


Article

Modeling Hourly Soil Temperature Using Deep BiLSTM Neural Network

Cong Li ^{1,2,3}, Yaonan Zhang ^{1,3,*}  and Xupeng Ren ²

¹ Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China; shuimulicong@lut.edu.cn

² College of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China; renxp@lut.edu.cn

³ University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: yaonan@lzb.ac.cn

Received: 11 June 2020; Accepted: 15 July 2020; Published: 17 July 2020



Abstract: Soil temperature (ST) plays a key role in the processes and functions of almost all ecosystems, and is also an essential parameter for various applications such as agricultural production, geothermal development, and their utilization. Although numerous machine learning models have been used in the prediction of ST, and good results have been obtained, most of the current studies have focused on daily or monthly ST predictions, while hourly ST predictions are scarce. This paper presents a novel scheme for forecasting the hourly ST using weather forecast data. The method considers the hourly ST prediction to be the superposition of two parts, namely, the daily average ST prediction and the ST amplitude (the difference between the hourly ST and the daily average ST) prediction. According to the results of correlation analysis, we selected nine meteorological parameters and combined two temporal parameters as the input vectors for predicting the daily average ST. For the task of predicting the ST amplitude, seven meteorological parameters and one temporal parameter were selected as the inputs. Two submodels were constructed using a deep bidirectional long short-term memory network (BiLSTM). For the task of hourly ST prediction at five different soil depths at 30 sites, which are located in 5 common climates in the United States, the results showed the method proposed in this paper performs best at all depths for 30 stations (100% of all) for the root mean square error (RMSE), 27 stations (90% of all) for the mean absolute error (MAE), and 30 stations (100% of all) for the coefficient of determination (R^2), respectively. Moreover, the method adopted in this study displays a stronger ST prediction ability than the traditional methods under all climate types involved in the experiment, the hourly ST produced by it can be used as a driving parameter for high-resolution biogeochemical models, land surface models and hydrological models and can provide ideas for an analysis of other time series data.

Keywords: soil temperature; machine learning; weather forecasting data; BiLSTM; soil depths

1. Introduction

Among the many soil factors, soil temperature (ST) is the most important, affecting the processes and functions of almost all ecosystems [1], such as soil infiltration rates [2], soil respiration [3], soil organic matter accumulation and degradation [4], soil chemical and physical reactions [5], land surface hydrological processes, and land atmosphere interactions [6]. ST is extremely important to the growth of crops [7], affecting their germination, root growth, and the absorption of nutrients [8]. Dang [9] reported that an increase in ST will reduce the growth cycle of crops and significantly increase the yield. The ST is also an essential important parameter in a geothermal exploration [10], ground

heat exchanger applications [11], thermal energy storage [12], ground source heat pump (GSHP) systems [13], and other geothermal development and utilization applications.

The ST measurement system represents one of the most accurate and direct methods of ST acquisition in the world. However, ST measurements are costly and time-consuming [14], and it is not easy to obtain different large-scale depths of ST in real-time.

A machine learning method is an effective way to estimate the ST. This type of method predicts the ST at different depths by establishing a nonlinear relationship between the input and output data. The input parameters are usually easily obtained meteorological parameters involving a simple calculation process and offering a high prediction accuracy. Feng [15] adopted four machine learning algorithms for predicting the half-hourly ST at four different depths. The results show that the selected machine learning method performs well for predicting the half-hourly ST at all depths. Xing [16] designed a daily ST prediction model using a support vector machine (SVM) algorithm, which performs well in the daily ST prediction tasks of 16 observation stations distributed throughout the United States. Artificial neural network (ANN) are frequently used in ST prediction and have achieved good results. Tabari [17] developed models for 1 day ahead ST predictions based on an ANN. Its performance is in terms of the Nash–Sutcliffe co-efficient of efficiency values of >0.94 and a correlation co-efficient of >0.96 . Bilgili [18] compared the capability of linear regression, nonlinear regression, and an ANN at different depths, and found that the accuracy of an ANN is better than that of the other two methods. Wu [19] used an ANN to predict the monthly average ST at a depth of 10 cm. Compared with multiple linear regression, the developed ANN achieves a good performance and robustness. In [20], gene expression programming (GEP), an ANN, and an adaptive neuro-fuzzy inference system (ANFIS) were used to estimate the monthly mean ST. The comparison shows that the ANN and ANFIS achieve a better performance than the GEP model. Kisi [21] estimated the monthly ST using a multi-layer perceptron (MLP), radial basis neural network (RBNN), generated regression neural network (GRNN), and multiple linear regression (MLR) at different depths. The results show that the RBNN algorithm achieves the highest prediction accuracy at depths of 5 and 10 cm, and an MLR and a GRNN are better at ST prediction at depths of 50 and 100 cm.

Although many machine learning models have been used in the prediction of ST and good results have been obtained, there are still some inherent problems. First, in a large number of studies, the input parameters of the model were selected through a correlation analysis, and the ST was considered as the output of the model. However, the correlation between meteorological parameters and the ST is not the same owing to the difference in climate at different sites. As a result, a model that considers a single correlation between meteorological parameters and the ST across different sites will suffer from obvious limitations. Second, the input of hourly soil temperature can help the model to obtain high-resolution simulation results in related studies such as using terrestrial biogeochemical models to simulate the dynamic changes of N_2O , CO , N_2 , and CH_4 in soil [22–24], simulating soil respiration by land surface models [25], and quantifying hydrological and biological processes in hydrological models [26,27]. However, most of the current studies have focused on daily or monthly ST predictions, and hourly ST predictions remain scarce. Third, as a branch of machine learning, deep learning technology performs well in nonlinear data processing, and therefore, has achieved success in predicting the solar radiation [28–30], wind speed [31–33], and soil moisture [34], whereas studies on ST prediction are relatively few in number.

In this research, we treat the hourly ST prediction as the sum of the daily average ST prediction and ST amplitude prediction, and verify them at 30 sites located under different climates. The prediction model was built using a bidirectional long short-term memory network (BiLSTM), which is a deep learning technique for sequence learning, and although less commonly applied to ST predictions, is inherently suitable for this domain.

The remainder of this article is organized as follows. Section 2 introduces the data applied in this research. The algorithm for predicting the hourly soil temperature is introduced in Section 3.

The experimental results and their analysis are presented in Section 4, and finally, Section 5 provides some concluding remarks regarding this research.

2. Data

The measured ST were obtained from the National Water and Climate Center of the US Department of Agriculture (<https://www.wcc.nrcs.usda.gov/scan/>). The agency has installed more than 200 sites throughout the country to collect data on meteorology, soil, and solar radiation. We selected the hourly data for 30 measurement sites, which are located in five common climate zones. Figure 1 shows the identification numbers and geographic locations of the 30 measurement sites used in the study.

