



## Machine Learning Based Modeling of Thermospheric Mass Density

### Key Points:

- Based on observations from Swarm C and machine learning approach we constructed the MBiLE model for predicting thermosphere mass density
- Despite using only Swarm C data at about 450–500 km altitude, the MBiLE model predictions well with satellite observations within 230–500 km
- Our MBiLE model exhibits superior altitude generalization, exceeding earlier models based on multivariate least-square-fitting approach

### Correspondence to:

C. Xiong,  
xiongchao@whu.edu.cn

### Citation:

Pan, Q., Xiong, C., Lühr, H., Smirnov, A., Huang, Y., Xu, C., et al. (2024). Machine learning based modeling of thermospheric mass density. *Space Weather*, 22, e2023SW003844. <https://doi.org/10.1029/2023SW003844>

Received 21 DEC 2023

Accepted 9 APR 2024

### Author Contributions:

**Conceptualization:** Qian Pan, Yuyang Huang, Chunyu Xu, Yang Hu  
**Data curation:** Qian Pan  
**Formal analysis:** Qian Pan, Hermann Lühr, Artem Smirnov, Yuyang Huang, Chunyu Xu, Yunliang Zhou, Yang Hu  
**Investigation:** Qian Pan  
**Methodology:** Qian Pan, Hermann Lühr, Artem Smirnov, Yuyang Huang, Chunyu Xu, Xu Yang, Yang Hu  
**Supervision:** Hermann Lühr  
**Validation:** Qian Pan, Hermann Lühr, Xu Yang  
**Visualization:** Qian Pan  
**Writing – original draft:** Qian Pan  
**Writing – review & editing:** Qian Pan, Hermann Lühr, Artem Smirnov, Chunyu Xu, Xu Yang, Yunliang Zhou

Qian Pan<sup>1</sup>, Chao Xiong<sup>1</sup> , Hermann Lühr<sup>2</sup>, Artem Smirnov<sup>2,3</sup> , Yuyang Huang<sup>1</sup> , Chunyu Xu<sup>1</sup> , Xu Yang<sup>4</sup>, Yunliang Zhou<sup>1</sup> , and Yang Hu<sup>1</sup>

<sup>1</sup>Department of Space Physics, Electronic Information School, Hubei LuoJia Laboratory, Wuhan University, Wuhan, China, <sup>2</sup>Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany, <sup>3</sup>Department of Earth and Environmental Sciences, Ludwig Maximilian University of Munich, Munich, Germany, <sup>4</sup>College of Meteorology and Oceanography, National University of Defense Technology, Changsha, China

**Abstract** In this study, we propose a machine learning based approach to construct an empirical model of thermospheric mass densities, based on the MultiLayer Perceptron and bi-directional Long Short-Term Memory for ensemble learning model (MBiLE). The MBiLE model was trained by using only the thermospheric mass density from Swarm C satellite at ~450 km altitude. To assess the performance of the MBiLE model, the model predictions were compared with observations from several satellites, namely, the Swarm C, the Challenging Minisatellite Payload (CHAMP) and the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) satellites. The determination coefficients ( $R^2$ ) for the three satellites are 0.98, 0.99, and 0.98, respectively. The MBiLE model predicts the thermospheric mass density well not only at Swarm C altitude but also at lower altitudes. Earlier empirical models based on multivariate least-square-fitting approach failed to achieve this good altitude generalization (e.g., Liu et al., 2013, <https://doi.org/10.1002/jgra.50144>; Xiong et al., 2018a, <https://doi.org/10.5194/angeo-2018-25>). Further tests have been made by checking the MBiLE model prediction deviations in relation to magnetic local time, day of year, solar flux level, and magnetic activities. No obvious dependences are found for these parameters. Comparing with the NRLMSIS-2.0 model, the MBiLE model improves prediction accuracy by 91%, 66%, and 56% at the three satellites altitudes. The results indicate that the MBiLE model has the ability to predict well the thermospheric mass density over a wide altitude range, for example, from 224 to 528 km, offering potential for atmospheric research applications.

**Plain Language Summary** Accurate prediction of thermospheric mass density is of significant importance for various space activities, for example, the safe operation of low Earth orbital satellites. To address the challenge of predicting thermospheric mass density over a broader range of altitudes, we propose an integrated model based on machine learning algorithms. By training the model solely with data from only the Swarm C satellite, we achieve high-quality predictions of thermospheric mass density covering a wider altitude range. Remarkably, the model demonstrates robustness against other parameters and exhibits excellent stability. In comparison with the latest version of NRLMSIS model, the prediction from our machine learning model has improved the accuracy by about 91%, 66% and 56% when taking the satellite measurements from Swarm C, CHAMP and GOCE as reference. These results highlight the efficacy and potential of our integrated learning model in accurately predicting thermospheric mass density, thereby benefiting future atmospheric research and enhancing space mission planning.

## 1. Introduction

The thermosphere is an important layer of the Earth's atmosphere, which extends from the top of the mesosphere to about 800 km (Doornbos, 2012). Various approaches have been used to determine the thermospheric mass density, including the observations of low Earth orbit (LEO) object trajectory variations (Lechtenberg et al., 2013), airborne accelerometers measurements (Doornbos et al., 2010), ballistic coefficient estimates (Bowman, 2002) and drag balancers (Santoni et al., 2010) etc. Even though, the thermospheric mass density variations are far from being well understood, as numerous factors can influence the changes, including the solar ultraviolet radiation, energetic particle precipitation from the magnetosphere and solar wind, Joule heating, as well as the various wave and tidal drivers from the lower atmosphere (e.g., Billett et al., 2021; Doornbos, 2012; Emmert et al., 2021; Lühr et al., 2004; Zhou et al., 2009). On the other side, predicting the density is crucial for a safe operation of LEO satellites, as atmospheric drag can deorbit them during extreme space weather changes,

© 2024. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

resulting in collision hazards (e.g., Zhang et al., 2022). Therefore, studying, modeling and predicting the thermospheric mass density is crucial for practical applications.

Different models have been developed in recent decades to calculate the thermospheric mass density. Due to a lack of observations, earlier thermospheric mass density distribution was initially calculated theoretically. For example, an expression for the mass density of the thermosphere was derived by Chapman and Cowling (1990), by considering processes such as molecular collisions and dissociation in the atmosphere. They introduced the Chapman model, which effectively described the density distribution in the atmosphere and thus laid the foundation for subsequent studies. Though the theoretical models can simulate in principle the atmospheric evolution processes and physical mechanisms, their output sometimes significantly differs from observations. Different from the theoretical models, empirical models, based on parametric empirical equations and observation-driven parameter fitting, have also been developed to predict the thermospheric mass densities. The latter exhibit in practical application usually better performance than the theoretic models. Mostly used empirical models include the Mass Spectrometer Incoherent Scattering Radar Extension (MSISE) series (Emmert et al., 2022; Hedin, 1983), the Drag Temperature Model (DTM) series (Bruinsma, 2015; Bruinsma & Boniface, 2021), the Jacchia-Bowman 2008 (JB2008) series (Bowman et al., 2008) and the high accuracy satellite drag model series (Doornbos, 2012; Storz et al., 2005). While empirical models have been continuously improved by adding more observational data and modifying the parameters, uncertainties still exist in their prediction when compared with the observations, for example, the prediction errors are around 15% during geomagnetically quiet periods and even larger during magnetic storms (Sutton, 2018). Specific empirical models based on multi-years observations from the satellites have been developed so far, by using multivariate least-square-fitting to the thermospheric density observations (e.g., Liu et al., 2013; Xiong et al., 2018a). However, these models work only for certain altitude ranges covered by the satellite data set. Weng et al. (2017) constructed the escape layer temperature model (ETM) by characterizing the global atmospheric temperature. The thermospheric density is further derived based on the relation between temperature and density. The model can represent well the variation characteristics of the thermospheric density, such as the seasonal variation and hemispheric asymmetry. It is worth to mention that the seasonal variations, diurnal variations etc., have been studied in depth with the parameters of time (diurnal, annual), P10.7 and geomagnetic activity separated by Müller et al. (2009). But the thermospheric density derived from ETM works in general only well within a certain altitude range.

During the past few years, machine learning approaches have been successfully applied to space weather related studies. For example, image completion of TEC maps (Chen et al., 2019), auroral image classification (Kvammen et al., 2020), plasmasphere modeling (Zhelavskaya et al., 2021), and long-term trends in ionospheric electron density (Cai et al., 2019; Smirnov et al., 2023). Among them, there is a growing number of studies on predicting the thermospheric mass density by using deep learning methods. Wang et al. (2014) applied an artificial neural network method for the first time to investigate the intra-annual variation of the global mean thermospheric density at 400 km altitude over the period of 1996–2006. This pioneering study represents the initial utilization of deep learning techniques for analyzing thermal layer mass density, marking a significant advancement in the field. Subsequently, Weng et al. (2020) used artificial neural networks to predict long-term trends in thermospheric mass density by fusing the solar radiation flux,  $Kp$  and other indices, but the model only performed well during low solar activity years. The long-term trend in thermospheric mass density has no apparent solar flux dependence. Additionally, Bonasera et al. (2021) improved the thermospheric mass density prediction method by introducing an integrated learning approach to the thermospheric mass density uncertainty estimation, which can better capture and quantify the thermospheric mass density uncertainty. However, the generalization of the thermospheric mass density prediction model and the prediction effectiveness during magnetic storms need to be further evaluated and discussed. Generalizability is an important metric to evaluate the merit of a model, indicating its ability to predict new data. Hence, Wang et al. (2022) further proposed a model based on an ensemble learning algorithm for long short-term memory (LSTM) and noted that the model has certain generalization. Obviously, the accuracy and generalization of the predicted thermospheric mass density remains an important issue, it has great significance to ensure the safe operation of LEO satellites.

Of particular interests for us is that the utilization of observational data from one satellite to train a deep learning model for accurately predicting a broader altitude range of thermospheric mass density, which can effectively demonstrate the model's generalization capability to the maximum extent. In this study, we propose the MBiLE model for predicting the thermospheric mass density and further illustrated the model's generality and the correlation analysis of parameters. The results show a dramatic improvement in the fitting between predicted and

observed values, even during geomagnetic storms. The MBiLE model was trained with observations from ESA's satellite Swarm C, and the developed model shows also quite nice agreement with observations from satellites flying at lower altitudes, for example, the Challenging Minisatellite Payload (CHAMP) and the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) satellites, demonstrating its reliable spatial generalization capability. In addition, we have further compared the MBiLE model prediction with the estimations of NRLMSIS-2.0 (Emmert et al., 2021). Overall, this study offers valuable insights into satellite detection methodologies and advancements in thermosphere modeling techniques.

## 2. Data Description

The Swarm mission, which was launched on 22 November 2013, consists of three satellites operated by the ESA. During the first operational mission phase, starting on 17 April 2014, Swarm A and C flew side-by-side, separated by 1.4° in longitude, at an altitude of about 460 km and an inclination of 87.3°. Swarm B is orbiting the Earth about 50 km higher with an inclination of 88° (Lühr et al., 2016). Two types of thermospheric mass densities have been provided as the Level-1B products of Swarm. One utilizes the precise orbit determination (POD) of Swarm, while the other used the atmospheric drag experienced on the satellite from the onboard accelerometers, to estimate the thermospheric densities. Compared to the density derived from the POD, the accelerometer provides direct and more detailed information for inferring thermospheric density, enabling a better capture of the smaller-scale variations. Thus, in this study, only the neutral densities derived from the onboard accelerometer of Swarm C are used in the deep learning to construct the MBiLE model.

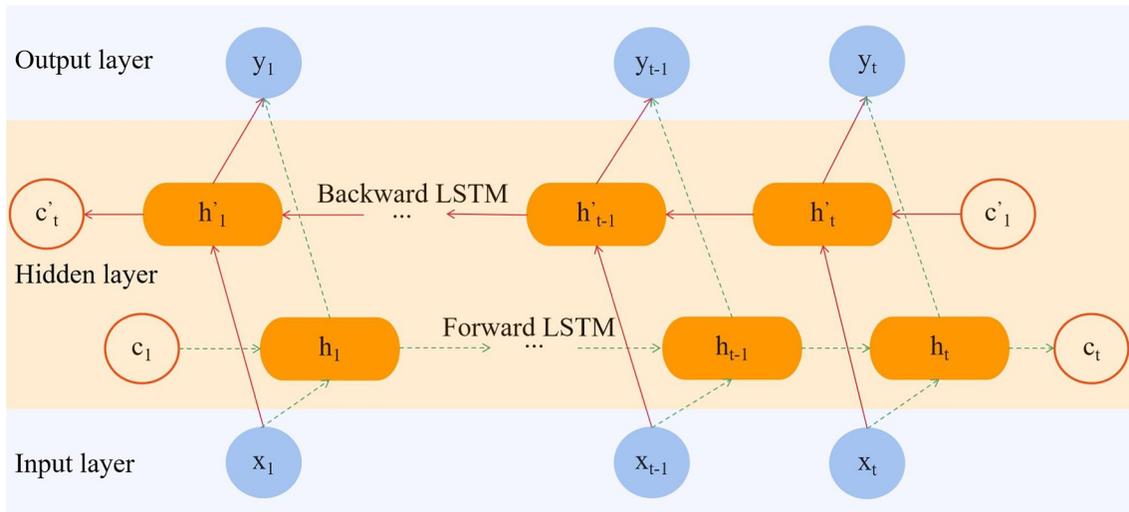
The CHAMP satellite was launched into a near-circular polar orbit on 15 July 2000, with an initial altitude of 456 km, and by the end of its mission on 19 September 2010, the orbit had decayed to approximately 250 km (Xiong et al., 2018b). The GOCE satellite was launched also by ESA, on 17 March 2009, and it reentered the Earth's atmosphere on 11 November 2013. During its mission period, GOCE flew mainly at an altitude range between 225 and 295 km. For both satellites, reliable thermospheric mass density data were derived from the onboard accelerometers, that play an important role in advancing our understanding of the thermosphere (e.g., Doornbos et al., 2009; Liu et al., 2010; Lühr et al., 2004). The thermospheric mass density data from the two satellites can be downloaded from the website <http://thermosphere.tudelft.nl/>.

