Relativistic Electron Flux Prediction at Geosynchronous Orbit Based on the Neural Network and the Quantile Regression Method

Hui Zhang1, Suiyan Fu1, Lun Xie1, Duo Zhao2, Chao Yue1, Zuyin Pu1, Ying Xiong1, Tong Wu1, Shaojie Zhao1, Yixin Sun1, Bo Cui1, and Zhekai Luo1

1School of Earth and Space Sciences, Peking University, Beijing, China, 2Institute of Applied Physics and Computational Mathematics, Beijing, China

Abstract Geosynchronous satellites are exposed to the relativistic electrons, which may cause irreparable damage to the satellites. The prediction of the relativistic electron flux is therefore important for the safety of the satellites. Unlike previous works focusing on the single-value prediction of relativistic electron flux, we predict the relativistic electron flux in a probabilistic approach by using the neural network and the quantile regression method. In this study, a feedforward neural network is first designed to predict average daily flux of relativistic electrons (>2 MeV), or the expectation of the flux from the probabilistic perspective, at geosynchronous orbit 1 day in advance. The neural network performs well, with the average root mean squared error, the average prediction efficiency, and the average linear correlation coefficient between observations and predictions reaching 0.305, 0.832, and 0.916, respectively, during the periods of 2011–2017. We then combine the quantile regression method with the feedforward neural network to predict the quantiles of relativistic electron flux by applying a special loss function to the neural network. We use the multiple-quantiles prediction model to predict flux ranges of the relativistic electrons and the corresponding probabilities, which is an advantage over the single-value prediction. Moreover, it appears to be for the first time that the approximate shape of the probability density function of relativistic electron flux is predicted.

1. Introduction

Geosynchronous orbit (GEO), which is located close to the outer radiation belt, is a critical region since hundreds of satellites for telecommunication, navigation, and other applications operate in this region. Nonetheless, these satellites are exposed to the relativistic electrons with energy reaching MeV, which can cause irreparable damage to the satellites (Horne et al., 2013; Lai et al., 2018; Pilipenko et al., 2006; Wrenn et al., 2002). Therefore, predicting the flux of relativistic electrons at GEO is essential for the safety of the satellites.

In the past decades, great progress has been made in understanding the acceleration mechanism of relativistic electrons and providing techniques in predicting the relativistic electron flux at GEO. There are mainly two kinds of theories that explain the electron acceleration mechanism, namely, the radial transport (Li et al., 2001) and the local acceleration (Horne & Thorne, 1998; Roth et al., 1999; Summers et al., 1998; Summers & Ma, 2000). Based on these two theories, some physical models are proposed to make predictions. Li et al. (2001) and Li (2004) used solar wind parameters, Dst, and historical flux values of MeV electrons to predict the average daily flux of MeV electrons 1–2 days in advance, based on the standard radial diffusion equation. The prediction efficiencies of their works reached around 0.6 and 0.72, respectively. Turner and Li (2008) used the flux of low- and high-energy electrons to predict the average daily flux of high-energy electrons several days in advance, based on the strong correlation between the flux of low-energy electrons and the flux of high-energy electrons. A combination of this model and the radial transport model produced an average prediction efficiency (PE) of 0.76 for the Years 1995–2006. Meanwhile, Lyatsky and Khazanov (2008) developed a similar model focusing on the sources and losses of relativistic electrons at GEO to make predictions 1 day in advance by using the solar wind data as the only input parameters.

In addition to the physics-based models, there are also many empirical models and data-based models such as the linear prediction filters (Chen et al., 2016; Tsutai et al., 1999), the support vector machine (Wang &
ZHANG ET AL. 2013

The neural network of their work is the long short time memory network, which is an advantage over the single value prediction. However, it is based on the statistical result and does not consider the temporal correlation of the data. Moreover, by making multiple quantile predictions, we show other capabilities of the combined model. It can predict flux ranges of the relativistic electrons and the corresponding probabilities, which is an advantage over the single-value prediction. And it appears to be for the first time that the probability density function of the relativistic electron flux is predicted. The data and codes of this study are available online (https://github.com/studyalltheway/Relativistic-Electron-Flux-Prediction).

2. Data

In this study, six solar wind parameters, four geomagnetic indices, and fluxes of medium and high energy (tens keV to MeV) electrons are used to predict relativistic electron flux. The solar wind parameters include magnetic field intensity (B), solar wind temperature (T), velocity (V), pressure (P), proton density (N), and electric field intensity (Ey). The geomagnetic indices consist of Kp, Dst, ap, and AE indices. The electron fluxes are composed of fluxes of >2 MeV, >0.8 MeV, 475 keV, 275 keV, 150 keV, 75 keV, and 40 keV electrons. The relativistic electron flux refers to the flux of >2 MeV electrons.

