Space Weather

RESEARCH ARTICLE
10.1029/2024SW004070

A Global Thermospheric Density Prediction Framework Based on a Deep Evidential Method

Yiran Wang[1] **and Xiaoli Bai**[1]

[1]Department of Mechanical and Aerospace Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

Key Points:

- A global thermospheric density prediction framework is presented based on a deep evidential method
- The proposed framework predicts thermospheric density at the required time and geographic position with given geomagnetic and solar indices
- The proposed framework predicts thermospheric density more accurately than two empirical models and provides reliable uncertainty estimations

Correspondence to:
X. Bai,
xiaoli.bai@rutgers.edu

Citation:
Wang, Y., & Bai, X. (2024). A global thermospheric density prediction framework based on a deep evidential method. *Space Weather*, *22*, e2024SW004070. https://doi.org/10.1029/2024SW004070

Received 11 JUL 2024
Accepted 12 NOV 2024

Author Contributions:
Conceptualization: Yiran Wang, Xiaoli Bai
Formal analysis: Yiran Wang, Xiaoli Bai
Funding acquisition: Xiaoli Bai
Investigation: Yiran Wang
Methodology: Yiran Wang, Xiaoli Bai
Project administration: Xiaoli Bai
Resources: Xiaoli Bai
Software: Yiran Wang
Supervision: Xiaoli Bai
Validation: Yiran Wang
Visualization: Yiran Wang
Writing – original draft: Yiran Wang
Writing – review & editing: Yiran Wang, Xiaoli Bai

Abstract Thermospheric density influences the atmospheric drag and is crucial for space missions. This paper introduces a global thermospheric density prediction framework based on a deep evidential method. The proposed framework predicts thermospheric density at the required time and geographic position with given geomagnetic and solar indices. It is called global to differentiate it from existing research that only predicts density along a satellite orbit. Through the deep evidential method, we assimilate data from various sources including solar and geomagnetic conditions, accelerometer-derived density data, and empirical models including the Jacchia-Bowman model (JB-2008) and the Naval Research Laboratory Mass Spectrometer and Incoherent Scatter Radar Extended (NRLMSISE-00) model. The framework is investigated on five test cases along various satellites from 2003 to 2015 involving geomagnetic storms with Disturbance Storm Time (Dst) values smaller than −50 nT. Results show that the proposed framework can generate density with higher accuracy than the two empirical models. It can also obtain reliable uncertainty estimations. Global density estimations at altitudes from 200 to 550 km are also presented and compared with empirical models on both quiet and storm conditions.

Plain Language Summary Thermospheric density affects atmospheric drag, which is important for many space missions. This paper presents a global thermospheric density prediction framework based on a deep evidential method. It incorporates data from empirical models JB-2008 and NRLMSISE-00, solar and geomagnetic indices, and density derived from satellite accelerometers. Test cases are designed from 2003 to 2015 under quiet and storm conditions based on various satellites. The proposed framework demonstrates higher accuracy and reliable uncertainty estimates compared to the empirical models.

1. Introduction

In recent years, active satellites in Earth's orbit have greatly increased. As of February 2024, there are 9,401 operational satellites in orbits and 8,299 of them are in low Earth orbit (LEO) (OrbitingNow, 2024). Atmospheric drag is the dominant perturbation force and is the most difficult to predict for orbit propagation of LEO satellites (Marcos et al., 1998). It is influenced by several causes, including the thermospheric density, drag coefficient, the area of the satellite facing the fluid, and the relative velocity between the satellite and the surrounding atmosphere. Fluctuations in thermospheric density may result in alterations to the trajectories of satellites, impacting communication systems (Chartier et al., 2013), orbital lifetimes, and overall mission planning. Significant evolutions have occurred in the study of thermospheric density prediction in space weather, driven by the demands for accuracy and reliability.

Earth's thermospheric density is a complex system affected by various factors, such as solar flares, Coronal Mass Ejections (CMEs), high-speed solar winds, seasonal and solar cycle variations, etc. Besides, the location and the corresponding temperature also affect the thermospheric density (Hall, 2021). Traditionally, thermospheric density prediction mainly relied on physical laws and observational data. The commonly used empirical models, such as the Naval Research Laboratory Mass Spectrometer and Incoherent Scatter Radar Extended (Picone et al., 2002) and the Jacchia-Bowman (Bowman et al., 2008) models can produce considerable errors during periods with high-level solar and geomagnetic activities. The uncertainties extend over 100% under such extreme conditions (Group, 2012). On the other side, the principled, physics-based models such as the Thermospheric General Circulation Models (UCAR, 2016) and the Global Ionosphere Thermosphere Model (Aaron et al., 2024) can provide a more realistic representation of the thermosphere but with high computing requests. In addition, it is still challenging for these models to outperform empirical models because of the large uncertainties in inputs and boundary conditions.

Studies indicate that the thermospheric density experiences strong fluctuations during geomagnetic storms (Burke et al., 2007; Burns et al., 1995, 2004), subsequently causing large drag variability on satellites (Oliveira & Zesta, 2019). The factors that influence the storms mainly include the solar winds, CMEs, geomagnetic activities, and other related phenomena (Gonzalez et al., 1994; Qian & Solomon, 2012). The Disturbance Storm Time (Dst) magnetic activity index serves as a common metric for categorizing the occurrence of geomagnetic storms and defining the duration of the storms. In this paper, we use Dst to define quiet and storm conditions (Richardson et al., 2006; Y. Wang & Bai, 2023b; Zhang et al., 2007): the quiet time is defined as the period when the minimum Dst is larger than −50 nT; a moderate storm occurs when the Dst value is between −50 and −100 nT; an intense storm happens when Dst is from −100 to −250 nT, and the super-storm is defined when the Dst is smaller than −250 nT.

Over the past decades, researchers have attempted to model thermospheric density using a variety of new approaches, including improving empirical models, exploring physics-based models, and developing data-driven models. Emmert (2009) analyzed a long-term global averaged total mass density database based on two-line element (TLE) data, which covers the years from 1967 to 2007, and the range of altitude is from 200 to 600 km. They combined about 5,000 objects into a single-density database to analyze the global average absolute density. Weimer et al. (2011) computed the total Poynting flux flowing into both polar hemispheres as a function of time. They compared the densities at two altitudes obtained from accelerometers on the CHAMP and GRACE satellites. Elvidge et al. (2016) proposed a multi-model ensembles (MME) strategy that combined different and independent models for forecasting the density. The results indicate that during solar minimum test scenarios using the MME to initialize Thermosphere Ionosphere Electrodynamics General Circulation Model (TIE-GCM) shows a reduction in Root Mean Squared Error (RMSE) of 60% Weimer et al. (2018) investigated the amplitudes and timings of combined, annual, and semiannual variations in thermospheric neutral density. This study involved a comparison of these variations with measurements of the infrared emissions from carbon dioxide and nitric oxide in the thermosphere. The study also revealed a strong correlation between semiannual variations in density and observed the changes in infrared emissions from carbon dioxide and nitric oxide during periods of high solar activity. Forootan et al. (2022) proposed a method called "C/DA-NRLMSISE-00" to use the density estimated from the GRACE to re-calibrate the NRLMSISE-00's parameters based on Ensemble Kalman Filter and data assimilation (C/DA) technique. The estimated density from the method was compared with the density from the original NRLMSISE-00 model along the GRACE and SWARM satellites's orbits. The results indicated that the proposed method provided smaller RMSE values and uncertainties than the NRLMSISE-00 model. Mutschler et al. (2023) introduced a method named SoleiTool which incorporated the position of CubeSats for estimating the density.

Several studies have used dynamic reduced-order models (ROM) based on satellite data to improve the thermospheric density prediction accuracy. Mehta et al. (2019) used modal decomposition to extract a reduced-order representation. They presented both the density and the temperature with different basis functions. Gondelach and Linares (2020) used the ROM method with the TLE data to predict the density for real-time. Their results had a smaller bias and RMSE than the empirical models. Mehta and Linares (2020) extended the dynamic ROM with satellite position measurements. The results demonstrated the effectiveness of the Unscented Kalman filter in estimating position, velocity, and density. Gondelach and Linares (2021) further used the ROM with radar and GPS tracking data to estimate and validate the density compared with the SWARM satellites. Licata and Mehta (2023) used the ROM to reduce the dimensionality of the TIE-GCM model and employed recurrent neural networks to model the thermosphere with a quicker calculation speed than the numerical model.

In recent years, machine learning (ML) has increasingly been used to predict thermospheric density (Nateghi, 2021). Pérez and Bevilacqua (2015) use neural networks to reduce the error in the density estimated by three empirical models. Weng et al. (2020) proposed an Artificial Neural Network Model (ANNM) to predict the thermospheric density in long-term trends (1967–2013). They compared the estimated density from the ANNM, the NRLMSISE-00 model, and the Global Mass Density Model to the satellite drag-derived densities. Bonasera et al. (2021) used the deep ensembles neural network with dropout to estimate the thermospheric density and the uncertainty. The inputs for the network included density data from satellite accelerometers, orbital parameters, and solar and geomagnetic indices. Licata, Mehta, Tobiska, and Huzurbazar (2022) used the Principal Component Analysis method to build the model. The predicted density showed optimal results than the JB-2008 model along the satellite orbit. Acciarini et al. (2024) used the Karman package to build a data-driven ML model to investigate

prediction accuracy. Their model was trained on precise orbit determination-derived thermospheric density data from satellites.

In previous studies Gao et al. (2020), Y. Wang and Bai (2022), and Y. Wang et al. (2021), we have developed and improved a thermospheric density prediction framework that integrates data from empirical models, including the JB-2008 and NRLMSISE-00, solar and geomagnetic indices, and accelerometer-derived density data along a particular satellite to predict thermospheric density. A recent study Y. Wang and Bai (2023b) successfully demonstrated the effectiveness of the deep evidential method in predicting thermospheric density along a satellite, achieving high accuracy and reliability uncertainty estimation.

Building upon this success, our current work aims to expand this approach into a global thermospheric density prediction framework based on the deep evidential method. This upgraded framework can estimate the thermospheric density at any geographic position and at any time with the given space weather condition. The deep evidential method can account for both aleatoric and epistemic uncertainty, which is crucial for ensuring reliable predictions. Initial results were obtained in predicting density for two different satellites within a close altitude range (Y. Wang & Bai, 2023a). In this paper, we extend this work to test the thermospheric density for more satellites with wider altitude ranges.

The paper makes the following contributions. First, the framework proposed in this paper shows great potential in predicting global density at a given time and location while only requiring the corresponding space weather indices as the external input. Second, experimental results demonstrate that the proposed framework generates thermospheric density with high accuracy and reliable uncertainty estimation for various satellites. Even when the test periods are not part of the training database but are in the future, the model exhibits a high level of accuracy. Third, test results indicate the robustness of the proposed framework in different conditions, including both quiet and storm conditions. Fourth, the framework shows its adaptability as the performance of the predicted results improves with more density information, even if the additional information is limited. Finally, the estimated density from the proposed framework shows similar patterns as the empirical models in the quiet condition and different patterns in the storm condition. Additionally, the predicted density is closer to the satellite accelerometer-derived densities than the two empirical models in both quiet and storm conditions.

