Sunspot Prediction Model Based on BiLSTM Network

Shubin Zhang*

School of Mathematics and Statistics, Wuhan University, Wuhan, China

*Corresponding author: 2022302011142@whu.edu.cn

Keywords: Sunspot Prediction; BiLSTM; AP Index; Solar Activity Cycle

Abstract: This paper presents an improved sunspot prediction model based on the Bidirectional Long Short-Term Memory (BiLSTM) network, which integrates solar physics parameters (e.g., the AP index) with historical sunspot data. The study reveals a significant correlation between the AP index and sunspot numbers, both exhibiting 11-year cyclic fluctuations, with the AP index peak lagging behind the sunspot number peak. Utilizing BiLSTM's ability to capture bidirectional time series features, the model considers both past and future data for predictions. The results demonstrate that the BiLSTM model significantly outperforms traditional LSTM in terms of prediction accuracy, as evidenced by metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). Specifically, the model predicts the peak number of sunspots and their occurrence time for the next solar activity cycle, indicating a peak of 185.4558 in July 2024 and 204.0451 in March 2036. This study provides a reliable tool for accurate solar activity cycle prediction, which is crucial for space weather research and applications.

1. Introduction

Sunspots are an important indicator of solar activity, and their quantity and distribution are closely related to changes in the solar magnetic field, directly affecting space weather, the ionospheric state of the Earth, and shortwave radio communication. [9]The solar activity cycle is about 11 years, with its peak and trough periods corresponding to the maximum and minimum periods of solar activity, respectively. However, the solar activity cycle is not completely stable, and the changes in peak intensity, time, and duration are difficult to accurately predict. This makes the prediction of sunspot numbers and their periodic changes an important topic in scientific research.

Accurately predicting the activity of sunspots has profound significance for both scientific research and practical applications. The existing sunspot prediction models often rely on methods such as time series analysis and neural network modeling[2], but their prediction accuracy is often limited. This is because these methods remain at the level of phenomenological analysis at the data level, lacking in-depth exploration of physical mechanisms. In models that rely solely on historical data, errors are particularly significant when predicting the maximum value and its duration.

In order to improve the accuracy of predictions, this paper proposes an improved BiLSTM[5] grid prediction model based on AP-SSN, which combines solar physics parameters (such as AP index) with historical sunspot data. The AP index, as a precursor physical parameter, can effectively characterize the intensity of future solar activity cycles and improve the reliability of predictions through a linear fitting relationship with the number of sunspots. Combining the capabilities of bidirectional long short-term memory networks (BiLSTM), this paper constructs a prediction framework that can simultaneously consider past and future data information. In the validation experiment, the BiLSTM model significantly outperformed traditional methods in terms of root mean square error (RMSE) and other metrics, verifying its predictive performance.

2. Model establishment

2.1 Factor analysis: The relationship between AP index and solar activity cycle

The AP index reflects the intensity of geomagnetic disturbances, calculated based on variations in

  DOI: 10.25236/icceme.2025.010

Earth's magnetic field. During the solar activity minimum, the Sun's dipole magnetic field plays a crucial role in the magnetic field reversal process, influencing the development and intensity of subsequent solar cycles. Changes in the AP index serve as a predictive parameter for the intensity and peak characteristics of upcoming solar cycles. This study identifies a strong correlation between the AP index and sunspot numbers, with both displaying pronounced 11-year cycles. Notably, the AP index peak lags behind the sunspot number peak, illustrating the indirect effect of solar magnetic field variations on Earth's magnetic field. A linear relationship is observed between the maximum sunspot numbers and the minimum polar field strength, with a high fitting degree ($R^2 = 0.8269$). This highlights a significant linkage between these key indicators of solar activity.

