Shim Ja Soon (Orcid ID: 0000-0001-8943-3192)
Tsagouri Ioanna (Orcid ID: 0000-0002-3810-7028)
Goncharenko Larisa, P. (Orcid ID: 0000-0001-9031-7439)
Rastaetter Lutz (Orcid ID: 0000-0002-7343-4147)
Bilitza Dieter (Orcid ID: 0000-0001-6551-2929)
Codrescu Mihail, V. (Orcid ID: 0000-0001-7216-9858)
Coster Anthea, J. (Orcid ID: 0000-0001-8980-6550)
Solomon Stanley, C. (Orcid ID: 0000-0002-5291-3034)
Fedrizzi Mariangel (Orcid ID: 0000-0001-8420-8623)
Förster Matthias (Orcid ID: 0000-0002-1104-2471)
Gardner Larry, C. (Orcid ID: 0000-0003-1127-3218)
Huba Joseph (Orcid ID: 0000-0002-1930-0076)
Prokhorov Boris, E. (Orcid ID: 0000-0003-1751-5361)
Scherliess Ludger (Orcid ID: 0000-0002-7388-5255)
Schunk Robert, W. (Orcid ID: 0000-0001-8137-0617)
Sojka Jan, J (Orcid ID: 0000-0001-7894-6785)

Validation of Ionospheric Specifications During Geomagnetic Storms: TEC and foF2 during the 2013 March Storm Event

J. S. Shim[1*], I. Tsagouri[2], L. Goncharenko[3], L. Rastaetter[4], M. Kuznetsova[4], D. Bilitza[5], M. Codrescu[6], A. J. Coster[3], S. C. Solomon[7], M. Fedrizzi[6], M. Förster[8], T. J. Fuller-Rowell[6], L. C. Gardner[9], J. Huba[10], A. A. Namgaladze[11], B. E. Prokhorov[8], A. J. Ridley[12], L. Scherliess[9], R. W. Schunk[9], J. J. Sojka[9], L. Zhu[9]

1. The Catholic University of America, NASA GSFC, Greenbelt, MD, USA,

2. National Observatory of Athens, Penteli, Greece

3. Haystack Observatory, Westford, MA, USA,

4. NASA GSFC, Greenbelt, MD, USA,

5. Department of Physics and Astronomy, George Mason University, Fairfax, Virginia, USA,

6. NOAA SWPC, Boulder, CO, USA,

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2018SW002034

7. High Altitude Observatory, NCAR, Boulder, CO, USA,

8. Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany,

9. Utah State Univ. Logan, UT, USA

10. Plasma Physics Division, Naval Research Laboratory, Washington, D. C., USA,

11. Murmansk Arctic State University, Murmansk, Russia,

12. Space Physics Research Laboratory, Univ. of Michigan, Ann Arbor, MI, USA,

Abstract

To address challenges of assessing space weather modeling capabilities, the CCMC (Community Coordinated Modeling Center) is leading a newly established "International Forum for Space Weather Modeling Capabilities Assessment." This paper presents preliminary results of validation of modeled foF2 (F2-layer critical frequency) and TEC (Total Electron Content) during the first selected 2013 March storm event (17 March, 2013). In this study, we used eight ionospheric models ranging from empirical to physics-based, coupled ionosphere-thermosphere and data assimilation models. The quantities we considered are TEC and foF2 changes and percentage changes compared to quiet-time background, and the maximum and minimum percentage changes. In addition, we considered normalized percentage changes of TEC. We

compared the modeled quantities with ground-based observations of vertical GNSS (Global Navigation Satellite System) TEC (provided by MIT Haystack Observatory) and foF2 data (provided by GIRO, Global Ionospheric Radio Observatory) at the 12 locations selected in middle latitudes of the American and European-African longitude sectors. To quantitatively evaluate the models' performance, we calculated skill scores including Correlation Coefficient (CC), Root Mean Square Error (RMSE), ratio of the modeled to observed maximum percentage changes (Yield), and Timing Error. Our study indicates that average RMSEs of foF2 range from about 1 MHz to 1.5 MHz. The average RMSEs of TEC are between ~5 TECU and ~10 TECU. dfoF2[%] RMSEs are between 15 % and 25 %, which is smaller than RMSE of dTEC[%] ranging from 30% to 60 %. The performance of the models varies with the location and metrics considered.

1. Introduction

In the past few decades, our general understanding of average behavior of the ionosphere during both quiet and disturbed times and the physical processes responsible for it has been established. However, there is still a lack of knowledge of how solar eruptive phenomena affect the interplanetary space and the near-Earth space environment, including the Earth's upper atmosphere. Modeling is crucial to not only advancing our understanding of causes and consequences of the space weather, but also mitigating storm impacts on our modern daily lives.

Therefore, ionosphere-thermosphere models have been developed to predict intensity, occurrence, and duration of ionospheric space weather effects on both ground and space-based systems, including satellite communication and navigation systems, electric power grids and pipelines.

To address the needs and challenge of assessment of our current knowledge about space weather effects on ionosphere-thermosphere (IT) system and current state of IT modeling capabilities, the Community Coordinated Modeling Center (CCMC) has been supporting community-wide model validation projects, including CEDAR and GEM-CEDAR Modeling Challenges. The CCMC initiated the CEDAR ETI (Electrodynamics Thermosphere Ionosphere) Challenge in 2009 focusing on the evaluation of basic IT system parameters modeled, such as electron and neutral densities, the ionospheric F2-layer peak electron density (NmF2) and peak height (hmF2), and vertical drift [*Shim et al.*, 2011, 2012, 2014] during both quiet and storm periods. The CEDAR-GEM Challenge that followed in 2011 focuses on assessing the prediction of geomagnetic storm impacts on the ionosphere-thermosphere, including Joule Heating [*Rastätter et al*., 2016], neutral density [*Kalafatoglu Eyiguler et al*, 2018], and TEC [*Shim et al.*, 2017a]. *Shim et al.* [2017a] reported the results of evaluation of TEC prediction during the 2006 AGU storm (14-15 Dec) in eight 5°-wide longitude sectors using 15 simulations obtained from eight ionospheric models.

Since last year, the CCMC has been leading a newly established "International Forum for Space Weather Modeling Capabilities Assessment" to define metrics to assess the current

state of space weather modeling capabilities and to track the scientific progress in models that feed into operations. As part of this international effort, Ionosphere-Thermosphere Working Group consisting of four working teams was formed [*Scherliess et al*., 2018]. One of them, "Ionosphere Plasmasphere Density Working Team", focuses on the evaluation of storm time TEC and foF2 predictions, which are the most important ionospheric characteristic parameters to assess space weather impact on radio propagation. The "Ionosphere Plasmasphere Density Working Team" performed its first validation study of foF2 and TEC predictions using various Ionosphere/Thermosphere models. This study is the first quantitative assessment of both parameters based on a set of metrics, unlike a number of validation studies done previously [*Burns et al.*, 2008; *Anderson et al*., 1998; *Fuller-Rowell et al*., 2000; *Araujo-Pradere et al*., 2007; *Feltens et al*., 2011; *Orús et al*., 2002, 2003; *Zhu et al*., 2006; *Perlongo et al*., 2017].

In this paper, we present preliminary results of the evaluation of TEC and foF2 predictions during the 2013 March storm (17 March, DOY (Day Of Year) 076), which is one of the storm events selected by the Ionosphere-Thermosphere Working Group [*Scherliess et al*., 2018] to investigate differences and similarities in the performance of the models. We evaluated performance of eight IT models in middle latitudes first, which are known to be the best understood ionospheric region largely due to the relatively simple physics and reasonably good coverage of measurements. Moreover, the good coverage of measurements in this region is an additional advantage for model-data comparisons that we attempt here. In Section 2, we briefly describe the GNSS TEC and foF2 measurements, models, and metrics used for this study. In

Section 3, the results of model-data comparisons and performance of the models are presented. Finally, the summary and conclusions are in Section 4.