2.1. Data Sources

The solar wind parameters are obtained from the OMNIWeb, and the geomagnetic indices are obtained from the World Data Center for Geomagnetism, Kyoto, with 1-hr resolution. The integral electron flux data of seven energy channels (>2 MeV, >0.8 MeV, 475 keV, 275 keV, 150 keV, 75 keV, and 40 keV) are obtained from GOES 13. The time interval is from 7 January 2011 to 14 December 2017 since the flux of the medium-energy electrons are available only during this period. The high-energy electron flux (>2 and
0.8 MeV) are 1-min resolution data, while the medium-energy electron flux (475, 275, 150, 75, and 40 keV) are 5-min resolution data.

### 2.2. Data Preprocessing

For all of these 17 variables, the raw data are adjusted to the 1-day resolution. As for the solar wind parameters and geomagnetic indices, the daily data are calculated as the mean values of the raw data every day. If there is not any available hourly data in a day, this day will be marked as the missing point. As for the high-energy electron flux, the raw data are first adjusted to 1-hr resolution data and then adjusted to the average daily data. After the daily data are constructed, we supplement some missing data by linear interpolation method (around 5.8% of all the >2 MeV data) and then calculate the log10 values. The daily medium-energy electron flux data are obtained in the same way, while the raw data of medium-energy electron flux are 5-min resolution data. Through this process, 2,423 data points are available eventually.

### 2.3. Correlation Analysis

The Spearman correlation coefficients analysis is used to select the inputs of the neural network. The result of the Spearman correlation coefficients based on the data from 2011–2017 is shown in Figure 1. It shows that the correlation coefficients are quite high between the relativistic electron flux and these input variables in the last 5 days. Therefore, the data of the last five consecutive days are selected to predict relativistic electron flux 1 day in advance.

### 2.4. Data Sets

Three data sets, training set, validation set, and test set are made, with the data of approximate 1 year as a test set, while the remaining 6-year data as the training and validation set. Moreover, we adopt the 10 cross-validation method. In this way, the 6-year data are divided into 10 chunks, with nine chunks to train the model and the remaining one chunk to validate the result. With this process running 10 times, 10 submodels with each submodel corresponding to different validation set are acquired. The final prediction of the test set is the mean value of the predictions of these 10 submodels.

---

**Figure 1.** Correlation analysis between relativistic electron flux and input variables with different offset days. $|L_C|$ is the absolute value of the linear correlation coefficient. (a) Solar wind parameters. (b) Geomagnetic indices. (c) Electron fluxes of seven energy channels.
3. Method

3.1. The FFN

The FFN is a widely used type of neural networks, where the information moves only in one direction, forward, from the input nodes through the hidden nodes and then to the output nodes. Usually, the FFN is composed of fully connected layers, which we call the fully connected FFN (FCN). The FCN consists of an input layer, one or more hidden layers, and an output layer, as shown in Figure 2a. In the FCN, each node in one layer connects with a certain weight to every node in the following layer, which means it is “fully connected.” Fundamentally, the FCN approximates a continuous function through a composite function. An FCN with one hidden layer can approximate any continuous function with arbitrary accuracy as long as it has enough nodes (Hornik et al., 1989). The mathematical form of the neural network can be presented as

\[ x_{j}^{k+1} = f_{\text{activation}} \left( \sum_{i=1}^{N_k} w_{ij}^{k} \cdot x_{i}^{k} + b_{j}^{k} \right) \]  

where \( x_{j}^{k+1} \) is the output of the \( j \)th node of layer \( k+1 \), \( x_{i}^{k} \) is the output of the \( i \)th node of layer \( k \), \( w_{ij}^{k} \) and \( b_{j}^{k} \) are corresponding weights and bias, which are training parameters, and \( f_{\text{activation}} \) is the activation function.

The outputs of nodes in the output layer can be seen as composite functions with many parameters. Correspondingly, the training of the network is to find the best parameters to make the outputs approach the true values. The method to adjust the parameters is “backward propagation” based on the gradient descent theory (Rumelhart et al., 1986). By back-propagating the “error” from the output layer through the hidden layers to the input layer, the parameters of the network are adjusted along the negative gradient descent direction. The error is the value of the loss function, which evaluates the difference between the outputs and the true values. Repeating this process many times can make the error small and may obtain a good result.

3.2. The Optimization of the Neural Network and Our FFN Model

We adopt a FFN in this study, which is a two-hidden layer FCN with an external head or the scaling transformation layer, as shown in Figure 2b. Some training techniques are applied to optimize the neural network. The techniques include scaled exponential linear unit (SELU) and sigmoid activation functions, the L1 and L2 regularization, the grid search method to find the best hyperparameters, and a scaling transformation layer with L1 regularization added on the FCN.

Activation functions are used to fit nonlinear functions by applying to the hidden layers. In this study, we adopt two activation functions, sigmoid function and SELU function, defined as follows:

\[ \text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \]  
\[ \text{selu}(x) = \begin{cases} x & x > 0 \\ \lambda (x - \alpha e^x) & x \leq 0 \end{cases} \quad (\lambda = 1.051, \alpha = 1.673) \]
The sigmoid function is widely used as the activation function in neural networks, which maps the input into the interval (0, 1). However, in multiple-layer neural networks, the sigmoid function may cause the problem of vanishing gradients, which will impede the training of the neural network and obtain a bad result. Rectified linear unit (RELU) activation function is thus proposed to solve the problem in deep neural networks, which is defined as $\text{relu}(x) = \max(0, x)$ (Glorot et al., 2011). More recently, Klambauer et al. (2017) proposed the SELU activation function, which is similar to the RELU activation function but reported to obtain better results compared with RELU.

