doi:10.6041/j.issn.1000-1298.2025.07.025

MPMS-SGH: Multi-parameter Multi-step Prediction Model for Solar Greenhouse

JI Ronghua　WANG Wenxuan　AN Dong　QI Shaotian　LIU Jincun
(*College of Information and Electrical Engineering*, *China Agricultural University*, *Beijing* 100083, *China*)

Abstract: Accurately predicting environmental parameters in solar greenhouses is crucial for achieving precise environmental control. In solar greenhouses, temperature, humidity, and light intensity are crucial environmental parameters. The monitoring platform collected data on the internal environment of the solar greenhouse for one year, including temperature, humidity, and light intensity. Additionally, meteorological data, comprising outdoor temperature, outdoor humidity, and outdoor light intensity, was gathered during the same time frame. The characteristics and interrelationships among these parameters were investigated by a thorough analysis. The analysis revealed that environmental parameters in solar greenhouses displayed characteristics such as temporal variability, non-linearity, and periodicity. These parameters exhibited complex coupling relationships. Notably, these characteristics and coupling relationships exhibited pronounced seasonal variations. The multi-parameter multi-step prediction model for solar greenhouse (MPMS-SGH) was introduced, aiming to accurately predict three key greenhouse environmental parameters, and the model had certain seasonal adaptability. MPMS-SGH was structured with multiple layers, including an input layer, a preprocessing layer, a feature extraction layer, and a prediction layer. The input layer was used to generate the original sequence matrix, which included indoor temperature, indoor humidity, indoor light intensity, as well as outdoor temperature and outdoor light intensity. Then the preprocessing layer normalized, decomposed, and positionally encoded the original sequence matrix. In the feature extraction layer, the time attention mechanism and frequency attention mechanism were used to extract features from the trend component and the seasonal component, respectively. Finally, the prediction layer used a multi-layer perceptron to perform multi-step prediction of indoor environmental parameters (i.e. temperature, humidity, and light intensity). The parameter selection experiment evaluated the predictive performance of MPMS-SGH on input and output sequences of different lengths. The results indicated that with a constant output sequence length, the prediction accuracy of MPMS-SGH was firstly increased and then decreased with the increase of input sequence length. Specifically, when the input sequence length was 100, MPMS-SGH had the highest prediction accuracy, with RMSE of 0.22℃, 0.28%, and 250 lx for temperature, humidity, and light intensity, respectively. When the length of the input sequence remained constant, as the length of the output sequence increased, the accuracy of the model in predicting the three environmental parameters was continuously decreased. When the length of the output sequence exceeded 45, the prediction accuracy of MPMS-SGH was significantly decreased. In order to achieve the best balance between model size and performance, the input sequence length of MPMS-SGH was set to be 100, while the output sequence length was set to be 35. To assess MPMS-SGH's performance, comparative experiments with four prediction models were conducted: SVR, STL-SVR, LSTM, and STL-LSTM. The results demonstrated that MPMS-SGH surpassed all other models, achieving RMSE of 0.15℃ for temperature, 0.38% for humidity, and 260 lx for light intensity. Additionally, sequence decomposition can contribute to

Received date: 2024-09-12　Accepted date: 2024-12-12
Foundation item: Beijing's Modern Agricultural Industry Technology System (BAIC01-2024-19)
First author: JI Ronghua, associate professor. E-mail: jessic1212@cau.edu.cn
Corresponding author: LIU Jincun, associate professor. E-mail: liujincun@cau.edu.cn

enhancing MPMS-SGH's prediction performance. To further evaluate MPMS-SGH's capabilities, its prediction accuracy was tested across different seasons for greenhouse environmental parameters. MPMS-SGH had the highest accuracy in predicting indoor temperature and the lowest accuracy in predicting humidity. And the accuracy of MPMS-SGH in predicting environmental parameters of the solar greenhouse fluctuated with seasons. MPMS-SGH had the highest accuracy in predicting the temperature inside the greenhouse on sunny days in spring (R^2 = 0.91), the highest accuracy in predicting the humidity inside the greenhouse on sunny days in winter (R^2 = 0.83), and the highest accuracy in predicting the light intensity inside the greenhouse on cloudy days in autumm (R^2 = 0.89). MPMS-SGH had the lowest accuracy in predicting three environmental parameters in a sunny summer greenhouse.

Key words: solar greenhouse; environmental parameter; time series; multi-step prediction
CLC number: S265 **Document code**: A **Article ID**: 1000-1298(2025)07-0265-14 **OSID**:

0 Introduction

The solar greenhouse is a distinctive greenhouse structure in northern China. As of 2022, the total area of China's facility horticulture exceeded 2.8 million hectares, accounting for more than 80% of the world's total facility area. Among them, solar greenhouses covered 0.81 million hectares, accounting for 29%. According to the statistics from Beijing Facility Agriculture Information Management System, by the end of 2022, Beijing had a total of 184 000 facilities for agriculture, of which 99 000 were solar greenhouses, accounting for 53.8% of the total. Adjusting the key environmental parameters in solar greenhouses to best meet crop growth requirements can improve crop quality. Currently, the solar greenhouse environmental parameters were primarily regulated relying on growers' experience or single-parameter, single-step prediction models[1-3]. However, these methods were inadequate for achieving precise solar greenhouse environmental control. The primary approaches for predicting environmental parameters in solar greenhouses were traditional machine learning algorithms[4-10] and deep learning algorithms[11-19]. Specifically, traditional machine learning algorithms can achieve high accuracy when predicting single parameter. When using traditional machine learning algorithms for predicting multiple parameters of greenhouse environment, the accuracy of multi-step prediction is low due to error accumulation[17-19]. Deep learning algorithms, with their robust nonlinear approximation and deep feature extraction capabilities, can more effectively uncover the correlations and dependencies among multivariate series data, achieving higher prediction accuracy in multivariate series prediction tasks in related fields[20-21].

Greenhouse environmental parameters were characterized by nonlinearity, complex couplings, and temporal variability. Specifically, the distribution characteristics of greenhouse environmental parameters vary in different seasons. A multi-parameter multi-step prediction model for solar greenhouse (MPMS-SGH) was proposed, aiming at overcoming the challenges of limited prediction accuracy and poor seasonal adaptability in existing models for greenhouse environmental parameters. MPMS-SGH was employed to forecast temperature, humidity, and light intensity simultaneously, all of which were essential environmental factors influencing crop growth in solar greenhouses[22-23]. A variety of environmental data was collected from inside and outside the solar greenhouse through the monitoring platform. Based on a thorough analysis of the data characteristics and their interrelationships, the structure of MPMS-SGH was designed. Various experiments were conducted to verify the performance of MPMS-SGH. MPMS-SGH was designed specifically for multi-step prediction of various parameters of solar greenhouse environments. MPMS-SGH would provide support for the development of intelligent regulation strategies for greenhouse environments.

1 Materials and methods

1.1 Data collection

Environmental parameters were continuously monitored in solar greenhouses A5 and B5 at Beijing Fangshan Hongke Farm (39°62′N, 115°97′E) from December 1st, 2022, through December 1st, 2023.

Each solar greenhouse is 80 m in length from east to west, 9.2 m in width from north to south, and has a ceiling height of 4 m. Equipment such as windows, fans, wet curtains, shading curtains, and heating systemswas used to regulate the solar greenhouse environment, ensuring that it was the most suitable for crop growth.

The greenhouses were equipped with Kunlun Coast JZH－025－12 sensors, which monitored temperature, humidity, and light intensity. In addition, Kunlun Coast JHW－504 sensors were placed outside to record meteorological data, including air temperature, humidity, and light intensity. The specifications of the sensors were detailed in Table 1. To ensure that the collected environmental parameters in the greenhouse can effectively reflect their spatial distribution. The sensors were arranged as shown in Fig. 1. The data were collected continuously from 00:00 to 24:00 every day, with a sampling interval of 10 mins.