2. Methods

2.1 Observations

In this study, we first selected 12 ionosonde stations in middle latitudes of the American and European-African longitude sectors (see Figure 1 and Table 1) including 4 stations located in the Southern Hemisphere in order to investigate longitudinal dependence and hemispheric asymmetry of the performance of the models. We used foF2 observations provided from Global Ionosphere Radio Observatory (GIRO) (http://giro.uml.edu/) [*Reinisch and Galkin,* 2011] and GNSS vertical TEC (vTEC) data provided by MIT Haystack Observatory (http://cedar.openmadrigal.org/, http://cedar.openmadrigal.org/cgi-bin/gSimpleUIAccessData.py) [*Rideout and Coster*, 2006]. For TEC model-data comparison, at any given location, we used 15 min × 3º lat × 5º lon running average every 5 minutes using 1º lat × 1º lon × 5 min resolution TEC data. The averaged error of the TEC data over all 12 locations for this selected time interval, DOY 076, including 30 quiet days, is about 1 TECU (1 total electron content unit = 10^{16} el m^{-2}). The number of data points used at each location is 288. For GIRO foF2 observations, we used 45-min running average of 15-min resolution data at all stations except for Port Stanley for which we used 60 min running average due to relatively coarse data coverage.

The number of data points at Port Stanley is 48, the other locations have more than 90 data points with a maximum of 96.

2.2 Models

The simulations used in this study were obtained from the latest version of the models available at the CCMC [*Webb et al.*, 2009], with the exception of UAM simulation submitted by the model developer. The models used are: IRI (International Reference Ionosphere) storm model, empirical; SAMI3 (Sami3 is Also a Model of the Ionosphere), IFM (ionosphere Forecast Model), physics-based ionosphere; CTIPe (Coupled Thermosphere Ionosphere Plasmasphere Electrodynamics Model), GITM (Global Ionosphere Thermosphere Model), TIE-GCM (Thermosphere Ionosphere Electrodynamics General Circulation Model), UAM-P (Upper Atmosphere Model - Potsdam Version), physics-based coupled ionosphere-thermosphere; and USU-GAIM (Utah State University Global Assimilation of Ionospheric Measurement), physics-based ionosphere data assimilation model. Table 2 shows the version of the models, input data used for the simulations, and models used for lower boundary forcing and high latitude electrodynamics. We used unique model setting identifiers to distinguish the simulations used in this study from those, obtained by using different version, input drivers, and/or different boundary conditions, used in our previous studies [*Shim et al.*, 2011, 2012, 2014, 2017a].

Additional information on the models and model setting identifiers is available in *Shim et al*. [2011] (please refer to all references included) and at http://ccmc.gsfc.nasa.gov/challenges/GEM-CEDAR/tags_list.php.

It is worth pointing out the factors in each model simulation that influence foF2 and TEC responses to geomagnetic storms. The IRI storm model [*Fuller-Rowell et al.*, 1999, 2000], based on the large number of storms covered by the long record of data from the worldwide network of ionosondes, describes the ratio of storm-time foF2 to monthly average foF2 in a first order approach that captures the summer hemisphere mid-latitude ionospheric response. 4_IRI was driven by the time history of the 3-hour ap index over the preceding 30 hours with a weighting function deduced from physically based modeling. The two physics-based ionosphere model simulations, 1_IFM and 1_SAMI3 used the same empirical inputs for the neutral atmosphere and E×B drift. There are notable differences between them: the version of the SAMI3 currently hosted at the CCMC does not include high-latitude ionospheric electrodynamics (HLIE, e.g., the auroral precipitation and the convection electric field pattern), while 1_IFM used empirical models for the HLIE; 1_SAMI3 includes the plasmasphere, whereas 1_IFM does not. All four coupled IT models can be coupled to various models of the HLIE. For the simulations used in this study, 1_UAM-P was driven by FAC (Field Aligned Current). The other three simulations, 11_CTIPE, 11_TIE-GCM, and 6_GITM used the same empirical model for high-latitude electric potential (Weimer-2005 [*Weimer*, 2005]); however, to drive Weimer-2005, 11_CTIPE and 6_GITM used the IMF (Interplanetary Magnetic Field) and solar wind speed and density from

the ACE 1-minute resolution data (http://www.srl.caltech.edu/ACE/ASC), while 11_TIE-GCM used the 15-minute trailing average (lagged by 5 minutes and sampled at 5 minutes) derived from the 1-minute OMNI data set (http://omniweb.gsfc.nasa.gov); each of them was driven by a different model for the auroral precipitation. 6_GITM used FISM (Flare Irradiance Spectral Model, http://lasp.colorado.edu/lisird/data/fism) solar EUV irradiance [*Chamberlin et al*., 2007], while the others used F10.7 to specify the solar EUV flux using EUVAC (EUV flux model for Aeronomic Calculations) [*Richards et al.,* 1994]. 1_UAM-P and 11_CTIPE include the plasmasphere, while 6_GITM and 11_TIE-GCM [*Solomon et al*., 2018] do not. The data assimilation model simulation, 1 _USU-GAIM currently used only GNSS TEC data between -60 and +60 geographic latitudes.

2.3 Metrics

As the quiet-time reference (TEC_quiet) to quantify impacts of the storm, we used 30-day median value at a given time. The 30 days consist of 15 days before (03/01-03/15/2013) and 15 days after (03/22-04/05/2013) the storm. The quantities considered to assess how well the models produce foF2 and TEC during the storm are, at any given location, (1) shifted values by subtracting a minimum value of the quiet time reference in 24 hours (e.g., TEC*(DOY, UT) = TEC(DOY, UT) - minimum of TEC_quiet = TEC(DOY, UT) – minimum 30-day median, (2) changes due to the storm with respect to the quiet time reference (e.g., dTEC(DOY, UT) =

TEC(DOY, UT) – TEC_quiet(UT)), and (3) percentage changes (e.g., dTEC[%](DOY, UT) = 100*dTEC(DOY, UT)/TEC_quiet(UT)).

Potential systematic uncertainties in the models and observations and baseline differences between the models and between models and observations can be reduced by considering the shifted values and changes from the their own quiet-time background values. In addition, the impact of differing upper boundaries for TEC calculations is likely reduced by using these quantities, since plasmaspheric TEC variations with geomagnetic activity are negligible in middle latitudes [*Shim et al*., 2017b].

For TEC validation, we also considered normalized percentage changes of TEC (dTEC[%]_norm = [dTEC[%] -ave_dTEC[%]]/std_dTEC[%], where ave_dTEC[%] is the average of percentage changes of TEC (dTEC[%]) at a given time and at a given location over the quiet 30 days described above, and std_dTEC[%] is the standard deviation of the average percentage change). The normalized percentage change, dTEC[%]_norm, has the advantage to exclude seasonal, local time, and latitudinal dependences of TEC variability by normalizing the percentage variation using its statistical standard deviation, and thus should have mainly storm-induced variation [*Nishioka et al*., 2017]. Figure 2 shows an example of dTEC[%] (in red) and dTEC[%]_norm (in blue) at Athens on DOY 076 (between 06UT and 22UT). The difference between local daytime (around 12 UT) and nighttime (around 21 UT) in dTEC[%] peak value has disappeared in dTEC[%]_norm.

To quantify the performance of the models, we calculated the following skill scores: including (1) Correlation Coefficient (CC), (2) Root-Mean-Square Error (RMSE $=\sqrt{\frac{\sum(x_{obs}-x_{mod})^2}{N}}$, where x_{obs} and x_{mod} are observed and modeled values), (3) the ratio of the peak of modeled percentage change to that of the observed one (Yield = $\frac{(x_{mod})_{max}}{(x_{obs})_{max}}$), e.g., [max. (min.) dTEC[%] of the model simulations]/[max. (min.) dTEC[%] of GNSS vTEC for positive (negative) phase] during the main phase of the storm, and (4) Timing Error, which is the difference between the modeled peak time and observed peak time (TE = t_peak_model – t_peak_obs.).

Correlation Coefficient measures how well the observed and modeled values are linearly correlated (in phase) with each other, and RMSE measures how different the values are on average over the time interval considered. We calculated CC and RMSE for the error values below 95th percentile in order to remove the effect of outliers of the predicted values, which is one of the main concerns of using CC and RMSE. In addition to CC and RMSE, we considered Yield and Timing Error to measure the models' capability to capture peak disturbances during the storm.

