



Pentad-mean air temperature prediction using spatial autocorrelation and attention-based deep learning model

Lei Xu¹ · Xi Zhang¹ · Wenying Du¹ · Hongchu Yu² · Zeqiang Chen¹ · Nengcheng Chen¹

Received: 17 July 2023 / Accepted: 22 November 2023 / Published online: 29 November 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2023

Abstract

Abnormal changes in air temperature cause natural disasters such as droughts, hailstorms, and storms, thereby affecting the normal lives of human beings. Consequently, timely and accurate air temperature prediction is essential for human production and livelihood. Traditional air temperature prediction methods are less accurate and less consider the spatial relationship between air temperature in different regions. In this paper, we propose a new deep learning model, convolutional long short-term memory based on channel attention and spatial autocorrelation (ConvLSTM-CASA), which focuses on the spatial correlation between ambient air temperatures and can effectively capture the interaction of air temperatures in different regions. The results show that the ConvLSTM-CASA model has an average R^2 of 0.954 and MSE of 5.245 for pentad-mean temperature prediction over the Yangtze River basin. Compared with baseline forecasting models, the MSE accuracy by the ConvLSTM-CASA model improved by 72.45%, 48.95%, 48.97%, 47.79%, and 22.63% over the decision tree regression (DTR), multiple linear regression (MLR), random forests (RF), long short-term memory (LSTM), and ConvLSTM models, respectively. The ConvLSTM-CASA model is expected to outperform the ConvLSTM model over 90% of the area, suggesting robust prediction skill improvement over space. The ConvLSTM-CASA model provides new insights for data-driven pentad-mean air temperature prediction by including elaborate channel and spatial feature modeling, which aid individuals in comprehending the intricate patterns of air temperature fluctuations.

1 Introduction

Air temperature is a key state variable in climate, hydrology, and ecosystems. Changes in air temperature play an important role in various human activities and are closely related to the way humans produce, live, and practice socially (Skripnuk and Samylovskaya 2018). Changes in air temperature can affect human health in several ways. When the temperature is too high, the human adjustment system cannot effectively regulate body temperature. It causes a large amount of heat in the body to be discharged, and heat stroke, sunburn, and even shock can occur. When the temperature is too low, the body temperature drops sharply. This

will cause the human body's immune ability to decrease, leading to colds, respiratory disease, frostbite, and other diseases (Bunker et al. 2016; Chou et al. 2012). Global warming will accelerate the melting of glaciers, causing sea level rise and flooding of coastal areas. The rise in global air temperature will also increase the possibility of storms. Superstorms combined with sea level rise will make flooding more frequent and pose a great threat to coastal areas. A rising global air temperature will also lead to extreme drought events. As the climate warms, soil moisture can be affected to the extent that crop yields decline. Shortages of water and food resources due to drought events can directly endanger people's lives (Nastos and Matzarakis 2012; Ostberg et al. 2013; Singh and Singh 2012; Xu et al. 2021a). Spatial and temporal variations in air temperature are critical for meteorology, climate, and hydrology. They contribute to improved knowledge of water, energy and carbon cycles and extreme climate predictions. Therefore, accurate air temperature prediction helps protect the ecological environment, safeguarding human security and reducing economic losses. It is of great importance in many fields, such as earth sciences,

✉ Lei Xu
xulei10@cug.edu.cn

¹ National Engineering Research Center for Geographic Information System, China University of Geosciences (Wuhan), Wuhan 430074, China

² School of Navigation, Wuhan University of Technology, Wuhan 430063, China

atmospheric sciences and ecological sciences (Ji et al. 2014; Xu et al. 2021b).

Current air temperature prediction methods can be divided into three categories. The first category is empirical methods, in which future air temperature is predicted based on historical temperature data, prior experience, and available information on weather conditions (Galacgac and Balisacan 2009; Rautela and Karki 2015), as air temperature generally shows a cyclical pattern. For example, Ayal et al. (2015) found that Borana herders make weather forecasts based on morphological changes in flora and fauna. Anandaraja et al. (2008) found that Coimbatore district farmers have weather forecasting calendars based on experience, which predict future weather conditions based on the previous year's weather conditions. Okonya and Kroschel (2013) found that people in the Uganda region invented some climate evaluation indicators based on the actions of insects and migratory birds as a method for determining the trend in climate change. However, these methods rely on only personal experience, and it is difficult to predict the impact of sudden weather extremes on air temperature. These methods have a low prediction accuracy and are suitable for fewer scenarios.

The second type is mathematical statistical methods, which rely on various mathematical function models. They can analyze the influence of various meteorological factors on air temperature and explore the intrinsic laws of air temperature change. Tektaş (2010) used an adaptive network-based fuzzy inference system (ANFIS) and autoregressive integrated moving average (ARIMA) mathematical statistical models based on the Istanbul, Turkey, daily average temperature, barometric pressure and wind speed from 2000 to 2008 and predicted the daily average temperature during July December 2008. Eredics (2009) studied short-term temperature prediction for the exterior of small greenhouses based on several effective mathematical methods. It is concluded that the accuracy of the above methods is better for 1–2 h of prediction and worse for a longer time. Bauer et al. (2015) argued that weather prediction can be considered a mathematical-physical problem. Future climate conditions can be efficiently predicted by understanding the physical mechanisms of climate formation and developing mathematical models that consider the effects of dynamics, physics, chemistry, and other fields.

The third category is machine learning and deep learning methods, which provide an effective method for elucidating complex features and predicting variable relationships. The good data fitting and generalization capabilities of artificial intelligence (AI) methods ensure accurate prediction of air temperature while enriching input characteristics (Choe and Yom 2020; Gong et al. 2022; Tabrizi et al. 2021; Xu et al. 2019). With the development of computer technology, deep learning has become the most popular tool at present

for time-series prediction (Lim and Zohren 2021; Xu et al. 2022). For example, Shen et al. (2020) fused remote sensing data, station data, and socioeconomic data to estimate the daily maximum temperature in mainland China by the deep belief network (DBN) model. The experimental results showed that the deep learning approach could better consider the nonlinear relationships among multiple variables. Liu et al. (2022) proposed a spatial temporal long short-term memory based on a self-attention (ST-LSTM-SA) deep learning model, which was proven to show good results in short-term extreme rainfall prediction. Zhang et al. (2020) proposed the spatiotemporal orthogonal regularization residual convolutional neural network (ST-OR-ResNet) model to capture the complex spatiotemporal relationships in meteorological data and used the model to accurately predict air quality changes in Beijing. An increasing number of fields, such as rainfall prediction, soil moisture prediction, and temperature prediction, are using deep learning methods to achieve good results (Li et al. 2022; Seong 2021; Wang et al. 2020; Xu et al. 2020). Among the deep learning methods, convolutional long short-term memory (ConvLSTM) can fit the air temperature variation well in large regions because spatial features are extracted by convolutional operations and temporal features are captured by LSTM (Shi et al. 2015). Lin et al. (2019) constructed a temperature forecast deviation dataset using numerical weather prediction (NWP) and then obtained more accurate forecasts than NWP with a modified ConvLSTM model. Xiao et al. (2019) used the traditional ConvLSTM to predict the sea surface temperature in a region of the East China Sea and confirmed that the model outperformed the support vector regression (SVR) and LSTM models through comparative experiments. In addition, attention mechanisms are widely used in deep learning models. It can help the model assign different weights to each part of the input and extract more critical information, and make more accurate judgments without affecting the computation and storage of the model (Guo et al. 2022; Niu et al. 2021; Woo et al. 2018). Yuan et al. (2021) proposed a spatiotemporal attention-based LSTM model to fit complex nonlinear variations in industrial process data for accurate simulations considering the correlated effects of spatial variables and temporal samples. Rao et al. (2023) added non-local attention and self-attention to DenseNet to focus more on the role of non-local areas in cellular traffic prediction. Tong et al. (2020) achieved significant results in connecting channel attention and DenseNet models to enhance the weights of important channels for remote sensing image classification work. Wu et al. (2022) combined ConvLSTM and the convolutional block attention module (CBAM) to perform convolutional operations on the data for local area temperature prediction and achieved high accuracy and stability.

Although existing deep learning methods have been well developed, there are some inadequate considerations

regarding the specific air temperature prediction problem. First, air temperature is spatially continuous, and the temperature of one region often shows some correlations with that of neighboring regions (Fallah Ghalhari and Dadashi Roudbari 2018). Two disjointed regions may also show some correlations due to similar geographic environments (Di Cecco and Gouhier 2018; Zhu et al. 2018). Second, the commonly used convolution calculation can consider only the spatial correlations of local adjacencies and lack consideration of spatial relationships at a distance, which can lead to certain deviations in air temperature prediction in large regions. Therefore, it is essential to consider global spatial relationships in large regional air temperature prediction (Shi et al. 2021; Song and Song 2022; Yang et al. 2018).