Regularization is an important method to prevent overfitting of neural networks. The L1 and L2 regularization methods are widely used in neural networks. Both L1 and L2 regularization restrain the training parameters (weights) by adding a penalty term to the loss function. The penalty term of the L1 regularization is the sum of the absolute values of the weights, while the penalty term of the L2 regularization is the sum of the squares of the weights. A critical feature of the L1 regularization is that it makes parameters more likely to become zeros, making it a good method for feature selection (Tibshirani, 1996). In this study, the L1 regularization is applied to the scaling transformation layer and the L2 regularization to the other hidden layers, and the final loss function is

$$L = \frac{1}{N} \sum_{i=1}^{N} (f(x_i) - y_i)^2 + k_1 \sum_{k=1}^{N_1} |w_k| + k_2 \sum_{j=1}^{N_2} w_j^2$$

where $N$ is the size of the data set, $N_1$ is the number of weights applied with L1 regularization, $N_2$ is the number of weights applied with L2 regularization, $f(x_i)$ is the prediction of the neural network on the input $x_i$, $y_i$ is the corresponding true value, $w$ is the weight, and $k_1$ and $k_2$ are hyperparameters. The first term on the right of Equation 4 is about the difference between the model outputs and true values, which is calculated as the mean square error to estimate the expectation of the flux from the probabilistic perspective, which we will explain in detail in section 3.4. The second term is the L1 regularization, and the third term is the L2 regularization.

For the techniques applied to the neural network, a scaling transformation layer with L1 regularization is an efficient technique we propose to improve the prediction results. The proposed scaling transformation layer is a “one-one” connected layer, which makes a scaling transformation of the input ($wx + b$), as shown in Figure 2b. The coefficients of the scaling transformation layer are training parameters of the neural network rather than specific values set manually. The reason that we add a scaling transformation layer with L1 regularization on an FCN is that we make this an in-built way of feature selection. The explanatory variables selected in this study may contain many unimportant features. The L1 regularization will make the scaling coefficients ($w$) more likely to become zeros, thus dramatically eliminating the influence of the unimportant features.

The final structure of our FFN model is as follows, shown in Figure 3.

Input (17 × 5 = 85 nodes) —— scaling transformation layer (85 nodes, L1 regularization) —— hidden layer (25 nodes, L2 regularization, SELU activation) —— hidden layer (25 nodes, L2 regularization, sigmoid activation) —— output (one node).

All inputs are normalized by subtracting their respective mean values and dividing by their respective standard deviations. The optimizer selected for the neural network is the Adam optimizer, which uses historical information of derivatives. To show the reasonability of the scaling transformation layer and SELU activation function, the experiment results of neural networks with different structures are listed in Table 1, which may be more comprehensible after section 3.3.
### 3.3. Performance Metrics

Generally, the metrics to evaluate a model’s performance include the root mean squared error (RMSE), the PE, and the linear correlation coefficient (LC) between observations and predictions. The RMSE, PE, and LC are defined as follows:

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - f_i)^2} \tag{5}
\]

\[
PE = 1 - \frac{\sum_{i=1}^{N} (y_i - f_i)^2}{\sum (y_i - \bar{y})^2 - 1} - \frac{RMSE^2}{\text{Var}(y)} \tag{6}
\]

\[
LC = \frac{\text{Cov}(y, f)}{\sqrt{\text{Var}(y) \text{Var}(f)}} \tag{7}
\]

where \(N\) is the size of the data set, \(y_i\) is the observation, \(\bar{y}\) is the mean value of \(y\), \(f_i\) is the prediction, \(\text{Var}(y)\) is the variance of \(y\), and \(\text{Cov}(y, f)\) is the covariance of \(y\) and \(f\). The smaller the RMSE is and the closer the PE and LC are to 1, the better the performance is.

### 3.4. Quantile Regression Method

From the probabilistic perspective, the relativistic electron flux of a single day is not a single value but a random variable with a probability density function. The expectation, median, and quantiles of the flux can be obtained from this perspective. The uncertainty of the prediction can be presented by the probability density function. We introduce the quantile regression method here to predict quantiles and the approximate shape of the probability density function of the flux.

Quantile regression is a method that can not only predict a single value but also many quantiles of the response variable (Hallock & Koenker, 2001; Koenker & Bassett, 1978). Quantiles are cut points dividing the range of a probability distribution into continuous intervals with equal probabilities. The \(\tau\)th quantile indicates the probability that the flux is less than the quantile is \(\tau\).

