agronomy

MDPI

Article

Evaluation of the Potential of Using Machine Learning and the Savitzky–Golay Filter to Estimate the Daily Soil Temperature in Gully Regions of the Chinese Loess Plateau

Wei Deng [1], Dengfeng Liu [1,*], Fengnian Guo [1], Lianpeng Zhang [2], Lan Ma [1], Qiang Huang [1], Qiang Li [3], Guanghui Ming [4] and Xianmeng Meng [5]

[1] State Key Laboratory of Eco-Hydraulics in Northwest Arid Region of China, School of Water Resources and Hydropower, Xi'an University of Technology, Xi'an 710048, China; dengw@sehemodel.club (W.D.)
[2] School of Water Conservancy & Civil Engineering, Northeast Agricultural University, Harbin 150030, China
[3] Center for Ecological Forecasting and Global Change, College of Forestry, Northwest A&F University, Yangling, Xianyang 712100, China
[4] Key Laboratory of Water Management and Water Security for Yellow River Basin (Ministry of Water Resources), Yellow River Engineering Consulting Co., Ltd., Zhengzhou 450003, China
[5] School of Environmental Studies, China University of Geosciences, Wuhan 430074, China
* Correspondence: liudf@xaut.edu.cn

Abstract: Soil temperature directly affects the germination of seeds and the growth of crops. In order to accurately predict soil temperature, this study used RF and MLP to simulate shallow soil temperature, and then the shallow soil temperature with the best simulation effect will be used to predict the deep soil temperature. The models were forced by combinations of environmental factors, including daily air temperature (T_{air}), water vapor pressure (P_w), net radiation (R_n), and soil moisture (VWC), which were observed in the Hejiashan watershed on the Loess Plateau in China. The results showed that the accuracy of the model for predicting deep soil temperature proposed in this paper is higher than that of directly using environmental factors to predict deep soil temperature. In testing data, the range of MAE was 1.158–1.610 °C, the range of RMSE was 1.449–2.088 °C, the range of R^2 was 0.665–0.928, and the range of KGE was 0.708–0.885 at different depths. The study not only provides a critical reference for predicting soil temperature but also helps people to better carry out agricultural production activities.

Keywords: soil temperature; soil moisture; long short-term memory; Savitzky–Golay filter

check for updates

Citation: Deng, W.; Liu, D.; Guo, F.; Zhang, L.; Ma, L.; Huang, Q.; Li, Q.; Ming, G.; Meng, X. Evaluation of the Potential of Using Machine Learning and the Savitzky–Golay Filter to Estimate the Daily Soil Temperature in Gully Regions of the Chinese Loess Plateau. *Agronomy* **2024**, *14*, 703. https://doi.org/10.3390/agronomy14040703

Academic Editor: Gniewko Niedbała

Received: 15 March 2024
Revised: 25 March 2024
Accepted: 26 March 2024
Published: 28 March 2024

1. Introduction

Promoting the sustainable development of agriculture is one of the United Nations Sustainable Development Goals (SDGs) [1]. However, an extraordinary challenge in achieving Sustainable Development Goal 2 (SDG2) is the food problem [2]. Reasonable and effective agricultural production activities can help to meet this challenge. Soil environment plays a vital role in human agricultural production activities [3]. As one of the key parameters of the soil environment, soil temperature directly affects the germination of seeds and the growth of crops [4]. In addition, soil temperature (T_s) plays an important role in many critical processes [5]. It strongly influences a wide range of biotic and abiotic processes, plays an important role in the exchange of energy and matter between the soil and the air, and even affects the local climate [6–8]. T_s is usually influenced by many factors [9], such as meteorological and topographical conditions [10]. To more accurately estimate soil temperatures, scientists have developed three primary methods, including statistical models [11], physical models [12,13], and machine learning [14].

With the development of computer science, machine learning methods have been widely used in many fields [15–20], including agriculture [21,22]. Soil temperature research based on machine learning has also received much attention in recent years [14]. At present, there are more than a dozen machine learning methods for simulating soil temperatures, including wavelet neural network (WNN) [23], long short-term memory (LSTM) [24], extreme learning machine (ELM) [7,25], and random forest (RF) [26,27]. Due to the excellent ability of machine learning methods to handle multiple complex data and nonlinear relationships, studies can make full use of the data for simulations [28,29]. In addition, many researchers have demonstrated that coupling multiple machine learning methods can effectively improve a simulation's accuracy and increase the model's stability [30–33]. Thus, the coupling of multiple machine learning methods has become one of the leading research directions.

Machine learning methods have substantially improved our understanding of T_s, proving the potential of machine learning methods using meteorological and environmental factors to simulate T_s. Many studies have simulated the predicted temperature of surface or shallow soil, but not many simulations of deep soil temperatures have been carried out. However, setting up deep soil monitoring equipment requires a large amount of labor and material costs [34]. Thus, the simulation of deep soil temperatures provides a critical data source for agricultural and land surface management.

Therefore, the objectives of this study were: (1) to simulate the shallow soil temperature by using environmental factors as input and evaluate the performance difference between RF and MLP; (2) to construct deep soil temperature prediction models based on the simulated shallow soil temperature and the air temperature; and (3) to evaluate the performance of deep soil temperature prediction models.

This study explores the feasibility of using environmental factors to simulate shallow soil temperature and using shallow soil temperature to simulate deep soil temperature. It will not only provide a reference for simulations of deep soil temperatures but will also be conducive to better agricultural work on the Loess Plateau in China. For farmers, accurate prediction of soil temperature based on machine learning is also conducive to helping them make decisions in a timely manner and reduce financial losses [35].

2. Materials and Methods

2.1. Study Area

The study area is located in gully regions of the Loess Plateau in China, which has a continental monsoon climate [36]. The Chunhua Ecohydrology Experimental Station in the Hejiashan watershed in Chunhua County, Xianyang City, Shaanxi Province, was constructed for on-site observations, and its location is shown in Figure 1a. The station is equipped with a conventional meteorological observation system, a soil temperature observation system, and an eddy correlation system, which can continuously observe various meteorological elements and soil temperatures at different depths in the field. The detail of the observation system was described by Guo et al. [37]. The soil type in the area is dominated by loess soil, and the average elevation is about 1330 m. In the Chunhua Ecohydrology Experimental Station, meteorological observation tower and equipment are shown in Figure 1b, three-component soil sensors (CS655, Campbell Scientific, Inc., Logan, UT, USA) have been buried alongside the meteorological observation tower, and these can be used to measure the actual soil temperature (T_s) and soil moisture (VWC) at different depths. On the tower, an air temperature sensor (HMP155A, Vaisala, Vantaa, Finland) can be used to measure air temperature (T_{air}), relative humidity, and daily water vapor pressure (P_w), and a four-component radiation sensor (CNR4, Kipp&Zonen, Delft, The Netherlands) can be used to measure net radiation (R_n).

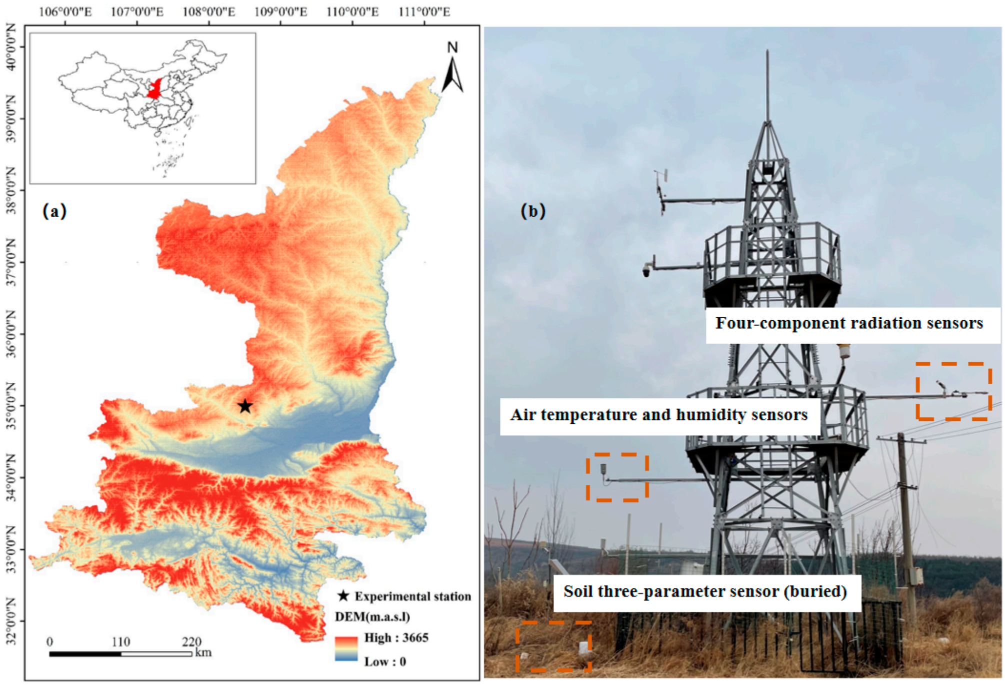

Figure 1. The location of the experimental station and the observation system. (**a**) DEM of Shaanxi Province and the location of Chunhua Ecohydrology Experimental Station. (**b**) Meteorological observation tower and equipment at the station.