3. Performance of the Models in Predictions of foF2 and vTEC on 17 March 2013

There have been a number of extensive studies on geospace responses to the 2013 March storm (see Figure 3 for solar wind and geomagnetic conditions) together with the 2015 March

storm [*Zhang et al.*,2017, please refer to the references included]. Figure 4 shows foF2 and TEC during the main phase of the 2013 (in red) along with the 30-day medians (in blue) at the 12 locations. As reported by *Yue et al.* [2016], the storm-enhanced density (SED) in middle-high latitudes and its source, the positive ionospheric storm effects, in the lower latitudes were observed around 12 UT in the European sector (EU) and around 20UT in the North American sector (NA), while the negative storm effects on the west side of the NA sector were also observed (at Idaho and Boulder) in the right two columns in Figure 4. In the Southern Hemisphere (SH), the positive TEC changes were observed in both South America and South Africa. The foF2 responses to the storm are similar to the TEC responses with larger change in TEC at the selected locations except for Chilton and Grahamstown in which there were noticeable increases in TEC, while foF2 hardly increases. We found that most model simulations do not reproduce the difference between eastern and western parts of the NA sector; for example, TEC increases at Millstone Hill and decreases at Idaho and Boulder around 20UT, and foF2 and TEC show different responses in Chilton and Grahamstown (see Figure S1 in supporting information).

Figure 5 shows scatter plots of the observed (x-axis) and modeled (y-axis) shifted foF2 (foF2* in the 1st and 3rd columns) and percentage change of foF2 (dfoF2[%] in the 2nd and 4th columns) during the storm (03/17/2013) for all 12 locations grouped into 4 sectors: North America (NA, denoted in green), Europe (EU, blue), South Africa (SAF, red), and South America (SAM, black). For most simulations, after shifting by subtracting the minimum of 30-

day median, the modeled foF2* agrees better with the observed one than before shifting (see Figure S2 in supporting information). For example, before shifting most foF2 data points of 1_GITM and 1_USU-GAIM are below and above the line with slope 1 (black solid line), respectively. This indicates that 1_GITM underestimates foF2, while 1_USU-GAIM overestimates it. They each produce the smallest (0.8 MHz) and largest (4.5 MHz) averaged minimum of 30-day median over all 12 locations compared to the observed one (3.6 MHz) (see Figure S1). Most models appear unable to produce foF2* larger than about 7 MHz during the storm. The modeled dfoF2[%] shows less agreement with the observed values than the modeled foF2* does. The modeled dfoF2 (not shown here) and dfoF2[%] show similar relationship with the observed values.

In Figure 6, we show scatter plots of the observed (x-axis) and modeled (y-axis) shifted TEC and percentage change of TEC, different colors are associated with different sectors (same as Figure 5). It is found that the averaged minimum of 30-day median of the observed TEC is about 5 TECU, and the lowest and highest modeled averages are 0.2 TECU (6_GITM) and 7.8 TECU (1_UAM-P). The relationships between the modeled and observed TEC* are similar to those of foF2* with the tendency for most of the simulations to underestimate TEC* larger than around 30-40 TECU. This tendency is more pronounced in 11_CTIPE and 6_GITM simulations.

Most models appear to not reproduce the large enhancements of dTEC[%] (about 200 %) at Port Stanley in South America. 1_USU-GAIM agrees better with GNSS vTEC than the other simulations do due to the fact that 1_USU-GAIM assimilated GNSS observations, which are

included to produce the MIT TEC data set. Like the predicted dfoF2[%], the modeled dTEC[%] shows less agreement with the observed value than the modeled TEC* does. From now on, foF2 and TEC will represent shifted foF2 (foF2*) and shifted TEC (TEC*), respectively.

3.1 Correlation Coefficient (CC)

To quantitatively assess the models' performance of TEC and foF2 predictions, we first calculated correlation coefficient (CC) between the modeled and observed foF2 and TEC for DOY 076. In Figure 7, the simulations were first grouped based on the type of the model and then arranged alphabetically in each group (foF2 in the left panel and TEC in the right panel). Four CC were displayed for each simulation. First three scores correspond to the average CC over Europe (EU), North America (NA), Southern Hemisphere (SH refers to SAF and SAM combined), and the last one is the average of all 12 locations. Different colors denote different quantities: Blue denotes shifted foF2 and TEC, green and red the change and percentage changes, and black normalized percentage change. The closer the circles are to the horizontal line of 1, the better the model performs. The modeled foF2 and TEC (blue dots) are highly correlated with the observed values.

The average CC values, over all 12 locations, between the modeled and observed foF2 (TEC) are about 0.8-0.9 (0.7- 0.95). However, the modeled changes (green), percentage changes (red) and normalized percentage changes (black only applicable for TEC) are much less correlated (closer to uncorrelated) with the observed values, especially for dfoF2 and dfoF2[%] (

~ -0.1 < average CC < 0.4). There is no big difference between dTEC[%] and dTEC[%]_norm based on the average values for each simulation.

For the changes and percentage changes, the physics based coupled IT and data assimilation models have better CC than the ionosphere models have. The differences in CC among locations are smaller for foF2 and TEC than for the changes and percentage changes. Most simulations, except for 6_GITM, show lower CC for dfoF2 and dTEC in NA. It seems to be caused by the opposite responses to the storm in the eastern (positive phase) and western (negative phase) parts of NA not being captured, producing either positive (most physics-based models) or negative (4_IRI) storm in all locations of NA instead. 6_GITM shows the opposite responses relatively well although it underestimated the magnitude of the change.

4_IRI produces CC comparably to the physics-based models that perform best for both foF2 (11_CTIPE) and TEC (11_TIE-GCM) predictions, although it does not for dTEC and dTEC[%].

To compare the two physics-based ionosphere models, 1_IFM performs about the same as 1_SAMI3 for all the considered quantities based on the average CC, but 1_SAMI3 shows less correlation with the observed foF2 and TEC in EU, compared to 1_IFM.

Four physics-based coupled ionosphere-thermosphere models also produce similar average CC. However, 11_CTIPE (11_TIE-GCM) shows high correlation for all three sectors for foF2 (TEC). Except for 6_GITM, the other three simulations show lower CC for dfoF2 and dTEC in North America (NA) than the other sectors. 11_TIE-GCM and 6_GITM shows the most dependence of CC for dTEC and dfoF2 on location, respectively.

3.2 Root Mean Square Error

Figure 8 shows RMSE of foF2 and dfoF2, in the left panel, and TEC and dTEC, in the right panel. For foF2 (blue) and dfoF2 (green) predictions, average RMSE values range from ~1 MHz (1_IRI and 6_GITM) to 1.5 MHz (1_SAMI3 and 1_UAM-P). 11_CTIPE has the lowest and highest RMSE for foF2: 0.7 in NA and 2 MHz in SH.

Average RMSE of TEC are between 5 TECU (1_USU-GAIM) and 10 TECU (11_CTIPE and 6_GITM). However, the minimum RMSE for TEC is 3 TECU in NA (11_TIE-GCM), and the maximum RMSE is about 13 TECU in SH (11_CTIPE and 6_GITM). For dTEC, average RMSE varies from about 4 TECU (1_USU-GAIM) and 7 TECU (1_UAM-P).

As seen in CC, RMSE also varies highly with location. Most models appear to predict foF2 and/or TEC better in NA and worse in SH (e.g., 11_CTIPE has its lowest of 5 TECU in NA and highest of 13 TECU in SH). Some models have the same tendency of better performance in NA for dTEC (e.g., 4_IRI, 11_CTIPE, 6_GITM), while most models show the opposite tendency for dfoF2 prediction (e.g., 1_IFM, 1_SAMI3, 11_CTIPE, 6_GITM, 11_TIE-GCM 1_UAM-P). However, 4_IRI and 1_USU-GAIM performed best for foF2 and TEC predictions, respectively, show less location dependence of RMSE than the others for foF2 and TEC predictions.

Shim et al. [2017a] found that the longitudinally averaged RMSE for both TEC* and dTEC (using TEC one day prior to the storm as a quiet time reference) in middle latitude during 2006 Dec. event ranges from about 2 to 8 TECU, which are about 2-3 TECU smaller than average

RMSE for 2013 March event. Overall, most models perform somewhat worse for 2013 March event than for 2006 December event. This is possibly partly attributed to TEC prediction at the 4 stations in the South Atlantic Anomaly (SAA) region (see Figure S3), which is not considered by *Shim et al.* [2017a]. In the SAA region, TEC enhancement occurred due to the energy input from the outer radiation belts [*Dmitriev et al.*, 2017], however, most models do not predict the large enhancement of TEC (e.g., dTEC[%] at Port Stanley) as we mentioned above.