Relying solely on historical data of sunspots for prediction can easily lead to significant errors, especially when predicting the maximum value and duration of solar activity, showing limitations. The introduction of the AP index has made up for this deficiency: it provides additional physics based data and enhances the scientificity of predictions; By fitting the relationship between AP index and sunspot number, the accuracy and stability of the prediction model for the maximum value of future cycles can be improved.[6]

Based on the above analysis, the AP index is not only a key parameter in the physics background, but also an important factor affecting the periodic changes of sunspots. In subsequent models, a predictive framework considering time series and causal relationships can be constructed by combining historical data of AP index and sunspot number using a bidirectional long short-term memory network (BiLSTM). This not only captures periodic patterns, but also enhances the model's predictive ability for key nodes in future cycles.

2.2 Data analysis

This article selects the data on the number of sunspots and changes in AP index from 1976 to 2024, including: sunspot number data, representing the intensity of solar activity, monthly average summary of data, including detailed changes within historical cycles; AP index data, representing the intensity of geomagnetic disturbances, also recorded as monthly averages, is used to reveal the relationship between solar activity and geomagnetic disturbances.

2.2.1 Data processing

To ensure data quality, the following steps are taken to preprocess the raw data:

Outlier detection and removal: In this paper ,we use box plots to detect outliers in the data. Th is paper calculate the quartiles of the data and define values that fall within 1.5 times or less of the interquartile range as outliers. For data above or below the interval range, use the upper and lower quartiles respectively (as shown in Figure 1).

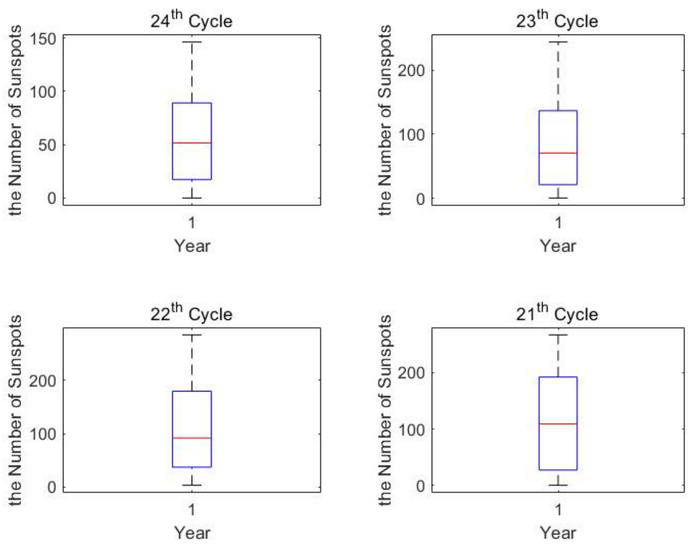


Fig.1 Sample Data Box Line Diagram

Then, the Kalman filter is used to smooth the data to reduce the interference of noise on the analysis.

2.2.2 Data visualization

Periodic variation of sunspot count: By displaying the fluctuation of sunspot count in different cycles through a bar chart, it is found that it mainly varies with a cycle of about 11 years (as shown in Figure 2).