2.2. Data Analysis and Processing

The data used in this study were all measured at the Chunhua Ecohydrology Experimental Station, and the dataset is named as Chunhua Ecohydrology Experimental Station Dataset (CEESD) including the daily soil temperatures at different depths (T_{s20cm}, T_{s40cm}, T_{s80cm}, T_{s120cm}, T_{s160cm}, T_{s200cm}), daily soil moisture at 20 cm depth (VWC_{20cm}), daily air temperature at a height of 2 m (T_{air}), daily water vapor pressure (P_w), and daily net radiation (R_n). The data used in this study were measured from 1 March 2020 to 30 October 2023. There are 1340 valid daily data. For outliers and missing values in the measured data, linear interpolation was used for processing. After that, SG filter was applied to the data, and the processed data are shown in Figure 2.

As shown in Figure 2a,b,d, the daily changes in temperature, daily water vapor pressure, and daily net radiation are very drastic. But both of soil temperature at different depths, air temperature, and net radiation all have strong seasonal patterns of change. All show high values in summer and low values in winter. During the spring and summer, shallow soil temperatures are higher than deep soil temperatures. In the autumn and winter, deep soil temperatures were higher than shallow soil temperatures. Figure 2c shows that the peak of VWC_{20cm} is concentrated from May to August each year, which is the same as the time when there is more rainfall in the region. In addition, the study area receives less precipitation in winter, resulting in a decreasing trend in VWC_{20cm} during winter.

In general, the selection of input variables should be based on a simple relationship to achieve high accuracy in the simulation. The Pearson's correlation coefficients among the variables (Table 1) indicated a high correlation between soil temperature and the soil temperature in the adjacent layers. The air temperature and water vapor pressure had the highest correlation with soil temperature, followed by net radiation. All of them are very suitable as input variables for soil temperature simulation. Although the linear correlation between soil moisture and soil temperature is low, soil moisture can affect temperature by controlling evaporation from the soil surface. Dry soils will reduce evaporation and thus increase surface temperatures. Conversely, wet soils usually make surface temperatures cooler [38]. Therefore, soil moisture is important for simulating shallow soil temperature, and it is also used as one of the input variables for simulating soil temperature at a depth of 20 cm (T_{s20cm}).

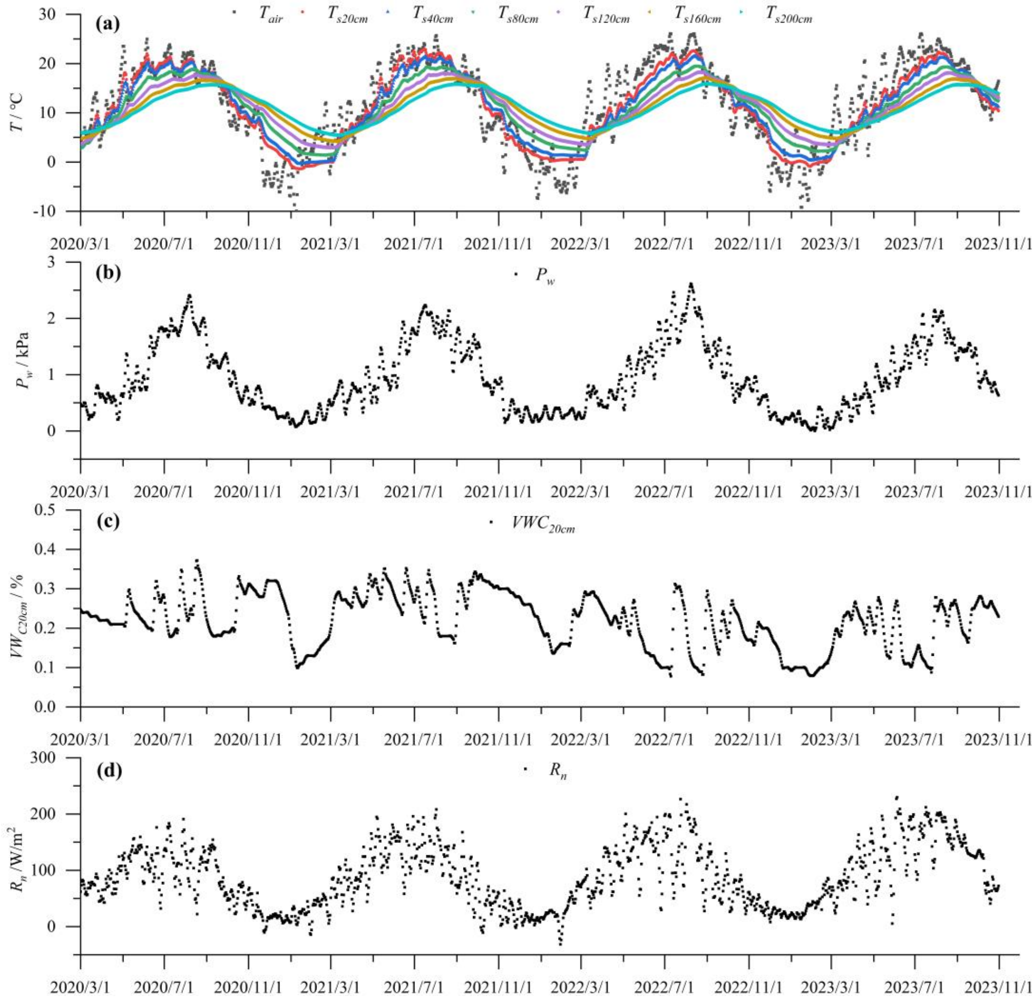

Figure 2. Time series of data. (**a**) The daily soil temperatures at different depths and daily air temperature at a height of 2 m. (**b**) The daily water vapor pressure. (**c**) The daily soil moisture at 20 cm depth. (**d**) The daily net radiation.

Table 1. Pearson's correlation coefficients among the variables.

Statistics	T_{s20cm}	T_{s40cm}	T_{s80cm}	T_{s120cm}	T_{s160cm}	T_{s200cm}
T_{air}	0.92	0.89	0.81	0.73	0.63	0.48
P_w	0.91	0.91	0.89	0.84	0.77	0.66
VWC_{20cm}	0.14	0.16	0.17	0.18	0.18	0.16
R_n	0.74	0.69	0.60	0.50	0.38	0.24
T_{s20cm}	1.00	0.99	0.95	0.89	0.79	0.66
T_{s40cm}	0.99	1.00	0.98	0.93	0.85	0.74
T_{s80cm}	0.95	0.98	1.00	0.99	0.94	0.85
T_{s120cm}	0.89	0.93	0.99	1.00	0.98	0.93
T_{s160cm}	0.79	0.85	0.94	0.98	1.00	0.98
T_{s200cm}	0.66	0.74	0.85	0.93	0.98	1.00

To analyze the statistical information of the selected measured data, the mean (x_{mean}), maximum (x_{max}), minimum (x_{min}), standard deviation (x_{std}), variation coefficient (C_v), skewness (C_S), and kurtosis (C_k) of each data series were calculated in this study. The results in Table 2 show that the maximum and minimum soil temperatures at different depths varied considerably, but the mean values of the soil temperatures at different depths did not differ much from those of the air temperatures, as soil temperatures are mainly influenced by air temperatures [26]. Both the air and soil temperatures at different depths were negatively skewed. Still, the degree of skewness of soil temperature became closer to a normal distribution as the depth of soil increased. Compared with shallow soils, deeper soil temperatures were less susceptible to strong influences from soil surface temperatures and seasonal fluctuations in temperature [39]. The results of x_{std}, C_v, and C_s also show that deeper soil temperatures were more stable and less volatile.

Table 2. Summary of the descriptive statistics of the soil temperature and climate data at different depths.

Data	x_{mean}	x_{max}	x_{min}	x_{std}	C_v	C_S	C_k
T_{air} (°C)	11.66	26.07	−11.44	8.64	0.74	−0.37	−0.90
P_w (kPa)	0.95	2.62	0.00	0.62	0.65	0.50	−0.83
R_n (W/m^2)	87.68	229.85	−31.50	54.80	0.63	0.30	−0.89
VWC_{20cm} (%)	0.22	0.37	0.08	0.07	0.31	−0.26	−0.84
T_{s20cm} (°C)	11.27	22.96	−1.38	7.56	0.67	−0.17	−1.34
T_{s40cm} (°C)	11.23	21.50	−0.36	6.93	0.62	−0.18	−1.36
T_{s80cm} (°C)	11.14	19.56	1.49	5.80	0.52	−0.16	−1.40
T_{s120cm} (°C)	11.09	18.24	2.95	4.96	0.45	−0.14	−1.43
T_{s160cm} (°C)	11.10	17.09	4.21	4.18	0.38	−0.11	−1.46
T_{s200cm} (°C)	11.04	15.90	5.45	3.46	0.31	−0.07	−1.48

The first 80% of the data were used as the training set for training the model, and the last 20% were used as the set for testing the model. Since the observational data showed large variations, all variables used in the model were first normalized as follows:

$$x_{normal} = \frac{x - x_{mean}}{x_{std}} \tag{1}$$

where x_{normal} is the normalized series of the variables, x is the original series of the observed variables, x_{mean} is the mean of the series of the corresponding variable and x_{std} is the standard deviation of the series of the corresponding variable.

