Shim Ja Soon (Orcid ID: 0000-0001-8943-3192)
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)
Förster Matthias (Orcid ID: 0000-0002-1104-2471)
Gardner Larry, C. (Orcid ID: 0000-0003-1127-3218)
Goncharenko Larisa, P. (Orcid ID: 0000-0001-9031-7439)
Huba Joseph (Orcid ID: 0000-0002-1930-0076)
McDonald Sarah, E (Orcid ID: 0000-0001-9235-409X)
Mannucci Anthony, J. (Orcid ID: 0000-0003-2391-8490)
Scherliess Ludger (Orcid ID: 0000-0002-7388-5255)
Schunk Robert, W. (Orcid ID: 0000-0001-8137-0617)

CEDAR-GEM Challenge for Systematic Assessment of Ionosphere/Thermosphere Models in Predicting TEC during the 2006 December Storm Event

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

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

2. NASA GSFC, Greenbelt, MD, USA,

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

4. NOAA SWPC, Boulder, CO, USA,

5. Haystack Observatory, Westford, MA, USA,

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

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

8. Utah State Univ. Logan, UT, 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: 10.1002/2017SW001649

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

10. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA,

11. Murmansk Arctic State University, Murmansk, Russia,

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

Abstract

In order to assess current modeling capability of reproducing storm impacts on TEC, we considered quantities such as TEC, TEC changes compared to quiet time values, and the maximum value of the TEC and TEC changes during a storm. We compared the quantities obtained from ionospheric models against ground-based GPS TEC measurements during the 2006 AGU storm event (14-15 Dec., 2006) in the selected eight longitude sectors. We used 15 simulations obtained from eight ionospheric models, including empirical, physics-based, coupled ionosphere-thermosphere and data assimilation models. To quantitatively evaluate performance of the models in TEC prediction during the storm, we calculated skill scores such as RMS error, Normalized RMS error (NRMSE), ratio of the modeled to observed maximum increase (Yield), and the difference between the modeled peak time and observed peak time. Furthermore, to investigate latitudinal dependence of the performance of the models, the skill scores were calculated for five latitude regions. Our study shows that RMSE of TEC and TEC changes of the model simulations range from about 3 TECU (in high

latitudes) to about 13 TECU (in low latitudes), which is larger than latitudinal average GPS TEC error of about 2 TECU. Most model simulations predict TEC better than TEC changes in terms of NRMSE and the difference in peak time, while the opposite holds true in terms of Yield. Model performance strongly depends on the quantities considered, the type of metrics used, and the latitude considered.

1. Introduction

Our daily lives are increasingly dependent on space-based technological infrastructure, such as satellites used for communications and navigations. Therefore, we are greatly affected by space weather. In mitigating any harmful effect, theory and modeling play a critical role in our quest to understand the connection between solar eruptive phenomena and their impacts in interplanetary space and in the near-Earth space environment, including the Earth's upper atmosphere. To evaluate the current state of space weather modeling capability and to track improvements of space weather models, it is important to assess model performance quantitatively. In an effort to address the needs and challenges of the quantitative assessment of modeling capabilities, the Community Coordinated Modeling Center (CCMC) initiated a series of community-wide model validation projects: SHINE,

GEM [*Pulkkinen et al*., 2010, 2011, 2013; *Rastätter et al*., 2011, 2013, 2016], CEDAR and GEM-CEDAR Modeling Challenges. The CEDAR ETI (Electrodynamics Thermosphere Ionosphere) Challenge initiated in 2009 focused on the capability of ionosphere-thermosphere (IT) models to reproduce basic IT system parameters, such as electron and neutral densities, NmF2, hmF2, and vertical drift [*Shim et al.*, 2011, 2012, 2014]. Model-data time series comparisons were performed for a set of selected events with different levels of geomagnetic activity (quiet, moderate, storm conditions). Since 2011, the follow-on CEDAR-GEM Challenge aims to quantify models' performance in predicting geomagnetic storm impacts on the ionosphere-thermosphere system parameters, including Joule Heating [*Rastätter et al*., 2016], TEC, and neutral density. TEC is one of the critical parameters in the description of ionospheric variability that affects heavily the accuracy of navigation and communication. There have been many validation studies to estimate the accuracy of TEC prediction of ionosphere models [e.g., *Orús et al*., 2002, 2003; *Zhu et al*., 2006; *Araujo-Pradere et al*., 2007; *Feltens et al*., 2011]. However, for the first time we performed metric studies for various Ionosphere/Thermosphere models, including empirical, physics-based, coupled and data assimilation models.

In this paper, we present results of the assessment of the models' performance for reproducing storm impacts on TEC during the 2006 AGU storm (14-15 Dec. 2006) in eight 5-degree wide longitude sectors. Although this study has a few shortcomings, such as neglecting differences in TEC calculations among models (e.g., upper boundary), considering only one storm event, and using only one-day TEC values as a quiet reference, the results of this systematic assessment of IT models in predicting TEC changes due to the geomagnetic

storm along with our previous findings provide a baseline for future validation studies using new models and improved models.

The GPS TEC measurements and the model simulations used for this study are briefly described in Section 2 and 3, respectively. In Section 4, the results of the analysis are presented with the details of the improvements introduced in different model simulations. Finally, the summary and conclusions are presented in Section 5.

2. GPS measurements

For this study, we selected eight 5-degree wide longitude sectors (25°-30°, 90°-95°, 140°-145°, 175°-180°, 200°-205°, 250°-255°, 285°-290°, and 345°-350°E) distributed over the globe.

First, in these eight sectors during the time interval from 13 to 15 Dec. 2006 (see Figure 1 for Dst and Kp values), we compared three GPS TEC data sets: MIT vertical TEC provided by MIT Haystack Observatory (http://cedar.openmadrigal.org/, http://cedar.openmadrigal.org/cgi-bin/gSimpleUIAccessData.py) [*Rideout and Coster*, 2006], JPL vertical TEC [*Mannucci et al.*, 1998], and IGS (International GNSS service) vertical TEC map data [*Hernandez-Pajares et al.*, 2009] . The top panel of Figure 2 shows vertical TEC from the three data sets at 30° S in the 285° - 290°E longitude sector. The similarity between JPL (denoted in blue) and IGS (green) TEC may be due in part to the fact that JPL TEC is part of the IGS product; IGS TEC is the weighted average of TEC values, which are obtained by using models that fill in the data gaps, from 4 analysis centers including CODE, ESOC, JPL, and UPC. However, the difference between the MIT (red) and JPL/IGS TEC data is noticeable. The MIT TEC values

appear to be smaller than JPL/IGS TEC in most cases (not shown here) for this study. The difference between the data sets likely resulted from different mapping functions that convert the line-of- sight TEC to vertical TEC, different pierce point altitudes, and/or different error processing schemes that account for instrument biases and multi-path corrections [*Rideout and Coster*, 2006; *Hernandez-Pajares et al.*, 2009].

However, it appears that the difference between MIT and JPL TEC data (in the middle panel of Figure 2) decreases after baseline subtraction by using their minimum value of a quiet day, which is one day prior to the storm in this study (12/13/2006, doy 347) (dTEC_m = TEC - TEC_min). In addition, the two data sets show similar differences between TEC at a given UT after the storm occurs and TEC at the same UT on the quiet day (dTEC_q = TEC - TEC_quiet) as shown in the bottom panel of Figure 2. The IGS dTECs, which have a coarser time resolution (2 hours) than the MIT and JPL dTECs (15 min), also match relatively well with the other two data sets most of the time.

To quantify storm impact and to reduce the dependence of model performance on a selection of measurement data set as a ground truth, we considered dTEC_q, dTEC_m and maximum values of the dTECs during the storm. Note that throughout this paper dTEC_m will be denoted as TEC*, which is the shifted TEC after subtracting minimum of TEC values of doy 347.

We used the MIT vertical TEC data set as a ground truth, because it is produced by using more GPS TEC data points (from more than 2,000 ground stations) than the others without any interpolation to fill in the data gaps. This guarantees that our results are purely data driven and not affected by the assumptions made in the interpolation schemes used in

generating the IGS TEC Maps. Furthermore, the MIT vertical TEC data were obtained by using a pierce point height of 350 km and a 7° elevation cutoff. For the eight longitude sectors, we used the data binned into 5° lat. × 5° lon. grid cells. The averaged error over all bins of the eight longitude sectors, for this selected time interval, is about 2 TECU. The average number of data points in a bin is about 15.

3. Models

A total of 15 model simulations were used for this study. The simulations were produced from eight models: IRI, empirical; SAMI3, USU-IFM, physics-based ionosphere; CTIPe, GITM, TIE-GCM, UAM, physics-based coupled ionosphere-thermosphere; and USU-GAIM, a physics-based ionosphere data assimilation model. The model outputs were either submitted by model developers through the CCMC online submission interface, which was developed for this and other model validation studies, or generated by the CCMC using ionosphere-thermosphere (IT) models hosted at the CCMC [*Webb et al.*, 2009]. The submissions of the model outputs are listed in Table 1. Multiple output submissions from one model using different input drivers and/or different boundary conditions are distinguished by unique model setting identifiers. For example: 1_IRI and 2_IRI were simulated with different models for topside electron density; 1_CTIPE and 2_CTIPE were obtained from different version of the model and slightly different input values; three UAM simulations (1_UAM ~ 3_UAM) used different high latitude electric potential models to model the energy input from the magnetosphere; and 1_TIE-GCM ~ 4_TIE-GCM were driven by different high latitude electric potentials, different tides at the low boundary, and/or with different resolutions (see

Table 1). The model setting identifier marked with "a" in Table 1 denotes that model results are submitted by the CCMC. Additional information on the models and the model submissions 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.

