Urban land surface temperature prediction using parallel STL-Bi-LSTM neural network

Xing Huo,[a] Guangpeng Cui,[a] Lingling Ma,[b,*] Bohui Tang[c]
Ronglin Tang,[d] Kun Shao,[a] and Xinhong Wang[b]
[a]Hefei University of Technology, Hefei, China
[b]Chinese Academy of Sciences, Aerospace Information Research Institute, Key Laboratory of Quantitative Remote Sensing Information Technology, Beijing, China
[c]Kunming University of Science and Technology, Kunming, China
[d]Institute of Geographic Sciences and Natural Resources Research Chinese Academy of Sciences, State Key Laboratory of Resources and Environment Information System, Beijing, China

Abstract. Accurate temperature prediction is of great significance to human life and social economy. A series of traditional methods and machine learning methods have been proposed to achieve temperature prediction, but it is still a challenging problem. We propose a temperature prediction model that combines seasonal and trend decomposition using loess (STL) and the bidirectional long short-term memory (Bi-LSTM) network to achieve high-accuracy prediction of the daily average temperature of China cities. The proposed model decomposes the temperature data using STL into trend component, seasonal component, and remainder component. Decomposition components and the original temperature data are input into the two-layer Bi-LSTM to learn the features of the temperature data, and the sum of prediction of three components and the original temperature data prediction result are added using learnable weights as the prediction result. The experimental results show that the average root mean square error and mean absolute error of the proposed model on the testing data are 0.11 and 0.09, respectively, which are lower than 0.35 and 0.27 of STL-LSTM, 2.73 and 2.07 of EMD-LSTM, 0.39 and 0.15 of STL-SVM, achieving a higher precision temperature prediction. © *The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JRS.16.034529]

1 Introduction

Temperature is closely related to human life, agricultural production, and social economy, and it affects all aspects of life. Studies showed that the increase in temperature reduced crop yields.[1,2] Doi et al.[3] presented that temperature change had an impact on deep-sea biodiversity. Dottori et al.[4] proposed river floods with warming would increase human and economic loss. Temperature change also increased building energy consumption.[5] The most serious issue is that temperature change would have an impact on the spread of diseases and endanger people's lives.[6,7] Accurate prediction of temperature is significant to protect people's lives and property and maintain stable economic development. However, temperature prediction is very challenging due to various uncertain relevant factors.

A series of forecasting methods including conventional methods and machine learning methods were proposed to predict temperature. Wang et al.[8] proposed an improved support vector machine (SVM) to predict the daily minimum temperature; Babu et al.[9] used different autoregressive integral moving average (ARIMA) models to predict the average global temperature;

*Address all correspondence to Lingling Ma, mall@aircas.ac.cn

Jallal et al.[10] proposed an artificial neural network (ANN) with delayed exogenous input to forecast air temperature on a half-hour scale. With the appearance of recurrent neural network (RNN), more and more methods based on RNN are used to solve the problem of temperature prediction. The long short-term memory (LSTM) network is one of the most popular methods. Li et al.[11] provided temperature prediction results every half an hour using stacked LSTM network. Qi and Guo[12] predicted the next hour's temperature using the LSTM network by fully considering the historic temperature and meteorological condition. Huang et al.[13] proposed multistep temperature prediction model using the LSTM network based on temperature data of surrounding cities. Sadeque and Bui[14] provided cascaded LSTM network for weather forecasting that can outperform some of the existing well-known models. Joanna et al.[15] presented an outdoor air-temperature time-series prediction model for a multifamily building using ANNs and obtained outstanding prediction results by selecting the best combination of predictors and the optimal number of neurons in a hidden layer. Wang et al.[16] proposed the development and evaluation of a new algorithm based on pattern approximate matching to predict the temperature of five cities in China. Yu et al.[17] presented an air temperature forecasting framework based on graph attention network and the gated recurrent unit, which overcame the flaw of the conventional graph network and achieved the best performance. Hrachya et al.[18] implemented a weather prediction technique based on machine learning to improve the hourly air temperature prediction for up to 24 h. Toni et al.[19] combined LSTM and prophet model to forecast 5-year daily air temperatures in Bandung; the results showed that the combination of two networks performed well for the prediction of low temperature and high temperature.

For time series data prediction, time series decomposition methods have an inspiring effect on improving the accuracy of time series data forecasting. Zhang[20] achieved foreign exchange rate forecasting using the combined model of empirical mode decomposition (EMD) and LSTM network. Jin et al.[21] suggested a vegetable price forecasting model using seasonal and trend decomposition using loess (STL) and LSTM network. Wang and Lou[22] proposed a hydrological time series forecast model based on wavelet denoising and ARIMA-LSTM that can be well adapted to the hydrological time series forecast and has the best forecast effect. Duan et al.[23] forecasted base station traffic using STL-LSTM networks, which have better performance compared with the other algorithms. Huo et al.[24] solved the problem of long-term span traffic prediction using STL and LSTM model. Yin et al.[25] updated the STL-LSTM model using an attention mechanism to achieve high accuracy of vegetable price forecasting. Chen et al.[26] forecasted the short-term metro ridership using the STL-LSTM model and proved it can achieve high accuracy.