2.3. Methods

2.3.1. Principles of RF

Random forest (RF) is a machine learning algorithm proposed by Breiman in 2001 based on methods such as classification, regression trees, and random subspaces [40,41]. Random forest is composed of multiple decision trees, where each regression tree is trained on a subset of data and a subset of explanatory variables that together determine the predicted values [42]. The construction process of random forest is shown in Figure 3. Random forest can effectively reduce the risk of overfitting due to its high stability and feature robustness [43]. It is now widely used in stochastic classification and stochastic regression.

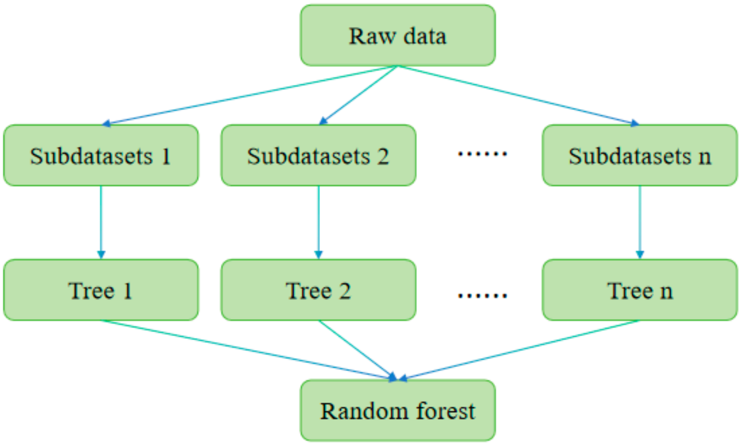

Figure 3. Structure of RF.

2.3.2. Principles of MLP

Perceptron was first proposed by Frank to solve the classification problem [40]. Multi-layer perceptrons (MLP) are formed by connecting several perceptrons. MLP is a multilayer artificial neural network that can handle nonlinear relationships [44]. MLP is a forward feedback artificial neural network with good nonlinear global effect and high parallel ability. It can be used to solve classification and regression problems, and its basic structure is composed of input layer, hidden layer, and output layer [45,46], as shown in Figure 4.

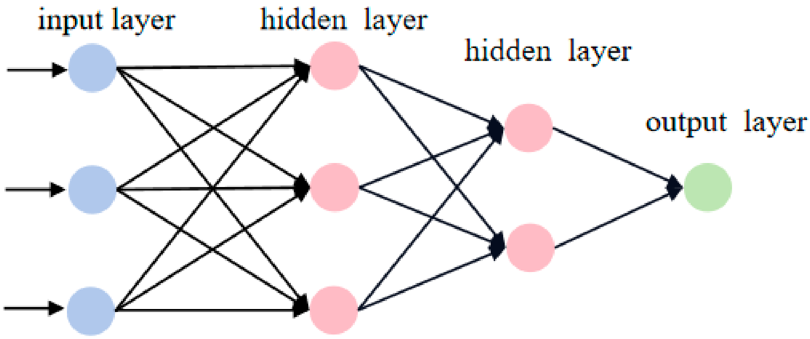

Figure 4. Structure of MLP.

2.3.3. Principles of LSTM

LSTM is widely used in simulation [47] and was first proposed by Hochreiter and Schmidhuber [48]. It is a good solution to the problems of insufficient long-term memory capacity, gradient explosion, and gradient vanishing that exist in traditional RNN [49]. It solves these problems by setting up forgetting gates, input gates, and output gates [50]. Its conventional unit structure is shown in Figure 5.

Its operation process can be expressed as follows:

$$\begin{aligned}
i_t &= \sigma(W_{hi}h_{t-1} + W_{xi}x_t + W_{ci}x_{t-1} + b_i) \\
f_t &= \sigma(W_{hf}h_{t-1} + W_{xf}x_t + W_{cf}x_{t-1} + b_f) \\
c_t &= f \otimes c_{t-1} + i_t \otimes tanh(W_{hc}h_{t-1} + W_{xc}x_t + b_i) \\
o_t &= \sigma(W_{ho}h_{t-1} + W_{xo}x_t + W_{co}x_t + b_0) \\
h_t &= o_t \otimes tanh(c_t) \\
y_t &= W_{hy}h_t + b_0
\end{aligned} \tag{2}$$

where x_t and y_t are the inputs and outputs of LSTM at moment t; i_t, f_t, c_t and o_t are the input gates, forgetting gates, memory cell states, and output gates, respectively, at moment t; w and b are, respectively, the weight coefficient matrices and bias terms for the corresponding moments and corresponding gates; h_t is the recursive input at moment t; σ is the sigmoid activation function and tanh is the hyperbolic tangent activation function.

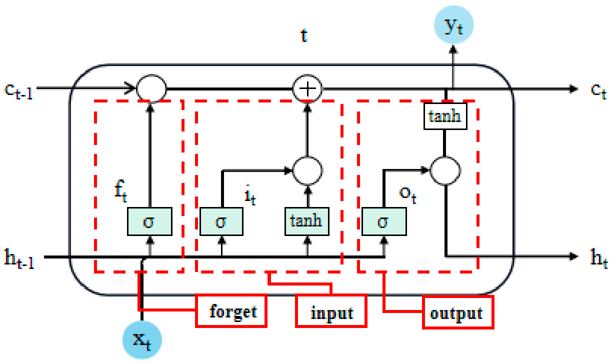

Figure 5. Unit structure of LSTM.

2.3.4. Schematic Workflow of Deep Soil Temperature Prediction

Figure 6 shows the schematic workflow of the methodology used in this study. The workflow consists of three main components: constructing input portfolios, model development, and model evaluation.

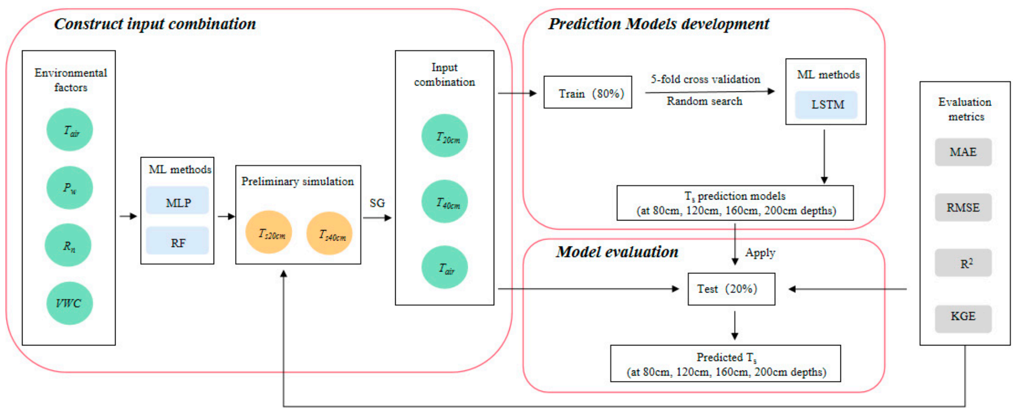

Figure 6. Schematic workflow of this study. (RF, MLP, and LSTM are different machine learning models. RF = random forest, MLP = multilayer perceptron, LSTM = long short-term memory).

In constructing the input combination, pre-processed air temperature data (T_{air}), water vapor pressure data (P_w), net radiation (R_n), and soil moisture data (VWC_{20cm}) are used as inputs for preliminary simulation of soil temperature at 20 cm depth (T_{s20cm}) and 40 cm depth (T_{s40cm}). In this section, both MLP and RF are used to simulate T_{s20cm} and T_{s40cm}. A model with better performance (evaluation metrics with better results on the test set) will be selected. After that, the simulated T_{s20cm}, T_{s40cm}, and T_{air} (preprocessed by SG filter) were used as the input combination to simulate soil temperature at other depths (T_{s80cm}, T_{s120cm}, T_{s160cm}, T_{s200cm}). In the model development, the first 80% of the input data were used as the training set for training the model. Five-fold cross-validation is used on the training dataset for performance evaluation of the model parameters. Random search is used to find optimal hyperparameters. The built model can predict the soil temperature of the target layer in the previous seven days based on the soil temperature and air temperature. In the model evaluation, each model predicts the concentration of Ts from the held-out testing dataset (last 20%), which was separated in the beginning and is only used for validation. The performance differences of LSTM in simulating deep soil temperature using simulated shallow soil temperature (T_{s20cm}, T_{s40cm}) and observed environmental factors (T_{air}, P_w) are mainly evaluated.