4. Results

Figure 3 shows an example of the observed and modeled vertical TEC in the 285° - 290°E longitude sector, where the GPS TEC data coverage is better than in other longitude sectors (see Figure 4). The two topmost-left panels show dTEC_m and dTEC_q of the GPS TEC data as a function of geographic latitude and time for the three days (doy 347- doy 349) of the Dec. 2006 event including one quiet day (doy 347) prior to the storm.

From top to bottom, the modeled TEC* (dTEC_m) and TEC changes (dTEC_q) obtained from empirical, physics-based ionosphere, coupled ionosphere-thermosphere, and data assimilation model simulations are shown. Figure 3 provides a comparative overview of the models' performance, showing qualitative differences in the TEC prediction among the simulations. For example, IRIs, 1_USU-IFM, 4_GITM, and 1_USU-GAIM, compared to other simulations, produce larger local daytime TEC* (dTEC_m) in southern low and middle latitudes during the initial and main phases of the storm and even during quiet time (see Figure 1). 1_USU-IFM has also larger TEC* in southern high latitudes most of the time. 2_IRI (with NeQuick for topside electron density) has larger TEC* than 1_IRI (with IRI-corr) in southern low and middle latitudes. 2_CTIPE with the improved version has larger TEC than 1_CTIPe in low and northern high latitudes. Among TIE-GCMs, 1_TIE-GCM (driven

by Heelis high latitude electric potential model) differs from the others, 2~4 TIE-GCM (driven by Weimer2005), and produces larger TEC* in high latitudes during the main phase and in low latitudes during the recovery phase. 2_UAM and 3_UAM, driven by AMIE and Weimer2005 high latitude electric potential model, respectively, produce similar TEC* and dTEC_q, while 1_UAM, driven by FAC (Field Aligned Current), has larger TEC* in northern high latitudes.

Most simulations, including 1_SAMI3_HWM93, which includes plasmasphere TEC, tend to underestimate TEC* (dTEC_m) and dTEC_q as well, although 4_GITM tends to overestimate TEC* (dTEC_m) and TEC changes from the quiet reference (dTEC_q) especially in northern low latitudes during the recovery phase. 1_USU-GAIM, obtained by assimilating GPS vertical TEC, tends to agree best with the observed GPS TEC. This is partially due to the fact that slant TEC values from about 350 of the 2000 stations, which are used to produce the MIT TEC data set, were assimilated for 1_USU-GAIM. One of possible causes of the tendency for the simulations to underestimate could be the difference between the height of the upper boundary for TEC calculation of the models (especially for the models with upper boundary of about 600 km, see Table 1) and the height of GPS satellites (20,200 km). However, the impact of differing upper boundaries (between the models and between models and observations) is likely reduced by using differential TEC (dTEC_m and dTEC_q).

For definite comparisons, we quantified the model performance using skill scores: including (1) 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), (2) RMSE normalized by the mean absolute value of the

observed TEC (NRMSE=$\frac{\text{RMSE}}{\sum |x_{obs}|/N}$), (3) the ratio of the maximum modeled TEC to that of the observed TEC (Yield = $\frac{(x_{mod})max}{(x_{obs})max}$), e.g., [maximum dTEC of the model simulations]/[maximum dTEC of GPS TEC] during two days of the storm (doy 348 ~ doy 349) , and (4) time difference between the modeled peak time and observed peak time (dt_max = t_max_model – t_max_obs) . In our Yield calculation, we focus on ionospheric positive storm effects (positive Yield) during the two days since the long-duration positive storm effects were reported for this storm event [*Pedatella et al.*, 2009].

Figure 5 shows scatter plots of RMSE and NRMSE of the simulations for all eight longitude and five latitude sectors: RMSE in the left and NRMSE in the right panel for each simulation. In each plot, x and y axes correspond to skill scores for dTEC_m and dTEC_q predictions, respectively. To investigate latitudinal dependence of model performance, the skill scores were calculated for five latitude regions: low (|lat| < 25°; red circles), northern middle (25° < lat < 50°; green squares), southern middle (green triangles), northern high (lat > 50°; blue squares), and southern high (blue triangles) latitudes. It should be noted that data coverage varies with latitudes and longitude sectors (see Figure 4). In southern middle and high latitudes, there are fewer data points than in other latitude regions, and the data are not evenly distributed over the eight longitude sectors. Therefore, comparing the performance in the different latitude regions requires caution.

We found that, using only the error values below 95th percentile for all cases, RMSE is reduced by a maximum of 40% and 10 % on the average percentage. This indicates that RMSE is not severely affected by a few upper outliers in the distribution.

Most of the model simulations show similar RMSE for the dTEC_m and dTEC_ q predictions (close to the dotted line of slope 1). All simulations have the largest RMSE in the low latitude region (especially in 175°-180°E, 200°-205°E longitude sectors) where TEC values are larger than in higher latitudes. However, NRMSE for the dTEC_q is larger than NRMSE for dTEC_m for most cases especially in low latitudes. Note that scale of y-axis of NRMSE plots for UAMs and 4_GITM are different than that of the plots for other simulations. NRMSE for dTEC_m (TEC) prediction of most simulations ranges between about 0.5 and 1.5, while NRMSE for dTEC_q (TEC changes) prediction ranges between about 1 and 2.5~3 (between 1 and 4~5 for UAMs and 4_GITM). This indicates that in terms of NRMSE the models predict TEC* better than TEC disturbances.

To investigate how the uneven data coverage mentioned above affects globally averaged model performance, we rank the model simulations using four different single values of RMSE in Table 2.1 ~ 2.2 and NRMSE in Table 2.3 ~ 2.4. The four scores include two latitudinal average scores and two global scores. To obtain the two latitudinal average scores, we first calculated longitudinal means of RMSE and NRMSE, which are averages of the values shown in the same color and figures in Figure 5, in each latitude sector. The longitudinal means were averaged again over (1) only three (low and middle) latitude sectors (3^{rd} column) and (2) all five latitude sectors (5^{th} column). The two global scores include (1) one global value (of RMSE and NRMSE) obtained by using errors of the simulations for only low and middle latitudes of all longitude sectors (7^{th} column) and (2) the other global value for all regions (9^{th} column). For both dTEC_m and dTEC_q predictions, all simulations produce the smallest RMSE average over all latitude sectors and the largest global RMSE for

low and middle latitudes. The global RMSE for low and middle latitudes are larger than that for all latitude regions. On the contrary, for dTEC_m predictions, NRMSE for low and middle latitudes are larger than that for all latitude regions. For dTEC_q predictions, 2_TIE-GCM (with Weimer2005) and 2_UAM (with AMIE) show reduced NRMSE including high latitudes, while CTIPEs (with Weimer2005) and 3_UAM (with AMIE) show the opposite and the others hardly show any changes. Ranking of the models appears not to depend heavily on the selection of any particular skill score among the four values.

Figure 6 shows the longitudinal average RMSE (upper panel) and NRMSE (lower panel), for dTEC_m (left) and dTEC_q (right) in each of the five latitude sectors. Four groups of lines with figures correspond to empirical, physics-based ionosphere, coupled ionosphere-thermosphere, and data assimilation model simulations from the left to the right. Note that 1_SAMI3_HWM93 data at high latitudes were excluded due to lack of reliability, since SAMI3 does not include high latitude driving forces (e.g., the auroral precipitation and the convection electric field pattern). Therefore, in Figure 6, the performance of the simulations is ranked based on the average RMSE and NRMSE over three latitude (low and middle) sectors (see Table 2.1~ 2.4). The best performing simulation is located in the extreme left in each group.

Simulations using the empirical model IRI, 1_IRI and 2_IRI, perform similar to each other to predict TEC* (dTEC_m) and dTEC_q for most cases, especially in middle latitudes. Note that different scales are used for dTEC_m and dTEC_q.

Two physics-based ionospheric model simulations, 1_USU-IFM and 1_SAMI3_HWM93 also show similar scores when considering scores available from both

simulations (e.g., low and middle latitudes). 1_USU-IFM performs slightly better than 1_SAMI3_HWM93 in low and northern middle latitudes for dTEC_m, while the opposite is true in northern middle latitudes for dTEC_q. However, the differences in averaged RMSE and NRMSE of the two simulations for low and middle latitudes do not exceed 0.5 TECU and 0.04, respectively (see Table 2.1~ 2.4). Both models use the same empirical models for the thermosphere wind and compositions, and low latitude electric fields. After including scores for high latitudes of 1_USU-IFM, the differences in averaged RMSE become larger (~1.4 TECU). Although the averaged NRMSE differences remain similar (~0.03), the rank of the two simulations for dTEC_m is reversed; 1_SAMI3_HWM93 shows slightly better performance than 1_USU-IFM.