The rest of this paper is organized as follows. Section 2 presents the methodology in detail, including the key concepts of the deep evidential method and how we design the neural network structures. We also introduce the proposed global thermospheric density prediction framework and the metrics used to evaluate performance. Section 3 first describes the databases we use for training and test cases for evaluations. It then presents and analyzes the results for all the cases. Section 4 presents and compares the predicted density with the density obtained from the empirical models JB-2008 and NRLMSISE-00 globally for a quiet and a storm condition. Additionally, the density from the satellite accelerometers at specific locations at the same time is used to verify the results. Conclusions are presented in the last section.

2. Methodology

2.1. Deep Evidential Method

Amini et al. (2020) introduced a deep evidential regression method to learn the uncertainty by placing evidential priors over the original Gaussian likelihood function and training the neural network to infer the hyperparameters of the evidential distribution. Consider a data set comprising input variables $\mathbf{X} = \{\boldsymbol{x}_i\}_{i=1}^{n} \in \mathbb{R}^{n\times d}$ and its corresponding output $\mathbf{y} = \{y_i\}_{i=1}^{n} \in \mathbb{R}^{n}$, where n is the number of samples in the data set and d is the dimension of the input. Assume the output $\mathbf{y}$ follows a Gaussian distribution with unknown mean and variance $\mathcal{N}(\mu, \sigma^2)$. An unknown mean is subjected to a Gaussian prior, while an unknown variance is subjected to an Inverse-Gamma prior. These assumptions can be represented as:

$$\begin{aligned} (y_1, \ldots, y_N) &\sim \mathcal{N}(\mu, \sigma^2) \\ \mu \sim \mathcal{N}(\gamma, \sigma^2 v^{-1}) \quad \sigma^2 &\sim \Gamma^{-1}(\alpha, \beta) \end{aligned} \tag{1}$$

where $\Gamma(\cdot)$ is the gamma function. The hyper-parameters now can be defined as $\boldsymbol{m} = (\gamma, v, \alpha, \beta)$ and $\gamma \in \mathbb{R}$, $v > 0$, $\alpha > 1$, $\beta > 0$.

The model is trained using a novel loss function defined in Equation 2 (Amini et al., 2020) that measures the distance and the discrepancy between the predicted and the true values so that the network can make predictions as well as provide uncertainty estimations.

$$\mathcal{L}_i(\boldsymbol{w}) = \mathcal{L}_i^{\mathrm{NLL}}(\boldsymbol{w}) + \lambda \mathcal{L}_i^{\mathrm{R}}(\boldsymbol{w}) \tag{2}$$

where $\mathcal{L}_i^{\mathrm{NLL}}(\boldsymbol{w})$ is the negative logarithm of model evidence, and the definition of $\mathcal{L}_i^{\mathrm{NLL}}(\boldsymbol{w})$ is shown in Equation 3. An objective for training a neural network is to maximize the model evidence to fit the observations by using the normal inverse gamma distribution.

$$\mathcal{L}_i^{\mathrm{NLL}}(\boldsymbol{w}) = \frac{1}{2}\log\left(\frac{\pi}{v}\right) - \alpha \log(2\beta(1+v)) + \left(\alpha + \frac{1}{2}\right)\log\left(\left(y_i - \gamma\right)^2 v + 2\beta(1+v)\right) + \log\left(\frac{\Gamma(\alpha)}{\Gamma\left(\alpha + \frac{1}{2}\right)}\right) \tag{3}$$

To minimize the errors in evidence, a regularization term is defined in Equation 4, which balances the loss function through the regularization coefficient λ in Equation 2.

$$\mathcal{L}_i^{\mathrm{R}}(\boldsymbol{w}) = \left|y_i - \mathbb{E}[\mu_i]\right| \cdot (2v + \alpha) = \left|y_i - \gamma\right| \cdot (2v + \alpha) \tag{4}$$

The prediction, aleatoric, and epistemic uncertainties are calculated as follows:

$$\text{Prediction: } \mathbb{E}[\mu] = \gamma \tag{5}$$

$$\text{Aleatoric Uncertainty: } \mathbb{E}[\sigma^2] = \frac{\beta}{\alpha - 1} \tag{6}$$

$$\text{Epistemic Uncertainty: } \mathrm{Var}[\mu] = \frac{\beta}{v(\alpha - 1)} \tag{7}$$

2.2. Global Thermospheric Density Prediction Framework Definition

In this paper, we propose a data-driven framework based on the deep evidential method. The framework takes the input of time indices, geographic positions, empirical models, and geomagnetic and solar indices. Explicitly, the inputs and the output of the model are described as Equation 8, the basic form of which has been studied in our previous papers (Y. Wang & Bai, 2022, 2023a, 2023b; Y. Wang et al., 2021). In this upgraded framework defined in Equation 8, we add time indices and geographic positions as new variables to enhance the model's capability to predict density at specific positions and time. Additionally, we do not include satellite-specific density information as we did in the previous studies because such information may be unavailable and the model is a global model.

$$\lg(\hat{\rho}(t)) = f\begin{pmatrix} t_1(t),\ t_2(t),\ t_3(t),\ t_4(t),\ t_5(t),\ t_6(t) \\ lat_{sat}(t),\ lon_{sat}(t),\ height_{sat}(t), \\ \lg(\rho_{JB}(t)),\ \ldots\ \lg(\rho_{JB}(t - T_{JB})) \\ \lg(\rho_{NRL}(t)),\ \ldots\ \lg(\rho_{NRL}(t - T_{NRL})) \\ F_{10.7}(t - 1d),\ F_{10.7A}(t - 1d),\ F_{30}(t),\ Ap(t) \\ Dst(t),\ Dst(t - 1hr),\ \ldots,\ Dst(t - T_{Dst}), \\ SymH(t),\ SymH(t - 1min),\ \ldots,\ SymH(t - T_{SymH}) \end{pmatrix} \tag{8}$$

The $\hat{\rho}$ on the left side of the equation is the predicted density from the evidential model at time t. Here we assume the density from the satellite accelerometer as the truth as it is regarded as the most accurate information among all

Table 1
Framework Variable Definitions

Variable	Definition
$F_{10.7}$	Daily value of 10.7 cm solar radio flux with a 1-day lag
$F_{10.7A}$	Ninety-one days mean of daily $F_{10.7}$ a 1-day lag
F_{30}	Daily value of 30 cm solar radio flux
Ap	Derived from the 3-hr geomagnetic index K_p
Dst	An index for the strength of the ring current with 1-hr resolution
SymH	One-minute resolution of the longitudinally symmetric component of the magnetic field disturbances (Iyemori, 1990)

the available information, and has been widely used in literature for performance validation (Bonasera et al., 2021; Gao et al., 2020; Pérez & Bevilacqua, 2015; Pérez et al., 2014). Additionally, for the following experiments, we use the information from CHAMP, GRACE-A provided by Mehta et al. (2017) and Mehta (2017). We also use the information from GOCE (ESA, 2024a), and SWARM (ESA, 2024b) provided by the ESA. Refereed to Licata, Mehta, Weimer, et al. (2022), t_1, t_2, …, t_6 are defined as:

$$t_1 = \sin\left(\frac{2\pi doy}{365.25}\right) \quad t_2 = \cos\left(\frac{2\pi doy}{365.25}\right) \tag{9}$$

$$t_3 = \sin\left(\frac{2\pi UT}{24}\right) \quad t_4 = \cos\left(\frac{2\pi UT}{24}\right) \tag{10}$$

$$t_5 = \sin\left(\frac{2\pi LST}{24}\right) \quad t_6 = \cos\left(\frac{2\pi LST}{24}\right) \tag{11}$$

where t_1 and t_2 represent the annual variations using the day of year *doy* as the variable. t_3 and t_4 represent diurnal variations using the universal time *UT*. *LST* is the local solar time transferred at the current time t of the current position. $lat_{sat}(t)$, $lon_{sat}(t)$, $height_{sat}(t)$ are the latitude, longitude, and height to specify the position of the satellite at the current time t. ρ_{JB} and ρ_{MSIS} are the densities estimated by the two empirical models JB2008 and NRLMSISE-00.

The meanings of the other variables are stated in Table 1.

We established parameters for time delays through a trial and error process. The delays in JB-2008 or NRLMSISE-00 expressed as T_{JB} and T_{MSIS}, respectively, are set as 15 min. A 3-hr time delay is applied to Dst, denoted by T_{Dst}. The time delay of SymH, represented by T_{SymH}, is set as 15 min. Previous studies Y. Wang and Bai (2022) and Y. Wang et al. (2021) have shown that providing these inputs is essential for the model to make accurate predictions.

2.3. Neural Network Structure

The neural network is constructed using the Keras toolbox (Keras Team, n.d.) in Python 3.9. We design several models with different training databases as outlined in Table 4 in the next section, and the network structures differ accordingly. Our approach involves an initial optimization of the neural network structures using the KerasTuner (O'Malley et al., 2019). The tuner settings are outlined in Table 2.

Based on the initial results from the KerasTuner, we further adjust the network structure to obtain better results. Table 3 shows the final structure of Model-1, which is set as the base model, and its training database will be explained in the next section. To reduce the effect of randomness and provide a robust and reliable assessment of the model's performance, we execute each experiment five times, employing different random seeds for each run. The final results are the average outcomes across these five runs.

Table 2
Hyperparameter Search for NN Model

Parameter	Settings
Num of hidden layer	1–5
Neurons	min = 64, max = 512, and step = 32
Activation function	linea, relu, elu
Optimizer	adam, sgd

2.4. Performance Metrics

We select four metrics to evaluate the performance of the proposed framework on both prediction accuracy and uncertainty estimation.

To evaluate the accuracy of the predictions, we use the Pearson correlation coefficient (R) and the RMSE. R and RMSE can be mathematically expressed as Equations 12 and 13:

$$\mathrm{R} = \frac{\sum_{i=1}^{n}(\rho_i - \bar{\rho})(\hat{\rho}_i - \bar{\hat{\rho}})}{(n-1)\sigma_\rho \sigma_{\hat{\rho}}} \tag{12}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\rho}_i - \rho_i)^2} \tag{13}$$

where ρ_i and $\hat{\rho}_i$ are the true density and predicted density. $\bar{\rho}$ represents the mean value of the density. σ_ρ and $\sigma_{\hat{\rho}}$ are the standard deviations of the truth and the predictions, and n is the size of the data that are used for evaluation. A good performance shall have an R close to one and an RMSE as small as possible.

To evaluate the uncertainty estimation performance of the proposed framework, we count the coverage rate of the 2σ area (Cov Rate). The coverage rate is defined as Equation 14:

$$\text{Cov Rate} = \frac{k}{n} \times 100\% \tag{14}$$

where k is the number of the true density that is within the 2σ uncertainty boundaries estimated by the evidential model. A good performance shall have a coverage rate close to 95%.

