5-1-2008

An event study to provide validation of TING and CMIT geomagnetic middle-latitude electron densities at the F2 peak

A. G. Burns

W. Wang

M. Wiltberger

S. C. Soloman

Harlan E. Spence
Boston University, harlan.spence@unh.edu

See next page for additional authors

Recommended Citation

Burns, A. G., W. Wang, M. Wiltberger, S. C. Solomon, H. Spence, T. L. Killeen, R. E. Lopez, and J. E. Landivar (2008), An event study to provide validation of TING and CMIT geomagnetic middle-latitude electron densities at the F2 peak, J. Geophys. Res., 113, A05310, doi:10.1029/2007JA012931.

An event study to provide validation of TING and CMIT geomagnetic middle-latitude electron densities at the F_2 peak

A. G. Burns,[1] W. Wang,[1] M. Wiltberger,[1] S. C. Solomon,[1] H. Spence,[2] T. L. Killeen,[1] R. E. Lopez,[3] and J. E. Landivar[3]

Received 10 November 2007; revised 21 December 2007; accepted 18 February 2008; published 17 May 2008.

[1] The coupled thermosphere-ionosphere magnetosphere (CMIT) model and the Thermosphere Ionosphere Nested Grid (TING) model have been run to simulate the 15 May 1997 interplanetary coronal mass ejection's (ICME) effects on the Earth's ionosphere and thermosphere. Comparisons were made between these model runs, the IRI-2007 model, and geomagnetic middle-latitude ionosonde data (NmF_2) from the World Data Center to determine how well the models simulated the event and to understand the causes of model-data disagreement. The following conclusions were drawn from this study: (1) skill scores were more often negative than positive on average; (2) the best and the worst skill scores occurred on the recovery day; (3) the line plots comparing models to data look better than the skill scores might suggest; (4) skill scores are significantly affected by timing issues and large, short-duration variability; (5) skill scores give an indication of the relative ability of one model relative to another, rather than an absolute statement of model accuracy; (6) the models capture negative storm effects better than they capture positive storm effects; (7) the TING model captured many short duration features seen in the data at high middle latitude stations that result from changes in the size of the auroral oval; (8) CMIT overestimates the energy driving changes in NmF_2, whereas TING provides approximately the correct energy input as a result of the saturation effects on potential that are included in TING; and (9) both TING and CMIT electron densities decreased too rapidly after sunset.

Citation: Burns, A. G., W. Wang, M. Wiltberger, S. C. Solomon, H. Spence, T. L. Killeen, R. E. Lopez, and J. E. Landivar (2008), An event study to provide validation of TING and CMIT geomagnetic middle-latitude electron densities at the F_2 peak, *J. Geophys. Res.*, *113*, A05310, doi:10.1029/2007JA012931.

1. Introduction

[2] The Center for Integrated Space weather Modeling (CISM) is developing a suite of models, all of which need validation and verification [*Spence et al.*, 2004]. This validation and verification has two purposes. The first is to assess model accuracy using metrics and validation. Section 4 describes the metrics that are currently being used for ionospheric modeling in CISM and the rationale for developing them. Later in the paper we apply these metrics to a simulation of the 15 May 1997 geomagnetic storm and discuss ways to improve their values. The second purpose of model validation and verification is to find out what the model's strengths and weaknesses are and to suggest ways to improve the model's representation of the region or physical processes being studied.

[3] A number of similar studies have been undertaken previously using general circulation models. All of the early studies involved comparing model output with data visually, often looking only to compare model output with averaged behavior of the data. For example, *Roble et al.* [1984], *Rees et al.* [1985], and *Killeen et al.* [1986] made comparisons between model runs and averaged satellite data to build up a picture of the typical quiet time behavior of the high latitudes. More direct comparisons were made between data and models during specific events not long after this time. *Crowley et al.* [1989] simulated an equinox storm and collected various data to compare with this model run. *Burns et al.* [1995] compared a number of storms with a storm model run to understand the behavior of the lower middle latitudes.

[4] A new type of study was undertaken by *Anderson et al.* [1998], who compared a number of different models with data from Millstone Hill. This numerical assessment of multiple theoretical models in the ionosphere using the same data set had not been tried before. *Fuller-Rowell et al.* [2000] applied numerical criteria to the issue by comparing model output with a Southern and Northern Hemisphere ionosonde using the same sort of skill scores used in this paper. Their concepts, and those developed as

[1] High Altitude Observatory, NCAR, Boulder, Colorado, USA.
[2] Department of Astronomy, Boston University, Boston, Massachusetts, USA.
[3] Department of Physics and Space Sciences, Florida Institute of Technology, Melbourne, Florida, USA.

0148-0227/08/2007JA012931

metrics for the United States National Space Weather Program, are extended here to attempt to obtain a more global view of metrics and validation using a larger number of ground stations.

[5] The main aim of this study is develop an insight into the weaknesses of coupled magnetosphere-ionosphere-thermosphere models in order to improve their forecast capabilities. Such insight can only be developed in the context of an understanding of the physics and chemistry of the ionosphere and thermosphere. Geomagnetic storms provide a well-ordered, dynamical, global evolution of the physics and chemistry in these coupled regions, and therefore are the focus of our study.

[6] The effects of geomagnetic storms on the ionosphere have been studied for a long time. The most striking features equatorward of the auroral oval are the large changes in electron density near the F_2 peak called positive and negative storm effects. *Burns et al.* [2007] recently summarized our knowledge of these phenomena, but it is worth considering them and their causes here as they directly impact the results of this study. Electron densities decrease considerably in the high latitudes and parts of the middle latitudes of both hemispheres. Their latitudinal extent is greatest in the post midnight hours in magnetic local time and they slowly taper off as they corotate into the daylight [*Prölss*, 1981].

[7] It is generally accepted that the main cause of negative storm effects are changes in neutral composition. The idea was first suggested by *Seaton* [1956], who showed that temperature changes could only account for one quarter of the negative storm effects. The contributions of temperatures to these negative storm effects are often not considered in explaining these effects today. In addition to the compositional and thermal causes of negative storm effects another process occurs in the evening just equatorward of the auroral oval. In this region, flux tubes move in the opposite direction to the Earth's rotation. The net effect is that flux tubes are trapped for long periods at night, leading to enhanced recombination and hence decreased electron densities [*Schunk and Nagy*, 2000].

[8] The mechanisms which cause changes in neutral composition in the regions where there are negative storm effects are relevant to this paper as they determine where we should expect to see these regions of electron density decrease. Joule heating causes upwelling of air that is rich in molecular nitrogen and molecular oxygen near the auroral oval. This air is then redistributed by the horizontal winds that are driven by the high-latitude convection pattern and by heating [e.g., *Burns et al.*, 1991; *Fuller-Rowell et al.*, 1994, 1996]. The former is the dominant mechanism in the magnetic high latitudes, so the neutral wind pattern assumes a shape that is similar to the ion convection pattern [e.g., *Killeen and Roble*, 1988]. The antisunward winds across the polar cap are important in determining the redistribution of neutral composition because, unlike the ion drifts which are constrained by interactions with the Earth's magnetic field, these winds are not restricted to the high latitudes and so continue to blow into the middle latitudes near midnight. Thus air that is rich in N_2 and O_2 is advected into the middle latitudes preferentially near midnight. This air then corotates with the Earth, producing the tail of N_2 and O_2 rich air that in turn drives ion-electron recombination and produces the *Prölss* [1981] pattern of negative storm effects.

[9] At low latitudes (and in the middle latitudes in some sectors) electron densities can increase during geomagnetic storms; these are called positive storm effects. Their morphology has not been described in the same detail as negative storm effects have. However, some general statements can be made about them. First, they occur at higher latitudes in the dayside than they do at night. Second, they extend over a larger area in the winter hemisphere than in the summer one.

[10] Their cause is still the subject of much dispute, as there are three competing mechanisms that are invoked to explain them. The three mechanisms are electric field effects [*Mendillo et al.*, 1992], wind effects [*Rishbeth*, 1975], and neutral compositional effects [*Prölss*, 1980]. Electric fields can change electron densities when penetration electric fields enhance the fountain effect. Neutral wind driven changes occur when winds blow up field lines driving the plasma upwards into regions where recombination is slower. Neutral compositional effects involve air at low latitudes blowing downward through constant pressure surfaces increasing O/N_2 and hence electron densities. Thermosphere General Circulation Models currently include the second and third of these processes. Early efforts have been made to incorporate penetration electric fields [e.g., *Peymirat et al.*, 2000; *Fuller-Rowell et al.*, 2002], but these studies are in their infancy.

[11] The aim of this paper is twofold. First to describe our metrics and verification procedures for CMIT's and TING's ionosphere and second to try to put comparisons between the data and the models into a context that will enable us to understand why the models succeed or fail in reproducing positive and negative storm effects that occur in the middle latitudes during geomagnetic storms. The following sections are arranged to meet these goals. The models are described in the next section. The May 1997 storm is described briefly in section 3; the validation techniques are described in section 4; results are presented in section 5; these results are discussed with reference to our understanding of the physics of the ionosphere in section 6; and conclusions and future work are described in section 7.

2. Models

2.1. TING

[12] The Thermosphere Ionosphere Nested Grid (TING) model [*Wang et al.*, 1999] is an extension of the National Center for Atmospheric Research—Thermosphere/Ionosphere General Circulation Model and Thermosphere/Ionosphere Electrodynamics General Circulation Model (NCAR-TIEGCM) [*Roble et al.*, 1988; *Richmond et al.*, 1992] that permits one or more nested grids to be included in the model. This feature was not used here. The NCAR-/TIEGCM is a time dependent, three-dimensional model that solves the fully coupled, nonlinear, hydrodynamic, thermodynamic, and continuity equations of the neutral gas self-consistently with the ion energy, ion momentum, and ion continuity equations using a finite differencing scheme for spatial and temporal variations [*Roble et al.*, 1988; *Richmond et al.*, 1992]. It has 25 constant-pressure levels in the vertical extending from approximately 97 km to 500 km in altitude

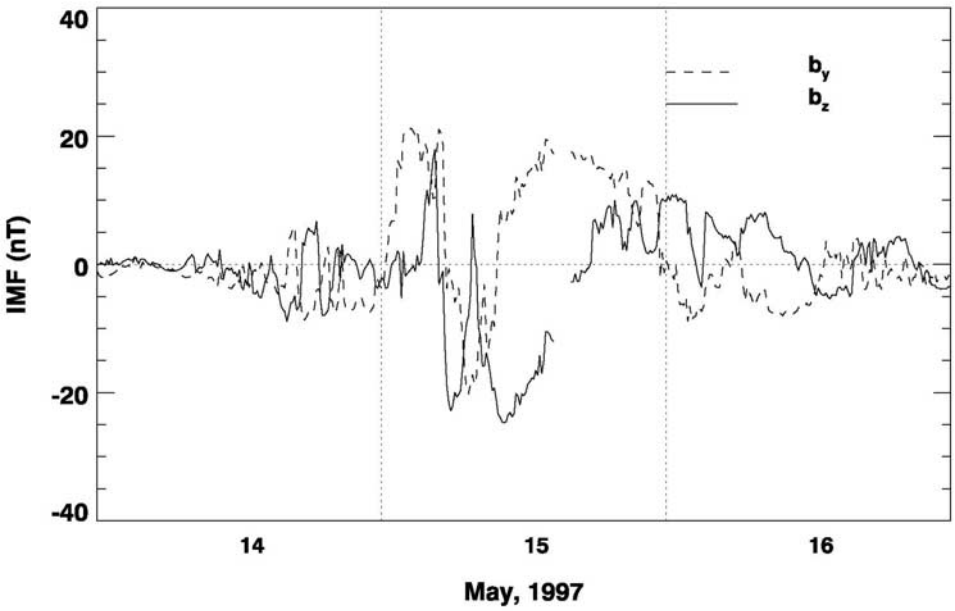

Figure 1. Interplanetary magnetic field variations from 14 May 1997 to 16 May 1997.

and a 5° × 5° degree latitude and longitude grid in its base configuration.

