The Short-time Prediction of Thermospheric Mass Density Based on Ensemble-Transfer Learning

Peian Wang[1,2,3], Zhou Chen[2,4,5,*], Xiaohua Deng[1,2], Jingsong Wang[5], Rongxing Tang[2,3], Haimeng Li[2], Sheng Hong[2], and Zhiping Wu[6]

[1] School of Resources & Environmental, Nanchang University, Nanchang, 330031, China

[2] Institute of Space Science and Technology, Nanchang University, Nanchang, 330031, China

[3] Jiangxi Provincial Key Laboratory of Interdisciplinary Science, Nanchang University, Nanchang, 330021, China

[4] Department of Physics, University of Texas at Arlington, Arlington TX 76019, USA

[5] Key Laboratory of Space Weather, National Center for Space Weather, China Meteorological Administration, Beijing 100081, China

[6] Computing Institute of Jiangxi Province, Nanchang, Jiangxi, China

*Corresponding Author: Zhou Chen, chenzhou760@foxmail.com

Key points:

- Based on three different machine-learning algorithms, we present a robust short-time prediction for thermospheric mass density.
- We evaluate the model performances using data from CHAMP, GOCE, and SWARM-C to compare with NRLMSISE-00, and the RMSE decrease is up to 94.8%.
- The LightGBM ensemble model of Bi-LSTM and Transformer outperforms others and can provide reliable transfer predictions for new satellites.

Abstract

Reliable short-time prediction of thermospheric mass density along the satellite orbit is always essential but challenging for the operation of Low-Earth orbit (LEO) satellites. In this paper, three machine-learning prediction algorithms are investigated, including the Bidirectional Long Short-Term Memory (Bi-LSTM), the Transformer, and the Light Gradient Boosting Machine (LightGBM) ensemble model of the above models. We use satellite data from CHAMP, GOCE, and SWARM-C to evaluate the robustness and accuracy of different density variations. The comparison demonstrates that all models achieve compelling predictions and are much better than NRLMSISE-00. The LightGBM ensemble model (LE-model) consistently outperforms others in accuracy and stability. Furthermore, when the obtained density data from the newly launched satellites are limited, the trained LE-model can provide a valid prediction for the new satellite orbit by transfer learning. This study offers a promising insight into the short-time prediction of thermospheric mass density using ensemble-transfer learning and may be advantageous to future research on space whether.

Plain Language Summary

Low-Earth orbit (LEO) satellites play an important role in many aspects, such as navigation, aerospace, military industry, and so on. The LEO satellites suffer atmospheric drag caused by thermospheric mass density. Therefore, we present three different machine-learning algorithms to achieve a robust short-time prediction for thermospheric mass density. All models can provide effective results from testing with

three satellite data, and the ensemble model always outperforms others. Then, when the obtained density data from the newly launched satellites are very limited, the pre-trained ensemble model is also useful for the new satellite orbit by transfer learning. We offer a good insight into the short-time prediction of thermospheric mass density and assistance for aerospace digitization and intellectualization.

1. Introduction

Thermospheric mass density is an essential parameter in the earth's thermosphere and can cause atmospheric drag force on LEO satellites. As a typical non-conservative force, atmospheric drag is the fundamental perturbation source of satellites, and it has a constant and considerable effect on the satellites, such as orbit determination, collision warning, and motion safety (Doornbos, 2012). Meanwhile, extreme space environment conditions can lead to complex and variable responses of thermosphere mass density, significantly impacting satellites' orbit (Lei et al., 2013; Emmert, 2015). For instance, in February 2022, 49 Starlink satellites suffered from a moderate geomagnetic storm (Kp=5) during orbit raising. In a short time, atmospheric drag increased by 50% because of increased density and declined orbit. In the end, this event caused 40 satellites to fall out of orbits. Therefore, with the growing importance of LEO satellites in navigation, communication, aerospace, military industry, emergency response, and commercial applications, short-time prediction of thermosphere mass density along the satellite orbit is indispensable.

With the rapid development of artificial intelligence (AI) technology and the continuous increase of observations in space physics, machine learning, as one kind of

data-driven methods, provides a promising way for researchers. Thereinto, deep learning technology (Lecun et al., 2021) based on neural network (NN) has shown its powerful data-learning ability and has been widely applied in various aspects of the space weather forecast, such as geomagnetic index (Shprits et al., 2019; Xu et al., 2020; Tan et al., 2018; Gruet et al., 2018), solar activity (Fang et al., 2019; Tang et al., 2021(a); Tang et al., 2021(b)), Total Electron Content (TEC) (Chen et al., 2019; Pan et al., 2020; Liu et al., 2020; Tang et al., 2020; Chen et al., 2022), electron flux (Pires de Lima et al., 2020; Tang et al., 2022), as well as NO emission (Chen et al., 2021). While related studies are far more than those listed above, these works have proved that machine-learning methods have been intensively studied in space weather forecasts.

As for thermospheric mass density, NN also has been gradually used in recent years. Wang et al. (2014) utilized Artificial Neural Network (ANN) to investigate intra-annual variations at a fixed altitude for ten years. Weng et al. (2020) revisited the average variation trend during either 1967-2005 or 1967-2013 from 250 km to 575 km based on ANN. These works utilized the basic NN to focus on the long-term trend and states by using large-scale physical parameters. On the other hand, Perez et al. (2014) proposed the orbit-concerned prediction by ANN, and then Perez and Bevilacqua (2015) presented two time-delay prediction approaches based on NN. These works mainly tried to use current external parameters and density values to realize the window prediction along the satellite orbit. Recently, Wang et al. (2022) used near-real-time parameters to provide a deep-learning algorithm based on the

LSTM (Long Short-Term Memory)-based ensemble learning, which paid more attention to storm-time prediction.

Different from the past methods, in this work, we try to present a short-time prediction with high-precision and high-fidelity for thermospheric mass density. We design three algorithms based on the Bi-LSTM, the Transformer, and the LightGBM ensemble model of the above models using ensemble learning for two kinds of multi-step prediction. The prediction performances are compared with CHAMP, GOCE, and SWARM-C to verify the robustness and accuracy. The results demonstrate that reliable predictions can be achieved in all models, and the LE-model consistently outperforms other standalone models in terms of accuracy and stability. Subsequently, when the available density data from the newly launched satellites are limited, the trained LE-model is also suitable for neutral density prediction along new satellite orbit using transfer learning. This lightweight method saves time and computing resources while maintaining accuracy and offers beneficial assistance for satellite-borne intellectualization research. Additionally, the NRLMSISE-00 has been used in the comparative analysis of the whole verification process. This study gives a promising insight into the short-time prediction of thermospheric mass density based on machine-learning algorithms, which can be critical for practical applications to the related research of LEO satellites.

This paper's structure is as follows: The descriptions of the datasets are briefly introduced in Section 2. Section 3 illustrates the model algorithms. In Section 4, the corresponding experimental results and analyses are introduced. Finally, the

conclusion and directions for future work are summarized in Section 5.

2. Data Description

In this study, we utilize the accelerometer-derived (ACC) data from TU Delft (http://thermosphere.tudelft.nl/) as datasets, which are quasi-instantaneous along orbits and closer to the actual values than the precise orbit determination (POD) data (Doornbos, 2012; Siemes et al., 2016; March et al., 2019). The ACC data used comprise the ACC density and corresponding local parameters (altitude, longitude, and latitude). Moreover, the external physical parameters ($F_{10.7}$, Ap, and DOY) from OMNI data (https://spdf.gsfc.nasa.gov/pub/data/omni/low_res_omni/) are added to datasets since the dynamic natural variabilities need to be considered. To verify the algorithm's usability and broader applicability, three different datasets from CHAMP, GOCE and SWARM-C are utilized. Each dataset is partitioned three times to get test-1/2/3, which there are 9 tests in total. The tests are sequential periods that keep the same test samples (100000 for CHAMP, 120000 for GOCE, and 300000 for SWARM-C). Table 1 depicts the detailed information of three satellites in the datasets used. Each test-1 period encompasses one moderate storm. The storm event is selected when the minimum Dst value is less than -50 nT during the main phase, and the Dst value of -10 nT is used as the threshold for the start of the initial phase and the ending of the recovery phase.

Table 1 The detailed information of three satellites in the datasets used.

	CHAMP	GOCE	SWARM-C
Altitude range (km)	257-487	224-295	433-528

Temporal resolution (sec)	30	10	10
Data points	9579665	13547435	17364276
Training-1 numbers	9479665	13427435	17064276
Test-1 numbers	100000	120000	300000
Training-1 time range	Start: 05/04/2001 End: 07/31/2010	Start: 11/01/2009 End: 10/06/2013	Start: 02/01/2014 End: 08/24/2020
Test-1 time range	Start: 07/31/2010 End: 09/04/2010	Start: 10/06/2013 End: 10/20/2013	Start: 08/24/2020 End: 09/30/2020
Training-2 numbers	9379665	13307435	16764276
Test-2 numbers	100000	120000	300000
Training-2 time range	Start: 05/04/2001 End: 06/26/2010	Start: 11/01/2009 End: 09/20/2013	Start: 02/01/2014 End: 07/15/2020
Test-2 time range	Start: 06/26/2010 End: 07/31/2010	Start: 09/20/2013 End: 10/06/2013	Start: 07/15/2020 End: 08/24/2020
Training-3 numbers	9279665	13187435	16464276
Test-3 numbers	100000	120000	300000
Training-3 time range	Start: 05/04/2001 End: 05/23/2010	Start: 11/01/2009 End: 09/06/2013	Start: 02/01/2014 End: 06/06/2020
Test-3 time range	Start: 05/23/2010 End: 06/26/2010	Start: 09/06/2013 End: 09/20/2013	Start: 06/06/2020 End: 07/15/2020

For the time-step setting, the serial 10 data samples (10 time steps) are prepared as the historical data, almost 300 seconds long for CHAMP and 100 seconds long for GOCE and SWARM-C. Then, the serial N samples (N=6/10 time steps) are predicted, which roughly correspond to 180/300 seconds for CHAMP, 60/100 seconds for GOCE, and SWARM-C backward. Therefore, when the current time is T, the historical samples contain a total of 10 data samples from T-10 to T-1, while the predictions contain N data samples from T to T+N-1. Figure 1 depicts the detailed time-step settings for the three satellites based on their temporal resolutions.