Among ten physics-based coupled ionosphere-thermosphere simulations, 2_CTIPE and three TIE-GCM runs (2_TIE-GCM through 4_TIE-GCM, referred to as 2~4_TIEGCMs) perform better than the other six simulations. 2_CTIPE tends to perform better than 2~4_TIE-GCMs for the dTEC_q prediction. The opposite holds true for the dTEC_m prediction. The four TIE-GCM simulations show similar RMSE and NRMSE in southern middle latitudes. However, in the other latitude sectors, the dTECs prediction of 1_TIE-GCM, which is driven by the Heelis electric potential with constant critical cross-over latitudes, is less satisfactory than that of the other three TIE-GCM simulations, which are driven by the Weimer-2005 electric potential with dynamic critical cross-over latitudes. This is caused mainly by a poorer performance in high latitudes. Among the three simulations, 2~4_TIE-GCMs, which show similar accuracy predicting dTECs, 4_ TIE-GCM (driven by SABER/TIDI lower boundary conditions) does better reproducing low latitude TEC. With regard to the two CTIPE

simulations, 2_CTIPE performs better than 1_CTIPE for all cases, although 1_CTIPE shows better scores for dTEC_q prediction at low latitudes. Differences between the two CTIPE simulations are caused by different input data sets for the solar wind parameters and different values for the solar flux (see Table 1). 2_CTIPE shows better performance than 1_CTIPE, especially at high latitudes. The model version used in 1_CTIPE simulation is an older version of 2_CTIPE, and required the use of trimmed IMF (Interplanetary Magnetic Field) values. For example, the magnitude of IMF has a saturated value of about 12 nT when Bz is less than -10 nT. 4_GITM shows similar performance in predicting middle and high latitude TEC, but it performs worse for low latitude TEC prediction than the other simulations. Three UAM simulations driven by different high-latitude electric potential models also show similar skill scores. However, 1 and 3_UAM perform better than 2_UAM for most cases, 1_UAM and 3_UAM appear to be better for low-latitude dTECs and high-latitude dTEC_q prediction, respectively.

The only simulation using a data assimilation model, 1_USU-GAIM outperforms other models for all cases excluding high-latitude TEC prediction. The average RMSE of 1_USU-GAIM for TEC* (dTEC_m) prediction is about 4 TECU, which is twice the average GPS TEC error of 2 TECU.

The RMSE of TEC* (dTEC_m) for all simulations and all cases ranges from about 3 TECU in northern high latitudes to about 13 TECU in low latitudes. Even the lowest RMSE of the simulations is larger than average GPS TEC error of about 2 TECU.

In Figure 7, we present Yield (ratio of modeled to observed maximum difference) and differences in time (dt_max = t_max_model – t_max_obs), considering only the maximum of

dTECs that occurs during the storm time, between the simulations and observation. Yield and dt_max are shown in left and right panels, respectively, in each simulation group. The x and y axes corresponds to the values of dTEC_m and dTEC_q, respectively. Colors and figures indicate different latitudes, as the same in Figure 5.

The largest maximum increase in GPS TEC due to the storm occurs in low latitude regions in the 200° - 205°E longitude sector during the main phase; dTEC_q reaches up to about 70 TECU (not shown here), which is about five times the quiet time TEC. The largest maximum percentage increase in TEC occurs in the 175° - 180°E longitude sector in low latitude regions during the main phase; dTEC_q of 50 TECU is about seven times larger than the quiet time TEC. These facts possibly explain why most model simulations have the largest RMSE in the low latitude region in 175°-180°E and 200°-205°E longitude sectors due to the tendency of the models to underestimate TEC (as described above), while RMSEs in the other longitude sectors are relatively similar to each other.

Most models have the tendency to underestimate the peak of dTECs. However, 4_GITM and TIE-GCMs appear to overestimate dTECs' peak in more cases in comparison to the other model simulations. Two simulations, 1_IRI and 2_IRI, of the empirical model, which represents average ionospheric conditions, produce the smallest Yield of dTEC_q (< 0.5 for most cases), even though the performance of the model in terms of the RMSE and/or NRMSE is comparable to that of the physics-based models.

1_USU-GAIM has a smaller spread of data points of time difference, dt_max, compared to other models. 2_IRI and 1_USU-IFM have smaller spread of data points of time difference than 1_IRI and 1_SAMI3_HWM93, respectively. TIE-GCM runs have more data points

closer to the line dt_max=1 than other coupled physics-based model simulations. 1_IRI, 2_UAM, and CTIPE runs tend to reach the maximum values of both dTEC_m and dTEC_q more often before the observed maximum occurs.

Table 3.1 and 3.2 show global ranking of the simulations in terms of two averages (as the same as in Table 2.1~2.4) of Yield and dt_max: over (1) low and middle latitudes and (2) all five latitudes. With high latitudes included, the following results are observed: 1_IRI and 2_IRI produce worse Yield and dt_max for dTEC_m prediction, but they produce better or similar scores for dTEC_q prediction; I_USU-IFM shows better Yield and dt_max for both dTECs' predictions; 4_GITM does not much change in the scores but produces better Yield for dTEC_q and dt_max for dTEC_m; 1~3_UAM produce worse scores for most cases, however 2_UAM and 3_UAM show better dt_max for dTEC_q; 2_CTIPE has better Yield and dt_max for dTEC_q, while 1_CTIPE has worse scores for both dTECs; four TIE-GCMs produce better Yield and dt_max, especially dt_max for dTEC_m, it caused noticeable ranking changes for dt_max prediction. However, with respect to Yield and dt_max, the global ranking of the simulations seems not to be affected strongly by including high latitudes in other cases.

In Figure 8, the upper panels show the longitudinal average Yield (modeled dTEC_max/observed dTEC_max) of dTEC_m (left) and dTEC_q (right) sectors, and the bottom panels show the longitudinal average time differences (average of |dt_max|) in the five latitudes. The longitudinal average scores were obtained from the values shown in the same color and figures in Figure 7. The results of the simulations are grouped according to the type of the models and the best simulation is located at the leftmost in each group as the

same as in Figure 6. The performance of the simulations is ranked based on the average scores over three latitude (low and middle) sectors (see Table 3.1 and 3.2). All simulations tend to underestimate dTECs for all cases with the exceptions of 4_GITM and 1_TIE-GCM, which are ranked first and second place in their group based on the average Yield for dTEC_q over low and middle latitudes. 4_GITM produces best Yield among all simulations, while 1_USU-GAIM predicts best the time at which the maximum dTECs occur for most cases (about 5 hour difference on the average). Most models tend to produce larger dt_max for low latitudes (red circles).

The two IRI runs produce similar scores of Yield and dt_max, but 2_IRI slightly outperforms 1_IRI for all cases. The IRI runs perform worse in predicting Yield of dTEC_q than that of TEC (dTEC_m) with one exception (2_IRI in southern high latitudes), although there is relatively small difference in average dt_max between dTEC_m and dTEC_q predictions of the two simulations (see Table 3.1~3.2).

1_USU-IFM shows better Yield and dt_max of dTEC_m than 1_SAMI3_HWM93, which shows better Yield of dTEC_q. 1_SAMI3_HWM93 appears to predict better Yield for low latitudes than 1_USU-IFM, but the opposite holds true for dt_max. The two simulations produce similar Yield and dt_max in northern middle latitudes.

The physics-based coupled models make better prediction of peak values of dTECs in middle and high latitudes of the northern hemisphere than in the southern hemisphere. They perform better in predicting Yield of dTEC_q than that of dTEC_m for most cases, which is opposite to the empirical and data assimilation model simulations. It indicates that the physics-based coupled model simulations tend to predict TEC increases (dTEC_q) somewhat

better than TEC itself (dTEC_m). However, in most of the simulations, dt_max values of dTEC_q are larger than or similar to those of dTEC_m.

The average values of Yield of 4_GITM and 1_TIE-GCM are closer to 1, with a larger spread of Yield around the mean compared to other simulations. Four TIE-GCM runs, 1-4_TIE-GCM, rank higher for predicting both Yield and dt_max, while 4_GITM ranks first for Yield but ranks lower for predicting dt_max.

The four TIE-GCM runs show similar performance especially in low and middle latitudes, and among them, 2~4_TIE-GCM driven by Weimer2005 show similar performance for most cases at high latitudes. 1_TIE-GCM driven by Heelis appears to overestimate Yield of dTEC_q in both southern and northern high latitudes, whereas the other three TIE-GCMs tend to slightly underestimate Yield in southern high latitudes and produce almost perfect Yield in northern high latitudes. Compared to other physics-based coupled model simulations, 2~4_TIE-GCM capture better the peak time of dTECs in northern middle (about 1 hour for dTEC_m and 3 hours for dTEC_q).

The three UAMs and two CTIPEs rank lower than others for most cases, however, 2_CTIPE and 1_UAM (driven by FAC) rank higher for dt_max of dTEC_m and dTEC_q, respectively. 2_CTIPE produces better ratios and time differences for all cases than 1_CTIPe. 1_UAM, which performs better than 2-3_UAM for most cases, hardly shows differences in dt_max between dTECs, while 2-3_UAM produce noticeable differences in dt_max between dTECs in most latitude regions.

The data assimilation model, I_USU-GAIM produces Yield of dTECs comparable to those of the highest-ranked physics-based coupled model simulations, and predicts best the

time at which the maximum dTECs occur in low and middle latitudes (about 0.4 ~ 3-4 hours for dTEC_m and 1~ 6 hours for dTEC_q).

5. Summary and Conclusion

We quantitatively assessed the performances of ionospheric models in predicting geomagnetic storm impact on TEC in the selected eight longitude sectors during 2006 Dec. storm event (doy 347 ~ 350). For this study, the modeled values obtained from 15 simulations using eight ionospheric models were compared with the observed ground-based GPS TEC values. We considered TEC* (dTEC_m) during the storm, TEC changes (dTEC_q) due to the storm, and maximum values of the dTECs.

The performance of the models was quantified using skill scores such as RMS error (RMSE), Normalized RMS error (NRMSE), the ratio of maximum value of the models to the observation (Yield), and the differences (dt_max) in peak time (dt_max = t_max_model – t_max_obs) between the simulations and the observations. Using Yield, we evaluated the capability of the models to produce peak values of TEC increase during the storm. The skill scores were calculated for five latitude regions to investigate the latitudinal dependence of the performance of the models. The simulations were grouped based on the type of model and ranked first based on the average values of the skill scores over all eight longitude and three latitude sectors (low and middle latitudes). This was done because (1) 1_SAMI3_HWM93 data at high latitudes were excluded due to lack of reliability and (2) the relatively low data coverage of GPS TEC in high latitudes may have an effect on the overall scores. Therefore, we also ranked the simulations according to average skill scores over all longitude and

latitude regions, and global RMSE and NRMSE for the error values over (1) only low and middle latitudes and (2) all latitudes. Ranking of the models in terms of RMSE and NRMSE appears not to depend heavily on the selection of any particular skill score among the four values. With respect to Yield and dt_max, the global ranking of the simulations also does not seem to be strongly affected by including high latitudes in most cases, although there are noticeable ranking changes for dt_max of dTEC_m among four TIE-GCMs with better scores, and 2_UAM and 1_CTIPE with worse scores in high latitudes.