A simple explanation of the quantile regression theory is provided as follows. Let \(Y\) be a real-valued random variable with cumulative distribution function \(F_Y(y) = P(Y \leq y)\). The \(\tau\)th quantile of \(Y\) is defined by \(Q_Y(\tau) = F_Y^{-1}(\tau) = \inf \{y : F_Y(y) \geq \tau\}\), where \(\tau \in (0,1)\).

Define the loss function:

\[
Loss(y, f) = \begin{cases} 
\tau(y - f) & y \geq f \\
(1 - \tau)(f - y) & y < f 
\end{cases} \tag{8}
\]

The \(\tau\)th quantile can be found by minimizing the expected loss of \(y-f\) with respect to \(f\) (\(f\) can be seen as the output of the neural work):

\[
\min_f E(Loss(y, f)) = \min_f \left\{ (\tau - 1) \int_{-\infty}^{\infty} (y - f) dF_Y(y) + \tau \int_{-\infty}^{f} (y - f) dF_Y(y) \right\} \tag{9}
\]

By setting the derivative of the expected loss function (9) to 0, we can get the solution:

\[
\frac{\partial E(Loss(y, f))}{\partial f} = \begin{cases} 
\frac{\partial}{\partial f} \left( (\tau - 1) \int_{-\infty}^{\infty} (y - f) dF_Y(y) \right) & y \geq f \\
\frac{\partial}{\partial f} \left( \tau \int_{-\infty}^{f} (y - f) dF_Y(y) \right) & y < f 
\end{cases}
\]

\[
\frac{\partial}{\partial f} \left( (\tau - 1) \int_{-\infty}^{\infty} (y - f) dF_Y(y) \right) = (\tau - 1) \int_{-\infty}^{f} dF_Y(y) = (\tau - 1) (\tau \cdot y - f)
\]

\[
\frac{\partial}{\partial f} \left( \tau \int_{-\infty}^{f} (y - f) dF_Y(y) \right) = \tau \int_{-\infty}^{f} dF_Y(y) = \tau \cdot \tau \cdot y - f
\]

Table 1

| Neural networks with different structures | Performance metrics | Hyperparameters |
|----------------------------------------|---------------------|---------------|
|                                        | RMSE    | PE    | LC   | Learning rate | Training epochs | Batch size | Nodes | \(k_1, k_2\) |
| FCN (two hidden layers, sigmoid + sigmoid) | 0.320   | 0.814 | 0.908 | 0.07         | 1,600           | 300  | 20, 20 | \(k_2 = 0.001\) |
| FCN (two hidden layers, SELU + sigmoid) | 0.320   | 0.816 | 0.909 | 0.016        | 1,500           | 300  | 25, 25 | \(k_2 = 0.001\) |
| FFN (FCN plus a head, sigmoid + sigmoid) | 0.311   | 0.826 | 0.912 | 0.014        | 2,000           | 300  | 50, 50 | \(k_1 = 0.0005, k_2 = 0.001\) |
| FFN (FCN plus a head, SELU + sigmoid) | 0.305   | 0.832 | 0.916 | 0.012        | 1,600           | 300  | 25, 25 | \(k_1 = 0.001, k_2 = 0.001\) |

Note: The models are implemented in Python programming language and built upon TensorFlow. The RMSE, PE, and LC are mean values on the seven test sets during the periods of 2011–2017. The hyperparameters are obtained by the grid search method.
The expected loss function is minimized to obtain the \( q_\tau \)th quantile of \( Y \), which is consistent with the training process of neural networks. Therefore, the neural network can be combined with quantile regression. By setting the loss function of the neural network as the form of the loss function of the quantile regression, the \( q_\tau \)th quantile can be predicted by the neural network. This model is referred to as the single-quantile prediction model.

Moreover, from the probabilistic perspective, the expectation of the flux can be obtained by minimizing the mean square error loss function (the first term of function (4)), which is explained as follows:

\[
\min_f E[\text{Loss}(y, f)] = \min_f \int_{-\infty}^{\infty} (y - f)^2 dF_Y(y)
\]

By setting the derivative of the expected loss function (11) to 0, we can get the solution:

\[
\frac{dE}{df} = \int_{-\infty}^{\infty} 2(f - y) dF_Y(y) = 0
\]

Let the \( f_0 \) be the solution. Equation 12 reduces to \( f_0 = \int_{-\infty}^{\infty} y dF_Y(y) \), which means \( f_0 \) is the expectation of \( Y \).

### 3.5. Multiple-Quantiles Prediction Method

In addition to predicting a specific quantile of relativistic electron flux, the combination of the neural network and the quantile regression can also be used to predict multiple quantiles at the same time, which can further predict flux ranges and the corresponding probabilities. Moreover, the approximate shape of the probability density function of the flux can be predicted.

