Potential of regional ionosphere prediction using a long short-term memory (LSTM) deep-learning algorithm specialized for geomagnetic storm period

Jeong-Heon Kim[1], Young-Sil Kwak[1, 2*], YongHa Kim[3], Su-In Moon[3], Se-Heon Jeong[1, 3], JongYeon Yun[4]

[1] Korea Astronomy and Space Science Institute (KASI), 776 Daedeok-daero, Yuseong-gu Daejeon, South Korea

[2]Department of Astronomy and Space Science, Korea University of Science and Technology (UST), 776 Daedeok-daero, Yuseong-gu, Daejeon, Korea

[3] Department of Astronomy, Space Science and Geology, Chungnam National University (CNU), 99 Daehak-ro, Yuseong-gu, Daejeon, South Korea

[4] Korea Space Weather Center (KSWC), 198-6 Gwideok-ro Hallim-eup Jeju, South Korea

* Corresponding author: yskwak@kasi.re.kr

Keywords: Ionosphere, prediction model, LSTM deep-learning algorithm, Geomagnetic storm period

Key Points:

- We developed a new LSTM specialized for geomagnetic storm periods by training examples of past geomagnetic storm events.
- Our LSTM storm model improves performance for foF2 (hmF2) by about 32%, 34%, and 37% (10%, 17%, and 5%) compared to the LSTM quiet, SAMI2, and IRI-2016 models.
- We propose that the prediction model less than 3 hours using the deep-learning method can effectively forecast the ionosphere state.

Abstract

In our previous study (Moon et al., 2020), we developed a Long Short Term Memory (LSTM) deep-learning model for geomagnetic quiet days (LSTM-quiet) to perform effective long-term predictions for the regional ionosphere. However, their model could not predict geomagnetic storm days effectively at all. This study developed an LSTM model suitable for geomagnetic storms using the new training data set and re-designing input parameters and hyper-parameters. We collected 131 days of geomagnetic storm cases from 1 January 2009 to 31 December 2019, provided by the Japan Meteorological Agency's Kakioka Magnetic Observatory, and obtained the IMF Bz, Dst, Kp, and AE indices related to the geomagnetic storm corresponding to each storm date from the OMNI database. These indices and F2 parameters (foF2 and hmF2) of Jeju ionosonde (33.43˚N, 126.30˚E) were used as input parameters for the LSTM model. To test and verify the predictive performance and the usability of the LSTM model for geomagnetic storms developed in this manner, we created and diagnosed the 0.5, 1, 2, 3, 6, 12, and 24-hour predictive LSTM models. According to the results of this study, the LSTM storm model for 24-hour developed in this study achieved a predictive performance during the geomagnetic storms about 32% (10%), 34% (17%), and 37% (5%) better in RMSE of foF2 (hmF2) than the LSTM quiet model (Moon et al., 2020), SAMI2, and IRI-2016 models. We propose that the short-term predictions of less than 3 hours are sufficiently competitive compared to other traditional ionospheric models. Thus, this study suggests that our model can be used for short-term prediction and monitoring of the regional mid-latitude ionosphere.

1 Introduction

In the 20th century, when radio communication started to be used, it became essential to monitor and predict changes in peak frequency (foF2) and peak height (hmF2) of F2 layer among the various ionospheric parameters. Especially, the long-range high frequency (HF) communication is dominantly affected by reflection from the ionosphere F2 layer. Although numerous models have been developed to simulate and predict changes in the ionosphere of the F2 layer, it is relatively recent that the deep-learning technology began to be utilized in the ionospheric models. (e.g., Williscroft & Poole, 1996; Altinay et al., 1997; McKinnell & Poole, 2000; Wintoft & Cander, 2000; Poole & Poole, 2002; Oyeyemi & Poole, 2005; Yue et al., 2006; Nakamura et al., 2007; Athieno et al., 2017; Wichaipanich et al., 2017; Gowtam & Ram, 2017; Hu & Zhang, 2018; Ram et al., 2018; Fan et al., 2019; Moon et al., 2020; Kim et al., 2020).

Ionospheric modeling studies using artificial neural network (ANN), the basic deep-learning method, began in the mid-1990s. Williscroft & Poole (1996) developed a simple ANN model trained with foF2 observed over ten years in Grahamstown Ionosonde (33.3°S, 26.5°E), sunspot number, and Ap index. However, the model was limited to predict only noontime foF2 daily. Altinay et al. (1997) developed a 1-hour prediction model for foF2 using a multi-layer perceptron layer model. Their ANN model was trained with the 10-year foF2 data observed by Poitiers ionosonde in central Europe and Kp indices. On average, their model predicted foF2 within a factor of two for geomagnetic quiet days, using only the Kp index as a space environment index. Two studies developed ANN models utilizing data from Grahamstown Ionosonde for 24 years and predicted foF2 after 1, 2, 3, 4, and 25 hours only for geomagnetically quiet days (McKinnell & Poole, 2000; Poole & Poole, 2002). Later, Oyeyemi & Poole (2005) extended the ANN model using 59 global ionosonde data from 1964 to 1986, with more input

parameters, such as the day number (day of the year), hour, sunspot number, Ap index, the meridian angle, magnetic dip angle, magnetic declination angle, and solar zenith angle. Their model improved the prediction performance by 15~16%, but the performance for geomagnetic storms was not presented. Yue et al. (2006) developed an ANN model capable of long-term prediction of foF2 by training it with 19 ionosonde data located in the Asia/Pacific region. However, it was also a model that could only predict the foF2 parameter accurately during geomagnetic quiet days. More recently, Wichaipanich et al. (2017) presented a neural network (NN) model for foF2 by utilizing data from three ionosonde stations near the magnetic equator of Southeast Asia. The model showed better predictive performance for the RMSE of foF2 in Chiang Mai, Chumphon, and Kodotabang than the IRI model, but their results were also analyzed only on geomagnetically quiet days. All of these neural network models have focused on predicting the foF2 parameter only for geomagnetically quiet conditions.

There have also been studies that developed NN models including geomagnetic storm cases. Wintoft & Cander (2000) developed an NN model for foF2 prediction by using Slough ionosonde (51.5°N, 0.6°W) data and the AE index as an input parameter. The model calculated the hourly prediction values from 1 hour to 25 hours later. Nakamura et al. (2017) developed the foF2 predictive ANN model using the Kokubunji ionosonde (35.71°N, 139.49°E) data in Japan and the local K index as an input parameter. Athieno et al. (2017) developed a similar ANN model by using 21-year data from Resolute (74.7°N, 265.1°E) located at the polar cap region. The model was limited to hourly prediction and its input parameters included the day number, hour, polar cap index, Ap index, and F10.7 index values. Fan et al. (2019) used an Elman neural network (ENN) algorithm different from ANN and tested a short-term foF2 prediction during the geomagnetic storm. The ENN algorithm is more advantageous for time-series analysis in which

it stores the past state that it may compensate for the weak points of ANN. All of these models attempted to predict foF2 including geomagnetic storm cases, but hmF2 was not attended.

A very few studies dealt with both foF2 and hmF2 in the development of NN models. Gowtam & Ram (2017) developed an ANN-based model that can predict both foF2 and hmF2 by using GPS radio-occultation (RO) observation data from the Formosa Satellite Mission 3 (FORMOSAT-3) - the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) satellites. The model was trained with inputs neutral wind field information of the Horizontal Wind Model – 14 (HWM14) model, in addition to the usual input parameters, F10.7 and Ap index. However, the model results of their study were also limited to geomagnetic quiet days. Ram et al. (2018) improved the ANN-based model by using data from the Challenging Minisatellite Payload (CHAMP), the Gravity Recovery and Climate Experiment (GRACE) RO, and global ionosonde network together. Because these ANN-based models can not consider past sata for a specific period from the present, the prediction ability may not be appropriate for phenomena that are affected by the previous state further back than the specific period.

To overcome the disadvantage of the ANN algorithm, a technique is needed to memorize past data and reflect them in prediction, such as the long-short term memory (LSTM) algorithm. Hu & Zhang (2018) used a bi-LSTM technique to memorize data characteristics in both forward and backward directions, and developed a model for 1 hour hmF2 prediction. Our previous work (Moon et al., 2020) predicted both foF2 and hmF2 parameters using the LSTM model. Kim et al. (2020) assimilated the F2 parameter predicted from Moon et al. (2020) into a physics-based model to predict the mid-latitude ionosphere up to 24 hours. Although the LSTM

based model showed reasonably good predictive performance on geomagnetically quiet days, the model prediction was poor for geomagnetic storm days. The reason for poor prediction for geomagnetic storm days may be due to the fact that the training data for the LSTM model were obtained mostly for geomagnetic quiet days so that the model was biased to the quiet condition.

Therefore, in this study, we attempt to overcome the problems of the new LSTM model during geomagnetic storms reported in the previous study (Moon et al., 2020). For this purpose, we collected 69 geomagnetic storm events from 1 January 2009 to 31 December 2019, considering the period in which data on Jeju ionosonde (33.43˚N, 126.30˚E) in South Korea can be used. The new LSTM model was trained with geomagnetic activity input parameters, including the interplanetary magnetic field (IMF) Bz component, Dst, Kp, and AE indices. In this way, we have developed the new LSTM model that applies only to storm periods. In addition, we have re-modeled different LSTM models specialized for 0.5, 1, 2, 3, 6, 12, and 24-hour prediction, considering that the performance of the LSTM model varies depending on the prediction target.

In this study, we present the results of the short-term prediction for foF2 and hmF2 during the geomagnetic storms and discuss the limitations of rapidly changing ionosphere states. Section 2 provides a detailed description of the LSTM algorithm and data used in this study for the geomagnetic storm cases. Section 3 presents the LSTM model results, and section 4 discusses the limitations and possibilities of predicting the rapidly changes ionosphere during storm times. Finally, section 5 summarizes and concludes our study.

2 Methodology

2.1 Long-Short Term Memory (LSTM)

Our model has a similar design to the LSTM model developed previously (Moon et al., 2020). Figures 1 (a) and (b) show diagrams of the new LSTM model developed in this study. Figure 1 (a) shows the detailed calculation flow included in one LSTM cell. The LSTM model consists of three steps, and each part has its own characteristics (e.g., Hochreiter & Schmidhuber, 1997). The first step here is the "forget gate layer" that controls what data is left behind from the previously calculated output (h_{t-1}) and the current input data (x_t). Here, the used sigmoid function (σ) receives the previous output and the present input and gives a weight by getting a value of 0 to 1. A value of σ closer to 0 means that the result of this function does not affect future results, and conversely, a value closer to 1 means the opposite. The sigmoid function (σ) expressed by Equation (1) is multiplied by the calculated value, f_t, by the LSTM cell state (C_{t-1}) to determine which value to discard.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

The second is the "input gate layer" step, which determines whether new information will be stored in the cell state. In this step, it decides which values (i_t) to update by the sigmoid layer (Equation 2), and generates a new candidate value ($\tilde{C}_t$) that can be added to the cell state (C_{t-1}) through the nonlinear function tanh layer (Equation 3), where W is the weight and b is the bias. As shown in Equation 4, the new candidate value ($\tilde{C}_t$) is multiplied by the updated information (i_t) and added to the existing cell state $(f_t \times C_{t-1})$ calculated in the previous step.

Here, this process solves the vanishing gradient problem raised in the Recurrent Neural Network (RNN) model by adding the updated value to the existing value.

$$C_t = (f_t \times C_{t-1}) + (i_t \times \tilde{C}_t) \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \times \tanh(C_t) \quad (6)$$

In the final step, the "output gate layer," it decides what to output. In other words, it is the process of determining what value to output (o_t) from the current state (h_{t-1}) and input value (x_t) in the sigmoid function (Equation 5). The updated cell state (C_t) is normalized to a value between -1 and 1 through the tanh function and then is multiplied by the output value (o_t) from the sigmoid function (Equation 6). Through this overall process in a one LSTM cell, we can obtain the result value (h_t) at the current time t. This single-cell serves as one hidden layer, and the optimal number was found by adjusting from 2 to 50 in 2 intervals in this study.

2.2 Training, Validation, and Test Data Sets

Previous studies (Moon et al., 2020; Kim et al., 2020) performed good predictions during geomagnetically quiet periods, but did not predict well during geomagnetic storms. It was pointed out that most of the long-term training data included geomagnetically quiet days. Actually, during the five-year training period used in Moon et al. (2020), there were only 57 cases and 105 days of geomagnetic storms. For the new LSTM we collected measured data under geomagnetic storm condition for a longer period of time, and excluded data for quiet days. We searched for geomagnetic storm events on the Japan Meteorological Agency/Kakioka Magnetic Observatory website (kakioka-jma.go.jp/en/) during the period when Jeju ionosonde data were available (from 1 January 2010 to August 2017, for a total of 7.7 years). We were able to find a

total of 71 events (138 days) associated with available ionosonde data during this period, as shown in Table 1. We divided these events into a training and validation set at a ratio of 9:1 to check the performance of the model. In addition, we used Bz, AE, Dst, and Kp indices as geomagnetic indices corresponding to te events in Table 2. Since the ionosphde data are observed every 15 minutes, all indices have been interpolated every 15 minutes.

The test period (final prediction target) in this study is three days from 6 September to 8 September 2017, the geomagnetic storm case used in the previous study (Moon et al., 2020; Kim et al., 2020). Figure 2 shows the space environment changes during the geomagnetic storm of test event #1 (see Figure 2 in Kim et al., 2020). Two more test events were analyzed to evaluate the performance of the LSTM model in this study. These events are detailed in Table 1. Ionosonde data observed during the three test periods were downloaded from the Korean space weather center's web-page (https://spaceweather.rra.go.kr/observation/service/iono), and the geomagnetic indices corresponding to each event period are obtained from the OMNI web (https://omniweb.gsfc.nasa.gov/ow.html).

2.3 Optimal Hyper Parameter Options

When making predictions using the LSTM algorithm, it is important to control the hyper-parameters of the model. In other words, we had to find the optimal model state by adjusting the hyper-parameters. Among the many hyper-parameters, the number of hidden layers, which means LSTM cells described above, and the range of historical data used for prediction are important. In our LSTM algorithm, the range of historical data can be adjusted with a hyper-parameter called lookback. Since each data point is every 15 minutes, when the lookback

parameter is 1, the past 15 minutes of data are reflected in the current state. If the lookback is 4, data from the past 1 hour would be remembered.

Another hyper-parameter, the batch size can set the data interval for updating the weights, and it can also be adjusted every 15 minutes. In this study, because we aim to predict a maximum of one day (data points = 96), the lookback and batch size are set and validated to 6 hours (data points = 24), 12 hours (data points = 48), and 24 hours (data points = 96). Since we have to validate the prediction performance of 0.5, 1, 2, 3, 6, 12, and 24-hours, we set the value of lookahead (prediction target) to 2, 4, 8, 12, 24, 48, and 96. Then we calculated the root mean square error (RMSE) values for the validation data sets using the combination of each hyper-parameter. The combination with the lowest value was chosen and adopted as the design of the final LSTM model. We performed all the steps for both foF2 and hmF2 models, respectively.

Figure 3 and 4 show the RMSE values for the foF2 and hmF2 validation set of the prediction target model, respectively, according to each hyper-parameter combination. We have marked the best combination for each plot with the black arrow compared with different prediction targets. Also, we summarized the best options in the plots in Tables 3 and 4. Here, we were able to find some interesting aspects in the tables. For both foF2 and hmF2, the best options were determined based on a certain prediction target. In the case of foF2, the LSTM structure with 24 lookbacks, 24 batch sizes, and 30 hidden layers showed the best performance in the prediction target of 3 hours or less. On the other hand, 48 lookbacks, 24 batch sizes, and 32 hidden layers were the best combinations for the 6 hours or more prediction target. For hmF2, the LSTM structure of 96 lookbacks, 24 batch sizes, and 44 hidden layers showed the best

performance in 3 hours or less, and 96 lookbacks, 48 batch sizes, and 46 hidden layers were the best options for 6 hours or more. In Tables 3 and 4, the optimal combinations corresponding to the short-term forecasting targets are indicated by a gray box and the long-term forecasting ones by a gold box, to improve readability. We trained the LSTM models using these selected best hyper-parameters and evaluated their performances during the testing periods.

3 Results and Discussions

The LSTM models trained for geomagnetic storm events in this study are first compared with the LSTM quiet model (Moon et al., 2020). Then, we analyze the performance of the LSTM storm models in terms of short- and long-term predictions for foF2 and hmF2. Moreover, we discuss the possibility and limitation that the LSTM-storm models can be used for forecasting services in the regional ionosphere.

3.1 Comparison of LSTM-Storm and LSTM-Quiet models

First, we compared the 24-hour prediction results of the LSTM storm model with those of the LSTM quiet-time model (Moon et al., 2020). Since the 24-hour prediction model results need to be compared, the LSTM model with the hyper-parameter corresponding to 24 hours in Tables 3 and 4 was used. In addition, we compare the performance of the SAMI2 model and the IRI-2016 model. The SAMI2 (physics-based model) and IRI-2016 (international reference ionosphere model) were used for standard comparison because they are widely used as ionospheric models around the globe.

The model predictions were made during a storm event (Test #1), which occurred on DOY 251 in 2017. During the storm, the Kp index rose up to 8, as presented in Figure 5, where the geomagnetic indices are shown for three days around the storm event. Figure 5 also displays the predicted foF2 and hmF2 values of the LSTM-storm models (LSTM-quiet models) with solid blue (red) lines, respectively. To compare with the observed values of Jeju Ionosonde, the SAMI2 and IRI-2016 model results are also shown with orange and green lines, respectively. It is evident from Figure 5 (a) that the LSTM storm model predicts the increased foF2 on day 251 better than the LSTM-quiet model during the test storm period. However, on DOY of 249 and 250, the predicted foF2's of the LSTM storm model are significantly larger than the LSTM-quiet model, thus overestimating the observed values. The predicted foF2's of the LSTM-quiet model seem to behave similarly for all three days (249, 250, and 251), regardless of the storm event on DOY of 251. It may be because the LSTM quiet model was mainly trained with foF2 data of quiet periods. On the contrary, the LSTM storm model trained with only storm period data predicts better the increased foF2 during the storm day but overestimates foF2's during quiet days. In Figure 5 (b), the observed hmF2 elevated sharply on day 251, but none of the models predicts the sharp elevation of the F2 layer. The prediction of hmF2, especially for stormy nighttime, remains a challenge to data-based and physics-based models.