We want to note that the local time coverage of Swarm C and CHAMP slowly changed during their mission period, which took about 133 and 131 days to cover the 24 hr, when considering both their ascending and descending orbits. While the GOCE satellite, with a dawn-dusk sun-synchronous orbit, it covered mainly the 07 and 19 magnetic local time (MLT) at the low and middle latitudes, and only at polar region its orbit coverage extended to other MLT hours.

In addition to the thermospheric mass density, a variety of geomagnetic and solar activity related indices have also been used to develop the MBiLE model, which can be accessed at the website of OMNI ([https://spdf.gsfc.nasa.gov/pub/data/omni/low\\_res\\_omni/](https://spdf.gsfc.nasa.gov/pub/data/omni/low_res_omni/)). In total, there are 62 parameters used as inputs, and the details are shown in the Table A1 of the Appendix. Within this study, we focused more on the model construction method based on machine learning. It is noted that the thermospheric mass density derived from the Swarm, CHAMP and GOCE have a time resolution of 10 s, we therefore used indices with also a time resolution of 10 s.

## 3. Model Construction

The Swarm C thermospheric mass density data used for developing the model cover the period from 1 February 2014 to 30 September 2020, and the data were divided into three sets: training, validation, and testing, following the common practice in machine learning with an 8:1:1 ratio. The training set consisted of 13,891,423 samples, including data from 1 February 2014 to 21 February 2019. Both the validation and test sets consist of 1,736,427 samples containing data from 21 February 2019 to 24 October 2019, and from 24 October 2019 to 30 September 2020, respectively. The training set facilitated model learning, while the validation set was used for hyperparameter tuning and performance evaluation during training. The testing set served as an independent benchmark to assess the model's generalization performance on unseen data. This data division strategy aimed to strike a balance between the adequacy of training data and reliable performance evaluation, ensuring the robustness and credibility of the proposed model. Additionally, the trained model was also used directly to predict the thermospheric mass density of the CHAMP satellite from 4 May 2001 to 4 September 2010, and the GOCE satellite from 1 November 2009 to 20 October 2013. The process is also



**Figure 1.** Structure of the bi-directional long short-term memory model.

known as testing the spatial generalizability of the model. It is commonly believed that the hallmark of a good deep learning model is the ability to generalize it well also for other data, allowing the model to make better predictions on conditions that has never been included in the training data. Therefore, to provide a more intuitive representation of the model's performance, we showed the predicted thermospheric mass density for each satellite in the analysis in Section 4.

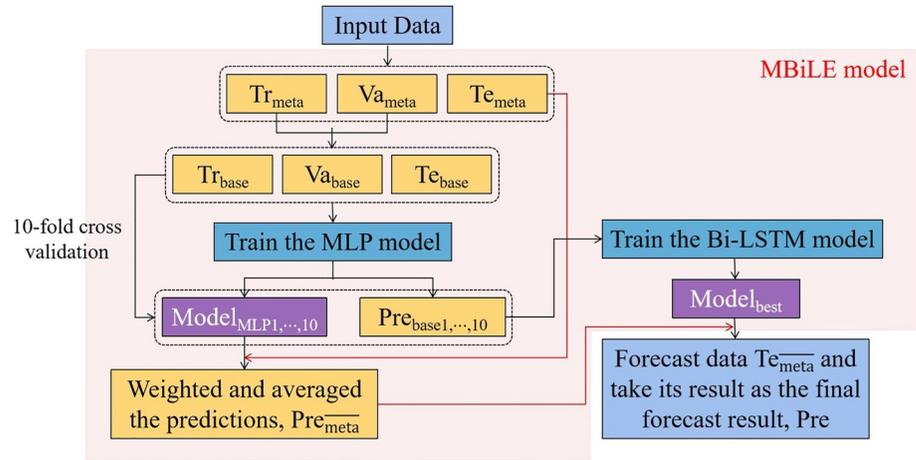
There are three commonly used methods for model ensembling: bagging, boosting, and stacking. Among these techniques, the stacking, also known as stacked generalization, is selected as the preferred approach for the model construction in this study (Cagnini et al., 2023). Stacking is considered an extension of other ensemble methods (Schwenker, 2013). It leverages the predictions from multiple models to make combined predictions, offering improved performance. The fundamental idea behind model ensembles is that the various models, trained independently, may excel for different reasons. Each model brings in a unique perspective to the data, contributing partial insights toward capturing the underlying truth. To ensure the effectiveness of the ensemble approach, it is crucial to incorporate diverse estimators, utilizing different architectures to enhance model stability and generalizability (Ganaie et al., 2022). Thus, by employing stacking and leveraging the diverse perspectives of individual models, this study aims to enhance the predictive capabilities and overall performance of the ensemble model.

Hochreiter and Schmidhuber (1997) proposed a new recursive network structure, the LSTM, consisting of the input, output and forget gates. The input gate determines how much information about the state of the network at the current moment needs to be saved to the internal state, while the forget gate determines how much information about the past state can be discarded, and finally the output gate determines how much information about the internal state at the current moment needs to be output to the external state. Compared to the LSTM, the bi-directional long short-term memory (Bi-LSTM) model acquires feature data at time  $t$  with both past and future information, and this bi-directional structure improves the long-term dependence of learning and the accuracy of the model (Graves & Schmidhuber, 2005). The Bi-LSTM model contains a bi-directional LSTM structure, where the forward LSTM structure is applied to the input sequence and the reverse form of the input sequence is fed back into the LSTM structure. The used model structure is shown in Figure 1. Assuming that the current moment is  $t$ , the internal formulation of the Bi-LSTM model is:

$$h_t = f(uh_{t-1} + wx_t) \quad (1)$$

$$h'_t = f(u'h'_{t-1} + w'x'_t) \quad (2)$$

$$c_t = f(Ux_t + Wc_{t-1}) \quad (3)$$



**Figure 2.** The technology guideline and MBiLE model structure of this research.

$$c'_i = f(U'x_i + W'c'_{i+1}) \quad (4)$$

$$y_i = g(Vc_i + V'c'_i) \quad (5)$$

where  $h_i$  and  $h'_i$  denote the network cells,  $c_i$  and  $c'_i$  denote the hidden layer outputs,  $x_i$  and  $y_i$  denote the model inputs and outputs, respectively,  $u, u', w, w', U, U', W, W', V$  and  $V'$  denotes the weighting coefficients. In this experiment, we have abundant data regarding the features of physical parameters linked to the long-term variations of thermospheric mass density, so the Bi-LSTM model is chosen as one of the integrated models. Another underlying model used in the integrated model proposed in this study is the Multilayer Perceptron (MLP). The activation function used in the MLP is sigmoid, and a Dropout structure is added to prevent overfitting. The MLP is a forward-structured neural network that maps input vectors to output vectors, and can also be viewed as a directed graph, consisting of multiple node layers, each fully connected to the next layer (Delashmit & Manry, 2005). For these reasons, we have chosen to integrate the MLP model with the Bi-LSTM model to propose the MBiLE model for predicting thermospheric mass density.

The technical roadmap and MBiLE model architecture for this study are shown in Figure 2. The input to the training is the 62 parameters at moments  $t_{n-5}$  to  $t_{n-1}$ , and the output is the thermal layer mass density data at moment  $t_n$ . Normalization is an important step for many machine learning estimators during the data pre-processing phase, especially when dealing with the deep learning. At the same time, it is ideal for large data scenarios, considering that the standard normalization is less susceptible to outliers and is more stable when there are already enough samples (Montavon et al., 2012). Therefore, we pre-processed the data using standard normalization prior to training, fitting a data set consisting of a total of 62 parameters and also the thermospheric mass density data from Swarm C. It is worth noting that the maximum and minimum features of the validation and test sets should be normalized by using the results of the training set. When we employ a base learner to learn and solve a problem, if the base learner learns a region of the featured space incorrectly, the meta-learner corrects this error appropriately by combining the learned behaviors of other base learners. This represents the fundamental strategy of the stacking ensemble. The training set for the MLP model is designated as  $Tr_{base}$ , the validation set as  $Va_{base}$ , and the test set as  $Te_{base}$ . The training set for the Bi-LSTM model is denoted as  $Tr_{meta}$ , the validation set as  $Va_{meta}$ , and the test set as  $Te_{meta}$ . The specific process of modeling is as follows. First, MLP is used for training, and K-fold cross-validation ( $k = 10$ ) is used to enhance the evaluation of the training effect of the model during training. After obtaining the model generated during training, it is saved as  $Model_{MLP1,...,10}$ , and the output values of the predictions made by the model during each cross-validation are saved in  $Pre_{base1,...,10}$ . It should be noted that  $Pre_{base1,...,10}$  corresponds to the input data of  $Tr_{base1,...,10}$ . Subsequently,  $Model_{MLP1,...,10}$  is used to predict  $Te_{meta}$  data and its predictions are weighted to take the mean value to obtain, which is seen to correspond to the input data. In addition, the Bi-LSTM model was used for the  $Tr_{base1,...,10}$  data to be trained 20 times, and the best model  $Model_{best}$  is determined based on the

**Table 1**  
The Values of the Parameters in MBiLE Model

Parameters	Values
Num of MLP model	5
Activation	sigmoid
Dropout	0.2
K-fold	10
Batch_size	128
Loss	RMSE
Optimizer	Adamax
Epochs	20
Num of Bi-LSTM model	2
Num of dense layers	1
Length of the input sequence	5
Length of the output sequence	1

magnitude of the root mean square error (RMSE) in the training results. Finally,  $Model_{best}$  is used to predict the data and the results are used as the final prediction of the MBiLE model, which is denoted as  $Pre$ . In the experiments, the number of MLP models is five and the number of Bi-LSTM models is two, the activation function used throughout is sigmoid, and the Dropout is 0.2. The specific parameter settings are shown in Table 1. It is important to highlight that our modeling utilizes the Tensorflow and Keras frameworks, and the training of the models takes place on the super-computing system at the Wuhan University Supercomputing Center. The computing time required to train the model within this study is approximately 6.5 hr.

In general, RMSE and the coefficient of determination ( $R^2$ ) are commonly used metrics for comparing and evaluating model results in machine learning and statistical modeling. They have wide applications in assessing model performance.

The RMSE metric quantifies the differences between model results and observed data. It calculates the RMSE between the predicted values of the model and the observed values, providing a measure of predictive accuracy. A smaller RMSE value indicates higher precision in model predictions. The RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (pre_i - obs_i)^2}{N}} \quad (6)$$

where  $obs_i$  denotes the observed value,  $pre_i$  denotes the predicted value, and  $N$  denotes the total number of samples.

The  $R^2$  metric measures the goodness of fit between model results and observed data, with values ranging from 0 to 1. A higher  $R^2$  value indicates a better fit of the model to the observed data, suggesting that the model can explain the variability in the data. However, a lower  $R^2$  value may indicate underfitting or errors in the model. Using  $\overline{pre}$  and  $\overline{obs}$  to denote the mean of the predicted and observed values, respectively,  $R^2$  is calculated as follows:

$$R^2 = \left( \frac{\sum_{i=1}^N (pre_i - \overline{pre})(obs_i - \overline{obs})}{\sqrt{\sum_{i=1}^N (pre_i - \overline{pre})^2} \sqrt{\sum_{i=1}^N (obs_i - \overline{obs})^2}} \right)^2 \quad (7)$$

By considering both  $R^2$  and RMSE together, we can comprehensively assess the model's goodness of fit and predictive accuracy, thus gaining a more comprehensive understanding of its performance. In order to better analyze the influence of MLT, F10.7 (10.7 cm solar radio flux), DoY (Day of Year) and ap index on our MBiLE model predictions, we introduced the error rate indicator with the following formula.

$$Error\ rate = \frac{\sum_{i=1}^N \left( \frac{obs_i - pre_i}{obs_i} \right)}{N} \quad (8)$$

Quantifying the extent of enhancement brought about by the MBiLE model in comparison to NRLMSIS-2.0 is of particular significance. This assessment can be effectively conducted using the Prediction Efficiency (PE) metric. The formulation for this metric is as follows:

$$PE = 1 - \frac{\sum_{i=1}^N (\text{obs}_i - M_i)}{\sum_{i=1}^N (\text{obs}_i - S_i)} \quad (9)$$

where  $M_i$  denotes MBiLE model values, and  $S_i$  represents baseline (NRLMSIS-2.0) model outputs.