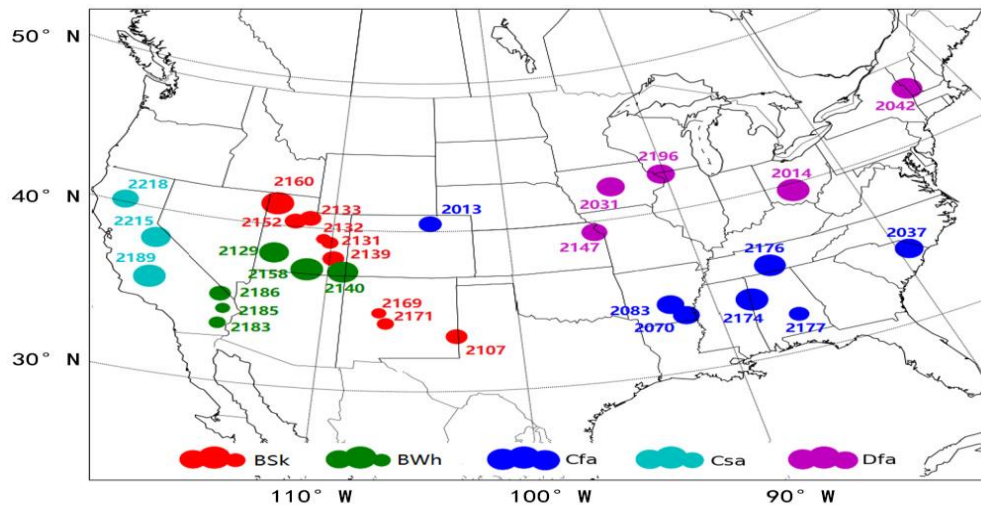


Figure 1. Location and climate types of 30 measurement stations used in the present study.

The data-collection times and elevations of these stations are given in Table 1, and are classified based on the Koppen climate classification method. This method involves a representation based on three letters indicating the climatic conditions. The first letter indicates the major climate type, which is one of the following: (A) equatorial, (B) arid, (C) warm temperature, (D) snowfall, and (E) polar climates. The second and third letters indicate the precipitation conditions and temperature, respectively. For example, BSk indicates that the region is located in an arid, cold temperature with a stepped climate. For a detailed description of the climate classification, please refer to [35].

Table 1. Measurement stations used in the study.

Climate Type	Site id	Elevation (m)	Period	Site id	Elevation (m)	Period
BSk	2107	1221	2011~2019	2152	1322	2010~2019
	2131	1251	2011~2019	2160	1779	2010~2019
	2132	1718	2011~2019	2169	1477	2009~2019
	2133	1558	2013~2019	2171	1595	2010~2019
	2139	2131	2011~2019			
BWh	2129	1526	2010~2019	2183	169	2012~2019
	2140	1620	2011~2019	2185	806	2012~2019
	2158	1628	2010~2019	2186	1110	2012~2019
	2013	235	2013~2019	2083	71	2011~2019
Cfa	2037	37	2011~2019	2174	49	2010~2018
	2070	50	2011~2017	2177	82	2015~2017
	2076	215	2011~2019			
Csa	2189	822	2012~2019	2218	1129	2015~2019
	2215	2385	2015~2019			
Dfa	2014	323	2010~2019	2147	336	2010~2015
	2031	327	2010~2019	2196	328	2013~2019
	2042	742	2010~2014			

There will be missing values in the data set, which are caused by sensor failure, memory damage, and other reasons. In order to avoid destroying the continuity of the time series data, we use MissForest algorithm [36] to interpolate missing values. This method is a nonparametric interpolation method based on random forest, which is suitable for both discrete variables and continuous variables, and can be well applied to nonlinear data. After processing missing values, the valid data of 30 stations totaled 1,577,714 groups, including the hourly weather conditions such as the temperature, dew point, relative humidity, wind speed, solar radiation, and soil temperature at 5, 10, 20, 50, and 100 cm, respectively. Based on the existing data, we further calculated the month, day of the month, hour of the day, and ST amplitude as supplementary data. The ST amplitude is a parameter defined in this paper, which is used to describe the fluctuation of hourly ST relative to the average ST of the day, it is obtained by subtracting daily average ST from each hourly ST.

We first determined the inputs of the model with a Pearson correlation coefficient [16], which can reflect the degree of correlation between the two variables x and y , based on Equation (1). It is generally believed that the absolute value of r is 0.0–0.2 indicates an extremely weak or no correlation between x and y , 0.2–0.4 indicates a weak correlation, 0.4–0.6 indicates a moderate correlation, 0.6–0.8 indicates a strong correlation, and 0.8–1.0 indicates an extremely strong correlation.

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}} \quad (1)$$

Figure 2 shows the correlation coefficient between the ST amplitude and meteorological parameters for the different depths, in which the correlation coefficient greater than 0 indicates a positive correlation between x and y , and less than 0 indicates a negative correlation. The parameters appearing between -0.2 and 0.2 are month, day of the month, air temperature observed, wind direction average, wind speed maximum, relative humidity, and dew point temperature, indicating that these parameters have no correlation or have extremely weak correlation with ST amplitude for most depths.

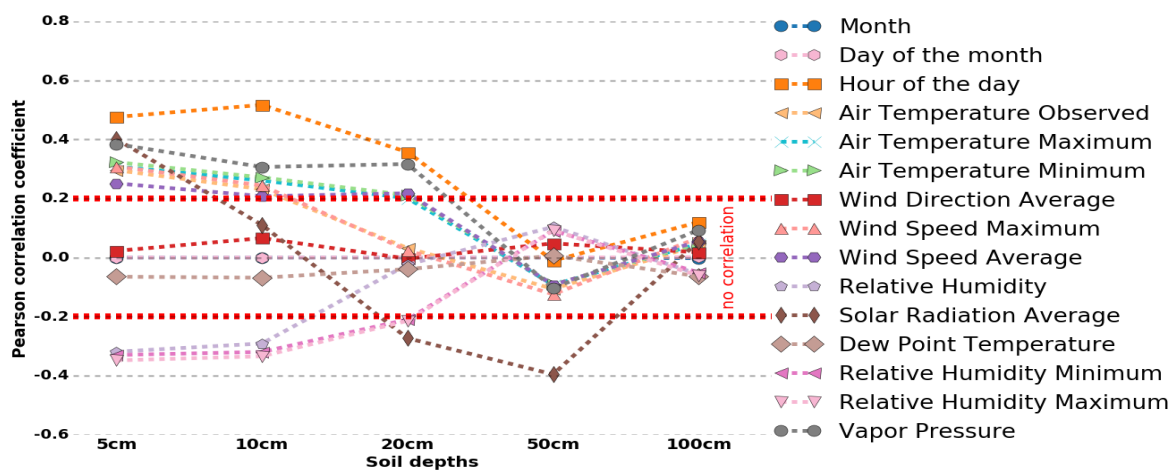


Figure 2. Correlation coefficients of (x) daily weather parameters and (y) soil temperature (ST) amplitude for different depths.

Figure 3 shows the correlation coefficients between the daily average ST and the meteorological parameters for the different depths. It can be seen that, except for the wind direction average, wind speed maximum, and relative humidity, the other parameters are moderately or strongly correlated with the daily average ST for most depths.

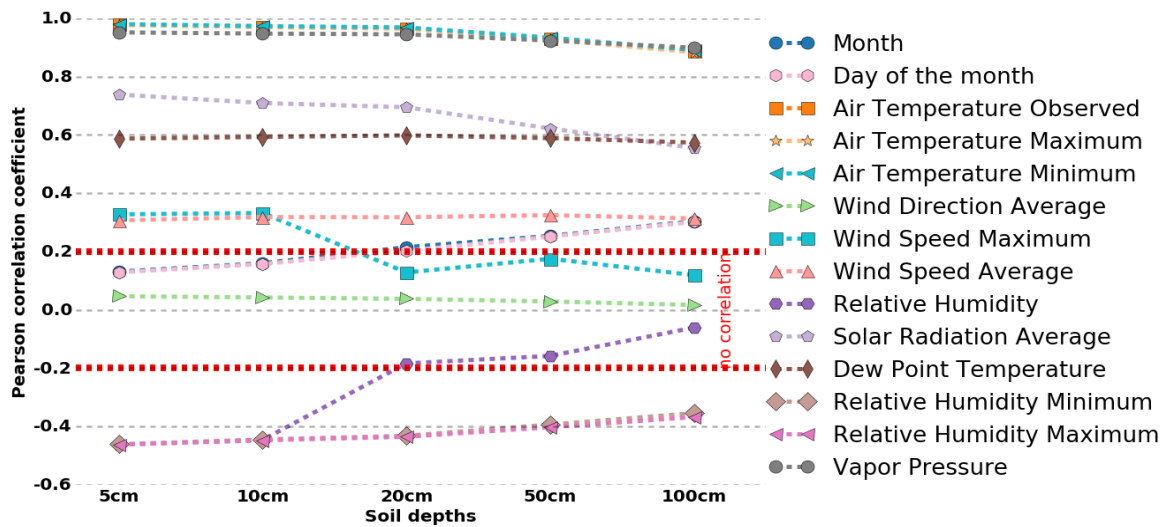


Figure 3. Correlation coefficients of daily average (x) weather parameters and (y) ST for different depths.