We also evaluate the confidential level and calculate the Mean Absolute Calibration Error (MACE) (Anderson et al., 2020) to evaluate the reliability of the uncertainty, which is defined as Equation 15.

$$\mathrm{MACE} = \frac{1}{n_C}\sum_{k=1}^{n_C}|C[k] - P[k]| \tag{15}$$

where $C[k]$ is the confidence interval level, and in this paper we define it as $C[k] = [5\%, 10\%, \ldots, 95\%, 99\%]$. n_c is the number of the confidence interval levels. $P[k]$ is the percentage of the predicted data within the uncertainty boundaries associated with the $C[k]$.

The MACE calculates the average difference between the ideal and actual probability in the test set and helps us evaluate the overall reliability of the model's prediction. A lower value of the MACE indicates a better-calibrated model.

3. Case Studies and Results

3.1. Experiments Design

To reduce the running time while including enough information and exploring our models' predictive capability, we did not use all the accessible data from CHAMP and GRACE-A, but instead selected several periods as summarized in Table 4. The data spans from year 2002 to 2006, with the selected months that contain intense geomagnetic storms, identified by Dst values smaller than −100 nT. The details of the storms can be referred to Y. Wang and Bai (2023b). We emphasize that the months of October and November in 2003, during which the Halloween storm occurred, have been excluded from the training database. This deliberate omission is to demonstrate the density prediction capabilities of the proposed framework under extreme geomagnetic storm conditions.

Table 3
Model-1 Network Structure

Hyperparameters	Settings
Network structure	Input - 128 - 256 - Output
Activation function	relu
Optimizer	adam

Table 4
Training Periods of CHAMP and GRACE

Year	Month	Year	Month
2002	August–October	2005	January, May, June, August, and September
2003	May–August	2006	April and December
2004	February, April, July, August, and November		

Overall, we design the following models to explore the predicted performance based on different databases. The definitions of these models with their corresponding databases are shown in Table 5.

In Model-1, the training database contains information from both CHAMP and GRACE-A, during the specific periods defined in Table 4. Model-2 includes information exclusively from CHAMP. As a comparison of similar situations, Model-3 contains the information only from GRACE-A. Model-4 contains the data from CHAMP and GRACE-A similar to Model-1 but with an additional 1-month data from 01 January 2010 to 31 January 2010 on GOCE. In a parallel fashion, Model-5 contains the data from CHAMP and GRACE-A the same as Model-1, and a shorter duration of additional data from 01 February 2014 to 15 February 2014 from SWARM-A.

The well-trained models will be applied in different test cases as defined in Table 6. We will assess the performance of Model-1, 2, and 3 in Test-1, spanning from 27 October 2003 to 03 November 2003 on both CHAMP and GRACE-A. This period is the famous Halloween Storm, which contains a large geomagnetic storm during which the minimum Dst is −432 nT. This test will be used to evaluate the performance of the models derived from the same satellites but the periods are not included in the training database.

Tests 2 and 3 are designed to evaluate the performance along GOCE under moderate storms when the Dst values during the two test periods are smaller than −50 nT but larger than −100 nT. We will also evaluate the same period on CHAMP and GRACE-A. With different satellites included in the training database, Model-1, 2, 3, and 4 will be applied to these two test cases. We highlight that the test periods are in the future time of all the training databases, and the GOCE is an unknown satellite for Model-1, 2, and 3.

The choice of the test date on Test-4 and 5 is motivated by the occurrence of intense storms happened during the specific periods. We will evaluate Model-1, 2, 3, and 5 on these two test cases. Both test periods are in the further future time beyond the training database, and these two SWARM satellites are not included in Model-1, 2, and 3.

The evaluations in Test-2 and 3, as well as Test-4 and 5, will assess the model's efficacy in predicting thermospheric density on a new satellite using the proposed framework. The evaluation of Model-4 and Model-5 is to incorporate information from the new satellites as extended data sources during training, but the available data periods on the new satellites are limited. The goal is to investigate the model's effectiveness in predicting thermospheric density on previously untrained satellites. We will also compare the performances of the models with two empirical models, JB-2008 and NRLMSISE-00. Since the empirical models cannot provide the uncertainty, we will focus on accuracy metrics for both empirical models. Collectively, these carefully designed evaluations will provide a comprehensive view of the capability of the proposed framework.

Table 5
Model Definition

Model	Database
1	CHAMP and GRACE-A
2	CHAMP
3	GRACE-A
4	CHAMP, GRACE-A, and 01 January 2010–31 January 2010 on GOCE
5	CHAMP, GRACE-A, and 01 February 2014–15 February 2014 on SWARM-A

3.2. Results

3.2.1. Test-1: From 27 October 2003 to 03 November 2003

In the test period of Test-1, we evaluate model performances on both CHAMP and GRACE-A. The numerical results of Model-1, 2, and 3 on CHAMP are presented in Table 7 and the results on GRACE-A are presented in Table 8. The accuracy metrics of the two empirical models are also presented after Model-3.

For the accuracy of the predictions, Model-1 exhibits the highest R value of 0.9286, indicating a strong relationship between the true value and the predictions of the proposed model, which has been trained with both CHAMP and GRACE-A. Model-1 also shows the smallest RMSE value as $1.1179 \times 10^{-12} \mathrm{kg/m^3}$. The second best accurate performance is from

Table 6
Test Period

Test case	Test period	Min Dst (nT)
1	27 October 2003–03 November 2003 on CHAMP and GRACE-A	−432
2	03 April 2010–07 April 2010 on CHAMP and GOCE	−76
3	11 October 2010–12 October 2010 on GRACE-A and GOCE	−79
4	20 February 2014 on SWARM-A and SWARM-C	−112
5	15 March 2015–26 March 2015 on SWARM-B and SWARM-C	−234

Model-2, in which training data is from CHAMP, with an R value of 0.9093 m and RMSE value of 1.3764 $\times 10^{-12}$kg/m^3. The training data of Model-1 and Model-2 both contain information from CHAMP. When the training data only contains information from GRACE-A, as Model-3 shows, the predicted accuracy becomes worse than the first two models. The R value in Model-3 is 0.85, with a larger RMSE value 2.5572 $\times 10^{-12}$kg/m^3. The R value from the JB-2008 model is 0.8404, which is lower than the three proposed models. The RMSE value is 2.4116 $\times 10^{-12}$kg/m^3, smaller than Model-3, but larger than Model-1 and Model-2. As for the results from the NRLMSISE-00 model, the accuracy performance is the worst among all the results.

We plot the error histogram between the truth and the predicted value from JB-2008, NRLMSISE-00, and Model-1 in Figure 1. The blue section represents the error between the truth and the JB-2008 model, red section is the error of the NRLMSISE-00 model, and the yellow section represents the error of Model-1.

It can be seen from the histogram in Figure 1 that compared with the two empirical models, the errors produced by the base model have smaller mean and deviation values, indicating the base model is more accurate than the two empirical models.

As for the uncertainty estimation, Model-1 shows the most reliable uncertainty estimations as it provides the highest coverage rate, which reaches 93.46%, with the smallest MACE value, indicating its reliability in capturing the observed data. Model-2 shows the second-best coverage rate, which is beyond 92%. The MACE value is larger in Model-2. Model-3 shows the relatively worst performance among the three models. The coverage rate in Model-3 is 80%, which is much lower than the other two models. The corresponding MACE value in Model-3 is the largest reaching 0.1059.

We next present the test results on GRACE-A from 27 October 2003 to 03 November 2003 from Model-1, 2 and 3 in Table 8.

Again, the best accuracy performance is from Model-1, in which the database contains information from both CHAMP and GRACE-A. The R value in this test case is 0.8904 with the RMSE value as 3.0459 $\times 10^{-13}$kg/m^3. The performance on GRACE-A is similar to the test performance on CHAMP, as the second-best performance is from Model-3, in which the training database contains information only from GRACE-A. The R value in Mode-3 reaches 0.8809, with RMSE value 3.6164 $\times 10^{-13}$kg/m^3. The worst performance on accuracy in this test period is on Model-2, in which the R value is 0.8286 with a larger RMSE value of 5.7379 $\times 10^{-13}$kg/m^3. The R value in this case from the JB-2008 model is slightly larger than the R value in Model-2, but the RMSE value from the JB-2008 model is larger than Model-2. Except for this, the accuracy metrics of the three models based on the proposed framework are better than the two empirical models.

Table 7
Test on CHAMP: 27 October 2003–03 November 2003

Model	1	2	3	JB-2008	MSIS
R	0.9286	0.9093	0.8565	0.8404	0.7803
RMSE $\times 10^{-12}$	1.1179	1.3764	2.5572	2.4116	3.4837
Cov rate	0.9346	0.9235	0.8095		
MACE	0.0287	0.0435	0.1059		

When we look at the estimated uncertainty in Test-1 on GRACE-A, the most reliable uncertainty estimation is from Model-1, as the coverage rate reaches 90.34% with the smallest MACE value of 0.0472. The coverage rate in Model-3 is close to 90% as the second-best performance, and the MACE value is also the second smallest among the three models. Model-2 in this case shows relatively worst performance, for the uncertainty coverage rate reaching 86.31%, with a MACE value of 0.0889.

The storm case has been widely studied by other researchers, with some existing results available for comparison. Liu et al. (2011) studied the storm case and reported the correlation coefficients of 0.88 for CHAMP and 0.89 for

Table 8
Test on GRACE-A: 27 October 2003–03 November 2003

Model	1	2	3	JB-2008	MSIS
R	0.8904	0.8286	0.8809	0.8318	0.7142
RMSE $\times 10^{-13}$	3.0459	5.7379	3.6164	6.6409	10.572
Cov rate	0.9034	0.8631	0.8943		
MACE	0.0472	0.0889	0.0567		

GRACE-A. The correlation coefficients from our base model are 0.92 for CHAMP and 0.89 for GRACE-A. Forootan et al. (2023) also tested their model during the storm period for GRACE-A, yielding an RMSE value of $6.16 \times 10^{-13}\mathrm{kg/m^3}$. The RMSE value from our base model is $3.04 \times 10^{-13}\mathrm{kg/m^3}$, which is notably smaller.

In general, the results in this section indicate that the model containing more satellites in the training database produces superior accuracy and uncertainty estimation results. Specifically, CHAMP and GRACE-A data can be used together to enhance the model's predictive capabilities. The results of the two test cases demonstrate the effectiveness of the proposed framework in predicting thermospheric density for both CHAMP and GRACE-A, even when the test periods are not included in the training data.

3.2.2. Test-2: From 03 April 2010 to 07 April 2010

In Test-2, we will evaluate Model-1, 2, and 3 which are trained with information from CHAMP and GRACE-A, and Model-4 which contains additional 1-month information from GOCE, and Model-1, 2, 3, and 4 will be applied for Test-2 on both CHAMP and GOCE from 03 April 2010 to 10 April 2010. Note the period in Test-2 is the future time of the training range in the four models.