In this study, we propose an air temperature prediction model, ConvLSTM with channel attention and spatial autocorrelation (ConvLSTM-CASA). The model considers the spatial correlation of air temperature distribution in large regions by calculating the spatial autocorrelation index and the channel attention mechanism. The ConvLSTM-CASA model focuses more on the effects of spatial relationships of air temperature distribution and learns the characteristics of similar air temperature trends at long distances. The main contributions of the work in this paper are as follows:

1. The ConvLSTM-CASA method introduces the spatial autocorrelation index, which fully accounts for the correlation of air temperature on the inhomogeneous spatial distribution. It can effectively reduce the random error and uncertainty in air temperature prediction, which can more accurately predict the trend of air temperature changes in different locations and improve the stability of the prediction results.
2. The ConvLSTM-CASA method adds the channel attention mechanism into the deep learning model to further enhance the role of the spatial correlation index. The different channels are weighted by channel attention to highlight the contribution of the spatial autocorrelation index so that the model can fully consider the spatial heterogeneity of air temperature distribution and spatial correlation.
3. A comparative experimental evaluation of the ConvLSTM-CASA model with decision tree regression (DTR) (Safavian and Landgrebe 1991), multiple linear regression (MLR) (Uyanik and Güler 2013), random forests (RF) (Breiman 2001), long short-term memory (LSTM) (Gers et al. 2000), and ConvLSTM models was carried out, which verified the superior predictive performance of the proposed model.

2 Study area and data description

2.1 Study area

In this study, the Yangtze River basin (YRB) in China, which is located between latitudes 24°N and 36°N and between longitudes 90°E and 122°E (Fig. 1), is selected as the study area. The YRB region, which refers to the vast area through which the mainstream and tributaries of the Yangtze River flow, spans three major economic regions in eastern, central, and western China. The Yangtze River flows through mountains, plateaus, basins, hills, and plains over a total of 19 provinces, municipalities, and autonomous regions and is the world's third-longest river. The watershed covers a total area of 1.8 million square kilometers, accounting for 18.8% of China's land area. The YRB region is located at the junction of subtropical and temperate zones, where the air temperature is influenced by topography, seasons, and climate. The air temperature characteristics in the whole region are highly spatially heterogeneous and spatially autocorrelated, with the annual average temperature higher in the east and south and lower in the west and north.

2.2 Data description

The air temperature data selected for this study are the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) data from the American National Aeronautics and Space Administration (<https://disc.gsfc.nasa.gov/>). MERRA-2 is a global climate reanalysis dataset developed by NASA's Goddard Space Flight Center to analyze climate change and trends over the past several decades (Reichle et al. 2017). It covers a wide range of meteorological parameters, such as temperature, humidity, wind speed, precipitation and radiation flux. With a spatial resolution of 0.5°×0.625° and an hourly temporal resolution, this dataset can be used to study climate change on long time scales and weather events on short time scales.

The original MERRA-2 dataset contains 15,500 days of daily air temperature data from January 1, 1980 to June 8, 2022, and is averaged every five days to obtain 3100 images of pentad-mean air temperature in TIFF format, with the size of each image being 48×23. To avoid the influence of scale differences on model training, the data are normalized to [0,1] by maximum-minimum normalization according to the following equation:

$$T' = \frac{T - T_{min}}{T_{max} - T_{min}} \quad (1)$$

where T is the pentad-mean air temperature, T_{max} is the maximum pentad-mean air temperature value, T_{min} is the

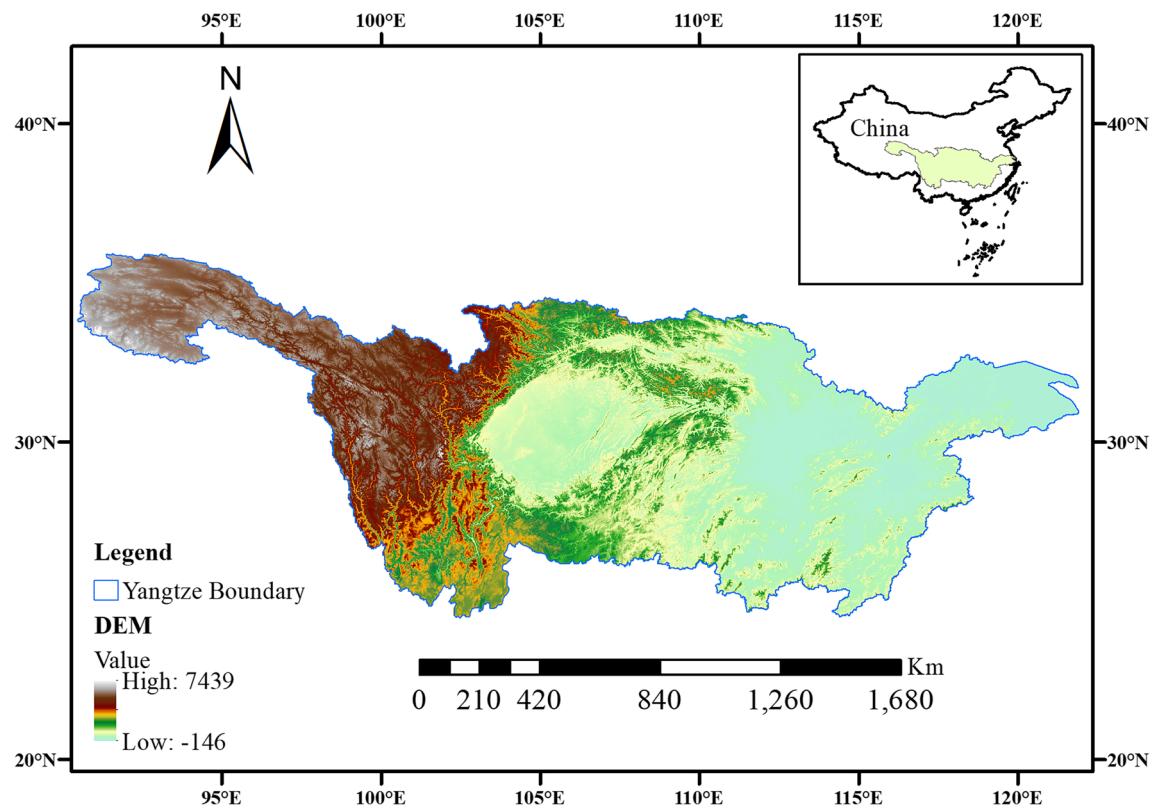


Fig. 1 Geographic location and digital elevation model of the YRB region

minimum pentad-mean air temperature value, and T' is the normalized pentad-mean air temperature.

After the normalization process, the data were divided into a training set and a test set in time order, where the first 2600 images were the training set and the last 500 images were the test set. To verify the prediction ability of the model for long sequences, the image data were then sliced using a sliding window of 10 frames in length, of which the first 5 consecutive frames were used as the input frames and the last 5 were used as the prediction frames.

3 Methodology

3.1 The ConvLSTM model

ConvLSTM is a combined model based on CNN and LSTM that combines the spatial feature extraction capability of the CNN and the temporal modeling capability of the LSTM (Shi et al. 2015). In spatiotemporal forecasting, it can handle both spatial data and capture long-term dependencies and short-term changes in time-series data. The convolution operation in ConvLSTM also has parameter sharing

capability, which can reduce the number of model parameters and improve the generalization ability of the model. This can avoid problems such as overfitting and data sparsity.

In the ConvLSTM model, the cell state is used to add or remove information through the gate structure. The cell state is used to store long-term information. The gate structure controls the cell state and decides which information to keep and which to discard. The main gate structures are the input gate, output gate and forget gate. A gate is a structure that allows selective passage of information, usually consisting of a sigmoid function and a pointwise product operation. The sigmoid function has an input value range of $[0, 1]$, with 0 representing a complete discard and 1 representing a complete pass. The forget gate is the output of the previous cell and the input of this cell through the sigmoid function. The function produces a value within $[0, 1]$ to control the degree to which the previous cell state is forgotten. The input gate determines the value to be retained and the value to be updated. The sigmoid function is applied to determine what to update and then the tanh function creates a new candidate values vector. The output gate requires the application of a sigmoid layer to determine the output part of the cell state, then process the cell state through tanh (to obtain a value between -1 and

1) and multiply it with the sigmoid gate output to finally output the part to be determined. The complete equation is shown in Eqs. (2)–(7). The interpretation of all variables and symbols is shown in Table 1.

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + b_f) \quad (2)$$

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + b_i) \quad (3)$$

$$\tilde{c}_t = \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (4)$$

$$C_t = f_t * C_{t-1} + i_t \odot \tilde{c}_t \quad (5)$$

$$O_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \odot C_t + b_o) \quad (6)$$

$$H_t = O_t \odot \tanh(C_t) \quad (7)$$

3.2 Spatial autocorrelation index

In general, there will be some kind of connection between geographically close phenomena or units, which makes their observations appear correlated. Spatial autocorrelation is a method and indicator used to measure the strength of the association of neighboring things (Getis 2007). Specifically, the spatial autocorrelation method deals with the covariance of variables between neighboring observation units, comparing the similarity of observations and the similarity between their spatial locations. There are two types of spatial autocorrelation: global and local measures. Global spatial autocorrelation describes the overall spatial relationships of all units within the study space. Local spatial autocorrelation describes how spatial relationships are distributed and vary in space (Ord and Getis 1995; Tiefelsdorf and Boots 1997). The full-domain Moran's *I* statistic is a very widely used full-domain spatial autocorrelation statistic, which is calculated as:

Table 1 Definition and interpretation of all variables and symbols

Variables and symbols	Definition and interpretation
f_t	Forget gate
i_t	Input gate
O_t	Output gates
σ	Sigmoid activation function
tanh	Tanh activation function
*	Convolution calculation
\odot	Pointwise product calculation
X_t	Current cell input
H_{t-1}	Previous cell output
H_t	Current cell output
C_{t-1}	Previous cell state
C_t	Current cell state
\tilde{c}_t	Information stored in the current cell state C_t
W_{xf}	Matrix of weight coefficients between current cell input X_t and forget gate
W_{hf}	Matrix of weight coefficients between previous cell output H_{t-1} and forget gate
W_{cf}	Matrix of weight coefficients between previous cell state C_{t-1} and forget gate
W_{xi}	Matrix of weight coefficients between current cell input X_t and input gate
W_{hi}	Matrix of weight coefficients between previous cell output H_{t-1} and input gate
W_{ci}	Matrix of weight coefficients between previous cell state C_{t-1} and input gate
W_{xc}	Matrix of weight coefficients between current cell input X_t and current cell state C_t
W_{hc}	Matrix of weight coefficients between previous cell output H_{t-1} and current cell state C_t
W_{xo}	Matrix of weight coefficients between current cell input X_t and output gate
W_{ho}	Matrix of weight coefficients between previous cell output H_{t-1} and output gate
W_{co}	Matrix of weight coefficients between previous cell state C_{t-1} and output gate
b_f	The bias of the forget gate
b_i	The bias of the input gate
b_c	The bias of the updated cell state
b_o	The bias of the output gate

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \sum_{i=1}^n (x_i - \bar{x})^2} \quad (8)$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, x_i denotes the observed value of region i , n is the number of regions and w is the spatial weight matrix.