In agreement with our previous study [*Shim et al.*, 2012, 2014], we found that model performance depends on the type of metrics and latitude as well. For instance, although our results based on the average skill scores show that the performance of the data assimilation model, 1_USU-GAIM, is superior to the other models' performance in most cases, especially in low- and middle-latitude TEC predictions, other physics-based model simulations (e.g., TIE-GCMs) are better than or comparable to 1_USU-GAIM in southern high-latitude dTECs prediction with respect to RMSE and NRMSE. In in northern high-latitudes, TIEGCMs, except for 1_TIE-GCM, show slightly better Yield than 1_USU-GAIM due to the fact that 1_USU-GAIM used for the study only assimilates GPS observations within ±60° geographic latitudes. For low latitude TEC simulations, 4_GITM and 1_USU_GAIM produce the best Yield and peak time differences (dt_max), respectively.

In terms of RMS and NRMS errors, the empirical model simulations, 1_IRI and 2_IRI, show comparable performance to the physics-based models, which are ranked higher in their group. However, the two IRI simulations are inferior to other physics-based model

simulations based on Yield and peak time differences since they represent the average ionospheric conditions rather than storm-time perturbations.

Overall, for TEC* (dTEC_m) and TEC disturbance (dTEC_q) prediction, the simulations have similar RMSE, which is larger than GPS TEC error for all cases (the average error over all bins of the eight longitude sectors is about 2 TECU). However, model performance also depends on the considered quantities. NRMSE and the difference in peak time are smaller for dTEC_m than for dTEC_q prediction for most model simulations, while the opposite holds true for Yield.

Our results for the comparisons among the same types of models suggest that two physics-based ionospheric models, 1_ SAMI3_HWM93 and 1_USU-IFM show similar performance in general as shown in our previous validation studies [*Shim et al*., 2011, 2012, 2014]. These two simulations were driven by the same empirical models for thermospheric inputs and low latitude electric fields, but for 1_SAMI3_HWM93, an updated version of the models for neutral composition and winds were used. 1_SAMI3_HWM93 has better Yield and worse dt_max of dTEC_q than 1_USU-IFM, and major differences occur in the low latitude regions.

Among the ten physics-based coupled ionosphere-thermosphere model simulations, in terms of RMSE and NRMSE, 2~4_TIE-GCM and 2_CTIPE appear to perform similar to each other and better than the other six simulations for most cases. However, 4_GITM and 1_TIE-GCM, which appear to overestimate dTECs in northern high latitudes, take the first and second places among the 10 simulations based on the average Yield over all five latitude

regions. 1_UAM driven by FAC, which is the only simulation that includes penetration electric fields physically, ranks higher for dt_max of dTEC_q.

It is worth pointing out the improvements of model performance caused by enhanced and/or more complex input drivers and/or more accurate input data. 1_IRI (IRI-2007 with NeQuick model for topside Ne) and 2_IRI (IRI-2012 with IRI-corr model for topside Ne) show similar performance, although 2_IRI is slightly better than 1_IRI for all cases except for low latitude dTEC_m.

From the comparison of 1_CTIPE and 2_CTIPE, it is found that 2_CTIPE has better scores for most cases, but 1_CTIPE is slightly better for low-latitude dTEC_q predictions based on all four skill scores. In most cases 1_CTIPE performs worse than 2_CTIPE in northern middle and high latitudes. The better performance of 2_CTIPE appears to be mainly caused by non-trimmed IMF data from the ACE satellite. The improved version of CTIPe used for 2_CTIPe simulation is more robust during geomagnetic storm conditions, and has been further tuned to better reflect the energy input to the system. The ratio between the Joule Heating contribution of the main E-field and its variability has been changed based on total mass density comparisons between CTIPe and CHAMP and GRACE measurements [*Fedrizzi et al*., 2012].

Only 1_TIE-GCM, among four TIE-GCM simulations, was obtained from TIE-GCM1.93 driven by the Heelis high latitude electric potential model, and the other three simulations, 2~4_TIE-GCM, used TIE-GCM1.94 with Weimer-2005. Although the skill scores for the four TIE-GCM simulations are more or less similar for all cases, 2~4_TIE-GCM (with Weimer-2005) performs better than 1_TIE-GCM (with Heelis) especially in

northern high latitudes. 4_TIE-GCM (TIE-GCM1.94 with Weimer-2005 and SABER/TIDI lower boundary conditions in double resolution) and 1_TIE-GCM (TIE-GCM1.93 driven by Heelis) performs the best and worst, respectively, in most cases. 2_TIE-GCM (with default resolution) and 3_TIE-GCM (with double resolution) displays small differences in their performance. The improvement of 2_TIE-GCM compared to 1_TIE-GCM in predicting ionospheric parameters during the strong storm was also shown in *Shim et al*. [2011, 2012, 2014].

The three UAM simulations obtained from UAM with different high latitude ionospheric drivers, show no significant differences in their performance in predicting TEC during the storm in general. However, 1_UAM driven by FAC (with double resolution in longitude) tends to perform somewhat better than the other two simulations in low latitude TEC predictions for all cases except for the Yield of dTEC_q for which 2_UAM (with AMIE electric potentials) performs better. Also, 1_UAM performs better in northern high latitudes all cases except for NRMSE of dTEC_q for which 3_UAM (with Weimer-2005) performs better.

Along with our earlier results [*Shim et al.*, 2011, 2012, 2014], our findings of this systematic assessment of TEC change prediction of IT models during the geomagnetic storm provide a baseline for future validation studies using new models and improved models, although this study considered only the eight 5-degree wide longitude sectors and has the shortcomings described in the introduction. In the near future, we will extend our study to overcome the shortcomings. For example, we will perform regional TEC validation (e.g., North American Sector and European sector) for more storm events with longer quiet time

references (e.g., the median for the 30 days prior to storms and the average of the five quietest days within the 30 days). Furthermore, we will investigate the local time dependence of storm impacts on TEC and how well the local time dependence can be predicted by the ionosphere models. We will also study in more detail the effects of high-latitude drivers (e.g., electric potential and auroral particle precipitations) on TEC changes during geomagnetic storm events. For future studies, we will use additional TEC data sets (e.g., Low Earth orbit satellite-based TEC) to overcome the limitation of data coverage of ground-based GPS TEC over high latitude regions (> 60° ~70°) and over the oceans.

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.

Acknowledgement

The vertical TEC data were provided by MIT Haystack Observatory (http://www.openmadrigal.org). This work is supported by grants from the National Science Foundation (NSF) Space Weather Program. Portions of this research were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. This model validation study is supported by the Community Coordinated Modeling Center (CCMC) at the Goddard Space Flight Center.

References

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., (1990), International Reference Ionosphere 1990, 155 pages, National Space Science Data Center, NSSDC/WDC-A-R&S 90-22, Greenbelt, Maryland, November 1990.

Bilitza D., (2001), International Reference Ionosphere 2000. *Radio Sci*., 2001, 36(2): 261-275.

Bilitza, D., and B. W. Reinisch (2008), International reference ionosphere 2007: Improvements and new parameters, *Adv. Space. Res.*, 42, 599–609.

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.

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.

Coisson, P., Radicella, S.M., Leitinger, R., Nava, B. Topside electron density in IRI and NeQuick: features and limitations. *Adv. Space Res.* 37 (5), 934–937, 2006.

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.

Fedrizzi, M., T. J. Fuller-Rowell, and M. V. Codrescu (2012), Global Joule heating index derived from thermospheric density physics-based modeling and observations, *Space Weather,* 10, S03001, doi:10.1029/2011SW000724.

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.

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. (1987), MSIS-86 thermospheric model. *J Geophys. Res.*, 92(A5): 4649–4662.

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.

Hernandez-Pajares, M., J. M. Juan, J. Sanz, R. Orus, A. Garcia-Rigo, J. Feltens, A. Komjathy, S. Shaer, (2009), The IGS VTEC maps: a reliable source of ionospheric information since 1998, *J. Geod*, *83*, 263-272, doi:10.1007/s00190-008-0266-1.

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.

Mannucci, A. J., B. D. Wilson, D. N. Yuan, C. H. Ho, U. J. Lindqwister, and T. F. Runge (1998), A global mapping technique for GPS-derived ionospheric total electron content measurements, *Radio Sci*., 33(3), 565–582, doi:10.1029/97RS02707.

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.

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.

Papitashvili, V. O, F. Christiansen, and T. Neubert, (2002), A new model of field-aligned currents derived from high-precision satellite magnetic field data, *Geophys Res Lett*, Vol. 29 (N.14, doi:10.1029/2001GL014207.

Pedatella, N. M., J. Lei, K. M. Larson, and J. M. Forbes, (2009), Observations of the ionospheric response to the 15 December 2006 geomagnetic storm: Long-duration positive storm effect, *J. Geophys. Res*., 114, A12313, doi:10.1029/2009JA014568.

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.

