ISSN 2616-5775 Vol. 9, Issue 5: 1-8, DOI: 10.25236/AJCIS.2026.090501

A Multi-Scale ConvLSTM with Seasonal Encoding for Monthly Land Surface Temperature Prediction

Yajun Xiao*

School of Information Science and Technology, Yunnan Normal University, Kunming, China
**Corresponding author: 2324100053@ynnu.edu.cn*

Abstract: *Land Surface Temperature (LST), as a key indicator for characterizing surface energy exchange and urban thermal environments, plays a crucial role in climate change research and sustainable urban development. To address the challenges of strong spatial heterogeneity, pronounced temporal periodicity, and difficulty in modeling long-term dependencies in monthly LST sequences, this study proposes an MSG-ConvLSTM model that integrates multi-scale spatial feature extraction with a periodic awareness mechanism. The model enhances spatial representation through parallel multi-scale convolutions and incorporates a sinusoidal positional encoding–based periodic gating mechanism to explicitly capture seasonal variations, thereby improving its ability to model long-term dependencies. Experimental results based on the Pearl River Delta dataset from 2003 to 2023 demonstrate that the proposed model outperforms several mainstream methods in multi-step forecasting tasks. Compared with ConvLSTM, the proposed model reduces MSE by approximately 10%–15%; compared with CNN-BiLSTM, by 12%–18%; compared with Swin-Transformer, by 8%–12%; and compared with PredRNN, by 9%–14%. Similar improvements are also observed in terms of MAE. Notably, in medium- and long-term forecasting, the model exhibits slower error accumulation and stronger stability. Ablation studies further verify the effectiveness of the multi-scale feature extraction and periodic modeling mechanisms in improving prediction accuracy and mitigating error propagation. This study provides an effective and robust approach for monthly LST forecasting using only historical LST data.*

Keywords: *Land Surface Temperature, Spatiotemporal Time Series Forecasting, Pearl River Delta, ConvLSTM*

1. Introduction

Land Surface Temperature (LST) is a key indicator of the Earth's surface energy balance and plays an important role in climate change studies, agricultural disaster warning, and urban heat island monitoring[1]. With rapid advances in remote sensing, long-term, high-coverage monthly LST datasets have become increasingly available, providing a solid basis for understanding thermal environment evolution[2]. However, monthly LST series exhibit significant spatial heterogeneity and temporal periodicity, influenced by land cover, topography, and clear annual cycles with long-term trends. This poses dual challenges for forecasting: capturing local spatial patterns and modeling long-term interannual dependencies, which limits progress in time series prediction.

Early studies relied on statistical and traditional machine learning methods. ARIMA models[3] require manually constructed lag variables and differencing operations, increasing modeling complexity. To simplify modeling, Atik et al.[4] used a multilayer perceptron to predict LST in Bogor, Indonesia, though based on only five years of data. Later, XGBoost, Random Forest (RF), and SVM were applied for seasonal LST prediction in the Yangtze River Delta with reasonable accuracy[5]. With time series forecasting advances, neural networks and LSTM models have been widely adopted[6][7][8], but they struggle to capture complex nonlinear spatiotemporal dependencies.

Recent deep learning methods improve prediction performance. Li et al.[9] used a Pix2pixHD conditional GAN framework integrating urban green space and morphology to improve LST prediction. Xu et al.[10] combined Variational Mode Decomposition (VMD) with a Memory-In-Memory (MIM) network for multimodal sea surface temperature forecasting, showing hybrid methods' effectiveness. However, these approaches mainly use temporal features of the series itself. Multivariate time series can capture inter-variable interactions, improving accuracy. Xin et al.[11] integrated an encoder-only architecture with adaptive multi-head attention into a Transformer, enhancing robustness and prediction for multivariate series. ConvLSTM[12], modeling temporal and spatial dependencies simultaneously, has

ISSN 2616-5775 Vol. 9, Issue 5: 1-8, DOI: 10.25236/AJCIS.2026.090501

been applied in precipitation nowcasting, sea surface temperature prediction, and cloud evolution. Yet most models rely on gating or attention mechanisms to capture temporal dependencies, learning periodicity and trends implicitly. Many studies also use low-temporal-resolution annual or interannual data, mixing intra-annual seasonal variations with long-term trends, limiting accurate LST dynamic modeling.

