

Assessment of the predictive capability of IT models at the Community Coordinated Modeling Center

Ja Soon Shim^{1*}, Lutz Rastätter², Maria M. Kuznetsova², Emine C Kalafatoglu³, Yihua Zheng²

¹CUA/NASA GSFC, Greenbelt, MD, USA,

²NASA/GSFC, Greenbelt, MD, USA,

³Istanbul Technical University, Turkey

Abstract

The Community Coordinated Modeling Center (CCMC) is an interagency partnership with the goal of bridging the gap between science and space weather operations while enhancing research, supporting development of next-generation space weather models, and disseminating knowledge to the broader community. The CCMC hosts the largest assembly of state-of-the-art physics-based space weather models and has developed a variety of tools for space weather visualization and analysis. In addition to providing the community an easy access to these modern space research models and tools to support science research, one of the primary goals of the CCMC is to test and validate models for transition from research to operations. In this paper, we focus on the community-wide Ionosphere/Thermosphere (IT) model validation efforts led by the CCMC and present results of the assessment of the models for reproducing storm impacts on IT parameters such as TEC, neutral density, and Joule Heating. In order to quantify storm impacts on the parameters, we consider several quantities: changes compared to quiet time (the day before storm), difference in 24-hour intervals, and maximum increase during the storm. We compare the calculated quantities from the models with the observed values, including ground-based GPS TEC measurements in several longitude sectors where data coverage is relatively better, orbital averages of neutral densities along the CHAMP satellite track and the DMSP Poynting Flux during storm events (e.g., 2006 AGU storm). Model output and observational data used for the challenge will be permanently posted at the CCMC website (<http://ccmc.gsfc.nasa.gov>) as a resource for the space science communities to use.

1. Introduction

Our daily lives, increasingly dependent on technological infrastructure such as satellites used for communications and navigations, 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 address the needs of our space science, the CCMC (Community Coordinated Modeling Center) has effectively established an "Open Model Policy" that gives access to modern space science simulations through one-of-a-kind Runs-on-Request (RoR) system [Webb *et al.*, 2009] for anyone interested in the subject. In addition to the provision of the easy access to these modern space research models, the CCMC's another primary goal is to test and validate models for transition from research to operations [Pulkkinen *et al.*, 2013]. To evaluate the current state of space weather modeling capability and to track improvements of the models, it is important to assess model performance quantitatively. In an effort to address needs and challenges of the quantitative assessment of modeling capabilities, the CCMC initiated a series of

community-wide model validation projects: SHINE, GEM [Pulkkinen *et al.*, 2010, 2012; Rastätter *et al.*, 2013], CEDAR and GEM-CEDAR Modeling Challenges. The CEDAR ETI (Electrodynamics Thermosphere Ionosphere) Challenge focuses on the ability 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]. Model-data time series comparisons were performed for a set of selected events with different levels of geomagnetic activity (quiet, moderate, storms). The follow-on CEDAR-GEM Challenge aims to quantify geomagnetic storm impacts on the IT system. In this paper, we present results of the assessment of the model performance for reproducing storm impacts on the key IT parameters, including TEC, neutral density, and Joule Heating. In order to quantify storm impacts on the parameters, we calculate several quantities: changes compared to quiet time (the day before storm), difference in 24-hour intervals, and maximum changes during the storm. We compare the calculated quantities from the models with the observed values, including ground-based GPS TEC measurements in several longitude sectors, orbital averages of neutral densities along the CHAMP satellite track and the DMSP Poynting Flux during 2006 AGU storm.

2. TEC

Ionospheric variability affects navigation and communication systems. Total electron content (TEC) is one of the key parameter in the description of the ionospheric variability that has influence on the accuracy of navigation and communication. As a first step to assess the IT model performance of reproducing global TEC variations during storms, we compared the modeled TEC with observations in eight 5-degree wide longitude sectors (25°-30°E, 90°-95°, 140°-145°, 175°-180°, 200°-205°, 250°-255°, 285°-290°, 345°-350°) during AGU storm in 2006 (12/13 -15, doy 347-349). Each longitude has 36 latitude bins of 5 degrees each from -90 to +90. For this validation study, we used more than 10 different simulations of IT models ranging from empirical (2 simulations), physics-based ionospheric (2 simulations), and coupled IT models (7 simulations) to physics based data assimilation model (1 simulation). The simulations were either submitted by model developers or generated by the CCMC using the IT models hosted at the CCMC [Webb *et al.*, 2009]. We used GPS TEC measurements as ground truth observations, provided by Drs. Coster and Goncharenko at MIT Haystack Observatory. During the 2006 AGU storm, Kp index increased up to about 9 and Dst index decreased to about -150 nT (Fig. 1).