In this paper, we propose a prediction model combining the STL method and bidirectional long short-term memory (Bi-LSTM) neural network to achieve the prediction of daily average temperature. Because temperature is affected by a variety of uncertain factors, it is difficult to obtain satisfactory results in temperature prediction directly using deep learning models. The time series decomposition method can be used to decompose periodic time series into trend components, seasonal components, and residual components. In a general periodic time series, the trend component generally represents the low-frequency variation, whereas the seasonal component represents the periodic variation. For temperature data, the seasonal component represents the periodic fluctuation of temperature with seasonal changes. The trend component represents changes in temperature that are influenced by other factors, such as increase in carbon dioxide. This paper first uses the STL method to decompose temperature data into trend component, seasonal component, and remainder component and then inputs the decomposition components and the original data into the two-layer Bi-LSTM neural network for training. The final output of the network is the predicted temperature.

The structure of the paper is as follows: Sec. 1 is the introduction; Sec. 2 introduces the study area; Sec. 3 is the related works, including the STL method, LSTM model, and Bi-LSTM model; Sec. 4 shows the structure of the proposed model; Sec. 5 is the experimental results; Sec. 6 is the conclusion.

2 Study Area

China is located in eastern Asia and on the western coast of the Pacific Ocean. It has a vast territory with a total land area of ~9.6 million km^2. The terrain of China is higher in the west

and lower in the east and is distributed in a stepped manner. The combination of temperature and precipitation is diverse, forming a diverse climate. There are a total of 34 provincial-level administrative units, including 23 provinces, 5 autonomous regions, 4 municipalities, and 2 special administrative regions.

3 Materials and Methods

We introduce the materials and related methods including seasonal and trend decomposition using loess, LSTM model, and Bi-LSTM model.

3.1 Materials

The temperature data for training and testing are land surface temperature, which are acquired from the weather station around the capital cities of China's 34 provincial-level administrative regions through Ref. 27. In the official documentation, the daily average temperature refers to the mean temperature calculated from the temperature of the day for 24 h in degrees Fahrenheit to tenths. We downloaded the daily average temperature of the mentioned 34 cities from 2010 to 2020 for training and testing, and there are about 136,510 pieces of temperature data, in which 95% of these data are used for training and the rest of them are used for testing. To test the robustness of the proposed model, additional testing data are added in the testing stage. According to the climatic conditions and geographic location, China is regionalized into four regions: north region, south region, northwest region, and Tibetan region (four regions).[28] Fifteen cities were selected from each region evenly to test the robustness of the model to the temperature of different regions. About 240,900 pieces of daily average temperature, data from 2010 to 2020 were downloaded and all of these data are used for model testing.

3.2 Methods

3.2.1 Seasonal and trend decomposition using loess

STL is a time series decomposition method proposed by Cleveland et al.,[29] which can decompose a time series into trend, seasonal, and remainder components based on loess. Suppose there is a temperature time series X_v, STL can decompose X_v into three addictive components: trend component (T_v), seasonal component (S_v), and remainder component (R_v), which can be expressed as follows:

$$X_v = T_v + S_v + R_v. \tag{1}$$

Suppose x_i and y_i are measurements of an independent and dependent variables for $i = 1$ to n. $g(x)$ is a smoothing of y given x that can be computed for any value x along the scale of the independent variable. That is, loess is defined everywhere and not just at the x_i; as we shall see, this is an important feature that in STL will allow us to deal with missing values and detrend the seasonal component in a straightforward way. That is, loess can be used to smooth y as a function of any number of independent variables, but for STL, only the case of one independent variable is needed. $g(x)$ is computed as follows: given a positive integer q, when $q \leq n$, the q values of the x_i that are closest to x are selected and each is given a neighborhood weight based on its distance from x. Let $\lambda_q(x)$ be the distance of the q'th farthest x_i from x. Let W be the tricube weight function:

$$W(u) = \begin{cases} (1 - u^3)^3 & \text{for } 0 \leq u < 1 \\ 0 & \text{for } u \geq 1 \end{cases}. \tag{2}$$

The neighborhood weight for any x_i is

$$v_i(x) = W\left(\frac{|x_i - x|}{\lambda_q(x)}\right). \tag{3}$$

The next step is to fit a polynomial of degree d to the data with weight $v_i(x)$ at (x_i, y_i). The value of the locally fitted polynomial at x is $g(x)$.

STL consists of two procedures: one is an inner loop and the other is an outer loop. The inner loop is nested inside the outer loop. The inner loop is used to update the seasonal component and trend component; the process of k'th epoch is as follows:

(1) Detrending. Calculate a detrended series by subtracting trend series from original series: $X_v - T_v^{(k)}$.
(2) Cycle-subseries smoothing. Each cycle-subseries obtained from step (1) is smoothed by loess and the result is the preliminary seasonal series $C_t^{(k+1)}$, consisting of $N + 2\times$ frequency values that range from $v = -\text{frequency} + 1$ to $N + \text{frequency}$, in which N is the length of data.
(3) Low-pass filtering of smoothed cycle-subseries. The preliminary seasonal series of step (2) is processed by a low-pass filter consisting of moving average of length frequency and the remainder trend series $L_t^{(k+1)}$ is obtained by loess.
(4) Detrending of smoothed cycle-subseries. Calculate seasonal series $S_t^{(k+1)} = C_t^{(k+1)} - L_t^{(k+1)}$.
(5) Deseasonalizing. Get deseasonalized series: $Y_t - S_t^{(k+1)}$.
(6) Trend smoothing. The deseasonalized series of step (5) is smoothed by loess to obtain the trend component $T_t^{(k+1)}$.