Based on the average RMSE, for TEC and dTEC, 1_SAMI3 (1_IFM) tends to perform worse in EU (SH) than the other sectors. Among the four physics-based coupled models, 11_TIE-GCM predicts TEC and dTEC better, 11_CTIPE and 6_GITM perform better for dfoF2, and 1_UAM-P predicts foF2 better.

Figure 9 shows RMSE of percentage changes of foF2 (blue) and TEC (red) and normalized percentage changes of TEC (black). RMSEs of dfoF2[%] based on the average over the 12 locations are between 15 % and 25 %, which is smaller than RMSE of dTEC[%] ranging from 30% to 60 %. Difference in the performance among locations is more noticeable in dTEC[%] and dTEC[%]_norm than in dfoF2[%]. All simulations, except 6_GITM, have smaller RMSE of dTEC[%] in NA than in the other two sectors. 4_IRI produces comparable dfoF2[%] and dTEC[%] to those of the physics based models for most cases. However, 4_IRI has larger RMSE in dTEC[%]_norm than the others do possibly due to the small standard deviation of dTEC[%]. It needs to be noted that IRI represents average conditions of the ionosphere and has a limitation to model relatively short-term disturbances during geomagnetic activities.

Differences in the performance based on RMSE of dTEC[%] among the simulations appear do not match those based on RMSE of dTEC[%]_norm. For example, 1_SAMI3 shows slightly larger average RMSE of dTEC[%], but smaller RMSE of dTEC[%]_norm than 1_IFM. 11_TIE-GCM shows smaller average RMSE of dTEC[%], but larger RMSE of dTEC[%]_norm than the other three coupled models. This is probably due to the difference in the normalization factor, standard deviation of dTEC[%], between the simulations.

3.3 Yield and Timing Error (TE)

In order to measure how well the models capture the degree of TEC and foF2 disturbances during the main phase, we calculated Yield and Timing Error of dfoF2[%], dTEC[%], and dTEC[%]_norm. In most of the 12 locations two peaks were observed as seen in Figure 2, therefore, we considered two time intervals, 06-15UT and 15-22UT, which correspond to about 06-17LT and 15-24LT in European-African longitude sector and about 22-11LT and 08-18LT in American longitude sector (see Figure S4 for Yield and Timing Error for all locations for each model). In each time interval, one Yield value and one TE value were calculated as defined in Sec. 2.3.

The Yield is positive if the model correctly predicts the storm phase. In each sector, we calculate average Yield and average Timing Error (TE) over the number of stations where the Yield is positive. Table 3 shows the total number of stations where the models have correct storm phase, either positive or negative. The numbers in bold are the highest among the

simulations. The coupled IT models and data assimilation model predict the storm phase better than the other three simulations, 4_IRI, 1_IFM and 1_SAMI3.

In Figure 10, we present average Yield (left) and average of absolute values of TE (right) over the two time intervals: dfoF2[%] in blue, dTEC[%] in red, and dTEC[%]_norm in black. Concerning the average of all 12 locations, Yields of dfoF2[%] and dTEC[%] range from 0.5 to 1.5 with larger variation for dTEC[%] with location. For dTEC[%]_norm, Yield values are between 0.8 and 3. Some of the simulations give better ratio for dTEC[%]_norm (closer to the horizontal line of 1) than for dTEC[%] (e.g., 11_CTIPE, 1_SAMI3, and 1_UAM-P), some of them give worse ratio for dTEC[%]_norm (e.g., 11_TIE-GCM in EU and 4_IRI in NA, which shows the largest difference in Yield between dTEC[%] and dTEC[%]_norm). In most cases, 11_CTIPE and 6_GITM appear to underestimate percentage changes, while 1_UAM-P and 11_TIE-GCM tend to overestimate.

Average Timing Errors of dfoF2[%] and dTEC[%]_norm are between 1 and 4 hours, and TE of dTEC[%] are about 1-5 hours. 1_USU-GAIM has smaller TE for dTEC[%] (~1.5 hrs) than for dfoF2[%] (~3 hrs). The physics-based models have 1-2 hour average TE for dfoF2[%], with the smallest of ~0.5 hrs in EU predicted by 11_CTIPE, and 1.5-2.5 hour average TE for dTEC[%] and dTEC[%]_norm.

1_SAMI3 gives slightly better Yield and TE for most cases than 1_IFM, and shows better Yield and worse TE of dTEC[%]_norm compared to dTEC[%].

4. Summary and Conclusions

We performed a systematic assessment of TEC and foF2 predictions of ionosphere-thermosphere models during the 2013 March storm event (DOY 076). We compared modeled foF2 and TEC obtained from eight ionospheric models with the ground-based GIRO foF2 and GNSS vertical TEC data for the selected 12 locations in middle latitudes of the American and European-African longitude sectors.

The quantities considered for model-data comparison are (1) shifted foF2 and TEC (e.g., TEC*) that were obtained by subtracting minimum of quiet-time reference, 30-day median, (2) foF2 and TEC changes (e.g., dTEC) with respect to the quiet-time reference, (3) percentage changes of foF2 and TEC (e.g., dTEC[%]), and (4) normalized percentage change of TEC (dTEC[%]_norm) by standard deviation of dTEC[%].

To quantify the performance of the models, we calculated skill scores, including Correlation Coefficient (CC), RMS error (RMSE), the ratio of maximum (minimum) percentage change of the models to the observation (Yield), and the Timing Error (TE: peak time difference between the simulations and the observations = t_peak_model – t_peak_obs). The skill scores were calculated for the three sectors: EU (Europe), NA (North America), and SH (Southern Hemisphere) to investigate the longitudinal and hemispheric dependence of the performance of the models.

Comparisons of the eight simulations indicate that most simulations tend to underestimate enhanced foF2 and TEC and appear to not reproduce large enhancements of dTEC[%] (e.g.,

about 200 % TEC increase at Port Stanley in the SAA region). Most of them also appear to not capture opposite responses to the storm in the eastern and western parts of NA. This is what in part causes lower CC for dfoF2 and dTEC in NA. However, with respect to RMSE, most models tend to predict foF2 and/or TEC better in NA and worse in SH. Some models have the same tendency of better performance in NA for dTEC (e.g., 4_IRI, 11_CTIPE, 6_GITM). Averaged RMSEs of foF2, over all 12 locations, range from about 1 MHz to 1.5 MHz. The average RMSEs of TEC are between about 5 TECU and 10 TECU. dfoF2[%] RMSEs are between 15 % and 25 %, which is smaller than RMSE of dTEC[%] ranging from 30% to 60 %. In terms of CC, the physics-based coupled IT (11_CTIPE, 6_GITM, 11_TIE-GCM, and 1_UAM-P) and data assimilation (1_USU-GAIM) models predict the changes and percentage changes better than the three ionosphere models (4_IRI, 1_IFM, and 1_SAMI3). In addition, the coupled IT models and data assimilation model predict the storm phase better than the other three simulations, 4_IRI, 1_IFM and 1_SAMI3.

As rather expected based on the fact that the empirical model IRI represents average conditions of the ionosphere, the IRI performs comparably to the physics-based models concerning RMSE and CC, which measure the performance of the models on average over the time interval considered. 4_IRI produces the smallest RMSE for foF2 prediction with less location dependence. However, the IRI has a limitation in modeling relatively short-term disturbances during geomagnetic activities, therefore it performs relatively poorly in its prediction of the maximum and minimum of percentage changes (Yields and Timing Error,

especially in NA, for dTEC[%]_norm). Compared to the physics-based model simulations, 4_IRI has larger RMSE in dTEC[%]_norm due to the small standard deviation of dTEC[%]. 1_USU-GAIM, which assimilated only GNSS TEC data, performs best for TEC prediction in most cases with the least location dependence of RMSE. 1_USU-GAIM makes better predictions of TEC related quantities than foF2 related ones. This suggests that the integration of additional observations, such as ionosonde profiles, is necessary to make better prediction of foF2.

