Deep Temporal Convolutional Networks for F10.7 Radiation Flux Short-Term Forecasting

Luyao Wang[1,2] Hua Zhang[3,4,5] Xiaoxin Zhang[4,5] Guangshuai Peng[3] Zheng Li[2,3] Xiaojun Xu[2]

[1]School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing, China

[2]State Key Laboratory of Lunar and Planetary Science, Macau University of Science and Technology, Macau, China

[3]Institute of Space Weather, Nanjing University of Information Science & Technology, Nanjing, China

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

[5]Innovation Center for Feng Yun Meteorological Satellite (FYSIC), China Meteorological Administration, Beijing, China

Correspondence to: Hua Zhang（289534957@qq.com）

Abstract. F10.7, the solar flux at a wavelength of 10.7 cm (F10.7), is often used as an important parameter input in various space weather models and is also a key parameter for measuring the strength of solar activity levels. Therefore, it is valuable to study and forecast F10.7. In this paper, the temporal convolutional network (TCN) approach in deep learning is used to predict the daily value of F10.7. The F10.7 series from 1957 to 2019 are used. The data during 1957–1995 are adopted as the training dataset, the data during 1996–2008 (solar cycle 23) are adopted as the validation dataset, and the data during 2009–2019 (solar cycle 24) are adopted as the test dataset. The leave-one-out method is used to group the data set for multiple validations. The prediction results for 1-3 days ahead during solar cycle 24 have a high correlation coefficient (R) of 0.98 and a root mean square error (RMSE) of only 5.04~5.18 sfu. The overall accuracy of the TCN forecasts is better than the autoregressive (AR) model (it only takes past values of the F10.7 index as inputs) and the results of the US Space Weather Prediction Center (SWPC) forecasts, especially for 2 and 3 days ahead. In addition, the TCN model is slightly better than other neural network models like back propagation neural network (BP) and long short term memory network (LSTM) in terms of the solar radiation flux F10.7 forecast. The TCN model predicted F10.7 with a lower root mean square error, a higher correlation coefficient, and a better overall model prediction.

1 Introduction

Solar activity has a significant impact on the Earth's climate, electromagnetic fields and communication systems, among other things. F10.7 (2800 MHz, 10.7 cm solar flux) is a typical parameter for characterizing solar activity levels, representing the cyclical variability of solar activity (Tapping,2013). The F10.7 index is an important parameter for predicting atmospheric density for spacecraft orbits and ionospheric forecasts affecting communication. For example, F10.7 is used as a control parameter in ionospheric models to calculate the variation of radio signal properties (Ortikov et al.,2003). F10.7 is also widely used for satellite, navigation, communication, and terrestrial climate (Huang et al., 2009; Yaya et al., 2017). Therefore, accurate forecasting of F10.7 is not only of great value for the conduct of application but is also of comparative importance in the scientific study of space weather forecasting (Katsavrias et al.,2021; Simms et al.,2023).

F10.7 has a clear periodicity, e.g. 27 days, 11 years. But the cycles are not simply repetitive, but have similar but different fluctuations, so the core of the F10.7 prediction problem for time series data is to uncover the potential patterns of historical data and predict the future data as far as possible (Lampropoulos et al., 2016). The F10.7 index forecast model is based on a time series model. Many researchers have used different methods to build predictive models for F10.7. Mordvinov et al. (1986) used a multiplicative autoregressive model to forecast the monthly mean of F10.7, but the model had a large error in predicting the monthly mean F10.7.Warren et al. (2017) built optimized independent models for each forecast date, and the results showed that this approach typically predicted better than autoregressive methods. Zhong et al. (2010) utilized the singular spectrum analysis signal processing technique to predict the F10.7 index of solar activity for the next 27 days. The research result indicated that the method performed well in predicting the periodic variations of the F10.7 index. Henney et al. (2012) predicted F10.7 using the global solar magnetic field generated by the energy transport model, with a Pearson correlation coefficient of 0.97 for 1-day ahead. Liu et al. (2018) applied two models by Yeates (Yeates et al., 2007) and Worden (Worden & Harvey, 2000) to predict short-term variability in F10.7. During low levels of solar activity, the predicted values of the model were closer to the observed values.

With the rapid development of machine learning and neural networks. Researchers are increasingly intrigued by the powerful learning capabilities of machine learning and neural networks, using them to study variations in solar activity. The support vector machine regression method was used by Wang et al. (2009) to predict daily values of solar activity F10.7. Xiao et al. (2017) used back propagation neural network (BP) to forecast the daily mean index F10.7 of solar activity for short-term prediction. The results showed that using BP neural networks to predict the solar activity daily index F10.7 was superior to the results of Wang et al(2009). Luo et al. (2020) proposed a method for predicting 10.7 cm radio flux in multiple steps. The method is a combination of the Empirical Mode Decomposition (EMD) and back propagation neural network (BP) to construct an EMD-BP model for predicting F10.7 values. The method significantly reduces the prediction error for high levels of solar activity compared to support vector machine regression (SVR) and backward propagation neural network (BP). Zhang et al. (2020) proposed a short-term forecast of the solar activity daily mean index F10.7 by a long short-term memory network (LSTM) method. The forecast had a high correlation coefficient (R) of 0.98 and a low root mean square error (RMSE) range of 6.20-6.35 sfu. Although the above recurrent neural network (RNN)-based architecture and its variants achieved good prediction accuracy of F10.7, the training process of a model often spends a significant amount of time and computational memory, and also frequently encounters issues such as gradient explosion or vanishing gradients during network training(Zachary et al.,2015; Yang et al.,2021). To this end, Bai et al. (2018) proposed a neural network called temporal convolutional network (TCN), in which long input sequences can be processed as a whole in the TCN. TCN uses convolutional operations for efficient parallel computation. In addition, the back propagation path of TCN is different from the time direction of the sequence, which makes TCN avoid the gradient problem in RNN. Given the above advantages and for the variability characteristics of F10.7 time-series data, this paper introduces machine learning-based TCN-related theories and techniques into the forecasting of F10.7 and compares the results of TCN prediction with other classical models to verify the effectiveness and feasibility in the short-term forecasting.

2 Data and Method

2.1 Data source and Data processing

F10.7 represents the solar radiation flux at a wavelength of 10.7 cm, and the magnitude of this index describes the intensity of solar activity. The 10.7cm solar flux is given in solar flux units (a sfu = 10^{-22}W m^{-2} Hz^{-1}). The 10.7 cm daily solar flux data were obtained from the website of the National Oceanic and Atmospheric Administration. Three flux determinations are made each day. Each 10.7cm Solar Flux measurement is expressed in three values: the observed, adjusted, and URSI Series D values (absolute values).The observed value is the number measured by the solar radio telescope. This is modulated by two quantities: the level of solar activity and the changing distance between the Earth and Sun. Since it is a measure of the emissions due to solar activity hitting the Earth, this is the quantity to use when terrestrial phenomena are being studied (Tapping, 1987). When studying the Sun, it is undesirable to have the annual modulation of the 10.7cm Solar Flux caused by the changing distance between the Earth and Sun. However, during the ephemeris calculations required for the solar flux monitors to accurately acquire and track the Sun, one of the by products obtained is the distance between the Sun and the Earth. Therefore, we generate an additional value called the adjusted value, which takes into account the variations in the Earth-Sun distance and represents the average distance. Absolute measurements of flux density are quite difficult. Astronomers attempt to match the solar flux density data at various frequencies with a frequency spectrum by applying a scale factor. By combining each wavelength with the calibrated spectrum, a series of D Flux is obtained, where D Flux equals 0.9 multiplied by the adjusted flux(Tanaka et al.,1973).

