Global ionospheric sporadic E intensity prediction from GNSS RO using a novel stacking machine learning method incorporated with physical observations

Tianyang Hu[1], Xiaohua Xu[1,2], Jia Luo[1,3], Jialiang Hou[1], Haifeng Liu[4]

[1]School of Geodesy and Geomatics, Wuhan University, Wuhan, 430079, China
[2]Collaborative Innovation Center for Geospatial Technology, Wuhan, 430079, China
[3]Key Laboratory of Geospace Environment and Geodesy, Ministry of Education, Wuhan, 430079, China
[4]School of Surveying and Geoinformation Engineering, East China University of Technology, Nanchang, 330013, China

Correspondence to: Xiaohua Xu (xhxu@sgg.whu.edu.cn)

Abstract. Sporadic E (Es) layers, the irregularities of enhanced electron density commonly occurring in ionospheric E region, are affected by the interactions between distinct atmospheric layers. Es intensity (EsI) is a crucial parameter to describe Es layer characteristics, while there still lacks the method for high-precision EsI prediction due to its complex spatiotemporal variation and physical driving mechanisms. We propose a novel stacking machine learning (SML) method for global EsI prediction, which combines the advantages from different ML models to obtain better performance than by using a single ML model. Various Es-related physical observations, including vertical ion convergence, gravity wave potential energy, solar and geomagnetic indices are incorporated as the inputs of SML together with the EsI derived from global navigation satellite system (GNSS) radio occultation (RO) measurements. SML performs well in both long-term and short-term EsI prediction and characteristics reconstruction. SML-predicted EsI is in good agreement with GNSS RO-derived EsI, with the mean error (ME) of 0.032 TECU km^{-1} and root mean square error (RMSE) of 0.158 TECU km^{-1}. Taking ionosonde observations as reference, SML has the RMSE of 1.064 MHz, which is reduced by 20.1%–40.5% compared to existing prediction methods. The higher accuracy of our method than those not incorporating physical observations illustrates the significance of considering multiple related physical factors when constructing the Es prediction model. The proposed method can be expected to provide valuable information for not only ionospheric irregularities monitoring and space weather forecasting, but also the mechanisms of Es layer formation and atmospheric coupling.

1 Introduction

Ionospheric sporadic E (Es) layers are thin-layer structures with abnormally sharp enhanced densities of electrons and metal ions, occurring frequently in ionospheric E region with a major altitude range of 90–130 km. Existing studies show that the occurrence of Es layers is driven by various physical mechanisms in the lower atmosphere, mesosphere-lower thermosphere (MLT), ionosphere, and space environment, such as the neutral wind shear (Chu et al., 2014), upward propagating gravity waves (GWs) (Qiu et al., 2023), atmospheric tides (Tang et al., 2022a), electric fields (Resende et al., 2016, 2021), solar

activity (Yu et al., 2021), and geomagnetic field (Luo et al., 2021a; Moro et al., 2022), resulting in its highly uncertain and irregular spatial and temporal characteristics. Specifically, the neutral wind shear theory is widely accepted for the Es layer formation in mid-latitudes, and the low latitude Es layers are well related to wind shear and equatorial electrojet plasma irregularities (Forbes, 1981; Raghavarao et al., 2002; Resende et al., 2013). In comparison, the particle precipitation during geomagnetic activity and upward propagation of gravity waves are more efficient in concentrating the ions of Es layers in high latitudes (Batista and Abdu, 1977; Kirkwood and Nilsson, 2000; MacDougall et al., 2000a, b). Es layers may cause scintillations on the propagations of radio signals and severely affect the radio communication and satellite navigation. Therefore, the high-precision modeling and prediction of Es layers is not only crucial for ionospheric irregularities monitoring and space weather forecasting, but also provides solutions for understanding the mechanisms of Es layer formation and coupling of distinct atmospheric layers, while it still remains a challenging task due to its complex patterns and influencing factors.

Traditionally, Es layers are detected by ground-based devices such as ionosonde and incoherent scatter radar (Leighton et al., 1962; Mathews, 1998; Heinselman et al., 2000). With the development of global navigation satellite system (GNSS) over the past two decades, the space-borne GNSS radio occultation (RO) missions have been widely applied to investigate the climatology of Es occurrence rates and Es intensity (EsI) in recent years (Arras et al., 2008; Chu et al., 2014; Yu et al., 2019; Xu et al., 2022; Liu et al., 2024a). Most RO missions consist of one or several low earth orbit (LEO) satellites, which can acquire all-weather and wide-area ionospheric observations with high vertical resolution. Based on the large amount of accumulated RO data, some empirical models were established using statistical method to describe the global EsI distribution (Hu et al., 2022; Yu et al., 2022; Niu and Fang, 2023). However, since the high uncertainty and irregularity of EsI, these empirical models are difficult to predict the localized and short-term Es variations. Besides, these numerically established models are not incorporated with physical observations such as wind shears, GWs, solar flux and geomagnetic indices, thus cannot estimate EsI based on the Es-related physical mechanisms. To overcome this limitation, some scholars compared the morphologies of RO-derived EsI and physical data to analyze their relationships (Qiu et al., 2019; Yu et al., 2019; Yamazaki et al., 2022), but these analyses were only qualitative and did not quantitatively reveal their correlations.

Recently, the advances in artificial intelligence algorithms provide new perspectives for data analysis in geosciences and many other fields. Machine learning (ML) is a powerful tool fitting complex nonlinear relationships between multiple variables, which has been employed to resolve the regression and classification problems in geoscience and is proven to have better performance compared with traditional methods. Currently, there are few studies on EsI prediction based on ML methods. Although Emmons et al. (2023) and Tian et al. (2023) used ML models for the detection and reconstruction of Es layers, there were little data on Es-related physical mechanisms used for model training in their studies. The incorporation of data about the background neutral wind, GWs, solar and geomagnetic activity should contribute to better performance of Es layer reconstruction and prediction. In addition, although a large number of ML models have been utilized in ionospheric predictions, the performance of single models is limited by their respective shortcomings. For example, neural network (NN)

performs well in learning complex patterns from large quantities of available data, but often tends to easily overfit in the analysis of limited datasets (Hinton et al., 2012). Besides, due to the "black box" nature of NN, it is difficult to investigate potential relationships between inputs and inner structure of NN, so that NN outputs are often lacking of interpretability (Hastie et al., 2009). Comparatively, the bagging and boosting ML models like random forest (RF) and gradient boosting decision tree (GBDT) tend to show better performance and interpretability than NN on some small datasets (Zhukov et al., 2021; Han et al., 2022). To bridge this gap, the stacking machine learning (SML) method is implemented to obtain better accuracy and generalization than a single model. SML uses the stacking strategy (Wolpert, 1992), which combines several base models and incorporates their outputs into a meta model, to leverage the efficiency of different models together, thus the possible errors of a single model can be complemented by other models. Compared with other ensemble method like bagging and boosting, it has better ability to reduce both variance and bias. SML has been successfully applied to solve some ionospheric predictive problems. For example, Asamoah et al. (2024) proposed a stacked model combining three ML models to predict total electron content (TEC) over a single station, and demonstrated its better performance than the single models. Liu et al. (2024b) utilized a hybrid ensemble model to forecast ionospheric irregularities over Brazilian sector. However, to our knowledge, SML method has not been applied for EsI prediction till now.

Hence, in this article, we present an SML model for global ionospheric EsI prediction, where a variety of observations representing Es-related physical mechanisms are used as inputs of our model, including the vertical ion convergence (VIC) driven by neutral wind shear, the GW activity, and the solar and geomagnetic indices. They are incorporated together with EsI derived from GNSS RO measurements into the proposed SML model. SML method selects five widely adopted ML models, RF, light gradient boosting machine (LightGBM), eXtreme Gradient Boosting (XGBoost), support vector machine (SVM), and back propagation neural network (BPNN) as the base models, and a multilayer perceptron neural network (MLPNN) is utilized as the meta model, i.e., the second part of SML model to optimally integrate the predictions generated by base models and generate the final prediction. The prediction performance of SML is validated using RO-derived EsI and ionosonde data under different space weather conditions and is also compared with other prediction methods.

2 Data and Materials

2.1 GNSS RO-derived EsI

Constellation Observing System for the Meteorology, Ionosphere, and Climate (COSMIC) is a global RO mission launched in April 2006, with the goal of providing GNSS RO data for operational weather prediction, climate analysis, and space weather forecasting (Syndergaard et al., 2006). COSMIC consists of six LEO satellites with orbital altitude of 800 km and inclination of 72 °, and provides more than 2000 globally distributed TEC profiles per day during its full operation stage. COSMIC TEC profiles are the profiles of the calibrated TEC below COSMIC LEO satellite altitude, and they have a high vertical resolution of better than 2 km, thus are suitable for the detection of Es layers which are with small vertical scales. In

this study, COSMIC TEC profiles during 2006–2019 are collected as the data source for deriving EsI. Since the altitude range of Es layers considered in this paper is 90–130 km, a quality control process is taken at first to remove the profiles with negative TEC values and the bottom heights higher than 90 km, and 93.45% of raw TEC profiles passed quality control and are retained for further analysis. Fig. 1 shows the traces of qualified COSMIC TEC profiles on Aug 1, 2007. It indicates that COSMIC TEC data has a dense global coverage that is a great data source for global ionospheric investigations.

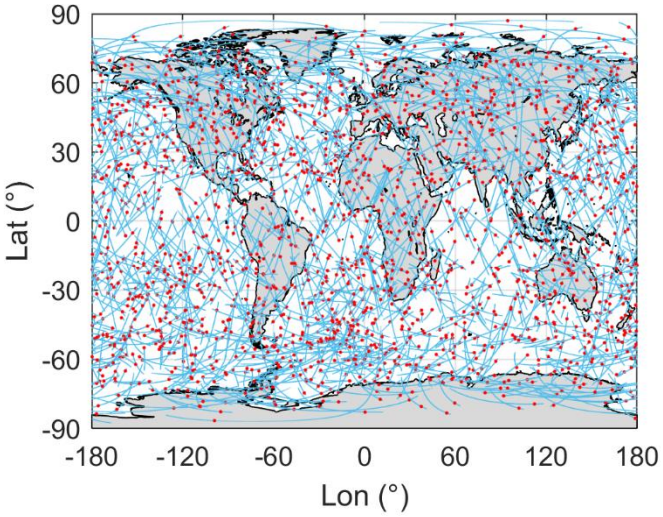


Figure 1: The traces of qualified COSMIC TEC profiles on Aug 1, 2007 (blue lines). The red points represent the locations corresponding to Smax.

We use single spectrum analysis (SSA) method as described in Hu et al. (2022) to obtain Smax index (unit: TECU km^{-1}), a proxy for EsI, from qualified TEC profiles. The TEC disturbances caused by Es layers are extracted from original TEC profiles using SSA. Smax is defined as the maximum vertical gradient of TEC disturbances in the altitude range of 90–130 km, and the corresponding altitude is designated as the altitude of Es layer. The reasonableness and effectiveness of Smax as the proxy of EsI have been verified by Niu et al. (2019) and Hu et al. (2022).

2.2 Ionosonde EsI

Es critical frequency (foEs) is a conventional parameter characterizing EsI. Ionosondes provide reliable ground-based local foEs observations. In this study, foEs data downloaded from National Earth System Science Data Center of China (NESSDC) and UK Solar System Data Center (UKSSDC) are used to validate the prediction results of SML. The foEs observations at 1-hour intervals, all manually scaled, are obtained from 15 ionosondes. Fig. 2 shows the spatial distribution of ionosonde stations.

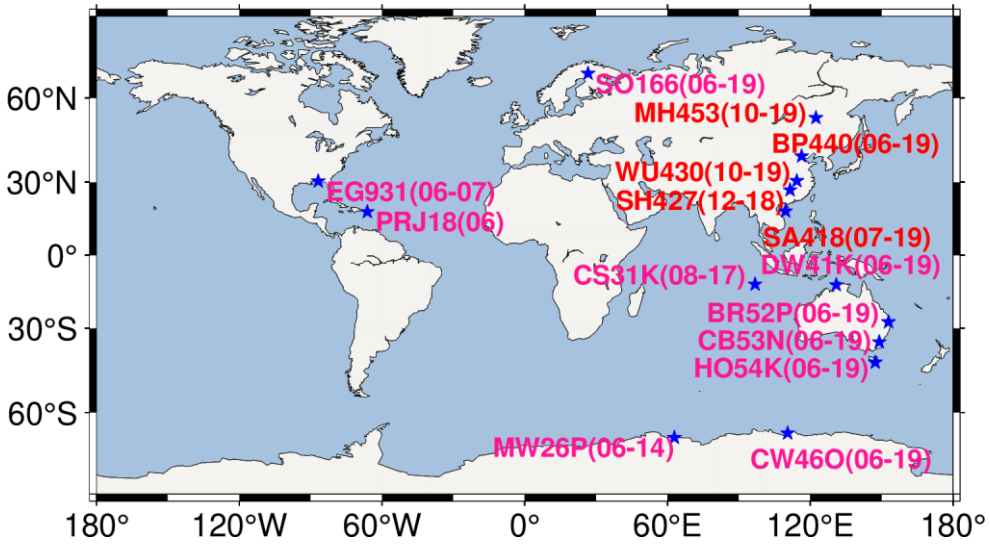


Figure 2: Spatial distribution of 15 ionosonde stations. The red and pink stations represent those from NESSDC and UKSSDC, respectively. In the bracket behind each station name, the corresponding time period of data is presented in the format of yy-yy.