The values of Moran's I range from $-1 \leq I \leq 1$. The closer I is to 1, the stronger the degree of spatial positive correlation between regions. The closer I is to -1 , the stronger the degree of spatial negative correlation between regions. The closer I is to 0, the less spatial autocorrelation between regions. In the significance test of Moran's I index, it is often assumed that the variables obey a normal distribution, so the spatial correlation between regions is judged by the z -statistic that obeys a standard normal distribution. In the actual calculation, the spatial weight matrix W is used based on the adjacency relationship. The element w_{ij} in W takes the value of 1 when the i th region is adjacent to the j th region and 0 otherwise. The local spatial autocorrelation statistic is used to identify the different spatial association patterns that may exist with different spatial locations, allowing observation of spatial local unevenness and detection of spatial heterogeneity between data. At spatial location i , the formula for the local Moran's I_i statistic is shown in Eq. (9).

$$I_i = \sum w_{ij} z_i z_j \quad (9)$$

$$I_i = \frac{\sum_i^n \sum_{j \neq i}^n w_{ij} z_i z_j}{S^2 \sum_i^n \sum_{j \neq i}^n w_{ij}} = \frac{1}{n} \sum_i^n \left(z_i \sum_{j \neq i}^n w_{ij} z_j \right) = \frac{1}{n} \sum_i^n I_i \quad (10)$$

where z_i and z_j are normalized observations and w_{ij} is the spatial weight matrix element after row normalization.

I_i is equal to the product between the observations at position i and the weighted average of its neighbors' observations. The relationship between the global Moran's I and the local Moran's I_i is shown in Eq. (10). The global Moran's I takes values from -1 to 1 , while the local Moran's I_i is not limited to -1 to 1 . The local Moran's I_i is interpreted as follows: if I_i is significantly positive and z_i is greater than 0, it indicates that the observations of location i and its neighbors are high compared to the sample average, which is High-High Cluster. If I_i is significantly positive and z_i is less than 0, it indicates that the observations of location i and its neighbors are relatively low, which is Low-Low Cluster. If I_i is significantly negative and z_i is greater than 0, it indicates that the observations of its neighbors are much lower than the observations of location i , which is High-Low Outlier. If I_i is significantly negative and z_i is less than 0, it means that the observations of its neighbors are much higher than the observations of location i , which is Low-High Outlier.

Based on the temperature dataset of this study, the local Moran's I_i of the YRB is calculated, as shown in Fig. 2. The

blue area represents the Low-Low Cluster, and the red area represents the High-High Cluster. The Low-Low Cluster area is the upper Yangtze River area. Geographical location is generally located in the Qinghai-Tibetan Plateau with high-altitude terrain, so the low-temperature agglomeration phenomenon is significant. The High-High Cluster area is the middle and lower Yangtze River area, which is flat and located in the western Pacific subtropical high pressure zone, so the high temperature concentration phenomenon is significant. Overall, the spatial autocorrelation of air temperature in the YRB region is significant, and the spatial relationship is obvious.

3.3 The proposed ConvLSTM-CASA model

The YRB region is a vast area with obvious climatic features and significant spatial relationships. Its air temperature prediction is greatly affected by spatial heterogeneity and spatial correlation. Thus, this study uses pentad-mean air temperature data combined with a distance weight matrix to calculate the corresponding local Moran's I_i , which is added to the temperature data as a feature. To highlight the dependence of the spatial autocorrelation index in air temperature prediction, the channel attention mechanism was introduced in this study. The mechanism is a technique used to enhance the CNN feature representation capability. It dynamically adjusts the importance of each channel by weighting the input feature maps in the channel dimension, thus improving the accuracy and generalization ability of the model. The core idea of the channel attention mechanism is that different channels may have different contributions for different tasks and scenarios. For example, some channels may focus more on the texture features of an image when dealing with an image classification task, while others focus more on image color features. The channel attention mechanism dynamically adjusts the contribution of each channel by learning the weights so that each channel can play the maximum role.

The overall structure of the ConvLSTM-CASA model is shown in Fig. 3. The pentad-mean air temperature data of MERRA-2 are combined with the spatial weight matrix to calculate the corresponding local Moran's I_i , which is taken as input to the ConvLSTM-CASA model as one feature channel together with the pentad-mean air temperature data. The weight of the two channels (temperature and Moran's index) is dynamically adjusted by channel attention so that the whole model can fully consider the spatial heterogeneity and spatial correlation of air temperature distribution. The channel attention module compresses the image and extracts the features by taking the input feature map through global max pooling and global average pooling based on width and height. Then, the weights and parameters are shared by the shared multilayer perceptron (MLP). The output features of the shared MLP are subjected to an elementwise summation

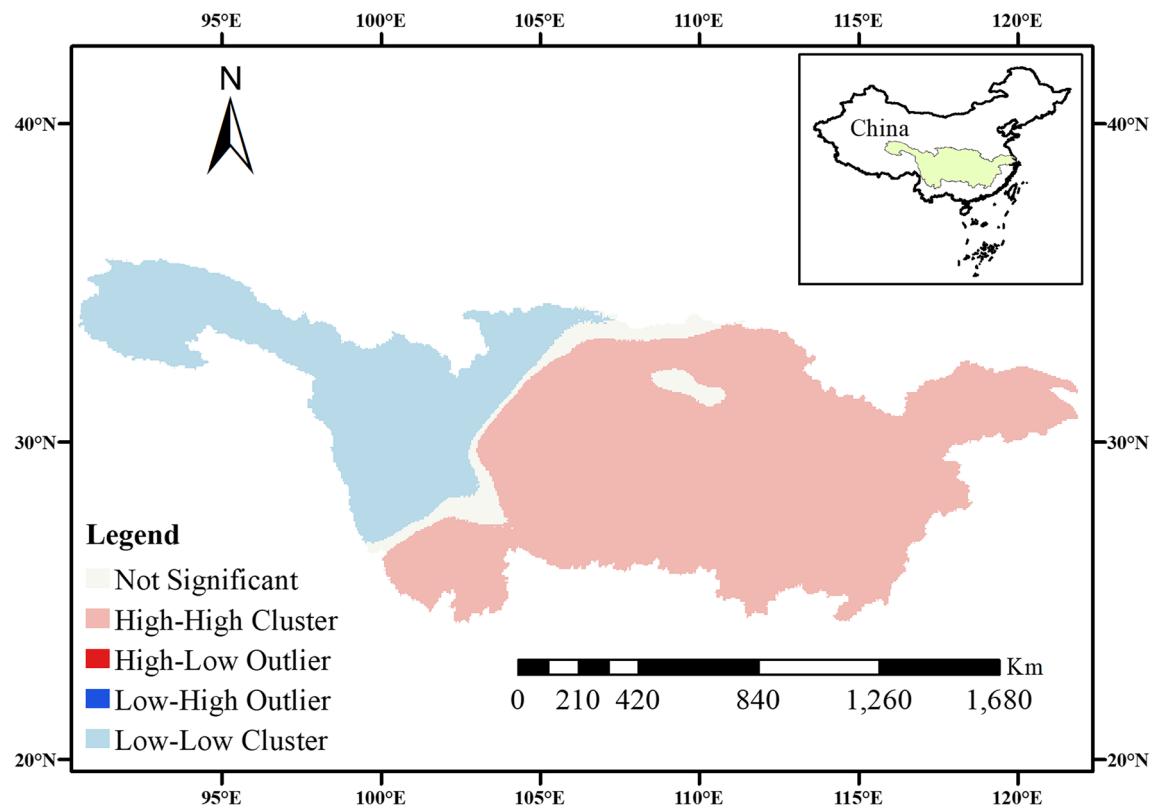


Fig. 2 Spatial clustering was calculated from the average air temperature in the YRB region from January 1, 1980 to June 8, 2022