The numerical results of CHAMP are presented in Table 9. The best accuracy performance in this test case is from Model-4, whose training database contains information from three satellites. It has the highest R value of 0.9155 with the smallest RMSE value of $2.1680 \times 10^{-12}\mathrm{kg/m^3}$. The second-best result is from Model-1, whose training database contains information from two satellites, with an R value of 0.8787 and an RMSE value of $2.2322 \times 10^{-12}\mathrm{kg/m^3}$. The performance in Model-2 is close to that of Model-1, as both the training and test data are along the CHAMP satellite. Its R value reaches 0.865, with RMSE value of $2.3361 \times 10^{-12}\mathrm{kg/m^3}$. Model-3, which contains the density values from GRACE-A provided by Mehta et al. (2017) and Mehta (2017) in the training database, which is not the test satellite, shows the worst performance in this test case. The R value is 0.8049, and the RMSE value is bigger than the others. The accuracy of the four proposed models in this case is better than the two empirical models.

The best uncertainty estimation performance in this test case is from Model-4, as it provides a reasonable coverage rate with the smallest MACE value. The coverage rate in Model-1 is close to that of Model-4 but slightly lower, and the MACE value in Model-1 is larger. The coverage rate in Model-2 is very high as it goes over 99%. However, we note that this is because the uncertainty value provided by the model is large when the test period is out of the training range. The large MACE value indicates the model's uncertainty prediction is not as reliable as Model-1 or Model-4. The coverage rate in Model-3 is 76%, which means the predicted uncertainty boundaries cannot cover the true values.

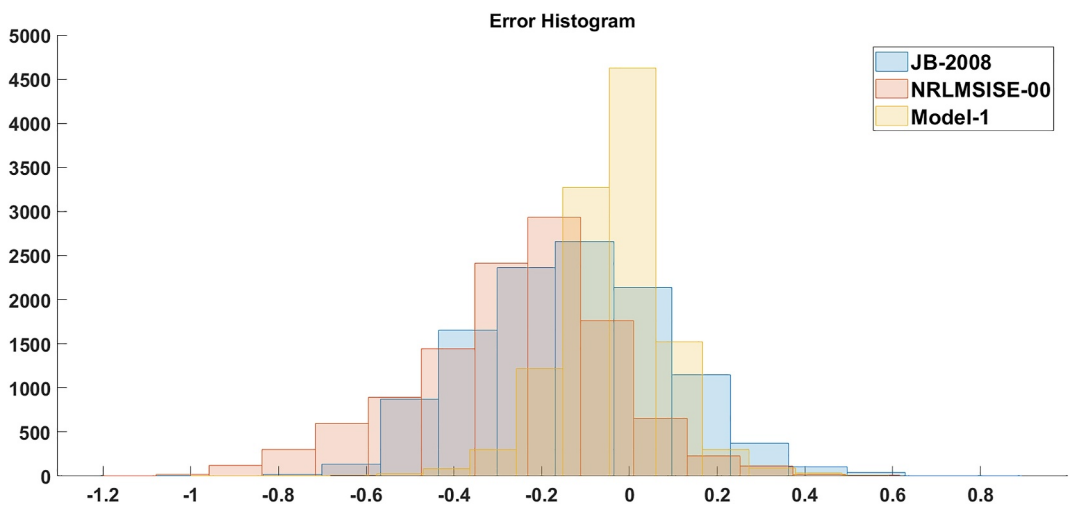

Figure 1. Test-1 on CHAMP: Error histogram.

Table 9
Test on CHAMP: 03 April 2010–07 April 2010

Model	1	2	3	4	JB-2008	MSIS
R	0.8787	0.8657	0.8049	0.9155	0.7714	0.5799
RMSE $\times 10^{-12}$	2.2322	2.3361	3.1686	2.1680	3.9063	4.2667
Cov rate	0.8527	0.9980	0.7656	0.8584		
MACE	0.0876	0.2752	0.1287	0.0554		

Table 10 shows the results when testing on GOCE from 03 April 2010 to 07 April 2010. Model-4 shows the most accurate predictions among all the models as it has the highest R value of 0.9494 and the smallest RMSE value of 2.8724×10^{-12}kg/m^3. The second-best performance on accuracy is from Model-1, which contains two satellites. The R value in Model-1 is 0.9449, which is very close to the value in Model-4. The RMSE value in Model-1 is 3.6339×10^{-12}kg/m^3. The performances of Model-2 and Model-3 are worse than Model-1 and Model-4. The R value on Model-2 is 0.8657, while on Model-3 R value is reduced to 0.8049. The RMSE values are also much larger than the values in Model-1 or Model-4. The predicted density from the JB-2008 and NRLMSISE-00 model in this test case shows a larger R value than Model-3, while the RMSE values are larger than all the proposed models. The same test case for GOCE has been studied in Forootan et al. (2023) with reported RMSE value of 4.29×10^{-12}kg/m^3, which is larger than the RMSE value achieved in the base model.

Model-4 performs exceptionally well for both test cases when we look at the uncertainty performance. The coverage rate in Model-4 reaches 97.79%, which is the closest to the ideal 2σ coverage rate, and the MACE value in Model-4 is the smallest, which indicates the reliability of the uncertainty estimation. The coverage rate in Model-1 is slightly larger than Model-4, as it reaches 98.09%, thus with a larger MACE value. Model-2 in this case shows the largest coverage rate of 99.81%, which is beyond the criteria of the perfect calibration curve, due to the uncertainty values from Model-2 being too large that they are overestimated. The MACE value in Model-2 is the largest, indicating less reliability of the uncertainty estimations. Model-3 in this case shows the smallest coverage rate and a larger MACE value than the values in Model-1 and Model-4.

The R metric can evaluate the strength of the connections between the predicted values and the ground truth based on overall performance. This allows for a general comparison across the same satellites for a closely matched test period. For example, Chen and Sang (2016) reported R values ranging from 0.81 to 0.89 for CHAMP density predictions during quiet periods. In contrast, our base model, tested during storm periods, shows higher R values, indicating the improved performance of the proposed model.

3.2.3. Test-3: From 11 October 2010 to 12 October 2010

Test-3 is from 11 October 2010 to 12 October 2010, and we will evaluate the performance of Model-1, 2, 3, and 4 on GRACE-A and GOCE. The numerical results of GRACE-A are presented in Table 11 and the results of GOCE are shown in Table 12.

For the test results on GRACE-A in Table 11, all four models provide reasonable predicted results as the R values exceed 0.87. The best result on the prediction accuracy is from Model-4, which has the highest R value of 0.9027, and the smallest RMSE value of 4.6180×10^{-14}kg/m^3. The performances of Model-1 and Model-3 are very close. Model-1 shows a larger R value while the RMSE value in Model-3 is smaller. The worst performance is from Model-2, in which R value is 0.8700 and the RMSE value is 6.7014×10^{-14}kg/m^3 since the training database in Model-2 contains information only from CHAMP. In this case, the empirical models show much larger RMSE values than the proposed models indicating a worse density estimation than the models based on the proposed framework.

On the other hand, for the uncertainty estimation, Model-4 has the most reasonable coverage rate, which is 93.84% with the smallest MACE value of 0.0268. This indicates that its uncertainty estimates are highly reliable. The second-best performance on uncertainty estimation in this case is from Model-1, as it provides the second-highest coverage rate 85.58% with the MACE value of 0.0544. The coverage rate on Model-3 is 82.57% and the MACE value is 0.0784. The metrics in Model-3 are slightly worse than the values in Model-1.

Table 10
Test on GOCE: 03 April 2010–07 April 2010

Model	1	2	3	4	JB-2008	MSIS
R	0.9449	0.8657	0.8049	0.9494	0.8717	0.8309
RMSE $\times 10^{-12}$	3.6339	4.1231	4.5297	2.8724	4.6833	4.9274
Cov rate	0.9809	0.9981	0.9274	0.9737		
MACE	0.0365	0.2252	0.0532	0.0207		

As for the results on GOCE in Table 12, the best-predicted value is from Model-4, which contains information from CHAMP and GRACE-A and a short period on GOCE. The R value is 0.8985, and the RMSE value is 2.3903×10^{-12}kg/m^3. The second best is from Model-1, with the R value of 0.8627 and the RMSE value of 2.5373×10^{-14}kg/m^3. The results on Model-2 and Model-3 are worse. The R value in Model-2 is 0.8542, and the RMSE value

Table 11
Test on GRACE-A: 11 October 2010–12 October 2010

Model	1	2	3	4	JB-2008	MSIS
R	0.9015	0.8700	0.9008	0.9027	0.8388	0.7649
RMSE $\times 10^{-14}$	4.9420	6.7014	4.6504	4.6180	11.918	13.609
Cov rate	0.8558	1.0000	0.8257	0.9384		
MACE	0.0544	0.2887	0.0784	0.0268		

increases to 3.5717 $\times 10^{-12}$kg/m^3. The accuracy in Model-3 is the worst, with the smallest R value of 0.8306 and the RMSE value is 3.8469 $\times 10^{-12}$kg/m^3. In this case, the R value of the JB-2008 model is larger than the value in Model-2 and Model-3, while the RMSE value in the JB-2008 model is larger than the proposed models. Besides this, the models based on the proposed framework show more optimal accuracy predictions than the two empirical models. A comparison of the results based on GOCE data is referenced in P. Wang et al. (2022), using the RMSE value as a benchmark. In their study, the RMSE for GOCE was reported as 4.849 $\times 10^{-12}$kg/m^3. This value is slightly larger than the results obtained from our base model.

The uncertainty estimation in Model-4 is the best, as the coverage rate of 95.42% is the most reasonable value. The smallest MACE value 0.0113 indicates the reliability of the uncertainty predictions. The coverage rates from Model-1 and Model-3 are 93% and 91%, with small MACE values of 0.02 and 0.01. For Model-2, the uncertainty values are very high, resulting in overly wide uncertainty bounds and coverage of 100%, and the corresponding MACE value is as high as 0.3432. Therefore, the uncertainty estimation in Model-2 is poor because of the lack of information.

As a summary of Test 2 and Test 3, for the two test periods in 2010 on CHAMP, GRACE-A, and GOCE, Model 4 consistently outperforms other models, demonstrating its outstanding performances on accuracy and the reliability of uncertainty estimations for the test cases. The results also indicate that including additional satellite information, even a short period, contributes to the effectiveness of capturing the complexities of thermospheric density variations. On the other hand, our base model, Model-1, shows the second-best performance for the test cases in 2010, with better accuracy than the two empirical models. Based on the proposed framework, the base model can predict the density along different satellites with high accuracy and reliable uncertainty estimation, even if the test satellite is not included in the training database.

3.2.4. Test-4: On 20 February 2014

For this case, we evaluate the models' performances on satellites SWARM-A and SWARM-C on 20 February 2014. We note that the database provided by ESA does not cover this period for SWARM-B, thus it is not included in this test. Model-1, 2, 3, and 5 will be applied and tested on SWARM satellites.

The results of SWARM-A are presented in Table 13. The most accurate predicted results in this test case are from Model-5, whose training database contains information from CHAMP, GRACE-A, and 15-day's information from SWARM-A. The R value in Model-5 is 0.8907, which is the highest, with the RMSE value of 3.0703 $\times 10^{-13}$kg/m^3. The second best model is Model-1, which contains information including satellites CHAMP and GRACE-A. The R value in Model-1 is 0.8512, and the RMSE value is 3.1736 $\times 10^{-13}$kg/m^3. The prediction results of Model-2 and Model-3 are relatively worse. The RMSE in Model-3 is larger than the value from the JB-2008 model. Except for this, the accuracy metrics of the three models based on the proposed framework are better than the values of the empirical models.