We wondered if the storm-only LSTM model developed in this study predicted positive storm well for foF2 but not well for hmF2. We speculated that the answer to this would be found in the training data and confirmed how the deep-learning model results were biased according to the training data. As such, we closer looked into the training dataset of 61 storm events listed in Table 1. We then confirmed the patterns statistically whether the ionospheric storm developed

into a positive or negative storm for foF2 values. In addition, we checked whether the hmF2 values were elevated or lowered during each event period.

In order to analyze each storm patterns, we characterized the 61 ionospheric storm patterns by setting a reference as the average value for a total of 10 days, that is ±5 days around each event date. Naturally, the 10 day period includes the geomagnetic quiet condition. The reference values of the ionospheric parameter have been utilized in various studies to distinguish the effects of a geomagnetic storm on the ionosphere. Szuszczewicz et al. (1998) set the reference with the average value for 3 days, and Adebiyi et al. (2014) and Lissa et al. (2020) did with the 5 days average values. The 7-day and 10-day average values were used by Fagundes et al. (2015) and Jin et al. (2017), respectively. It is expected that the longer the window size for the mean value is, the more suitable to make the quiet reference state. The above-mentioned studies also defined a positive ionospheric storm when the storm period value was higher than the reference value. In the opposite case is a negative storm. In the same way, we recorded a positive (negative) storm as the observed values higher (lower) than the reference line for the duration of all training events. For each storm event, we added all the time span for positive (negative) storm and computed the ratio of the added positive to negative storm spans, called the P/N ratio. In this way, if the P/N ratio of a storm event is larger (smaller) than1, it means that the positive (negative) storm is dominant. Using these factors, we were able to quantify how each training event could affect the entire training data set. In hmF2, we defined the rising (falling) of hmF2 to be positive (negative).

In Figure 6, we present how the positive/negative spans are computed for foF2 and hmF2 data of the training storm event #1. The positive spans are in red, and the negatives in blue.

Besides, the geomagnetic Kp index is bar-plotted at the bottom of Figure 6 to check the time zone of the geomagnetic storm. Especially, bars with Kp of 4 or more are shown in gold. As shown in the dominant red color region, the foF2 values are mostly higher than the reference (dotted line) during the storm period of training event #1. For the hmF2, positive and negative spans are more or less evenly distributed. For both foF2 and hmF2, the P/N ratios are greater than 1, and we could surmise that the positive storm dominated in training event #1. The other 60 training events all showed different ionospheric storm patterns, and we plotted these results in support information (Figure S1-S8).

We expect that the distribution of the P/N ratios for the total of 61 storm events will affect the deep learning model developed in this study. Figure 7 (a) briefly summarizes the P/N ratios of 61 storm events, and histograms in Figure 7 (b) show what types of ionospheric storms affect the LSTM storm model. As can be seen from the histogram, the foF2 training dataset has about twice as many positive storms as the negative storm, whereas the hmF2 dataset has about half and half for positive and negative storms. It can be indirectly inferred that the LSTM storm model for foF2 is more specialized to positive storms, which may explain that the LSTM storm model predicts well the foF2 positive storm for test event #1, as in Figure 5(a). However, the LSTM storm model seems not trained well for the case of a positive hmF2 storm, explaining the failure of predicting the elevated hmF2 in Figure 5(b).

Figure 8 compares the model predictions for test event #2 (a-c) and #3 (d-f) in the same way as in Figure 5. As shown in Figure 8(c), test event #2 includes a geomagnetic storm on day 111 of 2018, but there was little change in the ionosphere of the Jeju location. On this day, foF2 was slightly higher than the previous day (day 110) in the nighttime but slightly lower in the

daytime. In the case of hmF2, it also did not change on the day of the geomagnetic storm. Test event #3 with a geomagnetic storm on the day 254 day of 2018 did not significantly change the ionosonde observations, similar to test event #2. As shown in Figure 8 (b) and (e), both the LSTM storm and quiet models seem to predict reasonably close values to the observations. In the absence of a positive storm, as in test event #1, the performance of the two models is practically the same.

To quantitatively evaluate the performance of our model for the storm days in the three test events, we calculated three indices: correlation coefficient (CC), root means square error (RMSE), and mean absolute percentage error (MAPE).

$$\mathrm{CC} = \frac{\sum_1^n (model - \overline{model})(ionosonde - \overline{ionosonde})}{\sqrt{\sum_1^n (model - \overline{model})^2} \cdot \sqrt{\sum_1^n (ionosonde - \overline{ionosonde})^2}} \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_1^n (model - ionosonde)^2}{n}} \tag{8}$$

$$\mathrm{MAPE}\ (\%) = \frac{100\ \%}{n} \times \sum_1^n \left| \frac{model - ionosonde}{ionosonde} \right| \tag{9}$$

Table 5 summarizes the skill scores of all the models compared in Figure 5 and 8. As in Table 5, the LSTM storm model predicted mostly the best among all the models for the three test events. Quantitatively, our prediction model shows better predictive performance for foF2 (hmF2) by about 32%, 34%, and 37% (10%, 17%, and 5%) compared to the LSTM quiet, SAMI2, and IRI-2016 models. Although more test events are needed to evaluate the statistically meaningful performance, we could not find events related to geomagnetic storms in the Jeju Ionosonde observation data since 2018. Especially, the LSTM storm model needs to be evaluated more for negative storm events that are lacking for the current study.

Since ionospheric positive storms dominated the training data statistics of foF2, the LSTM storm model is expected to predict higher foF2 than the LSTM quiet model in all test events. However, the results of test events #2 and #3 overturned expectations. This may be because the LSTM algorithm remembers a specific feature of past data and uses it for prediction. In other words, the LSTM storm model may have recognized a feature for the positive storm from input data in the previous and current days, able to predict the increased foF2 for test event #1, but in the case of tests #2 and #3, it did not catch the feature for the positive storm so that it simply predict typical foF2s as the LSTM-quiet model does. It is important for a deep-learning model to be trained to catch features in the time window to predict the storm. However, it is not easy to collect training data because even with the same geomagnetic storm, ionospheric storms can appear differently depending on geographic location, season, and local time. Therefore, it may be better for regional ionosphere prediction to develop a deep-learning model using observed datasets from a specific local region rather than global datasets.

Furthermore, to use the LSTM storm model in a forecasting service, it is needed to know when the forecaster changes the LSTM quiet model to the storm model timely. This is a difficult issue because it requires predicting a geomagnetic storm. One practical solution to this issue is to use predicted Kp indices from global space weather services, where all the space weather data, including satellite data for solar wind, are collected and analyzed for the prediction. For example, Kp index greater than 4 is predicted over the next day, then the regional ionospheric forecaster adopts the LSTM storm model rather than the LSTM quiet model. We expect that the LSTM storm model developed in this study can be useful in this way. Besides, our research group plans to use these ideas for forecasting and surveillance work.

3.2 Performances of Short-term and Long-term Predictions

In this section, we evaluated the performance of the LSTM storm model for 1, 2, 3, 6, 12, and 24-hour predictions. During the test event #1 period, the models predict foF2 and hmF2 for next target times. For example, the 1-hour prediction model computes the value for 1 hour later than the current time by using the observed data up to now. In other words, it is easy to understand that a 1-hour prediction model makes predictions every hour.

Figure 9 shows the histogram, which includes the predictive ability of each model along with performance skill scores. The correlation coefficient is colored in gold, the slope plotted in 1-to-1 correspondence is in red, and the RMSE is in blue. We additionally used the slope obtained by a 1-to-1 correspondence in order to intuitively know the relationship between the observed value and each model predicted value. It is evident from Figure 9 that RMSE's for both F2 parameters increase with the prediction time except for the 0.5 hr prediction. On the contrary, the correlation coefficient and 1:1 slope decrease as the prediction time increases. It is congruent with what we initially predicted. What is unique here is that the 1-hour prediction model performs better than the 30-minute prediction model. The best RMSEs of the 1-hour prediction are 0.3 MHz and 17 km for foF2 and hmF2, respectively, which are encouraging results for the practical application of the model.

For practical application, we regard the LSTM models up to 3 hr prediction as short-term prediction models and longer than 3 hours as long-term prediction models. As shown in Figure 9, the RMSE values of the short-term prediction models for foF2 are significantly less than 1 MHz, while that of the long-term models exceeds 1 MHz. The performance of the short-term models is comparable to other deep learning models (McKinnell & Poole, 2000; Athieno et al., 2017; Fan

et al., 2019), although it is not easy to compare with the RMSE from these models because they have been developed for targeting different local, date, local time, and space environment conditions. For hmF2, the short-term prediction models were not easy to significantly improve performance. Nevertheless, Based on the results presented in Figure 9 and Table 5, we argue that prediction models of less than 3 hours are sufficiently competitive.

4 Conclusions and Summary

In this study, we have developed a new LSTM model for predicting foF2 and hmF2 to overcome the weaknesses of the LSTM quiet model developed by Moon et al. (2020). We collected 61 geomagnetic storm events from 1 January 2010 to August 2017 (about 7.7 years) for the training dataset. Also, the space environment indices related to geomagnetic storms were used together as training sets. Optimal hyper-parameters were searched for the LSTM models with each prediction target time. Three test events were selected to evaluate the performance of a geomagnetic storm-specific LSTM model (LSTM storm model), and correlation coefficient, RMSE, and MAPE scores were calculated and diagnosed for each model (LSTM storm, LSTM quiet, SAMI2, and IRI). We also discussed the performances and predictability of short-term (up to 3 hours) and long-term (longer than 3 hours up to 24 hours) prediction models.

Our results are summarized as follows:

1. In ionospheric positive storm (test event #1), the LSTM storm model for foF2 (hmF2) showed better performance in terms of RMSE by 32 (10)%, 34 (17)%, and 37 (5)% than the LSTM quiet, SAMI2, and IRI-2016 models, respectively.

2. Even if the LSTM storm model is specialized for storm conditions, our model predicts well enough for geomagnetic storms with no change in the ionosphere (test event #2 and #3).

3. We suggest that the deep-learning prediction model should be developed as the regional model using local observation data because it varies the region, date, local time, and space environment conditions.

4. Although the performance degrades significantly with increasing prediction times, short-term predictions (up to 3 hours) show an RMSE of less than ~1MHz for foF2 and 25 km for hmF2. therefore, we propose that short-term prediction models are sufficiently competitive.

This study used deep learning techniques to predict ionospheric storms during geomagnetic storm periods to overcome the limitations of previous learning techniques. Most significantly, this is the first attempt to develop a deep learning model by collecting only geomagnetic storm cases. Besides, our study is also meaningful in that it presents several possibilities for each prediction target model. Finally, we are confident that the results of this study will provide a means to find a way to respond to the risk factors from the ionospheric storm in the aspect of space weather.

Acknowledgments

This research was supported by basic research funding from Korea Astronomy and Space Science Institute (KASI). The Jeju ionosonde data were obtained from the Korean Space Weather Center homepage (https://spaceweather.rra.go.kr/observation/service/iono). The geomagnetic indices data can be downloaded from the OMNI online system (https://omniweb.gsfc.nasa.gov/ow.html). Also, the information of the geomagnetic storm events is obtained from the Japan Meteorological Agency/Kakioka Magnetic Observatory webpage (kakioka-jma.go.jp/en/). The SSLab team (Su-In Moon, Se-Heon Jeong, and YongHa Kim) at Chungnam National University helped design the skills of LSTM deep learning algorithms. We would also like to thank the Naval Research Laboratory (NRL) for providing the SAMI2 model (https://www.nrl.navy.mil/ppd/branches/6790/sami2) to be used and the IRI model developer for opening the IRI-2016 model (http://irimodel.org/) as open-source code. For the deep learning algorithm, we used the deep learning toolbox, a MATLAB code. Our analyzed data in this paper have been uploaded to Github (https://github.com/Jeongheon-Kim/2021SW_LSTM-storm).

References

Adebiyi, S. J., Adimula, I. A., Oladipo, O. A., Joshua, B. W., Adebesin, B. O., & Ikubanni, S. O. (2014). Ionospheric response to magnetic activity at low and mid-latitude stations. *Acta Geophysica*, *62*(4), 973–989. https://doi.org/10.2478/s11600-014-0205-x

Altinay, O., Tulunay, E., & Tulunay, Y. (1997). Forecasting of ionospheric critical frequency using neural networks. *Geophysical Research Letters*, *24*(12), 1467–1470. https://doi.org/10.1029/97GL01381

Athieno, R., Jayachandran, P. T., & Themens, D. R. (2017). A neural network-based foF2 model for a single station in the polar cap. *Radio Science*, *52*(6), 784–796. https://doi.org/10.1002/2016RS006192

Fagundes, P. R., Cardoso, F. A., Fejer, B. G., Venkatesh, K., Ribeiro, B. A. G., & Pillat, V. G. (2016). *Journal of Geophysical Research : Space Physics of March 2015 over the Brazilian sector*. 1–13. https://doi.org/10.1002/2015JA022214.Received

Fan, J., Liu, C., Lv, Y., Han, J., & Wang, J. (2019). A short-term forecast model of foF2 based on Elman neural network. *Applied Sciences (Switzerland)*, *9*(14). https://doi.org/10.3390/app9142782

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

Hu, A., & Zhang, K. (2018). Using bidirectional long short-term memory method for the height of F2 peak forecasting from ionosonde measurements in the Australian Region. *Remote Sensing*, *10*(10). https://doi.org/10.3390/rs10101658

Jin, S., Jin, R., & Kutoglu, H. (2017). Positive and negative ionospheric responses to the March 2015 geomagnetic storm from BDS observations. *Journal of Geodesy*, *91*(6), 613–626. https://doi.org/10.1007/s00190-016-0988-4

Kim, J. H., Kwak, Y. S., Kim, Y. H., Moon, S. I., Jeong, S. H., & Yun, J. Y. (2020). Regional Ionospheric Parameter Estimation by Assimilating the LSTM Trained Results Into the SAMI2 Model. *Space Weather*, *18*(10). https://doi.org/10.1029/2020SW002590

Lissa, D., Srinivasu, V. K. D., Prasad, D. S. V. V. D., & Niranjan, K. (2020). Ionospheric response to the 26 August 2018 geomagnetic storm using GPS-TEC observations along 80° E and 120° E longitudes in the Asian sector. *Advances in Space Research*, *66*(6), 1427–1440. https://doi.org/10.1016/j.asr.2020.05.025

McKinnell, L. A., & Poole, A. W. V. (2000). The development of a neural network based short term foF2 forecast program. *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial and Planetary Science*, *25*(4), 287–290. https://doi.org/10.1016/S1464-1917(00)00018-0

Moon, S., Kim, Y. H., Kim, J. H., Kwak, Y. S., & Yoon, J. Y. (2020). Forecasting the ionospheric F2 Parameters over Jeju Station (33.43°N, 126.30°E) by Using Long Short-Term Memory. *Journal of the Korean Physical Society*, *77*(12), 1265–1273. https://doi.org/10.3938/jkps.77.1265

Nakamura, M., Maruyama, T., & Shidama, Y. (2009). Using a neural network to make operational forecasts of ionospheric variations and storms at Kokubunji, Japan. *Journal of the National Institute of Information and Communications Technology*, *56*(1–4), 391–406.

