Space Weather®

Key Points:

- A global exospheric temperature model with uncertainty estimation capabilities is developed for NRLMSIS 2.0—called MSIS-UQ
- MSIS-UQ provides an avenue to obtain calibrated uncertainty estimates from NRLMSIS 2.0 density predictions
- MSIS-UQ allows for NRLMSIS 2.0 to capture effects of post-storm thermospheric overcooling

MSIS-UQ: Calibrated and Enhanced NRLMSIS 2.0 Model With Uncertainty Quantification

Richard J. Licata[1] [ID]**, Piyush M. Mehta[1]** [ID]**, Daniel R. Weimer[2]** [ID]**, W. Kent Tobiska[3]** [ID]**, and Jean Yoshii[3]**

[1]Department of Mechanical and Aerospace Enginering, West Virginia University, Morgantown, WV, USA, [2]Center for Space Science and Engineering Research, Virginia Tech, Blacksburg, VA, USA, [3]Space Environment Technologies, Pacific Palisades, CA, USA

Abstract The Mass Spectrometer and Incoherent Scatter radar (MSIS) model family has been developed and improved since the early 1970's. The most recent version of MSIS is the Naval Research Laboratory (NRL) MSIS 2.0 empirical atmospheric model. NRLMSIS 2.0 provides species density, mass density, and temperature estimates as function of location and space weather conditions. MSIS models have long been a popular choice of thermosphere model in the research and operations community alike, but—like many models—does not provide uncertainty estimates. In this work, we develop an exospheric temperature model based in machine learning that can be used with NRLMSIS 2.0 to calibrate it relative to high-fidelity satellite density estimates directly through the exospheric temperature parameter. Instead of providing point estimates, our model (called MSIS-UQ) outputs a distribution which is assessed using a metric called the calibration error score. We show that MSIS-UQ debiases NRLMSIS 2.0 resulting in reduced differences between model and satellite density of 25% and is 11% closer to satellite density than the Space Force's High Accuracy Satellite Drag Model. We also show the model's uncertainty estimation capabilities by generating altitude profiles for species density, mass density, and temperature. This explicitly demonstrates how exospheric temperature probabilities affect density and temperature profiles within NRLMSIS 2.0. Another study displays improved post-storm overcooling capabilities relative to NRLMSIS 2.0 alone, enhancing the phenomena that it can capture.

Plain Language Summary Uncertainty quantification (UQ) in modeling can be thought of as providing a range of values that can contain a solution as opposed to only providing a point estimate. UQ is currently not common-place when it comes to atmospheric modeling even while we use imperfect models. One of the most commonly used atmospheric model families is the Mass Spectrometer and Incoherent Scatter radar (MSIS) series. The current iteration in the series is the Naval Research Laboratory (NRL) MSIS 2.0 empirical atmosphere model. NRLMSIS 2.0 provides estimates of multiple atmospheric constituents and temperature which have a wide array of applications. Using data derived from satellite sensors, we develop a model called MSIS-UQ that predicts a parameter (exospheric temperature) which can be input to NRLMSIS 2.0 to provide calibrated estimates. Furthermore, MSIS-UQ predicts a distribution of temperatures as opposed to a point estimate. We show how the MSIS-UQ temperature predictions improve NRLMSIS 2.0 estimates of density by 25% and explore the usefulness of the uncertainty estimation capabilities of the model. We also perform a study where the magnetic model drivers are varied to investigate evidence of atmospheric overcooling after strong storms.

1. Introduction

The thermosphere is the neutral portion of the upper atmosphere where many satellites, debris, and space assets reside. It is primarily heated and influenced by solar extreme ultraviolet (EUV) and far ultraviolet emissions which vary with the solar cycle (Emmert, 2015). Geomagnetic storms—often caused by coronal mass ejections (CMEs) and coronal holes—can cause sudden increases in mass density with little warning. In early 2022, SpaceX had 38 satellites fail to reach their desired orbits due to a density response to a CME which resulted in a minor geomagnetic storm (Hapgood et al., 2022). While the occurrence could have been avoided with adequate neutral density forecasts, current models carry high errors and uncertainty during geomagnetic events (Oliveira et al., 2021).

Although geomagnetic storms receive considerable attention due to their immediate consequences, density model performance also varies as a function of solar EUV output. Bowman et al. (2008) compared density ratios

and error standard deviations of thermosphere models with respect to the High Accuracy Satellite Drag Model (HASDM) which is the operational model used by the United States Space Force (Storz et al., 2005). Their results showed considerable variability as a function of solar activity with strong underestimation (density ratio of 0.7–0.8) during solar minimum. Furthermore, the error standard deviation of these models can be up to 45% during solar minimum and no less than 10% during solar maximum. This points to the need for innovative solutions to provide overall improvements to our thermospheric mass density models.

One such solution is to develop a correction-based model using satellite density estimates (Choury et al., 2013; Licata, Mehta, Weimer, & Tobiska, 2021; Ruan et al., 2018; Weimer et al., 2016, 2020). Over the last few decades, there have been a growing number of satellites with high-fidelity onboard accelerometers such as CHAllenging Minisatellite Payload (CHAMP) (Lühr et al., 2002), Gravity Recovery and Climate Experiment (GRACE) (Tapley et al., 2004), and Swarm (Friis-Christensen et al., 2006). Some researchers have taken the accelerometer data from the satellites, removed other acceleration sources (e.g., solar radiation pressure), and isolated drag acceleration therefore estimating density (Bruinsma & Biancale, 2003; Liu et al., 2005; Sutton, 2008; Doornbos, 2012; Calabia & Jin, 2016; Mehta et al., 2017; van den IJssel et al., 2020). These density sources have a high temporal cadence and, when combined, provide coverage over a wide array of locations and space weather conditions.

Although these correction-based approaches help calibrate the original model, they still only provide point estimates. Point estimates do not provide the user with information on the potential error associated with the output, or density. Uncertainty quantification (UQ), with respect to thermosphere modeling, is concerned with providing information on the uncertainty associated with estimates of temperature or density (Boniface, Claude and Bruinsma, Sean, 2021; Licata, Mehta, Tobiska, & Huzurbazar, 2022; Licata & Mehta, 2022). This information—often provided as a Gaussian distribution—can be used for analysis or potentially in determining satellite position uncertainty (Bussy-Virat et al., 2018; Licata, Mehta, & Tobiska, 2021).

In this work, we develop an exospheric temperature model based on satellite estimates using machine learning (ML) to feed into NRLMSIS 2.0 (Emmert et al., 2021). This model (called MSIS-UQ) provides a distribution in its exospheric temperature predictions therefore incorporating uncertainty quantification (UQ) capabilities to NRLMSIS 2.0. MSIS-UQ differs from EXospheric TEMperatures on a PoLyhedrAl gRid—ML (EXTEMPLAR-ML) by: (a) using true locations (no grid) for training, (b) using the newest Mass Spectrometer and Incoherent Scatter radar (MSIS) model, and (c) providing uncertainty estimates. The manuscript is organized as follows. We describe the data, model development, and validation approaches. We then show results for overall model performance and a demonstration of its uncertainty capabilities. Last, we evaluate MSIS-UQ during a storm and perform a study to test its response to geomagnetic activity.