To compare the same types of models, two physics-based ionosphere models, 1_IFM and 1_SAMI3 show similar performance in terms of average CC and RMSE. 1_SAMI3 shows slightly worse CC and RMSE for foF2 and TEC in EU, and slightly better Yield and Timing Error for most cases. Overall, the average performances of the four physics-based coupled IT models over the 12 locations are comparable to each other. Among the four simulations, only 6_GITM captures the opposite responses in the eastern and western parts of NA relatively well, although it underestimates the magnitude of the change. Therefore, unlike the other simulations, 6_GITM produces better CC for dfoF2/dTEC in NA than in the other two sectors. Based on the average RMSE, 11_TIE-GCM predicts TEC and dTEC better, 11_CTIPE and 6_GITM perform better for dfoF2, and 1_UAM-P predicts foF2 better. In terms of CC, 11_CTIPE and 11_TIE-GCM show high correlation for all three sectors for foF2 and TEC, respectively. In most cases, 11_CTIPE and 6_GITM tend to underestimate percentage changes, while 1_UAM-P and 11_TIE-GCM overestimate them. The differences in performance among these four simulations could be caused by inherent differences among the models, for example, different methods to

solve for chemistry and advection, and different ways to treat eddy diffusion and vertical transport [*Fuller-Rowell et al., 1996; Ridley et al.*, 2006; *Solomon et al.*, 2012; *Prokhorov et al.* 2018; *Perlongo et al*.,2017]. In addition, the performance differences could also be caused by combination of different input data and different models used for lower boundary forcing and high-latitude electrodynamics, since each simulation was obtained by using its default input data and drivers. To investigate actual causes of these differences, detailed studies will be required, which are beyond the scope of this paper.

Furthermore, our findings suggest that in order to accurately model ionosphere-thermosphere disturbances during geomagnetic storms, it is crucial to advance our knowledge on the coupling of ionosphere-thermosphere with magnetosphere to correctly describe penetration electric fields and enhancement of energy deposition, which leads to composition changes, winds, and electron density changes [*Huba et al*., 2015; *Maruyama et al*., 2005]. Theoretical modeling will help us better understand the processes responsible for the observed characteristics and features during disturbed conditions, and this will improve our space weather forecasting capabilities.

This is the first systematic study to evaluate current capability of modeling foF2 and TEC during geomagnetic storms, performed by " Ionosphere Plasmasphere Density Working Team" of the International Forum for Space Weather Modeling Capabilities Assessment. The findings of this study will provide a baseline for future validation studies using new models and improved models, along with earlier results [*Shim et al.*, 2011, 2012, 2014, 2017a] obtained through

CEDAR ETI and GEM-CEDAR Modeling Challenges. We will extend our study to include more geomagnetic storm events selected by the Ionosphere-Thermosphere Working Group [*Scherliess et al*., 2018], including the 2015 March and 2017 September events, to investigate differences and similarities in the performance of the models. In addition, we will also include foF2 and TEC predictions in high and low latitudes.

Acknowledgement

The vertical TEC data were provided by MIT Haystack Observatory and can be obtained through CEDAR Madrigal database (http://cedar.openmadrigal.org). We thank the operators of the digisondes for sharing their data through http://giro.uml.edu/. Data from the South African Ionosonde network is made available through the South African National Space Agency (SANSA), who are acknowledged for facilitating and coordinating the continued availability of data. This work is supported by grants from the National Science Foundation (NSF) Space Weather Program. This model validation study is supported by the Community Coordinated Modeling Center (CCMC) at the Goddard Space Flight Center. Data processing and research at MIT Haystack Observatory are supported by cooperative agreement AGS-1242204 between the U.S. National Science Foundation and the Massachusetts Institute of Technology. The National Center for Atmospheric Research is sponsored by the National Science Foundation. Model output and observational data used for the study will be permanently posted at the CCMC

website (http://ccmc.gsfc.nasa.gov) and provided as a resource for the space science community to use in the future.

References

Anderson, D. N., et al. (1998), Intercomparison of physical models and observations of the ionosphere, *J. Geophys. Res.*, *103*, 2179 – 2192, doi:10.1029/97JA02872.

Araujo-Pradere, E. A., T. J. Fuller-Rowell, P. S. J. Spencer, and C. F. Minter, (2007), Differential validation of the US-TEC model, *Radio Sci.*, *42*, RS3016, doi:10.1029/2006RS003459.

Bilitza, D., et al, (2014), The International Reference Ionosphere 2012 – a model of international collaboration, *J. Space Weather Space Clim.*, *4*, A07, 1-12, doi:10.1051/swsc/2014004.

Bilitza, D., D. Altadill, V. Truhlik, V. Shubin, I. Galkin, B. Reinisch, and X. Huang (2017), International Reference Ionosphere 2016: From ionospheric climate to real-time weather predictions, *Space Weather*, 15, 418 –429, doi:10.1002/2016SW001593.

Burns, A. G., W. Wang, M. Wiltberger, S. C. Solomon, H. Spence, T. L. Killeen, R. E. Lopez, and J. E. Landivar (2008), An event study to provide validation of TING and CMIT geomagnetic

middle-latitude electron densities at the F2 peak, *J. Geophys. Res.*, *113*, A05310, doi:10.1029/2007JA012931.

Chamberlin, P. C., Woods, T. N., & Eparvier, F. G. (2007). Flare Irradiance Spectral Model (FISM): Daily component algorithms and results. *Space Weather, 5,* S07005. https://doi.org/10.1029/2007SW000316

Codrescu, M. V., T. J. Fuller-Rowell, J. C. Foster, J. M. Holt, and S. J. Cariglia, (2000), Electric field variability associated with the Millstone Hill electric field model, *J. Geophys. Res*., 105, 5265–5273, doi:10.1029/1999JA900463.

Chamberlin, P. C., Woods, T. N., & Eparvier, F. G. (2007). Flare Irradiance Spectral Model (FISM): Daily component algorithms and results. *Space Weather, 5,* S07005. https://doi.org/10.1029/2007SW000316

Dmitriev A. V., V. Suvorova, M. V. Klimenko, V. V. Klimenko, K. G. Ratovsky, R. A> Rakhmatulin, V. A. Parkhomov, (2017) Predictable and unpredictable ionospheric disturbances during St. Patrick's Day magnetic storms of 2013 and 2015 and on 8–9 March 2008. *J Geophys Res Space Phys 122*:2398–2423. doi:10.1002/2016JA0232

Drob, D. P., et al., (2008), An empirical model of the Earth's horizontal wind fields: HWM07, *J. Geophys. Res.*, 113, A12304, doi:10.1029/2008JA013668.

Feltens, J., M. Angling, N. Jackson-Booth, N. Jakowski, M. Hoque, M. Hernández-Pajares, A. Aragón-Àngel, R. Orús, and R. Zandbergen, (2011), Comparative testing of four ionospheric models driven with GPS measurements, *Radio Sci., 46*, RS0D12, doi:10.1029/2010RS004584.

Fuller -Rowell, T. J., and D. S. Evans, (1987), Height-Integrated Pedersen and Hall Conductivity Patterns Inferred From the TIROS-NOAA Satellite Data, *J. Geophys. Res.*, 92(A7), 7606–7618.

Fuller-Rowell, T. J., Rees, D., Quegan, S., Moffett, R. J., Codrescu, M. V., and Millward, G. H., A coupled thermosphere-ionosphere model (CTIM), *in Handbook of Ionospheric Models*, edited by R. W. Schunk, Utah State Univ., Logan, Utah, 217–238, 1996

Fuller-Rowell, T., M. Codrescu, E. Araujo-Pradere, and I. Kutiev, (1999), Progress in developing a storm-time ionospheric correction model, *Adv. Space Res., 22*(6), 821-827, doi:10.1016/S0273-1177(98)00105-7.

Fuller-Rowell, T. J., M. C. Codrescu, and P. Wilkinson (2000), Quantitative modeling of the ionospheric response to geomagnetic activity, *Ann. Geophys., 18*, 766– 781, doi:10.1007/s00585-000-0766-7.

Fuller-Rowell, T., E. Araujo-Pradere, and M. Codrescu, (2000), An empirical ionospheric storm-time correction model, *Adv. Space Res.*, *25*(1), 139-148, doi:10.1016/S0273-1177(99)00911-4.

Hagan, M. E., M. D. Burrage, J. M. Forbes, J. Hackney, W. J. Randel, and X. Zhang, (1999), GSWM-98: results for migrating solar tides. *J. Geophys. Res.* 104: 6813–6828.