2.3.5. Evaluation Metrics

In this study, mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and Kling–Gupta efficiency coefficient (KGE) were chosen to evaluate the results of the simulation. The formulas for calculating these are as follows:

$$\begin{aligned}
&\mathrm{MAE} = \frac{\sum_1^N |y_i - y_{si}|}{N} \\
&\mathrm{RMSE} = \sqrt{\frac{\sum_1^N (y_i - y_{si})^2}{N}} \\
&\mathrm{R}^2 = 1 - \frac{\sum_1^N (y_i - y_{si})^2}{\sum_1^N (y_i - y_m)^2} \\
&\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + \left(\frac{\mu_s}{\mu_0} - 1\right)^2 + \left(\frac{\sigma_s/\mu_s}{\sigma_0/\mu_0} - 1\right)^2}
\end{aligned} \tag{3}$$

where y_{si} and y_i are the simulated values and measured values, y_m is the mean of the series of measured values, N is the number of samples, μ_s and μ_0 are the mean square deviation of the simulated values and the series of measured values, σ_s and σ_0 are mean square deviation of the simulated value and the series of measured values, and r is the linear correlation coefficient between the simulated values and the series of measured values. Among the four model evaluation metrics, a smaller MAE and RMSE mean better performance, while when R^2 and KGE are closer to 1, the result of the simulation is better.

3. Results

3.1. Input Combination of Shallow Soil Temperature

The daily air temperature (T_{air}), daily water vapor pressure data (P_w), net radiation (R_n), and soil moisture data (VWC_{20cm}) were selected as input variables in this study. On the basis of the results of the correlation analysis (Table 2), four different input combinations were set up (Table 3). The correlation between air temperature, water vapor pressure, and shallow soil temperature is high. Therefore, the air temperature is used as combination 1, and the combination of air temperature and water vapor pressure is used as combination 2. Then, other variables are added to become combination 3 and combination 4.

Table 3. Combinations of different input variables.

Combination No.	Input Variables
1	T_{air}
2	T_{air}-P_w
3	T_{air}-P_w-R_n
4	T_{air}-P_w-R_n-VWC_{20cm}

The input combinations in Table 3 were used as input data in the RF and MLP to simulate the daily soil temperature at different depths (20 cm, 40 cm) and determine the optimal combination of input. In this preliminary simulation process, the first 80% of the data was selected for training the model, and the last 20% of the data was used for testing. Five-fold cross-validation is used on the training dataset for evaluation of the model parameters. Random search is used to find optimal hyperparameters.

3.2. Evaluation of the Results of Different Combinations of Input

The evaluation metrics of RF and MLP for simulating shallow soil temperature with different combinations of input are shown in Table 4. Both the training dataset and testing dataset reported acceptable and close results to each other according to the following: the minimum value of R^2 was 0.81, the minimum value of KGE was 0.88 for the training dataset, also, the minimum value of R^2 was 0.75, the minimum value of KGE was 0.80 for the testing dataset. The range of results showed a reasonable accuracy for shallow soil temperature simulation. According to the results of the evaluation metrics of different input combinations, the worst evaluation metrics of RF and MLP appear in the simulation with input combination 1, and the best evaluation metrics mostly appear in the simulation with input combination 4. The addition of meteorological factors improved the model's performance further, and the evaluation metrics of the simulation mostly improved. The addition of meteorological factors improved the model's performance further, and the evaluation metrics mostly improved. In the simulation of soil temperature at 40 cm depth by MLP, the MAE of input 4 is 1.1 °C lower than that of input 1, and the RMSE of input 4 is 1.47 °C lower than that of input 1. Overall, the results of the training dataset and the testing dataset all showed that input 4 produced the best simulation of the daily average temperature of the soil at different depths, so it was used as the optimal combination of the input and can be a more accurate simulation to the shallow soil temperature for the simulation of deep soil temperature.

Table 4. The evaluation metrics of soil temperature simulated by RF and MLP with different combinations of input at shallow depths.

Model	Depths	Input	Training Dataset				Testing Dataset			
			MAE (°C)	RMSE (°C)	R^2	KGE	MAE (°C)	RMSE (°C)	R^2	KGE
RF	20 cm	1	1.924	2.565	0.889	0.918	2.125	2.924	0.804	0.867
		2	1.266	1.693	0.952	0.947	1.309	1.725	0.932	0.958
		3	1.109	1.484	0.963	0.952	1.173	**1.497**	**0.949**	0.925
		4	0.771	1.088	0.980	0.972	**1.161**	1.538	0.946	**0.963**
	40 cm	1	2.131	2.823	0.839	0.877	2.415	3.151	0.750	0.804
		2	1.437	1.935	0.924	0.927	1.491	1.879	0.911	0.915
		3	1.280	1.719	0.940	0.929	1.434	1.782	0.920	0.892
		4	0.857	1.236	0.969	0.962	**1.379**	**1.769**	**0.921**	**0.958**
MLP	20 cm	1	2.111	2.731	0.874	0.916	2.104	2.783	0.823	0.876
		2	1.339	1.796	0.946	0.967	1.203	1.551	0.945	0.955
		3	1.319	1.778	0.947	0.938	1.118	1.436	0.953	0.958
		4	1.241	1.614	0.956	0.935	**1.041**	**1.404**	**0.955**	**0.973**
	40 cm	1	2.323	3.032	0.814	0.896	2.375	3.113	0.756	0.834
		2	1.523	2.064	0.914	0.880	1.389	1.803	0.918	0.891
		3	1.476	1.997	0.919	0.952	1.296	1.652	0.931	0.959
		4	1.384	1.879	0.929	0.960	**1.278**	**1.639**	**0.932**	**0.966**

Note that the optimal models are boldfaced.

According to the scatterplots of the results from the models' simulations versus the measured values, as shown in Figure 7, the two models simulated the soil temperature well at shallow depths (R^2 was close to 1), and the accuracy of the simulations was high. Figure 8 is the time series of soil temperature measured and simulated soil temperature by RM and MLP models during the study period. It also shows an acceptable soil temperature simulated by RF and MLP however, simulating the steady change in soil temperature is a challenging task for them. The main reason is that the change in the air temperature and water vapor pressure in the input combination is very severe, while the change in soil temperature is relatively gentle. Even the air temperature after SG filtering changes much more violently than the soil temperature. It is worth noting that the simulation performance of RF on soil temperature in winter is better than that of MLP, which can simulate the steady change in soil temperature in winter. This is also the main reason why the r^2 of RF is better than MLP. Figure 8 is the time series of soil temperature measured and predicted by the RF and MLP model during the study period (with the optimal combination of input). As a whole, the trend in soil temperature can be better simulated, however, the two methods have a certain underestimation of soil temperature in October.

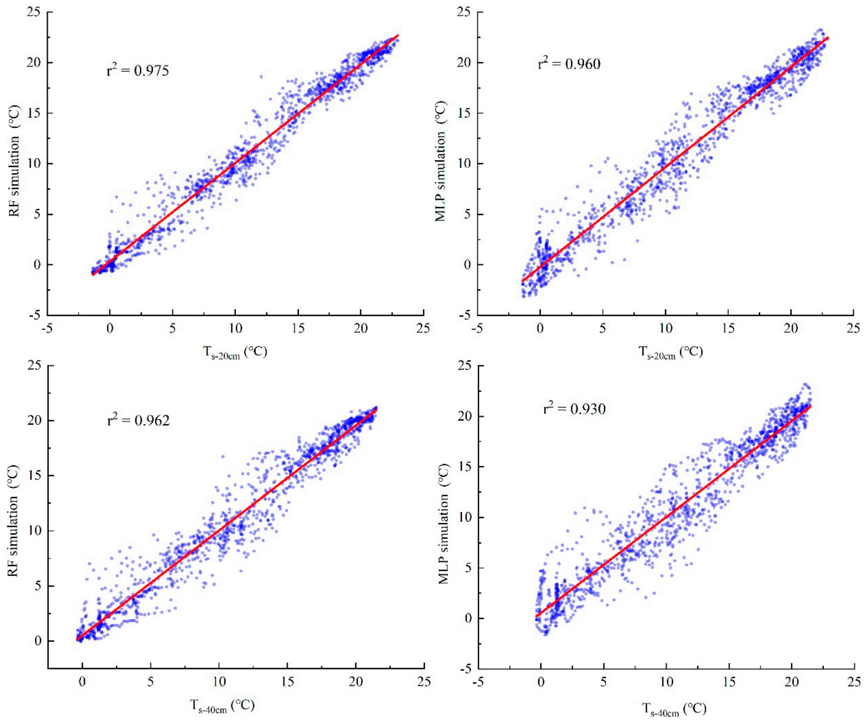

Figure 7. Scatterplots of the simulated and measured shallow soil temperatures.