2. Data and Methods

2.1. Exospheric Temperature Estimates

Exospheric temperatures were obtained through a binary search, deriving the exospheric temperature required in NRLMSIS 2.0 to match satellite density estimates (Weimer et al., 2016, 2021). The density estimates originate from the following sources—CHAMP 2001: Doornbos (2012), CHAMP 2002–2010: Mehta et al. (2017), GRACE 2002–2009: Mehta et al. (2017), GRACE 2010: Sutton (2008), Swarm A 2013–2018: van den IJssel et al. (2020), and Swarm B 2012–2016: van den IJssel et al. (2020). Note that only GRACE-A measurements are used due to the similarity of the GRACE-A and GRACE-B orbits. The CHAMP and GRACE density estimates originate from accelerometer measurements and span an altitude range of 300–535 km while the Swarm A and B density estimates are obtained from GPS data and span an altitude range of 437–545 km. The cadence of the satellite estimates are 10, 5, 30, and 30 s for CHAMP, GRACE, Swarm A, and Swarm B, respectively. There are over 82 million samples total for model development and evaluation. This data set was originally derived for the EXTEMPLAR exospheric temperature correction model for NRLMSISE-00 and NRLMSIS 2.0 which consisted of linear models for 1,620 polyhedral grid cells (Weimer et al., 2020, 2021). The interfacing between a separate exospheric temperature model and MSIS was crucial for the development of MSIS-UQ.

2.2. Model Drivers

To run NRLMSIS 2.0, the following space weather inputs are required: $F_{10.7}$, F_{81c}, and Ap. $F_{10.7}$ is a measure of the 10.7 cm solar radio flux and acts as a good proxy for solar EUV emissions while F_{81c} is merely an 81-day centered average of the proxy. The Ap index is a measure of daily global geomagnetic activity. NRLMSIS 2.0 also has a storm-time flag that allows additional 3-hourly ap values: ap, ap_3, ap_6, ap_9, ap_{12-33}, and ap_{36-57}. The subscripts indicate the number of hours prior to epoch that ap value represents; when there are two numbers (e.g., 36–57), the 3-hourly index values are averaged over this many hours prior to epoch. NRLMSIS also has temporal inputs (e.g., time of day, day of year) and spatial inputs (e.g., local time, longitude, latitude (LAT), and altitude).

To develop a machine-learned exospheric temperature model with a high temporal cadence, we expand the input set. To account for solar activity, the model receives $F_{10.7}$, $S_{10.7}$, $M_{10.7}$, and $Y_{10.7}$, all accounting for different forms of solar emissions that affect different regions of the thermosphere (Bowman et al., 2008; Tobiska et al., 2008). The ML model also uses inputs from EXTEMPLAR, particularly S_N, S_S, ΔT. The two S inputs are Poynting flux totals in the Northern and Southern hemispheres (Weimer, 2005a, Weimer, 2005b). ΔT is a parameter derived by Weimer et al. (2020, 2021) and represents exospheric temperature perturbations; it is a function of Poynting flux and simulated nitric oxide emissions.

For further representation of geomagnetic activity, a time history of SYM-H is used, similar to the use of storm-time ap indices for NRLMSIS 2.0. *SYM-H* represents disturbances to the geomagnetic field and has similar characteristics to Dst. A benefit over Dst is its 1-min cadence (Iyemori, 1990). The SYM-H inputs are as follows: SYM-H, SYM-H$_{0-3}$, SYM-H$_{3-6}$, SYM-H$_{6-9}$, SYM-H$_{9-12}$, SYM-H$_{12-33}$, SYM-H$_{33-57}$. Due to the model's use of in-situ satellite estimates, it takes location as an input. The local solar time (LST) is transformed using sine and cosine functions to make it continuous about midnight,

$$LST_1 = \sin\left(\frac{2\pi LST}{24}\right) \quad LST_2 = \cos\left(\frac{2\pi LST}{24}\right) \tag{1}$$

The model also uses satellite LAT. To account for temporal dependencies, day of year (t_1 and t_2) and universal time (UT, t_3 and t_4) are transformed through similar functions,

$$t_1 = \sin\left(\frac{2\pi doy}{365.25}\right) \quad t_2 = \cos\left(\frac{2\pi doy}{365.25}\right) \quad t_3 = \sin\left(\frac{2\pi UT}{24}\right) \quad t_4 = \cos\left(\frac{2\pi UT}{24}\right) \tag{2}$$

2.3. Data Preparation

The ML model drivers are the 21 space weather, spatial, and temporal inputs mentioned in the previous section. Each sample has an exospheric temperature estimate which will serve as the target. To reduce the variance of the output (T_∞), we use a logarithmic transformation making the output $\log_{10} T_\infty$. The final step to create ML-ready data is to perform standard normalization. This will center the data and provide a unit standard deviation for each input and output with respect to the statistics of the training data,

$$\tilde{x} = \frac{x - \mu}{\sigma} \tag{3}$$

where x represents each input and the output, μ is the mean of that quantity over the training set, and σ is the standard deviation of that quantity over the training set. μ and σ must be saved for downstream use of the model.

The 81 million samples are split into training, validation, and tests sets to achieve an 80%–10%–10% distribution. Licata, Mehta, Weimer, and Tobiska (2021) split a similar data set to have long segments—on the order of months or years. However, the resulting model was not well-generalized across the sets. When comparing EXTEMPLAR and a previous version of MSIS over the same three time periods, the error statistics also varied. Therefore, we use a different approach to data splitting. Starting with the first sample, 8 weeks are used for training, the following week is used for validation, and the subsequent week is used for testing. This pattern is repeated until the end of the data set. This approach forces similar solar cycle and spatial coverage across the three sets. Concurrently, there is a significant number of samples within each segment providing temporally disjoint segments throughout the 17 years time-span of the data set. Since each satellite has a different cadence and there are different numbers of satellites providing measurements at a given time, the number of samples in each segment varies. In the

Table 1
Hyperparameter Tuner Parameters (Left) and Search Space (Right)

Tuner option	Choice	Parameter	Values/range
Scheme	Bayesian optimization	Number of hidden layers	1–10
Total trials	150	Neurons	min = 16, max = 1,024, step = 4
Initial points	50	Activations	relu, softplus, tanh, sigmoid, softsign, selu, elu
Repeats per trial	3	Dropout	min = 0.10, max = 0.60, step = 0.01
Minimization parameter	Validation loss	Optimizer	RMSprop, Adam, Adadelta, Nadam

training, validation, and test sets, the number of samples varies from 22,454 to 1,450,380, 12,990 to 181,423, and 12,888 to 181,422, respectively. This means that there are between 25,878 and 362,845 samples separating training segments.

2.4. Model Development

A key consideration in ML model development is the choice for the loss (or objective) function. This will be the quantity that the algorithm will try to minimize or maximize during the training phase. For this work, we use the negative logarithm of predictive density loss function due to previous ML modeling results (Licata & Mehta, 2022; Licata, Mehta, Tobiska, & Huzurbazar, 2022) NLPD is given as

$$NLPD(y, \mu, \sigma) = \frac{(y - \mu)^2}{2\sigma^2} + \frac{\log\left(\sigma^2\right)}{2} + \frac{\log(2\pi)}{2} \tag{4}$$

where y, μ, and σ are the ground truth, mean prediction, and predicted standard deviation, respectively. This loss functions provides the capability for uncertainty prediction since the model can now make σ predictions and have them incorporated in the loss function without labels. The ML model will directly predict μ and σ for $\log_{10} T_\infty$ meaning it has two outputs. The output node for μ uses a linear activation function since the normalized values can take on any value (unbounded). The output node for σ must be positive, monotonically increasing, and have no upper-bound. Therefore, we proceed with the "softplus" activation function: $f(x) = \ln(1 + e^x)$.