The outer loop starts when the inner loop reaches the accuracy requirement. The remainder component $R_t^{(k+1)}$ is computed in the outer loop using the estimated trend and seasonal components:

$$R_t^{(k+1)} = Y_t - T_t^{(k+1)} - S_t^{(k+1)}. \tag{4}$$

3.2.2 LSTM model

The RNN is designed to process sequence information, such as speech recognition and machine translation. The disadvantage of this method is long-term dependencies which will lead to gradient disappearance. The emergence of LSTM fix this problem. Different from RNN, LSTM proposed by Hochreiter and Schmidhuber[30] adds structure of forget gate, input gate, and update gate to forget and update information to the cell state.

The structure of LSTM is shown in Fig. 1. The first step of LSTM is to determine what information of cell state needs to discard, which is handled by forget gate using a sigmoid function:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{5}$$

where x_t is the input of current step, h_{t-1} is the hidden state of the previous step, W_f is the learnable weight of the forget gate, b_f is the learnable bias, σ is the sigmoid function, and f_t is the output of the forget gate.

The input gate determines what new information obtained from the current input x_t can be saved in the current cell state C_t. This process has three steps. At first, determine what information to update from the current input x_t using a sigmoid function. Then, obtain a candidate vector $\tilde{C}_t$ using a *tanh* function. Finally, update current cell state C_t in terms of cell state of previous step C_{t-1} and candidate vector $\tilde{C}_t$. The operations are as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{6}$$

$$\tilde{C}_t = \tan h(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{7}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t. \tag{8}$$

The output gate determines the output state of current step:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{9}$$
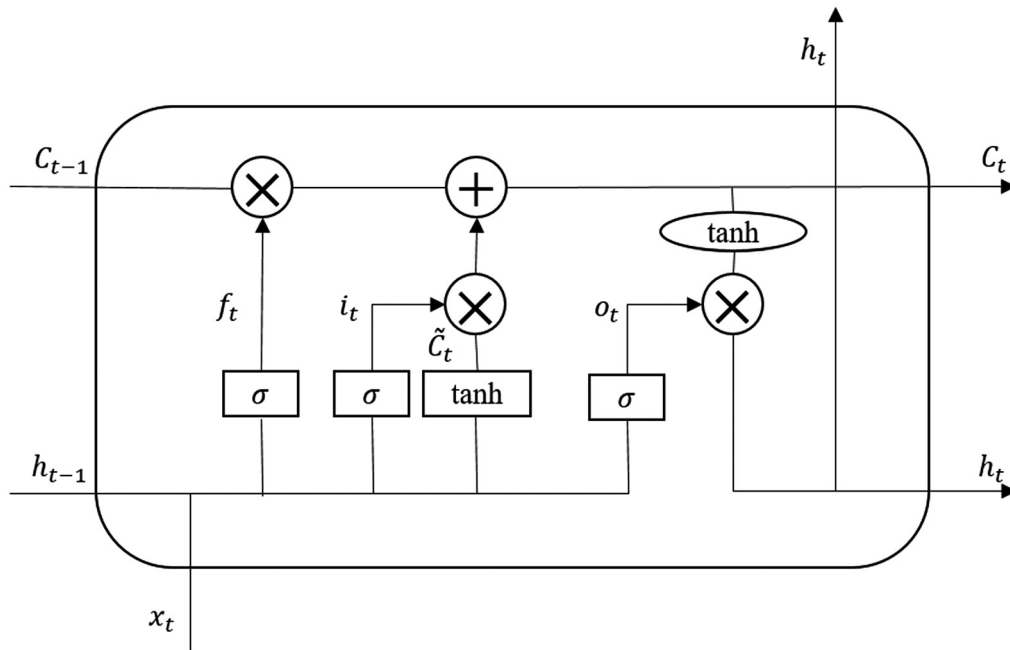
$$h_t = o_t * \tan h(C_t), \tag{10}$$
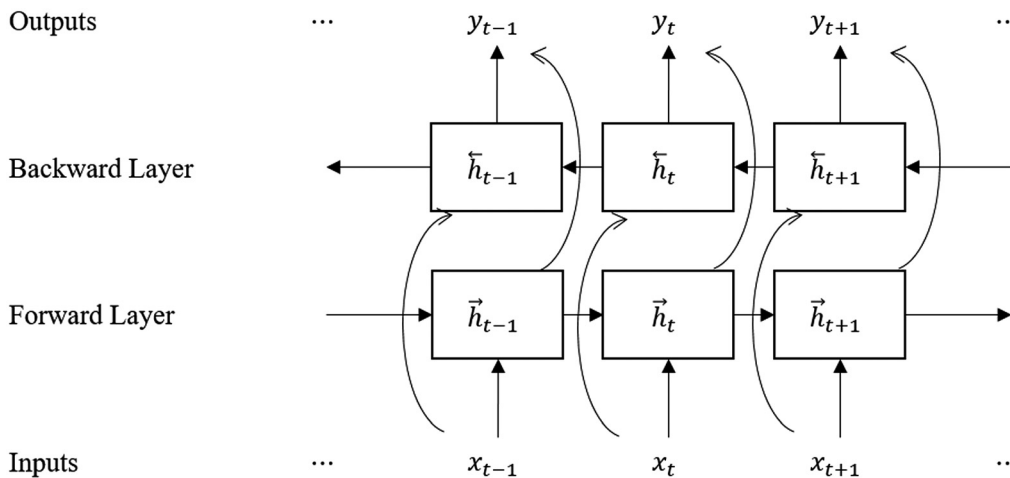
Fig. 1 Structure of LSTM model.

