

# CEDAR Electrodynamics Thermosphere Ionosphere (ETI) Challenge for systematic assessment of ionosphere/thermosphere models: Electron density, neutral density, NmF2, and hmF2 using space based observations

J. S. Shim,<sup>1</sup> M. Kuznetsova,<sup>2</sup> L. Rastätter,<sup>2</sup> D. Bilitza,<sup>3</sup> M. Butala,<sup>4</sup> M. Codrescu,<sup>5</sup> B. A. Emery,<sup>6</sup> B. Foster,<sup>6</sup> T. J. Fuller-Rowell,<sup>5</sup> J. Huba,<sup>7</sup> A. J. Mannucci,<sup>4</sup> X. Pi,<sup>4</sup> A. Ridley,<sup>8</sup> L. Scherliess,<sup>9</sup> R. W. Schunk,<sup>9</sup> J. J. Sojka,<sup>9</sup> P. Stephens,<sup>4</sup> D. C. Thompson,<sup>10</sup> D. Weimer,<sup>11</sup> L. Zhu,<sup>9</sup> and E. Sutton<sup>10</sup>

Received 25 August 2012; accepted 12 September 2012; published 18 October 2012.

[1] In an effort to quantitatively assess the current capabilities of Ionosphere/Thermosphere (IT) models, an IT model validation study using metrics was performed. This study is a main part of the CEDAR Electrodynamics Thermosphere Ionosphere (ETI) Challenge, which was initiated at the CEDAR workshop in 2009 to better comprehend strengths and weaknesses of models in predicting the IT system, and to trace improvements in ionospheric/thermospheric specification and forecast. For the challenge, two strong geomagnetic storms, four moderate storms, and three quiet time intervals were selected. For the selected events, we obtained four scores (i.e., RMS error, prediction efficiency, ratio of the maximum change in amplitudes, and ratio of the maximum amplitudes) to compare the performance of models in reproducing the selected physical parameters such as vertical drifts, electron and neutral densities, NmF2, and hmF2. In this paper, we present the results from comparing modeled values against space-based measurements including NmF2 and hmF2 from the CHAMP and COSMIC satellites, and electron and neutral densities at the CHAMP satellite locations. It is found that the accuracy of models varies with the metrics used, latitude and geomagnetic activity level.

**Citation:** 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.

## 1. Introduction

[2] There have been noticeable developments of many ionosphere/thermosphere (IT) models over the last 30 years [Schunk et al., 2002; American Institute of Aeronautics

and Astronautics, 2010] that deepen our understanding of the ionosphere/thermosphere (IT) system. All of the models, however, have errors associated with their predictions of climate and weather of the ionosphere and thermosphere. Therefore, it is important to assess the IT models

<sup>1</sup>Goddard Planetary Heliophysics Institute, University of Maryland Baltimore County, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

<sup>2</sup>NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.  
<sup>3</sup>School of Physics Astronomy and Computational Science, George Mason University, Fairfax, Virginia, USA.

<sup>4</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.

Corresponding author: J. S. Shim, Goddard Planetary Heliophysics Institute, University of Maryland Baltimore County, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. (jasoon.shim@nasa.gov)

<sup>5</sup>Space Weather Prediction Center, NOAA, Boulder, Colorado, USA.

<sup>6</sup>High Altitude Observatory, National Center for Atmospheric Research, Boulder, Colorado, USA.

<sup>7</sup>Plasma Physics Division, Naval Research Laboratory, Washington, D. C., USA.

<sup>8</sup>Space Physics Research Laboratory, University of Michigan, Ann Arbor, Michigan, USA.

<sup>9</sup>Center for Atmospheric and Space Sciences, Utah State University, Logan, Utah, USA.

<sup>10</sup>Air Force Research Laboratory, Albuquerque, New Mexico, USA.

<sup>11</sup>Center for Space Science and Engineering Research, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA.

**Table 1.** Events Studied in the CEDAR ETI Challenge

Event Name	Date (DOY) and Time	F10.7	Kp_max
<i>GEM Events</i>			
E.2006.348	2006/12/14(doy 348) 12:00 UT-12/16 (doy 350) 00:00 UT	91	8
E.2001.243	2001/08/31(doy 243) 00:00 UT-09/01 (doy 244) 00:00 UT	192	4
E.2005.243	2005/08/31 (doy 243) 10:00 UT-09/01 (doy 244) 12:00 UT	86	7
<i>Moderate Storms</i>			
E.2007.091	2007/04/01 (doy 091) 12:00 UT-04/02 (doy 092) 12:00 UT	72	5
E.2007.142	2007/05/22 (doy 142) 12:00 UT-05/25 (doy 145) 00:00 UT	74	6
E.2008.059	2008/02/28 (doy 059) 12:00 UT-03/01 (doy 061) 12:00 UT	69	5
<i>Quiet Periods</i>			
E.2007.079	2007/03/20 (doy 079) 00:00 UT-03/22 (doy 081) 00:00 UT	72	1
E.2007.190	2007/07/09 (doy 190) 00:00 UT-07/10 (doy 191) 00:00 UT	80	0
E.2007.341	2007/12/07 (doy 341) 00:00 UT-12/09 (doy 343) 00:00 UT	80	1

quantitatively in order to not only understand the strong and weak features in their prediction capabilities but also make improvements accordingly.

[3] The CEDAR (Coupling, Energetics, and Dynamics of Atmospheric Regions) community initiated the Electrodynamics Thermosphere Ionosphere (ETI) Challenge in 2009 to assess accuracy of a variety of IT models in predicting ionospheric-thermospheric parameters against measurements.

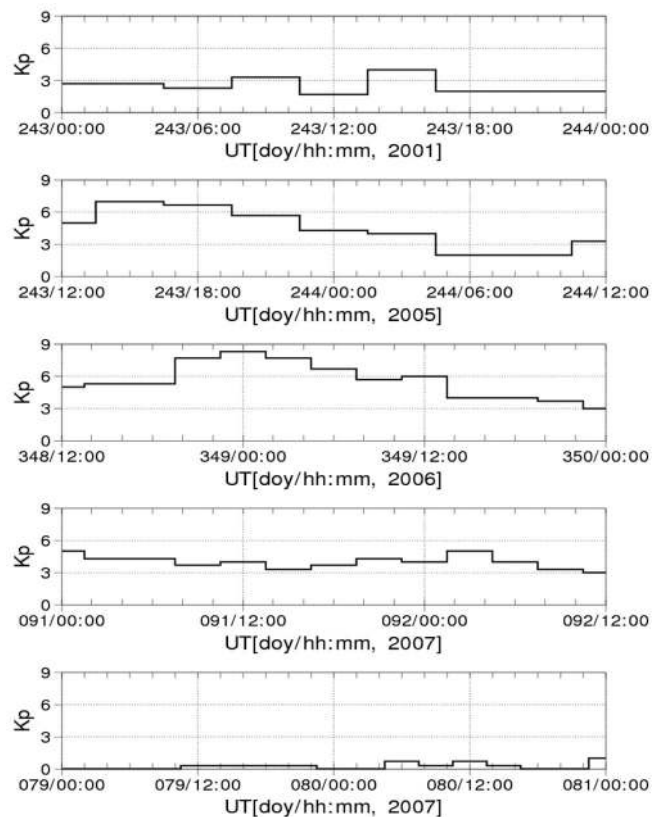
[4] The results of the CEDAR ETI Challenge using ground-based observations, such as vertical drift at Jicamarca, and NmF2 and hmF2 from ISR, have been presented by Shim *et al.* [2011]. Model simulations from up to 10 models were compared with the measurements during nine time intervals (two strong, four moderate geomagnetic storm events, and three quiet periods). The validation study was the first to quantitatively evaluate a wide variety of IT models ranging from empirical to physics-based, coupled IT and data assimilation models. However, the study focused on only the ionospheric parameters and was limited in latitude coverage. In this work, we demonstrate the Challenge results obtained by using space-based observations including the ionospheric parameters, NmF2 and hmF2 derived from radio occultation measurements by Low Earth Orbit (LEO) satellites (CHAMP and COSMIC), and electron density at the CHAMP (CHallenging Minisatellite Payload) locations, and the thermospheric parameter as well, such as neutral density at the CHAMP locations. The high inclination orbits of CHAMP and COSMIC give greater latitude coverage for the measurements. In this study, therefore, dependence of the model performance on latitude was investigated by calculating the four skill scores for three latitude regions, low, middle, and high geographic latitudes, during the selected nine time intervals.

[5] The CEDAR ETI Challenge is supported by the Community Coordinated Modeling Center (CCMC), which develops and provides tools used for a large number of model validation studies. All measurements and model simulation results used are available on the CCMC website for use by the space science communities.

## 2. Setup of the Challenge

[6] We chose all nine events, which were categorized into three levels of geomagnetic activity, to investigate the effect of geomagnetic activity on the IT model performance. Three time intervals were selected from the GEM (Geospace Environment Modeling) Challenge [Pulkkinen *et al.*, 2010, 2011; Rastätter *et al.*, 2011] (see Table 1). Three different geomagnetic activity levels, strong storm ( $Kp\_max \geq 7$ ), moderate storm events ( $4 \leq Kp\_max < 7$ ) and quiet periods ( $Kp\_max < 4$ ) were defined by the Kp index. The three GEM events were divided into moderate (E.2001.243) and strong (E.2005.243 and E.2006.348) storms. All but the moderate 2001 GEM event were for low solar flux conditions ( $F10.7 < 100$ ). Kp values for the selected events (the GEM events, one quiet and one moderate conditions) are shown in Figure 1.

