

Thermosphere model assessment for geomagnetic storms from 2001 to 2023

Sean Bruinsma* and Sophie Laurens

GET-CNES, Space Geodesy Office, 14 Av. Edouard Belin, 31400 Toulouse, France

Received 29 April 2024 / Accepted 14 August 2024

Abstract—We present an updated study for thermosphere model assessment under geomagnetic storm conditions, defined when the geomagnetic index $a_p = 80$ or larger. Comparisons between five empirical models, NRLMSISE-00, JB2008, and three versions of DTM2020, and CHAMP, GRACE, GOCE, Swarm-A, and GRACE-FO neutral density data sets for 152 storms are presented. The storms are categorized according to a_p , as single peak or multiple peaks. After applying a model debiasing procedure using the density data just before the onset of a storm, the models are on average only slightly biased, often only a few percent. This is an unexpected and reassuring result for these relatively simple models, which were fitted to different observations. The standard deviations of these averages are however up to 12% (1-sigma), which places the small biases into perspective. The smallest biases are achieved at the lowest altitude when comparing with GOCE data, and the highest for GRACE. The best results, i.e. smallest bias and standard deviation on average over all single-peak storms, over the entire 4-Phase storm period are obtained with DTM2020_Intermediate and DTM2020_Research models, while the oldest model, NRLMSISE-00, is the least precise. However, NRLMSISE-00 is the least biased when compared to multiple-peak storms. As could be expected, multiple-peak storms are reproduced with less precision than single-peak storms. The assessment reveals that model precision decreases with altitude, but that bias is independent of altitude, at least in the range covered by the data, 250–550 km.

Keywords: Thermosphere / Model assessment / Geomagnetic storm / Satellite drag

1 Introduction

Semi-empirical (SE) thermosphere models are used operationally in the determination and prediction of the ephemerides of active satellites and orbital debris. Conjunction analysis is becoming a major issue with the fast-growing number of objects in space. Notably for geomagnetic storms, good knowledge of thermosphere model performance, or the uncertainty, is therefore necessary. The accuracy of the determination and prediction of ephemerides of objects in Low Earth Orbit (LEO; altitudes lower than 1000 km) depends largely on the quality of the atmospheric drag modeling. The drag force varies mainly due to the fluctuations in total neutral density, and to a much lesser degree also due to satellite characteristics (Doornbos, 2011; Mehta et al., 2017) and atmospheric composition and temperature. The variability of the thermosphere is driven by the changing solar Extreme Ultra-Violet (EUV) emissions and the changing Joule heating and particle precipitation due to the interaction of the magnetosphere with the solar wind (commonly referred to as “solar activity” and “geomagnetic activity”,

respectively), whereas a small part is due to upward propagating perturbations that originate in Earth’s lower atmosphere. The accuracy of the forecast of the drivers is crucial to the accuracy of the density forecast. We do not address errors in the driver forecasts in this paper, but the model performance with observed drivers (or perfect forecasts of the proxies), i.e. we quantify the precision of the combination of the model and the proxies it employs.

In order to assess and track the progress over time of current and future thermosphere models, appropriate metrics and high-quality neutral density data sets, preferably with high-spatial resolution covering long intervals of time (i.e. years, and ideally a complete solar cycle), are required. In a previous paper (Bruinsma et al., 2021), specific storm-time metrics were introduced and described and then applied to a small number (13) of isolated storms. The neutral density data used in that study to verify model precision in space and time has been augmented with data from the missions Swarm-A (van den Ijssel et al., 2020) and GRACE-FO (Siemes et al., 2023), whereas a more precise GRACE dataset (Siemes et al., 2023) replaces the older one. In this study, a total of 152 storms from 2001 to 2023, defined as events with a K_p of 6, i.e., $a_p = 80$ (units: 2 nT),

*Corresponding author: sean.bruinsma@cnes.fr

and larger, is used to comprehensively assess the CIRA (COSPAR International Reference Atmospheres) models NRLMSISE-00 (hereafter MSIS00; Picone et al., 2002), JB2008 (Bowman et al., 2008), and DTM2020 (Bruinsma and Boniface, 2021). In this study, a modification of the storm-time metrics was necessary to accommodate also double or multiple-peaked storms, e.g., the 2003 Halloween storm.

The density data and the adopted metrics allow objective benchmarking of the thermosphere models, quantification, and detailed descriptions of the improved (or degraded) performance. A major goal of this study is to provide CCMC (Community Coordinated Modeling Center: <https://ccmc.gsfc.nasa.gov>) with the list of storms, and the density data, which will then be used to assess all thermosphere models that run there. Thermosphere model scorecards can then be published, applying consistent and always identical metrics, which can help users select the best model for their objective.

The following section briefly presents the models, the density data, and storm-time metrics with a small modification for double-peaked storms. The assessment results of the three models are presented in Section 3. After a summary, the conclusions are given in Section 4.

2 Model descriptions

2.1 Semi-empirical thermosphere models

Semi-empirical models are used in orbit computation and mission design and planning because they are easy to use and computationally fast. They provide density and temperature estimates for a single location and time (e.g., satellite positions along an orbit). These climatological (or “specification”) models have a low spatial resolution of the order of thousands of kilometers and a temporal resolution of 1–3 h that is imposed by the cadence of the geomagnetic indices. Some kind of storm pre-conditioning is taken into account through a combination of delayed and averaged geomagnetic indices, notably in MSIS00, and the resulting density estimate is based on the statistical fits to those drivers. SE models are constructed by optimally estimating the model coefficients to their respective underlying databases (some combination of total density, temperature, and composition measurements). As a consequence, the SE thermosphere models adopt the scales of the satellite models used in the respective databases (e.g., DTM2020 based on TU Delft satellite models, JB2008 based on HASDM satellite models) – and these are rarely consistent between modelers as was shown in Bruinsma et al. (2022).

The DTM2020-Operational model (DTM2020_Ops) was developed recently, using F10.7 and K_p as drivers. DTM2020_Int is an unpublished intermediate model using F30 but still K_p as drivers that were created in the development of the model DTM2020-Research (DTM2020_Res), which is too complicated to run operationally. The backbone of the DTM2020 models is the complete CHALLENGING Mini satellite Payload (CHAMP; Doornbos, 2011), Gravity Recovery and Climate Experiment (GRACE; Bruinsma, 2015) and Gravity field and steady-state Ocean Circulation Explorer (GOCE; Bruinsma et al., 2014) high-resolution accelerometer-inferred density datasets, as well as Swarm A densities (van den Ijssel et al., 2020). The JB2008 model was developed by fitting to

10 years of US Air Force daily-mean densities, 5 years of output of the U.S. Space Force High Accuracy Satellite Drag Model (HASDM; Storz et al., 2005), 5 years of CHAMP, and 4 years of GRACE data. The files with the solar drivers, which are computed and sometimes modified by the modelers, are regularly updated. The driver files (SOLFSGMY and DTGFILE) that were online in January 2024 were used for the assessment given in this paper. MSIS00 is used in this assessment because the difference with MSIS 2.0 (Emmert et al., 2020) is very small for geomagnetic storms and in the form of a bias (the geomagnetic model algorithm was not updated) and because it is widely used. MSIS00 was constructed with a mass spectrometer, incoherent scatter radar, historical accelerometer total densities, UV occultation, and various rocket measurements, along with global daily-mean orbit-derived thermospheric density and tabular lower atmospheric temperature. The time resolution and drivers of the models are listed in Table 1.