With the inputs/output(s) selected, normalization complete, loss function chosen, and output layer determined, the only remaining task for model development is to select an architecture and train a model. To accomplish this, we defined a scheme and hyperparameter space using Keras Tuner (O'Malley et al., 2019) as with previous work (Licata, Mehta, Weimer, & Tobiska, 2021, Licata, Mehta, Tobiska, & Huzurbazar, 2022; Licata & Mehta, 2022). Table 1 shows the parameters used and ranges provided to the tuner. Each hidden layer can have its own unique number of neurons, activation function, and dropout rate. Note: "trials" refers to number of model architectures, initial points is the number of randomly selected architectures prior to Bayesian optimization, and repeats refers to re-initializing weights and retraining.

Since there are over 65 million training samples in the data set, we only provide the tuner with a subset of this data. The tuner uses 1 million randomly selected samples from the training set and 200,000 randomly selected samples from the validation set. Each model trained by the tuner will run for 50 training iterations (or epochs) with a batch size of 4,096. Upon completion, the 10 best models are saved based on the validation loss values. All 10 models are evaluated (see Section 2.5), and the best performing one is used as a base architecture for full training. The best architecture from the tuner is displayed in Table 2.

Table 2
Model Architecture for the Best Model From the Tuner

	Neurons	Activation	Dropout rate
Layer 1	648	Relu	0.13
Layer 2	176	Tanh	0.21
μ Output	1	Linear	0.00
σ Output	1	Softplus	0.00

Note. There are 21 inputs for Layer 1, and it uses the NAdam optimizer https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Nadam.

2.5. Model Analysis

When comparing model to satellite densities, we use the mean absolute error (MAE) metric in percentage (relative) form. This metric is chosen due to its intuitive nature. To assess the quality of the ML uncertainty estimates, we use

a calibration error score (CES) (Licata, Mehta, Tobiska, & Huzurbazar, 2022; Licata & Mehta, 2022). This metric assesses how close the uncertainty estimates are to input prediction intervals (e.g., how close a 95% prediction interval is to containing 95% of true samples). The equations for MAE and CES are provided in Equations 5a and 5b, respectively.

$$\text{MAE} = 100\% \frac{|\rho_{satellite} - \rho_{model}|}{\rho_{satellite}} \tag{5a}$$

$$\text{CES} = \frac{100\%}{m \cdot n_{out}} \sum_{i=1}^{n_{out}} \sum_{j=1}^{m} |p(z_{i,j}) - p(\hat{z}_{i,j})| \tag{5b}$$

In Equation 5a, $\rho_{\text{satellite}}$ is the estimate of density from a given satellite measurement (e.g., from CHAMP, GRACE, etc.). ρ_{model} is the density from the model being evaluated. In Equation 5b, m is the number of prediction intervals being tested, n_{out} is the number of model outputs (for the ML model $n_{\text{out}} = 1$), $p(z_{i,j})$ is a given prediction interval, and $p(\hat{z}_{i,j})$ is the percentage of samples captured by the model's uncertainty estimates with the given prediction interval. In this work, the prediction intervals span from 5% to 99% in increments of 5%.

2.5.1. Comparison With NRLMSIS 2.0 and HASDM

To assess the validity of the model in terms of mean density prediction, its error with respect to the satellite estimates are compared to those of NRLMSIS 2.0 and HASDM. To get NRLMSIS 2.0 errors, the model is evaluated at all locations and times of the satellite measurements. For HASDM, the 3-dimension density grids from the SET HASDM density database are interpolated in log-scale to the satellite locations and times (Tobiska et al., 2021). We then break up the errors into the three sets used for ML model development (training, validation, and test). We do this to simultaneously test the generalization of our model while ensuring differences in performance across the sets is also seen with the other models. In addition to the error assessment, we also compute the CES for MSIS-UQ across the three sets (in terms of density). For information on the conversion from ML predicted exospheric temperature to NRLMSIS 2.0 adjusted density, see Weimer et al. (2020, 2021).

2.5.2. Uncertainty Demonstration

The reliability of the MSIS-UQ uncertainty estimates is demonstrated in Section 3, but we further establish the capabilities in Section 3.1. The ML model directly predicts the uncertainty into the exospheric temperature which is then incorporated into NRLMSIS 2.0. The probabilistic T_∞ values result in probabilistic local temperatures and species densities. We consider a given epoch (13 May 2007 at 21:42.50 UT) where CHAMP and GRACE are at very different locations; CHAMP is near the equator on the night-side while GRACE is at high LAT on the day-side. NRLMSIS 2.0 is provided probabilistic T_∞ values from the MSIS-UQ distribution at each location, and we consider the temperature, species densities, and mass density between 100 and 800 km altitude. The distributions are shown as a function of altitude, and the satellite estimates are provided for reference.

2.5.3. Post-Storm Cooling Capabilities

With overall performance investigated, we want to consider the 2003 Halloween storm along the CHAMP orbit. We interpolate global density grids for TIE-GCM (Qian et al., 2013), JB2008 (Bowman et al., 2008), and HASDM (Storz et al., 2005). In addition, we evaluate NRLMSIS 2.0, EXTEMPLAR, and MSIS-UQ directly at the CHAMP locations. The orbit-averaged densities are computed for all models (and CHAMP) for display and comparison. This can provide insight into storm responses and post-storm density depletion characteristics. In previous work, we investigated density ratios for NRLMSIS 2.0 and three ML models with time-lagged geomagnetic activity to examine post-storm density characteristics expected within the models (Licata, Mehta, Weimer, et al., 2022). While two of the machine-learned models exhibited significant density depletion, NRLMSIS 2.0 never showed evidence of thermospheric overcooling. We perform the same study—outlined below—to test if the exospheric temperature corrections can enforce the behavior present in the satellite datasets.

To start, all non-geomagnetic model drivers are kept to constant values. We set the solar indices to 120 solar flux units, and the time for the study is 00:00 UT during the fall equinox. Each of the time-history geomagnetic drivers will be increased individually while all others are kept at a constant value: $ap = 56$, SYM-H = $-50\ nT$. Since the ML model uses Poynting Flux totals and ΔT at epoch, they are kept constant at 200 GW and 120 K, respectively. This study is also conducted at four discrete locations: night equator (2 hr LST, 0° LAT), day equator (14 hr LST,

0° LAT), night pole (2 hr LST, 80° LAT), and day pole (14 hr LST, 80° LAT). The density ratios are reported with respect to the given geomagnetic driver being set to 0.

However, achieving this requires an additional consideration relative to the work of Licata, Mehta, Weimer, et al. (2022). MSIS-UQ uses SYM-H while NRLMSIS 2.0 uses *ap* for time-series geomagnetic drivers. To account for this distinction, we first fit a line between all SYM-H and *ap* values within our data set. Using this, we find the SYM-H value associated with the *ap* value that must be used to get density from NRLMSIS 2.0. Therefore, density ratios for MSIS-UQ use this simultaneous SYM-H and *ap* variation as opposed to only using *ap* variations with NRLMSIS 2.0 alone.

3. Results

Figure 1 shows the relative error distributions and MAE for NRLMSIS 2.0, HASDM, and MSIS-UQ with respect to density estimates from CHAMP, GRACE, Swarm A, and Swarm B. The calibration curve for MSIS-UQ is also displayed alongside the CES. This is separated by samples in the MSIS-UQ training, validation, and test sets. Similar figures are provided for each individual satellite in SM1. Note: these statistics are relative errors for MSIS-UQ since it was developed on this data set, but they could be interpreted as relative differences for NRLMSIS 2.0 and HASDM.