[7] Compared are the modeled values with observed values for (1) neutral and (2) electron densities at the

**Figure 1.** Kp values for five of the studied events.

**Table 2.** Models Submitted for the CEDAR Challenge

Model Setting ID	Model/Version	Grid (lat × lon × alt)
<i>Empirical Models</i>		
1_IRI <sup>a</sup>	IRI-2007, empirical ionospheric model	(50 km < alt < 2,000 km)
1_MSIS	NRLMSISE-00, empirical thermospheric model	
1_JB2008	JB2008, empirical thermospheric density model	
2_JB2008	JB2008 with thermosphere temperature correction derived from Weimer 2005 total Poynting fluxes	
<i>Physics Based Models</i>		
1_SAMI3_HWM93 <sup>a</sup>	SAMI3 with the neutral wind model HWM93 [Hedin et al., 1996]	120 × 90 × 160 (90 km < alt < 20,000 km)
1_USU-IFM <sup>a</sup>	IFM driven by F10.7, Kp and empirical inputs for the thermosphere parameters	60 × 49 × 73 (90 km < alt < 1,600 km)
<i>Coupled IT Models</i>		
1_CTIPe <sup>a</sup>	CTIPe driven by Weimer [2005]	91 × 20 × 15 (~90 km < alt < 500 km)
1_GITM	GITM	25 × 50 × 13
3_GITM	GITM with different O <sup>+</sup> reaction rates and collision frequencies, lower boundary conditions, equatorial electrojet, and potential solver from those for 1_GITM	25 × 50 × 13
1_TIE-GCM <sup>a</sup>	TIE-GCM1.93 driven by Heelis et al. [1982]	36 × 72 × 29 (~90 km < alt < 500 km)
2_TIE-GCM	TIE-GCM1.94 driven by Weimer [2005] with dynamic critical co-latitudes	36 × 72 × 29 (~90 km < alt < 500 km)
<i>Data Assimilation Models</i>		
1_JPL-GAIM	USC/JPL GAIM with ground based GPS data (−55° < lat < 55°) and COSMIC data	60 × 36 × every 40 km altitude
1_USU-GAIM <sup>a</sup>	USU-GAIM23 with GPS TEC observations from up to 400 ground stations (−60° < lat < 60°)	44 × 24 × 83 (90 km < alt < 1,400 km)

<sup>a</sup>The model results are submitted by the CCMC using the models hosted at CCMC. Different model setups are different model setting identification numbers.

CHAMP locations, and (3) NmF2 and (4) hmF2 derived from radio occultation measurements from CHAMP and COSMIC (Constellation Observing System for Meteorology, Ionosphere, and Climate) satellites during each event.

### 3. Submissions of Model Simulations

[8] For the comparison of neutral density at the CHAMP orbits, eight submissions of model simulations from five models (JB2008, NRLMSISE-00, CTIPe, GITM and TIE-GCM) were used. Nine submissions from eight models (IRI, SAMI3, USU-IFM, CTIPe, TIE-GCM, JPL-GAIM and USU-GAIM) were compared with NmF2, hmF2 and electron density measurements. The model outputs used for the study were either provided by modelers (via the submission interface at the CCMC website developed for a number of model validation studies) or simulated by the IT models hosted at the CCMC [Webb et al., 2009]. A unique model identifier was used to distinguish multiple submissions from the same model that were driven by different boundary conditions and/or inputs (e.g., 1\_JB2008 and 2\_JB2008) (see Table 2). In Table 2, model results generated by the CCMC are identified by a superscript “a.”

[9] We briefly described the models and the submissions of their simulations in Shim et al. [2011] (please refer to all references included) except for two empirical thermosphere models (JB2008, NRLMSISE-00). In the following sections, we provide short descriptions of the two models and their three submissions.

#### 3.1. JB2008 (1\_JB2008 and 2\_JB2008)

[10] Jacchia-Bowman 2008 (JB2008) is an empirical thermospheric density model that is developed as an improved revision to the Jacchia-Bowman 2006 model based on Jacchia’s diffusion equations [Jacchia, 1965, 1971; Bowman et al., 2008a, 2008b, 2008c]. New exospheric temperature and semiannual density equations are employed to represent the major thermospheric density variations. In addition to the F<sub>10.7</sub> index for radio flux, JB2008 and its predecessor use EUV flux from the SOHO satellite, and solar middle ultraviolet (MUV) flux from NOAA satellites. Daily and 81-day centered averages of these solar indices are used in a formula to derive the “global nighttime minimum exospheric temperature,” from which the neutral densities are calculated from the date, time, location, and altitude. JB2008 uses an additional correction to the exospheric temperature, due to geomagnetic activity, that is derived from the Dst index and based on results by Burke [2008]. Another means of calculating the correction to the average exospheric temperature was shown by Weimer et al. [2011]. This alternative method uses the total Poynting fluxes into the polar regions, calculated from empirical models that use the solar wind and IMF measured by the ACE satellite for input values [Weimer, 2005].

[11] Two submissions, 1\_JB2008 and 2\_JB2008, for the neutral density at the CHAMP orbits using JB2008 were used (see Table 2). 1\_JB2008 is JB2008 run with corrections to global nighttime minimum exospheric temperature due to auroral heating computed from the Dst index, while

2\_JB2008 is JB2008 run with the temperature corrections derived from the total Poynting fluxes [Weimer, 2005; Weimer et al., 2011].

### 3.2. NRLMSISE-00 (1\_MSIS)

[12] NRLMSISE-00 (Naval Research Laboratory Mass Spectrometer and Incoherent Scatter Extended) is an empirical model based on the earlier models MSIS-86 [Hedin, 1987] and MSISE-90 [Hedin, 1991]. NRLMSISE-00 and the associated NRLMSIS database include composition and temperature measurements by satellite, rocket, and incoherent scatter radar. They also include total mass density from satellite accelerometers and from orbit determination. The model covers the altitude range from the ground to the exobase (<1400 km), and provides altitude profiles of temperature, number densities of species (He, O, N<sub>2</sub>, O<sub>2</sub>, Ar, H, and N), total mass density, and the number density of a high-altitude anomalous oxygen component of total mass density [Picone et al., 2002]. NRLMSISE-00 accounts for the main drivers of the upper atmosphere: the solar EUV flux and geomagnetic heating. The 10.7-cm solar radio flux ( $F_{10.7}$ ) is the standard proxy for the solar EUV, while the daily  $A_p$  and the 3-hourly  $a_p$  geomagnetic indices measure the geomagnetic component of space weather.

[13] For the study, simulation results of NRLMSISE-00 (identified 1\_MSIS) were obtained from <http://sisko.colorado.edu/sutton/data.html> [Sutton et al., 2005].

## 4. Data

[14] We used the total neutral mass densities inferred from accelerometer measurements on the CHAMP satellite that are available at <http://sisko.colorado.edu/sutton/data.html> [Sutton et al., 2005]. CHAMP was orbiting the Earth with an inclination of 87.3° and taking measurements for ten years since its launch on 15 July 2000 [Reigber et al., 2002]. Due to the high inclination, CHAMP measurements cover almost all latitudes (see bottom plots in Figures 2a and 2b), while all local times are sampled roughly once every four months. The neutral densities are 3-degree latitude average values with a cadence of about 45 s. The absolute uncertainty of the CHAMP neutral densities derived from drag-based measurements is 10~15% [Bruinsma et al., 2004] and includes combined effects of several sources of errors such as accelerometer noise, accelerometer calibration, coefficient of drag, solar radiation pressure model, and neutral winds. The average error in the neutral density for the nine selected events ranges from about  $1.5 \times 10^{-13}$  to  $3 \times 10^{-13}$  kg/m<sup>3</sup>, which corresponds about 6 to 14%, for the altitudes around 340~390 km seen in most of this study. For the study, 1-min average observed values of electron and neutral densities were compared with modeled values of every minute.

[15] The in situ electron densities from the PLP (Planar Langmuir Probe) onboard the CHAMP satellite were used as ground truth. The PLP takes measurements of electron density at the satellite position every 15 s. The accuracy of the PLP measurements is within 10% [Liu et al., 2007]. The

CHAMP PLP data were provided by the Information System and Data Center (ISDC, <http://isdc.gfz-potsdam.de/>).

[16] We also used NmF2 and hmF2 obtained from the electron density profiles (EDPs) retrieved from the CHAMP and COSMIC GPS radio occultation (RO) measurements provided by the University Corporation for Atmospheric Research (UCAR) COSMIC Data Analysis and Archival Center (CDAAC) (<http://cosmic-io.cosmic.ucar.edu/cdaac>) [Schreiner et al., 2002]. The CDAAC radio occultation EDPs retrieved by Abel inversion have uncertainty due to assumptions and approximations used in the inversion method [Schreiner et al., 1999; Lei et al., 2007; Wu et al., 2009]. It was found that the NmF2 and hmF2 obtained with RO measurements and ground-based measurements show an agreement within about 10–30% [Hajj and Romans, 1998; Schreiner et al., 1999; Chu et al., 2010]. It was also found that the Abel retrieval method generally demonstrates good performance except at lower altitudes such as the E and F1 layers [Yue et al., 2010, 2011]. For the E.2001.243 event, there were no data available, and for the E.2005.243 events, only CHAMP measurements were used, since the COSMIC constellation of six satellites was launched on 14 April 2006 [Anthes et al., 2008]. The hmF2 values obtained from the CHAMP/COSMIC RO measurements at tangent points, which are on the line of sight between LEO and GPS satellites, range between about 200 and 350 km in most of the selected events except during the storm's main phase of the E.2006.348 (not shown here). Therefore, the electron densities at the CHAMP height (340~390 km), which is greater than the CHAMP/COSMIC hmF2, are smaller than NmF2 RO measurements for most cases. There is another group of hmF2 values near 100 km (not shown here). For the study, the hmF2 values less than 140 km were excluded that correspond to peak heights of E region. Also, the modeled hmF2 values obtained from EDPs by interpolation and those larger than 140 km were used.