2.2 Selected density data

The density variability in the thermosphere is very large, typically hundreds of percent, on long time scales (the solar cycle) but also on short time scales of hours to days in the event of geomagnetic storms. The variations in density depend on position, season, and the level of solar and geomagnetic activity. The high-resolution accelerometer-inferred densities and the GPS-inferred densities are in situ, along the satellite orbits, but a single satellite still provides a rather poor spatial distribution: the latitudinal extent depends on the orbital inclination, the local time coverage is essentially limited to the time of its ascending and descending pass and the precession of the orbital plane (negligible throughout a geomagnetic storm). While the measurement cadence typically is 10 s, a satellite passes the same latitude and local time (but at a *different longitude*) once per orbit of roughly 100 min.

Table 2 lists the essential information of the five selected datasets, CHAMP (Doornbos, 2011), GRACE and GRACE-FO (GFO; Siemes et al., 2023), GOCE (Bruinsma et al., 2014), and Swarm-A (van den Ijssel et al., 2020). These precise density datasets are selected because they are appropriate for storm-time evaluation, notably precise measurements from pole to pole. The Swarm-A densities are derived using precision orbit determination and have an along-track resolution of about 1000 km at best when solar and geomagnetic activity is high. The accelerometer-inferred densities have a constant resolution of about 600 km after lowpass filtering without data rejection.

The model assessment is performed by means of comparing to densities of 152 storms with a K_p of 6, i.e., $a_p = 80$, and higher. Table A1 lists the dates, minimum Dst, and maximum a_p of the storms, and which satellites monitored the storms.

2.3 Updated storm-time metrics for model-data comparison

Storms in this analysis are defined as $K_p = 6$ or larger ($a_p = 80$ or larger). In Bruinsma et al. (2021), the storms were divided into four phases of fixed length, two before and two after the minimum Dst associated with the storm defined through K_p : pre-storm (1), onset (2), recovery (3), and post-storm (4). An update of the above phases is necessary in this study in order to accommodate storms with more than one peak

Table 1. Thermosphere and upper atmosphere models used in the assessment. The temporal resolution is listed in the last column.

	Drivers (solar geomagnetic)	Resolution
NRLMSISE-00	F10.7 ap (array of 7 values)	3 h
JB2008	S10, F10.7, M10, Y10 Dst, ap	1 h
DTM2013	F30 Kp (24 h-mean and 3-h delayed)	3 h
DTM2020_Oper	F10.7 Kp (24 h-mean and 3-h delayed)	3 h
DTM2020_Interm	F30 Kp (24 h-mean and 3-h delayed)	3 h
DTM2020_Research	F30 Hp60 (array of 7 values)	1 h

Table 2. Datasets selected for the model assessment under storm conditions. (*i* is inclination, and LST is the local solar time coverage and the approximate period, in days, to cover 24 h).

	Period	Altitude (km)	<i>i</i>	LST24 h LST	Precision (%)
CHAMP	05/2001–08/2010	450–250	87°	0–24120–130	1–4%
GRACE	08/2002–07/2016	490–300	89°	0–24120–160	2–6%
GOCE	11/2009–10/2013	270–180	96°	6–8 & 18–20	1–3%
Swarm-A	01/2014–12/2023	510–435	87°	0–24115–135	6–30%*
GRACE-FO	06/2018–10/2023	510–490	89°	0–24160–165	3–8%

*The precision is strongly dependent on the level of drag, i.e., solar cycle and altitude. It is also for a resolution of a few thousand km vs 600 km for the other satellites.

in geomagnetic activity of Kp = 6 or larger in phases 3–4 (at least 9 h apart), of which the Halloween 2003 storm is a much-studied case. The intervals of such multiple-peaked storms can be unambiguously defined using ap/Kp, which was preferred over a more complicated definition based on both Dst and ap. For consistency, we modified the definition of t0 of the storm phases for all storms based on ap reaching 80, and Dst is no longer used. Figure 1 (top) illustrates the four phases of a single-peak (SP) storm. Phase 1, the pre-storm interval, is used to de-bias the models with respect to the observations via a scale factor, which is then applied to the model densities in all four phases. This de-biasing procedure is used to minimize the effect of non-storm-related model differences and errors on the assessment. Density data for the 108 single-peak (SP) storms are selected 30-h before to 48-h after the time ap reaches 80, which now defines t0 and the end of Phase 2 (storm onset). Phase 3 is the main and recovery phase, and Phase 4 is the post-storm phase. Figure 1 (bottom) illustrates the phases for double or multiple-peaked (MP) storms, which in this example is for 10–16 July 2004. The t0 for the 44 MP storms is defined for SP storms as the time when ap reaches 80. Phase 3 for MP storms, in Figure 1 (bottom) due to the second time ap = 80 is reached at $t = 1.4$, is extended by a variable duration based on when ap falls below 80 again (at t almost 3.0 in this example) plus 36-h. Phase 4 then extends 12 h after the end of Phase 3. The phases and their durations for SP and MP (italic) storms are computed as listed below with respect to t0.

Phase 1: t0 minus 30-h to t0 minus 18-h	<i>t0 minus 30-h to t0 minus 18-h</i>
Phase 2: t0 minus 18-h to t0	<i>t0 minus 30-h to t0 minus 18-h</i>
Phase 3: t0 plus 36-h	<i>t0 plus variable duration plus 36-h</i>
Phase 4: t0 plus 36-h to t0 plus 48-h	<i>end of Phase 3 plus 12-h</i>

Because of the very large range in density values due to differences in altitude and solar activity (i.e. phase of the solar cycle), it is difficult to analyze and interpret a model's

performance using absolute values. Therefore, relative precision is preferred in the form of ratios that are easy to comprehend. The models and data are compared by computing the observed-to-computed density ratios (O/C). The mean μ and standard deviation σ of the observed-to-computed density ratios, due to their distribution, are computed in log space following Sutton (2018):

$$\text{mean} \frac{O}{C} = \mu = \exp \left(\frac{1}{N} \sum_{n=1}^N \ln \frac{O_n}{C_n} \right)$$

$$\text{standard deviation} \frac{O}{C} = \sigma = \sqrt{ \frac{1}{N} \sum_{n=1}^N \left(\ln \frac{O_n}{C_n} - \ln \text{mean} \left(\frac{O}{C} \right) \right)^2 } \quad (1)$$

where N is the number of observations. A ratio of one indicates that on average the observations are reproduced perfectly, i.e. an unbiased model. Deviation from unity signifies under (larger than one) or overestimation (smaller than one) of the model. The standard deviation (StD) of the observed-to-computed density ratios σ , computed as percentage by multiplying by 100, represents the combination of the ability of the model to reproduce observed density variations, and the geophysical noise (e.g. waves, the short duration effect of large flares) and instrumental noise in the observations. The Pearson correlation coefficients R of observed and modeled densities are also computed, which are independent of model bias; R^2 represents the fraction of observed variance reproduced by the model.

The number of SP and MP storms identified in this study taking the available satellite density data into account is 108 and 44, respectively. Storms for which data gaps are present in any of the four phases were rejected. Storms are often observed by more than one spacecraft, and the total number of SP and MP storm comparisons is 193 (CHAMP 48, GRACE 72, GOCE 16, Swarm-A 42, and GFO 15) and 79 (CHAMP 31, GRACE 35, GOCE 2, Swarm-A 9, and GFO 2), respectively.