In summary, existing LST spatiotemporal prediction studies face several limitations: (1) Single-scale convolution kernels cannot simultaneously capture local texture and large-scale patterns, limiting spatial feature representation; (2) Long-term trends and periodic information are mostly learned implicitly, lacking explicit sequence modeling; (3) Some models use limited yearly samples, providing insufficient temporal data to capture long-term trends.

The main contributions of this paper are:

(1) We propose an MSG-ConvLSTM framework integrating multi-scale spatial feature extraction with a state space model. A multi-scale convolution module addresses single-scale limitations in ConvLSTM, and a selective state update mechanism inspired by Mamba effectively captures long-term trends while maintaining linear complexity.

(2) A sinusoidal position encoding–based periodic gating mechanism (SG-ConvLSTM) is introduced. By incorporating month-based sinusoidal encoding into the ConvLSTM gating, a periodic gating factor modulates memory cell updates, explicitly injecting annual periodic priors.

(3) The model is validated on a real-world monthly LST dataset. Experiments on the Pearl River Delta (2003–2023) show it significantly outperforms six baseline models in multi-step forecasting.

2. Related Work

2.1. Study Area and Data

The Pearl River Delta (PRD) in south-central Guangdong, China, adjacent to the South China Sea(Figure 1.), exhibits temporal trends and periodic patterns in land surface temperature (LST), with spatial gradients decreasing from urban cores to peripheries[13]. This makes PRD suitable for evaluating LST prediction models in complex urban environments. The LST data come from the Aqua MYD11A1.061 dataset, converted from Kelvin to Celsius[14], with missing values filled via cubic spline–KNN interpolation, then cropped and resampled to 1 km resolution[15].

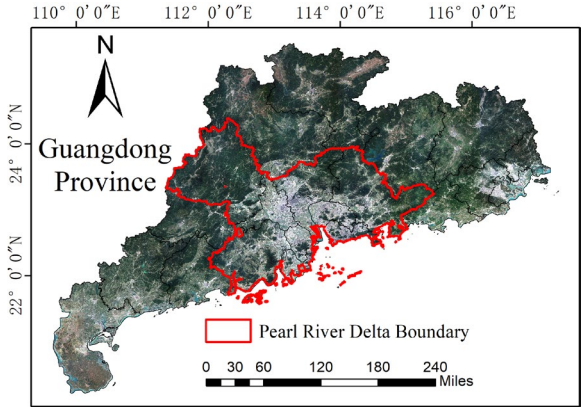

Figure 1: Geographic Location of the Pearl River Delta

2.2. Methods

Based on the ConvLSTM baseline model, this study proposes an improved time series forecasting framework (MSG-ConvLSTM) that integrates multi-scale spatial feature extraction, sinusoidal positional encoding, and a state space model. The proposed model enhances the accuracy of land surface temperature prediction through the collaboration of multiple modules. As illustrated in Figure 2, the overall architecture mainly consists of a multi-scale spatial feature extraction module, a periodic-aware ConvLSTM temporal modeling module, and a Mamba-based long-term dependency enhancement module. The structure of the model is shown as follows.