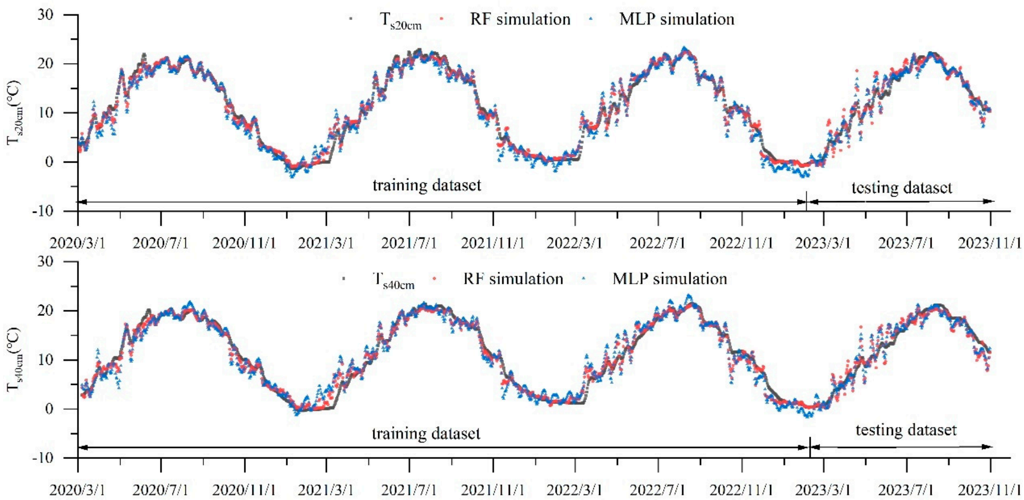

Figure 8. Simulated and measured shallow soil temperatures (with the optimal combination of input).

3.3. Evaluating the Performance of LSTM Prediction of Deep Soil Temperature

The simulated shallow soil temperature (T_{s20cm}, T_{s40cm}) and air temperature (T_{air}) will be used for deep soil temperature (T_{s80cm}, T_{s120cm}, T_{s160cm}, T_{s200cm}) prediction. In the model development, five-fold cross-validation is used on the training dataset for evaluation of the model parameters, and random search is used to find optimal hyperparameters. Figure 9 shows scatterplots of measured and predicted deep soil temperatures (LSTM). Their linear correlation of r^2 decreases with increasing soil depth, with a maximum value of 0.928 and a minimum value of 0.625. With the increase in soil depth, the prediction error increases gradually.

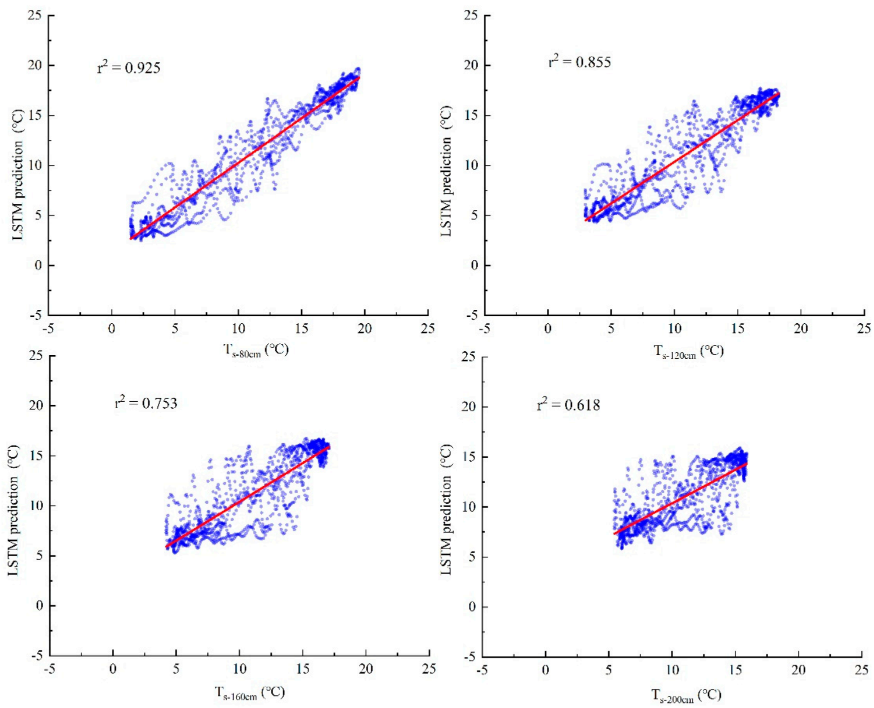

Figure 9. Scatterplots of the predicted and measured deep soil temperatures (LSTM).

Table 5 displays the evaluation metrics results. The MAE and RMSE metrics among the four evaluated metrics increased when soil depth increased, whereas R^2 and KGE decreased. Both results of the training dataset and testing dataset were reported to be acceptable according to the following: the minimum value of R^2 was 0.60, the minimum value of KGE was 0.74 for the training dataset, also, the minimum value of R^2 was 0.66, the minimum value of KGE was 0.71 for the testing dataset. The range of results showed a reasonable accuracy for deep soil temperature prediction.

Table 5. The evaluation metrics of soil temperature predicted by LSTM at deep depths.

Depths	Training Dataset				Testing Dataset			
	MAE (°C)	RMSE (°C)	R^2	KGE	MAE (°C)	RMSE (°C)	R^2	KGE
80 cm	1.192	1.637	0.921	0.910	1.158	1.449	0.928	0.885
120 cm	1.418	1.913	0.848	0.869	1.436	1.773	0.868	0.815
160 cm	1.538	2.098	0.741	0.827	1.554	1.971	0.787	0.775
200 cm	1.561	2.155	0.600	0.740	1.610	2.088	0.665	0.708

The time series of the soil temperature observed and the predicted soil temperature using the LSTM model during the study period are shown in Figure 10. As a whole, the trend in soil temperature can be better predicted after using the shallow soil temperature and air temperature from seven days prior. It also shows that the LSTM model simulates high and low temperatures more accurately and that the simulation error-prone places are distributed close to the temperature change. The soil temperature is higher in summer and lower in winter. One of the primary reasons the soil temperature predicted is more precise and constant is that the summertime temperature is a continuous hot temperature, and the wintertime temperature is also a continuous low temperature.

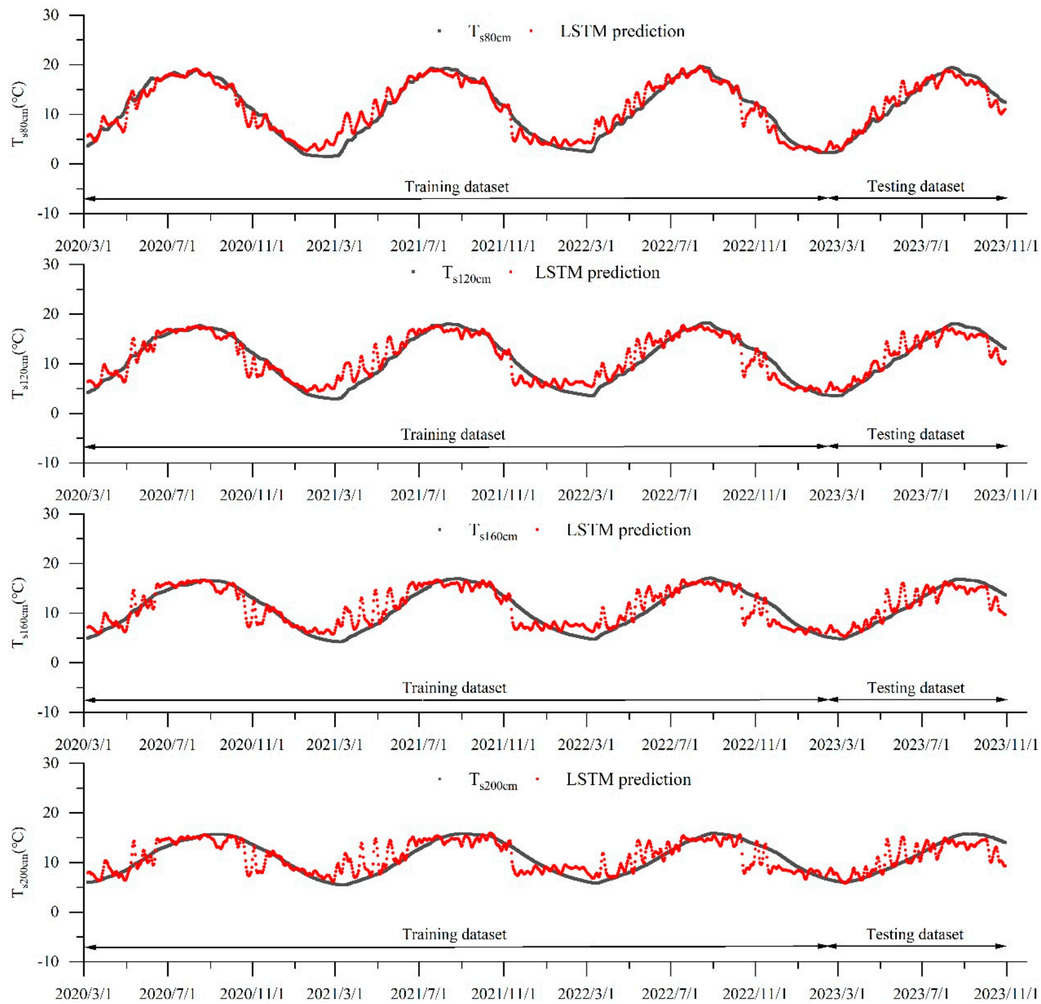

Figure 10. Predicted and measured deep soil temperatures.

The air temperature, water vapor pressure, net radiation, and soil moisture can also be directly used to predict deep soil temperature. The evaluation metrics are shown in Table 6 (prediction using environmental factors from seven days ago). In the testing dataset, the prediction results are worse than the prediction method proposed in this study, especially at 200 cm, where r^2 is only 0.579.