[13] The input parameters for the TING model are the solar EUV and UV spectral fluxes, parameterized by the $F_{10.7cm}$ index, auroral particle precipitation, an imposed magnetospheric electric field, and the amplitudes and phases of tides from the lower atmosphere. In this simulation, the particle precipitation and high-latitude electric field were calculated by applying the appropriate inputs to empirical models (adaptations of *Heelis et al.* [1982] and *Roble and Ridley* [1987]). This is called the stand-alone TING model in this paper. The location of the auroral oval is based on a parameterization by B. Emery et al. (unpublished manuscript, 1990). The width of the oval is based on hemispheric power. The formula used has been adjusted empirically over the years to better represent the observed positions of the auroral oval.

[14] The boundary conditions for the TING and CMIT models have been described by *Wang et al.* [1999]. The condition that is important to this study is the upper boundary condition for O^+. During the daytime no upper boundary flux is provided for O^+. At night there is a net downward flux that is constant at all locations and times, except near the magnetic equator; it is gradually reduced to zero at the magnetic equator. Although the magnitude of this flux is fixed in standard runs, various changes have been tested. One nonstandard version of the model includes an upward flux during the daytime.

2.2. CMIT

[15] The Lyon-Fedder-Mobarry (LFM) code solves the ideal magnetohydrodynamic (MHD) equations for the magnetosphere in a conservative form using the Partial Interface Method on a distorted spherical mesh and Yee type grid [*Lyon et al.*, 2004]. Its domain extends a distance of 30 R_e from the Earth toward the Sun; and a distance of 300 R_e away from both the Sun and the Earth. Its other boundaries (Y and Z directions) extend to ±100 R_e from the Earth in both the y and z planes. There is an additional boundary around the Earth that is typically specified at 3 Earth radii, but it can be lowered to 2 Earth radii.

[16] The coupled version of the LFM and TING models is treated as a separate model called the Coupled Magnetosphere Ionosphere Thermosphere (CMIT) model [*Wang et al.*, 2004; *Wiltberger et al.*, 2004]. Coupling is bidirectional: the LFM model provides the TING model with a high-latitude potential pattern, a characteristic energy, and a number flux of precipitating particles; the TING model provides the LFM model with Pedersen and Hall conductivities. A recent development has involved the inclusion of the effects of thermospheric neutral winds on the magnetosphere, which is now included in the model [*Wang et al.*, 2007], but was not used in the run described here. LFM is run by inputting solar wind parameters and a representation of conductivities. For this run, LFM provided TING with polar cross cap potentials and precipitation, whereas TING provided conductivities for the LFM model. This version of the TING model is called version 1. A new version is under testing which includes an interactive dynamo. This is designated version 2.

3. Storm of 15 May 1997

[17] The 15 May geomagnetic storm was the geospace response to a coronal mass ejection (CME) that was launched from the Sun on 12 May 1997. The storm was geoeffective, producing significant negative and positive storm effects (see section 5). It occurred during a time of low solar EUV radiative flux ($F_{10.7}$ was about 75 and was approximately constant throughout the storm period).

[18] The interplanetary magnetic field (IMF) plot (Figure 1) shows the development of the IMF disturbance from very quiet conditions at the beginning of 14 May through more active conditions late on that day when B_z and B_y were highly variable (albeit with fairly small amplitudes). This field has not been time shifted to account for

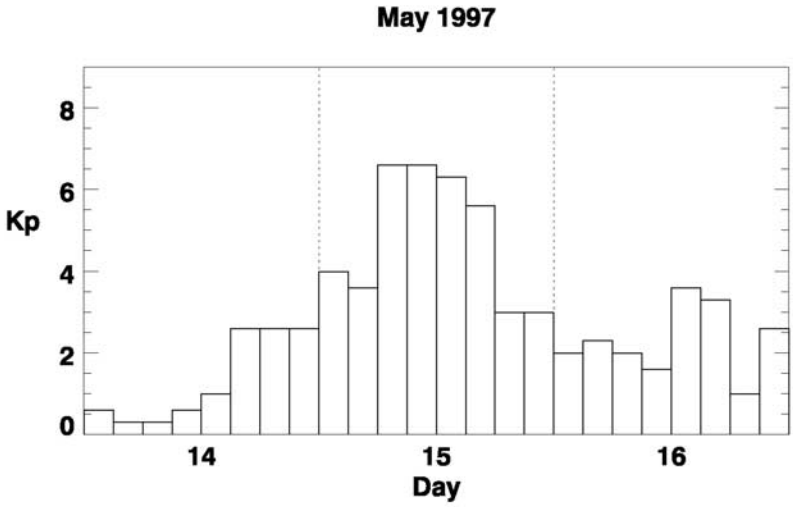

Figure 2. Kp variations for the same period.

the 40 min propagation time from the L1 point to the Earth. The main phase of the geomagnetic storm occurred at about 0600 Universal Time (UT) on 15 May. B_z was strongly southward, with one short northward excursion, until about 1700 UT when it turned northward. During this period B_y shifted from negative to positive at about 1100 UT. Some IMF changes continued throughout 16 May. However, B_z was northward through most of this day and its amplitude often exceeded that of B_y so the IMF suggests that only moderate activity should have occurred on 16 May.

[19] The interpretation of the IMF plot is supported by the changes in K_p during these three days (Figure 2). Kp values were very low at the beginning of 14 May as they had been for the previous week. Values of K_p were consistent with moderate activity in the second half of 14 May, which was in turn consistent with the behavior of the IMF. K_p values of greater than 6 occurred for 9 h after 0600 UT on 15 May, and they remained above 5 for the next 3 h. Only moderate activity occurred in the last 6 h of 15 May coinciding with the northward turning of B_z. K_p values were generally lower than three on 16 May, indicating low to moderate geomagnetic activity, although there was a 6 h period when Kp was at 4^- and 3^+. This corresponded to a period when B_z was southward, but small. In summation, there was a geomagnetic storm on 15 May and then fairly small forcing on the next day.

[20] Figure 3 shows the CMIT response to the IMF values at the L1 point. Cross Polar cap potential and hemispheric power are the inputs that drive the thermosphere and ionosphere in the model. The main feature of this plot lies in the averaged cross polar cap potential and power in CMIT being very high. These values were limited in TING a number of years ago because the model was producing unreasonable storm effects and also as a result of the calculations of *Siscoe et al.* [2002], who showed that the maximum potential is probably no more than a little over 200 kV. The potential in CMIT is beyond this saturation value. This was a large storm, but it was not a truly great storm, so these even potentials of 200–220 kV are not reasonable, hence the restriction of TING to 160 KV. The power that is calculated for TING is commensurate with this potential. It is not practicable to limit CMIT in this way as it would interfere with the magnetosphere, ionosphere thermosphere interaction in unknown ways.

[21] The third (middle) element of this plot is the Joule heating for CMIT and TING. At low levels of cross cap potential these are similar, although there is one large departure on the first day. At high levels of potential and power there are major differences between the two with CMIT Joule heating being much larger than TING's. Much of this difference can be attributed to the saturation that was applied to TING, although there is evidence from other work that the potential in CMIT is much higher than the observed potentials in the very high activity that occurred in the middle of this storm.

4. Data Sets and Metrics Techniques Used

4.1. Data

[22] Results from both the TING and CMIT models were compared with ionosonde data obtained from the World Data Center in Boulder, Colorado in order to assess their performance for the 14 to 17 May 1997 period. A number of different sets of ionosondes data were considered, but after careful study it was concluded that it would be best only to consider geomagnetic middle-latitude stations that showed good repeatability over several quiet days prior to the 15 May 1997 storm. A number of factors lead to this selection. First, high-latitude ionosonde data can be contaminated by auroral activity, so these were not considered. Second, low magnetic latitude data are sensitive to rapid changes in electric field and small changes in distance between locations can lead to large changes in electron density. Thus, there is a spatial and temporal ambiguity about models' performance in this region in that after performing metrics and finding poor performance, it cannot be ascertained whether the problem is that the model has timing issues or whether it is producing features at the right time but at the wrong place. It was decided to not use low-latitude ionosonde data for metrics and validation at present because of this ambiguity.

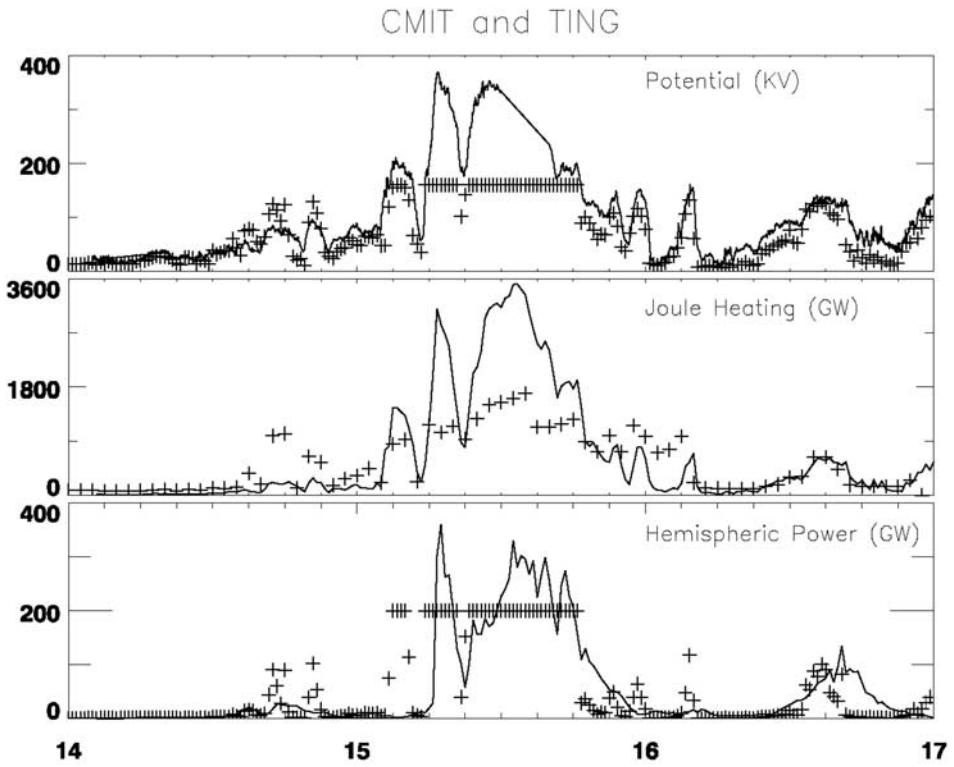

Figure 3. Cross Polar cap potentials, Joule heating, and hemispheric power for the study period. CMIT results are denoted by the lines; TING results are denoted by the crosses. Note that TING is saturated at 160 kV cross cap potential and 200 GW hemispheric power. This value was selected to prevent excess potential and power going into the TING model.

[23] The remaining data come from geomagnetic middle-latitude ionosondes. The data show considerable variations at some stations during quiet times. It could not be determined if these data variations were caused by geophysical variations or by data quality issues. Therefore we eliminated those stations from consideration. There is still a fairly large set of data available for the May 1997 event, certainly enough to provide a good basis for understanding the needs for metrics and to provide some physical insight into the causes of the model's performance successes and limitations. The stations that were used are shown in Figure 4 and their locations are given in Table 1. Note that two of the stations used are in fact just in the low latitudes, but for the sake of simplicity they are called geomagnetic middle-latitude stations here.

[24] These stations were distributed fairly uniformly over geomagnetic middle-latitude land areas (see Figure 4), although there is a preponderance of Northern Hemisphere

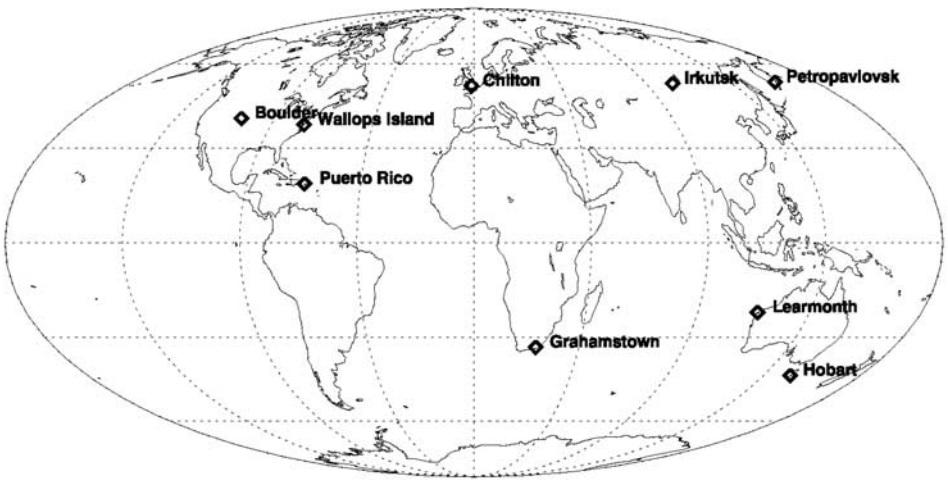

Figure 4. A map of the location of the ionosondes used in this study.