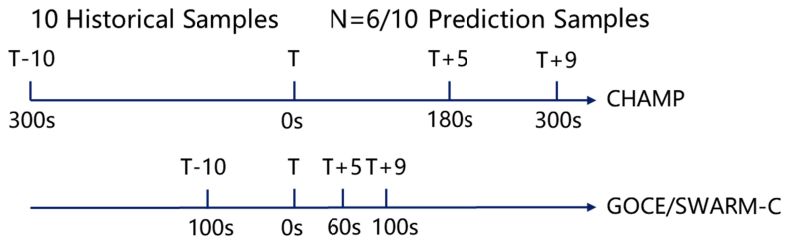


Fig 1 The detailed settings of the time steps according to the temporal resolutions of three different satellites.

In the test of transfer learning, the pre-trained model uses the LE-model from GOCE data, and the retraining data are from SWARM-C data. Two different cases are chosen to manifest the reliable prediction results for the quiet and storm periods. The quiet case possesses 45000 data samples for retraining (about seven days) and 10000 data samples for testing. The storm case has the same data samples for retraining, but it has 22148 data samples for the test. This case is a moderate storm (the minimum value of Dst is -56 nT), including the main and recovery phases. Table 2 provides the detailed information of two cases for the tests of transfer learning.

Table 2 The detailed information of two cases for the tests of transfer learning.

	Quiet case	**Storm case**
Retraining numbers	45000	45000
Test numbers	10000	22148
Retraining time range	Start: 17:45:50,03/02/2020 End: 17:47:00,03/08/2020	Start: 15:28:30,02/17/2015 End: 05:00:00,02/23/2015
Test time range	Start: 17:47:10,03/08/2020 End: 23:59:50,03/09/2020	Start: 05:00:10,02/23/2015 End: 18:35:50,02/25/2015

3. The Model Algorithms

The algorithms used in this study are respectively based on the Bi-LSTM, the Transformer, and the LightGBM ensemble model by using ensemble learning. And

transfer learning is utilized to predict SWARM-C data using the model trained by GOCE data. Figure 2 (a) provides the details of the three models, and Figure 2 (b) illustrates the transfer learning process in our study.

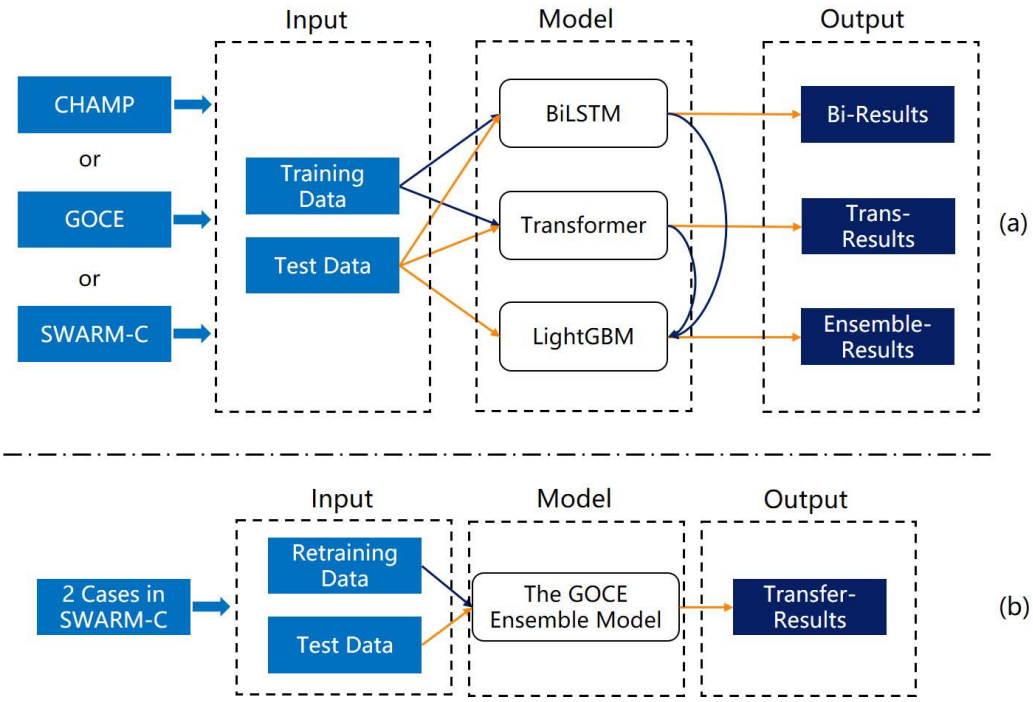


Fig 2 (a) The details of three models. (b) The process of transfer learning in this study.

The Bi-LSTM network (Siami et al., 2019; Chen et al., 2022), as an improved version of LSTM, adds the reverse LSTM to realize the bidirectional function in the learning process. If there is just normal LSTM, the outputs only learn the one-sided evolutionary laws from the time steps of the input sequence. As for Bi-LSTM, besides forwarding processing, it makes the prediction information pass and update from the back end to the beginning of the input sequence, which can learn the two-sided evolutionary laws. The additional traversal training can effectively increase the amount of information the network can use. Therefore, the Bi-LSTM can provide

169 better predictions compared to LSTM (Siami et al., 2019), and has been successfully
170 applied in related space weather forecasting, like TEC prediction (Chen et al., 2022).
171 In our Bi-LSTM model, there are three Bi-LSTM layers and three dense layers.

172 Compared to the Bi-LSTM, the Transformer is an entirely different deep neural
173 network with better parallelism. It relies on self-attention to compute representations
174 of inputs and outputs without using sequence-aligned RNNs or convolution (Vaswani
175 et al., 2017). In essence, the Transformer is the Encoder-Decoder architecture with
176 positional encoding. The Encoder layer includes a multi-head attention sublayer and a
177 Feed Forward Neural Network (FFNN) sublayer. After each sublayer, there is a
178 data-processing module named Add&Norm, which means the residual transformation
179 and layer normalization. The Decoder layer is similar to the Encoder layer and has
180 three sublayers with three Add&Norm modules, including the masked multi-head
181 attention, the multi-head attention, and the FFNN. Therefore, the Transformer can
182 meet the diversity requirement of the base learners in stacking ensemble learning.
183 Meanwhile, it also has good accuracy in time sequence prediction (Lim et al., 2021;
184 Wu et al., 2020; Wu et al., 2020). In our Transformer model, there are three Encoder
185 layers and three Decoder layers.

186 The ensemble model uses heterogeneous stacking ensemble learning based on
187 LightGBM (Liang et al., 2022). The heterogeneous stacking ensemble learning can
188 build several base learners using different base-learning models (Bi-LSTM and
189 Transformer in our study), then the outputs of different base learners are used as the
190 inputs of the meta learner for ensemble training using another model (LightGBM).

Generally, the ensemble model can learn respective advantages from different valid base learners to get better performance than one single model. But sometimes, on the contrary, if there are large performance gaps among the base learners, the ensemble model may become mediocre. And the LightGBM, a lightweight gradient elevator, can divide sequential floating-point features into k discrete values. Then, a histogram with the width of k is created, and the cumulative statistics of each discrete value are calculated, which converts the data traversing to the histogram traversing. Thus, the LightGBM can reduce the computation time and storage space of the algorithm on the premise of ensuring accuracy (Tang et al., 2022).

Transfer learning (Pan and Qiang, 2010; Farrens et al., 2021) is a machine-learning method that can transfer the source domain's knowledge to the target domain to achieve a better learning effect. It aims to solve the problems that the training data or the data labels are arduous to obtain in the target domain. Meanwhile, the trained source model can save training time and computing resources with the help of the trained source model. According to the form of knowledge to be transferred, there are four kinds of transfer learning: instance-based, feature-representation, model (parameter)-based, and relational-knowledge. This study utilizes model-based transfer learning to offer valid short-time predictions based on pre-trained models from past satellite data when the new satellite data are limited.

4. The Prediction Results and Discussions

The empirical model, NRLMSISE-00, is used as a unified reference during the whole model comparison process. The parameter inputs consist of the DOY, seconds

since the start of the day, longitude, latitude, altitude, $F_{10.7}$, $F_{10.7A}$ (81-day sliding averaged value of $F_{10.7}$), and Ap. Furthermore, to accurately assess the models, the root mean square error (RMSE), and the coefficient of determination (R^2) are utilized as evaluating standards in this study. They are defined in the following investigations:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum(P_i - O_i)^2} \quad (1)$$

$$R^2 = 1 - \frac{\sum_i (P_i - O_i)^2}{\sum_i (O_i - \bar{O})^2} \quad (2)$$

In detail, P_i is the predicted density, and O_i is the observed density. $\bar{P}$ and $\bar{O}$ are respectively the mean density values of P_i and O_i. And n is the sample number.

4.1. The Results of Model Comparison

In this section, to verify the robustness and accuracy of the model algorithms, the prediction performances are detailly compared with three different satellite data, including CHAMP, GOCE, and SWARM-C.

The test-1 of CHAMP is from July 31 to September 4, 2010, in which the altitude of satellite orbit is from 247 km to 295 km, and the minimum value of the Dst index is -74 nT. During this period, the solar activity is low since $F_{10.7}$ is from 73.3 sfu to 93 sfu (the average value is 81.3 sfu). As shown in Figure 3, the predicted results from three kinds of machine-learning models (see the red points) and NRLMSISE-00 (see the blue points) are compared with observation. The prediction from NRLMSISE-00 is as a reference at the leftmost, then the columns from left to right depict the predictions from the Bi-LSTM model (Bi-model), the Transformer model (Trans-model), and the LE-model, respectively, while the first row is for N=6 time steps, and the second is for 10 time steps. It can be observed that the three kinds of

models all obtain good prediction performances when N=6 or 10 time steps since most of the data points are exceedingly close to the black line, which is much better than NRLMSISE-00. From the evaluating indicators, the advantage of machine learning is overwhelmingly apparent since it can learn the inherent features of complicated response variations from lots of satellite data. For the value of RMSE, all the machine-learning models are no more than 1.529×10^{-12}, but NRLMSISE-00 is close to 7.254×10^{-12}. For the value of R^2, these models are larger than 0.960, but NRLMSISE-00 is just 0.109. On the other hand, if the results are checked in rows, better prediction performance is always obtained when N=6 time steps. In general, the prediction error increases along with the increasing time step in the multi-step prediction. As for the more detailed comparison among the machine-learning models, the performance difference among the three machine-learning models is very small. But the LE-model takes a little advantage because of the decrease of RMSE value and the increase of R^2 value from left to right. In addition, from the accuracy advantage of the LE-model, more results of model fusion can be expected.