ISSN 2616-5775 Vol. 9, Issue 5: 1-8, DOI: 10.25236/AJCIS.2026.090501

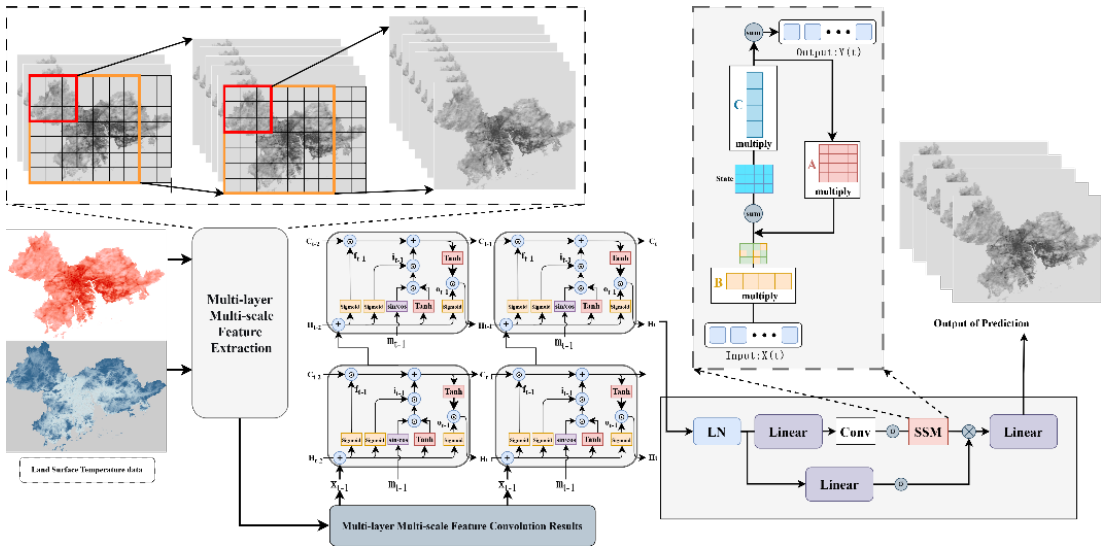

Figure 2: Overview of the MSG-ConvLSTM Network Architecture.

2.2.1. Multi-Level Multi-Scale Feature Extraction Module

In this study, a multi-scale convolutional feature extraction module is introduced before the ConvLSTM to enhance the model's ability to perceive spatial structures at different scales through parallel convolutions with multiple kernel sizes. This module provides richer and more stable spatial feature representations for subsequent temporal modeling. Assume that the input LST sequence is represented as: $X = \{X_t \mid t = 1,2,...,T\}$, $X_t \in R^{H \times W \times T \times C}$ .The T denotes the temporal length, H and W represent the spatial dimensions of the image, and C is the number of input channels.

The multi-scale module uses 3×3 and 5×5 convolutions in parallel at each time step, then concatenates the resulting feature maps along the channel dimension to form fused multi-scale features.

$$F_t = Concat(\sigma(W^{(3)} * X_t + b^{(3)}), \sigma(W^{(5)} * X_t + b^{(5)})) \tag{1}$$

where "*" denotes the convolution operation, $W^{(k)}$ and $b^{(k)}$ represent the weights and biases of convolution kernels, respectively, and $\sigma(\cdot)$ is the *Sigmoid* activation function.

2.2.2. SG-ConvLSTM Model Structure

The gating mechanism in the standard ConvLSTM applies a uniform update strategy across all time steps and does not explicitly consider the inherent annual periodicity in LST sequences. In monthly LST data, different months correspond to distinct stages of thermal environment evolution. Ignoring such periodic positional information may lead to noise interference or over-smoothing during memory updates.

To address this issue, a Seasonal Gate is introduced into the ConvLSTM structure (Figure 3), which modulates the memory update process using periodic positional information of months. Specifically, for a given time step t , the corresponding month index $m_t = \{0,\ 1,\ \ldots\ ,\ 11\}$ is used to construct sinusoidal periodic encoding[16]:

$$P_t = \left[\sin(2\pi m_t/12),\ \cos(2\pi m_t/12)\right]^T \in \mathbb{R}^2 \tag{2}$$

where P_t represents a temporal phase vector, and $2\pi m_t/12$ is the mapping projects the time step t each month onto a circular space, the two-dimensional sinusoidal encoding based on $\sin$ and $\cos$ functions uniquely determines the phase within the cycle.