Table 1. Geographic and Geomagnetic Coordinates of the Ionosonde Stations Used in This Study

Station	Geographic Latitude	Geomagnetic Latitude	Geographic Longitude	Geomagnetic Longitude
Boulder	40.	48.6	−106	−33.
Wallops Island	38.	48.	−76	2.4
Ramey	18.5	28.	−61.2	18.1
Chilton	51.5	52.6	−0.6	90.2
Irkutsk	52.5	42.6	110.	−173.4
Petropavlovsk	53.	47.6	164.7	−127.5
Hobart	−42.9	−56.5	153.3	−142.5
Learmonth	−22.	−37.6	114.	166.
Grahamstown	−33.	−29.2	27.	71.5

stations in the study. There are three North American stations, one in Europe and two that are distributed over Siberia. There are no South American stations, one in South Africa, and two in Australia. Thus the distribution of the stations should give a fairly broad picture of how the middle-latitude F_2 peak ionosphere responds to the geomagnetic storm.

4.2. Metrics and Validation

[25] Metrics were used to quantify the model's performance in this study. It was decided to use a simple mean square error test to perform this comparison. The base model used was IRI-90 [*Bilitza*, 2007]. This version was used because of the need to provide a thoroughly tested model as a baseline model. It does not provide a representation of geomagnetic variations, but this is an advantage in this context as it provides a stable, repeatable baseline for this and subsequent metrics studies without worrying about how the IRI's representation of geomagnetic storms might change in different conditions. Later it will be shown that it still outperforms the two versions of the TING model used here in the metric calculations most of the time. In large part this does not represent errors within TING; rather it indicates issues with the metrics that favor smooth, climatological models rather than models that attempt to do "weather." This issue will be discussed further later in the paper. The technique used was to calculate the mean square error for each day:

$$\mathrm{RMSE} = \sqrt{\frac{\sum(\mathrm{model} - \mathrm{data})^2}{\mathrm{N}}}$$

where model is the modeled value, data is the measured value, and N is the number of these pairs in the day. A similar calculation is made for IRI and the skill score is obtained by

$$\mathrm{skillscore} = \left(1 - \frac{\mathrm{RMSE_{TING}}}{\mathrm{RMSE_{IRI}}}\right)$$

These skill scores are presented later in this paper in Tables 2 and 3. The prime aim of skill scores is to provide a very limited subset of numbers that enables the reader to assess the models' overall performance (ideally one), that can be studied over time to assess how models have improved. Thus a small number of values are needed, so that a reader can immediately assess whether models have improved over time. For this reason we have chosen nine stations to assess skill scores and then divided the data in these stations into three parts: a quiet day; a storm day, and a recovery day.

[26] The aim of this paper is to determine how well CMIT and TING perform using the most commonly used skill score and what the strengths and weaknesses of this skill score is. This skill score has been developed from the one that was established as a result of community consensus, so we feel obliged to understand how our models performed using this technique and what its limitations are. A future paper will address the issue of better skill scores for ionospheric and thermospheric applications.

5. Comparisons Between Observed Electron Densities and Those Calculated by CMIT and the Stand-Alone TING Model

[27] Figure 5 shows the comparison between TING, CMIT (version 1.0) and ionosonde data for three stations over North America (including Puerto Rico). The two North American stations (Boulder and Wallops Island) have a brief positive phase (increases of NmF_2 relative to quiet time) for a short period on 14 May, followed by a prolonged negative phase throughout the daylight hours of 15 to 16 May. Daylight values of NmF_2 recover and may have become slightly positive by 16 May at this time. The nighttime changes in NmF_2 are negative (NmF_2 decreases relative to its quiet time values) during 15 and 16 May, although the decreases are much smaller at Wallops Island than they are at Boulder. Negative storm effects are seen at Puerto Rico during both nighttime and daytime on 15 May and during nighttime on 16 May. Electron densities have recovered (although there may be weak positive storm effects) during the daytime on 16 May.

[28] TING and CMIT behave in the same way during the quiet period before the storm. This is expected as there is little difference in the high-latitude forcing during quiet times. Also, the forcing that does occur may have little

Table 2. Skill Scores for Six Northern Hemisphere Ionosonde Stations

	Boulder		Wallops Island		Puerto Rico		Chilton		Irkutsk		Petropavlovsk	
	TING	CMIT	TING	CMIT	TING	CMIT	TING	CMIT	TING	CMIT	TING	CMIT
14 May	−0.34	−0.20	−1.35	−1.15	−0.03	−0.27	−1.88	−1.01	−2.18	−2.28	0.92	0.94
15 May	−9.22	−1.72	0.00	−0.18	0.72	0.73	−0.85	−0.26	−0.41	−2.40	−0.76	−1.54
16 May	0.67	0.22	0.30	−3.71	−8.37	−11.8	−0.57	−1.86	0.68	0.83	0.53	0.38

Table 3. Skill Scores for Three Southern Hemisphere Ionosonde Stations

	Hobart		Learmonth		Grahamstown	
	TING	CMIT	TING	CMIT	TING	CMIT
14 May	−6.32	−8.74	0.25	0.35	−1.56	−2.13
15 May	−0.92	−1.50	0.15	0.48	−1.37	−2.35
16 May	−3.39	0.32	−4.14	−10.20	−6.87	−22.72

impact on middle-latitude electron densities. TING and CMIT behave in a similar fashion for both Boulder and Wallops Island on both 15 May and 16 May with one striking exception at both Boulder and Wallops Island before the middle of the UT day on 15 May. At this time it is night in Boulder, but the TING electron densities are higher than they are in the daytime. The reasons for this excursion will be considered later. Apart from this deviation, both TING and CMIT electron densities are lower than quiet time values throughout 15 May and 16 May at Boulder, although the TING electron densities are larger than those of CMIT. They both also produce negative storm effects at Wallops Island on 15 May and 16 May as well. However, TING values in the latter half of 16 May reproduce the recovery seen in the data far better than those from CMIT do.

[29] The situation is similar at Puerto Rico, except that both CMIT and TING greatly underestimate quiet time values of NmF_2. Both TING and CMIT electron densities reflect the negative storm effects seen in the data on 15 May, although they underestimate electron densities at night (i.e., the negative storm effects that they calculate at night are too strong). TING electron densities recover on 16 May, but the CMIT electron densities do not. In both cases the models underestimate the observations, much as they did before the storm started.

[30] Chilton is the only European ionosonde discussed in this paper (see Figure 6 for the European and Asian comparison between ionosondes data and the models) because the other European ionosondes did not provide stable electron density measurements (the peaks on these quiet days varied greatly, which may have been a physical result, but may also have indicated data issues) during the quiet days before the storm. NmF_2 values at Chilton increased greatly from their quiet time values late on 14 May, presumably in response to the changes in IMF that resulted from the arrival of the shock that can be seen in Figure 1. The positive storm effects lasted for several hours until the beginning of 15 May. NmF_2 values at Chilton fell at or below their quiet time values for the first 4 to 6 h of 15 May. They were weaker than their quiet time values throughout the rest of 16 May. They also showed strong variability on this day at Chilton producing two strong spikes and one weaker one. NmF_2 values exhibited negative storm effects at the beginning of 16 May, but they had returned to their quiet time values by the middle of the morning and surpassed them later in the day.

[31] CMIT does not reproduce the positive storm effects seen at Chilton on 14 May. Negative storm effects are produced by CMIT on 15 May. The agreement is mostly good early on this day, but CMIT tends to underestimate NmF_2 values later in the day. CMIT calculates low values of electron density on 16 May, in contrast to the data, which indicate that NmF_2 values have recovered or are slightly positive on this day, rather than the negative values that CMIT calculates. Note that CMIT values are generally smoother than those of the stand-alone TING model.

[32] The TING model performed better at Chilton than it did at any other location in this study (this is not reflected in the skill scores, a point that will also be discussed later), except for failing to capture the positive effect on 14 May and the short, sharp increase in electron densities that it calculated early on 15 May. The TING model even picks up spikes (the model and data agreement suggest that the spike in the data are not just an artifact) that were seen in the ionosonde data on 15 May. It also captures the positive effects that were seen in Chilton ionosondes data on 16 May, although the electron density from the TING model peaks earlier than that seen in the ionosonde data.

[33] Figure 6b compares the ionosonde NmF_2 data from Irkutsk with model calculations of the same parameter at the same location. Daytime values on 14 May are similar to those on the previous 2 d and nighttime values are very similar to those on the previous day. Positive storm effects occur on 15 May during the daytime. The results show considerable structure in this phase. Negative storm effects occur at night on 16 May with very small values of NmF_2 occurring at this time. These effects continue through the day and the night of 16 May, although NmF_2 has partially recovered during the daytime on 17 May.

[34] Both TING and CMIT reproduce the daytime values of NmF_2 on 14 May at Irkutsk fairly well but greatly underestimate the nighttime values. This underestimation of nighttime values during quiet geomagnetic conditions was an issue at all the previously mentioned stations, with the exception of Wallops Island. There is a general tendency for both the stand-alone TING and the CMIT values of NmF_2 to fall to their overnight low values more quickly than the Irkutsk data on every day during the simulated event.

[35] The most surprising thing about the ionosonde's measurements at Petropavlovsk (Figure 6c) is the very strong positive effect that occurs early on 15 May. NmF_2 values increase by 3 times for a period of a few hours at this time. This feature is not reproduced by either CMIT or the stand-alone TING model. It is not a single impulse; it has at least one other structure associated with it. CMIT goes into a negative phase (instead of a positive one) during this event. The stand-alone TING model captures some structure during the positive phase in the data but does not show any positive phase itself.

[36] After 1800 UT on 15 May, CMIT tracks the data fairly well. Negative storm effects are seen in the data and in CMIT for the next 18 h. The stand-alone TING model indicates a large increase in electron density from about 1200 to 1800 UT on 15 May. This increase is not seen in either the data or the CMIT run. It is similar to the

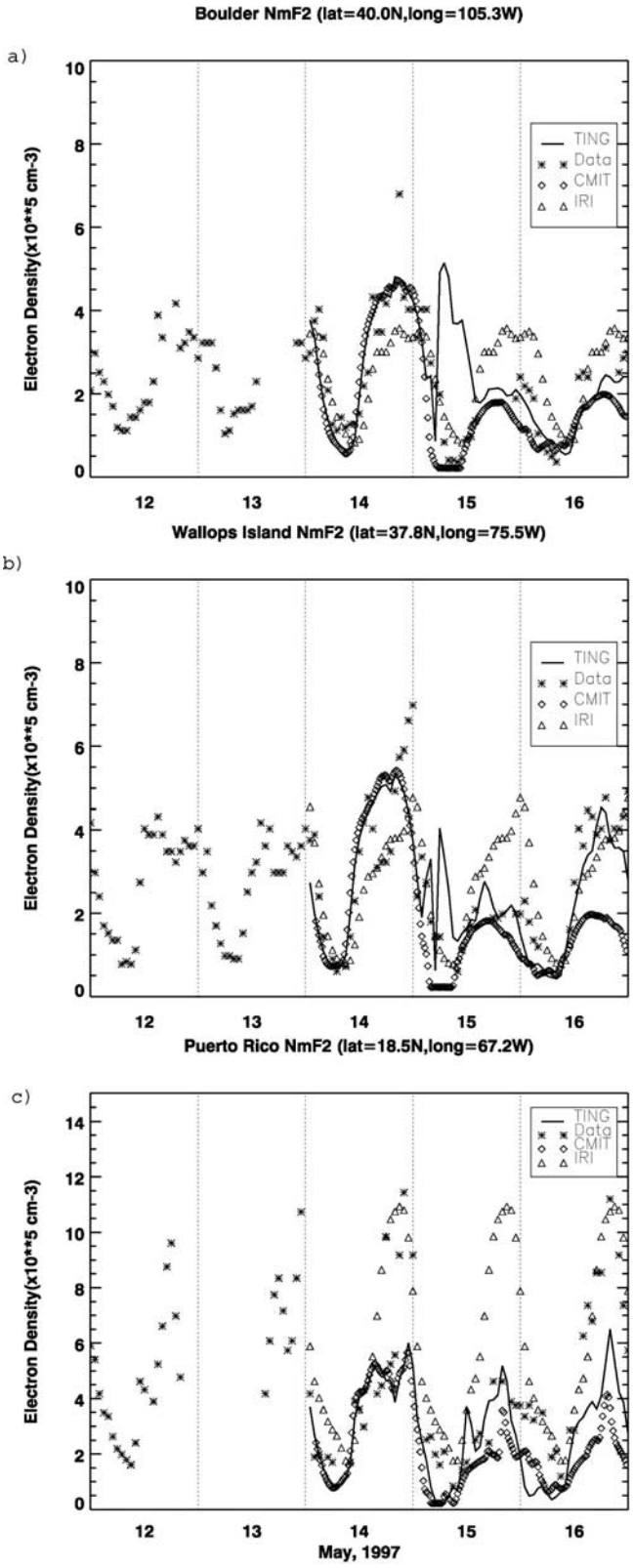

Figure 5. Variations of NmF_2 for three North American ionosondes during the period from 12 May 1997 to 16 May 1997. The two versions of the TING model and IRI-90 results are also plotted here for the same stations. (a) Boulder, USA; (b) Wallops Island, USA; and (c) Ramey, Puerto Rico.

excursions that were seen at Boulder, Wallops Island, and Chilton.