As for the estimated uncertainty, the coverage rates in the four models are all higher than 93%, which means most of the truth can be covered within the uncertainty boundaries. The best performance is from Model-5 with the highest coverage rate of 97.23% and the smallest MACE value 0.0410. The uncertainty performance of Model-1, 2, and 3 are close, but the results in Model-1 are optimal than the results in the other two models.

Table 12
Test on GOCE: 11 October 2010–12 October 2010

Model	1	2	3	4	JB-2008	MSIS
R	0.8627	0.8542	0.8306	0.8985	0.8564	0.7783
RMSE $\times 10^{-12}$	2.5373	3.5714	3.8469	2.3903	4.7608	6.2180
Cov rate	0.9305	1.0000	0.9158	0.9542		
MACE	0.0239	0.3432	0.0437	0.0113		

In the case of SWARM-C, as shown in Table 14, the best R value 0.9149 is from Model-5 with the smallest RMSE value of 1.3438 $\times 10^{-13}$kg/m^3. The results in Model-1 are slightly worse than Model-5, but the R value in Model-1 can reach 0.8719, and the RMSE value is 1.5889 $\times 10^{-13}$kg/m^3. The RMSE values in Model-3 in this case is larger than the value from the MSIS model.

As for the uncertainty prediction performance, the coverage rate in Model-5 is not the highest, but it shows the smallest MACE value, indicating the reliability of the uncertainty estimation. The coverage rates from Model-1 and

Table 13
Test on SWARM-A: 20 April 2014

Model	1	2	3	5	JB-2008	MSIS
R	0.8512	0.8479	0.8167	0.8970	0.7851	0.7856
RMSE $\times 10^{-13}$	3.1736	3.3017	4.5243	3.0703	4.4510	3.5465
Cov rate	0.9485	0.9378	0.9460	0.9723		
MACE	0.0602	0.0613	0.0684	0.0410		

Model-2 are close as both reach 98%. The MACE value in Model-1 is slightly larger than the value of Model-4 but smaller than the value of Model-2. Thus, the uncertainty estimation from Model-1 is the second best one. The coverage rate of Model-3 is the smallest, and the MACE value is larger than Model-1 and Model-5.

3.2.5. Test-5: From 15 March 2015 to 26 March 2015

We evaluate Model-1, 2, 3, and 5 from 15 March 2015 to 26 March 2015 on SWARM-B and SWARM-C. Both satellites in this test case are not among the training satellites and the test period is in the future time of the training periods. The numerical results of SWARM-B are presented in Table 15 and the results of SWARM-C are shown in Table 16.

When we look at the predicted values, Model-5 provides the highest R value, with the RMSE value of 1.4993×10^{-13}kg/m^3. Model-1 shows the second largest R value among all the models, while it gives the smallest RMSE value as 1.2350×10^{-13}kg/m^3. Model-3 provides the third-best results with an R value of 0.8078 and an RMSE value of 1.5438×10^{-13}kg/m^3. The worst prediction is from Model-2 as the R value is 0.7875, with the largest RMSE value of 3.2018×10^{-13}kg/m^3 among all the proposed models. The JB-2008 model in this case shows better estimations than Model-2. Except for this, the accuracy of the two empirical models is worse than the models based on the framework. Fernandez-Gomez et al. (2022) also studied this storm and reported an R value of 0.82. In comparison, our base model demonstrates a higher correlation coefficient of 0.84, indicating improved predictive accuracy.

Model-5 also shows advantages in the uncertainty estimations as the coverage rate in Model-5 is beyond 95% with the MACE value as 0.0410. The result suggests a high level of reliability and accuracy in the model's ability to estimate uncertainties associated with its predictions. Model-1 in this case shows the second-best uncertainty estimations. The coverage rate in Model-1 is 87.35% with the MACE value of 0.0991. Model-2 and Model-3 in this condition fail to produce reasonable uncertainty estimations.

The results on SWARM-C are presented in Table 16. Model-5 shows the most accurate predictions in this test case as the R value is 0.8044 with RMSE value 2.7533×10^{-13}. The predicted results from Model-1 and Model-5 are very close, as the R values are close to 0.80 with RMSE values around 2.8×10^{-13}. The predicted errors from Model-2 and Model-3 are more obvious in this case since the RMSE values in the two models are much larger.

Model-1 and Model-5 can provide reasonable uncertainty estimations as the coverage rate exceeds 80%. Model-5 shows the best results with 88% of the coverage rate and the MACE value of 0.0258. The uncertainty estimation in Model-2 and Model-3 cannot be considered reasonable. The coverage rate in Model-2 is much lower and the rate in Model-3 is much higher than being reasonable. The corresponding MACE values for the two models are also much larger.

In general, based on the analysis of SWARM satellites in Test-3, 4, and 5, it is evident that Model-5, trained with multiple satellites, consistently outperforms other models. On the other hand, our base model, Model-1, which is trained with the information from two satellites regularly performs well, especially showing better accuracy than the empirical models JB-2008 and NRLMSISE-00. Model-1 also demonstrates robust performances even in the absence of test satellite data. Comparatively, the performances of Model-2 and Model-3 lag behind those of Model-1 and Model-5.

Table 14
Test on SWARM-C: 20 April 2014

Model	1	2	3	5	JB-2008	MSIS
R	0.8719	0.8637	0.8315	0.9149	0.8047	0.7993
RMSE $\times 10^{-13}$	1.5889	2.1863	2.8397	1.3438	3.3785	2.6075
Cov rate	0.9825	0.9829	0.9693	0.9533		
MACE	0.0564	0.0645	0.0611	0.0417		

The critical insight here is the utilizing more satellite data in constructing the global model leads to better overall performances. Furthermore, the inclusion of even a small amount of the test satellite data in the training database contributes to performance improvement.

4. Global Density Prediction

To further demonstrate the capability of the proposed framework for predicting thermospheric density globally, we generate density at different geographic locations on various altitudes from Model-1, which is the base

Table 15
Test on SWARM-B: 15 March 2015–26 March 2015

Model	1	2	3	5	JB-2008	MSIS
R	0.8475	0.7875	0.8078	0.8627	0.8086	0.5852
RMSE $\times 10^{-13}$	1.2350	3.2018	1.5438	1.4992	1.7148	1.8803
Cov rate	0.8735	0.8087	0.8098	0.9561		
MACE	0.0991	0.1158	0.1195	0.0410		

model. We compare the predicted density with the density obtained from the empirical models JB-2008 and NRLMSISE-00 globally, and also for specific locations where the accelerometer-derived density data is available.

4.1. Test on 01 July 2007, 00:00 UTC

The test period is on 01 July 2007, at 00:00 UTC, during which the Dst value is −3 nT, indicating a quiet period. We set the spatial resolution to 5° in both latitude and longitude. Altitude values are considered in the range of 200–550 km, with increments set at 50 km intervals. We aim to demonstrate the model's ability to predict global thermospheric density across a wide range of geophysical settings.

The distribution of the two empirical models, along with the predicted values from Model-1 are plotted in Figures 2 and 3. The first column is the estimated density from the JB-2008 model at each altitude. The second column presents the estimated value derived from the NRLMSISE-00 model. The final column shows the estimated density from Model-1. The color bars of the density values for each altitude are shown on the right side of each row.

From Figures 2 and 3 we can see that once the altitudes are higher than 250 km, the patterns from the three models show similar patterns. Model 1 is not reliable for very low altitudes as the altitudes of the training satellites are in the range of 350–528 km. Overall, the results in this case show the model's ability to produce the thermospheric density across a broad altitude range with reasonable values.

To analyze the differences between the density estimated from Model-1 and the two empirical models, we calculate the mean and standard deviation of the difference between the predicted density from Model-1 and the values from the two empirical models as shown in Table 17. The mean values of the differences between Model-1 and JB-2008 are found to be smaller compared to the differences between Model-1 and the NRLMSISE-00 model. This is consistent with the previous results that the JB-2008 model in the test cases usually provides higher R values and smaller RMSE values.

Although there is no true density for calculating the real prediction errors globally, fortunately, densities derived from the accelerometers of CHAMP and GRACE-A exist. We refer to the density from CHAMP and GRACE-A at the same time and compare it with the density derived from Model-1 and the two empirical models. The results are shown in Table 18. The difference between the CHAMP and our base model is 2.812×10^{-13} with a relative error of 19.11%. The difference between CHAMP and the JB-2008 model is 3.858×10^{-13}, and the corresponding relative error is 26.22%. The difference between CHAMP and NRLMSISE-00 model is 5.665×10^{-13}, and the relative error is 38.51%. As for the predicted results on GRACE-A, the difference of the base model is 3.982×10^{-14}, and the relative error in this case is 28.96%. The difference of the JB-2008 model is 9.229×10^{-14} with the relative error of 67.12%. The difference of the NRLMSISE-00 model is 1.0356×10^{-13} and the corresponding relative error is 75.31%.

The numerical results indicate that for this quiet condition and at the two locations where we have satellite data, the density from our base model is closer to the density derived from the satellite accelerometers than the two empirical models.

Table 16
Test on SWARM-C: 15 March 2015–26 March 2015

Model	1	2	3	5	JB-2008	MSIS
R	0.7918	0.7871	0.5359	0.8044	0.7843	0.5980
RMSE $\times 10^{-13}$	2.8692	10.622	29.216	2.7553	3.3709	3.6451
Cov rate	0.8273	0.2013	1.0000	0.8807		
MACE	0.1378	0.5445	0.2758	0.0258		

4.2. Test on 17 March 2015, 22:00 UTC

We test another case on 17 March 2015, at 22:00 UTC. A storm was happening at this time with a Dst value of −234 nT. The altitude values span from 200 to 550 km with a resolution of 50 km. The distributions are plotted in Figures 4 and 5.