Then, during the computation, it is broadcasted to $\mathbb{R}^{B \times C \times H \times W}$, thereby applying a consistent temporal modulation weight to each spatial location. Through a linear mapping followed by a sigmoid activation function, the seasonal gating factor is generated:

ISSN 2616-5775 Vol. 9, Issue 5: 1-8, DOI: 10.25236/AJCIS.2026.090501

$$s_t = \sigma(W_s * P_t + b_g) \qquad s_t \in \mathbb{R}^{B\times C\times H\times W} \tag{3}$$

where $s_t \in [0,1]$ is the seasonal gate vector, which modulates the influence of the current temporal phase on memory updates.

With the seasonal gate, the memory cell update of ConvLSTM is modified as:

$$C_t = f_t \odot C_{t-1} + i_t \odot (s_t \odot \tilde{C}_t), \qquad H_t = o_t \odot tanh(C_t) \tag{4}$$

where $\odot$ denotes the Hadamard (element-wise) product.

This structure integrates spatial features with temporal positional encoding, focusing on modeling the dynamic transitions between adjacent months. Through convolutional gating, the model captures local thermal evolution patterns while maintaining spatial consistency, providing high-quality spatiotemporal representations for subsequent long-term modeling.

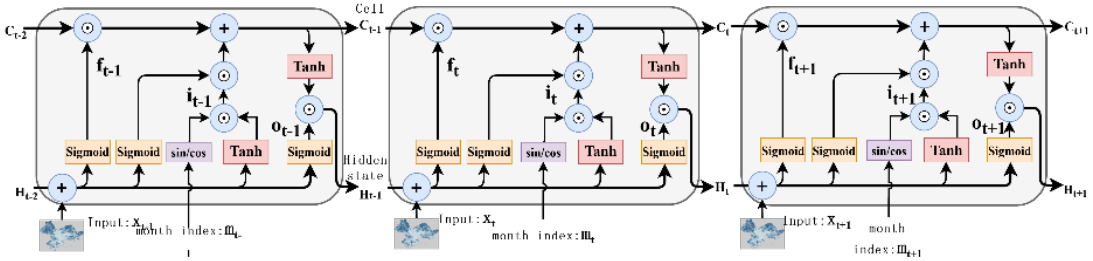

Figure 3: SG-ConvLSTM Network Structure.

2.2.3. Mamba-Based State Space Temporal Modeling

To further enhance the model's ability to capture long-term temporal dependencies, a Mamba model based on selective state space mechanisms is introduced after the ConvLSTM module for deeper temporal modeling. State Space Models (SSMs)[17] explicitly define the evolution of system states and can efficiently model long-sequence dependencies through linear recurrence.

In the model pipeline, the hidden states $\{H_t\}_{t=1}^{T}$ output by ConvLSTM at each time step are first aggregated across spatial dimensions and transformed into a pure temporal sequence:

$$Z = \{z_t \in \mathbb{R}^D \mid t = 1,2,...,T\} \tag{5}$$

where z_t represents the high-dimensional temporal feature vector at time step t. This sequence is then used as input to the Mamba module to model long-term temporal evolution.

The state update process of Mamba can be abstractly formulated as:

$$s_t = As_{t-1} + B(u_t)z_t, \qquad y_t = Cs_t \tag{6}$$

where s_t is the hidden state, A, and C are learnable state transition and output matrices, and $B(\cdot)$ is an input-dependent selective mapping function that dynamically adjusts the contribution of input features at different time steps. This mechanism enables the model to emphasize important information while suppressing redundant noise, effectively mitigating error accumulation in long-term forecasting.

3. Experiments

3.1. Dataset Construction and Experimental Settings

The study uses monthly land surface temperature (LST) time-series images from 2003 to 2023 as the primary data source, covering a continuous 21-year period. A sliding window strategy was applied to construct the dataset. In this study, the window length was set to 42, with the first 36 time steps as model inputs and the remaining 6 time steps as prediction targets. The first 80% of the samples in chronological order were used as the training set, and the remaining 20% as the test set, avoiding information leakage from random splitting. Additionally, the images were cropped into local patches of 32×32 as model inputs,

ISSN 2616-5775 Vol. 9, Issue 5: 1-8, DOI: 10.25236/AJCIS.2026.090501

resulting in a total of 54,144 samples.

