Space Weather

RESEARCH ARTICLE
10.1029/2025SW004782

Comprehensive and Open Assessment of Thermospheric Models During Geomagnetic Storm Times Within CCMC Framework

Jack C. Wang[1,2], Jia Yue[1,2], Sean Bruinsma[3], Masha Kuznetsova[2], Joseph Sypal[1], Richard Mullinix[2], Chiu Wiegand[2], Paul Dimarzio[1], Min-Yang Chou[1,2], Maksym Petrenko[2], Christian Siemes[4], and Sophie Laurens[3]

[1]Catholic University of America, Washington, DC, USA, [2]NASA Goddard Space Flight Center, Greenbelt, MD, USA, [3]CNES, Space Geodesy Office, Toulouse, France, [4]Delft University of Technology, Delft, The Netherlands

Key Points:

- Thermospheric models hosted at the Community Coordinated Modeling Center are assessed against neutral density observations across 151 geomagnetic storms from 2001 to 2023
- DTM2020 shows the best overall skill. Among physics-based models, WACCMX-Heelis performs most consistently
- DTM2020 and JB2008 are recommended as empirical models for satellite drag computations during geomagnetic storm times, replacing MSIS

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. C. Wang,
jack.c.wang@nasa.gov

Citation:

Wang, J. C., Yue, J., Bruinsma, S., Kuznetsova, M., Sypal, J., Mullinix, R., et al. (2026). Comprehensive and open assessment of thermospheric models during geomagnetic storm times within CCMC framework. *Space Weather*, *24*, e2025SW004782. https://doi.org/10.1029/2025SW004782

Received 20 OCT 2025
Accepted 8 APR 2026

Author Contributions:

Conceptualization: Jack C. Wang, Jia Yue, Sean Bruinsma, Masha Kuznetsova, Richard Mullinix, Min-Yang Chou
Data curation: Jack C. Wang, Sean Bruinsma, Min-Yang Chou, Maksym Petrenko, Christian Siemes, Sophie Laurens
Formal analysis: Jack C. Wang
Funding acquisition: Jia Yue, Masha Kuznetsova
Investigation: Jack C. Wang, Sean Bruinsma

Abstract Thermospheric neutral density controls satellite drag in low Earth orbit and varies strongly during geomagnetic storms. We present a multi-mission, phase-resolved assessment of empirical and physics-based thermosphere models using 151 storms from 2001 to 2023, including MSIS, DTM, JB2008, WAM-IPE, WACCM-X, GITM, CTIPe, and TIE-GCM. High-resolution densities from satellites are compared to model outputs using a pre-storm debiasing procedure. Skill is quantified using the debiased mean observed-to-modeled ratio ($\mu(o/m)$), standard deviation ($\sigma(o/m)$), and the Pearson correlation (R), evaluated separately for onset, main and recovery, and post-storm phases. Across all phases, DTM2020 shows the best performance (ratios near unity, lowest $\sigma(o/m)$, highest R), followed by JB2008 and DTM2013. MSIS models systematically underestimate density during the main and recovery and post–storm phases by ∼20%–30% and 10%, respectively. Based on these findings, we recommend DTM2020 and JB2008 as empirical models for satellite drag computations during the geomagnetically active periods, replacing MSIS as the standard reference. Among physics-based models, WACCMX-Heelis performs most reliably (ratios closest to unity, smallest $\sigma(o/m)$). GITM shows the best mean ratio, but its standard deviation of $\mu(o/m)$ and $\sigma(o/m)$ are nearly twice that of other models, suggesting that its density is wider spread than others. Sensitivity to electrodynamic forcing is evident: Heelis-driven runs of WACCM-X and TIE-GCM generally outperform Weimer-driven counterparts. Latitude-local time maps reveal persistent high-latitude underestimation of neutral density and diurnal structure in several physics-based models. Results are synthesized into model scorecards and delivered via the CCMC ITMAP platform to support open and reproducible, storm-time validation and future operational benchmarking.

Plain Language Summary The upper atmosphere heats and expands during geomagnetically active periods, increasing atmospheric drag on satellites. This causes their orbits to decay faster and complicates collision avoidance and mission planning. Understanding and accurately modeling neutral density is critically important because it directly impacts satellite safety, lifetime, and operational costs. In this study, we evaluated the performance of the thermospheric models hosted at the CCMC using satellite measurements from five missions, covering 151 geomagnetic storms over more than 20 years. Model performance are evaluated using the debiased mean observed-to-modeled ratio, standard deviation, and the Pearson correlation. DTM2020 shows the best overall performance, followed by JB2008 and DTM2013. Based on these findings, we recommend using DTM2020 and JB2008, instead of MSIS, as the benchmark empirical models for satellite drag computations during the geomagnetically active periods. Among physics-based models, WACCM-X with the Heelis driver performed most consistently. However, many physics-based models still underestimate densities at high latitudes and exhibit day–night variations that differ from observations. These results are available through NASA's CCMC ITMAP platform, which highlight where current models succeed and where improvements are needed. This web-based system will help guide the development of better storm-time drag forecasts that protect satellites and space infrastructure.

1. Introduction

The thermosphere, extending from roughly 90–600 km altitude, is composed mainly of atomic oxygen (O), molecular nitrogen (N_2), and molecular oxygen (O_2) at lower altitudes, with lighter species such as helium (He) and hydrogen (H) becoming dominant above ∼500 km. The mass density of these neutral constituents (hereafter: neutral density) directly governs aerodynamic drag on satellites and debris in low Earth orbit and is therefore a

Methodology: Jack C. Wang, Jia Yue, Sean Bruinsma, Sophie Laurens
Project administration: Masha Kuznetsova, Chiu Wiegand
Resources: Jia Yue
Software: Jack C. Wang, Joseph Sypal, Richard Mullinix, Paul Dimarzio, Maksym Petrenko
Supervision: Jia Yue, Sean Bruinsma, Masha Kuznetsova
Validation: Jack C. Wang, Jia Yue
Visualization: Jack C. Wang, Joseph Sypal, Richard Mullinix, Paul Dimarzio
Writing – original draft: Jack C. Wang
Writing – review & editing: Jia Yue, Sean Bruinsma, Christian Siemes

critical parameter for orbit prediction, collision risk assessment, and space situational awareness (Doornbos, 2012; Vallado & Finkleman, 2014).

Thermospheric neutral density exhibits variability on timescales ranging from minutes to decades. On short timescales, geomagnetic storms can cause rapid and substantial density enhancements, particularly at high latitudes, through processes such as Joule heating, particle precipitation, and associated auroral dynamics (Fuller-Rowell et al., 1994; Lu et al., 2016). These storm-time energy inputs heat and expand the upper atmosphere, increasing density at a given altitude and leading to elevated satellite drag and accelerated orbital decay (e.g., Fang et al., 2022; Oliveira & Zesta, 2019; Sutton et al., 2005). On longer timescales, variability is modulated by solar extreme ultraviolet (EUV) flux (Lean et al., 2011), seasonal cycles, and quasi-decadal solar activity, as well as coupling from the lower atmosphere through tides, planetary waves, and gravity waves (e.g., Oberheide et al., 2009; H. Liu, 2016).

Accurate specification and forecasting of thermospheric neutral density during geomagnetic storms remain challenging. Empirical models such as NRLMSIS (Emmert et al., 2021; Picone et al., 2002) and JB2008 (Bowman et al., 2008) provide climatological estimates but often underestimate the magnitude of storm-time density enhancements (Bruinsma & Laurens, 2024; Sutton, 2018). Physics-based general circulation models, including the Thermosphere-Ionosphere-Electrodynamics GCM (TIE-GCM; Qian et al., 2014) and the Whole Atmosphere Community Climate Model with thermosphere-ionosphere extension (WACCM-X; H.-L. Liu et al., 2018) explicitly simulate the response to high-latitude electrodynamic forcing and thermosphere-ionosphere coupling. However, their performance remains sensitive to uncertainties in storm-time energy inputs (Knipp et al., 2013; Weimer, 2005), model parameterizations (Qian et al., 2014), and the representation of upward-propagating wave forcing from the lower atmosphere (Jones et al., 2014; Yamazaki & Richmond, 2013). In particular, biases vary with geomagnetic latitude and local time (LT), and model skill tends to decrease during periods of strong and rapidly changing drivers. These discrepancies limit the reliability of both scientific investigations of upper atmospheric dynamics during storms and operational space weather forecasts.

Satellite accelerometer missions such as CHAMP (Reigber et al., 2002), Gravity Recovery and Climate Experiment (GRACE) (Tapley et al., 2004), GOCE (Floberghagen et al., 2011), Swarm-A and Swarm-B (Olsen et al., 2013), and GRACE-FO (Kornfeld et al., 2019) have provided unprecedented high-resolution measurements of thermospheric neutral density, enabling rigorous model validation during both quiet and disturbed conditions. Yet storm-time comparisons between models and observations were often limited to a relatively small number of events, focusing on a few well-observed intense storms, thereby restricting the scope for evaluating model performance across a broader range of geomagnetic activity levels. Bruinsma and Laurens (2024) conducted the most comprehensive assessment to date, analyzing 152 storms from 2001 to 2023 using updated storm-time metrics to benchmark five semi-empirical thermosphere models. Their study quantified biases, precision, and correlation over storm phases, providing an objective baseline for future model intercomparisons and paving the way for standardized model "scorecards".

Building on the work of Bruinsma and Laurens (2024), this study extends the analysis to include not only semi-empirical models but also physics-based models hosted at NASA's Community Coordinated Modeling Center (CCMC). The results are synthesized into comprehensive model scorecards and disseminated through the CCMC Ionosphere–Thermosphere Model Assessment and Validation (ITMAP) platform. The CCMC-hosted thermospheric models are evaluated unbiasedly against observations during a set of well-characterized geomagnetic storms. We quantify model biases using the debiased mean observed-to-computed density ratio, along with the associated standard deviation and correlation coefficient, to assess model performance during 151 geomagnetic storms from 2001 to 2023. The analysis is based on multi-satellite observations, enabling evaluation of both the magnitude and variability of modeled storm-time density responses across different storm phases. By identifying systematic model-data differences and their likely sources, we provide new insight into the processes controlling thermospheric density during disturbed conditions and offer guidance for improving the fidelity of both empirical and physics-based models for space weather applications.

2. Observations

This study makes use of high-resolution accelerometer-inferred density data sets in addition to GPS-derived densities. These data sets are well suited for storm-time assessment, providing near pole-to-pole coverage with consistent accuracy. Figure 1 illustrates the mission lifetimes and the distribution of sample altitudes used in this

10.1029/2025SW004782

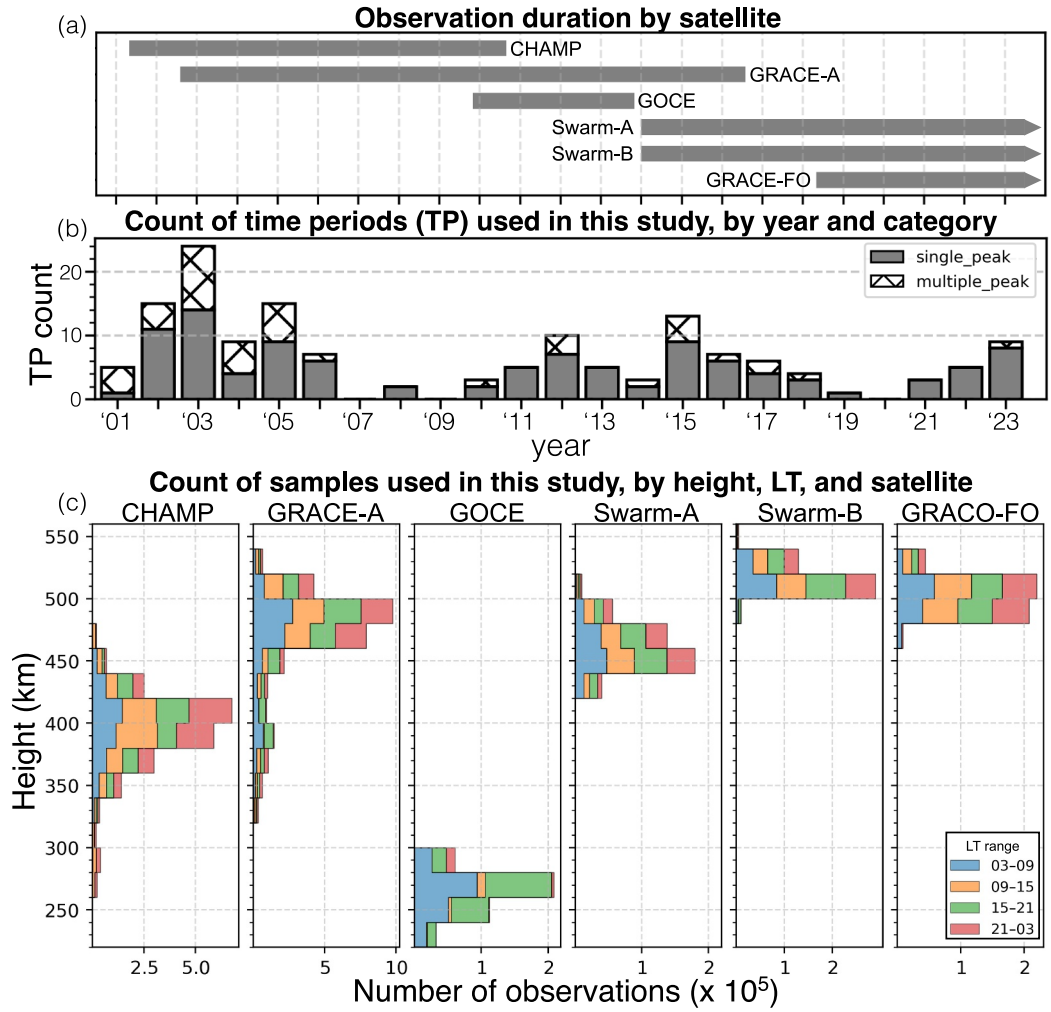


Figure 1. (a) Data set availability over time. (b) Annual count of time periods (TPs) used in this study. Filled and hatched bars represent TPs with single-peak and multiple-peak ap index profiles, respectively. (c) Number of samples as a function of height and local time (LT) for each satellite. Columns from left to right correspond to CHAMP, GRACE-A, GOCE, Swarm-A, Swarm-B, and GRACE-FO 1 satellite, respectively. Blue bars represent LT 03–09, orange bars represent LT 09–15, green bars represent LT 15–21, and red bars represent LT 21-03.

study, along with the specific assessment time periods (TPs), counted by year and categorized as either single-peak or multiple-peak events based on the ap index (equivalent planetary amplitude; Matzka et al., 2021) time series within each given TP. Note that, in this study, the 3-hr ap index is used to identify and define storm phases, while the daily Ap index is used when driving the NRLMSIS 2.0 model. Here, the lowercase ap denotes the three-hour planetary geomagnetic index derived from Kp, whereas the capitalized Ap represents the daily mean of the eight ap values following the standard geomagnetic index definitions (Matzka et al., 2021). Both ap and Ap indices are assigned a unit of 2 nT. In total, there are 151 events analyzed in the study with 107 single-peak and 44 multiple-peak events from 2001 to 2023. The historical TPs are selected based on the availability of the data sets and the requirement that the ap index reaches the storm-time threshold of 80. The observational altitude range spans from about 220 to 545 km, providing a total of approximately 2.4 million samples for the assessment.