Table 6. The evaluation metrics of soil temperature predicted by LSTM at deep depths (using environmental factors prediction).

Depths	Training Dataset				Testing Dataset			
	MAE (°C)	RMSE (°C)	R^2	KGE	MAE (°C)	RMSE (°C)	R^2	KGE
80 cm	1.144	1.619	0.923	0.901	1.249	1.647	0.908	0.949
120 cm	1.146	1.663	0.886	0.927	1.586	2.042	0.825	0.888
160 cm	1.243	1.769	0.816	0.869	1.786	2.330	0.703	0.779
200 cm	1.158	1.763	0.732	0.821	1.666	2.339	0.579	0.741

3.4. Impact of Sliding Panes on Prediction Accuracy

The LSTM model was used to predict deep soil temperature. In this study, shallow soil temperatures are the important input variables for prediction, and their simulation accuracy certainly has an impact on the prediction accuracy. Not only do the input variables have an effect on the prediction accuracy, but the length of the data used for prediction (the size of the sliding panes) also has an effect on the prediction accuracy. The sliding panes refer to dividing the time series data into continuous windows during the training process and using these panes to train the model. In this section, setting different sizes of the sliding panes are used for prediction and their performance differences are compared. The different sizes of the sliding panes are shown in Table 7.

Table 7. The different sizes of the sliding panes.

Models	Different Size of the Sliding Pane
LSTM3	3
LSTM7	7
LSTM10	10
LSTM14	14
LSTM21	21

The prediction results of the different models in Table 7 are shown in Table 8. The results show that all evaluation metrics become better with the increase in the size of the sliding panes. It was also found that the size of the sliding pane was directly related to the magnitude of the fluctuations in the predicted values. The larger the size of the sliding pane, the smoother the predicted values. This suggests that when the size of the sliding panes is too small, the model is likely to fail to capture long-term data features. Thus, choosing the right sliding window size is equally important. The optimal sliding pane size is often achieved through continuous experimentation.

Table 8. The evaluation metrics of soil temperature predicted by different LSTM.

Model	Depths	Training Dataset				Testing Dataset			
		MAE (°C)	RMSE (°C)	R^2	KGE	MAE (°C)	RMSE (°C)	R^2	KGE
LSTM3	80 cm	1.383	1.900	0.894	0.923	1.371	1.732	0.901	0.881
	120 cm	1.589	2.173	0.805	0.866	1.599	2.013	0.835	0.808
	160 cm	1.711	2.327	0.683	0.808	1.740	2.203	0.738	0.751
	200 cm	1.768	2.314	0.541	0.697	1.842	2.292	0.598	0.653
LSTM7	80 cm	1.192	1.637	0.921	0.910	1.158	1.449	0.928	0.885
	120 cm	1.418	1.913	0.848	0.869	1.436	1.773	0.868	0.815
	160 cm	1.538	2.098	0.741	0.827	1.554	1.971	0.787	0.775
	200 cm	1.561	2.155	0.600	0.740	1.610	2.088	0.665	0.708
LSTM10	80 cm	1.040	1.448	0.938	0.951	0.926	1.215	0.948	0.935
	120 cm	1.233	1.720	0.877	0.917	1.161	1.525	0.900	0.889
	160 cm	1.387	1.933	0.780	0.879	1.310	1.800	0.820	0.839
	200 cm	1.421	1.997	0.655	0.794	1.385	1.937	0.709	0.772
LSTM14	80 cm	0.957	1.289	0.951	0.975	0.829	1.051	0.960	0.971
	120 cm	1.132	1.531	0.902	0.949	0.991	1.269	0.929	0.914
	160 cm	1.262	1.754	0.818	0.902	1.166	1.510	0.870	0.848
	200 cm	1.255	1.699	0.774	0.776	1.280	1.825	0.711	0.836
LSTM21	80 cm	0.724	0.979	0.972	0.977	0.679	0.833	0.972	0.981
	120 cm	0.752	1.052	0.954	0.967	0.797	1.007	0.952	0.957
	160 cm	0.802	1.171	0.918	0.950	0.850	1.135	0.923	0.949
	200 cm	0.795	1.238	0.866	0.909	0.842	1.221	0.880	0.895

3.5. Effect of Savitzky–Golay Filter on Prediction Accuracy

The Savitzky–Golay filter is commonly used in data preprocessing processes to eliminate data noise and reduce data fluctuations. Compared to simple moving average filtering, SG filter can better retain the overall trend of the data while smoothing it. SG filter is also used in this study for post-processing the prediction data of the LSTM7 model (projections using data from the previous seven days) to explore whether it can further improve prediction accuracy. The postprocessed model is named LSTM7-SG. The results of the evaluation indicators are shown in Table 9. It shows that SG filter postprocessing of LSTM7 can improve the results of some evaluation metrics, but the improvement is very limited. This also suggests that the SG filter is more suitable for data preprocessing.

Table 9. The evaluation metrics of soil temperature predicted by LSTM7 and LSTM7-SG.

Model	Depths	Training Dataset				Testing Dataset			
		MAE (°C)	RMSE (°C)	R^2	KGE	MAE (°C)	RMSE (°C)	R^2	KGE
LSTM7	80 cm	1.192	1.637	0.921	0.910	1.158	1.449	0.928	0.885
	120 cm	1.418	1.913	0.848	0.869	1.436	1.773	0.868	0.815
	160 cm	1.538	2.098	0.741	0.827	1.554	1.971	0.787	0.775
	200 cm	1.561	2.155	0.600	0.740	1.610	2.088	0.665	0.708
LSTM7-SG	80 cm	1.186	1.626	0.922	0.909	1.152	1.435	0.930	0.885
	120 cm	1.409	1.896	0.851	0.869	1.430	1.755	0.871	0.814
	160 cm	1.526	2.076	0.747	0.827	1.543	1.948	0.792	0.774
	200 cm	1.548	2.128	0.610	0.742	1.598	2.061	0.673	0.707

4. Discussion

From the perspective of energy exchange, heat transfer occurs between the temperature of the air and the soil, and the air temperature greatly affects the soil temperature [51]. The change in soil moisture can control the partitioning of surface energy between sensible and latent heat fluxes through evapotranspiration, and they jointly drive the change in soil temperature [52–54]. Shallow soil temperature plays a significant role in the land–air heat exchange that determines the underground temperature of deep soil [55]. This shows that both air temperature and soil moisture have the potential to simulate soil temperature, and shallow soil temperature also has the potential to simulate deep soil temperature. Currently, a number of researchers have found that air temperature, solar radiation, and rainfall in combination may accurately simulate the soil temperature of different areas that have different climatic and geographical circumstances [7,25,56,57]. In our research, air temperature, water vapor pressure, net radiation, and soil moisture were used to simulate soil temperature. Air temperature has a direct impact on soil temperature. The magnitude of water vapor pressure depends on the amount of water vapor in the atmosphere [58]. Soil moisture directly indicates the condition of water content in the soil. The difference between downward and upward (sun and earth) radiation is referred to as net radiation [59]. To an extent, soil moisture and water vapor pressure can reflect the amount of rainfall. In contrast to solar radiation, net radiation takes into account the impact of upward radiation on the Earth. The selection of input variables is an essential task in time series prediction, and the choice of variables is dependent on the quality and correlation of the data. Rainfall has a far lower correlation with shallow soil temperature than water vapor pressure and soil moisture. Based on the findings of previous research, this is a new attempt to simulate soil temperature using several factors, which may potentially have broader applicability. However, the experimental site is only Chunhua Experimental Station in the Hejiashan watershed, its applicability in other regions still needs to be verified.

For the machine learning methods used in this study, RF may reduce the danger of overfitting for the machine learning methods utilized in this work [60]. MLP is quite good at modeling nonlinearities [23]. Both of them are more skilled at using a wide range of environmental inputs to determine the temperature of the shallow soil and improve the simulation's impact. Furthermore, neither of the two models will lose data and their simulation speeds are quicker than LSTM. Therefore, to simulate the shallow soil temperature, RF and MLP have been used. Long-term dependencies in sequence data may be efficiently captured by LSTM, which can also thoroughly mine the link between input data and target variables. It is excellent for forecasting [55].

Accurately estimating soil temperature is critical to carry out agricultural planting activities. To a certain extent, the planting time of field crops and greenhouse crops all depends on the optimal soil temperature for seed germination and seedling emergence [34,61]. Different crops demand different temperatures for optimal growth. In order to help farmers better plan when to sow their crops, this study proposes a model that can accurately predict soil temperature. This could improve the crop's survival rate, which will raise output and boost farmers' income. In addition, the exploration of input combinations can also identify the number of environmental factors, so as to select the most suitable number of equipment to monitor research areas and reduce the cost of site equipment.

5. Conclusions

With on-site observation data, including the measured temperature of the soil and different environmental factors, RF and MLP are used to simulate shallow soil temperature, and LSTM is used to predict deep soil temperature in the gully areas of the Loess Plateau in China. The main conclusions are as follows:

(1) For different combinations of input variables, the inclusion of relevant environmental factors can improve the model's performance. When the daily temperature of the air is at a height of 2 m (T_{air}), daily water vapor pressure data (P_w), net radiation (R_n), and soil moisture data (VWC_{20cm}) were jointly used as inputs for all the simulations at 20 cm and 40 cm depths, the results of RF and MLP were the best. Both RF and MLP can simulate shallow soil temperature well, but the performance of MLP is better than that of RF.