Oyeyemi, E. O., Poole, A. W. V., & McKinnell, L. A. (2005). On the global model for foF2 using neural networks. *Radio Science*, *40*(6), 1–15. https://doi.org/10.1029/2004RS003223

Poole, A. W. V., & Poole, M. (2002). Long-term trends in f0F2 over Grahamstown using Neural Networks. *Annals of Geophysics*, *45*(1), 155–162. https://doi.org/10.4401/ag-3485

Sai Gowtam, V., & Tulasi Ram, S. (2017). An Artificial Neural Network-Based Ionospheric Model to Predict NmF2 and hmF2 Using Long-Term Data Set of FORMOSAT-3/COSMIC Radio Occultation Observations: Preliminary Results. *Journal of Geophysical Research: Space Physics*, *122*(11), 11,743-11,755. https://doi.org/10.1002/2017JA024795

Szuszczewicz, E. P., Lester, M., Wilkinson, P., Blanchard, P., Abdu, M., Hanbaba, R., Igarashi, K., Pulinets, S., & Reddy, B. M. (1998). A comparative study of global ionospheric responses to intense magnetic storm conditions. *Journal of Geophysical Research: Space Physics*, *103*(A6), 11665–11684. https://doi.org/10.1029/97ja01660

Tulasi Ram, S., Sai Gowtam, V., Mitra, A., & Reinisch, B. (2018). The Improved Two-Dimensional Artificial Neural Network-Based Ionospheric Model (ANNIM). *Journal of Geophysical Research: Space Physics*, *123*(7), 5807–5820. https://doi.org/10.1029/2018JA025559

Wichaipanich, N., Hozumi, K., Supnithi, P., & Tsugawa, T. (2017). A comparison of neural network-based predictions of foF2 with the IRI-2012 model at conjugate points in Southeast Asia. *Advances in Space Research*, *59*(12), 2934–2950. https://doi.org/10.1016/j.asr.2017.03.023

Williscroft, L. A., & Poole, A. W. V. (1996). Neural networks, foF2, sunspot number and magnetic activity. *Geophysical Research Letters*, *23*(24), 3659–3662. https://doi.org/10.1029/96GL03472

Wintoft, P., & Cander, L. R. (2000). Ionospheric foF2 storm forecasting using neural networks. *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial and Planetary Science*, *25*(4), 267–273. https://doi.org/10.1016/S1464-1917(00)00015-5

Yue, X., Wan, W., Liu, L., Ning, B., & Zhao, B. (2006). Applying artificial neural network to derive long-term foF2 trends in the Asia/Pacific sector from ionosonde observations. *Journal of Geophysical Research: Space Physics*, *111*(10), 1–11. https://doi.org/10.1029/2005JA011577

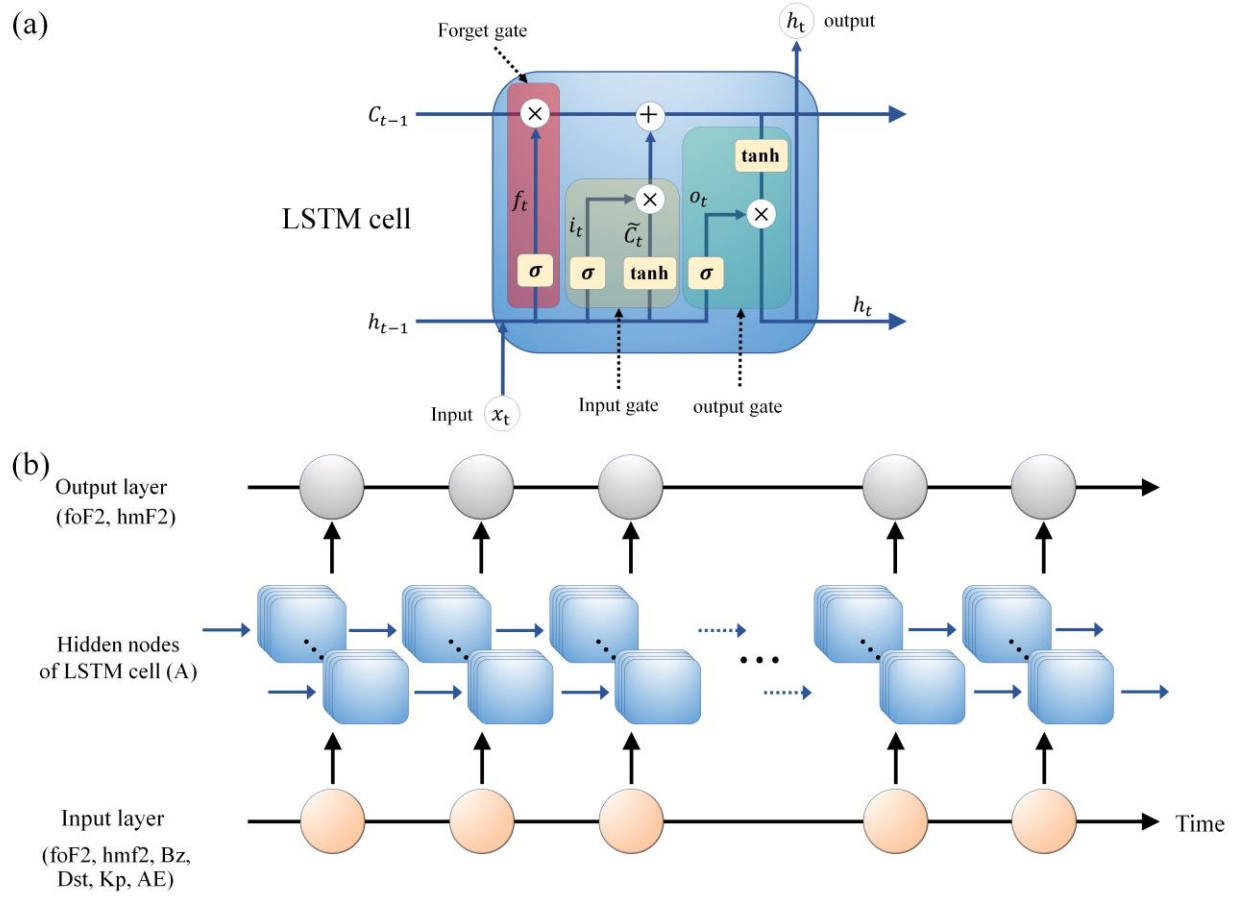


Figure 1. Detailed schematic diagrams of the newly developed LSTM model. (a) one LSTM cell (b) system architectures of the overall model. X_t *and* h_t are the input and output values. A LSTM cell includes sigmoid (σ) layers and hyperbolic tangent functions (tanh).

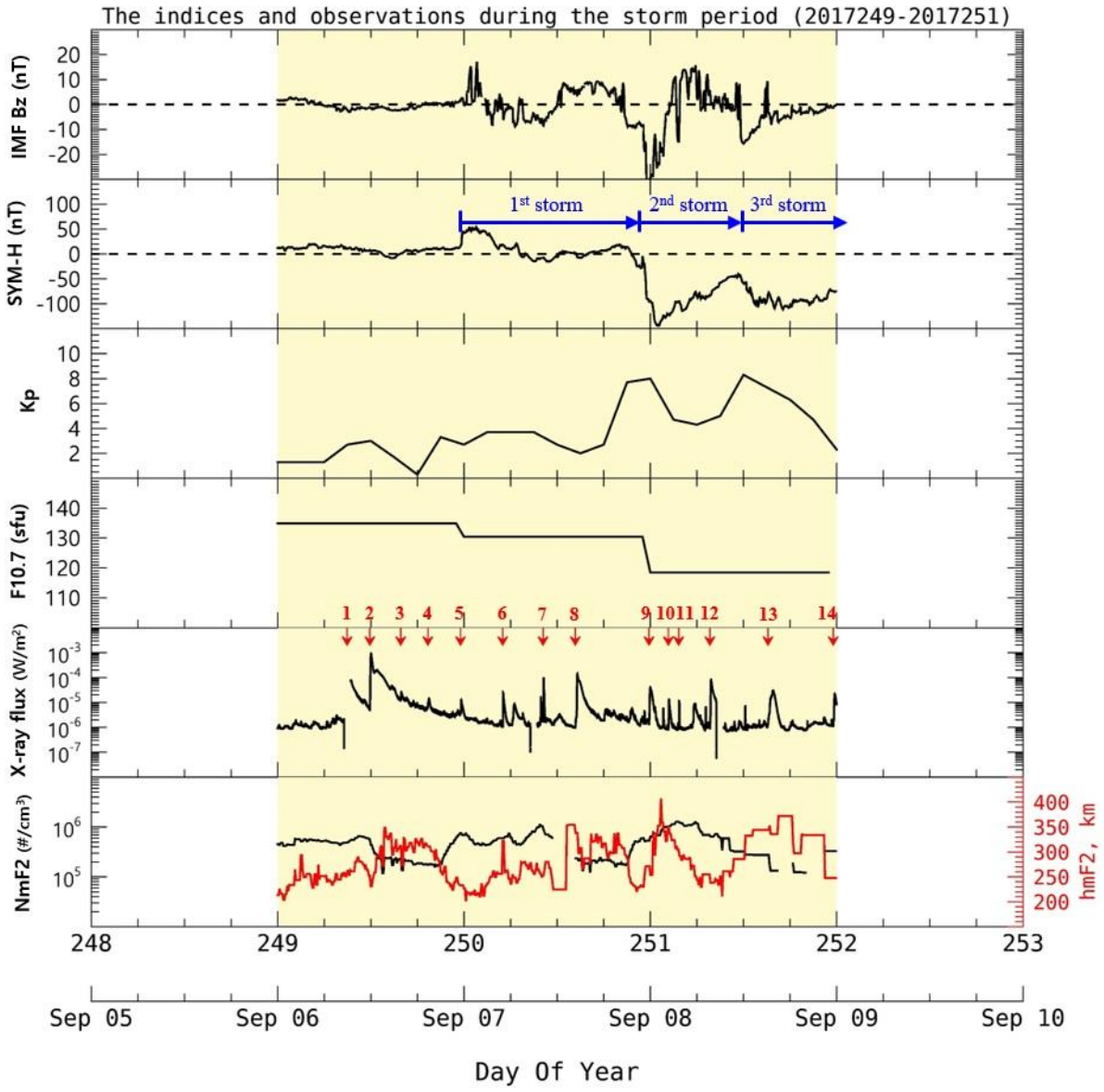


Figure 2. The IMF Bz, SYM-H, Kp, F10.7, GOES X-ray data, and ionospheric parameters (NmF2, hmF2) observed by ionosonde at Jeju station during the geomagnetic storm periods (The shaded box indicates the period of the storm test set, from 06 September through 08 September in 2017). The blue arrows in the second panel indicate the durations of three geomagnetic storms. The red arrows and numbers in the fifth panel represent the solar flare events (> M class). It is taken from Figure 2 of the *Kim et al.* (2020).

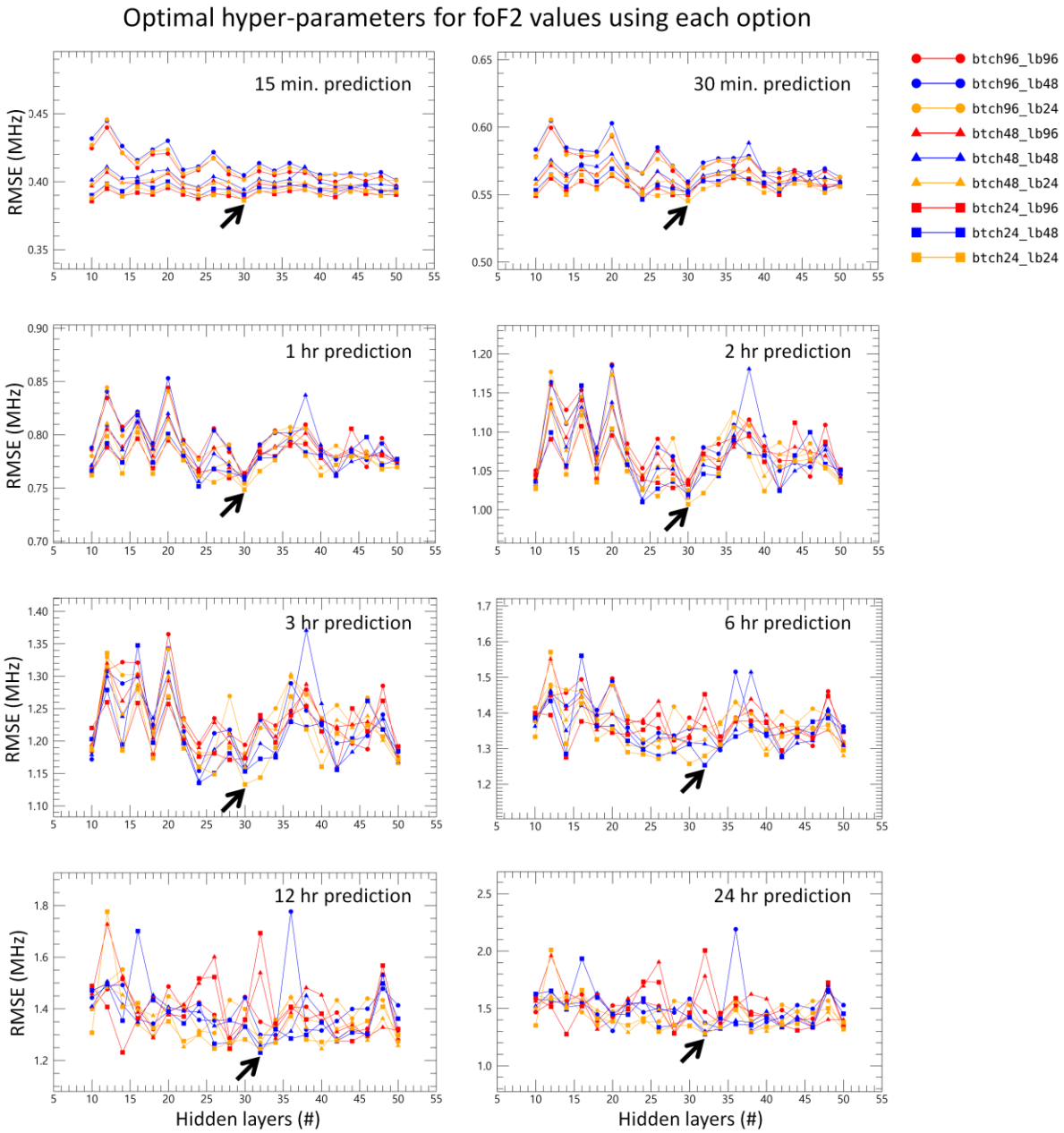


Figure 3. Optimal hyper-parameters for foF2 values using each option. The best options for each prediction model are indicated by black arrows. (btch = batch size, lb = lookback)

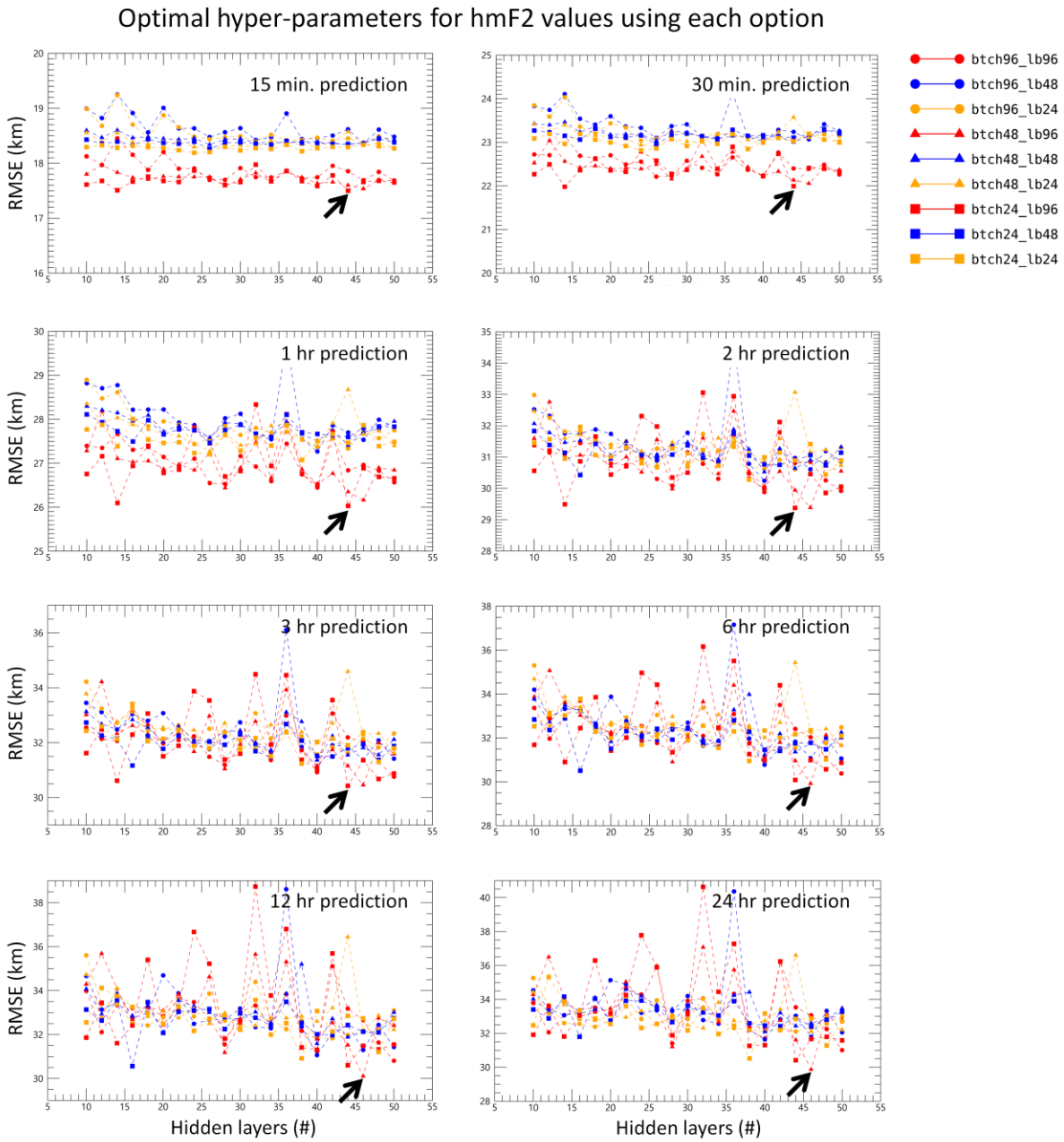


Figure 4. Same formats in Figure 3 for hmF2 values using each option. The best options for each prediction model are indicated by black arrows. (btch = batch size, lb = lookback)