Fig. 2 Structure of Bi-LSTM model.

3.2.3 *Bi-LSTM model*

Bi-LSTM is based on LSTM and combines the information of the input sequence in both the forward and backward directions, which can better capture the two-way semantic dependence. The structure of Bi-LSTM is shown in Fig. 2. The LSTM unit has limitations, and it is able to make predictions using past data but not future data. Bi-LSTM overcomes the limitations of LSTM, it consists of two different LSTM hidden layers with opposite output directions. Under this structure, both backward and forward information can be utilized in the output layer. Bi-LSTM has advantages that LSTM does not have, so we choose the Bi-LSTM network as the temperature prediction network.

4 Proposed Model

To improve the accuracy of temperature prediction, we built a temperature prediction model combining STL and Bi-LSTM named parallel STL-Bi-LSTM neural network.
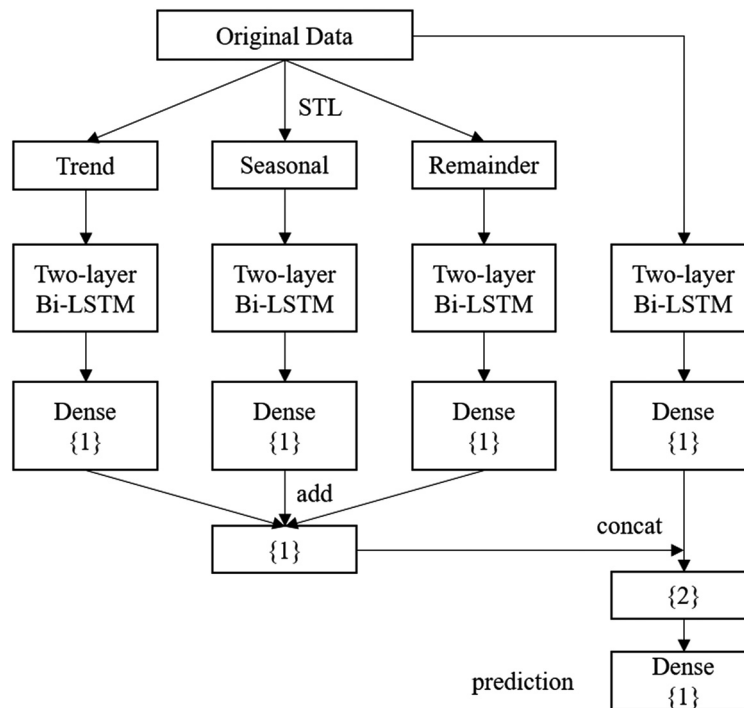
Fig. 3 Structure of parallel STL-Bi-LSTM model.

As shown in Fig. 3, the proposed model mainly has five steps: at first, original temperature data are decomposed into three components: trend component, seasonal component, and remainder component using the STL method. Then, original temperature data, trend component, seasonal component, and remainder component are input into two-layer Bi-LSTM neural network. The three-part decomposed data obtain one predicted value through the two-layer Bi-LSTM network and the fully connected layer, respectively, then adding the three predicted values to obtain one predicted value of the decomposed data. The origin temperature data obtain one predicted value through the two-layer Bi-LSTM network and a fully connected layer. The two predicted values are merged to obtain a feature containing two predicted values. The final predicted value is obtained by adding the two values through a fully connected layer using learnable weights.

5 Experiment

The experiment has three parts: first, the original temperature data are processed to input into the proposed model. Second, we determine the optimal parameters of the proposed model based on the performance of different parameters on the same testing data. Third, the proposed model is evaluated on testing data and the prediction results are compared with other prediction models.