## 4. Results

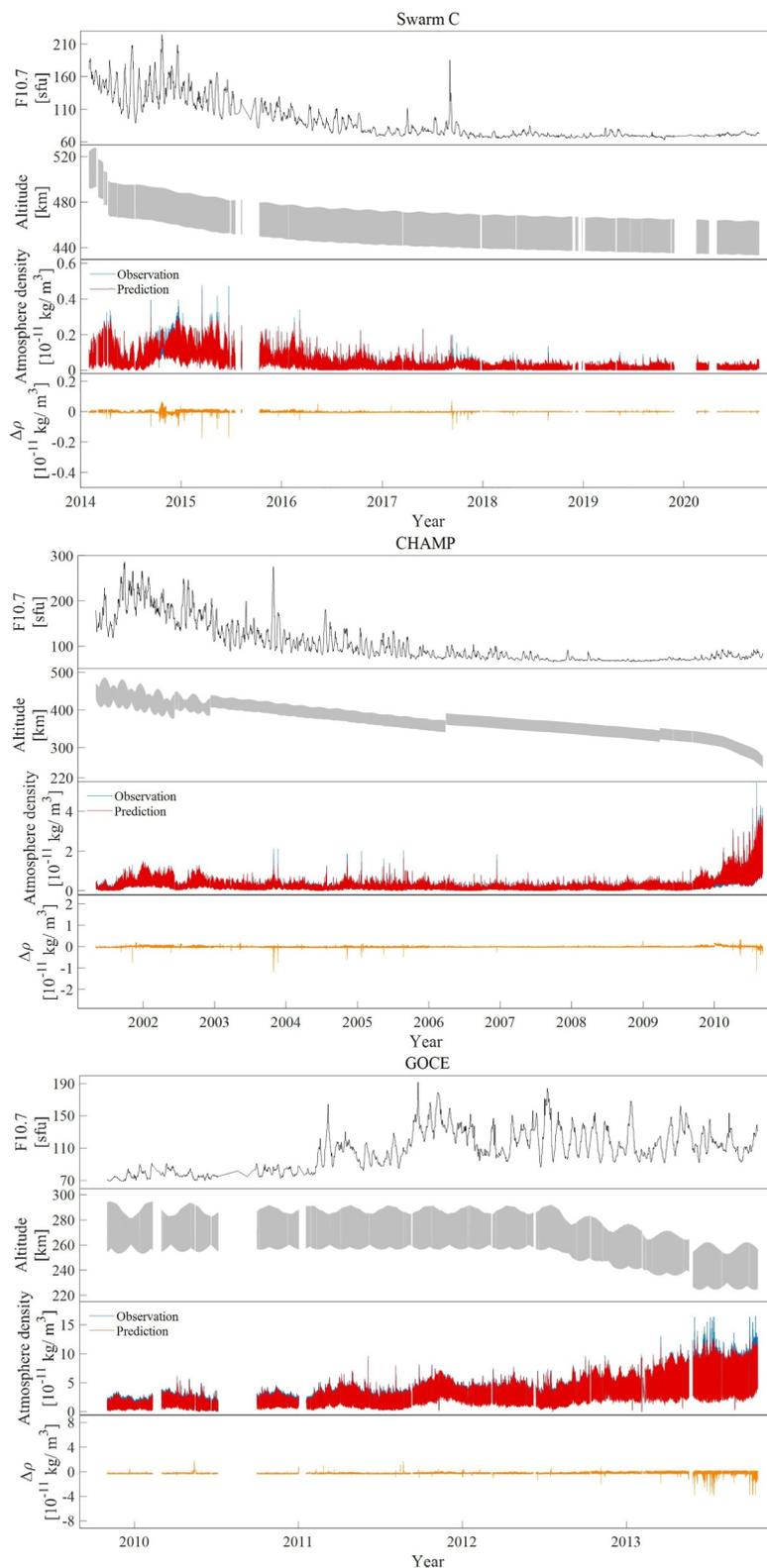
### 4.1. Comparison of MBiLE Model Prediction With Observations From the Swarm C, CHAMP and GOCE Satellites

In addition to the Swarm C data, we used also the trained MBiLE model to predict the thermospheric mass density of the CHAMP and GOCE satellites, to check if the model has a strong generalization capability. Figure 3 shows the MBiLE model predictions of thermospheric mass density, separately for Swarm C, CHAMP and GOCE satellites over their mission periods. In addition, the variations of solar flux index F10.7, the satellite orbital altitude, as well as the thermospheric mass density difference, the model prediction minus satellite observations, are also shown. In general, the MBiLE model predictions show very consistent variations to the satellite observations, and their difference are in general very small. In addition, there are also “spikes” with relatively larger amplitude seen in the difference. After checking the geomagnetic indices, these “spikes” happened mainly during geomagnetically disturbed periods. For example, a strong magnetic storm occurred from 6 to 7 September 2017, and it resulted in significant disturbances of the Earth's atmosphere and ionosphere (e.g., Lei et al., 2018; Xiong et al., 2019). As seen in Figure 3a, the differences between the MBiLE model predictions and Swarm C measurements during this magnetic storm period are slightly more pronounced compared to the preceding and subsequent periods, but still the maximum difference is less than  $0.12 \times 10^{-11} \text{ kg/m}^3$ . Similar results, underestimation of storm effects, are found in the differences between the model predictions and observations by the CHAMP satellite. Figure 4 shows in the top panel the variations of thermospheric mass density from the GOCE satellite as well as model predictions during a strong magnetic storm on 17–18 March 2013, while in the bottom two panel the corresponding variations of SYM-H Kp indices are shown. The RMSE between the GOCE observations and MBiLE and NRLMSIS-2.0 model predictions was  $2.73 \times 10^{-12} \text{ kg/m}^3$  and  $1.47 \times 10^{-11} \text{ kg/m}^3$ , respectively. The result revealed that our MBiLE model can reflect well the drastic fluctuations of thermospheric mass density caused by geomagnetic storms.

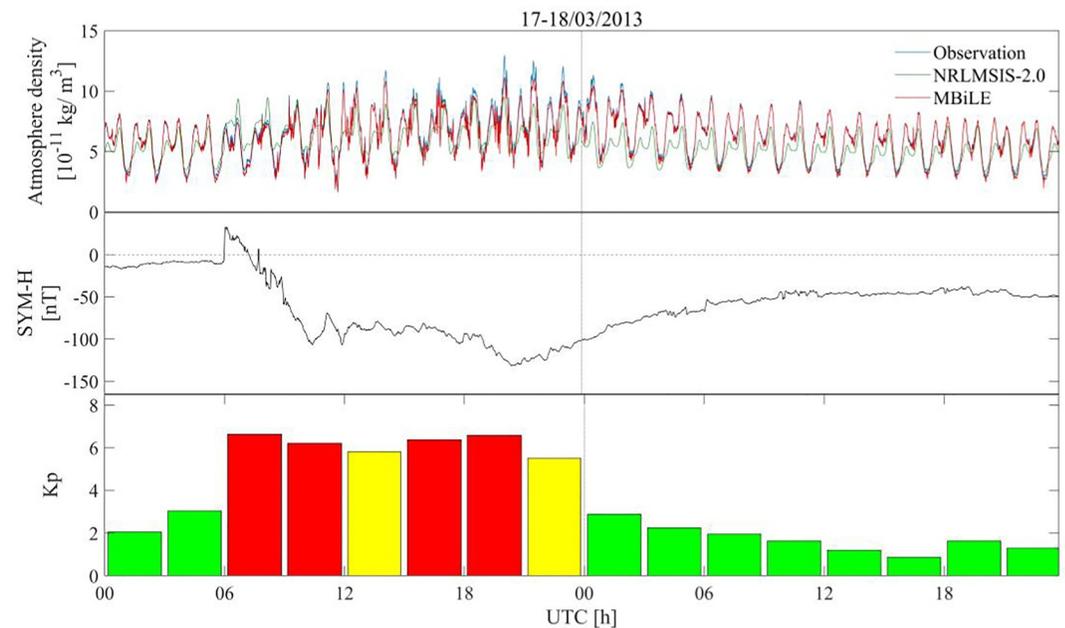
Another feature we want to note is that the CHAMP satellite began to experience prominent orbital decay in 2010, due to the strong increased influence of atmospheric drag at lower altitudes. The satellite's interaction with the denser atmosphere caused a decrease of its orbital altitude, and with that CHAMP lost potential energy. Therefore, over time, the CHAMP altitude decreased more rapidly until it reentered the atmosphere and burned up, marking the mission end. By comparing the predicted results of the MBiLE model with the actually observed data during CHAMP's orbital decay, the model's capability during this low-orbit phase can be assessed. The MBiLE model accurately predicts the increase in thermospheric mass density experienced by CHAMP during 2010. Obtained results are shown in Figure 3b.

Similar to CHAMP, the GOCE satellite also underwent significant orbital decay prior to the end of its in-orbit operations, albeit due to distinct reasons. After 2011, the GOCE satellite gradually entered the period of the solar activity peak, with the F10.7 index reaching a maximum of approximately 191.6 sfu. As solar activity increased, the heating caused by solar EUV radiation lifted up the atmosphere, leading to an increase in thermospheric mass density at satellite altitudes. This, in turn, increases the drag of the GOCE satellite, resulting in an orbital decay. From Figure 3c, it can be seen that even during periods of strong solar activity, the MBiLE model is still able to accurately predict the increasing trend of thermospheric mass density, and it reflects a strong correlation between solar activity and thermospheric mass density in the predicted results.

To further evaluate the MBiLE model prediction, Figure 5 shows the variation of the thermospheric mass density derived from the NRLMSIS-2.0 model, as well as the difference between the NRLMSIS-2.0 prediction minus satellite observations. At first glance, the difference between the NRLMSIS-2.0 model predictions and the



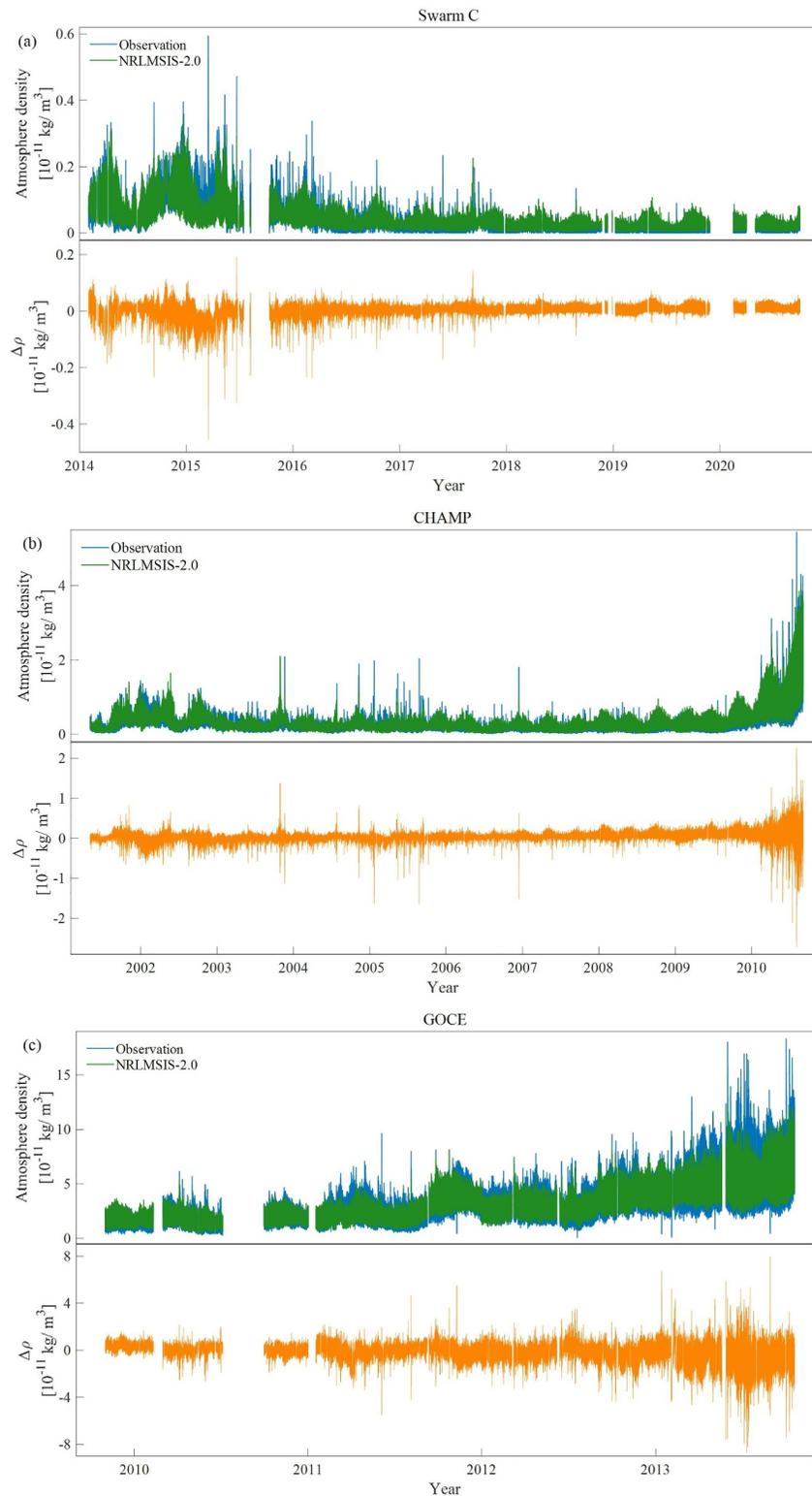
**Figure 3.** The F10.7 index, orbital altitude, thermospheric mass density, and prediction difference over time.  $\Delta\rho$  denotes the difference, which is the prediction minus the observation.