## 5. Metrics

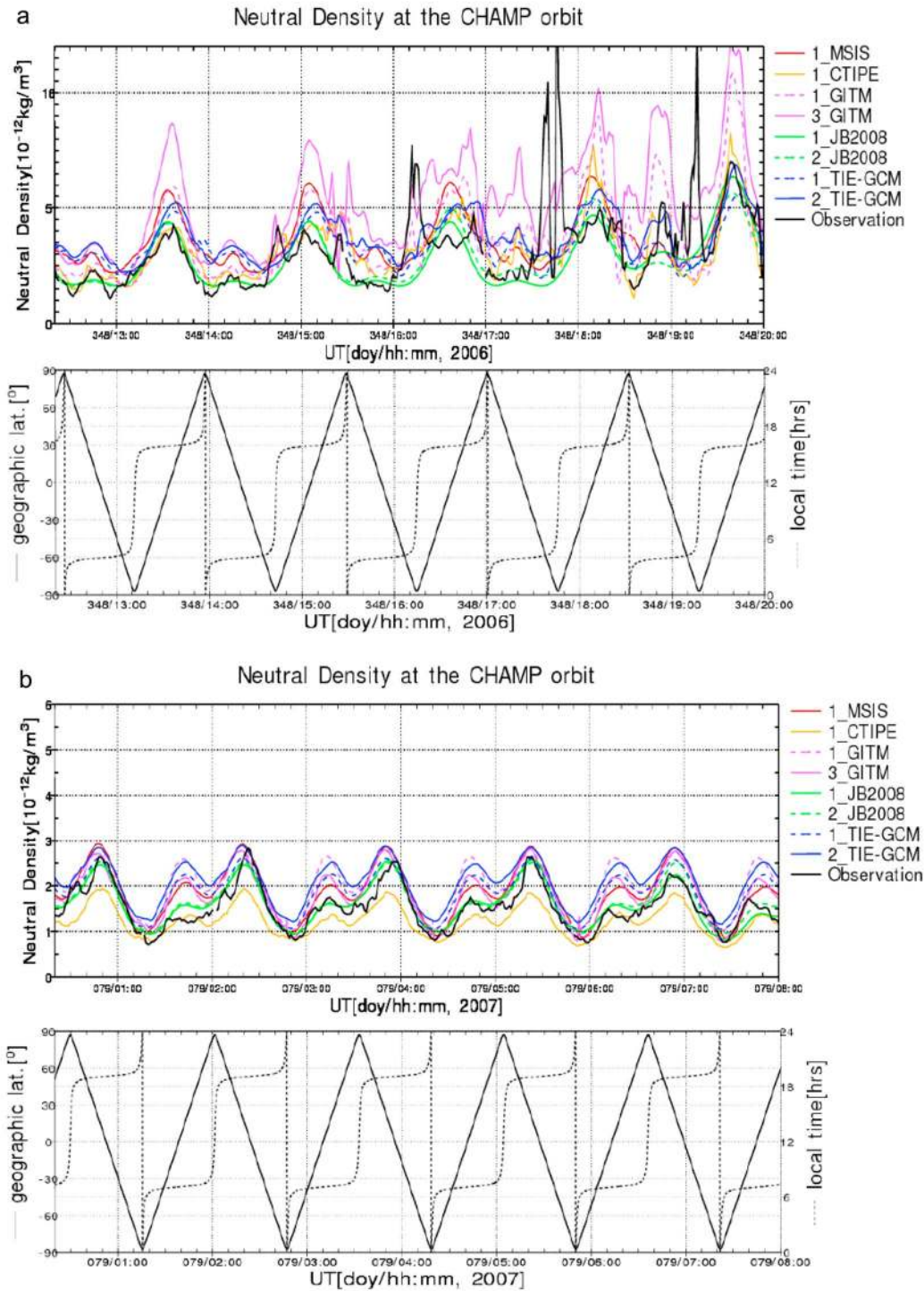
[17] We used four metrics in order to quantify the model accuracy. Here, 'metric' means functions, which give one real number (skill score) for one set of modeled and observed data.

### 5.1. Root-Mean Square (RMS) Difference

[18] For quantitative model assessment, the root-mean square difference is a widely used method to determine the differences between the observed and model values defined as

$$RMS = \sqrt{\frac{\sum (x_{obs} - x_{mod})^2}{N}}$$

where  $x_{obs}$  and  $x_{mod}$  are values obtained from observation and model prediction, respectively. RMS errors of 0 indicate perfect agreement of modeled values with observations, hence the RMS error approaches 0 as the model



**Figure 2.** Examples of the observed and modeled neutral densities at the CHAMP orbit (a) for the E.2006.348 strong storm event, and (b) for the E.2007.079 quiet time event. In the top plots in Figures 2a and 2b, the black curves represent observation data, and the colored curves correspond to modeled values every 1-min. The bottom plots show the CHAMP orbit track as a function of local time (dashed lines) and latitude (solid lines).

prediction gets better. Note that the unit of the RMS error is the same as the unit of observed and modeled values.

## 5.2. Prediction Efficiency (PE)

[19] Prediction efficiency against the mean observed value is also used for assessment of models:

$$PE = 1 - \frac{RMS_{mod}}{RMS_{ref}} = 1 - \sqrt{\frac{\sum (x_{obs} - x_{mod})^2 / N}{\sum (x_{obs} - \langle x_{obs} \rangle)^2 / N}}$$

where  $x_{obs}$  and  $x_{mod}$  are again values from observation and model prediction, respectively, and  $\langle x_{obs} \rangle$  is the mean value of the observed data. In this study, we used the observed mean value  $\langle x_{obs} \rangle$  as a reference model rather than any empirical model. PE can vary from negative infinity to 1 (PE = 1 means perfect prediction). A value of 0 means that the model accuracy is comparable to the variation of the measurements about their mean in an aggregate sense. Negative values indicate that model error in an RMS sense is larger than the variation of the observations about their mean and imply that the observed mean is a better predictor of the observations than the model. Local time dependence of the selected physical parameters (electron and neutral densities, NmF2 and hmF2) was considered by calculating PE for the daytime (06:00–18:00 LT) and the nighttime (18:00–06:00 LT) separately using daytime and nighttime mean values of observed data. Physical conditions corresponding to less IT variability (e.g., quiet time) require corresponding increases in model accuracy to achieve comparable skill scores to more variable times (e.g., storms).

## 5.3. Ratios of the Maximum Change in Amplitudes and Maximum Amplitudes

[20] We also used metrics based on ratio in order to measure capability of models to predict maximum amplitudes or short-term temporal changes during a certain time interval, although accuracy of models may be low with respect to the RMS error and/or PE that measure how well modeled values are correlated with observed values. Two ratios were considered: the ratio of the maximum change (max – min) and the ratio of the maximum (max) modeled values to observed values:

$$ratio(\max - \min) = \frac{(x_{mod})_{\max} - (x_{mod})_{\min}}{(x_{obs})_{\max} - (x_{obs})_{\min}},$$

$$ratio(\max) = \frac{(x_{mod})_{\max}}{(x_{obs})_{\max}}$$

where  $(x_{obs})_{\max}$  and  $(x_{mod})_{\max}$  are the observed and modeled maximum amplitudes for a certain time window. A ratio of 1 indicates perfect model prediction, while the ratio(max – min) and the ratio(max) greater than 1 suggest overestimation of maximum variations and maximum values by models.

[21] In *Shim et al.* [2011], it was shown that selecting an appropriate time window length is crucial in calculating the two ratios, which is dependent on the time window length. It was found for this study that a suitable length of time window is 1 h, compared to 4 and 7 h for calculating the ratios of the vertical drifts at Jicamarca. The ratios of max – min of the model, which performed the best in predicting the vertical drift variability, move away from 1 as the time window length increases from 1 to 7, while the opposite holds true for the models that performed worse.

[22] For this study, we selected a 90-min time window length that is close to the period of the CHAMP (about 94 min) and COSMIC (about 100 min) satellites. Due to the daytime and nighttime alternation during the 90 min, the ratio(max – min) represents the ratio of the diurnal variation (difference between daytime maxima and nighttime minima).

## 6. Results

### 6.1. Neutral Density

[23] Figure 2 displays the observed and modeled neutral mass densities along the CHAMP track during the first eight hours for the E.2006.348 strong storm event (Figure 2a) and for the E.2007.079 quiet time event (Figure 2b). Top plots in Figures 2a and 2b use black and colored curves to show observation data and modeled values. In the bottom plots, the CHAMP orbit track is shown as a function of local time (dashed lines) and latitude (solid lines).

[24] In Figure 2, during the E.2007.079 quiet time event and before storm onset for the E.2006.348 event, most models produce diurnal and latitudinal variations of neutral mass density similar to those observed, however, differences in the model performance are clearly seen. For the E.2007.079 event, the results from the empirical model JB2008, 1\_JB2008 and 2\_JB2008 are almost identical and agree with the measurements better than the results from the others, which tend to underestimate (e.g., CTIPE) or overestimate (e.g., NRLMSISE-00, GITM, and TIE-GCM) the neutral density at the CHAMP orbit. During the E.2006.348 event, none of the models succeed in producing the observed abrupt increases in neutral density in the morning sector at high latitudes in the southern hemisphere, although both GITM models show some increases. However, from Figure 2, only limited qualitative conclusions can be obtained.

