agriculture

MDPI

Article

Interpreting Conv-LSTM for Spatio-Temporal Soil Moisture Prediction in China

Feini Huang [1], Yongkun Zhang [1], Ye Zhang [1], Wei Shangguan [1,*], Qingliang Li [2], Lu Li [1] and Shijie Jiang [3]

[1] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences, Sun Yat-sen University, Zhuhai 519082, China; huangfn3@mail2.sysu.edu.cn (F.H.)
[2] College of Computer Science and Technology, Changchun Normal University, Changchun 130032, China
[3] Department of Computational Hydrosystems, Helmholtz Centre for Environmental Research, 04318 Leipzig, Germany
* Correspondence: shgwei@mail.sysu.edu.cn; Tel.: +86-13466654629

Abstract: Soil moisture (SM) is a key variable in Earth system science that affects various hydrological and agricultural processes. Convolutional long short-term memory (Conv-LSTM) networks are widely used deep learning models for spatio-temporal SM prediction, but they are often regarded as black boxes that lack interpretability and transparency. This study aims to interpret Conv-LSTM for spatio-temporal SM prediction in China, using the permutation importance and smooth gradient methods for global and local interpretation, respectively. The trained Conv-LSTM model achieved a high R2 of 0.92. The global interpretation revealed that precipitation and soil properties are the most important factors affecting SM prediction. Furthermore, the local interpretation showed that the seasonality of variables was more evident in the high-latitude regions, but their effects were stronger in low-latitude regions. Overall, this study provides a novel approach to enhance the trust-building for Conv-LSTM models and to demonstrate the potential of artificial intelligence-assisted Earth system modeling and understanding element prediction in the future.

Keywords: explainable artificial intelligence; deep learning; soil moisture prediction; interpretation

check for updates

Citation: Huang, F.; Zhang, Y.; Zhang, Y.; Shangguan, W.; Li, Q.; Li, L.; Jiang, S. Interpreting Conv-LSTM for Spatio-Temporal Soil Moisture Prediction in China. *Agriculture* **2023**, *13*, 971. https://doi.org/10.3390/agriculture13050971

Academic Editor: Gerard Arbat

Received: 10 March 2023
Revised: 24 April 2023
Accepted: 26 April 2023
Published: 27 April 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the development of smart agriculture, soil moisture (SM) is crucial for smart agriculture, as it affects agricultural cultivation and management. In order to improve water use efficiency, accurate soil moisture prediction is of great importance for crop water supply and management. Furthermore, climate change intensifies land–atmosphere coupling related to SM, which may compromise crop yields [1,2].

Typically, SM exhibits spatial and temporal variability due to the complex coupling between land surface (especially vegetation) and atmospheric processes and the non-linear relationship between the water cycle and thermal transformations. This poses a challenge for accurate SM prediction. In recent decades, artificial intelligence (AI) and machine learning (ML) have provided practical tools for water resources management [3]. Particularly, deep learning (DL) has emerged as one of the most advanced AI paradigms that can help capture more information hidden in datasets and is of interest to weather forecasters and climate policymakers [4]. Among DL techniques, long short-term memory (LSTM) provides a recurrent architecture and a special gate design for temporal prediction. This unique structure has enabled the successful application of LSTM for SM prediction [5,6]. On the other hand, spatial models such as convolutional neural networks (CNNs) have advantages in extracting the spatial distribution of hydrometeorological elements [7] and have been successfully used to improve spatial prediction performance [8,9]. Considering the advantages of LSTM and CNN, convolutional LSTM (Conv-LSTM) can describe temporal variations and extract spatial features simultaneously [10], which has been demonstrated

superior forecasting performance with respect to hydrometeorological variables, such as SM [11], air temperature [12], wind speed [13] and evapotranspiration [14].

However, for pure AI models, there remains a trade-off between model accuracy and interpretability [15,16]. Interpretability refers to a passive characteristic of a model referring to the level at which a given model makes sense for a human observer [17]. Enhancing interpretability means that we can extract more relevant information from an AI model regarding relationships either contained in the data or learned by the model [15]. This "black-box nature" hinders the further application of DL models in Earth system science. Understanding the rationale behind the decisions of DL models is fundamental to increasing our confidence in their use and building a system of trust for the demands of the regulatory environment [18,19]. To address this problem, explainable artificial intelligence (XAI) techniques provide a practical approach to peeking into the black box and understanding its internal logic [20], which may facilitate DL prediction in Earth system science.

