Astronomical Techniques and Instruments, Vol. 3, January 2026, 43-54

Article** | **Open Access

2026-06-11

ChinaRxiv

Abstract

The solar cycle (SC), a phenomenon caused by the quasi-periodic regular activities in the Sun, occurs approximately every 11 years. Intense solar activity can disrupt the Earth' s ionosphere, affecting communication and navigation systems. Consequently, accurately predicting the intensity of the SC holds great significance, but predicting the SC involves a long-term time series, and many existing time series forecasting methods have fallen short in terms of accuracy and efficiency. The Time-series Dense Encoder model is a deep learning solution tailored for long time series prediction. Based on a multi-layer perceptron structure, it outperforms the best previously existing models in accuracy, while being efficiently trainable on general datasets. We propose a method based on this model for SC forecasting. Using a trained model, we predict the test set from SC 19 to SC 25 with an average mean absolute percentage error of 32.02, root mean square error of 30.3, mean absolute error of 23.32, and R2 (coefficient of determination) of 0.76, outperforming other deep learning models in terms of accuracy and training efficiency on sunspot number datasets. Subsequently, we use it to predict the peaks of SC 25 and SC 26. For SC 25, the peak time has ended, but a stronger peak is predicted for SC 26, of 199.3, within a range of 170.8–221.9, projected to occur during April 2034.

Full Text

Astronomical Techniques and Instruments, Vol. 3, January 2026, 43–54

Article | Open Access

Forecasting solar cycles using the time-series dense encoder deep learning model

Cui Zhao[1], **Shangbin Yang**[2,3], **Jianguo Liu**[1*], **Shiyuan Liu**[1]

[1] *College of Applied Science and Technology, Beijing Union University, Beijing 102200, China*
[2] *State Key Laboratory of Solar Activity and Space Weather, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China*
[3] *University of Chinese Academy of Sciences, Beijing 100049, China*

*Correspondence: Liujianguoniat@163.com

Received: July 3, 2025; Accepted: September 17, 2025; Published Online: September 18, 2025; https://doi.org/10.3724/ati2025052; https://cstr.cn/32083.14.ati2025052

Citation: Zhao, C., Yang, S. B., Liu, J. G., et al. 2026. Forecasting solar cycles using the time-series dense encoder deep learning model. *Astronomical Techniques and Instruments*, **3**(1): 43-54. https://doi.org/10.3724/ati2025052.

Abstract: The solar cycle (SC), a phenomenon caused by the quasi-periodic regular activities in the Sun, occurs approximately every 11 years. Intense solar activity can disrupt the Earth' s ionosphere, affecting communication and navigation systems. Consequently, accurately predicting the intensity of the SC holds great significance, but predicting the SC involves a long-term time series, and many existing time series forecasting methods have fallen short in terms of accuracy and efficiency. The Time-series Dense Encoder model is a deep learning solution tailored for long time series prediction. Based on a multi-layer perceptron structure, it outperforms the best previously existing models in accuracy, while being efficiently trainable on general datasets. We propose a method based on this model for SC forecasting. Using a trained model, we predict the test set from SC 19 to SC 25 with an average mean absolute percentage error of 32.02, root mean square error of 30.3, mean absolute error of 23.32, and R^2 (coefficient of determination) of 0.76, outperforming other deep learning models in terms of accuracy and training efficiency on sunspot number datasets. Subsequently, we use it to predict the peaks of SC 25 and SC 26. For SC 25, the peak time has ended, but a stronger peak is predicted for SC 26, of 199.3, within a range of 170.8-221.9, projected to occur during April 2034.

Keywords: Solar cycle; Forecasting; TiDE; Deep learning

1. INTRODUCTION

The solar cycle (SC), also referred to as the solar magnetic activity cycle or the sunspot activity cycle, represents a quasi-periodic variation in the count of sunspots and other solar activity[1-3]. The periodic change spans approximately 11 years and includes a rising phase, a peak phase, and a declining phase. During the peak phase, a multitude of sunspots emerge on the solar surface, accompanied by an upsurge in solar activity, while the intensity of solar radiation increases to a level that can potentially interfere with the ionosphere and affect terrestrial communication and navigation systems. Moreover, solar storms can inflict damage on infrastructure, such as power grids[4]. Strong solar activity may also impact Earth' s climate. For instance, during the peak of the SC, the Earth may experience more extreme weather events such as heat waves, droughts, and geomagnetic storms, as well as space-weather-related hazards[5-8]. By predicting the SC, the intensity and distribution of solar radiation can be predicted, which can offer valuable guidance for the design and operation of communication and navigation systems, strengthening power grid protection, and ensuring the safety and stability of human society. Additionally, it allows for a more precise prediction of future climate change, providing an important reference for Earth' s climate research.