[25] The model performances were quantified by using metrics to make explicit comparisons. Figure 3 shows the ranking of eight model simulations using four different metrics for neutral density along the CHAMP track; RMS error, PE, ratio(max – min), and ratio(max) (from top to bottom). To find out model performance dependency on latitude, the skill scores using the four metrics were calculated for three latitude regions, which are low ( $|\text{lat}| < 25^\circ$ ), middle ( $25^\circ < |\text{lat}| < 50^\circ$ ), and high ( $|\text{lat}| > 50^\circ$ ) geographic latitudes (from left to right in Figure 3). In Figure 3, squares, circles, and triangles indicate the average values for strong storms, moderate storms, and quiet periods, respectively. Ranking of the model performance is based on the multievent

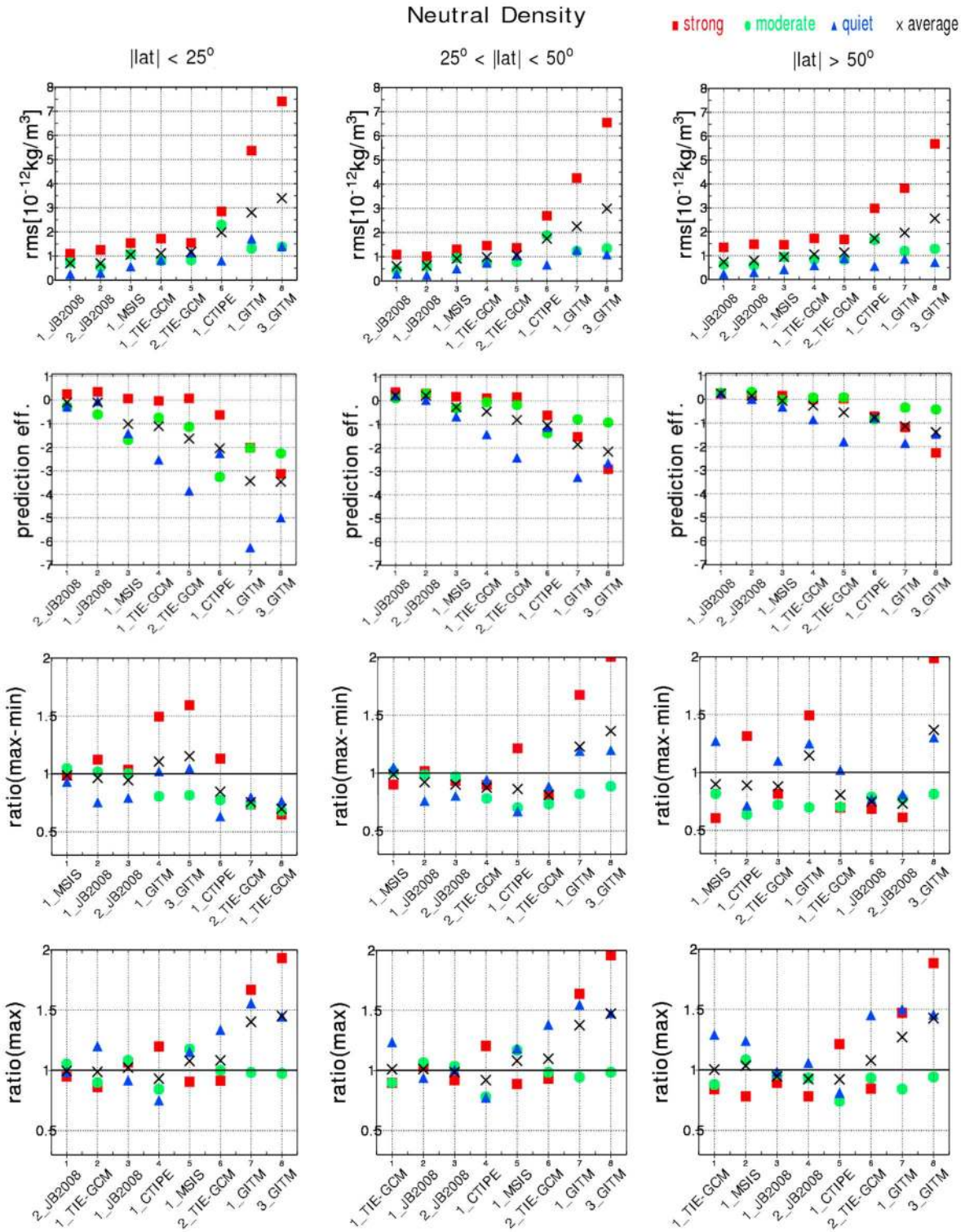


Figure 3

average (denoted by crosses) of the three geomagnetic activity levels. The model that performs best is placed in the extreme left. In the bottom two panels, the models nearer the thin black horizontal line (ratio = 1) perform better than the others located farther above (below) the line that overestimate (underestimate) the maximum changes and/or maximum values.

[26] The RMS differences for the neutral densities seem to get larger with increasing geomagnetic activity. During the strong storms, all models show the largest RMS error. Highly ranked models including 1\_JB2008, 2\_JB2008, 1\_MSIS, TIE-GCMs (1\_TIE-GCM and 2\_TIE-GCM), show similar differences in RMS error between strong storms and quiet periods that are about  $1 \times 10^{-12} \text{ kg/m}^3 \sim 2 \times 10^{-12} \text{ kg/m}^3$ . However, relatively low ranked models such as 1\_CTIPE, GITMs (1\_GITM and 3\_GITM) show larger differences up to about  $6 \times 10^{-12} \text{ kg/m}^3$ . PE also shows dependency on geomagnetic activity. Most models tend to have better PE during storms than quiet times in all latitudes.

[27] The two submissions, 1\_JB2008 and 2\_JB2008, rank at the top and are followed by another empirical model 1\_MSIS. The model rankings based on the RMS error and PE are similar but not the same (see first and second rows in Figure 3) due to the fact that the ranking in terms of PE is obtained by using the average of daytime and nighttime PE, while the RMS error does not depend on local time. In addition, PE is normalized by the standard deviation of the observations, whereas the RMS error is not normalized.

[28] Compared to the three empirical model results, TIE-GCMs show rather worse performance, but they show comparable performance in terms of RMS. 1\_CTIPE produces worse RMS differences and PE than TIE-GCMs for the storm events, although the two scores of 1\_CTIPE are better than theirs for the quiet events. GITMs perform worse than the other models especially during the strong storm events. 3\_GITM shows the largest RMS errors and negative PE during the strong storms. However, for moderate storms, GITMs show comparable performance to TIE-GCMs, and better performance than 1\_CTIPE.

[29] 1\_MSIS shows the best ratio(max – min) in low and middle latitudes for all geomagnetic levels. 2\_JB2008 and 1\_JB2008 show the best ratio(max) in low and middle/high latitudes, respectively, although 1\_TIE-GCM has a ratio(max) closest to 1 as a result of the counterbalance of the overestimation during quiet periods and underestimation during storm events in an average sense. As for the ratios, therefore, the model ranking for each geomagnetic condition is more focused than the ranking based on the

average over the all three geomagnetic conditions in this paper.

[30] The two JB2008s, for the quiet periods in the three latitude regions, produce the ratio(max) close to 1 and the ratio(max – min) smaller than 1. This indicates that the JB2008s predict well daytime maximum neutral density, but overestimate nighttime minimum values. The two GITMs, for moderate storms in all latitudes, show the same features. In terms of ratios, during the strong storms, among the physics-based coupled IT models, TIE-GCMs, which tend to underestimate ratios, are better than 1\_CTIPE and GITMs, which overestimate ratios. However, during the moderate storm events, the coupled models produce similar ratios (less than 1).

## 6.2. Electron Density

[31] Figure 4 displays the observed and modeled electron densities at the CHAMP orbit during the first five hours for the E.2005.243 strong storm event (Figure 4a) and for the E.2007.079 quiet time event (Figure 4b). In the top plots in Figures 4a and 4b, the black and colored curves represent values from observations and models, respectively. The bottom plots in Figures 4a and 4b show the CHAMP orbit track as a function of local time (dashed lines) and latitude (solid lines). During the E.2007.079 quiet time event, 1\_GITM tends to overestimate electron densities for almost all latitudes and all local times, while most of the other models show a similar tendency to overestimate with a small peak at low latitudes in the morning sector (Figure 4b). Some models produce the ionospheric equatorial anomaly better than the others (not shown clearly here). In higher latitudes, the observed values show better agreements with the results of the empirical model 1\_IRI and physics-based coupled model such as 1\_CTIPE and 2\_TIE-GCM than the results of the physics-based ionosphere models and the data assimilation models, which have limited latitude coverage. It is also shown that, in middle latitude regions, most models produce relatively well the electron densities during the quiet time, while the simulation results of most models noticeably differ from the measurements during the storm.

[32] Figure 5 shows the ranking of the models based on the four skill scores for electron density at the CHAMP orbit. Figure 5 is the same as Figure 3 but for predicating electron density. It should be noted 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

**Figure 3.** Model ranking for predicting neutral density along the CHAMP trajectory based on RMS error (first row), prediction efficiency (second row), ratio(max – min) (third row), and ratio(max) (fourth row) for low (first column), middle (second column), and high (third column) latitudes. Squares denote the average values for strong storm cases (E.2005.243 and E.2006.348), circles and triangles indicate the average of moderate storms (E.2001.243, E.2007.091, E.2007.142, and E.2008.059) and the average of quiet periods (E.2007.079, E.2007.190, and E.2007.341). Crosses are the average of the all three geomagnetic activity levels, and the ranking of the performance of the models are arranged by the final average values. The best performing model is located in the extreme left.