Together, these data sets span a broad range of altitudes and solar cycles, enabling multi-mission assessments across 151 geomagnetic storms. This combination allows us to evaluate model performance consistently from the lower thermosphere into the upper thermosphere under disturbed conditions. The following sections outline the data sets retrieved from the respective instruments onboard each satellite in the study.

2.1. CHAMP (CHAllenging Minisatellite Payload)

CHAMP operated from 2000 to 2010 in a near-polar (inclination of 87.2°), low-Earth orbit with altitudes decaying from ~450 km to below 300 km over its lifetime (Reigber et al., 2002). Most of the density data were collected between 360 and 440 km. Equipped with a high-precision STAR accelerometer (Spatial Triaxial Accelerometer for Research), CHAMP provided one of the first long-term, pole-to-pole thermospheric density data sets. The sampling rate of the desnity data is 10 s, which corresponds to a spatial resolution of ~80 km along the track (Doornbos, 2012; Siemes et al., 2023).

2.2. GRACE (Gravity Recovery and Climate Experiment) and GRACE-FO (GRACE Follow-On)

GRACE, a twin-satellite mission launched in 2002, operated until 2017 at altitudes of ~480–510 km with an orbital inclination of 89° (Tapley et al., 2004). Each spacecraft carried a highly sensitive SuperSTAR electrostatic accelerometer, enabling the derivation of thermospheric neutral densities along their orbital tracks (Kenyon et al., 2012). In this study, only densities from GRACE-A are used because the two satellites flew in close formation (Siemes et al., 2023).

GRACE-FO, launched in 2018, continues the GRACE legacy with improved instrumentation and a similar near-polar orbit (Kornfeld et al., 2019). The twin-satellite mission extends the thermospheric density record seamlessly into the present, with measurement capabilities and orbital characteristics comparable to GRACE. The onboard accelerometer was slightly updated and is referred to as SuperSTAR-FO (Christophe et al., 2015). We use neutral density from GRACE-FO 1 satellite in the study. The density data from GRACE and GRACE-FO have a cadence of 10 s, corresponding to an along-track resolution of ~85 km (Siemes et al., 2023).

2.3. GOCE (Gravity Field and Steady-State Ocean Circulation Explorer)

GOCE was launched in 2009, flew at an exceptionally low altitude (~270 km) in a nearly sun-synchronous dawn-dusk orbit until 2013 (Floberghagen et al., 2011). Its drag-free control system and accelerometer provided density data with high precision in the lower part of the thermosphere, valuable for storm-time analysis despite its limited LT coverage. Densities are inferred from the combination of ion propulsion, the main contributor, and accelerometer data. The sampling rate of the density data is 10 s, which corresponds to a spatial resolution of ~70 km along the track (Bruinsma et al., 2014).

2.4. Swarm

The Swarm constellation mission launched in November 2013, flies in a near-polar, circular orbits with initial altitudes and inclinations of ~450 km and 87.4° for Swarm-A and Swarm-C, and ~530 km and 86.8° for Swarm-B (Friis-Christensen et al., 2008; Olsen et al., 2013). Each Swarm satellite carries an accelerometer (ACC) and a GPS receiver as part of its scientific payload. In principle, the accelerometers directly measure the non-gravitational accelerations acting on the spacecraft at high temporal resolution, enabling thermospheric neutral density retrieval following the approach successfully applied to CHAMP, GRACE, and GOCE (Bruinsma et al., 2004; Doornbos, 2012; Sutton et al., 2007).

However, the Swarm accelerometer measurements are affected by instrumental disturbances, including sudden bias changes (referred to as “steps” in Siemes et al. (2016)), higher-than-expected temperature sensitivity in the onboard accelerometers, thruster spikes, and Error Detection And Correction related discontinuities, which complicate neutral density retrieval using accelerometers (Siemes et al., 2016). Consequently, ACC-derived densities require substantial correction and calibration that impact the fidelity of the accelerometer-derived density (Siemes et al., 2016).

In this study, thermospheric mass densities derived from precise orbit determination (POD) based on onboard GPS observations from Swarm-A and Swarm-B are utilized (van den IJssel et al., 2020). The POD-derived density data set from Swarm-A and Swarm-B is provided at a cadence of 30 s. The uncertainty of the GPS-derived densities depends on solar activity level. During low solar activity conditions, relative density errors can reach ~60% for the higher-altitude Swarm-B satellite and ~19% for the lower-altitude Swarm-A, whereas during higher solar activity the relative uncertainties reduce to ~7% and ~4%, respectively (van den IJssel et al., 2020).

Table 1
Comparison of Physics-Based Models in This Study and Their Key Specifications Used in CCMC Runs-On-Request Service

	WAM-IPE	WACCM-X	CTIPe	GITM	TIE-GCM
Horizontal resolution (lon. × lat.)	4° × 2°	1.25° × 0.95°	18° × 2°	4° × 2°	2.5° × 2.5°
Vertical resolution	$\frac{1}{4}$H	$\frac{1}{4}$H	1H	$\frac{1}{3}$H	$\frac{1}{4}$H
Vertical domain	Surface–3 × 10^{-7} Pa (500 km)	Surface–4 × 10^{-10} hPa (600 km)	1–3 × 10^{-7} hPa (80–500 km)	100–900 km	5.5 × 10^{-4}–4.6 × 10^{-10} hPa (97–600 km)
Resolved major species	O, O_2, N_2	O, O_2, N_2, H	O, O_2, N_2	O, O_2, N_2, He, H	O, O_2, N_2, He
Typical time step	WAM: 1 min, IPE: 5 min	5 min	1 min	1 min	20 s
CPU usage	112 cores, 90 min/day	112 cores, 120 min/day	1 core, 10 min/day	100 cores, 160 min/day	128 cores, 15 min/day
Model version	v1.2.5	v2.2.0	v4.1	v23.01.01	v2.0

Note. H denotes a scale height.

3. Models

In the following sections, we introduce the semi–empirical and first–principles physics–based models evaluated in this study, highlighting their key configurations, drivers, and boundary conditions. The semi–empirical models (e.g., DTM2013, DTM2020, JB2008, MSIS) combine large volumes of observational data with parameterized physical formulations to represent global density and temperature variability efficiently. In contrast, the physics–based general circulation models (e.g., TIE-GCM, WACCM-X, GITM, CTIPe, WAM-IPE) explicitly solve the coupled continuity, momentum, and energy equations for the thermosphere and ionosphere, allowing them to capture nonlinear feedback and transient storm–time dynamics. Table 1 provides a concise overview of the key specifications of the physics–based models hosted at the CCMC and used in this assessment.

3.1. NRLMSIS

The MSIS series of empirical models describes composition, mass density, and neutral temperature in Earth's atmosphere through an analytical algorithm based on parameterized equations. The MSIS series has evolved from early versions (MSISE-83, -86, -90) based primarily on in situ mass spectrometer measurements and incoherent scatter radar profiles. With NRLMSISE-00 (Picone et al., 2002), drag-derived densities from orbit determination and accelerometers, as well as molecular oxygen observations, were incorporated. The latest version used in this study is NRLMSIS 2.0 (Emmert et al., 2021), which assimilates atomic oxygen and hydrogen observations in the middle atmosphere, SABER and Odin/OSIRIS measurements between 50 and 100 km, orbit-derived global mean densities, and reanalysis temperature profiles below 50 km. Above 120 km, the model relaxes to orbit-derived mass densities. Note that a newer version, NRLMSIS 2.1 (Emmert et al., 2022), additionally includes nitric oxide (NO). For this study, NRLMSISE-00 and NRLMSIS 2.0 are driven by F10.7 and daily Ap index.

3.2. DTM (Drag Temperature Model)

DTM2013 is an updated semi-empirical thermosphere model developed by CNES (Bruinsma, 2015). It is based on DTM2009 but incorporates an improved formulation and an expanded data set. In addition to the density and temperature data used in earlier versions, DTM2013 assimilates high-resolution accelerometer-inferred densities from GOCE, and extending CHAMP and GRACE data set up to 2011. The model provides thermospheric density and temperature as functions of location, solar radiation (represented by the 30 cm solar radio flux, F30), and geomagnetic activity (Km index, but Kp can also be used to drive the DTM2013). Key advances in DTM2013 include better representation of storm-time density response and reduced biases in the low-to mid-latitude thermosphere.

DTM2020 is the latest release of the semi-empirical DTM series (Bruinsma & Boniface, 2021), which assimilates an expanded set of accelerometer-inferred density data (CHAMP, GRACE, and GOCE) as well as the POD-derived density from Swarm-A up to 2019. Compared to DTM2013, it offers more accurate storm-time responses, reduced latitudinal and seasonal biases, and improved consistency across solar cycles, providing reliable

global density estimates up to 1,000 km altitude. Two versions of the model are available: an operational version for orbit determination and routine space weather service and a research version used for model development and testing. In this study, only the operational version is employed. The operational model is driven by the F10.7 and Kp indices for solar and geomagnetic activity. The research version, although generally more accurate, uses the F30 and hourly Hp60 indices, which are not yet accredited for operational use.

3.3. JB2008 (Jacchia-Bowman 2008)

The JB2008 empirical thermospheric density model (Bowman et al., 2008) was developed as an operational model for satellite drag applications. JB2008 extends the earlier Jacchia 1970 series by assimilating extensive satellite orbit data, including drag-derived densities from 1970 to 2007. The model incorporates three key modifications: (a) an improved exospheric temperature formulation, (b) new density corrections based on geomagnetic storm-time activity, and (c) inclusion of a semiannual variation derived from multi-decadal observations. The JB2008 model was using four primary density data sources: Air Force daily density values (1997–2007), Air Force HASDM (High Accuracy Satellite Drag Model) model (2001–2005), CHAMP accelerometer-derived densities (2001–2005), and GRACE accelerometer-derived densities (2002–2005).

Inputs to JB2008 include the daily and 81-day averaged F10.7 solar flux, S10 solar EUV emission, M10 solar MUV radiation, Y10 Solar X-ray emission, the planetary geomagnetic index *ap*, and storm-time corrections derived from the Dst index. It has to be stressed that the input driver is provided by the model developer, and the values can sometimes differ substantially from the calibrated input used to drive the other models in this assessment at the CCMC. In addition, the input driver values available on the model distribution website are subject to occasional corrections and adjustments over time, rather than being fixed at the time of release. The driver used in the study was downloaded on 15 November 2024 (SOLFSMY and DTCFILE). The model outputs total mass density and temperature at altitude (i.e., not exospheric T and not constituents). JB2008 has been validated against independent satellite accelerometer data, demonstrating improved storm-time density response and long-term climatological consistency relative to earlier Jacchia and MSIS models.

3.4. GITM (Global Ionosphere-Thermosphere Model)

The GITM is a three-dimensional, physics-based, nonhydrostatic model of the coupled thermosphere-ionosphere system (Ridley et al., 2006). It solves the continuity, momentum, and energy equations for both neutral and ionized species on a user-specified vertical grid stretched in altitude. Unlike most thermospheric models, GITM does not assume hydrostatic equilibrium in its vertical velocity solver. This enables it to resolve large vertical winds and storm-time upwelling, both of which play a key role in shaping thermospheric neutral densities. Each neutral species has a separate vertical velocity, with coupling of the velocities through a frictional term.

Neutral density variability in GITM is driven by solar EUV heating, particle precipitation, ion-neutral interactions, and high-latitude electrodynamics. Different parameterizations of high-latitude electric fields, auroral particle precipitation, and energy deposition can be applied in GITM. In this study, the solar EUV is parameterized by the EUVAC empirical model (EUV flux model for Aeronomic Calculations; Richards et al., 1994) based on F10.7 proxy. The (Weimer (2005) model driven by solar wind inputs specifies the high-latitude electric potential. These inputs come from upstream solar wind and IMF (interplanetary magnetic field) observations from OMNIWeb (https://omniweb.gsfc.nasa.gov/). The auroral particle heating is provided by the Fuller-Rowell and Evans (Fuller-Rowell & Evans, 1987) model driven by the Hemispheric Power Index (HPI). Through self-consistent ion-neutral coupling, these drivers strongly modulate the thermospheric temperature and density structure, allowing GITM to capture storm-time enhancements and composition changes.

The simulation has a horizontal resolution of $4° \times 2°$ or $3.3° \times 3.3°$ in longitude and latitude depending on which on-premise machine at the CCMC is used to run the model. This occurs because the GITM is a parallel code and uses a 2D domain decomposition. The size of the horizontal grid in the simulation is determined by the number of processors on the given machine, in order to optimally configure and utilize computational resources. This issue has been fixed in the GITM RoR, and the output will be provided at $4° \times 2°$ (longitude × latitude) for future runs. The model lower boundary is located at 100 km, with two additional vertical levels extended downward to specify the lower boundary condition from MSIS. The upper boundary in these runs is located near 900 km. The vertical resolution is set to one-third of a scale height, with a total of 50 levels. The output cadence is 10 min.

3.5. TIE-GCM (The Thermosphere-Ionosphere Electrodynamics General Circulation Model)

TIE-GCM is a physics-based, three-dimensional, nonlinear model of the coupled thermosphere-ionosphere system with self-consistent electrodynamics. It solves the ion and neutral momentum, energy, continuity, and composition equations on constant pressure surfaces, with a lower boundary near 97 km and an upper boundary extending to approximately 500 km under solar minimum conditions (R. G. Roble & Ridley, 1987; R. G. Roble et al., 1988; Richmond et al., 1992). The electrodynamics module self-consistently calculates electric fields and ion drifts at low and middle latitudes, driven by the neutral-wind dynamo (Richmond et al., 1992). At high latitudes, the electric potential is imposed by an empirical model by default. Two commonly used empirical specifications of high-latitude ion convection in TIE-GCM are the Heelis model (Heelis et al., 1982) and the Weimer model (Weimer, 2005). Both configurations are employed in the assessment, which are referred to as TIEGCM-Heelis and TIEGCM-Weimer hereafter. The two-way ion-neutral coupling in TIE-GCM allows thermospheric neutral dynamics and energetics to be modulated by plasma motion and its associated electric fields through ion drag (V. M. Hsu et al., 2016), which has a broader implication on composition. In addition, Helium has been included as one of the major neutral constituents starting from version 2.0 (Sutton et al., 2015), along with N_2, O_2, and O. At the model lower boundary, migrating tides DW1 and SW2 in the neutral atmosphere are specified using monthly climatologies from the Global Scale Wave Model (Hagan & Forbes, 2002, 2003).