5.1 Data Preparation

The temperature data need to be processed before inputting into the neural network. For the input of the proposed model, the original temperature data and the trend component, seasonal component, and remainder component need to be converted to input and output pairs for training. The process is as follows: decompose original data using the STL method into three components: trend component, seasonal component, and remainder component. Figure 4 shows an example of STL decomposition results. Frequency is the number of observations in each period, or cycle, of the seasonal component. The trend component in different years is different, because the temperature in different years is different and the decomposed trend data will also be different. The seasonal component obtained by STL decomposition is variation in the data at or near the
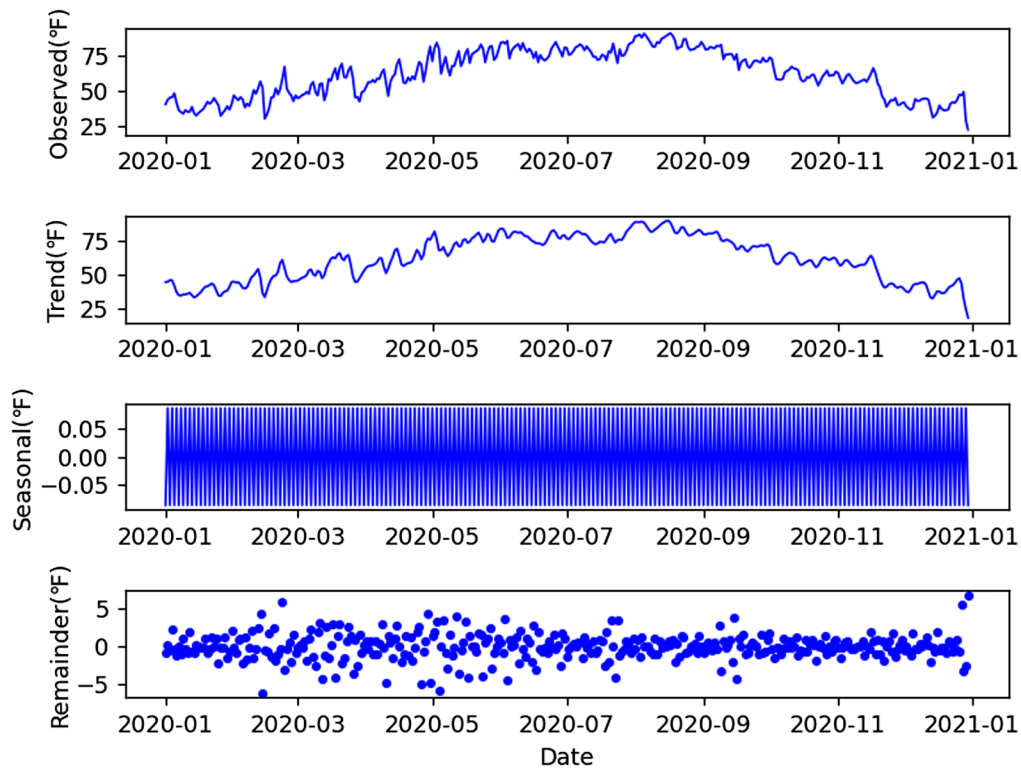
Fig. 4 Decomposition results of temperature data when the frequency is 2. The first line is original temperature series, other lines are trend component, seasonal component, and residual component.

seasonal frequency.[29] After using the curve connection, it looks like a continuous waveform, but the actual meaning is still regularly changing data. The seasonal component is the regular data of the decomposition period, which mainly analyzes the regular change of the irregularly changing data. Then set time_step, which denotes how much data are used to predict the next value. Use a sliding window of length time_step to traverse the data to generate the training set, and select the value of time_step + 1 as the corresponding predicted value. Generate the input and output pairs of original data and the trend, seasonal, and remainder components for the proposed model.

5.2 Evaluation Metrics

To quantitatively evaluate the performance of the proposed model in temperature prediction, we used root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination R^2 for evaluation and comparison. The equation of each evaluation index is as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=0}^{n} (y_i' - y_i)^2}{n}}, \tag{11}$$

$$\text{MAE} = \frac{\sum_{i=0}^{n} |y_i' - y_i|}{n}, \tag{12}$$

$$R^2 = 1 - \frac{\sum_{i=0}^{n} (y_i' - y_i)^2}{\sum_{i=0}^{n} (y_i - \overline{y})^2}, \tag{13}$$

where y is the true value, y' is the predicted value, $\overline{y}$ is the mean value, and n is the number of data in the testing set.

Table 1 Comparison of RMSE (°F) of different decomposition frequencies.

Period	Beijing	Shanghai	Guangzhou	Wulumuqi
2	**0.110**	**0.104**	**0.092**	**0.138**
3	1.555	1.502	1.206	1.671
4	8.810	5.549	5.483	6.908
5	7.893	7.952	7.633	9.408

Note: Bold values represent that the proposed model has the smallest RMSE on testing set when frequency is 2.

Table 2 Comparison of MAE (°F) of different decomposition frequencies.

Period	Beijing	Shanghai	Guangzhou	Wulumuqi
2	**0.092**	**0.085**	**0.076**	**0.110**
3	1.189	1.099	0.878	1.240
4	6.779	4.280	3.990	5.366
5	6.257	6.203	5.797	7.586

Note: Bold values represent that the proposed model has the smallest MAE on testing set when frequency is 2.

5.3 Parameters Setting

The proposed model consists of four parallel two-layer Bi-LSTM and fully connection layer. The units of two-layer Bi-LSTM are set to 50. The activation function is *tanh*. To avoid overfitting, the dropout rate is set to 0.2. And we use Adam optimizer with a learning rate of 0.0001. Our model and compared models are trained on GeForce RTX 2070 Super using python = 3.6.1 and keras = 2.2.4.