Experiments were implemented using the PyTorch framework and conducted on an NVIDIA GeForce RTX 3060 GPU. The training configuration was as follows: the Adam optimizer was used with an initial learning rate of 0.0001, batch size of 64, and 20 epochs. Multiple runs were conducted, and the results were averaged.

3.2. Experimental Results Analysis

Under the above dataset and experimental settings, the proposed model is trained and evaluated. Multiple evaluation metrics are adopted to quantitatively assess prediction performance. Since LST prediction is a continuous regression problem, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2) are selected as the primary evaluation metrics.

Table 1 Experimental Results

Step	MSE	MAE	RMSE	R2
Step1	2.2013	1.1021	1.4836	0.8760
Step2	2.2586	1.1145	1.5028	0.8728
Step3	2.1779	1.0970	1.4757	0.8773
Step4	1.7967	0.9361	1.3404	0.8987
Step5	2.0466	0.9914	1.4305	0.8845
Step6	2.2422	1.1107	1.4973	0.8737

The experimental results demonstrate that the proposed model can effectively capture both the overall trend and seasonal fluctuations of land surface temperature on the test set in Table 1. The short-term predictions (Step1–Step2) exhibit relatively low errors and show high consistency with the ground truth. Due to the use of a sliding window with stride = 6 and the sequence starting from January 2003, different prediction steps statistically correspond to fixed month combinations (e.g., Step1 corresponds to January and July). Consequently, the errors at each step not only reflect model performance but are also influenced by seasonal characteristics. Land surface temperature exhibits strong annual periodicity, and variations differ across months in terms of amplitude and stability. Transitional seasons such as spring and autumn are relatively stable with weaker nonlinearity, resulting in lower prediction errors, whereas winter and peak summer show larger fluctuations and higher uncertainty, leading to relatively higher errors. Therefore, the MSE across prediction steps does not increase monotonically but instead exhibits certain fluctuations. This phenomenon essentially reflects the seasonal heterogeneity and sample distribution characteristics of the land surface temperature time series, indicating that prediction performance depends not only on the forecast horizon but also on the climatic stability of the corresponding months. Meanwhile, the overall similarity of errors across different steps suggests that the model maintains relatively stable predictive capability across various seasonal stages. Figure 4 presents the comparison between predicted results and ground truth, along with the corresponding error maps.

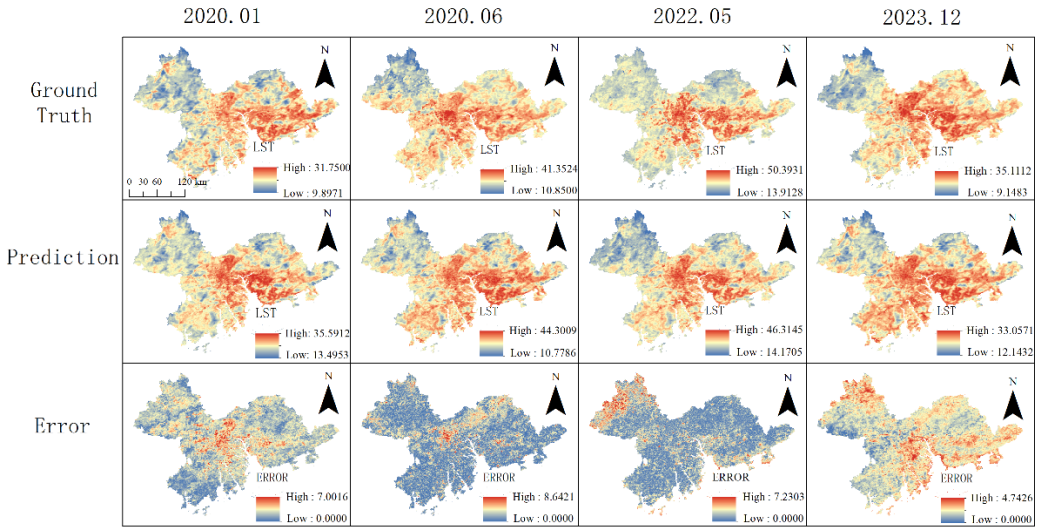

Figure 4: Prediction results and error maps.