Through the analysis of Figures 2 and 3, we delete the parameter in which the absolute value of correlation coefficient is less than 0.2, and choose eight parameters as the input vector for predicting ST amplitude. For the forecasting task of daily average ST, eleven parameters are selected as input. Details of the input parameters are shown in Table 2, in which meteorological parameters are collected by sensors at intervals of one hour.

Table 2. Input variables for the forecast models.

ST Amplitude Forecast Task	Daily Average ST Forecast Task
Hour of the day	Month
Air Temperature Maximum	Day of the month
Air Temperature Minimum	Air Temperature Observed
Wind Speed Average	Air Temperature Maximum
Solar Radiation Average	Air Temperature Minimum
Relative Humidity Minimum	Wind Speed Average
Relative Humidity Maximum	Solar Radiation Average
Vapor Pressure	Dew Point Temperature
	Relative Humidity Minimum
	Relative Humidity Maximum
	Vapor Pressure

3. Methodology

3.1. BiLSTM Networks

The structure of BiLSTM is shown in Figure 4a, and the structure of long short-term memory network (LSTM) is shown in Figure 4b. In the LSTM structure, the output result at time t is only related to time $t - 1$, whereas in the BiLSTM structure, the output result at time t is related to both $t - 1$ and $t + 1$, and this structure can effectively use all information of the past and future for training [37]. In [38] and [39], BiLSTM was used for time-series data processing, and it was found that BiLSTM is faster and more accurate than LSTM and standard recurrent neural networks.

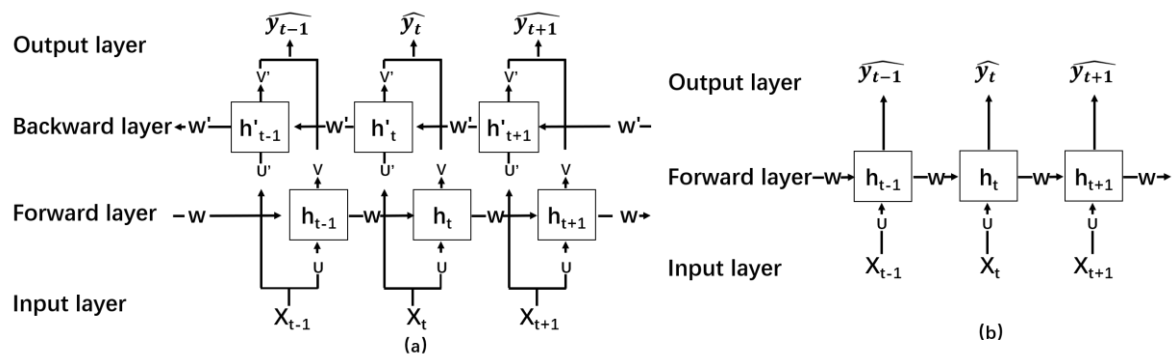


Figure 4. Structures of bidirectional long short-term memory network (BiLSTM) (a) and LSTM (b).

Equations (2)–(5) describe the calculation process of BiLSTM, where h_t is the output result of the forward layer at time t , h'_t is the output result of the backward layer at time t , \hat{y}_t is the network output that combines h_t and h'_t , f and g are nonlinear activation functions, and the goal of network training is to minimize the mean square error (MSE) function s based on the target value y_t .

$$h_t = f(UX_t + Wh_{t-1}) \tag{2}$$

$$h'_t = f(U'X_t + W'h_{t+1}) \tag{3}$$

$$\hat{y}_t = g(V'h'_t + Vh_t) \tag{4}$$

$$s = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n} \tag{5}$$

3.2. Integrated BiLSTM Model

We proposed a novel method to predict the hourly ST, which includes two sub-models: The daily average ST prediction model and the ST amplitude prediction model. The hourly ST estimation is the summation of the daily average ST prediction results and ST amplitude prediction results, because in the Section 2, we have defined that ST amplitude was obtained by subtracting the daily average ST from the hourly ST. A flow chart of the model is shown in Figure 5, and the input parameters of the two submodels are given in Table 2.

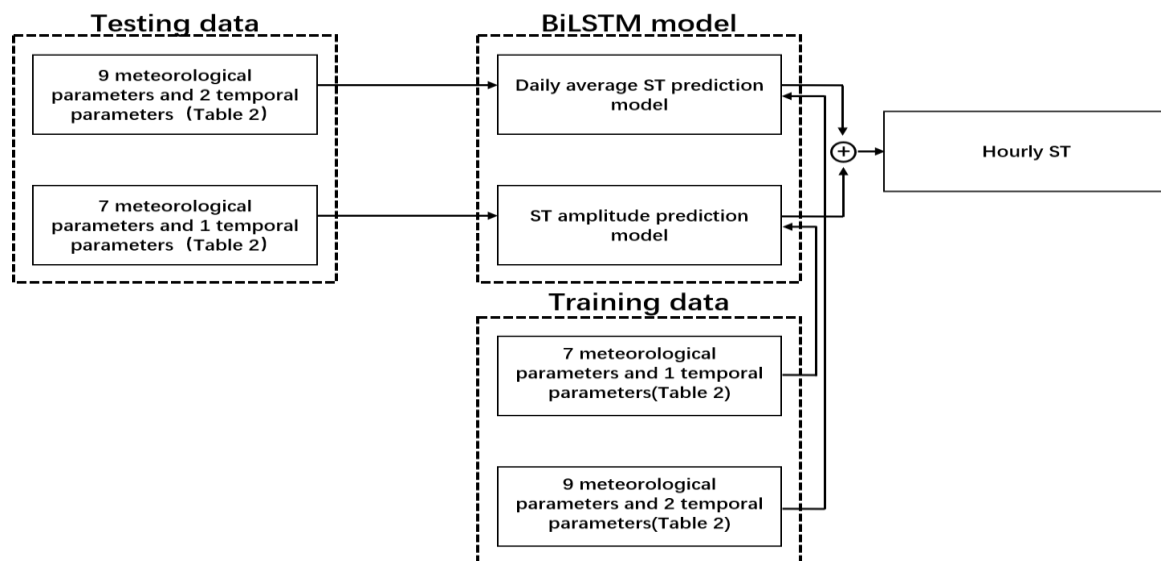


Figure 5. Flowchart of integrated BiLSTM ST prediction model.

The topology of each sub-model is set to six layers, namely, an input layer with 12 features and 1 timestep; 4 hidden layers, in which the numbers of neurons of each layer are 14, 14, 14, and 6, respectively; and an output layer with 1 neuron, the loss function of which is the MSE. In addition, an adaptive moment estimation algorithm (Adam) [40] is selected as the optimizer.