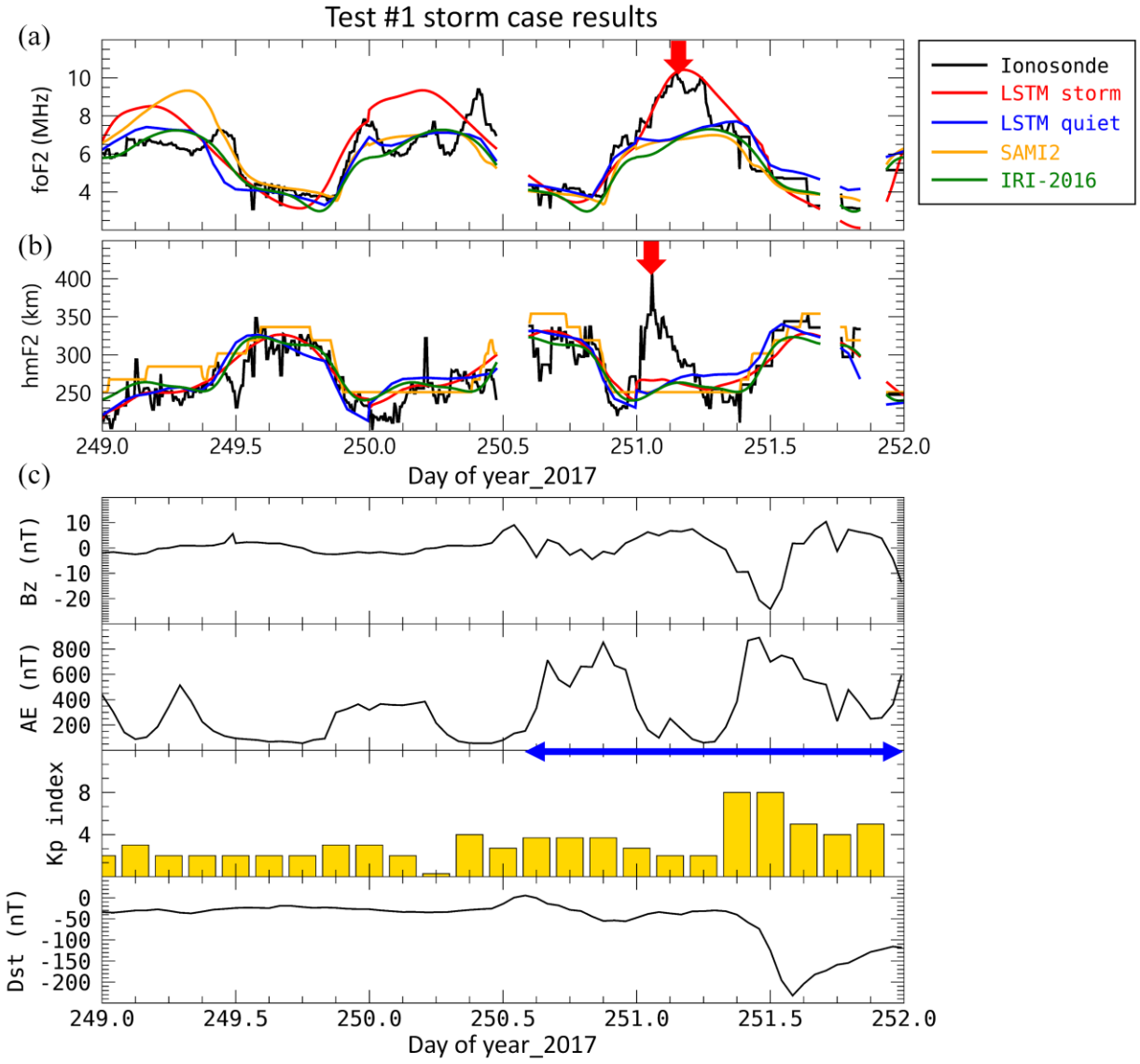


Figure 5. (a)-(b) Prediction results of foF2 and hmF2 at Jeju location of each model for test #1 storm case. The black line indicates the Jeju ionosonde observations, and red (blue) ones mean the LSTM storm (quiet) model values. The orange (green) lines are the SAMI2 and IRI-2016 model. The LSTM quiet model is taken from the results of *Moon et al. (2020)*. (c) The IMF Bz, AE, Kp and Dst indices. The blue arrow in the panel (c) indicate the storm period, and the vertical red ones in the panel (a) and (b) show the ionospheric positive storm and elevated F2 layer, respectively.

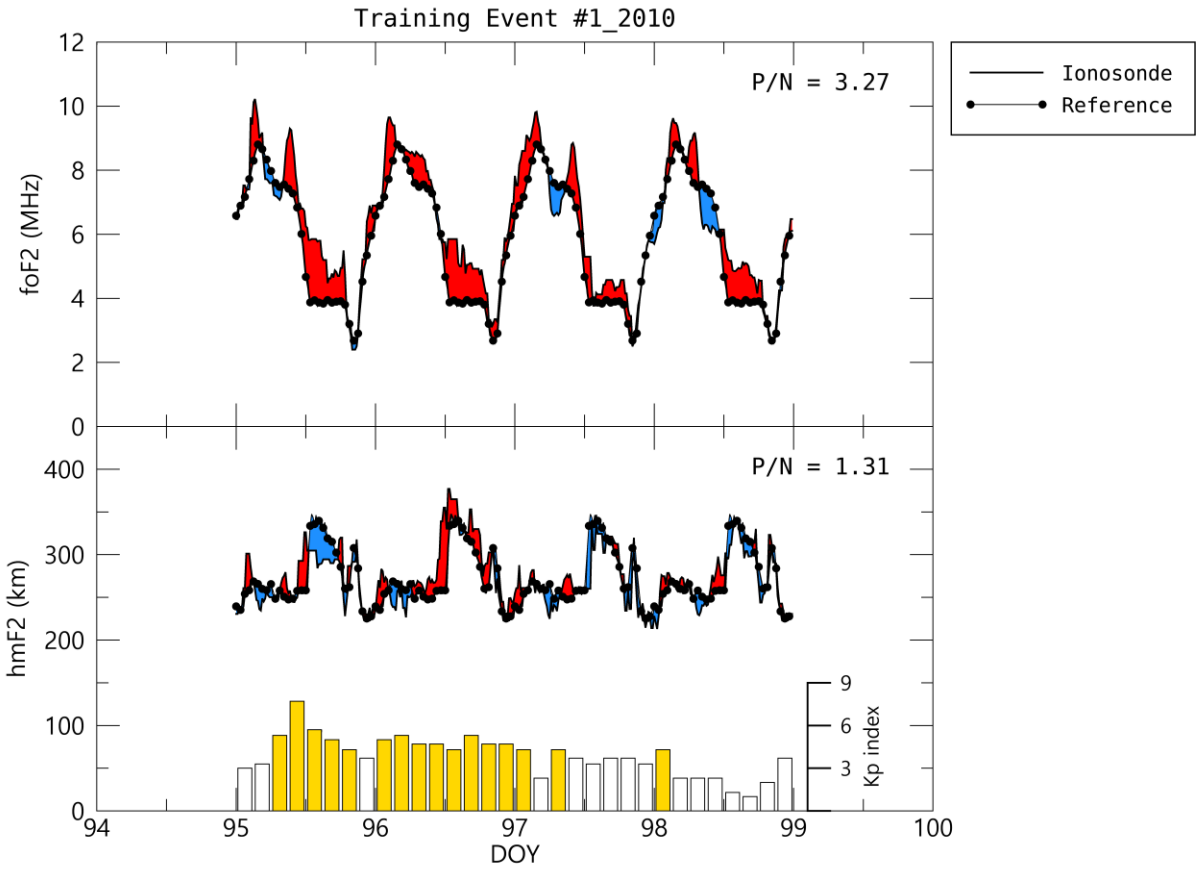


Figure 6. Ionosonde observations (foF2 and hmF2) for training events #1 in 2010 and the 10 days averaged reference line for the quiet condition. The black (dotted) line means the observations (reference). The red (blue) colored regions indicate the positive (negative) storm. Also, we marked the bar-plot in the bottom figure for the geomagnetic activity and colored bars in gold which indicate the Kp of 4 or higher.

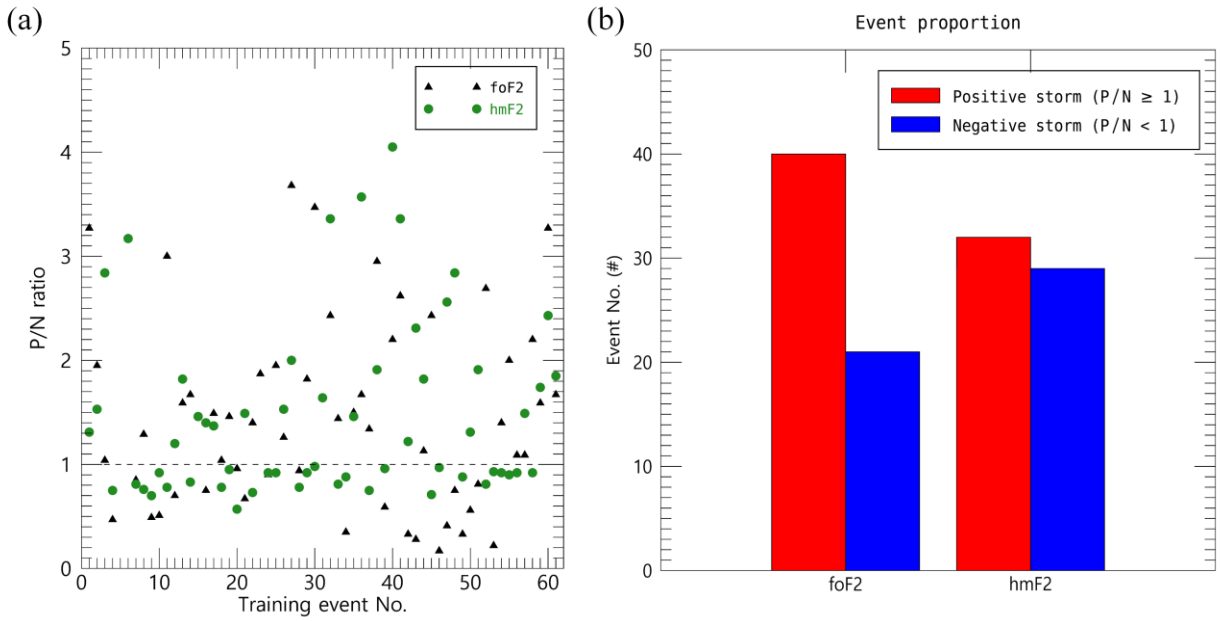


Figure 7. (a) The positive and negative (P/N) ratios of each training event. The black triangle (green circle) symbols indicate the foF2 (hmF2) values. A value greater than 1 (= dashed line) means a positive event, and a value of less than 1 means a negative event. (b) The event proportions as the plotted histogram of 61 storm events. The positive storm colored red, and the negative one colored blue.

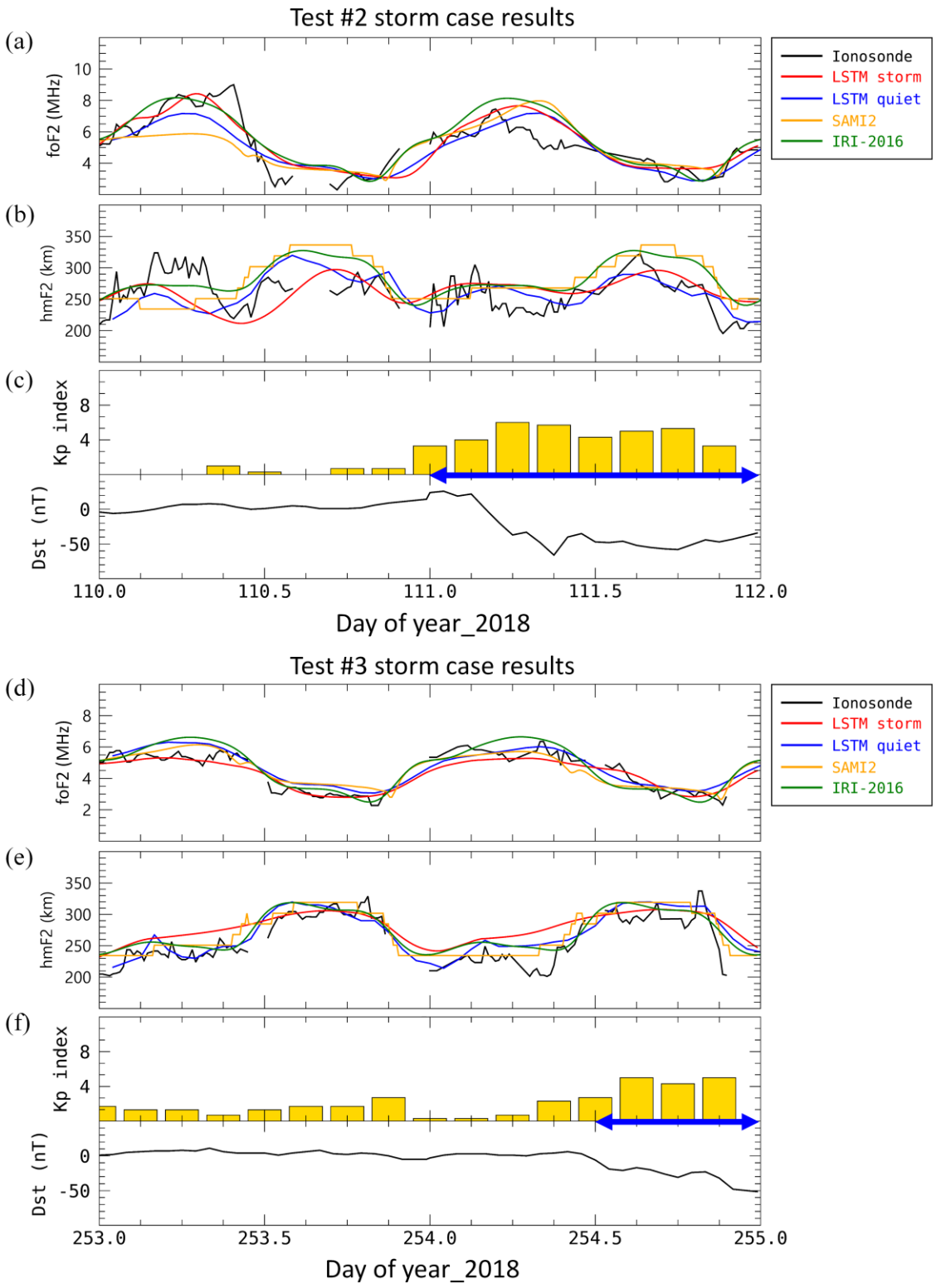


Figure 8. Prediction results at Jeju location of each model for test #2 (a-c) & #3 (d-f) storm cases. Same formats in Figure 5.

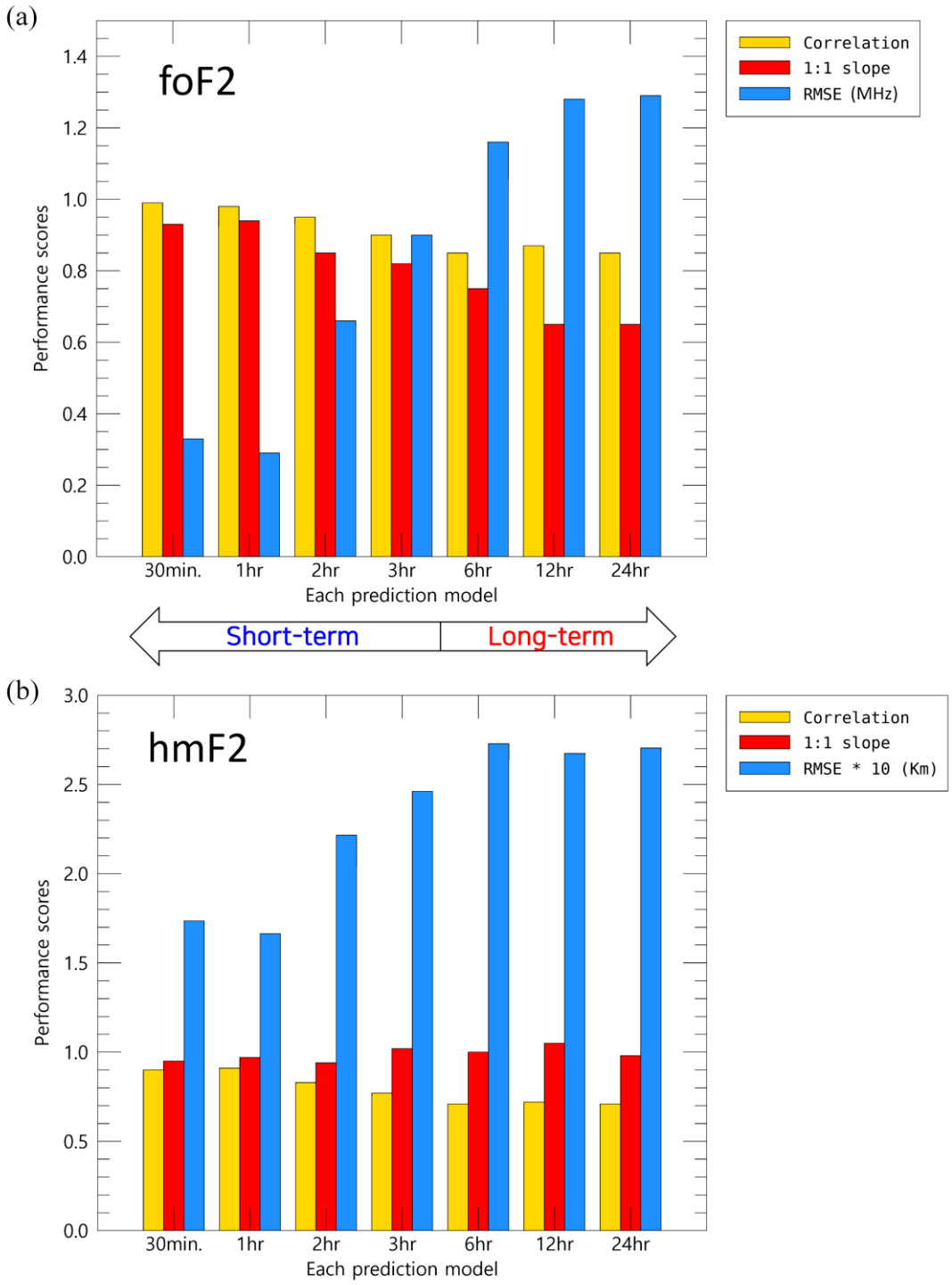


Figure 9. The performance skill scores for (a) foF2 and (b) hmF2 of each LSTM prediction model. The gold, red, and blue histograms mean the correlation coefficient, 1:1 slope, and the RMSE values, respectively.

Table 1. The list of geomagnetic storm events.