(2) It is feasible to use LSTM to predict the deep soil temperature with the simulated shallow soil temperature and the measured air temperature as input.

(3) The accuracy of soil temperature prediction is different at different depths. With the increase in soil depth, the accuracy of soil temperature prediction decreases. The simulation accuracy of shallow soil temperature directly affects the prediction accuracy of deep soil temperature. In addition, the size of the sliding pane of the LSTM model also affects the prediction accuracy.

(4) The SG filter is more suitable for data preprocessing, and its ability to post-process prediction results is very limited.

This study evaluated the feasibility of simulating the daily soil temperature using conventional machine learning techniques (MLR and MLP) for shallow soil temperature. The combination of input variables in the shallow soil simulation was mostly determined by the variables' physical relevance and by doing numerous experiments. More attention could be paid to the interrelationships among the environmental factors and other environmental factors in future studies to further improve the stability and accuracy of the simulation. LSTM is used to predict deep soil temperature. In the process of prediction, due to the small amount of data, the predicted soil temperature still has some instability. In addition, the design of the sliding pane size of the prediction model also only depends on repeated experiments. In the future, we can focus on optimizing the setting of the window size and the setting of the input combination.

Author Contributions: Conceptualization, D.L.; methodology, D.L. and W.D.; software, W.D. and L.M.; validation, F.G.; formal analysis, W.D.; investigation, D.L., W.D., F.G., L.Z. and L.M.; resources, D.L. and Q.H.; data curation, D.L., W.D., F.G. and L.Z.; writing—original draft preparation, W.D.; writing—review and editing, D.L., Q.L., G.M., F.G., L.M. and X.M.; visualization, W.D. and L.M.; supervision, Q.H.; project administration, D.L.; funding acquisition, D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (NSFC) (Grant Nos. 52279025, 42071335, and 52109031) and the National Key Research and Development Program of China (2022YFF1302200).

Data Availability Statement: The experimental data used in this study were obtained from the Chunhua Ecohydrology Experimental Station of Xi'an University of Technology and the dataset is named Chunhua Ecohydrology Experimental Station Dataset (CEESD), which can be obtained from the corresponding author (Dengfeng Liu, liudf@xaut.edu.cn) upon personal reasonable request.

Conflicts of Interest: Author Guanghui Ming was employed by the company Yellow River Engineering Consulting Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Veldhuizen, L.J.; Giller, K.E.; Oosterveer, P.; Brouwer, I.D.; Janssen, S.; van Zanten, H.H.; Slingerland, M.A. The Missing Middle: Connected action on agriculture and nutrition across global, national and local levels to achieve Sustainable Development Goal 2. *Glob. Food Secur.* **2020**, *24*, 100336. [CrossRef]
2. Linghu, L.; Sun, P.; Zhang, M.; Wu, Y. Data-Driven Projections Demonstrate Non-Farming Use of Cropland in Non-Major Grain-Producing Areas: A Case Study of Shaanxi Province, China. *Agronomy* **2023**, *13*, 2060. [CrossRef]
3. Furtak, K.; Gawryjołek, K.; Marzec-Grządziel, A.; Niedźwiecki, J. The Influence of Human Agricultural Activities on the Quality of Selected Fluvisols from the Vistula River Valley, Poland—Preliminary Research. *Agronomy* **2024**, *14*, 480. [CrossRef]
4. Zeynoddin, M.; Ebtehaj, I.; Bonakdari, H. Development of a linear based stochastic model for daily soil temperature prediction: One step forward to sustainable agriculture. *Comput. Electron. Agric.* **2020**, *176*, 105636. [CrossRef]
5. Seyfried, M.S.; Flerchinger, G.N.; Murdock, M.D.; Hanson, C.L.; Van Vactor, S. Long-Term Soil Temperature Database, Reynolds Creek Experimental Watershed, Idaho, United States. *Water Resour. Res.* **2001**, *37*, 2843–2846. [CrossRef]
6. Kramer, P.J. Effects of Soil Temperature on the Absorption of Water by Plants. *Science* **1934**, *79*, 371–372. [CrossRef] [PubMed]
7. Alizamir, M.; Kisi, O.; Ahmed, A.N.; Mert, C.; Fai, C.M.; Kim, S.; Kim, N.W.; El-Shafie, A.; Lin, L. Advanced machine learning model for better prediction accuracy of soil temperature at different depths. *PLoS ONE* **2020**, *15*, e231055. [CrossRef] [PubMed]
8. Ganeshi, N.G.; Mujumdar, M.; Takaya, Y.; Goswami, M.M.; Singh, B.B.; Krishnan, R.; Terao, T. Soil moisture revamps the temperature extremes in a warming climate over India. *npj Clim. Atmos. Sci.* **2023**, *6*, 12. [CrossRef]
9. Onwuka, B.; Mang, B. Effects of soil temperature on some soil properties and plant growth. *Adv. Plants Agric. Res.* **2018**, *8*, 34–37. [CrossRef]
10. Yin, X.; Arp, P.A. Predicting forest soil temperatures from monthly air temperature and precipitation records. *Can. J. Forest Res.* **1993**, *23*, 2521–2536. [CrossRef]
11. Zhao, H.; Sassenrath, G.F.; Kirkham, M.B.; Wan, N.; Lin, X. Daily soil temperature modeling improved by integrating observed snow cover and estimated soil moisture in the USA Great Plains. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 4357–4372. [CrossRef]
12. Mihalakakou, G. On estimating soil surface temperature profiles. *Energy Build.* **2002**, *34*, 251–259. [CrossRef]
13. Qi, J.; Li, S.; Li, Q.; Xing, Z.; Bourque, 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]
14. Padarian, J.; Minasny, B.; McBratney, A.B. Machine learning and soil sciences: A review aided by machine learning tools. *Soil* **2020**, *6*, 35–52. [CrossRef]
15. Recknagel, F.; French, M.; Harkonen, P.; Yabunaka, K. Artificial neural network approach for modelling and prediction of algal blooms. *Ecol. Model.* **1997**, *96*, 11–28. [CrossRef]
16. Lin, X.; Duan, X.; Jacobs, C.; Ullmann, J.; Chan, C.; Chen, S.; Cheng, S.; Zhao, W.; Poduri, A.; Wang, X.; et al. High-throughput brain activity mapping and machine learning as a foundation for systems neuropharmacology. *Nat. Commun.* **2018**, *9*, 5142. [CrossRef] [PubMed]
17. Hulbert, C.; Rouet-Leduc, B.; Johnson, P.A.; Ren, C.X.; Rivière, J.; Bolton, D.C.; Marone, C. Similarity of fast and slow earthquakes illuminated by machine learning. *Nat. Geosci.* **2019**, *12*, 69–74. [CrossRef]
18. Fang, K.; Kifer, D.; Lawson, K.; Shen, C. Evaluating the potential and challenges of an uncertainty quantification method for long short-term memory models for soil moisture predictions. *Water Resour. Res.* **2020**, *56*, e2020WR028095. [CrossRef]
19. Cui, Q.; Ammar, M.E.; Iravani, M.; Kariyeva, J.; Faramarzi, M. Regional wetland water storage changes: The influence of future climate on geographically isolated wetlands. *Ecol. Indic.* **2021**, *120*, 106941. [CrossRef]
20. Zhong, L.; Lei, H.; Gao, B. Developing a Physics-Informed Deep Learning Model to Simulate Runoff Response to Climate Change in Alpine Catchments. *Water Resour. Res.* **2023**, *59*, e2022WR034118. [CrossRef]
21. Nabavi-Pelesaraei, A.; Shaker-Koohi, S.; Dehpour, M.B. Modeling and optimization of energy inputs and greenhouse gas emissions for eggplant production using artificial neural network and multi-objective genetic algorithm. *Int. J. Adv. Biol. Biomed. Res.* **2013**, *4*, 170–183.
22. Sándor, R.; Barcza, Z.; Acutis, M.; Doro, L.; Hidy, D.; Chy, M.K.; Minet, J.; Lellei-Kovács, E.; Ma, S.; Perego, A. Multi-model simulation of soil temperature, soil water content and biomass in Euro-Mediterranean grasslands: Uncertainties and ensemble performance. *Eur. J. Agron.* **2017**, *88*, 22–40. [CrossRef]
23. Samadianfard, S.; Ghorbani, M.A.; Mohammadi, B. Forecasting soil temperature at multiple-depth with a hybrid artificial neural network model coupled-hybrid firefly optimizer algorithm. *Inf. Process. Agric.* **2018**, *5*, 465–476. [CrossRef]