Pulkkinen, A., L. Rastätter, M. Kuznetsova, M. Hesse, A. Ridley, J. Raeder, H. J. Singer, and A. Chulaki, (2010), Systematic evaluation of ground and geostationary magnetic field predictions generated by global magnetohydrodynamic models, *J. Geophys. Res*., 115, A03206, doi:10.1029/2009JA014537.

Pulkkinen, A., et al. (2011), Geospace Environment Modeling 2008–2009 Challenge: Ground magnetic field perturbations, *Space Weather*, 9, S02004, doi:10.1029/2010SW000600.

Pulkkinen, A., et al., (2013), Community-wide validation of geospace model ground magnetic field perturbation predictions to support model transition to operations, *Space Weather*, *11*, 369–385, doi:10.1002/swe.20056

Rastätter, L., M. M. Kuznetsova, A. Vapirev, A. Ridley, M. Wiltberger, A. Pulkkinen, M. Hesse, and H. J. Singer, (2011), Geospace Environment Modeling 2008–2009 Challenge: Geosynchronous magnetic field, *Space Weather*, 9, S04005, doi:10.1029/2010SW000617.

Rastätter, L., et al., (2013), Geospace environment modeling 2008–2009 challenge: Dst index, *Space Weather, 11*, 187–205, doi:10.1002/swe.20036.

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.

Richmond, A. D., (1992), Assimilative mapping of ionospheric electrodynamics, *Adv. Space Res*., Vol 12, No 6, 59-68.

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., and E. C. Ridley, (1987), An auroral model for the NCAR thermospheric general circulation model (TGCM), *Annales Geophys.*, 5A, 369-382.

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.

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.

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. (1995), Models of the high-latitude electric potential derived with a least error fit of spherical harmonic coefficients, *J. Geophys*. Res., 100, 19,595–19,608.

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

Wu, Q., D. A. Ortland, B. T. Foster, and R. G. Roble, (2012), Simulation of nonmigrating tide influences on the thermosphere and ionosphere with a TIMED data driven TIEGCM, *J. Atmos. Sol. Terr. Phys*., **90**, 61-67, doi:10.1016/j.jastp.2012.02.009.

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.

Table 1. Models submitted for this study

Model Setting ID	Model Version	Drivers			Upper boundary for TEC calculation/ Resolution
		Input data	Models used for thermosphere, tides from lower boundary, and high latitude electrodynamics		
Empirical Model					
1_IRI [a]	IRI-2007 [*Bilitza*, 1990, 2001; *Bilitza and Reinisch*, 2008]	F10.7, R12, IG12, and ap	NeQuick model for topside Ne [*Coisson et al.*, 2006], MSIS-86 neutral composition [*Hedin*, 1991]		~2,000 km
2_IRI [a]	IRI-2012 [*Bilitza et al*., 2014]		IRI-corr model for topside Ne, NRLMSISE00 neutral composition [*Picone et al*., 2002]		
Physics Based Ionosphere Model					
1_SAMI3_HWM93[a]	SAMI3 [*Huba et al.*, 2000, 2008]	F10.7 and ap	NRLMSISE00 neutral composition, HWM93 wind [*Hedin et al*., 1991; *Drob et al*., 2008], E×B drift [*Scherliess and Fejer*, 1999]		~20,000 km, 1.5° lat × 4° lon
1_USU-IFM[a]	IFM [*Schunk et al*., 1997, 2002]	F10.7, daily Ap, and Kp	MSIS-90 neutral composition [*Hedin*, 1991], HWM neutral wind [*Hedin et al*., 1991], Scherliess and Fejer E×B drift, high latitude electric fields [*Heppner and Maynard*, 1995], Hardy aurora precipitation [*Hardy et al.*, 1985]		~1,600 km, 3° lat × 7.5° lon
Physics-based Coupled Ionosphere-Thermosphere Model					
			Tides	High Latitude Electrodynamics	
1_CTIPE[a]	CTIPe_2.0 [*Codrescu et al*., 2000; *Millward et al*., 2001]	F10.7, ACE IMF data and solar wind speed and density, NOAA POES Hemispheric Power data	migrating diurnal and semidiurnal tides	Weimer-2005 high latitude electric potential [*Weimer*, 2005], Fuller-Rowell and Evans aurora precipitation [1987]	~2,000 km, 2° lat × 18° lon
2_CTIPE	CTIPe_2.1 run at NOAA/SWPC with Weimer-2005	½*(F10.7_81ave_current_day + F10.7_previous_day), ACE IMF data and WIND solar wind speed and density, NOAA POES Hemispheric Power data			
4_GITM[a]	GITM2.0 [*Ridley et al.*, 2006]	F10.7, ACE IMF data and solar wind speed and density, NOAA POES Hemispheric Power data	GSWM[Hagan et al., 1999] migrating diurnal and semidiurnal tides	Weimer-2005 high latitude electric potential, Fuller-Rowell and Evans aurora precipitation [1987]	~600 km, 2.5° lat × 5° lon

1_TIE-GCM[a]	TIE-GCM1.93 [*Roble et al.*, 1988; *Richmond et al*., 1992]	F10.7, Kp	GSWM migrating diurnal and semidiurnal tides	Heelis high latitude electric potential model [*Heelis*, 1982], Roble and Ridley aurora precipitation [1987]	~600 km, 5° lat × 5° lon
2_TIE-GCM	TIE-GCM1.94	F10.7, Kp, OMNI IMF data and solar wind speed and density		Weimer-2005 high latitude electric potential Roble and Ridley aurora precipitation [1987]	
3_TIE-GCM	TIE-GCM1.94 with double resolution				~600 km, 2.5° lat × 2.5° lon
4_TIE-GCM	TIE-GCM1.94 with double resolution		SABER/TIDI migrating diurnal and semidiurnal tides		
1_UAM	Upper Atmosphere Model (UAM) [*Namgaladze et al*., 1988, 1991]	F10.7, Ap, Kp, AE, IMF data and solar wind speed and density	Equal low boundary conditions with no tides	FACs using Papitashvili et al. model [2002] and the Hardy model	~2,000 km 2° lat (irregular) × 7.5° lon
2_UAM	UAM with AMIE			AMIE electric potential model [*Richmond,*1992], Hardy auroral precipitation	~2,000 km 2° lat (irregular) × 15° lon
3_UAM	UAM with Weimer2005			Weimer-2005 high latitude electric potential, Hardy auroral precipitation	
Physics-based Data Assimilation Model					
1_USU-GAIM[a]	USU-GAIM2.3 [*Schunk et al*., 2004; *Scherliess et al*., 2004, 2006]	F10.7, daily Ap, and Kp, GPS TEC observations from more than 350 ground stations (-60° < lat < 60°)	The IFM background physics-based ionosphere model, the same models used for IFM		~1,400 km, 3~5° lat (irregular) × 15° lon

[a]The model results are submitted by the CCMC using the models hosted at the CCMC. Different mode setups are referred as different model setting identification number.

Table 2-1 Model ranking w.r.t. RMSE of dTEC_m

TEC* (dTEC_m)	average RMSE (TECU)				global RMSE (TECU)			
	low and middle		all latitudes		low and middle		all latitudes	
	ranking	RMSE	ranking	RMSE	ranking	RMSE	ranking	RMSE
Empirical								
1_IRI	1	6.0	1	5.4	1	6.3	1	6.1
2_IRI	2	6.2	2	5.5	2	6.5	2	6.3
Physics-based Ionosphere Model								
1_USU-IFM	1	6.7		5.7	1	6.5		6.2
1_SAMI3	2	7.0			2	7.1		
Physics-based Coupled Ionosphere-Thermosphere Model								
4_TIE-GCM	1	6.3	1	5.4	1	6.6	1	6.1
3_TIE-GCM	1	6.3	1	5.4	2	6.8	2	6.3
2_TIE-GCM	3	6.4	1	5.4	4	6.9	3	6.4
2_CTIPE	3	6.4	4	5.5	2	6.8	3	6.4
1_TIE-GCM	5	6.6	5	5.6	4	6.9	3	6.4
1_UAM	6	7.0	6	5.9	6	7.4	6	6.9
3_UAM	7	7.5	7	6.3	8	7.9	7	7.3
2_UAM	8	7.6	8	6.4	9	8.0	8	7.4
1_CTIPE	8	7.6	8	6.4	7	7.9	8	7.4
4_GITM	10	8.8	10	7.1	10	9.5	10	8.6
Physics-based Data Assimilation Model								
1_USU-GAIM		3.3		3.6		2.8		3.5

Table 2-2 Model ranking w.r.t. RMSE of dTEC_q

dTEC_q	average RMSE (TECU)				global RMSE (TECU)			
	low and middle		all latitudes		low and middle		all latitudes	
	ranking	RMSE	ranking	RMSE	ranking	RMSE	ranking	RMSE
Empirical								
2_IRI	1	6.6	1	5.5	1	7.4	1	7.2
1_IRI	2	6.7	2	5.5	2	7.5	2	7.2
Physics-based Ionosphere Model								
1_USU-IFM	1	7.4	1	6.0	1	7.8	1	7.3
1_SAMI3	1	7.4			2	8.0		
Physics-based Coupled Ionosphere-Thermosphere Model								
2_CTIPE	1	6.3	1	5.3	1	7.2	1	6.8
4_TIE-GCM	2	6.4	1	5.3	2	7.4	2	6.9
2_TIE-GCM	3	6.6	3	5.4	5	7.8	4	7.3
3_TIE-GCM	3	6.6	3	5.4	4	7.7	3	7.2
1_CTIPE	5	6.7	5	5.8	3	7.6	4	7.3
1_TIE-GCM	6	6.9	6	6.0	6	7.8	6	7.4
1_UAM	7	8.1	8	6.6	7	9.5	7	8.8
3_UAM	8	8.2	7	6.5	8	9.9	8	9.1
2_UAM	9	8.6	9	6.9	9	10.3	9	9.5
4_GITM	10	9.9	10	8.0	10	11.3	10	10.3