**Figure 4.** The top panel shows the comparison of thermospheric mass density observations from the GOCE satellite, as well as NRLMSIS-2.0 and MBIle model predictions during the geomagnetic storm on 17–18 March 2013. The bottom two panels show the corresponding variations of SYM-H and Kp indices during this storm.

satellite observations are in general larger than the difference between our MBIle model predictions and observations. In Figure 5a, substantial differences are noticeable between the calculated outcomes derived from NRLMSIS-2.0 and the satellite observations, particularly around the year 2015, where the maximum discrepancy surpasses  $-0.45 \times 10^{-11} \text{ kg/m}^3$ . Generally, the observations are larger. In Figure 5b, the maximum difference between the NRLMSIS-2.0 derived results and the satellite observations exceeds  $-2 \times 10^{-11} \text{ kg/m}^3$ , with notably higher discrepancies observed for GOCE in Figure 5c, surpassing  $\pm 8 \times 10^{-11} \text{ kg/m}^3$ . This indicates that the NRLMSIS-2.0 model's performance may need further improvement for accurately capturing the influences of higher solar and geomagnetic activity, as well as the effect of the rapid satellite orbital decay. In comparison, the MBIle model predictions better captured these influences on the thermospheric mass density, demonstrating the excellent generality of our model.

To further analyze the distribution of the prediction results and observations of the MBIle model, as well as the derived results from the empirical model NRLMSIS-2.0, we screened the data for low magnetic activity,  $a_p < 12$ , for comparative analysis. As shown in Figure 6a, the MBIle model predictions follows much better the observations of Swarm C, compared to the NRLMSIS-2.0 model predictions. In Figure 6b, when the thermospheric mass density observed by CHAMP is below  $3.5 \times 10^{-11} \text{ kg/m}^3$ , the scatter plot of predicted results from the MBIle model are concentrated near the diagonal line, while the inversion results of NRLMSIS-2.0 remain relatively scattered. Particularly in Figure 6c, when the observed values from GOCE are relatively small, the inversion results of the thermospheric mass density using NRLMSIS-2.0 are evidently dispersed. It is worth noting that for relatively large observed values, both the MBIle model and the empirical model NRLMSIS-2.0 exhibit thermospheric mass density values that are smaller than the observations. For example, in Figure 6a, when the observed thermospheric mass density of the Swarm C satellite is greater than  $0.25 \times 10^{-11} \text{ kg/m}^3$ , the MBIle model results start to fall short. By referring to Figure 3a, it can be observed that the MBIle model exhibits higher prediction deviations around the year 2015 compared to other periods. This may be attributed to an insufficient learning and training of the MBIle model on the characteristics of thermospheric mass density during periods of high magnetic and solar activity. In Figure 6b, when the observed thermospheric mass density of the CHAMP satellite is greater than  $3.5 \times 10^{-11} \text{ kg/m}^3$ , the scatter plot of the MBIle model's prediction results starts to deviate from the diagonal line. Combining this with Figure 3b, it can be noted that at 2010, the observed thermospheric mass density exhibits a rapid increase due to the starting solar activity and the decay of the satellite's orbit. During



**Figure 5.** Results of empirical model NRLMSIS-2.0 for predicting the thermospheric mass density of Swarm C, Challenging Minisatellite Payload and GOCE satellites are presented. The  $\Delta\rho$  denotes the difference, which is the predictions minus the observations.

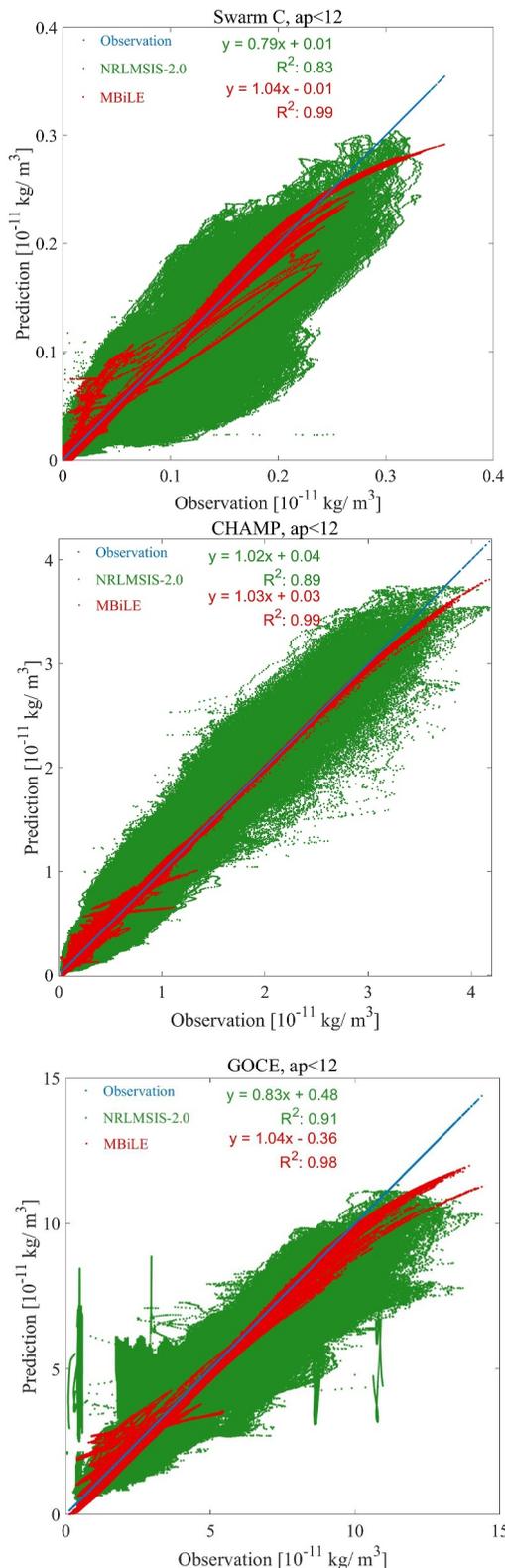


Figure 6.

this period, the MBIle model's prediction difference for the thermospheric mass density of the CHAMP satellite is higher. Our model lacks training for predicting accurately the thermospheric mass density during this specific period.

Overall, Figure 6 reflects a higher degree of dispersion in the distribution of thermospheric mass density obtained from the empirical model NRLMSIS-2.0, while our model's results are much more concentrated around the observations. Further, the performance of the model was analyzed in more detail using  $R^2$ , RMSE and prediction error standard deviation (PESD) metrics. All performance evaluations of the model regarding the thermospheric mass density from the Swarm C satellite are conducted only on test data set spanning from 24 October 2019 to 30 September 2020. This approach provides a more realistic, objective, and rigorous evaluation of the model's performance. For the sake of comparison, the  $R^2$  and RMSE metrics calculated for Swarm C with respect to the empirical model NRLMSIS-2.0 are based on the same range of data. Then, when  $ap < 12$ , the MBIle model achieved  $R^2$  values of 0.99, 0.99, and 0.98 for predicting the thermospheric mass densities of the Swarm C, CHAMP, and GOCE satellites, respectively. In contrast, the inversion results of the empirical model NRLMSIS-2.0 yielded  $R^2$  values of 0.83, 0.89, and 0.91 in comparison with the observed data. A more comprehensive statistical analysis is presented in Table 2. This shows the MBIle model achieves an overall high  $R^2$  between the predictions and observations with values of 0.98, 0.99, and 0.98 for the three satellites, respectively, and these coefficients are higher than those obtained with respect to the NRLMSIS-2.0 empirical model, which are 0.61, 0.89, and 0.90, respectively. Compared to the NRLMSIS-2.0 model, the MBIle model shows a substantial decrease in RMSE by an order of magnitude for the Swarm C and CHAMP satellites. And the RMSE for the GOCE satellite decreased from  $6.153 \times 10^{-12}$  to  $2.727 \times 10^{-12}$  kg/m<sup>3</sup>. Overall, the prediction accuracy for these three satellites have been improved by 91%, 66%, and 56%, respectively. To provide a clearer demonstration of the superiority of the MBIle model, we have introduced the PESD (standard deviation of the difference between observed and predicted values) metric. This metric quantifies the average level of variation between the model's predicted values and the actual observed values. In comparison to the NRLMSIS-2.0 model, the MBIle model showcases a reduction in the PESD of the Swarm C satellite by an order of magnitude. Moreover, the PESD of the CHAMP satellite decreases from  $8.676 \times 10^{-13}$  to  $2.650 \times 10^{-13}$  kg/m<sup>3</sup>, and the PESD of the GOCE satellite decreases from  $5.693 \times 10^{-12}$  to  $1.207 \times 10^{-12}$  kg/m<sup>3</sup>. This decline in the PESD values further underscores the enhanced performance of the MBIle model. The PE metric serves to quantify the advancements achieved by the MBIle model in comparison to the NRLMSIS-2.0. As indicated in Table 2, the outcomes distinctly illustrate that the MBIle model consistently enhances the efficacy of the NRLMSIS-2.0 model by an impressive margin ranging from 78% to 93%. Particularly noteworthy is the remarkable

Figure 6. Distribution of observations (light blue dots) and the prediction of our MBIle model (red dots) as well as the NRLMSIS-2.0 (green dots) for the thermospheric mass densities from the Swarm C, Challenging Minisatellite Payload and GOCE satellites during geomagnetic quiet period ( $ap < 12$ ). The linear regressions between observations and MBIle model predicts, between observations and NRLMSIS-2.0 predicts, are provided in red and green, respectively, and their correlation coefficients ( $R^2$ ) are also provided with corresponding colors.

**Table 2**  
The Overall RMSE,  $R^2$ , PESD and PE Between the Thermospheric Mass Density Predictions From the MBIle/NRLMSIS-2.0 Models With Respect to the Satellite Observations

		Swarm C	CHAMP	GOCE
RMSE	NRLMSIS-2.0	$1.315 \times 10^{-13}$	$1.003 \times 10^{-12}$	$6.153 \times 10^{-12}$
	MBiLE	$1.195 \times 10^{-14}$	$3.43 \times 10^{-13}$	$2.727 \times 10^{-12}$
$R^2$	NRLM2.0	0.61	0.89	0.90
	MBiLE	0.98	0.99	0.98
PESD	NRLM2.0	$1.6453 \times 10^{-13}$	$8.676 \times 10^{-13}$	$5.693 \times 10^{-12}$
	MBiLE	$4.081 \times 10^{-14}$	$2.650 \times 10^{-13}$	$1.207 \times 10^{-12}$
PE		93%	89%	78%

amelioration observed in predicting the thermosphere mass density of the Swarm C satellite.

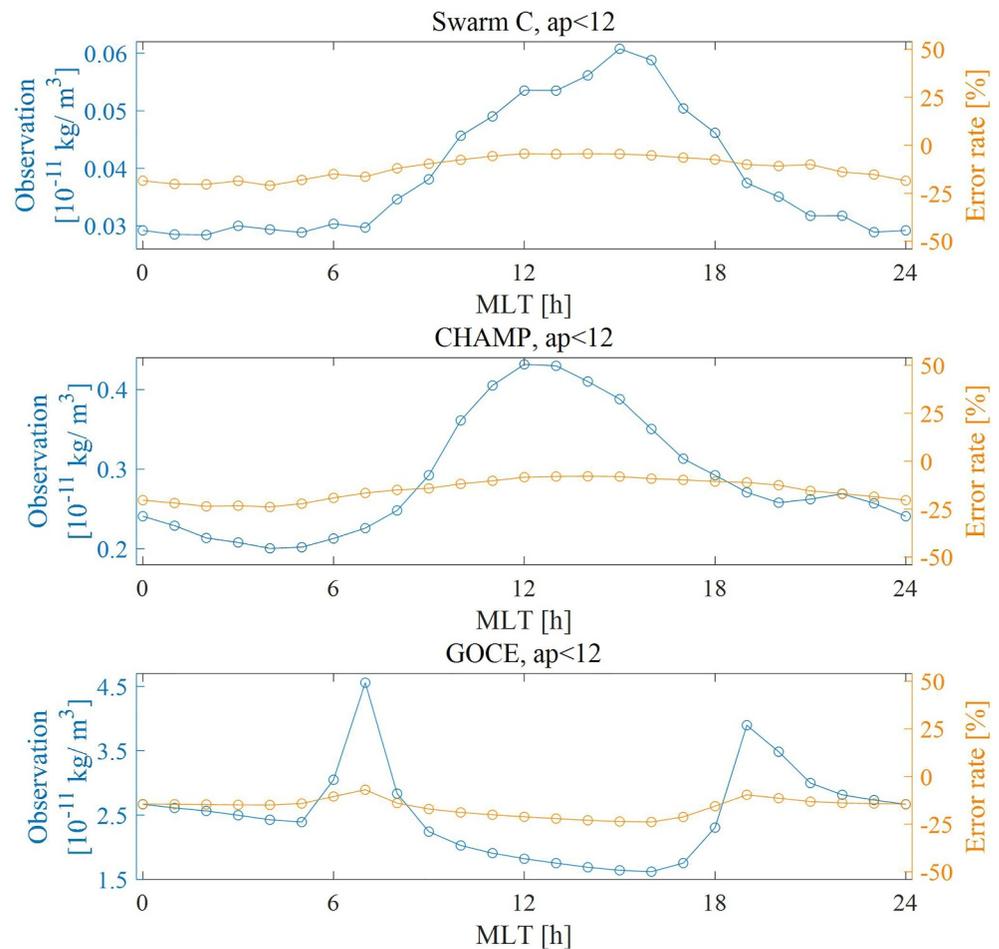
#### 4.2. Parameter Dependences Analysis

After conducting a detailed comparative analysis of the observation results from Swarm C, CHAMP, and GOCE satellites with the results from an empirical model and the MBIle model, we further examined the correlation between MLT, F10.7, DoY, ap index, and the satellite observations, as well as the error rates. It should be noted that the MBIle model, driven by 62 feature parameters, has been fully trained and tested, and its loss function value has become very stable (as shown in Figure A1), indicating that the model has reached its optimal state. Based on this, we conducted research on whether the model's performance is affected by changes in MLT, F10.7, DoY, and ap index. In order to minimize the influence of geomagnetic storms on the results, we first specifically analyzed the variations of satellite observations and

error rates with respect to MLT for cases where the ap index was less than 12 (calm period). Figure 7 shows that at daytime, the thermospheric mass density from Swarm C and CHAMP satellites initially increases and then decreases, reaching peak values at 15:00 and 13:00 MLT, respectively. The bin size of the MLT in Figure 7 is 1 hr. For Swarm C and CHAMP, the local time coverages of their orbits slowly change with time, therefore the data from these two satellites provide a rather even distribution over MLT. However, due to the sun-synchronous orbit of GOCE, at low and middle latitudes it covers mainly the dawn (0700 MLT) and dusk (1900 MLT) sectors, and only at high latitudes the GOCE covers the other MLT sectors. For a constant altitude, the neutral densities at low and middle latitudes are generally higher than that at high latitudes, we therefore see the observations (blue line) show larger values around dawn and dusk hours, which is biased by the orbital coverage but not the real diurnal variation of the neutral density at GOCE altitude. The error rates of the thermospheric mass density predictions from the MBIle model for the three satellites show very small fluctuations with MLT and exhibit similar trends as the observed values of the mass density. Overall, the relative error rate ranges from  $-25\%$  to 0, indicating a rather small difference between the predicted and true values. This suggests the stability of the MBIle model in terms of its ability to reproduce accurately the diurnal variations.