Between March and October measurements are made at 1700, 2000 (local noon) and 2300UT. However, the combination of location in a mountain valley and a relatively high latitude makes it impossible to maintain these times during the rest of the year. Consequently, from November to February, the flux determination times are changed at 1800, 2000, and 2200, so that the Sun is high enough above the horizon for a good measurement to be made .Therefore, we chose the adjusted flux value of F10.7 measured at 8:00 p.m. UT (local noon). The data during 1957–1995 are adopted as the training dataset, the data during 1996–2008(solar cycle 23) are adopted as validation set and the data during 2009–2019(solar cycle 24) are adopted as test set. Figure 1 shows the data. The black line represents the training dataset, the red line represents the validation dataset and the blue line is the testing dataset.

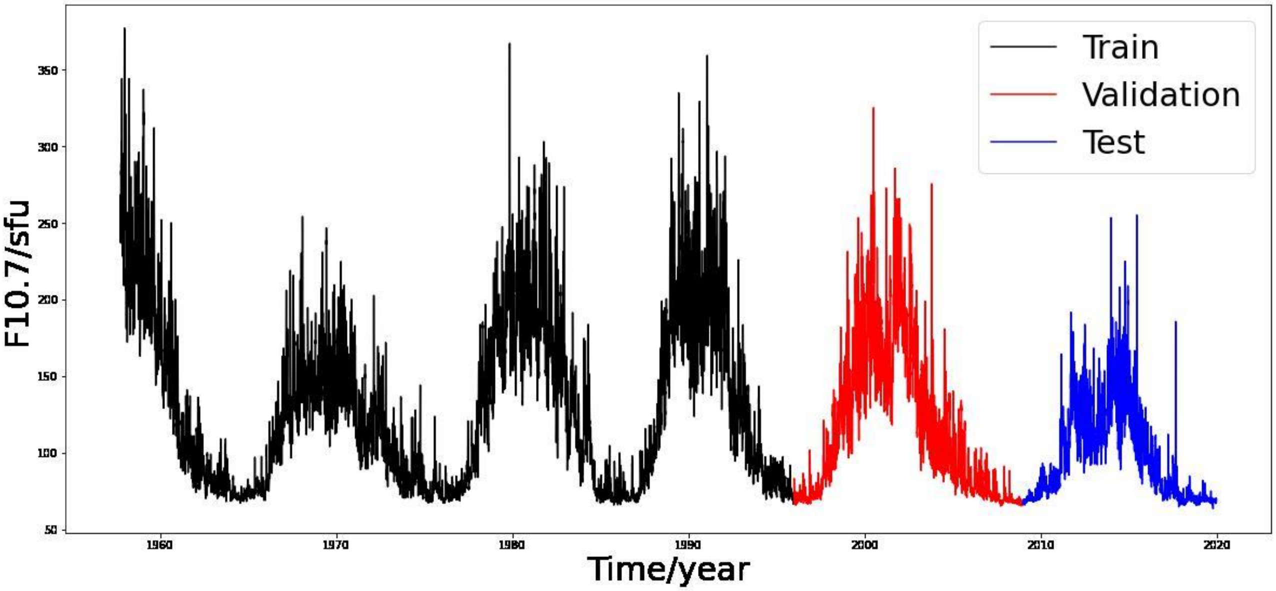

Figure 1: The daily values of F10.7 index from 1957 to 2019.Where the black line represents the training set(solar cycles 19-22), red represents the validation set(solar cycle 23), and blue represents the test set(solar cycle 24)

In this paper, the hardware environment used for the solar radiation flux F10.7 experiment is NVIDIA GeForce 940MX, CPU is Inter(R) Core(TM) i5-6200. We build a model by Python and utilize some efficient frameworks including Pandas,Matplotlib, Tensorflow,and Sklearn. Pandas is a powerful data analysis library that provides several methods for processing and analysing the parameters of the solar flux F10.7, such as selective sorting, merging and aggregating, etc. Matplotlib is a plotting Python library that provides a rich set of customisation options for this paper to visualise the predicted results of the solar flux F10.7 as well as to analyse the related results. Sklearn is an open-source, third-party library for machine learning model training and big data mining that provides a unified interface for many machine learning algorithms and a number of tools for evaluating model performance and tuning hyperparameters.TensorFlow is also an open-source machine learning library for building, training, and deploying a variety of model types, including regression, classification, convolutional neural network and recurrent neural network construction, among others. In this paper, the network construction, training, parameter tuning, and evaluation of the prediction model for solar radiation flux F10.7 are based on the above two machine learning libraries.

2.3 Method

TCN was proposed by Bai et al. (2018). Some scholars have demonstrated that TCN not only achieves better performance but also reduces the computational cost for training, compared to that of RNN (Lea et al.,2016; Bai et al.,2018; Dieleman et al.,2018). TCN combines both RNN and convolutional neural network (CNN) architectures and is a convolutional neural network variant designed to handle time series modelling problems. TCN is well adapted to the temporal nature of the data by using both causal and extended convolutional structures to extract feature information. The convolutions in TCN are causal, meaning there is no information leakage from future time steps. This distinguishes TCN from other recurrent neural networks such as LSTM, GRU, which require gate mechanisms. As a result, TCN achieves higher accuracy and longer memory without the need for gate mechanisms. Long input sequences can be processed as a whole in TCN. TCN does not have the advantages of gradient disappearance and gradient explosion problems. Here, TCN is introduced to model the prediction of F10.7.

For the prediction of a univariate time series, the TCN model takes lagged observations of the time series as inputs and predicts future F10.7 sequence values as outputs. Each set of input patterns consists of moving a fixed length window in the time series. The principle of forecasting is represented in Fig.2. The original F10.7 data is lengthy, and during training, a continuous subsequence needs to be inputted. The output and input lengths of the temporal convolutional network (TCN) are equal, meaning the length of the output sequence generated by the TCN is equal to the specified input length. To meet the prediction requirements, the specific number of steps to forecast (referred to as output length) should not exceed input length, allowing for partial overlap between the input and output sequences.