Panel (a) shows the altitudes for each satellite used in this analysis showing over a 200 km span over 15 years of measurements. The left-most panels (b), (d), and (f) indicate that MSIS-UQ provides much more accurate density predictions than both NRLMSIS 2.0 alone and HASDM. All three models have a tendency to overpredict density although MSIS-UQ has the smallest bias. The MAE values highlight the ~25% error reduction from NRLMSIS 2.0 and the ~11% error reduction from HASDM. Across the three sets, MSIS-UQ is well-generalized with density prediction errors ranging <1.5%. With respect to its uncertainty estimates (panels (c), (e), and (g)), MSIS-UQ has a CES <5% across the three sets. It has a tendency to underestimate in the middle prediction intervals (between 20% and 80%) but is well-calibrated at prediction intervals >90%.

3.1. Uncertainties as a Function of Altitude

Figure 2 contains uncertainty profiles for MSIS-UQ at CHAMP and GRACE locations on 13 May 2007. There are panels for species density, temperature, mass density, relative uncertainty, and satellite position. Please reference the figure caption and Section 2.5.2 for details.

Panels (a) and (b) show the species density profiles at CHAMP and GRACE positions, respectively. The uncertainty bounds provide valuable information on the impact of exospheric temperature uncertainty on the uncertainty of local species. For example, one can investigate the Oxygen (O) to Helium (He) transition for various locations and conditions. Panel (a) shows that at CHAMP's position, this transition is occurring somewhere in the region of 507–552 km (1-σ) while at GRACE's position, the transition may occur between 688 and 738 km (1-σ). Other insights can be gained such as which species are most impacted by exospheric temperature at a given location/altitude. Note: only 1-σ bounds are shown here to prevent artifacts at low-values caused by the semi-logarithmic scale. The scale also causes the bounds to appear to be not-centered about the mean.

Panels (c), (d), and (e) provide information on the local temperature and mass density with uncertainty. In panel (c), MSIS-UQ severely shifts the exospheric temperature prediction and brings it closer to the estimates of CHAMP and GRACE; in both cases NRLMSIS 2.0 overpredicts temperature. The uncertainty in temperature is unobservable below 130 km and grows until it reaches the asymptotic temperature between 250 and 300 km. The uncertainty bounds and mean remain unchanged above these altitudes. Note that the CHAMP and GRACE temperature estimates are for T_∞ but we show them at their current altitude as the temperature has converged. In panel (d), we see different trends in mass density. Again the uncertainty is minimal below approximately 200 km and begins to increase for a few hundred kilometers. The overprediction of temperature in NRLMSIS 2.0 results in higher than observed density by CHAMP and GRACE around 350 and 475 km, respectively. MSIS-UQ provides a more accurate density predictions at the satellite locations. Panel (e) shows the 1-σ uncertainty with respect to mean density. This shows different model behavior between the two locations. At CHAMP's location, the uncertainty increases to 24% around 460 km and decreased until around 700 km where it settles to 9%. At GRACE's location, the uncertainty continues to increase until it reaches 18% at 600 km where it begins to decrease.
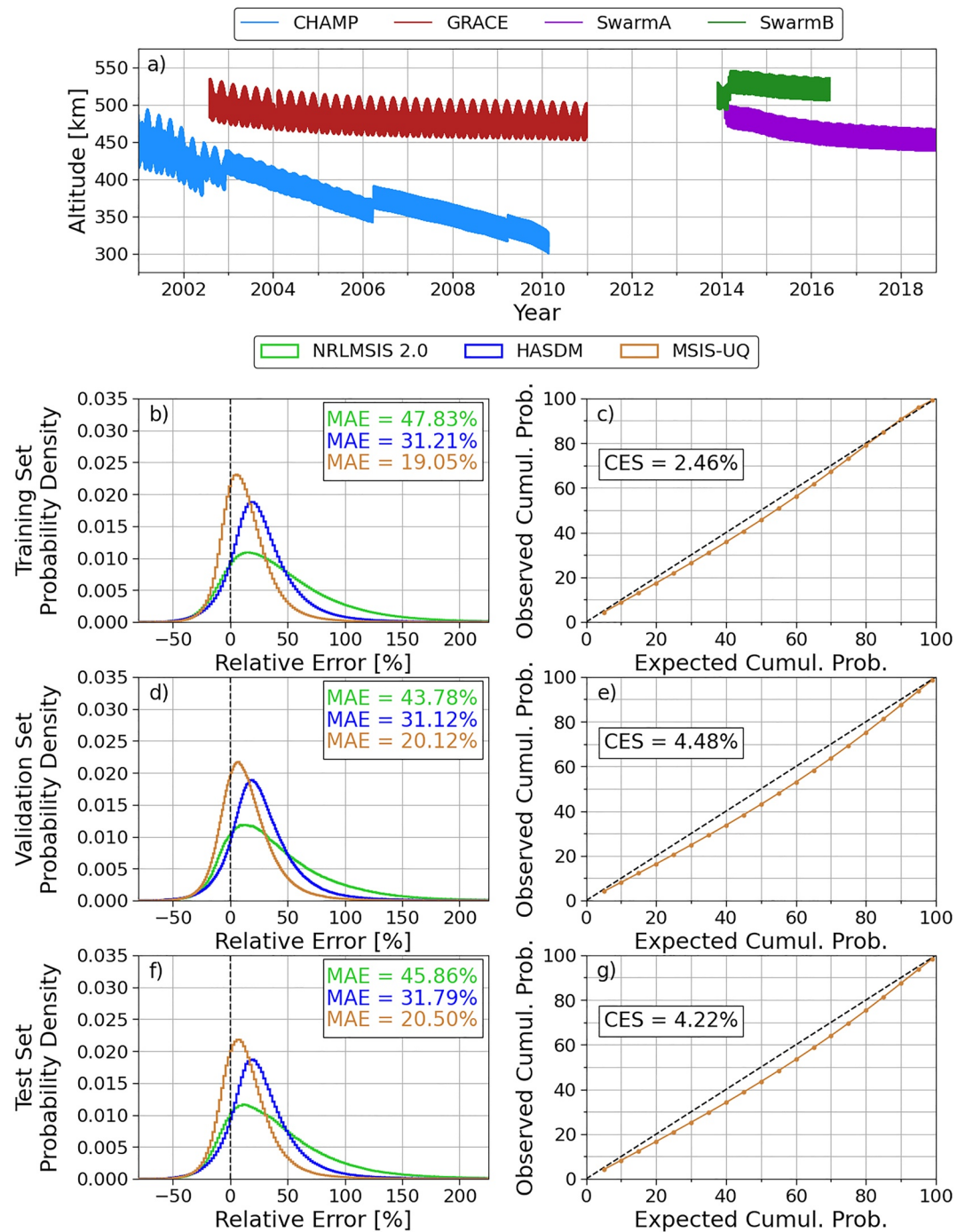
Figure 1. Altitudes of the satellites used for temperature and density estimates (a), relative error histograms (b, d, and f), and MSIS-UQ calibration curves (c, e, and g).