In recent years, XAI has been applied to various agriculture-related predictions such as hailstorms [21], precipitation [22], streamflow [23], SM [24], and droughts [25]. These techniques aim to make AI models more transparent and interpretable for different applications. The concepts, methods, and risks of different XAI techniques have been discussed in Earth sciences, especially in the field of meteorology [26]. XAI methods can be grouped along different axes: ante-hoc vs. post-hoc, global vs. local, and model-specific vs. model-agnostic [15,27]. Table S1 provides an overview of these categories and Table S2 presents a flowchart to help select a type of interpretation method based on the intended users (referred to [28]. For SM prediction in agriculture, the needs and expectations of the intended user should be considered when applying XAI.

This study aims to investigate global and local post-hoc approaches to XAI techniques for SM prediction. Global approaches, such as permeability importance (PI) [29], partial dependency diagrams (PDP) [30], and aggregated Shapley's additional interpretation (SHAP) [31] can provide a conformity assessment by extracting the sensitivity of variables from a trained model while disrupting the input data. PI is recognized as the most practical of these global explanatory techniques due to its low computational cost and reliable principle. On the other hand, local approaches can help forecasters evaluate how SM prediction is influenced by different input features at specific locations and times. For Conv-LSTM models, gradient-based methods such as Saliency map (SA) [32], integrated gradients (GI) [33], gradient input (GI) [34], square gradient (SqG) [35], VarGrad (VG) [36] and smooth gradient (SG) [37], can compute the gradient of the prediction with respect to the input features, which offers the interpretation.

In this study, we interpret a Conv-LSTM model for spatio-temporal SM prediction in China and explore how it can build trust among policymakers and users. We compare a Conv-LSTM model using meteorological variables and static variables (CL-S), with a Conv-LSTM model without static variables (CL). We interpret the CL-S globally by using PI and locally by using gradient-based methods. We analyzed the evolution of spatial and temporal information hidden in the network and discover how they predict SM. We aim to address two research questions: (i) How to offer interpretability for highly complex DL models with high accuracy using XAI technology in SM prediction? (ii) What interpretations can be offered for SM prediction of the Conv-LSTM model? Overall, this study exemplifies the potential of using XAI to enhance DL prediction and supports the expectation that DL may eventually lead to a fundamental change in practical approaches in agricultural management.

2. Materials and Methods

2.1. Study Area

China, located at 3–54° N, 72–126° W, was restricted as the study domain, which covers about 96 million km^2. Geographically, China is in central and eastern Asia on the west coast of the Pacific Ocean, and it is affected by the monsoon climate and has

important monsoon climate characteristics, which is recognized as the region of strong land–atmosphere coupling [38]. In order to explore the spatial and temporal patterns of SM throughout China, we followed [39] to divide China into six regions based on elevations, rainfall, topography, and hydrogeology (Figure 1). These regions are the northeast monsoon region (NEM), the north China monsoon region (NCM), the south China monsoon region (SCM), the southwest humid region (SWH), the northwest arid region (NWA), and the Qinghai–Tibet Plateau region (QTP). Figure 1 shows the annual SM in these regions from 2016 to 2018. Most of the NCM and NWA were in an arid state with annual SM below 0.3 m^3/m^3. By contrast, the SWH and SCM were relatively wet with annual SM > 0.4 m^3/m^3 throughout the examined years. Spatial SM variability at QTP was high with annual SM varying from 0 to a maximum of around 0.7 m^3/m^3.