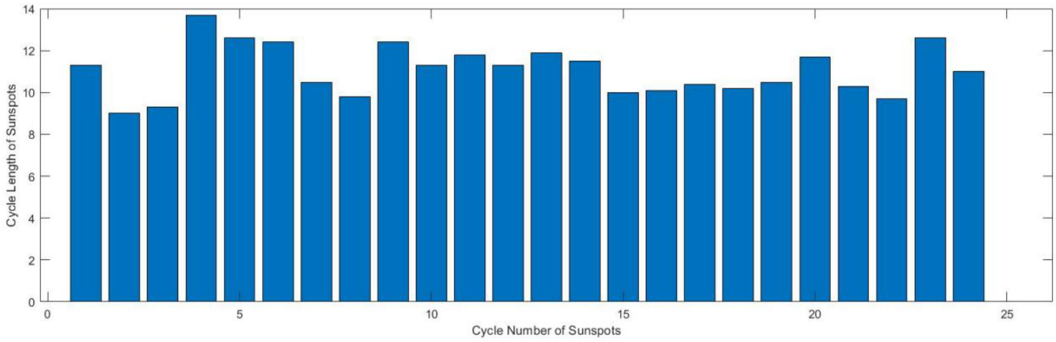


Fig.2 Bar chart of sunspot cycle variation

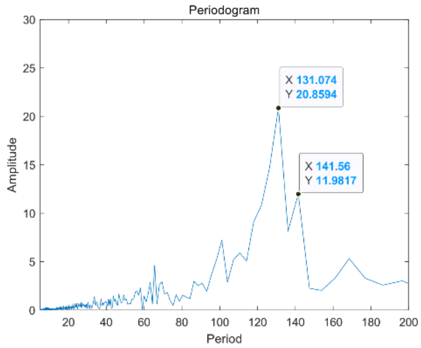
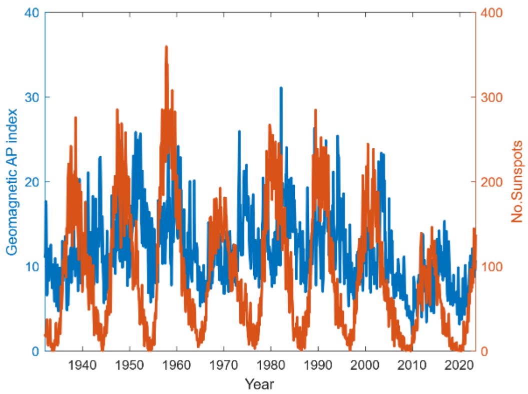


Fig.3 Periodic chart

Frequency domain analysis: Perform discrete Fourier transform on the smoothed sunspot count data to obtain a frequency spectrum and label the main periods. The results indicate that the number of sunspots has significant peaks at 10.92 and 11.79 years, respectively, further verifying the average period of about 11 years (as shown in Figure 3).

The relationship between AP index and sunspot count: By plotting the trend of the two (before and after smoothing), it was found that the peak of AP index lagged behind the peak of sunspot count on average. This lag relationship provides a theoretical basis for predicting the number of sunspots based on the AP index (as shown in Figure 4, Figure 5 and Figure 6).



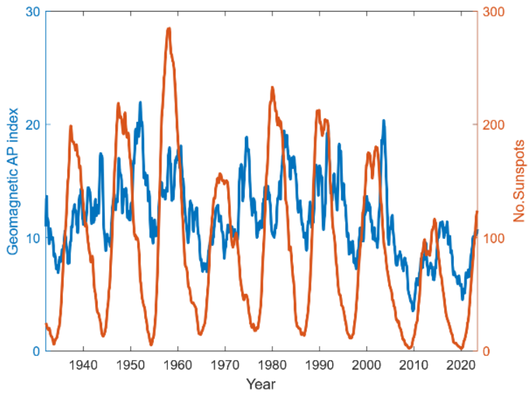


Fig.4 Unsmoothed AP&SSN trend chart Fig.5 Smooth AP&SSN trend chart

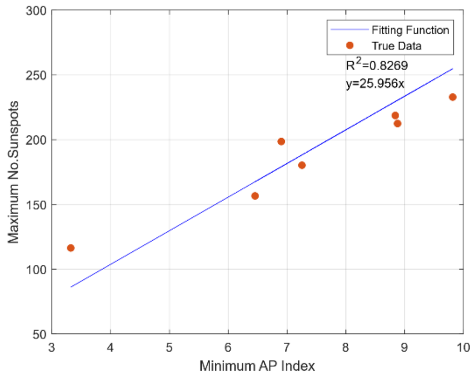


Fig.6 SSN&AP Index Fitting Relationship Diagram

2.2.3 Data analysis summary

Data analysis shows that the periodic variation of sunspot numbers is significant and closely related to changes in the AP index. As a physical parameter of geomagnetic disturbance, the minimum value of AP index can serve as an important basis for predicting the maximum number of sunspots in the next solar activity cycle. Combining linear fitting relationships with time series models can provide scientific support for model establishment.