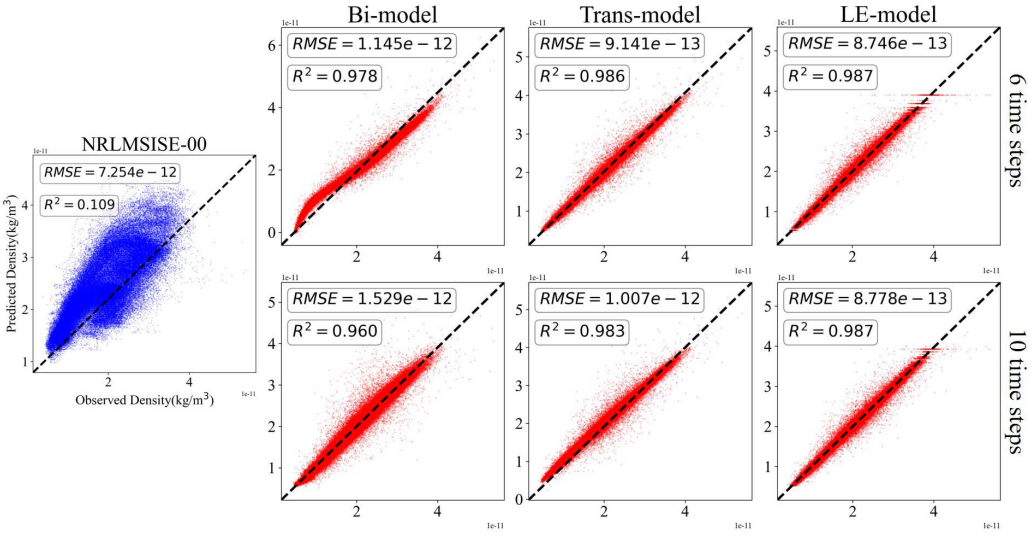


Fig 3 The comparison of the machine-learning models (red) and NRLMSISE-00 (blue) with CHAMP observation during the period (-74 nT≤Dst≤44 nT) from July 31 to September 4, 2010.

Similarly, the models display their corresponding prediction results of GOCE's test-1 in Figure 4. The test period (from October 6 to 20, 2013) also includes a moderate storm in that the minimum Dst value is -69 nT, but the solar activity becomes moderate level, which $F_{10.7}$ is from 106.5 sfu to 138.8 sfu (the averaged value is 124 sfu). Meanwhile, the altitude of satellite orbit (from 224 km to 261 km) is slightly lower than the CHAMP test. The GOCE results confirm that all machine-learning models can supply effective prediction results, and the evaluating indicators have even improved than those in the CHAMP test since most of the data points are closer to the black line. For the value of RMSE, the machine-learning models are in the range of 1.666×10^{-12} to 2.581×10^{-12}, but NRLMSISE-00 is 1.199×10^{-11}. As for the value of R^2, these models are from 0.988 to 0.995, but NRLMSISE-00 is just 0.731. Expectedly, NRLMSISE-00 is still far behind the machine-learning models, although it has a significant improvement from before, which can be proved by the increase of R^2 value (from 0.109 to 0.731). Importantly, the conclusions in the CHAMP results are suggested again in the GOCE results. On the one hand, when N=6 time steps, the prediction performance is always better whether for the Bi-model, the Trans-model, or the LE-model. On the other hand, although the results are quite close, the LE-model invariably has the best prediction performance. In addition, it should be noted that the improvement of the LE-model becomes more limited in this test than in the CHAMP test since the results of the

Bi-model and the Trans-model are good sufficiently.

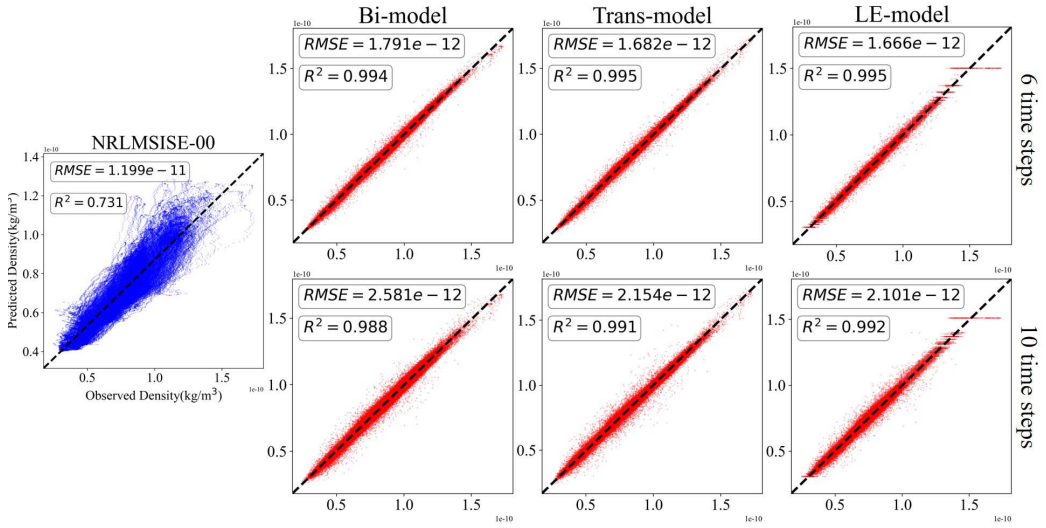


Fig 4 The comparison of the machine-learning models (red) and NRLMSISE-00 (blue) with GOCE observation during the period (-69 nT≤Dst≤17 nT) from October 6 to 20, 2013.

The test-1 of SWARM-C is from August 24 to September 30, 2020, and the minimum value of the Dst index (-59 nT) is still similar to the above two tests. During this period, the solar activity is very low, which $F_{10.7}$ is from 69.5 sfu to 74.4 sfu (the average value is 71.5 sfu). However, the altitude of satellite orbit has roughly risen 200 km, which reaches the range of 433 km to 464 km. Consequently, the density values decrease several orders of magnitude compared with before. Figure 5 displays the detailed comparison. In this test, the NRLMSISE-00 gives worse results (RMSE: 1.527×10^{-13}, R^2: -1.948) than before in CHAMP (R^2: 0.109) or GOCE (R^2: 0.731). Oppositely, the machine-learning models still maintain a superior level of performance (RMSE: 1.077×10^{-14} - 1.697×10^{-14}, R^2: 0.964 - 0.985). In more detail, different from the above tests, the Bi-model first precedes the Trans-model when N=6 time steps. But the final comparison result has not changed that the LE-model of 6

time steps continuously outperforms other models in terms of accuracy and stability.

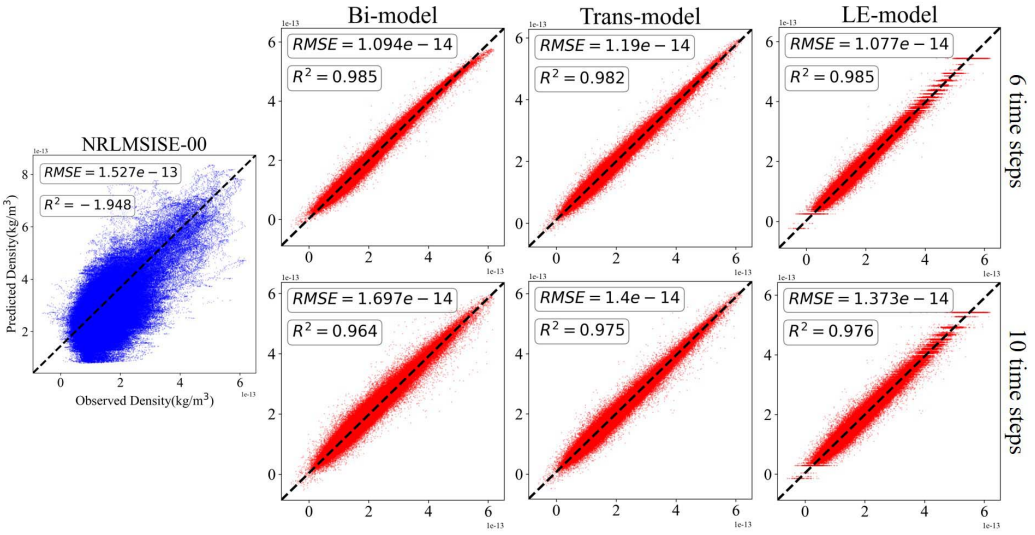


Fig 5 The comparison of the machine-learning models (red) and NRLMSISE-00 (blue) with SWARM-C observation during the period (-59 nT≤Dst≤14 nT) from August 24 to September 30, 2020.

From the qualitative analysis of three satellite datasets, the machine-learning models give satisfactory prediction results with different periods, altitudes, and response variations. Subsequently, it is also necessary to further verify the prediction performances by statistical analysis. In Figure 6, there is the RMSE decrease between each model and NRLMSISE-00 in the three different tests. It is visible that these models are much more advanced than NRLMSISE-00, in which all decrease values of RMSE are larger than 78.5%, and the best value is 93.0%. More importantly, the LE-model invariably has the best performance for the two kinds of multi-step predictions. In detail, the RMSE decreases of the LE-model are followed by GOCE (86.1%, 82.5%)<CHAMP (87.9%, 87.8%)<SWARM-C (93.0%, 91.0%) for N=6/10 time step. However, on the contrary, the LE-model performances (R^2 as the example) from above tests are followed by SWARM-C (0.985, 0.976)<CHAMP (0.987,

0.987)<GOCE (0.995, 0.992) for N=6/10 time step. Actually, it needs to be noted that this sorting order of the RMSE decrease is under the huge influence of the NRLMSISE-00 results that are followed by SWARM-C (R^2: -1.948)<CHAMP (R^2: 0.109)<GOCE (R^2: 0.731) from the above tests. Although NRLMSISE-00 constantly updates space weather indices to conduct model correction, it still often gives unsatisfactory simulation results like these.

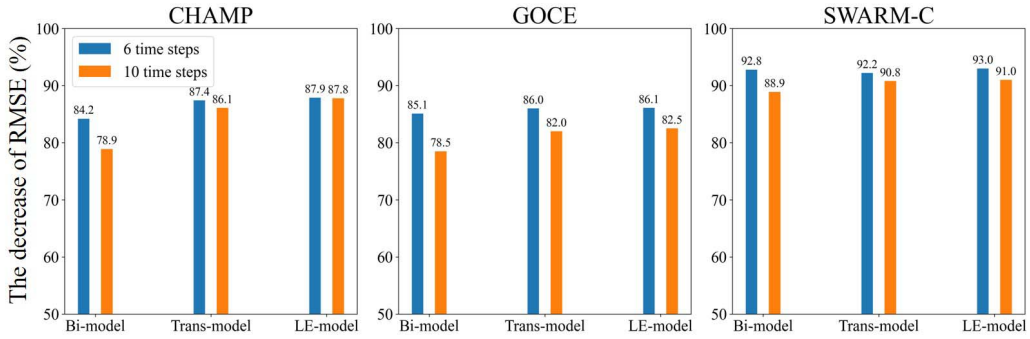


Fig 6 The decrease of RMSE (%) between each machine-learning model and NRLMSISE-00 for N=6/10 time steps in different satellite datasets during the corresponding test period.