Hardy, D. A., M. S. Gussenhoven, and E. Holeman (1985), A statistical model of auroral electron precipitation, *J. Geophys*. Res., 90, 4229–4248.

Hedin, A. E. (1991), Extension of the MSIS thermospheric model into the middle and lower atmosphere, *J. Geophys. Res*., 96, 1159–1172.

Hedin, A. E., et al. (1991), Revised global model of thermospheric winds using satellite and ground-based observations, *J. Geophys. Res*., 96, 7657–7688.

Heelis, R. A., J. K. Lowell, and R. W. Spiro, (1982), A Model of the High-Latitude Ionospheric Convection Pattern, *J. Geophys. Res*. 87, 6339.

Heppner, J.P. and N.C. Maynard (1987), Empirical High Latitude Electric Field Models, *J. Geophys. Res.* **92**, 4467–4489.

Huba, J., G. Joyce, and J. Fedder, (2000), Sami2 is Another Model of the Ionosphere (SAMI2): A new low-latitude ionosphere model, *J. Geophys. Res.*, 105, 23,035.

Huba, J. D., G. Joyce, and J. Krall, (2008), Three-dimensional equatorial spread F modeling, *Geophys. Res. Lett.*, 35, L10102, doi:10.1029/2008GL033509.

Huba, J. D., S. Sazykin, and A. Coster (2017), SAMI3-RCM simulation of the 17 March 2015 geomagnetic storm, J. Geophys. Res. Space Physics, 122, 1246–1257, doi:10.1002/2016JA023341.

Iijima, T., and T. A. Potemra, (1976), Field-aligned currents in the dayside cusp observed by Triad, *J. Geophys. Res*., 8, 5971.

Kalafatoglu Eyiguler E. C., J. S. Shim, M. Kuznetsova, Z. Kaymaz, B. R. Bowman, M. Codrescu, S. Solomon, T. Fuller-Rowell, A. Ridley, P. Mehta, E. Sutton, and D. Drob, (2018), Quantifying the storm-time thermospheric neutral density variations using model and observations, *on this issue*

Maruyama, N., A. D. Richmond, T. J. Fuller-Rowell, M. V. Codrescu, S. Sazykin, F. R. Toffoletto, R. W. Spiro, and G. H. Millward (2005), Interaction between direct penetration and disturbance dynamo electric fields in the storm-time equatorial ionosphere, Geophys. Res. Lett., 32, L17105, doi:10.1029/2005GL023763.

Millward, G. H., I. C. F. Müller-Wodrag, A. D. Aylward, T. J. Fuller-Rowell, A. D. Richmond, and R. J. Moffett, (2001), An investigation into the influence of tidal forcing on F region

equatorial vertical ion drift using a global ionosphere-thermosphere model with coupled electrodynamics, *J. Geophys. Res*., 106, 24,733–24,744, doi:10.1029/2000JA000342.

Namgaladze A. A., Y. N. Korenkov, V. V. Klimenko, I. V. Karpov, F. S. Bessarab, V. A. Surotkin, T. A. Glushchenko, N. M. Naumova, (1988), Global model of the thermosphere-ionosphere-protonosphere system, *Pure Appl. Geophys*. Vol.127 (N.2/3): 219–254, doi:10.1007/BF00879812.

Namgaladze, A. A., Y. N. Korenkov, V. V. Klimenko, I. V. Karpov, V. A. Surotkin, N. M. Naumova, (1991), Numerical modeling of the thermosphere-ionosphere-protonosphere system, *J. Atmos. Terr. Phys*, Vol.53 (N.11/12):1113–1124, doi:10.1016/0021-9169(91)90060-K.

Newell, P. T., T. Sotirelis, and S. Wing (2009), Diffuse, monoenergetic, and broadband aurora: The global precipitation budget, *J. Geophys. Res*., 114, A09207, doi: 10.1029/2009JA014326.

Newell, P.T., and J.W. Gjerloev (2011), Substorm and magnetosphere characteristic scales inferred from the SuperMAG auroral electrojet indices, *J. Geophys. Res*., 116, A12232, doi:10.1029/2011JA016936.

Nishioka, M., T. Tsugawa, H. Jin, and M. Ishii (2017), A new ionospheric storm scale based on TEC and foF2 statistics, *Space Weather, 15*, doi:10.1002/ 2016SW001536.

Orús, R., M. Hernández-Pajares, J. M. Juan, J. Sanz, and M. García-Fernández, (2002), Performance of different TEC models to provide GPS ionospheric corrections, *J. Atmos. Sol. Terr. Phys*., 64, 2055–2062.

Orús, R., M. Hernández-Pajares, J. M. Juan, J. Sanz, and M. García-Fernández, (2003), Validation of the GPS TEC maps with TOPEX data, *Adv. Space Res*., 31, 621–627.

Perlongo, N. J., Ridley, A. J., Cnossen, I., and Wu, C. (2018). A year-long comparison of GPS TEC and global ionosphere-thermosphere models, *Journal of Geophysical Research: Space Physics*, *123*, 1410–1428. https://doi.org/10.1002/2017JA024411.

Picone, J. M., A. E. Hedin, D. P. Drob, and A. C. Aikin (2002), NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues, *J. Geophys. Res*., 107, doi:10.1029/2002JA009430.

Prokhorov, B. E., M. Förster, V. Lesur, A. A. Namgaladze, M. Holschneider, C. Stolle, (2018) Modeling of the Ionospheric Current System and Calculating Its Contribution to the Earth's Magnetic Field. Magnetic Fields, *in Magnetic Fields the Solar System, Astrophysics and Space Science Library*, vol. 448. Springer, Cham, doi: 10.1007/978-3-319-64292-5_10.

Rastäetter, L., et al., (2016), GEM-CEDAR Challenge: Poynting Flux at DMSP and modeled Joule Heat, *Space Weather, 14*, 113–135, doi:10.1002/2015SW001238.

Reinisch, B., and I. Galkin, (2011). Global Ionospheric Radio Observatory (GIRO). *Earth, Planets, and Space. 63*. 377-381. 10.5047/eps.2011.03.001.

Richards, P. G., Fennelly, J. A., and Torr, D. G. (1994). EUVAC: A solar EUV Flux Model for aeronomic calculations. *Journal of Geophysical Research, 99*(A5), 8981–8992. https://doi.org/10.1029/94JA00518.

Richmond, A. D., E. C. Ridley and R. G. Roble, (1992), A Thermosphere/Ionosphere General Circulation Model with coupled electrodynamics, *Geophys. Res. Lett.*, **19**, 601-604.

Rideout, W., and A. Coster, (2006), Automated GPS processing for global total electron content data, GPS Solution, doi:10.1007/s10291-006-0029-5.

Ridley, A. J., Y. Deng, and G. Toth, (2006), The global ionosphere-thermosphere model, *J. Atmos. Sol. Terr. Phys*., 68, 839-864.

Roble, R. G., E. C. Ridley, A. D. Richmond, and R. E. Dickinson, (1988), A coupled thermosphere/ionosphere general circulation model, *Geophys. Res. Lett*., 15, 1325–1328, doi:10.1029/GL015i012p01325.

Scherliess, L., *and* B. G. Fejer (1999), Radar and satellite global equatorial *F*-region vertical drift model, *J. Geophys. Res*., 104, 6829–6842, *doi:*10.1029/1999JA900025.

Scherliess, L., R. W. Schunk, J. J. Sojka, and D. C. Thompson, (2004), Development of a physics-based reduced state Kalman filter for the ionosphere, *Radio Sci.*, 39, RS1S04, doi:10.1029/2002RS002797.

Scherliess, L., R. W. Schunk, J. J. Sojka, D. C. Thompson, and L. Zhu, (2006), Utah State University Global Assimilation of Ionospheric Measurements Gauss-Markov Kalman filter model of the ionosphere: Model description and validation, *J. Geophys. Res.*, 111, A11315, doi:10.1029/2006JA011712.

Scherliess, L., I. Tsagouri, E. Yizengaw, S. Bruinsma, J. S. Shim, A. Coster, and J. M. Retterer, (2018), The iCCMC Ionosphere/Thermosphere Modeling Capabilities Assessment: Overview and Initial Results, on this issue.

Schunk, R. W., J. J. Sojka, and J. V. Eccles, (1997), Expanded capabilities for the ionospheric forecast model, Rep. AFRL-VS-HA-TR-98-0001, Air Force Res. Lab., Hanscom Air Force Base, Mass., December.