ISSN 2616-5775 Vol. 9, Issue 5: 1-8, DOI: 10.25236/AJCIS.2026.090501

3.3 Comparative Experiments

To comprehensively evaluate the performance of the proposed model, several representative spatiotemporal prediction models were selected for comparison, including ConvLSTM, SwinLSTM, Swin-Transformer, and PredRNN. All models were trained and evaluated under the same dataset, input-output settings, and training strategies to ensure fairness. The results are presented in Table 2.

Table 2: Comparison results

Model		Step1	Step2	Step3	Step4	Step5	Step6
ConvLSTM	MSE:	3.8282	3.8831	3.1260	3.0463	3.8889	3.4672
	MAE:	1.3765	1.3884	1.2259	1.2088	1.2813	1.2976
	RMSE:	1.9565	1.9705	1.7680	1.7453	1.8408	1.8620
	R^2	0.7841	0.7810	0.8238	0.8283	0.8089	0.8045
Swin-LSTM	MSE:	2.8350	2.4744	2.9340	2.7156	2.3782	2.9198
	MAE:	1.1612	1.0831	1.1827	1.1354	1.0623	1.1796
	RMSE:	1.6837	1.5730	1.7129	1.6479	1.5421	1.7087
	R^2	0.8402	0.8605	0.8347	0.8470	0.8660	0.8355
Swin-Transformer	MSE:	2.5134	2.5675	2.6301	2.1963	2.4244	2.6791
	MAE:	1.1322	1.1437	1.1551	1.1257	1.1741	1.2271
	RMSE:	1.5854	1.6023	1.6214	1.4820	1.5570	1.6368
	R^2	0.8583	0.8522	0.8497	0.8762	0.8633	0.8489
PredRNN	MSE:	2.4665	2.7148	2.6077	2.0307	2.3575	**2.0933**
	MAE:	1.1021	1.1352	1.1120	0.9871	1.0578	**1.0241**
	RMSE:	1.5705	1.6476	1.6148	1.4250	1.5354	**1.4468**
	R^2	0.8510	0.8470	0.8422	0.8856	0.8672	**0.8803**
MSG-ConvLSTM	MSE:	**2.2013**	**2.2586**	**2.1779**	**1.7967**	**2.0466**	2.2422
	MAE:	**1.0814**	**1.1145**	**1.0970**	**0.9361**	**0.9914**	1.1107
	RMSE:	**1.4836**	**1.5028**	**1.4757**	**1.3404**	**1.4305**	1.4973
	R^2	**0.8760**	**0.8728**	**0.8773**	**0.8987**	**0.8845**	0.8737

Table 2 compares multi-step monthly LST prediction performance. Overall, MSG-ConvLSTM achieves the best or second-best results across most steps, showing clear advantages in error control and long-term forecasting. It outperforms ConvLSTM, and several Transformer models in MSE, RMSE, and R^2, with the advantage more pronounced in medium- to long-term steps (Step4–Step6). The multi-scale spatial feature extraction captures both local variations and large-scale structures, enhancing feature representation. Sinusoidal positional encoding introduces explicit temporal periodic priors, while the Mamba state space model efficiently models long-term dependencies with linear complexity. Compared with PredRNN, MSG-ConvLSTM shows lower errors in most steps, particularly Step4–Step6, indicating stable long-term trend modeling. Short-term differences are minor, but traditional models show error

ISSN 2616-5775 Vol. 9, Issue 5: 1-8, DOI: 10.25236/AJCIS.2026.090501

growth due to memory decay, whereas MSG-ConvLSTM maintains stable error propagation.