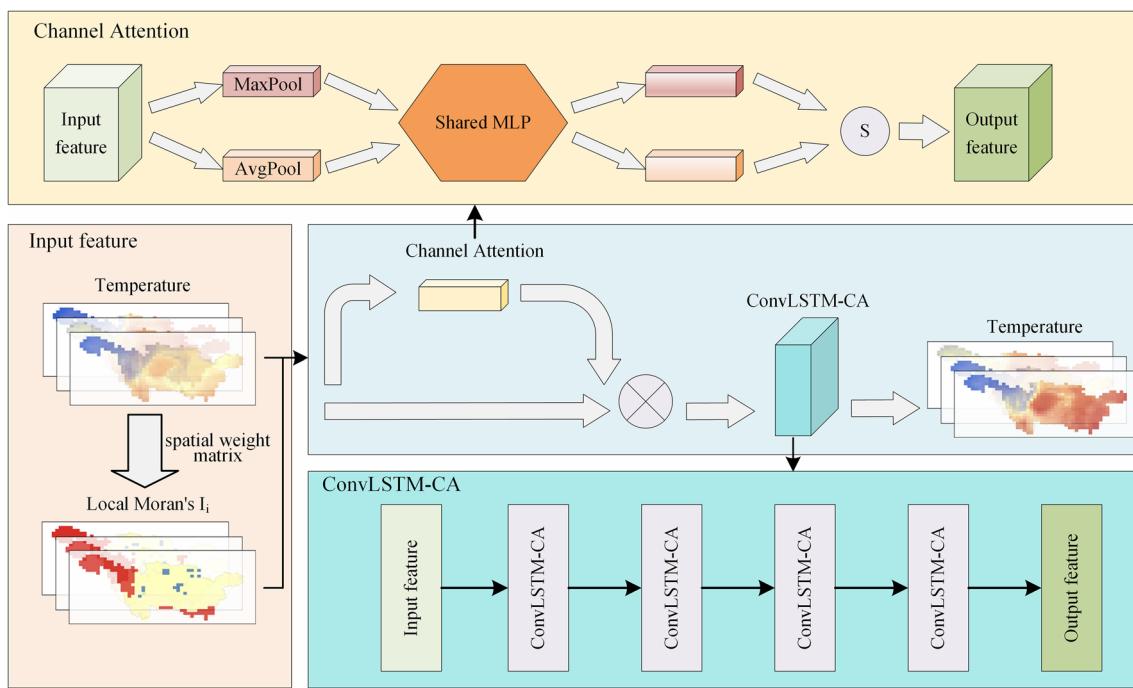


Fig. 3 Overall structure of the ConvLSTM-CASA model

operation and then a sigmoid activation operation to generate the final channel attention feature map. Then, the channel attention output feature is passed through four ConvLSTM-CA layers, and the ConvLSTM-CA cell in this model is shown in Fig. 4. The previous cell state and the output of the previous cell are passed through channel attention to adjust the weights of different channel features. The temporal data are then subjected to long- and short-term memory and spatial convolution operations via ConvLSTM cells to better extract spatiotemporal features and model the temporal data.

4 Experiments and discussion

4.1 Experimental setup

In this study, we compare the proposed ConvLSTM-CASA model with the ConvLSTM, DTR, MLR, RF, and LSTM models. As the structures of the ConvLSTM-CASA and ConvLSTM models are similar, the two models are set to the same hyperparameters during tuning. The loss function is set to L2 loss. The batch size of each iteration is set to 10. The number of iterations is set to 600. The initial learning rate is 0.001. The optimizer is Adam. The convolutional kernel size is 3×3. The model contains four layers of ConvLSTM-CA with 64, 64, 64, and 64 units, and the model is tuned to ensure that it converges by validating its performance on the test set. The other machine learning models are also parameter-tuned to ensure that optimal experimental results are obtained for all of the above models.

The R-square (R^2), mean squared error (MSE), root mean squared error (RMSE), and Pearson correlation coefficient (PCC) are used as model evaluation indicators. The R^2 is a statistical indicator used to reflect the degree of regression fit of the regression model, which can be defined as the ratio of

the fitted value of the independent variable to its true value in the sample regression model. The closer the R^2 value is to 1, the higher the stability and goodness of fit of the model. The MSE is the ratio of the square of the deviation between the predicted and true values to the number of observations n. The RMSE is the square root of the MSE, both of which are used to measure the deviation of the predicted value from the true value. The closer the MSE and RMSE values are to zero, the stronger the predictive power of the model. PCC is used to measure the strength and direction of the linear correlation between the predicted and true values. Its value is between -1 and 1. The closer its value is to 1, the stronger the positive correlation. The closer it is to 0, the weaker the correlation. The closer it is to -1, the stronger the negative correlation. The calculation formulas are as follows:

$$R^2 = 1 - \frac{\sum_i (y_i - \bar{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (11)$$

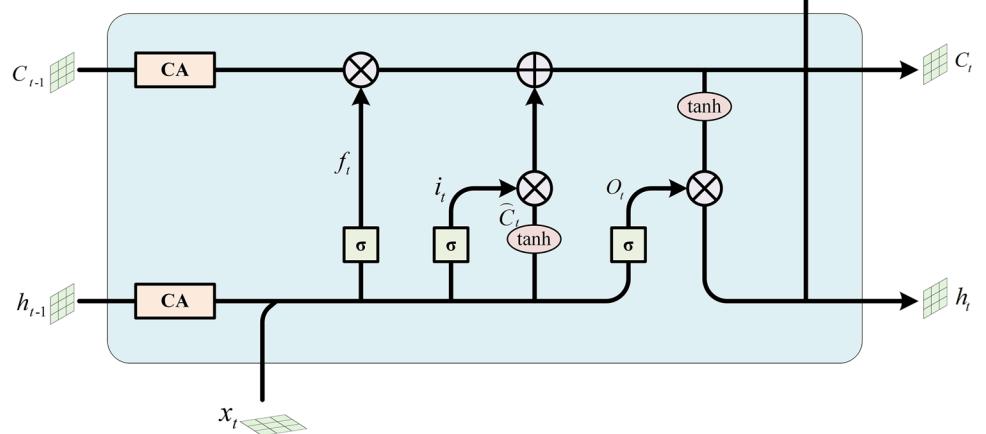
$$\text{MSE} = \frac{\sum_i (y_i - \bar{y}_i)^2}{n} \quad (12)$$

$$\text{RMSE} = \sqrt{\frac{\sum_i (y_i - \bar{y}_i)^2}{n}} \quad (13)$$

$$\text{PCC} = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)(\bar{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \sqrt{\sum_{i=1}^n (\bar{y}_i - \bar{y})^2}} \quad (14)$$

where y_i and \bar{y}_i are the true and predicted values of sample i , respectively, n is the total number of samples, \bar{y} is the average of y_i , and \bar{y}_i is the average of y_i .

Fig. 4 Internal schematic of the ConvLSTM-CA cell, where CA stands for channel attention



4.2 Overall predictive accuracy

To verify that the model proposed in this paper can show good air temperature prediction performance, the ConvLSTM-CASA model was compared with the DTR, MLR, RF, LSTM, and ConvLSTM models. To explore the separate roles of the spatial autocorrelation index and channel attention, ablation experiments adding channel attention to ConvLSTM (ConvLSTM-CA), and adding the spatial autocorrelation index to ConvLSTM (ConvLSTM-SA) were also conducted. The accuracy results are shown in Table 2. The results show the ConvLSTM-CASA model is significantly better than the other models in all accuracy indices. Among all models, DTR performs the worst, and all accuracy indices are inferior to those of the other models. MLR and RF have almost the same effects, with equal accuracy on R^2 and PCC, but MLR is slightly better than RF on MSE and

Table 2 In the comparison experiments of different models and the results of ablation experiments, the lower the MSE and RMSE the better, the higher the R^2 and PCC the better. Bold denotes the best evaluate index among all models

Models	MSE	RMSE	R^2	PCC
DTR	19.037	4.363	0.835	0.918
MLR	10.275	3.205	0.911	0.955
RF	10.279	3.206	0.911	0.955
LSTM	10.046	3.170	0.913	0.956
ConvLSTM	6.779	2.604	0.941	0.970
ConvLSTM-CA	6.468	2.543	0.944	0.972
ConvLSTM-SA	5.751	2.398	0.950	0.975
ConvLSTM-CASA	5.245	2.290	0.954	0.977

RMSE. LSTM is better than the above two models, but the advantage is very small. ConvLSTM significantly improved the accuracy compared with all the previous models, and the MSE index is the most obvious. The proposed ConvLSTM-CASA model improved all accuracy metrics compared to the ConvLSTM model, with a 72.45% improvement in MSE, 47.51% improvement in RMSE, 14.25% improvement in R^2 and 6.43% improvement in PCC compared to those of the least effective DTR model. The results of the ablation experiments also show that adding the spatial autocorrelation index alone and channel attention alone can improve the prediction effect, among which the spatial autocorrelation index improved more significantly. This indicates that the spatial autocorrelation index does have an impact on pentad-mean air temperature prediction.