Methods for predicting SC can be categorized into two main types: numerical simulation methods grounded in physical laws and those based on machine learning. Numerical simulation methods aim to numerically replicate the physical processes within the Sun, such as variations in the solar magnetic field and motion in the convection zone. For example, Upton and Hathaway[9] employed the Advective Flux Transport model to generate an early prediction for SC 25, indicating that it would be similar in strength to that of SC 24, with an uncertainty of about 15%. Schatten and Pesnell[10] presented a solar dynamo of

SC 24. Kitiashvili and Kosovichev[11] used an Ensemble Kalman Filter (EnKF) to predict the SC properties. This approach accounted for uncertainties inherent in both the dynamo model and the measurements. Their prediction suggested that the next sunspot cycle would be significantly weaker, approximately 30% less intense than the previous cycle. McIntosh et al.[12] extracted a relationship between the temporal spacing of terminators and the magnitude of sunspot cycles, deducing that sunspot SC 25 could have a magnitude that rivaled the top few since records began. Bhowmik and Nandy[13] used an observational data assimilated, century-scale calibrated Surface Flux Transport (SFT) model, with an output coupled to a solar internal dynamo model, to predict that SC 25 would have a peak SSN of 118 and a range from 109–155. Jiang et al.[14] predicted SC 25 with a solar SFT model, predicting a peak SSN of 125 $\pm$ 32. Pesnell and Schatten[15] used the Solar Dynamo (SODA) index to create a precursor of future solar activity, predicting a maximum SSN of 135 $\pm$ 25 occurring in 2025.2 $\pm$ 1.5. Petrovay[16] determined that the prediction methods form three main groups, including Precursor methods (Nagovitsyn et al.[17], Kilcik et al.[18]), solar dynamo theory (Hawkes and Berger[19] and Gopalswamy et al.[20]), and Model-based approaches (Petrovay and Talafha[21]; Petrie and Ettinger[22]; Whitbread et al.[23] and Mandal et al.[24]). Jiang et al.[25] compared 7 physics-based prediction models from 4 aspects, and demonstrated the large effect of assimilated magnetograms on predictions, using two SFT numerical tests. Bhowmik et al.[26] determined that the most successful prediction approaches support the Babcock-Leighton (B-L) solar dynamo mechanism as the primary driver of SC variability. Other physical methods were also employed by various researchers, including Dikpati et al.[27], Pesnell[28], and Hathaway and Upton[29].

Machine-learning-based SC prediction methods leverage machine learning and deep learning technologies to automatically learn, mine and analyze the data of solar activity cycles. For instance, Lee[30] used the hybrid Empirical Mode Decomposition method and Long Short-Term Memory network (EMD-LSTM) approach to model the SSN series. The EMD-LSTM model predicted that for SC 25, the SSN would be both later and larger than what was predicted by the LSTM model alone. Benson et al.[31] employed a combination of WaveNet and LSTM networks to forecast the SSN. Their results suggested that the upcoming SC 25 would have a maximum SSN around 106, indicating that the cycle would be slightly weaker than SC 24. Wang et al.[32] presented a LSTM deep learning model to forecast SC 25, predicting that the peak amplitude would occur around 2023, with a peak value of approximately 114.3. Prasad et al.[33] employed a deep learning model to predict the strength and peak time of SC 25. This model predicted that SC 25 would peak with an SSN amplitude of 171.9, being 47% stronger than SC 24, and would reach its peak in August 2023. Dai et al.[34] established a prediction model based on the Temporal Convolutional Network (TCN) with phase space reconstruction to predict SC 25. Their findings indicated that the maximum SSN would be 139.55, occurring in April 2024. Su et al.[35] proposed the XGBoost-SCINet (XG-SN) ensemble model for making predictions regarding known SC. Other deep learning models for SC prediction have been proposed by Pala and Atici[36], Dang et al.[37], Su et al.[38], and Cao et al.[39–40].

Because of the complexity of solar dynamos, the underlying mechanisms governing the Sun' s internal activity remain incompletely understood. As a result, numerical simulation methods grounded in physical laws have perpetually posed a significant challenge. Machine learning boasts robust learning capabilities and the capacity to rapidly model complex nonlinear systems, which has led to its extensive application in time series forecasting. Nonetheless, when it comes to predicting SC, a long time series problem, many contemporary deep learning methods, such as Recurrent Neural Networks (RNNs) and transformer architectures, have demonstrated less than satisfactory predictive prowess. For example, RNN models yield inaccuracies when forecasting long-term solar activity[30]. Moreover, because of the complexity of the transformer model, it often requires prolonged training times and struggles to achieve effective outcomes[41]. In recent years, research efforts aimed at modeling long temporal dependencies have led to further enhancements, giving rise to the emergence of several more reliable deep learning approaches. Among these, the Time-series Dense Encoder (TiDE) deep learning model[42] has been proposed as a deep learning solution tailored for long time series prediction. This model is based on a multi-layer perceptron (MLP) structure. Across experiments conducted on multiple available benchmark datasets, it not only outperforms other previously employed models in terms of accuracy, but also boasts an inference speed and training efficiency 5–10 times faster than those of the Transformer model. Applied to the Electricity Transformer Temperature-Hourly with 1 Variate (ETTh1) dataset, it gives a prediction error 19.3% lower than that of Informer.

Here, we propose a SC forecasting model based on the TiDE network and use it to predict SC from 15 to 25. When compared with some other previous deep learning models, it shows better performance with smaller errors and high computational efficiency. The TiDE was also checked using other SSN datasets. Finally, we use it to predict the peaks of SC 25 and 26 and present it alongside the data and the experimental procedures used.

2. DATA

Sunspots serve as a prominent indicator of the intensity of solar activity. Commonly, the SSN is employed to represent the intensity of sunspots. Since 1749, the Royal Observatory of Belgium, Brussels has been conducting long-term, reliable sunspot observations. In this study, we used these data for our experiments, selecting the 13-

smoothed monthly total SSN. The data spanned July 1749 to November 2024, and Table 1 shows the descriptive statistical analysis of the data we used in the experiments. Data preprocessing involved removing outliers and normalizing the data. Because the first 6 records contained invalid SSN values “−1.0” , they were regarded as outliers and removed. The values were then normalized to [0, 1], to facilitate model training and convergence. Subsequently, the dataset was partitioned into a training set and a test set at a ratio of 0.75, with January 1956 as the partition point. The training set totaled 2479 months from July 1749 to January 1956, and the training set totaled 826 months from February 1956 to November 2024.