2.3 VIC simulated by HWM14

The wind shear theory has been proven to be the most significant factor influencing the formation of mid-latitude Es layers (Chu et al., 2014; Luo et al., 2021a). In the procedure of VIC driven by vertical wind shears in horizontal neutral winds, the metal ions are compressed into a thin layer by Lorentz force, and then the electrons drift along the magnetic field lines and converge to form an Es layer (Mathews 1998). We employ Horizontal Wind Model 2014 (HWM14) (Drob et al., 2015) to simulate VIC at the location of Es layer. The vertical ion drift velocity w caused by the horizontal neutral wind shear can be written as follows:

$$w = \frac{r\cos I}{1+r^2}U + \frac{\cos I \sin I}{1+r^2}V \tag{1}$$

where U and V are the zonal and meridional velocities of horizontal neutral wind, respectively, which are calculated by HWM14; I is the geomagnetic inclination angle; $r = v_i/w_i$ is the ratio of the ion-neutral collision frequency to ion gyrofrequency (Nygrén et al. 2008), $w_i = eB/M$, e and M are the mass and charge of ion, respectively, and B is the magnetic field strength. Qiu et al. (2019) and Yu et al. (2019) indicated that Es layers tend to appear in regions with negative vertical zonal wind shears ($\partial w/\partial z < 0$, where z is the altitude). Therefore, it is assumed that only positive values of VIC contribute to Es layer formation:

$$VIC = \begin{cases} -\frac{\partial w}{\partial z}, & -\frac{\partial w}{\partial z} > 0 \\ 0, & -\frac{\partial w}{\partial z} \le 0 \end{cases} \quad (2)$$

2.4 GW activity extracted from GNSS RO data

Recent studies have shown that the upward propagating GWs can transport energy and momentum from lower atmosphere to the mesosphere and lower thermosphere, causing ionospheric disturbances and contributing to the formation of Es layers (Qiu et al., 2023; Seid et al., 2023). COSMIC temperature profiles are able to cover a wide altitude range of 0–60 km with the vertical resolution better than 0.1 km, which are ideal data sources for studying GW activity. The proxy for GW activities, GW potential energy (Ep), can be calculated by following equations:

$$E_p = \frac{1}{2}\frac{g^2}{N^2}\left(\frac{T'}{\bar{T}}\right)^2 \quad (3)$$

$$N^2 = \frac{g}{\bar{T}}\left(\frac{\partial \bar{T}}{\partial z} + \frac{g}{c_p}\right) \quad (4)$$

where g is the gravitational acceleration, N is the buoyancy frequency, $\bar{T}$ is the background temperature, $T' = T - \bar{T}$ is the temperature perturbation caused by GWs, and $c_p = 1004.5$ J $(\mathrm{kg \cdot K})^{-1}$ is the isobaric heating capacity. To derive background temperature $\bar{T}$ that represents longer waves like tides and planetary waves, global daily temperature profiles are binned into $10° \times 15°$ latitude-longitude grids with an altitude interval of 0.1 km to obtain daily mean temperature maps at each altitude level, and S-transform is applied on each zonal component of mean temperature maps to derive gridded data of $\bar{T}$. The temperature perturbation T' is obtained by subtracting $\bar{T}$ from original temperature profiles, and then GW Ep is calculated using Eq. (3) and (4). The detailed procedure for extracting GW Ep is described in Luo et al. (2021b).

2.5 Solar and geomagnetic indices

EsI is also affected by solar and geomagnetic conditions (Yu et al., 2021; Tang et al., 2022b). The solar radiation flux F10.7 and Dst indices are used to represent the solar and geomagnetic activity. We choose Dst index rather than 3-h Kp or Ap indices because Dst has a higher time resolution of 1 h. Fig. 3 shows the variations of F10.7 and Dst indices during 2006–2019.

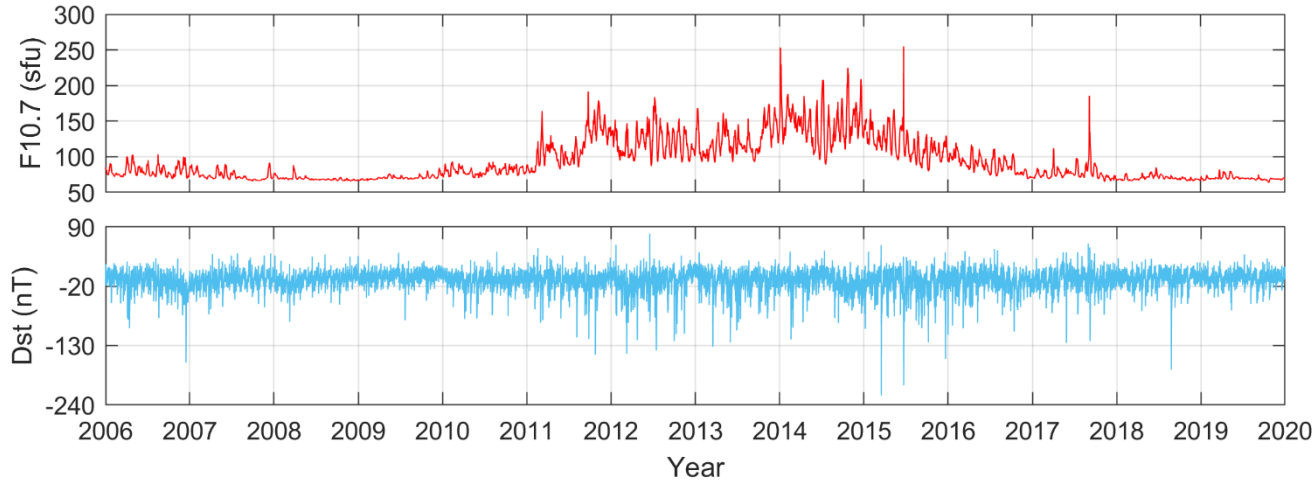


Figure 3: Variations of F10.7 and Dst indices during 2006–2019.

3 Methodology

3.1 Accuracy evaluation metrics

The mean error (ME), root mean square error (RMSE), and correlation coefficient (CC) are used as metrics to evaluate the accuracy of prediction results, which are calculated as:

$$\begin{cases} ME = \dfrac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right) \\ RMSE = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \\ CC = \dfrac{\mathrm{cov}\left(y_i, \hat{y}_i\right)}{\sigma_{y_i} \cdot \sigma_{\hat{y}_i}} \end{cases} \tag{5}$$

where n is the total number of prediction results; y_i and $\hat{y}_i$ are the predicted and observed EsI, respectively; $\mathrm{cov}\left(y_i, \hat{y}_i\right)$ is the covariance between y_i and $\hat{y}_i$; σ_{y_i} and $\sigma_{\hat{y}_i}$ are the standard deviations of y_i and $\hat{y}_i$, respectively. The units of both ME and RMSE are TECU km^{-1}. CC has no unit.

3.2 Dataset configuration and segmentation

The proposed EsI prediction method aims to build a nonlinear functional model between the target (EsI) and inputs (spatiotemporal information and physical observations). Therefore, the time, latitude, longitude, and altitude corresponding

to each RO-derived EsI (Smax), as well as the VIC, GW Ep, F10.7, and Dst, are formed into samples and fed into SML. To reduce the input feature complexity and modeling costs, the time of each sample is expressed as follows:

$$Time = Year + (Doy + UT / 24) / 365.25 \quad (6)$$

where Doy and UT are day of year and universal time, respectively.

In SML method, the training and validation set is used for the training of the base models, and their prediction values become the training and validation set of the meta model. Since the cross-validation (CV) strategy is utilized for optimization of SML model (see Sect. 3.3.2), samples of the entire dataset collected during 2006–2019 are divided into two groups: training and validation set (80%, from 22 April 2006 to 31 December 2013), and testing set (20%, from 01 January 2014 to 31 December 2019). Note that there are fewer samples after 2014 due to the decline in the number of measurements caused by the aging and loss of COSMIC satellites. Nevertheless, it in turn allows for a longer time period of testing set and a more comprehensive evaluation of SML performance.

3.3 SML model development

3.3.1 ML models

SML combines the advantages from different ML models to obtain better performance than a single ML model. Diverse types of ML models should be selected to make SML fully incorporate their strengths, which require the selection of appropriate base models to simultaneously reduce the bias and variance. In this study, five ML models are utilized as base models, including RF, LightGBM, XGBoost, SVM, and BPNN. RF is a tree-based parallel ensemble ML algorithm using the bagging technique that is widely applied for classification and regression problems in GNSS and remote sensing tasks. RF is effective in reducing the variance of the model, and has an improved robustness to outliers. Comparatively, LightGBM and XGBoost are sequential ensemble ML models based on the boosting technique. They use the gradient boosting technique with outstanding performance in reducing bias of numerous datasets. SVM is an ML method based on the principle of structural risk minimization. It utilizes the structural risk minimization theory to suppress the overfitting problem and minimize empirical risk and confidence interval. BPNN is a widely used NN model with high adaptability and learning ability for regression problems. It comprises three types of fully connected layers, i.e., the input layer, hidden layer(s), and output layer, which help to better capture complex nonlinear relationships.

Furthermore, MLPNN is also a common NN model to solve regression problems. It has a similar structure with BPNN, while the main difference between MLPNN and BPNN lies in their activation functions. Here we use MLPNN as the meta model to find the optimal combination of base models, and the structures of the base models and meta model are shown in Fig. 4.

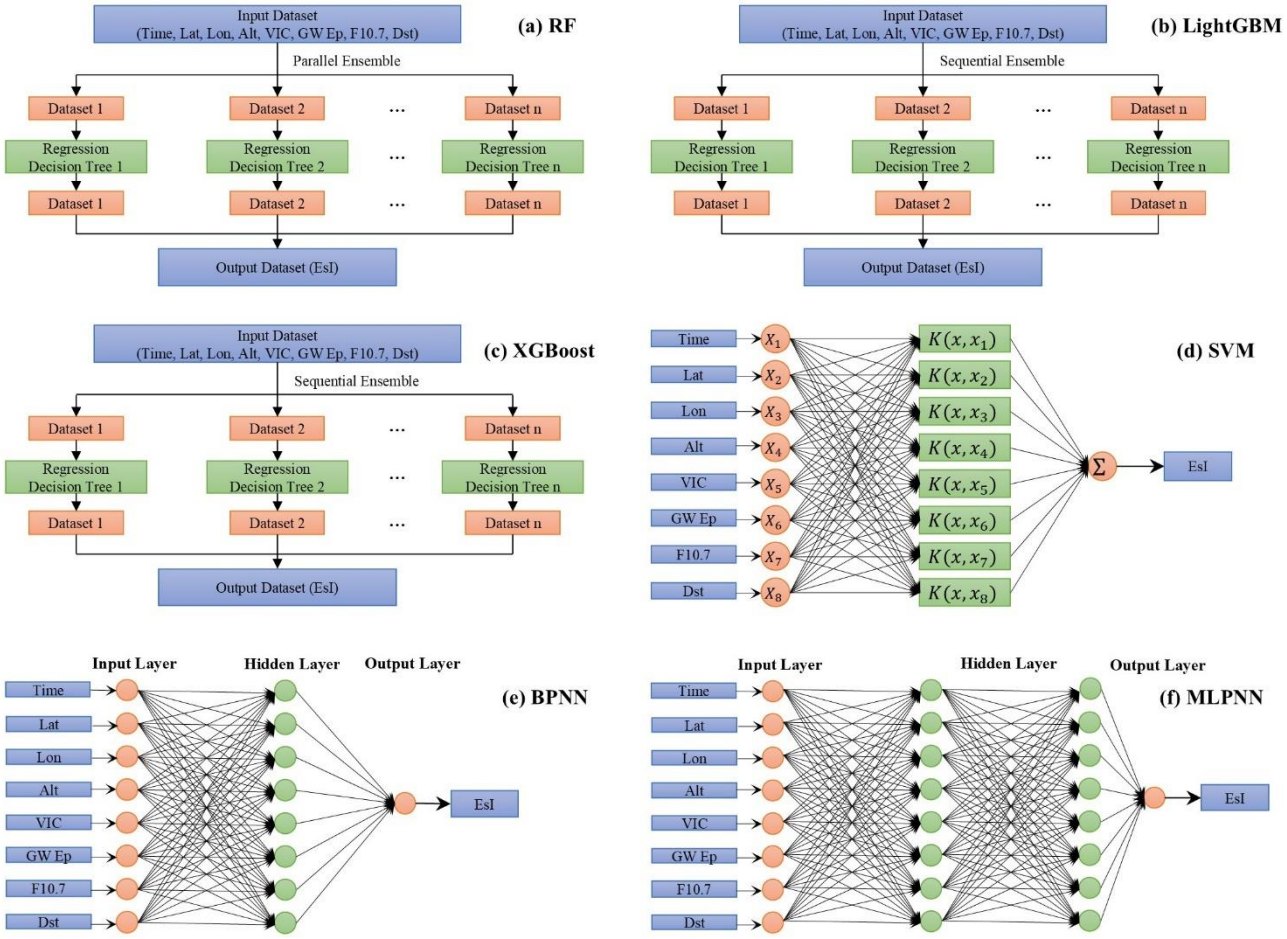


Figure 4: The structures of RF, LightGBM, XGBoost, SVM, BPNN, and MLPNN.