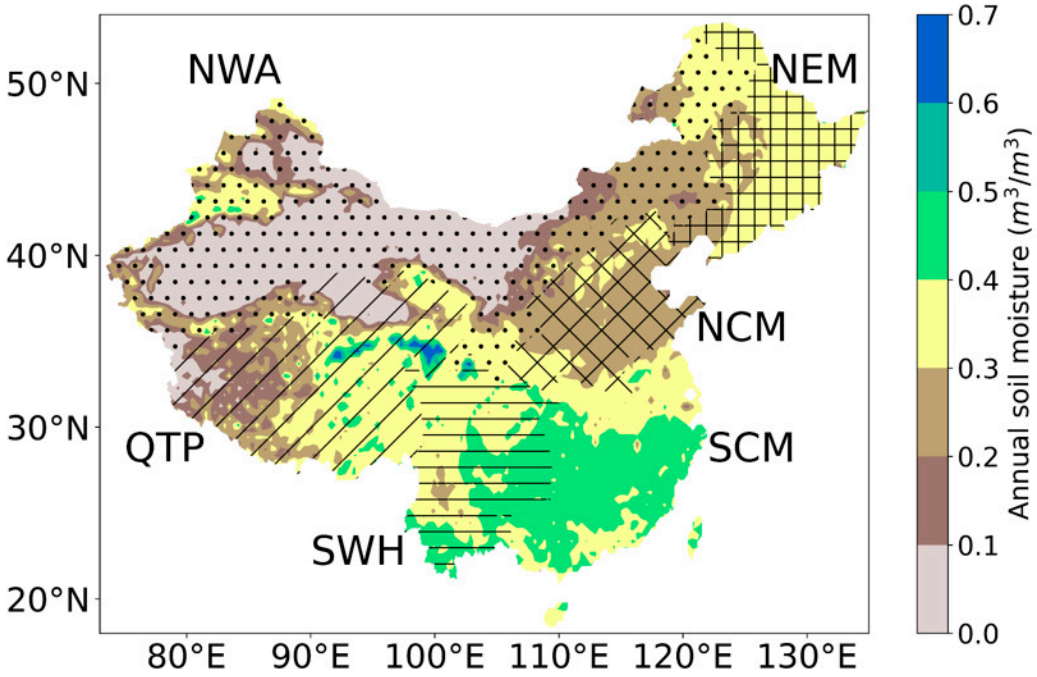


Figure 1. Annual soil moisture in six geographic-climatic regions of China. The base map is derived from the Resource and Environment Science and Data Center of the Chinese Academy of Sciences (http://www.resdc.cn/ (accessed on 26 March 2023)).

2.2. Data Source

In this study, we used ERA5-Land as the source dataset for establishing a model to estimate SM in China. ERA5-land is a state-of-the-art reanalysis dataset that provides high-resolution SM data along with other meteorological variables [40]. We selected ERA5-Land because it has several advantages over other products: it is more refined, it has a longer time span, and it combines multiple observations based on physical laws. Previous studies have reported that ERA5-Land has higher accuracy than most other SM datasets from model output, remote sensing, and reanalysis [41]. We used six input meteorological variables for our model: total precipitation (P), 10 m U-wind component (U), 10 m V-wind component (V), 2 m temperature (TA), net surface solar radiation (SR), and net surface thermal radiation (TR), and they were obtained from the ERA5-Land (9 km). These variables are effective because they influence SM dynamics through natural processes such as evaporation, transpiration, infiltration, and runoff. The raw data period for ERA5-Land used in this study is from 1 January 2016 to 31 December 2018 (1096 daily samples), and the spatial resolution is 9 km. The training target is the surface SM (0–7 cm). ERA5-land dataset is split into two parts: the training and validating sets and the testing set. The training and validating sets are from 1 January 2016 to 30 November 2017 and the testing set is from 1 December 2017 to 31 December 2018.

To account for the spatial heterogeneity of land surface, we used several static variables as additional inputs for our model. These variables were: soil properties (sand (SAND), silt (SILT), clay (CLAY) content, and bulk density (BULK)) extracted from the China soil dataset for land surface modeling [42], land cover type (LAND) extracted from the United States geological survey [43] and digital elevation model (DEM) [44]. All data are interpolated to the lower resolution of 36 km (102 × 167 grid points) to match our target domain.

2.3. Soil Moisture Prediction with Convolution Long Short-Term Memory

We used a (DL model called Conv-LSTM to extract features from ERA5-land data and predict SM. Conv-LSTM combines CNN and LSTM to extract the temporal-spatial information from time-series graphical data and achieves high performance. Unlike LSTM, Conv-LSTM preserves the spatial information and dimension of the input data by replacing matrix multiplication with convolution operations in its recurrent layer. Therefore, the data that flows through the Conv-LSTM cells retain the input dimension and preserve all the spatial information. The key idea of Conv-LSTM is to apply a convolution operator at every gate in state-to-state and input-to-state transitions so that the future state of each cell depends on the inputs and past state of its local neighbors [10]. The equations in the Conv-LSTM cell are as follows:

$$\begin{gathered} i_t = \sigma(W_{xi} * \chi_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \\ f_t = \sigma\Big(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f\Big) \\ \hat{C}_t = tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \\ C_t = f_t \circ C_{t-1} + i_t \circ \hat{C}_t \\ o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \\ H_t = o_t \circ tanh(C_t) \end{gathered} \tag{1}$$

where χ_t donates cell input at t moment, C_t donates cell output, H_t donates hidden state. W are the convolutional kernels. i, f, o are the input, forget, and output gates of the Conv-LSTM cell, respectively. Additionally, "$*$" is the convolution operation and "$\circ$" is the Hadamard product, and "σ" is the sigmoid function.

2.4. Assessment of Model Performance

We applied 10-fold cross-validation to tune the hyper-parameters and avoid overfitting, using a validating set. We then evaluated the trained models on a test set with multiple metrics: explained variance (or determination coefficient, R^2 mean absolute error (MAE) and root mean squared error ($RMSE$). These statistical metrics are calculated as follows:

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

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{4}$$

where y_i is the measured value of the i-th sample, $\hat{y}_i$ is the corresponding predicted value, n is the sample size, and $\overline{y}$ is the average value of the measured samples. R^2 ranges from $-\infty$ to 1, where 1 depicts an exact match. A Low $RMSE$ or MAE means the model performs well.