3.2. Enhanced Storm and Post-Storm Modeling

Geomagnetic storms remain a challenge in modeling the thermosphere. Even further, post-storm characteristics vary between models, and understanding post-storm thermospheric overcooling has been a major focus in the community (Knipp et al., 2017; Lei et al., 2012; Weimer et al., 2015). To highlight this, we show orbit-average density along CHAMP's orbit during the 2003 Halloween storm for CHAMP, TIE-GCM, NRLMSIS 2.0, JB2008,
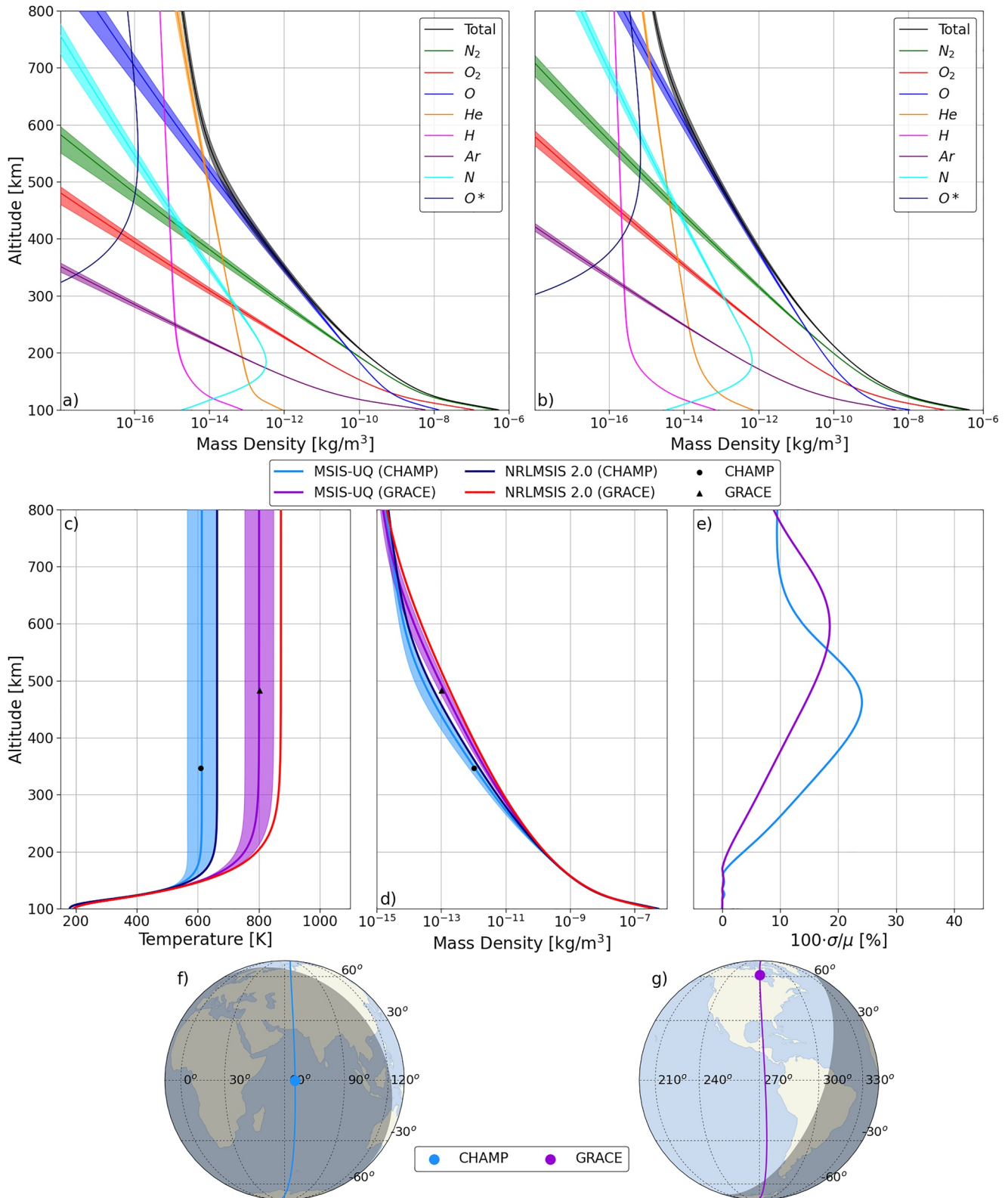
Figure 2. MSIS-UQ species density profiles for CHAMP (a) and GRACE (b) locations with 1-σ bounds, temperature profiles with 2-σ bounds (c), total mass density profiles with 2-σ bounds (d), 1-σ uncertainty normalized by the mean prediction (e), and the paths for CHAMP (f) and GRACE (g) with the current location denoted by the markers. This was conducted for 13 May 2007 at 21:42.50 UT.
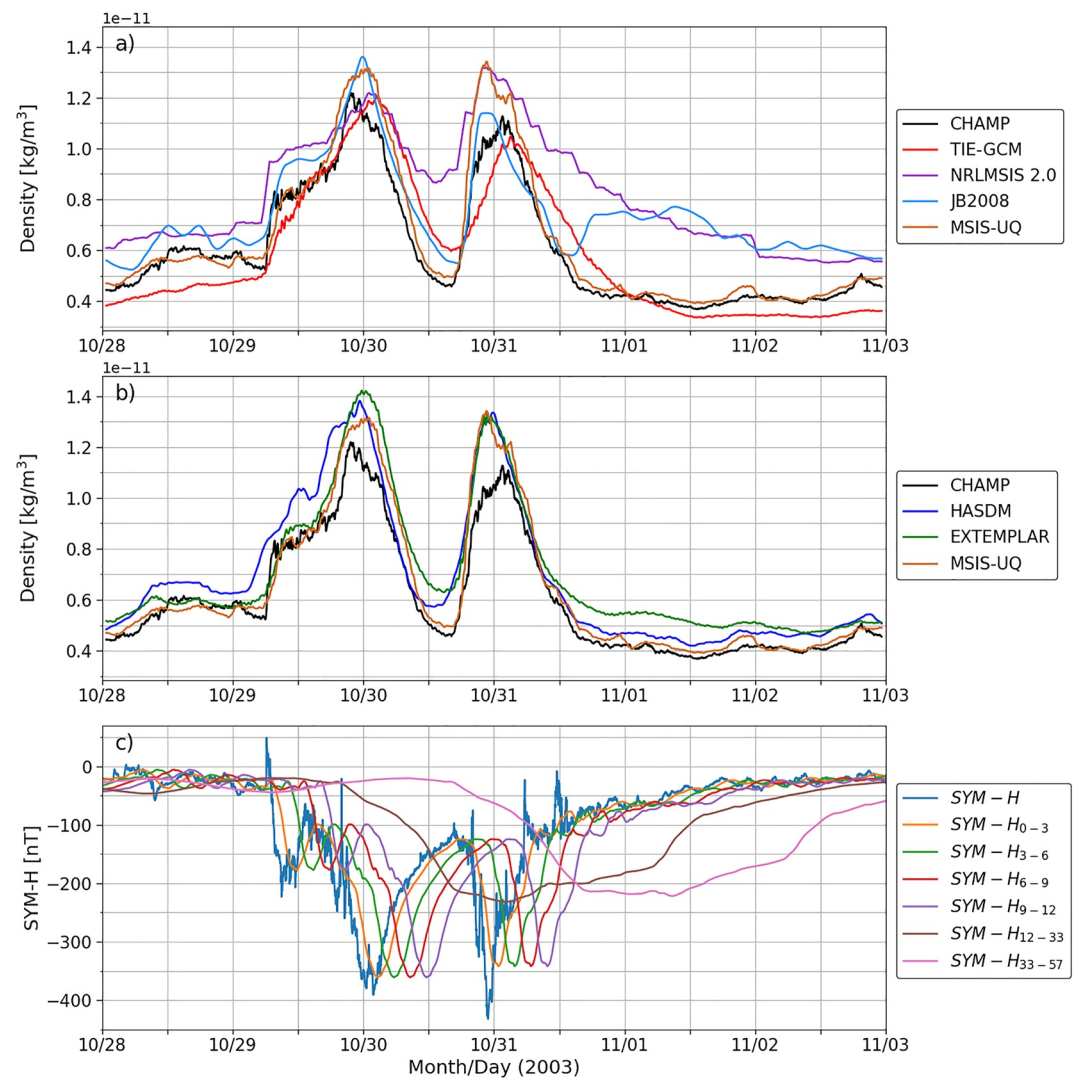
Figure 3. Orbit-average density for TIE-GCM, NRLMSIS 2.0, JB2008, MSIS-UQ, and CHAMP (a), orbit-average density for HASDM, EXTEMPLAR, MSIS-UQ, and CHAMP (b) and the associated SYM-H time-series inputs (c).