The horizontal resolution of output is 2.5° in latitude and longitude. The vertical resolution is one-quarter scale height. The solar UV radiation input is specified by the EUVAC model driven by F10.7 proxy (Richards et al., 1994). Kp is used to parameterize the high-latitude electric potential in TIEGCM-Heelis. For TIEGCM-Weimer, IMF components (Bx, By, Bz), solar wind density and speed, and dipole tilt angle are used to specify the ion convection pattern. The nominal model time step is 30 s, and the cadence of model output is 20 min.

3.6. WACCM-X (Whole Atmosphere Community Climate Model eXtended)

WACCM-X is a configuration of the National Center for Atmospheric Research Community Earth System Model version 2.2 (CESM 2.2.0, Hurrell et al., 2013), and released for community use in September 2021 (H.-L. Liu et al., 2018; J. Liu et al., 2018). The WACCM-X Run-on Request service is available to community in August 2022, which is the first whole atmosphere model at the CCMC. The model physics and dynamical core are based upon Community Atmosphere Model version 6 (CAM 6; Neale et al., 2010) and Whole Atmosphere Community Climate Model version 6 (WACCM6; Gettelman et al., 2019; Marsh et al., 2004). The model incorporates the chemical, dynamical, and physical processes to simulate the whole atmosphere spanning from the ground to about 4×10^{-10} hPa (~500 km geometric height in low solar flux years). A fourth-order divergence damping has been implemented in the WACCM-X simulations to remove small-scale numerical noise and instabilities. This procedure of damping is necessary to reduce non-physical and numerical effects in the model. Additional advances in the momentum and energy equation solvers have been improved to account for variable mean molecular mass and specific heat, which are important for accurately simulating the thermospheric condition.

The ionospheric electrodynamics in WACCM-X is adapted from the TIE-GCM. For this study, two widely used empirical specifications of high-latitude ion convection, the Heelis (Heelis et al., 1982) and Weimer (2005) models, are employed. Both ion convection empirical model driven by the same inputs described in Section 3.5. These configurations are hereafter referred to as WACCMX-Heelis and WACCMX-Weimer, respectively. Although the ionospheric electrodynamics in WACCM-X are adapted from TIE-GCM, the two models differ in several thermospheric parameterizations. In particular, the formulations of NO cooling and Joule heating differ between TIE-GCM and the current WACCM-X configuration. These differences in energy input and cooling processes can lead to different thermospheric temperature and density responses during geomagnetic storms. Updated formulations have been implemented for the next WACCM-X release in CESM3, where the earlier implementation has been corrected. A manuscript describing these updates is currently in preparation.

The horizontal resolution of WACCM-X output is 0.95° in latitude and 1.25° in longitude in the current study. The vertical resolution is one-quarter scale height above the stratosphere with totally 130 vertical levels. The finite volume dynamical core is implemented in the current model simulation. The model is free-running with an input of F10.7 proxy to drive the EUVAC model for solar radiation (Richards et al., 1994). The nominal model time step is 5 min, and the cadence of model output is 10 min.

3.7. CTIPe (Coupled Thermosphere Ionosphere Plasmasphere Electrodynamics)

The CTIPe is a global, three-dimensional, time-dependent, physics-based model of the upper atmosphere (Codrescu et al., 2012; Fedrizzi et al., 2012). CTIPe self-consistently solves the coupled momentum, energy, and composition equations for the neutral thermosphere, including O, O_2, and N_2, together with ionospheric electrodynamics and plasmaspheric transport. The neutral component explicitly accounts for solar EUV and UV heating, high-latitude Joule heating, auroral particle precipitation, and infrared cooling by NO, thereby capturing the primary drivers of thermospheric neutral density variability during geomagnetic disturbances. Neutral density output from CTIPe has been validated against CHAMP and GRACE accelerometer data, showing good agreement on climatological timescales and during geomagnetic storms (Fedrizzi et al., 2012). The inclusion of realistic high-latitude Joule heating has proven essential for improving storm-time density specification.

The model uses a latitude resolution of 2°, longitude resolution of 18°, and 15 pressure levels from a lower boundary at ∼80 km (1 Pa) to altitudes about 500 km. High-latitude ion convection is specified by the Weimer model driven by upstream solar wind conditions from the OMNIWeb. Auroral precipitation is parameterized using the HPI (Fuller-Rowell & Evans, 1987). The model time step is 1 min and the output cadence is 15 min.

3.8. WAM-IPE (Whole Atmosphere Model-Ionosphere Plasmasphere Electrodynamics)

WAM is a physics-based whole-atmosphere model developed from the Global Forecast System GCM, which is used by the U.S. National Weather Service for medium-range weather prediction. The model upper boundary is extended to approximately 3×10^{-9} hPa (∼600 km) with a total of 150 vertical layers (Akmaev et al., 2008; Fuller-Rowell et al., 2008). IPE is a physics-based ionosphere and plasmasphere model up to approximately 10,000 km (Maruyama et al., 2016). WAM is coupled one-way to IPE, meaning that the thermospheric parameters calculated by WAM are fed into IPE for calculating the responses in the ionosphere. It is important to emphasize that ionospheric variability in IPE does not feed back into WAM; instead, the ionospheric impact on the thermosphere is represented empirically (e.g., ion drag on neutral winds). In addition, the NO number density is prescribed through parameterized functions based on Marsh et al. (2004) rather than solved prognostically, and its radiative cooling effect is therefore not treated self-consistently in the simulation. These limitations may constrain the realism of neutral density representation in the current WAM-IPE framework.

Currently, WAM-IPE serves as the operational space weather model at the National Oceanic and Atmospheric Administration Space Weather Prediction Center (NOAA SWPC), providing forecasts of thermospheric neutral density and ionospheric conditions. The operational WAM lower atmosphere incorporates a three-dimensional variational assimilation scheme (3DVAR) with a 6-hr assimilation cycle to capture realistic lower-atmospheric variability. As demonstrated during the February 2022 storm event that caused the loss of ∼40 Starlink satellites, WAM-IPE captured the sudden density enhancements at altitudes near 200–300 km that drove excessive satellite drag (Fang et al., 2022).

In the CCMC Runs-on-Request (RoR) setting and this study, WAM-IPE is free-running without data assimilation. There is nothing constraining the lower atmosphere as in the operational WAM-IPE Forecast System. The version number of WAM-IPE employed in the study is 1.2.5. The model is driven by F10.7, IMF By and Bz, solar wind wind density and speed, and Kp index. The Weimer ion convection model and the TIROS auroral empirical model are employed to specify the external energy input from the magnetosphere (Fang et al., 2022). The output cadence is 20 min. The spatial resolution is 4° in longitude, 2° in latitude, and approximately one-quarter scale height in the vertical from the stratosphere and above.

4. Methodology

4.1. Construction of Fixed-Height Neutral Density Cubes and Flythrough

Physics-based model outputs are typically archived on pressure-coordinate grids, except for GITM. In the first step, model outputs are interpolated from their native pressure or altitude coordinates to fixed geometric altitudes to enable flythrough sampling along satellite tracks. For each model run, the major species (O, O_2, N_2, and He when explicitly resolved) are vertically log-interpolated onto a fixed-altitude grid spanning 100-1,000 km in 10 km increments at each column.

For models that do not explicitly resolve helium (WACCM-X, CTIPe, and WAM-IPE), an effective He ratio is derived from the NRLMSIS 2.0 driven by $F_{10.7}$ solar flux, daily Ap index, and the simulation date and time as

$$q^*_{\mathrm{He}} = \frac{n^{\mathrm{MSIS}}_{\mathrm{He}}}{n^{\mathrm{MSIS}}_{\mathrm{O}} + n^{\mathrm{MSIS}}_{\mathrm{O_2}} + n^{\mathrm{MSIS}}_{\mathrm{N_2}}}, \tag{1}$$

where $n^{\mathrm{MSIS}}_{\mathrm{He}}$, $n^{\mathrm{MSIS}}_{\mathrm{O}}$, $n^{\mathrm{MSIS}}_{\mathrm{O_2}}$, and $n^{\mathrm{MSIS}}_{\mathrm{N_2}}$ are number densities from MSIS.

The effective He ratio $\left(q^*_{\mathrm{He}}\right)$ is mapped onto the model grid and used to reconstruct the helium number density before recomputing the total mass density in WACCM-X, CTIPe, and WAM-IPE:

$$\hat{n}^{\mathrm{model}}_{\mathrm{He}} = q^*_{\mathrm{He}}\left(n^{\mathrm{model}}_{\mathrm{O}} + n^{\mathrm{model}}_{\mathrm{O_2}} + n^{\mathrm{model}}_{\mathrm{N_2}}\right), \tag{2}$$

where $\hat{n}^{\mathrm{model}}_{\mathrm{He}}$ is the estimated helium number density and n^{model}_i are the model-simulated number densities of O, O_2, and N_2.

The total neutral mass density is computed as $\rho = \sum_i n_i m_i$, where n_i and m_i are the number density and molecular mass of each constituent (O, O_2, N_2, and reconstructed He). The mass density are then combined across latitude, longitude, and time to form a four-dimensional neutral density field. The construction is carried out column by column to preserve local structure and temporal variability. This approach follows previous studies that adopted MSIS-based composition ratios to estimate helium densities when extending model thermospheric density fields to higher altitudes (e.g., Fang et al., 2022), and is also consistent with current NOAA/SWPC operational procedures for constructing thermospheric density fields used in satellite drag applications.

The resulting density cubes are written to NetCDF format, with time steps corresponding to the native model output cadence. Each file contains the full 4D neutral density field, enabling direct spatiotemporal sampling along satellite orbits (hereafter referred to as flythrough). In this approach, model densities are interpolated in space and time to the exact position of satellite measurements, producing time series directly comparable to observations. This standardized procedure provides a consistent framework for validation and skill-score evaluation across different thermosphere models.

4.2. Storm Phases

Geomagnetic storms are identified when $ap \geq 80$ (equivalent to $Kp \geq 6o$) and are included in the analysis only when observation data are available. Each event is divided into four phases following the procedure of Bruinsma and Laurens (2024). Figure 2 provides an illustration of these definitions, showing an example *ap* time series with shaded areas in different colors corresponding to each storm phase.

Phase 1 (pre–storm) extends from t_1 ($t_0 - 30$ hr) to t_2 ($t_0 - 18$ hr), where t_0 is the time when *ap* first reaches 80; this interval is used for model debiasing as described in the next section. Phase 2 (onset) covers the interval from t_2 to t_0. Phase 3 (main and recovery) is defined differently for single–peak (SP) in given ap time series and multiple–peak (MP) storms: for SP storms it spans t_0 to t_3 ($t_0 + 36$ hr), while for MP storms it extends until *ap* decreases below 80 (defined as t^*_0 in Figure 2b) and continues for an additional 36 hr. Phase 4 (post–storm) covers the subsequent period, extending 12 hr beyond the end of Phase 3 from t_3 to $t_4 = t_0 + 48$ hr for SP storms, and for MP storms $t_4 = t^*_0 + 48$ hr. This definition, based solely on the *ap*/*Kp* indices, provides a consistent framework for analyzing both single– and multiple–peak events.

4.3. Debiasing

There are many factors potentially contributing to the deviation between modeled and observed mass density. These include uncertainties in the solar and geomagnetic forcing proxies that drive the models, limitations in empirical representations of high-latitude electrodynamics and particle precipitation, and the absence or simplification of key physical processes such as gravity wave drag, composition transport, or ion-neutral coupling. On the observational side, errors in accelerometer calibration, drag coefficient assumptions, or orbit determination can also introduce biases (e.g., Mehta et al., 2023). Together, these model and measurement uncertainties compound, leading to systematic differences that vary with altitude, latitude, LT, and storm phase.

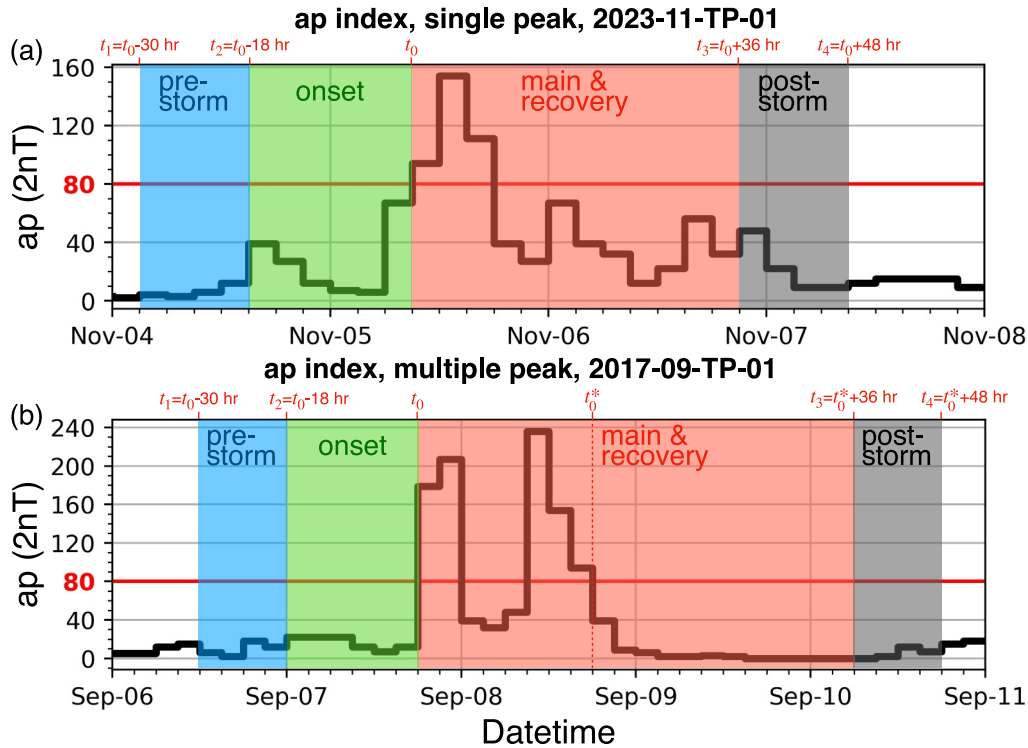


Figure 2. Example time series of the 3-hourly ap index for (a) single-peak and (b) multiple-peak cases. The horizontal red line at 80 (equivalent to Kp = 6o) marks the storm threshold used in this analysis. Storms are divided into four phases: pre-storm (blue shaded area), onset (green), main and recovery (red), and post–storm (gray). For single-peak events, the start and end times of each phase are defined with respect to t_0, when the ap index first reaches 80. The main and recovery phase spans t_0 to $t_3 = t_0 + 36$ hr, followed by the post-storm phase extending to $t_4 = t_0 + 48$ hr. For multiple-peak cases, the main and recovery phase continues until the ap index falls below 80 (defined as t_0^* in panel (b)) and extends for an additional 36 hr ($t_3 = t_0^* + 36$ hr). The post-storm phase then extends to $t_4 = t_0^* + 48$ hr. See text for details.