To make the statistical analysis more convincible, the results of test-2 and test-3 in each satellite dataset also show in Table 3 and Table 4. In the test-2, the R^2 is from 0.928 to 0.995, and the decrease of RMSE is from 79.0% to 92.9%. In the test-3, the R^2 is from 0.944 to 0.997, and the decrease of RMSE is from 81.7% to 94.8%. Importantly, the LE-model consistently has the best average prediction performance since it can outperform the other two models when the test conditions are constantly changing.

Table 3 The evaluation results in test-2 of three satellite datasets.

Test-2	Time Step	Model	R^2	RMSE	Decrease(%)
CHAMP	6	Bi-model	0.983	7.163E-13	89.6
		Trans-model	0.985	6.696E-13	90.3
		LE-model	0.986	6.482E-13	90.6
	10	Bi-model	0.947	1.260E-12	81.7
		Trans-model	0.949	1.228E-12	82.2
		LE-model	0.954	1.168E-12	83.1

GOCE	6	Bi-model	0.993	1.903E-12	79.0
		Trans-model	0.993	1.765E-12	80.6
		LE-model	0.995	1.499E-12	83.5
	10	Bi-model	0.993	1.897E-12	79.1
		Trans-model	0.994	1.655E-12	81.8
		LE-model	0.995	1.597E-12	82.4
SWARM-C	6	Bi-model	0.959	1.425E-14	92.2
		Trans-model	0.964	1.350E-14	92.6
		LE-model	0.966	1.299E-14	92.9
	10	Bi-model	0.928	1.899E-14	89.6
		Trans-model	0.944	1.669E-14	90.8
		LE-model	0.958	1.443E-14	92.1

Table 4 The evaluation results in test-3 of three satellite datasets.

Test-3	Time Step	Model	R^2	RMSE	Decrease(%)
CHAMP	6	Bi-model	0.985	4.509E-13	90.5
		Trans-model	0.985	4.449E-13	90.6
		LE-model	0.986	4.414E-13	90.7
	10	Bi-model	0.944	8.718E-13	81.7
		Trans-model	0.949	8.318E-13	82.5
		LE-model	0.956	7.675E-13	83.9
GOCE	6	Bi-model	0.994	1.530E-12	85.1
		Trans-model	0.995	1.388E-12	86.5
		LE-model	0.997	9.666E-13	90.6
	10	Bi-model	0.993	1.666E-12	83.8
		Trans-model	0.994	1.539E-12	85.0
		LE-model	0.995	1.381E-12	86.5
SWARM-C	6	Bi-model	0.964	1.470E-14	93.8
		Trans-model	0.967	1.401E-14	94.0
		LE-model	0.974	1.234E-14	94.8
	10	Bi-model	0.952	1.682E-14	92.9
		Trans-model	0.957	1.602E-14	93.2
		LE-model	0.964	1.468E-14	93.8

4.2. The Results of Transfer Learning

For the newly launched LEO satellites, the obtained data along the new satellite orbits are limited, and the available models could not be trained with sufficient data to perform accurate predictions. To solve the problem, based on transfer learning, we utilize the effective pre-trained machine-learning model by large amounts of data

from the past satellites to retrain the obtained data from the new satellite with fine-tuning. On the one hand, this lightweight method can provide a new viewpoint for the short-time prediction of thermospheric mass density along orbits, which reduces computation complexity and dramatically saves time and computing resources while maintaining accuracy. On the other hand, it may supply helpful assistance for future satellite-borne applications of AI technologies. With the in-depth development of aerospace digitization and intellectualization, satellite-borne software systems are improving to support on-board data processing, AI algorithms, and independent adjustment and decision-making, like Project Blackjack (Keller, 2018). Meanwhile, smart microchips (such as CPU, FPGA, DSP, and SoC) for satellite payloads are under accelerated studying, enhancing computing performance while having radiation-resistant ability, low energy consumption, and small physical dimensions. It is foreseeable that machine-learning methods, deep-learning methods, and other AI technologies will soon be applied to related payloads carried by LEO satellites and other spacecrafts.

Therefore, in this section, we choose the pre-trained LE-model from GOCE data with 6 time steps to test two cases from SWARM-C data by transfer learning. In both cases (a quiet case and a storm case), about seven days of data from the new satellite (the temporal resolution is 10 seconds) are prepared as retrained data. The whole quiet case is shown in Figure 7, which describes the compared details of predicting variation from March 8 to 9, 2020. During this quiet period, the Dst is between a minimum value of -7 nT and a maximum value of 10 nT, and the $F_{10.7}$ is around 69 sfu.

Moreover, the altitude range is from 434 km to 465 km. From Figure 7, the LE-model (red line) achieves positive transfer results that are more consistent with the measured data (black line) than NRLMSISE-00 (blue line). Obviously, the estimations of NRLMSISE-00 are much larger than the measured data almost all the time. So, its evaluating indicators are seriously affected by this situation, which NRLMSISE-00 (RMSE: 1.260×10^{-13}, R^2: -0.987) is far behind the LE-model (RMSE: 1.088×10^{-14}, R^2: 0.985). The difference in RMSE is more than an order of magnitude, and the RMSE decrease of the LE-model is 91.4%. Meanwhile, the prediction performance has hardly any reduction since R^2 of LE-model, in this case, is extremely close to that in the test of SWARM-C (both are around 0.985). However, the training time has a massive reduction from about 3.5 hours to 5 minutes since transfer learning just needs a little new data to retrain, which saves a large amount of training time and computing resources. The pre-trained LE-model has an excellent prediction performance for the fundamental variation features of thermospheric mass density, especially the details of peaks and valleys. Therefore, the GOCE LE-model can rapidly learn valuable features of density variations along the new satellite orbits during quiet periods.

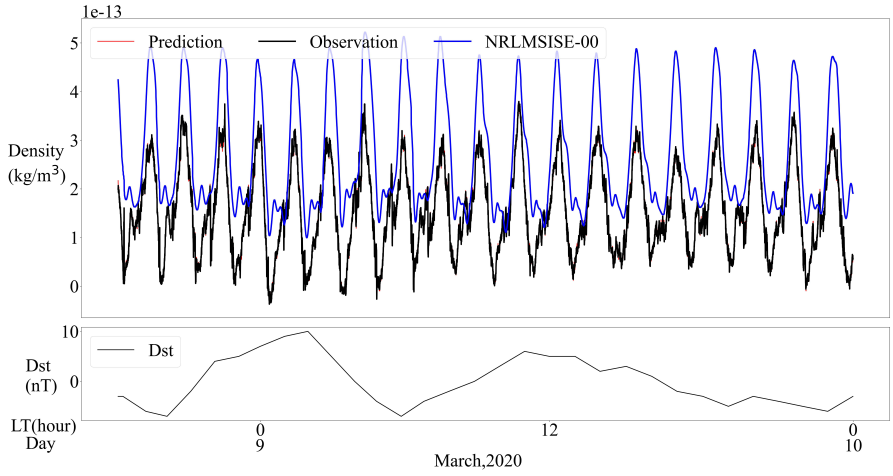


Fig 7 The detailed comparison of the LE-model prediction (red), the observation (black), and NRLMSISE-00 (blue) during a quiet period (-7 nT≤Dst≤10 nT) from March 8 to 9, 2020.

The storm case in transfer learning is from February 23 to 25, 2015, and the altitude range (from 456 km to 487 km) has marginally raised. During this moderate storm, the Dst index is between the minimum value of -56 nT and the maximum value of -10 nT, as well as the solar activity is also moderate since the F10.7 is around 111.7 sfu. Figure 8 describes a detailed comparison of this storm, including the main phase and recovery phase. It can be seen that, unlike the regularity of the quiet case above, the densities experience apparent disturbances. The NRLMSISE-00 estimations are much smaller than most measured data, which affects NRLMSISE-00 accuracy since the RMSE is 3.349×10^{-13} and the R^2 is -0.979. At this time, the LE-model still gets satisfactory results for storm responses in RMSE (3.847×10^{-14}) and R^2 (0.974), which shows powerful performance prediction advantages because of the RMSE decrease of 88.5%. And, as rapid as the transfer prediction of the above quiet case, the training time is just 6 minutes.

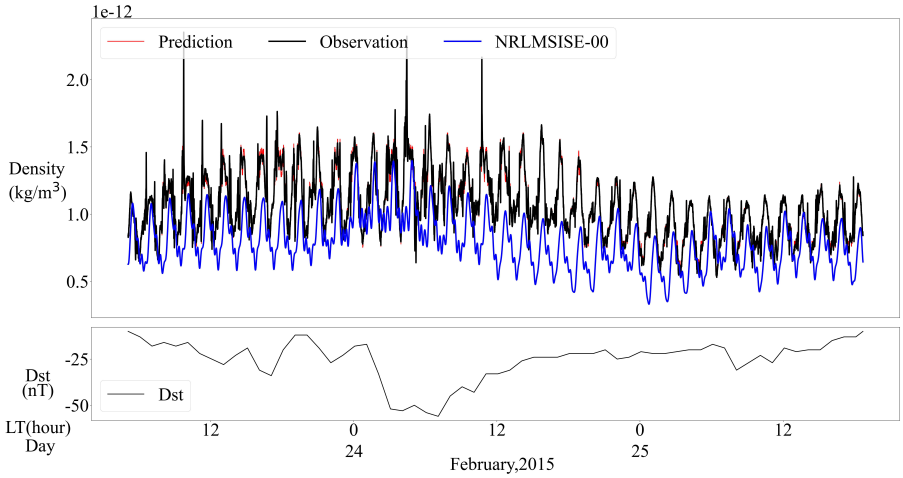


Fig 8 The detailed comparison of the LE-model prediction (red), the observation (black), and NRLMSISE-00 (blue) during a moderate storm period (-56 nT≤Dst≤-10 nT) from February 23 to 25, 2015.