3.4 Ablation Study

To evaluate the contribution of each component, a series of ablation experiments were conducted by removing or replacing key modules, including: X. Multi-scale convolution module (MS Conv): replaced with single-scale convolution; Y. Seasonal gating mechanism (SG-Gate): removing sinusoidal positional encoding; Z. Mamba module: replaced with a linear layer.

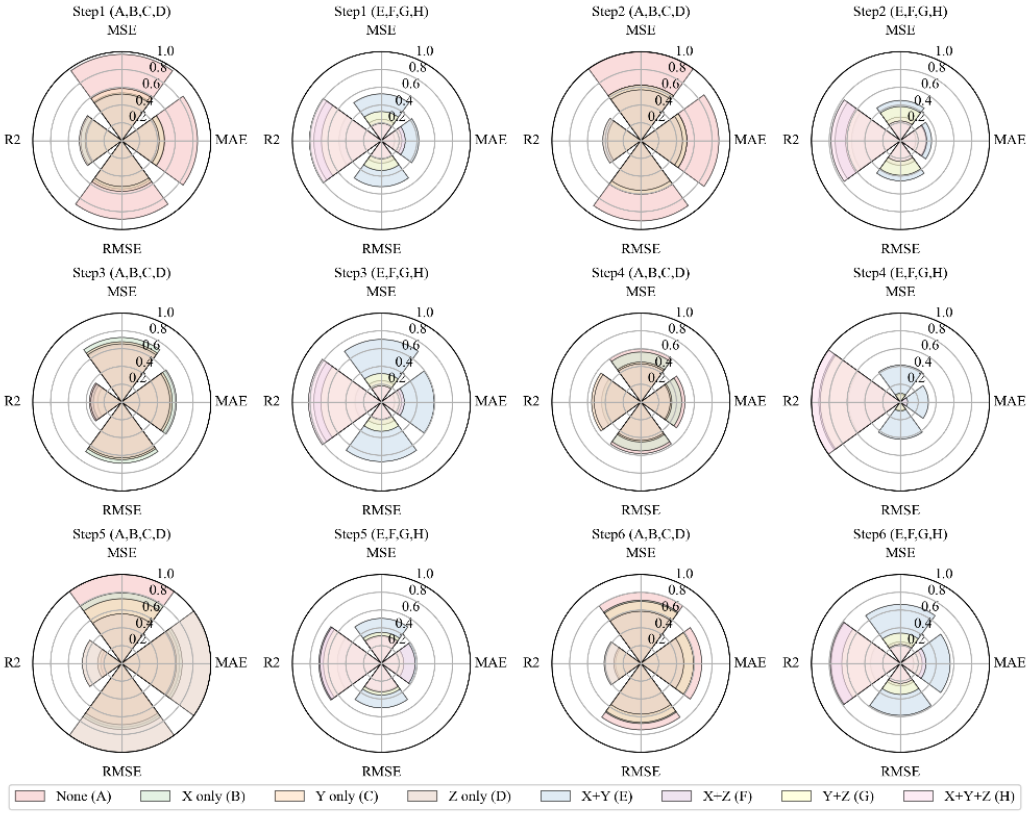

Figure 5: Ablation results

The results (Figure 5) show that removing any key module leads to performance degradation. Specifically, removing the multi-scale convolution module significantly affects spatial prediction accuracy, indicating that single-scale convolution cannot adequately capture spatial patterns at different scales. Removing the seasonal gating mechanism also increases prediction errors, demonstrating the importance of periodic information in monthly prediction tasks. Without the Mamba module, error accumulation becomes more pronounced at longer horizons, indicating its critical role in modeling long-term dependencies and suppressing error propagation.