We performed an exploratory analysis on the dataset and checked its periodicity. Time series decomposition is a common preprocessing method in time series forecasting, used to decompose the time series into trend and periodicity variations. Trend reflects a tendency or state of continuous development and change over a long period of time, and periodicity describes a cyclical pattern in the data. Fig. 1 shows a data visualization of SSN and its time series decomposition, where the first subgraph is the original SSN, and the second and third subgraphs are the trend and periodicity. The data have clearly apparent periodicity with a distribution of approximately 11 years. This is confirmed by the autocorrelation function (ACF). As

Table 1. Descriptive statistical analysis of SSN data

Count	Mean	Std	Min	Max	Skewness	Kurtosis
3305	81.93	63.07	0	285.0	0.76	−0.19

Fig. 1. Exploratory analysis of 13-month smoothed monthly total SSN datasets. (A) The original SSN data distribution. (B) The general trend of the data. (C) The periodicity of the data. (D) ACF check for the data, for which the line in orange indicates that the abscissa value of the peak is 126.

shown in Fig. 1D, the peak indicates the existence of an annual periodic trend.

3. TiDE MODEL

The TiDE model focuses on solving multivariate long-term time series prediction tasks. Its structure comprises an MLP, which solves the problem of linear models being unable to establish non-linear dependencies between the prediction window and the historical window, and addresses the issue of being unable to effectively handle external covariates.

The basic layer in the architecture is the residual block. It consists of a Dense and Reliable Label Utilization (RLU) layer, a Dense Linear layer and a Layer norm. All other layers of TiDE are built by this layer. Overall, the model can be divided into four parts: feature projection, dense encoder, dense decoder, and temporal decoder (see Fig. 2). The data fed into the model are the past of the time-series labeled variable y, static attributes labeled variable a, and all the past and future projected covariates labeled variable x. For example, to make load forecasts, y represents the load power, which is the data to be predicted; x represents dynamic covariates, such as temperature and wind speed, which change with time; a represents variables such as region and altitude that do not change with time.

Text appearing in Fig. 2: Predictions $\hat{y}^{(i)}_{L+1:L+H}$; Residual; Temporal decoder (per time-step); Unflatten; $g^{(i)}$; $d^{(i)}_{L+1:L+H}$; $\tilde{x}^{(i)}_{L+1:L+H}$; Stack; Decoder $\times n_d$; $e^{(i)}$; Encoder $\times n_e$; Concat; Flatten; $y^{(i)}_{1:L}$; $a^{(i)}$; $\tilde{x}^{(i)}_{1:L+H}$; Lookback; Attributes; Feature projection (per time-step); $x^{(i)}_{1:L+H}$; Dynamic covariates; Residual block; Layer Norm; Skip connection; Dropout; Dense (linear); Dense (ReLU).

Fig. 2. Overview of the TiDE architecture. The model can be divided into four parts: feature projection, dense encoder, dense decoder, and temporal decoder. The data are input into the model from bottom and the prediction results are output from the top.

Feature projection maps variable x to a low-dimensional vector, using a residual block. The dense encoder concatenates variable y, variable a, and the low-dimensional vector mapped by variable x, and maps them using multiple layers of residual block to finally obtain an encoding result e. By using the same multiple layers of residual block, the dense decoder maps the encoding result e to obtain the variable g, and reshapes g into a $[p, H]$ matrix. Among them, H is the length of the prediction window, and p is the output dimension of the decoder, which is equivalent to obtaining a vector at each moment of the prediction window. The temporal decoder concatenates g and x from the previous step according to the time dimension and uses a residual block to map the output results at each moment. Subsequently, the direct mapping results of the historical sequence are added to the residual connection to obtain the final prediction result.

Although the traditional MLP structure has high parameter efficiency, it may still be prone to overfitting, because the network layer width is too large, resulting in the expansion of the parameter space. The decoder complexity is also too high to match the limited training data. TiDE maintains the generalization ability of the model through the methods and strategies with a feature projection layer designed to actively reduce the dimensionality of the output from the model. The high-dimensional features are then mapped to a lower-dimensional space through mathematical transformation. The model is forced to discard feature dimensions with low correlation (usually corresponding to noise), retaining the principal components with high information content. The time series decoder adopts a narrow layer design, forming an information bottleneck. In the decoder, the number (width) of neurons in the middle layer is deliberately designed to be

smaller than the input/output dimensions. The information can only be transmitted through the "narrow" middle layer during the decoding process, retaining the most core patterns, such as periodicity and trend, rather than local noise. Next, skip connections are designed. Part of the encoder output is directly passed to the decoder output layer to reduce the mapping complexity that the decoder needs to learn. Finally, the method used in the learning strategy initially freezes some layer parameters before gradually unfreezing them.