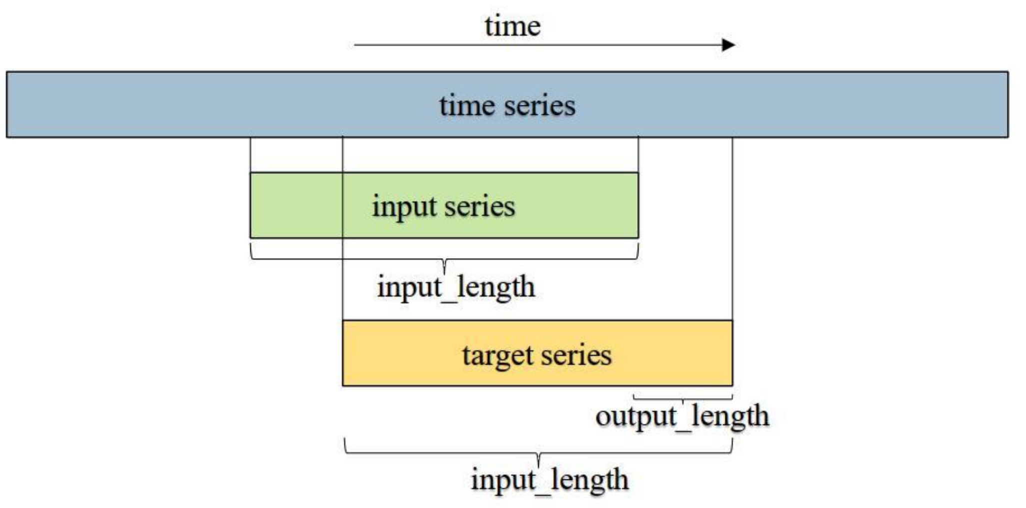

Figure 2: Diagram of F10.7 sequence data prediction. The blue part represents the original sequence, the green part represents the input subsequence, and the orange part represents the overlap and the actual predicted lengths

Supposed the input of F10.7 is $x = (x_0, x_1, \ldots, x_T)$, the desired output sequence is $y = (y_0, y_1, \ldots, y_T)$, where the two sequences x, y satisfy the causal relationship.The input $x_0, x_1, \ldots, x_{t-1}$ observed at the previous moment be used to predict the output y_t at moment t. The modelling objective of the TCN network is to generate any hidden function mapping, which means that the prediction of the F10.7 sequence can be represented as:

$$\hat{y}_1, \ldots, \hat{y}_{T+1} = f(x_0, x_1, \ldots, x_i, \ldots, x_T) \tag{1}$$

where x_i and $\hat{y}_i$ are the observed and predicted values of F10.7 at time i, respectively, and f is the mapping of the function trained by the TCN network.

TCN is one of the algorithms developed on the basis of convolutional neural network (CNN). That uses a one-dimensional convolutional network, consisting of an inflated causal convolution and a residual module.

One-dimensional convolution operates on time series and extracts various features, but as the length of the time series grows, a regular convolutional network requires more convolutional layers to receive longer sequences. Extended convolution, on the other hand, improves on convolution by allowing interval sampling of the input for convolution with a number of layers L and a convolution kernel of size k with an acceptance domain of:

$$r = 2^{(L-1)}k \tag{2}$$

The causal extended convolution operation F for element s in a time series is defined as:

$$F(s) = (x * f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \tag{3}$$

where: $x = (x_0, x_1, \ldots, x_T)$ is the input vector, d is the expansion factor, $*$ is the causal expansion convolution operator, f is the convolution kernel vector, k is the convolution kernel size, and $s - d \cdot i$ indicates the past direction of the input.

The expanding causal convolution for a convolution kernel size k of 3 is illustrated in Fig. 3. where the output y_1 at moment t is determined by the current input as well as the previous inputs. That is $x_0, x_1, \ldots, x_n$. it shows that the predicted output is not affected by future information and therefore avoids information leakage. In addition, the introduction of the expansion factor d to the input of the convolutional layer matrix is sampled at intervals. In the first hidden layer, the sampling interval rate $d = 1$ which represents each point of the input is sampled. In the second hidden layer, the sampling interval rate $d = 2$ i.e., every two points are taken, ignoring one neuron. At higher layers using the d grows exponentially, thus allowing for fewer layers to achieve a larger receptive field with fewer layers.The expanding causal convolution can be adjusted by varying the number of layers, perceptual field size, convolution kernel size, and expansion coefficient. This helps to address the challenge in CNNs where the length of temporal modelling is limited by the size of the convolution kernel. Compared to traditional neural networks like LSTM and BP, TCN overcomes issues such as gradient vanishing and exploding. At the same time, TCN possesses advantages such as lower memory consumption, stable gradient, improved parallelism, and flexible perceptual field.

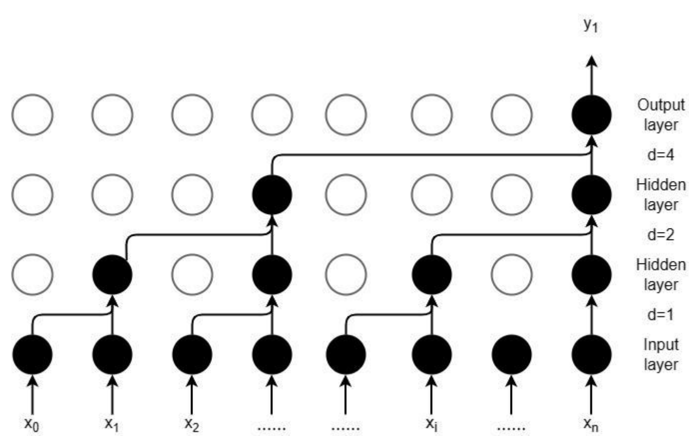

Figure 3: Expansion causal convolutional structure diagram

The structure of the residual module in the TCN is shown in Fig.4. Using ReLU as an activation function. To avoid the problem of gradient explosion, add the weight normalisation layer. To avoid overfitting, a dropout layer is added for regularisation. The residual links allow the network to pass information across layers, thus avoiding information loss due to too many layers. Residual convolution is introduced for layer hopping and 1 × 1 convolution is performed to ensure that the input and output remain consistent.