In this study, we design two more complex combined neural networks to predict multiple quantiles at the same time. The first combined neural network is designed with three outputs, which are for predicting the \((0.5 - \alpha)\)th, the 0.5th, and the \((0.5 + \alpha)\)th quantiles, as shown in Figure 4. It consists of a scaling transformation layer, a fully connected hidden layer, and a sparsely connected hidden layer. From another
For the single-quantile prediction model introduced in section 3.4, we can predict different quantiles by setting \( \tau \) to different values in the loss function (8). But the predicted quantiles may not be in the right order. To solve this problem, we add a penalty term in the loss function of the multiple-quantiles prediction model, as shown from Equations 13 to 17.

The loss function of the combined neural network is designed as

\[
L = \sum_{i=1}^{3} L_i + \lambda \left[ \sigma \left( \tilde{y}_{0.5} - \alpha - y \right) + \sigma \left( \tilde{y}_{0.5} - \tilde{y}_{0.5} + \alpha \right) \right]
\]

\[
L_1 = \{(0.5 - \alpha) \tilde{y}_{0.5} - y \}, \quad \tilde{y}_{0.5} - \alpha \geq y(0.5 + \alpha)(y - \tilde{y}_{0.5} - \alpha) \quad \tilde{y}_{0.5} - \alpha < y
\]

\[
L_2 = 0.5 \tilde{y}_{0.5} - y
\]

\[
L_3 = \{(0.5 + \alpha) \tilde{y}_{0.5} + \alpha - y \}, \quad \tilde{y}_{0.5} + \alpha \geq y(0.5 - \alpha)(y - \tilde{y}_{0.5} + \alpha) \quad \tilde{y}_{0.5} + \alpha < y
\]

\[
\sigma(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}
\]

The first term on the right side of Equation 13 represents three loss functions to predict three quantiles, while the second term on the right side is designed as a penalty term to keep three quantiles in the right order. The 0.5th quantile is the median of the relativistic electron flux. The \( (0.5 - \alpha) \)th and \( (0.5 + 0.5 \alpha) \)th quantiles can determine a flux range \( [\tilde{y}_{0.5} - \alpha, \tilde{y}_{0.5} + \alpha] \) with a probability of \( 2 \alpha \).

The other combined neural network is similar to the three-quantiles prediction network, while it has nine outputs, which are for predicting 0.1th, 0.2th, 0.3th, ... 0.9th quantiles, as shown in Figure 4b. Similar penalty terms to that in Equation 13 are added in the combined loss function to avoid the order inconsistency of different quantiles. From the nine quantiles, the approximate shape of the probabilistic density function of the flux can be obtained.

**4. Results**

**4.1. Result of the Single-Value Prediction by the FFN With a Scaling Transformation Layer**

The relativistic electron fluxes (>2 MeV) are predicted 1 day in advance with a total of 85 inputs, which consist of five consecutive daily averaged data of 17 input variables. To evaluate the model thoroughly, the time interval is divided into seven continuous periods with each period corresponding to 1 year approximately. For each year, the other 6-year data are used to train the model and predict the relativistic electron flux of that year subsequently. The performance of the model is evaluated for each year as well as the whole period. The results of the predictions in seven years are shown in Figure 5.

It shows that the performances of the model are quite different in different years. The Years 2015 and 2016 show the best performance, with the PE getting the largest in 2015 and LC getting the largest in 2016. Though the RMSE gets the least in 2014, the PE and LC in 2014 show relatively poor performance, which may be due to that the mean flux value of 2014 is much lower than the mean flux value of the whole period. The Years 2012 and 2017 show relatively better results than the Years 2013 and 2014, and the Year 2011 shows the poorest result. On average, the RMSE, PE, and LC reach 0.305, 0.832, and 0.916 during the periods of 2011 to 2017.

One of the causes that the model performances are different in different years may be due to the differences of the flux data distributions in different years. It seems that if the flux data distribution of a specific year (presented by the mean value and variance) deviates from the flux data distribution of the whole period more, the poorer the performance of the model will be in this year. The flux data distributions of 2011, 2013, and 2014 are more inconsistent with the flux data distribution of the whole period, with much lower mean flux values and variances, which result in relatively worse performances than the other years. If the larger difference there is between the flux data distribution of a specific year and the whole period, the
larger inconsistency there is between the data distribution of the training set (other 6-year data) and the test set (1-year data), which may cause a poorer result. The solar cycle may also have an influence on the performance of the model. For the Years 2011 and 2013, which are two special years with two peaks of sunspot numbers, the performances are worse than other years. Qian et al. (2019) pointed out the nonstationary nature of the relativistic electron flux time series, which causes it to be irregular and difficult to predict. It is consistent with our result, as the mean flux values are quite different in different years.

Furthermore, we compare our model with several other models, as shown in Table 2. We list the performance results of the persistence model and the multiple linear regression model (linear model) in the same period as the baseline models. The results of the empirical mode decomposition model (Qian et al., 2019), the nonlinear autoregressive moving average with exogenous inputs (Boynton et al., 2015), the multivariate autoregressive model plus Kalman filter (Sakaguchi et al., 2013), and the support vector machine (Wang & Shi, 2012) are also listed in Table 2. It shows that our FFN model performs well.