This study aims to focus specifically on biases associated with geomagnetic storms. We seek to distinguish biases that arise from the models' storm-time response from those linked to background climatology or solar cycle variability. To quantify these biases in a consistent manner, observed and modeled neutral densities are directly compared along each satellite track for a given TP. Bias factors in each TP are estimated in log space over the pre-storm phase using the geometric mean of the observed-to-modeled density ratio:

$$r_m = \exp\left(\left\langle \ln\frac{\rho_{\text{obs}}}{\rho_m}\right\rangle_{t_1:t_2}\right), \tag{3}$$

where ρ_{obs} and ρ_m denote the observed and modeled densities, respectively. $\langle\rangle_{t_1:t_2}$ denotes the discrete average over the interval $[t_1, t_2]$ as follows,

$$\langle f\rangle_{t_1:t_2} \equiv \frac{1}{N}\sum_{i=1}^{N} f(t_i), \qquad t_i \in [t_1, t_2], \tag{4}$$

The debiased model time series over the full window is then

$$\rho'_m(t) = r_m\,\rho_m(t), \qquad t \in [t_1, t_4]. \tag{5}$$

so that

$$\exp\left(\left\langle \ln\frac{\rho_{\mathrm{obs}}}{\rho'_m}\right\rangle_{t_1:t_2}\right)=1, \tag{6}$$

ρ'_m is referred to as debiased modeled mass density hereafter. The debiasing procedure is designed to minimize non-storm effects related to model errors as well as systematic offsets in accelerometer-derived density estimates. Therefore, the analyses presented in the following section focus on isolating storm-time responses while minimizing the influence of background biases.

4.4. Validation Metrics

Model performance is quantified using three complementary metrics, each evaluated using the observed and debiased modeled mass density for a given TP.

4.4.1. Debiased Mean Observed-to-Modeled Ratio

The primary metric is the debiased mean ratio between observed (ρ_{obs}) and debiased modeled mass densities (ρ'_m), defined in log space (Sutton, 2018) as

$$\mu(o/m)=\exp\left(\left\langle \ln\frac{\rho_{\mathrm{obs}}}{\rho'_m}\right\rangle_{t_a:t_b}\right), \tag{7}$$

The choice of $[t_a, t_b]$ depends on the storm phase: it corresponds to the onset phase, the combined main and recovery phase, or the post–storm phase, as specified in Section 4.2. A value of $\mu(o/m)=1$ indicates perfect agreement, while $\mu(o/m)>1$ indicates systematic model underestimation and $\mu(o/m)<1$ implies a model overestimation.

4.4.2. Standard Deviation

The scatter of the observed-to-modeled ratios is quantified in log space as

$$\sigma(o/m)=\sqrt{\left\langle\left(\ln\frac{\rho_{\mathrm{obs}}}{\rho'_m}-\ln\mu(o/m)\right)^2\right\rangle_{t_a:t_b}}, \tag{8}$$

which measures variability about the geometric mean ratio. For convenience, $\sigma(o/m)$ is often reported as a percentage (e.g., Sutton, 2018), calculated as $100\times(\exp(\sigma_m)-1)\%$.

4.4.3. Correlation Coefficient

Temporal coherence between the observed and modeled densities is quantified using the Pearson correlation coefficient, applied to the debiased model output ρ'_m and ρ_{obs}. Values close to unity indicate strong agreement in temporal variability.

These three metrics together characterize model bias, amplitude of variability, and phase agreement with observations, providing a comprehensive evaluation framework (Bruinsma & Laurens, 2024; Sutton, 2018).

5. Results

5.1. Scorecard

5.1.1. Debiased Mean Observed-to-Modeled Ratio

Figures 3–5 summarize the averaged debiased mean observed-to-modeled density ratios for all events, separated by storm phase, providing a phase-resolved view of model performance. The empirical models exhibit smaller biases than the physics-based models. During the storm onset (Figure 3), most models show only modest underestimation relative to observations, except for MSIS. MSIS20 provides the closest agreement overall ($\mu(o/m)=1.00\pm18.5\%$), followed by MSISE00 ($\mu(o/m)=0.99\pm18.6\%$). However, the standard deviation

10.1029/2025SW004782

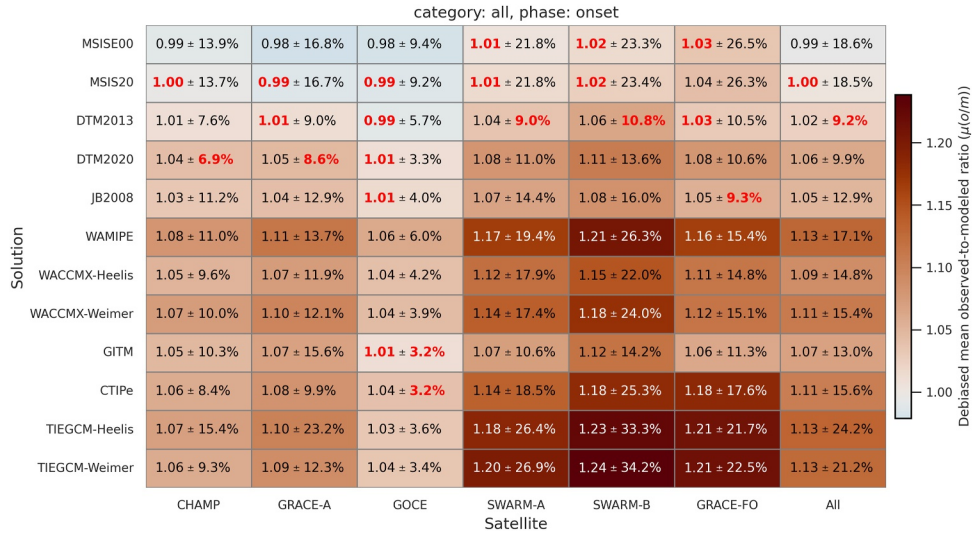


Figure 3. Scorecard of the debiased mean observed-to-modeled ratio ($\mu(o/m)$) by model solution and satellite, averaged over all 151 events during the onset phase. Each cell reports the mean ± standard deviation of $\mu(o/m)$. Cell background colors are scaled to the mean values using a diverging colormap centered at 1 (white). Values below 1 are mapped toward blue hues and values above 1 toward dark red. The best value in each satellite group (column), for either mean or standard deviation, is highlighted in red. White text is used solely to improve readability against darker backgrounds.

of $\mu(o/m)$ in MSIS is large compared to most models in the assessment, with the exception of TIE-GCM. Such a result suggests that the debiased mean ratio of MSIS during the onset is more widely spread than others, making it relatively more difficult to correct the model density using the mean ratio. GITM performs best among the physics-based models during the onset phase aggregated over all satellite observations, yielding $\mu(o/m) = 1.07 \pm 13.0\%$. The averaged debiased mean ratio and its standard deviation tend to increase with observation altitude (see Figure 1 for the mean observed altitude), a dependence particularly evident when comparing results from a given model between Swarm-A and Swarm-B, as Swarm-B orbits at a higher altitude while sampling nearly the same longitudes and latitudes at the same time.

In the main and recovery phase (Figure 4), biases increase across nearly all models except for DTM2013 and JB2008 as density enhancements grow stronger. The MSIS models underestimate densities by roughly 20%–30%, resulting in the largest overall bias among the models considered. On average, DTM2013 and JB2008 yield

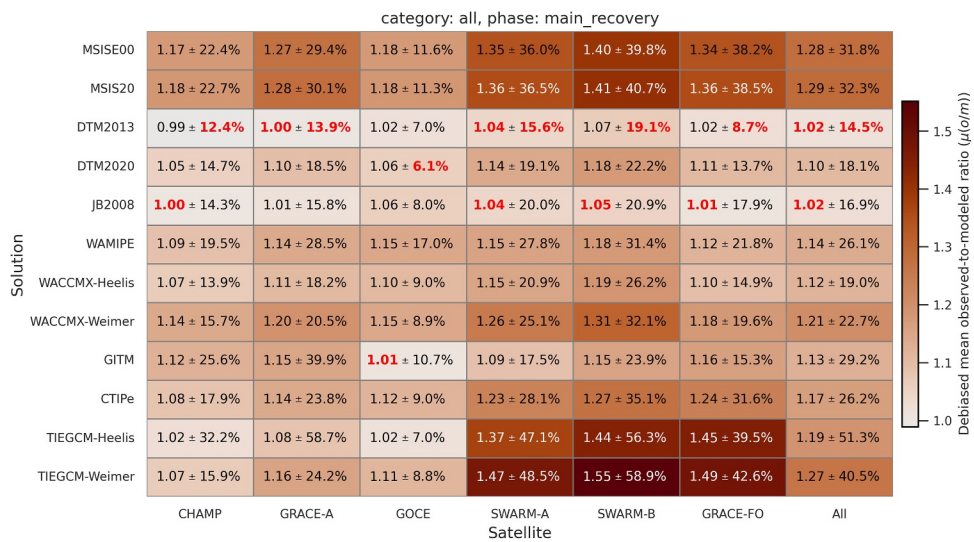


Figure 4. Same as Figure 3 but for the main and recovery phase.

10.1029/2025SW004782

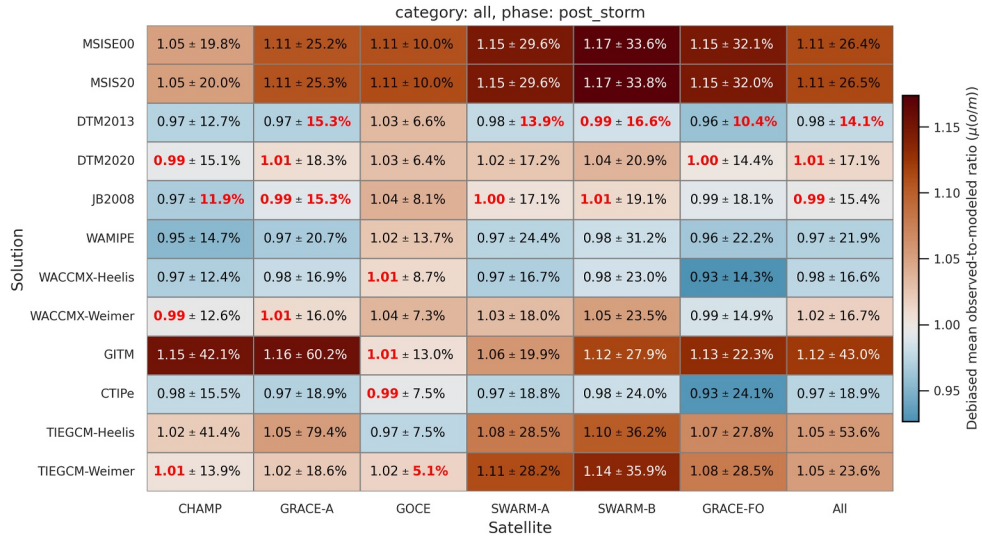


Figure 5. Same as Figure 3 but for the post–storm phase.

$\mu(o/m) = 1.02 \pm 14.5\%$ and $\mu(o/m) = 1.02 \pm 16.9\%$ respectively, obtaining the best performance in the group. Among the physics-based models, WACCMX-Heelis (best averaged mean ratio and standard deviation) outperforms the rest of the group, while all the physics-based model yield a better performance than MSIS models which are considered as the baseline. It is also interesting to note that WACCMX-X and TIE-GCM driven by the Heelis electric potential model both show smaller bias than their respective Weimer-driver counterpart, which is roughly 10% closer to unity across all satellite observations.

The difference in the debiased mean ratios between the TIE-GCM and WACCM-X simulations using the Heelis and Weimer specifications likely reflects differences in how these empirical ion convection specifications represent magnetospheric forcing. The Heelis convection pattern parameterizes ionospheric convection patterns using Kp indices, and has been extensively tuned over many years to better represent high-latitude magnetospheric inputs in NCAR-developed models, compared to the Weimer model (e.g., Burns et al., 2012). In contrast, the Weimer model derives empirical electric potentials based on upstream solar wind and IMF measurements. Differences between these formulations can lead to variations in the spatial distribution and magnitude of ionospheric convection, Joule heating, and ion drag (e.g., Richmond et al., 1992), particularly during geomagnetic storms when high-latitude forcing becomes highly variable. Because Joule heating and ion drag are among the primary drivers of storm-time thermospheric circulation and density variability, differences in the electrodynamic forcing specification can propagate into the simulated thermospheric response. For example, Joule heating rate produced by the Weimer convection pattern show a negative bias in the TIE-GCM simulations when comparing to European incoherent scatter (EISCAT) radar measurements (Bag et al., 2024), which in principle leads to a cooler modeled thermosphere and correspondingly lower neutral densities (equivalent to higher $\mu(o/m)$). However, we note that the relatively better performance of the Heelis-driven simulations in this study may reflect a thermospheric response i.e. closer to the observed density variability, rather than necessarily indicating a more accurate representation of the high-latitude electric potential or the associated electrodynamic forcing applied to the upper atmosphere in the NCAR-based models. Such differences in the underlying physical and chemical processes warrant further investigation in future studies.

Following storm main and recovery phase, most models revert toward unity ratio in the post–storm phase shown in Figure 5 with several even slightly overestimating densities. DTM and JB2008 show excellent agreement with $\mu(o/m) \approx 1.00$ and the standard deviation is smaller than other model. WACCM-X with Heelis driver shows the best score among the physics-based models with the best averaged mean ratio and smallest standard deviation. A weak altitude dependence is still apparent. The models tend to produce slightly larger biases at the higher altitudes sampled by Swarm-B and GRACE-FO, although the spread is narrower than in the onset and main/recovery phases. This indicates that model biases partly relax toward climatological conditions during the post-storm phase.

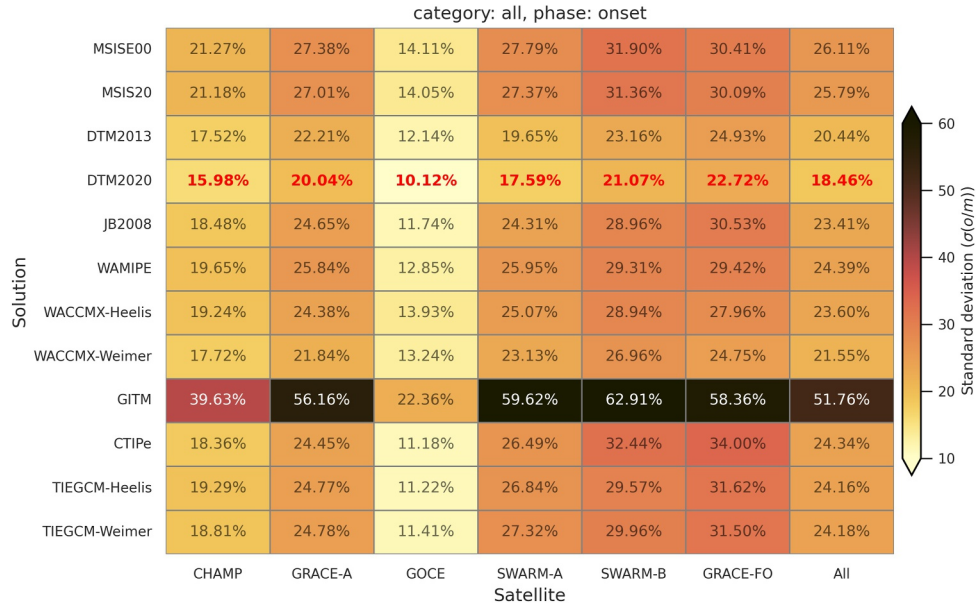


Figure 6. Scorecard of the standard deviation of the observed-to-modeled density ratio, $\sigma(o/m)$, by model solution and satellite, averaged over all 151 events during the storm onset. Each cell reports $\sigma(o/m)$ in percent. Cell background colors are scaled to $\sigma(o/m)$ using a sequential colormap, with lighter shades indicating smaller scatter (better performance). The lowest $\sigma(o/m)$ value within each satellite group is highlighted in red. White text is used solely to improve readability against darker backgrounds.