2.5. Model Interpretation Techniques

2.5.1. Permutation Importance

Permutation importance (PI) is a straightforward approach to estimating variable importance. It measures how much a feature x_j contributes to the performance of a trained model on the target y, by randomly permuting x_j over all of the examples and

comparing the model performance on the original and permuted data. The larger the drop in performance, the more important the feature is. *PI* was first introduced by [45] for random forests (RF) and later generalized by [29]. The formula for *PI* is:

$$PI_j = loss\left(\hat{f}\left(x_j^*\right), y\right) - l\left(\hat{f}(x), y\right) \quad (5)$$

where x_j^* is created by permutation of the values in the j-th feature, x is the un-permuted input feature matrix, y is the observation dataset, $\hat{f}$ donates the prediction function and *loss* donates the error function using the values of *RMSE*.

2.5.2. Smooth Gradient

To interpret DL models locally, gradient-based techniques are adopted. Generally, the gradient $\frac{\partial F(x)}{\partial x}$ in DL models at a given point can vary rapidly and cause misinterpretations of the model behavior. To address this issue, a new gradient-based approach that smooths the gradient with a Gaussian kernel was presented [37]. The smooth gradient (SG) method works as follows: first, some random noise is added to the input image, second, the pixel attribution is obtained by a saliency map of the noisy image, and third, the heatmap is averaged over multiple noisy images. The equation for SG is:

$$SG(x) = \frac{1}{n}\sum_1^n \frac{\partial F\left(x + \chi\left(0, \sigma^2\right)\right)}{\partial x} \quad (6)$$

where x is the input in a neighborhood, n is the number of samples, $\chi\left(0, \sigma^2\right)$ is Gaussian noise with standard deviation σ.

2.6. Experimental Design

In this study, we aim to interpret a deep-learning model for soil moisture prediction in China. We collected data on meteorological variables, static variables, and SM. To investigate the effects of static variables on SM prediction, we built two deep-learning models: CL and CL-S. CL was a Conv-LSTM with only six dynamic attributes. CL-S was similar to CL but also included six static attributes. We compared their prediction performance and explained CL-S using the SG technique.

Figure 2 shows our network, consisting of four Conv-LSTM layers with a processing procedure of batch normalization and dropout regularization, and ending with a three-dimensional convolution (Conv3D) layer. Batch normalization helps speed up the optimization process and allows for higher learning rates t by normalizing the internal values of the neural network [46]. Dropout regularization [47], which randomly sets some inputs to zero during training, reduces overfitting by promoting independence among the weights. We reduced the number of convolutional filters by half in each successive Conv-LSTM layer to avoid excessive complexity. We tuned our hyperparameters using a random search tool named Keras tuner [48] that generalized well to unseen testing data. To avoid overfitting, the validating set is randomly selected from 20% of the training set. The initial number of filters varied between 128 and 512. The kernel sizes were 7, 5, 3, and 1, respectively, for each Conv-LSTM layer, respectively. In the last Conv3D layer, the number of filters was set to 1 and the kernel size was set to cubic 3. The dropout rates were 25%. The ReLU activation function was selected because it is simple and robust. The Adam [49] optimizers were selected with a learning rate of 0.001. All networks were trained for 125 epochs with a batch size of 5 examples. These hyperparameter settings were not exhaustive, and further tuning might yield better results.