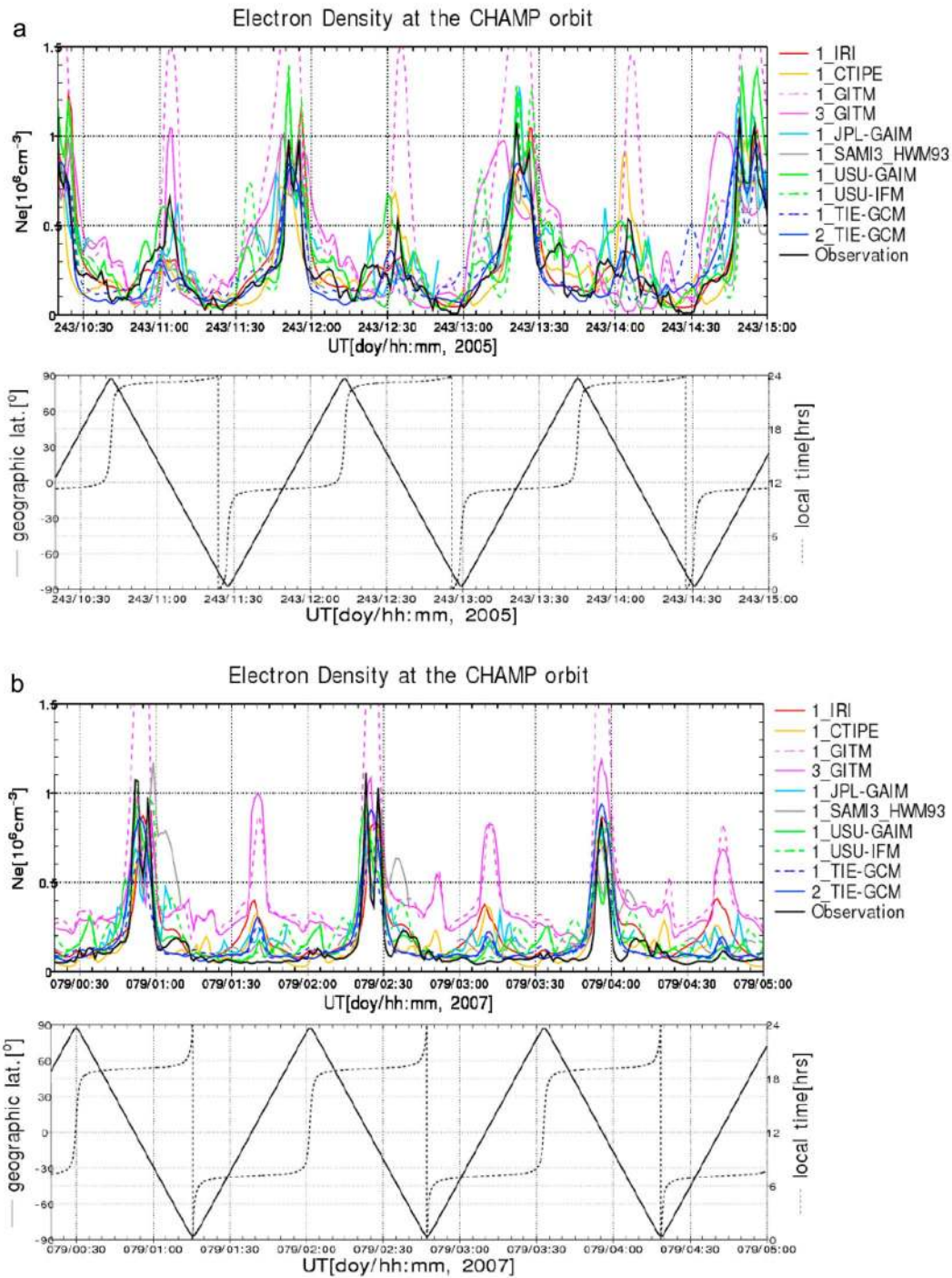


Figure 4. Same as Figure 2 but for electron density along the CHAMP trajectory and for E.2005.243 strong storm event (Figure 4a).

electric field pattern), and 1\_GITM has data only for one strong (E.2005.243), two moderate storm events (E.2007.091 and E.2007.142), and two events (E.2007.079 and E.2007.190) during quiet conditions. Therefore, comparing 1\_GITM

with other models based on the averaged performance taken over the events needs caution.

[33] All models, except GITMs in low latitudes, show the RMS increases due to increases of geomagnetic activity from low (quiet period) to medium (moderate storm) and

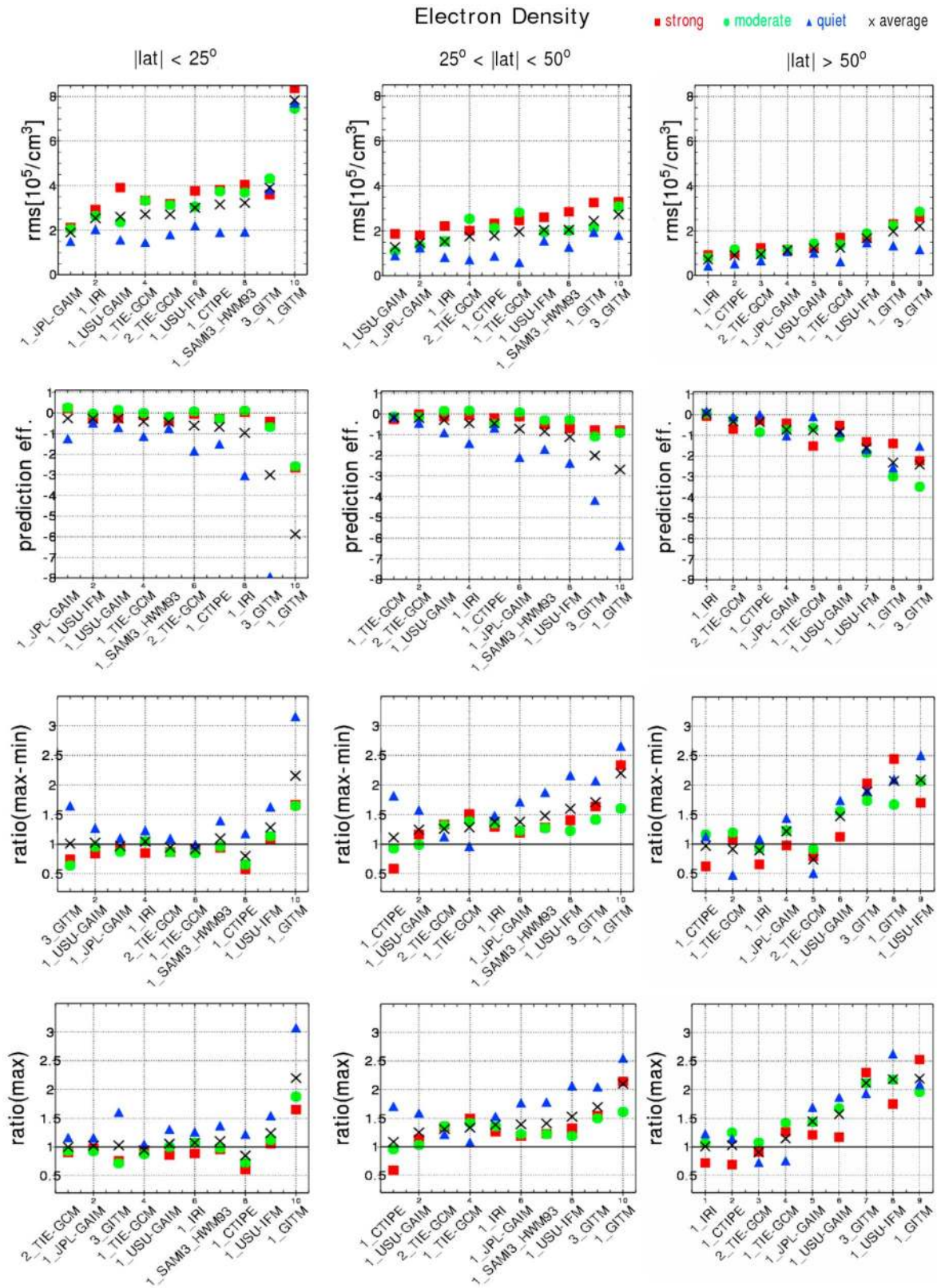


Figure 5. Same as Figure 3 but for predicting electron density along the CHAMP trajectory.

to high (strong storm) levels. For example, the RMS increase reaches up to about  $2.5 \times 10^5/\text{cm}^3$ . However, a change of geomagnetic activity from medium to high level appears to produce smaller increases in the RMS error than the above mentioned increases in geomagnetic activity especially in high latitudes. All models, except GITMs in low latitudes, show best performance during the quiet periods in terms of RMS error. Although PE and ratios show less systematic dependency on geomagnetic activity than the RMS error does, most models tend to have better PE and ratios during storms than quiet times in low and middle latitudes.

[34] In terms of the both RMS error and PE, 1\_JPL-GAIM, 1\_USU-GAIM, and 1\_IRI rank at the top in low, middle, and high latitudes, respectively, and they are followed by coupled and physics based models. In low and middle latitudes most models produce better PE (near 0) during the storms than during quiet time periods as mentioned above. However, at high latitudes, all models have better (worse) PE than in low and middle latitudes during the quiet period (storms), which results in reduced differences in PE between geomagnetic activity levels. For data assimilation models, 1\_JPL-GAIM and 1\_USU-GAIM, the inconsistency in ranking at different latitudes (ranking near middle at high latitudes and higher in lower latitudes) is possibly due to limited data assimilation. For this study, 1\_JPL-GAIM and 1\_USU-GAIM submissions were obtained using ground based GPS TEC data between  $\pm 55^\circ$  geomagnetic, and between  $\pm 60^\circ$  geographic latitudes, respectively.