4. EXPERIMENTS AND RESULTS

4.1. Models Architecture and Selection of Hyperparameters

In our experiment, we selected the data from the first t time steps as the input to predict the subsequent n time steps. Specifically, the input length was set to 240, and the output length was set to 120. This meant that data from the first 240 months were used to predict the data for the next 120 months (i.e., the first two SC were employed to forecast the next SC). For example, for the training set, the first sample input $[1, 2, \ldots, 240]$ months SSN predicts the next $[241, \ldots, 360]$ months SSN. For the testing set, the first sample input $[2480, 2481, \ldots, 2719]$ predicts the next $[2720, \ldots, 2839]$ months. These were optimal values obtained through parameter tuning. Dropout is a regularization technique that aims to enhance the generalization ability of the model and prevent overfitting by randomly setting the activation values of some neurons to zero[29,30]. In this experiment, the dropout rate was set to 0.1. The term "N epochs" refers to the number of times the entire dataset is used for training. This is to ensure that the model comprehensively learns the features within the training data set. After experimentation, we determined that setting N epochs to 100 was optimal.

In the TiDE structure, for the residual block configuration, each residual block contains two fully connected layers (with the dimensions [`input_dim`, `hidden_dim`] and [`hidden_dim`, `output_dim`]) with Rectified Linear Unit (ReLU) activation and optional Layer Normalization. Output dimensions vary per module. The feature projection layer was implemented as a specialized residual block for dimensionality reduction when `input_dim` is larger than `output_dim`. Taking wind speed prediction as an example, y is the target variable (i.e., the data to be predicted, such as power load), x is time-varying dynamic covariates (e.g., temperature, wind speed), and a is static features (e.g., geographic location, altitude). Dimensionality reduction is achieved when the input dimension of x exceeds the output dimension of the feature projection layer. In our current study, the model only processes the target variable y, without incorporating dynamic covariates x or static features a. In the training stage, to prevent overfitting and enable the model to converge faster, the projection layer and encoding layer were frozen by setting their parameters to not be updated, training only the decoder. The optimized architectural parameters of the TiDE model are listed in Table 2, and the optimization parameters of comparison models are Neural Basis Expansion Analysis for Interpretable Time Series (N-BEATS), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU) and Bi-directional Long Short-Term Memory (BiLSTM), which are listed in Table 3.

Table 2. The optimized architectural parameters for the TiDE model, as determined through hyperparameter tuning

Parameters	Explanation	Value
`input_chunk_length`	Number of time steps in the past to take as a model input	240
`output_chunk_length`	Number of time steps predicted at once (per chunk) by the internal model	120
`num_encoder_layers`	Number of stacked residual blocks in the encoder	1
`num_decoder_layers`	Number of stacked residual blocks in the decoder	1
`decoder_output_dim`	The number of output components of the decoder	16
`hidden_size`	The width of the hidden layers in the encoder/decoder residual blocks	128
`temporal_decoder_hidden`	The width of the hidden layers in the temporal decoder	32
`use_layer_norm`	Whether to use layer normalization in the residual blocks	True
`dropout`	Dropout probability	0.1

Table 3. Optimized architectural parameters of other models in experiments

Model	Optimized architectural parameters
N-BEATS	`num_layers = 4`, `num_blocks = 1`, `layer_widths = 256`, `dropout = 0.1`
CNN	`hidden_size = 128`, `kernel_size = 3`, `dropout = 0.4`
LSTM	`hidden_dim = 19`, `n_rnn_layers = 1`, `dropout = 0.2`
GRU	`hidden_dim = 19`, `n_rnn_layers = 2`, `dropout= 0.2`
BiLSTM	`hidden_dim = 19`, `n_rnn_layers = 2`, `dropout= 0.2`

The model evaluation metrics employed are the mean absolute percentage error (MAPE), the mean absolute error (MAE) and root mean square error (RMSE). The smaller these values are, the better the performance of the model is. In addition, the coefficient of determination, R^2, is also used for evaluation, for which ideal values are those that approach 1. The experiments were carried out on a computer with an Intel Core i5-7200U CPU, using the PYTORCH software to construct the deep learning model.

4.2. Fitting Effect on the Test Set from SC 19 to SC 25

To evaluate the proposed TiDE model, we applied the trained model to predict the test set from SC 19 to SC 25. Fig. 3 illustrates the comparative analysis between observed solar activity and model predictions. The visualization reveals that TiDE achieves satisfactory phase matching across most SC, with the notable exception of SC 19. This discrepancy likely stems from data limitations and model underfitting, caused by the limited number of training samples available; there are only 2119 samples from SC 1 to SC 18, of which only SC 8 shows a similar degree of extreme activity to that of SC 19. The data volume for SC 8 accounts for 5% of the training sample, fundamentally constraining the ability of the model to learn extreme activity patterns. The model regards SC 19 as an outlier. Obtaining richer observation data and incorporating multi-source solar activity records (such as radio, extreme ultraviolet, and magnetic data) are potential methods to address this.

We quantified the prediction accuracy of our model with multiple metrics (results are summarized in the last column of Table 4). The model demonstrates performance with an average R^2 of 0.76 across all cycles,

Fig. 3. Fitting effect on the test set from SC 19 to SC 25. The black curves show the observed solar activity (true value), the red lines are the predictions by our model, and the dotted lines are predictions by other models.

though SC 19 shows significant deviation from this trend. Aggregate error metrics yield MAPE = 32.02, MAE = 23.32, and RMSE = 30.3, reflecting the baseline predictive capability of the model. We expect that more observational data will help the model to make better predictions and achieve better accuracy.