[37] Comparisons between the models and NmF_2 values were also made for Southern Hemisphere stations (Figure 7). The southern ionosphere in May was a typical winter one, with positive storm effects occurring frequently in the lower middle latitude stations, while negative storm effects were seen through much of the storm at Hobart (Figure 7a).

[38] The NmF_2 data at Hobart increased above their quiet time values before 1200 UT on 15 May and then declined back to their quiet time values over the next local night. There were large decreases in electron density during the local daytime (between 15 and 16 May UT day) on the next day, with electron densities reaching a peak only about half their quiet time values early on 16 May. There was a spike in electron densities during the local night at about 1400 UT on 16 May UT day.

[39] CMIT overestimated the positive storm effects at the beginning of 15 May by about 20%. However, it did capture both the drop off of electron densities at about 1200 UT and the subsequent low electron densities during the following local night. It also captured much of the negative storm effects seen at the beginning of 16 May, although electron densities were somewhat higher on average in CMIT. There was a spike of increased electron densities at about 1400 UT in the data. The sense of this was captured by CMIT, but the model greatly underestimated the magnitude of the spike. The stand-alone TING model did a poor job of capturing changes in electron densities on both 15 May and the first half of 16 May. It captured the magnitude of the positive storm effects at the beginning of 15 May but did not capture the details. There was an extended structure of high electron densities from about 1100 UT on 15 May to about 1800 UT on that day in TING. This seems to be similar to the increases in electron density that were seen at several Northern Hemisphere stations at about this time. The stand-alone TING model results showed a peak at about 1400 UT; this peak also occurred in the data although it was slightly smaller than the one seen in the TING results.

[40] The Learmonth results are shown in Figure 7b. Strong positive storm effects were seen in the data during local daytime on 15 May. The following local nighttime electron densities were slightly lower than those of the previous quiet days. Similar behavior occurred during the next local night. Electron density changes were slightly positive during local daytime from 0000 UT to 1200 UT on 16 May. There was a twin peaked structure in the electron densities during this period, indicating that there was a brief period of negative storm effects during this daylight period.

[41] CMIT also produced positive storm effects from 0000 UT to 1200 UT on 15 May, but the magnitude of these effects were larger than those seen in the data. At local night on 15 May the CMIT model calculated deeper electron density depletions than those that occurred in the data for most of the night. The lowest electron densities also occurred earlier in the CMIT run than they did in the data. CMIT underestimated electron densities during local daytime on 16 May. However, CMIT does produce a two-peaked NmF_2 trace at this time. The stand-alone TING model also produced positive storm effects on 15 May. Like the CMIT results, they were also larger than those seen in the data. The local nighttime changes on 15 May were of

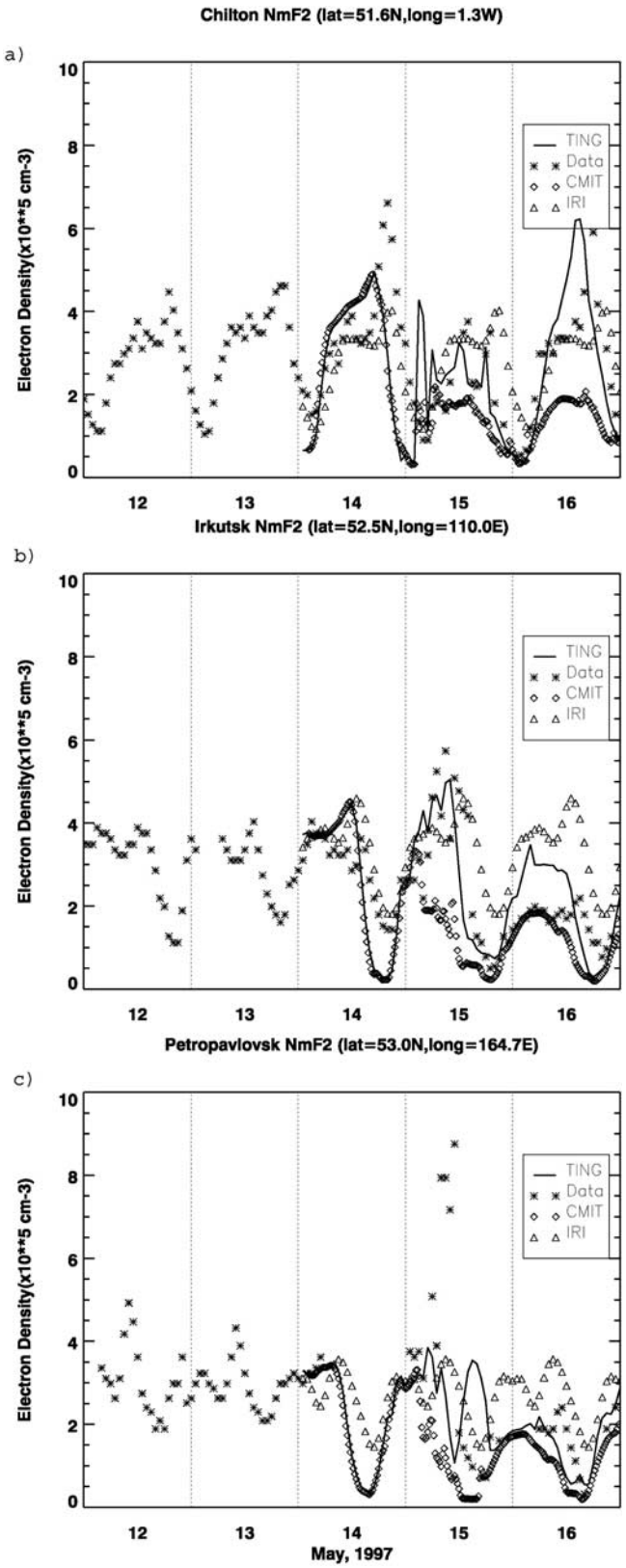

Figure 6. Variations of NmF_2 for three Eurasian ionosondes during the period from 12 May 1997 to 16 May 1997. The two versions of the TING model and IRI-90 results are also plotted here for the same stations. (a) Chilton, England; (b) Irkutsk, Russia; and (c) Petropavlovsk, Russia.

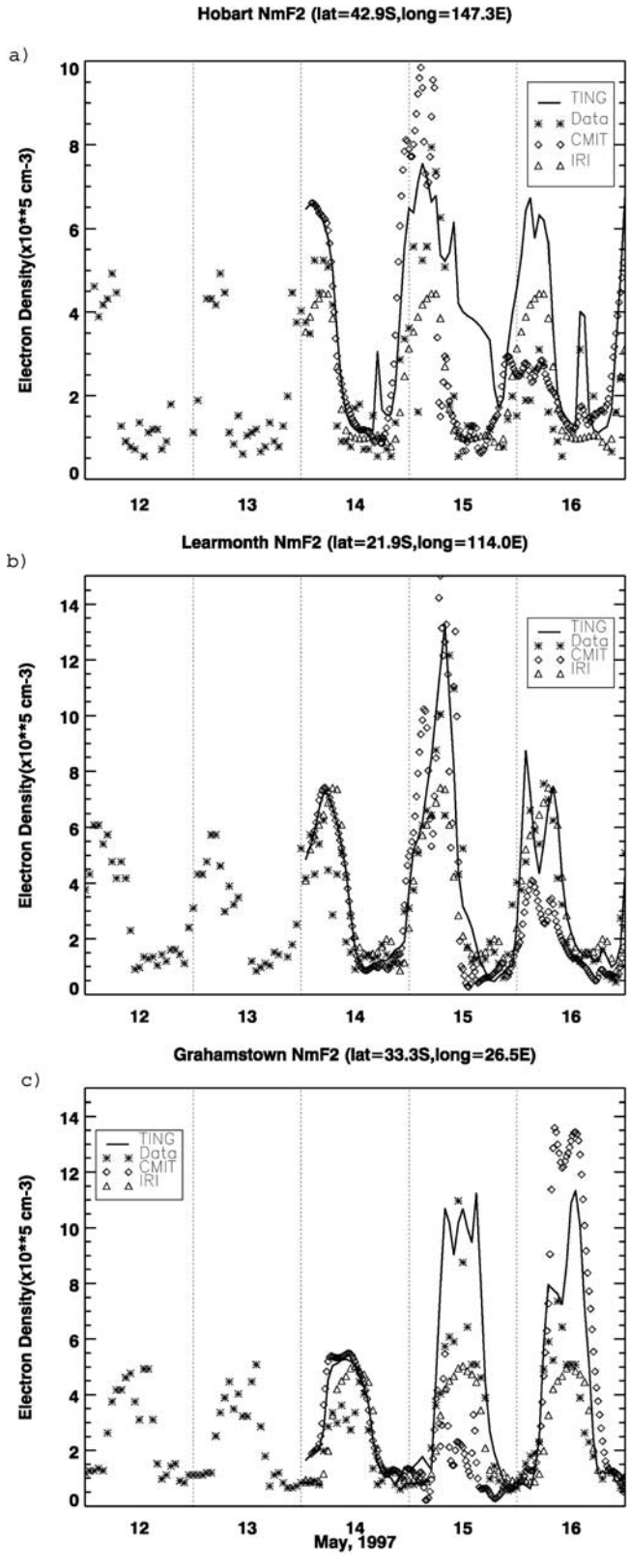

Figure 7. Variations of NmF_2 for three Southern Hemisphere ionosondes during the period from 12 May 1997 to 16 May 1997. The two versions of the TING model and IRI-90 results are also plotted here for the same stations. (a) Hobart, Australia; (b) Learmonth, Australia; (c) Grahamstown, South Africa.

roughly the same magnitude as the data, but the model output was smooth, whereas the data varied greatly. The twin peaked structure seen in the data during local daytime was reproduced well in the stand-alone TING model as was the detailed structure seen during the local night on 16 May.

[42] Positive storm effects predominated during local daytime at Grahamstown (Figure 7c). Daytime electron densities were enhanced during the storm and recovery days. On 15 May the enhancement was as much as 100% and it was as much as about 50% on the other 2 d. Local nighttime electron densities were the same or slightly smaller than their quiet time values on the day of the main phase of the storm and on all subsequent days. Electron densities at Grahamstown were highly structured, both during the quiet times before the storm, on the day of the main phase of the storm, and on subsequent days.

[43] The CMIT model run did not produce positive storm effects during the local daytime at Grahamstown on 15 May. Instead, negative storm effects were seen in the CMIT output through most of the daylight hours of this day. The CMIT model also produced positive storm effects in the local daylight hours on 16 May, but the magnitude of these effects was much greater than those seen in the data. The local nighttime simulation at the end of the day was similar to that seen in the data. The stand-alone TING model produced positive storm effects on 15 May, but they had a longer lived triple peak structure, so the integrated electron density for the day was much larger than the observed electron densities. The lowest values of electron density during local night on this day were also reproduced by the stand-alone TING model.