To determine the optimal parameters of proposed model, different parameters were used to train the model. We evaluate the performance of different parameters based on RMSE and MAE to obtain best model.

For the frequency of STL decomposition, we decomposed temperature data using different frequencies to test the forecasting result. Tables 1 and 2 show RMSE and MAE on part of testing set with different decomposition frequencies. When decomposition frequency is 2, the proposed model has the best prediction result. We think that a small decomposition frequency strengthens the learning of features at adjacent time points, and as the decomposition frequency increases, the effect weakens.

At the same time, we also test the influence of different time_step and batch-size on the prediction results. The time_step is set to 30 and batch-size is set to 32.

The decomposition frequency and the time_step are not parameters of the network, but two parameters affect the input of the network. The final result also shows that the frequency of 2 has the best training effect, and using a small time_step may not get a good prediction result. Using a large time_step will get an outstanding prediction result, but the training time will increase, so choosing a moderate time_step is required. Regarding the parameters of the network, after the above two parameters are determined, the optimal training results are obtained by adjusting the learning rate and batch_size.

5.4 Model Evaluation

The training set and testing set are separated from temperature data of 34 cities. Additional testing data of four regions are added to test the robustness of the proposed model. We tested the model using the testing set and the temperature data of four regions and calculated the average RMSE and MAE to compare with other prediction models.

The proposed model can be divided into two parts according to the input: the input part using STL decomposition components (part 1) and the original temperature data input part (part 2).

Table 3 Comparison of RMSE (°F) on models of different input.

Model	Testing set	North region	South region	Northwest region	Tibetan region
Our	**0.108**	**0.127**	**0.069**	**0.139**	**0.103**
Part 1	0.155	0.150	0.152	0.181	0.149
Part 2	4.230	4.8	3.475	4.967	3.409

Note: Bold values represent that the proposed model has the smallest RMSE on testing set than two part models.

Table 4 Comparison of MAE (°F) on models of different input.

Model	Testing set	North region	South region	Northwest region	Tibetan region
Our	**0.089**	**0.103**	**0.080**	**0.115**	**0.085**
Part 1	0.128	0.112	**0.080**	0.136	0.117
Part 2	3.208	3.627	2.622	3.387	2.613

Note: Bold values represent that the proposed model has the smallest MAE on testing set than two part models.

To prove that the proposed model combines the advantages of the two parts to improve the prediction effect, we used the same parameters to train the model using the STL decomposition components input and the model using the original data input.

It is shown in Tables 3 and 4 that the proposed model achieves good prediction results for the temperature in different regions, and the proposed model is better than the other two comparative models in the prediction results. Compared with the model that uses the original temperature data, the proposed model greatly improves the prediction accuracy. At the same time, the high accuracy of the model that uses STL decomposition components input also shows that the combination of the time series decomposition method can improve the accuracy of the model.

The proposed model was compared with other time series prediction models composed of time series decomposition methods combined with LSTM neural network. At the same time, the proposed model was compared with the linear regression model (LR) and the nonlinear regression model SVM, in which SVM uses temperature data and temperature data combined with STL decomposition data as input. Finally, we compare our model with the Bi-LSTM model used by Liang et al. for atmospheric temperature prediction.[31]

The comparison results are shown in Tables 5–7. The average of RMSE and MAE of the proposed model, STL-LSTM, and EMD-LSTM are 0.11 and 0.09, 0.35 and 0.27, 2.73 and 2.07, respectively. We can find that the RMSE and MAE of the proposed model are significantly lower

Table 5 Comparison of RMSE (°F) on different prediction models.

Model	Testing set	North region	South region	Northwest region	Tibetan region
Our	**0.108**	**0.127**	**0.096**	**0.139**	**0.103**
STL-LSTM	0.338	0.359	0.323	0.403	0.306
EMD-LSTM	2.357	3.492	2.245	2.918	2.626
LR	4.343	4.82	3.483	5.141	3.496
SVM	4.276	4.882	3.424	5.095	3.541
STL-SVM	0.323	0.757	0.114	0.548	0.217
Bi-LSTM	4.0537	4.6882	3.4058	4.7769	3.3438

Note: Bold values represent that our model has the smallest RMSE among all comparison models.

Table 6 Comparison of MAE (°F) on different prediction models.

Model	Testing set	North region	South region	Northwest region	Tibetan region
Our	**0.089**	**0.103**	0.08	**0.115**	**0.085**
STL-LSTM	0.262	0.279	0.26	0.31	0.252
EMD-LSTM	1.804	2.576	1.569	2.333	2.086
LR	3.222	3.650	2.572	3.872	2.664
SVM	3.111	3.608	2.475	3.782	2.667
STL-SVM	0.116	0.216	**0.071**	0.25	0.108
Bi-LSTM	3.0815	3.5447	2.5454	3.6915	2.5445

Note: Bold values represent that our model has the smallest MAE among all comparison models.

Table 7 Comparison of R^2 on different prediction models.