Physics-based Data Assimilation Model								
1_USU-GAIM	2.6	2.6	3.1	3.1	2.6	2.6	3.3	3.3

Table 2-3 Model ranking w.r.t. NRMSE of dTEC_m

TEC* (dTEC_m)	average NRMSE				global NRMSE			
	low and middle		all latitudes		low and middle		all latitudes	
	ranking	RMSE	ranking	RMSE	ranking	RMSE	ranking	RMSE
Empirical								
1_IRI	1	0.77	1	0.91	1	0.80	1	0.88
2_IRI	2	0.79	1	0.91	2	0.83	2	0.90
Physics-based Ionosphere Model								
1_USU-IFM	1	0.87		0.94	1	0.82		0.89
1_SAMI3	2	0.91			2	0.90		
Physics-based Coupled Ionosphere-Thermosphere Model								
4_TIE-GCM	1	0.79	1	0.87	1	0.83	1	0.89
3_TIE-GCM	1	0.79	1	0.87	2	0.86	2	0.91
2_TIE-GCM	3	0.80	1	0.87	4	0.87	3	0.92
2_CTIPE	3	0.80	4	0.90	2	0.86	3	0.92
1_TIE-GCM	5	0.85	5	0.95	4	0.87	5	0.93
1_UAM	6	0.90	6	0.99	6	0.94	6	0.99
3_UAM	7	0.97	7	1.05	7	1.00	8	1.06
2_UAM	8	0.98	8	1.06	7	1.00	8	1.06
1_CTIPE	9	0.99	9	1.07	7	1.00	7	1.05
4_GITM	10	1.14	10	1.18	10	1.21	10	1.24
Physics-based Data Assimilation Model								
1_USU-GAIM		0.44	0.67	0.67	0.36	0.36	0.50	0.50

Table 2-4 Model ranking w.r.t. NRMSE of dTEC_q

dTEC_q	average NRMSE				global NRMSE			
	low and middle		all latitudes		low and middle		all latitudes	
	ranking	RMSE	ranking	RMSE	ranking	RMSE	ranking	RMSE
Empirical								
2_IRI	1	1.44	1	1.45	1	1.71	1	1.74
1_IRI	2	1.46	2	1.47	2	1.72	2	1.75
Physics-based Ionosphere Model								
1_SAMI3	1	1.66			2	1.91		
1_USU-IFM	2	1.70	1	1.65	1	1.86	1	1.88
Physics-based Coupled Ionosphere-Thermosphere Model								
2_CTIPE	1	1.42	2	1.47	1	1.73	4	1.87
4_TIE-GCM	1	1.42	1	1.46	2	1.76	2	1.78
2_TIE-GCM	3	1.45	3	1.47	5	1.87	1	1.75
3_TIE-GCM	4	1.48	4	1.49	4	1.84	3	1.85
1_CTIPE	5	1.57	5	1.71	3	1.82	8	2.33
1_TIE-GCM	6	1.61	6	1.78	5	1.87	6	1.89
3_UAM	7	1.90	7	1.80	8	2.36	9	2.45
1_UAM	8	1.95	8	1.93	7	2.28	7	2.26
2_UAM	9	2.06	9	2.00	9	2.47	5	1.88
4_GITM	10	2.42	10	2.41	10	2.69	10	2.64
Physics-based Data Assimilation Model								
1_USU-GAIM		0.64	0.99	0.99	0.63	0.63	0.85	0.85

Table 3-1 Model ranking w.r.t. Yield and dt_max of dTEC_m

TEC* (dTEC_m)	average Yield				average |dt_max| (hrs)			
	low and middle		all latitudes		low and middle		all latitudes	
	ranking	Yield	ranking	Yield	ranking	dt_max	ranking	dt_max
Empirical								
2_IRI	1	0.66	1	0.60	1	8.9	1	9.4
1_IRI	2	0.62	2	0.51	2	11.9	2	13.2
Physics-based Ionosphere Model								
1_USU-IFM	1	0.65		0.70	1	13.6		11.6
1_SAMI3	2	0.58			2	14.9		
Physics-based Coupled Ionosphere-Thermosphere Model								
4_GITM	1	0.92	1	0.88	10	11.3	7	10.1
4_TIE-GCM	2	0.70	2	0.67	2	7.3	1	6.5
1_TIE-GCM	3	0.69	2	0.79	8	10.8	5	9.5
3_TIE-GCM	4	0.67	4	0.66	6	10.0	2	8.4

2_TIE-GCM	4	0.67	5	0.65	9	11.0	4	9.3
1_UAM	6	0.55	6	0.54	5	9.6	6	9.9
2_CTIPE	7	0.50	7	0.52	1	6.8	3	8.6
2_UAM	8	0.48	8	0.44	4	9.5	10	11.5
3_UAM	9	0.43	9	0.41	7	10.4	8	10.4
1_CTIPE	10	0.39	10	0.34	3	8.4	9	11.4
Physics-based Data Assimilation Model								
1_USU-GAIM		0.87		0.82		2.3		4.3

Table 3-2 Model ranking w.r.t. Yield and dt_max of dTEC_q

dTEC_q	average Yield				average |dt_max| (hrs)			
	low and middle		all latitudes		low and middle		all latitudes	
	ranking	Yield	ranking	Yield	ranking	dt_max	ranking	dt_max
Empirical								
2_IRI	1	0.10	1	0.21	1	12.3	1	10.6
1_IRI	2	0.08	2	0.08	2	15.7	2	13.2
Physics-based Ionosphere Model								
1_SAMI3	1	0.55			2	20.8		
1_USU-IFM	2	0.34		0.45	1	13.3		11.7
Physics-based Coupled Ionosphere-Thermosphere Model								
4_GITM	1	1.05	2	0.99	6	11.6	7	12.8
1_TIE-GCM	2	0.82	1	1.05	3	9.7	1	9.2
3_TIE-GCM	3	0.78	3	0.83	4	10.2	4	10.3
2_TIE-GCM	4	0.75	4	0.83	5	11.0	5	10.8
1_UAM	5	0.73	6	0.70	1	9.1	3	9.8
4_TIE-GCM	6	0.71	5	0.81	2	9.1	2	9.6
2_UAM	7	0.70	8	0.60	7	13.6	6	12.7
2_CTIPE	8	0.64	7	0.67	9	15.5	9	14.0
3_UAM	9	0.60	9	0.54	8	15.2	8	13.6
1_CTIPE	10	0.45	10	0.43	10	17.7	10	18.4
Physics-based Data Assimilation Model								

1_USU-GAIM		0.78		0.73	3	3.1	5	4.8

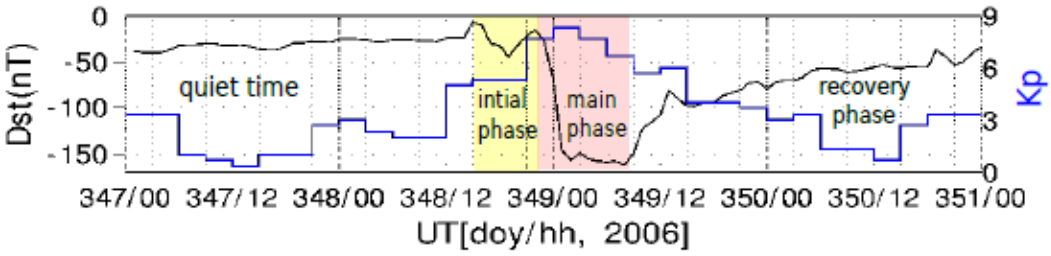

Figure 1. Dst (black) and Kp (blue) values for the 2006 Dec. event.

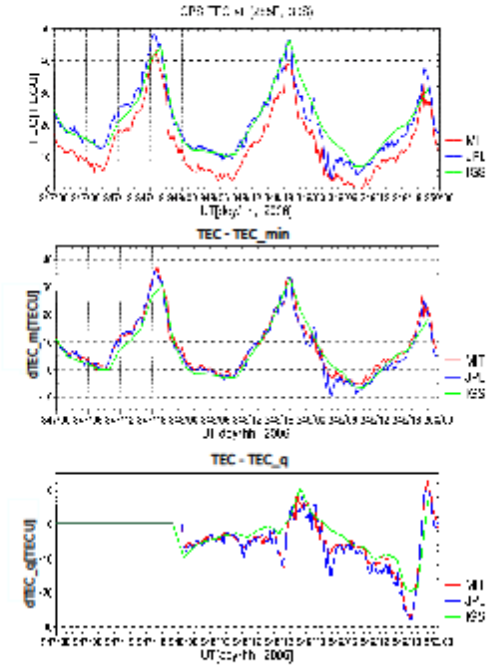

Fig. 2

Figure 2. Example of the observed vertical TEC from the three data sets (MIT GPS in red, JPL TEC in blue, and IGS TEC in green) at (285° E, 30° S) during Dec. 13 – 15 (doy 347-349), 2006. The upper, middle and lower panels show vertical TEC, dTEC_m (TEC – minimum TEC of doy 347), and dTEC_q (TEC – TEC on doy 347), respectively.