5. Conclusion

This paper aims to apply machine-learning algorithms to provide a robust short-time prediction for thermospheric mass density with high fidelity. Therefore, we develop three different algorithms based on the Bi-LSTM, the Transformer, and the LightGBM ensemble model of the above two models using ensemble learning. The model performances for different density variations are verified using data from three satellites, including CHAMP, GOCE, and SWARM-C. Compared with NRLMSISE-00, the results reveal that these machine-learning models can achieve much better predictions with two kinds of multi-step prediction, and the LE-model consistently outperforms others in terms of accuracy and stability. Subsequently, when the obtained density data are very limited from the new LEO satellites, it is demonstrated that the pre-trained LE-model from GOCE data can accurately supply

transfer predictions for the new satellite orbits by transfer learning. This lightweight method with satisfactory accuracy can greatly reduce training time and computing resources and offer beneficial assistance for future research on satellite-borne smart systems and microchips. This study gives a promising way for the short-time prediction of thermospheric mass density based on ensemble and transfer learning, which may be critical for practical applications to the related research of space weather. The main conclusions are as follows:

1. We present three different machine-learning algorithms to achieve the short-time prediction for thermospheric mass density, respectively, based on the Bi-LSTM, the Transformer, and the LightGBM ensemble model of the above two models using ensemble learning.
2. Compared with NRLMSISE-00, these machine-learning models with N=6/10 time steps offer much better predictions for different density variations from CHAMP, GOCE, and SWARM-C. The decreased values of RMSE are from 78.5%-94.8%, and the LE-model always has the best performance.
3. In two different cases, the pre-trained LE-model with 6 time steps can provide accurate transfer predictions for the new satellite orbits by transfer learning. The lightweight method dramatically saves time and computing resources and benefits satellite-borne intellectualization research.

In future work, we attempt to find suitable weights of local parameters and external physical parameters to improve the model accuracy. Then, we will try to merge more network advantages by updating or adding new neural-network structures

to enrich the model algorithm, which may enhance learning ability. We also intend to lengthen the predicting time steps by testing multiple multi-step predictions. And, the forecasting lead time will be lengthened, like 1-hour or even longer time, to provide more references for practical applications. The horizontal lines when the density is high in the LE-model results are also future work. We will find out the reasons and optimize the model algorithm. The model robustness is important to be verified in the future for periods of strong geomagnetic storms with much sharper changes in the external physical parameters. And, the model methods should compare with previous deep-learning models to increase the scientific relevance. Meanwhile, we will add other models, such as GITM, TIEGCM, to produce simulation values as reference. Furthermore, we will try to simulate the more accurate global distributions of thermospheric mass density using machine-learning methods to compare with empirical and theoretical models.

Acknowledgments

National Natural Science Foundation of China supports this work under grants 41974183 and 42064009. The authors thank these grants for underpinning. The ACC data used in this paper are from TUDelft (http://thermosphere.tudelft.nl/). The OMNI data are from the Goddard Space Flight Center, online at https://spdf.gsfc.nasa.gov/pub/data/omni/. The authors thank these organizations for making their data public.

References

[1] C. Siemes, J. de Teixeira da Encarnação, E. N. Doornbos, J. van den IJssel, J. Kraus, R. Pereštý, L. Grunwaldt, G. Apelbaum, J. Flury and P. E. Holmdahl Olsen (2016) Swarm accelerometer data processing from raw accelerations to thermospheric neutral densities. Earth, Planets and Space, 68: 92. https://doi.org/10.1186/s40623-016-0474-5

[2] Chen, X., Lei, J., Ren, D., & Wang, W. (2021). A deep learning model for the thermospheric nitric oxide emission. Space Weather, 19, e2020SW002619.

[3] 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 (TEC) maps: image completion. Journal of Geophysical Research: Space Physics. doi:10.1029/2018ja026167

[4] Chen, Z., Liao, W., Li, H., Wang, J., Deng, X., Hong, S. (2022), Prediction of Global Ionospheric TEC Based on Deep Learning, Space weather, DOI: 10.1029/2021SW002854

[5] Doornbos, E. N. (2012). Thermospheric Density and Wind Determination from Satellite Dynamics. Springer Berlin Heidelberg.

[6] Emmert, J. T., Picone, J. M., & Meier, R. R. (2008). Thermospheric global average density trends, 1967–2007, derived from orbits of 5000 near-earth objects. Geophysical Research Letters, 35(5), 84-102.

[7] Emmert, J. T. (2015). Thermospheric mass density: a review. Advances in Space Research, 56(5), 773-824.

[8] Fang, Yuanhui; Cui, Yanmei; Ao, Xianzhi (2019): Deep Learning for Automatic Recognition of Magnetic Type in Sunspot Groups. In Advances in Astronomy 2019, pp. 1–10. DOI: 10.1155/2019/9196234.

[9] Farrens, S. , Lacan, A. , Guinot, A. , & Vitorelli, A. Z. . (2021). Deep transfer learning for blended source identification in galaxy survey data.

[10] Gruet, M. A., Chandorkar, M., Sicard, A., & Camporeale, E. (2018). Multiple hours ahead forecast of the Dst index using a combination of Long Short-Term Memory neural network and Gaussian Process. Space Weather. doi:10.1029/2018sw001898

[11] Keller, J. . (2018). Darpa blackjack to develop small, secure military satellites for low-earth orbit. Military & Aerospace Electronics, 29(6), 4-6.

[12] Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436.

[13] Lei, J., Chen, G., Xu, J., & Dou, X. (2013). Impact of Solar Forcing on Thermospheric Densities and Spacecraft Orbits from CHAMP and GRACE.

[14] Liang, J. , Bu, Y. , Tan, K. , Pan, J. , Yi, Z. , & Kong, X. , et al. (2022). Estimation of Stellar Atmospheric Parameters with Light Gradient Boosting Machine Algorithm and Principal Component Analysis.

[15] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748-1764.

[16] Liu, L. , Zou, S. , Yao, Y. , & Wang, Z. . (2020). Forecasting global ionospheric tec using deep learning approach. Space Weather, 18(11).

[17] March, G.; Doornbos, E.N.; Visser, P.N.A.M. (2018). High-fidelity geometry models for improving the consistency of CHAMP, GRACE, GOCE and Swarm thermospheric density data sets. Advances in Space Research, (), S0273117718305714–. doi:10.1016/j.asr.2018.07.009

[18] Pan, S. J. , & Qiang, Y. . (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.

[19] Pan, Y., Jin, M., Zhang, S., & Deng, Y. (2020). TEC map completion using DCGAN and Poisson blending. Space Weather, e2019SW002390. doi:10.1029/2019sw002390

[20] Perez, D., & Bevilacqua, R. (2015). Neural network based calibration of atmospheric density models. Acta Astronautica, 110(may-jun.), 58-76.

[21] Perez, D. , Wohlberg, B. , Lovell, T. A. , Shoemaker, M. , & Bevilacqua, R. . (2014). Orbit-centered atmospheric density prediction using artificial neural networks. Acta Astronautica, 98(may-jun.), 9-23.

[22] Pires de Lima, R., Chen, Y., & Lin, Y. (2020). Forecasting Megaelectron-Volt Electrons Inside Earth's Outer Radiation Belt: PreMevE 2.0 Based on Supervised Machine Learning Algorithms. Space Weather-the International Journal of Research and Applications, 18(2).

[23] Shprits, Y. Y., Vasile, R., & Zhelayskaya, I. S. (2019). Nowcasting and Predicting the Kp Index Using Historical Values and Real-Time Observations. Space Weather-the International Journal of Research and Applications, 17(8), 1219-1229.

[24] Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. 2019 IEEE International Conference on Big Data (Big Data). doi:10.1109/bigdata47090.2019.9005997

[25] Tan, Y., Hu, Q., Wang, Z., & Zhong, Q. (2018). Geomagnetic index Kp forecasting with LSTM. Space Weather, 16(4), 406-416.

[26] Tang, R., Zeng, F., Chen, Z., Wang, J. S., & Wu, Z. (2020). The comparison of predicting storm-time ionospheric tec by three methods: arima, lstm, and seq2seq. Atmosphere, 11(4), 316.

[27] Tang, R., Liao, W., Chen, Z., Zeng, X., Wang, J. S., Luo, B., ... & Wu, Z. (2021). Solar Flare Prediction Based on the Fusion of Multiple Deep-learning Models. The Astrophysical Journal Supplement Series, 257(2), 50.

[28] Tang, R., Zeng, X., Chen, Z., Liao, W., Wang, J., Luo, B., ... & Wu, Z. (2021). Multiple CNN Variants and Ensemble Learning for Sunspot Group Classification by Magnetic Type. The Astrophysical Journal Supplement Series, 257(2), 38.

[29] Tang, R., Tao Y., Li J., Chen Z., Deng X., Li H. (2022), The Short-time Prediction of the Energetic Electron Flux in the Planetary Radiation Belt Based on Stacking Ensemble-Learning Algorithm, Space weather, DOI:10.1029/2021SW002969

[30] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

[31] Wang, S., Weng, L., Fang, H., Xie, Y., & Yang, S. (2014). Intra-annual variations of the thermospheric density at 400km altitude from 1996 to 2006. Advances in

Space Research, 54(3), 327–332. doi:10.1016/j.asr.2013.12.011
[32] Weng, L., Lei, J., Zhong, J., X Dou, & Fang, H. (2020). A machine-learning approach to derive long-term trends of thermospheric density. Geophysical Research Letters, 47.
[33] Wu, N., Green, B., Ben, X., & O'Banion, S. (2020). Deep transformer models for time series forecasting: The influenza prevalence case. arXiv preprint arXiv:2001.08317.
[34] Wu, S., Xiao, X., Ding, Q., Zhao, P., Wei, Y., & Huang, J. (2020). Adversarial sparse transformer for time series forecasting. Advances in neural information processing systems, 33, 17105-17115.
[35] Xu, S. B., Huang, S. Y., Yuan, Z. G., Deng, X. H., & Jiang, K. (2020). Prediction of the Dst Index with Bagging Ensemble-learning Algorithm. Astrophysical Journal Supplement Series, 248(1).

Figure 1.

