqwister, U. J., & Runge, T. F. (1998). A global mapping technique for GPS-derived ionospheric total electron content measurements. *Radio Science*, *33*, 565–582. https://doi.org/10.1029/97RS02707
Matsuo, T., Fedrizzi, M., Fuller-Rowell, T. J., & Codrescu, M. V. (2012). Data assimilation of thermospheric mass density. *Space Weather*, *10*, 05002. https://doi.org/10.1029/2012SW000773
Nava, B., Coïsson, P., Miró Amarante, G., Azpilicueta, F., & Radicella, S. M. (2005). A model assisted ionospheric electron density reconstruction method based on vertical TEC data ingestion. *Annales de Geophysique*, *48*(2). https://doi.org/10.4401/ag-3203
Nava, B., Coïsson, P., & Radicella, S. M. (2008). A new version of the NeQuick ionosphere electron density model. *Journal of Atmospheric and Solar-Terrestrial Physics*, *70*, 1856–1862. https://doi.org/10.1016/j.jastp.2008.01.015
Nava, B., Radicella, S. M., & Azpilicueta, F. (2011). Data ingestion into NeQuick 2. *Radio Science*, *46*, RS0D17. https://doi.org/10.1029/2010RS004635
Nava, B., Radicella, S. M., Leitinger, R., & Coïsson, P. (2006). A near-real-time model-assisted ionosphere electron density retrieval method. *Radio Science*, *41*, RS6S16. https://doi.org/10.1029/2005RS003386
Nigussie, M., Radicella, S. M., Damtie, B., Yizengaw, E., Nava, B., & Roininen, L. (2016). Validation of NeQuick TEC data ingestion technique against C/NOFS and EISCAT electron density measurements. *Radio Science*, *51*, 905–917. https://doi.org/10.1002/2015RS005930
Pi, X., Mannucci, A. J., Iijima, B. A., Wilson, B. D., Komjathy, A., Runge, T. F., et al. (2009). Assimilative modeling of ionospheric disturbances with FORMOSAT-3/COSMIC and ground-based GPS measurements. *Terrestrial, Atmospheric and Oceanic Sciences*, *20*, 273–285.
Pi, X., Wang, C., Hajj, G. A., Rosen, G., Wilson, B. D., & Bailey, G. J. (2003). Estimation of E× B drift using a global assimilative ionospheric model: An observation system simulation experiment. *Journal of Geophysical Research*, *108*, 1075. https://doi.org/10.1029/2001JA009235
Radicella, S. M., & Leitinger, R. (2001). The evolution of the DGR approach to model electron density profiles. *Advances in Space Research*, *27*, 35–40. https://doi.org/10.1016/S0273-1177(00)00138-1
Rawer, K. (1963). Propagation of decameter waves (HF band). In *Meteorological and Astronomical Influences on Radio Wave Propogation B Landmark* (pp. 221–250).
Rideout, W., & Coster, A. (2006). Automated GPS processing for global total electron content data. *GPS Solutions*, *10*(3), 219–228. https://doi.org/10.1007/s10291-006-0029-5
Schaer, S. (1999). Mapping and Predicting the Earth's Ionosphere Using the Global Positioning System, (Phd dissertation). Astronomical Institute, University of Berne, Berne, Switzerland, 25 March.
Scherliess, L., Schunk, R. W., Sojka, J. J., & Thompson, D. C. (2004). Development of a physics-based reduced state Kalman filter for the ionosphere. *Radio Science*, *39*, RS1S04. https://doi.org/10.1029/2002RS002797
Scherliess, L., Schunk, R. W., Sojka, J. J., Thompson, D. C., & Zhu, L. (2006). Utah State University global assimilation of ionospheric measurements Gauss-Markov Kalman filter model of the ionosphere: Model description and validation. *Journal of Geophysical Research*, *111*, A11315. https://doi.org/10.1029/2006JA011712
Schunk, R. W., Scherliess, L., Eccles, V., Gardner, L. C., Sojka, J. J., Zhu, L., et al. (2014). Ensemble modeling with data assimilation models: A new strategy for space weather specifications, forecasts, and science. *Space Weather*, *12*, 123–126. https://doi.org/10.1002/2014SW001050
Schunk, R. W., Scherliess, L., Sojka, J. J., Thompson, D. C., Anderson, D. N., Codrescu, M., et al. (2004). Global assimilation of ionospheric measurements (GAIM). *Radio Science*, *39*, RS1S02. https://doi.org/10.1029/2002RS002794
Schunk, R. W., Scherliess, L., Sojka, J. J., Thompson, D., & Zhu, L. (2005). Ionospheric weather forecasting on the horizon. *Space Weather*, *3*, S08007. https://doi.org/10.1029/2004SW000138
Singer, W., & Dvinskikh, N. I. (1991). Comparison of empirical models of ionospheric characteristics developed by means of different mapping methods. *Advances in Space Research*, *11*, 3–6. https://doi.org/10.1016/0273-1177(91)90311-7
Vierinen, J., Coster, A. J., Rideout, W. C., Erickson, P. J., & Norberg, J. (2016). Statistical framework for estimating GNSS bias. *Atmospheric Measurement Techniques*, *9*, 1303–1312. https://doi.org/10.5194/amt-9-1303-2016

Wang, C., Hajj, G., Pi, X., Rosen, I. G., & Wilson, B. (2004). Development of the global assimilative ionospheric model. *Radio Science*, *39*, RS1S06. https://doi.org/10.1029/2002RS002854

Yu, X., Zhen, W., Xiong, B., She, C., Ou, M., Xu, J., et al. (2015). The performance of ionospheric correction based on NeQuick 2 model adaptation to Global Ionospheric Maps. *Advances in Space Research*, *55*, 1741 – 1747. https://doi.org/10.1016/j.asr.2015.01.011

Yue, X., Schreiner, W. S., Kuo, Y.-H., Hunt, D. C., Wang, W., Solomon, S. C., et al. (2012). Global 3-D ionospheric electron density reanalysis based on multisource data assimilation. *Journal of Geophysical Research*, *117*, A09325. https://doi.org/10.1029/2012JA017968

Yue, X., Schreiner, W. S., Lin, Y.-C., Rocken, C., Kuo, Y.-H., & Zhao, B. (2011). Data assimilation retrieval of electron density profiles from radio occultation measurements. *Journal of Geophysical Research*, *116*, A03317. https://doi.org/10.1029/2010JA015980

Zhu, L., Schunk, R., Scherliess, L., & Eccles, V. (2012). Importance of data assimilation technique in defining the model drivers for the space weather specification of the high-latitude ionosphere. *Radio Science*, *47*, RS0L24. https://doi.org/10.1029/2011RS004936