3.3. Benchmark Models

We chose six benchmark algorithms to prove the relative advantages of the proposed method, namely, three deep learning methods, i.e., LSTM, BiLSTM and deep neural network (DNN), and three traditional machine learning methods, random forest (RF), support vector regression (SVR), and linear regression (LR). A brief introduction and the parameter selection of each algorithm are as follows.

The LSTM network is a special type of deep neural network proposed by Hochreiter and Schmidhuber [41] in 1997. Because LSTM can solve tasks with a long time-lag well, it is widely used in time-series data processing [42–46]. The topology of the LSTM is consistent with the subnetwork described in Section 3.2.

The topology of BiLSTM, introduced in Section 3.1, is consistent with the subnetwork described in Section 3.2.

DNN is a neural network with multiple hidden layers, which has strong nonlinear fitting ability and has been successfully applied in prediction and classification tasks [47–50]. In order to ensure fair comparison, we set the number of layers of DNN and the number of neurons in each layer to be consistent with the subnetwork described in Section 3.2, and choose rectified linear units (RELU) [51] as the activation function.

RF is an excellent algorithm used in machine learning. It has the advantages of a fast learning speed, insensitivity to noise, less overfitting, and can achieve good results without numerous parametric adjustments. Through a grid search of the search variable space, the number of decision trees was set to 10, and the tree depth was set to 20.

An SVR is one of the most important algorithms in machine learning. It can deal with high-dimensional, heterogeneous, and scarcely labeled datasets extremely efficiently, and can also successfully adapt to specific applications [52]. In our experiment, the kernel function and penalty factor are set to a radial basis function (RBF) and 1, respectively, which are determined through a grid search of the variable search space.

LR is a regression method widely used in practical applications and does not require a setting of the parameters.

The RF, SVR, and LR were built using scikit-learn, whereas the LSTM, BiLSTM, and DNN networks, based on Theano, were developed using Keras.

In the experiment, the data from each site were divided into two parts, with the first 80% of the data serving as the training set and the second 20% as the test set. When executing the ST amplitude prediction model, all input data need to be normalized to $[-1, 1]$, and, when running the daily average ST prediction model, all input data need to be normalized to $[0, 1]$. The month, hour of the day, air temperature maximum, air temperature minimum, solar radiation, relative humidity minimum, relative humidity maximum, and vapor pressure are used as the inputs of the benchmark algorithm, and the ST temperature of different depths is the output of the model. When executing the benchmark algorithm, the data need to be normalized to $[0, 1]$. Because the performances of the integrated BiLSTM, BiLSTM, and LSTM are significantly affected by the initialization weight matrix, we ran these algorithms ten times on the same data and computed the mean of the results to ensure a fair comparison with traditional machine learning methods.

4. Results

In this paper, a novel integrated BiLSTM model is proposed that combines the daily average ST prediction and the ST amplitude prediction to obtain the hourly ST. Compared with the traditional

machine learning model, the integrated BiLSTM exhibits obvious advantages in predicting the hourly ST for various climates.

Our results are presented in four aspects. First, the proposed algorithm was used to predict the ST under different observations, and the results were compared with those of other traditional algorithms. Second, the ST prediction performances of each algorithm under different soil depths were compared. Third, the ST prediction performances of each algorithm under different climate types were compared. Finally, the algorithm proposed in this paper was compared with the algorithms in other literatures.

4.1. Model Comparisons

The prediction performance of different models at each observation station are shown in Figures 6–8. As presented in Figure 6, the best performance is obtained by the integrated BiLSTM model, the root mean squared error (RMSE) of which at all 30 sites is lower than that of the benchmark algorithm, whereas the worst performance is obtained by the LR algorithm, the RMSE of which is the highest at all sites. According to the statistics, based on the RMSE index, the integrated BiLSTM is 4.8–18.7%, 8.8–23.6%, 20.9–25.7%, 12.8%–25.1%, 17.0–45.0%, and 42.0–65.1% more accurate than LSTM, BiLSTM, DNN, RF, SVR, and LR, respectively.

The mean absolute error (MAE) of every model for each observation station is given in Figure 7. As can be seen, integrated BiLSTM performs better than the benchmark models at 27 observation sites (90% of the total number of stations), and based on the MAE index is 2.6–18.7%, 5.7–21.5%, 20.3–25.7%, 9.8–22.1%, 17.5–44.1%, and 42.2–65.1% more accurate than LSTM, BiLSTM, DNN, RF, SVR, and LR, respectively. At station 2215, the BiLSTM algorithm obtains the best MAE value of 1.06 °C, whereas the MAE value of the integrated BiLSTM algorithm is 1.07 °C, which is only 0.01 °C higher than that of the BiLSTM. By contrast, LSTM obtained the lowest MAE of 1.14 °C and 1.12 °C at the two stations 2031 and 2147, which are 0.06 °C and 0.02 °C lower than that of the integrated BiLSTM, respectively. In other words, the difference between the integrated BiLSTM and the best algorithm is not obvious at these three stations.

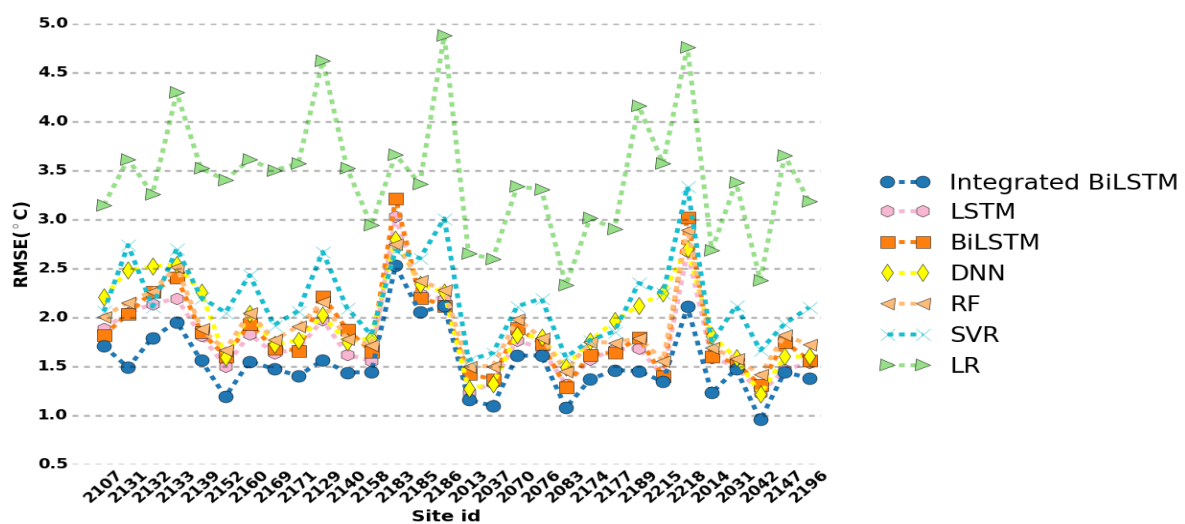


Figure 6. Performance comparison of forecasting models for 30 sites in terms of root mean squared error (RMSE).

Figure 8 shows the R^2 generated by every model at each observation, and as the figure indicates, the integrated BiLSTM was the best model with the highest R^2 value at each observation station. According to statistics, based on the R^2 index, the integrated BiLSTM is 1.3–4.8%, 2.4–8.0%, 5.1–8.3%, 3.2–9.3%, 4.4–45.0%, and 23–57.4% more accurate than LSTM, BiLSTM, DNN, RF, SVR, and LR, respectively. The performance of the LSTM model is slightly better than that of the BiLSTM, and the performance of the LR model is still not ideal. We also noted that the R^2 value of the LR is less than

zero at the two sites labeled 2218 and 2147, which indicates that LR model is unsuitable for processing data with a nonlinear correlation.