Table 1　Sensor specifications

Model number	Parameter	Range	Accuracy
JZH－025－12	Temperature/℃	－20～60	±0.5
	Humidity/%	0～100	±3
	Light intensity/lx	0～200 000	±5 000
JHW－504	Temperature/℃	－20～60	±0.5
	Humidity/%	0～100	±3
	Light intensity/lx	0～200 000	±5 000

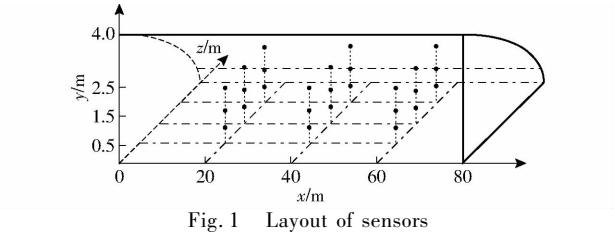

Fig. 1　Layout of sensors

1.2　Data processing

The data obtained from 27 indoor sensors indicated that the spatial distribution of environmental parameters in solar greenhouses was relatively uniform in spring, autumn, and winter. However, notable variation of the spatial distribution of environmental parameters in solar greenhouses was observed in the summer, particularly in terms of vertical variation. During summer, temperature was observed to be higher at the top and lower near the bottom, with humidity exhibiting an inverse trend. In addition, it was noted that the top had the highest light intensity, which gradually decreased as the height decreased. Therefore, different data selection strategies were adopted for the sample set data according to different seasons. In spring, autumn, and winter, the average value of all sampling points was selected for environmental data. In summer, data from the central sampling point was selected.

In order to ensure the accuracy and reliability of environmental parameters in solar greenhouses, missing data needed to be processed before normalizing the dataset. According to the actual situation during the data collection process, there were two types of missing data. The first type was continuously missing data, lasting up to 7 h, caused by unexpected nighttime power outages. The second type was isolated missing data, which was caused by unpredictable sensor issues.

Continuous missing data was filled with corresponding data from neighboring greenhouse, while isolated missing data was filled with the average of two adjacent time points. Considering the significant differences in the dimension of temperature, humidity, and light intensity, formula (1) was used for data normalization to eliminate the impact of dimensional differences on MPMS-SGH.

$$x^{(i)} = \frac{x^i - x^i_{\min}}{x^i_{\max} - x^i_{\min}} \tag{1}$$

Where　$x^{(i)}$—The i_{th} environmental parameter after normalization

x^i—The original value of the i_{th} environmental parameter collected. The subscript i took on the values 1, 2, 3, 4, 5, and 6, which corresponded to indoor temperature, indoor humidity, indoor light intensity, outdoor temperature, outdoor light intensity, and outdoor humidity, respectively

$x^i_{\min}$—The minimum value of the i_{th} environmental parameter

$x^i_{\max}$—The maximum value of the i_{th} environmental parameter

1.3　Data analysis

Understanding the change rules of environmental parameters within solar greenhouses in Beijing was crucial for enhancing the adaptability of forecasting models across different seasons. The division of the

four seasons (spring, summer, autumn, and winter) was defined by weather fluctuations, precipitation trends, monsoon influences, and the actual state of agricultural production.

According to the meteorological industry standards of Beijing Meteorological Bureau and climate data for 2022, the four seasons of 2022 were divided as follows: spring from March 20th to June 3rd, summer from June 4th to September 14th, autumn from September 15th to November 15th, and winter from November 16th to March 19th.

The environmental parameters in the solar greenhouse were analyzed statistically for each season, with the results presented in Table 2.

Table 2 Statistical information of environmental parameters in solar greenhouse during different seasons

Environmental parameter	Season	Mean	Variance
Temperature	Spring	22.62℃	40.46℃^{2}
	Summer	31.72℃	63.66℃^{2}
	Autumn	24.01℃	75.70℃^{2}
	Winter	16.14℃	89.06℃^{2}
Humidity	Spring	62.80%	352.54%^{2}
	Summer	63.55%	440.73%^{2}
	Autumn	58.97%	510.85%^{2}
	Winter	74.13%	362.18%^{2}
Light intensity	Spring	7 310 lx	211.68 lx^2
	Summer	2 000 lx	40.98 lx^2
	Autumn	8 200 lx	71.06 lx^2
	Winter	5 080 lx	153.82 lx^2

The highest temperatures were observed during summer, while the lowest temperatures were recorded during winter. Winter exhibits the most pronounced temperature fluctuations. The humidity was he highest in winter and lowest in autumn. The humidity changed greatly in winter. The light intensity reached its maximum in summer. The light intensity changed significantly in autumn. The environmental parameters within solar greenhouses were the most stable during the spring season. The greatest fluctuations in humidity and light intensity occurred in autumn, while temperature variations were the most pronounced in winter.

Visualized variations in the three environmental parameters within the solar greenhouse following data preprocessing were shown in Fig. 2. Fig. 2a showed the changes in environmental parameters within one day, while Fig. 2b showed the changes in environmental parameters for five consecutive days.

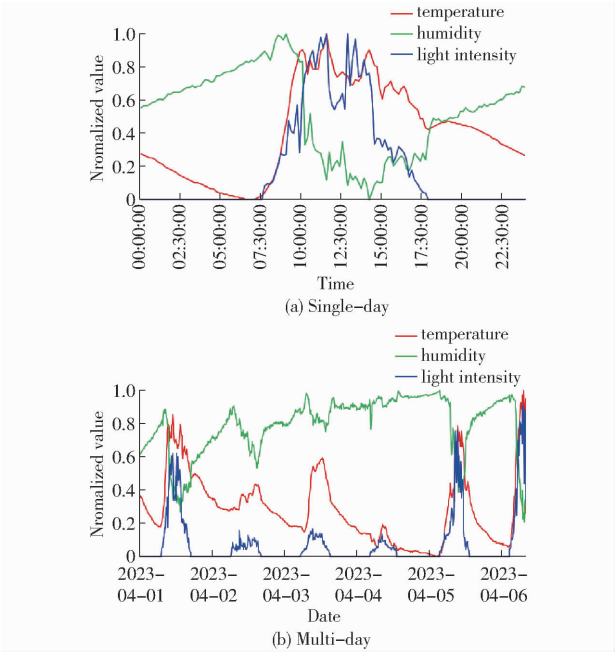

Fig. 2 Distribution of the solar greenhouse environmental parameters

The environmental parameters of solar greenhouses exhibited nonlinear changes over time. Each parameter change exhibited a certain degree of periodicity, with a period of one day. Within a day, the temperature and light intensity in a solar greenhouse exhibited similar patterns of change, while humidity showed the opposite pattern of change.

The STL algorithm was used to decompose the environmental parameters inside the solar greenhouse into three components: trend component, seasonal component, and residual component. The trend component was used to capture the overall trend of environmental parameter over time. The seasonal component was used to obtain periodic changes in environmental parameters, while the residual component was the part of the environmental parameter data other than the trend component and seasonal component.

Fig. 3 showed the changes in environmental parameters and their decomposed components. The first subgraph showed the changes in the raw data for each parameter, the second subgraph showed the variation of the trend component of the parameter, the third subgraph showed the variation of the seasonal component, and the fourth subgraph showed the variation of the residual component.