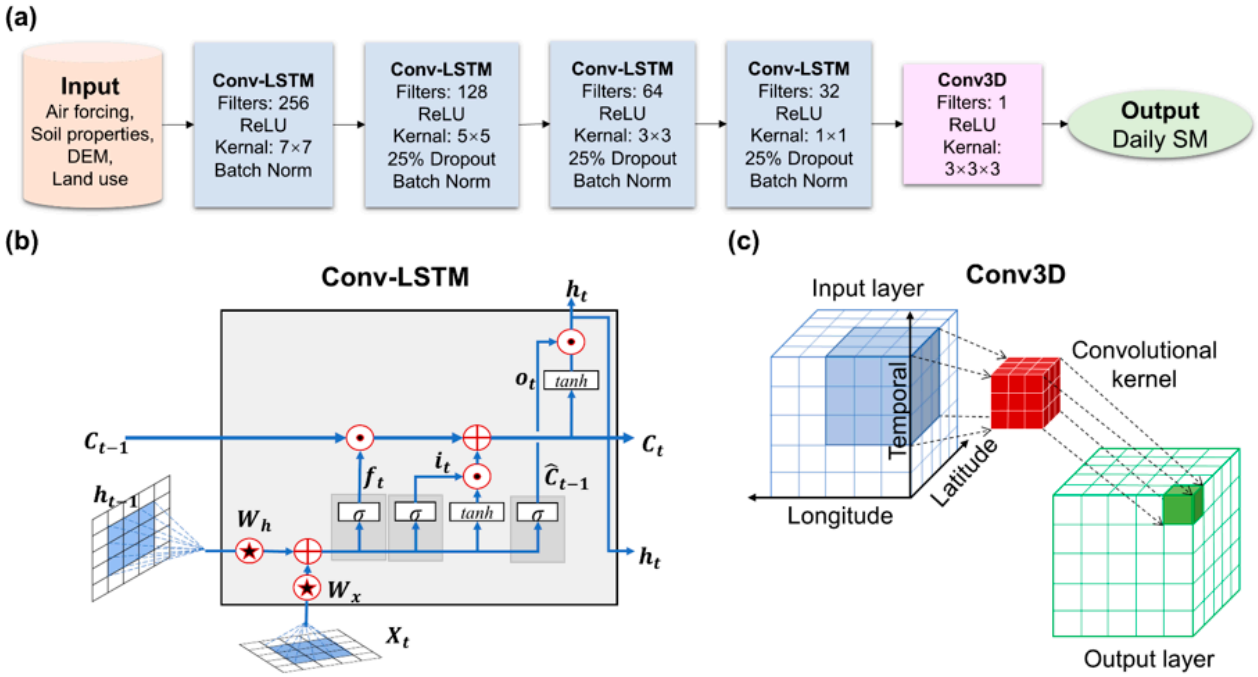


Figure 2. Convolutional long short-term memory (Conv-LSTM) network for soil moisture prediction. (**a**) schematic network setup. The network architecture and hyperparameters yield the best validation scores. (**b**) The internal structure of a Conv-LSTM cell. (**c**) A diagram of how 3D convolution operates on the input data.

The XAI techniques were implemented to gain insight into how our models work and what features they use to make predictions. We implemented PI as a global interpretation method to measure the importance of each input variable for the model output. We also applied SG to interpret CL and CL-S locally. We compared SG with other gradient-based methods including IG, GI, SA, SqG, and VG (see Supplementary material). The details and results of these methods are shown in Supplementary material. We selected SG as our preferred interpretation method because it showed physical consistency in several examples (See Section 3.2.2). We implemented CL and CL-S using Keras backend on a GPU (Tesla T4) with RAM: 128 GB running on an Anaconda platform. CUDA technology (Version 11.4) was used to accelerate the computation.

3. Results

3.1. Model Performance

Figure 3 shows the density scatter plots of the predicted SM and observed SM to evaluate model performances using three metrics: R^2, MAE, and RMSE. CL-S had better performance than CL, with R^2 at 0.92, MAE at 0.028 m^3/m^3, and RMSE at 0.040 m^3/m^3. These results indicate that including both meteorological forcing and static variables improved the accuracy of SM prediction. By contrast, CL had a worse performance with R^2 at 0.84, MAE at 0.039 m^3/m^3, and RMSE at 0.055 m^3/m^3. In Figure 3a, many dots were far from the ideal line (y = x), especially those with high SM values, which indicates that CL was not capable of predicting high SM values without the static variables. Therefore, we chose CL-S as our final model for further interpretation in the following sections because it performed well and provided a sound basis for XAI's interpretation.