Overall, the model performance varies systematically across storm phases, with the largest discrepancies occurring during the main and recovery phase when thermospheric density enhancements peak. DTM2013 and JB2008 consistently provide the closest agreement to observations across all phases, with biases typically within a few percent of unity and the standard deviation of $\mu(o/m)$ is considerably smaller than others. DTM2020 also produces moderate biases across all storm phases. Physics-based models exhibit larger storm-time biases but generally outperform the MSIS models, which underestimate densities during the storm peak. Among physics-based models, WACCMX-Heelis, and GITM demonstrate relatively strong performance, but the standard deviation of $\mu(o/m)$ in GITM is about 2 times larger than WACCMX-Heelis. The WACCM-X solutions driven by the Heelis convection pattern systematically outperform their Weimer-driven counterparts, whereas TIEGCM-Heelis shows better performance than TIEGCM-Weimer only during the main and recovery phase. In the post-storm phase, nearly all models return toward climatological conditions, yielding ratios closer to unity and reduced spread, indicating partial relaxation of storm-time biases.

The altitude dependence of $\mu(o/m)$ in the physics-based models throughout the three phases likely reflects a combination of factors. At higher altitudes, the aerodynamic drag signal becomes weaker relative to other non-gravitational forces such as radiation pressure, which increases the uncertainty in density retrievals and makes the derived densities more sensitive to modeling assumptions (e.g., radiation pressure or aerodynamic force coefficient modeling) (e.g., Bruinsma et al., 2023; Mehta et al., 2023; Siemes et al., 2023). In addition, changes in thermospheric composition with altitude, including the increasing contribution of helium, may introduce additional uncertainties either in physics-based model simulations or in the density retrieval process, whereas the latter particularly through assumptions in the gas–surface interaction model used to estimate the aerodynamic force coefficient. These factors may partly explain the larger deviations between physics-based model results and observations for higher-altitude satellites such as Swarm-B and GRACE-FO. By contrast, at lower altitudes (e.g., GOCE), the stronger drag signal generally results in more robust density retrievals and smaller observational uncertainties. A detailed investigation of these effects is beyond the scope of the present study, though this warrants further investigation.

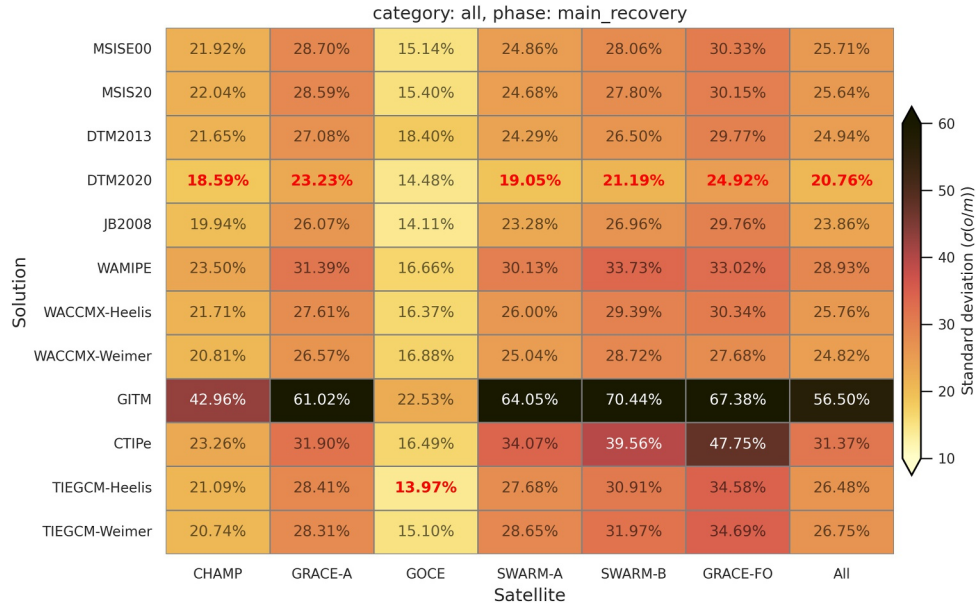


Figure 7. Same as Figure 6 but for the main and recovery phase.

5.1.2. Standard Deviation

Figures 6–8 present the standard deviation, $\sigma(o/m)$, for all events, separated by storm phase. Smaller values of $\sigma(o/m)$ indicate reduced scatter and thus better agreement in the temporal and spatial variability between modeled and observed densities.

During the storm onset phase (Figure 6), most models produce $\sigma(o/m)$ values in the range of 20%–30%, with DTM2020 showing the lowest values across all satellites ($\sigma(o/m) = 18.46\%$ overall), followed by DTM2013. Among the physics-based models, WACCMX-Weimer achieves the smallest onset standard deviation ($\sigma(o/m) = 21.55\%$), whereas GITM exhibits the largest spread ($\sigma(o/m) > 50\%$), particularly for GRACE and Swarm satellites where fly in relatively high altitudes.

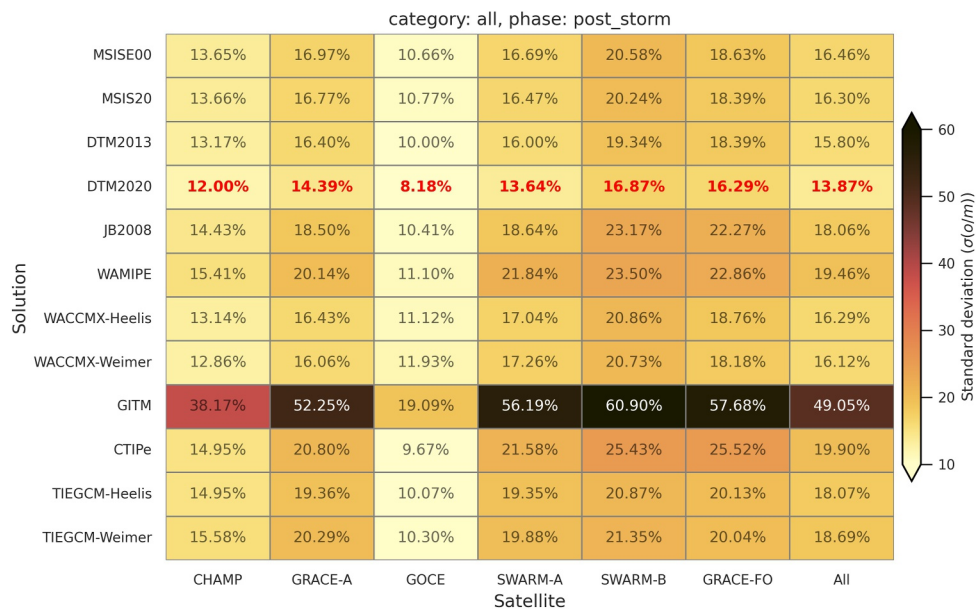


Figure 8. Same as Figure 6 but for the post–storm phase.

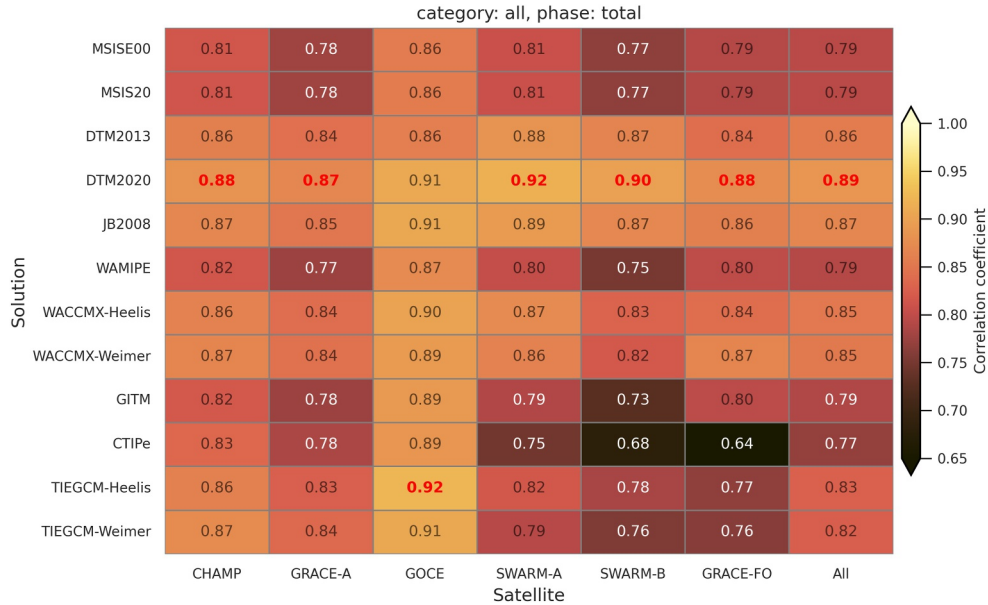


Figure 9. Scorecard of the Pearson correlation coefficient R between observed and modeled neutral densities, by model solution and satellite, averaged over all 151 events from the onset to post–storm phases. Cell background colors are scaled to R using a sequential colormap, with lighter shades indicating stronger agreement. The highest R within each satellite group is highlighted in red. White text is used solely to improve readability against darker backgrounds.

In the main and recovery phase (Figure 7), $\sigma(o/m)$ slightly increases from the onset phase for nearly all models as density enhancements intensify. DTM2020 again provides the best performance, with $\sigma(o/m) = 20.76\%$ overall and consistently low standard deviation across satellites. JB2008, DTM2013, and WACCMX-Weimer also perform well, maintaining $\sigma(o/m) < 25\%$. GITM shows the highest values, exceeding 60% for GRACE-A and more than 70% for Swarm-B, suggesting that it captures strong variability but tends to overshoot observed fluctuations. TIE-GCM solutions show moderate performance, with slightly smaller scatter for Weimer-driven runs compared to Heelis-driven runs.

During the post-storm phase (Figure 8), most models revert to substantially lower standard deviation as the thermosphere relaxes toward quieter conditions. DTM2020 continues to outperform all other models with $\sigma(o/m) = 13.87\%$. The physics-based models improve compared to the storm phases but still yield $\sigma(o/m)$ values around 16%–20%, yet GITM remains the most fluctuated than observation ($\sigma(o/m) \approx 49\%$ overall).

Across all storm phases, DTM2020 consistently exhibits the lowest standard deviation, indicating the most stable agreement with observations and the smallest variability. DTM2013 and JB2008 also perform well, maintaining relatively low $\sigma(o/m)$ throughout. Among the physics-based models, WACCMX–Weimer and WACCMX-Heelis show the smallest variability, while TIE-GCM solutions achieve moderate performance. GITM displays the largest $\sigma(o/m)$ across all phases, especially at higher altitudes (GRACE and Swarm-B). In the post–storm phase, most models converge toward significantly lower $\sigma(o/m)$, reflecting a relaxation of the thermosphere toward quiet conditions, but GITM remains noticeably more variable than the others.

5.1.3. Correlation Coefficient

Figure 9 shows the Pearson correlation coefficient R between observed and debiased modeled densities, averaged over all storm phases (onset, main/recovery, and post–storm). Overall, the DTM2020 solution achieves the highest correlations across most satellites ($R = 0.89$ when averaged over all observations), followed closely by JB2008 and DTM2013. Among physics–based models, WACCMX and TIEGCM solutions driven by the Heelis and Weimer convection patterns exhibit moderate correlations ($R \approx 0.82$–0.85), while WAM-IPE, GITM, and CTIPe yield systematically lower correlations, especially at higher altitudes (e.g., Swarm-B, GRACE-FO). The empirical models generally outperform the physics–based models in reproducing the temporal variability of density, indicating that their storm–time responses are well synchronized with the observations. The results also