Model	Testing set	North region	South region	Northwest region	Tibetan region
Our	**0.9999**	**0.9999**	**0.9999**	**0.9999**	**0.9999**
STL-LSTM	0.9996	0.9994	0.9991	0.9997	0.9996
EMD-LSTM	0.9857	0.9577	0.9656	0.8995	0.9459
LR	0.9551	0.9574	09477	0.9595	0.9668
SVM	0.9564	0.9563	0.9495	0.9602	0.9659
STL-SVM	0.9998	0.9989	**0.9999**	0.9995	0.9998
Bi-LSTM	0.9411	0.9410	0.9133	0.9590	0.9511

Note: Bold values represent that our model has the largest R^2 among all comparison models.
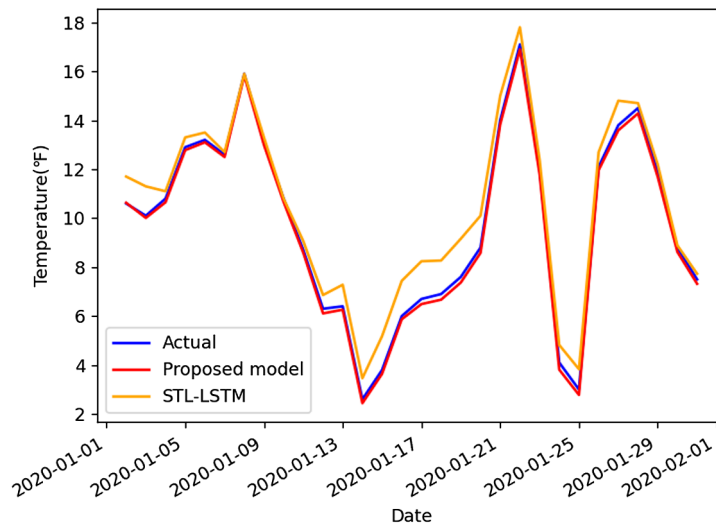

Fig. 5 Visualization of predicted temperature and actual temperature.

than the comparison networks STL-LSTM and EMD-LSTM, indicating that the proposed model improves the accuracy of temperature prediction and can predict temperature more accurately. From the comparison results, the neural network-based methods outperform the regression-based methods in temperature prediction. Table 7 shows that the proposed model has the highest coefficient of determination, proving that the proposed model has outstanding performance and

successfully realize the city temperature prediction of China. Figure 5 shows the visualization of predicted and actual temperature of Jixi using the proposed model, which also indicates that the proposed model can obtain outstanding prediction results.

6 Conclusion

This paper proposed a parallel STL-Bi-LSTM model, which combined STL and Bi-LSTM model to realize the prediction of city daily average temperature. Experiments showed that the prediction model combined with the STL can greatly improve the prediction accuracy. Moreover, the prediction loss of the proposed model is smaller than that of STL-LSTM and EMD-LSTM, which proved that the proposed model is very suitable for temperature prediction. The prediction model proposed in this paper can theoretically be used for temperature prediction in other countries, provided that sufficient temperature data of the country is used for network training, and the optimal network weights are obtained to realize the temperature prediction of the country. In this study, we only predicted temperature using decomposition components and original temperature series, more impacting factors will be considered into the model to improve the prediction accuracy in future studies.

Acknowledgments

Code, Data, and Materials Availability

The temperature data are acquired from Ref. 27.

References

1. C. Zhao et al., "Temperature increase reduces global yields of major crops in four independent estimates," *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9326–9331 (2017).
2. K. M. Barlow et al., "Simulating the impact of extreme heat and frost events on wheat crop production: a review," *Field Crops Res.* **171**, 109–119 (2015).
3. H. Doi, M. Yasuhara, and M. Ushio, "Causal analysis of the temperature impact on deep-sea biodiversity," *Biol. Lett.* **17**, 20200666 (2021).
4. F. Dottori et al., "Increased human and economic losses from river flooding with anthropogenic warming," *Nat. Clim. Change* **8**(9), 781–786 (2018).
5. X. Li et al., "Urban heat island impacts on building energy consumption: a review of approaches and findings," *Energy* **174**, 407–419 (2019).
6. L. Lambrechts et al., "Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*," *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7460–7465 (2011).
7. J. Wang et al., "Impact of temperature and relative humidity on the transmission of COVID-19: a modelling study in China and the United States," *BMJ Open* **11**(2), e043863 (2021).
8. G. Wang, Y. Qiu, and H. Li, "Temperature forecast based on SVM optimized by PSO algorithm," in *Int. Conf. Intell. Comput. and Cognitive Inf.*, IEEE Computer Society, pp. 259–262 (2012).
9. C. N. Babu and B. E. Reddy, "Predictive data mining on average global temperature using variants of ARIMA models," in *IEEE-Int. Conf. Adv. in Eng., Sci. and Manage.*, pp. 256–260 (2012).
10. M. A. Jallal et al., "Air temperature forecasting using artificial neural networks with delayed exogenous input," in *Int. Conf. Wireless Technol., Embedded and Intell. Syst. (WITS)*, pp. 1–6 (2019).

11. C. Li, Y. Zhang, and G. Zhao, "Deep learning with long short-term memory networks for air temperature predictions," in *Int. Conf. Artif. Intell. and Adv. Manuf. (AIAM)*, pp. 243–249 (2019).