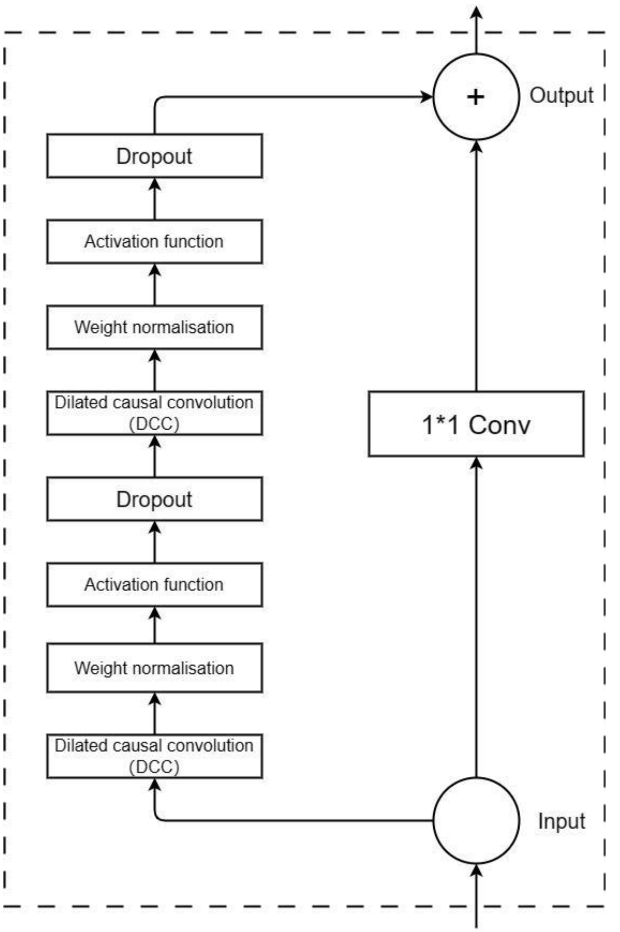

Figure 4: Expansion causal convolutional structure diagram

2.4 Selection of training parameters

A key component of the machine learning model training process is called the loss function, which gives direction to the optimization of the model by measuring the difference between the model output $\hat{y}$ and the observation y. The smaller the loss function, and the better the robustness of the model. The L1 norm loss function is extensively utilized in deep learning tasks (Zhao et al.,2017). It possesses a notable advantage of being insensitive to outliers and exceptional values, consequently avoiding the gradient explosion issue. Moreover, The loss function provides a more robust solution by offering stability.

Therefore, the L1 loss function is chosen to construct the loss function for the predicted and observed values of the F10.7 sequence. The function is defined as:

$$L(\hat{y}, y) = \sum_{i=0}^{n} |\hat{y}_i - y_i| \tag{4}$$

where $\hat{y}_i$ is the predicted value of F10.7 at moment i, and y_i is the observed value of F10.7 at moment i.

To build the TCN model that is not merely a linear regression model, it is essential to introduce non-linearity by adding a ReLU activation function at the top of the convolutional layers. The function is defined as:

$$f(x) = \max(0, x) \tag{5}$$

where: $x = (x_0, x_1, \dots, x_T)$ is the input vector.

To counteract the problem of gradient explosion, weights are normalized at each convolutional layer. To prevent overfitting, each convolutional layer is followed by a dropout for regularization. After several training sessions, the optimal parameters for model training are shown in Table 1:

Table 1. Training parameters of the TCN model

Parameter	Value	Parameter explaination
batch_size	None	Batch size
time_steps	20	Step length
epochs	30	Number of training sessions
input_dim	1	Dimension
input_shape	20	Input shape size
tcn_layer.receptive_field	/	The perceptual wildness of the convolutional layer
Dense(1)	/	Fully connected layer
optimizer	adam	Optimizer
loss	L1	Loss function
activation=	relu	Activation function
filters	64	Number of channels for the input and output of the convolution kernel
kernel_size	3	Convolution kernel size
stacks	1	Determining the depth of the network
dilations	{1,2,4,8,16,32}	Expansion coefficient
padding	causal	Fill factor

2.5. Forecast evaluation criteria

In order to quantify the forecast performance of the model. We chose five evaluation metrics. The chosen performance metrics include the mean absolute error (MAE), the mean absolute percentage error(MAPE) and the root mean square error(RMSE) for accuracy, the correlation coefficient(R) for association, and the error(σ) for bias. Five commonly used model evaluation metrics for assessing predictive performance (Liemohn et al.,2021).

$$MAE = \frac{1}{\mathrm{N}}\sum_{i=1}^{N} | f_i - F_i | \tag{6}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - F_i)^2} \tag{7}$$

$$MAPE = \frac{1}{N}\frac{\sum_{i=1}^{N} | f_i - F_i |}{f_i} \tag{8}$$

$$R = \frac{\sum_{i=1}^{N}(f_i - \bar{f})(F_i - \bar{F})}{\sqrt{\sum_{i=1}^{N}(f_i - \bar{f})^2}\sqrt{\sum_{i=1}^{N}(F_i - \bar{F})^2}} \tag{9}$$

$$\sigma = f_i - F_i \tag{10}$$

Where MAE denotes mean absolute error, MAPE denotes mean absolute percentage error, RMSE denotes root mean square error, R denotes linear correlation coefficient, N denotes number of samples, f_i denotes forecast and F_i denotes observation, $\bar{f}$ is the mean of f_i, and $\bar{F}$ the average of F_i. Each indicator evaluates the model in a different perspective. Among them, MAE represents the average absolute error between predicted values and actual values. RMSE represents the root mean square error between predicted values and actual values. R represents the degree of trend fitting between predicted values and actual values. σ represents the error between predicted values and actual values. Therefore, the smaller the MAE , MAPE, and RMSE and the larger the R, the better the model prediction.

3. Results and Discussions

The TCN model is used to predict the values of F10.7 for 1-3 days ahead. Table 2 shows the evaluation metrics of TCN model predictions compared to observations for different years of the 24 solar cycle. The table represents the performance of the TCN model in different years. In Table 2, it can be seen that the TCN model predicts F10.7 with a root mean square error (RMSE) ranging from 1 to 9 sfu for 1-day ahead, and an average absolute error (MAE) ranging from 0 to 6 sfu. The highest correlation coefficient reaches up to 0.98. For 2 and 3 days ahead, the RMSE ranges from 1 to 9 sfu, the MAE ranges from 1 to 6 sfu, and the highest correlation coefficient remains at 0.98. Irrespective of the lead time, be it one, two, or three days, the TCN model demonstrates consistent performance with relatively small ranges of root mean square error and mean absolute error, accompanied by a consistently high correlation coefficient. The results demonstrate the stability of the TCN model. However, the magnitude of prediction errors for 1-3 days ahead forecasts varies across different years. For example, the RMSE for a 1-day ahead forecast is 1.09 sfu in 2009, while its value is 8.88 sfu in 2014. Li et al. (2023) defined the years in which the mean value of F10.7 is greater than 110sfu as high solar activity, and the years in which the mean value is less than 110sfu as low solar activity. In this paper, the annual average of F10.7 from 2011 to 2015 is greater than 110sfu, so the years from 2011 to 2015 are called high solar activity years and the remaining years are called low solar activity years. Table 2 shows that solar activity has a periodic effect, and the prediction accuracy of the model is negatively correlated with the intensity of solar activity. The magnitude of error is related to the year of high and low solar activity.