Although the RF model is not as good as the integrated BiLSTM, LSTM, and BiLSTM with respect to the three indicators, RMSE, MAE, and R^2 , it can be clearly seen from Figures 6–8 that favorable agreements exist between the results of the RF model and these three deep learning models, and the prediction results at multiple sites are better than DNN, which confirms the potential of using the RF model for an estimation of the hourly ST.

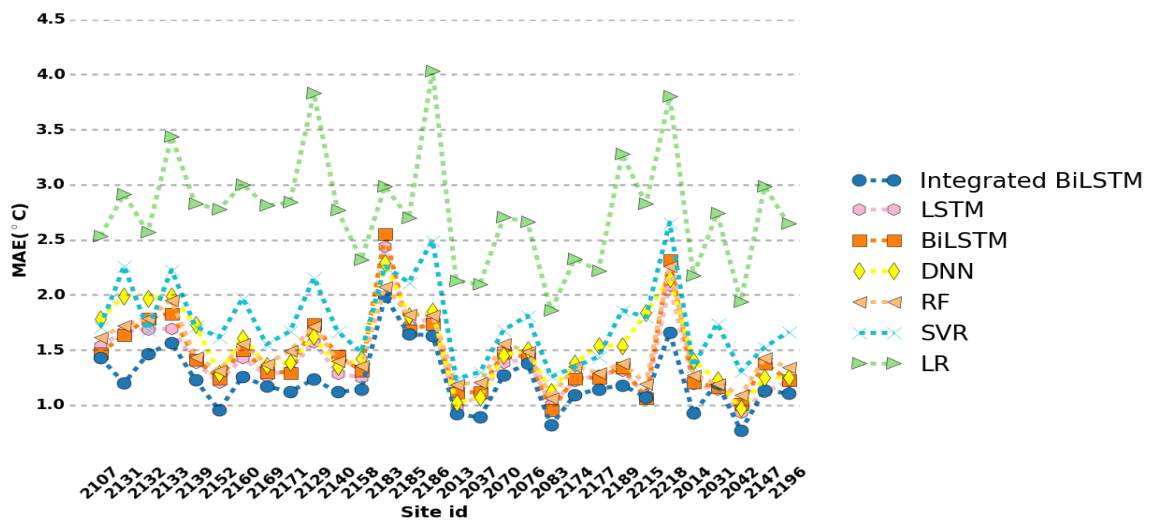


Figure 7. Performance comparison of forecasting models for 30 sites in terms of mean absolute error (MAE).

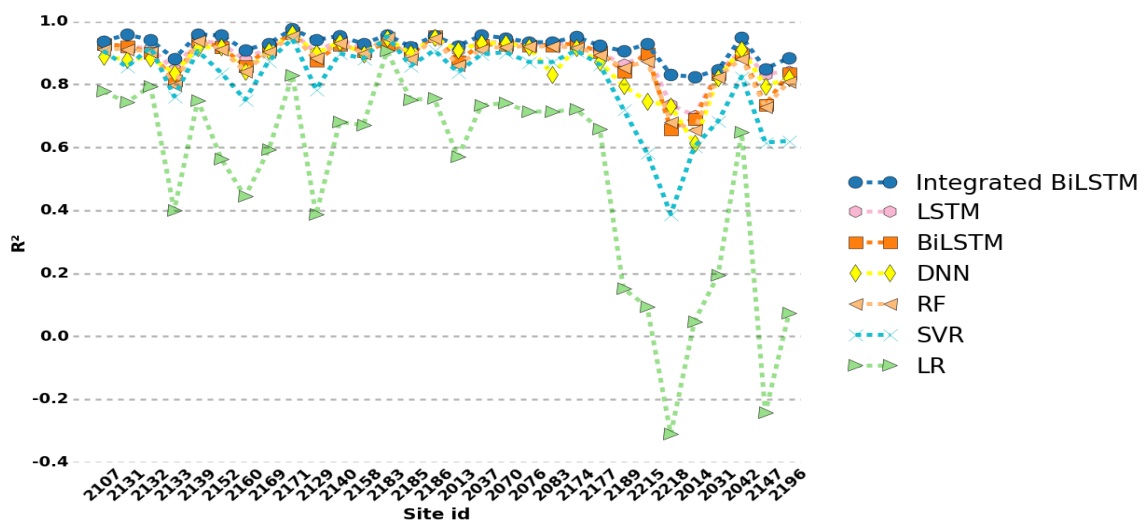


Figure 8. Performance comparison of forecasting models for 30 sites in terms of R^2 .

The best and worst statistical results for each model, and the identification number of the observation sites that produced these results, are given in Tables 3 and 4, respectively. The RMSE, MAE, and determination coefficient (R^2) are used as the evaluation criteria. From Tables 3 and 4, we can clearly see that the performance of the integrated BiLSTM model developed in this study showed the best results. For the 30 observation stations involved in the experiment, the RMSE, MAE, and R^2 values obtained for the integrated BiLSTM are within the range of 0.95–2.53 °C, 0.76–1.99 °C, and 0.823–0.976, respectively.

Among the deep learning technology, DNN algorithm is not as good as integrated BiLSTM, LSTM and BiLSTM for predicting hourly ST. As mentioned above, the results of DNN are sometimes even worse than the RF, as can be seen from Table 4, the worst R^2 obtained by DNN is 0.613, while the worst R^2 of RF is 0.656.

Table 3. Best statistical results of each model.

Method	RMSE		MAE		R^2	
	Best	Site id	Best	Site id	Best	Site id
Integrated BiLSTM	0.95 °C	2042	0.76 °C	2042	0.976	2171
LSTM	1.22 °C	2042	0.93 °C	2042	0.964	2171
BiLSTM	1.29 °C	2083	0.95 °C	2083	0.965	2171
DNN	1.22 °C	2042	0.97 °C	2042	0.961	2171
RF	1.42 °C	2042	1.07 °C	2083	0.954	2171
SVR	1.56 °C	2013	1.24 °C	2013	0.953	2183
LR	2.33 °C	2083	1.86 °C	2083	0.906	2183

Table 4. Worst statistical results of each model.

Method	RMSE		MAE		R^2	
	Worst	Site id	Worst	Site id	Worst	Site id
Integrated BiLSTM	2.53 °C	2183	1.99 °C	2183	0.823	2014
LSTM	3.03 °C	2183	2.44 °C	2183	0.701	2014
BiLSTM	3.22 °C	2183	2.55 °C	2183	0.658	2218
DNN	3.21 °C	2028	2.80 °C	2028	0.613	2014
RF	2.90 °C	2218	2.26 °C	2218	0.656	2014
SVR	3.33 °C	2218	2.66 °C	2218	0.388	2218
LR	4.87 °C	2186	4.03 °C	2186	−0.311	2218

The statistic performance of each model for an estimation of the hourly ST is shown in Table 5, which are the average performance of each model after 10 runs. According to Table 5, the maximum R^2 , and the minimum RMSE and MAE (the best results), with values of 0.923, 1.53 °C, and 1.22 °C, respectively, were obtained using the integrated BiLSTM. By contrast, the minimum R^2 , and the maximum RMSE and MAE values, were found to be 0.518, 3.43 °C, and 2.76 °C when using the LR. The integrated BiLSTM, BiLSTM, and LSTM perform better than the DNN, RF, SVR, and LR. DNN is inferior to LSTM-based method in processing time series data. Among the deep learning methods, the integrated BiLSTM method achieves the best prediction results for the hourly ST, whereas within the range of traditional machine learning, the random forest model achieves the best performance.