12. Y. Qi and C. Guo, "Deep learning-based hourly temperature prediction: a case study of mega-cites in north China," in *ICBDT 2020: 2020 3rd Int. Conf. Big Data Technol.*, pp. 93–96 (2020).

13. Z. Huang et al., "Multi-step temperature prediction model based on surrounding cities and long-term memory neural networks," in *14th Int. Conf. Comput. Sci. and Educ. (ICCSE)*, pp. 518–522 (2019).

14. Z. A. Sadeque and F. M. Bui, "A deep learning approach to predict weather data using cascaded LSTM network," in *IEEE Can. Conf. Electr. and Comput. Eng. (CCECE)*, pp. 1–5 (2020).

15. K. S. Joanna et al., "Neural approach in short-term outdoor temperature prediction for application in HVAC systems," *Energies* **14**(22), 7512 (2021).

16. Y. Y. Wang et al., "Short time air temperature prediction using pattern approximate matching," *Energy Build.* **244**, 111036 (2021).

17. X. Yu, S. X. Shi, and L. Y. Xu. "A spatial-temporal graph attention network approach for air temperature forecasting," *Appl. Soft Comput.* **113**, 107888 (2021).

18. A. Hrachya et al., "Air temperature forecasting using artificial neural network for Ararat valley," *Earth Sci. Inf.* **14**(2), 711–722 (2021).

19. T. Toni et al., "Employing long short-term memory and Facebook prophet model in air temperature forecasting," *Commun. Stat.-Simul. Comput.* 1–24 (2021).

20. B. Zhang, "Foreign exchange rates forecasting with an EMD-LSTM neural networks model," *J. Phys. Conf.* **1053**, 012005 (2018).

21. D. Jin et al., "Forecasting of vegetable prices using STL-LSTM method," in *6th Int. Conf. Syst. and Inf. (ICSAI)*, pp. 866–871 (2019).

22. Z. Wang and Y. Lou, "Hydrological time series forecast model based on wavelet de-noising and ARIMA-LSTM," in *IEEE 3rd Inf. Technol., Netw., Electron. and Autom. Control Conf. (ITNEC)*, pp. 1697–1701 (2019).

23. Q. Duan et al., "Base station traffic prediction based on STL-LSTM networks," in *24th Asia-Pacific Conf. Commun. (APCC)*, pp. 407–412 (2018).

24. Y. Huo et al., "Long-term span traffic prediction model based on STL decomposition and LSTM," in *20th Asia-Pacific Netw. Oper. and Manage. Symp. (APNOMS)*, pp. 1–4 (2019).

25. H. Yin et al., "STL-ATTLSTM: vegetable price forecasting using STL and attention mechanism-based LSTM," *Agriculture* **10**, 612 (2020).

26. D. Chen, J. Zhang, and S. Jiang, "Forecasting the short-term metro ridership with seasonal and trend decomposition using loess and LSTM neural networks," *IEEE Access* **8**, 91181–91187 (2020).

27. National Climatic Data Center, NESDIS, NOAA, U.S. Department of Commerce, "Global summary of the day," https://www.ncei.noaa.gov/maps/alltimes/ (2021).

28. Y. He and M. Wang, "China's geographical regionalization in Chinese secondary school curriculum (1902–2012)," *J. Geogr. Sci.* **23**, 370–383 (2013).

29. R. B. Cleveland et al., "STL: a seasonal-trend decomposition procedure based on loess," *J. Off. Stat.* **6**(1), 3–73 (1990).

30. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.* **9**(8), 1735–1780 (1997).

31. S. Liang et al., "Method of bidirectional LSTM modelling for the atmospheric temperature," *Intell. Autom. Soft Comput.* **29**(3), 701–714 (2021).

Xing Huo is a professor working at the School of Mathematics, Hefei University of Technology. Her research focuses on image processing, remote sensing information processing, and medical data analysis.

Guangpeng Cui is currently working toward his MS degree in mathematics at the School of Mathematics, Hefei University of Technology, Anhui, China. His research interests include remote sensing information processing and medical data analysis.

Lingling Ma received her PhD from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2008. She is currently a professor at the Aerospace Information Research Institute, Chinese Academy of Sciences. Her research interests include the calibration, validation, and quality assurance in remote sensing.

Bohui Tang received his PhD in cartography and geographical information system from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, in 2007. He is currently a professor at the Kunming University of Science and Technology, Kunming. His research mainly includes the retrieval and validation of surface net radiation and surface temperature.

Ronglin Tang received his PhD in cartography and geographical information system from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, in 2011. He is currently a professor at the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. His research interests include the remote sensing retrieval and validation of land surface evapotranspiration and soil moisture.

Kun Shao is an associate professor at the School of software, Hefei University of Technology. His research interests include software modeling and development, software requirements analysis and modeling, graphics and image processing, etc.

Xinhong Wang received his PhD in cartography and geography information system from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (CAS), in 2008. He is an associate professor at the Aerospace Information Research Institute, CAS. His research interests include performance evaluation of remote sensors and quantitative infrared remote sensing.