10.1029/2025SW004782

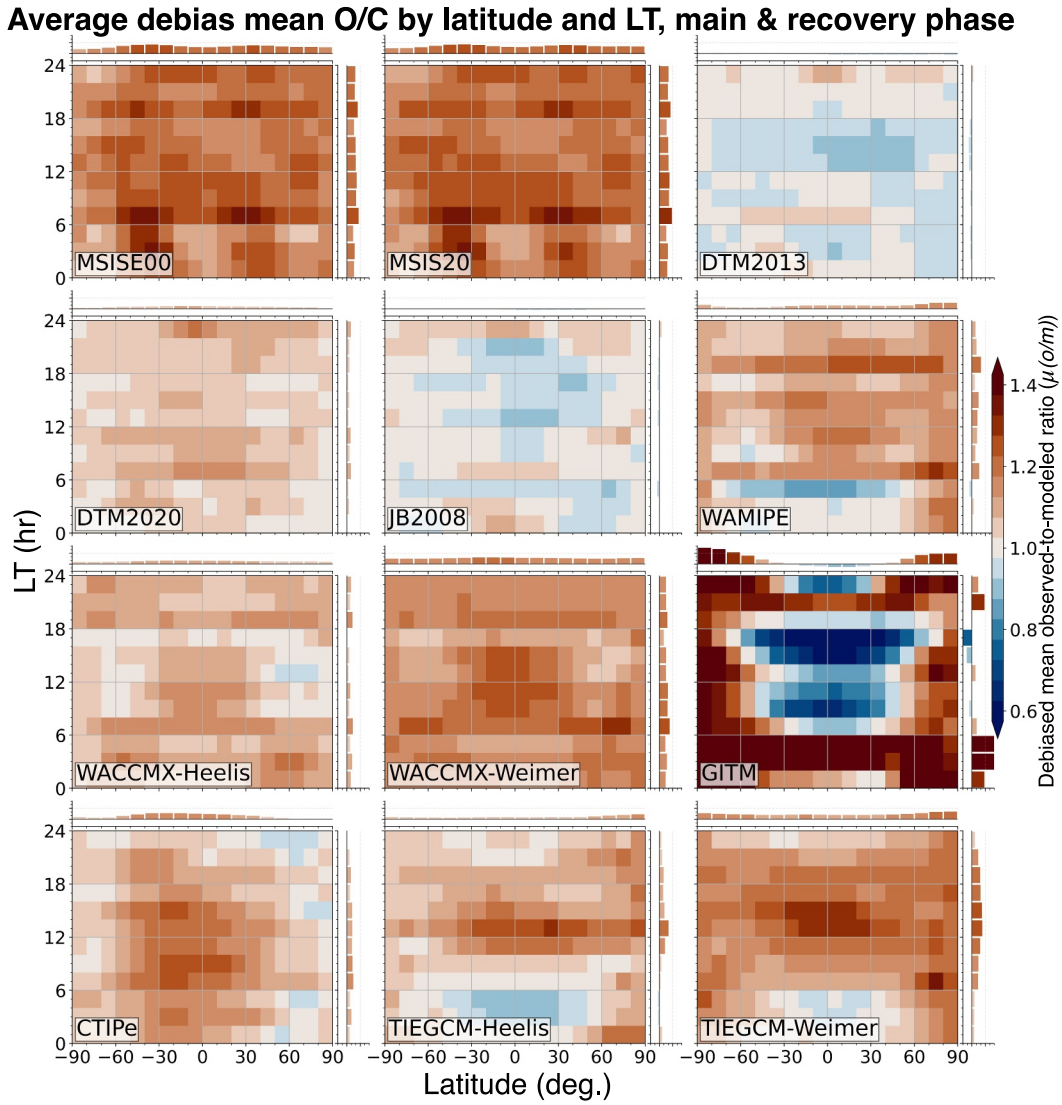


Figure 10. Mean ratio of debiased observed to model-computed neutral mass density (debiased mean O/C) as a function of local time (LT) and latitude during the main and recovery phase for (a) MSISE00, (b) MSIS20, (c) DTM2013, (d) DTM2020, (e) JB2008, (f) WAM-IPE, (g) WACCMX-Heelis, (h) WACCMX-Weimer, (i) GITM (j) CTIPe, (k) TIEGCM-Heelis, and (l) TIEGCM-Weimer. Color are shown only for bins containing at least 10 samples. Each panel includes two marginal inset plots: (top) arithmetic mean of the debiased mean O/C versus latitude, aggregated over all LT; (right) arithmetic mean of the debiased mean O/C versus LT, aggregated over all latitude. The solid black lines in the marginal plots indicate a debiased mean O/C of 1, representing perfect agreement between observations and models. The height of the bars in the marginal plots represents the deviation of the debiased mean O/C from 1. The minor tick interval in the marginal inset plots is 0.1. Red shading and bars indicate underestimation (debiased mean O/C $>$ 1), while blue shading and bars indicate overestimation (debiased mean O/C $<$ 1).

suggest that while physics–based models capture the general storm–time density evolution, their representation of short–timescale variability remains less consistent with observations compared to latest empirical models.

In summary, across all metrics, model performance exhibits a clear phase dependence, with the largest discrepancies occurring during the main and recovery phase when thermospheric density enhancements peak. Empirical models (DTM2020, DTM2013, JB2008) consistently provide the best overall agreement with observations, yielding debiased mean ratios closest to unity, the smallest scatter $\sigma(o/m)$, and the highest correlations R.

Among the physics–based models, WACCMX–Heelis emerges as the most reliable performer, producing the smallest storm–time bias, moderate standard deviation, and robust correlation. In the post–storm phase, nearly all

10.1029/2025SW004782

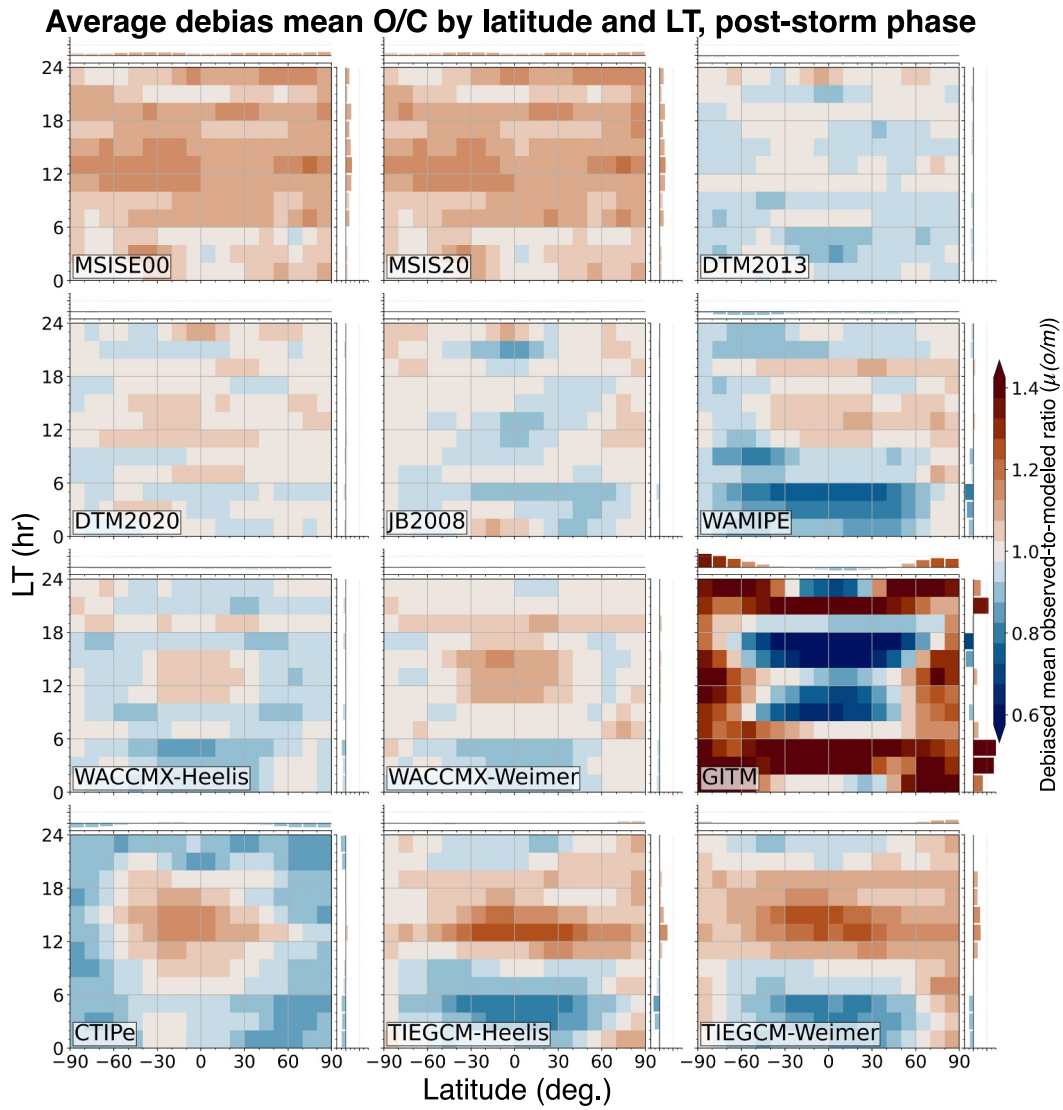


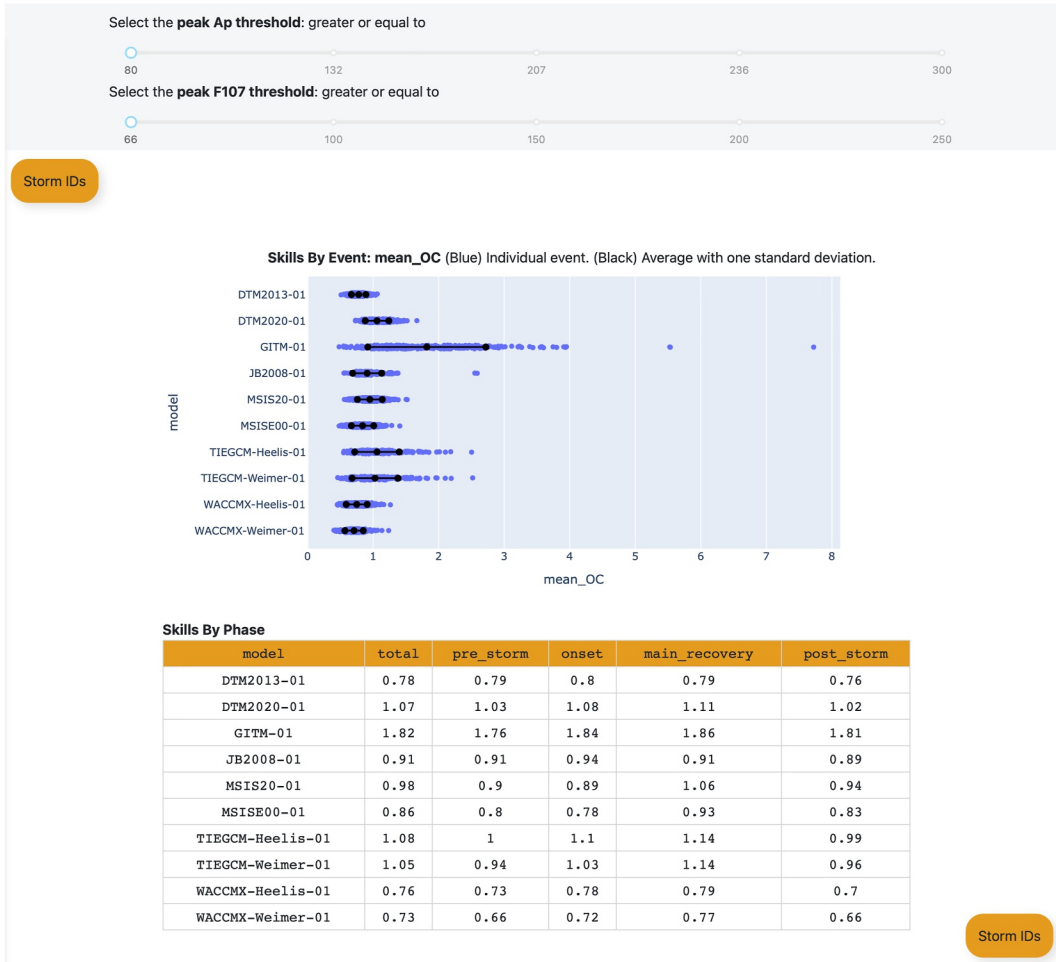
Figure 11. Same as Figure 10 but for the post–storm phase.

models relax toward climatological conditions, reducing both bias and scatter. Meanwhile, MSIS models systematically underestimate storm–time densities by 20%–30%, reinforcing the need for bias correction when using them as baselines.

5.2. Latitude-Local Time Dependence of Biases

Figures 10 and 11 present the debiased mean observed-to-modeled density ratios, $\mu(o/m)$, as a function of latitude and LT for the main and recovery phase and the post–storm phase, respectively. These maps reveal systematic spatial patterns that highlight where models perform best and where significant discrepancies persist from model development perspective.

During the main and recovery phase (Figure 10), a pronounced underestimation is observed at high latitudes for WAM-IPE, GITM, and TIE-GCM, particularly within the auroral zones (60°–80° in both hemispheres). This bias likely reflects underestimated or mis-timed storm-time energy inputs, such as Joule heating and auroral particle precipitation, in the high-latitude electrodynamic drivers. At mid-latitudes, most physics-based models perform better, with the notable exception of CTIPe, which exhibits increasing bias toward lower latitudes.



model	total	pre_storm	onset	main_recovery	post_storm
DTM2013-01	0.78	0.79	0.8	0.79	0.76
DTM2020-01	1.07	1.03	1.08	1.11	1.02
GITM-01	1.82	1.76	1.84	1.86	1.81
JB2008-01	0.91	0.91	0.94	0.91	0.89
MSIS20-01	0.98	0.9	0.89	1.06	0.94
MSISE00-01	0.86	0.8	0.78	0.93	0.83
TIEGCM-Heelis-01	1.08	1	1.1	1.14	0.99
TIEGCM-Weimer-01	1.05	0.94	1.03	1.14	0.96
WACCMX-Heelis-01	0.76	0.73	0.78	0.79	0.7
WACCMX-Weimer-01	0.73	0.66	0.72	0.77	0.66

Figure 12. Screenshot of the ITMAP (Ionosphere–Thermosphere Model Assessment Platform) interface. The platform allows users to filter storm events by solar and geomagnetic thresholds and compare model skill scores by event and storm phase, highlighting differences in accuracy and variability across empirical and physics-based models.

A clear LT dependence also emerges in the physics-based models, though the behavior differs between models. Most physics-based solutions display a diurnal variation in $\mu(o/m)$, with larger values during daytime equatorward of 30°, except for GITM, which shows an opposite behavior. Interestingly, both whole-atmosphere models, WAM-IPE and WACCM-X, display enhanced biases in the dawn and dusk sectors. Meanwhile, TIE-GCM and WACCM-X driven by the Weimer convection pattern show stronger LT and latitude dependence and generally larger biases compared to their respective Heelis-driven runs.

In contrast, the empirical models show more uniform behavior across latitude and LT. DTM2020 and JB2008 in particular demonstrate relatively small latitudinal and LT dependence, which is not surprising given that these models are assimilated or tuned using long-term, globally distributed density data and therefore inherently incorporate a portion of the storm-time variability already present in the observational data sets in the study. MSIS models continue to show a pronounced systematic underestimation across nearly all regions, reinforcing their known limitations during storm-time conditions.

Following the storm peak, during the post-storm phase (Figure 11), high-latitude biases decrease in magnitude but are not fully eliminated. Mid-latitude and equatorial regions largely return to $\mu(o/m) \approx 1.0$ on average, indicating a relaxation toward climatological conditions. At low latitudes, however, the diurnal variation in $\mu(o/m)$ becomes more pronounced in the physics-based models, with a clear maximum near 1400 LT for WAM-IPE, WACCM-X, CTIPe, and TIE-GCM. In contrast, GITM exhibits the opposite diurnal pattern. The TIE-GCM results show hemispheric differences in the local-time dependence of the biases, with distinct LT structures

between the Northern and Southern Hemispheres and generally larger bias magnitudes in the Northern Hemisphere for both the Heelis- and Weimer-driven simulations.

GITM reproduces the debiased mean density ratios during the onset, main, and recovery phases (Figures 3–5); however, it exhibits a standard deviation nearly twice that of the other models and a comparatively lower correlation in the overall assessment (Figures 6–8). In addition, a pronounced local-time and latitude dependence is evident in the debiased mean ratios, with enhanced daytime values and stronger deviations at high latitudes (Figures 9–11). This "accurate mean but larger standard deviation" characteristic may be related to global circulation changes during geomagnetic storms, which are driven by enhanced high-latitude Joule heating and ion drag drive (Rees et al., 1983; R. G. Roble et al., 1987). The circulation pattern changes typically induce upwelling in the auroral regions, balanced by equatorward downwelling in the thermosphere (R. Roble, 1986). The associated adiabatic cooling and heating in the high-latitude and subsiding regions lead to changes in the neutral temperature structure and consequently redistribute mass density throughout the thermosphere.