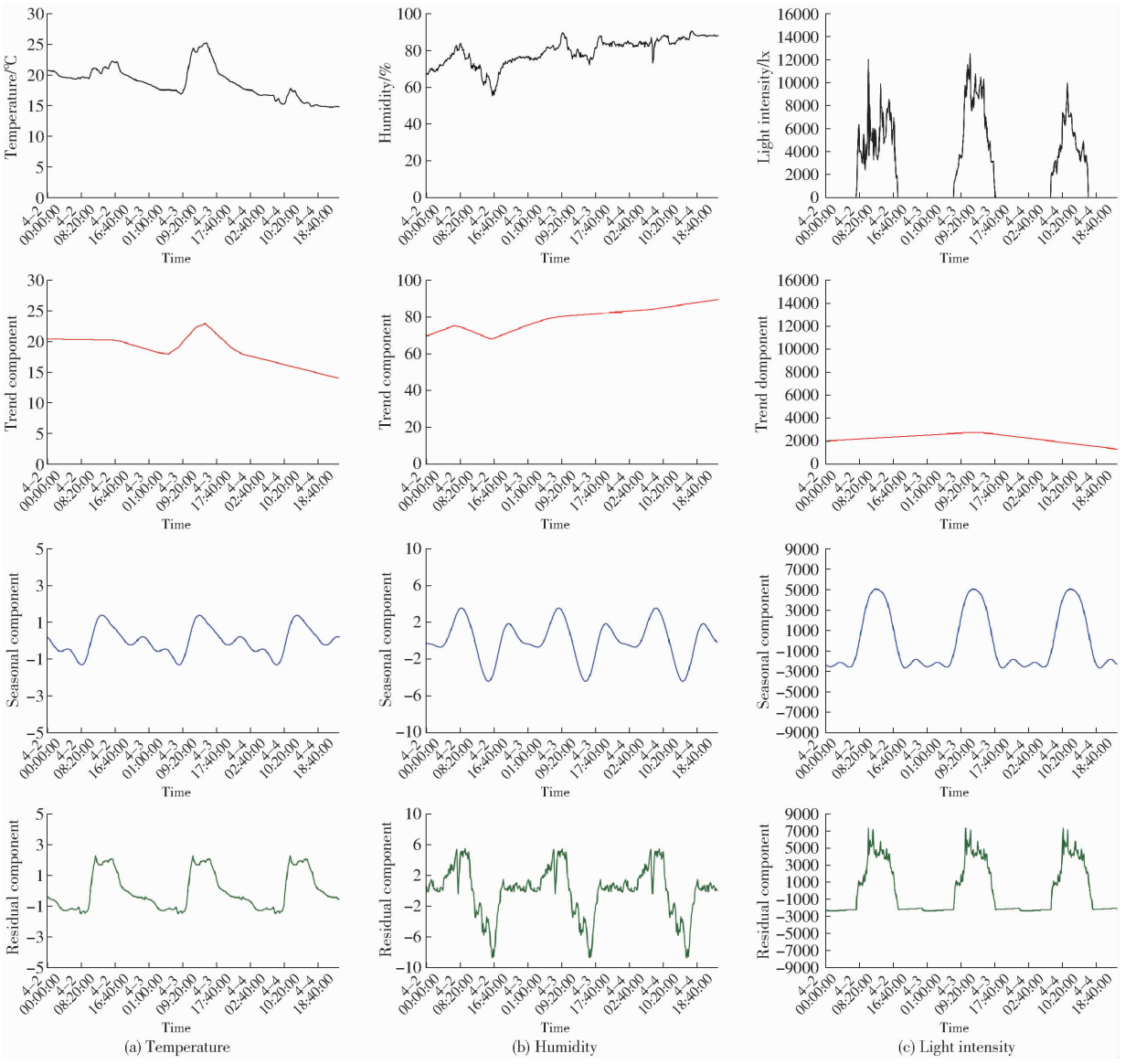

Fig. 3　Distribution of solar greenhouse environmental parameters after decomposition

From Fig. 3, it can be seen that after applying STL to decompose the environmental parameters of the solar greenhouse into three components, each component had its unique characteristics: the trend component showed a relatively simple stable trend, the seasonal component showed significant periodic features, and the residual component had a similar change pattern to the original data with a decrease in amplitude.

In order to further explore the interrelationships between the internal environmental parameters of the greenhouse and their coupling effects with outdoor meteorological data, the Spearman correlation coefficients between the three environmental parameters inside the solar greenhouse and the Spearman correlation coefficients between three environmental parameters and outdoor meteorological data in each season were calculated. The result was shown in Fig. 4.

The results indicated that the mutual relationship between indoor environmental parameters was stronger than the coupling relationship between indoor environmental parameters and outdoor meteorological data. And the coupling relationship among these data differed across various seasons.

In spring and autumn, the distribution of correlation coefficients between parameters was similar. The parameter paired with a correlation coefficient greater than 0.5 included: indoor temperature and indoor humidity, indoor temperature and indoor light intensity, indoor temperature and outdoor temperature, indoor light intensity and outdoor light intensity. The correlation coefficient between indoor humidity and indoor light intensity was less than 0.5, indicating that

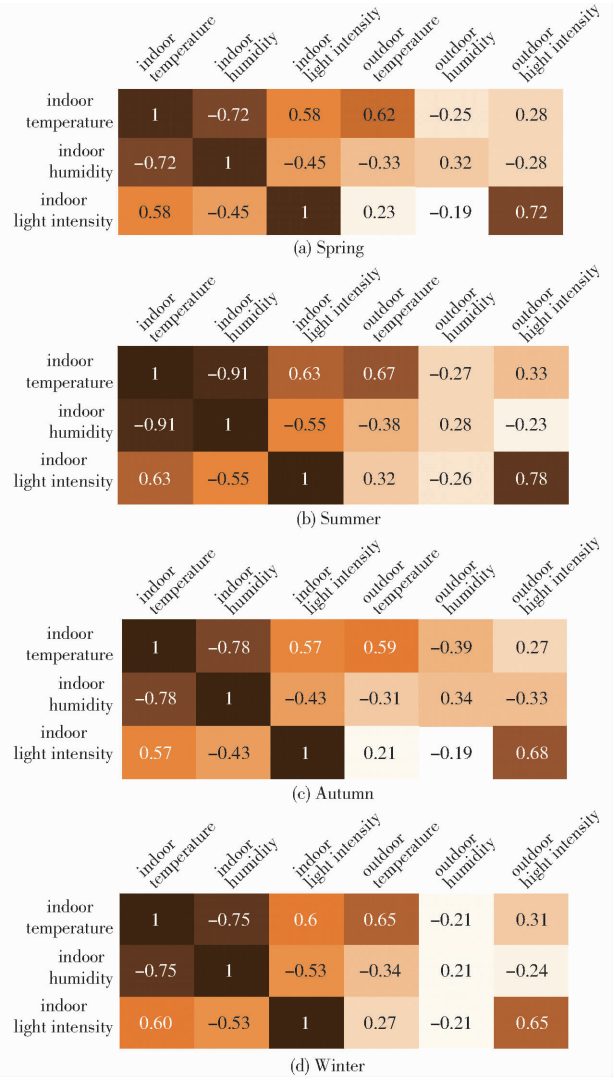

Fig. 4 Spearman correlation coefficients of parameters of solar greenhouse in various seasons

the coupling relationship between the two was weakened in spring and autumn.

The distribution of correlation coefficients between parameters was similar in summer and winter. The parameter paired with a correlation coefficient greater than 0.5 were following: indoor temperature and indoor humidity, indoor temperature and indoor light intensity, indoor humidity and indoor light intensity, indoor temperature and outdoor temperature, indoor light intensity and outdoor light intensity. The above parameter pairs have higher correlation coefficients in summer than in winter, indicating a tighter coupling relationship between the parameters in summer.

The correlation coefficient between indoor temperature and outdoor temperature was greater than 0.5, indicating that changes in outdoor temperature had an undeniable impact on indoor temperature. The correlation coefficient between indoor and outdoor light intensity was greater than 0.5, indicating that changes in outdoor light intensity affect indoor light intensity to some extent.