6. Global Variations in NmF_2 From CMIT and the Stand-Alone TING Model

[44] No attempt is made in this paper to provide an all-encompassing picture of how NmF_2 changed globally in the model runs. Instead, a few, representative plots of NmF2 have been made for select times in an attempt to shed some light on why certain features are seen in the comparison plots in the previous section and to put these comparison plots in a more global context.

[45] The global distribution of NmF_2 at 0700 UT on 15 May 1997 that was calculated by CMIT is shown in Figure 8a. This time was near the peak of the storm. Low electron densities were seen over much of the globe. There was evidence of the ions produced by auroral precipitation in the north, which extended across much of Canada and New England, but did not extend as far south as Boulder or Wallops Island. These electron density enhancements, that were associated with the auroral oval, occurred below 200 km and were rather weak in this case. Another noticeable feature of this plot is that the values of NmF_2 were low throughout most of the Northern Hemisphere and were smooth in these regions at these times. Large values of NmF_2 occurred during daylight in the middle and low latitudes of the Southern Hemisphere. NmF_2 was enhanced over its quiet time values in these regions at these times. There were significant, relatively small-scale structures associated with these regions and times of enhanced NmF_2. Hobart was near the boundary between regions of enhanced and depleted NmF_2, but the other Southern Hemisphere stations discussed in this paper were located near some of the biggest enhancements of NmF_2. The region of depleted NmF_2 in the dayside in the Southern Hemisphere was also quite structured, suggesting that the Hobart observations were likely to be difficult to reproduce using CMIT.

[46] Figure 8b gives values of NmF_2 that were calculated by the stand-alone TING model at 0700 UT on 15 May 1997 during the height of the storm. This plot has several similarities to Figure 8a, but there are noticeable differences as well. The feature of most relevance to the observations in section 5 is the band of enhanced NmF_2 that were seen over North America and across the Atlantic to England. Like those seen in the CMIT results, these bands were not enhancements at the F_2 peak, rather they represented enhancements of electron density in the auroral oval at altitudes between about 120 km and 180 km. Such enhancements are often seen in model output at solar minimum, when NmF_2 values are sufficiently small to permit them to be exceeded by the electron densities associated with particle precipitation. Note that these bands extended over a number of the ionosondes used in this study.

[47] Some other features in Figure 8b are of relevance to interpreting the comparison figures. First, small-scale features dominate the plot of NmF_2 over Australia and South Africa. These differ considerably from those seen in the CMIT results, especially insofar as there were increases in electron density extending from Australia over Antarctica in the TING results and that Hobart was in a region and time of enhanced density in this case. The region of enhanced densities extended further north in the Pacific than those calculated by CMIT did and the enhancements were not as strong over the Indian Ocean and Australia. There were fairly smooth depletions in electron density over much of the Northern Hemisphere, but these depletions were much weaker than the ones that occurred in the CMIT run.

[48] Figure 9a gives values of NmF_2 from CMIT at 1300 UT on 15 May. The storm had continued for a number of hours by this time, so the figures represent some of the greatest effects of the storm on NmF_2. Depletions of NmF_2 were very deep in the high and middle latitudes of both hemispheres. Their morphology was smooth in these regions at this time, suggesting that differences between CMIT and ionosonde data could not be explained by small differences in the location of the disturbance, except in those regions that are near boundaries between areas of depletions and areas of enhancements. Four such stations are Hobart, Grahamstown, Wallops Island, and Boulder. There were significant enhancements of NmF_2 at low latitudes. The Grahamstown ionosonde was located near the boundary between this region of enhancements and the region of depletions, so relatively small changes in the position of the structured regions associated with these enhancements could have resulted in major changes in NmF_2. This probably explains why increases of NmF_2 occurred in the Grahamstown ionosonde data, but decreases were seen in the CMIT model output at this UT. This point and others like it will be considered further in section 8.

[49] TING model outputs are presented for 1300 UT in Figure 9b. Values of NmF_2 were not as low in regions of depletion as they were in the CMIT output. However, the regions of depleted NmF_2 were extensive in both hemi-

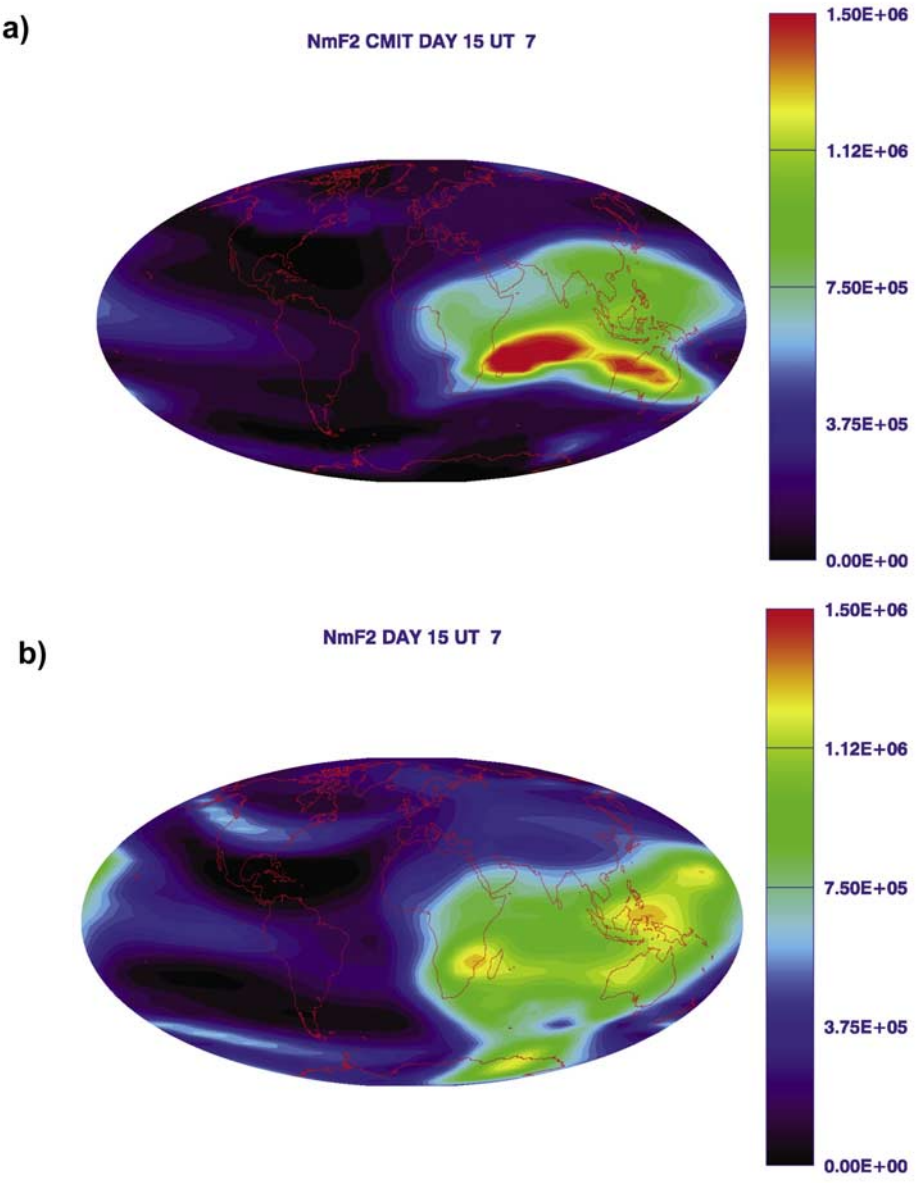

Figure 8. Global plots of modeled NmF_2 at 0700 UT on 15 May 1997. (a) CMIT output and (b) stand-alone TING output.

spheres and they were smooth except near the auroral regions, when precipitation caused sharp boundaries in NmF_2. Strong precipitation occurred over Hobart and Petropavlovsk in this model run and weaker precipitation occurred over Boulder. The regions of enhanced NmF_2 were much larger than they were in the CMIT run, but their magnitude was smaller, and they were smoother. There were boundaries between decreases and increases of electron density over South Africa at this UT, which indicates that NmF_2 was difficult to model at Grahamstown.

[50] The main phase of the storm had been over for a day by 1600 UT on 16 May, but the IMF was still disturbed (and B_z was negative at this time) and there were strong storm signatures in the NmF_2 model results. Figure 10a shows the CMIT calculations for this time. There were still extensive areas of depleted NmF_2, particularly over the Pacific Ocean, Australia, and Central Asia. Depletions were very deep in these nighttime regions. There were also depletions in the daytime over North America and over the North Atlantic. All of these areas of depletion were smooth, in agreement with the smooth nature of the ionosondes profiles that were simulated by CMIT in these locations at this time. The region of enhanced NmF_2 values that CMIT calculated was small at this time. It was centered over the South Atlantic and extended over parts of southern Africa and over South America into the South Pacific. The extension over Africa suggests that simulations of Grahamstown NmF_2 values should be difficult to predict because of the proximity of this station to the boundary between regions of enhancement and regions of depletion.

[51] Figure 10b is a similar plot for the stand-alone TING model at 1600 UT on 16 May. Depletions were not as deep

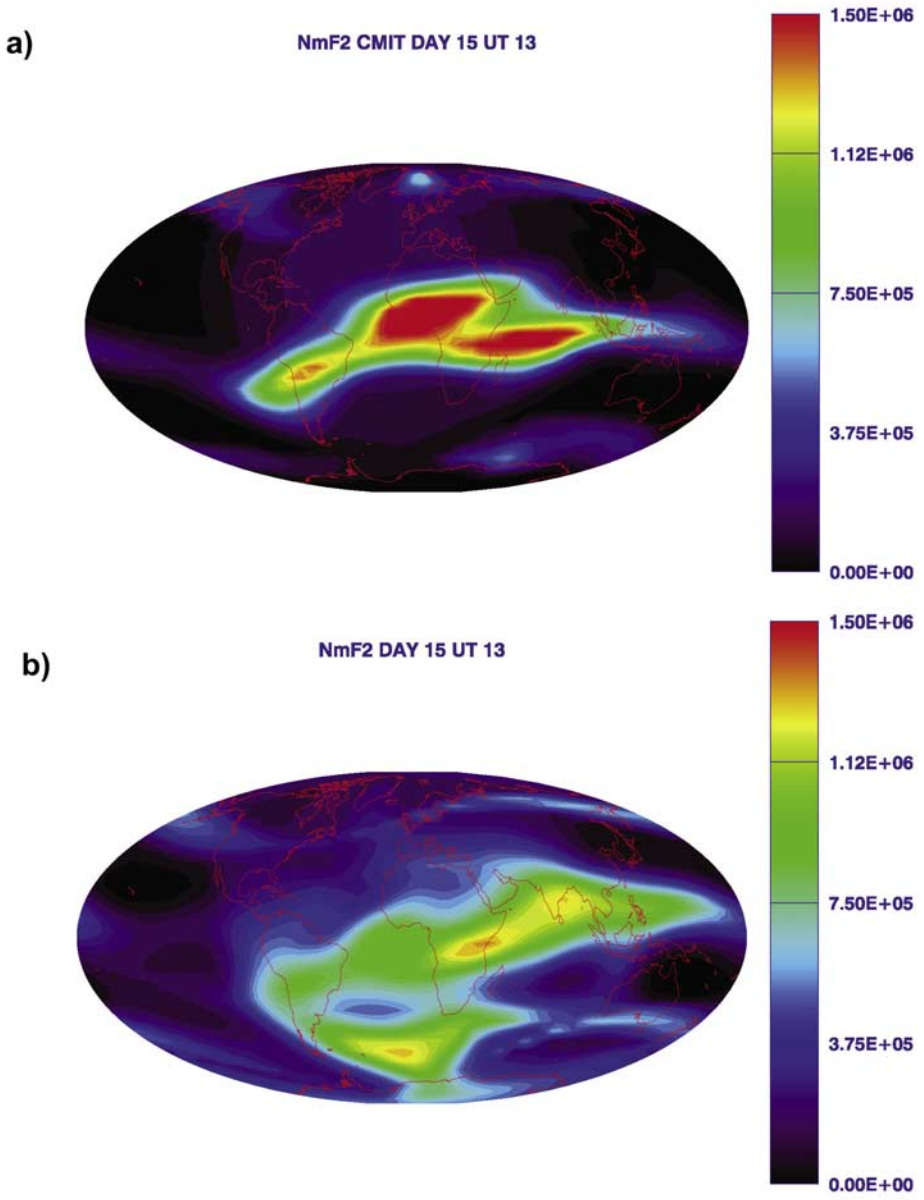

Figure 9. Global plots of modeled NmF_2 at 1300 UT on 15 May 1997. (a) CMIT output and (b) stand-alone TING output.

in this simulation and occurred over a smaller area. They were still deep over eastern Siberia, which was also reflected in the Petropavlovsk ionosonde data. Irkutsk was near the boundary between regions of enhancement and depletion and was thus more difficult to simulate. Hobart was in a region of lower NmF_2 at this time, but these were the normal nighttime electron densities and so they did not represent a depletion produced by the storm. There was a region of relatively enhanced NmF_2 just south of Hobart at this time. This was a manifestation of the auroral oval. Enhanced values of NmF_2 were seen at Hobart in both the ionosonde data and the TING model output (and to a lesser extent in the CMIT output) just prior to this time. This suggests that this bump probably represented a northward excursion of the auroral oval. The region of enhanced NmF_2 was much more extensive in the stand-alone TING model output than it was in the CMIT output at this time. It was also strong over a much larger area. The feature of these regions of enhancement that is of most relevance to this paper was its boundary over southern Africa. Grahamstown was near this boundary, so it likely that accurate model simulations of Grahamstown NmF2 values were difficult at this time.