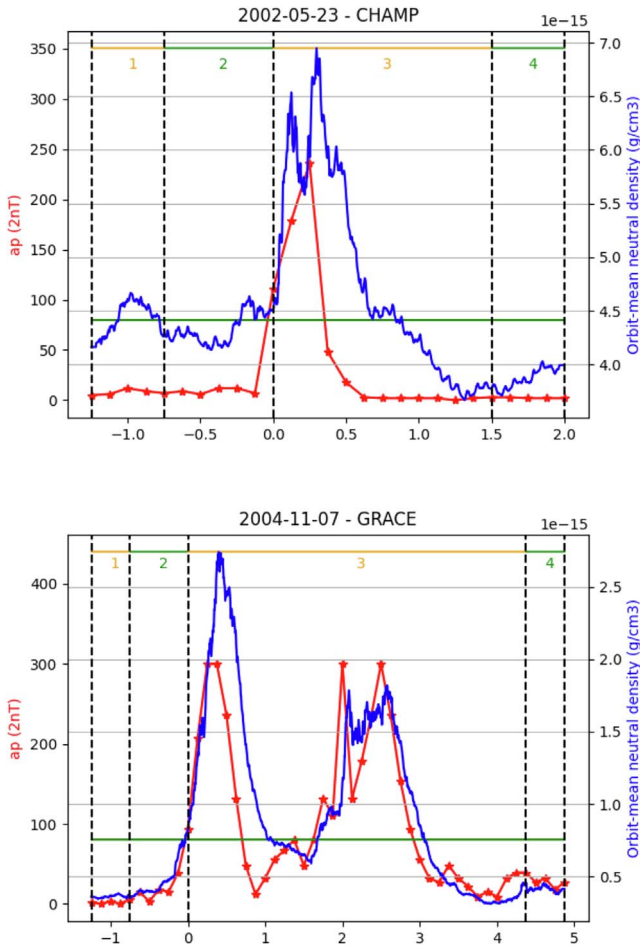


Figure 1. The four phases of the assessment interval for SP (top) and MP (bottom) storms, with t_0 centered on the time of the first peak in a_p with a minimum of 80.

The comparisons of each model with satellite density data are computed in three steps:

1. Calculate the model debiasing scale factor with density data in Phase 1 per storm and per satellite (e.g., 48 scale factors for CHAMP and SP storms);
2. Apply debiasing and calculate $\mu_{i,s}$, $\sigma_{i,s}$, and $R_{i,s}$, where $i = 1$ to the number of storms per satellite s (e.g., i is 48 for CHAMP and SP storms) for the four phases separately and all four together. Observations that are ten times smaller or bigger than model densities are rejected;
3. Calculate the average of the mean O/C and standard deviation of the $\mu_{i,s}$ per satellite again with Eq. (1) (so calculated with the 48 $\mu_{i,s}$ in case of CHAMP and SP storms), referred to hereafter as $\langle O/C \rangle$ and $\text{StD } \langle O/C \rangle$, respectively. Calculate the arithmetic average of the StD of O/C (i.e., of the $\sigma_{i,s}$), and the arithmetic average of $R_{i,s}$ as computed in step 2.

3 Storm-time assessment results

Scale offsets between models are mainly due to the ingested density data. Therefore, assessment of the performance of

models over geomagnetic storms requires debiasing. Table 3 presents the average of the mean O/C ($\langle O/C \rangle$) and the StD of all mean O/C ($\text{StD } \langle O/C \rangle$) of Phase 1, i.e. the mean of the scale factors applied to the models and their StD. The scale factors for the DTM models are close to unity because they were fitted to most of these datasets. MSIS00 and JB2008 were not fitted to the TU Delft densities, and consequently, their scale factors are considerable, which demonstrates the essential nature of debiasing to compare fairly.

For users of thermosphere models, the main result of the assessment is the model performance over entire storms, i.e. from phase 1 to phase 4. The average of the 193 mean O/C ($\langle O/C \rangle$) and the StD of all mean O/C ($\text{StD } \langle O/C \rangle$) is displayed in Figure 2 for all SP storms and per satellite for the five models. On average, the models perform rather well during storms, and the $\langle O/C \rangle$ per satellite density dataset are within 5% of unity for all models excepting MSIS00. Overall, the models are only slightly underestimating during storms, by 1–5%. So on first sight model performance is fairly equal. However, the $\text{StD } \langle O/C \rangle$ indicated by the bars in Figure 2 shows that the spread around the biases is not equal. The smallest $\text{StD } \langle O/C \rangle$ are obtained with DTM2020_Int and DTM2020_Res. Figure 3 shows the arithmetic average StD of O/C for all SP storms and per satellite for the five models, and the bars represent the StD of the StD of O/C. The StD of O/C increases from GOCE to CHAMP to GRACE/Swarm-A/GFO, which clearly shows the dependency of model performance on altitude. It also clearly shows that the DTM models reproduce the storm variations with significantly smaller errors than MSIS00 and JB2008. However, a better StD of O/C does not necessarily improve the model performance when calculating satellite orbits, which is provided by $\langle O/C \rangle$.

The average of the 79 mean O/C ($\langle O/C \rangle$) and the StD of all mean O/C ($\text{StD } \langle O/C \rangle$) is displayed in Figure 4 for all MP storms and per satellite for the five models. On average, the models perform well again during MP storms, and the $\langle O/C \rangle$ per satellite is within 6% of unity for all models. Contrary to what was observed for SP storms, the models are overestimating MP storms by a few percent (NB: only CHAMP and GRACE results are based on a large number of storms). The model performance is significantly reduced for MP storms, which is reflected by the larger $\text{StD } \langle O/C \rangle$ compared with SP storms (Fig. 2): for GRACE for example, it increases from 7.5% to 9.9% (10.1% to 12.1%) for DTM2020_Res (JB2008). The arithmetic average StD of O/C for all MP storms shown in Figure 5 is also significantly larger than those obtained for SP storms shown in Figure 3, which confirms that the MP storms are reproduced with lower fidelity than SP storms.

The complete statistics per phase are given in tables, which enable identifying the strong and weak points of the models more specifically, which is probably not very pertinent for operational issues, but interesting for modelers. Table 4 lists the $\langle O/C \rangle$ and $\text{StD } \langle O/C \rangle$ of the five models for the five density datasets and all single-peak storms. The results, per model, are listed for all four phases (bold), Phase 2, Phase 3, and Phase 4 (from top to bottom). Results for Phase 1 are not pertinent due to the debiasing and are not listed. The bold-faced numbers are displayed in Figure 2. The performance of DTM2020_Ops stands out for GFO, for which it is significantly more precise than DTM2020_Int and DTM2020_Res. The $\langle O/C \rangle$ per phase of DTM2020_Ops and DTM2020_Int do not present a specific

Table 3. The $\langle O/C \rangle$ and StD $\langle O/C \rangle$ for Phase 1, for debiasing, per model for the five density datasets and all 152 storms (single and multiple-peak storms).