The overall performance of ConvLSTM-CASA and the other five models are presented in Fig. 5. The x-axis is the observed value of pentad-mean air temperature, and the y-axis is the predicted value of pentad-mean air temperature, with red representing high scatter density and blue representing low scatter density. As a whole, DTR has the most scattered scatter, and LSTM, MLR, and RF have similar scatter distributions. ConvLSTM-CASA has the most concentrated scatter and is closest to the 1:1 line. This indicates that DTR has the worst fit, LSTM, MLR, and RF have a better fit with little difference and ConvLSTM-CASA has the best fit. Overall, the fit and overall accuracy of the six models remained consistent. The DTR is the worst, with an MSE of 19.037, RMSE of 4.363, and R^2 of 0.835. MLR, LSTM, and RF with little difference in accuracy. ConvLSTM-CASA has the best results, with an MSE of 5.245, RMSE of 2.290, and R^2 of 0.954. The results show that ConvLSTM-CASA can better simulate the air temperature variation trend compared

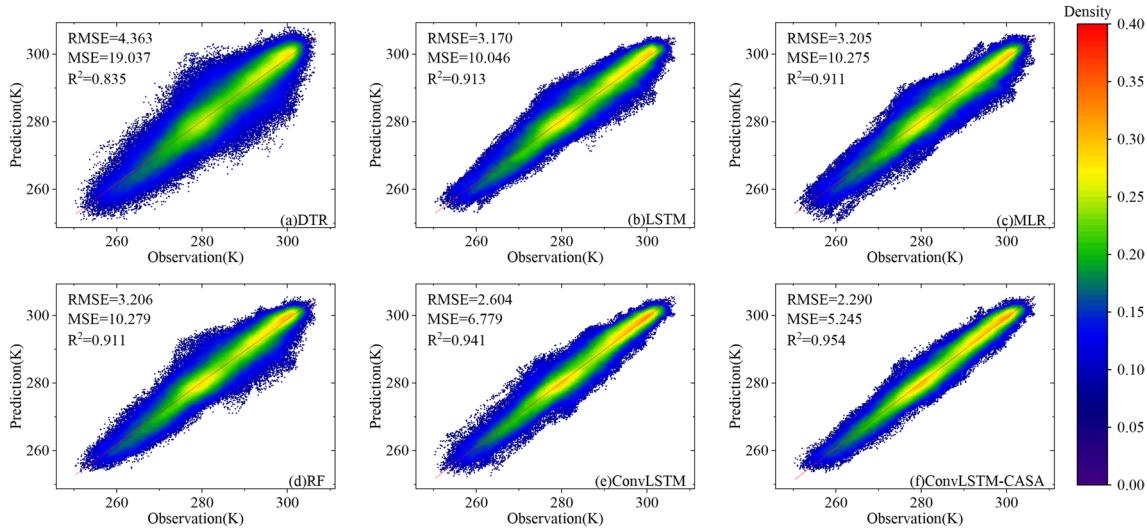


Fig. 5 Scatter density plots for different model comparisons: **a** DTR, **b** LSTM, **c** MLR, **d** RF, **e** ConvLSTM, **f** ConvLSTM-CASA

with the classical machine learning methods and deep learning methods.

4.3 Single-point accuracy evaluation

The dataset used in this paper is raster data, and each raster point is recorded with a pentad-mean air temperature value. Predictions are made for each raster point throughout the experiment, so the individual accuracy of each raster point is also calculated. The single-point accuracy plotted by the difference between the single-point accuracy predicted by the ConvLSTM-CASA model and the single-point accuracy predicted by the ConvLSTM model is shown in Fig. 6. Larger values of PCC and R^2 represent higher accuracy, so a larger difference between them represents a larger accuracy improvement. Smaller values of MSE and RMSE represent higher accuracy, so a smaller difference between them represents a larger accuracy improvement. The shades of orange are used in the figure to represent the single-point accuracy improvement. Four accuracy metrics show orange

color in the YRB, indicating that the ConvLSTM-CASA model has improved accuracy at all raster points compared with that of the ConvLSTM model. All four accuracy indicators are darker in orange in the lower YRB region and lighter in orange in the upper YRB region, suggesting that ConvLSTM-CASA model performs better in predicting pentad-mean air temperature in areas with lower altitudes and milder climates.

The single-point accuracy case for all models is reflected in Fig. 7. ConvLSTM-CASA has the highest mean and median values on PCC and has the smallest error fluctuations. ConvLSTM mean and median values are slightly lower than ConvLSTM-CASA but again, better than those of the other models. MLR and LSTM have the same mean and median values. RF has a slightly lower mean value than its median value, but the outliers are the least of all models. DTR has the lowest mean and median values and a more uneven distribution of values. The overall trend for R^2 is the same as for PCC. In the MSE, ConvLSTM-CASA has the lowest mean and median values and still maintains the

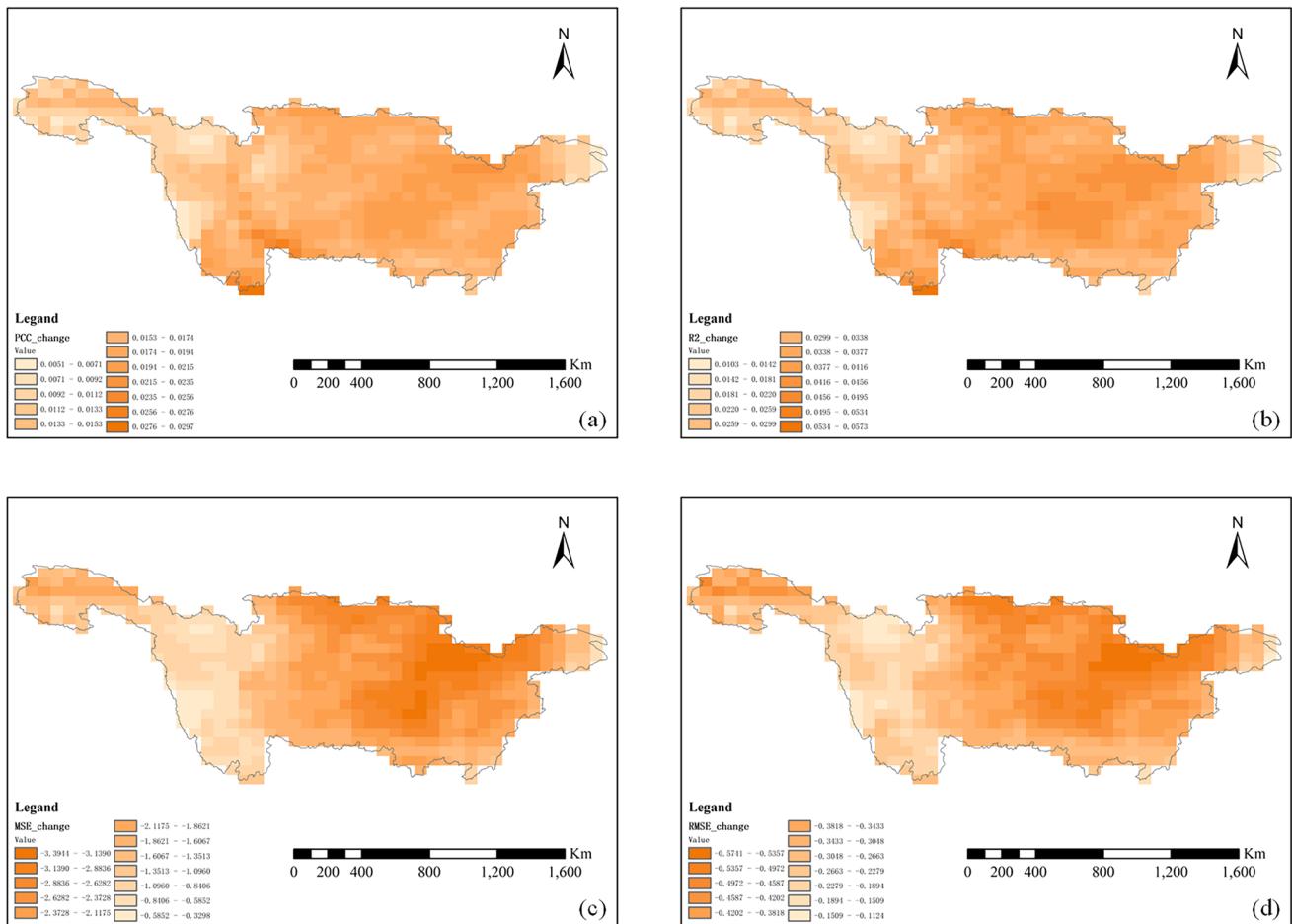
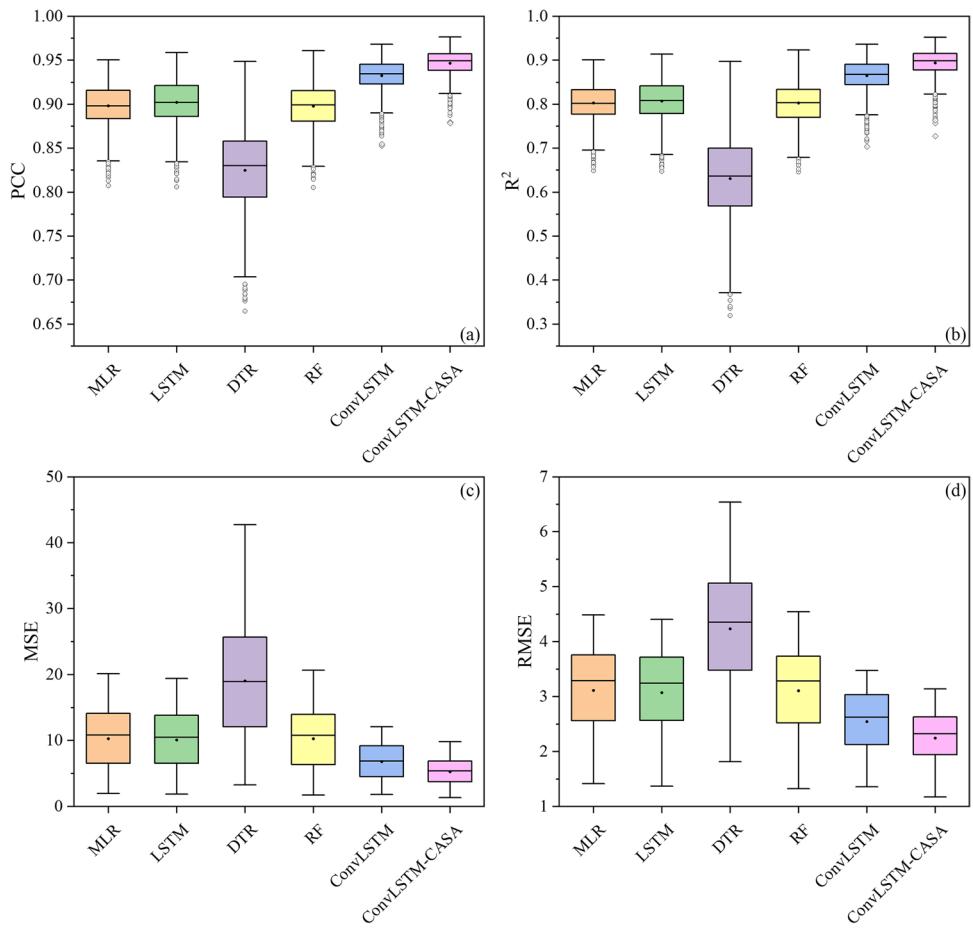