The mathematical expressions for all the ML models used are presented in Eq. (7)–(12):

$$RF(x) = \frac{1}{M}\sum_{m=1}^{M} T_m(x) \tag{7}$$

where x is the inputs, M is the number of trees, and $T_m(x)$ is the mth tree output.

$$LightGBM(x) = \frac{1}{M}\sum_{m=1}^{M} T_m(x) W_m \tag{8}$$

where W_m is the weight of the mth tree. LightGBM is similar to GBDT, but compared with the depth-wise tree growth approach, it grows trees using the leaf-wise approach that focuses on nodes with the highest loss change, which is better at handling large datasets and improving prediction accuracy.

$$XGBoost(x)=\frac{1}{M}\sum_{m=1}^{M}T_m(x)W_m \tag{9}$$

XGBoost is also similar to GBDT, while it offers a parallel tree boosting algorithm to improve computation efficiency. Actually, LightGBM and XGBoost are new optimized implementations for GBDT using different techniques.

$$\begin{aligned} &SVM(x)=w\varphi(x)+b \\ &s.t.\quad \min\left[\frac{1}{2}\|w\|^2+C\sum_{i,j=1}^{L}(\xi_i,\xi_j)\right] \end{aligned} \tag{10}$$

where w is the weight vector, φ is the nonlinear mapping function, b is the bias, L is the number of input samples, C is the penalty factor specifying the degree of penalty for outliers, and ξ_i and ξ_j are relaxation factors. Equation (9) can be solved by introducing Lagrange multipliers to obtain the regression function of SVM, in which φ is usually replaced by the radial basis function kernel $K(x_i,x_j)=\exp(-\gamma\|x_i-x_j\|^2)$, where γ is the kernel parameter. The details of SVM algorithm can be found in Yetilmezsoy (2019).

$$\begin{aligned} BPNN_k(x)&=f\left(\sum_{j=1}^{L}w_{kj}x_j+b_k\right) \\ MLPNN_k(x)&=g\left(\sum_{j=1}^{L}w_{kj}x_j+b_k\right) \end{aligned} \tag{11}$$

where $BPNN_k$ and $MLPNN_k$ are the outputs of the kth neuron, $f(\cdot)$ and $g(\cdot)$ are the activation functions of BPNN and MLPNN, respectively. We select the sigmoid and hyperbolic tangent functions as the activation functions of BPNN and MLPNN, respectively, which can be written as:

$$\begin{aligned} f(x)&=\frac{1}{1+e^{-x}} \\ g(x)&=\frac{2}{1+e^{-2x}}-1 \end{aligned} \tag{12}$$

3.3.2 Model optimization

Hyperparameters are the internal configuration parameters for ML models. The optimization of hyperparameters is important for improving the accuracy and generalizability of ML models. To determine the optimal hyperparameters while maintaining a relatively low computational cost, the grid search method is adopted to optimize the two hyperparameters with the greatest impact on the model. Specifically, for each model, a set of candidate values of the two hyperparameters to be optimized is defined in the parameter space, the model performance for each hyperparameter combination is evaluated, and the best performing hyperparameter combination is defined as the optimal hyperparameters. During the grid search process, a five-fold CV is utilized. The training and validation set is randomly divided into five non-overlapping folds. For each iteration of

training and evaluation, the *i*th fold ($i = 1, 2, \ldots, 5$) is used as the validation set, and the remaining folds are used as the training set. The average results of these iterations are denoted as the final performance evaluation.

Figure 5 shows the optimized hyperparameters, the candidate values, and the optimal values (denoted by red asterisks) of hyperparameters for each ML model. For RF (bagging model), the number of leaf nodes and the number of trees are selected as the hyperparameters to be optimized, and they are optimally determined as 200 and 200, respectively. While for LightGBM and XGBoost (boosting model), the number of leaves and the maximum depth play the more important roles in improving model performance, and they are optimally computed to 63 and 9, respectively. As shown in Eq. (9), the penalty factor C and the kernel parameter γ have significant impacts on the performance of SVM regressor. In this sense, they are selected for optimization with optimal values of 10 and 1, respectively. For BPNN and MLPNN, the number of hidden layer(s) and the neuron number in each hidden layer are key hyperparameters in determining the accuracy of the network. Since one hidden layer-based NN can approximate the arbitrarily small error among most bounded continuous functions (Hornik et al., 1989), we only use one hidden layer in BPNN, while two hidden layers are adopted in MLPNN to better combine the predictions from base models. The optimal neuron number in hidden layer(s) can be determined empirically based on the range from $2\sqrt{n} + \mu$ to $2n+1$, where n and μ are the neuron numbers in input and output layers, respectively. Therefore, the neuron numbers of BPNN and MLPNN are validated from 6 to 17 and from 4 to 11, respectively, and the optimal neuron numbers corresponding to the minimum RMSE are 16 for BPNN, 6 and 7 for MLPNN, respectively. By optimizing these ML models, their performance, generalizability and interpretability are improved so that they are more suitable for the specific task of EsI prediction in this study.

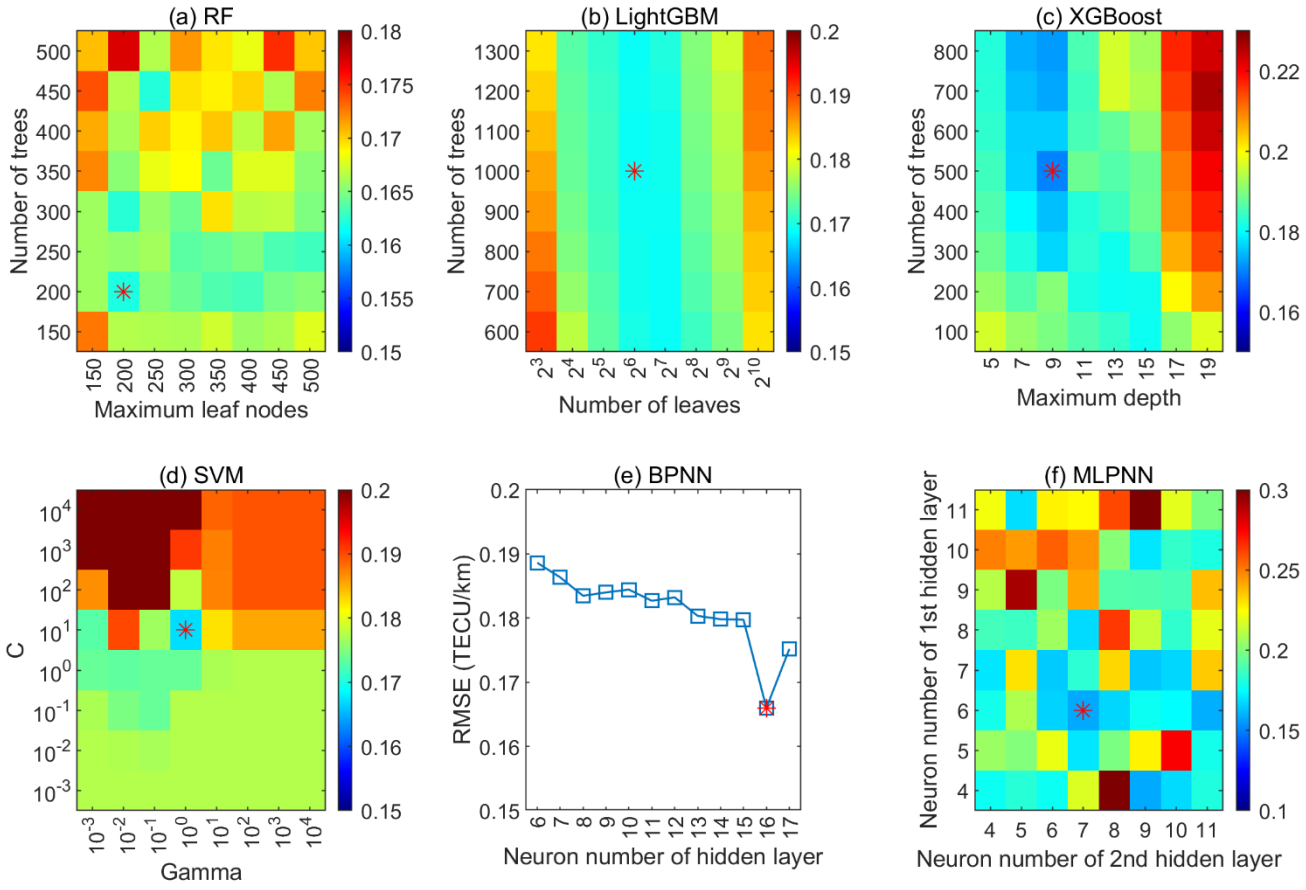


Figure 5: Grid search results (RMSE) for the optimization of RF, LightGBM, XGBoost, SVM, BPNN, and MLPNN. The red asterisks denote the optimal hyperparameter combination of each ML model.

3.3.3 SML model architecture

In the training stage of SML model, the outputs of RF, LightGBM, XGBoost, SVM, and BPNN on the training and validation set are fed into MLPNN as its input data, and its outputs are the final predicted EsI. This process is also similar in the test stage. The framework of SML is shown in Fig. 6.

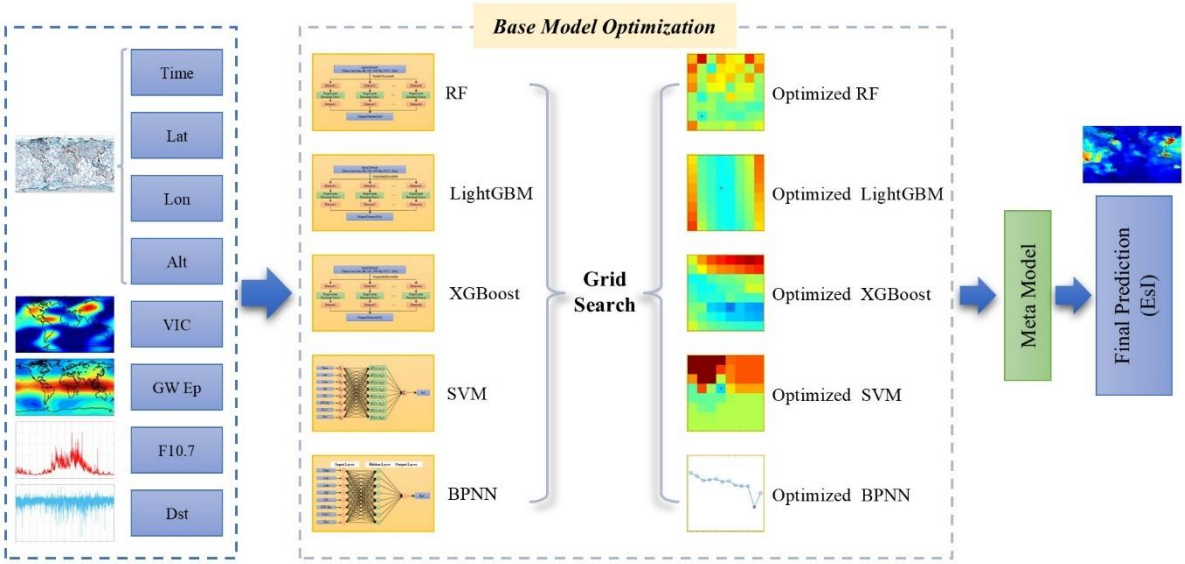


Figure 6: Framework of the SML model.