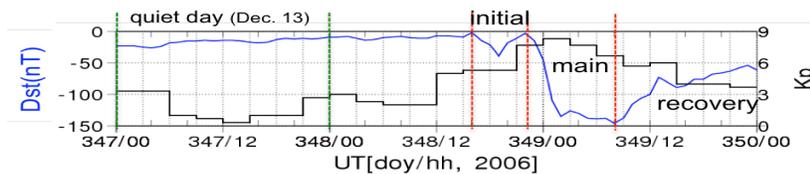


Figure 1. Kp and Dst index during 2006 AGU storm (Dec. 13-15, doy 347-349)

Figure 2 shows an example of TEC (a and b) and dTEC (c-e) in one of the eight longitude sectors (140°-145°E) as a function of latitude at 03:00 UT on doy 348 (before initial phase of the storm) and 349 (during main phase). dTEC includes (1) TEC change compared to quiet time (dTEC_q = TEC on current day – TEC on the day before storm, Dec. 13) and (2) difference in 24-hour intervals (dTEC_p = TEC on current day – TEC on previous day). Most models overestimate TEC in this longitude sector during quiet period (Figure 2a) despite the fact that upper limit of the models is lower than GPS satellite height (20,200 km). However, all models except a data assimilation model (brown line) fail to reproduce TEC enhancement at the right locations in the northern hemisphere during the main phase of the storm (Figure 2b).

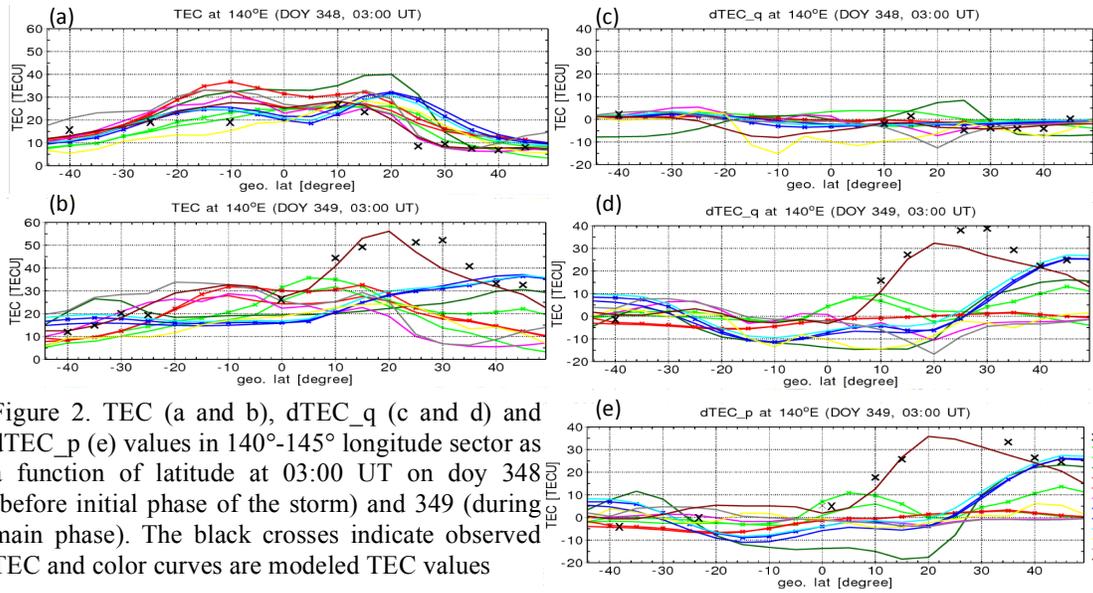


Figure 2. TEC (a and b), dTEC_q (c and d) and dTEC_p (e) values in 140°-145° longitude sector as a function of latitude at 03:00 UT on doy 348 (before initial phase of the storm) and 349 (during main phase). The black crosses indicate observed TEC and color curves are modeled TEC values

To quantify models' capability to reproduce storm impact on TEC, we calculated dTEC_q and dTEC_p. In 140° E, there is no noticeable difference between two observed dTECs since TEC maximum enhancement occurs on doy 349. Physics-based coupled models (blue and green lines in Figure 2d and 2e) produce dTEC_q and dTEC_p peak values better than TEC peak in higher latitudes, even though the location of the peak does not agree with the GPS TEC peak latitude.