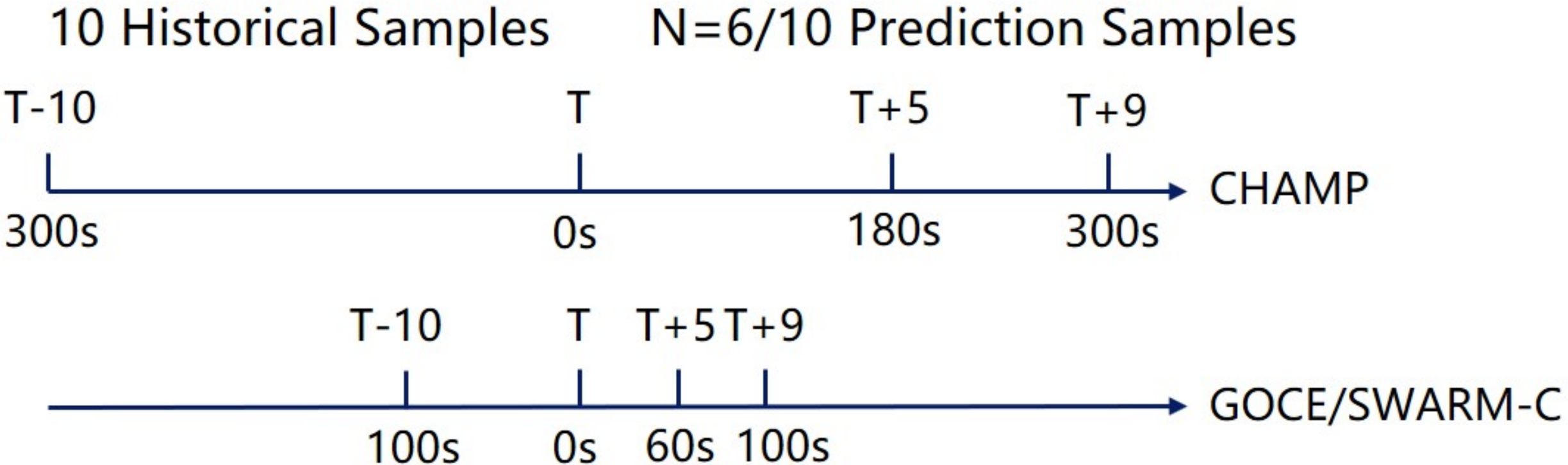
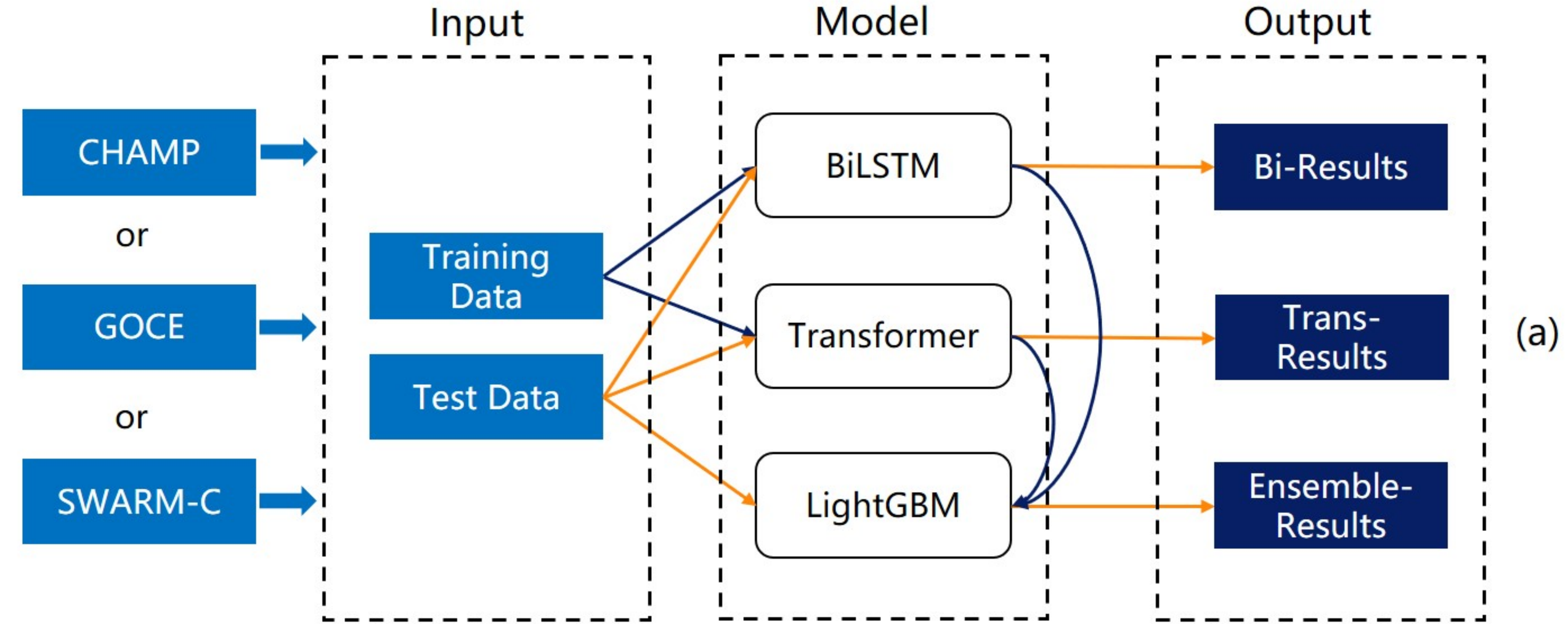
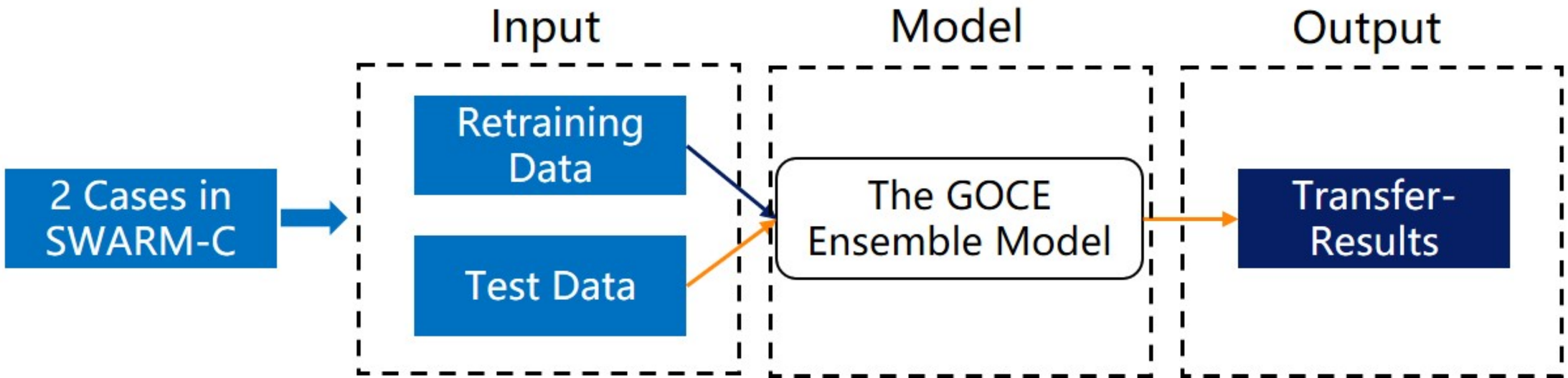

10 Historical Samples
N=6/10 Prediction Samples
T-10
T
T+5
T+9
CHAMP
300s
0s
180s
300s
T-10
T
T+5
T+9
GOCE/SWARM-C
100s
0s
60s
100s

Figure 2.


Input
Model
Output
CHAMP
or
GOCE
or
SWARM-C
Training Data
Test Data
BiLSTM
Transformer
LightGBM
Bi-Results
Trans-Results
Ensemble-Results

(a)


Input
Model
Output
2 Cases in SWARM-C
Retraining Data
Test Data
The GOCE Ensemble Model
Transfer-Results

(b)

Figure 3.

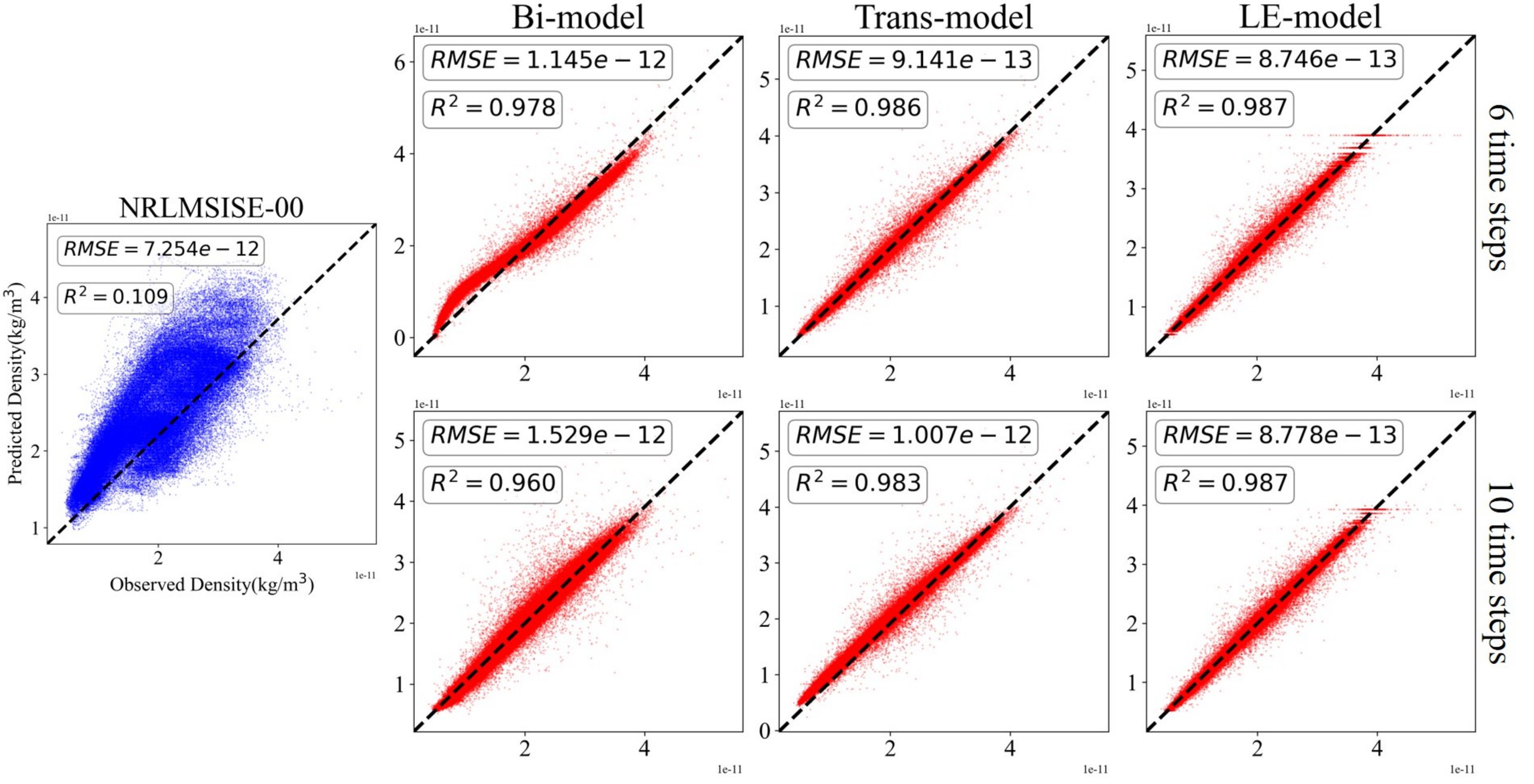

NRLMSISE-00
Bi-model
Trans-model
LE-model
6 time steps
10 time steps
RMSE = 7.254e − 12
R² = 0.109
RMSE = 1.145e − 12
R² = 0.978
RMSE = 9.141e − 13
R² = 0.986
RMSE = 8.746e − 13
R² = 0.987
RMSE = 1.529e − 12
R² = 0.960
RMSE = 1.007e − 12
R² = 0.983
RMSE = 8.778e − 13
R² = 0.987
1e-11
Predicted Density(kg/m³)
Observed Density(kg/m³)

Figure 4.