Schunk, R. W., L. Scherliess, and J. J. Sojka, (2002), Ionospheric specification and forecast modeling, *J. Spacecr. Rockets*, 39(2), 314–324.

Schunk, R. W., et al., (2004), Global Assimilation of Ionospheric Measurements (GAIM), *Radio Sci*., 39, RS1S02, doi:10.1029/2002RS002794.

Shim, J. S., et al., (2011), CEDAR Electrodynamics Thermosphere Ionosphere (ETI) Challenge for systematic assessment of ionosphere/thermosphere models: NmF2, hmF2, and vertical drift using ground-based observations, *Space Weather*, *9*, S12003, doi:10.1029/2011SW000727.

Shim, J. S., et al., (2012), CEDAR Electrodynamics Thermosphere Ionosphere (ETI) Challenge for systematic assessment of ionosphere/thermosphere models: Electron density, neutral density, NmF2, and hmF2 using space based observations, *Space Weather*, *10*, S10004, doi:10.1029/2012SW000851.

Shim, J. S., et al., (2014), Systematic Evaluation of Ionosphere/Thermosphere (IT) Models: CEDAR Electrodynamics Thermosphere Ionosphere (ETI) Challenge (2009-2010), in *Modeling the Ionosphere-Thermosphere System*, AGU Geophysical Monograph Series.

Shim, J. S., Rastätter, L., Kuznetsova, M., Bilitza, D., Codrescu, M., Coster, A. J., … Zhu, L. (2017a). CEDAR-GEM challenge for systematic assessment of Ionosphere/thermosphere models in predicting TEC during the 2006 December storm event. *Space Weather*, *15*, 1238–1256. https://doi.org/10.1002/ 2017SW001649

Shim, J. S., G. Jee, and L. Scherliess (2017b), Climatology of plasmaspheric total electron content obtained from Jason 1 satellite, J. Geophys. Res. Space Physics, 122, 1611–1623, doi:10.1002/2016JA023444.

Solomon, S. C., A. G. Burns, B. A. Emery, M. G. Mlynczak, L. Qian, W. Wang, D. R. Weimer, and M. Wiltberger (2012). Modeling studies of the impact of high-speed streams and co-rotating interaction regions on the thermosphere-ionosphere. *J. Geophys. Res.*, *117,* A00L11, doi:10.1029/2011JA017417

Solomon, S. C., L. Qian, and A. J. Mannucci, (2018). Ionospheric electron content during solar cycle 23. *J. Geophys. Res. Space Physics*, *123*, 5223, doi:10.1029/2018JA025464

Webb, P. A., M. M. Kuznetsova, M. Hesse, L. Rastaetter, and A. Chulaki, (2009), Ionosphere-thermosphere models at the Community Coordinated Modeling Center, *Radio Sci*., 44, RS0A34, doi:10.1029/2008RS004108.

Weimer, D. R., (2005), Improved ionospheric electrodynamic models and application to calculating Joule heating rates, *J. Geophys. Res.*, *110,* A05306, doi:10.1029/2004JA010884.

Yue, X., W. Wan, L. Liu, J. Liu, S. Zhang, W. S. Schreiner, B. Zhao, and L. Hu (2016), Mapping the conjugate and corotating storm-enhanced density during 17 March 2013 storm through data assimilation, *J. Geophys. Res. Space Physics*, *121*, 12,202–12,210, doi:10.1002/2016JA023038.

Zhang, S.-R., Y. Zhang, W. Wang, and O. P. Verkhoglyadova (2017), Geospace system responses to the St. Patrick's Day storms in 2013 and 2015, *J. Geophys. Res. Space Physics, 122,* 6901–6906, doi:10.1002/2017JA024232.

Zhu, L., R. W. Schunk, G. Jee, L. Scherliess, J. J. Sojka, and D. C. Thompson, (2006), Validation study of the Ionosphere Forecast Model using the TOPEX total electron content measurements, *Radio Sci., 41,* RS5S11, doi:10.1029/2005RS003336.