In summary, the three environmental parameters (temperature, humidity, and light intensity) in a solar greenhouse exhibited time-varying, nonlinear, and periodic characteristics, as well as complex coupling relationships. In addition, outdoor temperature and outdoor light intensity had a significant impact on indoor parameters, and their effects vary by season.

1.4 MPMS-SGH architecture

Considering the different distribution characteristics of temperature, humidity, and light intensity in solar greenhouses, a multi-step prediction model (MPMS-SGH) specifically designed for predicting various parameters in solar greenhouses was proposed. The results predicted by MPMS-SGH would provide data support for precise regulation of solar greenhouse environment. MPMS-SGH consisted of four layers as follows: input layer, preprocessing layer, feature extraction layer, and prediction layer. The overall structure was shown in Fig. 5.

Indoor temperature, indoor humidity, indoor light intensity, outdoor temperature, and outdoor light intensity were selected as inputs for MPMS-SGH. The input of MPMS-SGH was denoted as the original sequence matrix ($\boldsymbol{X}$), as shown in formula (2).

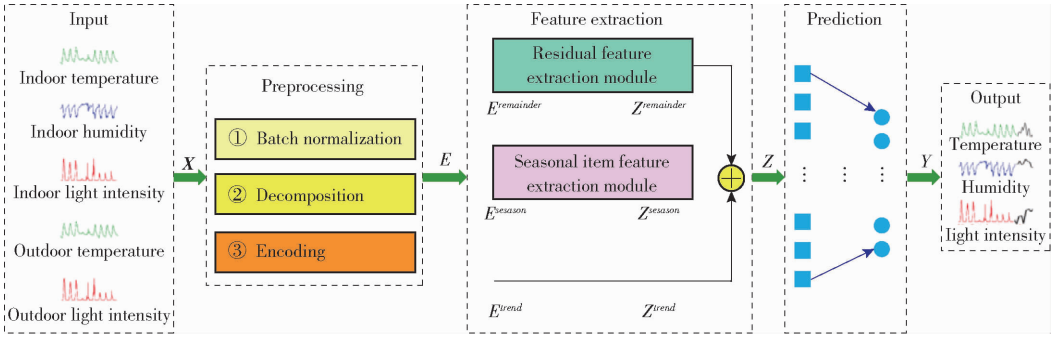

Fig. 5 Overall structure of MPMS-SGH

$$\boldsymbol{X}=\begin{bmatrix} x_{t-j-1,1} & \cdots & x_{t,1} \\ x_{t-j-1,2} & \cdots & x_{t,2} \\ x_{t-j-1,3} & \cdots & x_{t,3} \\ x_{t-j-1,4} & \cdots & x_{t,4} \\ x_{t-j-1,5} & \cdots & x_{t,5} \end{bmatrix} \tag{2}$$

The size of $\boldsymbol{X}$ was $5\times j$, where j was the length of the input sequence. $x_{t-j-1,i}$ was the value of the i_{th} environmental parameter at $t-j-1$. $i=1$ standed for indoor temperature, $i=2$ for indoor humidity, $i=3$ for indoor light intensity, $i=4$ for outdoor temperature, $i=5$ for outdoor light intensity.

The preprocessing layer converted $\boldsymbol{X}$ into a data format suitable for MPMS-SGH, including normalization, decomposition, and encoding.

(1) Normalization. By standardizing the distribution of input data, MPMS-SGH training can be accelerated and performance can be improved. After normalizing $\boldsymbol{X}$, B was obtained, as shown in formula (3).

$$B=\mathrm{BN}(\boldsymbol{X}) \tag{3}$$

Where BN standed for batch normalization.

(2) Decomposition. B was decomposed into the trend component B^{trend}, seasonal component B^{season}, and residual component $B^{remainder}$ by the STL algorithm[22].

$$\begin{bmatrix} B^{trend} \\ B^{season} \\ B^{remainder} \end{bmatrix}=decomp(B) \tag{4}$$

Where *decomp* referred to the decomposition of B.

(3) Encoding. The environmental parameters in solar greenhouses fluctuated over time. Time played a certain role in predicting the environmental parameters at the current time. Therefore, it was considered to incorporate time information (i. e. location) into the environmental parameter sequence to improve the performance of MPMS-SGH.

There were various methods for encoding position information, including sine-cosine-based encoding[23] and encoding using a learnable coding matrix [24]. The sine-cosine coding approach was adopted, which allowed MPMS-SGH to interpret the distance between positions based on the differences in coding values. This not only captured the periodic features of the environmental parameters, but also enabled MPMS-SGH to understand the relative positional relationships between the data, thus providing a more comprehensive understanding of the dynamic relationships between the parameters. The positional information matrix was generated by using the sine-cosine coding method, with the calculation formula detailed below.

$$E_{pos(X,2i)}=\sin\frac{\boldsymbol{X}}{10\,000^{\frac{2i}{l_X}}} \tag{5}$$

$$E_{pos(X,2i+1)}=\cos\frac{\boldsymbol{X}}{10\,000^{\frac{2i}{l_X}}} \tag{6}$$

Where E_{pos} was the position information matrix; $\boldsymbol{X}$ was the original sequence matrix; i was the i_{th} position, and l_X was the length of the sequence.

After decomposition, E_{pos} was added to each component to obtain the encoded matrix (E). The calculation was as shown in formula (7).

$$E=\begin{bmatrix} E^{trend} \\ E^{season} \\ E^{remainder} \end{bmatrix}=\begin{bmatrix} B^{trend}+E_{pos} \\ B^{season}+E_{pos} \\ B^{remainder}+E_{pos} \end{bmatrix} \tag{7}$$

The output of the preprocessing layer was the encoded matrix (E), consisting of three components, as shown in formulas (8) ~ (10).

$$E^{trend}=\begin{bmatrix} e^{trend}_{t-j-1,1} & \cdots & e^{trend}_{t,1} \\ e^{trend}_{t-j-1,2} & \cdots & e^{trend}_{t,2} \\ e^{trend}_{t-j-1,3} & \cdots & e^{trend}_{t,3} \\ e^{trend}_{t-j-1,4} & \cdots & e^{trend}_{t,4} \\ e^{trend}_{t-j-1,5} & \cdots & e^{trend}_{t,5} \end{bmatrix} \tag{8}$$

$$E^{season}=\begin{bmatrix} e^{season}_{t-j-1,1} & \cdots & e^{season}_{t,1} \\ e^{season}_{t-j-1,2} & \cdots & e^{season}_{t,2} \\ e^{season}_{t-j-1,3} & \cdots & e^{season}_{t,3} \\ e^{season}_{t-j-1,4} & \cdots & e^{season}_{t,4} \\ e^{season}_{t-j-1,5} & \cdots & e^{season}_{t,5} \end{bmatrix} \tag{9}$$

$$E^{remainder}=\begin{bmatrix} e^{remainder}_{t-j-1,1} & \cdots & e^{remainder}_{t,1} \\ e^{remainder}_{t-j-1,2} & \cdots & e^{remainder}_{t,2} \\ e^{remainder}_{t-j-1,3} & \cdots & e^{remainder}_{t,3} \\ e^{remainder}_{t-j-1,4} & \cdots & e^{remainder}_{t,4} \\ e^{remainder}_{t-j-1,5} & \cdots & e^{remainder}_{t,5} \end{bmatrix} \tag{10}$$

The feature extraction layer was responsible for extracting and constructing the feature matrix (Z) for greenhouse environmental parameters. It consisted of two modules: a residual feature extraction module and a seasonal feature extraction module. The residual feature extraction module used the time-domain attention mechanism to extract and construct features from the residual component of each parameter,

resulting in $Z^{remainder}$, as depicted in Fig. 6. The seasonal feature extraction module, on the other hand, was specifically designed to extract and construct seasonal features, resulting in Z. The frequency-domain attention mechanism can effectively predict time series with strong periodicity[25], thus the seasonal feature extraction module employed it for seasonal component feature extraction, as shown in Fig. 7. Firstly, Fourier transform was applied to the seasonal components of environmental parameters, then the top 5 main frequencies were selected and frequency-domain attention was applied to them. Finally, inverse Fourier transform was applied to convert the seasonal components from frequency-domain to time-domain.