To evaluate performance of the models quantitatively, we calculated skill scores, including RMS error and Yields (ratio of the maximum modeled TEC and dTEC_q to that of the observed values). Figure 3 shows eight longitude averaged RMS errors (top row) and Yields (bottom row). To investigate latitudinal dependence of model performance, the skill scores were calculated for five latitude regions: low, southern middle, northern middle, southern high, and northern high latitudes (left to right). In each plot, x and y axes correspond to skill scores for TEC and dTEC_q predictions

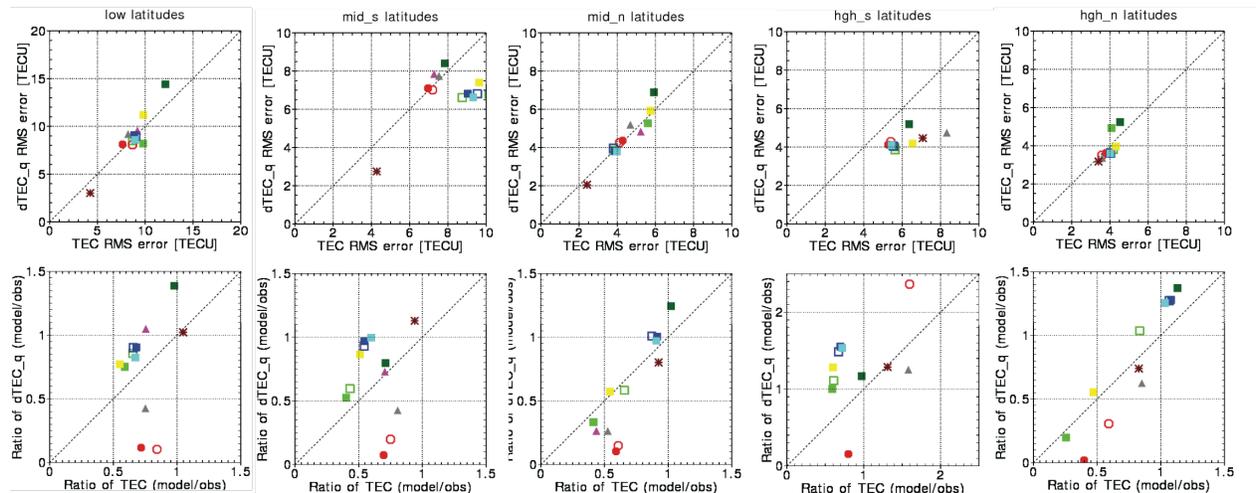


Figure 3. RMS errors (upper row) and Yields (ratios of maximum of TEC (x-axis) and dTEC_q modeled values to maximum of observed values) (lower row). Circles indicate empirical model simulations, triangles physics-based ionospheric, squares physics-based coupled ionosphere thermosphere, and asterisk physics based data assimilation models.

respectively. Most of the model simulations show similar scores for the $dTEC_q$ and $dTEC_p$ predictions (not shown here). The model simulations show similar RMS errors for TEC and $dTEC_q$ in the low and northern middle and high latitudes, while RMS errors are larger for TEC than for $dTEC_q$ in southern middle and high latitudes, especially for physics-based coupled models (squares). For most cases, the performance of the data assimilation model (brown asterisk) in predicting TEC and $dTEC_q$ (located close to (0, 0) and (1, 1) in the plots in top and bottom row, respectively) is superior to other models' performance. In terms of Yield, the physics based coupled and data assimilation models have better scores than the empirical (circles) and physics-based ionosphere models (triangles) in middle latitudes. The empirical model simulations are inferior to others in all latitude regions in terms of Yield, whereas the empirical simulations have smaller RMS errors than physics-based model simulations. It should be noted that, in addition to the value of Yield, we should take into account the timing and location of $dTEC$ peak. There is no one model that scores high in all cases, including timing and location of peak (not shown here), than others. More detailed results of the study will be discussed in the CCMC CEDAR-GEM Modeling Challenge report [Shim *et al.*, 2015].