	DTM20_Ops		MSIS00		JB2008		DTM20_Int		DTM20_Res	
GOCE	0.97	6.6	0.83	5.9	0.80	6.0	0.97	4.5	0.98	4.0
CHAMP	1.02	14.2	0.76	11.6	0.83	9.6	1.06	12.5	1.04	12.3
GRACE	1.06	19.2	0.77	13.8	0.87	11.3	1.09	15.8	1.06	15.2
SwarmA	1.07	15.5	0.76	15.4	1.03	32.5	1.09	21.7	1.06	20.3
GFO	1.06	19.1	0.78	18.0	1.28	52.0	1.11	19.6	1.08	19.0

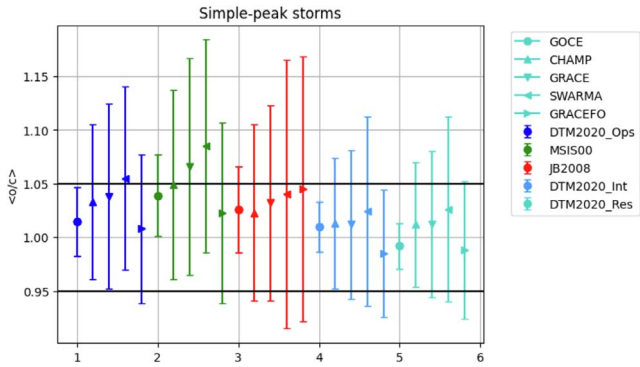


Figure 2. $\langle O/C \rangle$ and StD $\langle O/C \rangle$ of phases 1–4 of the 108 SP storms, per satellite for the five models.

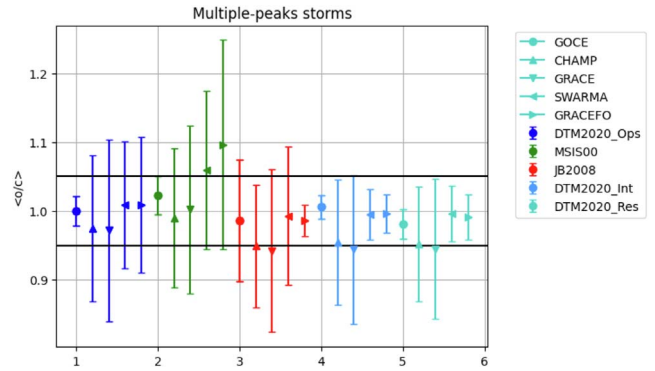


Figure 4. $\langle O/C \rangle$ and StD $\langle O/C \rangle$ of phases 1–4 of the 44 MP storms, per satellite for the five models.

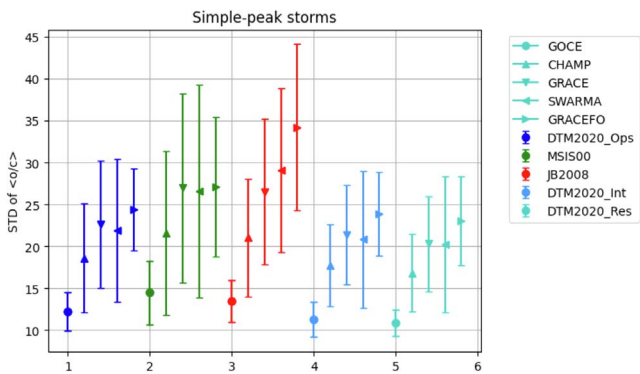


Figure 3. The arithmetic average StD of $\langle O/C \rangle$ of the 108 SP storms and per satellite for the five models.

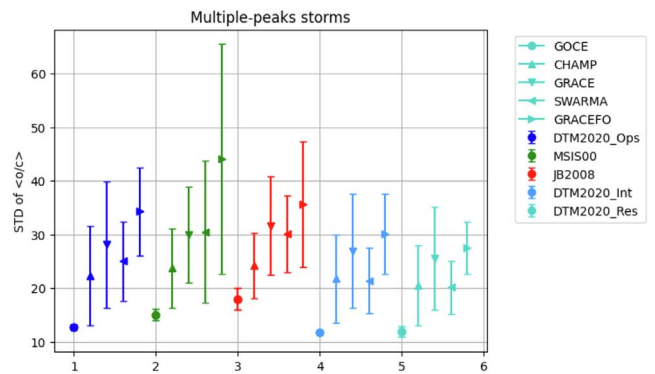


Figure 5. The arithmetic average StD of $\langle O/C \rangle$ of the 44 MP storms and per satellite for the five models.

tendency for Phases 2–4. MSIS00 on the other hand underestimates Phase 3 (recovery) and then overestimates in Phase 4 (post-storm). Except for GOCE, JB2008 underestimates Phase 2, the onset. DTM2020_Res presents an overestimation in Phase 4. The StD $\langle O/C \rangle$ is predominantly smallest during Phase 2 (onset) for all models. It is mostly largest in Phase 3 for MSIS00 and Phase 4 for JB2008 and the DTM models.

Table 5 lists the average StD of $\langle O/C \rangle$ and the average correlation coefficients of the five models for the five density datasets and all single-peak storms. Results are listed again for all four phases (bold), Phase 2, Phase 3 and Phase 4 (from top to bottom). The bold-faced average StD is displayed in Figure 3. DTM2020_Res now clearly has the smallest StD and highest correlations, also for GFO (not used in DTM), even if DTM2020_Ops results are very close. The smallest StD of

$\langle O/C \rangle$ is most often seen in Phase 4 for all models, i.e. in the post-storm phase, and the highest mostly in Phase 2 for the DTM models and JB2008, while it is seen in Phase 3 for MSIS00. For all models, the highest correlation is always seen for Phase 4, and lowest for Phase 2. The StD and correlation also reveal a small dependency on altitude, i.e. in the order GOCE-CHAMP-GRACE. While slightly higher, Swarm-A StD and correlations results are better again than for GRACE, but this is due to the different nature of the densities, i.e. low-resolution GPS-inferred densities versus high-resolution accelerometer-inferred densities. GFO StD and correlation results are also better than for GRACE, and this is most likely due to the much smaller number of storms for GFO.

Table 6 lists the $\langle O/C \rangle$ and StD $\langle O/C \rangle$ of the five models for the five density datasets and all multiple-peak storms.

Table 4. The $\langle O/C \rangle$ and StD $\langle O/C \rangle$ per model for the five density datasets and the 108 single-peak storms. Per model are given the results of all four phases (bold), Phase 2, Phase 3, and Phase 4.

	DTM20_Ops		MSIS00		JB2008		DTM20_Int		DTM20_Res	
GOCE	1.01	3.2	1.04	3.8	1.03	4.0	1.01	2.3	0.99	2.1
	1.01	3.3	1.00	3.9	1.00	4.2	1.01	3.5	0.99	3.7
	1.02	5.9	1.08	6.7	1.05	6.1	1.01	5.0	0.98	4.4
	1.02	5.8	1.02	6.2	1.03	6.6	1.02	4.2	1.01	4.1
CHAMP	1.03	7.2	1.05	8.8	1.02	8.2	1.01	6.1	1.01	5.8
	1.03	6.0	1.01	6.2	1.04	10.7	1.02	5.7	1.01	5.7
	1.05	11.1	1.11	14.9	1.03	12.1	1.02	9.5	1.02	9.0
	1.01	12.2	1.00	13.0	1.00	10.8	0.99	10.2	0.99	9.7
GRACE	1.04	8.6	1.07	10.1	1.03	9.1	1.01	6.9	1.01	6.8
	1.03	8.0	1.01	7.8	1.05	12.3	1.02	7.3	1.01	7.4
	1.07	13.1	1.14	16.9	1.04	12.9	1.03	10.4	1.03	10.2
	1.01	14.8	1.02	15.7	1.02	14.1	0.99	12.6	0.99	12.3
SwarmA	1.05	8.5	1.08	9.9	1.04	12.5	1.02	8.8	1.03	8.6
	1.06	8.6	1.03	8.8	1.07	12.8	1.04	8.1	1.03	7.7
	1.08	12.7	1.17	16.6	1.04	17.7	1.04	13.5	1.04	13.3
	1.02	13.8	1.02	14.7	1.02	16.4	0.99	12.9	0.99	12.8
GFO	1.01	6.9	1.02	8.4	1.04	12.3	0.98	5.9	0.99	6.4
	1.02	8.4	0.98	7.6	1.09	10.3	1.00	7.6	1.00	7.4
	1.02	9.4	1.08	12.8	1.03	17.1	0.98	8.0	0.99	8.9
	0.98	15.1	0.94	14.4	1.04	19.0	0.97	12.8	0.96	12.6

Table 5. The average StD of the O/C and correlation coefficient per model for the five density datasets of the 108 single-peak storms. Per model are given the results of all four phases (bold), Phase 2, Phase 3 and Phase 4. Green indicates best performance.