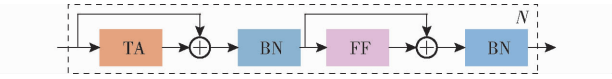

Fig. 6 Structural of residual feature extraction module

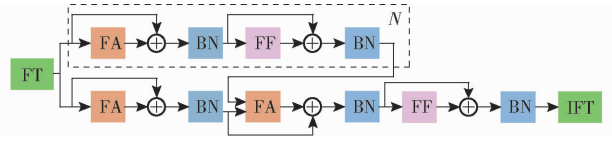

Fig. 7 Structural of seasonal feature extraction module

Where TA standed for time-domain attention, BN for batch normalization, FF for feedforward network, FT for the Fourier transform, FA for frequency-domain attention, and IFT for the inverse Fourier transform.

The characteristic matrix (Z) of greenhouse environmental parameters resulted from the summation of residual feature, seasonal feature, and trend feature.

$$Z = EBX^{trend} + Z^{season} + Z^{remainder} = Z^{trend} + Z^{season} + Z^{remainder} \quad (11)$$

Where $Z^{remainder}$ standed for the output of the residual feature extraction module, Z^{season} for the output of the seasonal feature extraction module, Z^{trend} for the trend component of Z.

The prediction layer was implemented y using a multi-layer perceptron (MLP). MLP consisted of two hidden layers, with 60 and 85 neurons respectively.

The output of MPMS-SGH was the temperature, humidity, and light intensity within the solar greenhouse from $t+1$ to $t+i$. Where i was the length of the output.

1.5 MPMS-SGH training

This dataset contained 52 560 samples. Each sample included indoor temperature, indoor humidity, indoor light intensity, outdoor temperature, and outdoor light intensity. Considering the seasonal variation of environmental parameters in the solar greenhouse, in order to maintain consistency in data distribution between train set and test set, data from specific dates listed in Table 3 were selected as test set samples, while data from other dates were selected as train set samples. The train set consisted of 43 488 samples and the test set consisted of 9 072 samples.

Table 3 Collection date for test set

Season	Date
Spring	3.20, 3.29, 3.31, 4.9, 4.10, 4.19, 4.20, 4.29, 5.9, 5.10, 5.19, 5.20, 5.25
Summer	6.4, 6.5, 6.17, 6.18, 6.28, 6.29, 7.8, 7.9, 7.18, 7.19, 7.28, 7.29, 8.7, 8.8, 9.5
Autumn	9.15, 9.16, 9.25, 9.26, 10.5, 10.6, 10.15, 10.16, 10.25, 10.26, 11.7
Winter	11.16, 11.17, 11.28, 11.29, 12.8, 12.9, 12.18, 12.19, 12.28, 12.29, 1.8, 1.9, 1.18, 1.19, 1.28, 1.29, 2.2, 2.7, 2.8, 2.17, 2.18, 2.28, 3.9, 3.10, 3.19

MPMS-SGH was implemented with PyTorch 1.7.0. The hardware setup includes an Intel Core i7 – 7700K CPU operating at 4.20 GHz, with 16 GB of RAM, and a GeForce RTX3090 GPU boasting 24 GB of memory.

The primary hyperparameters for MPMS-SGH training were set as follows. The training optimizer was Adam, the batch size was 32, the weight decay was set to 0.000 1, and the learning rate of 0.001. MPMS-SGH was trained for 300 epochs. The early stopping criterion (early-stopping) was to terminate model training if there was no loss value decrease within 10 epochs.

1.6 Evaluation metrics

To assess the prediction performance of the models on the test set, three distinct evaluation metrics were employed: the coefficient of determination (R^2), the root mean square error (RMSE), and the relative prediction error (RPD). The coefficient of determination served as a primary metric in prediction analyses, reflecting the models's fit to the data. It varied between 0 and 1, with values closer to 1 indicating superior model fit and predictive accuracy. The root mean square error inversely related to R^2; as R^2 was increased, RMSE was decreased, signifying

better prediction accuracy. Relative prediction error, calculated as the reciprocal of R^2, evaluated the model's stability. A higher RPD suggested greater model stability, whereas a lower RPD indicated reduced stability.

2 Results and discussion

Evaluating and validating the performance of MPMS-SGH through parameter selection experiments, ablation experiments, comparative experiments, and performance experiments. The models used in the experiment were trained and evaluated based on the configuration described in Section 1.5. The corresponding experimental instructions and results were as follows.

2.1 Parameter selection experiments

The length of the input and output sequences was critical parameters for MPMS-SGH. Correctly setting these lengths can reduce the complexity and lower the risk of overfitting while maintaining the prediction accuracy of MPMS-SGH. MPMS-SGH was tested with various combinations of input and output sequence lengths to find their optimal combinations. Specifically, the output sequence length was set to be 5, 15, 25, 35, and 45, while the input sequence length increased uniformly from 10 to 160 in increments of 10. The models were tested using test set to assess their prediction accuracy for three greenhouse environmental parameters.

By gradually increasing the input sequence length from 10 to 160 and fixing the output sequence length to 5, a series of MPMS-SGH models were obtained. Fig. 8a showed the accuracy changes (measured by RMSE) of these models predicting three indoor environmental parameters (temperature, humidity, and light intensity). It was evident that when the length of the input sequence reached 100, the prediction accuracy dropped to its lowest point. Specifically, the RMSE for temperature was 0.22℃, the RMSE for humidity was 0.28%, and the RMSE for light intensity was 250 lx.

Observation showed that when the output sequence length was a fixed value, the prediction accuracy of the model was increased with the increase of the input sequence length, reaching its peak at a length of 100. As the length of the input sequence continued to increase, the prediction accuracy of the model began to decrease. This may be due to the fact that as the length of the input sequence was increased, the current data in the sequence was gradually less affected by its previous data. The impact of the first 100 data points on the current data can be ignored.

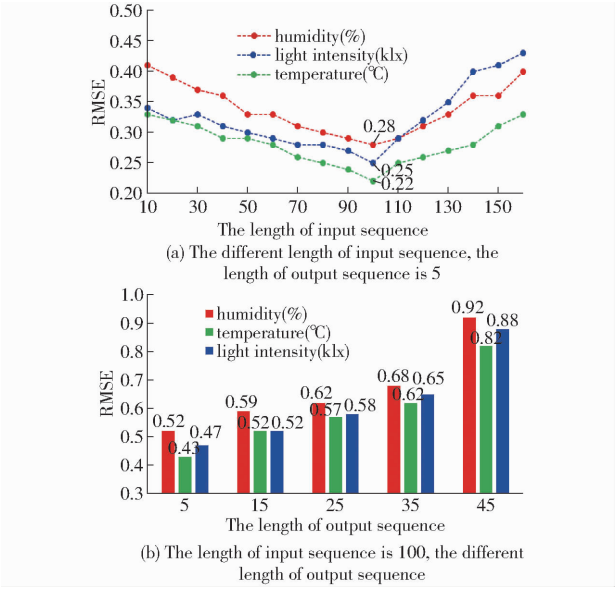

Fig. 8 Changes in prediction accuracy of MPMS-SGH under different combinations of input sequence length and output sequence length

By keeping the input sequence length fixed at 100 and gradually increasing the output sequence length from 5 to 45 in increments of 10, a set of MPMS-SGH models was obtained. Fig. 8b showed the changes in prediction accuracy of these models when predicting three indoor environmental parameters (temperature, humidity, and light intensity). When the input sequence length was 100 and the output sequence length was 5, the prediction accuracy reached its highest, and the RMSE of temperature was 0.43℃. The RMSE of light intensity was 470 lx. The RMSE of humidity was relatively low, at 0.52%. When the input sequence length was fixed, as the output sequence length was increased, the model's prediction accuracy for all three environmental parameters was slightly decreased. When the length of the output sequence reached 45, the prediction accuracy of the model was significantly decreased. This may be due to the accumulation of errors in the model during multi-step sequence prediction. When the length of the output sequence reached 45, the accumulated error was too large, resulting in a significant decrease in the prediction accuracy of the model.