Data type	Event No.	Begin (yyyy-mm-dd hr)	Main phase (yyyy-mm-dd hr)	Last phase (yyyy-mm-dd hr)	End (yyyy-mm-dd hr)	Max K index
Training	1	2010-04-05 08.4	2010-04-05 09.3	2010-04-06 07.7	2010-04-08 14	7
Training	2	2010-08-03 17.7	2010-08-03 20.8	2010-08-04 00.8	2010-08-05 05	6
Training	3	2011-02-18 01.5	2011-02-18 04.8	2011-02-18 08.7	2011-02-18 18	6
Training	4	2011-03-01 05.2	2011-03-01 09.5	2011-03-01 14.6	2011-03-02 21	6
Training	5	2011-03-10 06.5	2011-03-10 23.5	2011-03-11 07.0	2011-03-12 03	5
Training	6	2011-04-06 09.5	2011-04-06 10.4	2011-04-06 18.4	2011-04-07 00	5
Training	7	2011-05-28 06.5	2011-05-28 06.8	2011-05-28 11.9	2011-05-30 00	5
Training	8	2011-06-04 20.7	2011-06-04 21.8	2011-06-05 00.9	2011-06-06 03	5
Training	9	2011-08-05 17.8	2011-08-05 21.6	2011-08-06 03.5	2011-08-06 15	6
Training	10	2011-09-09 12.7	2011-09-09 14.3	2011-09-10 01.5	2011-09-11 01	5
Training	11	2011-09-17 03.7	2011-09-17 08.3	2011-09-17 14.9	2011-09-18 01	6
Training	12	2011-09-26 12.5	2011-09-26 15.3	2011-09-27 00.5	2011-09-28 17	6
Training	13	2011-10-05 07.5	2011-10-05 08.9	2011-10-05 14.5	2011-10-06 01	5
Training	14	2011-10-24 18.5	2011-10-24 21.8	2011-10-25 01.2	2011-10-25 21	6
Training	15	2011-11-01 09.1	2011-11-01 09.4	2011-11-01 13.3	2011-11-02 00	5
Training	16	2012-01-22 06.2	2012-01-22 10.3	2012-01-22 12.5	2012-01-23 09	5
Training	17	2012-01-24 15.0	2012-01-25 08.6	2012-01-25 10.3	2012-01-25 18	5
Training	18	2012-03-07 04.3	2012-03-07 05.5	2012-03-07 09.0	2012-03-08 00	5
Training	19	2012-03-08 11.0	2012-03-08 15.4	2012-03-09 08.2	2012-03-10 19	7
Training	20	2012-03-12 09.2	2012-03-12 11.3	2012-03-12 12.7	2012-03-13 00	6
Training	21	2012-04-23 03.3	2012-04-23 21.1	2012-04-24 01.1	2012-04-26 16	5
Training	22	2012-07-14 18.2	2012-07-15 05.5	2012-07-15 09.5	2012-07-17 12	6
Training	23	2012-09-30 11.5	2012-09-30 14.0	2012-10-01 01.2	2012-10-01 16	5
Training	24	2012-10-08 05.3	2012-10-08 05.5	2012-10-08 12.1	2012-10-10 00	6
Training	25	2012-10-13 03.4	2012-10-13 03.5	2012-10-13 11.0	2012-10-14 22	5
Training	26	2012-10-31 15.6	2012-10-31 04.4	2012-11-01 12.2	2012-11-02 00	5
Training	27	2012-11-13 23.2	2012-11-13 23.0	2012-11-14 09.6	2012-11-14 19	5
Training	28	2013-01-17 13.2	2013-01-17 14.0	2013-01-18 0.00	2013-01-18 06	6
Training	29	2013-03-01 08.0	2013-03-01 08.0	2013-03-01 10.8	2013-03-02 09	5
Training	30	2013-03-17 06.0	2013-03-17 06.6	2013-03-17 12.5	2013-03-18 12	6
Training	31	2013-04-30 09.8	2013-05-01 05.9	2013-05-01 13.2	2013-05-02 12	5
Training	32	2013-10-02 01.9	2013-10-02 05.0	2013-10-02 06.0	2013-10-03 03	6
Training	33	2013-10-08 20.4	2013-10-08 23.7	2013-10-09 02.0	2013-10-10 18	5
Training	34	2013-11-07 04.7	2013-11-07 04.7	2013-11-07 11.9	2013-11-07 17	5
Training	35	2013-11-09 03.5	2013-11-09 03.6	2013-11-09 11.0	2013-11-11 19	5
Training	36	2013-12-07 22.5	2013-12-08 03.0	2013-12-08 08.6	2013-12-08 22	6
Training	37	2013-12-25 04.9	2013-12-25 07.8	2013-12-25 12.6	2013-12-25 21	5
Training	38	2014-02-07 17.0	2014-02-07 17.1	2014-02-08 00.0	2014-02-10 00	5
Training	39	2014-02-18 13.9	2014-02-18 15.3	2014-02-19 09.0	2014-02-19 23	6
Training	40	2014-02-20 03.3	2014-02-20 03.8	2014-02-20 13.1	2014-02-21 00	6

Data type	Event No.	Begin (yyyy-mm-dd hr)	Main phase (yyyy-mm-dd hr)	Last phase (yyyy-mm-dd hr)	End (yyyy-mm-dd hr)	Max K index
Training	41	2014-02-27 16.8	2014-02-27 16.9	2014-02-28 00.0	2014-02-28 12	5
Training	42	2014-08-27 03.0	2014-08-27 04.2	2014-08-27 16.2	2014-08-30 15	5
Training	43	2014-09-12 15.9	2014-09-12 22.0	2014-09-13 03.7	2014-09-14 00	5
Training	44	2014-11-10 02.3	2014-11-10 10.4	2014-11-10 17.5	2014-11-10 24	5
Training	45	2015-01-07 06.3	2015-01-07 08.4	2015-01-07 11.0	2015-01-08 18	6
Training	46	2015-03-17 04 8	2015-03-17 07.0	2015-03-17 23.0	2015-03-21 15	7
Training	47	2015-05-12 22.0	2015-05-13 05.3	2015-05-13 06.8	2015-05-14 01	5
Training	48	2015-06-08 05.0	2015-06-08 05.6	2015-06-08 08.0	2015-06-09 00	6
Training	49	2015-07-23 03.0	2015-07-23 03.7	2015-07-23 08.8	2015-07-23 16	5
Training	50	2015-09-08 22.3	2015-09-09 00.0	2015-09-09 09.7	2015-09-10 03	5
Training	51	2015-09-11 04.2	2015-09-11 07.5	2015-09-11 09.6	2015-09-11 19	6
Training	52	2015-09-20 06.0	2015-09-20 06.7	2015-09-20 11.1	2015-09-20 21	6
Training	53	2015-10-07 04.4	2015-10-07 04.7	2015-10-07 09.3	2015-10-10 02	6
Training	54	2015-10-18 06.1	2015-10-18 07.4	2015-10-18 10.0	2015-10-18 20	6
Training	55	2015-11-03 01.6	2015-11-03 08.0	2015-11-03 12.4	2015-11-04 21	6
Training	56	2015-11-06 18.3	2015-11-06 23.5	2015-11-07 08.2	2015-11-07 21	6
Training	57	2015-12-14 13.3	2015-12-14 17.0	2015-12-14 19.1	2015-12-15 19	6
Training	58	2015-12-19 16.3	2015-12-19 04.7	2015-12-21 02.1	2015-12-22 02	6
Training	59	2015-12-31 00.8	2015-12-31 08.2	2015-12-31 14.1	2016-01-01 16	6
Training	60	2016-01-20 03.2	2016-01-20 08.1	2016-01-20 16.0	2016-01-21 08	5
Training	61	2016-02-16 08.3	2016-02-16 08.5	2016-02-18 00.0	2016-02-19 01	5
Validation	62	2016-03-06 13.3	2016-03-06 14.4	2016-03-07 08.3	2016-03-08 04	5
Validation	63	2016-03-14 17.2	2016-03-14 19.9	2016-03-15 07.4	2016-03-15 14	5
Validation	64	2016-04-14 07.7	2016-04-14 09.8	2016-04-14 12.1	2016-04-14 22	5
Validation	65	2016-09-29 04.4	2016-09-29 06.7	2016-09-29 09.3	2016-09-29 21	6
Validation	66	2016-10-12 22.2	2016-10-13 07.2	2016-10-13 15.9	2016-10-14 22	5
Validation	67	2017-03-27 03.4	2017-03-27 05.4	2017-03-27 10.3	2017-04-01 18	5
Validation	68	2017-08-31 05.6	2017-08-31 07.1	2017-08-31 11.8	2017-08-31 21	6
Test #1	69	2017-09-06 23.7	2017-09-07 05.1	2017-09-08 00.7	2017-09-09 07	7
Test #2	70	2018-04-20 00.3	2018-04-20 04.1	2018-04-20 09.9	2018-04-21 20	6
Test #3	71	2018-09-10 10.5	2018-09-10 12.3	2018-09-11 09.7	2018-09-11 21	5

Table 2. Detailed structures of the training, validation, and test data sets in our newly LSTM model.

	Training	Validation	Test
Target	Observed Dst, Kp, AE, Bz indices, hmF2 & foF2 of Jeju ionosonde	hmF2, foF2	hmF2, foF2
Period	05 Apr 2010 ~ 18 Feb 2016	06 Mar 2016 ~ 31 Aug 2017	#1: 06 Sep 2017 ~ 08 Sep 2017 #2: 20 Apr 2018 ~ 21 Apr 2018 #3: 10 Sep 2018 ~ 11 Sep 2018
Data points	117 days = 11,232 (89.3 %)	14 days = 1,344 (10.7 %)	7 days = 672
Storm event	61	7	3

Table 3. The selected optimal hyper-parameters for foF2 values. The gray (gold) background boxes indicate the best options for short-term (long-term) prediction.

Prediction target	15 min	30 min	1 hr	2 hr	3 hr	6 hr	12 hr	24 hr
Look ahead	1	2	4	8	12	24	48	96
Look back	24	24	24	24	24	48	48	48
Batch size	24	24	24	24	24	24	24	24
Hidden layer	30	30	30	30	30	32	32	32
RMSE (MHz)	0.38	0.54	0.74	1.00	1.13	1.25	1.23	1.27

Table 4. The selected optimal hyper-parameters for hmF2 values. The gray (gold) background boxes indicate the best options for short-term (long-term) prediction.

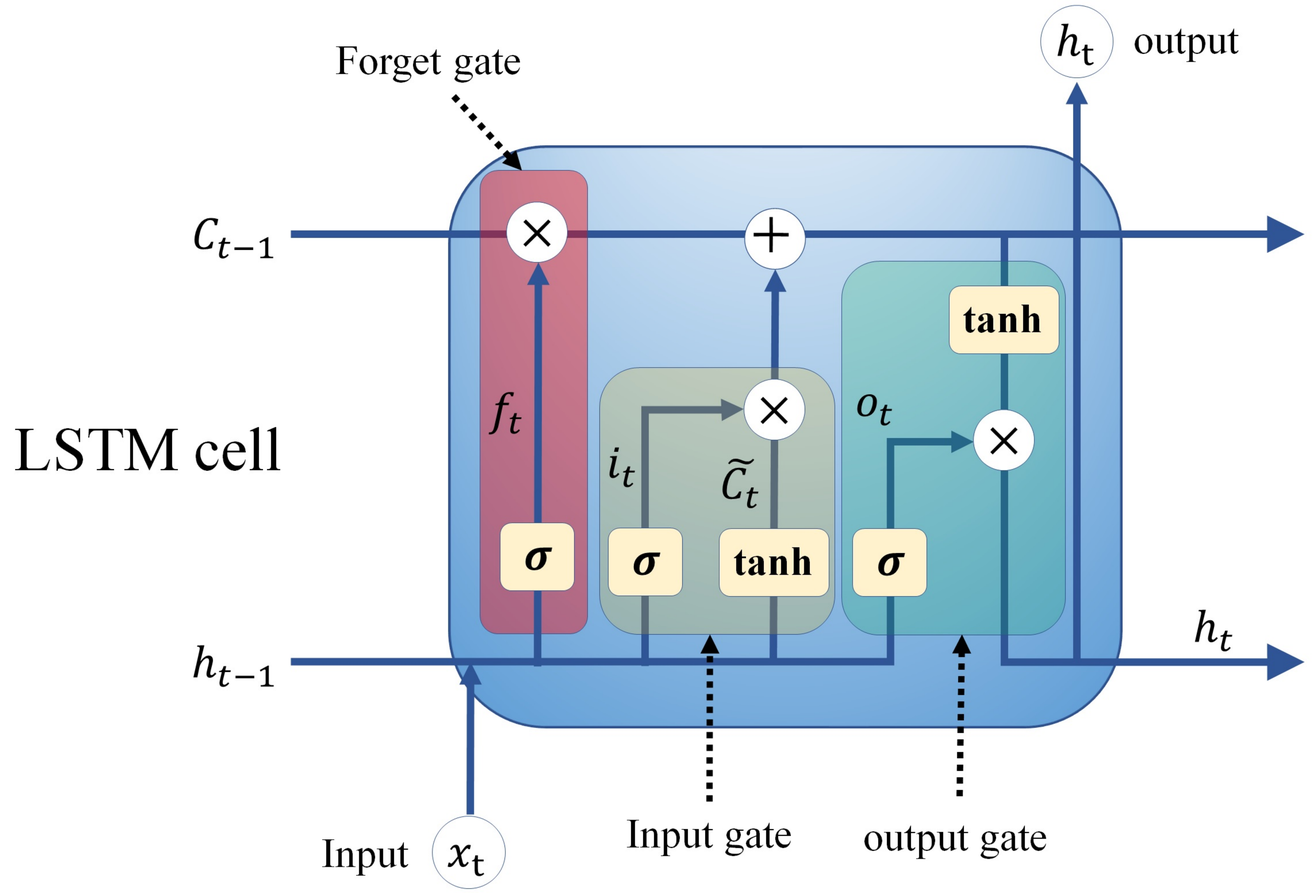
Prediction target	15 min	30 min	1 hr	2 hr	3 hr	6 hr	12 hr	24 hr
Look ahead	1	2	4	8	12	24	48	96
Look back	96	96	96	96	96	96	96	96
Batch size	24	24	24	24	24	48	48	48
Hidden layer	44	44	44	44	44	46	46	46
RMSE (km)	17.49	21.99	26.03	29.37	30.42	29.92	30.10	29.87

Table 5. The performance skill scores of three test events. (The units of foF2 and hmF2 are MHz and km, respectively. The gray boxes mean the best model.)

Scores \ Model		LSTM quiet		LSTM storm		SAMI2		IRI-2016	
		foF2	hmF2	foF2	hmF2	foF2	hmF2	foF2	hmF2
Test #1	CC	0.84	0.54	**0.95**	**0.61**	0.91	0.56	0.88	0.61
	RMSE	1.21	35.39	**0.82**	**31.70**	1.25	37.99	1.30	33.46
	MAPE	14.57	8.71	**12.65**	8.00	13.61	9.06	14.36	**7.73**
Test #2	CC	0.81	**0.34**	**0.81**	0.29	0.81	0.2	0.78	0.28
	RMSE	1.09	41.18	**1.07**	**29.42**	1.12	34.15	1.14	33.17
	MAPE	**18.5**	11.74	19.8	**9.47**	22.2	11.12	21.65	10.81
Test #3	CC	0.91	0.87	**0.95**	0.87	0.89	0.90	0.91	**0.91**
	RMSE	0.57	**21.62**	**0.47**	30.59	0.56	21.81	0.69	22.24
	MAPE	12.02	**6.75**	**8.60**	10.59	11.76	7.26	12.38	7.81

Figure 1.

(a)


Forget gate
h_t output
C_{t-1}
LSTM cell
f_t
i_t
$\widetilde{C}_t$
o_t
tanh
σ
σ
tanh
σ
h_t
h_{t-1}
Input x_t
Input gate
output gate

(b)

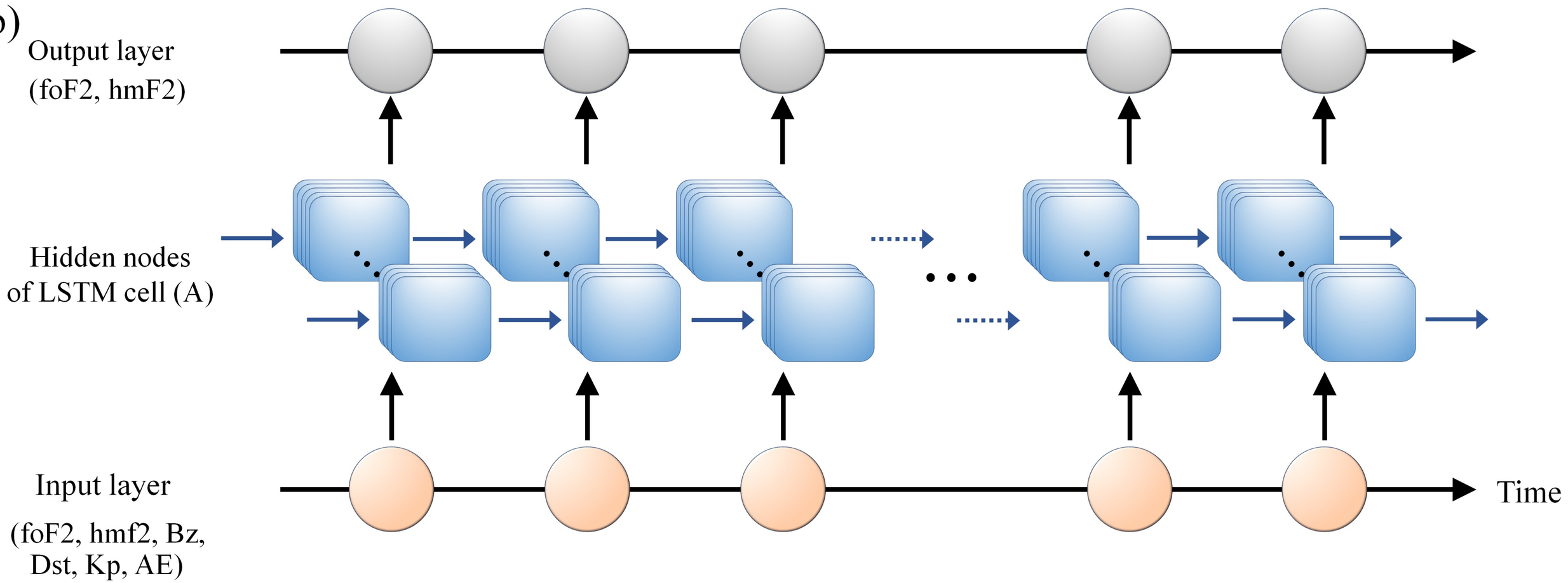

Output layer
(foF2, hmF2)
Hidden nodes
of LSTM cell (A)
Input layer
(foF2, hmf2, Bz,
Dst, Kp, AE)
Time

Figure 2.