3. Neutral Density

One of the main sources of errors in orbit determination is thermospheric density variation. Accurate specification and prediction of neutral density is necessary to minimize orbital drag errors. To evaluate how accurately the IT models reproduce neutral density, the density from the IT models averaged along the orbit are compared with the observed density along the CHAMP trajectories during the same AGU storm event in 2006 as in the TEC study above. We used two empirical model simulations and three physics-based coupled IT model simulations and calculated the orbit averaged neutral density with OAT (Orbit Average Tool) developed by the CCMC [Kalafatoglu *et al.*, 2015]. In this paper, to assess the models' capability to capture storm effects on neutral density, two shifting approaches were performed: (1) modeled densities were shifted to CHAMP data by subtracting from them the quiet time mean difference between the observation and modeled values, and (2) both modeled and observed data were shifted to zero by point-to-point subtraction of the quiet time neutral density obtained from the quiet day runs of each model (12/13, doy 347). Figure 4a shows one example of the orbit averaged neutral density along the CHAMP after shifting to zero (2). Although two of the model simulations underestimate the increments to the undisturbed thermosphere, all five show similar features during the storm.

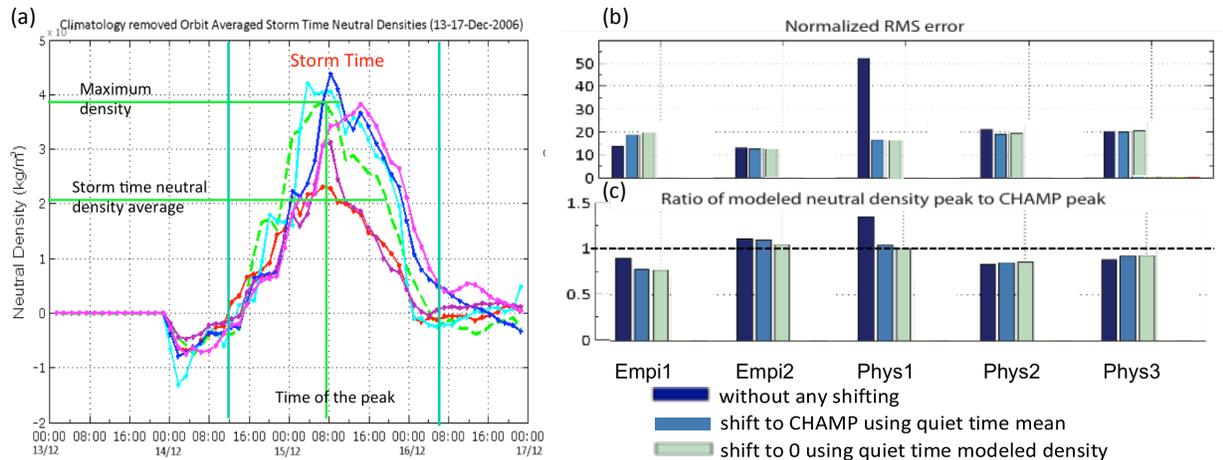


Figure 4. (a) Observed (dashed green line) and modeled orbital averaged neutral density along the CHAMP and skill scores, (b) normalized RMS error, and (c) ratios of modeled peak density to the observed values.

To compare the performance of the models quantitatively, we calculated normalized RMS errors (Figure 4b) and Yield (modeled peak value/observed peak value) (Figure 4c). In Figure 4a and 4b, navy-colored bar indicates the original modeled neutral density without any shift, blue and green bars indicate density with shift (1) and (2), respectively. Five groups of color bars correspond to two empirical (Empi1 and 2) and three physics-based model simulations (Phys1-3) from the left to the right. Skill scores without any shift of Empi1 and Empi2 are better than those of Phys1-3. Empi1 and Empi2 without shift show similar performance; however, Empi1 tends to underestimate and Empi2 overestimate the neutral density peak during the storm. Empi1 has better scores before shifting data, and Phys1 has better scores after shifting data, while the other three hardly show any difference between the scores without shifting and those after the two shifting approaches. Phys1 after the shifts performs the best in reproducing peak density; however, it has larger timing error than the empirical models. Empi1, which produces the worst Yield, shows the best performance in terms of timing of peak (Figure 5).

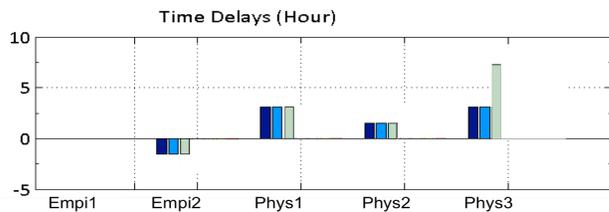


Figure 5. Timing error (observed peak time – modeled peak time).