	DTM20_Ops		MSIS00		JB2008		DTM20_Int		DTM20_Res	
GOCE	12.3	0.92	14.5	0.91	13.5	0.92	11.3	0.93	10.9	0.94
	10.0	0.95	12.6	0.92	12.1	0.92	10.0	0.94	9.8	0.95
	15.2	0.89	17.3	0.90	15.6	0.90	13.8	0.91	13.3	0.92
	10.0	0.97	12.2	0.94	12.3	0.94	8.7	0.97	8.3	0.97
CHAMP	18.6	0.90	21.6	0.89	21.0	0.89	17.7	0.91	16.8	0.92
	17.0	0.90	18.3	0.89	22.0	0.87	16.5	0.90	15.9	0.91
	21.3	0.88	25.6	0.88	22.8	0.87	20.1	0.89	18.9	0.90
	16.5	0.94	18.5	0.94	17.8	0.92	15.3	0.94	14.7	0.94
GRACE	22.6	0.90	27.0	0.88	26.5	0.87	21.4	0.90	20.3	0.91
	21.5	0.88	22.9	0.88	28.4	0.83	20.7	0.89	19.8	0.90
	25.6	0.87	32.3	0.85	28.4	0.85	23.9	0.87	22.4	0.89
	20.0	0.94	22.7	0.93	22.8	0.90	18.9	0.94	18.4	0.94
SwarmA	21.9	0.92	26.6	0.89	29.1	0.89	20.8	0.92	20.2	0.93
	21.5	0.90	23.6	0.89	30.3	0.82	19.7	0.92	19.1	0.92
	23.4	0.90	30.4	0.87	30.9	0.88	22.5	0.90	21.8	0.92
	20.5	0.93	23.4	0.92	26.8	0.88	19.9	0.93	19.5	0.93
GFO	24.4	0.91	27.1	0.90	34.2	0.88	23.9	0.91	23.1	0.92
	23.8	0.90	25.2	0.90	34.4	0.86	23.1	0.91	21.9	0.92
	26.9	0.89	30.5	0.89	37.1	0.87	26.4	0.89	25.6	0.90
	23.5	0.95	25.8	0.93	30.9	0.90	22.6	0.94	21.8	0.95

The Phase 3 results listed here for the multiple-peak storms are based on longer intervals (see Fig. 1). The bold-faced numbers are displayed in Figure 4. Note that GOCE and GFO results are only based on 2 multiple-peak storms, and therefore not statistically significant. CHAMP and GRACE have the smallest biases now with MSIS00 (NB: where they are largest for the SP storms), and Swarm-A with JB2008 and the DTM models. All models reveal a clear offset between CHAMP and GRACE compared with Swarm-A. The StD $\langle O/C \rangle$ for MP storms is noticeably higher than for SP storms.

Table 7 lists the average StD of O/C and the average correlation coefficients of the five models for the five density datasets and all multiple-peak storms. The bold-faced average StD is displayed in Figure 5. DTM20_Res again clearly has smallest StD and highest correlations. The smallest StD of O/C is as for SP storms most often found for Phase 2, while a clear tendency cannot be gleaned from Phases 3 and 4. For all models, the highest correlation is as for SP storms always seen in Phase 4, it is lowest in Phase 2 for CHAMP and GRACE, and mostly lowest in Phase 3 for Swarm-A.

Table 6. The <O/C> and StD <O/C> per model for the five density datasets and the 44 multiple-peak storms. Per model are given the results of all four phases (bold), Phase 2, Phase 3, and Phase 4.

	DTM20_Ops		MSIS00		JB2008		DTM20_Int		DTM20_Res	
GOCE	1.00	2.2	1.02	2.8	0.99	8.8	1.01	1.7	0.98	2.2
	0.98	0.8	0.95	0.3	1.01	1.6	0.99	2.5	0.99	3.5
	1.02	2.9	1.06	4.0	0.99	12.7	1.02	2.0	0.98	2.6
	0.94	3.7	0.95	6.4	0.99	12.2	0.95	2.7	0.94	3.4
CHAMP	0.97	10.6	0.99	10.1	0.95	8.9	0.95	9.1	0.95	8.3
	1.02	5.3	0.99	5.6	1.00	8.1	1.01	5.3	1.00	5.4
	0.97	13.7	1.00	14.0	0.94	12.1	0.94	12.0	0.94	10.9
	0.93	14.4	0.93	11.9	0.90	9.5	0.92	11.5	0.91	11.0
GRACE	0.97	13.2	1.00	12.2	0.94	11.8	0.94	10.8	0.94	10.2
	1.02	7.2	0.99	7.9	1.00	11.5	1.00	7.3	1.00	7.1
	0.97	17.1	1.02	16.8	0.93	15.6	0.93	14.3	0.94	13.5
	0.91	18.2	0.92	14.7	0.91	12.9	0.90	12.9	0.89	12.8
SwarmA	1.01	9.2	1.06	11.5	0.99	10.0	0.99	3.7	1.00	4.0
	1.06	8.5	1.02	8.6	1.03	12.7	1.06	7.3	1.05	5.9
	1.03	14.1	1.12	18.9	0.99	14.2	0.99	7.5	0.99	7.9
	0.88	9.6	0.92	12.1	0.93	8.8	0.92	6.3	0.92	5.7
GFO	1.01	9.8	1.10	15.2	0.99	2.3	1.00	2.8	0.99	3.3
	1.08	9.2	1.06	11.0	0.98	13.3	1.08	12.5	1.07	11.2
	1.03	20.0	1.17	30.0	1.02	8.2	1.02	9.2	1.00	9.2
	0.85	7.8	0.97	15.1	0.85	4.9	0.84	3.0	0.84	4.4

Table 7. The average StD of the O/C and correlation coefficient per model for the five density datasets and the 44 multiple-peak storms. Per model are given the results of all four phases (bold), Phase 2, Phase 3, and Phase 4.