Table 5. Statistic performance of each model for estimation of hourly soil temperature.

Method	RMSE(°C)	MAE(°C)	R^2
Integrated BiLSTM	1.53	1.22	0.923
LSTM	1.76	1.39	0.896
BiLSTM	1.85	1.44	0.882
DNN	1.99	1.58	0.865
RF	1.92	1.49	0.874
SVR	2.18	1.75	0.808
LR	3.43	2.76	0.518

4.2. Model Performance at Different Depths

Figure 9 shows the performance of each model in estimating the hourly ST at different soil depths. With respect to RMSE, MAE, and R^2 , the integrated BiLSTM models generally show the best performance at all soil depths, and the LSTM model is the second-best prediction algorithm.

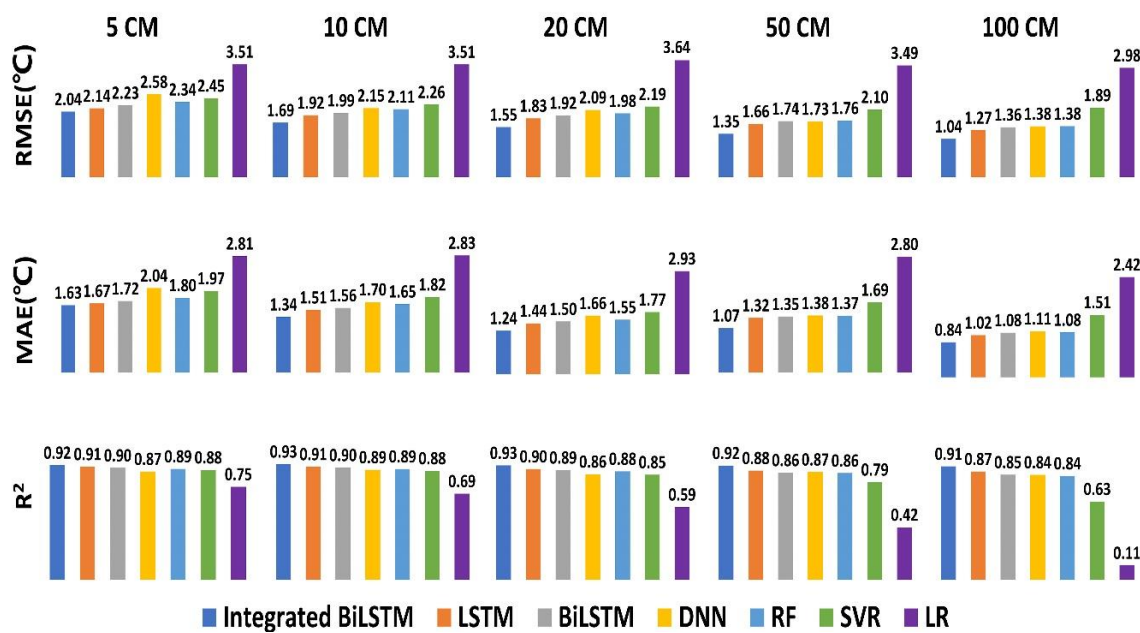


Figure 9. Performance of each model for prediction of hourly ST at different soil depths.

Except for RF, the accuracy (based on the RMSE and MAE) of the other models increases with an increase in depth. Taking the integrated BiLSTM model as an example, the highest RMSE and MAE values (the worst results) are obtained at 5 cm, whereas the lowest values (the best results) are obtained at 100 cm. With respect to R^2 , the accuracy of the integrated BiLSTM model increases from 5 to 20 cm, and decreases from 50 to 100 cm, whereas the R^2 values of the other models decrease with an increase in depth. This result differs from the conclusions in [21] and [53], the experimental results of which show that the accuracy of the machine learning algorithm will gradually decrease at a soil depth of 10 to 100 cm. This may be related to the different models used and the different climatic conditions.

We noticed that the lowest RMSE value obtained by the integrated BiLSTM is 2.04 °C at 5 cm, which is 0.1 °C lower than that obtained by the second-best model. The difference gradually increases to 0.31 °C at 50 cm, and then decreases to 0.23 °C at 100 cm. In other words, from 5 to 50 cm, the difference between the integrated BiLSTM model and the second-best model widens, indicating that the integrated BiLSTM has potential for deeper ST prediction tasks.

4.3. Model Performance at Different Climates

The statistic performance of each algorithm for predicting the hourly ST under different climates is shown in Tables 6–8. As can be seen from these tables, the best performances were generally obtained using the integrated BiLSTM. This produced the best results for four climates (80% of the total number) for RMSE, five climates (100% of the total number) for MAE, and five climates (100% of the total number) for R^2 , respectively. The LR was the worst model, producing the minimum R^2 , and the maximum RMSE and MAE, under all climate types involved in the experiment. With respect to the RMSE, the BiLSTM performed slightly better than the integrated BiLSTM under the BWh climate.

For the five models, namely, integrated BiLSTM, LSTM, BiLSTM, DNN, and RF, the lowest RMSE and lowest MAE were obtained under the Dfa climate type, whereas the highest RMSE and MAE were generated under the BWh or Csa climate, which does not mean that these models perform better under the Dfa climate. Based on the statistics, the average soil temperature measured in the Dfa climate is 12.21 °C, whereas the average soil temperatures of Bwh and Csa are 21.68 °C and 21.72 °C, respectively. If the relative error (the ratio of the error to the true value) is calculated, we can see that the relative error of each algorithm is the highest in the Dfa climate. In other words, the performance of each model

under a Dfa climate (snow areas) is not as good as that in warm or dry areas (BSk, BWh, Cfa, and Csa). It is also clear from Table 8 that the R^2 of these five models is the lowest under the Dfa climate type.

Table 6. Average RMSE of each model for different climate types ($^{\circ}\text{C}$).

Method	BSk	BWh	Cfa	Csa	Dfa
Integrated BiLSTM	1.56	1.82	1.34	1.63	1.25
LSTM	1.86	2.09	1.51	1.92	1.44
BiLSTM	1.92	1.75	1.57	2.07	1.54
DNN	2.13	2.16	1.60	2.35	1.56
RF	2.03	2.17	1.67	2.08	1.64
SVR	2.25	2.46	1.82	2.64	1.90
LR	3.56	3.81	2.88	4.16	3.04

Table 7. Average MAE of each model for different climate types ($^{\circ}\text{C}$).

Method	BSk	BWh	Cfa	Csa	Dfa
integrated BiLSTM	1.27	1.44	1.07	1.30	1.00
LSTM	1.47	1.67	1.20	1.50	1.11
BiLSTM	1.50	1.74	1.22	1.57	1.19
DNN	1.68	1.72	1.26	1.84	1.22
RF	1.58	1.69	1.30	1.61	1.26
SVR	1.82	2.02	1.44	2.11	1.51
LR	2.86	3.09	2.29	3.30	2.48

Table 8. Average determination coefficient (R^2) of each model for different climate types.

Method	BSk	BWh	Cfa	Csa	Dfa
Integrated BiLSTM	0.94	0.94	0.94	0.89	0.87
LSTM	0.91	0.92	0.92	0.83	0.83
BiLSTM	0.91	0.91	0.91	0.81	0.80
DNN	0.89	0.92	0.90	0.78	0.77
RF	0.89	0.92	0.89	0.80	0.78
SVR	0.86	0.88	0.87	0.56	0.67
LR	0.65	0.69	0.67	-0.02	0.15

4.4. Compared with the Other Literatures

Feng [15] uses four machine learning methods to predict the half hourly ST of a maize field in northern China. The results show that the four methods can provide ideal ST prediction at all depths, and the extreme learning machine (ELM) performs best. We compared the prediction results of our proposed algorithm at 5 cm, 10 cm, and 20 cm (only these three depths are involved in both studies) with Feng's optimal prediction results. The results are shown in Table 9. Our proposed method is less accurate than Feng's when the soil depth is 5 cm, but when the soil depth is 10 cm and 20 cm, it is more accurate than the optimal method mentioned in his article.