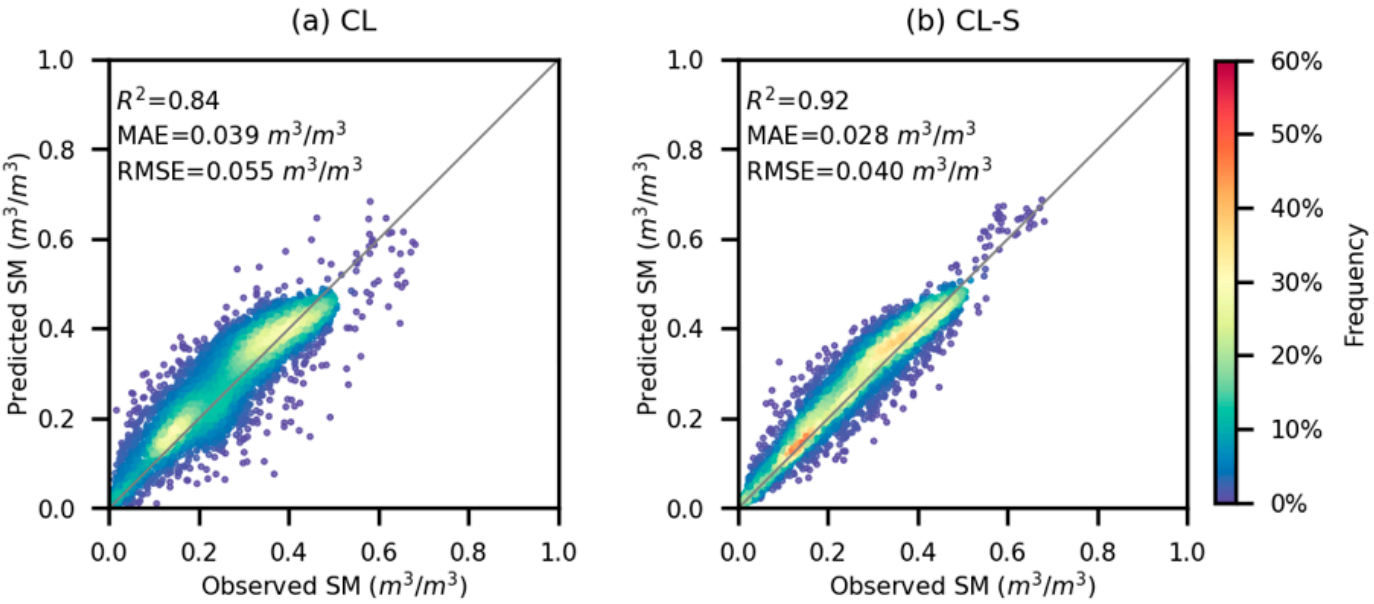


Figure 3. Density scatter plot of the observed and predicted SM. (**a**) Conv-LSTM with only dynamic attributes (CL), (**b**) Conv-LSTM with dynamic and static attributes (CL-S). The grey line is the 1:1 line indicating perfect agreement between observed and predicted values. The red line is the regression line showing the best fit for each model.

3.2. Model Interpretation

3.2.1. Global Interpretation by Permutation Importance

Figure 4 shows the daily distribution of variable importance obtained by averaging the PI for CL-S in all grids during the training period. The PI is a measure of how important the variable is to the model. Among the variables used in CL-S, precipitation (P), which is the most crucial meteorological forcing variable, had the highest importance in predicting SM (median > 20%). Other meteorological variables, such as TR and TA, also had a significant effect on predicting SM. The importance of TR ranged from about 7% to 20%, indicating a great contribution to the seasonal variability of a year. The overall SR importance was less than P, TR, and TA but with more outliers. As expected, wind speed components U and V had the least influence on SM. For the static variables, soil properties including BULK, CLAY, and SILT ranked behind the P. This suggests that SM was closely related to soil properties. However, the SAND presented had less impact on predicting SM than other soil properties. DEM was one of the factors that affected SM distribution, but it was not a key variable in this case. Similarly, LAND had a low contribution to SM modeling.

3.2.2. Local Interpretation by Smooth Gradient

We used gradient-based methods to locally interpret CL-S and verify the model correctly learns the variables' effects. Here, we focused on the SG interpretation, which was the most suitable for CL-S (see Supplementary material for a comparison of different methods). Figure 5 shows how gradients change with variable values for each grid and each time interval. The absolute slope of the regression line indicates how fast the predicted SM changes as a variable change. A positive slope means that as the variable increases, the predicted SM increases, while a negative slope means the opposite. The variable values were standardized to better assess their relationships regardless of their magnitudes. P had the most significant positive effect (y = 2.8504x, Figure 5f) causing a rapid increase in the predicted SM as it increased. Other meteorological variables such as TR (Figure 5e) and TA (Figure 5c) also showed significant impacts on SM prediction and their high contribution matched well with the results of PI. In Figure 5c, where higher TA values and higher SM appeared simultaneously, the negative gradients declined dramatically. This indicates that as TA increases, the predicted SM was more likely to decrease in these environments. On

the contrary, this change slowed down in the hot and dry environments. The effect of the TR was similar to that of the TA but differed in the direction and magnitude of the gradients. Wind speed (U and V) and SR had weak negative effects on SM prediction (Figure 5a,b,d). Among static variables, the BULK (Figure 5h) and CLAY (Figure 5i) had stronger positive effects than other variables. The positive relationship between BULK and SM was more evident in wetter environments with lower BULK values (the gradients were around 0.5). However, in most of the arid areas, the gradients of BULK were near zero, which means that BULK had little effect on SM prediction. This effect was also found in other static variables (i.e., LAND, CLAY, SAND, SILT, and DEM). These results demonstrate that SG can capture the local effects of variables on SM prediction.