HASDM, EXTEMPLAR, and MSIS-UQ in Figure 3. Time history SYM-H inputs for MSIS-UQ are displayed in panel (c) for reference.

In the pre-storm period (28 October–29 October 2003), there is significant variability in model outputs. During the first peak of the storm, the models show similar trends but are mostly above the CHAMP density estimates. In the lull between the two peaks (30 October–31 October), all models show density decreases but to varying extents. NRLMSIS 2.0 shows the smallest density decay during this period, but due to the exospheric temperature corrections in MSIS-UQ, the density falls to CHAMP levels. For the second storm, all models overestimate density with the exception of TIE-GCM and JB2008. In the post-storm period (3 October–3 November), NRLMSIS 2.0 does not follow the trend observed by CHAMP—a sudden and severe decrease in density. JB2008 shows this briefly before overpredicting density to the extent of NRLMSIS 2.0. The MSIS-UQ follows the post-storm overcooling trends observed by CHAMP, showing the impact the exospheric temperature corrections to NRLMSIS 2.0 have. We now conduct the overcooling study from Licata, Mehta, Weimer, et al. (2022) to observe the characteristics in the model without effects from other drivers. These results are shown in Figure 4, and correspond to Figure 2 in Licata, Mehta, Weimer, et al. (2022).

In Figure 4, the curves represent the density ratio of a given model when one of the time-lagged geomagnetic drivers are increased with respect to their value set to zero. In panel (e), for example, the orange curve for ap_3 is
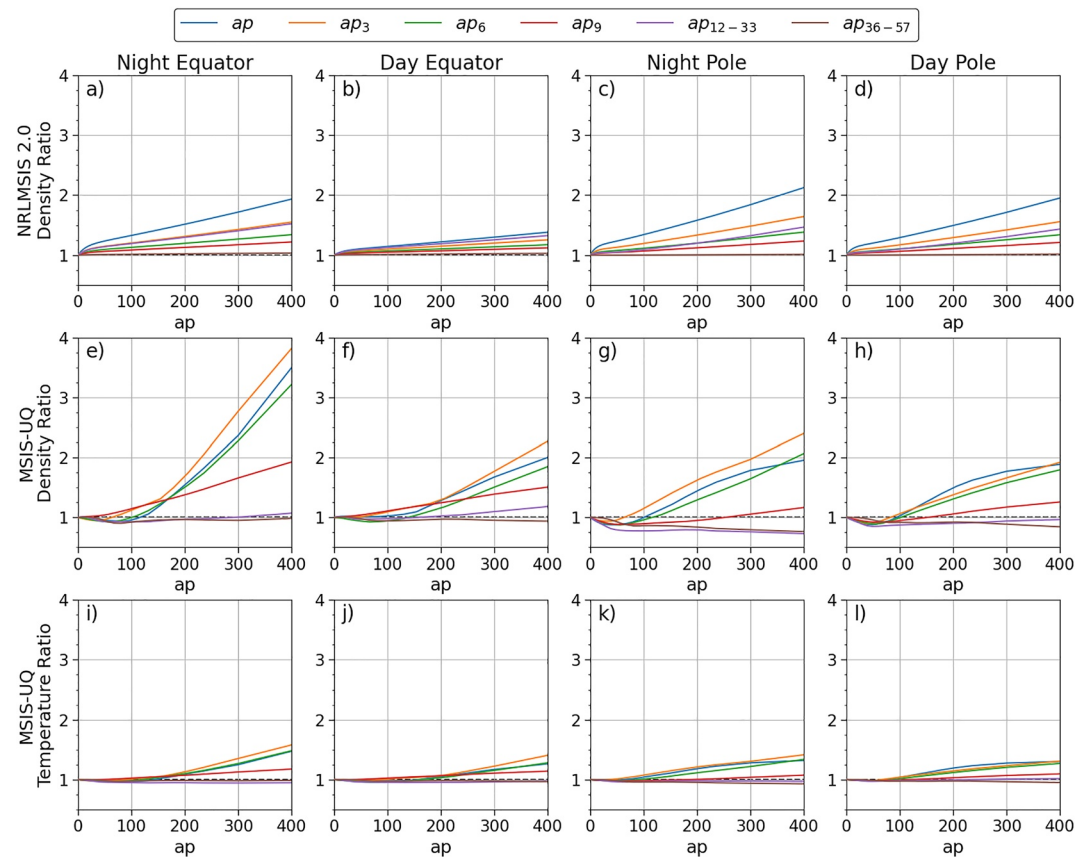
Figure 4. Density ratios for NRLMSIS 2.0 (a–d) and MSIS-UQ (e–h) with the corresponding temperature ratios for MSIS-UQ (i–l).

approximately 3.75 when $ap_3 = 400$. This indicates that the density is 3.75 times the value when $ap_3 = 0$. Panels (a–d) show the density ratios for NRLMSIS 2.0 while increasing each time-series ap value independently. We observe linear relationships between ap and density for all of the time-series inputs. The current 3-hourly ap value has the strongest impact on density and, of these four locations, never generates a density ratio greater than 2.25. These results show that in NRLMSIS 2.0, the relative importance of each ap value becomes less important the further back from epoch it is with ap_{36-57} having minimal impact on density. The exception to this statement is ap_{12-13}. At the day equator (panel (b)), it causes the second-most positive density ratio. When ap_{12-13} is very large, it causes a strong relative increase at all locations considered here. This could explain the overprediction observed in Figure 3. The second row (panels (e–h)) shows the results for MSIS-UQ. The trends shown in these panels contradict many observations when using NRLMSIS 2.0 alone. For example, at 3/4 locations, ap_3 causes the largest density ratios, even being nearly twice as large at the night equator. MSIS-UQ also shows a nonlinear relationship between geomagnetic activity and density which is not as clearly seen in NRLMSIS 2.0 alone.