### 4.2. Result of the Single-Quantile Prediction

We combine the quantile regression method with the FFN by adopting the loss function of the quantile regression to predict the quantiles of relativistic electron flux. A quantile is predicted by a specific neural network. The Year 2011 is selected to present the result, as shown in Figure 6. Two periods of 2011 are selected to show the prediction result in detail. It shows that the predicted values of high quantiles are usually higher than the observations, while the predicted values of low quantiles are usually lower than the observations, which is consistent with the definition of the quantiles. For the \( \tau \)th quantile, the predicted frequency that

### Table 2

Comparison of Our FFN Model With Other Models

| Models      | FFN   | Persistence model | Linear model | EMD (Qian et al., 2019) | NARMAX (Boynton et al., 2015) | AR-KLM (Sakaguchi et al., 2013) | SVM (Wang & Shi, 2012) |
|-------------|-------|-------------------|--------------|--------------------------|--------------------------------|---------------------------------|------------------------|
| Test time   |       |                   |              |                          |                                |                                 |                        |
| 2011–2017 average | 2001–2006 average | April 2010 to June 2012 | 2007–2008 | 2008                     |
| RMSE        | 0.305 | 0.445             | 0.353        | —                        | —                              | —                               | —                      |
| PE          | 0.832 | 0.647             | 0.777        | 0.80                     | 0.786                          | 0.82                            | 0.71                   |
| LC          | 0.916 | 0.824             | 0.885        | —                        | 0.894                          | —                               | 0.85                   |

---

*Figure 5.* The single-value prediction result of the FFN. The unit is \( \log_{10} \text{cm}^{-2} \text{day}^{-1} \text{sr}^{-1} \). (a) Plots of the relativistic electron flux of the seven years and the flux data distribution of the whole period. The mean value and variance of the flux in each year are shown in the titles of the subplots. (b) Scatter plots of the observations and the predictions of each year.
the flux is less than the quantile $\tau$. The number of the “correct predictions,” which means the observed flux is less than the predicted quantile, and the corresponding observed frequency, which is the ratio between the “correct predictions” and the total data, are shown in parentheses in Figure 6a. It shows that the single-quantile prediction model performs well in 2011. Moreover, in order to thoroughly evaluate the reliability of the quantile regression method, the reliability diagram is plotted in Figure 7. The performances of the model in the seven years are presented by the consistency between the predicted frequencies and the corresponding observed frequencies that the flux is less than the quantile. The black diagonal line indicates perfect reliability. The closer the black line is to the colored line, the more accurate the quantile regression is. The results of different years are quite different. The model performs well in the Years 2011, 2012, 2013, and 2015, with the mean absolute frequency errors being 0.010, 0.014, 0.027, and 0.013, respectively. However, the predictions of 2014 tend to overestimate the flux quantiles, while the predictions of 2016 and 2017 tend to underestimate the flux quantiles. It may be due to that the mean flux value of the Year 2014 is the smallest in all seven years, while the mean flux values of the Years 2016 and 2017 are much larger than other years. The mean absolute frequency error of the seven years is 0.050.

4.3. Result of the Multiple-Quantiles Prediction

The reliability diagrams of the combined neural network for predicting three quantiles are shown in Figure 8, where the Years 2011 and 2014 are selected to show the results. The red dashed line with red triangles is the result of the single-quantile prediction. The results of the three outputs of a specific neural network are shown by the three dots of the same color. Different colored dots correspond to different neural networks. The first float number in the parentheses is the
mean absolute error between predicted frequencies and observed frequencies over \((0.5 - \alpha)\)th, 0.5th, and \((0.5 + \alpha)\)th quantiles. The second float number in the parentheses is the frequency that the observed flux is in the interval \([y_{0.5 - \alpha}, y_{0.5 + \alpha}]\). It shows that there are no great changes in the model performance compared with the single-quantile prediction model. More specifically, if the interval between the high quantile \(y_{0.5 + \alpha}\) and the low quantile \(y_{0.5 - \alpha}\) is large, the model will perform well. However, if the interval is very small, such as \(\alpha = 0.05\) in this case, the prediction performance may fall quite a lot, as Figure 8a shows. We think the reason is as follows. If the interval between two quantiles were small and there had no penalty term, the two quantiles would be more probably in the wrong order. Hence, the penalty term in the loss function plays a more important role to ensure the order consistency of different quantiles with small intervals, and this will influence the prediction performance of the combined neural network.

The reliability diagram of the nine-quantiles prediction model for predicting 0.1th, 0.2th, ..., 0.9th quantiles is shown in Figure 9a. It shows similar characteristics of the reliability diagram of the single-quantile prediction (Figure 7), with the mean absolute frequency error being 0.057. From the nine quantiles, the approximate shape of the probabilistic density function of the flux can be obtained. The first 5 days of the Year 2011 are selected to show the predicted probability density functions of the flux. Most of them show unimodal structures. Moreover, we combine all the predicted probability density functions of the flux in a year to get the accumulated flux data distribution and compare it with the observed flux data distribution. In this way, we use the nine predicted quantiles to represent the probability density function of a single day, which can be seen as samplings of the probability density function. Then we combine all the predicted quantiles and plot the normalized histogram to show the accumulated flux data distribution. The Years 2011 and 2014 are selected to show the results, as shown in Figures 9c and 9d. It shows that the accumulated flux data distribution of the quantile is very close to the observed flux data distribution in the Year 2011, while the accumulated flux data distribution shows a deviation from the observed flux data distribution in the Year 2014. In conclusion, the combined neural networks for predicting multiple quantiles performs well.