5. DISCUSSION

5.1. Comparison with Traditional Deep Learning Models

To evaluate the capability of the TiDE model, we benchmarked it against the six traditional deep learning models: N-Beats, CNN, RNN, LSTM, BiLSTM and GRU. N-BEATS employed stacked, fully connected blocks with bidirectional residual connections for sequential prediction. The CNN model was configured with a kernel size of 3 and 6 filters. For RNN, we applied the "vanilla RNN" which is the basic form, without any complex structures, such as the gating mechanisms in LSTM or GRU. LSTM adopted a single-layer LSTM, while the BiLSTM adopted a dual-layer stacked architecture. GRU was a single-layer GRU architecture. Fig. 3 shows SC predictions made with different models, where the real data align more closely with the predictions of TiDE than other models. TiDE shows enhanced predictive accuracy across a majority of cycles, with its trajectory showing minimal deviation from observational data.

To quantitatively assess model performance, Table 4 summarizes evaluation metrics on the test set from SC 19 to SC 25, with MAPE, MAE, RMSE and R^2 serving as comparative metrics. The statistical analysis demonstrates that TiDE shows superior performance, exhibiting the lowest average errors across all cycles (MAPE = 32.02, MAE = 23.32, RMSE = 30.3, $R^2 = 0.76$). LSTM is the closest competitor (MAPE = 32.98, MAE = 26.45, RMSE = 34.7, $R^2 = 0.69$), while N-BEATS shows the highest error margins (MAPE = 68.24, MAE = 42.19, RMSE = 52.34, $R^2 = 0.34$). Fig. 4 shows the variations of RMSE across cycles in predictions from different models.

To further evaluate the model performance, we conducted an analysis of the temporal accuracy in peak predictions across SC. Specifically, we quantified the time differ-

Table 4. Evaluation errors of test sets from different models

	SC	N-BEATS	CNN	RNN	LSTM	BiLSTM	GRU	Proposed model
MAPE	SC 19	73.00	50.71	61.48	33.98	61.67	73.72	43.79
MAPE	SC 20	30.14	24.98	19.44	18.3	57.68	90.00	28.77
MAPE	SC 21	78.60	55.54	40.01	44.77	69.16	43.07	34.59
MAPE	SC 22	47.69	44.26	46.34	29.61	29.52	26.76	13.29

	SC	N-BEATS	CNN	RNN	LSTM	BiLSTM	GRU	Proposed model
MAPE	SC 23	70.16	40.56	42.08	26.5	38.56	27.44	31.05
MAPE	SC 24	64.75	74.99	69.35	40.82	65.91	29.55	37.91
MAPE	SC 25	113.33	52.32	31.9	36.86	42.8	19.81	34.72
AVG		68.24	49.05	44.37	32.98	52.19	44.34	32.02
MAE	SC 19	72.69	55.87	64.95	43.68	48.28	75.46	54.61
MAE	SC 20	20.25	14.86	15.03	10.98	21.84	48.42	22.27
MAE	SC 21	56.69	49.54	41.96	44.38	47.22	34.43	30.47
MAE	SC 22	35.14	37.23	42.08	25.05	18.49	23.66	10.67
MAE	SC 23	43.99	20.61	23.83	13.22	15.53	10.73	13.16
MAE	SC 24	19.79	40.25	28.28	17.91	24.26	10.79	13.33
MAE	SC 25	46.79	24.91	9.33	29.92	15.29	11.29	18.72
AVG		42.19	34.75	32.21	26.45	27.27	30.68	23.32
RMSE	SC 19	93.68	73.19	91.11	59.54	62.62	103.22	73.66
RMSE	SC 20	26.58	18.33	19.45	14.72	23.72	65.21	28.35
RMSE	SC 21	71.43	58.62	56.1	57.05	59.61	39.99	36.77
RMSE	SC 22	42.22	43.06	55.15	29.82	23.06	30.73	15.03
RMSE	SC 23	53.75	26.31	29.35	18.38	20.27	14.39	15.88
RMSE	SC 24	23.73	51.41	33.61	22.91	30.78	14.99	17.36
RMSE	SC 25	54.97	27.56	11.93	40.49	19.22	18.29	25.05
AVG		52.34	42.64	42.39	34.7	34.18	40.97	30.30
R^2	SC 19	0.04	0.41	0.09	0.61	0.57	-0.17	0.40
R^2	SC 20	0.70	0.86	0.84	0.91	0.76	-0.82	0.66
R^2	SC 21	0.15	0.43	0.48	0.46	0.41	0.73	0.78

	SC	N-BEATS	CNN	RNN	LSTM	BiLSTM	GRU	Proposed model
R^2	SC 22	0.67	0.66	0.44	0.84	0.90	0.83	0.96
R^2	SC 23	0.23	0.81	0.77	0.91	0.89	0.94	0.93
R^2	SC 24	0.61	−0.82	0.22	0.64	0.35	0.85	0.79
R^2	SC 25	−0.03	0.74	0.95	0.44	0.87	0.89	0.79
AVG		0.34	0.44	0.54	0.69	0.68	0.46	0.76

Fig. 4. Line chart of the RMSEs of test sets from different models. The red line shows the RMSE error between the true value and the value predicted by TiDE from SC 19 to SC 25. Other lines are the errors between the true value and values predicted by other models.

…ences between the observed peaks and peaks predicted by different models (see Table 5). The analysis encompasses SC 19 through SC 24 (excluding SC 25 because this solar cycle has not yet ended). The second row of Table 5 gives the actual peak timings, while subsequent rows present predictions given by each model, together with their corresponding timing errors (calculated as the absolute difference between predicted and observed times). The final column summarizes the average timing errors across all analyzed cycles. Our model achieves good performance, demonstrating an average timing error of 10.8 months. By contrast, the GRU model showed the largest prediction discrepancy, with an average error of 15.3 months.