Table 9. Comparison of the results of the two studies.

Soil Depth	Feng's Research		Our' Research	
	RMSE	MAE	RMSE	MAE
5 cm	1.85	1.44	2.04	1.63
10 cm	2.05	1.6	1.69	1.34
20 cm	2.47	1.91	1.55	1.44

5. Conclusions

Although numerous machine learning models have been used and have achieved good results for predicting the monthly or daily ST, hourly ST predictions have been scarce.

This paper presented a novel scheme for forecasting the hourly ST at five different soil depths in 30 sites, located under five common climates in the United States. Differing from other studies that consider meteorological parameters as the input of the model and the ST as the output, our proposed method includes two submodels for the daily average ST prediction and ST amplitude prediction. The hourly ST was obtained by superposing the prediction results of the two submodels. Through an input–output correlation analysis, we selected nine meteorological parameters and two temporal parameters as the inputs of the daily average ST prediction model, and seven meteorological parameters and one temporal parameter as the inputs of the ST amplitude prediction model. The experimental results show that the proposed model produced the best results for 30 stations (100%) for RMSE, 27 stations (90%) for MAE, and 30 stations (100%) for R^2 , respectively. We compared the prediction results of each algorithm for the ST at different depths, and demonstrated that the integrated BiLSTM developed in this study is better than the benchmark algorithm in terms of ST prediction at the five soil depths involved in the experiment, and has greater potential than the benchmark algorithm in terms of a deeper ST prediction. By comparing the prediction results according to the climate type, it was found that the algorithm proposed in this study performs better than the benchmark algorithm under all climate types involved in the experiment, especially in warm or dry areas.

Although we verified the effectiveness of the model with hourly ST predictions, our method is also applicable to the sequence learning of other data, such as the wind speed, humidity, or temperature.

In future work, we will improve the prediction accuracy of the model for ST by enriching data sets, and adjusting the structure of neural network.

Author Contributions: Conceptualization, C.L. and Y.Z.; methodology, C.L.; software, C.L.; original draft preparation, C.L.; review and editing of manuscript, C.L.; visualization, C.L. and X.R.; and data curation, X.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Data Sharing Fundamental Program for the Construction of the National Science and Technology Infrastructure Platform [grant no. Y719H71006].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yan, Y.; Yan, R.; Chen, J.; Xin, X.; Eldridge, D.J.; Shao, C.; Wang, X.; Lv, S.; Jin, D.; Chen, J.; et al. Grazing modulates soil temperature and moisture in a Eurasian steppe. *Agric. For. Meteorol.* **2018**, *262*, 157–165. [[CrossRef](#)]
2. Yin, X.; Arp, P.A. Predicting forest soil temperatures from monthly air temperature and precipitation records. *Can. J. For. Res.* **1993**, *23*, 2521–2536. [[CrossRef](#)]
3. Brooks, P.D.; McKnight, D.M.; Elder, K. Carbon limitation of soil respiration under winter snowpacks: Potential feedbacks between growing season and winter carbon fluxes. *Glob. Chang. Biol.* **2005**, *11*, 231–238. [[CrossRef](#)]
4. Schimel, J.P.; Bilbrough, C.; Welker, J.M. Increased snow depth affects microbial activity and nitrogen mineralization in two Arctic tundra communities. *Soil Biol. Biochem.* **2004**, *36*, 217–227. [[CrossRef](#)]
5. Kim, S.; Singh, V.P. Modeling daily soil temperature using data-driven models and spatial distribution. *Theor. Appl. Clim.* **2014**, *118*, 465–479. [[CrossRef](#)]
6. Hu, Q.; Feng, S. A Daily Soil Temperature Dataset and Soil Temperature Climatology of the Contiguous United States. *J. Appl. Meteorol.* **2003**, *42*, 1139–1156. [[CrossRef](#)]
7. Yilmaz, T.; Özbek, A.; Yilmaz, A.; Büyükalaca, O. Influence of upper layer properties on the ground temperature distribution. *J. Therm. Sci. Technol.* **2009**, *29*, 43–51.
8. Lahti, M.; Aphalo, P.J.; Finer, L.; Ryyppö, A.; Lehto, T.; Mannerkoski, H. Effects of soil temperature on shoot and root growth and nutrient uptake of 5-year-old Norway spruce seedlings. *Tree Physiol.* **2005**, *25*, 115–122. [[CrossRef](#)]