The indices and observations during the storm period (2017249-2017251)

IMF Bz (nT)

SYM-H (nT)

1st storm 2nd storm 3rd storm

Kp

F10.7 (sfu)

X-ray flux (W/m²)

1 2 3 4 5 6 7 8 9 10 11 12 13 14

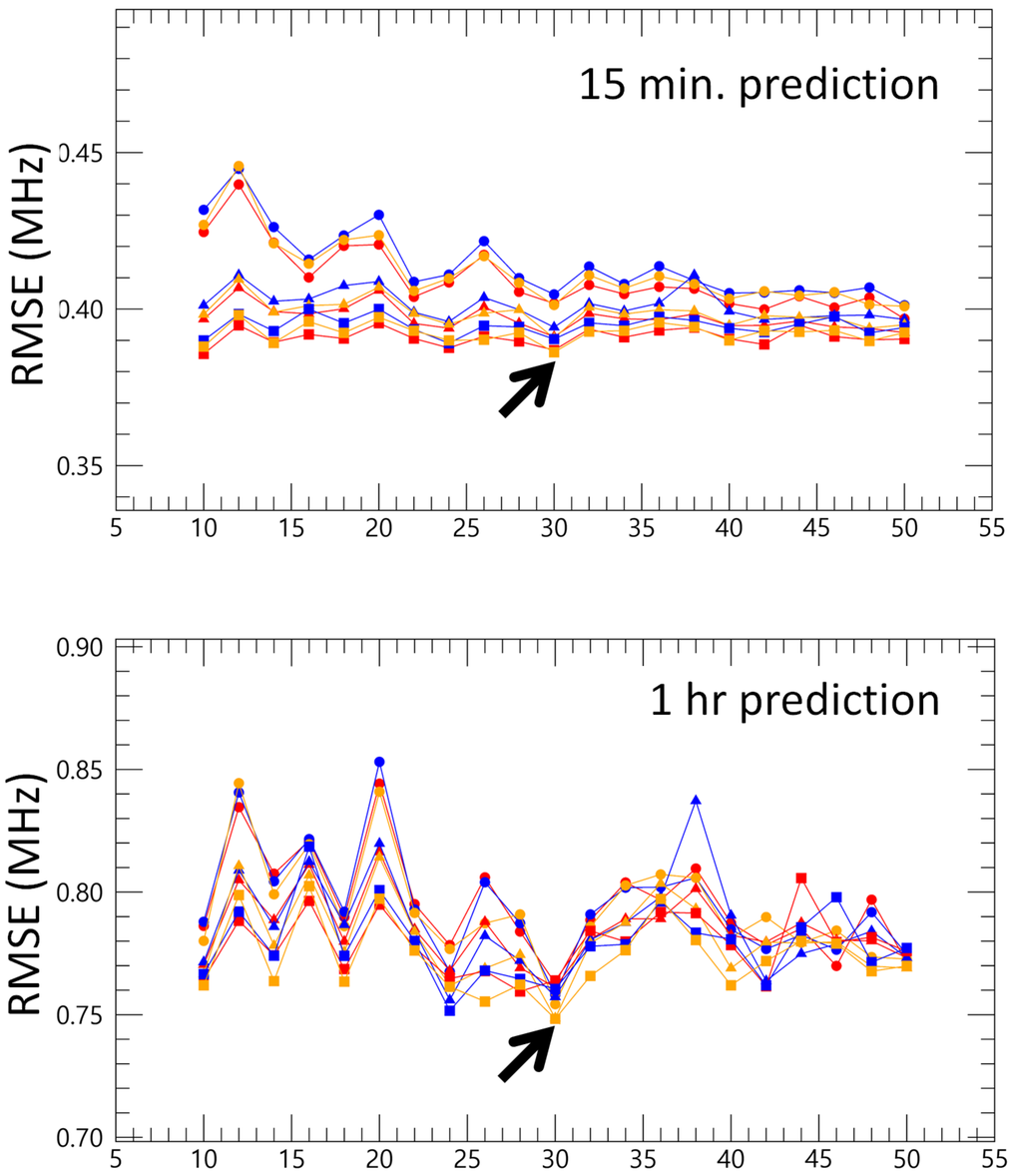
NmF2 (#/cm³)

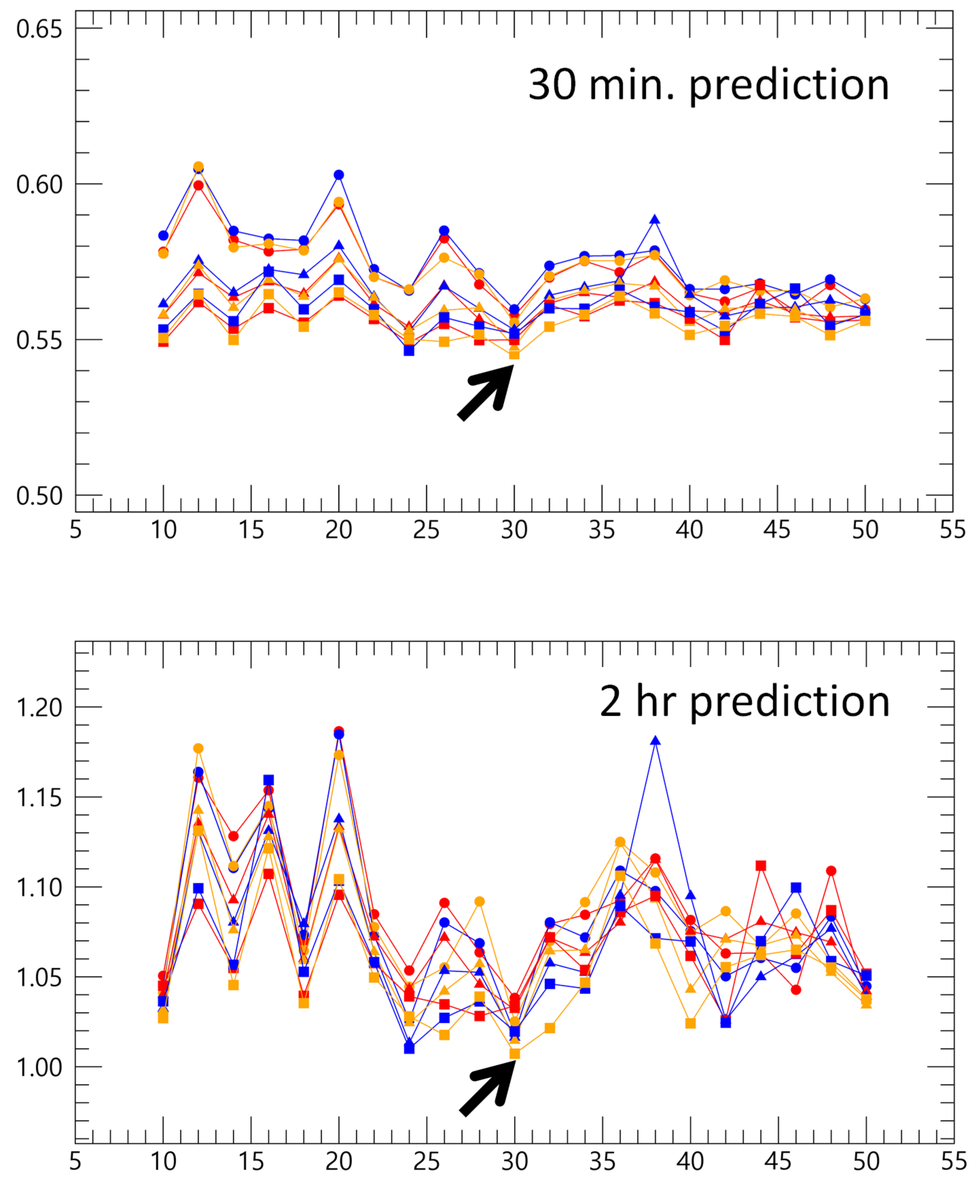
hmF2, km

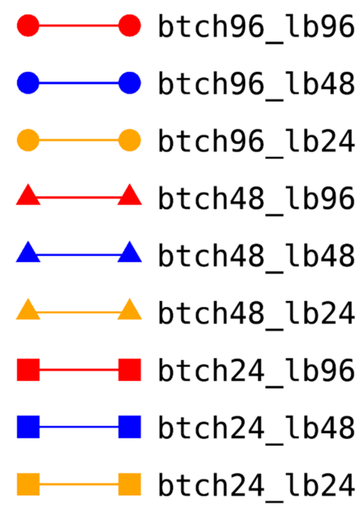
Day Of Year

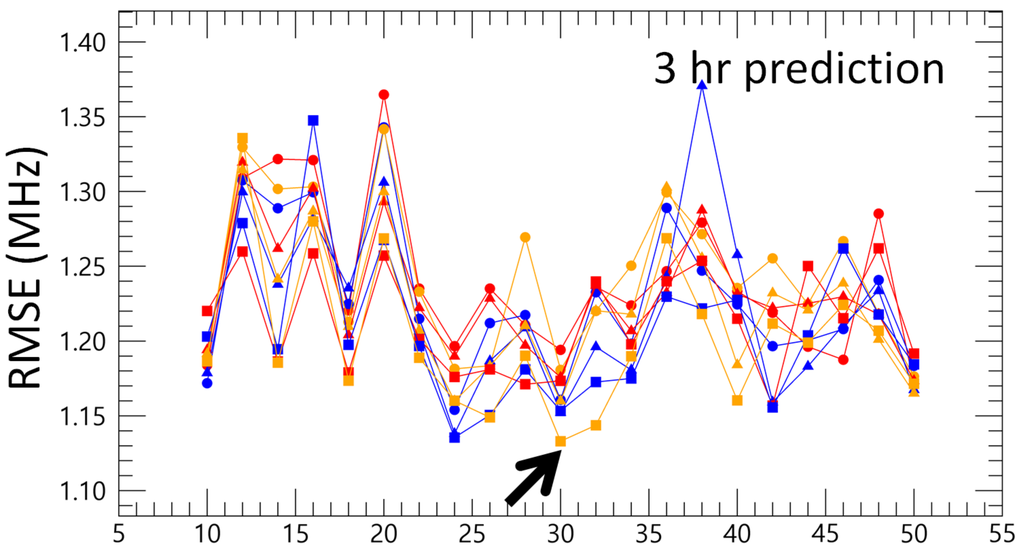
Figure 3.

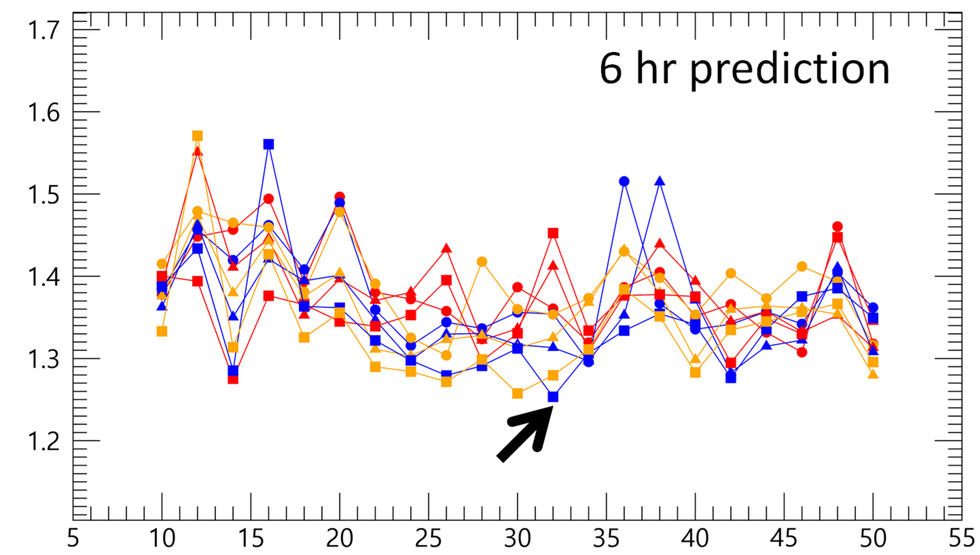
Optimal hyper-parameters for foF2 values using each option