7. Skill Scores

[52] CMIT has been developed to provide end-to-end predictions of the environment from the Sun to the Earth [*Luhmann et al.*, 2004]. These predictions are of little value unless there are objective methods of testing the models' performance [*Spence et al.*, 2004]. An initial attempt has been made here to develop skill scores for TING and CMIT.

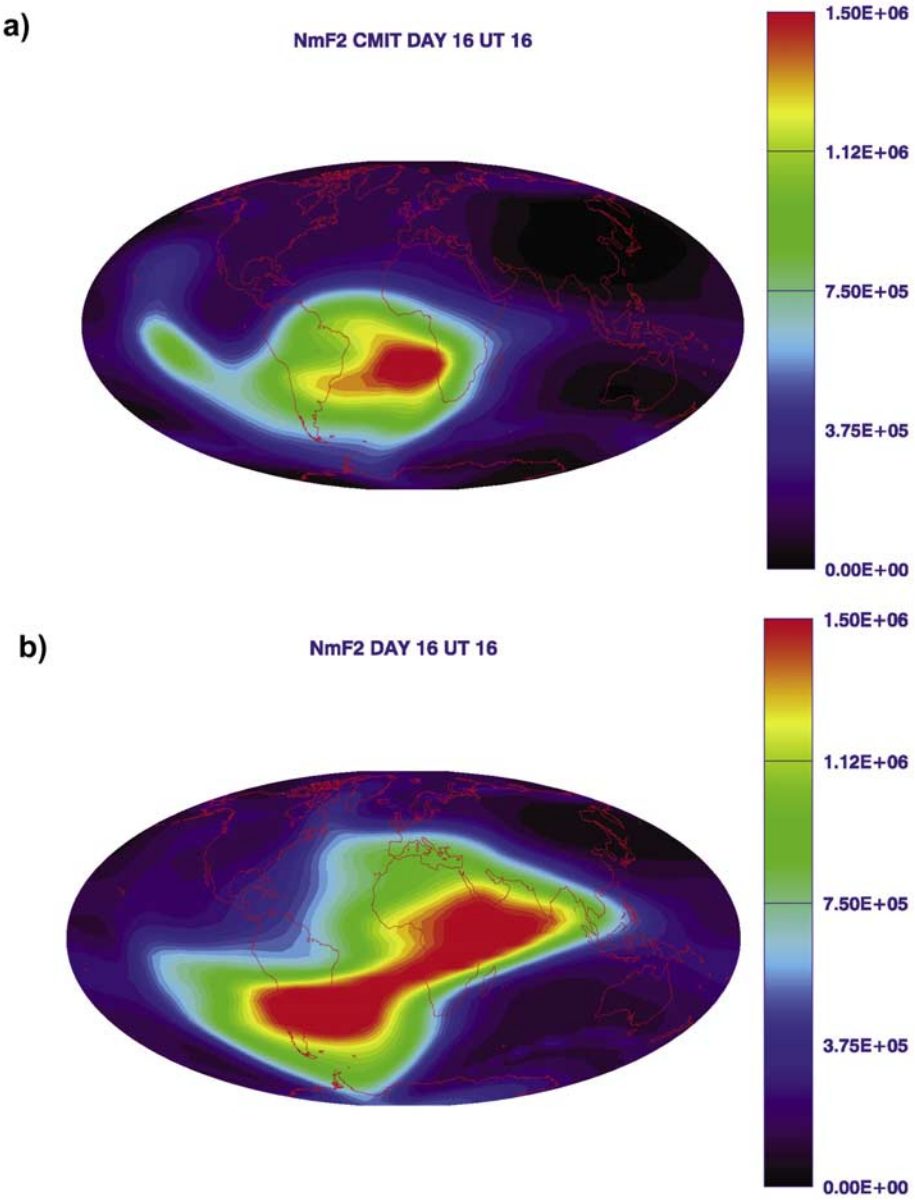

Figure 10. Global plots of modeled NmF_2 at 1600 UT on 16 May 1997. (a) CMIT output and (b) standalone TING output.

The skill scores are obtained by the method described in section 4.

[53] Tables 2 and 3 give the TING and CMIT skill scores for the various ionosondes used in this study. What stands out is that the skill scores were negative for most ionosondes on most days. In other words the two versions of the GCM made worse predictions than a model that assumed persistence of the quiet time behavior. This indicates either a weakness in the way that the skill scores are measured, or a weakness in the two versions of the model's ability to simulate geomagnetic storms, or, both.

[54] It is worth examining these skill scores in a systematic fashion and comparing them with the equivalent line plots to see if any conclusions can be drawn about why CMIT and TING simulations behaved the way that they did in comparison with the IRI model.

[55] The skill scores in Boulder are slightly negative for both models on 14 May, more negative on 15 May, when TING is much more negative than CMIT, and positive on 16 May. Comparing these results to Figure 5a leads to the following observations. Both TING and CMIT values of NmF_2 decreased more rapidly than the data and the IRI model as night fell on 14 May at Boulder and may have increased more rapidly in the morning. The differences at these times were partially offset by TING and CMIT better matching the data than the IRI model during the daytime.

[56] TING NmF_2 calculations were severely affected by the positioning of the auroral oval over Boulder in the model output on 15 May. A data gap late on this day when TING and model output predicted negative storm effects in much the same way that they probably occurred in the data skewed the results in favor of the IRI model. TING and

CMIT performance was better than the IRI-90 model on 16 May because the general agreement between the two theoretical models and the data early on the day was not sufficiently counterbalanced by the good agreement between IRI results and the data later on that day.

[57] Similar relationships occurred between the model results and the ionosonde data at Wallops Island (see Figure 5b). These gave rise to a similar behavior in the skill scores. TING and CMIT NmF_2s again decreased too rapidly after dusk and there was another period when the auroral oval was overhead. The main difference between the two versions of the model is that TING values of NmF_2 recovered late on 16 May, whereas CMIT results indicated that negative storm effects still existed at this time, giving rise to a positive skill score for TING and a fairly strong negative one for CMIT.

[58] TING and CMIT greatly underestimated NmF_2 values at Puerto Rico during the afternoon of the quiet day (14 May) at Puerto Rico. However, the skill scores were not bad (<-1) on this day because TING and CMIT tracked the data fairly well for the rest of the day. Negative storm effects were seen in the data and in the TING and CMIT models on 15 May, which gave rise to positive skill scores. The ionosonde NmF_2 values indicated that the data had recovered or became slightly positive on 16 May, resulting in poor skill scores for TING and very bad ones for CMIT. Negative storm effects prevailed in the latter model, which caused the poor skill score on 16 May.

[59] Chilton skill scores are surprising insofar as the TING model not only captured the magnitude of the electron density during the daytime on 15 May but also captured the short duration excursions. Despite this agreement, the skill scores for TING and CMIT were negative for all 3 d. One of the factors leading to these low skill scores was that the IRI model results were about 10% lower than the ionosonde data on 14 May, which was, for the most part, a quiet day. Thus the IRI calculations were close to the values of the observed electron densities during the storm. In such a case the skill scores estimated for the theoretical models during the storm will be poorer than they would be if the IRI electron densities were the same as the quiet time observed electron densities. NmF_2 ionosonde measurements showed evidence of weak negative storm effects on the next day, which resulted in NmF_2 values that were not greatly different from the IRI model values. CMIT overestimated the decrease of NmF_2 on 15 May, which caused its poor skill score, and TING's skill score was seriously affected by the peak of NmF_2 that it predicted early on this day. TING's skill score on 16 May was poor because it predicted that the daytime peak in electron density would occur 2 to 4 h before it actually did. This had a double effect on the skill scores insofar as both the contribution from the time of TING's peak and the contribution from the ionosonde data's peak adversely affected the skill score. CMIT values were far two low throughout this day; consequently the skill score was poor.

[60] Little additional information can be gained from Irkutsk. Skill scores for both TING and CMIT were good on 16 May because there were deep negative storm effects that both models predicted. They were poorer on the first day because both TING and CMIT NmF_2 values decreased too rapidly after dark and decreased to too low a value. On the second day CMIT's skill score was poor because its NmF_2 values indicated that a negative storm occurred in the daytime, when, in fact, electron densities were slightly enhanced as TING predicted.

[61] The skill scores on 14 May at Petropavlovsk are misleading in the absence of careful analysis. Ionosonde measurements were only available for the first few hours on this day. TING and CMIT agreed more closely with the data than the IRI model did for these hours, so the skill scores were good. Skill scores were also good on 16 May, when deep decrease of NmF_2 occurred in the data. These decreases also occurred in both the TING and CMIT model results. Skill scores were poor on 15 May, when CMIT and TING calculated a decrease in NmF_2, when there was a large increase. TING results were also compromised in the second half of this day, when its calculation showed large increases in electron density which were caused by the movement of the auroral oval to a position over Petropavlovsk, as was described in section 6.

[62] The skill scores of TING and CMIT were very poor on 14 May at Hobart. One of the main causes of these poor skill scores was that the agreement between IRI and the ionosonde data was very good on this day, so significant disagreements between the two models and the data at the beginning and end of the day were exacerbated. Note that there is no evidence that CMIT or TING electron densities decreased too quickly after dark at this station. Neither TING nor CMIT had good skill scores on 15 May. CMIT considerably overestimated the increase of NmF_2 at Hobart, whereas enhanced electron densities associated with the auroral oval were seen by TING in the second half of this day. CMIT had a positive skill score on 16 May, where it accurately modeled negative storm effects during the daytime. TING did not model these negative storm effects and consequently had a poor skill score.

[63] Skill scores for both TING and CMIT were positive at Learmonth on 14 May, the day before the storm. These positive scores can be explained easily using Figure 6b. TING, CMIT, and IRI were in fairly close agreement, but as night fell TING and CMIT were slightly more like the ionosonde data than the IRI model results were. Because these differences between all of the models and the data were small, the skill scores provided little insight on this day. The TING skill score was positive on the next day, but again this does not seem significant, as the major contribution to this positive score seemed to come from data that were measured as electron densities were decreasing as night fell. TING and CMIT both had negative skill scores on the next day and yet this was the day when Figure 6b suggests that TING performed best, insofar as it picked up the twin peaked NmF_2 structure and it mostly picked up the magnitude of these features. The problem that TING had was that it mistimed their occurrence by 1 or 2 h. The result was a poor skill score. CMIT also reproduced these structures, suggesting that they were a fundamental feature of the ionosphere over Learmonth on this day, but negative storm effects occurred in the CMIT simulation, resulting in a bad skill score.

[64] Both TING and CMIT had negative skill scores at Grahamstown on all three days. On the first day this resulted from the IRI-90 model producing better quiet time values of NmF_2. TING NmF_2 values had a positive tendency

on the next day, but greatly overestimated its magnitude, resulting in a negative skill score. CMIT's calculations indicated that NmF_2 values decreased on 15 May, also resulting in a negative skill score. Both models' calculations indicated that large positive storm effects occurred on 16 May, whereas relatively small ones were seen in the data. This caused very bad skill scores for both models on this day. There was no indication at Grahamstown, or at either of the other two Southern Hemisphere stations that TING and CMIT calculated too rapid a fall off of NmF_2 after dark. This is in contrast with the Northern Hemisphere stations.