Table 5. Peak timing predictions and errors between predicted timing and the true timing. The input data here are test sets predicted by different models

Models	SC 19	SC 20	SC 21	SC 22	SC 23	SC 24	Avg timing error
True peak timing	1958-03	1968-11	1979-12	1989-11	2001-11	2014-04	
N-BEATS	1959-09	1970-08	1981-02	1991-01	2002-03	2014-05	12
Error/month	18	21	14	14	4	1	12
CNN	1958-09	1968-09	1980-03	1990-10	2001-02	2011-11	10
Error/month	6	2	3	11	9	29	10

Models	SC 19	SC 20	SC 21	SC 22	SC 23	SC 24	Avg timing error
RNN	1959-04	1970-02	1981-03	1991-07	2000-10	2015-01	14.2
Error/month	13	15	15	20	13	9	14.2
LSTM	1958-04	1969-07	1980-11	1990-03	2000-10	2012-11	9
Error/month	1	8	11	4	13	17	9
BiLSTM	1958-08	1969-01	1981-08	1990-06	2000-08	2012-10	11.2
Error/month	5	2	20	7	15	18	11.2
GRU	1959-12	1971-12	1980-03	1989-09	2000-08	2013-02	15.3
Error/month	21	37	3	2	15	14	15.3
Proposed model	1958-09	1970-06	1980-05	1990-10	2001-02	2013-01	10.8
Error/month	6	19	5	11	9	15	10.8

5.2. Fitting Ability Validation of the Model

Overfitting occurs when a model adapts excessively to noise and idiosyncrasies in the training data, leading to significantly degraded performance on the test set. The key indicator is a large discrepancy between training and test errors (i.e., low training error but substantially higher test error). To rigorously assess overfitting risks, we compared the performance of our model on the training and test sets.

Experimental validation details: As data were split using 1956-01 as the partition point, the training set is from SC 1 to SC 18, and the test set is from SC 19 to SC 25. Extended analysis for the training set is included in Fig. 5, and a comparison of MAE errors in the training and test sets is given in Table 6. SC 1 and SC 2 cannot be predicted because of a lack of input data, and

Visible figure panels: SC 9, SC 10, SC 11, SC 12, SC 13, SC 14, SC 15, SC 16, SC 17, and SC 18. Axes: Year; SSN. Legends: Actual; TiDE.

Fig. 5. Model performance on the training set (from SC 3 to SC 18).

Table 6. Error comparison between the training set and test set

Training set	Training set	Test set	Test set
SC	MAE	SC	MAE
SC 3	36.08	SC 20	22.27
SC 4	16.96	SC 21	30.47
SC 5	8.57	SC 22	10.67

Training set	Training set	Test set	Test set
SC 6	22.30	SC 23	13.16
SC 7	12.96	SC 24	13.33
SC 8	18.35	SC 25	18.72
SC 9	23.02	AVG	18.10
SC 10	13.42		
SC 11	17.82		
SC 12	14.99		
SC 13	18.32		
SC 14	6.24		
SC 15	15.28		
SC16	11.32		
SC 17	22.84		
SC 18	15.27		
AVG	17.11		

SC 19 is treated as an outlier. The results show that the average MAE difference between the two sets is approximately 5%, indicating no significant overfitting. This consistency between training and test performance confirms the generalizability of the model.

5.3. Investigating the Impact of Model Regularization on Output Volatility

Compared with the raw observational data, the output predictions of the TiDE model show noticeable volatility. To investigate whether this fluctuation stems from model overfitting, we conducted experiments with three distinct dropout rates ($p = 0.1, 0.5,$ and 0.9) for rigorous validation. As shown in Fig. 6, the high regularization condition ($p = 0.9$), which would typically suppress overfitting-related fluctuations, did not significantly reduce output volatility. This empirical evidence suggests that the observed fluctuations likely reflect inherent characteristics of solar activity rather than model overfitting.

We maintain that the observed fluctuations inherently reflect the multiscale nature of solar activity (e.g., events such as flares or coronal mass ejections). While

Fig. 6. Model fitting performance, compared with observational data, under different dropout rates ($p = 0.1, 0.5, 0.9$) for representative SC 15–SC 17.

the 13-month smoothing suppresses measurement noise, it preserves physically genuine short-term activity signals. The architecture of the TiDE model demonstrates some capability in recovering these authentic, albeit smoothed, physical fluctuations. This gives TiDE an advantage in capturing true solar activity dynamics, rather than representing a model limitation.

5.4. Forecasting SC 25 and SC 26 using the Proposed Model

Finally, we used the trained model to forecast the ongoing SC 25 and the upcoming SC 26. The input for the model consisted of SSN data from the past 240 months, used to predict the SSN of the subsequent 120 months. Fig. 7 shows the prediction for SC 25 and SC 26. We implemented the Monte Carlo Dropout algorithm to check the uncertainty quantification in SC prediction. This is used during the prediction stage, in which dropout layers were kept active, enabling stochastic forward passes to generate varied outputs and quantify prediction uncertainty. Through 500 Monte Carlo Dropout iterations (n_samples = 500), we obtained uncertainty estimates. The peak for

Fig. 7. SSN predictions of SC 25 and SC 26 by our model. Major tick marks on the abscissa correspond to January for each year shown. The prediction horizon begins at December 2024, shown by the red vertical dashed line. The blue-shaded region represents the prediction distribution, and the solid blue line shows median predictions.