2.3 Establish a BiLSTM AP-SSN index prediction model

2.3.1 Advantages of BiLSTM model

Bidirectional Long Short Term Memory Network (BiLSTM) can capture temporal dependencies between the past and future by simultaneously processing forward and backward information of time series. Compared to traditional one-way LSTM models, BiLSTM can learn the features of time series more comprehensively and is an ideal choice for processing complex nonlinear time series data.

2.3.2 Model building process

BiLSTM is an improvement on the previous model LSTM, with its hidden layer consisting of both forward LSTM and backward LSTM. The BiLSTM memory network has two directional transport layers, with the front layer along the forward training time series and the back layer along the backward training time series.[8]

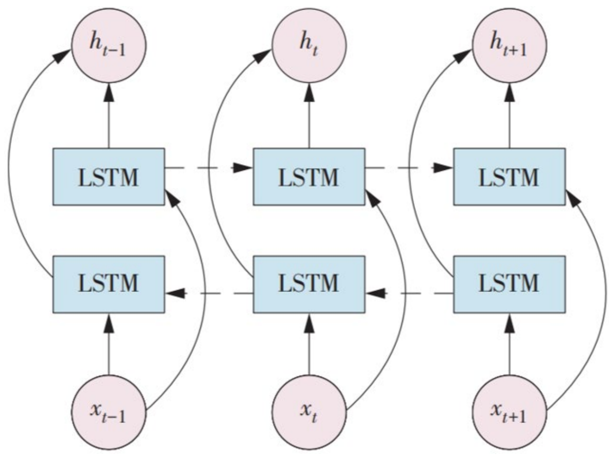


Fig.7 Schematic diagram of Bi LSTM structure[3]

As shown in Figure 7, the BiLSTM neural network structure model is divided into two independent LSTMs, and the input sequences are input to the two LSTM neural networks in positive and negative order for feature extraction. The two output vectors (i.e. the extracted feature vectors) are concatenated to form the final feature expression.[4] The design concept of Bi LSTM model is to enable the feature data obtained at time t to have both past and future information. In addition, the two LSTM neural network parameters in Bi LSTM are independent of each other.

To predict the start time and duration of peak solar activity in the next solar cycle, this study models solar activity using the change in sunspot numbers, which typically indicate solar activity intensity. The peak sunspot number marks the cycle's maximum, while its duration is determined by the interval from the initial rise to the peak and subsequent return to baseline. This paper proposes a BILSTM grid prediction model enhanced with AP-SSN improvements, integrating past and future data for the t-th solar cycle. Using the AP index as a precursor parameter, the BILSTM method predicts the AP index, which is then linearly correlated with the sunspot number (SSN) to estimate the maximum sunspot count. Combining these predictions with historical sunspot data, the model forecasts the maximum sunspot number for the t-th cycle.

3. Results

Using Matlab as the development platform, the design of Long Short Term Memory (LSTM) network uses two network layers, 600 hidden units, and Adam as the optimizer. The model is trained for a total of 300 cycles. This paper use RMSE as the loss function.

The Figures 8 and 9 show the fitting of the BiLSTM prediction model to the training set. Easy to understand, the model fits well and can be used to test the validation set to further verify the performance of the model.