We want to note that different from theoretic models, we could not totally separate the influences of these different parameters on the satellite measured neutral density. In this study, we are interested more if our model can work quite well under different conditions. From Figure 3, it is clear that there is a close relation between the solar activity level and the thermospheric mass density, we further analyzed its effect (using F10.7 index) on the observed values and error rates of the thermospheric mass density. First, from Figure 8, it can be observed that the thermospheric mass density of Swarm C and CHAMP satellites is as expected larger on the dayside (10:00 to 14:00 MLT) than on the nightside (00:00 to 02:00 MLT and 22:00 to 24:00 MLT), while for the GOCE satellite, the trend is opposite. Second, with an increase in the F10.7 index, the thermospheric mass density of Swarm C satellite shows a gradual increase, while for CHAMP and GOCE satellites, the density tends to first increase and then decrease, especially for the GOCE satellite, exhibiting distinct peaks. Overall, with an increase in the F10.7 index, the fluctuation of error rates in the predicted thermospheric mass density by the MBIle model for the three satellites is relatively small, ranging from  $-30\%$  to  $10\%$ . The error rates of Swarm C and CHAMP satellites on the dayside are smaller than on the nightside, with error rates ranging from approximately  $-10\%$  to  $10\%$ . This may be due to the fact that the thermosphere mass density is higher on the dayside than on the nightside. The error rates of the GOCE satellite are slightly higher than those of Swarm C and CHAMP satellites, ranging from  $-30\%$  to  $-5\%$ , and furthermore, the error rates of the GOCE satellite are lower on the nightside compared to the dayside.

We want to note that these curves shown here do not reflect true seasonal variations as they have not been normalized to a common F10.7, MLT, or latitude distributions. We intent to check further whether our model predictions have been biased by these parameters, due to the data coverage used for model construction. Figure 9 presents the characteristics of the thermospheric mass density and error rate with DoY. The thermospheric mass density of the Swarm C satellite is somewhat higher in the first half of the year compared to the second half. The error rates on the dayside range from  $-17\%$  to  $3\%$ , while on the nightside, they range from  $-30\%$  to  $10\%$ . Overall, the error rates show relatively small fluctuations with respect to DoY, and they also exhibit a rather even distribution throughout the year, implying our model can well reflect the thermospheric density variations over

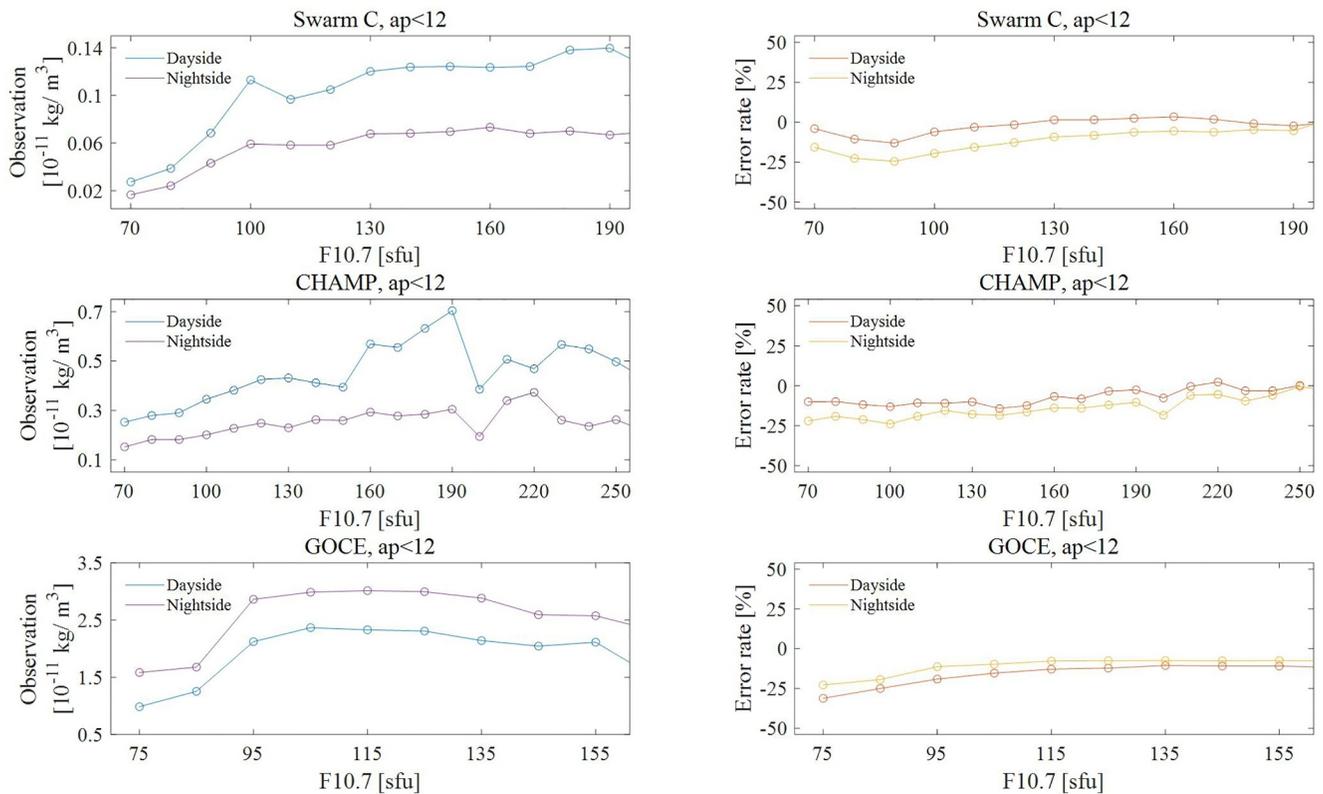


**Figure 7.** Variation of thermospheric mass density observation and error rate with magnetic local time (MLT) for Swarm C, Challenging Minisatellite Payload and GOCE satellites. The MLT has a bin size size of 1 hr.

seasons. For the CHAMP satellite, the variation in error rates is similar to that of Swarm C satellite, but with a smoother trend. As for the GOCE satellite, the thermospheric mass density is higher in summer and winter compared to spring and autumn. Additionally, the error rates on the dayside are slightly higher than on the nightside, with an overall range of approximately  $-35\%$ – $0\%$ .

Figure 10 shows the variation of thermospheric mass density and error rate with ap index for the Swarm C, CHAMP, and GOCE satellites on the dayside and nightside. It can be observed that as the ap index increases, the thermospheric mass density of Swarm C satellite initially increases and then decreases, while the error rates show a relatively smooth change. For the CHAMP satellite, both the thermospheric mass density and error rates increase with the increase in the ap index. The error rates exhibit a more noticeable variation during the daytime, starting around  $0\%$  and gradually increasing to approximately  $-50\%$ , while the nighttime error rates show a relatively gradual change. Similarly, the variations of thermospheric mass density and error rates of the GOCE satellite with respect to the ap index are similar to those of the Swarm C satellite. The error rates of GOCE satellite range from approximately  $-20\%$ – $1\%$ . For the Swarm, CHAMP, and GOCE satellites, there is a rather linear trend of increasing thermosphere mass density with increasing ap index, which is also as expected. However, one should keep in mind that when ap exceeds a certain value, the number of entries becomes very small, and the very low entries for the high activity level may make the results less reliable.

To better visualize the ability of the MBI model to predict the thermospheric mass density at different orbital altitudes and to further analyze the distribution of errors, we plotted the altitude profiles of the Swarm C, CHAMP, and GOCE satellites on the dayside and nightside. These profiles are shown in Figure 11. Specifically, the altitude

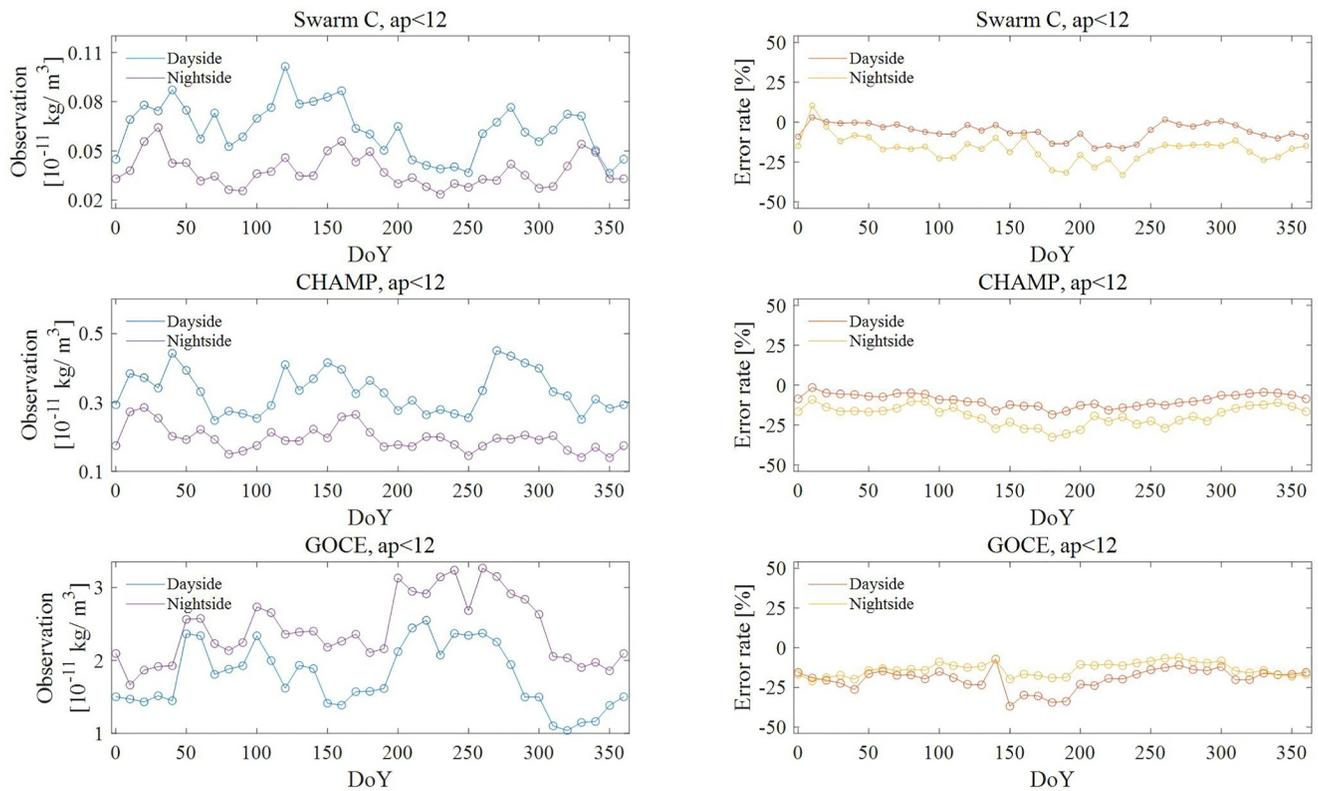


**Figure 8.** Variation of thermospheric mass density observation and error rate with F10.7 for Swarm C, Challenging Minisatellite Payload and GOCE satellites under dayside and nightside in the calm period.

profiles from the Swarm C satellite show that the MBiLE model predictions and the observed values of the mean thermospheric mass density in each altitude range are basically close to each other. In contrast, the results of the empirical model, NRLMSIS-2.0, differ from the observed values, and they exhibit significant variations and lack stability. The altitude profiles of the CHAMP satellite shows that below 325 km and above 475 km, the prediction results of the MBiLE model in general agree well with the observed values. Similarly, the error between the empirical model results and the observed values are also small. Between 325 and 475 km, the difference between the empirical model results and the observations is larger. The altitude profiles of the GOCE satellite shows that the predictions of the MBiLE model are very similar to the observed values, while the empirical model also achieves good results. Overall, the model results on the nightside are slightly worse than those on the dayside, as the atmosphere at night is more disturbed, which may lead to instability of the night data itself.