Table 2. The prediction errors (MAE, RMSE) and R of the TCN model for the F10.7 data during 2009–2019

Year	1-Day ahead			2-Days ahead			3-Days ahead		
	MAE (sfu)	RMSE (sfu)	R	MAE (sfu)	RMSE (sfu)	R	MAE (sfu)	RMSE (sfu)	R
2009	0.77	1.09	0.9279	1.05	1.29	0.9308	1.20	1.49	0.9274
2010	1.66	2.27	0.9131	1.59	2.14	0.9146	1.82	2.44	0.9093
2011	3.37	5.11	0.9774	3.37	5.11	0.9773	3.47	5.19	0.9771
2012	4.39	6.53	0.9375	4.40	6.60	0.9353	4.43	6.61	0.9363
2013	3.59	4.88	0.9690	3.62	4.96	0.9673	3.58	4.95	0.9677
2014	5.87	8.88	0.9460	5.87	8.96	0.9442	5.65	8.97	0.9451
2015	4.12	8.28	0.9099	4.13	8.25	0.9100	4.00	8.23	0.9107
2016	2.15	2.99	0.9662	2.24	2.99	0.9654	2.20	3.02	0.9662
2017	1.97	4.64	0.9072	2.16	5.51	0.8696	2.23	5.51	0.8778
2018	0.84	1.15	0.9323	1.26	1.50	0.9330	1.12	1.43	0.9292
2019	0.80	1.17	0.9044	1.19	1.52	0.9073	1.15	1.47	0.9048
Total	2.69	5.04	0.9860	2.80	5.16	0.9852	2.81	5.18	0.9854

To further validate the performance of the model, we use the leave-one-out method for cross validation (Aminalragia et al.,2020). We leave iteratively one solar cycle out as a test dataset and rerun the model each time (e.g. keep solar cycle 23 as test dataset and train the model with the remaining solar cycles, then keep solar cycle 22 as test dataset and train the model with the remaining solar cycles, etc.). The results of the tests are shown in the table 3. It can be seen that cycles with stronger solar activity are found to have larger model forecast errors. For cycles with weaker solar activity, the results are better. Solar cycles 20 and 24 have about the same intensity of solar activity and are both weaker. The model forecasts are better, Solar cycles 21 and 22 have about the same intensity of solar activity and are both stronger. The model forecasts are poorer. However, the overall average prediction results do not change much compared to solar cycle 24. The prediction accuracy of the model is negatively correlated with the intensity of solar activity.The results show that the change of prediction accuracy of the model is related to the intensity of solar activity.The F10.7 data has a solar cycle effect.The TCN model does not large affect the final F10.7 forecasts due to specific properties of the data.

Table.3 The prediction errors (MAE, RMSE) and R of the TCN model for the F10.7 data during different solar cycles.

Solar cycle	1-Day ahead			2-Days ahead			3-Days ahead		
	MAE (sfu)	RMSE (sfu)	R	MAE (sfu)	RMSE (sfu)	R	MAE (sfu)	RMSE (sfu)	R
19	4.35	9.03	0.9880	4.29	7.84	0.9908	4.42	8.51	0.9897
20	3.35	5.16	0.9924	3.86	5.76	0.9928	3.40	5.37	0.9926
21	4.59	7.51	0.9921	4.48	7.16	0.9927	4.65	7.45	0.9930
22	4.71	7.89	0.9908	5.36	8.57	0.9908	4.75	8.05	0.9903
23	3.76	6.46	0.9917	4.30	7.01	0.9915	3.91	6.73	0.9912
24	3.03	5.60	0.9846	2.78	5.49	0.9833	3.23	5.52	0.9850
Mean	3.97	6.94	0.9899	4.18	6.97	0.9903	4.06	6.94	0.9903

Figure 5 displays the frequency distribution of the difference between the observed value and the predicted value of the model.

To maintain the compactness of the histogram, differences greater than 15 sfu and smaller than -15 sfu are not displayed.As can be seen from Figure 5 that the prediction differences for 2-day ahead are skewed towards the right. The differences in predictions for 3-days ahead are skewed towards the left. Despite these differences, frequency is maximized when the difference between the observed and predicted values is in the vicinity of zero and most predictions (88.5% of the 1-3 days ahead forecast) were located within ±6sfu of error.

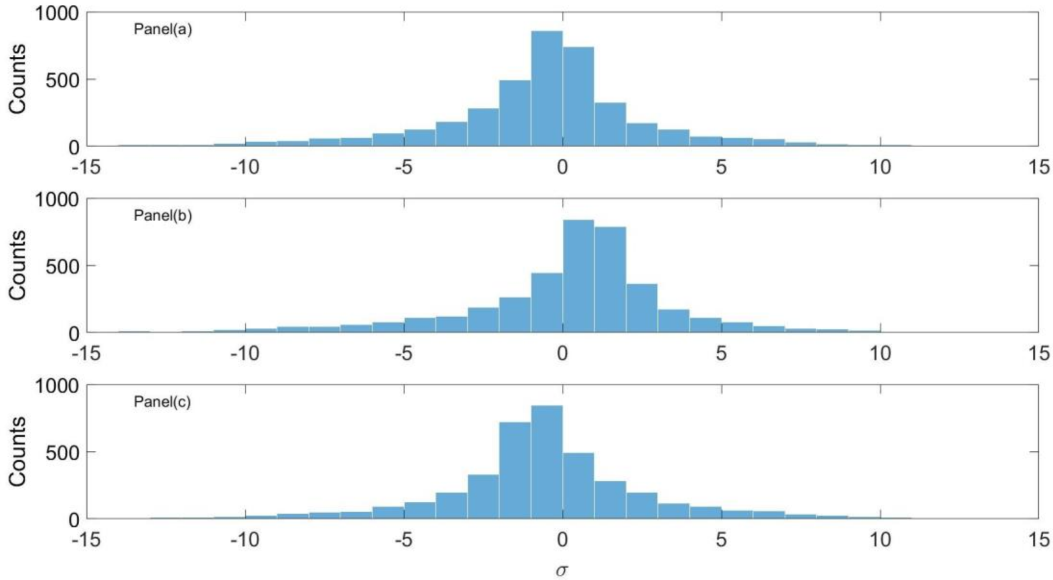

Figure 5: shows the frequency distribution of the difference between the observed values and the model predictions during 2009-2019 (solar cycle 24) for 1-day ahead (Panel (a)), 2-days ahead (Panel (b)), and 3-days ahead (Panel (c))

The high solar activity years of 2013- 2014, and the low solar activity year of 2018 are chosen for comparison in solar cycle 24. We chose predicted values from January 15 to February 15 in 2013, 2014, and 2018 to compare with observed values and improve image representation. Figure 6 shows the predicted effects for solar activity high years in the Panel (a)-(b) and solar activity low year in the panel(c) in solar cycle 24. The black line represents observed values, while the blue dots represent predicted values. As can be seen from Fig. 6, it shows that the TCN model effectively predicts the trend of F10.7 and exhibits good agreement in terms of magnitude between the observed and predicted values for the majority of the time. Especially during the peak of F10.7, the TCN model's predictions align well with the actual values, and it performs exceptionally well during periods of high solar activity.