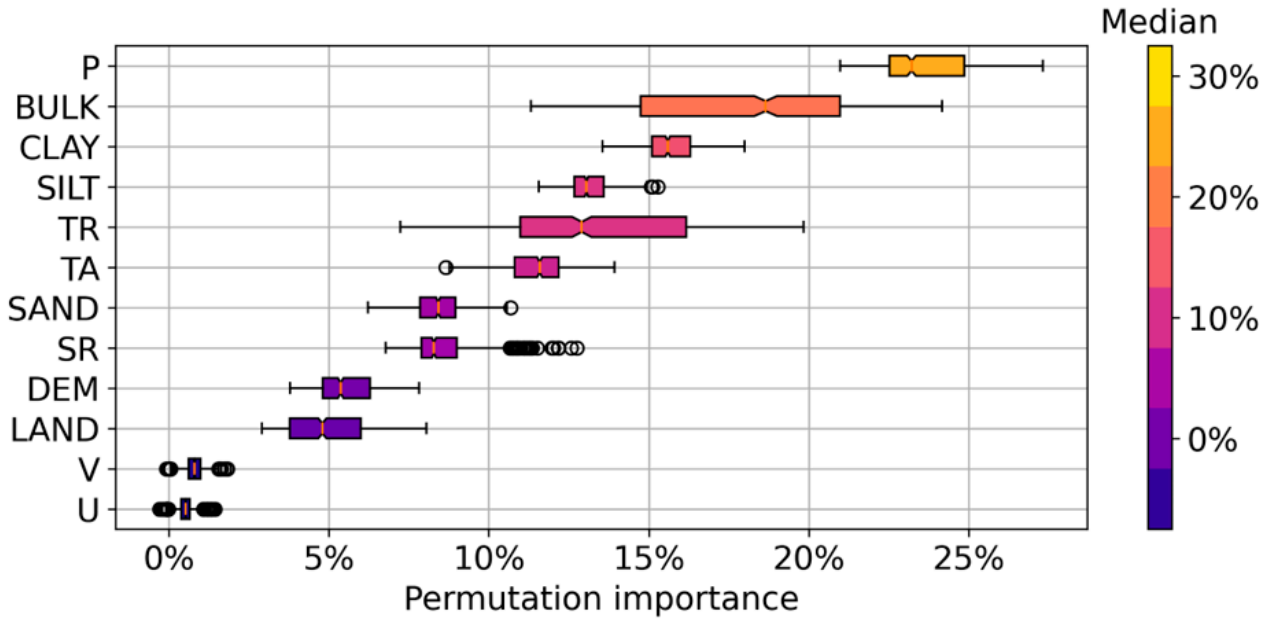


Figure 4. Effect of variable permutation on CL-S model performance. The figure shows the variable permutation importance in terms of the decrease in RMSE for the CL-S model in the training set. The whiskers indicate the 5% and 95% percentiles of the distribution of the daily mean decrease of RMSE. The dynamic variables are total precipitation (P), 10 m U-wind component (U), 10 m V-wind component (V), 2 m temperature (TA), net surface solar radiation (SR), and net surface thermal radiation (TR). The static parameters of the land surface are sand (SAND), silt (SILT), clay (CLAY) content, bulk density (BULK), land cover type (LAND), and digital elevation model (DEM).

In this study, we discussed the importance of the meteorological variables (P, TR, and TA) in terms of both time and space, respectively. Figure 6 shows how the gradients of important variables (P, TR, and TA) vary over time for the test year obtained by SG. We divided China into six regions based on their climatic characteristics: NWA, NCM, NWQ, QTP, SWH, and SCM (see Figure 1). P and TR had positive effects on SM prediction in all regions, while TA had negative effects. The gradients of P (Figure 6a–f) were higher in wetter regions (SWH, SCM, and NEM) than in drier regions (NWA, NCM, and NWQ), suggesting that P was more important for SM prediction in humid areas. The gradients of P also showed a seasonal pattern in higher latitudes (NWA and NEM), with lower values in winter and spring and higher values in summer and autumn. We also found regional and seasonal differences in the gradients of TR (Figure 6g–l). SWH and SCM had larger gradients than other regions. All regions exhibited a clear seasonal variation with higher values in summer than in autumn. The patterns of the northern regions (NWA, NEM, and NCM) and QTP showed a peak rapid rise in summer. By contrast, the TR gradients of SWH and SCM had less seasonal variation. TA played a more important role in SM prediction in SWH and SCM than in other regions (Figure 6m–r). It varied seasonally in NWA, NEM, and QTP with more negative in summer than in autumn. Overall, P, TR, and TA had stronger

effects on SM from drier regions to wetter regions, but their seasonal impacts were more evident in drier regions.