[35] In terms of ratios, in low latitudes, 1\_JPL-GAIM and TIE-GCMs show good agreements with observations for the all three geomagnetic conditions. In middle latitudes, 1\_TIE-GCM shows the best performance in producing diurnal variations and maximum values of electron density for the quiet periods, while 1\_USU-GAIM ranks at the top for storm cases. At high latitudes, 1\_IRI and 1\_CTIPE produce better ratios for quiet and moderate conditions, whereas TIE-GCMs, 1\_JPL-GAIM and 1\_USU-GAIM show better ratios during the strong storms than the others.

[36] For quiet conditions, most models tend to overestimate diurnal variations and the daytime maximum of electron density in all three latitude regions, except for TIE-GCMs in high latitudes. In low and middle latitudes, 1\_GITM, 3\_GITM, 1\_CTIPE and 1\_USU-IFM show relatively larger differences in the ratios between the three levels of geomagnetic activity than the others.

[37] For predicting electron density, data assimilation models and the IRI empirical model show better scores than physics-based ionosphere and coupled IT models in low and middle latitudes especially in terms of RMS difference and PE. In high latitudes, the IRI empirical model and physics-based coupled IT models rank higher than the others.

[38] 1\_SAMI3-HWM93 and 1\_USU-IFM, which are physics-based ionospheric models, produce comparable performance in terms of RMS error and PE, although 1\_USU-IFM produces slightly smaller RMS error during the storms, and 1\_SAMI3-HWM93 shows relatively better

agreement with observations during the quiet periods, and it produces better diurnal variations and maximum values of electron density than 1\_USU-IFM for most cases.

[39] TIE-GCMs and 1\_CTIPE among five coupled model submissions show similar performance in producing electron density along the CHAMP tracks, and they show better performance than the other two, GITMs, for all cases. 2\_TIE-GCM is slightly better for the storms and worse for the quiet events than 1\_TIE-GCM. 1\_CTIPE performs worse than TIE-GCMs in low latitudes, while 1\_CTIPE is better for moderate storms in middle latitudes and for the quiet periods in high latitudes.

[40] Differences between performances of the two data assimilation models, 1\_JPL-GAIM and 1\_USU-GAIM, are hardly seen except for larger RMS error of 1\_USU-GAIM for strong storm in low latitudes and larger negative PE of 1\_JPL-GAIM for quiet events in low and middle latitudes.

### 6.3. NmF2 and hmF2

[41] Figures 6 and 7 show the model performance of predicting NmF2 and hmF2, respectively. The model scores are the averaged performance taken over the CHAMP and six COSMIC satellite measurements. In Figure 6, similar to electron density and neutral density, most models produce increasing RMS errors as geomagnetic activity grows with a few exceptions (e.g., 1\_JPL-GAIM and 1\_CTIPE for NmF2 predicting at high latitudes). Differences in the RMS error of NmF2 between geomagnetic activity levels decrease as latitude increases. For example, RMS error differences between low and high geomagnetic activity levels for 2\_TIE-GCM is about  $4 \times 10^5/\text{cm}^3$  at low latitudes, it decreases to about  $1.5 \times 10^5/\text{cm}^3$  at high latitudes. In terms of RMS error and PE, 1\_JPL-GAIM and 1\_IRI rank highly at all three latitude regions and for all three geomagnetic activity levels. 1\_JPL-GAIM is better than 1\_USU-GAIM in predicting NmF2 at low and high latitudes, while the two models show similar performance at middle latitudes based on RMS error and PE. 1\_JPL-GAIM also shows better agreement with NmF2 observations than 1\_USU-GAIM in terms of ratios. 1\_USU-GAIM has ratios greater than 1 for all cases, while 1\_JPL-GAIM tends to underestimate the ratios at high latitudes. 1\_SAMI3-HWM93 and 1\_USU-IFM produce similar scores, but 1\_SAMI3-HWM93 is slightly better than 1\_USU-IFM during the quiet periods in terms of RMS error and PE. Among the coupled physics based IT models, in terms of RMS and PE, the two TIE-GCMs are better than 3\_GITM and 1\_CTIPE, although 2\_TIE-GCM shows the worst performance in predicting NmF2 at low latitudes during the strong storms, while 3\_GITM is better than 1\_CTIPE in low and middle latitudes. In terms of ratios, for quiet conditions, 1\_CTIPE shows better ratios for most cases than the others, however during strong storms it has the worst ratios (less than 1) in middle latitudes. During strong storms, 2\_TIE-GCM produces better ratios in middle and high latitudes than 1\_TIE-GCM, while the opposite is true in low latitudes. 3\_GITM shows the best ratio(max) in middle latitudes during storms.

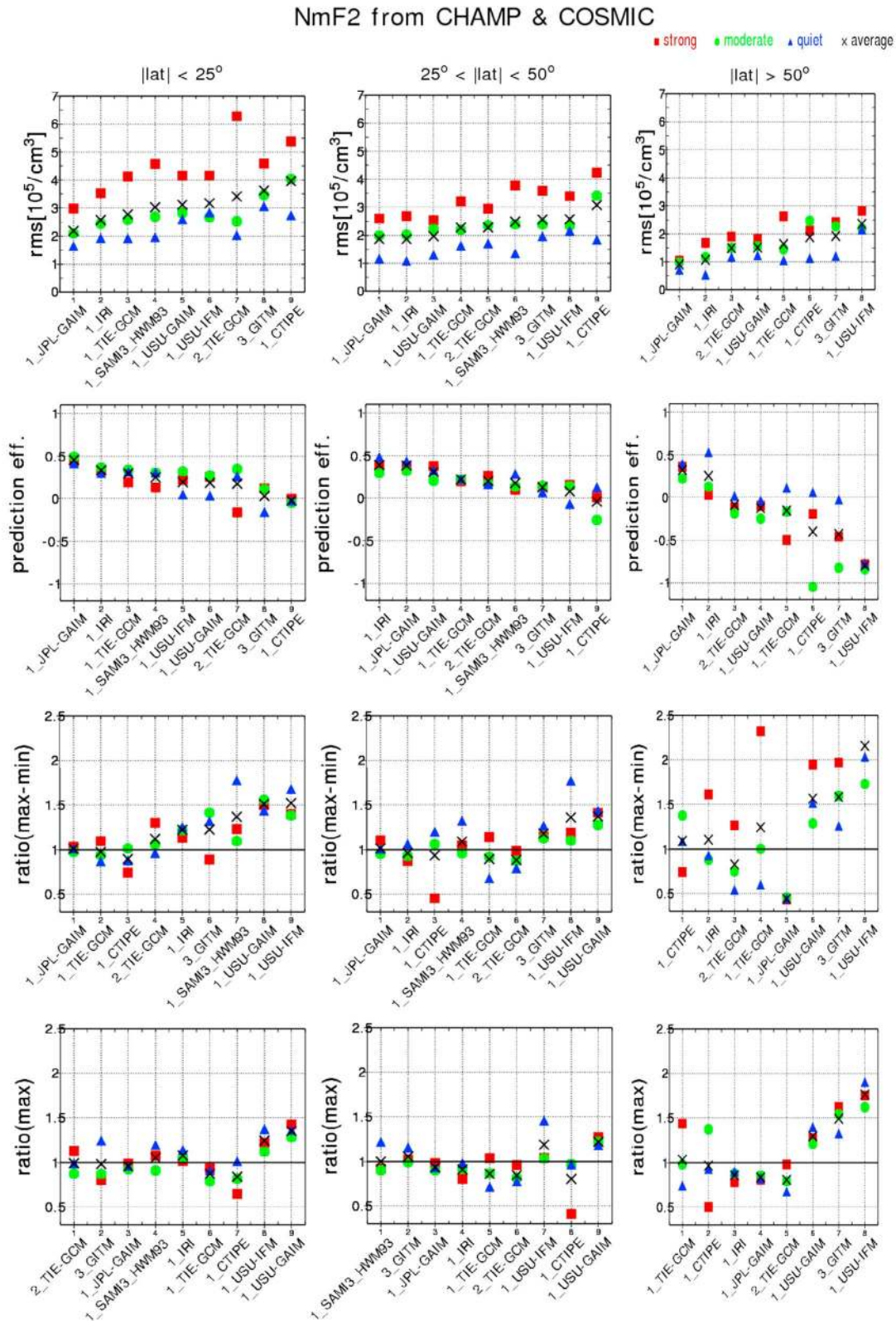


Figure 6. Same as Figure 3 but for predicting NmF2 from the CHAMP and COSMIC GPS RO measurements.