■ without any shifting
 ■ shift to CHAMP using quiet time mean
 ■ shift to 0 using quiet time modeled density

4. Poynting Flux

Several models of the Ionosphere-Thermosphere were run for the 12/14- 12/16/2006 time period. Physics-based models include 3-dimensional models of the Ionosphere/Thermosphere and 2-dimensional ionospheric electrodynamics modules of global magnetosphere MHD models. Empirical models specify Joule Heating from the electrodynamics of the ionosphere or the Poynting Flux directly. DMSP observations were used to compute Poynting Flux values (vertical component S_z) derived from in-track and cross-track electric and magnetic fields observed. All model results were interpolated onto the DMSP satellite track and timelines were analyzed in each pass of the auroral zone (i.e., satellite orbit segments within 45 degrees of the northern and southern magnetic poles) [Rastätter *et al.*, 2015]. Three passes through the auroral zone at the onset of the storm are shown in Figure 6 with the physics-based model Joule Heating on the left and Joule Heating and Poynting Flux from empirical models on the right.

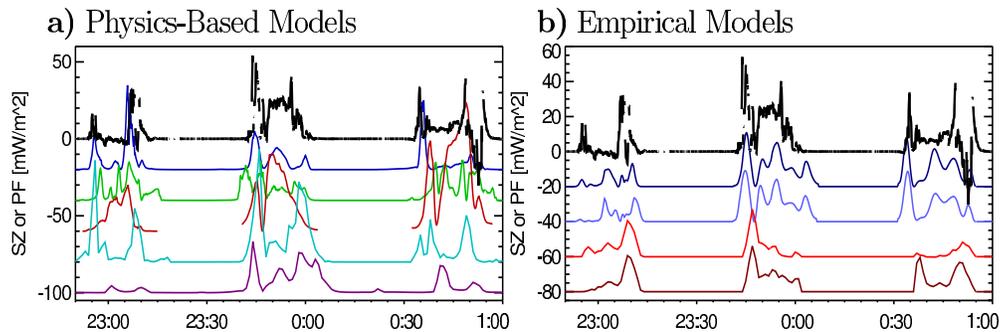


Figure 6. Observed vertical component of Poynting Flux (S_z) with sign flipped (negative S_z denoting energy flux into the ionosphere) to match positive Joule dissipation of a) physics-based models and b) empirical models. Adjacent traces in the stack plots are by -20 mW/m^2 .

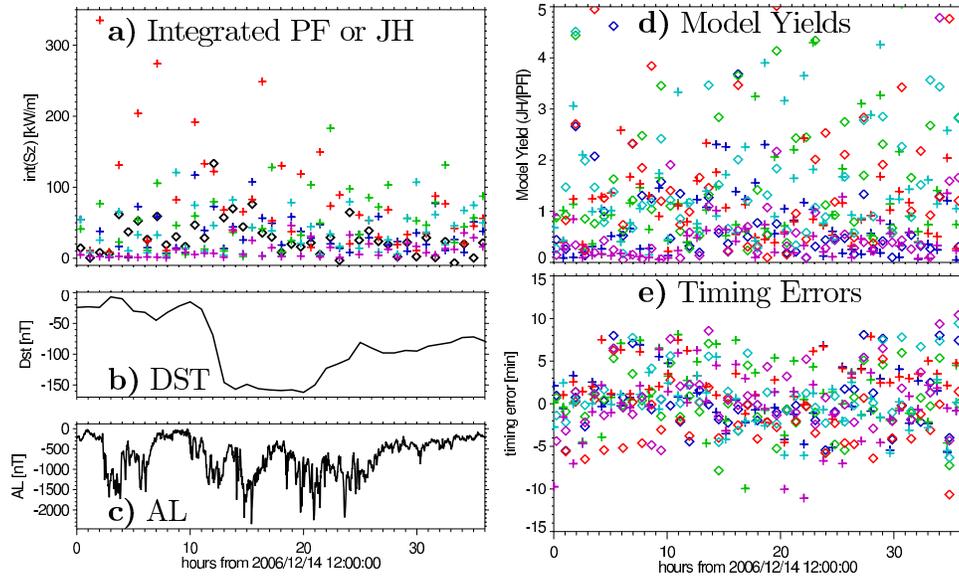


Figure 7. Poynting Flux (PF) and Joule Heating (JH) integrated along the DMSP-F15 satellite track during polar region crossing. On the left: a) Total Poynting Flux observed (black diamonds) and Joule Heat from five physics-based model runs (colors), b) Dst index, c) auroral electrojet AL index. Right: d) maximum Joule Heat value divided by maximum Poynting Flux observed, e) time difference in the occurrence of maximum value during segments of auroral passes (cross: inbound toward magnetic pole, diamonds: outbound away from pole).