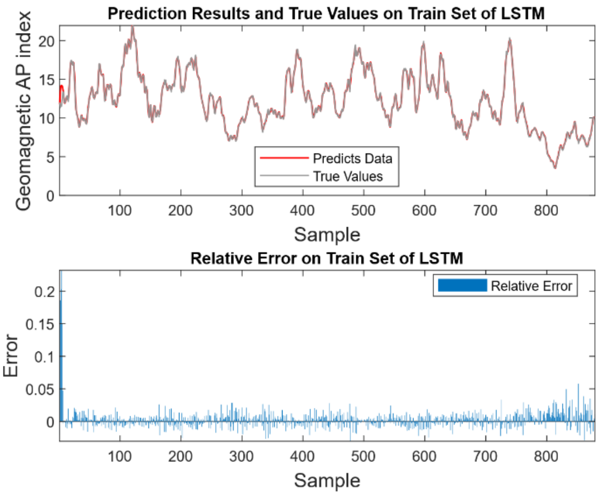


Fig.8 Fitting diagram of AP index LSTM prediction model training set

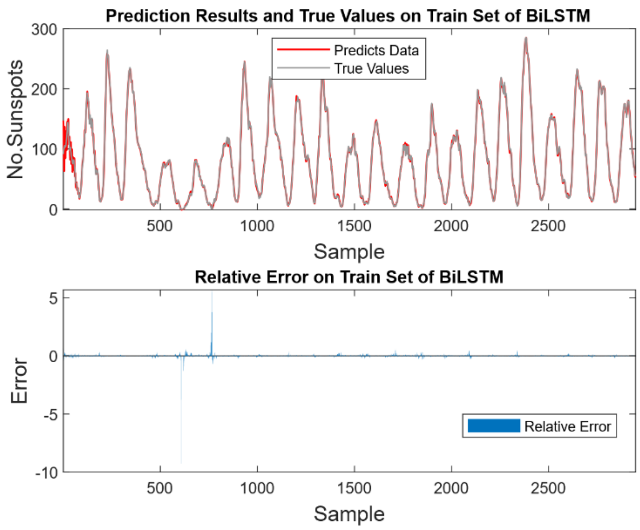
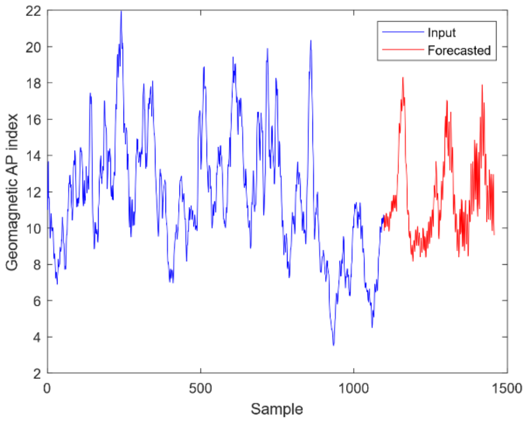


Fig.9 Fitting diagram of BiLSTM prediction model training set[7]

Table 1 Performance Results of AP Index LSTM&BiLSTM Prediction Models

Index	AP-LSTM	BiLSTM
MAE	2.6412	1.4735
MAPE	0.2030	0.1926
RMSE	3.2328	1.8563

From the above structure (Table 1), it can be seen that the performance indicators of the AP index LSTM&BiLSTM prediction models are relatively small, but BiLSTM has better prediction performance.



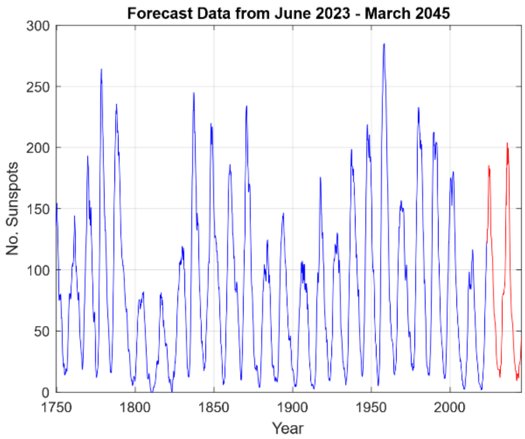


Fig.10 AP Index LSTM model prediction results Fig.11 Prediction results of BiLSTM model

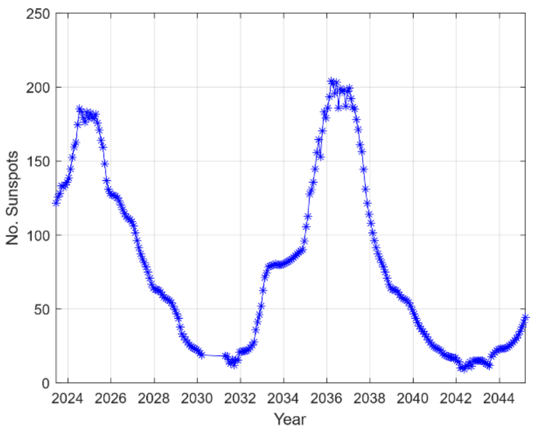


Fig.12 Enlarged BiLSTM Model Prediction Results

Based on this, this article uses an improved BILSTM grid prediction model based on AP-SSN to predict the peak number of sunspots in the next solar activity cycle. The AP index LSTM model prediction results are shown in Figure 10, and the BiLSTM model prediction results are shown in Figure 11. To better characterize the prediction results, they are enlarged as shown in Figure 12.[1]