4 Results

4.1 Comparison of SML model and base models

Both the SML model and the base models with the optimal hyperparameters are fitted on the training set. Then they are employed to make predictions on the testing set, and the prediction results are compared with the ground truth. Fig. 7 illustrates the histograms and the density scatter plots of the comparisons of EsI predicted by SML and base models with ground truth. SML shows the much more aggregated histogram than those of base models, which means that SML has the best agreement with ground truth among all ML models, with the minimum ME/RMSE of 0.032/0.158 TECU km^{-1}. Compared to the maximum ME and RMSE of 0.053 and 0.170 TECU km^{-1} for the base models, SML has the improvement of 39.6% and 7.1%, respectively. The density scatter plots show that SML also has the highest CC of 0.891. As mentioned above, RF is more robust for outliers and more effective in reducing variances, thus has a lower RMSE than other base models. Comparatively, the other base models, especially BPNN, play an important role in reducing bias, and they outperform RF in terms of the overall prediction accuracy of EsI, as demonstrated by their MEs. By combining the strengths of different types of ML models, SML is able to achieve predictions with both lower biases and lower variances compared with all base models.

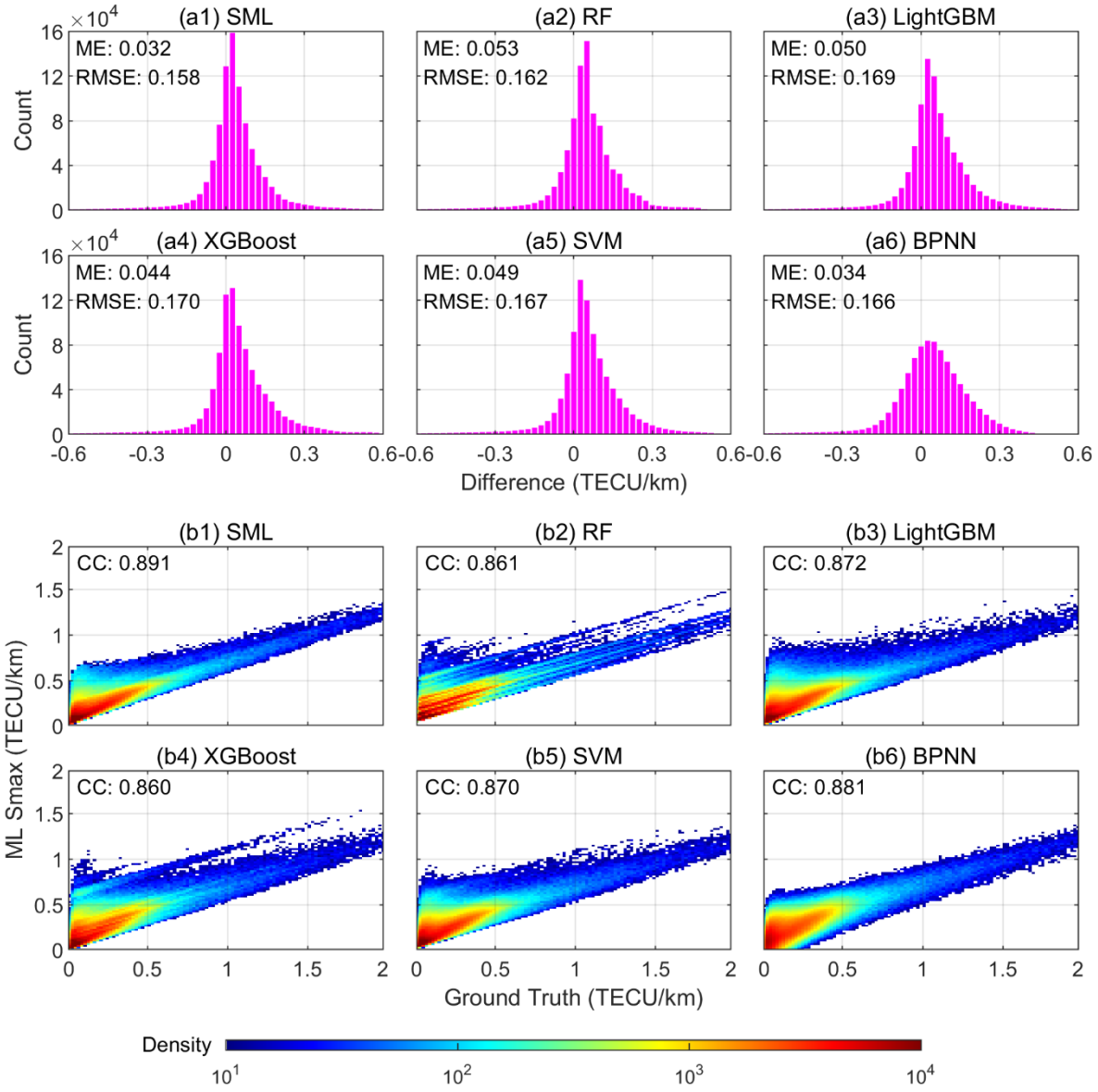


Figure 7: (a) Histograms and (b) density scatter plots of the comparisons of EsI predicted by SML and by base models with ground truth.

285 In addition to the overall performance, the global distributions of EsI predicted by the models are also compared. Figure 8 presents the latitude-longitude maps of the differences between ground truth and the EsI predicted by SML and by base models on the testing set. Here the EsI maps represent the average EsI in the latitude-longitude bin of 2.5 ° × 5 ° over the period of testing set. Specifically, RF has larger deviations than other models in North America (30 °N–50 °N and 80 °W–

120 °W), South American Magnetic Anomaly (SAMA) zone (30 °S–60 °S and 20 °W–40 °E), and near the geomagnetic equator (denoted by the red line in Fig. 8(a)), which are the regions with smaller EsI due to the near horizontal geomagnetic field lines (Yu et al., 2019; Luo et al., 2021a). RF also has a considerable underestimation of EsI in the Arctic region. The other base models, especially BPNN, have overall smaller biases than RF, while they have more outliers exhibited as localized small patches with larger prediction errors at mid- and low-latitudes. On the other hand, SML has the best prediction accuracies and the least number of outliers in the regions mentioned above, which is due to that SML avoids the shortcomings of different base models by selectively integrating their outputs.

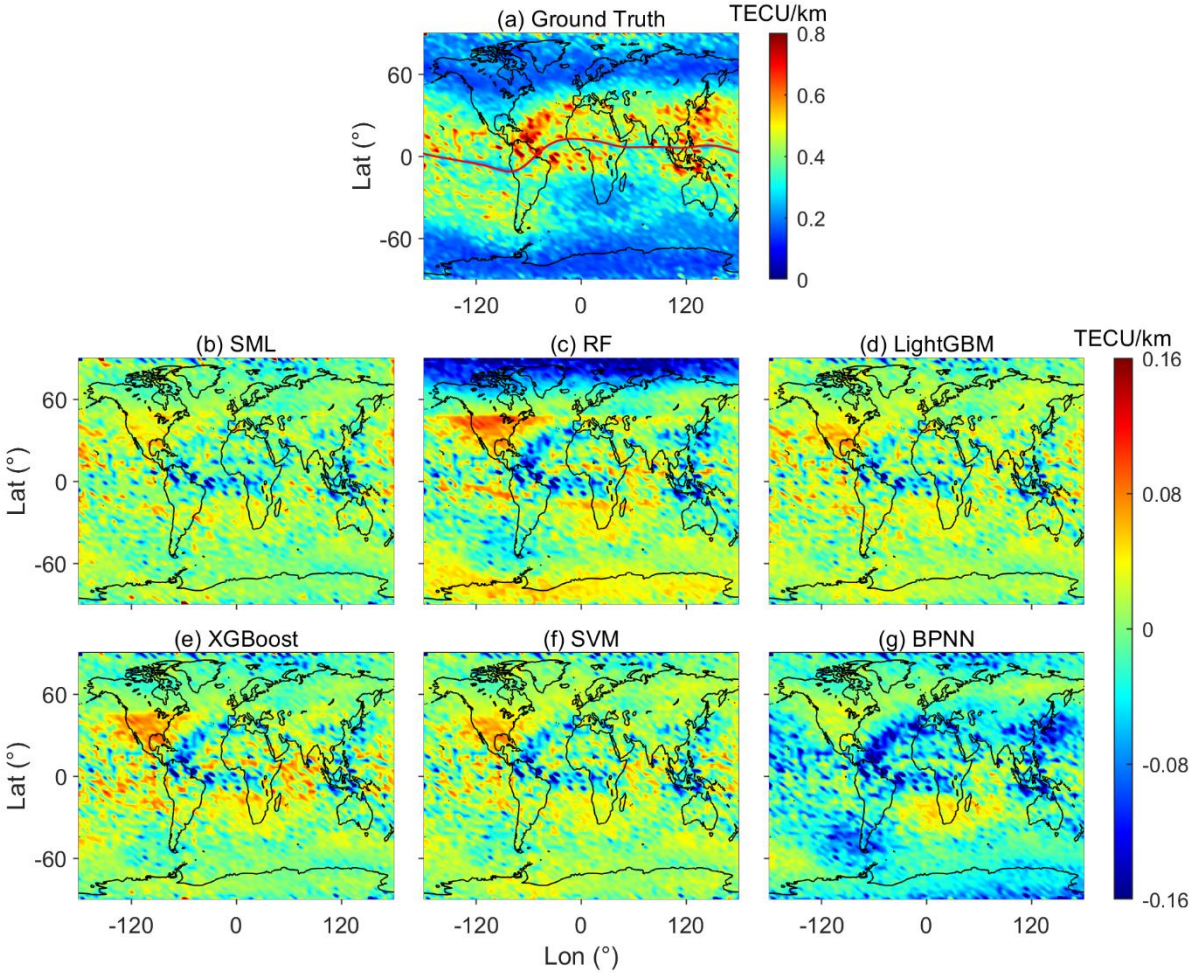
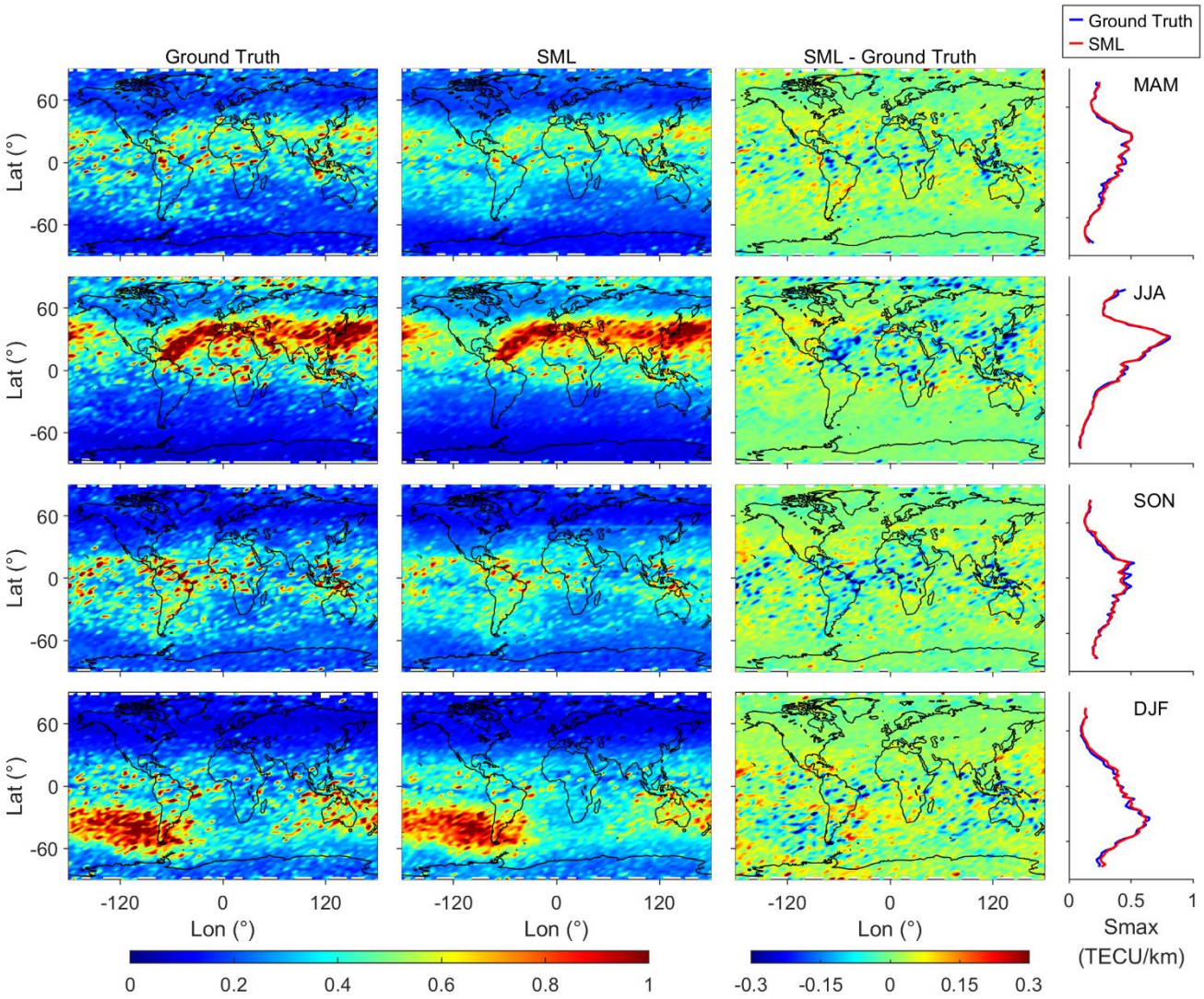