9. Dang, J.; Liang, W.; Wang, G.; Shi, P.; Wu, D. A preliminary study of the effects of plastic film-mulched raised beds on soil temperature and crop performance of early-sown short-season spring maize (*Zea mays* L.) in the North China Plain. *Crop. J.* **2016**, *4*, 331–337. [[CrossRef](#)]
10. Taussi, M.; Nisi, B.; Pizarro, M.; Morata, D.; Veloso, E.A.; Volpi, G.; Vaselli, O.; Renzulli, A. Sealing capacity of clay-cap units above the Cerro Pabellón hidden geothermal system (northern Chile) derived by soil CO₂ flux and temperature measurements. *J. Volcanol. Geotherm. Res.* **2019**, *384*, 1–14. [[CrossRef](#)]
11. Yener, D.; Ozgener, O.; Ozgener, L. Prediction of soil temperatures for shallow geothermal applications in Turkey. *Renew. Sustain. Energy Rev.* **2017**, *70*, 71–77. [[CrossRef](#)]
12. Zhang, D.; Gao, P.; Zhou, Y.; Wang, Y.; Zhou, G. An experimental and numerical investigation on temperature profile of underground soil in the process of heat storage. *Renew. Energy* **2020**, *148*, 1–21. [[CrossRef](#)]
13. Bi, Y.; Lyu, T.; Wang, H.; Sun, R.; Yu, M. Parameter analysis of single U-tube GHE and dynamic simulation of underground temperature field round one year for GSHP. *Energy* **2019**, *174*, 138–147. [[CrossRef](#)]
14. Hu, G.; Zhao, L.; Wu, X.; Li, R.; Wu, T.; Xie, C.; Qiao, Y.; Shi, J.; Cheng, G. An analytical model for estimating soil temperature profiles on the Qinghai-Tibet Plateau of China. *J. Arid. Land* **2015**, *8*, 232–240. [[CrossRef](#)]
15. Feng, Y.; Cui, N.; Hao, W.; Gao, L.; Gong, D. Estimation of soil temperature from meteorological data using different machine learning models. *Geoderma* **2019**, *338*, 67–77. [[CrossRef](#)]
16. Xing, L.; Li, L.; Gong, J.; Ren, C.; Liu, J.; Chen, H. Daily soil temperatures predictions for various climates in United States using data-driven model. *Energy* **2018**, *160*, 430–440. [[CrossRef](#)]
17. Tabari, H.; Hosseinzadeh Talaee, P.; Willems, P. Short-term forecasting of soil temperature using artificial neural network: ANN-based soil temperature forecasting. *Meteorol. Appl.* **2015**, *22*, 576–585. [[CrossRef](#)]
18. Bilgili, M. Prediction of soil temperature using regression and artificial neural network models. *Theor. Appl. Clim.* **2010**, *110*, 59–70. [[CrossRef](#)]
19. Wu, W.; Tang, X.-P.; Guo, N.-J.; Yang, C.; Liu, H.-B.; Shang, Y.-F. Spatiotemporal modeling of monthly soil temperature using artificial neural networks. *Theor. Appl. Clim.* **2012**, *113*, 481–494. [[CrossRef](#)]
20. Mehdizadeh, S.; Behmanesh, J.; Khalili, K. Evaluating the performance of artificial intelligence methods for estimation of monthly mean soil temperature without using meteorological data. *Environ. Earth Sci.* **2017**, *76*, 59. [[CrossRef](#)]
21. Kisi, O.; Tombul, M.; Zounemat-Kermani, M. Modeling soil temperatures at different depths by using three different neural computing techniques. *Theor. Appl. Clim.* **2014**, *121*, 377–387. [[CrossRef](#)]
22. Zhang, Z.; Xin, X.-P. Research Progress of Biogeochemistry Model DNDC in Carbon Dynamic Modeling. *Acta Agrestia Sinica* **2017**, *25*, 445–452.
23. Zhang, Y.; Li, C.; Trettin, C.; Li, H.; Sun, G. An integrated model of soil, hydrology, and vegetation for carbon dynamics in wetland ecosystems. *Glob. Biogeochem. Cycles* **2002**, *16*, 9–1–9–17. [[CrossRef](#)]
24. Li, Z.; Yang, J.; Drury, C.; Yang, X.; Reynolds, W.; Li, X.; Hu, C. Evaluation of the DNDC model for simulating soil temperature, moisture and respiration from monoculture and rotational corn, soybean and winter wheat in Canada. *Ecol. Model.* **2017**, *360*, 230–243. [[CrossRef](#)]
25. Shi, X.; Mao, J.; Wang, Y.-P.; Dai, Y.; Tang, X. Coupling a terrestrial biogeochemical model to the common land model. *Adv. Atmospheric Sci.* **2011**, *28*, 1129–1142. [[CrossRef](#)]
26. Qi, J.; Li, S.; Li, Q.; Xing, Z.; Bourque, C.P.-A.; Meng, F.-R. A new soil-temperature module for SWAT application in regions with seasonal snow cover. *J. Hydrol.* **2016**, *538*, 863–877. [[CrossRef](#)]
27. Zeng, S.; Xia, J.; Chen, X.; Zou, L.; Du, H.; She, D. Integrated land-surface hydrological and biogeochemical processes in simulating water, energy and carbon fluxes over two different ecosystems. *J. Hydrol.* **2020**, *582*, 124390. [[CrossRef](#)]
28. Wang, M.; Zang, H.; Cheng, L.; Wei, Z.; Sun, G. Application of DBN for estimating daily solar radiation on horizontal surfaces in Lhasa, China. *Energy Procedia* **2019**, *158*, 49–54. [[CrossRef](#)]
29. Qing, X.; Niu, Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* **2018**, *148*, 461–468. [[CrossRef](#)]
30. Zhao, X.; Wei, H.; Wang, H.; Zhu, T.; Zhang, K. 3D-CNN-based feature extraction of ground-based cloud images for direct normal irradiance prediction. *Sol. Energy* **2019**, *181*, 510–518. [[CrossRef](#)]
31. Liu, H.; Mi, X.; Li, Y.; Duan, Z.; Xu, Y. Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional Gated Recurrent Unit network and Support Vector Regression. *Renew. Energy* **2019**, *143*, 842–854. [[CrossRef](#)]

32. Li, F.; Ren, G.; Lee, J. Multi-step wind speed prediction based on turbulence intensity and hybrid deep neural networks. *Energy Convers. Manag.* **2019**, *186*, 306–322. [[CrossRef](#)]
33. Liu, H.; Mi, X.; Li, Y. Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network. *Energy Convers. Manag.* **2018**, *166*, 120–131. [[CrossRef](#)]
34. Fang, K.; Pan, M.; Shen, C. The Value of SMAP for Long-Term Soil Moisture Estimation With the Help of Deep Learning. *IEEE Trans. Geosci. Remote. Sens.* **2019**, *57*, 2221–2233. [[CrossRef](#)]
35. Kottek, M.; Grieser, J.; Beck, C.; Rudolf, B.; Rubel, F. World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* **2006**, *15*, 259–263. [[CrossRef](#)]
36. Stekhoven, D.J.; Bühlmann, P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **2011**, *28*, 112–118. [[CrossRef](#)]
37. Schuster, M.; Paliwal, K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [[CrossRef](#)]
38. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **2005**, *18*, 602–610. [[CrossRef](#)]
39. Zhao, H.; Hou, C.-N.; Alrobassy, H.; Zeng, X. Recognition of Transportation State by Smartphone Sensors Using Deep Bi-LSTM Neural Network. *J. Comput. Netw. Commun.* **2019**, *2019*, 1–11. [[CrossRef](#)]
40. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
41. Zhao, J.; Deng, F.; Cai, Y.; Chen, J. Long short-term memory—Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. *Chemosphere* **2019**, *220*, 486–492. [[CrossRef](#)]
42. Yang, B.; Sun, S.; Li, J.; Lin, X.; Tian, Y. Traffic flow prediction using LSTM with feature enhancement. *Neurocomputing* **2019**, *332*, 320–327. [[CrossRef](#)]
43. Li, Y.; Cao, H. Prediction for Tourism Flow based on LSTM Neural Network. *Procedia Comput. Sci.* **2018**, *129*, 277–283. [[CrossRef](#)]
44. Fischer, T.; Krauss, C. Deep learning with long short-term memory networks for financial market predictions. *Eur. J. Oper. Res.* **2018**, *270*, 654–669. [[CrossRef](#)]
45. Li, X.; Zhang, L.; Wang, Z.; Dong, P. Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and Elman neural networks. *J. Energy Storage* **2019**, *21*, 510–518. [[CrossRef](#)]
46. Lei, J.; Liu, C.; Jiang, D. Fault diagnosis of wind turbine based on Long Short-term memory networks. *Renew. Energy* **2019**, *133*, 422–432. [[CrossRef](#)]
47. Reddy, T.; Priya, R.M.; Parimala, M.S.P.; Chowdhary, C.L.; Reddy, P.K.; Hakak, S.; Khan, W.Z. A deep neural networks based model for uninterrupted marine environment monitoring. *Comput. Commun.* **2020**, *157*, 64–75. [[CrossRef](#)]
48. Achieng, K.O. Modelling of soil moisture retention curve using machine learning techniques: Artificial and deep neural networks vs. support vector regression models. *Comput. Geosci.* **2019**, *133*, 104320. [[CrossRef](#)]
49. Li, K.; Wu, Y.; Nan, Y.; Li, P.; Li, Y. Hierarchical multi-class classification in multimodal spacecraft data using DNN and weighted support vector machine. *Neurocomputing* **2017**, *259*, 55–65. [[CrossRef](#)]
50. Park, J.; Riaz, H.; Kim, H.; Kim, J. Advanced cover glass defect detection and classification based on multi-DNN model. *Manuf. Lett.* **2020**, *23*, 53–61. [[CrossRef](#)]
51. Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural networks. *J. Mach. Learn. Res.* **2011**, *15*, 315–323.
52. Salcedo-Sanz, S.; Rojo-Álvarez, J.L.; Martínez-Ramón, M.; Camps-Valls, G. Support vector machines in engineering: An overview. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2014**, *4*, 234–267. [[CrossRef](#)]
53. Nahvi, B.; Habibi, J.; Mohammadi, K.; Shamshirband, S.; Al Razgan, O.S. Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature. *Comput. Electron. Agric.* **2016**, *124*, 150–160. [[CrossRef](#)]