The error bars on top of each curve in fact represent the standard deviation, or dispersion of the thermospheric mass density within an altitude bin. A larger standard deviation indicates a more spread data distribution, while a smaller value indicating more concentrated distribution. Due to the significant differences in the orders of magnitude of the thermosphere mass density among the three satellites, we simply take the standard deviation of the thermosphere mass density values without any logarithmic transformation, as shown in Figures 11 and 12. This approach accurately depicts the true variations of the data, and facilitates a more reliable assessment and interpretation of the dispersion and consistency of the thermosphere mass density data for different altitude levels and satellite missions. The standard deviation of the MBiLE model results and observations is greater than that of the empirical model NRLMSIS-2.0 inversion results because the thermospheric mass density from the empirical model varies smoother at the different altitudes. However, in detail the comparison of the empirical model results with observations is very dispersed, as can be seen in Figure 6.

In order to show more clearly the distribution of thermospheric mass density with height on the dayside and nightside during the calm period, we have combined the distribution of thermospheric mass density over the



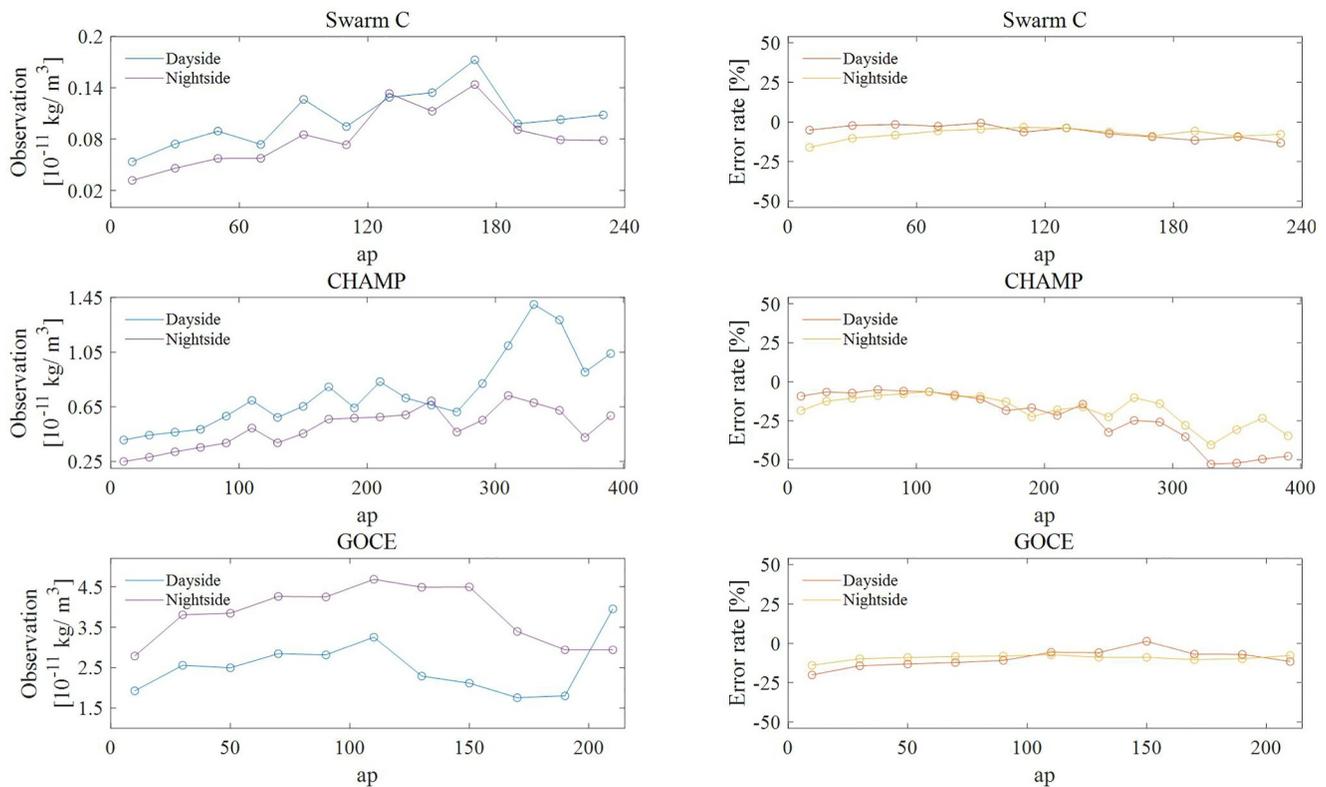
**Figure 9.** Variation of thermospheric mass density observation and error rate with DoY for Swarm C, Challenging Minisatellite Payload and GOCE satellites under dayside and nightside in the calm period.

altitude range from 224 to 528 km in Figure 12. As observed, within the altitude ranges of approximately 430–490 km and 250–295 km, there are two distinct curve branches originating from the different satellites, Swarm C and CHAMP. This feature arises due to the fact that the thermospheric mass density is influenced not only by the satellite altitudes but also by the other various factors, such as solar activity, season, local time, geomagnetic activity, as well as geographic latitude and longitude, etc. Though there is an overlap of height between Swarm C and CHAMP, the above-mentioned parameters, for example, solar flux, season, local time etc., are different. Therefore, the neutral densities measured by the two satellites are different, even when they flew at the same altitude. In particular, the distribution of the dayside in Figure 12 shows that the MBILE model results and the observations are in general agreement. However, at around 370 km, the MBILE model predictions are slightly smaller than the satellite observations, with similar deviations on the nightside, and slightly larger deviations on the nightside than on the dayside. In general, the predictions of the MBILE model show some deviations from satellite observations in terms of their standard deviations at various heights in the height profile of the thermospheric mass density during the calm period but the differences are not significant. The standard deviations of the empirical model NRLMSIS-2.0 are largely small, but differ significantly from the standard deviation of the satellite observations. It should be noted that the standard deviation of thermospheric mass density increases with decreasing altitude due to larger absolute values of mass density at lower altitudes. In particular, the standard deviation of GOCE satellite is much larger than that of Swarm C and CHAMP satellites, as the thermospheric mass density observed by GOCE satellite is higher by 1–2 orders of magnitude.

## 5. Discussions

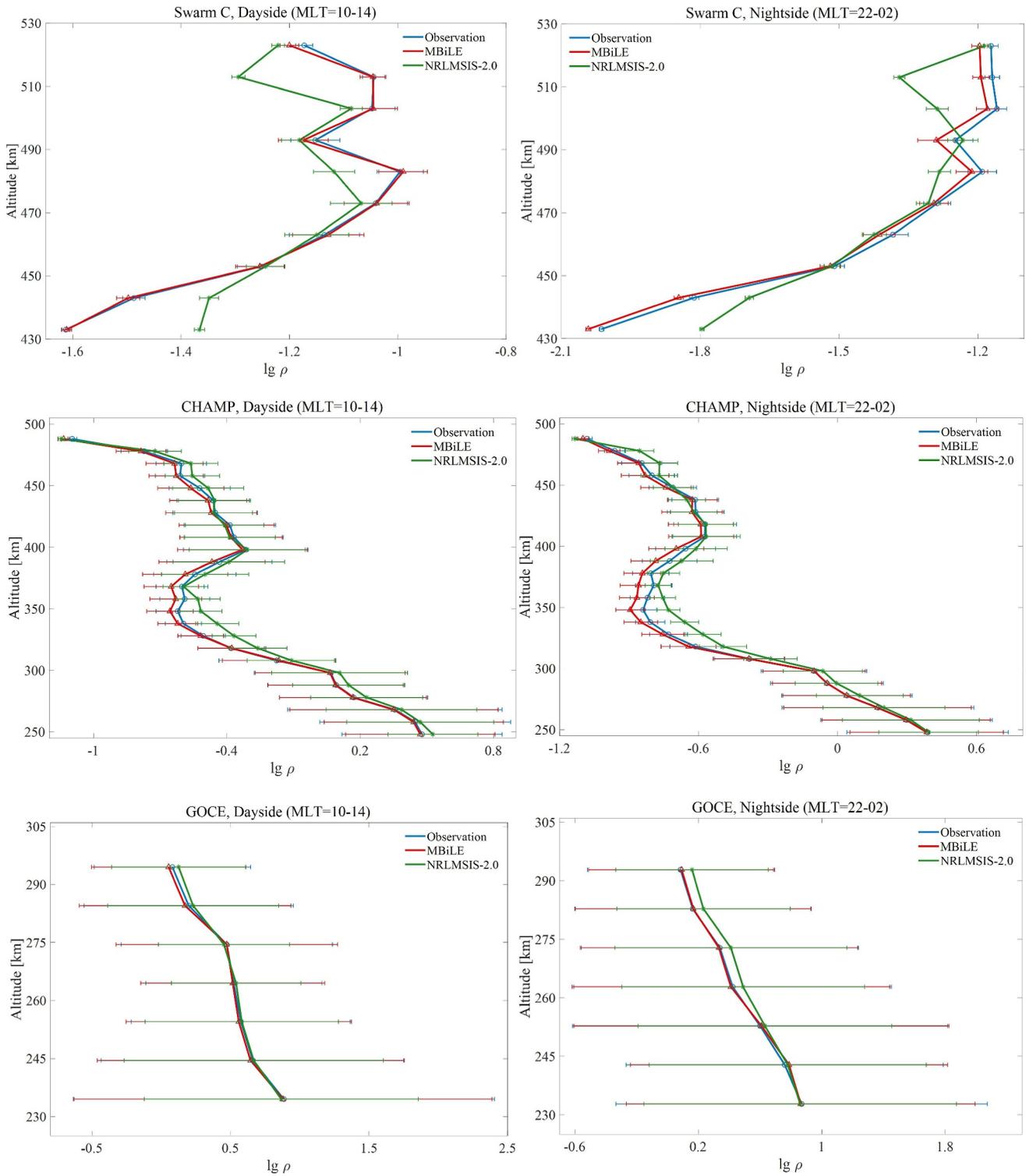
### 5.1. Statistical Analysis

The thermosphere is an important part of the solar-terrestrial space system, and the study of the thermosphere is of great scientific significance for understanding the interactions in the whole solar-terrestrial causal chain (e.g., Lei



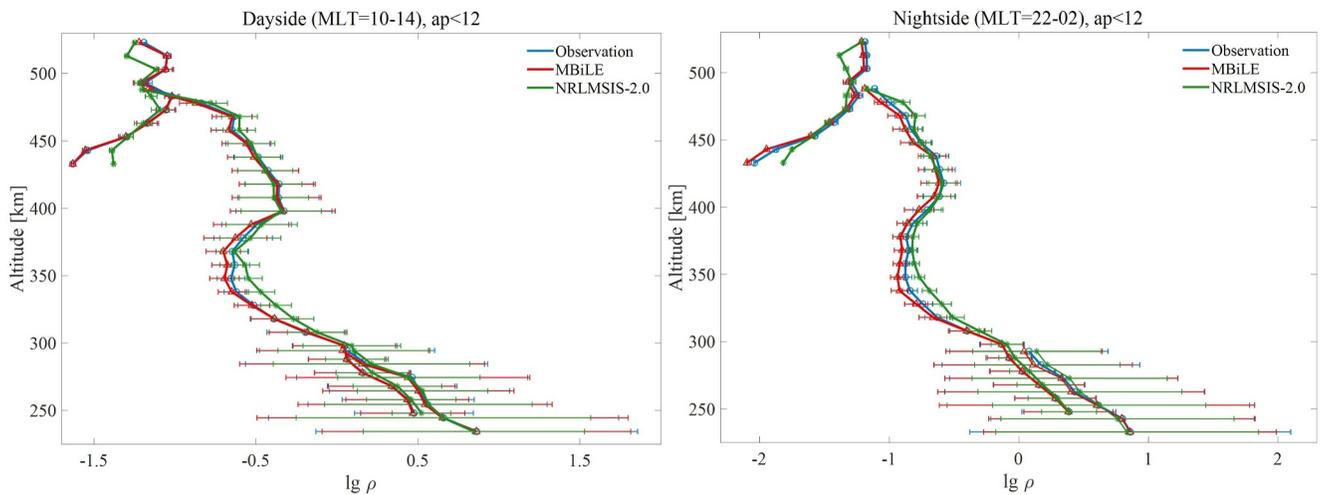
**Figure 10.** Thermospheric mass density observations and error rate variation with ap for Swarm C, Challenging Minisatellite Payload and GOCE satellites on the dayside and nightside.

et al., 2023). The current inversion methods of thermospheric mass density include four main types, which are theoretical model, empirical model, improved model, and artificial intelligence model. The theoretical model can simulate the atmospheric dynamic processes, which are consistent with the considered physical mechanisms, but the differences between the calculated values by such model and the real world can be large, so it is more suitable for studying the underlying mechanisms. An empirical model is based on parametric equations combined with physical laws for data fitting, which is more widely used, but the model prediction accuracy during the magnetic storm periods needs to be improved. Some improved models emerged meanwhile, basically confined to data within a certain orbital altitude, and they cannot achieve reliable prediction of thermospheric mass density across different orbital altitudes. Among the artificial intelligence models, neural network models are dominant, which play a crucial role in the prediction studies of the thermospheric mass density. Unlike traditional models (e.g., MSISE, JB2008, etc.), neural networks can construct parametric models to learn valuable features extracted from a large number of observations and reproduce relevant features using multiple neural network layers. Ensemble learning is a technique that combines multiple models to obtain more accurate and stable prediction results. In ensemble learning, combining different models can compensate for the shortcomings of individual models, thereby improving the overall performance. In this study, Bi-LSTM model and MLP, two different models, are combined to form a more powerful ensemble model MBiLE. This model not only captures contextual information but also learns features sufficiently, possessing strong and rich feature representation capabilities, enabling the ensemble model to better capture complex relationships in underlying data. Therefore, the ensemble model can reduce the risk of overfitting and improve generalization performance. Additionally, by collecting prediction results from the two models and processing them using the stacking algorithm, more accurate prediction results can be obtained. By training the MBiLE model using thermospheric mass density data from the Swarm C satellite at about 450 km, accurate predictions are made for different altitude ranges from 450 km down to 250 km. Such an altitude generalization has not been achieved by using traditional multi-variables least-square fitting (e.g., Liu et al., 2013; Xiong et al., 2018a). The machine learning approach learns the complex relationship of



**Figure 11.** Altitude profiles of the thermospheric mass density for Swarm C, Challenging Minisatellite Payload and GOCE satellites under dayside and nightside. Where error bar denotes the standard deviation and its unit is  $10^{-11} \text{ kg/m}^3$ .

thermospheric mass density on altitudes from observational data gathered by the Swarm C satellite, and such predictive capabilities has been confirmed by comparing to the observation from CHAMP and GOCE satellites flying at lower altitudes. It should be noted that the prediction results of the MBiLE model are somewhat low



**Figure 12.** Altitude profile of the thermospheric mass density on the dayside and nightside during the calm period. Where error bar denotes the standard deviation and its unit is  $10^{-11} \text{ kg/m}^3$ .

compared to the true values, which may be due to the normalized pre-processing of the data by the model before training, compressing most of the variability in the original data to a small range, making the model become less sensitive to changes in the data during training, thus leading to low prediction results (Behzad et al., 2009). It may also be because our model structure is not yet reasonable enough to make more accurate predictions of the thermal layer mass density.