Fig. 6 Difference between the single-point accuracy predicted by the ConvLSTM-CASA model and the single-point accuracy predicted by the ConvLSTM model, with darker colors representing a greater degree of improvement

Fig. 7 Box plots of single-point accuracy for different models: **a** PCC, **b** R^2 , **c** MSE, **d** RMSE



smallest error fluctuations. The graphs of MLR, LSTM, and RF are almost identical. DTR is the least effective, but all models have no outliers. The overall trend of RMSE is the same as that of MSE. It is concluded that ConvLSTM-CASA also has advantages in single-point prediction.

4.4 Spatiotemporal accuracy evaluation

Air temperature prediction has great differences in different spaces. For example, in high-altitude mountains and plateaus, where the air temperature is low year-round, even extremely cold temperatures will occur. In low-altitude basins and plains, climate change is obvious year-round. The higher the latitude, the lower the air temperature, and the lower the latitude, the higher the air temperature. Therefore, differences in spatial location will affect the air temperature prediction capability. Air temperature prediction also varies considerably in different periods. For example, the air temperature is generally lower in winter and the air temperature is generally higher in summer. To verify the air temperature prediction ability of the ConvLSTM-CASA model in different spaces and at different times, the spatial and temporal distributions of the ConvLSTM-CASA model

and other models for pentad-mean air temperature prediction in the YRB (Fig. 8) are plotted. All data are the predicted and observed values of the pentad-mean air temperature from April 17, 2014 to June 4, 2022. To highlight the difference between the four seasons, the data from March to May are classified as spring. The data from June to August are classified as summer. The data from September to November are classified as autumn. The data from December to February are classified as winter and the data are averaged according to the four seasons. Figure 8 shows a spatial trend of generally lower temperature in the Qinghai-Tibet Plateau and higher temperature in the Sichuan Basin and middle and lower reaches of the Yangtze River Plain. The temporal temperature trend of low temperature in winter and high temperature in summer is presented. All models can reflect the spatiotemporal temperature trend and perform well for both very high and extreme temperature predictions. For normal temperature prediction, the ConvLSTM-CASA model has the closest prediction to the observed value. The ConvLSTM-CASA model is good at capturing the intrinsic patterns of small changes in air temperature.

To verify the predictive capability of the ConvLSTM-CASA model on temporal dynamics, the predicted values

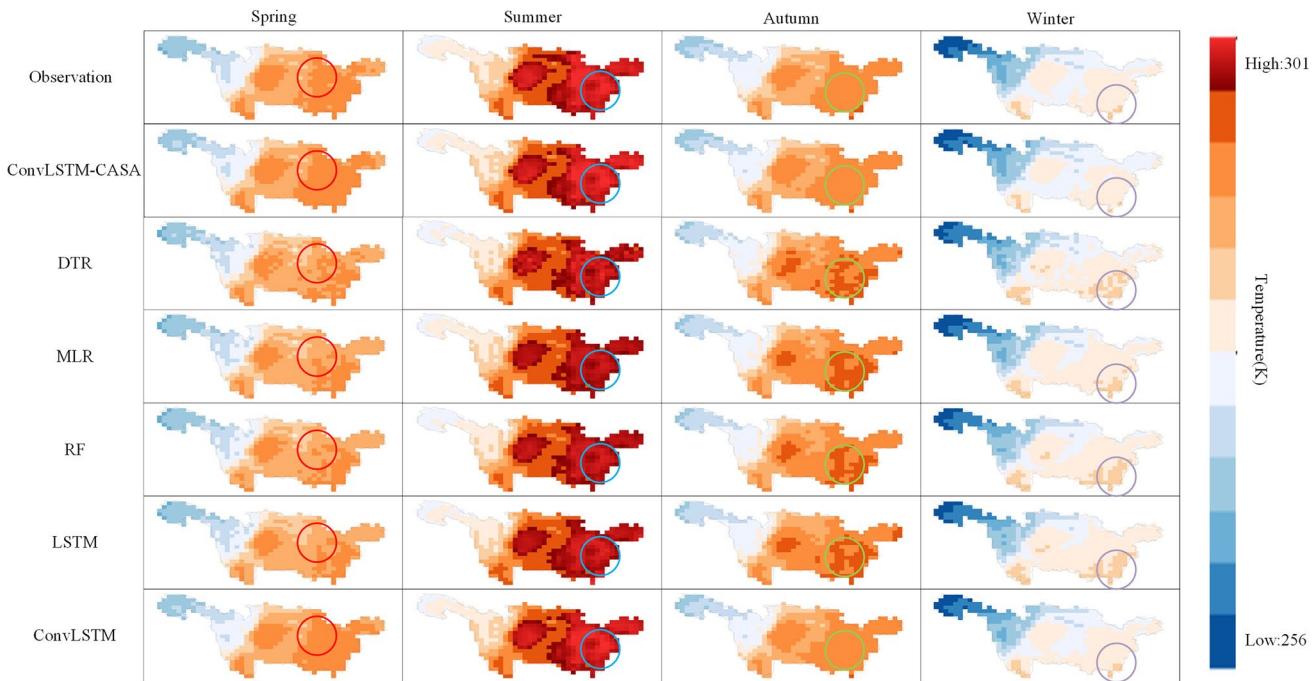


Fig. 8 Comparison between observed values and predicted values of different models in spatial and temporal distribution

of all raster points of each image predicted by ConvLSTM-CASA are averaged and the corresponding observations are averaged. The average predicted and observed values for the YRB overall are plotted as a line graph (Fig. 9). In general, the average predicted values are consistent with the trend of the average observed values. In Fig. 9a and b, other baseline models have weak predictions of deviation values and even predict wrong air temperature trends. ConvLSTM-CASA model predicts deviation values closer to observation. In Fig. 9c and d, the ConvLSTM-CASA model is stable and does not change much in the prediction. However, it has advantages in maximum and minimum values and overall trend prediction. The results demonstrate that the ConvLSTM-CASA model can effectively learn the intrinsic laws of air temperature change, making it outstanding in conventional temperature prediction.

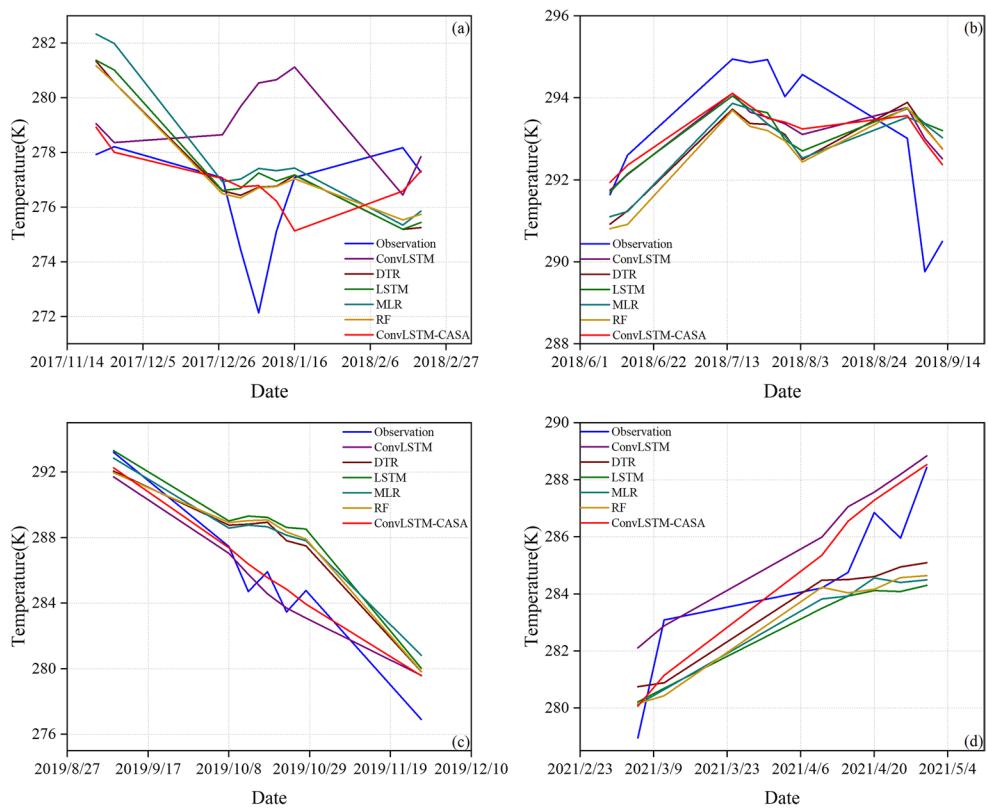
4.5 Strengths and limitations of the ConvLSTM-CASA model