Figure 8: Latitude-longitude distribution of (a) ground truth, and the difference between ground truth and the EsI predicted by (b) SML, (c) RF, (d) LightGBM, (e) XGBoost, (f) SVM, and (g) BPNN on the testing set. The red line denotes the geomagnetic equator.

Based on the above comparisons, SML has the highest CC and the smallest ME and RMSE, i.e., the best prediction performance. In the following sections, only SML model is selected to assess the ability in reconstructing the complex long-term and short-term characteristics of EsI morphology, and we also compare the SML predictions with ionosonde observations for external validation of the prediction performance.

4.2 Long-term evaluation of SML performance

The long-term evaluation of SML prediction performance is conducted on the whole testing set, i.e., 2014–2019. Figure 9 presents the latitude-longitude and latitudinal distributions of ground truth, SML-predicted EsI, and the corresponding error maps in four seasons, which are categorized as MAM (March, April, and May), JJA (June, July, and August), SON (September, October, and November), and DJF (December, January, and February). Visual inspection shows that SML accurately simulate the seasonal variation of EsI. SML successfully shows the larger EsI that peaks in the banded area at mid-latitudes of summer hemisphere and reaches the valley values in winter hemisphere, which is primarily dominated by the seasonal variation of meteor flux and the resulting metallic ion content, coupled with the neutral wind shear (Haldoupis et al., 2007). The weaker EsI in North America, SAMA zone, and along the geomagnetic equator due to the lower geomagnetic inclination angle is well reconstructed by SML predictions. Furthermore, the SML-predicted latitudinal distribution of EsI also agrees well with ground truth in all the four seasons. The larger EsI moves northward or southward with seasonal variations, which is under the control of wind shear at mid-latitudes. While at latitudes higher than 70 °, there is also strong EsI which is larger than that at 60 °, and this is no longer due to the wind shear but the vertical transport of ions and electrons caused by GWs propagating upward along near vertical geomagnetic field lines (Kirkwood and Nilsson, 2000). It indicates that SML model can clearly reconstruct and predict the larger EsI in mid and high latitudes dominated by different physical mechanisms, and can comprehensively consider the impact of multiple influencing factors. The SML prediction errors are between ±0.1 TECU km^{-1} in most areas, demonstrating the excellent prediction performance in long-term EsI prediction. The ME of SML predictions for the four seasons are 0.004/-0.005/0.001/0.012 TECU km^{-1}, and RMSE are 0.146/0.166/0.143/0.176 TECU km^{-1}, respectively. Nevertheless, we can see from the error maps that the areas of underestimation beyond ±0.1 TECU km-1 are mainly concentrated at the peak of EsI in the summer hemisphere. The possible explanation for this phenomenon is that larger EsI is the minority of the training set (predominantly occurs in the summer hemisphere only), and the trained model fits this part of data slightly worse than smaller EsI. Improvement in the prediction performance for larger EsI should be required for future study.



330 **Figure 9: Latitude-longitude and latitudinal distributions of ground truth, SML-predicted EsI, and the corresponding error maps in four seasons.**

Figure 10 shows the local time (LT)-day of year (DOY) distributions of ground truth and SML-predicted EsI at different latitude ranges in 2014–2019. Larger EsI mainly exists in daytime, increasing after sunrise and decreasing after 335 sunset. The EsI tidal signatures reconstructed by SML, mainly dominated by the wind shear and atmospheric tides (Yu et al., 2019), is consistent with ground truth, which can be identified in summer days with diurnal tides (start around 10 LT) occurring at low-latitudes (30 °N–30 °S) and semi-diurnal tides (start around 8 LT and 16 LT) occurring at mid-latitudes (30 °N–60 °N and 30 °S–60 °S). Although the EsI seasonal variation in the Southern Hemisphere (SH) is opposite to that in the

Northern Hemisphere (NH), the diurnal and semi-diurnal tides can still be discerned, only with a slightly lower peak intensity. Figure 10 indicates the effectiveness of SML in reconstructing tidal signatures of EsI.

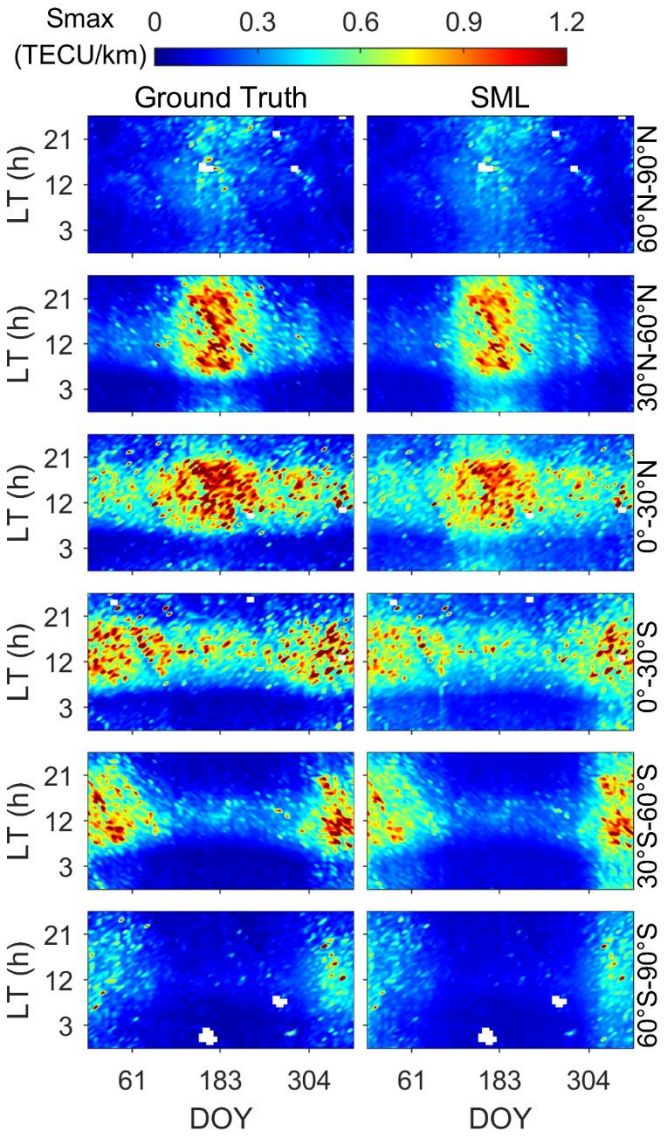


Figure 10: LT-DOY distributions of ground truth and SML-predicted EsI. The right label of the right panel of each row represents the latitude range of this row.

Figure 11 plots the latitudinal distribution of daily SML-predicted EsI and the daily RMSE on the whole testing set, with blank areas indicating the days without EsI data. The results present that the morphology characteristics of SML predictions are close to those of ground truth. SML succeeds in capturing the hemispheric asymmetry of EsI, i.e., EsI is

generally slightly higher in the NH summer than in the SH summer of the same year, which is also found by Luo et al. (2021a) and Xu et al. (2022). This is mainly due to the lower EsI in SAMA zone caused by the distribution of horizontal geomagnetic field, which diminishes the EsI over the corresponding latitude zones. Furthermore, the daily RMSE in Fig. 11 is generally stable below 0.2 TECU km^{-1}, with unusual sudden enhancements only in a few days. Overall, these results fully demonstrate the good ability of SML for stable EsI prediction and characteristics reconstruction in long-term periods.

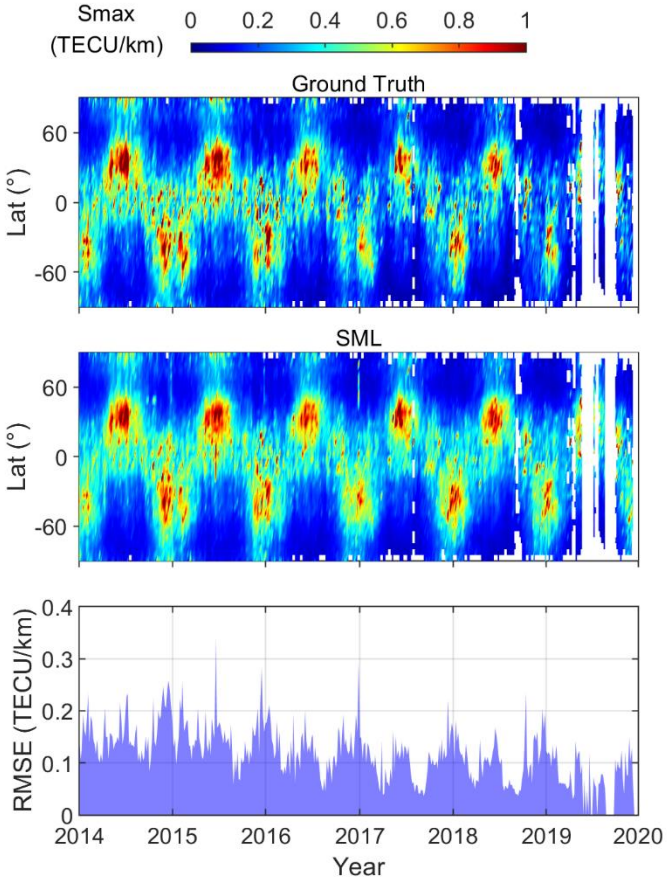


Figure 11: Latitudinal distribution of daily ground truth and SML-predicted EsI, as well as the daily RMSE on the testing set.

4.3 Short-term evaluation of SML performance

The response of EsI to geomagnetic storms has been widely reported (Resende et al., 2021; Moro et al., 2022; Tang et al., 2022b; Qiu and Liu, 2025), which is usually a combined effect of neutral wind and electric field variations. We have conducted two case studies to evaluate the short-term prediction performance of SML during the geomagnetic quiet and

storm time periods. Figure 12 shows the latitude-longitude and latitudinal distributions of ground truth, SML-predicted EsI, and the corresponding error maps on two quiet days, 04 Jul 2014 and 24 Jan 2018. Compared with ground truth, SML can effectively predict the general distribution of EsI, particularly the considerable agreement in latitudinal distribution, which is similar to that in Fig. 9. The prediction errors are mostly within ±0.1 TECU km^{-1}, with a small number of underestimates which are with larger errors mainly existing in the summer hemisphere. The ME/RMSE for the two quiet days are -0.006/0.183 TECU km^{-1} and 0.006/0.133 TECU km^{-1}, respectively.