Although the MBiLE model achieves good prediction results, there are still some errors, and we therefore develop a detailed error analysis of the model's prediction results in order to optimize the model subsequently. In this study, a comparison was made between the prediction results of the MBiLE model and the inversion results of the empirical model NRLMSIS-2.0, as well as the observed values. The case representing the Swarm C satellite is shown in Figures 3a and 5a. The MBiLE model exhibited an error range of  $-0.156 \times 10^{-12} \text{ kg/m}^3$  to  $0.130 \times 10^{-12} \text{ kg/m}^3$ , while the empirical model showed an error range of  $-0.247 \times 10^{-12} \text{ kg/m}^3$  to  $0.572 \times 10^{-12} \text{ kg/m}^3$ . The error rate of the MBiLE model was reduced by 91%. Even during several geomagnetic storm periods, such as 28 October 2003, and 4 November 2003, the results of the MBiLE model showed relatively high reliability compared to the empirical model NRLMSIS-2.0. Furthermore, Figure 6 provides detailed scatter plots of the distribution comparing observed mass densities with the prediction results from the MBiLE model, as well as from the empirical model NRLMSIS-2.0, demonstrating the reliability and strong generalization ability of the MBiLE model. The larger errors in the inversion results of the empirical model can be attributed to the non-uniformity and complexity of the thermospheric mass density distribution. The study also investigated the variations of observed thermospheric mass density and error rates with respect to the MLT, F10.7, DoY, and ap parameters.

Figure 7 illustrates the variations of observed thermospheric mass density and error rates with MLT. Overall, the thermospheric mass density of Swarm C and CHAMP satellites exhibits an increasing and then decreasing trend, while the GOCE satellite shows two peaks. It is noteworthy that the error rates do not vary significantly with the increase of MLT; they only slightly decrease at the peaks of thermospheric mass density. This indicates that the performance of the MBiLE model is minimally affected by MLT and demonstrates the model's stability. Subsequently, Figure 8 demonstrates the variations of observed thermospheric mass density and error rates with F10.7 index. Solar activity has a significant impact on thermospheric mass density, particularly during periods of high solar activity. Variations in solar radiation and solar wind lead to heating and disturbances in the thermosphere, thus affecting the distribution of thermospheric mass density. Therefore, during periods of high solar activity, the thermospheric mass density typically increases, leading to augmented drag on satellites and subsequent orbital decay, as exemplified in Figure 3 depicting alterations in the thermospheric mass density of the CHAMP satellite post-2010. Additionally, the fluctuation in error rates of the MBiLE model with F10.7 index is relatively small, ranging from  $-30\%$  to  $10\%$ , which falls within an acceptable range. Clearly, even during periods

of high solar activity, the prediction results of the MBiLE model exhibit a certain level of reliability. We further examined the variations of observed thermospheric mass density and error rates with DoY and ap parameters. Overall, the error rates show relatively small fluctuations with increasing DoY and exhibit a relatively uniform distribution throughout the year. Moreover, as the ap index increases, there is a noticeable change in thermospheric mass density, while the error rates for Swarm C and GOCE satellites demonstrate relatively stable variations. For the CHAMP satellite, the error rates show some increase with the ap index, but the overall error rate remains within  $-30\%$ , which is within an acceptable range. The above analysis indicates that the MBiLE model exhibits good stability and reliability.

## 5.2. Altitude Coverage of MBiLE Model

Satellite orbital altitude is the most important factor affecting the thermospheric mass density, as it varies exponentially with altitude. At higher orbital altitudes, such as for the Swarm C satellite (approximately 430–528 km) and the CHAMP satellite (approximately 220–500 km), the thermospheric mass density is lower. Consequently, these satellites are more influenced by atmospheric drag and require periodic adjustments to maintain the correct orbital altitude. In contrast, at lower orbital altitudes, such as for the GOCE satellite (approximately 224–300 km), the thermospheric mass density is relatively higher. In this study, the data from the Swarm C satellite was used to train the deep learning model MBiLE, which was then used to predict the thermospheric mass density for a broader range of satellites including CHAMP and GOCE. This was done to demonstrate the model's generalization capability and stability.

We examined the altitude profiles of thermospheric mass density for the Swarm C, CHAMP, and GOCE satellites separately during daytime and nighttime. A comparison was made between satellite observation data, the predicted results from the MBiLE model, and the inversion results from an empirical model. Figure 11 clearly shows that the agreement between the predicted thermospheric mass density profiles by the MBiLE model and satellite observations is slightly better in higher density regions, that is on the daytime side compared to the nighttime side, but for GOCE on the nighttime side (see Figure 8). Overall, the inversion results from the empirical model NRLMSIS-2.0 also yielded good results at certain altitudes but not at all altitudes. For example, in the altitude profiles of the Swarm C satellite, both on the daytime and nighttime sides, there were significant discrepancies observed at altitudes below 450 km, within the range of 470–490 km, and between 500 and 520 km. In the altitude profiles of the GOCE satellite the deviations are more pronounced on the nighttime side, particularly above 255 km. This indicates that the empirical model NRLMSIS-2.0 has a reasonable level of inversion capability but lacks stability. In order to eliminate the influence of geomagnetic storms on thermospheric mass density, we further studied the altitude distribution of thermospheric mass density during quiet periods on the daytime and nighttime sides, aiming to fully demonstrate the reliability of the MBiLE model in cross-orbital altitude predictions. In Figure 12, within the range of 224–528 km, the predicted results of the MBiLE model showed better agreement with the observed altitude profiles compared to the empirical model NRLMSIS-2.0, particularly between 300 and 400 km. It should be noted that the length of the error bars represents the size of the standard deviation, which indicates more the degree of data dispersion within an altitude bin. Throughout the entire altitude range, the standard deviation of the empirical model NRLMSIS-2.0 was essentially smaller than that of the MBiLE model and satellite observation data. While this suggests that the inversion results from the empirical model are smoother, but their mean values deviate partly significantly from the true distribution of satellite observations. On the other hand, within the range of 224–528 km, the standard deviation of the MBiLE model is comparable to that of the satellite observation data. These comparative analyses serve to support the demonstration of the robust generalization capability and stability of the MBiLE model.

The thermospheric mass density is closely related to the Earth's magnetic field, solar activity and other factors, and the changes of these factors will lead to the increase of uncertainty in the distribution of the thermospheric density, which causes the larger errors of the empirical model inversions. The MBiLE model, however, has better performance in the prediction task of thermal layer mass density, mainly due to its ability to take advantage of the Bi-LSTM and MLP learners, which can consider information in both forward and back directions, thus making better use of contextual information and improving the ability to model sequences, and improving the overall model stability and generalization through the stacking algorithm. Furthermore, it is noteworthy that the physical significance of machine learning lies in its ability to leverage complex multi-parameter relationships, and finally enable accurate estimations of the thermospheric mass density at a wider altitude range.

## 6. Summary

In this study we combine the ensemble learning algorithms of Bi-LSTM and MLP, and develop a reliable prediction model, MBiLE, for predicting the Earth's thermospheric mass density. Though with only the neutral density measurements from the Swarm C satellites at about 450 km to construct the model, the predictions of MBiLE can well represent the density variations down to 250 km. The main findings of our work are summarized as:

1. The MBiLE model predicted thermospheric mass densities show overall good agreements with the observations from Swarm C, CHAMP and GOCE satellites. The correlation coefficient,  $R^2$ , between predictions and measurements are 0.98, 0.99 and 0.98, respectively, for the three satellites. This result also confirms that the MBiLE model has good generalizability for representing the altitude variation of the thermospheric mass density.
2. Further detailed analysis of the MLT, F10.7, DoY, and ap parameters shows that the error rate of the MBiLE model predictions show almost no dependences on these parameters, which further confirms the reliability and robustness of the MBiLE model.
3. The standard deviation of NRLMSIS-2.0 is relatively small in the height profiles, but the difference of their inversion results from observations are rather scattered. Compared to that, the prediction accuracy of the MBiLE model to predict the thermospheric mass density observed by Swarm C, CHAMP and GOCE satellites is improved by 91%, 66% and 56%, respectively. A comparison of results along the altitude profiles shows that over a wider altitude range the MBiLE model can cover 224–528 km well, indicating that the MBiLE model has a good generalization performance.

In this study, we only utilized data from the Swarm C satellite for modeling purposes, to demonstrate the effectiveness of this approach in providing comprehensive altitude coverage. However, from the perspective of modeling completeness, incorporating observations from multiple satellites, including three or more, would be advantageous. In our future work, we plan to leverage a broader range of observational data to enhance the accuracy of our model predictions. Another weak point of our study is that we used 62 parameters as input for constructing our model. In fact, some of these parameters should be coupled. Therefore, performing a more detailed “sensitivity check” on these parameters is essential to make the model more convenient to the users, which is also the task for our next study.

## Appendix A

Specific information about the input data is given in Table A1 and Figure A1 gives the change in loss values for the training and test sets during the iteration process.

**Table A1**  
*The Detailed Information of the Input Data*

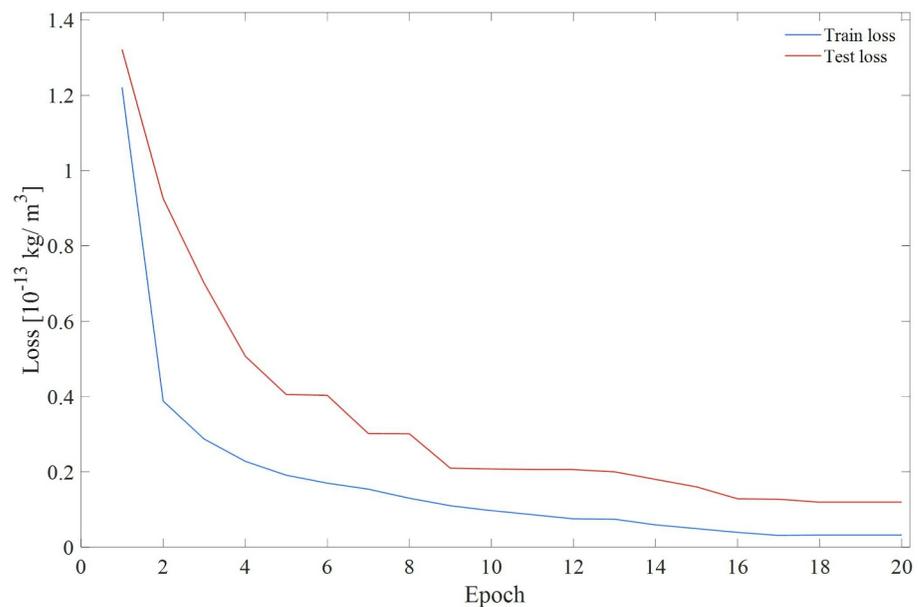
Swarm C data	1	Satellite altitude
	2	Geodetic longitude
	3	Geodetic latitude
	4	Local solar time
	5	Argument of latitude
OMNI web data	6	Bartels rotation number
	7	ID for IMF spacecraft
	8	ID for SW plasma spacecraft
	9	# of points in the IMF averages
	10	# of points in the plasma averages
	11	Field Magnitude Average IBI
	12	Magnitude of Average Field Vector

**Table A1**  
*Continued*

13	Lat.Angle of Aver. Field Vector
14	Long.Angle of Aver.Field Vector
15	Bx GSE, GSM
16	By GSE
17	Bz GSE
18	By GSM
19	Bz GSM
20	sigmaBI
21	sigma B
22	sigma Bx
23	sigma By
24	sigma Bz
25	Proton temperature
26	Proton Density
27	Plasma (Flow) speed
28	Plasma Flow Long. Angle
29	Plasma Flow Lat. Angle
30	Na/Np
31	Flow Pressure
32	sigma T
33	sigma N
34	sigma V
35	sigma phi V
36	sigma theta V
37	sigma-Na/Np
38	Electric field
39	Plasma beta
40	Alfven mach number
41	Kp
42	R
43	DST Index
44	AE-index
45	Proton flux (>1 Mev)
46	Proton flux (>2 Mev)
47	Proton flux (>4 Mev)
48	Proton flux (>10 Mev)
49	Proton flux (>30 Mev)
50	Proton flux (>60 Mev)
51	Flag
52	ap-index
53	F10.7_index
54	PC(N) index
55	AL-index, from Kyoto
56	AU-index, from Kyoto

**Table A1**  
*Continued*

	57	Magnetosonic Mach number = $V/\text{Magnetosonic\_speed}$
Solar angles data	58	solar zenith angle
	59	solar altitude angle
	60	solar azimuth angle
Time	61	year-month-second in the day
	62	hour



**Figure A1.** The size of the loss value varies with epoch.

### Data Availability Statement

The thermospheric mass density data derived from the accelerometer on board the Swarm C, CHAMP and GOCE satellites are processed and provided by the TU Delft (Doornbos, 2012; March et al., 2021; Siemes et al., 2016). The OMNI data can be downloaded from the Goddard Space Flight Center (Papitashvili & King, 2020). Detailed description of the NRLMSIS-2.0 model can be found and downloaded from <https://ccmc.gsfc.nasa.gov/models/NRLMSIS~2.0/>. The code for the model structure and the produced data set in this paper can be found at (Pan, 2024).