hmF2 from CHAMP & COSMIC

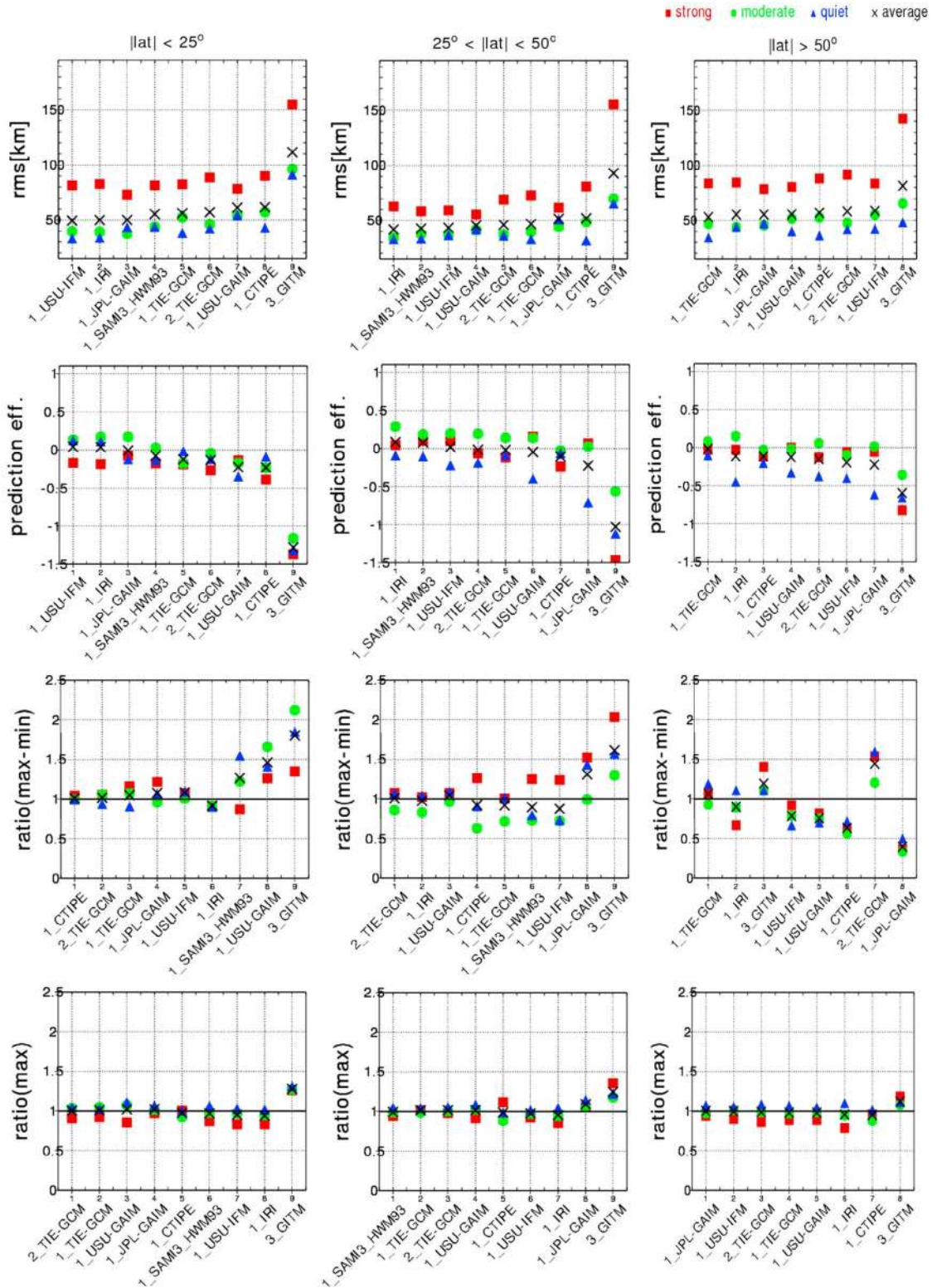


Figure 7. Same as Figure 3 but for predicting hmF2 from the CHAMP and COSMIC GPS RO measurements.

[42] As shown in Figure 7, RMS errors in predicting hmF2 tend to increase as geomagnetic activity increases like RMS errors in predicting NmF2. All models produce the largest RMS errors during the strong storms in all three latitude regions, however, the errors in middle latitudes are less than those in low and high latitudes. RMS error differences between quiet and moderate storm conditions are smaller than the difference between moderate and strong storms. Most models produce similar average RMS errors except for 3\_GITM, which has the largest RMS errors for most cases. Most models produce better PE during moderate storms than during quiet periods and strong storms, although the differences are not significant. Highly ranked models in terms of the ratio(max – min) in low and middle latitudes show similar performance for all geomagnetic conditions. Models show good agreement with observed ratios of max for all cases. 1\_SAMI3\_HWM93 and 1\_USU-IFM hardly show differences in the skill scores, however, 1\_USU-IFM shows better ratio(max – min) in low latitudes. 1\_JPL-GAIM appears slightly better (worse) than 1\_USU-GAIM in predicting hmF2 at low (middle) latitudes, while they show similar performance at high latitude based on RMS error and PE. 1\_JPL-GAIM also shows better agreement with hmF2 observations at low latitudes than 1\_USU-GAIM in terms of ratios, however, 1-USU-GAIM has better ratio(max – min) in middle and high latitudes. For most cases, two TIE-GCMs and 1\_CTIPe perform similarly and better than 3\_GITM, although 2\_TIE-GCM shows worse ratio(max – min) at high latitudes than 1\_CTIPe and 3\_GITM.

## 7. Discussion and Conclusions

[43] We quantified the accuracy of various Ionosphere/Thermosphere (IT) models in predicting electron density, neutral density, NmF2 and hmF2 against space-based measurements using four different metrics, which are RMS error, prediction efficiency (PE), ratio of the maximum change in amplitudes, and ratio of maximum amplitudes within a 90 min time window. In addition, dependence of the model performance on geomagnetic activity and latitude was investigated by calculating the four skill scores for three latitude regions, low, middle, and high geographic latitudes, during the selected nine time intervals. The nine events were binned into three geomagnetic levels by maximum value of Kp during the time interval. Measurements used as ground truth are electron and neutral densities obtained from CHAMP, and NmF2 and hmF2 derived from radio occultation measurements by the CHAMP and COSMIC satellites. The average value of skill scores over all nine events was used to rank the model performance.

[44] Our study indicates that the model accuracy depends on geomagnetic activity. The RMS errors increase with increase of geomagnetic activity for most cases, although RMS errors of a few models during quiet periods are slightly larger or similar to those during storms. This nonlinearity of the exceptional cases was also found in our earlier paper on IT model evaluation using ground-based observations [Shim

*et al.*, 2011]. Possible causes of the nonlinearity are the simultaneous changes in the external driving forces (for example, neutral wind, composition and temperature and electric fields), which vary with each storm, depending on the magnetospheric energy input into the ionosphere. The prediction efficiency and ratios also show dependency on geomagnetic activity, however, they show less systematic (and even opposite) dependency than RMS error does. For predicting electron and neutral densities at the CHAMP locations, most models tend to have better PE and ratios during storms than quiet times in low and middle latitudes. This is likely due to the increased observed variability during storms that tends to produce a higher skill score for PE. Also, for predicting hmF2, the same feature is seen in middle and high latitudes. More than half of the models tend to produce better ratios of electron density during the storm rather than the quiet time intervals, and the ratio(max) of hmF2 hardly depends on geomagnetic activity in all three latitudes. Most of models ranked higher tend to produce smaller differences in the skill scores among the different geomagnetic conditions.

[45] The results of the study also indicate that model accuracy varies with the type of metrics and latitude. For example, in the predictions of electron density at the CHAMP track, the ionospheric empirical model 1\_IRI and data assimilation models, 1\_JPL-GAIM and 1\_USU-GAIM rank at or near the top in terms of RMS error, while data assimilation models and coupled IT models rank higher in terms of ratio(max – min) especially during storms. The data assimilation models are ranked top with respect to RMS in low and middle latitudes. However, they perform worse than 1\_IRI in high latitudes probably due to the limitation of latitude coverage of simulation results and observations used for the data assimilation models. The submission of 1\_JPL\_GAIM for this study was generated by assimilating ground based GPS TEC measurements from about 200 stations located between  $\pm 55^\circ$  geomagnetic latitude, although COSMIC TEC measurements were also assimilated. The 1\_USU-GAIM used for the study only assimilates GPS TEC measurements between  $\pm 60^\circ$  geographic latitude, thus, high-latitude electron densities are the same as those from the physics based model USU-IFM, which is used as a background, without assimilating any data. For reproducing hmF2 from CHAMP and COSMIC, two physics based models, 1\_USU-IFM and 1\_SAMI3\_HWM93, and 1\_IRI performed better than others in terms of RMS and PE in low and middle latitudes, although the performance differences among the models are not significant. For thermosphere neutral density predictions, three empirical model submissions, 1\_JB2008, 2\_JB2008 and 1\_MSIS rank higher for almost all cases even during the storms. 1\_JB2008 and 2\_JB2008, however, have better RMS and PE, and worse ratios of max – min than 1\_MSIS.