2.2 Ablation experiment

Ablation experiments were performed to validate the influence of position encoding and the frequency-domain transformation module on the performance of MPMS-SGH. The input sequence length for MPMS-SGH was set to 100 and the output sequence length was set to 35. The results of the ablation experiments were shown in Table 4.

Table 4 Ablation experiment results

Model	Temperature			Humidity			Light intensity		
	RMSE/℃	R^2	RPD	RMSE/%	R^2	RPD	RMSE/lx	R^2	RPD
MPMS-SGH	0.15	0.91	7.81	0.38	0.81	6.65	260	0.89	7.63
-PE	0.38	0.77	5.21	0.62	0.61	4.51	470	0.78	4.91
-FC	0.44	0.65	4.62	0.63	0.58	4.22	540	0.64	4.42

Note: -PE represented the model generated after removing the position encoding, and -FC represented the model generated after removing the frequency-domain transformation module.

When the model predicted three environmental parameters after removing the position encoding module, the prediction accuracy was decreased. Among them, the decrease in prediction accuracy for humidity was the most significant, with R^2 decreased by 0.2, while the decrease in prediction accuracy for temperature and light intensity was relatively close, with R^2 decreased by about 0.1. This may be due to the more complex fluctuations in humidity over time, indicating that humidity was strongly influenced by position information. Therefore, position encoding played a crucial role in achieving accurate humidity prediction.

After removing the frequency-domain transformation module, the prediction accuracy of the model for three environmental parameters was significantly decreased, indicating that the frequency-domain transformation module had a significant impact on the performance of the model. Specifically, the prediction accuracy of temperature and light intensity was decreased by a similar amount, with RMSE increased by 280 lx, R^2 decreased by 0.25, and RPD decreased by 3.2. The decrease in humidity prediction accuracy was not severe, with RMSE increased by 0.25%, R^2 decreased by 0.23, and RPD decreased by 2.43. This can be explained by the stronger periodicity of temperature and light intensity sequences. At the same time, it indicated that the frequency-domain attention mechanism was more effective in extracting features from periodic sequences.

2.3 The comparative experiments

To evaluate the predictive performance of MPMS-SGH, it was compared with four models: SVR, decomposition-based SVR (STL-SVR), LSTM, and decomposition-based LSTM (STL-LSTM). Among them, support vector regression (SVR) was a powerful machine learning algorithm suitable for solving complex nonlinear regression problems. Long short-term memory (LSTM) solved the problem of gradient vanishing or exploding that traditional RNNs often encountered when processing long sequences by combining gating mechanisms. Further combined SVR and SVM with sequence decomposition to explore the impact of sequence decomposition on model performance. All models were trained using the same training dataset and hyperparameters in the software and hardware environment described in Section 1.5. The performance of these models was then evaluated using the same test dataset. The results of the comparative experiment were shown in Table 5.

Table 5 Comparative experimental results

Parameter	Model	RMSE	R^2	RPD
Temperature	MPMS-SGH	0.15℃	0.91	7.81
	SVR	0.35℃	0.69	6.23
	STL-SVR	0.29℃	0.78	7.16
	LSTM	0.35℃	0.72	4.23
	STL-LSTM	0.28℃	0.81	6.56
Humidity	MPMS-SGH	0.38%	0.81	6.65
	SVR	0.49%	0.61	3.56
	STL-SVR	0.45%	0.74	3.98
	LSTM	0.46%	0.72	4.23
	STL-LSTM	0.41%	0.74	5.14
Light intensity	MPMS-SGH	260 lx	0.89	7.63
	SVR	450 lx	0.63	4.73
	STL-SVR	410 lx	0.71	5.23
	LSTM	360 lx	0.71	5.39
	STL-LSTM	310 lx	0.79	5.87

The results indicated that MPMS-SGH exhibited

the most favorable prediction performance, while SVR demonstrated the least favorable. The predictive performance of LSTM and STL-SVR was quite comparable. Sequence decomposition helps improve model prediction performance. MPMS-SGH and STL-LSTM demonstrated the highest degree of accuracy in predicting temperature, with a slightly lower level of accuracy in predicting light intensity and the lowest accuracy in predicting humidity.

2.4 Performance experiments

According to the climate characteristics of Beijing, the environmental control measures inside the solar greenhouse should be adjusted accordingly with the seasons to ensure that the indoor environmental conditions were suitable for crop growth. In spring and autumn, the focus of greenhouse environmental regulation strategies was to maintain the optimal temperature while eliminating excessive humidity. In spring and autumn, due to unstable climate, the regulation strategy of solar greenhouse environment needed to be adjusted according to real-time weather conditions to effectively cope with the unpredictability of the climate. In summer, the focus of greenhouse environmental regulation strategies shifted to cooling and dehumidification to adapt to the high temperature and humidity characteristics of summer. In winter, the main focus was on increasing temperature inside the greenhouse. The environmental control measures for solar greenhouses usually remained relatively stable in summer and winter. Based on the previous data analysis results, it can be found that the environmental parameters inside the solar greenhouse exhibited different patterns of change in different seasons. Therefore, the focus of performance experiments was to examine the suitability of the model for seasons.

Several experiments were conducted to evaluate the seasonal adaptability of MPMS-SGH. Dividing the test set into four subsets by season. Use these four subsets to test the trained MPMS-SGH separately. Subsequently, statistical analysis on the prediction results was conducted and the average R^2 of the MPMS-SGH's prediction results was calculated under different weather conditions in seasons to evaluate the generalization ability of MPMS-SGH for different seasons. Table 6 listed the weather conditions and corresponding MPMS-SGH's prediction results in the test set.

Table 6　Prediction results of MPMS-SGH under different weather conditions in different seasons

Season	Weather condition	The mean of R^2 on test set		
		Temperature	Humidity	Light intensity
Spring	Sunny	0.91	0.81	0.85
	Cloudy	0.88	0.80	0.88
	Rainfall	0.87	0.79	0.84
Summer	Sunny	0.85	0.75	0.86
	Cloudy	0.86	0.77	0.87
	Rainfall	0.88	0.73	0.85
Autumn	Sunny	0.89	0.82	0.86
	Cloudy	0.87	0.79	0.89
	Rainfall	0.86	0.78	0.86
Winter	Sunny	0.90	0.83	0.87
	Cloudy	0.86	0.80	0.88
	Snowfall	0.86	0.77	0.86

It can be observed that MPMS-SGH exhibited the highest accuracy in predicting indoor temperature, the lowest accuracy in predicting indoor humidity. This difference may be caused by frequent fluctuations in humidity. MPMS-SGH can accurately predict indoor temperature in sunny spring ($R^2 = 0.91$), indoor humidity in sunny winter ($R^2 = 0.83$), and indoor light intensity in cloudy autumn ($R^2 = 0.89$). MPMS-SGH had the lowest accuracy in predicting three environmental parameters in a sunny summer greenhouse.