5. Conclusion

The neural networks are powerful data-based models that can capture and represent complex input/output correlations. In this study, a FFN with a scaling transformation layer is thus designed to predict relativistic electron flux at GEO 1 day in advance during the periods of 2011–2017. The explanatory variables consist of six solar parameters (T, N, V, P, B, and Ey), four geomagnetic indices (Kp, Dst, ap, and AE), and electron
fluxes of seven energy channels (tens keV to MeV) measured at GEO. The model performs well, with the average RMSE, PE, and LC reach 0.305, 0.832, and 0.916, respectively.

From the probabilistic perspective, the relativistic electron flux is not a single value but a random variable. The single-value prediction which adopts the mean square error loss function is to predict the expectation of the flux. Correspondingly, the quantiles of relativistic electron flux can be obtained by the combination of the FFN and the quantile regression method. By adopting a special loss function, we combine the FFN and the quantile regression method to predict the quantiles of relativistic electron flux, and it performs well. The combination of the neural network and the quantile regression is a powerful method. By designing multiple-quantiles prediction models for predicting multiple quantiles at the same time, the flux ranges of relativistic electrons can be predicted with the corresponding probabilities, which is an advantage over the single-value prediction. A nine-quantiles prediction model is designed to predict nine quantiles at the same time, which can further predict the probability density function of the flux.

![Figure 9. Results of the nine-quantiles prediction model. (a) Reliability diagram of the nine-quantiles prediction model. (b) The predicted probability density function of flux in the first to fifth day of 2011. (c) The accumulated flux data distribution of the predicted quantiles and the observed flux data distribution of the Year 2011. (d) The accumulated flux data distribution and the observed flux data distribution of the Year 2014.](image-url)
Acknowledgments
This research is supported by the National Natural Science Foundation of China (Grants 41731068, 41674164, and 41974391), China National Space Administration project D020303 and Ministry of Science and Technology project 2020YFE002010. We are grateful to National Geophysical Data Center, NOAA, for providing the GOES satellite data. We also thank the Space Physics Data Facility (SPDF) at NASA's Goddard Space Flight Center (https://omniweb.gsfc.nasa.gov/) and the World Data Center (WDC) for Geomagnetism, Kyoto, for providing the solar wind parameters and geomagnetic indices.