To further examine this behavior, a latitude-height contour plot of the debiased mean ratio is provided in the Supplement. The latitude-dependence is consistent with circulation-driven temperature variations associated with storm-time adiabatic processes, including cooling in high-latitude upwelling regions and warming in equatorward downwelling regions. THE latitude-, local-time-, and height-dependent variations in the debiased mean ratios may suggest that the simulated storm-time circulation in GITM is too strong, leading to enhanced adiabatic processes that produce the observed structure in the debiased mean ratio. Such spatially structured variability can account for the larger standard deviation seen in Figures 6–8, without implying a systematic bias in the mean storm-time response. A detailed validation of the modeled circulation and thermodynamic responses (neutral temperature, NO cooling, and Joule heating rate during geomagnetically storm times), however, is beyond the scope of the present study and warrants dedicated investigation in future work.

Overall, these latitude-LT maps highlight key strengths and limitations of the models in capturing storm-time thermospheric density variability. Physics-based models reproduce many of the expected large-scale spatial features of the storm response but still exhibit systematic high-latitude underestimation and notable diurnal structure in $\mu(o/m)$, suggesting that high-latitude energy inputs, circulation, and lower-thermospheric wave forcing remain areas for improvement. The sensitivity of TIE-GCM and WACCM-X to the choice of high-latitude electrodynamic driver (Heelis vs. Weimer) further underscores the importance of accurately specifying magnetospheric forcing. In contrast, empirical models such as DTM2020 and JB2008 provide more uniform performance across latitude and LT. These results emphasize the need for continued development and validation of physics-based models to reduce biases, particularly in auroral regions and during periods of strong geomagnetic activity, while retaining their ability to resolve dynamic storm-time responses.

5.3. CCMC Ionosphere-Thermosphere Model Assessment and Validation Platform (ITMAP)

A key challenge for the community has been the absence of a centralized, reproducible framework for tracking progress in model development. To address this, we developed the CCMC Ionosphere-Thermosphere Model Assessment and Validation Platform (ITMAP; https://ccmc.gsfc.nasa.gov/itmap/), an interactive web-based system that provides a unified environment for model-data intercomparison. ITMAP integrates a wide range of satellite data sets introduced in the study with IT models hosted at CCMC. Validation results are summarized in the form of a thermospheric model scorecard, which enables direct intercomparison of empirical and physics-based models across multiple geomagnetic storm events. Figure 12 shows a screenshot of the ITMAP (Ionosphere–Thermosphere Model Assessment Platform). The platform enables users to filter storm events by solar and geomagnetic thresholds and compare model skill scores both by event and storm phase, highlighting accuracy, variability, and performance differences across empirical and physics-based models. This capability supports robust model-to-data assessment for applications such as satellite drag computations.

The ITMAP web application was released in beta form in May 2025 and is now publicly accessible. Ongoing development is focused on expanding the number of storm events covered, integrating additional observational data sets, and enabling automated, reproducible scorecard generation. A major future objective is to establish a long-term validation campaign to track model performance across solar cycles without debiasing, supporting both scientific research and operational space weather forecasting.

6. Summary and Future Outlook

This study presents a multi-mission, phase-resolved evaluation of thermospheric neutral density during 151 geomagnetic storms (2001–2023), using high-resolution accelerometer- and POD-derived observations from CHAMP, GRACE-A, GRACE-FO, GOCE, and Swarm-A/B. Model skill is quantified with three complementary metrics applied in log space–debiased mean observed-to-modeled ratio, standard deviation, and correlation after pre-storm debiasing to minimize non-storm offsets.

Overall performance exhibits a clear phase dependence, with the largest discrepancies during the main and recovery phase when density enhancements peak. Among empirical models, DTM2020 performs best across all three metrics (ratios closer to unity, smallest $\sigma(o/m)$, highest R), followed by JB2008 and DTM2013. MSIS models systematically underestimate storm-time densities. Among physics-based models, WACCMX-Heelis emerges as the most reliable overall, with comparatively small storm-time bias, samllest standard deviation, and robust correlation.

GITM matches the debiased mean ratios during the onset and main and recovery phases, but produces nearly twice the standard deviation of other models. GITM's correlation is relatively lower compared to other models in the assessment. GITM LT–latitude dependence of the debiased mean ratio is also evident, with larger values during daytime and stronger biases at high latitudes. The "accurate mean but larger standard deviation" characteristic may be related to storm-time global circulation changes driven by enhanced high-latitude Joule heating and ion drag (Rees et al., 1983; R. G. Roble et al., 1987). The associated high-latitude upwelling and equatorward downwelling produce adiabatic cooling and heating that modify the neutral temperature structure and redistribute thermospheric mass density. These circulation-driven temperature perturbations alter the local scale height and can introduce latitude-, local-time-, and height-dependent variability, increasing the standard deviation without introducing a systematic bias in the mean storm-time response. This behavior may suggest that the simulated storm-time circulation in GITM is too strong, leading to enhanced adiabatic processes that produce the observed LT-latitude structure in the debiased mean ratio.

WAM-IPE and CTIPe show moderate performance. Their debiased mean ratios are reasonable during the onset phases but deteriorate in the main and recovery phase. Meanwhile, they exhibit larger values in standard deviation ($\sigma(o/m)$) and weaker correlations, indicating that their storm-time specification and dynamics still require refinement. TIE-GCM displays systematic high-latitude underestimation and moderate standard deviation. Its correlation coefficients ($R \approx 0.82$–0.84) indicate reasonable phase alignment with observations but reduced amplitude response, particularly during strong geomagnetic driving. Sensitivity to the specification of high-latitude electrodynamics is evident: Heelis-driven runs of both WACCM-X and TIE-GCM generally outperform their Weimer-driven counterparts. Spatial diagnostics further reveal persistent high-latitude underestimation and notable local-time structure in $\mu(o/m)$ for several physics-based models, while empirical models display more uniform latitude-LT behavior. A weak but recurring altitude dependence is present particularly in all physics-based models, with larger biases and standard deviation at the higher sampling altitudes of Swarm-B and GRACE-FO.

Building on these findings, we identify the following priorities for advancing storm-time thermospheric density specification and forecast skill:

- *High-latitude forcing.* Reduce biases linked to electrodynamic drivers by refining empirical convection patterns and auroral precipitation specifications, and by assimilating high-latitude observations (e.g., SuperDARN radars and ISR measurements) or state-of-art ionospheric electrodynamics outputs (e.g., MAGE, SWMF, AMGeO, and AMPERE) to better constrain Joule heating and ion convection patterns.
- *Lower-atmosphere coupling.* Improve representation of tides and planetary waves, and gravity-wave momentum deposition through nudging or data assimilation of reanalysis products (e.g., TIE-GCM version 3.0; Wu et al., 2025); quantify their role in the observed diurnal structure of $\mu(o/m)$.
- *Composition and helium.* Standardize helium treatment across models (self-resolved or MSIS-informed) and assess its impact on density at Swarm-B/GRACE-FO altitudes; extend composition validation with independent O/He proxies where available.
- *Uncertainty quantification.* Report metric uncertainties alongside scorecards by propagating observational errors (e.g., accelerometer calibration and C_D models) and driver uncertainties (e.g., OMNI solar wind inputs

or empirical convection patterns). These inputs can be perturbed within their estimated uncertainties to assess the sensitivity of modeled thermospheric responses (e.g., C. Hsu & Pedatella, 2025). In addition, develop ensemble-based model scorecards to improve operational relevance.

- *ITMAP roadmap.* Use the CCMC ITMAP platform as the community hub for reproducible, continuously updated validation: (a) expand the assessment to encompass full-year coverage, including both geomagnetically disturbed and quiet periods. (b) add new data sets (e.g., temperature and O/N^2 ratio from GOLD, neutral wind from ICON), (c) automate scorecard generation and API access for near-real-time evaluation, and (d) link to application-level tools (satellite drag and reentry, POD, such as in Waldron et al. (2024)).
- *Operations-ready benchmarks.* Establish a small set of open, versioned benchmark storms (quiet-to-severe) with frozen inputs/drivers to track model changes over time and to facilitate transition to operations for space-situational-awareness use cases.

In summary, rather than relying on MSIS for satellite drag computations, our results show that DTM2020, DTM2013, and JB2008 consistently outperform MSIS in both accuracy and robustness across all storm phases. We therefore strongly recommend adopting DTM and JB2008 as empirical models to compute satellite drag, providing a more reliable reference against which physics-based models can be assessed. While empirical models currently achieve closer agreement with observations, physics-based models show strong potential for capturing the dynamic storm-time responses when supported by improved drivers and coupling.

ITMAP, funded by NASA and NSF, will continue to track and document improvements in model performance, ensuring that progress is transparent and accessible to the community. With ITMAP now publicly available in beta, a long-term validation campaign spanning solar cycles will enable systematic assessment of model biases, more rigorous quantification of uncertainties, and ultimately more reliable storm-time density nowcasts to support orbit prediction and space weather operations.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

Neutral density data set is available in ASCII table format on the web server, designated as version 2 (V2) (https://thermosphere.tudelft.nl/data/data/; Siemes, 2024). Simulation outputs can be obtained through the CCMC Metadata Registry (https://kauai.ccmc.gsfc.nasa.gov/CMR/TimeInterval/viewAllTI; NASA Community Coordinated Modeling Center, 2026a). The catalog of validation events used in this study is provided as a CSV file in the Supporting Information S1 and archived as a data set on Zenodo (Wang, 2026). Each row corresponds to a specific TP ID (TPID), with associated timestamps marking the storm interval ($t0$–$t4$), satellite instrument (e.g., CHAMP, GRACE), and storm type (single–peak or multiple–peak). For each TPID, the CSV file also includes direct RoR URLs linking to the simulation outputs used in this study. An interactive web-based validation system, the CCMC Ionosphere–Thermosphere Model Assessment Platform (ITMAP, https://ccmc.gsfc.nasa.gov/itmap/), provides validation results as a thermospheric model scorecard interactively (NASA Community Coordinated Modeling Center, 2026b).

Acknowledgments

This research is supported by NSF Awards AGS-2140031, AGS-2409172, and NASA Goddard Space Flight Center through the Cooperative Agreement 80NSSC21M0180 to Catholic University, Partnership for Heliophysics and Space Environment Research (PHaSER). SB and SL were supported by CNES APR. JCW acknowledges Adam Kubaryk from NOAA Space Weather Prediction Center for developing and sharing the WAM-IPE helium correction code and extrapolation module used in this study. Resources supporting this work were providing by the NASA Advanced Supercomputing (NAS) Division at Ames Research Center. The scientific color maps (https://doi.org/10.5281/zenodo.1243862; Crameri et al., 2020; Crameri, 2023) are used in figures herein to prevent visual distortion of the data and exclusion of readers with color-vision deficiencies. The authors thank the editor and the reviewers for their careful reading of the manuscript; their comments considerably improved its quality.

References

Akmaev, R. A., Fuller-Rowell, T. J., Wu, F., Forbes, J. M., Zhang, X., Anghel, A. F., et al. (2008). Tidal variability in the lower thermosphere: Comparison of whole atmosphere model (WAM) simulations with observations from TIMED. *Geophysical Research Letters*, *35*(3), 2007GL032584. https://doi.org/10.1029/2007GL032584

Bag, T., Ogawa, Y., & Sivakumar, V. (2024). Thermospheric NO cooling during 2003 October "halloween storm": Revisited. *Journal of Geophysical Research: Space Physics*, *129*(7), e2024JA032805. https://doi.org/10.1029/2024JA032805

Bowman, B., Tobiska, W. K., 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 and exhibit*. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2008-6438

Bruinsma, S. (2015). The DTM-2013 thermosphere model. *Journal of Space Weather and Space Climate*, *5*, A1. https://doi.org/10.1051/swsc/2015001

Bruinsma, S., & Boniface, C. (2021). The operational and research DTM-2020 thermosphere models. *Journal of Space Weather and Space Climate*, *11*, 47. https://doi.org/10.1051/swsc/2021032

Bruinsma, S., Doornbos, E., & Bowman, B. (2014). Validation of GOCE densities and evaluation of thermosphere models. *Advances in Space Research*, *54*(4), 576–585. https://doi.org/10.1016/j.asr.2014.04.008

Bruinsma, S., & Laurens, S. (2024). Thermosphere model assessment for geomagnetic storms from 2001-2023. *Journal of Space Weather and Space Climate*, *14*, 28. https://doi.org/10.1051/swsc/2024027

Bruinsma, S., Siemes, C., Emmert, J. T., & Mlynczak, M. G. (2023). Description and comparison of 21st century thermosphere data. *Advances in Space Research*, *72*(12), 5476–5489. https://doi.org/10.1016/j.asr.2022.09.038

Bruinsma, S., Tamagnan, D., & Biancale, R. (2004). Atmospheric densities derived from CHAMP/STAR accelerometer observations. *Planetary and Space Science*, *52*(4), 297–312. https://doi.org/10.1016/j.pss.2003.11.004

Burns, A., Solomon, S., Qian, L., Wang, W., Emery, B., Wiltberger, M., & Weimer, D. (2012). The effects of corotating interaction Region/high speed stream storms on the thermosphere and ionosphere during the last solar minimum. *Journal of Atmospheric and Solar-Terrestrial Physics*, *83*, 79–87. https://doi.org/10.1016/j.jastp.2012.02.006

Christophe, B., Boulanger, D., Foulon, B., Huynh, P.-A., Lebat, V., Liorzou, F., & Perrot, E. (2015). A new generation of ultra-sensitive electrostatic accelerometers for GRACE Follow-on and towards the next generation gravity missions. *Acta Astronautica*, *117*, 1–7. https://doi.org/10.1016/j.actaastro.2015.06.021

Codrescu, M. V., Negrea, C., Fedrizzi, M., Fuller-Rowell, T. J., Dobin, A., Jakowsky, N., et al. (2012). A real-time run of the coupled thermosphere ionosphere plasmasphere electrodynamics (CTIPe) model. *Space Weather*, *10*(2), 2011SW000736. https://doi.org/10.1029/2011SW000736

Crameri, F. (2023). *Scientific colour maps*. Zenodo. https://doi.org/10.5281/zenodo.8035877

Crameri, F., Shephard, G. E., & Heron, P. J. (2020). The misuse of colour in science communication. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-19160-7

Doornbos, E. (2012). *Thermospheric density and wind determination from satellite dynamics*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-25129-0

Emmert, J. T., Drob, D. P., Picone, J. M., Siskind, D. E., Jones, M., Mlynczak, M. G., et al. (2021). NRLMSIS 2.0: A whole-atmosphere empirical model of temperature and neutral species densities. *Earth and Space Science*, *8*(3), e2020EA001321. https://doi.org/10.1029/2020EA001321

Emmert, J. T., Jones, M., Siskind, D. E., Drob, D. P., Picone, J. M., Stevens, M. H., et al. (2022). NRLMSIS 2.1: An empirical model of nitric oxide incorporated into MSIS. *Journal of Geophysical Research: Space Physics*, *127*(10), e2022JA030896. https://doi.org/10.1029/2022JA030896