	DTM20_Ops		MSIS00		JB2008		DTM20_Int		DTM20_Res	
GOCE	12.7	0.88	15.1	0.87	17.9	0.88	11.8	0.89	11.8	0.90
	8.1	0.90	12.1	0.89	9.2	0.88	9.0	0.90	9.0	0.90
	14.6	0.87	16.9	0.86	21.1	0.87	13.7	0.88	13.8	0.90
	10.6	0.95	12.7	0.94	16.8	0.92	9.4	0.94	10.0	0.94
CHAMP	22.3	0.88	23.7	0.87	24.2	0.88	21.7	0.89	20.5	0.90
	16.1	0.90	17.4	0.89	19.1	0.88	16.1	0.90	15.7	0.90
	24.9	0.87	26.7	0.86	27.0	0.87	24.3	0.88	22.5	0.90
	20.6	0.95	20.2	0.94	21.0	0.92	19.6	0.94	19.8	0.94
GRACE	28.1	0.87	29.9	0.85	31.6	0.85	26.9	0.87	25.5	0.89
	20.1	0.88	22.2	0.87	25.9	0.83	20.2	0.88	19.4	0.89
	31.4	0.85	33.8	0.83	35.2	0.84	30.2	0.86	28.3	0.88
	26.8	0.91	25.1	0.89	26.7	0.85	23.3	0.91	23.5	0.91
SwarmA	25.1	0.91	30.5	0.88	30.1	0.90	21.4	0.93	20.1	0.94
	21.6	0.94	23.6	0.92	30.0	0.89	19.8	0.95	18.9	0.95
	27.1	0.90	35.0	0.86	31.1	0.89	22.1	0.92	20.5	0.93
	24.0	0.96	25.6	0.94	26.6	0.91	20.0	0.96	19.3	0.96
GFO	34.2	0.83	44.0	0.79	35.7	0.82	30.1	0.85	27.6	0.88
	32.1	0.83	35.0	0.81	40.8	0.78	31.8	0.85	29.8	0.87
	38.2	0.85	53.4	0.80	35.4	0.83	31.0	0.86	27.4	0.90
	27.6	0.95	29.1	0.94	31.9	0.90	27.8	0.95	27.5	0.95

Table 8. The <O/C> and StD <O/C> per model for the five density datasets and all 152 storms (single and multiple-peak storms). Per model are given the results of all four phases.

	DTM20_Ops		MSIS00		JB2008		DTM20_Int		DTM20_Res	
GOCE	1.01	3.2	1.04	3.7	1.02	5.0	1.01	2.3	0.99	2.1
CHAMP	1.01	9.2	1.02	9.8	0.99	9.2	0.99	8.1	0.99	7.6
GRACE	1.02	11.0	1.04	11.3	1.00	11.0	0.99	9.1	0.99	8.8
SwarmA	1.05	8.9	1.08	10.3	1.03	12.2	1.02	8.2	1.02	8.1
GFO	1.01	7.3	1.03	9.7	1.04	11.7	0.99	5.6	0.99	6.1

Table 9. The average StD of the O/C and correlation coefficient per model for the five density datasets and all 152 storms (single and multiple-peak storms). Per model, the results of all four phases.

	DTM20_Ops		MSIS00		JB2008		DTM20_Int		DTM20_Res	
GOCE	12.3	0.92	14.5	0.91	14.0	0.92	11.3	0.93	11.0	0.94
CHAMP	20.1	0.90	22.4	0.89	22.3	0.88	19.3	0.90	18.3	0.91
GRACE	24.4	0.89	27.9	0.87	28.2	0.86	23.2	0.89	22.0	0.91
SwarmA	22.5	0.92	27.3	0.89	29.3	0.89	20.9	0.93	20.2	0.93
GFO	25.5	0.90	29.1	0.89	34.4	0.88	24.7	0.90	23.6	0.92

Table 8 lists the $\langle O/C \rangle$ and StD $\langle O/C \rangle$ of the five models for the five density datasets and all 152 storms (single and multiple-peak). In orbit prediction, the type of storm is not known beforehand, and these results are therefore most comprehensive. Since phases 2 and 3 do not represent exactly the same storm phases in case of SP and MP storms, only the assessments over the complete storms, i.e., for all four phases, is listed. The results are of course more or less (because there are more SP than MP) in between the SP and MP results. DTM2020_Res and DTM2020_Int have the smallest biases $\langle O/C \rangle$ of 2% maximum, and smallest variations StD $\langle O/C \rangle$. JB2008 has only slightly higher biases, a maximum of 4%, but their StD are up to twice as large. MSIS00 has the highest biases, but these are still only 2–8%.

Table 9 lists the average StD of O/C and the average correlation coefficients of the five models for the five density datasets and all storms (single and multiple-peak). DTM2020_Res always has the smallest StD of O/C and highest correlations, i.e., the storms are reproduced with highest fidelity of the five models assessed in this study.

4 Conclusions and outlook

The density data and indices for the 152 selected storms as well as updated metrics for thermosphere model assessment have been described in this paper. The mean and standard deviation of the density ratios and the correlations were computed using CHAMP, GRACE GOCE, Swarm-A and GFO density data and the CIRA models MSIS00, JB2008, and DTM2020_Ops/Int/Res.

Overall, taking all SP and MP storms into account, the DTM and JB2008 models are only slightly underestimating during storms, 2–4% maximum, which is an unexpectedly good performance for semi-empirical models. MSIS00 biases are somewhat higher and range from 2 to 8%. It is remarkable, and reassuring, that the MSIS, JB2008 and DTM models predict storm perturbations with comparably small biases, while they were fitted to different observations (very different for MSIS). That performance is reached on condition of debiasing the models before the storm onset, which option is not available to all thermosphere model users, nor is it always possible. However, excepting GOCE results, the StD $\langle O/C \rangle$ of all models ranges from 6 to 12%, i.e. the scatter is considerable and improvements are undeniably necessary.

The best results, i.e. smallest bias and standard deviation on average over all SP storms, over the entire 4-Phase storm period are obtained with DTM2020_Int and DTM2020_Res models, while the oldest model, MSIS00 (which unlike DTM and

JB2008 has not benefitted from an overhaul of its geomagnetic model algorithm), is the least precise. However, MSIS00 is the least biased model for MP storms. The assessment clearly revealed that model precision (StD of O/C) decreases with altitude, but that bias $\langle O/C \rangle$ is independent of altitude, at least in the range of 250–550 km.

The updated metric and the full list of SP and MP storms (Table A1) will eventually be implemented in CAMEL (Comprehensive Assessment of Models and Events based on Library tools) at NASA's Community Coordinated Modeling Center (CCMC). Then, assessment of not only the semi-empirical models tested in this study but also of many first principles models such as TIEGCM, WACCMX, GITM and WAM will be possible. This study showed that the average model bias can hardly be improved upon, but maybe the physical models succeed in driving down the standard deviations (StD $\langle O/C \rangle$). Presently, an incomplete test version is already available with fewer storms and not all model output (<https://webserver1.ccmc.nasa.gov/camel/NeutralDensity>). The final CAMEL version will also generate score cards per model, which will be based on selected reference storms and density data, so that model users will have access to objective and standardized model assessments.

Acknowledgments

We thank the reviewers for their time and help to improve the paper. We thank TU Delft (<http://thermosphere.tudelft.nl>) for making their total density data publicly available. We thank WDC Kyoto for the Dst index (<http://wdc.kugi.kyoto-u.ac.jp/>), and GFZ Potsdam for the Kp index (<https://www.gfz-potsdam.de/en/kp-index/>) and Hpo index (<https://www.gfz-potsdam.de/en/hpo-index/>). We received support from CNES/SHM. The editor thanks Christian Siemes and an anonymous reviewer for their assistance in evaluating this paper.