Air temperatures are spatially continuous, with neighboring regions showing a great correlation in air temperature. Regions that are farther away may also exhibit a high correlation of air temperatures due to similar geography. Ordinary convolution calculations can consider only local spatial correlations and lack consideration of global spatial relationships. The ConvLSTM-CASA model can consider the global spatial relationships of the YRB by adding a spatial autocorrelation index and highlighting the impact of spatial

correlation on temperature prediction through a channel attention mechanism. This allows the ConvLSTM-CASA model to perform better in predicting temperatures over large areas than ordinary models. However, the ConvLSTM-CASA model also has shortcomings. First, single variable prediction is limited. Changes in air temperature are influenced by a variety of factors, such as rainfall, solar radiation, atmospheric circulation and so on. Of course, natural disasters are also an important part of this. In the follow-up study, we will also consider the influence of multiple factors and carry out the experiment of multivariate prediction.

The second is the lack of consideration of time relationships. Air temperature also varies continuously over time, with air temperature usually showing a cyclical pattern of variation over time. The ConvLSTM-CASA model captures the temporal pattern of air temperature changes through LSTM, which loses some of the temporal information. There are also current studies that consider both temporal and spatial relationships in ConvLSTM. Wang et al. (2018) proposed a spatiotemporal attention ConvLSTM for traffic prediction. Lin et al. (2020) proposed a self-attention ConvLSTM for spatiotemporal prediction. Spatiotemporal attention is used to value spatiotemporal relationships through spatial attention and temporal attention, while ignoring secondary information. Self-attention is used to remember spatiotemporal long-term dependent features to reduce information loss. Compared to these two methods, ConvLSTM-CASA is superior in terms of spatial relations but inferior in terms of temporal relations. In a follow-up study, the

Fig. 9 ConvLSTM-CASA, RF, MLR, LSTM, DTR, and ConvLSTM models predicted values versus observed values in a line chart



ConvLSTM-CASA model can also incorporate a temporal attention mechanism or a temporal correlation index to better account for the temporal pattern of air temperature.

5 Conclusion

In this study, a deep learning model ConvLSTM-CASA that accounts for spatial autocorrelation is proposed and successfully applied to the pentad-mean air temperature prediction task. Based on the air temperature data provided by the global climate reanalysis dataset MERRA-2, the corresponding spatial autocorrelation index is calculated and added to the model as a feature channel to focus on the spatial relationship of air temperature. In addition, a channel attention mechanism is added, through which the weights of different feature channels are adaptively adjusted to give full play to the spatial autocorrelation index. This allows the model to fully consider the spatial correlation of air temperature.

The experimental results show that the model proposed in this paper can have better pentad-mean air temperature prediction performance with an average R^2 of 0.954 and MSE of 5.245. MSE accuracy improvements of 72.45%, 48.95%, 48.97%, 47.79%, and 22.63% compared to the DTR, MLR, RF, LSTM, and ConvLSTM models, respectively. Ablation experiments also confirm the effectiveness of the added spatial autocorrelation index and

channel attention mechanism. Machine learning models are less effective in single-point prediction, and the ConvLSTM-CASA model also has a greater advantage over traditional machine learning or deep learning models. The performance of the model under different spaces and times was analyzed. The ConvLSTM-CASA model can make more accurate predictions in both higher and lower altitude areas in both summer and winter. Overall, the ConvLSTM-CASA model proposed in this paper can better consider spatial and temporal characteristics in pentad-mean air temperature prediction and be applied to a variety of terrain and climate environments. When making predictions about air temperature, one must contend with significant levels of uncertainty that arise from the variability of meteorological conditions and the complexity of environmental dynamics. How to model and control the uncertainty is a challenge for deep learning methods in air temperature prediction. In future research, methods to reduce data and model uncertainty and improve the interpretability of air temperature prediction will be explored.

Author contribution Lei Xu and Xi Zhang designed the research. Lei Xu and Xi Zhang completed the experiment. Xi Zhang wrote the first draft of the manuscript. All authors conducted the manuscript review, edited, and wrote the final paper. All authors have read and agreed to the published version of the manuscript.

Funding This research was supported by the National Key Research and Development Program for Young Scientist (2021YFF0704400), China Postdoctoral Science Foundation (2022M722930), the National Natural Science Foundation of China (42201509, 42101429), and the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (162301212687).

Data availability The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Code availability Code will be available on request to the corresponding author.

Declarations

Ethics approval Not applicable.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication All authors agreed to let the paper published when considered for publication.

Competing interests The authors declare no competing interests.

References

- Anandaraja N, Rathakrishnan T, Ramasubramanian M, Saravanan P, Suganthi NS (2008) Indigenous weather and forecast practices of coimbatore district farmers of Tamil Nadu. Indian J Tradit Know 7(4):630–633
- Ayal DY, Desta S, Gebre G, Kinyangi J, Recha J, Radeny M (2015) Opportunities and challenges of indigenous biotic weather forecasting among the borena herders of southern ethiopia. Springerplus 4(1):1–11. <https://doi.org/10.1186/s40064-015-1416-6>
- Bauer P, Thorpe A, Brunet G (2015) The quiet revolution of numerical weather prediction. Nature 525(7567):47–55. <https://doi.org/10.1038/nature14956>
- Breiman L (2001) Random forests. Mach Learn 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Bunker A, Wildenhain J, Vandenbergh A, Henschke N, Rocklöv J, Hajat S, Sauerborn R (2016) Effects of air temperature on climate-sensitive mortality and morbidity outcomes in the elderly; A systematic review and meta-analysis of epidemiological evidence. Ebiomedicine 6:258–268. <https://doi.org/10.1016/j.ebiom.2016.02.034>
- Choe Y, Yom J (2020) Improving accuracy of land surface temperature prediction model based on deep-learning. Spat Inf Res 28(3):377–382. <https://doi.org/10.1007/s41324-019-00299-5>
- Chou C, Chen C, Tan P, Chen KT (2012) Mechanisms for global warming impacts on precipitation frequency and intensity. J Clim 25(9):3291–3306. <https://doi.org/10.1175/JCLI-D-11-00239.1>
- Di Cecco GJ, Gouhier TC (2018) Increased spatial and temporal auto-correlation of temperature under climate change. Sci Rep 8(1). <https://doi.org/10.1038/s41598-018-33217-0>
- Eredics P (2009) Short-term external air temperature prediction for an intelligent greenhouse by mining climatic time series. In: 2009 IEEE International Symposium on Intelligent Signal Processing. IEEE, pp 317–322. <https://doi.org/10.1109/WISP.2009.5286544>
- Fallah Ghahhari G, Dadashi Roudbari A (2018) An investigation on thermal patterns in iran based on spatial autocorrelation. Theor Appl Climatol 131(3-4):865–876. <https://doi.org/10.1007/s00704-016-2015-3>
- Galacgac ES, Balisacan CM (2009) Traditional weather forecasting for sustainable agroforestry practices in Ilocos Norte Province, Philippines. For Ecol Manage 257(10):2044–2053. <https://doi.org/10.1016/j.foreco.2009.01.002>
- Gers FA, Schmidhuber J, Cummins F (2000) Learning to forget: continual prediction with lstm. Neural Comput 12(10):2451–2471. <https://doi.org/10.1162/089976600300015015>
- Getis A (2007) Reflections on spatial autocorrelation. Reg Sci Urban Econ 37(4):491–496. <https://doi.org/10.1016/j.regsciurbeco.2007.04.005>
- Gong B, Langguth M, Ji Y, Mozaffari A, Stadtler S, Mache K, Schultz MG (2022) Temperature forecasting by deep learning methods. Geosci Model Dev 15(23):8931–8956. <https://doi.org/10.5194/gmd-15-8931-2022>
- Guo M, Xu T, Liu J, Liu Z, Jiang P, Mu T, Zhang S, Martin RR, Cheng M, Hu S (2022) Attention mechanisms in computer vision: a survey. Comput Vis Media (Beijing) 8(3):331–368. <https://doi.org/10.1007/s41095-022-0271-y>
- Ji F, Wu Z, Huang J, Chassaignet EP (2014) Evolution of land surface air temperature trend. Nat Clim Chang 4(6):462–466. <https://doi.org/10.1038/nclimate2223>
- Li Q, Zhu Y, Shangguan W, Wang X, Li L, Yu F (2022) An attention-aware lstm model for soil moisture and soil temperature prediction. Geoderma 409:115651. <https://doi.org/10.1016/j.geoderma.2021.115651>
- Lim B, Zohren S (2021) Time-series forecasting with deep learning: a survey. Philos Trans Royal Soc A 379(2194):20200209. <https://doi.org/10.1098/rsta.2020.0209>
- Lin H, Hua Y, Ma L, Chen L (2019) Application of convlstm network in numerical temperature prediction interpretation. In: Proceedings of the 2019 11th International Conference on Machine Learning and Computing, pp 109–113. <https://doi.org/10.1145/3318299.3318381>
- Lin Z, Li M, Zheng Z, Cheng Y, Yuan C (2020) Self-attention convlstm for spatiotemporal prediction Proceedings of the AAAI conference on artificial intelligence, pp 11531–11538. <https://doi.org/10.1609/aaai.v34i07.6819>
- Liu J, Xu L, Chen N (2022) A spatiotemporal deep learning model st-lstm-sa for hourly rainfall forecasting using radar echo images. J Hydrol (Amst) 609:127748. <https://doi.org/10.1016/j.jhydrol.2022.127748>
- Nastos PT, Matzarakis A (2012) The effect of air temperature and human thermal indices on mortality in athens, greece. Theor Appl Climatol 108:591–599. <https://doi.org/10.1007/s00704-011-0555-0>
- Niu Z, Zhong G, Yu H (2021) A review on the attention mechanism of deep learning. Neurocomputing 452:48–62. <https://doi.org/10.1016/j.neucom.2021.03.091>
- Okonya JS, Kroschel J (2013) Indigenous knowledge of seasonal weather forecasting: a case study in six regions of uganda. Agric Sci 2013. <https://doi.org/10.4236/as.2013.412086>
- Ord JK, Getis A (1995) Local spatial autocorrelation statistics: distributional issues and an application. Geogr Anal 27(4):286–306
- Ostberg S, Lucht W, Schaphoff S, Gerten D (2013) Critical impacts of global warming on land ecosystems. Earth Syst Dyn 4(2):347–357. <https://doi.org/10.5194/esd-4-347-2013>
- Rao Z, Xu Y, Pan S, Guo J, Yan Y, Wang Z (2023) Cellular traffic prediction: a deep learning method considering dynamic nonlocal spatial correlation, self-attention, and correlation of spatiotemporal feature fusion. Ieee Trans Netw Serv Manag 20(1):426–440. <https://doi.org/10.1109/TNSM.2022.3187251>
- Rautela P, Karki B (2015) Weather forecasting: traditional knowledge of the people of uttarakhand himalaya. J Geogr Environ Earth Sci Int 3(3):1–14