MSIS-UQ also enforces the idea of negative density ratios—density decreasing while the geomagnetic drivers are increasing. This is most pronounced at the two pole locations. When the least recent geomagnetic drivers (ap_{12-33} and ap_{36-57}) become large, the density becomes up to 25% lower than when they are set to zero. This overcooling was seen in Figure 3 and is observed in the satellite density data, therefore becoming present in MSIS-UQ. Another interesting trend is seen at low levels on geomagnetic activity particularly in panels (g) and (h). When any of the ap values increase from 0 up to 50–100, the density decreases. This seems counter-intuitive but could be caused by the approach of the study. When ap is being considered, for example, the ap and SYM-H values are set to 56 and −50 nT, respectively. When $ap = 0$, this would represent the time immediately after moderate geomagnetic activity while $ap = 56$ would represent sustained moderate geomagnetic activity. The model shows that when the conditions abruptly return to quiet values, the density increases—likely only temporarily. The bottom row (panels (i–l)) show the temperature ratios from MSIS-UQ corresponding to the middle row. The

general trends are the same between temperature and density; however, the difference comes from the magnitude. The relative changes in temperature result in much stronger changes in density. There are negative temperature ratios, but they are much less prominent due to the consistent scaling.

In SM2, we provide a movie of global density evolution for the 2003 Halloween storm at 400 km. This contains density maps for TIE-GCM, NRLMSIS 2.0, EXTEMPLAR, and MSIS-UQ. Prior to the storm, all models have a similar diurnal structure. This is quickly disrupted at the onset of the storm when both TIE-GCM and MSIS-UQ show strong auroral density enhancements. Throughout the storm, these are the only two models that display significant global disruptions to the structure of the thermosphere at this altitude. The thermosphere expands and contracts both longitudinally and latitudinally for these models. EXTEMPLAR—a linear model based on the same data set as MSIS-UQ—models an abnormal thermosphere, but the movement is more longitudinal or latitudinal with some wave-like oscillations about the equator. During this storm, NRLMSIS 2.0 exhibits general increases in density, but the structure is well-preserved. There are some slow-moving variations although it does not resemble the dynamics exhibited by the physics-based model.

4. Summary and Conclusions

In this work, we developed an exospheric temperature model for NRLMSIS 2.0 with UQ. When these exospheric temperatures are input to NRLMSIS 2.0, the errors with respect to satellite density estimates are reduced from approximately 45%–20% (Figure 1). The relative error distributions for MSIS-UQ have lower variance and bias when compared to both NRLMSIS 2.0 and HASDM. The MSIS-UQ uncertainty estimates proved to be well-calibrated to the satellite density data with a tendency to underestimate the uncertainty bounds.

The uncertainty estimates were closely examined for a given time where CHAMP and GRACE were at unique locations in terms of local time and LAT (Figure 2). The uncertainty bounds for species densities showed potential for scientific value when considering relative abundances or the uncertainty associated with the O-He transition region. Instead of having a specific altitude where He takes over as the dominant constituent, we observed a 1-σ interval of 45–50 km where this may occur, depending on geographical location. Other panels in Figure 2 highlight the improvement in temperature and density prediction with the MSIS-UQ T_∞ predictions. Not only is the biased reduced, the uncertainty estimates can be used to inform decision making (e.g., collision probability estimation). In panel (e), we observed the effect of uncertain exospheric temperature on the relative uncertainty in density as a function of altitude, highlighting the ability to provide different uncertainty ranges as a function of position.

We also investigated the relationship between geomagnetic activity and density. We compared the density of multiple models to CHAMP during the 2003 Halloween storm (Figure 3). MSIS-UQ portrayed similar characteristics to the satellite estimates, particularly after the storm when there was the most disagreement between the models. This was further explored in the time-lag study in Figure 4. By individually varying the time-series geomagnetic model drivers, we can observe the models' reaction in density. This showed NRLMSIS 2.0 has a very linear relationship between geomagnetic activity and density. Furthermore, the more recent the driver, the more it impacts density predictions. For MSIS-UQ, we observed nonlinear relationships between many of the time-lagged drivers and density, and the current *ap* value was rarely the most closely tied to density enhancements. We also observed post-storm overcooling effects due to the negative density ratios when the time-lagged drivers were increased. A video showing the evolution of density at 400 km during the 2003 Halloween (SM2) provided an example of how MSIS-UQ resembles behavior seen by a physics-based model during extremely nonlinear events, enhancing the capabilities of NRLMSIS 2.0.

MSIS-UQ has value in the community both from a research and operational standpoint. The model can be used to study uncertainty effects on species density, mass density, and temperature on a global scale. It can also be used to investigate global/local mean and uncertainty responses to geomagnetic storms. From an operational perspective, the uncertainty estimates are particularly valuable. MSIS-UQ combines the internal formulation of NRLMSIS 2.0 with calibrated density uncertainties which can be used to obtain more precise collision probability estimates and even uncertainty estimates on satellite re-entry time/location. Further work can be conducted to also tune other temperature profile parameters in NRLMSIS 2.0 to vary the profile in the lower thermosphere. One alternative approach to achieving this is described by Mehta et al. (2019).

Data Availability Statement

Requests can be submitted for full access to the SET HASDM density database at https://spacewx.com/hasdm/ and all reasonable requests for scientific research are accepted as explained in the rules of road document on the website. The historical space weather indices used in this study can be found at the following links: $F_{10.7}$: https://www.spaceweather.gc.ca/forecast-prevision/solar-solaire/solarflux/sx-en.php, ap: https://doi.org/10.5880/Kp.0001, and SYM-H: http://wdc.kugi.kyoto-u.ac.jp/aeasy/index.html. The remaining solar indices and proxies can be found at https://spacewx.com/jb2008/ in the SOLFSMY.TXT file. Free, one-time only registration is required to access the historical data while nowcasts and forecasts are provided by SET as a data service from data@spacewx.com. The Weimer (2019) Pointing flux data can be accessed at https://doi.org/10.5281/zenodo.3525166. For operational use of this model, the Poynting flux totals can be obtained through the W05 model (Weimer, 2005a, 2005b), and ΔT can be computed using equations in Weimer et al. (2020, 2021). W05 is driven in real-time by solar wind measurements. Programs and files used for ML model development are available at https://doi.org/10.5281/zenodo.7245164. Specifically, data used for training, validation, and testing is provided along with programs to train a model from scratch. The pre-trained model is also available with a program to evaluate it for a given set of inputs.

References

Boniface, C., & Bruinsma, S. (2021). Uncertainty quantification of the DTM2020 thermosphere model. *Journal of Space Weather and Space Climate*, *11*, 53. https://doi.org/10.1051/swsc/2021034

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*. AIAA 2008-6438.

Bruinsma, S., & Biancale, R. (2003). Total densities derived from accelerometer data. *Journal of Spacecraft and Rockets*, *40*(2), 230–236. https://doi.org/10.2514/2.3937

Bussy-Virat, C. D., Ridley, A. J., & Getchius, J. W. (2018). Effects of uncertainties in the atmospheric density on the probability of collision between space objects. *Space Weather*, *16*(5), 519–537. https://doi.org/10.1029/2017SW001705

Calabia, A., & Jin, S. (2016). New modes and mechanisms of thermospheric mass density variations from GRACE accelerometers. *Journal of Geophysical Research: Space Physics*, *121*(11), 11191–11212. https://doi.org/10.1002/2016JA022594

Choury, A., Bruinsma, S., & Schaeffer, P. (2013). Neural networks to predict exosphere temperature corrections. *Space Weather*, *11*(10), 592–602. https://doi.org/10.1002/2013SW000969

Doornbos, E. (2012). *Producing density and crosswind data from satellite dynamics observations* (pp. 91–126). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-25129-0_4

Emmert, J. (2015). Thermospheric mass density: A review. *Advances in Space Research*, *56*(5), 773–824. https://doi.org/10.1016/j.asr.2015.05.038