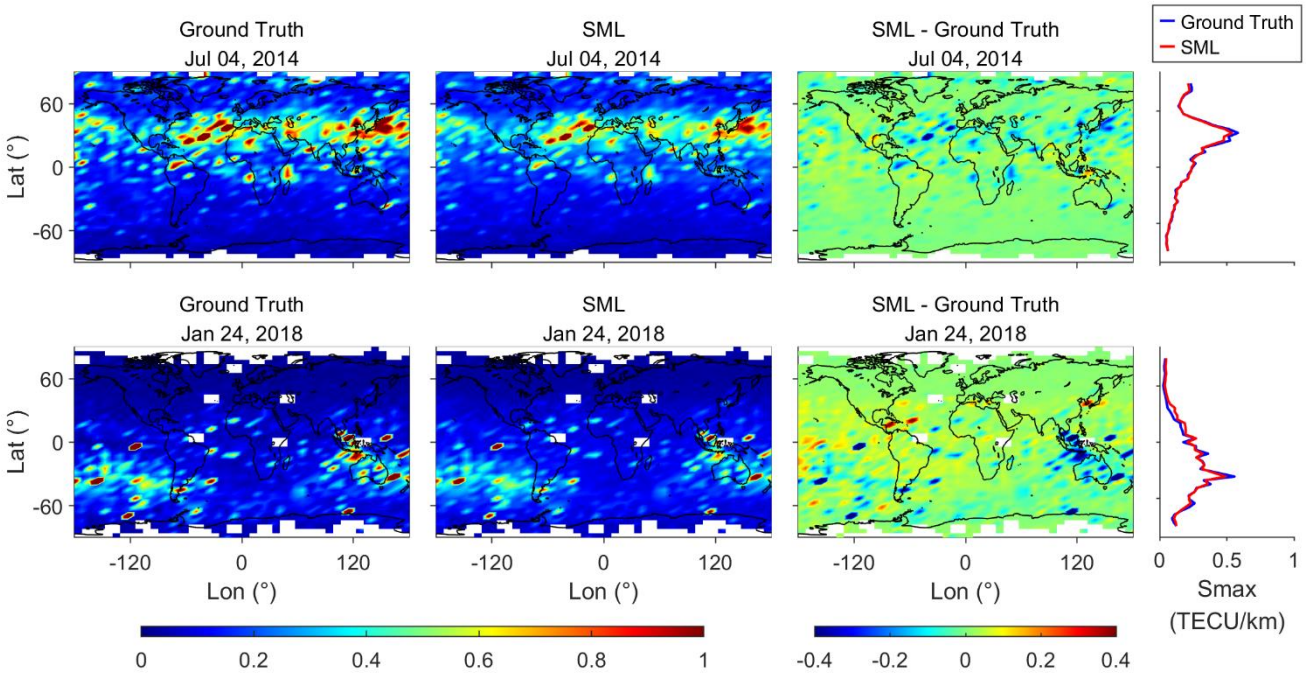


Figure 12: Latitude-longitude and latitudinal distributions of ground truth, SML-predicted EsI, and the corresponding error maps during geomagnetic quiet time on 04 Jul 2014 and 24 Jan 2018.

Figure 13 shows the latitude-longitude and latitudinal distributions of ground truth, SML-predicted EsI, and the corresponding error maps on two storm days, 07 Dec 2014 (moderate storm, Dst = -43 nT) and 22 Jun 2015 (major storm, Dst = -121 nT). Compared to quiet days, the EsI distributions during storm days are more complex, showing more irregular patches of EsI enhancement. The general distributions of SML predictions still agree well with ground truth, while there are more outliers in the summer hemisphere and at low latitudes compared to quiet days, as shown in the error maps. The ME/RMSE for the two storm days are 0.008/-0.004 TECU km^{-1} and 0.278/0.268 TECU km^{-1}, respectively. In addition, SML has more overestimations of EsI on 07 Dec 2014, while presents both overestimations and underestimations on 22 June 2015. Liu et al. (2022) reported that EsI usually has a decrease during moderate storms than during quiet time, and presents a

complex variation during major storms. Although their study is from a climatological perspective, it may explain our prediction results during geomagnetic storms. Nonetheless, Figs. 12 and 13 suggest that SML model has a reliable ability for short-term EsI prediction under different geomagnetic levels.

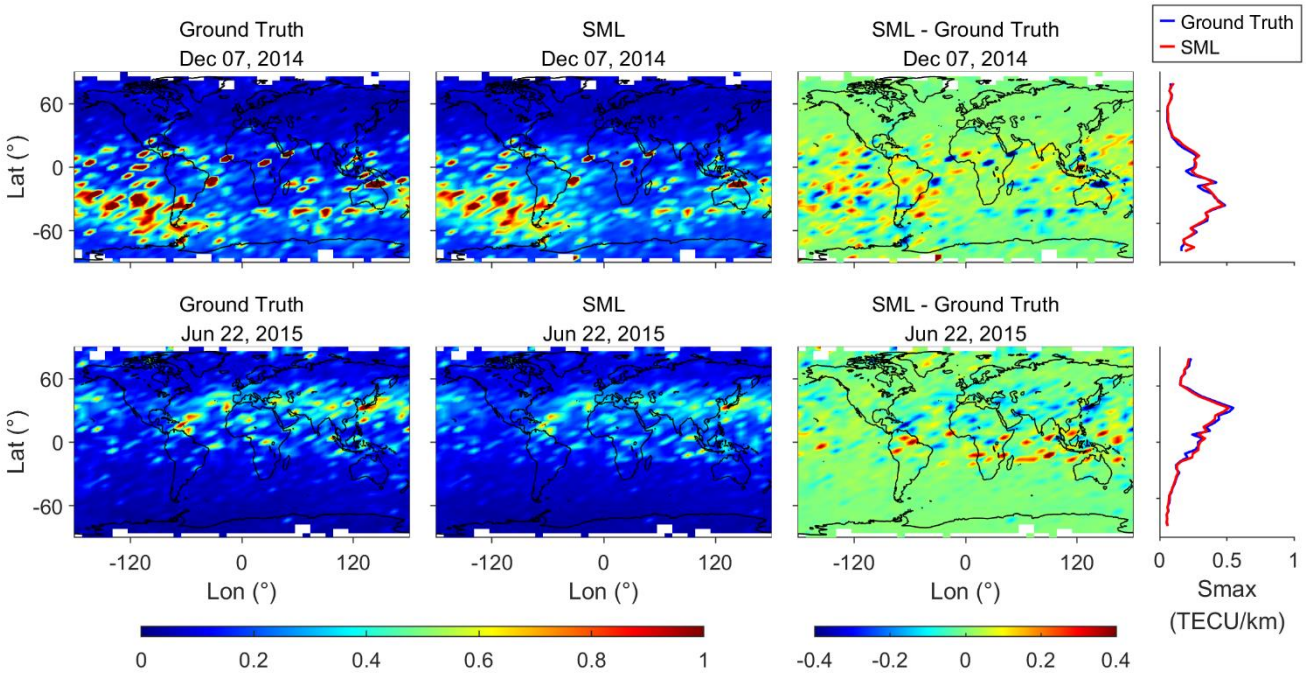


Figure 13: Latitude-longitude and latitudinal distributions of ground truth, SML-predicted EsI, and the corresponding error maps during geomagnetic storm time on 07 Dec 2014 and 22 Jun 2015.

4.4 External validation using ionosonde observations

In the previous empirical modeling of EsI, ionosonde foEs observations are usually used to verify the accuracy of RO measurements (Niu et al., 2019; Hu et al., 2022). To compare the SML-predicted EsI (Smax) and ionosonde foEs on the testing set, they should be matched under a specific spatiotemporal window. Luo et al. (2019) indicated that the influence of the increase in the spatial window on the matching results is much greater than that in the time window. Therefore, we adopt the window of (0.5 °, 0.5 °, 1 h) to ensure both the amount and consistency of the matched pairs. Fig 14 demonstrates the scatter plots of the matched ground truth and the SML-predicted EsI with ionosonde foEs. The results show that the fitted equation between SML predictions and foEs is much close to that between ground truth and foEs. The fitted RMSE of 0.122 TECU km^{-1} for SML is only slightly worse than that of the ground truth of 0.121 TECU km^{-1}, while the CC becomes even

better, from 0.716 for ground truth to 0.727 for SML. The high consistency between the metrics of SML predictions and ground truth indicates the good performance of SML.

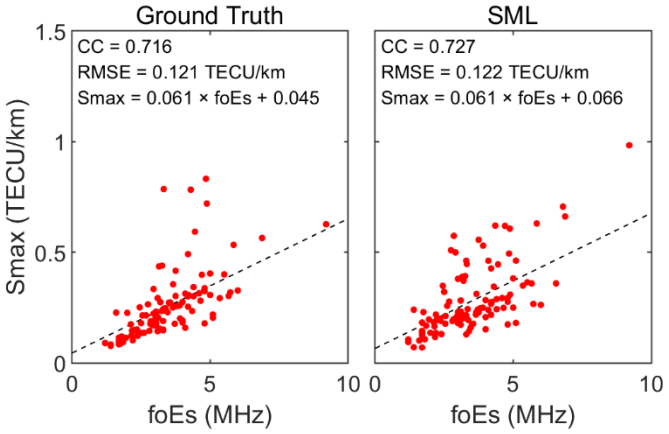


Figure 14: Scatter plots of the matched ground truth and the SML-predicted EsI with ionosonde foEs.

Furthermore, we verify the consistency of the long-term trends of SML results and ionosonde observations. Three ionosondes located at different latitudes, DW41K, BP440, and SO166, are selected for evaluation. Fig. 15 shows the daily maximum of the SML-predicted EsI and foEs over the selected ionosondes during 2014–2019. Note that here EsI and foEs are matched over the location of each ionosonde rather than within the spatiotemporal window. The climatological variations of SML-predicted EsI correspond well with those of the ionosonde foEs in low, mid and high latitudes. The CCs of SML-predicted EsI and ionosonde foEs over the three ionosondes are 0.613, 0.739, and 0.636, respectively.

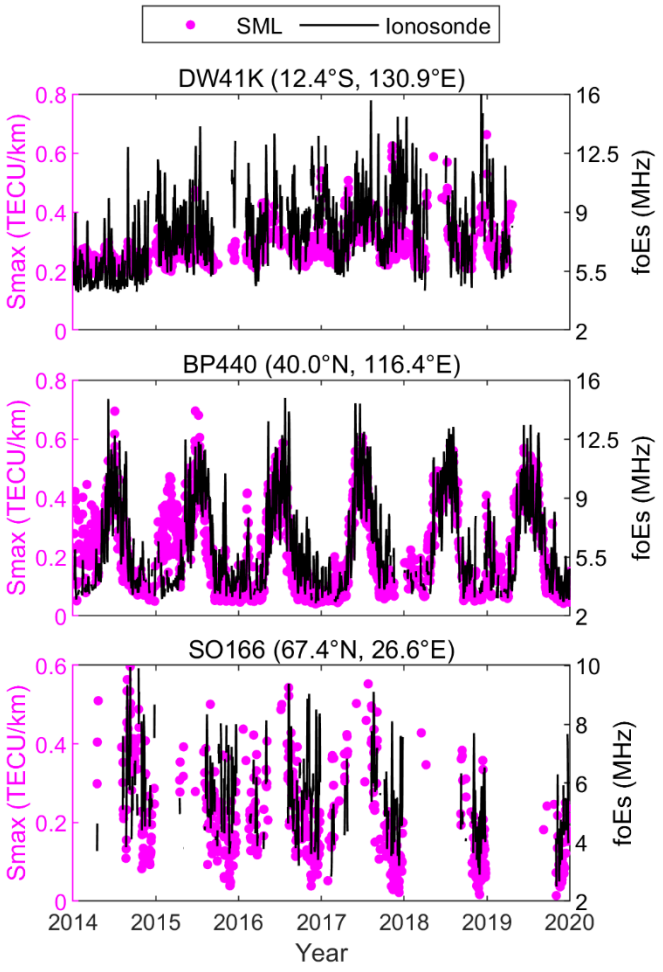


Figure 15: Time series of daily maximum of SML-predicted EsI and foEs over DW41K, BP440, and SO166 ionosondes during 2014–2019.