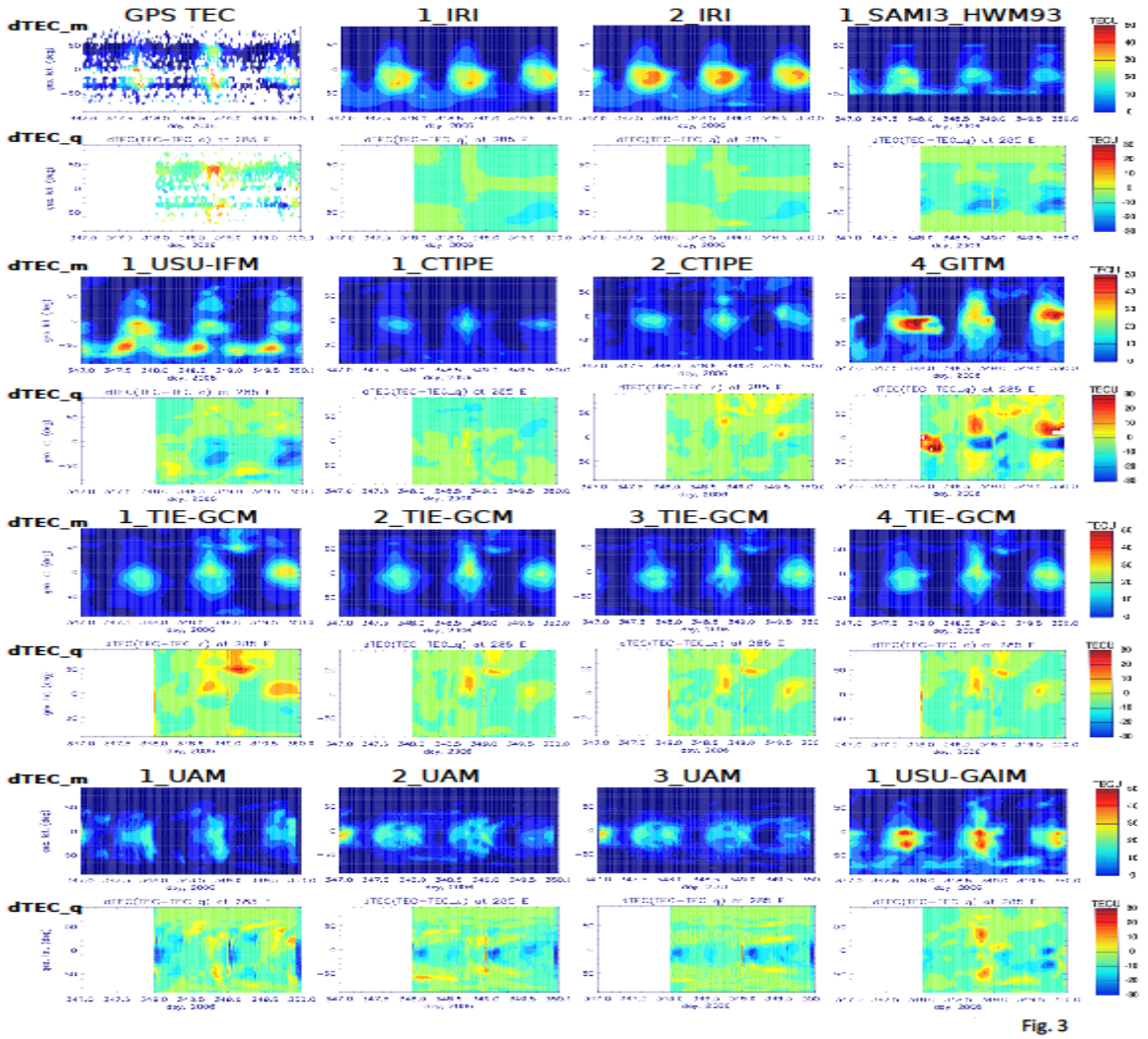

Figure 3. Example of the observed and modeled dTEC_m (TEC – minimum TEC of doy 347), and dTEC_q (TEC – TEC on doy 347) in 285° E as a function of geographic latitude and UT.

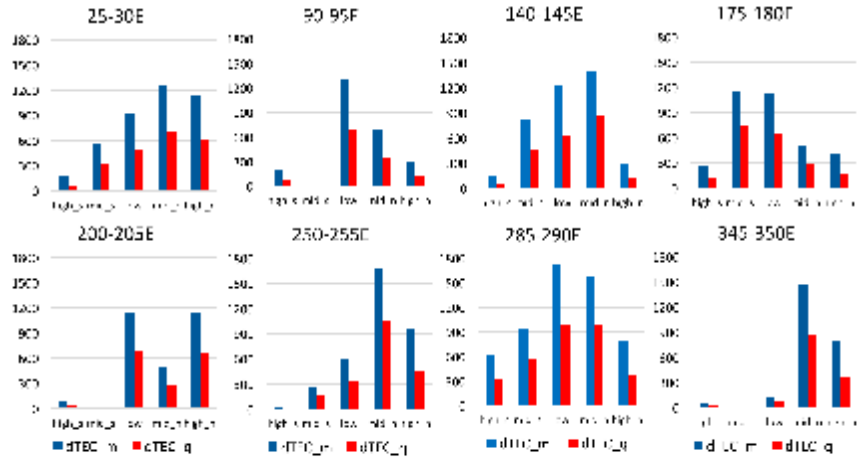

Fig. 4

Figure 4. The total number of data points used for RMSE and NRMSE calculations for dTEC_m (blue bars) and dTEC_q (red bars) predictions in each of eight longitude and five latitude sectors: southern high (lat < -50°) and middle (-50° < lat < -25°), low (|lat| < 25°), and northern middle (25° < lat < 50°) and high (lat > 50°) latitudes (from left to right on the x-axis).

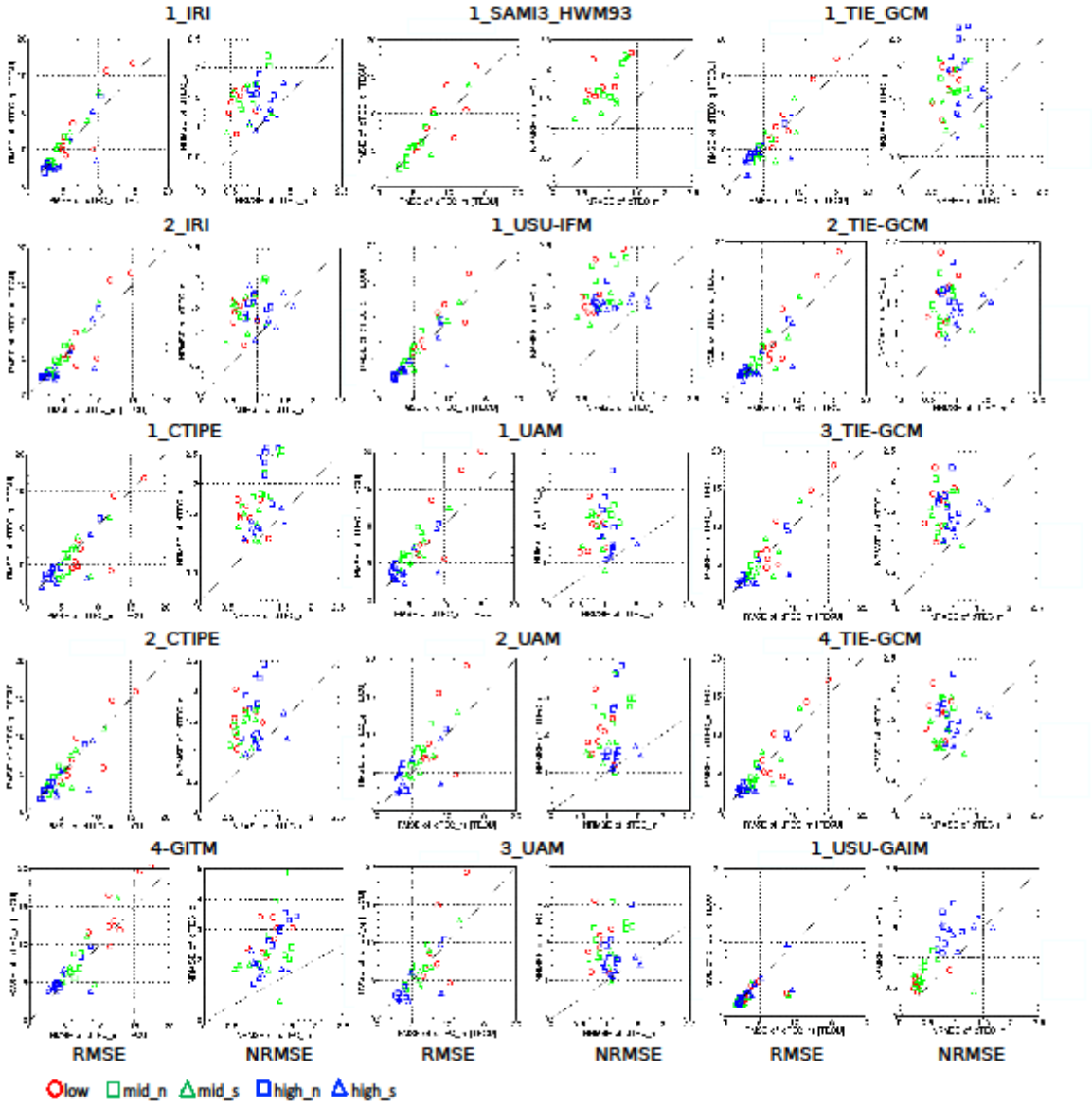

Figure 5. Scatter plots of RMSE and NRMSE of the 15 simulations for all 8 longitude sectors: RMSE of dTEC_m (x-axis) and dTEC_q (y-axis) in the left panel, and NRMSE in the right panel. Red circles indicate low latitudes, green squares northern middle, green triangles southern middle, blue squares northern high, and blue triangles southern high.