8. Discussion

[65] This study was undertaken to study how well two versions of the NCAR-TGCM models modeled NmF_2 data in the magnetic middle latitudes during a geomagnetic storm. It is clear from the skill scores that these models often underperform compared with IRI-90, a model that does not even include geomagnetic variations. However, the issues with TING and CMIT are often not as bad as these skill scores suggest. At many locations the comparisons between model and data were quite good, but TING and CMIT output varies considerably from the data for relatively short parts of the day. The causes of some of these excursions and their physical significance are considered in this section.

[66] The first of these issues is the sharp increases in electron density seen in the TING output on 15 May at several Northern Hemisphere stations (Boulder, Wallops Island, Chilton, and Petropavlovsk) and at Hobart. It is clear from Figures 5, 6, and 7 that these increases in electron density are the result of the auroral oval being too far south in this model run insofar as it is above these stations during the storm, whereas the data do not indicate that the auroral oval is this far south. This discovery has lead to a simple improvement in the specification of the expansion of the auroral oval in TING. The relationship between the expansion of the oval and geomagnetic activity has been adjusted to decrease the amount that the oval expands during geomagnetic storms. However, similar excursions, that are seen in the TING output and in the data but are only seen weakly or not at all in the CMIT output, lead us to conclude that accurately modeling the expansion of the auroral oval is critical. This modeling is needed to accurately describe many relatively short timescale variations of NmF_2 at high-middle-latitude stations. This observation is supported by the fact that TING uses smoothed empirical models for the high-latitude inputs, whereas CMIT uses much more variable coupling algorithms. TING energetic electron precipitation, and its convection pattern do vary as the IMF varies, but these inputs vary more in CMIT. The shape of the high-latitude inputs also varies more in CMIT. The only input in which TING shows more variability is in the size of the auroral oval, so this must be the variation that is responsible for the short-term variability in the high middle latitude NmF_2 values. It is possible that CMIT will capture this size variation better when the Rice Convection Model [*Wolf et al.*, 1991] is included. This model will drive a more realistic inner magnetosphere and may permit greater expansion and contraction of the auroral oval.

[67] Another feature of both TING and CMIT output is the tendency for NmF_2 values to decrease far too rapidly after dark in the Northern Hemisphere. There was no indication that this happened in the Southern Hemisphere. Such behavior was noticed previously by *Jee et al.* [2007]. It is not clear why NmF_2 values should fall off so rapidly and, in this case, only occur in the Northern Hemisphere (Jee et al. noticed the fall off in both hemispheres). Likely mechanisms that might drive this rapid decay are too high N_2 densities causing too rapid recombination; too high temperatures or too fast a recombination rate coefficient causing too rapid recombination; a downward particle flux from the plasmasphere that is too weak; an upper boundary that is too low, so the source population for the F2 peak ionosphere is too weak; too rapid ambipolar diffusion taking plasma down into regions where recombination can occur; and neutral winds driving plasma down (or not driving plasma up fast enough) into regions of recombination. All of these possible explanations have advantages and disadvantages.

[68] There is no evidence that N_2 densities are too high, and if they were, why were the nighttime electron densities at Boulder in agreement with data, whereas those at Petropavlosk were not? Some evidence has been put forward recently to indicate that there might be some issues with neutral densities in the model. *Lei et al.* [2007] showed that the winter anomaly in electron densities, which is usually attributed to increases in O density, was stronger in the NCAR-TIEGCM than it was in either the IRI model or in COSMIC data. This overestimation of the winter anomaly may be connected with why the overly rapid decrease in electron density was not seen in the Southern, winter, Hemisphere.

[69] Temperatures in TING and CMIT are in good agreement with MSIS [*Hedin*, 1991] at quiet times. Ion temperatures should be about the same as neutral temperatures after dark, so it is probable that temperatures cannot explain the discrepancy. The reaction rates for recombination have uncertainties associated with them, but there is no evidence that these could lead to the too rapidly declining electron densities after dark.

[70] The next three explanations are linked. TING and CMIT include a flux of ions at the top boundary that is downward at night. This flux was included to provide a better representation of the F_2 peak at night. W. Wang et al. (unpublished manuscript, 2006) have shown that the flux from the SUPIM [*Bailey et al.*, 1997] model is probably not different enough from the flux used here to account for the rapid post-dusk collapse of the CMIT and TING ionosphere. However, this requires further investigation. Variations in ambipolar diffusion are another plausible mechanism for the discrepancy between model and data. It is going to be dependent in part on the plasma temperature. If the heat flux from the plasmasphere is not correct then the electron temperature will not be correct and the ambipolar diffusion calculation will be in error. But it is not clear that this will improve or worsen the representation of nighttime electron densities in the model. The last mechanism, neutral winds, is unlikely insofar as the dusk region is one in which the meridional winds are weak [*Burns et al.*, 1995].

[71] Another issue of importance is the difference between CMIT and TING performance in regions of positive

and negative storm effects at any particular time. The best skill scores occurred in regions where the models and the data both showed negative storm effects, whereas the worst skill scores occurred in regions where either the models predicted positive storm effects or they were seen in data or both. The ability of the models to reproduce negative storm effects suggests two things. First, it is unlikely that the dusk error in the decrease of electron densities is the result of either composition or temperature effects. Composition and temperature changes are much greater where there are negative storm effects than they are for the transition from day to night. Therefore, any errors in the models representation of the compositional and thermal effects on electron density should have manifested itself as a major error in the electron densities in regions where negative storm effects occur in the TING runs. Such differences were not seen, so it is unlikely that the rapid drop in electron densities after dark is a direct result of errors in the composition and temperature calculations. Another explanation for the rapid decrease in electron density in the model after dusk may bee that there is insufficient mass or heat flux flowing into the top boundary of the model after dark. However, preliminary studies of these fluxes suggest that they are of the correct magnitude. This problem is one that we are continuing to work on.

[72] The second thing that the ability to correctly estimate negative storm effects suggest is the following. Figures 8, 9, and 10 show that variations of NmF_2 in the region of negative storm effects were relatively smooth spatially. Thus it was an easier task to reproduce the negative effects that were seen in the data than positive ones. CMIT generally has negative storm effects in most places during and after the storm. These negative storm effects are often deeper than those seen in the data. This suggests that CMIT is overestimating high-latitude energy inputs, although, given the agreement between CMIT output and ionosonde data at a variety of places and times, it does not seem that this overestimation is severe. This information has been used to adjust energy inputs from CMIT to better represent the observed behavior of the ionosphere. This new version of the model has been run recently for other storms and seems to represent high-latitude inputs more accurately.

[73] There is no reason to believe that TING energy input either overestimates or underestimates high-latitude energy inputs. Where negative storm effects occur during the daytime, TING reproduces the magnitude of these effects, suggesting that the energy required for upwelling and the subsequent horizontal transport of the compositional disturbance is reasonable. This suggests that the saturation effect in potential must be included in model runs to correctly evaluate the effects of these large events on the thermosphere and ionosphere.

[74] The issue of positive storms effects is complicated by their mainly occurring in the Southern Hemisphere, where only three ionosondes were available for this study. A further complication is that these ionosondes were frequently near borders between positive and negative storm effects. Such borders are hard to model as they require very accurate representations of complicated physical phenomena. When TING and CMIT do produce positive storm effects at Grahamstown and Learmonth that also occur in the ionosonde data, there is a tendency for the models to greatly overestimate the magnitude of NmF_2, hinting that the maximum strength of positive storm effects in the two models is overestimated. *Codrescu et al.* [1997] showed that a model that does not include an interactive low-latitude dynamo would behave this way.

[75] Another issue relates to when the models performed best and when they performed worst. The majority of positive skill scores occurred on 16 May but so did the largest negative skill scores. This is a little surprising as the storm occurred on 15 May. The thing that the three stations at which TING and CMIT both had positive skills scores had in common was that negative storm effects persisted at these stations into and in two cases through 16 May. This reinforces the point made earlier that the region of negative storm effects was broad and smooth and that the models performed best in this region. The stations at which the models performed worst were Learmonth, Grahamstown, and Puerto Rico. These are the three lowest-latitude stations used in the study. However, the reasons for the poor performance vary from station to station. At Grahamstown both models greatly overestimated the positive storm effects that did occur in the data. At Learmonth CMIT predicted that negative storm effects were occurring on 16 May, whereas positive storm effects occurred in the data. TING produced positive storm effects at Learmonth, but the timing was such that peaks occurred when valleys were seen in the model data and vice versa. At Puerto Rico both models' output suggested that strong negative storm effects were occurring at Puerto Rico, whereas the ionosonde data had recovered to its quiet time values. Both models underestimated electron densities at Puerto Rico on all 3 d, suggesting that the models calculations put Puerto Rico poleward of the equatorial anomaly, when it is actually in the northern portions of the anomaly.

[76] The large errors in the models at a number of locations on 16 May suggests that the recovery period after a geomagnetic storm is difficult to model and that it could be a significant source of errors for ionospheric and thermospheric space weather applications.

[77] Last, the techniques used to calculate skill scores in this study have weaknesses that need to be remedied. Particular problems include that too much importance is placed on a short period of bad comparisons and poor skill scores result when there are relatively small issues with the timing of events. We considered using ratios of the model results to the data when we were doing this work but then realized that this process contains both the weaknesses of the existing method and an additional one. The original weakness is that the data and model are time shifted in a nonregular way. We considered attempting to solve this problem by using ratios of data to models. The additional problem with this technique is that the process of getting ratios for the data places emphasis on places or times where the data are of small magnitude (i.e., mainly at night). Thus skill scores calculated this way will be dominated by night data when absolute differences between the model and the data are small and thus reflect less information about the model performance than our existing skill scores do. We also attempted to attack the problem of time shifts by doing correlations. We stopped doing this when we realized that the daily variations were overwhelming any other signals in these calculations. It may be possible to calculate skill

scores by filtering the data to remove the daily variations and then doing correlations, but this is creating a significant gap between the comparison of raw data and basic model results, which is against the essential philosophy of skill scores: they should be representative of the direct comparisons between models and data. At the moment we can think of no better way to do simple skill scores than the existing techniques. *Owens et al.* [2005] confronted similar shortcomings in skill score studies of solar wind model predictions at 1 AU and developed strategies specific to the character of those data to overcome the problems. Owens et al. used a similar technique as the one used here to provide an initial estimate of the skill. They pointed out the value of developing a scheme that provided a single number to estimate skill. They also developed an event-based approach that considered key features such as the arrival time of the shock and the magnetospheric response. They then defined methods that could be used to define these events. A similar approach should be considered in the thermosphere for future studies of event driven effects in the ionosphere.

9. Conclusions

[78] CMIT is a coupled magnetosphere-ionosphere-thermosphere model that self consistently calculates the interactions between these regions. The stand-alone TING model uses empirical models to provide information about these interactions. In this work CMIT and the stand-alone TING model output have been compared with geomagnetic midlatitude ionosonde data for the 15 May 1997 geomagnetic storm as part of the CISM validation effort. Skill scores have been developed using root mean square calculations and comparing these with the IRI-90 semiempirical model. Generally, the models produced mixed results in attempting to reproduce the ionosonde data. Some reasons for these mixed results are discussed in the following conclusions and are discussed in the previous section. We also point out deficiencies in the use of existing techniques in the skill scores here and in the previous section. A number of conclusions have been drawn from this study:

[79] 1. Skill scores were poor on all 3 d of the investigation. The best skill scores and the worst ones occurred on the recovery day after the storm. On this day skill scores were better where both models predicted negative storm effects and these also occurred in the data and worst in regions where the models predicted positive storm effects and these did not occur. It is clear that there are issues pertaining to the behavior of the ionosphere during storm recovery that need to be addressed in improved models.

[80] 2. Generally, superposed epoch comparisons between the data and the models look better than the skill scores indicated. In several cases skill scores were poor, yet the model traced the data very well for most of the day.

[81] 3. Small variations in the timing of the occurrence of peaks in the model output compared with the data had a disproportionate effect on the skill scores. Skill scores were also poor when the background model performs well. This is not necessarily a reflection on the absolute skill of the primary model; rather it is an indication of whether it can outperform the background model. Some care should be taken when skill scores are applied to ensure that they accurately reflect the performance of the model.