The overall density values in this case are higher compared to the previous case in 2007, attributable to the impact of the ongoing storm. The JB-2008 model, in this case, shows similar patterns at different altitudes. NRLMSISE-00 model shows distinctive patterns at low altitudes from 200 to 300 km. As the altitude increases, the patterns from the NRLMSISE-00 model

10.1029/2024SW004070

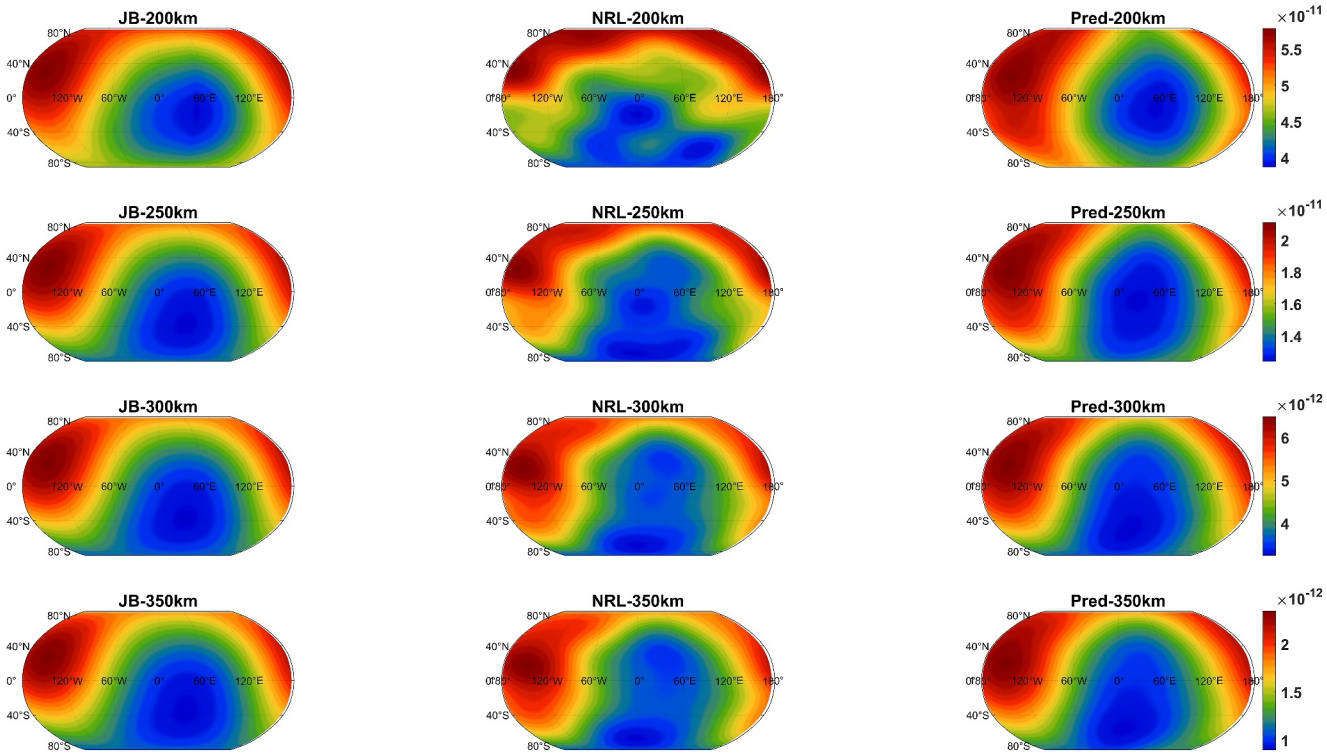

Figure 2. Global Density Prediction on 01 July 2007, 00:00 UTC: altitude from 200 to 350 km.

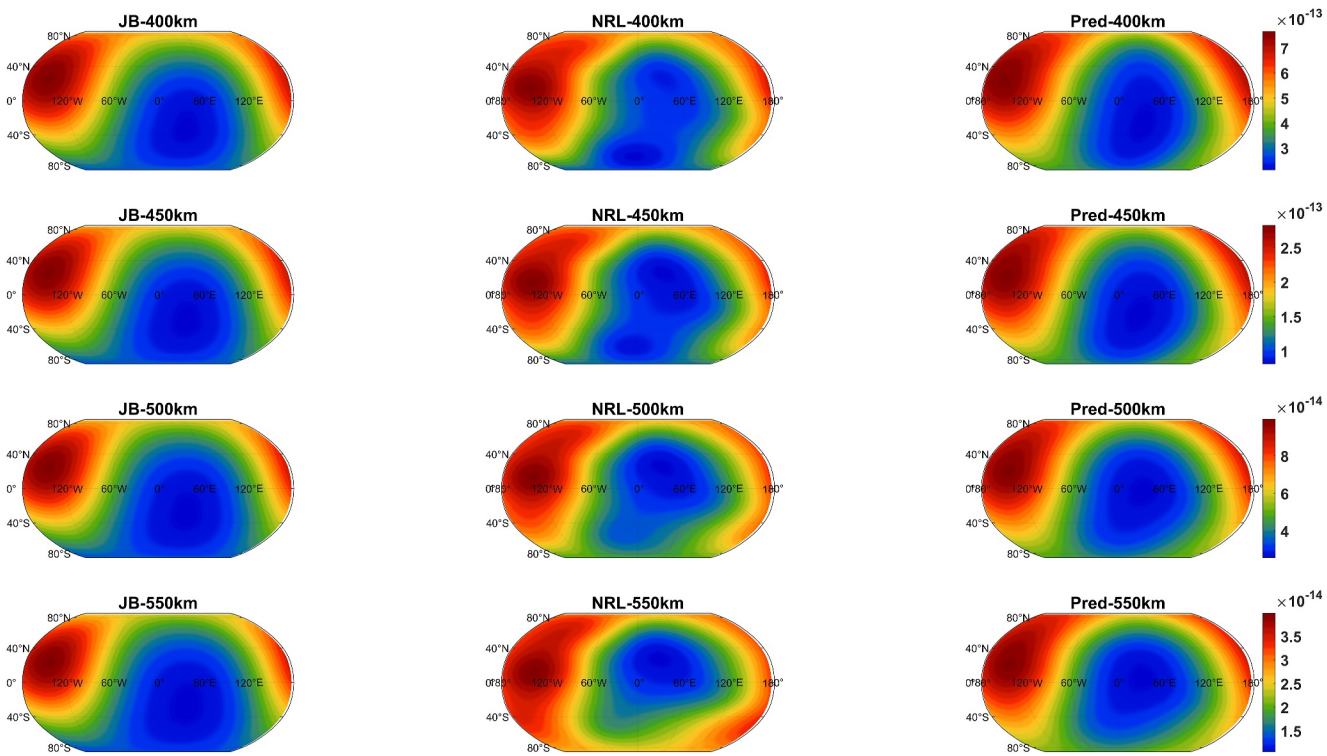

Figure 3. Global Density Prediction on 01 July 2007, 00:00 UTC: altitude from 400 to 550 km.

Table 17
The Features of Density Differences (With Unit of kg/m^3) on 01 July 2007

Altitude (km)	JB-diff-mean	JB-diff-std	MSIS-diff-mean	MSIS-diff-std
200	6.7967×10^{-11}	8.9243×10^{-12}	9.9220×10^{-11}	1.6811×10^{-11}
250	9.6863×10^{-12}	2.1328×10^{-12}	9.7681×10^{-12}	2.7884×10^{-12}
300	1.4965×10^{-12}	5.1507×10^{-13}	2.0869×10^{-12}	6.3484×10^{-13}
350	1.7928×10^{-13}	1.1913×10^{-13}	3.6402×10^{-13}	1.6588×10^{-13}
400	7.8382×10^{-14}	4.8324×10^{-14}	1.5159×10^{-13}	7.0074×10^{-14}
450	7.7768×10^{-15}	2.1557×10^{-14}	4.1063×10^{-14}	3.0714×10^{-14}
500	5.9841×10^{-15}	1.0301×10^{-14}	2.2409×10^{-14}	1.2806×10^{-14}
550	1.7810×10^{-15}	4.9541×10^{-15}	1.0497×10^{-14}	6.1891×10^{-15}

become similar. The NRLMSISE-00 model also shows that when the altitude is above 300 km, density near the equator in the eastern hemisphere is greater than in the upper and lower halves of the east. Furthermore, both empirical models consistently indicate that the density in the western hemisphere is larger than the density in the eastern hemisphere, which is also revealed in Model-1.

Table 19 presents the mean and standard values of the differences between the two empirical models and Model-1, in the unit of kg/m^3. The difference at various altitudes is at reasonable magnitudes. The difference based on the JB-2008 model and Model-1 is slightly smaller than the differences between the NRLMSISE-00 model and Model-1.

Again, although there are no true density values to calculate the prediction errors globally, fortunately, we have two accelerometer-derived density data at this time. To further verify the effectiveness of Model-1, we obtain the density derived from the accelerometers of SWARM-B and SWARM-C at the same time, and compare it with the density from Model-1 and the two empirical models. The results are shown in Table 20. The difference between SWARM-B and our based model is 2.689×10^{-14} with the relative error of 0.1234%. The difference between SWARM-B and JB-2008 is 1.4250×10^{-12} with the relative error of 65.45%. The difference between SWARM-B and the NRLMSISE-00 model is 1.6161×10^{-12} with the relative error of 74.22%. The difference between SWARM-C and Model-1 is 1.978×10^{-13}, and the relative error value is 11.88%. The difference between SWARM-c and JB-2008 model is 3.317×10^{-13} with the relative error of 19.92%. The difference between SWARM-C and NRLMSISE-00 model is 1.4922×10^{-12}, and the relative error is 89.64%.

Therefore, for this storm condition and at the two locations where we have satellite data, the estimated densities from Model-1 are much closer to the densities derived from satellite accelerometers than the empirical models. Furthermore, the improvements of Model-1 from the empirical models are more significant than in the quiet time.

5. Conclusion

This paper proposes a global thermospheric density prediction framework based on a deep evidential method. The framework incorporates solar and geomagnetic indices, empirical models including JB-2008 and NRLMSISE-00, and accelerometer-derived satellite density data. In the paper, five models based on different training databases

Table 18
Predicted Density (With Unit of kg/m^3) Compared on 01 July 2007

Satellite position	CHAMP	Model-1	JB	MSIS
[83.7, 118, 353.5]	1.4710×10^{-12}	1.1898×10^{-12}	1.8568×10^{-12}	2.0375×10^{-12}
Relative error		19.11%	26.22%	38.51%
Satellite position	GRACE-A	Model-1	JB	MSIS
[51.1, −178.1, 463.9]	1.3750×10^{-13}	1.7732×10^{-13}	2.2979×10^{-13}	2.4106×10^{-13}
Relative error		28.96%	67.12%	75.31%