### Acknowledgments

This work is supported by the National Key R&D Program of China (Grant 2022YFF0503700), and the special found of Hubei LuoJia Laboratory (220100011). Chao Xiong is supported by the Dragon 5 cooperation 2020–2024 (project no. 59236).

### References

- Behzad, M., Asghari, K., Eazi, M., & Palhang, M. (2009). Generalization performance of support vector machines and neural networks in runoff modeling. *Expert Systems with Applications*, 36(4), 7624–7629. <https://doi.org/10.1016/j.eswa.2008.09.053>
- Billett, D. D., Perry, G. W., Clausen, L. B. N., Archer, W. E., McWilliams, K. A., Haaland, S., et al. (2021). The relationship between large scale thermospheric density enhancements and the spatial distribution of poynting flux. *Journal of Geophysical Research: Space Physics*, 126(5), e2021J–e29205J. <https://doi.org/10.1029/2021JA029205>
- 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. Paper presented at the Bayesian Deep Learning workshop, NeurIPS 2021, Vancouver, Canada.
- Bowman, B. (2002). *True satellite ballistic coefficient determination for HASDM*. AIAA/AAS Astrodynamics Specialist Conference and Exhibit. <https://doi.org/10.2514/6.2002-4887>
- 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*. AIAA/AAS Astrodynamics Specialist Conference. <https://doi.org/10.2514/6.2008-6438>

- Bruinsma, S. (2015). The DTM-2013 thermosphere model. *Journal of Space Weather and Space Climate*, 5, A1. <https://doi.org/10.1051/swsc/2015001>
- Bruinsma, S., & Boniface, C. (2021). The operational and research DTM-2020 thermosphere models. *Journal of Space Weather and Space Climate*, 11, 41–47. <https://doi.org/10.1051/swsc/2021032>
- Cagnini, H. E. L., Das Dóres, S. C. N., Freitas, A. A., & Barros, R. C. (2023). A survey of evolutionary algorithms for supervised ensemble learning. *The Knowledge Engineering Review*, 38, e1. <https://doi.org/10.1017/S0269888923000024>
- Cai, Y., Yue, X., Wang, W., Zhang, S., Liu, L., Liu, H., & Wan, W. (2019). Long-term trend of topside ionospheric electron density derived from DMSP data during 1995–2017. *Journal of Geophysical Research: Space Physics*, 124(12), 10708–10727. <https://doi.org/10.1029/2019JA027522>
- Chapman, S., & Cowling, T. G. (1990). The mathematical theory of non-uniform gases: An account of the kinetic theory of viscosity, thermal conduction, and diffusion in gases. *The Mathematical Gazette*, 38(323), 63–64. <https://doi.org/10.2307/3609795>
- Chen, Z., Jin, M., Deng, Y., Wang, J. S., Huang, H., Deng, X., & Huang, C. M. (2019). Improvement of a deep learning algorithm for total electron content maps: Image completion. *Journal of Geophysical Research: Space Physics*, 124(1), 790–800. <https://doi.org/10.1029/2018JA026167>
- Delashmit, W. H., & Manry, M. T. (2005). Recent developments in multilayer Perceptron neural networks. Paper presented at the 7th Annual Memphis Area Engineering and Science Conference, MAESC.
- Doornbos, E. (2012). *Thermospheric density and wind determination from satellite dynamics*. Springer Science & Business Media.
- Doornbos, E., Förster, M., Fritsche, B., Helleputte, T. V., Ijssel, J. V. D., Koppenwallner, G., et al. (2009). Air density models derived from multi-satellite drag observations. Paper presented at the ESAs Second Swarm International Science Meeting, Potsdam, Germany.
- Doornbos, E., van den Ijssel, J., Luhr, H., Förster, M., & Koppenwallner, G. (2010). Neutral density and crosswind determination from arbitrarily oriented multiaxis accelerometers on satellites. *Journal of Spacecraft and Rockets*, 47(4), 580–589. <https://doi.org/10.2514/1.48114>
- Emmert, J. T., Drob, D. P., Picone, J. M., Siskind, D. E., Jones, M., Mlynczak, M. G., et al. (2021). NRLMSIS 2.0: A whole-atmosphere empirical model of temperature and neutral species densities. *Earth and Space Science*, 8(3), e20221–e30896J. <https://doi.org/10.1029/2020EA001321>
- Emmert, J. T., Jones, M., Siskind, D. E., Drob, D. P., Picone, J. M., Stevens, M. H., et al. (2022). NRLMSIS 2.1: An empirical model of nitric oxide incorporated into MSIS. *Journal of Geophysical Research: Space Physics*, 127(10), e2022J–e30896J. <https://doi.org/10.1029/2022JA030896>
- Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115, 105151. <https://doi.org/10.1016/j.engappai.2022.105151>
- Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5), 602–610. <https://doi.org/10.1016/j.neunet.2005.06.042>
- Hedin, A. E. (1983). A revised thermospheric model based on mass spectrometer and incoherent scatter data: MSIS-83. *Journal of Geophysical Research*, 88(A12), 10170–10188. <https://doi.org/10.1029/JA088iA12p10170>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Kvammen, A., Wickstrøm, K., McKay, D., & Partamies, N. (2020). Auroral image classification with deep neural networks. *Journal of Geophysical Research: Space Physics*, 125(10), e2020JA027808. <https://doi.org/10.1029/2020JA027808>
- Lechtenberg, T., McLaughlin, C. A., Locke, T., & Krishna, D. M. (2013). Thermospheric density variations: Observability using precision satellite orbits and effects on orbit propagation. *Space Weather*, 11(1), 34–45. <https://doi.org/10.1029/2012SW000848>
- Lei, J., Huang, F., Chen, X., Zhong, J., Ren, D., Wang, W., et al. (2018). Was magnetic storm the only driver of the long-duration enhancements of daytime total electron content in the Asian-Australian sector between 7 and 12 September 2017? *Journal of Geophysical Research: Space Physics*, 123(4), 3217–3232. <https://doi.org/10.1029/2017JA025166>
- Lei, J., Li, R., Ren, D., Weng, L., & Ruan, H. (2023). Recent progress on the retrieval and modeling of thermosphere mass density. *Reviews of Geophysics and Planetary Physics*, 54(4), 434–454. <https://doi.org/10.19975/j.dqyxx.2022-047>
- Liu, H., Hirano, T., & Watanabe, S. (2013). Empirical model of the thermospheric mass density based on CHAMP satellite observations: Empirical Thermospheric Mass Density Model. *Journal of Geophysical Research: Space Physics*, 118(2), 843–848. <https://doi.org/10.1002/jgra.50144>
- Liu, R., Lühr, H., Doornbos, E., & Ma, S. Y. (2010). Thermospheric mass density variations during geomagnetic storms and a prediction model based on the merging electric field. *Annales Geophysicae*, 28(9), 1633–1645. <https://doi.org/10.5194/angeo-28-1633-2010>
- Lühr, H., Huang, T., Wing, S., Kervalishvili, G., Rauberg, J., & Korth, H. (2016). Filamentary field-aligned currents at the polar cap region during northward interplanetary magnetic field derived with the Swarm constellation. *Annales Geophysicae*, 34(10), 901–915. <https://doi.org/10.5194/angeo-34-901-2016>
- Lühr, H., Rother, M., Köhler, W., Ritter, P., & Grunwaldt, L. (2004). Thermospheric up-welling in the cusp region: Evidence from CHAMP observations. *Geophysical Research Letters*, 31(6). <https://doi.org/10.1029/2003GL019314>
- March, G., van den Ijssel, J. A. A., Siemes, C., Visser, P. N. A. M., Doornbos, E. N., & Pilinski, M. (2021). Gas-surface interactions modelling influence on satellite aerodynamics and thermosphere mass density. *Journal of Space Weather and Space Climate*, 11, 54. <https://doi.org/10.1051/swsc/2021035>
- Montavon, G., Orr, G., & Müller, K. (2012). *Neural networks: Tricks of the trade*. Springer.
- Müller, S., Lühr, H., & Rentz, S. (2009). Solar and magnetospheric forcing of the low latitude thermospheric mass density as observed by CHAMP. *Annales Geophysicae*, 27(5), 2087–2099. <https://doi.org/10.5194/angeo-27-2087-2009>
- Pan, Q. (2024). [Dataset] machine learning based modeling of thermospheric mass density. *Zenodo*. <https://doi.org/10.5281/zenodo.10695041>
- Papitashvili, N. E., & King, J. H. (2020). OMNI hourly data. *NASA Space Physics Data Facility*. <https://doi.org/10.48322/1shr-ht18>
- Santoni, F., Piergentili, F., & Graziani, F. (2010). Broglio Drag Balance for neutral thermosphere density measurement on UNICubeSAT. *Advances in Space Research*, 45(5), 651–660. <https://doi.org/10.1016/j.asr.2009.10.001>
- Schwenker, F. (2013). Ensemble methods: Foundations and algorithms. *IEEE Computational Intelligence Magazine*, 8(1), 77–79. <https://doi.org/10.1109/MCI.2012.2228600>
- Siemes, C., De Teixeira Da Encarnação, J., Doornbos, E. N., van den Ijssel, J. A. A., Kraus, J., Perešty, R., et al. (2016). Swarm accelerometer data processing from raw accelerations to thermospheric neutral densities. *Earth Planets and Space*, 68(1), 1–16. <https://doi.org/10.1186/s40623-016-0474-5>
- Smirnov, A., Shprits, Y., Prol, F., Lühr, H., Berrendorf, M., Zhelavskaya, I., & Xiong, C. (2023). A novel neural network model of Earth's topside ionosphere. *Scientific Reports*, 13(1), 1303. <https://doi.org/10.1038/s41598-023-28034-z>
- Storz, M. F., Bowman, B. R., Branson, M. J. I., Casali, S. J., & Tobiska, W. K. (2005). High accuracy satellite drag model (HASDM). *Advances in Space Research*, 36(12), 2497–2505. <https://doi.org/10.1016/j.asr.2004.02.020>
- Sutton, E. K. (2018). A new method of physics-based data assimilation for the quiet and disturbed thermosphere. *Space Weather*, 16(6), 736–753. <https://doi.org/10.1002/2017SW001785>

- 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, S., Weng, L., Fang, H., Xie, Y., & Yang, S. (2014). Intra-annual variations of the thermospheric density at 400 km altitude from 1996 to 2006. *Advances in Space Research*, *54*(3), 327–332. <https://doi.org/10.1016/j.asr.2013.12.011>
- Weng, L., Lei, J., Sutton, E., Dou, X., & Fang, H. (2017). An exospheric temperature model from CHAMP thermospheric density. *Space Weather*, *15*(2), 343–351. <https://doi.org/10.1002/2016SW001577>
- 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>
- Xiong, C., Lühr, H., Schmidt, M., Bloßfeld, M., & Rudenko, S. (2018a). An empirical model (CH-Therm-2018) of the thermospheric mass density derived from CHAMP. *Annales Geophysicae*, 1–20. <https://doi.org/10.5194/angeo-2018-25>
- Xiong, C., Lühr, H., Schmidt, M., Bloßfeld, M., & Rudenko, S. (2018b). An empirical model of the thermospheric mass density derived from CHAMP satellite. *Annales Geophysicae*, *36*(4), 1141–1152. <https://doi.org/10.5194/angeo-36-1141-2018>
- Xiong, C., Lühr, H., & Yamazaki, Y. (2019). An opposite response of the low-latitude ionosphere at Asian and American sectors during storm recovery phases: Drivers from below or above. *Journal of Geophysical Research: Space Physics*, *124*(7), 6266–6280. <https://doi.org/10.1029/2019JA026917>
- Zhang, Y., Paxton, L. J., Schaefer, R., & Swartz, W. H. (2022). Thermospheric conditions associated with the loss of 40 starlink satellites. *Space Weather*, *20*(10), e2022SW003168. <https://doi.org/10.1029/2022SW003168>
- Zhelavskaya, I. S., Aseev, N. A., & Shprits, Y. Y. (2021). A combined neural network-and physics-based approach for modeling plasmasphere dynamics. *Journal of Geophysical Research: Space Physics*, *126*(3), e2020JA028077. <https://doi.org/10.1029/2020JA028077>
- Zhou, Y. L., Ma, S. Y., Lühr, H., Xiong, C., & Reigber, C. (2009). An empirical relation to correct storm-time thermospheric mass density modeled by NRLMSISE-00 with CHAMP satellite air drag data. *Advances in Space Research*, *43*(5), 819–828. <https://doi.org/10.1016/j.asr.2008.06.016>