References
Boydton, R. J., Bakhitkin, M. A., & Billings, S. A. (2015). Online NARMAX model for electron fluxes at GEO. Annales Geophysicae, 33(3), 405–411.
Boydton, R. J., Bakhitkin, M. A., Billings, S. A., Reeves, G. D., Ganushkina, N., Gedalin, M., et al. (2013). The analysis of electron fluxes at geosynchronous orbit employing a NARMAX approach. Journal of Geophysical Research: Space Physics, 118, 1500–1513. https://doi.org/10.1002/jgra.50192
Camporeale, E. (2019). The challenge of machine learning in space weather: Nowcasting and forecasting. Space Weather, 17, 1166–1207. https://doi.org/10.1029/2018SW00200
Chen, Y., Reeves, G. D., Cunningham, G. S., Redmon, R. J., & Henderson, M. G. (2016). Forecasting and remote sensing outer belt relativistic electrons from low Earth orbit. Geophysical Research Letters, 43, 1031–1038. https://doi.org/10.1002/2015GL067481
Fukata, M., Taguchi, S., Okuzawa, T., & Obara, T. (2002). Neural network prediction of relativistic electrons at geosynchronous orbit during the storm recovery phase. Effects of recurring substorms. Annales Geophysicae, 20(7), 947–951.
Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neurons. Journal of Machine Learning Research, 15, 315–324.
Guo, C., Xue, B., & Lin, Z. (2013). Approach for predicting the energetic electron flux in geosynchronous earth orbit. Chinese Journal of Space Science, 33(4), 418–426.
Hallock, K. F., & Koenker, R. W. (2001). Quantile regression. Journal of Economic Perspectives, 15(4), 143–156.
Horne, R. B., Glauer, S. A., Meredith, N. P., Boscher, D., Maget, V., Heynderickx, D., & Pitchford, D. (2013). Space weather impacts on satellites and forecasting the Earth's electron radiation belts with SPACECAST. Space Weather, 11, 169–186. https://doi.org/10.1029/2012SW000203
Horne, R. B., & Thorne, R. M. (1998). Potential waves for relativistic electron scattering and stochastic acceleration during magnetic storms. Geophysical Research Letters, 25(15), 3011–3014. https://doi.org/10.1029/98GL01002
Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks, 2(5), 359–366.
Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S. (2017). Self-normalizing neural networks. In Advances in neural information processing systems (pp. 971–980).
Koenker, R., & Bassett, G. Jr. (1978). Regression quantiles. Econometrica: Journal of the Econometric Society, 46(1), 33–50.
Koon, H. C., & Gorney, D. J. (1991). A neural network model of the relativistic electron flux at geosynchronous orbit. Journal of Geophysical Research, 96(A4), 5549–5559. https://doi.org/10.1029/90JA02380
Lai, S. T., Cahoy, K., Lohmeyer, W., Carlton, A., Aniceto, R., & Minow, J. (2018). Deep dielectric charging and spacecraft anomalies. In N. Buzulukova (Ed.), Extreme events in geospace: Origins, predictability, and consequences (pp. 419–432). US: Elsevier Inc.
Li, X. (2004). Variations of 0.7–6.0 MeV electron at geosynchronous orbit as a function of solar wind. Space Weather, 2, S03096. https://doi.org/10.1029/2003SW000017
Li, X., Temerin, M., Baker, D. N., Reeves, G. D., & Larson, D. (2001). Quantitative prediction of radiation belt electron at geostationary orbit based on solar wind measurements. Geophysical Research Letters, 28(9), 1887–1890. https://doi.org/10.1029/2000GL012681
Ling, A. G., Ginet, G. P., Hilmer, R. V., & Perry, K. L. (2010). A neural network–based geosynchronous relativistic electron flux forecasting model. Space Weather, 8, S09003. https://doi.org/10.1029/2010SW000576
Lytatsky, W., & Khazanov, G. V. (2008). A predictive model for relativistic electrons at geostationary orbit. Geophysical Research Letters, 35, L15108. https://doi.org/10.1029/2008GL034688
Miyoshi, Y., & Katoaka, R. (2008). Probabilistic space weather forecast of the relativistic electron flux enhancement at geosynchronous orbit. Journal of Atmospheric and Solar Terrestrial Physics, 70(2–4), 475–481.
Pilipenko, V., Yagova, N., Romanova, N., & Allen, J. (2006). Statistical relationships between satellite anomalies at geostationary orbit and high-energy particles. Advances in Space Research, 37(6), 1192–1205.
Qian, Y., Yang, J., Zhang, H., Shen, C., & Wu, Y. (2019). An hourly prediction model of relativistic electrons based on empirical mode decomposition. Space Weather, 17, https://doi.org/10.1029/2018SW002076
Roth, I., Tenerin, M., & Hudson, M. K. (1999). Resonant enhancement of relativistic electron fluxes during geomagnetically active periods. Annales Geophysicae, 17(5), 631–638.
Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323, 533–536.
Sakaguchi, K., Miyoshi, Y., Saito, S., Nagatsuma, T., Seki, K., & Murata, K. T. (2013). Relativistic electron flux forecast at geostationary orbit using Kalman filter based on multivariate autoregressive model. Space Weather, 11, 79–89. https://doi.org/10.1002/swe.20202
Shin, D. K., Lee, D. Y., Kim, K. C., Hwang, J., & Kim, J. (2016). Artificial neural network prediction model for geosynchronous electron fluxes: Dependence on satellite position and particle energy. Space Weather, 14, 313–321. https://doi.org/10.1002/2015SW001359
Summers, D., & Ma, C. Y. (2000). A model for generating relativistic electrons in the Earth's inner magnetosphere based on gyrosorant wave-particle interactions. Journal of Geophysical Research, 105(A2), 2625–2639. https://doi.org/10.1029/99JA00444
Summers, D., Thorne, R. M., & Xiao, F. (1998). Relativistic theory of wave-particle resonant diffusion with application to electron acceleration in the magnetosphere. Journal of Geophysical Research, 103, 20,487–20,500. https://doi.org/10.1029/98JA01740
Tsibhirini, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B: Methodological, 58(1), 267–288.
Tsutai, A., Mitsui, C., & Nagai, T. (1999). Geosynchronous electron environment with in situ magnetic field measurements. Earth, Planets and Space, 51(3), 210–233.
Turner, D. L., & Li, X. (2008). Quantitative forecast of relativistic electron flux at geosynchronous orbit based on low-energy electron flux. Space Weather, 6, S05005. https://doi.org/10.1029/2007SW000354
Wang, R., & Shi, L. (2012). Study on the forecasting method of relativistic electron flux at geostationary orbit based on support vector machine. Chinese Journal of Space Science, 32(3), 354–361.
Wei, L., Zhong, Q., Lin, R., Wang, J., Liu, S., & Cao, Y. (2018). Quantitative prediction of high-energy electron integral flux at geostationary orbit based on deep learning. Space Weather, 16, 903–916. https://doi.org/10.1029/2018SW001829
Wrenn, G. L., Rodgers, D. J., & Ryden, K. A. (2002). A solar cycle of spacecraft anomalies due to internal charging. Annales Geophysicae, 20(7), 953–956.