4. Conclusions

To address the difficulty of modeling periodic trends in monthly land surface temperature prediction, this study proposes the MSG-ConvLSTM model. Based on ConvLSTM, the model enhances spatial heterogeneity perception through multi-scale convolution, introduces sinusoidal positional encoding as an explicit periodic prior, and integrates a state space model to efficiently capture long-term dependencies. Experimental results demonstrate that the proposed model outperforms mainstream methods in multi-step prediction, particularly in medium- and long-term stages with lower errors. Ablation studies further verify the critical role of each module in improving accuracy and suppressing error accumulation. This work provides a robust solution for fine-grained monthly land surface temperature prediction based solely on historical data.

ISSN 2616-5775 Vol. 9, Issue 5: 1-8, DOI: 10.25236/AJCIS.2026.090501

References

[1] Gaur A, Deb C. Machine learning methods and approaches for Urban Heat Island (UHI) assessment: A comprehensive review[J]. Renewable and Sustainable Energy Reviews, 2026, 234: 116903.
[2] Hu L, Sun Y, Collins G, et al. Corrigendum to "Improved estimates of monthly land surface temperature from MODIS using a diurnal temperature cycle (DTC) model"[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 171: 118.
[3] Kesavan R, Muthian M, Sudalaimuthu K, et al. ARIMA modeling for forecasting land surface temperature and determination of urban heat island using remote sensing techniques for Chennai city, India[J]. Arabian Journal of Geosciences, 2021, 14(11): 1016.
[4] Nurwanda A, Honjo T. The prediction of city expansion and land surface temperature in Bogor City, Indonesia[J]. Sustainable Cities and Society, 2020, 52: 101772.
[5] Zhao H, Cui Y, Wang J, et al. A multiscale and seasonal model for urban surface temperature prediction based on landscape, land use and spectral indices[J]. Sustainable Cities and Society, 2025, 131: 106783.
[6] Suthar G, Singh S, Kaul N, et al. Prediction of land surface temperature using spectral indices, air pollutants, and urbanization parameters for Hyderabad city of India using six machine learning approaches[J]. Remote Sensing Applications: Society and Environment, 2024, 35: 101265.
[7] Zhang J, Xiao C, Liang X, et al. Machine learning based on a swarm intelligence algorithm and explainable AI for the prediction of reservoir temperature[J]. Energy, 2025, 341: 139412.
[8] Deo R C, Şahin M. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland[J]. Renewable and Sustainable Energy Reviews, 2017, 72: 828-848.
[9] Li Q, Zheng H. Prediction of summer daytime land surface temperature in urban environments based on machine learning[J]. Sustainable Cities and Society, 2023, 97: 104732.
[10] Xu S, Dai D, Cui X, et al. A deep learning approach to predict sea surface temperature based on multiple modes[J]. Ocean Modelling, 2023, 181: 102158.
[11] Xin N, Su J, Hasan M M. MMformer with adaptive attention: Advancing multivariate time series forecasting for environmental applications[J]. Applied Soft Computing, 2026, 186(Part B): 114090.
[12] Shi X, Chen Z, Wang H, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting[M]. MIT Press, 2015.
[13] Rajesh J, Pande C B. Estimation of land surface temperature for Rahuri Taluka, Ahmednagar District (MS, India) using remote sensing data and algorithm[M]. Climate Change Impacts on Natural Resources, Ecosystems and Agricultural Systems. Cham: Springer, 2023: 565-577.
[14] Pande C B. Land use/land cover and change detection mapping in Rahuri watershed area (MS), India using the Google Earth Engine and machine learning approach[J]. Geocarto International, 2022, 37(26): 13860-13880.
[15] Zhu J, Huang J, Li H, et al. Multi-scenario simulation of land use change and carbon stock assessment in the Pearl River Delta urban agglomeration[J]. Journal of South China Normal University (Natural Science Edition), 2025, 57(03): 62-73.
[16] Kim S H, Lee S-H, Chung C C. Phase shift calibration method in optical sinusoidal encoder signals applied to servo track writer[J]. IFAC-PapersOnLine, 2016, 49(21): 1-6.
[17] Karadag Y M, Talaz I, Dino I G, et al. ms-mamba: Multi-scale mamba for time-series forecasting[J]. Neurocomputing, 2026, 680: 133226.