SC 25 is predicted to have ended. For SC 26, we predict that the peak amplitude will be 199.3, with a 95% confidence interval (CI) of 170.8–221.9, expected to occur in April 2034.

6. CONCLUSION

We have trained an optimized TiDE network for SC forecasting. The TiDE model we employed abandons the complex structure of the Transformer and instead adopts the encoder-decoder architecture based on MLP. In our experiments, we made predictions on a test set from SC 19 to SC 25 that demonstrated consistency with true observational values. The average prediction errors, with MAPE = 32.02, MAE = 23.32, RMSE = 30.3, $R^2 = 0.76$, outperforming previously used deep learning models (N_BEATS, RNN, CNN, LSTM, BiLSTM, GRU) in accuracy. In the future, more data can be used to achieve better results and smaller errors. However, the model shows limitations in accurately capturing dual-peak structures and extreme solar cycles (e.g., SC 19). These discrepancies primarily stem from the data constraints (with only 2119 available training samples) and model underfitting issues. This can potentially be addressed by acquiring more observational data and incorporating multi-source solar activity records (including sources such as radio, extreme ultraviolet, and magnetic data) into the training framework. We have also forecasted SC 25 and SC 26. For SC 25, the peak has ended, and the peak of SC 26 is predicted to be stronger, at 199.3, varying between 170.8–221.9, and is projected to occur during April 2034.

Our work will help in the development of more accurate space weather early-warning systems, with the forecasted solar peak amplitude as a deterministic factor for the occurrence rate and magnitude of geomagnetic disturbances and solar energetic particle events. This can help to ensure the operational security of spacecraft in orbit and deep space exploration missions.

ACKNOWLEDGEMENTS

The SSN data in our experiments are provided by WDC-SILSO, Royal Observatory of Belgium, Brussels, for which we are very grateful. This research is supported by the Academic Research Projects of Beijing Union University (ZK20202204); the National Natural Science Foundation of China (12250005, 12073040, 12273059, 11973056, 12003051, 11573037, 12073041, 11427901, 11572005, 11611530679 and 12473052); the Strategic Priority Research Program of the China Academy of Sciences (XDB056000, XDA15052200, XDB09040200, XDA15010700, XDB0560301, and XDA15320102); and the Chinese Meridian Project (CMP).

AI DISCLOSURE STATEMENT

Deepseek was employed for language and grammar checks within the article. The authors carefully reviewed, edited, and revised the Deepseek-generated texts to their own preferences, assuming ultimate responsibility for the content of the publication.

AUTHOR CONTRIBUTIONS

Cui Zhao conceived the ideas, designed the study, and reviewed the paper. Shangbin Yang provided the research background and significance, and proposed the research ideas. Jianguo Liu performed the experiment and wrote the draft. Shiyuan Liu performed the statistical analysis and revised the paper. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

[1] Hathaway, D. H. 2015. The solar cycle. *Living Reviews in Solar Physics*, **12**(1): 4.

[2] Balogh, A., Hudson, H., Petrovay, K., et al. 2015. Introduction to the solar activity cycle: overview of causes and consequences. *Space Science Reviews*, **186**(1): 1-15.

[3] Schatten, K. 2003. Solar activity and the solar cycle. *Advances in Space Research*, **32**(4): 451-460.

[4] Campbell, R. J., Lowry, S. 2012. Weather-related power outages and electric system resiliency. Washington: Congressional Research Service.

[5] Patt, A., Pfenninger, S., Lilliestam, J. 2013. Vulnerability of solar energy infrastructure and output to climate change. *Climatic Change*, **121**: 93-102.

[6] Kappenman, J. G. 2005. Space weather and the vulnerability of electric power grids: An overview of the increasing vulnerability trends of modern electric power grid infrastructures and the potential consequences of extreme space weather environments. In Effects of space weather on technology infrastructure. Kluwer Academic Publishers, 257–286.

[7] Daglis, I. A. 2001. Space storms and space weather hazards. Berlin: Springer Science & Business Media.

[8] Teskey, R., Wertin, T., Bauweraerts, I., et al. 2015. Responses of tree species to heat waves and extreme heat events. *Plant, Cell & Environment*, **38**(9): 1699–1712.

[9] Upton, L. A., Hathaway, D. H. 2018. An updated solar cycle 25 prediction with AFT: The modern minimum. *Geophysical Research Letters*, **45**(16): 8091–8095.

[10] Schatten K, Pesnell W D. 2007. Solar cycle 24 and the solar dynamo. In Proceedings of the 20th International Symposium on Space Flight Dynamics.

[11] Kitiashvili, I., Kosovichev, A. 2008. Application of data assimilation method for predicting solar cycles. *The Astrophysical Journal*, **688**(1): 49.

[12] McIntosh, S. W., Chapman, S., Leamon, R. J., et al. 2020. Overlapping magnetic activity cycles and the sunspot number: Forecasting sunspot cycle 25 amplitude. *Solar Physics*, **295**: 163.

[13] Bhowmik, P., Nandy, D. 2018. Prediction of the strength

and timing of sunspot cycle 25 reveal decadal-scale space environmental conditions. *Nature Communications*, **9**(1): 5209.

[14] Jiang, J., Wang, J. X., Jiao, Q. R., et al. 2018. Predictability of the solar cycle over one cycle. *The Astrophysical Journal*, **863**(2): 159.

[15] Pesnell, W. D., Schatten, K. H. 2018. An early prediction of the amplitude of solar cycle 25. *Solar Physics*, **293**(7): 112.