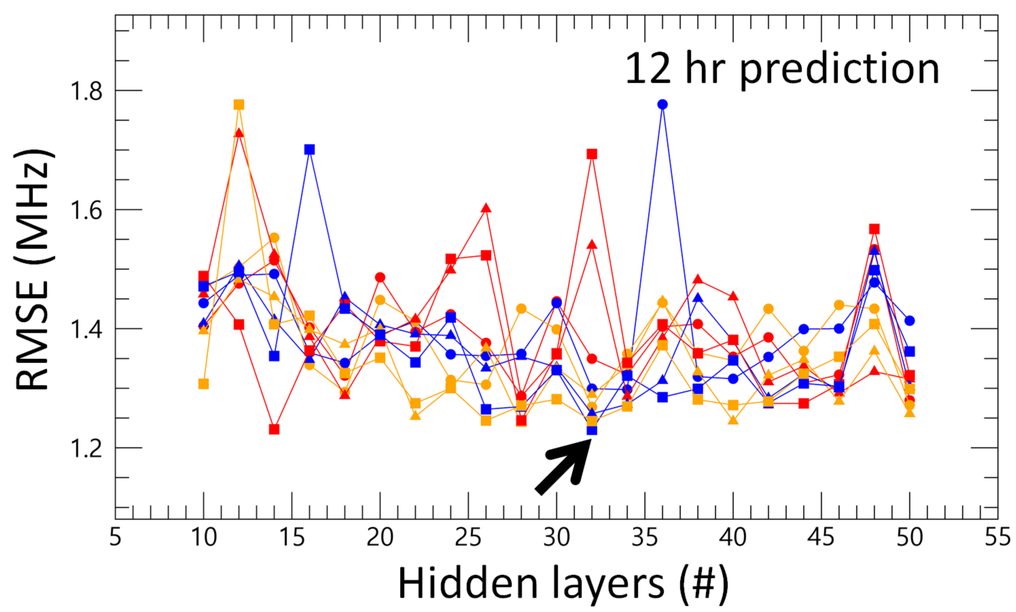












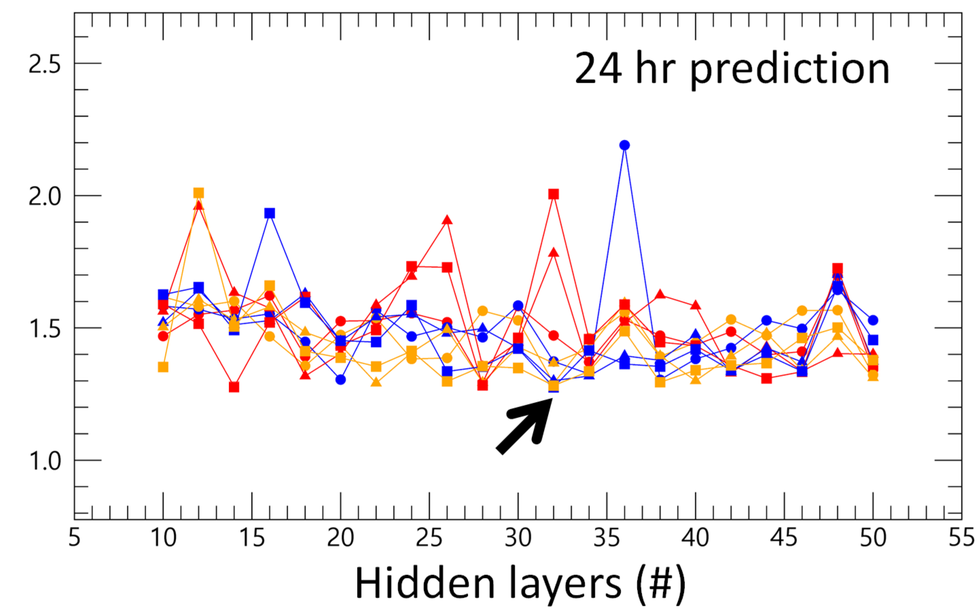
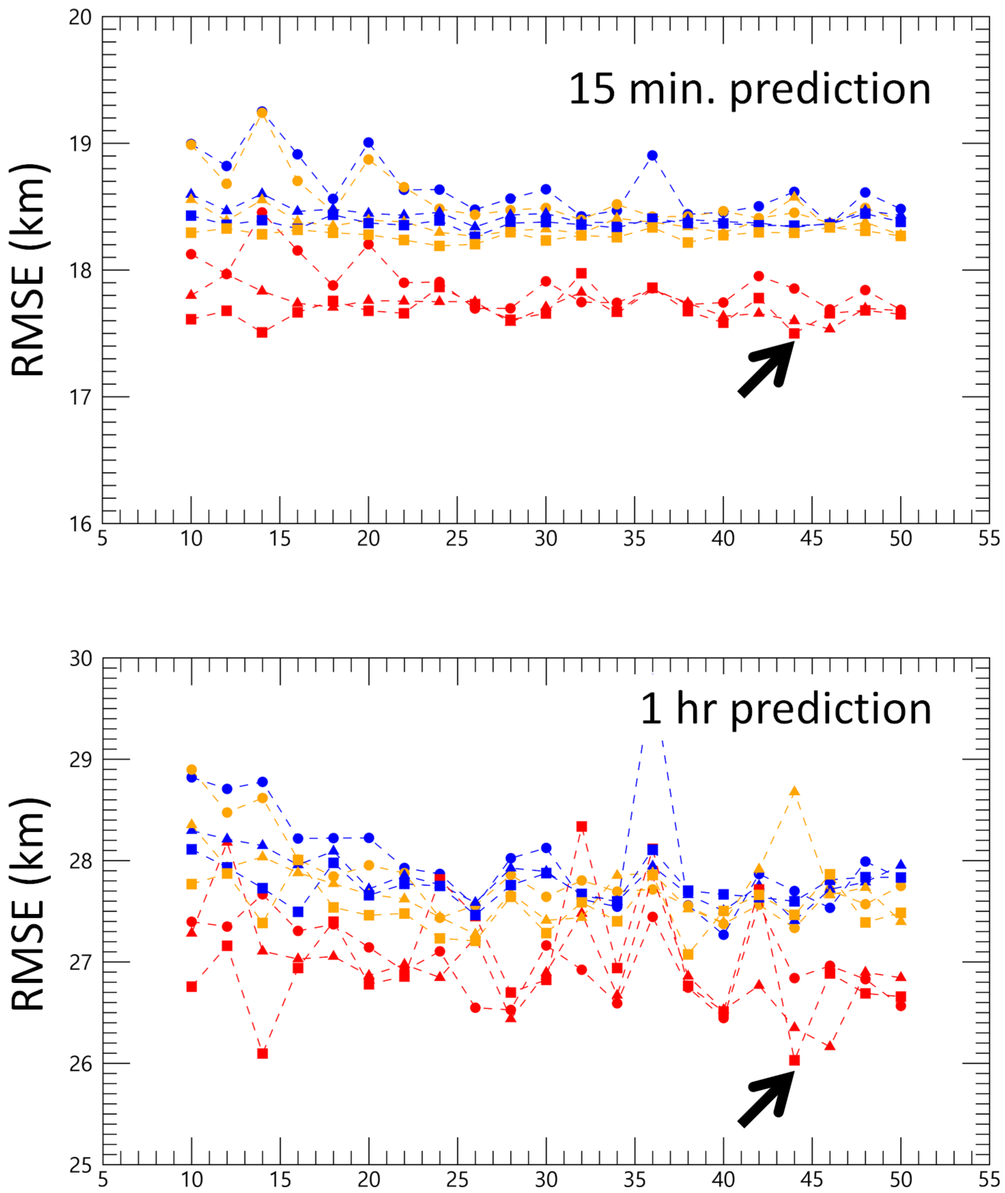
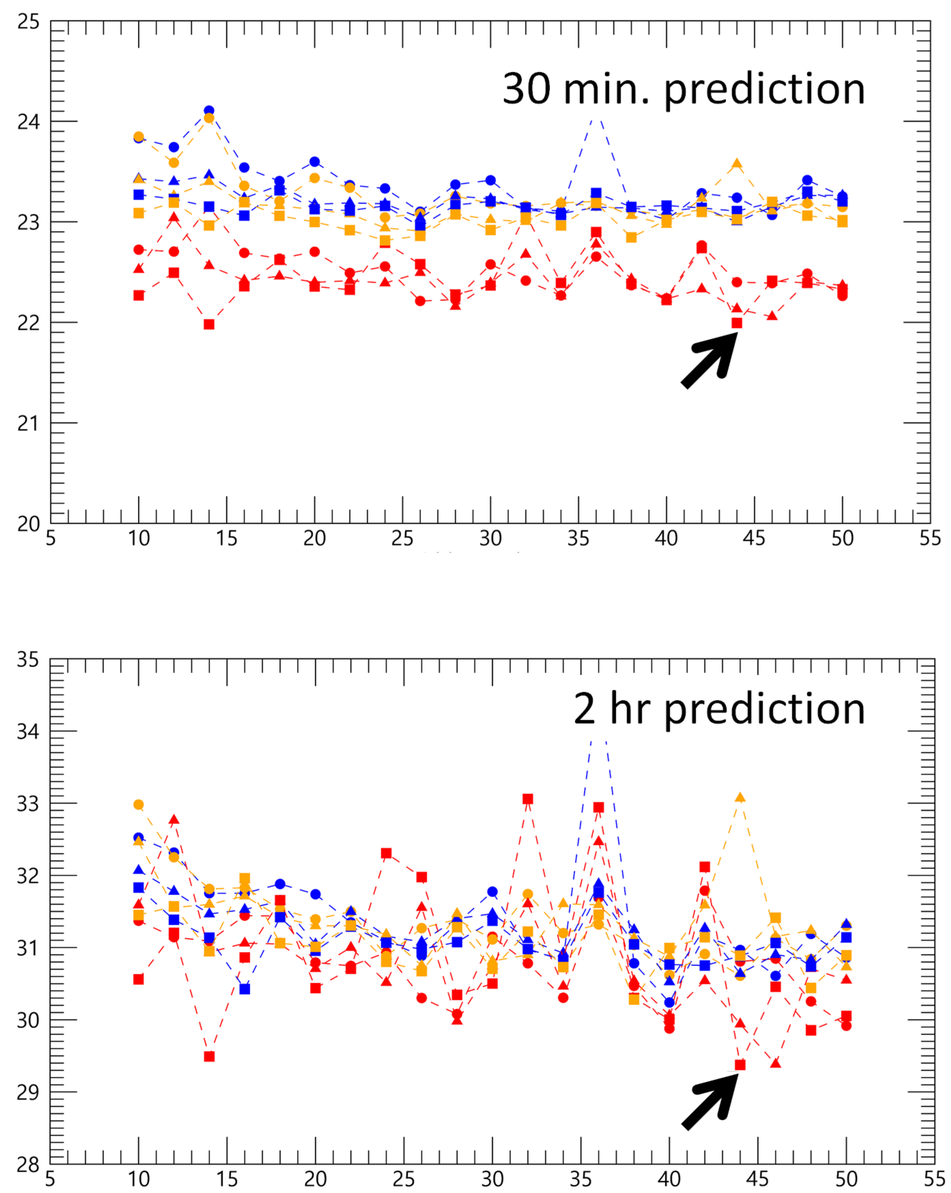
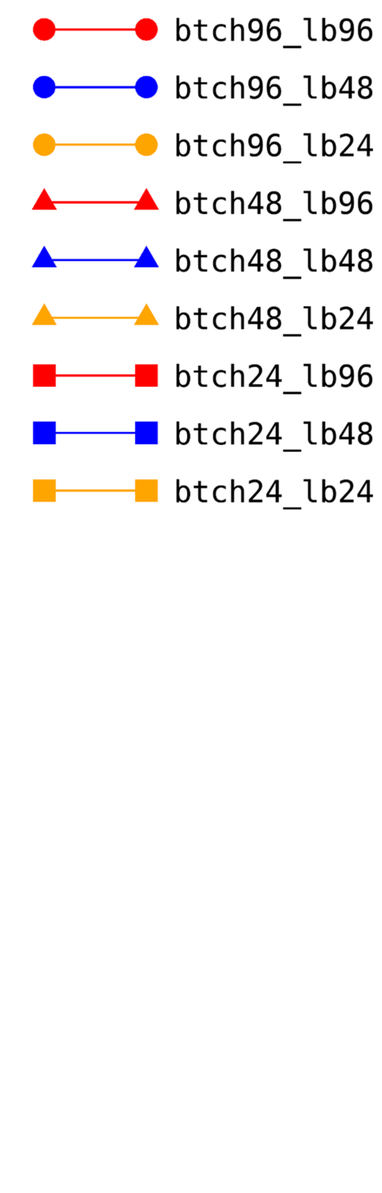


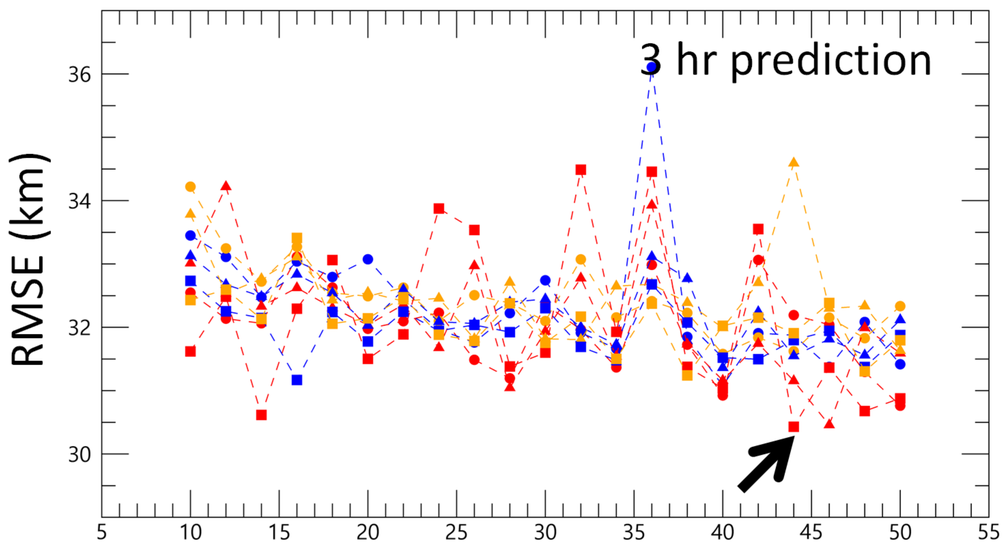
Figure 4.

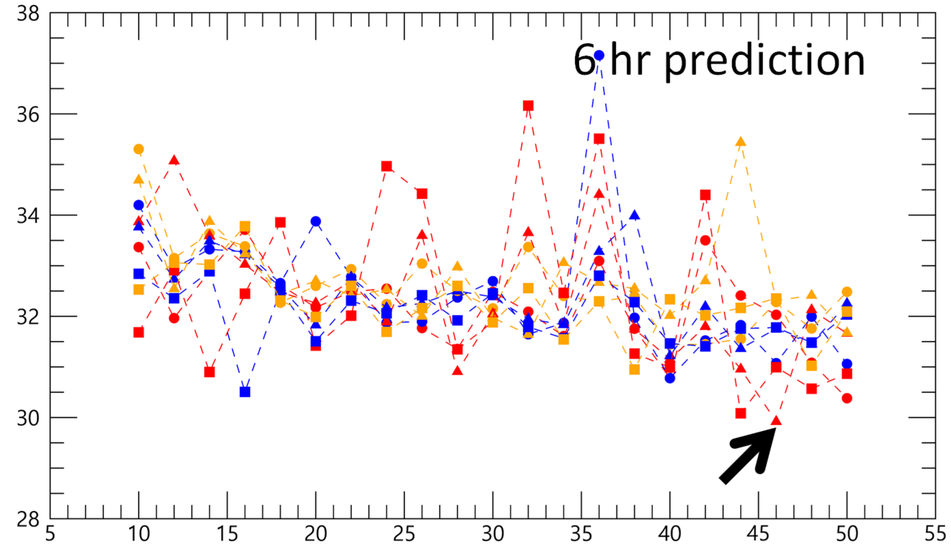
Optimal hyper-parameters for hmF2 values using each option

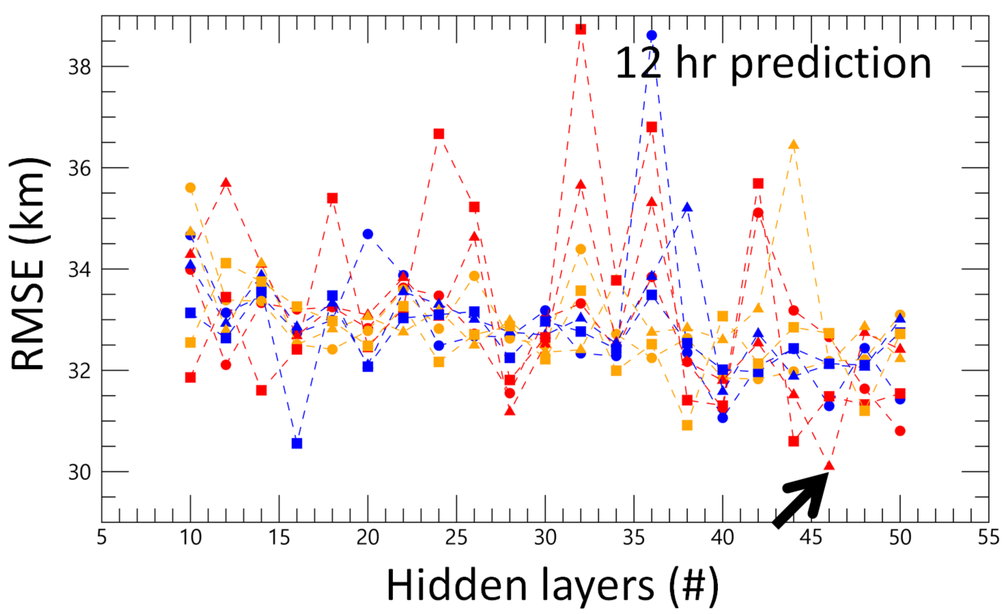












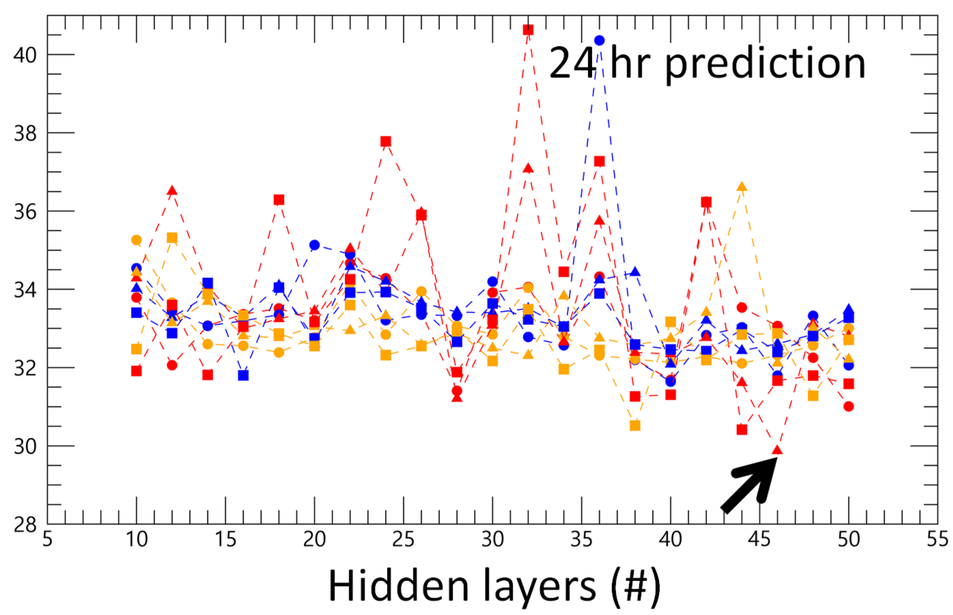


Figure 5.

Test #1 storm case results

(a)

foF2 (MHz)

(b)

hmF2 (km)

Ionosonde
LSTM storm
LSTM quiet
SAMI2
IRI-2016

Day of year_2017

(c)

Bz (nT)

AE (nT)

Kp index

Dst (nT)

Day of year_2017

Figure 6.

Training Event #1_2010

P/N = 3.27

P/N = 1.31

Ionosonde

Reference

foF2 (MHz)

hmF2 (km)

Kp index

DOY

Figure 7.