Fang, T., Kubaryk, A., Goldstein, D., Li, Z., Fuller-Rowell, T., Millward, G., et al. (2022). Space weather environment during the SpaceX starlink satellite loss in February 2022. *Space Weather*, *20*(11), e2022SW003193. https://doi.org/10.1029/2022SW003193

Fedrizzi, M., Fuller-Rowell, T. J., & Codrescu, M. V. (2012). Global joule heating index derived from thermospheric density physics-based modeling and observations. *Space Weather*, *10*(3), 2011SW000724. https://doi.org/10.1029/2011SW000724

Floberghagen, R., Fehringer, M., Lamarre, D., Muzi, D., Frommknecht, B., Steiger, C., et al. (2011). Mission design, operation and exploitation of the gravity field and steady-state ocean circulation explorer mission. *Journal of Geodesy*, *85*(11), 749–758. https://doi.org/10.1007/s00190-011-0498-3

Friis-Christensen, E., Lühr, H., Knudsen, D., & Haagmans, R. (2008). Swarm – An Earth observation mission investigating geospace. *Advances in Space Research*, *41*(1), 210–216. https://doi.org/10.1016/j.asr.2006.10.008

Fuller-Rowell, T. J., Akmaev, R. A., Wu, F., Anghel, A., Maruyama, N., Anderson, D. N., et al. (2008). Impact of terrestrial weather on the upper atmosphere. *Geophysical Research Letters*, *35*(9), 2007GL032911. https://doi.org/10.1029/2007GL032911

Fuller-Rowell, T. J., Codrescu, M. V., Moffett, R. J., & Quegan, S. (1994). Response of the thermosphere and ionosphere to geomagnetic storms. *Journal of Geophysical Research*, *99*(A3), 3893–3914. https://doi.org/10.1029/93JA02015

Fuller-Rowell, T. J., & Evans, D. S. (1987). Height-integrated pedersen and hall conductivity patterns inferred from the TIROS-NOAA satellite data. *Journal of Geophysical Research*, *92*(A7), 7606–7618. https://doi.org/10.1029/JA092iA07p07606

Hagan, M. E., & Forbes, J. M. (2002). Migrating and nonmigrating diurnal tides in the middle and upper atmosphere excited by tropospheric latent heat release. *Journal of Geophysical Research*, *107*(24), 1–15. https://doi.org/10.1029/2001JD001236

Hagan, M. E., & Forbes, J. M. (2003). Migrating and nonmigrating semidiurnal tides in the upper atmosphere excited by tropospheric latent heat release. *Journal of Geophysical Research*, *108*(A2), 4754. https://doi.org/10.1029/2002JA009466

Heelis, R. A., Lowell, J. K., & Spiro, R. W. (1982). A model of the high-latitude ionospheric convection pattern. *Journal of Geophysical Research*, *87*(A8), 6339–6345. https://doi.org/10.1029/JA087iA08p06339

Hsu, C., & Pedatella, N. M. (2025). Quantifying the impact of solar irradiance uncertainty on thermosphere-ionosphere variability using ensemble forecasts. *Space Weather*, *23*(12), e2025SW004612. https://doi.org/10.1029/2025SW004612

Hsu, V. W., Thayer, J. P., Wang, W., & Burns, A. (2016). New insights into the complex interplay between drag forces and its thermospheric consequences. *Journal of Geophysical Research: Space Physics*, *121*(10), 10417–10430. https://doi.org/10.1002/2016JA023058

Hurrell, J. W., Holland, M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., et al. (2013). The community Earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, 130204122247009. https://doi.org/10.1175/BAMS-D-12-00121

Jones, M., Forbes, J. M., & Hagan, M. E. (2014). Tidal-induced net transport effects on the oxygen distribution in the thermosphere. *Geophysical Research Letters*, *41*(14), 5272–5279. https://doi.org/10.1002/2014GL060698

Kenyon, S., Pacino, M. C., & Marti, U. (Eds.) (2012)., *Geodesy for planet Earth: Proceedings of the 2009 IAG symposium, Buenos Aires, Argentina, 31 August 31–4 September 2009* (Vol. 136). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20338-1

Knipp, D., Kilcommons, L., Hunt, L., Mlynczak, M., Pilipenko, V., Bowman, B., et al. (2013). Thermospheric damping response to sheath-enhanced geospace storms. *Geophysical Research Letters*, *40*(7), 1263–1267. https://doi.org/10.1002/grl.50197

Kornfeld, R. P., Arnold, B. W., Gross, M. A., Dahya, N. T., Klipstein, W. M., Gath, P. F., & Bettadpur, S. (2019). GRACE-FO: The gravity recovery and climate experiment Follow-On mission. *Journal of Spacecraft and Rockets*, *56*(3), 931–951. https://doi.org/10.2514/1.A34326

Lean, J. L., Woods, T. N., Eparvier, F. G., Meier, R. R., Strickland, D. J., Correira, J. T., & Evans, J. S. (2011). Solar extreme ultraviolet irradiance: Present, past, and future: Solar EUV irradiance. *Journal of Geophysical Research*, *116*(A1), A01102. https://doi.org/10.1029/2010JA015901

Liu, H. (2016). Thermospheric inter-annual variability and its potential connection to ENSO and stratospheric QBO 2. Aeronomy. *Earth Planets and Space*, *68*(1), 77. https://doi.org/10.1186/s40623-016-0455-8

Liu, H.-L., Bardeen, C. G., Foster, B. T., Lauritzen, P., Liu, J., Lu, G., et al. (2018a). Development and validation of the whole atmosphere community climate model with thermosphere and ionosphere extension (WACCM-X 2.0). *Journal of Advances in Modeling Earth Systems*, *10*(2), 381–402. https://doi.org/10.1002/2017MS001232

Liu, J., Liu, H., Wang, W., Burns, A. G., Wu, Q., Gan, Q., et al. (2018b). First results from the ionospheric extension of WACCM-X during the deep solar minimum year of 2008. *Journal of Geophysical Research: Space Physics*, *123*(2), 1534–1553. https://doi.org/10.1002/2017JA025010

Lu, G., Richmond, A. D., Lühr, H., & Paxton, L. (2016). High-latitude energy input and its impact on the thermosphere. *Journal of Geophysical Research: Space Physics*, *121*(7), 7108–7124. https://doi.org/10.1002/2015JA022294

Marsh, D. R., Solomon, S. C., & Reynolds, A. E. (2004). Empirical model of nitric oxide in the lower thermosphere. *Journal of Geophysical Research*, *109*(A7), 2003JA010199. https://doi.org/10.1029/2003JA010199

Maruyama, N., Sun, Y., Richards, P. G., Middlecoff, J., Fang, T., Fuller-Rowell, T. J., et al. (2016). A new source of the midlatitude ionospheric peak density structure revealed by a new ionosphere-plasmasphere model. *Geophysical Research Letters*, *43*(6), 2429–2435. https://doi.org/10.1002/2015GL067312

Matzka, J., Stolle, C., Yamazaki, Y., Bronkalla, O., & Morschhauser, A. (2021). The geomagnetic Kp index and derived indices of geomagnetic activity. *Space Weather*, *19*(5), e2020SW002641. https://doi.org/10.1029/2020SW002641

Mehta, P. M., Paul, S. N., Crisp, N. H., Sheridan, P. L., Siemes, C., March, G., & Bruinsma, S. (2023). Satellite drag coefficient modeling for thermosphere science and mission operations. *Advances in Space Research*, *72*(12), 5443–5459. https://doi.org/10.1016/j.asr.2022.05.064

NASA Community Coordinated Modeling Center. (2026a). CCMC metadata registry for runs-on-request simulations [dataset]. *Retrieved from*. Retrieved from https://kauai.ccmc.gsfc.nasa.gov/CMR/TimeInterval/viewAllTI

NASA Community Coordinated Modeling Center. (2026b). Ionosphere–thermosphere model assessment platform (ITMAP) web application [software]. Retrieved from https://ccmc.gsfc.nasa.gov/itmap/

Oberheide, J., Forbes, J. M., Häusler, K., Wu, Q., & Bruinsma, S. L. (2009). Tropospheric tides from 80 to 400 km: Propagation, interannual variability, and solar cycle effects. *Journal of Geophysical Research*, *114*(23), 1–18. https://doi.org/10.1029/2009JD012388

Oliveira, D. M., & Zesta, E. (2019). Satellite orbital drag during magnetic storms. *Space Weather*, *17*(11), 1510–1533. https://doi.org/10.1029/2019SW002287

Olsen, N., Friis-Christensen, E., Floberghagen, R., Alken, P., Beggan, C. D., Chulliat, A., et al. (2013). The swarm satellite constellation application and research facility (SCARF) and swarm data products. *Earth Planets and Space*, *65*(11), 1189–1200. https://doi.org/10.5047/eps.2013.07.001

Picone, J. M., Hedin, A. E., Drob, D. P., & Aikin, A. C. (2002). NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. *Journal of Geophysical Research*, *107*(A12), 1–16. https://doi.org/10.1029/2002JA009430

Qian, L., Burns, A. G., Emery, B. A., Foster, B., Lu, G., Maute, A., et al. (2014). The NCAR TIE-GCM. *Modeling the Ionosphere–Thermosphere System*, 73–83. https://doi.org/10.1002/9781118704417.ch7

Rees, D., Fuller-Rowell, T. J., Gordon, R., Killeen, T. L., Hays, P. B., Wharton, L., & Spencer, N. W. (1983). A comparison of wind observations of the upper thermosphere from the dynamics explorer satellite with the predictions of a global time-dependent model. *Planetary and Space Science*, *31*(11), 1299–1314. https://doi.org/10.1016/0032-0633(83)90067-3

Reigber, C., Lühr, H., & Schwintzer, P. (2002). CHAMP mission status. *Advances in Space Research*, *30*(2), 129–134. https://doi.org/10.1016/S0273-1177(02)00276-4

Richards, P. G., Fennelly, J. A., & Torr, D. G. (1994). EUVAC: A solar EUV flux model for aeronomic calculations. *Journal of Geophysical Research*, *99*(A5), 8981–8992. https://doi.org/10.1029/94JA00518

Richmond, A. D., Ridley, E. C., & Roble, R. G. (1992). A thermosphere/ionosphere general circulation model with coupled electrodynamics. *Geophysical Research Letters*, *19*(6), 601–604. https://doi.org/10.1029/92GL00401

Ridley, A., Deng, Y., & Tóth, G. (2006). The global ionosphere–thermosphere model. *Journal of Atmospheric and Solar-Terrestrial Physics*, *68*(8), 839–864. https://doi.org/10.1016/j.jastp.2006.01.008

Roble, R. (1986). Results of various studies made with the NCAR thermospheric general circulation model (TGCM). In *Nasa* (Vol. 2). Goddard space flight center thermosphere dynamics workshop.

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

Roble, R. G., Ridley, E. C., & Dickinson, R. E. (1987). On the global mean structure of the thermosphere. *Journal of Geophysical Research*, *92*(A8), 8745–8758. https://doi.org/10.1029/JA092iA08p08745

Roble, R. G., Ridley, E. C., Richmond, A. D., & Dickinson, R. E. (1988). A coupled thermosphere/ionosphere general circulation model. *Geophysical Research Letters*, *15*(12), 1325–1328. https://doi.org/10.1029/GL015i012p01325

Siemes, C. (2024). Thermospheric neutral density dataset, version 2 [dataset]. *TU Delft*. Retrieved from https://thermosphere.tudelft.nl/data/data/

Siemes, C., Bruinsma, S., Fernandez-Gomez, I., Hładczuk, N., Van den IJssel, J., Kodikara, T., et al. (2023). New thermosphere neutral mass density and crosswind datasets from CHAMP, GRACE, and GRACE-FO. *Journal of Space Weather and Space Climate*, *13*, 16. https://doi.org/10.1051/swsc/2023014

Siemes, C., De Teixeira Da Encarnação, J., Doornbos, E., Van den IJssel, J., Kraus, J., Pereštý, R., et al. (2016). Swarm accelerometer data processing from raw accelerations to thermospheric neutral densities. *Earth Planets and Space*, *68*(1), 92. https://doi.org/10.1186/s40623-016-0474-5

Sutton, E. K. (2018). A new method of physics-based data assimilation for the quiet and disturbed thermosphere. *Space Weather*, *16*(6), 736–753. https://doi.org/10.1002/2017SW001785

Sutton, E. K., Forbes, J. M., & Nerem, R. S. (2005). Global thermospheric neutral density and wind response to the severe 2003 geomagnetic storms from CHAMP accelerometer data. *Journal of Geophysical Research*, *110*(A9), 1–10. https://doi.org/10.1029/2004JA010985

Sutton, E. K., Nerem, R. S., & Forbes, J. M. (2007). Density and winds in the thermosphere deduced from accelerometer data. *Journal of Spacecraft and Rockets*, *44*(6), 1210–1219. https://doi.org/10.2514/1.28641

Sutton, E. K., Thayer, J. P., Wang, W., Solomon, S. C., Liu, X., & Foster, B. T. (2015). A self-consistent model of helium in the thermosphere. *Journal of Geophysical Research: Space Physics*, *120*(8), 6884–6900. https://doi.org/10.1002/2015JA021223

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters*, *31*(9), 2004GL019920. https://doi.org/10.1029/2004GL019920

Vallado, D. A., & Finkleman, D. (2014). A critical assessment of satellite drag and atmospheric density modeling. *Acta Astronautica*, *95*, 141–165. https://doi.org/10.1016/j.actaastro.2013.10.005

Van den IJssel, J., Doornbos, E., Iorfida, E., March, G., Siemes, C., & Montenbruck, O. (2020). Thermosphere densities derived from swarm GPS observations. *Advances in Space Research*, *65*(7), 1758–1771. https://doi.org/10.1016/j.asr.2020.01.004

Waldron, Z. C., Garcia-Sage, K., Thayer, J. P., Sutton, E. K., Ray, V., Rowlands, D. D., et al. (2024). Assessing thermospheric neutral density models using geodyn's precision orbit determination. *Space Weather*, *22*(2), e2023SW003603. https://doi.org/10.1029/2023SW003603

Wang, J. C. (2026). Catalog of thermosphere–ionosphere validation events used for thermospheric model assessment in the CCMC framework [dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.19055026

Weimer, D. R. (2005). Improved ionospheric electrodynamic models and application to calculating joule heating rates. *Journal of Geophysical Research*, *110*(A5), A05306. https://doi.org/10.1029/2004JA010884

Wu, H., Wang, W., Pham, K. H., Lin, D., Rao, N., Wiltberger, M. J., et al. (2025). The NCAR-TIEGCM version 3.0. *Journal of Geophysical Research: Space Physics*, *130*(9), e2025JA034219. https://doi.org/10.1029/2025JA034219

Yamazaki, Y., & Richmond, A. D. (2013). A theory of ionospheric response to upward-propagating tides: Electrodynamic effects and tidal mixing effects. *Journal of Geophysical Research: Space Physics*, *118*(9), 5891–5905. https://doi.org/10.1002/jgra.50487