To assess the adaptability of MPMS-SGH across various seasons, a representative date from each season within the test set was chosen to showcase its predictive capabilities. Table 7 outlined the weather conditions on these selected dates, the environmental control measures applied in the solar greenhouse, and the predictive accuracy of MPMS-SGH for indoor environmental parameters on those specific days.

From Table 7 and Fig. 9, it can be observed that MPMS-SGH was capable of accurately predicting environmental parameters in solar greenhouse during spring, autumn, and winter. The predictive ability of MPMS-SGH was decreased in summer, especially during sunny summer. In summer, the environmental control equipment in solar greenhouse was more complex to regulate, which meant that the solar greenhouse environment was subject to more external intervention in summer, which may be the reason for the decrease in the prediction accuracy of MPMS-SGH.

Table 7 Date explanation of test data set for four seasons in Fig. 9 and prediction accuracy

Test date	Weather condition			Whether the environmental control equipment in greenhouse was used				R^2		
	Weather	Outdoor temperature range/℃	Maximum outdoor light intensity/lx	Fan	Wet curtain	Sunshade	Heating	Temperature	Humidity	Light intensity
April 10th	Sunny	10 ~ 24	65 367	no	no	yes	no	0.89	0.81	0.87
July 18th	Sunny	25 ~ 35	71 329	yes	yes	yes	no	0.86	0.71	0.85
October 6th	Cloudy	11 ~ 22	54 374	no	no	no	no	0.9	0.83	0.89
January 2th	Cloudy	−7 ~ 2	31 384	no	no	no	yes	0.88	0.81	0.88

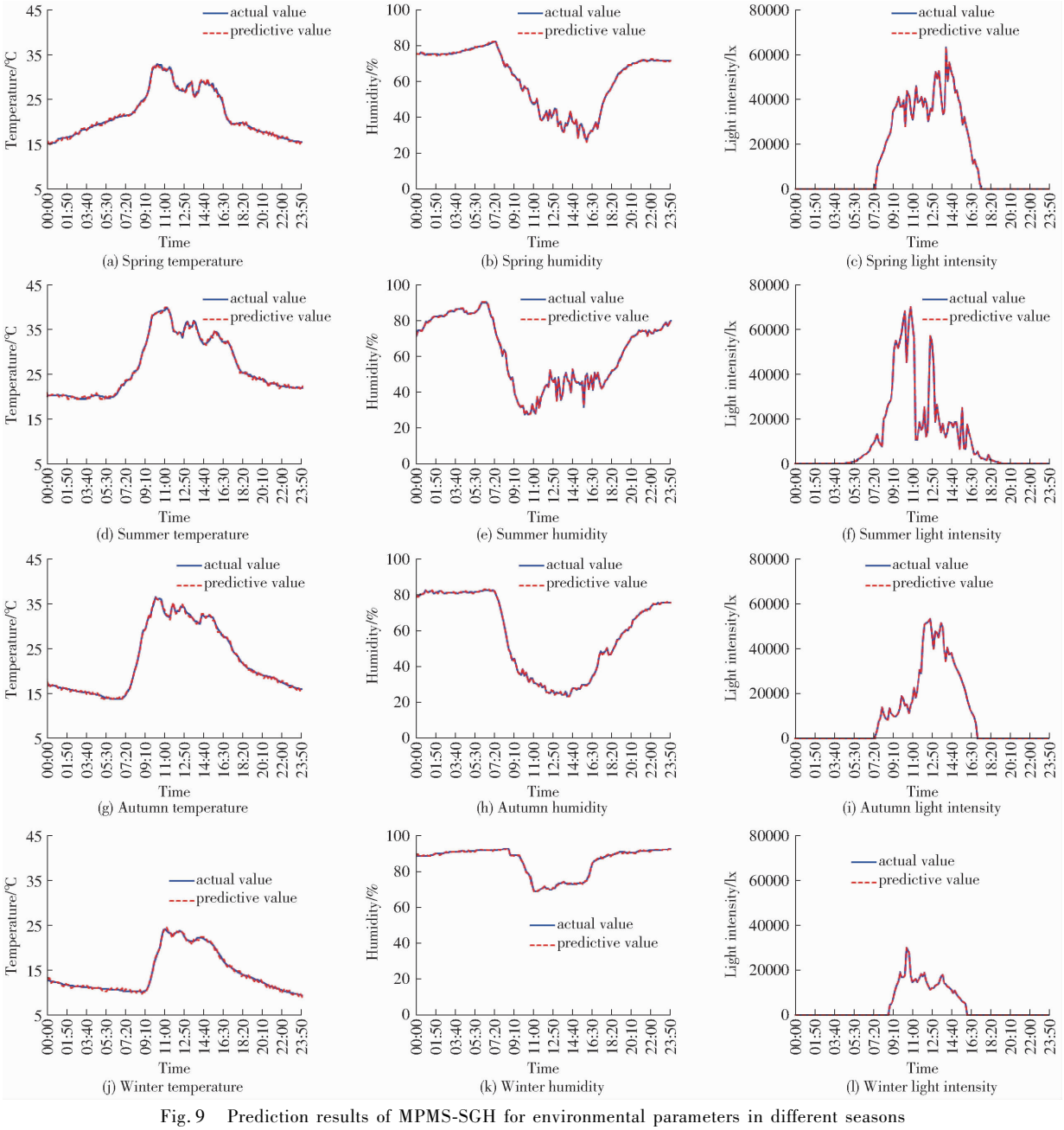

Fig. 9 Prediction results of MPMS-SGH for environmental parameters in different seasons

The environmental control measures in the daylight temperature had a certain impact on the performance of MPMS-SGH. Incorporating operational data of control devices into MPMS-SGH inputs in the future may enhance predictive capabilities.

3 Conclusion

Effective environmental control in solar greenhouses required precise prediction of environmental parameters. These parameters were

nonlinear and non-stationary time series with distinct seasonal patterns. In order to improve the prediction accuracy and seasonal adaptability of the prediction model, MPMS-SGH was introduced to the multi-parameters prediction of solar greenhouses. The main findings were as follows.

(1) The performance and complexity of MPMS-SGH were notably influenced by the length of input and output sequences. Notably, the current data in the environmental parameter sequence was significantly influenced by its first 100 data points. As the length of the output sequence was increased to 45, MPMS-SGH's prediction accuracy was significantly decreased. Therefore, when the input sequence length was less than 100 and the output sequence length was less than 45, the prediction accuracy of MPMS-SGH was higher.

(2) The position encoding module of MPMS-SGH had a significant impact on humidity prediction, while the frequency-domain transformation module helped to predict seasonal component with strong periodicity. Sequence decomposition helped to improve the prediction accuracy of the model for solar greenhouse environmental parameters.

(3) The accuracy of MPMS-SGH in predicting solar greenhouse environmental parameters fluctuated with seasons, which was consistent with the complexity of these parameters changing with seasons. MPMS-SGH showed the highest accuracy in predicting temperature and the lowest accuracy in predicting humidity. MPMS-SGH had the highest accuracy in predicting the temperature inside the greenhouse on sunny days in spring (R^2 = 0.91), the highest accuracy in predicting the humidity inside the greenhouse on sunny days in winter (R^2 = 0.83), and the highest accuracy in predicting the light intensity inside the greenhouse on cloudy days in autumn (R^2 = 0.89). MPMS-SGH had the lowest accuracy in predicting three environmental parameters in a sunny summer greenhouse. These results demonstrated the seasonal adaptability of MPMS-SGH.

References

[1] DU Taihang, LIU De, SUN Shuguang, et al. Optimal temperature control of solar greenhouse based on MPC method[J]. China Measurement & Test, 2022, 48(1): 85 – 91, 167.

[2] JIA W, WEI Z. Short term prediction model of environmental parameters in typical solar greenhouse based on deep learning neural network[J]. Applied Sciences, 2022, 12(24): 12529.