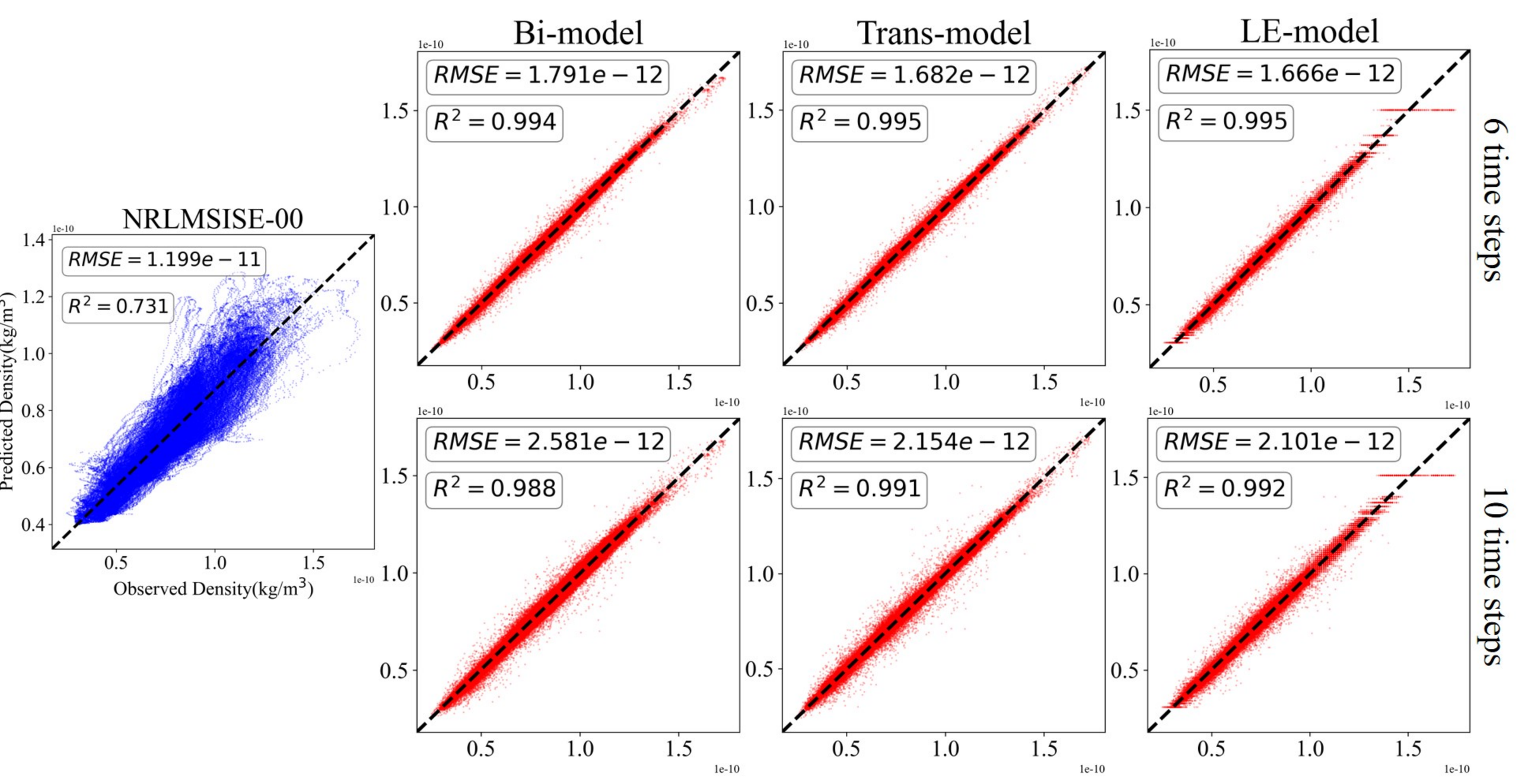

NRLMSISE-00
Bi-model
Trans-model
LE-model
6 time steps
10 time steps
RMSE = 1.199e − 11
R² = 0.731
RMSE = 1.791e − 12
R² = 0.994
RMSE = 1.682e − 12
R² = 0.995
RMSE = 1.666e − 12
R² = 0.995
RMSE = 2.581e − 12
R² = 0.988
RMSE = 2.154e − 12
R² = 0.991
RMSE = 2.101e − 12
R² = 0.992
1e-10
Predicted Density(kg/m³)
Observed Density(kg/m³)

Figure 5.

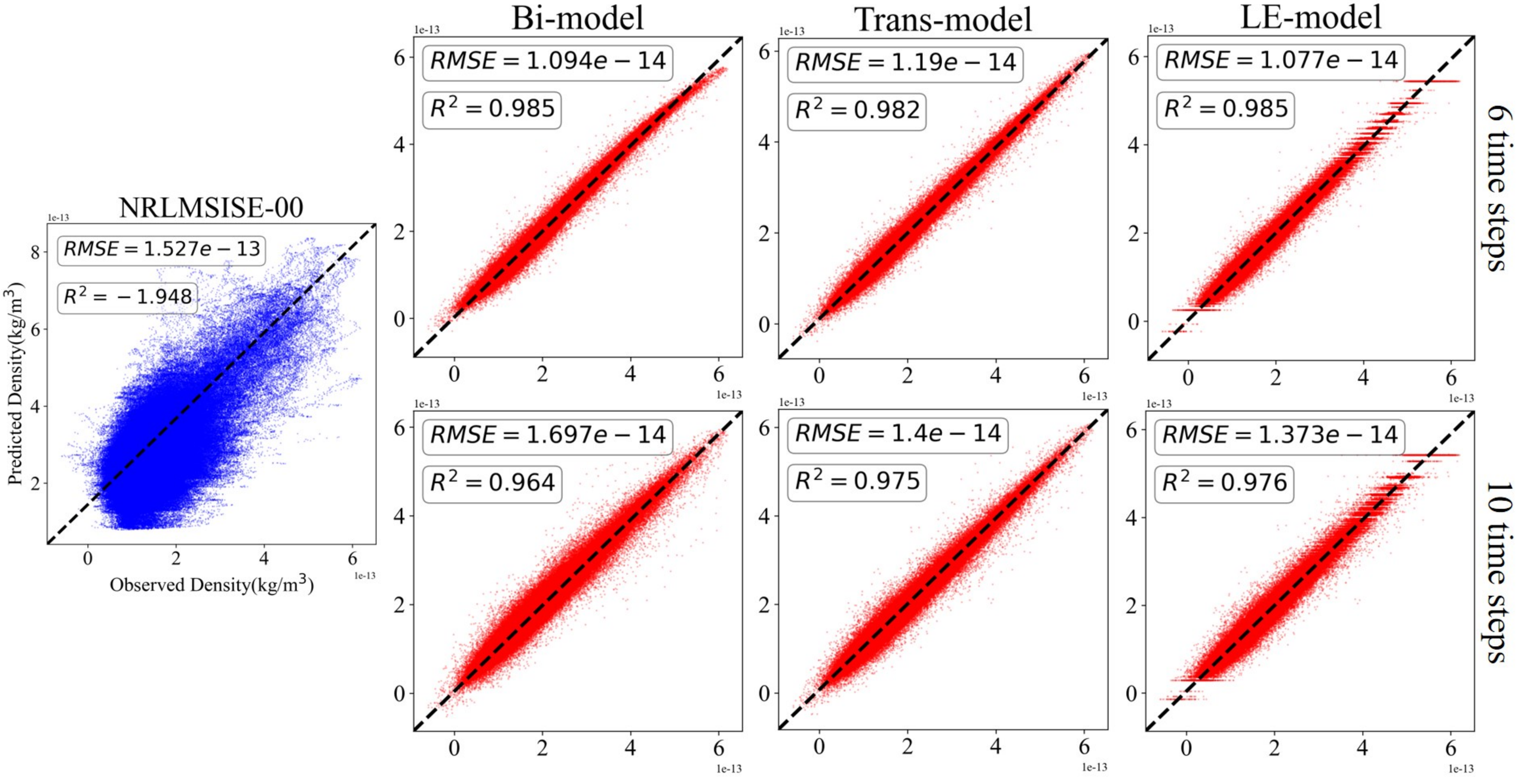

NRLMSISE-00
Bi-model
Trans-model
LE-model
6 time steps
10 time steps
1e-13
$RMSE = 1.527e-13$
$R^2 = -1.948$
$RMSE = 1.094e-14$
$R^2 = 0.985$
$RMSE = 1.19e-14$
$R^2 = 0.982$
$RMSE = 1.077e-14$
$R^2 = 0.985$
$RMSE = 1.697e-14$
$R^2 = 0.964$
$RMSE = 1.4e-14$
$R^2 = 0.975$
$RMSE = 1.373e-14$
$R^2 = 0.976$
Predicted Density(kg/m^3)
Observed Density(kg/m^3)

Figure 6.