5 Discussion

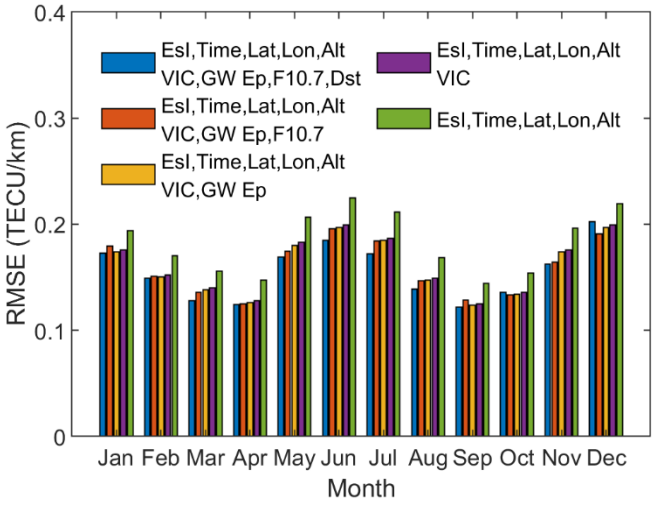
5.1 Advantage of incorporating physical observations in EsI prediction

To investigate the effect of incorporating physical observations on the prediction performance of SML model, we have evaluated the contributions of each physical parameter by removing them from the input variables one at a time. The sequence of removal is Dst, F10.7, GW Ep, and VIC. Table 1 represents the metrics of SML with different input variable combinations on the testing set, which are designated as Combinations 1–5. The models with Combinations 1–4 (with physical parameters) have ME/RMSE smaller than Combination 5 (without physical parameter), and the SML model with

Combination 1 has the smallest ME and RMSE compared to other combinations. We can see that after removing each parameter, ME and RMSE generally increase with different magnitudes. The increases are larger after removing VIC and F10.7 (ME of 0.009/0.006 TECU km^{-1} and RMSE of 0.018/0.002 TECU km^{-1}), while are smaller after removing GW Ep and Dst (ME of 0.002/0.000 TECU km^{-1} and RMSE of 0.002/0.002 TECU km^{-1}), indicating that the contributions of VIC and F10.7 on the model performance are more significant than those of GW Ep and Dst. The possible reason is that the effects of VIC and F10.7 on EsI are on long time scales, such as seasonal, annual and solar cycles, while GW Ep and Dst tend to impact the small-scale EsI distributions (e.g., hourly or during geomagnetic storms). For example, the relationship between VIC and seasonal/annual variation of Es layers has been widely revealed based on many observations and simulations (Shinagawa et al., 2017; Qiu et al., 2019; Yu et al., 2019; Luo et al., 2021a; Yamazaki et al., 2022; Ruan et al., 2025), and the variations of EsI and Es occurrence rate within one solar cycle have also been reported by Bergsson and Syndergaard (2022) and Fontes et al. (2024). In comparison, Qiu et al. (2023) and Liu et al. (2024c) indicated that the modulation of Es layers by GWs has the time scales comparable to the periods of GWs, and there is no significant consistency of the seasonal variation of Es occurrence with that of GWs. Liu et al. (2022) and Qiu and Liu (2025) pointed out that the downward impacts of geomagnetic storm on the Es layers mainly occur during the recovery phase. Furthermore, the monthly RMSE of SML with different input variables are shown in Fig. 16. It is evident that the monthly RMSE of SML with Combination 5 are larger than the SML models with other combinations. The SML with Combination 1 performs better than other models during January–November. The above results show the necessity of incorporating multiple related physical factors to consider the interactions of different atmospheric layers as a coupling system when constructing the Es prediction model, in which VIC and F10.7 are of significant contributions.

Table 1. Metrics of SML with different input variables

Input Variables	**ME (TECU km^{-1})**	**RMSE (TECU km^{-1})**
1. EsI, Time, Lat, Lon, Alt, VIC, GW Ep, F10.7, Dst	0.032	0.158
2. EsI, Time, Lat, Lon, Alt, VIC, GW Ep, F10.7	0.032	0.160
3. EsI, Time, Lat, Lon, Alt, VIC, GW Ep	0.038	0.162
4. EsI, Time, Lat, Lon, Alt, VIC	0.040	0.164
5. EsI, Time, Lat, Lon, Alt	0.049	0.184



445 **Figure 16: Monthly RMSEs of SML with different input variables.**

In recent years, some methods have been proposed for EsI modeling and prediction. Niu and Fang (2023) used COSMIC RO data to develop an empirical model that reproduces the climatological characteristics of EsI at low- and mid-latitudes with averaged deviation of 0.23 MHz. Emmons et al. (2023) presented two improved prediction model for EsI and 450 demonstrated better performance than those of the empirical models. Although the above methods achieved considerable EsI prediction performance, the lack of Es-related physical observations limited the further improvements of model accuracy. Tian et al. (2023) conducted the importance ranking of potential Es-related lower atmospheric parameters (zonal wind, geopotential, temperature, etc.), based on which they selected the input variables for their prediction model, but they did not consider VIC, the most important physical factor. In this study, we comprehensively incorporate VIC, GW Ep, F10.7 and 455 Dst that have been proven to have significant correlations with EsI. Hence, we have obtained a better performance than the previous models with only the EsI information as inputs.

5.2 Comparison of SML and other EsI estimation models

We have collected EsI estimation models using RO measurements from literatures in recent years. These models utilize different COSMIC RO products to derive various EsI proxies using statistical or ML methods. For intuitive comparison, all 460 EsI predictions are validated by ionosonde foEs observations on the same testing set using the same collocation window in Sect. 4.4, and their metric units are all converted to MHz. The comparison results are listed in Table 2. SML has the best prediction performance, and its RMSE is considerably smaller than all the other four models, with the improvements of 40.5%, 35.5%, 33.5%, and 20.1%, respectively. Emmons et al. (2023) used an ML model (SVM regression) to achieve a smaller RMSE than the other three empirical models, while it is still larger than the SML RMSE of this work. This may be 465 due to that the single SVM model is not as robust as SML.

Table 2. Comparison results of EsI estimation models.

	COSMIC product	EsI proxy	Method	RMSE (MHz)
Yu et al. (2022)	S4 index	S4max	Nonlinear least-squares fitting	1.787
Niu and Fang (2023)	TEC profiles	Smax	Multivariable functional fitting	1.650
Liu et al. (2024a)	Electron density profiles	Electron density	Linear fitting	1.601
Emmons et al. (2023)	50 Hz SNR profiles	Normalized SNR	SVM regression	1.331
This paper	TEC profiles	Smax	SML	1.064

6 Conclusions

This study proposes an SML method for global EsI prediction, in which a variety of Es-related physical observations are incorporated as inputs together with EsI derived from GNSS RO measurements. SML combines the strengths of the optimized base models to obtain lower prediction bias and variance. Taking RO-derived EsI as reference, the ME and RMSE of SML are 0.032 TECU km^{-1} and 0.158 TECU km^{-1}, respectively, and the reductions compared with the maximum ME and RMSE of base models are 39.6% and 7.1%, respectively. The evaluation results during 2014–2019 show that SML performs well in the prediction and the characteristics reconstruction of both long-term and short-term EsI variations. Taking ionosonde foEs observations as reference, SML shows better performance in EsI prediction compared to the existing methods, with the RMSE decreases of 20.1%–40.5%. Overall, this study presents an effective tool for high-precision global EsI prediction, which can be expected to provide valuable information for ionospheric irregularities monitoring and space weather forecasting. The method's incorporation of multiple Es-related physical factors is of significant contribution for deepening the understanding of complex interactions between lower atmosphere, thermosphere, ionosphere, and solar-terrestrial environment.

Author contributions

TH: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft, Visualization. XX: Investigation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. JL: Data curation, Writing – review & editing, Funding acquisition. JH: Methodology, Data curation. HL: Software, Formal analysis.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Code and data availability

The COSMIC RO data are available from COSMIC Data Analysis and Archive Center (CDAAC) at University Corporation for Atmospheric Research (UCAR) at https://data.cosmic.ucar.edu/gnss-ro/cosmic1/ (UCAR COSMIC Program, 2022). The ionosonde data are available from National Earth System Science Data Center of China (NESSDC) at http://wdc.geophys.ac.cn/dbList.asp?dType=IonoPublish (National Earth System Science Data Centre, 2025) and UK Solar System Data Center (UKSSDC) at https://www.ukssdc.ac.uk/wdcc1/ionosondes/secure/iono_data.shtml (UK Solar System Data Centre, 2025). The F10.7 and Dst indices are available from NASA/Goddard Space Flight Center's OMNIweb at https://omniweb.gsfc.nasa.gov/form/dx1.html (Papitashvili and King, 2025).

Acknowledgements

We kindly acknowledge the COSMIC Data Analysis and Archive Center (CDAAC) at University Corporation for Atmospheric Research (UCAR) for providing the COSMIC RO data, and NASA/Goddard Space Flight Center's OMNIweb for providing the F10.7 and Dst indices. The authors also acknowledge the use of HWM14 provided by National Space Science Data Center (NSSDC).

Financial Support

This study is supported by the National Natural Science Foundation of China under Grant Nos. 42074027, 42174017, 41774033, and 41774032.

References

Arras, C., Wickert, J., Beyerle, G., Heise, S., Schmidt, T., and Jacobi, C.: A global climatology of ionospheric irregularities derived from GPS radio occultation, Geophys. Res. Lett., 35, L14809, https://doi.org/10.1029/2008GL034158, 2008.

Asamoah, E. N., Cafaro, M., Epicoco, I., Franceschi, G. De, and Cesaroni, C.: A stacked machine learning model for the vertical total electron content forecasting, Adv. Sp. Res., 74, 223–242, https://doi.org/10.1016/j.asr.2024.04.055, 2024.

Batista, I. S. and Abdu, M. A.: Magnetic storm associated delayed sporadic E enhancements in the Brazilian Geomagnetic Anomaly, J. Geophys. Res., 82, https://doi.org/10.1029/JA082i029p04777, 1977.

Bergsson, B. and Syndergaard, S.: Global Temporal and Spatial Variations of Ionospheric Sporadic-E Derived From Radio Occultation Measurements, J. Geophys. Res. Sp. Phys., 127, e2022JA030296, https://doi.org/10.1029/2022JA030296, 2022.

Chu, Y. H., Wang, C. Y., Wu, K. H., Chen, K. T., Tzeng, K. J., Su, C. L., Feng, W., and Plane, J. M. C.: Morphology of sporadic E layer retrieved from COSMIC GPS radio occultation measurements: Wind shear theory examination, J. Geophys. Res. Sp. Phys., 119, 2117–2136, https://doi.org/10.1002/2013JA019437, 2014.

Drob, D. P., Emmert, J. T., Meriwether, J. W., Makela, J. J., Doornbos, E., Conde, M., Hernandez, G., Noto, J., Zawdie, K. A., McDonald, S. E., Huba, J. D., and Klenzing, J. H.: An update to the Horizontal Wind Model (HWM): The quiet time thermosphere, Earth Sp. Sci., 2, 301–319, https://doi.org/10.1002/2014EA000089, 2015.

Emmons, D. J., Wu, D. L., Swarnalingam, N., Ali, A. F., Ellis, J. A., Fitch, K. E., and Obenberger, K. S.: Improved models for estimating sporadic-E intensity from GNSS radio occultation measurements, Front. Astron. Sp. Sci., 10, 1327979, https://doi.org/10.3389/fspas.2023.1327979, 2023.

Fontes, P. A., Muella, M. T. A. H., Resende, L. C. A., and Fagundes, P. R.: Evidence of anti-correlation between sporadic (Es) layers occurrence and solar activity observed at low latitudes over the Brazilian sector, Adv. Sp. Res., 73, 3563–3577, https://doi.org/10.1016/j.asr.2023.09.040, 2024.

Forbes, J. M.: The Equatorial Electrojet, Rev. Geophys., 19, 469–504, https://doi.org/10.1029/RG019i003p00469, 1981.

Haldoupis, C., Pancheva, D., Singer, W., Meek, C., and MacDougall, J.: An explanation for the seasonal dependence of midlatitude sporadic E layers, J. Geophys. Res. Sp. Phys., 112, A06315, https://doi.org/10.1029/2007JA012322, 2007.

Han, Y., Wang, L., Fu, W., Zhou, H., Li, T., and Chen, R.: Machine Learning-Based Short-Term GPS TEC Forecasting During High Solar Activity and Magnetic Storm Periods, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 15, 115–126, https://doi.org/10.1109/JSTARS.2021.3132049, 2022.

Hastie, T., Friedman, J., and Tibshirani, R.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer, Berlin, Germany, 2009.

Heinselman, C. J., Thayer, J. P., and Watkins, B. J.: A high-latitude observation of sporadic sodium and sporadic E-layer formation, Geophys. Res. Lett., 25, 3059–3062, https://doi.org/10.1029/98GL02215, 1998.

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., and Kingsbury, B.: Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, IEEE Signal Process. Mag., 29, 82–97, https://doi.org/10.1109/MSP.2012.2205597, 2012.

Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, Neural Networks, 2, 359–366, https://doi.org/10.1016/0893-6080(89)90020-8, 1989.

Hu, T., Luo, J., and Xu, X.: Deriving Ionospheric Sporadic E Intensity From FORMOSAT-3/COSMIC and FY-3C Radio Occultation Measurements, Sp. Weather, 20, e2022SW003214, https://doi.org/10.1029/2022SW003214, 2022.