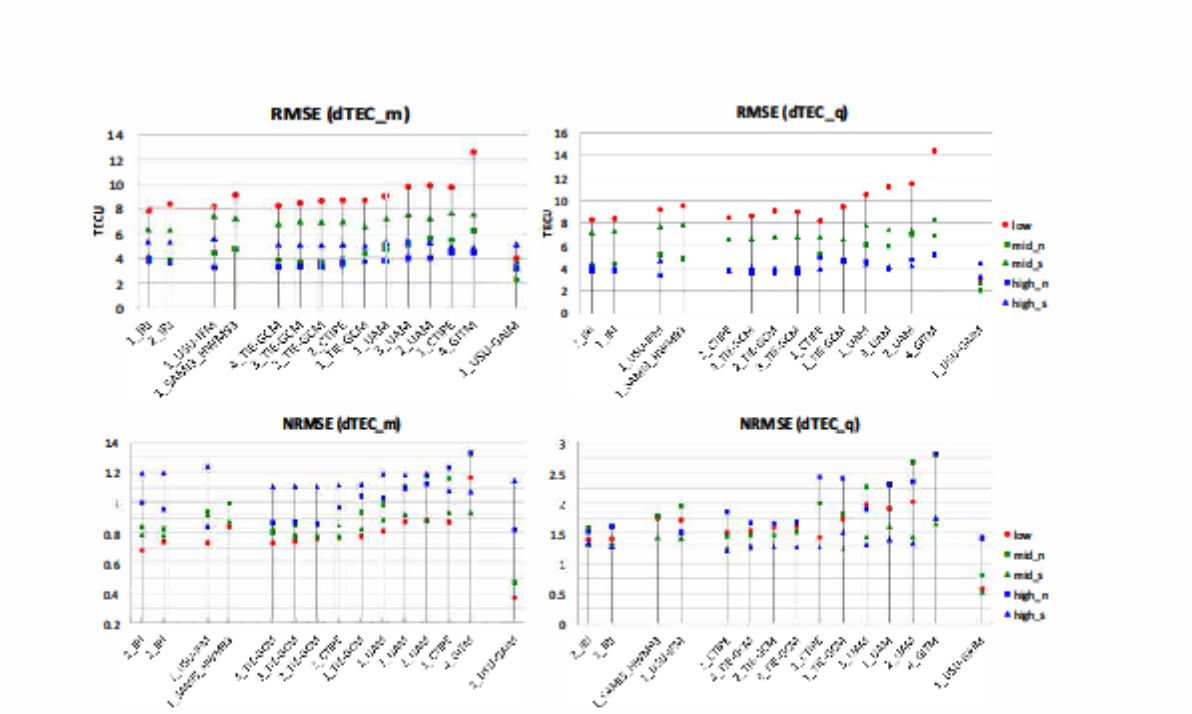

Figure 6. Model ranking for predicting dTEC_m (left panel) and dTEC_q (right panel) based on RMS error (upper panel), and NRMS error (lower panel). Red circles denote the average values of 8 longitude sectors for low latitudes (|lat| < 25°), green squares and triangles indicate the average of those for middle latitudes (25° < |lat| < 50°) in northern and southern hemispheres, blue squares and triangles indicate the average for high geographic latitudes (|lat| > 50°) in northern and southern hemispheres, respectively. The ranking of the models performance among the same types of the models are arranged by the average values of three latitude sectors (low and middle latitudes (mid_n and mid_s)) in each group of the models. The best performing model is located in the extreme left in each group.

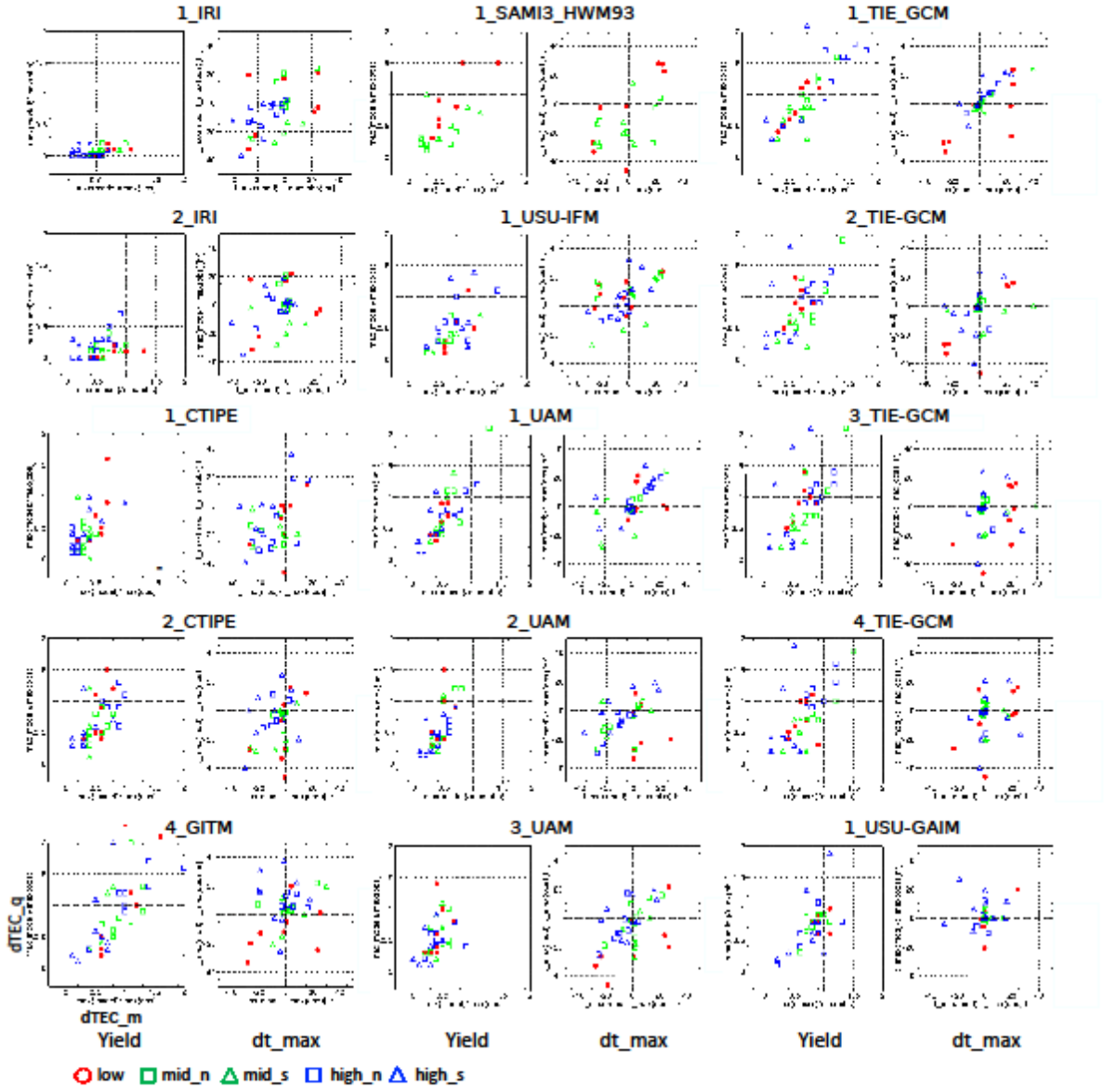

Fig. 7

Figure 7. Same as Figure 5 but for Yield (ratio) and differences (dt_max =t_max_model – t_max_obs) in time (at which the maximum of dTECs occurs during the storm time) between a simulation and observation.

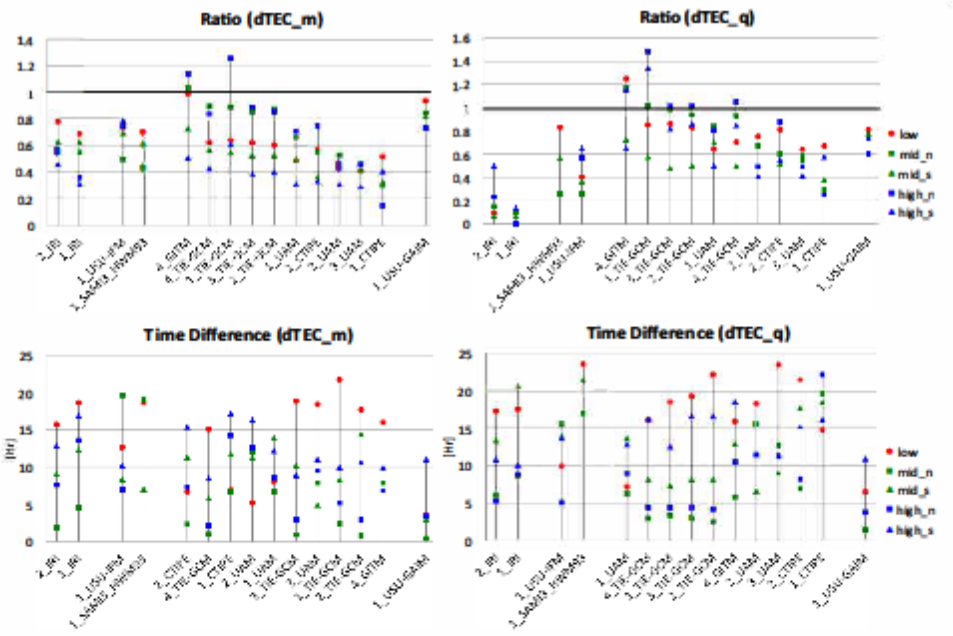

Fig. 8

Figure 8. Same as Figure 6 but for model ranking based on Yield (ratio of the maximum modeled TEC to that of the observed TEC) and time difference, which is average of time delay (dt_max > 0) and |time advance| (dt_max < 0), of dTEC_m (left) and dTEC_q (right) from top to bottom.