References

- Bowman B, Tobiska WK, Marcos F, Huang C, Lin C, Burke W. 2008. A new empirical thermospheric density model JB2008 using new solar and geomagnetic indices. In: *AIAA/AAS astrodynamics specialist conference*, AIAA, Honolulu, HI. <https://doi.org/10.2514/6.2008-6438>.
- Bruinsma S, Siemes C, Emmert JT, Mlynczak MG. 2022. Description and comparison of 21st century thermosphere data. *Adv Space Res* **72**(12): 5476–5489. <https://doi.org/10.1016/j.asr.2022.09.038>.
- Bruinsma S, Boniface C, Sutton E, Fedrizzi M. 2021. Thermosphere modeling capabilities assessment: geomagnetic storms. *J Space Weather Space Clim* **11**: 12. <https://doi.org/10.1051/swsc/2021002>.
- Bruinsma S, Boniface C. 2021. The DTM2020 thermosphere models. *J Space Weather Space Clim* **11**: 47. <https://doi.org/10.1051/swsc/2021032>.

- Bruinsma SL. 2015. The DTM-2013 thermosphere model. *J Space Weather Space Clim* **5**: A1. <https://doi.org/10.1051/swsc/2012005>.
- Bruinsma SL, Doornbos E, Bowman BR. 2014. Validation of GOCE densities and thermosphere model evaluation. *Adv Space Res* **54**: 576–585. <https://doi.org/10.1016/j.asr.2014.04.008>.
- Doornbos E. 2011. Thermospheric density and wind determination from satellite dynamics, *Ph.D. Dissertation*, University of Delft, 188 p. Available at <http://repository.tudelft.nl/>.
- Emmert JT, Drob DP, Picone JM, Siskind DE, Jones Jr. M et al. 2020. NRLMSIS 2.0: a whole-atmosphere empirical model of temperature and neutral species densities. *Earth Space Sci* **8**(3): e2020EA001321. <https://doi.org/10.1029/2020EA001321>.
- Mehta PM, Walker AC, Sutton EK, Godinez HC. 2017. New density estimates derived using accelerometers on board the CHAMP and GRACE satellites. *Space Weather* **15**: 558–576. <https://doi.org/10.1002/2016SW001562>.
- Picone JM, Hedin AE, Drob DP, Aikin AC. 2002. NRLMSISE-00 empirical model of the atmosphere: statistical comparisons and scientific issues. *J Geophys Res* **107**(A12): 1468. <https://doi.org/10.1029/2002JA009430>.
- Siemes C, Borries C, Bruinsma S, Fernandez-Gomez I, Hładczuk N, van den Ijssel J, Kodikara T, Vielberg K, Visser P. 2023. New thermosphere neutral density and crosswind datasets from CHAMP, GRACE and GRACE-FO. *J Space Weather Space Clim*. <https://doi.org/10.1051/swsc/2023014>.
- Storz MF, Bowman BR, Branson MJ, Casali SJ, Tobiska WK. 2005. High accuracy satellite drag model (HASDM). *Adv Space Res* **36**: 2497–2505. <https://doi.org/10.1016/j.asr.2004.02.020>.
- Sutton EK. 2018. A new method of physics-based data assimilation for the quiet and disturbed thermosphere. *Space Weather* **16**: 736–753. <https://doi.org/10.1002/2017SW00178>.
- Van den Ijssel J, Doornbos E, Iorfida E, March M, Siemes C, Montenbruck O. 2020. Thermosphere densities derived from Swarm GPS observations. *Adv Space Res* **65**: 7. <https://doi.org/10.1016/j.asr.2020.01.004>.

Appendix

Table A1. The 152 selected storm intervals from 2001 to 2023, storm class (cat: Single Peak or Multiple Peak), minimum Dst, maximum ap, and the satellite dataset (CH, GR, GO, SW, GFO = CHAMP, GRACE, GOCE, Swarm-A and GRACE-FO).

Start	End	Class	Min Dst	Max ap	Datasets	
2001-09-22	2001-09-27	MP	−102	154	CH	
2001-09-29	2001-10-05	MP	−166	132	CH	
2001-10-10	2001-10-14	SP	−71	80	CH	
2001-10-20	2001-10-24	MP	−187	179	CH	
2001-11-04	2001-11-08	MP	−292	300	CH	
2002-02-27	2002-03-02	SP	−71	80	CH	
2002-03-22	2002-03-26	SP	−100	80	CH	
2002-04-16	2002-04-22	MP	−149	154	CH	GR
2002-04-22	2002-04-25	SP	−57	80	CH	GR
2002-05-10	2002-05-13	SP	−110	111	CH	GR
2002-05-22	2002-05-25	SP	−109	236	CH	
2002-07-31	2002-08-04	SP	−102	80	CH	GR
2002-08-19	2002-08-22	SP	−106	94	CH	GR
2002-09-02	2002-09-06	SP	−109	94	CH	GR
2002-09-06	2002-09-09	SP	−181	154	CH	GR
2002-09-30	2002-10-06	MP	−176	154	CH	GR
2002-10-06	2002-10-10	SP	−115	80	CH	GR
2002-10-23	2002-10-27	MP	−98	94	CH	GR
2002-11-19	2002-11-23	MP	−128	111	CH	GR
2002-12-25	2002-12-29	SP	−68	80	CH	GR
2003-03-15	2003-03-18	SP	−60	80	CH	GR
2003-03-30	2003-04-02	SP	−78	94	CH	GR
2003-04-28	2003-05-03	MP	−78	111	CH	GR
2003-05-08	2003-05-12	SP	−84	111	GR	
2003-05-20	2003-05-23	SP	−73	80	CH	GR
2003-05-28	2003-06-01	MP	−144	236	CH	GR
2003-06-08	2003-06-11	SP	−50	80	CH	GR
2003-06-15	2003-06-20	MP	−141	111	CH	GR
2003-06-27	2003-06-30	SP	−34	94	CH	GR
2003-07-10	2003-07-14	MP	−105	111	CH	GR
2003-07-15	2003-07-18	SP	−90	80	CH	GR
2003-07-25	2003-07-28	SP	−57	80	CH	GR
2003-08-06	2003-08-09	SP	−61	111	CH	GR
2003-08-16	2003-08-20	SP	−148	154	CH	GR
2003-08-20	2003-08-23	MP	−68	80	CH	GR
2003-09-16	2003-09-20	MP	−65	132	CH	GR

(Continued on next page)

Table A1. (Continued)