[16] Petrovay, K. 2020. Solar cycle prediction. *Living Reviews in Solar Physics*, **17**(1): 2.

[17] Nagovitsyn, Y. A., Pevtsov, A. A., Osipova, A. A. 2017. Long-term variations in sunspot magnetic field-area relation. *Astronomische Nachrichten*, **338**(1): 26–34.

[18] Kilcik, A., Yurchyshyn, V., Donmez, B., et al. 2018. Temporal and periodic variations of sunspot counts in flaring and non-flaring active regions. *Solar Physics*, **293**(4): 63.

[19] Hawkes, G., Berger, M. A. 2018. Magnetic helicity as a predictor of the solar cycle. *Solar Physics*, **293**(7): 109.

[20] Gopalswamy, N., Mäkelä, P., Yashiro, S., et al. 2018. Long-term solar activity studies using microwave imaging observations and prediction for cycle 25. *Journal of Atmospheric and Solar-Terrestrial Physics*, **176**: 26–33.

[21] Petrovay, K., Talafha, M. 2019. Optimization of surface flux transport models for the solar polar magnetic field. *Astronomy & Astrophysics*, **632**: A87.

[22] Petrie, G., Ettinger, S. 2017. Polar field reversals and active region decay. *Space Science Reviews*, **210**(1): 77–108.

[23] Whitbread, T., Yeates, A. R., Muñoz-Jaramillo, A. 2018. How many active regions are necessary to predict the solar dipole moment? *The Astrophysical Journal*, **863**(2): 116.

[24] Mandal, S., Karak, B. B., Banerjee, D. 2017. Latitude distribution of sunspots: analysis using sunspot data and a dynamo model. *The Astrophysical Journal*, **851**(1): 70.

[25] Jiang, J., Zhang, Z. B., Petrovay, K. 2023. Comparison of physics-based prediction models of solar cycle 25. *Journal of Atmospheric and Solar-Terrestrial Physics*, **243**: 106018.

[26] Bhowmik, P., Jiang, J., Upton, L., et al. 2023. Physical models for solar cycle predictions. *Space Science Reviews*, **219**(5): 40.

[27] Dikpati, M., de Toma, G., Gilman, P. A. 2006. Predicting the strength of solar cycle 24 using a flux-transport dynamo-based tool. *Geophysical Research Letters*, **33**(5).

[28] Pesnell W D. 2012. Solar cycle predictions. *Solar Physics*, **281**(1): 507–532.

[29] Hathaway, D. H., Upton, L. A. 2016. Predicting the amplitude and hemispheric asymmetry of solar cycle 25 with surface flux transport. *Journal of Geophysical Research: Space Physics*, **121**(11): 10744–10753.

[30] Lee, T. 2020. EMD and LSTM hybrid deep learning model for predicting sunspot number time series with a cyclic pattern. *Solar Physics*, **295**(6): 82.

[31] Benson, B., Pan, W. D., Prasad, A., et al. 2020. Forecasting solar cycle 25 using deep neural networks. *Solar Physics*, **295**(5): 65.

[32] Wang, Q. J., Li, J. C., Guo, L. Q. 2021. Solar cycle prediction using a long short-term memory deep learning model. *Research in Astronomy and Astrophysics*, **21**(1): 012.

[33] Prasad, A., Roy, S., Sarkar, A., et al. 2022. Prediction of solar cycle 25 using deep learning based long short-term memory forecasting technique. *Advances in Space Research*, **69**(1): 798–813.

[34] Dai, S. C., Liu, Y. R., Meng, J. 2021. Sunspot forecast using Temporal Convolutional Neural (TCN) network based on phase space reconstruction. In

Proceedings of 2021 33rd Chinese Control and Decision Conference (CCDC), IEEE, 2895–2900.

[35] Su, X., Liang, B., Feng, S., et al. 2024. Solar cycle prediction using a combinatorial deep learning model. *Monthly Notices of the Royal Astronomical Society*, **527**(3): 5675–5682.

[36] Pala, Z., Atici, R. 2019. Forecasting sunspot time series using deep learning methods. *Solar Physics*, **294**(5): 50.

[37] Dang, Y. C., Chen, Z. Q., Li, H., et al. 2022. A comparative study of non-deep learning, deep learning, and ensemble learning methods for sunspot number prediction. *Applied Artificial Intelligence*, **36**(1): 2074129.

[38] Su, X., Liang, B., Feng, S., et al. 2023. Solar cycle 25 prediction using N-BEATS. *The Astrophysical Journal*, **947**(2): 50.

[39] Cao, J., Xu, T. T., Deng, L. H., et al. 2024. An improved prediction of solar cycles 25 and 26 using the informer model: Gnevyshev Peaks and North–South Asymmetry. *The Astrophysical Journal*, **969**(2): 120.

[40] Cao, J., Xu, T. T., Deng, L. H., et al. 2025. Hemispheric prediction of solar cycles 25 and 26 from multivariate sunspot time-series data via Informer models. *Astronomical Techniques and Instruments*, **2**(1): 16–26.

[41] Wen, Q. S., Zhou, T., Zhang, C. L., et al. 2022. Transformers in time series: A survey. *arXiv*: 2202.07125.

[42] Das, A., Kong, W., Leach, A., et al. 2023. Long-term forecasting with tide: Time-series dense encoder. *arXiv*: 2304.08424.

Note: Figure translations are in progress. See original paper for figures.