[82] 4. The TING model captured relatively small structures in the high middle latitude stations. This variability was related to the amount of expansion and contraction of the auroral oval, rather than to any other feature. This variability is attributed to the size of the auroral oval because this is the only parameter in which TING is more variable than CMIT is. The size of the auroral oval changes by only small amounts in CMIT, but it changes a lot in TING.

[83] 5. The TING model slightly overestimated the expansion of the auroral oval, modeling an auroral oval that occurred overhead at several high middle latitude stations, whereas the actual auroral oval was poleward of the station. This made a significant contribution to the poor skill scores that TING had on 15 September.

[84] 6. When negative storm effects occurred in the data and model on the day of the main phase of the storm, TING captured these depletions well, whereas CMIT overestimated them somewhat. This suggests that the total energy input in the TING run was approximately the amount needed to force changes in the F_2 region, whereas CMIT overemphasized the energy input that could affect the upper ionosphere. Much of this difference can be ascribed to the application of a saturation level in the TING model, which suggests that the use of saturation levels for potential like that suggested by *Siscoe et al.* [2002] is essential when modeling large events. As this cannot be undertaken readily in coupled magnetosphere-ionosphere-thermosphere models, the physics of this saturation is an important problem that must be solved and incorporated in these models.

[85] 7. Modeled NmF_2 values decreased too rapidly after dark at all six Northern Hemisphere stations but at none of the Southern Hemisphere stations. There are a number of possible causes of this problem, including overestimation of N_2 densities (and hence too much recombination after dark and incorrect mass and heat flux at the upper boundary. There are issues with both of these explanations that have been described in section 8. They will be considered further in future studies.

[86] 8. Both TING and CMIT placed Puerto Rico poleward of the equatorial anomalies, whereas the data indicated that it was within these anomalies throughout the study period.

[87] **Acknowledgments.** This work was supported by the Center for Integrated Space Weather Modeling (CISM), which is funded by the STC Program of NSF under agreement ATM-0120950 and a grant from the National Space Weather Program of NSF. The National Center for Atmospheric Research is sponsored by NSF. Ionosonde data were obtained from the NGDC World Data Center run by NOAA at Boulder, Colorado.

[88] Wolfgang Baumjohann thanks Dieter Bilitza and another reviewer for their assistance in evaluating this paper.

References

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

Bailey, G. J., N. Balan, and Y. Z. Su (1997), The Sheffield University plasmasphere-ionosphere model — A review, *J. Atmos. Sol. Terr. Phys.*, *59*, 1541–1552, doi:10.1016/S1364-6826(96)00155-1.

Bilitza, D. (2007), International Reference Ionosphere 2007, Natl. Space Sci. Data Cent., Greenbelt, Md. (Available at http://modelweb.gsfc.nasa.gov/ionos/iri.html)

Burns, A. G., T. L. Killeen, and R. G. Roble (1991), A theoretical study of thermospheric composition perturbations during an impulsive geomag-

netic storm, *J. Geophys. Res.*, *96*, 14,153–14,167, doi:10.1029/91JA00678.

Burns, A. G., T. L. Killeen, W. Deng, G. R. Carignan, and R. G. Roble (1995), Geomagnetic storm effects in the low- to middle-latitude upper thermosphere, *J. Geophys. Res.*, *100*, 14,673–14,691, doi:10.1029/94JA03232.

Burns, A. G., S. C. Solomon, W. Wang, and T. L. Killeen (2007), The ionospheric and thermospheric response to CMEs: Challenges and successes, *J. Atmos. Sol. Terr. Phys.*, *69*, 77–85, doi:10.1016/j.jastp.2006.06.010.

Codrescu, M. V., T. J. Fuller-Rowell, and I. S. Kutiev (1997), Modeling the *F* layer during specific geomagnetic storms, *J. Geophys. Res.*, *102*, 14,315–14,329, doi:10.1029/97JA00638.

Crowley, G., B. A. Emery, R. G. Roble, H. C. Carlson Jr., and D. J. Knipp (1989), Thermospheric dynamics during September 18–19, 1984: 1. Model simulations, *J. Geophys. Res.*, *94*, 16,925–16,944, doi:10.1029/JA094iA12p16925.

Fuller-Rowell, T. J., M. V. Codrescu, R. J. Moffett, and S. Quegan (1994), Response of the thermosphere and ionosphere to geomagnetic storms, *J. Geophys. Res.*, *99*, 3893–3914, doi:10.1029/93JA02015.

Fuller-Rowell, T. J., M. V. Codrescu, H. Rishbeth, R. J. Moffett, and S. Quegan (1996), On the seasonal response of the thermosphere and ionosphere to geomagnetic storms, *J. Geophys. Res.*, *101*, 2343–2353, doi:10.1029/95JA01614.

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

Fuller-Rowell, T. J., G. H. Millward, A. D. Richmond, and M. C. Codrescu (2002), Storm time changes in the atmosphere at low latitudes, *J. Atmos. Sol. Terr. Phys.*, *64*, 1383–1391, doi:10.1016/S1364-6826(02)00101-3.

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.

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, doi:10.1029/JA087iA08p06339.

Jee, G., A. G. Burns, W. Wang, S. C. Solomon, R. W. Schunk, L. Scherliess, D. C. Thompson, J. J. Sojka, and L. Zhu (2007), Duration of an ionospheric data assimilation initialization of a coupled thermosphere-ionosphere model, *Space Weather*, *5*, S01004, doi:10.1029/2006SW000250.

Killeen, T. L., and R. G. Roble (1988), Thermosphere dynamics driven by magnetospheric sources: contributions from the first five years of the Dynamics Explorer program, *Rev. Geophys. Space Phys.*, *26*(2), 329–367, doi:10.1029/RG026i002p00329.

Killeen, T. L., et al. (1986), Mean neutral circulation in the winter polar *F* region, *J. Geophys. Res.*, *91*, 1633–1649, doi:10.1029/JA091iA02p01633.

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.

Luhmann, J. G., S. C. Solomon, J. A. Linker, J. G. Lyon, Z. Mikic, D. Odstrcil, W. Wang, and M. Wiltberger (2004), Coupled model simulation of a Sun-to-Earth space weather event, *J. Atmos. Sol. Terr. Phys.*, *66*, 1243–1256, doi:10.1016/j.jastp.2004.04.005.

Lyon, J. G., J. A. Fedder, and C. M. Mobarry (2004), The Lyon-Fedder-Mobarry (LFM) global MHD magnetospheric simulation code, *J. Atmos. Sol. Terr. Phys.*, *66*, 1333–1350, doi:10.1016/j.jastp.2004.03.020.

Mendillo, M., X.-Q. He, and H. Rishbeth (1992), How the effects of winds and electric fields in F_2 layer storms vary with latitude and longitude: A theoretical study, *Planet. Space Sci.*, *40*, 595–606, doi:10.1016/0032-0633(92)90001-5.

Owens, M. J., C. N. Arge, H. E. Spence, and A. Pembroke (2005), An event-based approach to validating solar wind speed predictions: High-speed enhancements in the Wang-Sheeley-Arge model, *J. Geophys. Res.*, *110*, A12105, doi:10.1029/2005JA011343.

Peymirat, C., A. D. Richmond, and A. T. Kobea (2000), Electrodynamic coupling of high and low latitudes: Simulations of shielding/overshielding effects, *J. Geophys. Res.*, *105*, 22,991–23,003, doi:10.1029/2000JA000057.

Prölss, G. W. (1980), Magnetic storm associated perturbation of the upper atmosphere: Recent results obtained by satellite-borne gas analyzers, *Rev. Geophys. Space Phys.*, *18*, 183–202, doi:10.1029/RG018i001p00183.

Prölss, G. W. (1981), Latitudinal structure and extension of the polar atmospheric disturbance, *J. Geophys. Res.*, *86*, 2385–2396, doi:10.1029/JA086iA04p02385.

Rees, D., R. Gordon, T. J. Fuller-Rowell, M. Smith, G. R. Carignan, T. L. Killeen, P. B. Hays, and N. W. Spencer (1985), The compositions, structure, temperature, and dynamics of the upper thermosphere in the polar region during October to December, 1981, *Planet. Space Sci.*, *33*, 617–666, doi:10.1016/0032-0633(85)90047-9.

Richmond, A. D., E. C. Ridley, and R. G. Roble (1992), A thermosphere/ionosphere general circulation model with coupled electrodynamics, *Geophys. Res. Lett.*, *19*(6), 601–604, doi:10.1029/92GL00401.

Rishbeth, H. (1975), *F* region storms and thermospheric circulation, *J. Atmos. Terr. Phys.*, *37*, 1055–1064, doi:10.1016/0021-9169(75)90013-6.

Roble, R. G., and E. C. Ridley (1987), An auroral model for the NCAR thermospheric general circulation model (TGCM), *Ann. Geophys.*, *5*, 369–382.

Roble, R. G., B. A. Emery, R. E. Dickinson, E. C. Ridley, T. L. Killeen, P. B. Hays, G. R. Carignan, and N. W. Spencer (1984), Thermospheric circulation, temperature, and compositional structure of the Southern Hemisphere polar cap during October–November, 1981, *J. Geophys. Res.*, *89*, 9057–9068, doi:10.1029/JA089iA10p09057.

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*(12), 1325–1328, doi:10.1029/GL015i012p01325.

Schunk, R. W., and A. F. Nagy (2000), *Ionospheres: Physics, Plasma Physics, and Chemistry*, Cambridge Univ. Press, Cambridge, U.K.

Seaton, M. J. (1956), A possible explanation of the drop in F-region critical densities accompanying major ionospheric storms, *J. Atmos. Terr. Phys.*, *8*, 122–124, doi:10.1016/0021-9169(56)90102-7.

Siscoe, G. L., G. M. Erickson, B. U. Ö. Sonnerup, N. C. Maynard, J. A. Schoendorf, K. D. Siebert, D. R. Weimer, W. W. White, and G. R. Wilson (2002), Hill model of transpolar potential saturation: Comparisons with MHD simulations, *J. Geophys. Res.*, *107*(A6), 1075, doi:10.1029/2001JA000109.

Spence, H., D. Baker, A. Burns, T. Guild, C.-L. Huang, G. Siscoe, and R. Weigel (2004), Center for integrated space weather modeling metrics plan and initial model validation results, *J. Atmos. Sol. Terr. Phys.*, *66*, 1499–1508, doi:10.1016/j.jastp.2004.03.029.

Wang, W., T. L. Killeen, A. G. Burns, and R. G. Roble (1999), A high-resolution, three dimensional, time dependent, nested grid model of the coupled thermosphere-ionosphere, *J. Atmos. Terr. Phys.*, *61*, 385–397, doi:10.1016/S1364-6826(98)00079-0.

Wang, W., M. Wiltberger, A. G. Burns, S. Solomon, T. L. Killeen, N. Maruyama, and J. Lyon (2004), Initial results from the CISM coupled magnetosphere-ionosphere-thermosphere (CMIT) model: Thermosphere ionosphere responses, *J. Atmos. Sol. Terr. Phys.*, *66*, 1425–1442, doi:10.1016/j.jastp.2004.04.008.

Wang, W., A. G. Burns, M. Wiltberger, S. C. Solomon, and T. L. Killeen (2007), An analysis of neutral wind generated currents during geomagnetic storms, *J. Atmos. Sol. Terr. Phys.*, *69*, 159–165, doi:10.1016/j.jastp.2006.06.014.

Wiltberger, M., W. Wang, A. Burns, S. Solomon, J. G. Lyon, and C. C. Goodrich (2004), Initial results from the coupled magnetosphere ionosphere thermosphere model: Magnetospheric and ionospheric responses, *J. Atmos. Sol. Terr. Phys.*, *66*, 1411–1444, doi:10.1016/j.jastp.2004.03.026.

Wolf, R. A., R. W. Spiro, and F. J. Rich (1991), Extension of the Rice Convection Model into the high-latitude ionosphere, *J. Atmos. Terr. Phys.*, *53*, 817–829, doi:10.1016/0021-9169(91)90096-P.

A. G. Burns, T. L. Killeen, S. C. Solomon, W. Wang, and M. Wiltberger, High Altitude Observatory, NCAR, P. O. Box 3000, Boulder, CO 80307, USA. (aburns@ucar.edu)

J. E. Landivar and R. E. Lopez, Department of Physics and Space Sciences, Florida Institute of Technology, 150 W. University Boulevard, Melbourne, FL 32901, USA.

H. Spence, Department of Astronomy, Boston University, 625 Commonwealth Avenue, Boston, MA 02215, USA.