[46] In addition, by comparing the same types of models, we find that two physics-based ionospheric models, 1\_SAMI3\_HWM93 and 1\_USU-IFM perform similarly in general. Although 1\_USU-IFM performs slightly better in predicting electron density during the storms in terms of

RMS error, and 1\_SAMI3\_HWM93 produces relatively better diurnal variations (max – min) and maximum values of electron density. Among the five physics based coupled IT models, it appears that, for most cases, performance of 2\_TIE-GCM, 1\_TIE-GCM and 1\_CTIPE are similar to each other and better than GITMs, except the GITM models were the only models to even come close to reproducing the enhanced neutral densities in storms at high southern latitudes (see Figure 2a). For reproducing of electron density at the CHAMP orbit, 1\_CTIPE performs worse than TIE-GCMs (1\_TIE-GCM and 2\_TIE-GCM) in low latitudes, while 1\_CTIPE is better in middle latitudes for moderate conditions and in high latitudes for the quiet conditions. In reproducing the neutral density at the CHAMP orbit, 1\_CTIPE produces worse (better) RMS and PE than TIE-GCMs for the storm events (quiet events). 3\_GITM shows the largest RMS error and negative PE during the strong storms. However, for moderate storms, GITMs show comparable performance to TIE-GCMs, and perform better than 1\_CTIPE. 2\_TIE-GCM shows worst performance in predicting NmF2 at low latitudes during the strong storms in terms of RMS and PE. In general, two data assimilation models, 1\_JPL-GAIM and 1\_USU-GAIM show similar performance. However, for electron density prediction, 1\_USU-GAIM has larger RMS errors for strong storms in low latitudes and 1\_JPL-GAIM has larger negative PE for the quiet events in middle latitudes. 1\_JPL-GAIM appears slightly better (worse) than 1\_USU-GAIM in predicting hmF2 at low (middle) latitudes. However, at high latitudes, 1\_USU-GAIM has better ratio(max – min). The two submissions, 1\_JB2008 and 2\_JB2008, using the thermospheric empirical model JB2008 are slightly better than 1\_MSIS in terms of RMS and PE. 1\_MSIS shows the best ratio(max – min) in low and middle latitudes for all geomagnetic levels. 2\_JB2008 and 1\_JB2008 show the best ratio(max) in low and middle/high latitudes, respectively.

[47] It is worth pointing out improvements of model performance caused by enhanced and/or more complex input drivers. 2\_TIE-GCM (driven by Weimer high-latitude electric potential with dynamic critical crossover latitudes) is better than 1\_TIE-GCM (driven by Heelis high-latitude electric potential with constant critical crossover latitudes) in electron and neutral densities predicted during the storms for all latitudes. Systematic improvements in 2\_TIE-GCM, however, are not seen for quiet and moderate storm conditions. The improvement of 2\_TIE-GCM in predicting ionospheric parameters during the strong storm was also shown in Shim et al. [2011]. From the comparison of 1\_GITM and 3\_GITM, it is found that 3\_GITM performs better for electron density, whereas 1\_GITM performs better for neutral density. The differences in input and boundary conditions between 1\_GITM and 3\_GITM (see Table 2) produce the different performance of GITM. Two JB2008 runs for neutral density, 1\_JB2008 (with exospheric temperature corrections derived from the Dst index) and 2\_JB2008 (with the temperature corrections derived from Weimer [2005] total Poynting fluxes), show slight differences in skill scores. 2\_JB2008

shows better scores during moderate storm events for most cases, while 1\_JB2008 is better for the strong storms. Although not all performance caused by enhanced and/or more complex input drivers shows systematic improvement, the results of the comparison will help model improvement.

[48] Furthermore, the results of this systematic assessment of IT models provide a baseline for future validation studies using new models and improved models. Such assessments also provide a basis for understanding the role of data to improve assimilative models, and can suggest what observational systems are most useful for improving IT specifications and forecasts. All measurements and model simulation results used for the challenge are available on the CCMC website (<http://ccmc.gsfc.nasa.gov>) for use by the space science communities.

[49] **Acknowledgments.** The CHAMP neutral density data used in this study are obtained from <http://sisko.colorado.edu/sutton/data.html>. Portions of this research were performed at the Jet Propulsion Laboratory, California Institute of Technology under contract with the National Aeronautics and Space Administration.

## References

- American Institute of Aeronautics and Astronautics (2010), Guide to reference and standard atmosphere models, *Rep. G-003C-2010e*, Reston, Va.
- Anthes, R. A., et al. (2008), The COSMIC/FORMOSAT-3 mission: Early results, *Bull. Am. Meteorol. Soc.*, 89, 313–333, doi:10.1175/BAMS-89-3-313.
- Bowman, B. R., W. K. Tobiska, F. A. Marcos, and C. Valladares (2008a), The JB2006 empirical thermospheric density model, *J. Atmos. Sol. Terr. Phys.*, 70, 774–793, doi:10.1016/j.jastp.2007.10.002.
- Bowman, B. R., W. K. Tobiska, F. A. Marcos, C. Y. Huang, C. S. Lin, and W. J. Burke (2008b), A new empirical thermospheric density model JB2008 using new solar and geomagnetic indices, paper presented at AIAA/AAS Astrodynamics Specialist Conference, Am. Inst. of Aeronaut. and Astronaut., Honolulu, Hawaii.
- Bowman, B. R., W. K. Tobiska, and M. J. Kendra (2008c), The thermospheric semiannual density response to solar EUV heating, *J. Atmos. Sol. Terr. Phys.*, 70, 1482–1496, doi:10.1016/j.jastp.2008.04.020.
- Bruinsma, S., D. Tamagnan, and R. Biancale (2004), Atmospheric densities derived from CHAMP/STAR accelerometer observations, *Planet. Space Sci.*, 52, 297–312, doi:10.1016/j.pss.2003.11.004.
- Burke, W. J. (2008), Stormtime energy budgets of the global thermosphere, in *Mid-Latitude Ionospheric Dynamics and Disturbances*, *Geophys. Monogr. Ser.*, vol. 181, edited by P. M. Kintner et al., pp. 235–246, AGU, Washington, D. C., doi:10.1029/181GM21.
- Chu, Y. H., C. L. Su, and H. T. Ko (2010), A global survey of COSMIC ionospheric peak electron density and its height: A comparison with ground-based ionosonde measurements, *Adv. Space Res.*, 46(4), 431–439, doi:10.1016/j.asr.2009.10.014.
- Hajj, G. A., and L. J. Romans (1998), Ionospheric electron density profiles obtained with the Global Positioning System: Results from the GPS/MET experiment, *Radio Sci.*, 33(1), 175–190, doi:10.1029/97RS03183.
- Hedin, A. E. (1987), MSIS-86 thermospheric model, *J. Geophys. Res.*, 92, 4649–4662, doi:10.1029/JA092iA05p04649.
- Hedin, A. E. (1991), Extension of the MSIS thermosphere model into the middle and lower atmosphere, *J. Geophys. Res.*, 96, 1159–1172, doi:10.1029/90JA02125.
- Hedin, A. E., et al. (1996), Empirical wind model for the upper, middle, and lower atmosphere, *J. Atmos. Sol. Terr. Phys.*, 58, 1421–1447, doi:10.1016/0021-9169(95)00122-0.
- Heelis, R. A., J. K. Lowell, and R. W. Spiro (1982), A model of the high-latitude ionospheric convection pattern, *J. Geophys. Res.*, 87, 6339–6345.

- Jacchia, L. G. (1965), Static diffusion models of the upper atmosphere with empirical temperature profiles, *Smithson. Contrib. Astrophys.*, *8*, 215–257.
- Jacchia, L. G. (1971), Semiannual variation in the heterosphere: A reappraisal, *J. Geophys. Res.*, *76*, 4602–4607, doi:10.1029/JA076i019p04602.
- Lei, J., et al. (2007), Comparison of COSMIC ionospheric measurements with ground-based observations and model predictions: Preliminary results, *J. Geophys. Res.*, *112*, A07308, doi:10.1029/2006JA012240.
- Liu, H., C. Stolle, and S. Watanabe (2007), Evaluation of the IRI model using CHAMP observations in polar and equatorial regions, *Adv. Space Res.*, *39*, 904–909, doi:10.1016/j.asr.2006.08.006.
- 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*(A12), 1468, 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.
- 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.
- Reigber, C., H. Lühr, and P. Schwintzer (2002), CHAMP mission status and perspectives, *Eos Trans. AGU*, *81*, 48.
- Schreiner, W. S., S. V. Sokolovskiy, C. Rocken, and D. C. Hunt (1999), Analysis and validation of GPS/MET radio occultation data in the ionosphere, *Radio Sci.*, *34*, 949–966, doi:10.1029/1999RS900034.
- Schreiner, W. S., D. Hunt, C. Rocken, and S. V. Sokolovskiy (2002), Radio occultation data processing at the COSMIC Data Analysis and Archival Center (CDAAC), paper presented at First CHAMP Science Meeting, Dtsch. GeoForschungsZent., Potsdam, Germany.
- Schunk, R. W., L. Scherliess, and J. J. Sojka (2002), Ionospheric specification and forecast modeling, *J. Spacecr. Rockets*, *39*(2), 314–324, doi:10.2514/2.3815.
- 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.
- Sutton, E. K., J. M. Forbes, and R. S. Nerem (2005), Global thermospheric neutral density and wind response to the severe 2003 geomagnetic storms from CHAMP accelerometer data, *J. Geophys. Res.*, *110*, A09S40, doi:10.1029/2004JA010985.
- 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.
- Weimer, D. R., B. R. Bowman, E. K. Sutton, and W. K. Tobiska (2011), Predicting global average thermospheric temperature changes resulting from auroral heating, *J. Geophys. Res.*, *116*, A01312, doi:10.1029/2010JA015685.
- Wu, X., X. Hu, X. Gong, X. Zhang, and X. Wang (2009), Analysis of inversion errors of ionospheric radio occultations, *GPS Solut.*, *13*, 231–239, doi:10.1007/s10291-008-0116-x.
- Yue, X., W. S. Schreiner, J. Lei, S. V. Sokolovskiy, C. Rocken, D. C. Hunt, and Y.-H. Kuo (2010), Error analysis of Abel retrieved electron density profiles from radio occultation measurements, *Ann. Geophys.*, *28*, 217–222, doi:10.5194/angeo-28-217-2010.
- Yue, X., W. S. Schreiner, Y.-C. Lin, C. Rocken, Y.-H. Kuo, and B. Zhao (2011), Data assimilation retrieval of electron density profiles from radio occultation measurements, *J. Geophys. Res.*, *116*, A03317, doi:10.1029/2010JA015980.