[3] CHEN Xin, TANG Xianglu, LI Xiang, et al. Two-steps prediction method of temperature in solar greenhouse based on twice cluster analysis and neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(Supp.): 353 – 358. (in Chinese)

[4] MOHMED G, HEYNES X, NASER A, et al. Modelling daily plant growth response to environmental conditions in Chinese solar greenhouse using Bayesian neural network[J]. Scientific Reports, 2023, 13(1): 4379.

[5] GARCIA-VAZQUEZ F, PONCE-GONZALEZ J R, GUERRERO-OSUNA H A, et al. Prediction of internal temperature in greenhouses using the supervised learning techniques: linear and support vector regressions[J]. Applied Sciences, 2023, 13(14): 8531.

[6] LIU Hong, DANG Xiaodong, DU Quansheng, et al. Temperature prediction model in solar greenhouse based on stochastic forest algorithm[J]. Chinese Agricultural Science Bulletin, 2020, 36(25): 95 – 100.

[7] WEN Yongjing, LI Cun, XUE Qingyu, et al. Temperature and humidity prediction models in solar greenhouse: comparative analysis based on stepwise regression and BP neural network[J]. Chinese Agricultural Science Bulletin, 2018, 34(16): 115 – 125. (in Chinese)

[8] YANG J, REN G, WANG Y, et al. Environmental prediction model of solar greenhouse based on improved harris hawks optimization-catboost[J]. Sustainability, 2024, 16(5): 2021.

[9] QIAO Xiaodan, ZHENG Wengang, ZHANG Xin, et al. Study on solar greenhouse environment prediction method based on LSTM – GRU[J]. Jiangsu Agricultural Sciences, 2022, 50(16): 211 – 218. (in Chinese)

[10] LI Junwei, ZHU Tao, PAN Peng, et al. Research of multistep prediction model for multivariate whether based on MTGNN[J]. Water Resources and Hydropower Technology (Chinese and English), 2023, 54(Supp. 2): 133 – 140.

[11] HU Jin, LEI Wenye, LU Youqi, et at. Solar greenhouse temperature prediction model based on 1D CNN – GRU[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(8): 339 – 346. (in Chinese)

[12] GAO M, WU Q, LI J, et al. Temperature prediction of solar greenhouse based on NARX regression neural network[J]. Scientific Reports, 2023, 13(1): 1563.

[13] HAO J, LIU F. Improving long-term multivariate time series forecasting with a seasonal-trend decomposition-based 2-dimensional temporal convolution dense network[J]. Scientific Reports, 2024, 14(1): 1689.

[14] ZU Linlu, LIU Pingzeng, ZHAO Yanping, et at. Solar greenhouse environment prediction model based on SSA – LSTM[J].

Transactions of the Chinese Society for Agricultural Machinery,2023,54(2): 351 - 358. (in Chinese)
[15] JUNG D H, KIM H S, JHIN C, et al. Time-serial analysis of deep neural network models for prediction of climatic conditions inside a greenhouse[J]. Computers and Electronics in Agriculture, 2020, 173: 105402.
[16] ZHANG Yongfang, WANG Fang. Study on temperature and humidity prediction model of solar greenhouse based on SSA - RBF network[J]. Journal of Agricultural University of Hebei,2021,44(3):115 - 121. (in Chinese)
[17] XIN P, LI B, ZHANG H, et al. Optimization and control of the light environment for greenhouse crop production[J]. Scientific Reports, 2019, 9(1): 8650.
[18] LI Zhige, YANG Derek, ZHAO Li, et al. Individualized indicator for all: stock-wise technical indicator optimization with stock embedding[C]//KDD'19. New York, NY, USA,2019: 894 - 902.
[19] YANG Xuan, DAI Kan, ZHU Hu, et al. Progress and challenges of deep learning techniques in intelligent grid weather forecasting[J]. Acta Meteorologica Sinica, 2022, 80(5):649 - 667.
[20] YANG Chenglei. Research on monitoring system of key environmental factors and forecasting method of temperature and humidity in solar greenhouse[D]. Zibo: Shandong University of Technology, 2023. (in Chinese)
[21] ZONG Chengji, WANG Jianyu, SONG Weitang, et al. Construction and validation of hourly air temperature prediction model in solar greenhouse at night[J]. Transactions of the CSAE, 2022, 38(Supp.): 218 - 225. (in Chinese)
[22] CLEVELAND, STL R B. A seasonal-trend decomposition procedure based on loess[J]. Journal of Official Statistics, 1990, 6(1): 3 - 33.
[23] DEVLIN J, CHANG Mingwei, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT),2019: 4171 - 4186.
[24] DOSOVOTSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[C]//Proceedings of the 9th International Conference on Learning Representations (ICLR), 2021:1 - 21.
[25] ZHANG X, JIN X, GOPALSWAMY K, et al. First de-trend then attend: rethinking attention for time-series forecasting[J]. arXiv Preprint, arXiv: 2212.08151,2022.

MPMS-SGH:面向日光温室的多参数多步预测模型

冀荣华 王文轩 安 冬 齐劭天 刘金存
(中国农业大学信息与电气工程学院, 北京 100083)

摘要: 准确预测日光温室的环境参数对实现精准环境调控至关重要。在日光温室中,温度、相对湿度和光照强度是关键环境参数。利用监测平台采集了日光温室内 1 年环境数据,包括温度、相对湿度和光照强度,以及同时段的室外温度、室外湿度和室外光照强度等气象数据。通过深入分析,研究了这些参数的特征及相互关系。分析表明,日光温室环境参数具有时序性、非线性和周期性特征,且存在复杂耦合关系,且在不同季节存在着显著性差异。本研究提出日光温室多参数多步预测模型(MPMS-SGH),用于精确预测 3 种关键温室环境参数,且具有一定季节适应性。MPMS-SGH 包括输入层、预处理层、特征提取层和预测层。输入层用于生成包含室内温度、室内湿度、室内光照强度以及室外温度和室外光照强度的原始序列矩阵;预处理层对原始序列矩阵进行归一化、分解和位置编码;特征提取层分别使用时序注意力机制和频域注意力机制从趋势分量和季节分量中提取特征;预测层利用多层感知机对室内环境参数(温度、相对湿度和光照强度)进行多步预测。参数选择实验评估了在不同输入输出序列长度下 MPMS-SGH 的预测性能。结果表明:当输出序列长度固定时,MPMS-SGH 预测精度随输入序列长度增加呈先升后降趋势,输入序列长度为 100 时预测精度最高,此时温度、相对湿度和光照强度 RMSE 分别为 0.22℃、0.28% 和 250 lx;当输入序列长度固定时,随着输出序列长度增加,模型对 3 种环境参数的预测精度持续下降,输出序列长度超过 45 时 MPMS-SGH 预测精度显著降低。综合考虑模型的规模与性能,将 MPMS-SGH 的输入序列长度设为 100,输出序列长度设为 35。通过与 SVR、STL-SVR、LSTM 和 STL-LSTM 4 种预测模型对比实验表明,MPMS-SGH 性能最优,其温度、相对湿度和光照强度 RMSE 分别达到 0.15℃、0.38% 和 260 lx,且序列分解有助于提升预测性能。进一步测试 MPMS-SGH 在不同季节的预测精度发现:该模型对室内温度预测精度最高,对相对湿度预测精度最低,且预测精度随季节波动。其中,春季晴天温室温度预测精度最高(R^2 = 0.91),冬季晴天相对湿度预测精度最高(R^2 = 0.83),秋季阴天光照强度预测精度最高(R^2 = 0.89);夏季晴天 3 种环境参数预测精度均为最低。

关键词: 日光温室; 环境参数; 时间序列; 多步预测

基金项目: 北京市现代农业产业技术体系项目(BAIC01 - 2024 - 19)