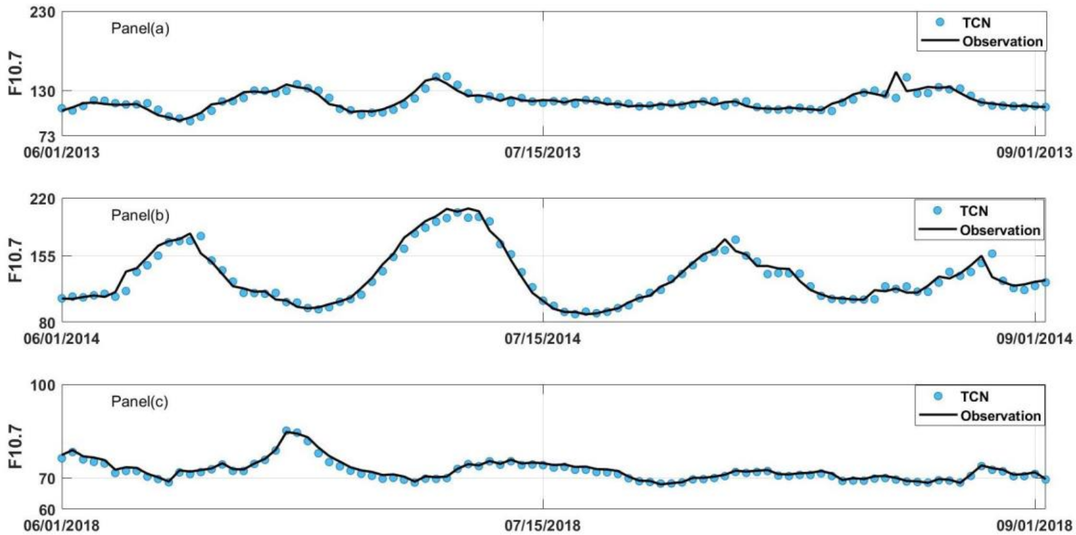

Figure 6: shows the predicted effects for solar activity high years in the Panel(a)-(b) and solar activity low years in the panel(c) for 1-day ahead in solar cycle 24. The black line represents observed values, while the blue dots represent predicted values

To assess the model's effectiveness, we compare the TCN model's forecasting results with those of the SWPC forecast (https://www.swpc.noaa.gov/sites/default/files/images/u30/F10.7%20Solar%20Flux.pdf) and the AR model (Du et al., 2020) for 1-3 days ahead. Furthermore, we compare the predictions with the BP model (Xiao et al., 2017) and LSTM (Zhang et al., 2020) for 3-days ahead.

Figure 7shows the prediction results of the SWPC compared to the TCN model for 1-day ahead in panel (a), 2-days ahead in panel (b) and 3-days ahead in panel(c). The blue bars represent the predicted outcome parameters for SWPC and the yellow bars represent the predicted outcome parameters for the TCN model. Figure 7 shows that the TCN model's predictions are generally better than the forecasts of the SWPC. Compared with F10.7 values for 1-3 days ahead, the TCN model's prediction for 1-day ahead is 0.07 sfu higher than the SWPC forecast only in 2012. While in other years, the TCN model consistently outperformed the SWPC forecast. Particularly for 2 and 3 days ahead predictions, the TCN model's performance is significantly better than the SWPC forecast. The RMSE of TCN is 5.11sfu for 1-day ahead, while the RMSE of the SWPC is 5.61 sfu in 2011. The RMSE of TCN is 0.50 sfu lower than SWPC, representing a relative decrease of 10%. For 2-days ahead prediction, the RMSE of TCN is 5.11sfu, while the SWPC of RMSE is 9.17 sfu in 2012. The RMSE of TCN is approximately 4.06 sfu lower than SWPC, representing a relative decrease of 79%. For 3-days ahead prediction in 2011, the RMSE of TCN is 5.19 sfu, while the RMSE of the SWPC is 11.46 sfu. The RMSE of TCN is approximately 6.27sfu lower than SWPC, representing a relative decrease of 120%. All these show that the TCN model proposed in this paper has better performance relative to the SWPC model.The TCN model is feasible for F10.7 prediction.

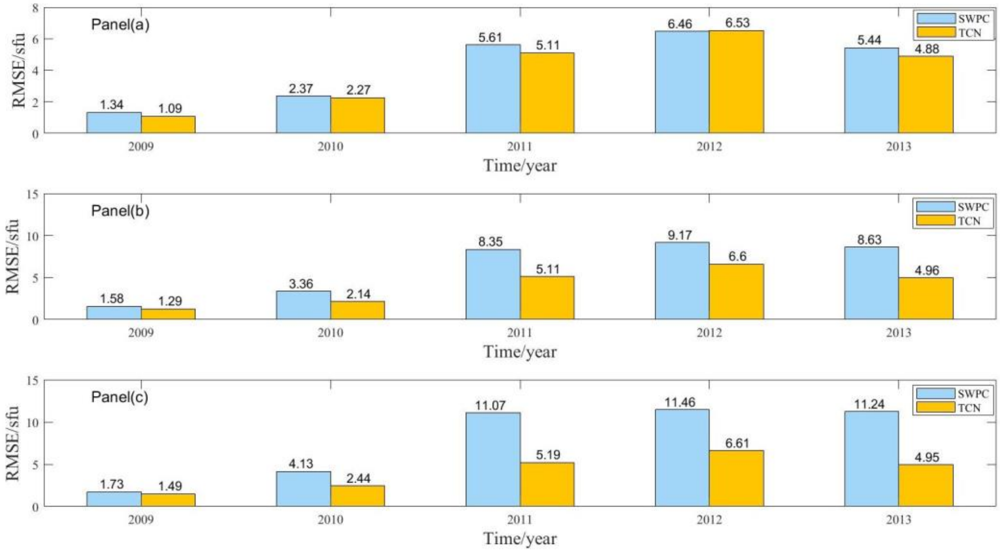

Figure 7: Comparison of the prediction performance of SWPC and TCN for Panel(a) for 1-day ahead, Panel (b) for 2-days ahead, and Panel (c) for 3-days ahead during different years.

Figure 8 shows the prediction results of the AR model compared to the TCN model for 1-day ahead in panel (a), 2-days ahead in panel (b) and 3-days ahead in panel(c). The blue bars represent the predicted outcome parameters for AR, and the yellow bars represent those for the TCN model. As can be seen in Fig.8, the TCN model outperforms the AR model overall in forecasting for 1-3 days ahead. The TCN model only has forecasts that are 0.96sfu and 0.04sfu larger than the AR model pattern for 1-day ahead in 2014 and 2019, respectively. In addition, the TCN model outperforms the AR model in forecasting for both 2 and 3 days ahead. The RMSE of TCN is only 5.19 sfu for predicting outcomes for 3-days ahead in 2011, while the RMSE of AR model is 10.43 sfu. The stability and prediction accuracy of the TCN model in predicting F10.7 is again verified.

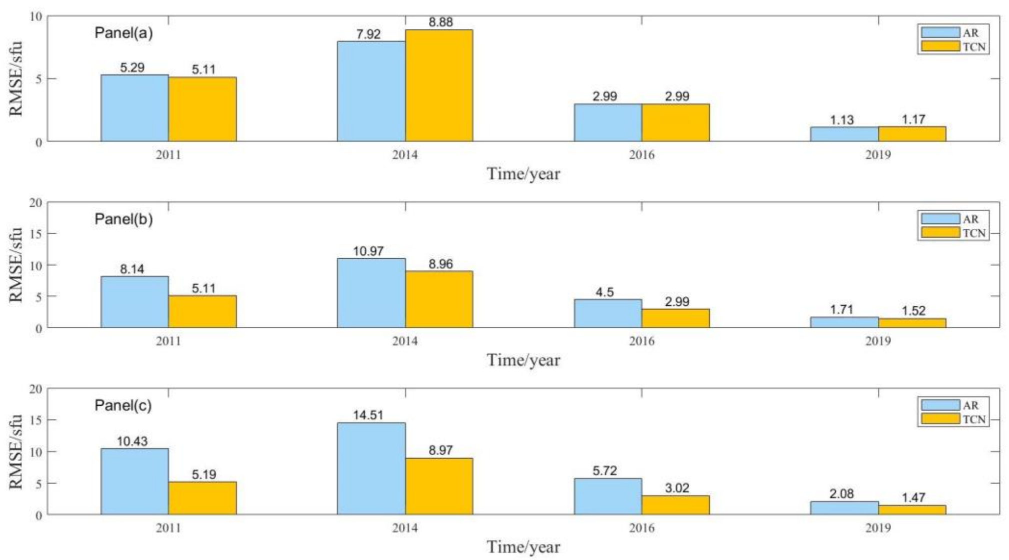

Figure 8: Comparison of the prediction performance of AR and TCN for Panel(a) for 1-day ahead, Panel (b) for 2-days ahead, and Panel (c) for 3-days ahead during different years.