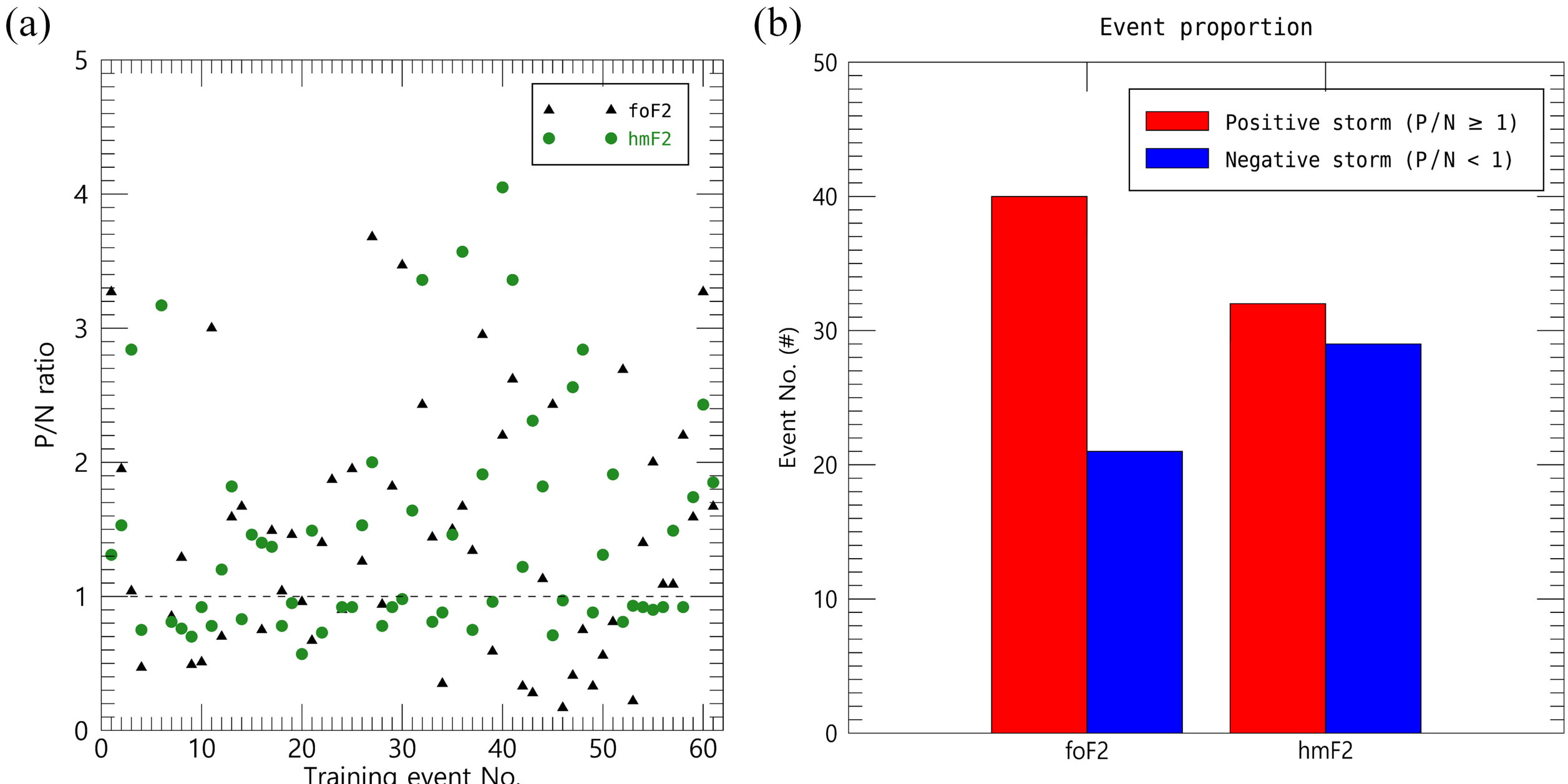
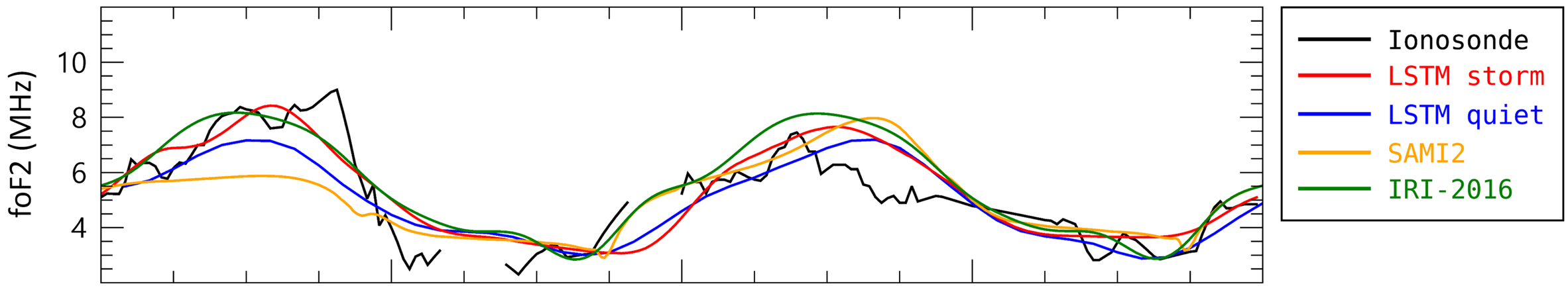

(a)
P/N ratio
Training event No.
foF2
hmF2
0
1
2
3
4
5
0
10
20
30
40
50
60
(b)
Event proportion
Event No. (#)
Positive storm (P/N ≥ 1)
Negative storm (P/N < 1)
0
10
20
30
40
50
foF2
hmF2

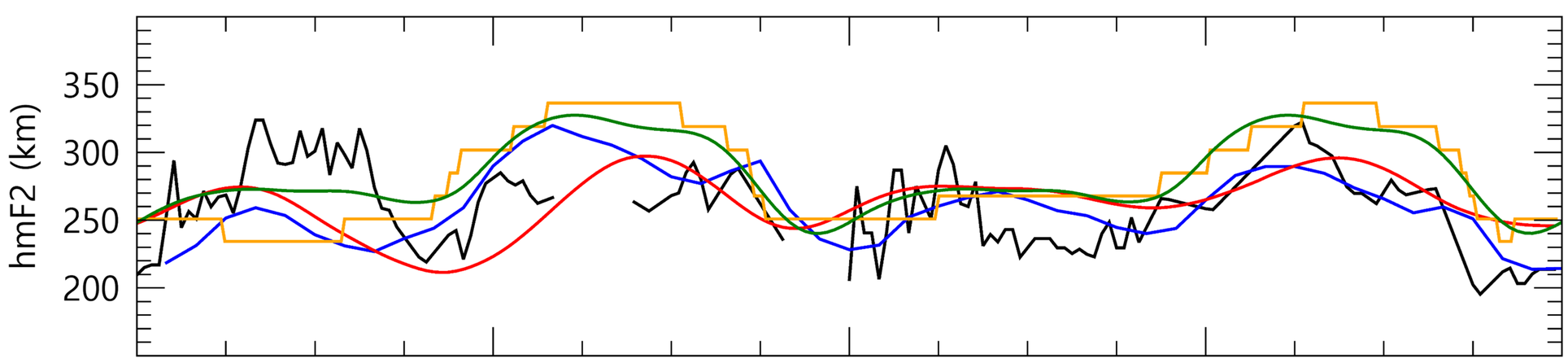
Figure 8.

Test #2 storm case results

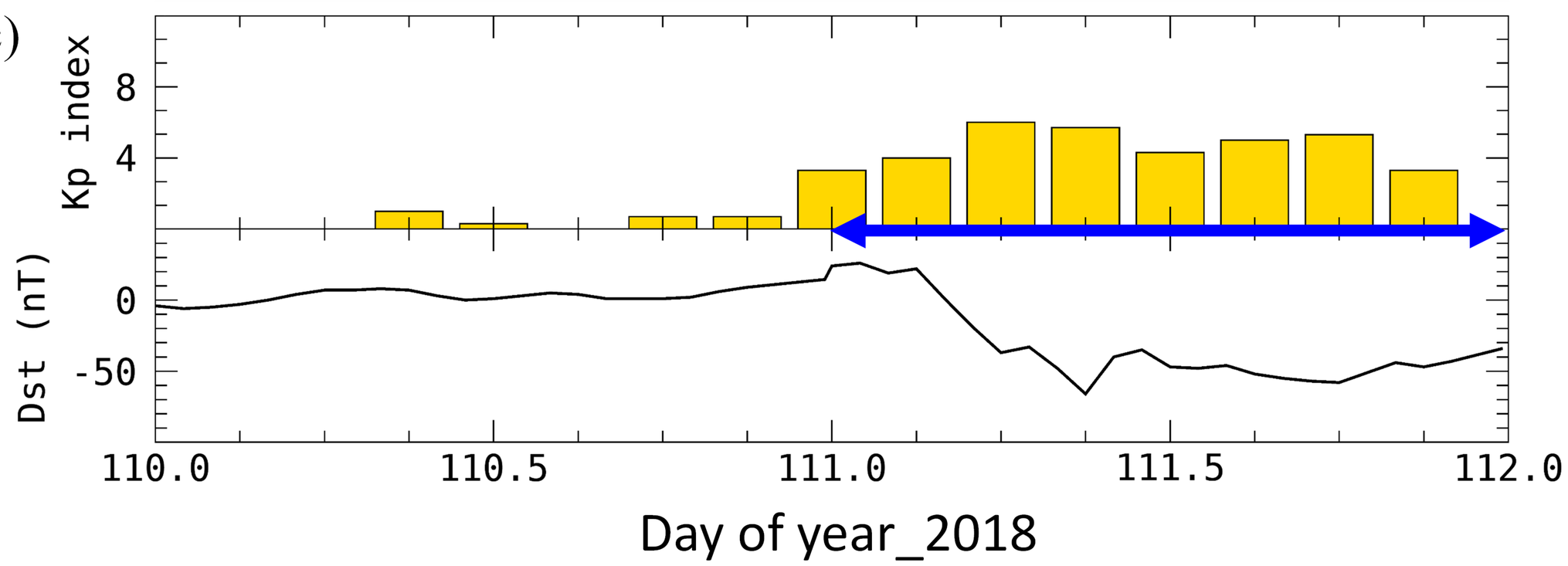
(a)



(b)

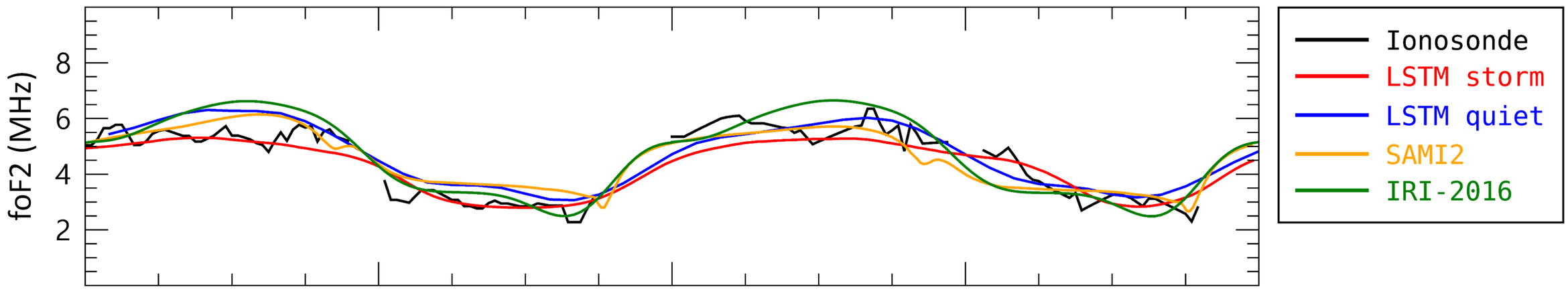


(c)

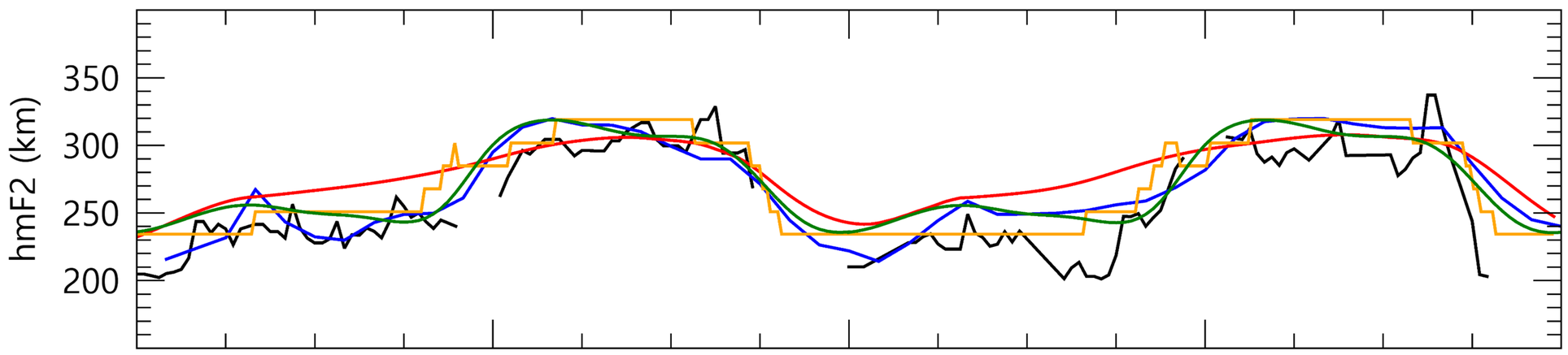


Test #3 storm case results

(d)



(e)



(f)

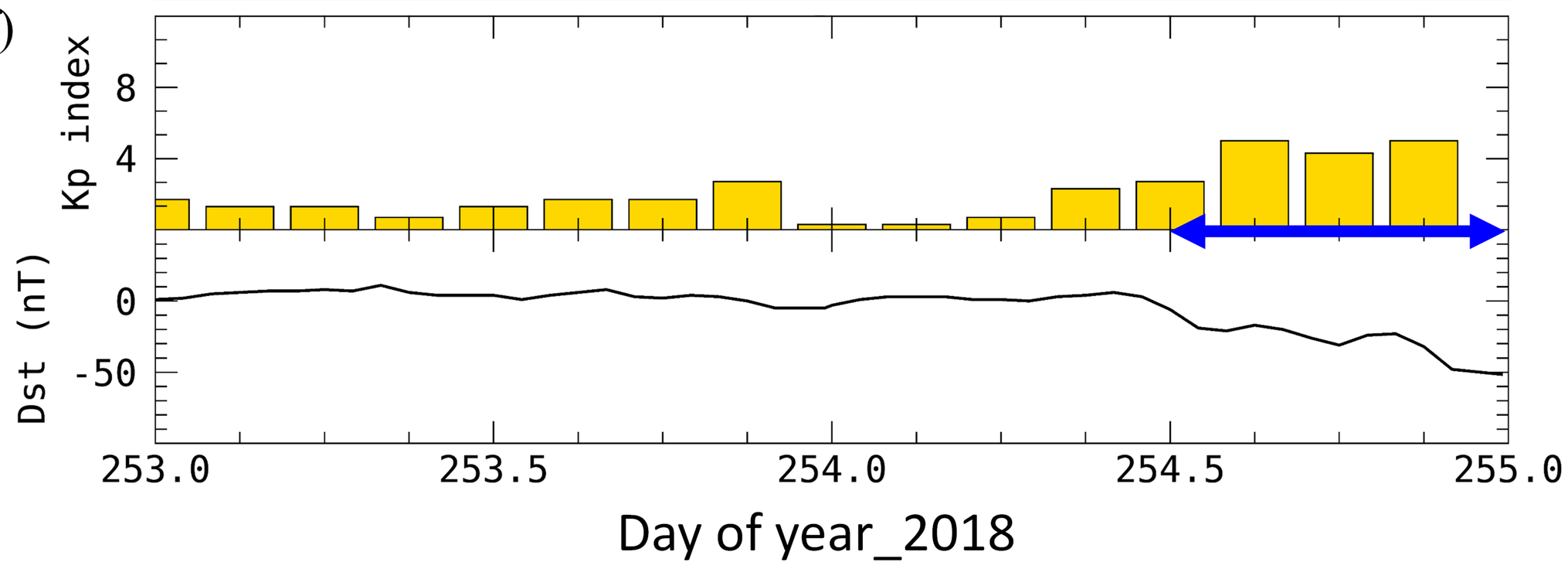
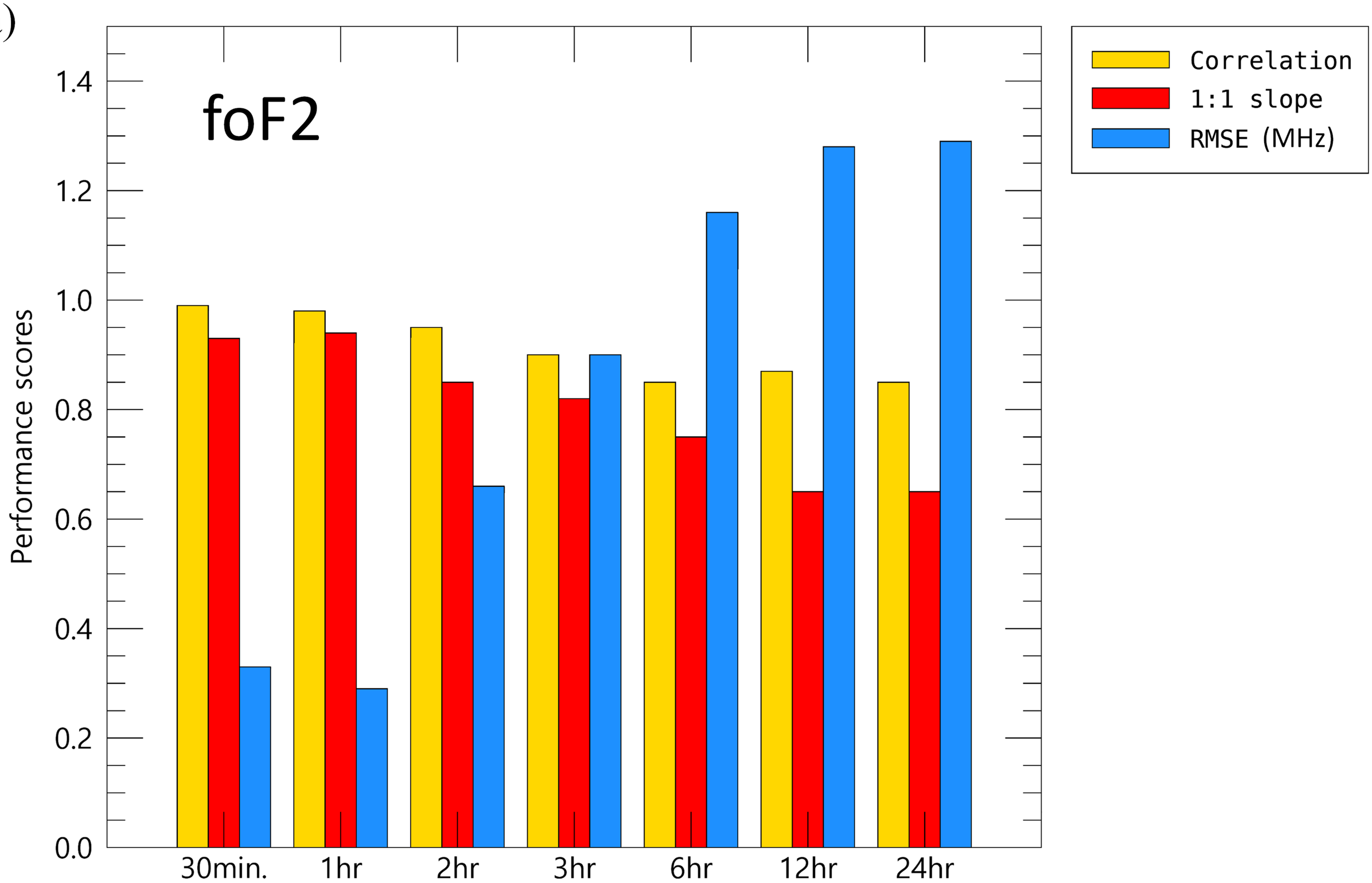


Figure 9.

(a)



Each prediction model

Short-term Long-term

(b)