24. Li, Q.; Hao, H.; Zhao, Y.; Geng, Q.; Liu, G.; Zhang, Y.; Yu, F. GANs-LSTM Model for Soil Temperature Estimation from Meteorological: A New Approach. *IEEE Access* **2020**, *8*, 59427–59443. [CrossRef]
25. 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]
26. Bayatvarkeshi, M.; Bhagat, S.K.; Mohammadi, K.; Kisi, O.; Farahani, M.; Hasani, A.; Deo, R.; Yaseen, Z.M. Modeling soil temperature using air temperature features in diverse climatic conditions with complementary machine learning models. *Comput. Electron. Agric.* **2021**, *185*, 106158. [CrossRef]
27. Tsai, Y.Z.; Hsu, K.S.; Wu, H.Y.; Lin, S.I.; Yu, H.L.; Huang, K.T.; Hu, M.C.; Hsu, S.Y. Application of random forest and ICON models combined with weather forecasts to predict soil temperature and water content in a greenhouse. *Water* **2020**, *12*, 1176. [CrossRef]
28. Recknagel, F. Applications of machine learning to ecological modelling. *Ecol. Model.* **2001**, *146*, 303–310. [CrossRef]
29. Massoud, E.C.; Hoffman, F.; Shi, Z.; Tang, J.; Alhajjar, E.; Barnes, M.; Braghiere, R.K.; Cardon, Z.; Collier, N.; Crompton, O.; et al. Perspectives on Artificial Intelligence for Predictions in Ecohydrology. *Artif. Intell. Earth Syst.* **2023**, *2*, e230005. [CrossRef]
30. Chan, W.S.; Recknagel, F.; Cao, H.; Park, H. Elucidation and short-term forecasting of microcystin concentrations in Lake Suwa (Japan) by means of artificial neural networks and evolutionary algorithms. *Water Res.* **2007**, *41*, 2247–2255. [CrossRef]
31. Li, C.; Zhang, Y.; Ren, X. Modeling Hourly Soil Temperature Using Deep BiLSTM Neural Network. *Algorithms* **2020**, *13*, 173. [CrossRef]
32. Tsai, W.P.; Feng, D.; Pan, M.; Beck, H.; Lawson, K.; Yang, Y.; Liu, J.; Shen, C. From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. *Nat. Commun.* **2020**, *12*, 5988. [CrossRef]
33. Li, X.; Zhang, L.; Wang, X.; Liang, B. Forecasting greenhouse air and soil temperatures: A multi-step time series approach employing attention-based LSTM network. *Comput. Electron. Agric.* **2024**, *217*, 108602. [CrossRef]
34. Khosravi, K.; Golkarian, A.; Barzegar, R.; Aalami, M.T.; Heddam, S.; Omidvar, E.; Keesstra, S.D.; López-Vicente, M. Multi-step ahead soil temperature forecasting at different depths based on meteorological data: Integrating resampling algorithms and machine learning models. *Pedosphere* **2023**, *33*, 479–495. [CrossRef]
35. Taki, M.; Abdanan Mehdizadeh, S.; Rohani, A.; Rahnama, M.; Rahmati-Joneidabad, M. Applied machine learning in greenhouse simulation; new application and analysis. *Inf. Process. Agric.* **2018**, *5*, 253–268. [CrossRef]
36. Zhang, K.; Liu, D.; Liu, H.; Lei, H.; Guo, F.; Xie, S.; Meng, X.; Huang, Q. Energy flux observation in a shrub ecosystem of a gully region of the Chinese Loess Plateau. *Ecohydrol. Hydrobiol.* **2022**, *22*, 323–336. [CrossRef]
37. Guo, F.; Liu, D.; Mo, S.; Huang, Q.; Ma, L.; Xie, S.; Deng, W.; Ming, G.; Fan, J. Estimation of daily evapotranspiration in gully area scrub ecosystems on Loess Plateau of China based on multisource observation data. *Ecol. Indic.* **2023**, *154*, 110671. [CrossRef]
38. Trok, J.T.; Davenport, F.V.; Barnes, E.A.; Diffenbaugh, N.S. Using Machine Learning with Partial Dependence Analysis to Investigate Coupling Between Soil Moisture and Near-Surface Temperature. *J. Geophys. Res. Atmos.* **2023**, *128*, e2022JD038365. [CrossRef]
39. Sahoo, M. Winter soil temperature and its effect on soil nitrate Status: A Support Vector Regression-based approach on the projected impacts. *Catena* **2022**, *211*, 105958. [CrossRef]
40. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386–408. [CrossRef]
41. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844.
42. Isles, P.D.F. A random forest approach to improve estimates of tributary nutrient loading. *Water Res.* **2024**, *248*, 120876. [CrossRef]
43. Han, T.; Jiang, D.; Zhao, Q.; Wang, L.; Yin, K. Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery. *Trans. Inst. Meas. Control* **2017**, *40*, 2681–2693. [CrossRef]
44. Mohanty, M.D.; Mohanty, M.N. Chapter 5—Verbal sentiment analysis and detection using recurrent neural network. In *Advanced Data Mining Tools and Methods for Social Computing*; De, S., Dey, S., Bhattacharyya, S., Bhatia, S., Eds.; Academic Press: Cambridge, MA, USA, 2022; pp. 85–106.
45. Abirami, S.; Chitra, P. Chapter Fourteen—Energy-efficient edge based real-time healthcare support system. In *Advances in Computers*; Raj, P., Evangeline, P., Eds.; Elsevier: Amsterdam, The Netherlands, 2020; Volume 117, pp. 339–368.
46. Abinaya, S.; Devi, M.K.K. Chapter 12—Enhancing crop productivity through autoencoder-based disease detection and context-aware remedy recommendation system. In *Application of Machine Learning in Agriculture*; Khan, M.A., Khan, R., Ansari, M.A., Eds.; Academic Press: Cambridge, MA, USA, 2022; pp. 239–262.
47. Rahmani, F.; Shen, C.; Oliver, S.; Lawson, K.; Appling, A. Deep learning approaches for improving prediction of daily stream temperature in data-scarce, unmonitored, and dammed basins. *Hydrol. Process.* **2021**, *35*, e14400. [CrossRef]
48. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
49. Huang, S.; Liu, Q.; Wu, Y.; Chen, M.; Yin, H.; Zhao, J. Edible Mushroom Greenhouse Environment Prediction Model Based on Attention CNN-LSTM. *Agronomy* **2024**, *14*, 473. [CrossRef]
50. Di, Y.; Gao, M.; Feng, F.; Li, Q.; Zhang, H. A New Framework for Winter Wheat Yield Prediction Integrating Deep Learning and Bayesian Optimization. *Agronomy* **2022**, *12*, 3194. [CrossRef]
51. Bai, Y.; Scott, T.A.; Min, Q. Climate change implications of soil temperature in the Mojave Desert, USA. *Front. Earth Sci.* **2014**, *8*, 302–308. [CrossRef]

52. Miralles, D.G.; Van, D.B.M.J.; Teuling, A.J.; De Jeu, R.A.M. Soil moisture-temperature coupling: A multiscale observational analysis. *Geophys. Res. Lett.* **2012**, *39*, 6. [CrossRef]
53. Zhang, T.; Huang, J.; Lei, Q.; Liang, X.; Lindsey, S.; Luo, J.; Zhu, A.; Bao, W.; Liu, H. Empirical estimation of soil temperature and its controlling factors in Australia: Implication for interaction between geographic setting and air temperature. *Catena* **2022**, *208*, 105696. [CrossRef]
54. Xu, Y.; Wang, P.; Lu, Y.; Ma, M.; Dong, G.; Tang, J. Convection-permitting regional climate simulation on soil moisture-heatwaves relationship over eastern China. *Atmos. Res.* **2024**, *301*, 107285. [CrossRef]
55. Amato, M.T.; Giménez, D. Predicting monthly near-surface soil temperature from air temperature and the leaf area index. *Agric. Forest Meteorol.* **2024**, *345*, 109838. [CrossRef]
56. Citakoglu, H. Comparison of artificial intelligence techniques for prediction of soil temperatures in Turkey. *Theor. Appl. Climatol.* **2017**, *130*, 545–556. [CrossRef]
57. Gao, S.; Wu, Q.; Zhang, Z.; Jiang, G. Simulating active layer temperature based on weather factors on the Qinghai–Tibetan Plateau using ANN and wavelet-ANN models. *Cold Reg. Sci. Technol.* **2020**, *177*, 103118. [CrossRef]
58. Gao, B.; Coon, E.T.; Thornton, P.E.; Lu, D. Improving the estimation of atmospheric water vapor pressure using interpretable long short-term memory networks. *Agric. Forest Meteorol.* **2024**, *347*, 109907. [CrossRef]
59. Bonachela, S.; Fernández, M.D.; Hernández, J.; López, J.C. Adaptation of standardised (FAO and ASCE) procedures of estimating net longwave and shortwave radiation to Mediterranean greenhouse crops. *Biosyst. Eng.* **2023**, *231*, 104–116. [CrossRef]
60. He, Z.; Wang, J.; Jiang, M.; Hu, L.; Zou, Q. Random Subsequence Forests. *Inf. Sci.* **2024**, *667*, 120478. [CrossRef]
61. Jiao, Y.; Chen, C.; Li, G.; Fu, H.; Mi, X. Research on the variation patterns and predictive models of soil temperature in a solar greenhouse. *Sol. Energy* **2024**, *270*, 112267. [CrossRef]