Start	End	Class	Min Dst	Max ap		Datasets	
2003-10-13	2003-10-17	MP	-85	154	CH	GR	
2003-10-20	2003-10-23	SP	-61	80	CH	GR	
2003-10-23	2003-10-26	SP	-49	111	CH	GR	
2003-10-28	2003-11-02	MP	-383	400	CH	GR	
2003-11-03	2003-11-06	SP	-69	132	CH	GR	
2003-11-10	2003-11-13	SP	-62	80	CH	GR	
2003-11-12	2003-11-17	MP	-59	80	CH	GR	
2003-11-19	2003-11-24	MP	-422	300	CH	GR	
2004-01-05	2004-01-08	SP	-50	80	CH	GR	
2004-01-21	2004-01-25	MP	-130	132	CH	GR	
2004-02-10	2004-02-13	SP	-93	94	CH	GR	
2004-03-08	2004-03-13	MP	-78	94	CH	GR	
2004-04-02	2004-04-07	MP	-117	94	CH	GR	
2004-07-15	2004-07-19	SP	-76	80	CH	GR	
2004-07-21	2004-07-29	MP	-170	300	CH	GR	
2004-08-29	2004-09-01	SP	-129	132	CH	GR	
2004-11-06	2004-11-12	MP	-374	300	CH	GR	
2005-01-06	2005-01-09	SP	-93	179	CH	GR	
2005-01-16	2005-01-21	MP	-103	179	CH	GR	
2005-01-20	2005-01-23	SP	-97	207	CH	GR	
2005-02-16	2005-02-20	SP	-80	94	CH	GR	
2005-04-03	2005-04-06	SP	-70	132	CH	GR	
2005-05-06	2005-05-10	MP	-110	236	CH	GR	
2005-05-14	2005-05-18	MP	-247	236	CH	GR	
2005-05-29	2005-06-01	SP	-113	179	CH	GR	
2005-06-11	2005-06-14	SP	-106	154	CH	GR	
2005-06-22	2005-06-25	SP	-85	132	CH	GR	
2005-07-09	2005-07-14	MP	-92	94	CH	GR	
2005-08-23	2005-08-26	SP	-184	300	CH	GR	
2005-08-30	2005-09-04	MP	-122	132	CH	GR	
2005-09-09	2005-09-15	MP	-139	179	CH	GR	
2005-09-14	2005-09-17	SP	-80	132	CH	GR	
2006-03-17	2006-03-21	SP	-48	94	CH	GR	
2006-04-08	2006-04-11	MP	-82	80	CH	GR	
2006-04-12	2006-04-16	SP	-98	132	CH	GR	
2006-07-26	2006-07-30	SP	-48	80	CH	GR	
2006-08-18	2006-08-21	SP	-79	80	CH	GR	
2006-11-09	2006-11-12	SP	-63	80	CH	GR	
2006-12-13	2006-12-16	SP	-162	236	CH	GR	
2008-09-02	2008-09-06	SP	-51	80	CH	GR	
2008-10-10	2008-10-13	SP	-54	94	CH	GR	
2010-04-04	2010-04-07	SP	-81	179	GO	CH	GR
2010-05-01	2010-05-04	SP	-71	80	GO	CH	GR
2010-08-02	2010-08-06	MP	-74	111	CH	GR	
2011-05-27	2011-05-30	SP	-80	94	GO	GR	
2011-06-03	2011-06-07	SP	-45	94	GO	GR	
2011-08-04	2011-08-07	SP	-115	179	GO	GR	
2011-09-25	2011-09-28	SP	-118	94	GR		
2011-10-23	2011-10-26	SP	-147	154	GO	GR	
2012-03-06	2012-03-11	MP	-145	207	GR		
2012-03-11	2012-03-14	SP	-64	94	GO	GR	
2012-03-14	2012-03-17	SP	-88	94	GO	GR	
2012-04-22	2012-04-26	SP	-120	111	GO		
2012-06-15	2012-06-19	MP	-86	94	GO	GR	
2012-07-08	2012-07-11	SP	-78	111	GO	GR	
2012-07-14	2012-07-17	SP	-139	132	GO	GR	
2012-09-29	2012-10-03	SP	-122	111	GO		
2012-10-07	2012-10-11	MP	-109	111	GO		
2012-11-12	2012-11-16	SP	-108	94	GO	GR	

(Continued on next page)

Table A1. (Continued)

Start	End	Class	Min Dst	Max ap		Datasets
2013-03-16	2013-03-19	SP	-132	111	GO	
2013-05-30	2013-06-03	SP	-124	132	GO	GR
2013-06-27	2013-07-01	SP	-102	94	GO	GR
2013-09-30	2013-10-04	SP	-72	179	GR	
2013-12-06	2013-12-10	SP	-66	80	SW	GR
2014-02-17	2014-02-22	MP	-119	94	GR	
2014-06-07	2014-06-10	SP	-37	94	SW	GR
2014-09-11	2014-09-14	SP	-88	94	SW	GR
2015-01-06	2015-01-09	SP	-107	94	SW	
2015-03-16	2015-03-19	SP	-234	179	SW	GR
2015-03-21	2015-03-24	SP	-65	94	SW	GR
2015-04-15	2015-04-18	SP	-88	80	SW	GR
2015-06-07	2015-06-10	SP	-67	80	SW	
2015-06-21	2015-06-24	SP	-198	236	SW	
2015-08-14	2015-08-17	SP	-98	94	SW	GR
2015-08-25	2015-08-29	MP	-103	94	SW	GR
2015-09-06	2015-09-13	MP	-105	132	SW	GR
2015-09-19	2015-09-22	SP	-81	132	SW	GR
2015-10-05	2015-10-09	MP	-130	154	SW	
2015-11-05	2015-11-09	SP	-87	80	SW	
2015-12-18	2015-12-23	MP	-166	111	SW	GR
2016-03-05	2016-03-08	SP	-99	94	SW	GR
2016-05-06	2016-05-11	MP	-95	94	SW	GR
2016-06-13	2016-06-16	SP	-17	94	SW	GR
2016-08-31	2016-09-04	SP	-59	80	SW	
2016-10-12	2016-10-15	SP	-110	94	SW	GR
2016-10-24	2016-10-27	SP	-65	94	SW	
2016-12-20	2016-12-23	SP	-39	80	SW	GR
2017-03-26	2017-03-29	SP	-70	94	SW	GR
2017-05-26	2017-05-30	SP	-125	132	SW	GR
2017-07-15	2017-07-18	SP	-72	80	SW	
2017-09-06	2017-09-10	MP	-122	236	SW	GR
2017-09-26	2017-09-29	SP	-56	111	SW	
2017-11-06	2017-11-10	MP	-73	94	SW	
2018-03-17	2018-03-20	SP	-50	80	SW	
2018-04-19	2018-04-22	SP	-66	80	SW	
2018-08-24	2018-08-28	MP	-175	154	GFO	SW
2018-09-10	2018-09-13	SP	-60	80	GFO	SW
2019-05-13	2019-05-16	SP	-65	94	GFO	SW
2021-05-11	2021-05-14	SP	-61	132	GFO	SW
2021-10-10	2021-10-14	SP	-65	94	GFO	SW
2021-11-02	2021-11-05	SP	-105	179	GFO	SW
2022-03-12	2022-03-15	SP	-85	94	GFO	SW
2022-04-08	2022-04-12	SP	-48	111	GFO	SW
2022-04-13	2022-04-16	SP	-81	80	GFO	SW
2022-08-16	2022-08-19	SP	-50	111	GFO	SW
2022-09-03	2022-09-06	SP	-72	94	GFO	SW
2023-02-25	2023-03-01	SP	-132	111	GFO	SW
2023-03-22	2023-03-25	SP	-163	207	GFO	SW
2023-04-22	2023-04-25	SP	-212	236	GFO	SW
2023-05-04	2023-05-08	SP	-67	80	GFO	SW
2023-08-03	2023-08-07	SP	-93	94	GFO	SW
2023-09-17	2023-09-21	MP	-85	94	GFO	SW
2023-11-04	2023-11-07	SP	-172	154	SW	
2023-11-24	2023-11-27	SP	-99	94	SW	
2023-11-30	2023-12-03	SP	-108	111	SW	

Cite this article as: Bruinsma S & Laurens S. 2024. Thermosphere model assessment for geomagnetic storms from 2001 to 2023. *J. Space Weather Space Clim.* **14**, 28. <https://doi.org/10.1051/swsc/2024027>.