Figure 1

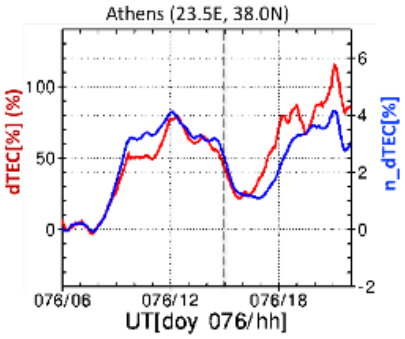


Figure 2

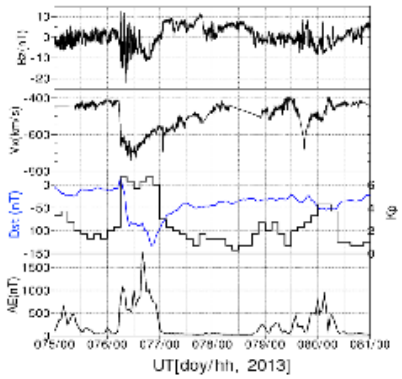


Figure 3

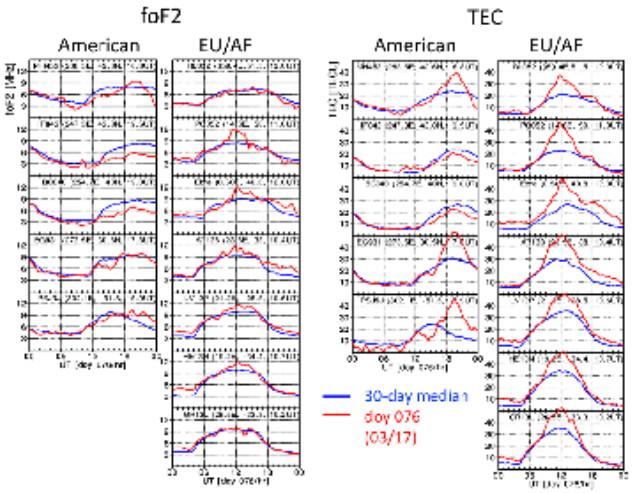


Figure 4

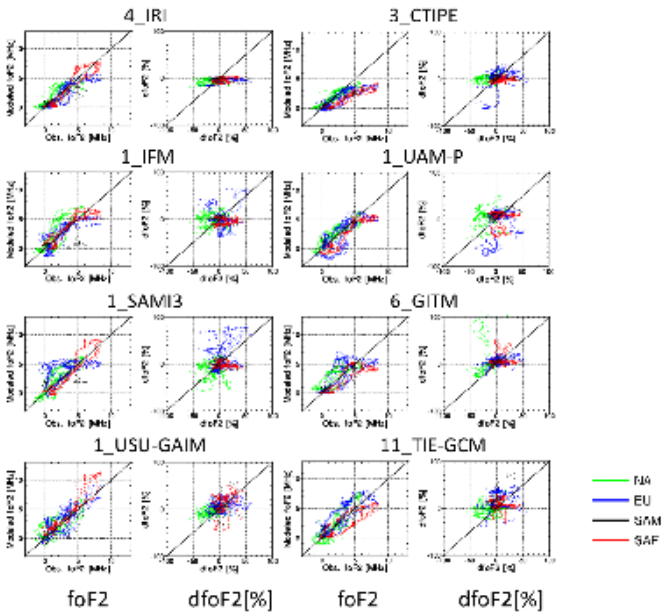


Figure 5

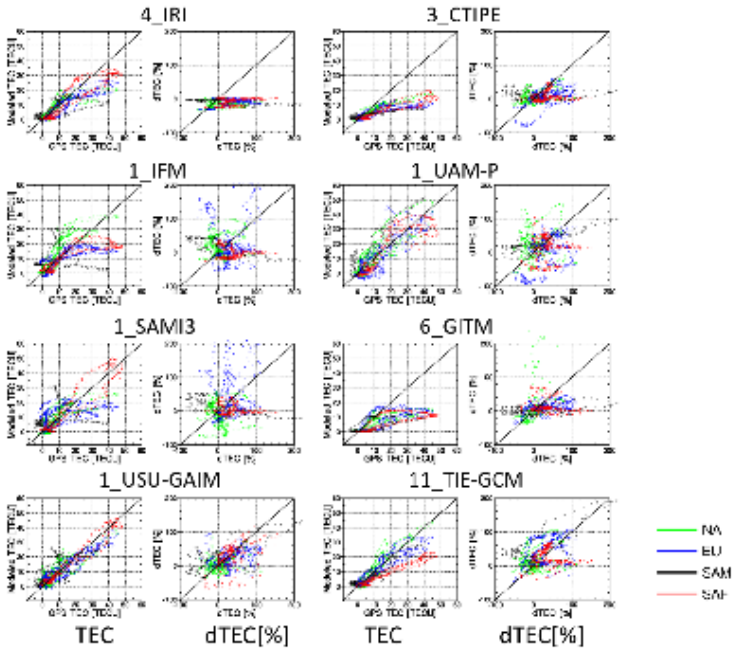


Figure 6

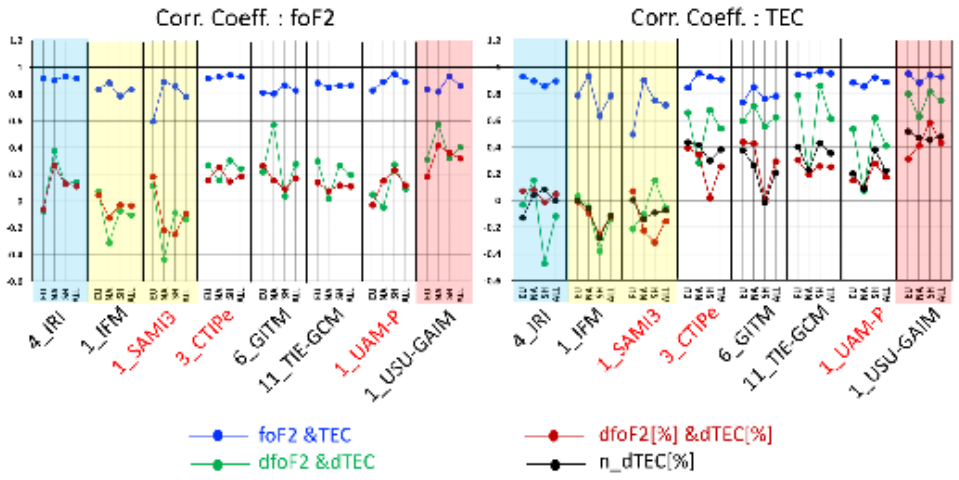
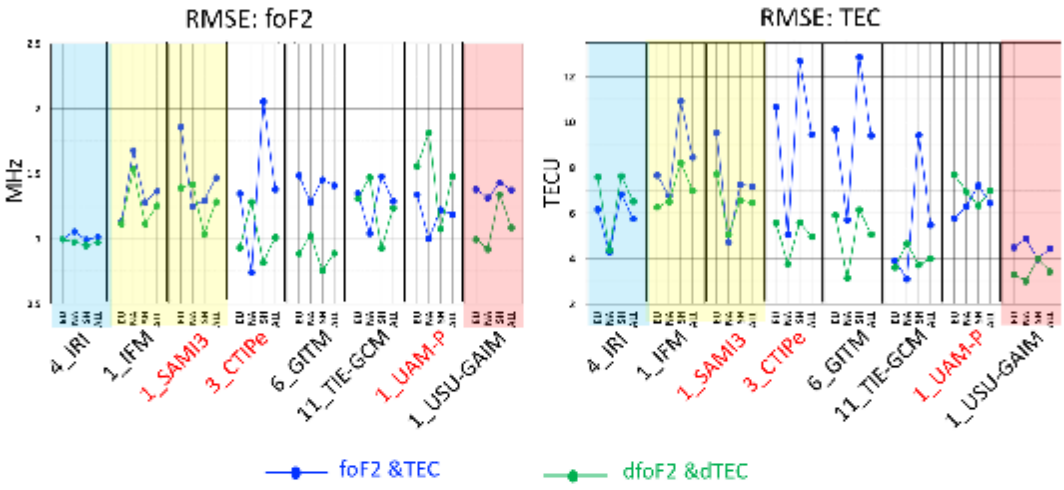


Figure 7



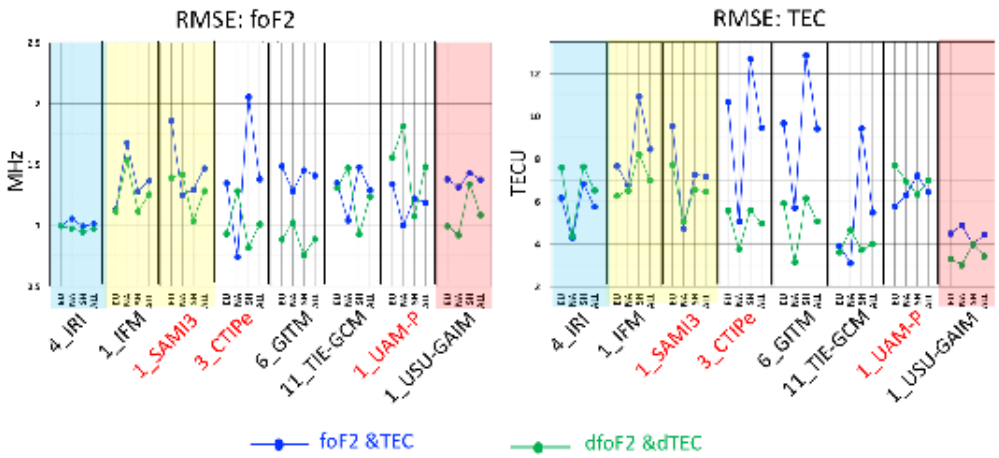


Figure 8

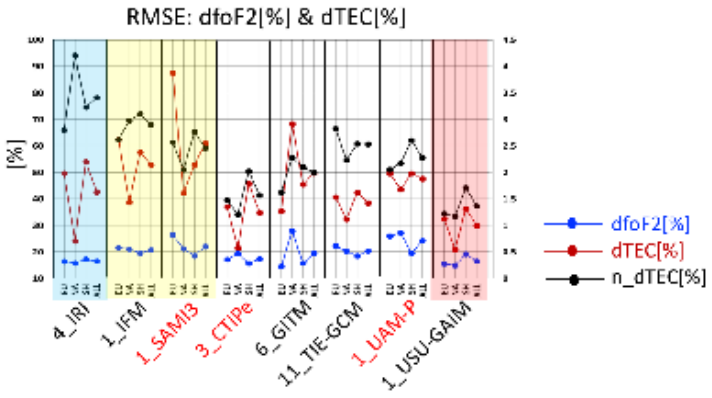


Figure 9

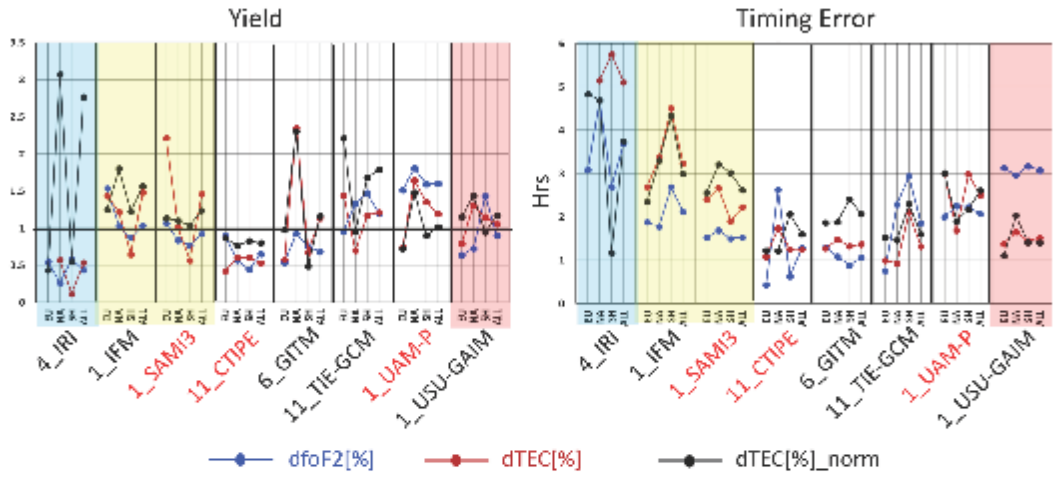


Figure 10