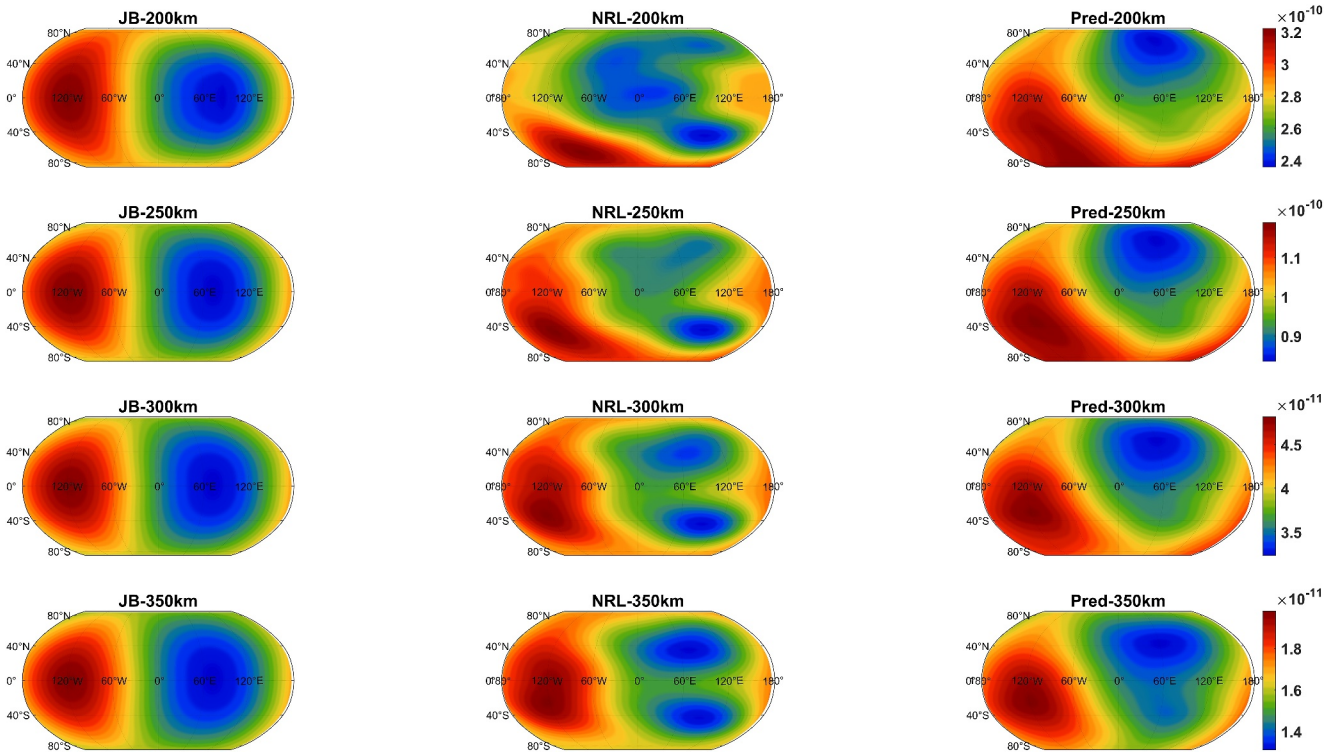

Figure 4. Global Density prediction on 17 March 2015, 22:00 UTC, altitude from 200 to 350 km.

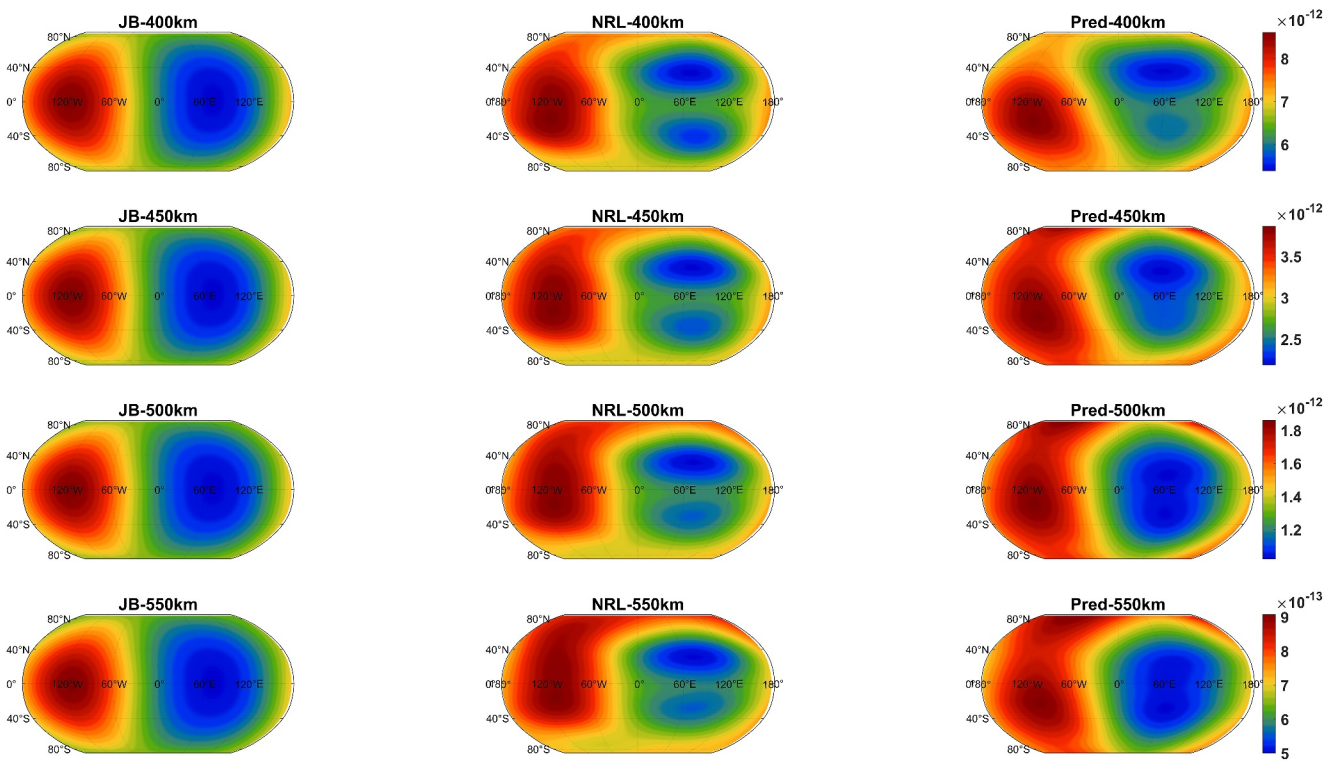

Figure 5. Global Density prediction on 17 March 2015, 22:00 UTC, altitude from 400 to 550 km.

Table 19
The Features of Density (With Unit of kg/m^3) Differences on 17 March 2015

Altitude (km)	JB-diff-mean	JB-diff-std	MSIS-diff-mean	MSIS-diff-std
200	$6.6045.6 \times 10^{-11}$	1.8989×10^{-11}	5.6351×10^{-11}	1.8499×10^{-11}
250	6.5952×10^{-12}	6.9760×10^{-12}	1.2337×10^{-11}	6.0674×10^{-12}
300	7.0165×10^{-12}	2.4405×10^{-13}	1.0347×10^{-11}	2.3456×10^{-12}
350	2.9729×10^{-12}	7.4355×10^{-13}	4.8577×10^{-12}	7.0420×10^{-13}
400	1.2300×10^{-12}	3.8548×10^{-13}	2.2764×10^{-12}	3.0299×10^{-13}
450	4.8039×10^{-13}	2.3952×10^{-13}	1.0536×10^{-12}	1.8342×10^{-12}
500	1.9532×10^{-13}	1.3486×10^{-13}	5.0682×10^{-13}	9.7796×10^{-14}
550	8.7870×10^{-14}	9.6215×10^{-14}	2.5652×10^{-13}	4.9876×10^{-14}

are constructed and evaluated on different satellites including CHAMP, GRACE-A, GOCE, and SWARM satellites.

For a comprehensive evaluation, we first select five test cases from 2003 to 2015, during which geomagnetic storms happened. The results demonstrate that our proposed framework can generate thermospheric density with both high accuracy and reliable uncertainty estimations, even in cases where the test periods are in the future time of the training database. The base model trained with CHAMP and GRACE-A data, Model-1 in this paper, consistently achieves more accurate thermospheric density predictions compared to the empirical models including JB-2008 and NRLMSISE-00. Furthermore, the experiment results reveal that incorporating more satellite data into the training database enhances the performance of the prediction, even if the amount of additional data is limited.

Additionally, we assess the performance of the proposed framework in predicting densities across global geographic positions at various altitudes ranging from 200 to 550 km. The comparisons include the estimated density from the base model whose training data are only from two satellites, and the JB-2008 and NRLMSISE-00 models during both a quiet time on January 2007, 00:00 UTC, and a storm time on 15 March 2015, 00:00 UTC. For the quiet time on 01 July 2007, the patterns from the three models are similar. We refer to the density from the accelerometer of CHAMP and GRACE-A at the same time and compare the density from the base model and the two empirical models. The relative errors from the base model are smaller than the two empirical models on both satellites, indicating the density from the base model is more accurate than the two empirical models for this case. During the storm time on 15 March 2015, the global density pattern from the base model is different from the two empirical models. We validate the effectiveness of the proposed model by comparing with the densities from two SWARM satellites at the same time. The results indicate that the estimated density values from the base model based on the proposed framework are again closer to the densities from satellite accelerometers than the two empirical models.

Table 20
Predicted Density (With Unit of kg/m^3) Compared With SWARM Satellites

Satellite position (deg, deg, km)	SWARM-B	Model-1	JB	MSIS
[31.2, −32.9, 457]	2.1772×10^{-12}	2.2041×10^{-12}	7.5218×10^{-13}	5.6106×10^{-13}
Relative error		0.1234%	65.45%	74.22%
Satellite position (deg, deg, km)	SWARM-C	Pred	JB	MSIS
[−72.9, 175, 536]	1.6646×10^{-12}	1.4668×10^{-12}	1.9963×10^{-12}	1.7237×10^{-12}
Relative error		11.88%	19.92%	89.64%

Data Availability Statement

All the data we use is from publicly available websites. The empirical models JB-2008 and NRLMSISE-00 can be accessed from the open-source MATLAB code (Mahooti, 2022a, 2022b). The space weather indices $F_{10.7}$, $F_{10.7\ A}$, Ap, and F_{30} are obtained from Kelso (2024). We refer to Cho (2024) for accessing the *Dst* data and obtain the *SymH* data from ISGI (2024). The density databases derived from the accelerometers of the satellites are collected from various sources. We use the information from CHAMP, GRACE-A provided by Mehta et al. (2017) and Mehta (2017). We also use the information from GOCE (ESA, 2024a) and SWARM (ESA, 2024b) provided by the European Space Agency (ESA).

Acknowledgments
The research has been supported by the National Science Foundation, United States, under Award 2149747, and NASA, United States, under Award 80NSSC24K0843.

References

Aaron, R., Josh, P., & Jack, W. (2024). Global ionosphere thermosphere model (GITM) [Website]. *Community Coordinated Modeling Center (CCMC)*. Retrieved from https://ccmc.gsfc.nasa.gov/models/GITM~21.11/

Acciarini, G., Brown, E., Berger, T., Guhathakurta, M., Parr, J., Bridges, C., & Baydin, A. G. (2024). Improving thermospheric density predictions in low-earth orbit with machine learning. *Space Weather*, *22*(2), e2023SW003652. https://doi.org/10.1029/2023sw003652

Amini, A., Schwarting, W., Soleimany, A., & Rus, D. (2020). Deep evidential regression. *Advances in Neural Information Processing Systems*, *33*, 14927–14937.

Anderson, G. J., Gaffney, J. A., Spears, B. K., Bremer, P.-T., Anirudh, R., & Thiagarajan, J. J. (2020). Meaningful uncertainties from deep neural network surrogates of large-scale numerical simulations. arXiv preprint arXiv:2010.13749.

Bonasera, S., Acciarini, G., Pérez-Hernández, J., Benson, B., Brown, E., Sutton, E., et al. (2021). Dropout and ensemble networks for thermospheric density uncertainty estimation. In *Bayesian deep learning workshop*. Neurips.

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* (p. 6438).