Emmert, J. T., Drob, D. P., Picone, J. M., Siskind, D. E., Jones, M., Jr., 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), e2020EA001. https://doi.org/10.1029/2020EA001321

Friis-Christensen, E., Lühr, H., & Hulot, G. (2006). Swarm: A constellation to study the Earth's magnetic field. *Earth, Planets and Space*, *58*(4), 351–358. https://doi.org/10.1186/BF03351933

Hapgood, M., Liu, H., & Lugaz, N. (2022). Spacex—sailing close to the space weather? *Space Weather*, *20*(3), e2022SW003. https://doi.org/10.1029/2022SW003074

Iyemori, T. (1990). Storm-time magnetospheric currents inferred from mid-latitude geomagnetic field variations. *Journal of Geomagnetism and Geoelectricity*, *42*(11), 1249–1265. https://doi.org/10.5636/jgg.42.1249

Knipp, D. J., Pette, D. V., Kilcommons, L. M., Isaacs, T. L., Cruz, A. A., Mlynczak, M. G., et al. (2017). Thermospheric nitric oxide response to shock-led storms. *Space Weather*, *15*(2), 325–342. https://doi.org/10.1002/2016SW001567

Lei, J., Burns, A. G., Thayer, J. P., Wang, W., Mlynczak, M. G., Hunt, L. A., et al. (2012). Overcooling in the upper thermosphere during the recovery phase of the 2003 October storms. *Journal of Geophysical Research*, *117*(A3). https://doi.org/10.1029/2011JA016994

Licata, R., Mehta, P., & Tobiska, W. K. (2021). Impact of driver and model uncertainty on drag and orbit prediction. In *Proceedings of the 31st AAS/AIAA space flight mechanics meeting*.

Licata, R. J., & Mehta, P. M. (2022). Uncertainty quantification techniques for data-driven space weather modeling: Thermospheric density application. *Scientific Reports*, *12*, 7256. https://doi.org/10.1038/s41598-022-11049-3

Licata, R. J., Mehta, P. M., Tobiska, W. K., & Huzurbazar, S. (2022). Machine-learned HASDM thermospheric mass density model with uncertainty quantification. *Space Weather*, *20*(4). https://doi.org/10.1029/2021SW002915

Licata, R. J., Mehta, P. M., Weimer, D. R., Drob, D. P., Tobiska, W. K., & Yoshii, J. (2022). Science through machine learning: Quantification of post-storm thermospheric cooling. *Space Weather*, *20*(9) e2022SW003189. https://doi.org/10.1029/2022SW003189

Licata, R. J., Mehta, P. M., Weimer, D. R., & Tobiska, W. K. (2021). Improved neutral density predictions through machine learning enabled exospheric temperature model. *Space Weather*, *19*(12), e2021SW002. https://doi.org/10.1029/2021SW002918

Liu, H., Lühr, H., Henize, V., & Köhler, W. (2005). Global distribution of the thermospheric total mass density derived from CHAMP. *Journal of Geophysical Research*, *110*(A4). https://doi.org/10.1029/2004JA010741

Lühr, H., Grunwaldt, L., & Forste, C. (2002). CHAMP reference systems, transformations and standards Tech. rep. CH-GFZ-RS-002. *GFZ-Potsdam*.

Mehta, P. M., Linares, R., & Sutton, E. K. (2019). Data-driven inference of thermosphere composition during solar minimum conditions. *Space Weather*, *17*(9), 1364–1379. https://doi.org/10.1029/2019SW002264

Mehta, P. M., Walker, A. C., Sutton, E. K., & Godinez, H. C. (2017). New density estimates derived using accelerometers on board the CHAMP and GRACE satellites. *Space Weather*, *15*(4), 558–576. https://doi.org/10.1002/2016SW001562

Oliveira, D. M., Zesta, E., Mehta, P. M., Licata, R. J., Pilinski, M. D., Tobiska, W. K., & Hayakawa, H. (2021). The current state and future directions of modeling thermosphere density enhancements during extreme magnetic storms. *Frontiers in Astronomy and Space Sciences*, *8*. https://doi.org/10.3389/fspas.2021.764144

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al. (2019). Keras Tuner. Retrieved from https://github.com/keras-team/keras-tuner

Qian, L., Burns, A., Emery, B., Foster, B., Lu, G., Maute, A., et al. (2013). The NCAR TIE-GCM: A community model of the coupled thermosphere/ionosphere system. *Geophysical Monograph Series*, *201*, 73–83. https://doi.org/10.1029/2012GM001297

Ruan, H., Lei, J., Dou, X., Liu, S., & Aa, E. (2018). An exospheric temperature model based on CHAMP observations and TIEGCM simulations. *Space Weather*, *16*(2), 147–156. https://doi.org/10.1002/2017SW001759

Storz, M., Bowman, B., & Branson, J. (2005). High accuracy satellite drag model (HASDM). In *AIAA/AAS astrodynamics specialist conference and exhibit*. https://doi.org/10.2514/6.2002-4886

Sutton, E. K. (2008). Effects of solar disturbances on the thermosphere densities and winds from CHAMP and GRACE satellite accelerometer data Ph.D. thesis. *University of Colorado at Boulder*.

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). https://doi.org/10.1029/2004GL019920

Tobiska, W. K., Bouwer, S. D., & Bowman, B. R. (2008). The development of new solar indices for use in thermospheric density modeling. *Journal of Atmospheric and Solar-Terrestrial Physics*, *70*(5), 803–819. https://doi.org/10.1016/j.jastp.2007.11.001

Tobiska, W. K., Bowman, B. R., Bouwer, D., Cruz, A., Wahl, K., Pilinski, M., et al. (2021). The SET HASDM density database. *Space Weather*, *19*(4), e2020SW002682. https://doi.org/10.1029/2020SW002682

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

Weimer, D. (2019). Improving neutral density predictions using exospheric temperatures calculated on a geodesic. *Polyhedral Grid*. https://doi.org/10.5281/zenodo.3525166

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

Weimer, D. R. (2005b). Predicting surface geomagnetic variations using ionospheric electrodynamic models. *Journal of Geophysical Research*, *110*(A12), A12307. https://doi.org/10.1029/2005JA011270

Weimer, D. R., Mehta, P. M., Tobiska, W. K., Doornbos, E., Mlynczak, M. G., Drob, D. P., & Emmert, J. T. (2020). Improving neutral density predictions using exospheric temperatures calculated on a geodesic, polyhedral grid. *Space Weather*, *18*(1), e2019SW002. https://doi.org/10.1029/2019SW002355

Weimer, D. R., Mlynczak, M. G., Hunt, L. A., & Tobiska, W. K. (2015). High correlations between temperature and nitric oxide in the thermosphere. *Journal of Geophysical Research: Space Physics*, *120*(7), 5998–6009. https://doi.org/10.1002/2015JA021461

Weimer, D. R., Sutton, E. K., Mlynczak, M. G., & Hunt, L. A. (2016). Intercalibration of neutral density measurements for mapping the thermosphere. *Journal of Geophysical Research: Space Physics*, *121*(6), 5975–5990. https://doi.org/10.1002/2016JA022691

Weimer, D. R., Tobiska, W. K., Mehta, P. M., Licata, R. J., Drob, D. P., & Yoshii, J. (2021). Comparison of a neutral density model with the SET HASDM density database. *Space Weather*, *19*(12), e2021SW002. https://doi.org/10.1029/2021SW002888