A comparison of the TCN model with other commonly used neural network models ,like BP model (Xiao et al., 2017) and LSTM model (Zhang et al., 2020) for 3-days ahead prediction is shown in Table 4. The RMSE of BP and LSTM models in predicting F10.7 in the high solar activity year of 2003 is 14.28sfu and 7.04sfu, respectively. However, the RMSE of TCN 3-days ahead is 4.71 sfu in 2003. The mean absolute percentage error(MAPE)of BP and LSTM models in predicting F10.7 in the low solar activity year of 2009 is 1.84 and 1.05, respectively. However, the MAPE of TCN 3-days ahead is 1.49 in 2009.Which is better than those of other classical models. The TCN model predicts F10.7 better than the LSTM and BP model's results.There could be three reasons for such results. Firstly, the TCN model use a structure of convolutional layers and residual connections, which enables it to better capture long-term dependencies in time series data (Bai et al., 2018). In comparison, although the LSTM model can also handle long-term dependencies in sequential data, its gated unit structure may not fully capture the complex nonlinear relationships in the data (Zhang et al., 2020). On the other hand, the BP model is simpler and lacks specialized structures for handling time series data, which may result in an ineffective capture of temporal features (Xiao et al., 2017). The residual connections in the TCN model can help mitigate the vanishing gradient problem and improve the stability of the model. This is particularly important for long-term prediction tasks, as the model needs to propagate gradients through multiple time steps. In contrast, the LSTM model may encounter issues of vanishing or exploding gradients in long-term prediction, leading to difficulties in training and unstable predictions (Zhang et al., 2022). The BP model, as a traditional feedforward neural network, may also face similar problems. The TCN model possesses higher flexibility and adaptability, being able to automatically learn appropriate feature representations based on the characteristics of the data. In comparison, the LSTM and BP models require manual feature design and selection, which may not fully leverage the information in the data. The adaptive nature of the TCN model helps it better adapt to different time series data and improve the accuracy of predictions. Therefore, it is precisely because of the advantages mentioned above that TCN performs better in F10.7 prediction.

Table 4. Results of the TCN model's forecast performance 3-days ahead compared to other models

Year	BP/TCN			LSTM/TCN		
	RMSE (sfu)	MAPE (%)	R	RMSE (sfu)	MAPE (%)	R
2003	14.82/4.71	8.15/3.63	0.9937/0.9704	7.04/4.71	3.70/3.63	/
2004	9.74/3.14	7.05/3.12	0.9960/0.9612	5.14/3.14	3.22/3.12	0.9603/0.9612
2008	2.15/1.22	2.11/1.11	0.9996/0.9198	1.22/1.22	1.20/1.11	0.9200/0.9198
2009	1.84/1.49	1.91/1.40	0.9996/0.9540	1.05/1.49	1.07/1.40	/

4. Conclusion

The F10.7 solar flux is an important indicator of solar activity. Its applications in solar physics include serving as an indicator of solar activity level and predicting solar cycle characteristics.In view of the long observation time and certain periodicity of F10.7, this paper introduces for the first time the theory and technique related to TCN based on machine learning into the F10.7

sequence prediction of space weather.

Firstly, we analyze the ability of the TCN model to predict daily F10.7 during solar cycle 24 using training samples from 1957 to 1995. In addition we use the leave-one-out method for cross validation. The results show that the change of prediction accuracy of the model is related to the intensity of solar activity .The TCN model does not large affect the final F10.7 forecasts due to specific properties of the data. This proves that the TCN model is robust to some extent.

Secondly, we compared the predictive performance of the TCN model with the SWPC forecast results and autoregressive (AR) model forecast results. The results show that the TCN model outperformed the SWPC and AR models in terms of prediction accuracy. The predictive accuracy of the TCN model do not significantly vary with the lead time of short-term forecasts (1-day, 2-days, and 3-days). This demonstrates the stability of the TCN model's predictions.

Thirdly, the TCN model has been compared to other classic models such as the BP model and the LSTM model. The TCN model outperformed these models with lower root mean square error (RMSE) and mean absolute percentage error(MAPE).This validates the effectiveness and reliability of the TCN model in predicting the F10.7 solar radio flux. The TCN model is capable of capturing sudden increases or decreases in F10.7, indicating extreme enhancements in solar activity. Therefore, the TCN model has significant implications in predicting F10.7, as it can help us better understand and forecast changes in solar activity.

Although the TCN method has proven to be a viable method for predicting F10.7, there is still room for further improvement in its predictive ability. Future work could attempt to introduce the variable of sunspot number into the model and use a more scientific approach to improve the generalization ability of the model.

Data availability

The data of F10.7 used in this study are available from the National Oceanic and Atmospheric Administration at https://spaceweather.gc.ca/forecast-prevision/solar-solaire/solarflux/sx-5-en.php.

Author contributions

ZL is responsible for data acquisition, processing, data analysis, and drafting the manuscript. LYW, HZ and GSP have made substantial and ongoing contributions to model development, interpretation, and manuscript writing. XXZ and XJX supervised the project, and reviewed and edited the paper.In addition to writing the article, they have also contributed to visualizing the observed results and providing explanations and discussions.

Competing interests

The authors declare no conflict of interests.

Acknowledgments

This work was supported by the National Key R&D Program of China (2021YFA0718600), the National Natural Science Foundation of China grant (No.41931073, 42074183), and State Key Laboratory of Lunar and Planetary Science(SKL-

LPS(MUST)-2021-2023). The author acknowledges the National Oceanic and Atmospheric Administration for providing the F10.7.

References

Aminalragia-Giamini, S., Jiggens, P., Anastasiadis, A., Sandberg, I., Aran, A., Vainio, R., … Dierckxsens, M.:Prediction of Solar Proton Event Fluence spectra from their Peak flux spectra. Journal of Space Weather and Space Climate, 10, 1. doi:10.1051/swsc/2019043, 2020.

Bai, S.J., Kolter, J.Z., Koltun, V., et al. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling[J]. ArXivPreprint, arXiv:1803.01271v2,2018.