Therefore, based on the analysis of the above results, it can be concluded that in July 2024, the number of sunspots reached its peak at 185.4558, entering the maximum period of solar activity cycle, which lasted for 14 months until August 2025 when the number of sunspots fell back to 160.3468 and began to exit the maximum period. In the next cycle, the number of sunspots will reach its maximum starting from March 2036, with a peak of 204.0451, lasting for 12 months until February 2037, when the number of sunspots will fall back to 192.201 and exit the maximum period.

4. Conclusion

This article focuses on the number of sunspots and AP index, and proposes an improved BiLSTM grid prediction model based on AP-SSN to address the shortcomings of traditional prediction methods. A dual input time series prediction framework was constructed by combining the AP index, an important physical precursor parameter, with historical data of sunspots.

During the model building process, Kalman filtering is used to smooth the data, eliminate outliers, and BiLSTM is utilized to capture the bidirectional features of the time series. At the same time, comparing the performance of the model with traditional LSTM, the results show that BiLSTM is significantly better than LSTM in terms of mean absolute error (MAE), root mean square error (RMSE) and other indicators, demonstrating stronger nonlinear fitting and generalization abilities.

In the end, BiLSTM successfully predicted the maximum number of sunspots and their occurrence time for the next solar activity cycle, with excellent results. This model provides a reliable tool for accurate prediction of solar activity cycles and is of great significance for the research and application of space weather.

References

[1] Mao N ,Sun K ,Zhang J .Monthly Maximum Magnitude Prediction in the North–South Seismic Belt of China Based on Deep Learning[J].Applied Sciences,2024,14(19):9001.

[2] Geng S ,Zhai S ,Ye J , et al.Decoupling and predicting natural gas deviation factor using machine learning methods[J].Scientific Reports,2024,14(1):21640.

[3] VishnupriyaG ,AnushaS ,KayikciS .An efficient and secure wearable sensor based remote healthcare monitoring system using adaptive dilated transformer Bi-LSTM with gated recurrent unit[J].Transactions on Emerging Telecommunications Technologies,2024,35(2).

[4] Qiyue X ,Lin M ,Yao L , et al.An improved SSA-BiLSTM-based short-term irradiance prediction model via sky images feature extraction[J].Renewable Energy,2023,219(P2).

[5] Xiaopei C ,Xueyang T ,Fei Y , et al.Estimation of turnout irregularities using vehicle responses with improved BiLSTM and Gaussian process regression[J].Measurement,2023,221.

[6] Hui X ,Liming S ,Tianjiao Z , et al.Effects of Different Spatial Resolutions on Prediction Accuracy of Thunnus alalunga Fishing Ground in Waters Near the Cook Islands Based on Long Short-Term Memory (LSTM) Neural Network Model[J].Journal of Ocean University of China,2023,22(5):1427-1438.

[7] Haobing C ,Guohui Z ,Xin Z , et al.Learning Emotion Assessment Method Based on Belief Rule Base and Evidential Reasoning[J].Mathematics,2023,11(5):1152.

[8] Yiqi Z ,Jinglin Z ,Yanping Z , et al.RBCN-Net: A Data-Driven Inertial Navigation Algorithm for Pedestrians[J].Applied Sciences,2023,13(5):2969.

[9] Carrasco S M V ,Vaquero M J ,Gallego C M .Analysing Spotless Days as Predictors of Solar Activity from the New Sunspot Number[J].Solar Physics,2017,292(11).