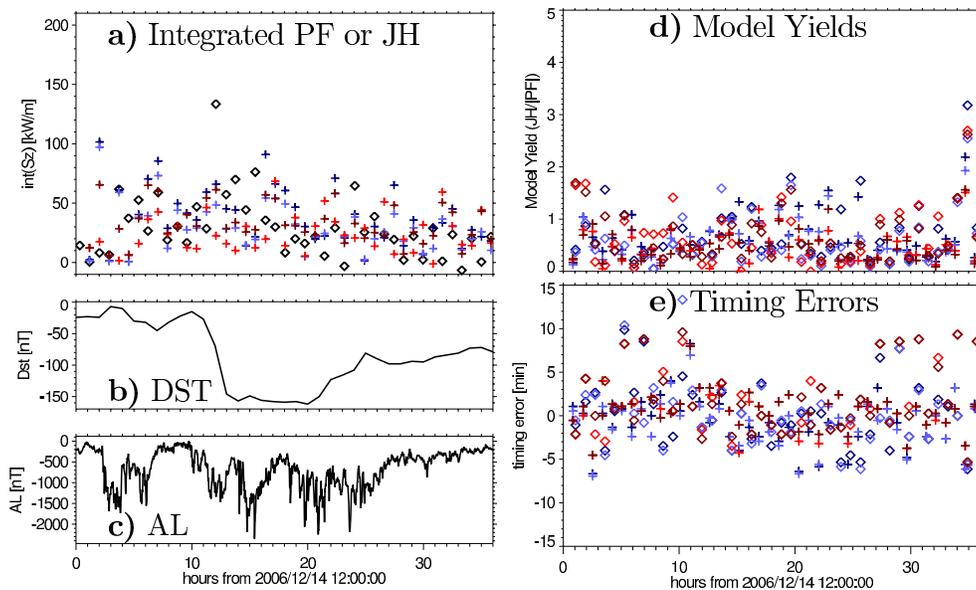


Figure 8. Analysis for 4 empirical model runs in the same format as the previous figure. Note the narrower spread experienced by these models in all three measures compared to the physics-based models

The figures show three measures of performance for the physics-based models (Figure 7) and the statistical or empirical models (Figure 8). In each polar cap/auroral pass, the Poynting Flux (in mW/m) is calculated and compared with the modeled Joule Heating (Poynting Flux for one specific empirical model). From the plots we immediately see that the empirical models exhibit a smaller spread in terms of Yield (ratio of JH/PF) and timing error compared to the physics-based models. Empirical models have Yields below 1 for the most part. Each model specifies the changes in size

and shape of the aurora between quiet and disturbed times differently, resulting in timing errors. Yields and timing errors observed in individual passes do not change significantly during the course of the storm event. Note that the Yields (d) are derived from different values than the track-integrated Poynting fluxes or Joule dissipation (a).

Model run	Avg. Yield1	Sigma Yield1	Avg. Yield2	Sigma Yield2	Avg. DT1	Sigma DT1	Avg. DT2	Sigma DT2
1-Phys	0.71	0.63	0.99	1.22	0.14	3.18	0.30	3.57
2-Phys	1.43	1.38	2.21	2.28	1.14	4.12	-0.10	3.51
3-Phys	1.05	0.61	1.86	1.63	2.70	3.89	-2.04	3.81
4-Phys	1.51	1.05	1.61	1.28	-0.32	2.77	1.58	3.98
5-Phys	0.39	0.33	0.62	0.83	-0.35	4.00	1.10	3.96
1-Empi	0.70	0.43	0.72	0.59	-0.84	3.07	0.09	3.69
2-Empi	0.56	0.38	0.56	0.50	-0.94	2.92	0.17	3.84
3-Empi	0.46	0.32	0.67	0.57	0.46	2.40	1.92	4.48
4-Empi	0.49	0.31	0.67	0.52	0.56	2.18	1.95	4.42

Table 1: Average (Avg.) and standard deviation (sigma) of Yield1, Yield2, Timing errors DT1 and DT2 for the five physics-based models (labeled “Phys”) and four empirical models (labeled “Empi”).