Lea, C., Flynn, M. D., Vidal, R., Reiter, A., & Hager, G. D. (2017). :Temporal Convolutional Networks for Action Segmentation and Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2017.113,2017.

Du, Z.: Forecasting the Daily 10.7 cm Solar Radio Flux Using an Autoregressive Model. Sol Phys 295, 125,doi: 10.1007/s11207-020-01689-x,2020.

Huang, C., Liu, D.-D., & Wang, J.-S.: Forecast daily indices of solar activity, F10.7, using support vector regression method. Research in Astronomy and Astrophysics, 9(6), 694–702, doi:10.1088/1674-4527/9/6/008,2009.

Henney, C. J., Toussaint, W. A., White, S. M., and Arge, C. N.: Forecasting F10.7 with solar magnetic flux transport modeling, Space Weather, 10, S02011, doi: 10.1029/2011SW000748,2012.

H. Zhao, O. Gallo, I. Frosio and J. Kautz.:Loss Functions for Image Restoration With Neural Networks,in IEEE Transactions on Computational Imaging, vol. 3, no. 1, pp. 47-57,doi: 10.1109/TCI.2016.2644865,2017.

Katsavrias, C., Aminalragia-Giamini, S., Papadimitriou, C., Daglis, I. A., Sandberg, I., & Jiggens, P.: Radiation belt model including semi-annual variation and solar driving (Sentinel). Space Weather,19,e2021SW002936, doi:https://doi.org/10.1029/2021SW002936,2021.

Lampropoulos, G., Mavromichalaki, H., & Tritakis, V.: Possible Estimation of the Solar Cycle Characteristic Parameters by the 10.7 cm Solar Radio Flux. Solar Physics, 291(3), 989–1002. doi:10.1007/s11207-016-0859-4,2016.

Liu, C., Zhao, X., Chen, T., & Li, H.: Predicting short-term F10.7 with transport models. Astrophysics and Space Science, 363(12). doi:10.1007/s10509-018-3476-x ,2018.

Luo, J., Zhu, H., Jiang, Y., Yang, J., & Huang, Y.: The 10.7-cm radio flux multistep forecasting based on empirical mode decomposition and back propagation neural network. IEEJ Transactions on Electrical and Electronic Engineering, doi:10.1002/tee.23092,2020.

Li X X, Li M, Jiang K C.: Forecasting the F10.7 of solar cycle 25 using a method based on "similar-cycle". Chinese J Geophys. (in Chinese),66(9):3623-3638,doi:10.6038/cjg2022Q0735,2023.

Mordvinov, A.V .: 1986, Prediction of monthly indices of solar activity F10.7 on the basis of a multiplicative autoregression model. Soln. Dannye, Byull. 12, 67.,1986.

M. W. Liemohn, A. D. Shane, A. R. Azari, A. K. Petersen, B. M. Swiger, and A. Mukhopadhyay.:RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics.J. Atmospheric Sol.-Terr. Phys., vol. 218, p. 105624. doi: 10.1016/j.jastp.2021.105624,2021.

Ortikov, M. Y., Shemelov, V. A., Shishigin, I. V., & Troitsky, B. V.: Ionospheric index of solar activity based on the data of measurements of the spacecraft signals characteristics. Journal of Atmospheric and Solar-Terrestrial Physics, 65(16-18), 1425–1430. doi:10.1016/j.jastp.2003.09.005,2003.

Si-qing, L., Qiu-zhen, Z., Jing, W., & Xian-kang, D.: Modeling Research of the 27-day Forecast of 10.7cm Solar Radio Flux (I). Chinese Astronomy and Astrophysics, 34(3), 305–315. doi:10.1016/j.chinastron.2010.07.006,2010.

Tapping, K. F.: The 10.7 cm solar radio flux (F10.7), Space Weather, 11, 394– 406, doi:10.1002/swe.20064, 2013.

Tapping, K. F., & DeTracey, B.: The origin of the 10.7 cm flux. Solar Physics, 127(2), 321–332. doi:10.1007/bf00152171, 1990.

Tapping, K. F. Recent solar radio astronomy at centimeter wavelengths.: The temporal variability of the 10.7-cm flux. Journal of Geophysical Research, 92(D1), 829. doi:10.1029/jd092id01p00829,1987.

Tanaka, H., Castelli, J.P., Covington, A.E. et al.: Absolute calibration of solar radio flux density in the microwave region. Sol Phys 29, 243–262. doi:https://doi.org/10.1007/BF00153452,1973.

Simms, L. E., Ganushkina, N. Y., Van der Kamp, M., Balikhin, M., & Liemohn, M. W.: Predicting geostationary 40–150 keV electron flux using ARMAX (an autoregressive moving average transfer function), RNN (a recurrent neural network), and logistic regression: A comparison of models. Space Weather, 21, e2022SW003263,doi: https://doi.org/10.1029/2022SW003263,2023.

S. Dieleman, A. van den Oord, and K. Simonyan.: The challenge ofrealistic music generation: Modelling raw audio at scale. arXiv:1806.10474, 2018.

Warren, H. P., Emmert, J. T., and Crump, N. A.: Linear forecasting of the F10.7 proxy for solar activity, Space Weather, 15, 1039– 1051, doi:10.1002/2017SW001637, 2017.

Worden, J., & Harvey, J.: An Evolving Synoptic Magnetic Flux map and Implications for the Distribution of Photospheric Magnetic Flux. Solar Physics, 195(2), 247–268. doi:10.1023/a:1005272502885,2000.

XIAO Chao, CHENG Guosheng, ZHANG Hua, RONG Zhaojin, SHEN Chao, ZHANG Bo, HU Hui.: Using Back Propagation Neural Network Method to Forecast Daily Indices of Solar Activity F10.7[J]. Chinese Journal of Space Science,37(1): 1-7, doi:10.11728/cjss2017.01.001,2017.

Yaya, Philippe & Hecker, Louis & Dudok de Wit, Thierry & Fèvre, Clémence & Bruinsma, Sean.: Solar radio proxies for improved satellite orbit prediction. Journal of Space Weather and Space Climate. 7. A35. doi:10.1051/swsc/2017032,2017.

YANG H J，SUN Y Q，ZHU W，et al.: Prediction method of dissolved gas concentration in transformer oil based on CEEMD-TCN model［J］.Electronic Devices，44(4) : 887-8922021, 2021．

Zhang, W.; Zhao, X.; Feng, X.; Liu, C.; Xiang, N.; Li, Z.; Lu, W.: Predicting the Daily 10.7-cm Solar Radio Flux Using the Long Short-Term Memory Method. Universe 2022, 8, 30, doi: 10.3390/universe8010030,2022.

Zhang, H., Xu, H. R., Peng, G. S., Qian, Y. D., Zhang, X. X., Yang, G. L., et al.: A prediction model of relativistic electrons at geostationary orbit using the EMD-LSTM network and geomagnetic indices. Space Weather, 20, e2022SW003126, doi: 10.1002/essoar.10511184.1, 2022.

Zachary C. Lipton, John Berkowitz, Charles Elkan.: A Critical Review of Recurrent Neural Networks for Sequence Learning. arXiv:1506.00019,2015.