Kirkwood, S. and Nilsson, H.: High-latitude sporadic-E and other thin layers - The role of magnetospheric electric fields, Space Sci. Rev., 91, 579–613, https://doi.org/10.1023/A:1005241931650, 2000.

Leighton, H.I., Shapley, A.H., and Smith, E.K.: The occurrence of sporadic E during the IGY, in: Ionospheric sporadic, edited by: Smith, E.K., and Matsushita, S., Pergamon, 166–177.

Liu, H., Xu, X., Luo, J., and Hu, T.: Using radio occultation-based electron density profiles for studying sporadic E layer spatial and temporal characteristics, Earth, Planets Sp., 76, 93, https://doi.org/10.1186/s40623-024-02038-z, 2024a.

Liu, H., Yang, P., Ren, X., Mei, D., and Le, X.: The Short-Term Prediction of Low-Latitude Ionospheric Irregularities Leveraging a Hybrid Ensemble Model, IEEE Trans. Geosci. Remote Sens., 62, 4100615, https://doi.org/10.1109/TGRS.2023.3346449, 2024b.

Liu, Y., Zhou, C., Xu, T., Deng, Z., Du, Z., Lan, T., Tang, Q., Zhu, Y., Wang, Z., and Zhao, Z.: Geomagnetic and Solar Dependencies of Midlatitude E-Region Irregularity Occurrence Rate: A Climatology Based on Wuhan VHF Radar Observations, J. Geophys. Res. Sp. Phys., 127, e2021JA029597, https://doi.org/10.1029/2021JA029597, 2022.

Liu, Y., Chen, Z., Fan, Z., Zhou, C., Wang, X., Zhang, Y., Zhou, Y., Lan, T., and Qing, H.: Statistical analysis on orographic atmospheric gravity wave and sporadic E layer, J. Atmos. Solar-Terrestrial Phys., 259, 106256, https://doi.org/10.1016/j.jastp.2024.106256, 2024c.

Luo, J., Wang, H., Xu, X., and Sun, F.: The influence of the spatial and temporal collocation windows on the comparisons of the ionospheric characteristic parameters derived from COSMIC radio occultation and digisondes, Adv. Sp. Res., 63, 3088–3101, https://doi.org/10.1016/j.asr.2019.01.024, 2019.

Luo, J., Liu, H., and Xu, X.: Sporadic E morphology based on COSMIC radio occultation data and its relationship with wind shear theory, Earth, Planets Sp., 73, 212, https://doi.org/10.1186/s40623-021-01550-w, 2021a.

Luo, J., Hou, J., and Xu, X.: Variations in Stratospheric Gravity Waves Derived from Temperature Observations of Multi-GNSS Radio Occultation Missions, Remote Sens., 13, 4835, https://doi.org/10.3390/rs13234835, 2021b.

MacDougall, J. W., Jayachandran, P. T., and Plane, J. M. C.: Polar cap sporadic-E: Part 1, observations, J. Atmos. Solar-Terrestrial Phys., 62, 1155–1167, https://doi.org/10.1016/S1364-6826(00)00093-6, 2000a.

MacDougall, J. W., Plane, J. M. C., and Jayachandran, P. T.: Polar cap Sporadic-E: Part 2, modeling, J. Atmos. Solar-Terrestrial Phys., 62, 1169–1176, https://doi.org/10.1016/S1364-6826(00)00092-4, 2000b.

Mathews, J. D.: Sporadic E: Current views and recent progress, J. Atmos. Solar-Terrestrial Phys., 60, 413–435, https://doi.org/10.1016/S1364-6826(97)00043-6, 1998.

Moro, J., Xu, J., Denardini, C. M., Resende, L. C. A., Da Silva, L. A., Chen, S. S., Carrasco, A. J., Liu, Z., Wang, C., and Schuch, N. J.: Different Sporadic-E (Es) Layer Types Development During the August 2018 Geomagnetic Storm: Evidence of Auroral Type (Esa) Over the SAMA Region, J. Geophys. Res. Sp. Phys., 127, e2021JA029701, https://doi.org/10.1029/2021JA029701, 2022.

National Earth System Science Data Centre: Ionosonde data, NESSDC [data set], http://wdc.geophys.ac.cn/, last access: 30 March 2025.

Niu, J. and Fang, H.: An Empirical Model of the Sporadic E Layer Intensity Based on COSMIC Radio Occultation Observations, Sp. Weather, 21, e2022SW003280, https://doi.org/10.1029/2022SW003280, 2023.

Niu, J., Weng, L., and Fang, H.: An attempt to inverse the ionospheric sporadic-E layer critical frequency based on the COSMIC radio occultation data, Adv. Sp. Res., 63, 1204–1213, https://doi.org/10.1016/j.asr.2018.10.029, 2019.

Nygren, T., Voiculescu, M., and Aikio, A. T.: The role of electric field and neutral wind in the generation of polar cap sporadic E, Ann. Geophys., 26, 3757–3763, https://doi.org/10.5194/angeo-26-3757-2008, 2008.

Papitashvili, N. E. and King, J. H.: OMNI Hourly Data Set, NASA Space Physics Data Facility [data set], https://doi.org/10.48322/1SHR-HT18, 2025.

Qiu, L., Zuo, X., Yu, T., Sun, Y., and Qi, Y.: Comparison of global morphologies of vertical ion convergence and sporadic E occurrence rate, Adv. Sp. Res., 63, 3606–3611, https://doi.org/10.1016/j.asr.2019.02.024, 2019.

Qiu, L., Yamazaki, Y., Yu, T., Becker, E., Miyoshi, Y., Qi, Y., Siddiqui, T. A., Stolle, C., Feng, W., Plane, J. M. C., Liang, Y., and Liu, H.: Numerical Simulations of Metallic Ion Density Perturbations in Sporadic E Layers Caused by Gravity Waves, Earth Sp. Sci., 10, e2023EA003030, https://doi.org/10.1029/2023EA003030, 2023.

Qiu, L. and Liu, H.: Sporadic-E Layer Responses to Super Geomagnetic Storm 10-12 May 2024, Geophys. Res. Lett., 52, e2025GL115154, https://doi.org/10.1029/2025GL115154, 2025.

Raghavarao, R., Patra, A. K., and Sripathi, S.: Equatorial E region irregularities: A review of recent observations, J. Atmos. Solar-Terrestrial Phys., 64, 1435–1443, https://doi.org/10.1016/S1364-6826(02)00107-4, 2002.

Resende, L. C. A., Denardini, C. M., and Batista, I. S.: Abnormal fb Es enhancements in equatorial Es layers during magnetic storms of solar cycle 23, J. Atmos. Solar-Terrestrial Phys., 102, 228–234, https://doi.org/10.1016/j.jastp.2013.05.020, 2013.

Resende, L. C. A., Batista, I. S., Denardini, C. M., Carrasco, A. J., Andrioli, V. D. F., Moro, J., Batista, P. P., and Chen, S. S.: Competition between winds and electric fields in the formation of blanketing sporadic E layers at equatorial regions, Earth, Planets Sp., 68, 201, https://doi.org/10.1186/s40623-016-0577-z, 2016.

Resende, L. C. A., Shi, J., Denardini, C. M., Batista, I. S., Picanço, G. A. S., Moro, J., Chagas, R. A. J., Barros, D., Chen, S. S., Nogueira, P. A. B., Andrioli, V. F., Silva, R. P., Carrasco, A. J., de Araujo, R. C., Wang, C., and Liu, Z.: The Impact of the Disturbed Electric Field in the Sporadic E (Es) Layer Development Over Brazilian Region, J. Geophys. Res. Sp. Phys., 126, e2020JA028598, https://doi.org/10.1029/2020JA028598, 2021.

Ruan, H., Qiu, X., Guo, X., and Wang, X.: Climatological Investigation of Ionospheric Es Layer Based on Occultation Data, Remote Sens., 17, 280, https://doi.org/10.3390/rs17020280, 2025.

Seid, C. M., Su, C. L., Wang, C. Y., and Chu, Y. H.: Interferometry Observations of the Gravity Wave Effect on the Sporadic E Layer, Atmosphere (Basel)., 14, 987, https://doi.org/10.3390/atmos14060987, 2023.

Shinagawa, H., Miyoshi, Y., Jin, H., and Fujiwara, H.: Global distribution of neutral wind shear associated with sporadic E layers derived from GAIA, J. Geophys. Res. Sp. Phys., 122, 4450–4465, https://doi.org/10.1002/2016JA023778, 2017.

Syndergaard, S., Schreiner, W. S., Rocken, C., Hunt, D. C., and Dymond, K. F.: Preparing for COSMIC: Inversion and analysis of ionospheric data products, in: Atmosphere and Climate: Studies by Occultation Methods, edited by: Foelsche, U., Kirchengast, G., and Steiner, A, Graz, Australia, 137–146, https://doi.org/10.1007/3-540-34121-8_12, 2006.

Tang, Q., Zhou, C., Liu, H., Du, Z., Liu, Y., and Zhao, J.: Global Structure and Seasonal Variations of the Tidal Amplitude in Sporadic-E Layer, J. Geophys. Res. Sp. Phys., 127, e2022JA030711, https://doi.org/10.1029/2022JA030711, 2022a.

Tang, Q., Sun, H., Du, Z., Zhao, J., Liu, Y., Zhao, Z., and Feng, X.: Unusual Enhancement of Midlatitude Sporadic-E Layers in Response to a Minor Geomagnetic Storm, Atmosphere (Basel)., 13, 816, https://doi.org/10.3390/atmos13050816, 2022b.

Tian, P., Yu, B., Ye, H., Xue, X., Wu, J., and Chen, T.: Ionospheric irregularity reconstruction using multisource data fusion via deep learning, Atmos. Chem. Phys., 23, 13413–13431, https://doi.org/10.5194/acp-23-13413-2023, 2023.

UCAR COSMIC Program: COSMIC-1 Data Products, UCAR/NCAR – COSMIC [data set], https://doi.org/10.5065/ZD80-KD74, 2022.

UK Solar System Data Centre: Ionospheric data, UKSSDC [data set], https://www.ukssdc.ac.uk/wdcc1/ionosondes/secure/iono_data.shtml, last access: 30 March 2025.

Wolpert, D.: Stacked Generalization, Neural Networks, 5, 241–259, https://doi.org/10.1016/S0893-6080(05)80023-1, 1992.

Xu, X., Luo, J., Wang, H., Liu, H., and Hu, T.: Morphology of sporadic E layers derived from Fengyun-3C GPS radio occultation measurements, Earth, Planets Sp., 74, 55, https://doi.org/10.1186/s40623-022-01617-2, 2022.

Yamazaki, Y., Arras, C., Andoh, S., Miyoshi, Y., Shinagawa, H., Harding, B. J., Englert, C. R., Immel, T. J., Sobhkhiz-Miandehi, S., and Stolle, C.: Examining the Wind Shear Theory of Sporadic E With ICON/MIGHTI Winds and COSMIC-2 Radio Occultation Data, Geophys. Res. Lett., 49, e2021GL096202, https://doi.org/10.1029/2021GL096202, 2022.

Yetilmezsoy, K.: Applications of Soft Computing Methods in Environmental Engineering, in: Handbook of Environmental Materials Management, edited by: Hussain, C., Springer, Cham, Germany, 2001–2046, https://doi.org/10.1007/978-3-319-73645-7_149, 2019.

Yu, B., Xue, X., Yue, X., Yang, C., Yu, C., Dou, X., Ning, B., and Hu, L.: The global climatology of the intensity of the ionospheric sporadic E layer, Atmos. Chem. Phys., 19, 4139–4151, https://doi.org/10.5194/acp-19-4139-2019, 2019.

Yu, B., Scott, C. J., Xue, X., Yue, X., Chi, Y., Dou, X., and Lockwood, M.: A Signature of 27 day Solar Rotation in the Concentration of Metallic Ions within the Terrestrial Ionosphere, Astrophys. J., 916, 106, https://doi.org/10.3847/1538-4357/ac0886, 2021.

Yu, B., Xue, X., Scott, C. J., Yue, X., and Dou, X.: An Empirical Model of the Ionospheric Sporadic E Layer Based on GNSS Radio Occultation Data, Sp. Weather, 20, e2022SW003113, https://doi.org/10.1029/2022sw003113, 2022.

Zhukov, A. V., Yasyukevich, Y. V., and Bykov, A. E.: GIMLi: Global Ionospheric total electron content model based on machine learning, GPS Solut., 25, 19, https://doi.org/10.1007/s10291-020-01055-1, 2021.