Table 1 shows the average Yields, timing errors, and their standard deviation (sigma). Physics-based models have Yields at or greater than unity (perfect score), and all empirical models have Yields below unity with proportionally smaller sigma. Timing errors exhibit large standard deviations and are indistinguishable between the classes of models. Most models have negative DT1 and positive DT2 (or $DT2 > DT1$), suggesting that the modeled Joule dissipation region in the auroral zone extends farther away from the magnetic pole than the observed Poynting flux pattern. Two of the physics-based models have $DT2 < DT1$ suggesting the opposite (modeled dissipation region closer to the pole than the observed Poynting fluxes). The absolute values of DT1 and DT2 usually differ in each pass, indicating that the modeled Joule Heating pattern is shifted in local time compared to the observed Poynting flux pattern.

5. Summary and Future Plans

In this paper, we performed model-data comparison of TEC, neutral density along the CHAMP satellite track, and the DMSP Poynting Flux for 2006 AGU storm event by using more than 10 model simulations of the Ionosphere-Thermosphere (IT), including 3-dimensional IT models and 2-dimensional ionospheric electrodynamics modules of global magnetosphere MHD models. To quantify storm impacts, we consider changes compared to quiet time, difference within 24-hour intervals, and maximum increase during the storm. We calculated skill scores such as RMS errors, ratio of maximum, and timing errors to quantify performance of the models.

We find that eliminating quiet time climatology gives a better way to determine the actual storm-time response and to remove baselines of both the modeled and the observed data. We also find that no single model simulation is always better than the others in terms of the skill scores. Model performance depends on metrics, parameters, latitudes, and data preparation approaches (e.g., shifting, averaging and etc.). Therefore, an ensemble of model simulations will allow for the models to supplement each other’s shortcomings.

In order to better understand the effects of geomagnetic storms on IT system, it is critical to investigate the effects of energy inputs from magnetosphere on the enhancement of TEC and neutral density in high latitudes. Therefore, through our next validation activity, we will examine the IT

parameters' response to high latitude drivers and how each latitude region (high/middle/low) of IT system responds to the external drivers.

Acknowledgements

GPS TEC measurements were provided by Drs. Coster and Goncharenko at MIT Haystack Observatory. CHAMP neutral density data were provided by Eric Sutton at AFRL. We wish to thank the model developers of Cosgrove, CTIPe, IRI, JB2008, LFM-MIX, NRLMSISE-00, OpenGGCM, SAMI3, SWMF, TIE-GCM, UAM, USU-GAIM, and Weimer 2005.

References

- Kalafatoglu Eyiguler E. C., J. S. Shim, M. M. Kuznetsova, Z. Kaymaz, B. R. Bowman, M. Codrescu, B. Emery, B. Foster, T. Fuller-Rowell, A. Ridley, (2015), Quantifying the storm-time thermospheric neutral density variations using model and observations, to be submitted to JGR, 2015
- 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., et al. (2013), Geospace environment modeling 2008–2009 challenge: Dst index, *Space Weather*, *11*, 187–205, doi:10.1002/swe.20036.
- Rastätter, L., J. S. Shim, M. M. Kuznetsova, M. Codrescu, T. Fuller-Rowell, B. Emery, D. Weimer, R. Cosgrove, P. Newell, M. Wiltberger, J. Raeder, T. I. Gombosi, GEM-CEDAR Challenge: Poynting Flux at DMSP and Modeled Joule Heating, to be submitted to *Space Weather*, 2015.
- Shim, J. S., et al., 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, 2011.
- Shim, J. S., et al., 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, 2012.
- Shim, J. S., et al., Systematic Assessment of Ionosphere/Thermosphere Models in Predicting TEC during 2006 Dec. Storm event, to be submitted to *Space Weather*, 2015.
- Webb, P. A., M. M. Kuznetsova, M. Hesse, L. Rastätter, and A. Chulaki (2009), Ionosphere-thermosphere models at the Community Coordinated Modeling Center, *Radio Sci.*, *44*, RS0A34, doi:10.1029/2008RS004108.