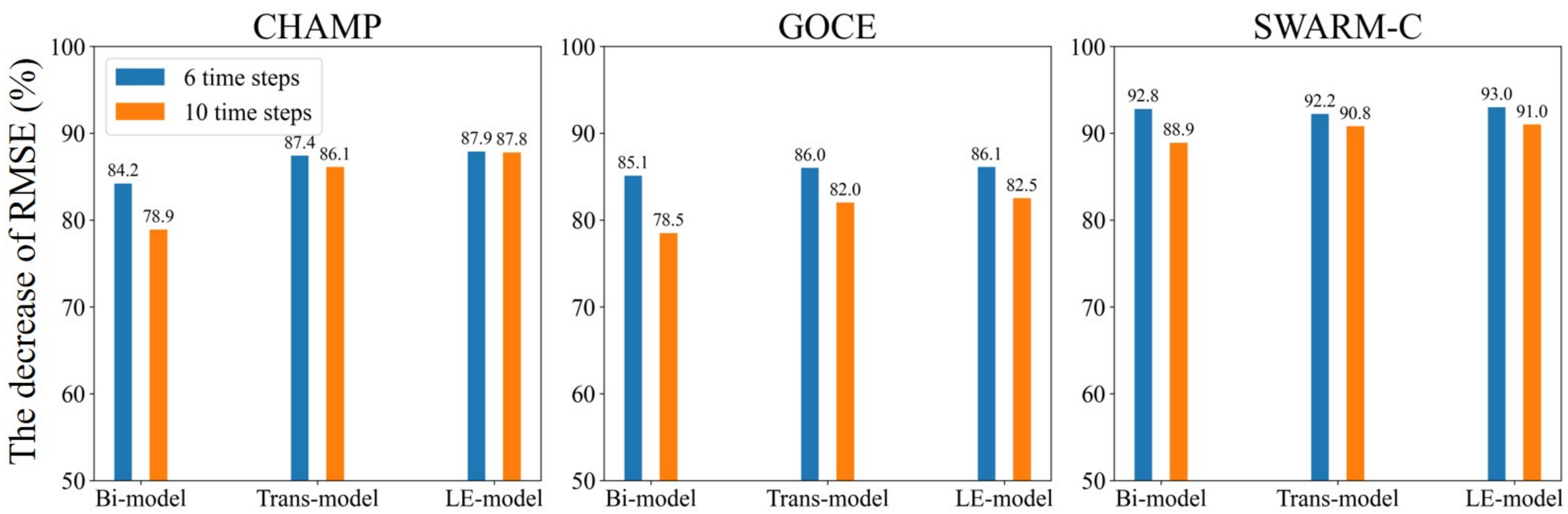

CHAMP
GOCE
SWARM-C
The decrease of RMSE (%)
6 time steps
10 time steps
100
90
80
70
60
50
84.2
78.9
87.4
86.1
87.9
87.8
85.1
78.5
86.0
82.0
86.1
82.5
92.8
88.9
92.2
90.8
93.0
91.0
Bi-model
Trans-model
LE-model

Figure 7.

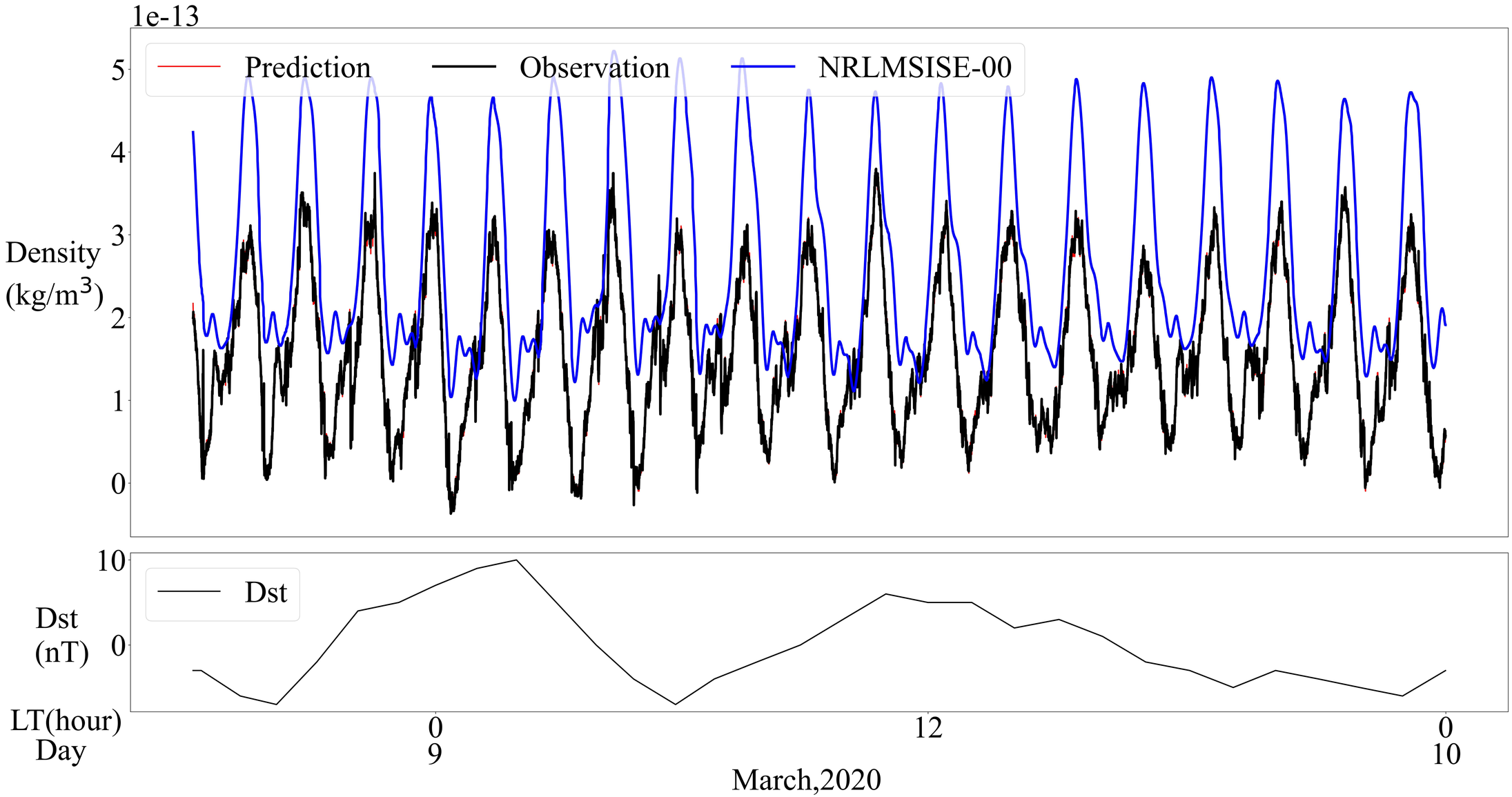
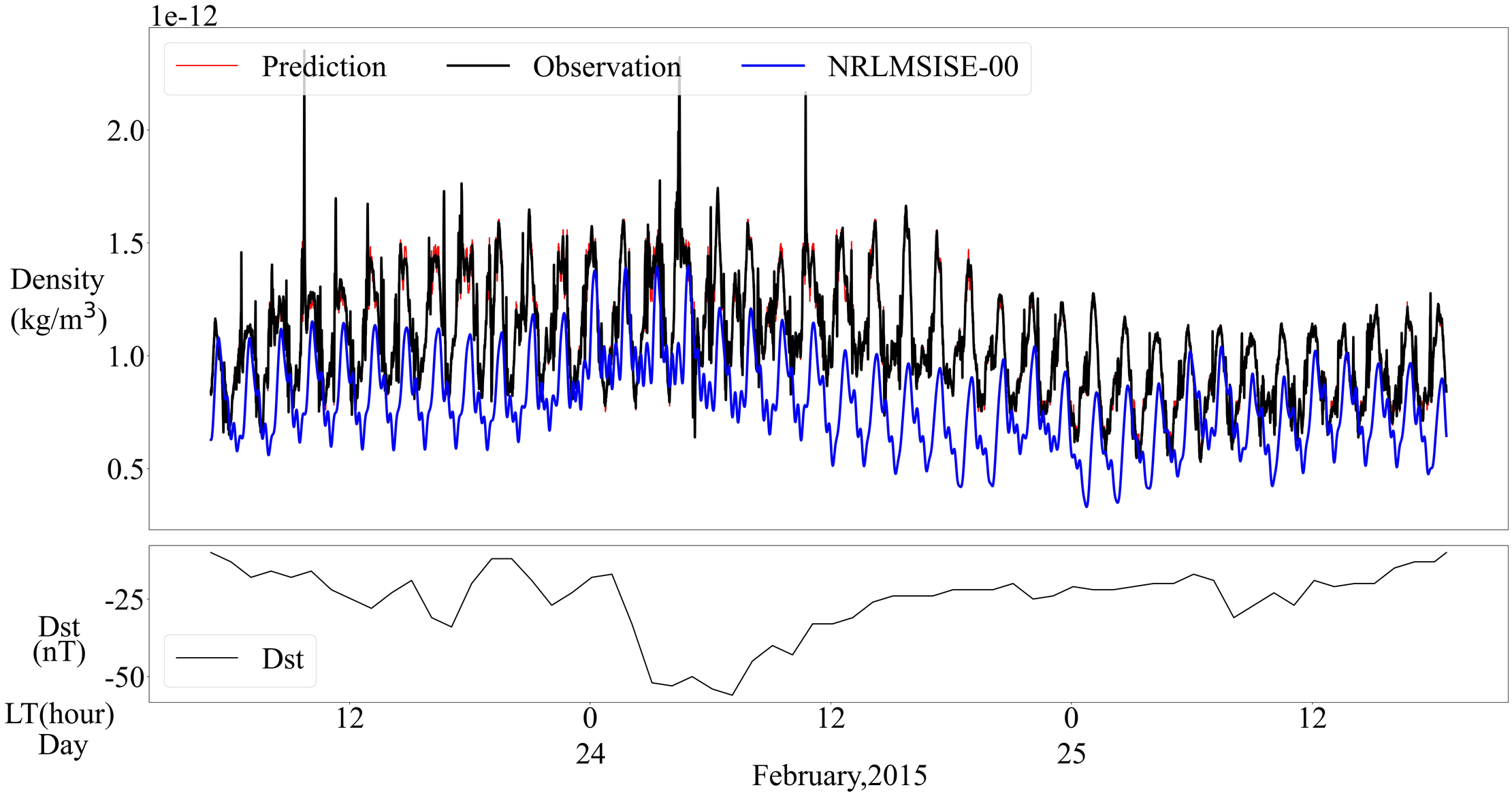

1e-13
Prediction
Observation
NRLMSISE-00
Density
(kg/m3)
5
4
3
2
1
0
Dst
Dst
(nT)
10
0
LT(hour)
Day
0
9
12
0
10
March,2020

Figure 8.


1e-12
Prediction
Observation
NRLMSISE-00
2.0
1.5
1.0
0.5
Density
(kg/m3)
Dst
-25
-50
Dst
(nT)
LT(hour)
Day
12
0
12
0
12
24
25
February,2015