Burke, W. J., Huang, C. Y., Marcos, F. A., & Wise, J. O. (2007). Interplanetary control of thermospheric densities during large magnetic storms. *Journal of Atmospheric and Solar-Terrestrial Physics*, *69*(3), 279–287. https://doi.org/10.1016/j.jastp.2006.05.027

Burns, A., Killeen, T., Deng, W., Carignan, G., & Roble, R. (1995). Geomagnetic storm effects in the low to middle latitude upper thermosphere. *Journal of Geophysical Research*, *100*(A8), 14673–14691. https://doi.org/10.1029/94ja03232

Burns, A., Killeen, T., Wang, W., & Roble, R. (2004). The solar-cycle-dependent response of the thermosphere to geomagnetic storms. *Journal of Atmospheric and Solar-Terrestrial Physics*, *66*(1), 1–14. https://doi.org/10.1016/j.jastp.2003.09.015

Chartier, A. T., Jackson, D. R., & Mitchell, C. N. (2013). A comparison of the effects of initializing different thermosphere-ionosphere model fields on storm time plasma density forecasts. *Journal of Geophysical Research: Space Physics*, *118*(11), 7329–7337. https://doi.org/10.1002/2013ja019034

Chen, J., & Sang, J. (2016). Thermospheric mass density measurement from precise orbit ephemeris. *Geodesy and Geodynamics*, *7*(3), 210–215. https://doi.org/10.1016/j.geog.2016.05.004

Cho, K.-O. (2024). Geomagnetic equatorial Dst index home page [Database]. *World Data Center for Geomagnetism*. Retrieved from https://wdc.kugi.kyoto-u.ac.jp/

Elvidge, S., Godinez, H. C., & Angling, M. J. (2016). Improved forecasting of thermospheric densities using multi-model ensembles. *Geoscientific Model Development*, *9*(6), 2279–2292. https://doi.org/10.5194/gmd-9-2279-2016

Emmert, J. (2009). A long-term data set of globally averaged thermospheric total mass density. *Journal of Geophysical Research*, *114*(A6), A06315. https://doi.org/10.1029/2009ja014102

ESA. (2024a). ESA GOCE online dissemination: GOCE thermosphere data [Database]. *European Space Agency*. Retrieved from https://goce-ds.eo.esa.int/oads/access/collection/GOCE_Thermosphere_Data

ESA. (2024b). Swarm data - ESA earth online [Database]. *European Space Agency*. Retrieved from https://earth.esa.int/eogateway/missions/swarm/data

Fernandez-Gomez, I., Kodikara, T., Borries, C., Forootan, E., Goss, A., Schmidt, M., & Codrescu, M. V. (2022). Improving estimates of the ionosphere during geomagnetic storm conditions through assimilation of thermospheric mass density. *Earth Planets and Space*, *74*(1), 121. https://doi.org/10.1186/s40623-022-01678-3

Forootan, E., Farzaneh, S., Kosary, M., Borries, C., Kodikara, T., & Schumacher, M. (2023). Predicting global thermospheric neutral density during periods with high geomagnetic activity. *Scientific Reports*, *13*(1), 20322. https://doi.org/10.1038/s41598-023-47440-x

Forootan, E., Kosary, M., Farzaneh, S., Kodikara, T., Vielberg, K., Fernandez-Gomez, I., et al. (2022). Forecasting global and multi-level thermospheric neutral density and ionospheric electron content by tuning models against satellite-based accelerometer measurements. *Scientific Reports*, *12*(1), 2095. https://doi.org/10.1038/s41598-022-05952-y

Gao, T., Peng, H., & Bai, X. (2020). Calibration of atmospheric density model based on Gaussian processes. *Acta Astronautica*, *168*, 273–281. https://doi.org/10.1016/j.actaastro.2019.12.014

Gondelach, D. J., & Linares, R. (2020). Real-time thermospheric density estimation via two-line element data assimilation. *Space Weather*, *18*(2), e2019SW002356. https://doi.org/10.1029/2019sw002356

Gondelach, D. J., & Linares, R. (2021). Real-time thermospheric density estimation via radar and GPS tracking data assimilation. *Space Weather*, *19*(4), e2020SW002620. https://doi.org/10.1029/2020sw002620

Gonzalez, W., Joselyn, J.-A., Kamide, Y., Kroehl, H. W., Rostoker, G., Tsurutani, B., & Vasyliunas, V. (1994). What is a geomagnetic storm? *Journal of Geophysical Research*, *99*(A4), 5771–5792. https://doi.org/10.1029/93ja02867

Group, C. W. (2012). *Cospar international reference atmosphere–2012 (Tech. Rep.)*. The Committee on Space Research.

Hall, N. (2021). Earth atmosphere model - Imperial units [Website]. *NASA*. Retrieved from https://www.grc.nasa.gov/www/k-12/airplane/atmos.html

ISGI. (2024). ASY/SYM indices [Database]. *International Service of Geomagnetic Indices*. Retrieved from https://isgi.unistra.fr/playground.php

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

Kelso, T. S. (2024). Space weather data documentation [Database]. *CelesTrak*. Retrieved from https://celestrak.org/SpaceData

Keras Team. (n.d.). Keras. Retrieved from https://keras.io

Licata, R. J., & Mehta, P. M. (2023). Reduced order probabilistic emulation for physics-based thermosphere models. *Space Weather*, *21*(5), e2022SW003345. https://doi.org/10.1029/2022sw003345
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), e2021SW002915. 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
Liu, R., Ma, S.-Y., & Lühr, H. (2011). Predicting storm-time thermospheric mass density variations at champ and grace altitudes. *Annales geophysicae*, *29*(3), 443–453. https://doi.org/10.5194/angeo-29-443-2011
Mahooti, M. (2022a). Jacchia-bowman atmospheric density model [Website]. *The MathWorks Inc*. Retrieved from https://www.mathworks.com/matlabcentral/fileexchange/56163-jacchia-bowman-atmospheric-density-model
Mahooti, M. (2022b). NRLMSISE-00 atmosphere model [Website]. *The MathWorks Inc*. Retrieved from https://www.mathworks.com/matlabcentral/fileexchange/56253-nrlmsise-00-atmosphere-model
Marcos, F. A., Rendra, M., Griffin, J., Bass, J., Larson, D., & Liu, J. (1998). Precision low earth orbit determination using atmospheric density calibration. *Journal of the Astronautical Sciences*, *46*(4), 395–409. https://doi.org/10.1007/bf03546389
Mehta, P. M. (2017). Champ and grace density data sets [Database]. Retrieved from https://drive.google.com/drive/folders/0BwtX8XEH-aEueHJiU1htLVo0cms?resourcekey=0-byxPMLbZSVC5Blxb_Rfl9g
Mehta, P. M., & Linares, R. (2020). Real-time thermospheric density estimation from satellite position measurements. *Journal of Guidance, Control, and Dynamics*, *43*(9), 1656–1670. https://doi.org/10.2514/1.g004793
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
Mutschler, S. M., Axelrad, P., Sutton, E. K., & Masters, D. (2023). Physics-based approach to thermospheric density estimation using CubeSat GPS data. *Space Weather*, *21*(1), e2021SW002997. https://doi.org/10.1029/2021sw002997
Nateghi, V. (2021). *Machine learning-based thermospheric density modeling and estimation for space operations*. Politecnico di Milano - School of Industrial and Information Engineering.
Oliveira, D. M., & Zesta, E. (2019). Satellite orbital drag during magnetic storms. *Space Weather*, *17*(11), 1510–1533. https://doi.org/10.1029/2019sw002287
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
OrbitingNow. (2024). Active satellite orbit data [Dataset]. *orbit.ing-now.com*. Retrieved from https://orbit.ing-now.com/
Pérez, D., & Bevilacqua, R. (2015). Neural network based calibration of atmospheric density models. *Acta Astronautica*, *110*, 58–76. https://doi.org/10.1016/j.actaastro.2014.12.018
Pérez, D., Wohlberg, B., Lovell, T. A., Shoemaker, M., & Bevilacqua, R. (2014). Orbit-centered atmospheric density prediction using artificial neural networks. *Acta Astronautica*, *98*, 9–23. https://doi.org/10.1016/j.actaastro.2014.01.007
Picone, J., Hedin, A., Drob, D. P., & Aikin, A. (2002). NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. *Journal of Geophysical Research*, *107*(A12), SIA15-1–SIA15-16. https://doi.org/10.1029/2002ja009430
Qian, L., & Solomon, S. C. (2012). Thermospheric density: An overview of temporal and spatial variations. *Space Science Reviews*, *168*, 147–173. https://doi.org/10.1007/978-1-4614-5677-3_4
Richardson, I., Webb, D., Zhang, J., Berdichevsky, D., Biesecker, D., Kasper, J., et al. (2006). Major geomagnetic storms (Dst≤-100 nT) generated by corotating interaction regions. *Journal of Geophysical Research*, *111*(A7), A07S09. https://doi.org/10.1029/2005ja011476
UCAR. (2016). The thermospheric general circulation models (TGCM's) [Website]. Retrieved from http://www.hao.ucar.edu/modeling/tgcm
Wang, P., Chen, Z., Deng, X., Wang, J., Tang, R., Li, H., et al. (2022). The prediction of storm-time thermospheric mass density by LSTM-based ensemble learning. *Space Weather*, *20*(3), e2021SW002950. https://doi.org/10.1029/2021sw002950
Wang, Y., & Bai, X. (2022). Comparison of Gaussian processes and neural networks for thermospheric density predictions during quiet time and geomagnetic storms. In *AAS/AIAA astrodynamics specialist conference*.
Wang, Y., & Bai, X. (2023a). Global thermospheric density prediction model based on deep evidential framework. In *AAS/AIAA astrodynamics specialist conference*.
Wang, Y., & Bai, X. (2023b). Thermospheric density predictions during quiet time and geomagnetic storm using a deep evidential model-based framework. *Acta Astronautica*, *211*, 316–325. https://doi.org/10.1016/j.actaastro.2023.06.023
Wang, Y., Peng, H., Bai, X., Wang, J. T., & Wang, H. (2021). Advance thermospheric density predictions through forecasting geomagnetic and solar indices based on Gaussian processes. In *AAS/AIAA astrodynamics specialist conference*.
Weimer, D. R., Bowman, B., Sutton, E., & Tobiska, W. (2011). Predicting global average thermospheric temperature changes resulting from auroral heating. *Journal of Geophysical Research*, *116*(A1), A01312. https://doi.org/10.1029/2010ja015685
Weimer, D. R., Mlynczak, M., Emmert, J., Doornbos, E., Sutton, E., & Hunt, L. (2018). Correlations between the thermosphere's semiannual density variations and infrared emissions measured with the saber instrument. *Journal of Geophysical Research: Space Physics*, *123*(10), 8850–8864. https://doi.org/10.1029/2018ja025668
Weng, L., Lei, J., Zhong, J., Dou, X., & Fang, H. (2020). A machine-learning approach to derive long-term trends of thermospheric density. *Geophysical Research Letters*, *47*(6), e2020GL087140. https://doi.org/10.1029/2020gl087140
Zhang, J., Richardson, I., Webb, D., Gopalswamy, N., Huttunen, E., Kasper, J., et al. (2007). Solar and interplanetary sources of major geomagnetic storms (Dst≤-100 nT) during 1996–2005. *Journal of Geophysical Research*, *112*(A10), A10102. https://doi.org/10.1029/2007ja012321