- Reichle RH, Draper CS, Liu Q, Girotto M, Mahanama SPP, Koster RD, De Lannoy GJM (2017) Assessment of merra-2 land surface hydrology estimates. *J Clim* 30(8):2937–2960. <https://doi.org/10.1175/JCLI-D-16-0720.1>
- Safavian SR, Landgrebe D (1991) A survey of decision tree classifier methodology. *Ieee Trans Syst Man Cybern* 21(3):660–674. <https://doi.org/10.1109/21.97458>
- Seong N (2021) Deep spatiotemporal attention network for fine particle matter 2.5 concentration prediction with causality analysis. *Ieee Access* 9:73230–73239. <https://doi.org/10.1109/ACCESS.2021.3080828>
- Shen H, Jiang Y, Li T, Cheng Q, Zeng C, Zhang L (2020) Deep learning-based air temperature mapping by fusing remote sensing, station, simulation and socioeconomic data. *Remote Sens Environ* 240:111692. <https://doi.org/10.1016/j.rse.2020.111692>
- Shi X, Chen Z, Wang H, Yeung D, Wong W, Woo W (2015) Convolutional lstm network: a machine learning approach for precipitation nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1, pp 802–810
- Shi L, Liang N, Xu X, Li T, Zhang Z (2021) Sa-jstn: self-attention joint spatiotemporal network for temperature forecasting. *Ieee J Sel Top Appl Earth Obs Remote Sens* 14:9475–9485. <https://doi.org/10.1109/JSTARS.2021.3112131>
- Singh BR, Singh O (2012) Study of impacts of global warming on climate change: rise in sea level and disaster frequency. In: Global Warming-Impacts and Future Perspective. IntechOpen, Rijeka
- Skripnuk DF, Samylovskaya EA (2018) Human activity and the global temperature of the planet IOP conference series: earth and environmental science. IOP Publishing, p 12021. <https://doi.org/10.1088/1755-1315/180/1/012021>
- Song Y, Song J (2022) Analysis of surface temperature in an urban area using supervised spatial autocorrelation and moran's i. *Earth Sci Inform* 15(4):2545–2552. <https://doi.org/10.1007/s12145-022-00856-x>
- Tabrizi SE, Xiao K, Van Griensven TJ, Saad M, Farghaly H, Yang SX, Gharabaghi B (2021) Hourly road pavement surface temperature forecasting using deep learning models. *J Hydrol (Amst)* 603:126877. <https://doi.org/10.1016/j.jhydrol.2021.126877>
- Tektaş M (2010) Weather forecasting using anfis and arima models. *Environ Res Eng Manag* 51(1):5–10
- Tiefelsdorf M, Boots B (1997) A note on the extremities of local moran's iis and their impact on global moran's i. *Geogr Anal* 29(3):248–257
- Tong W, Chen W, Han W, Li X, Wang L (2020) Channel-attention-based densenet network for remote sensing image scene classification. *Ieee J Sel Top Appl Earth Obs Remote Sens* 13:4121–4132. <https://doi.org/10.1109/JSTARS.2020.3009352>
- Uyanik GK, Güler N (2013) A study on multiple linear regression analysis. *Proc Soc Behav Sci* 106:234–240. <https://doi.org/10.1016/j.sbspro.2013.12.027>
- Wang D, Yang Y, Ning S (2018) Deepstcl: a deep spatio-temporal convlstm for travel demand prediction 2018 international joint conference on neural networks (IJCNN). IEEE, pp 1–8. <https://doi.org/10.1109/IJCNN.2018.8489530>
- Wang S, Cao J, Philip SY (2020) Deep learning for spatio-temporal data mining: a survey. *Ieee Trans Knowl Data Eng* 34(8):3681–3700. <https://doi.org/10.1109/TKDE.2020.3025580>
- Woo S, Park J, Lee J, Kweon IS (2018) Cbam: convolutional block attention module. *Proceedings of the European conference on computer vision (ECCV)*, pp 3–19. <https://doi.org/10.48550/arXiv.1807.06521>
- Wu S, Fu F, Wang L, Yang M, Dong S, He Y, Zhang Q, Guo R (2022) Short-term regional temperature prediction based on deep spatial and temporal networks. *Atmosphere (Basel)* 13(12):1948. <https://doi.org/10.3390/atmos13121948>
- Xiao C, Chen N, Hu C, Wang K, Xu Z, Cai Y, Xu L, Chen Z, Gong J (2019) A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. *Environ Model Softw* 120:104502. <https://doi.org/10.1016/j.envsoft.2019.104502>
- Xu L, Abbaszadeh P, Moradkhani H, Chen N, Zhang X (2020) Continental drought monitoring using satellite soil moisture, data assimilation and an integrated drought index. *Remote Sens Environ* 250:112028. <https://doi.org/10.1016/j.rse.2020.112028>
- Xu L, Chen N, Chen Z, Zhang C, Yu H (2021a) Spatiotemporal forecasting in earth system science: methods, uncertainties, predictability and future directions. *Earth Sci Rev* 222:103828. <https://doi.org/10.1016/j.earscirev.2021.103828>
- Xu L, Chen N, Yang C, Yu H, Chen Z (2022) Quantifying the uncertainty of precipitation forecasting using probabilistic deep learning. *Hydrol Earth Syst Sci* 26(11):2923–2938. <https://doi.org/10.5194/hess-26-2923-2022>
- Xu L, Chen N, Zhang X, Chen Z, Hu C, Wang C (2019) Improving the north american multi-model ensemble (nmme) precipitation forecasts at local areas using wavelet and machine learning. *Clim Dyn* 53(1–2):601–615. <https://doi.org/10.1007/s00382-018-04605-z>
- Xu L, Chen N, Zhang X, Moradkhani H, Zhang C, Hu C (2021b) In-situ and triple-collocation based evaluations of eight global root zone soil moisture products. *Remote Sens Environ* 254:112248. <https://doi.org/10.1016/j.rse.2020.112248>
- Yang Y, Dong J, Sun X, Lima E, Mu Q, Wang X (2018) A cfcc-lstm model for sea surface temperature prediction. *Ieee Geosci Remote Sens Lett* 15(2):207–211. <https://doi.org/10.1109/LGRS.2017.2780843>
- Yuan X, Li L, Shardt YAW, Wang Y, Yang C (2021) Deep learning with spatiotemporal attention-based lstm for industrial soft sensor model development. *Ieee Trans Ind Electron* 68(5):4404–4414. <https://doi.org/10.1109/TIE.2020.2984443>
- Zhang L, Li D, Guo Q (2020) Deep learning from spatio-temporal data using orthogonal regularization residual cnn for air prediction. *Ieee Access* 8:66037–66047. <https://doi.org/10.1109/ACCESS.2020.2985657>
- Zhu A, Lu G, Liu J, Qin C, Zhou C (2018) Spatial prediction based on third law of geography. *Ann Gis* 24(4):225–240. <https://doi.org/10.1080/19475683.2018.1534890>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”). Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com