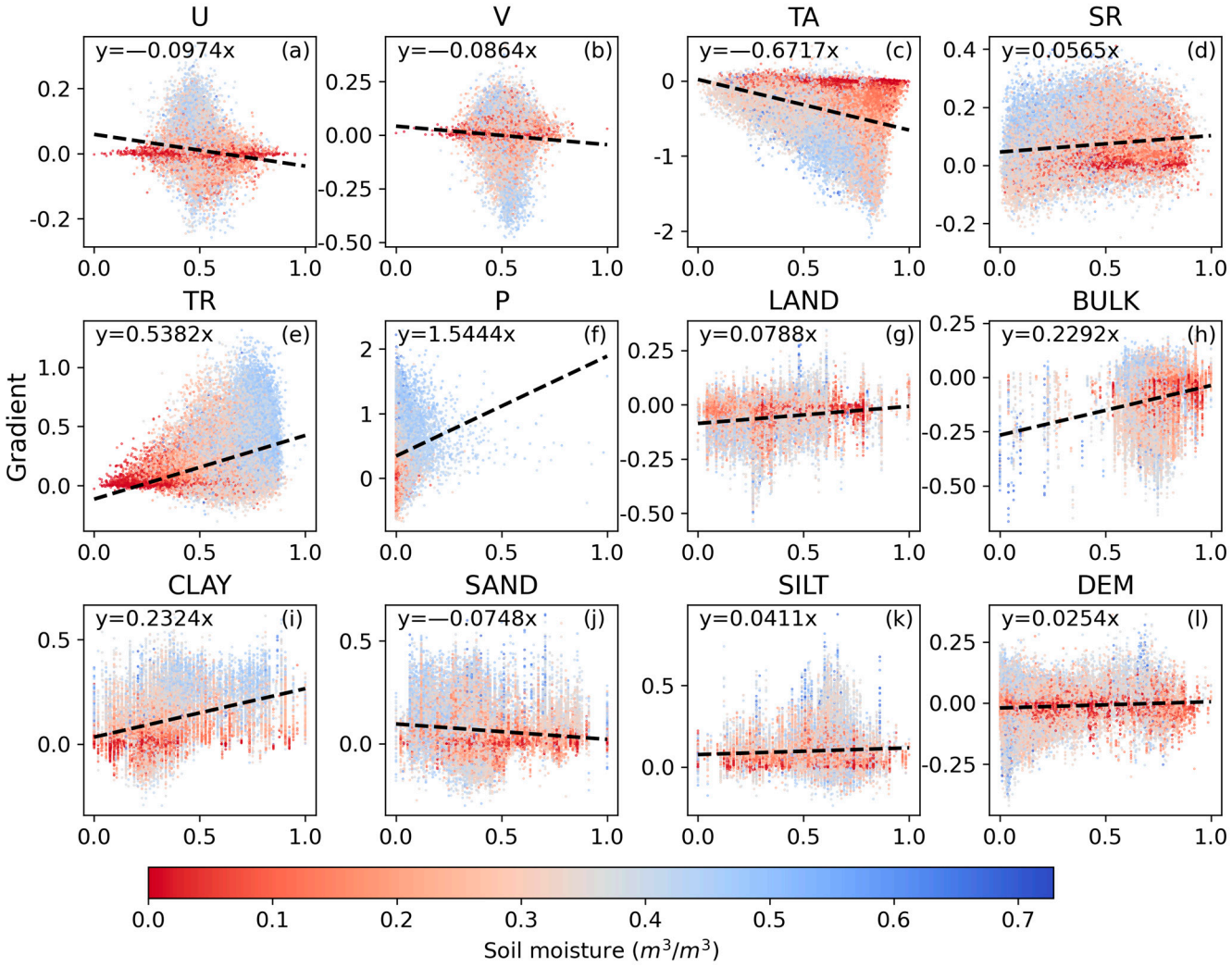


Figure 5. Relationship between gradients and standardized variables for soil moisture prediction. The scatterplot shows the gradients plotted against 12 variables for each plot in the testing set. The variables were standardized to range from 0 to 1. The dashed line represents the linear regression for each variable.

We analyzed the seasonal spatial gradients of P, TR, and TA to understand their spatial variability on SM prediction. Figure 7 shows the pixel-level trends of these gradients. P had positive gradients in most of the areas except for several regions of the arid west (Figure 7a–d). P gradients were comparatively stable in NWA and NEM (0 to 0.4), but significantly more in SWH and SCM. P effects decreased from winter to spring but increased in summer and autumn. In summer, high P effects extended into the interior of China. This suggests that for the same magnitude of P increase, the SM prediction was more likely to increase in southern regions than that in northern regions in dry seasons (autumn and winter). TR always had positive effects on SM prediction except in several high-latitude regions (Figure 7e–h). TR effects were more seasonal in the south than in the north. In the south, TR effects peaked in spring and reached the smallest in autumn. TA had a negative effect and was more pronounced than TR. The effect enhanced in most parts of the country from winter to spring and reached its peak in China's South-East coastal areas in spring (Figure 7j,k). The effect was then retarded from summer to autumn. Interestingly, some western regions had a positive TA impact, which was larger in cold environments and

smaller in warm ones. Overall, the spatial variability of P and TA's effect was much greater than that of TR and all of them were more distinct in the south.

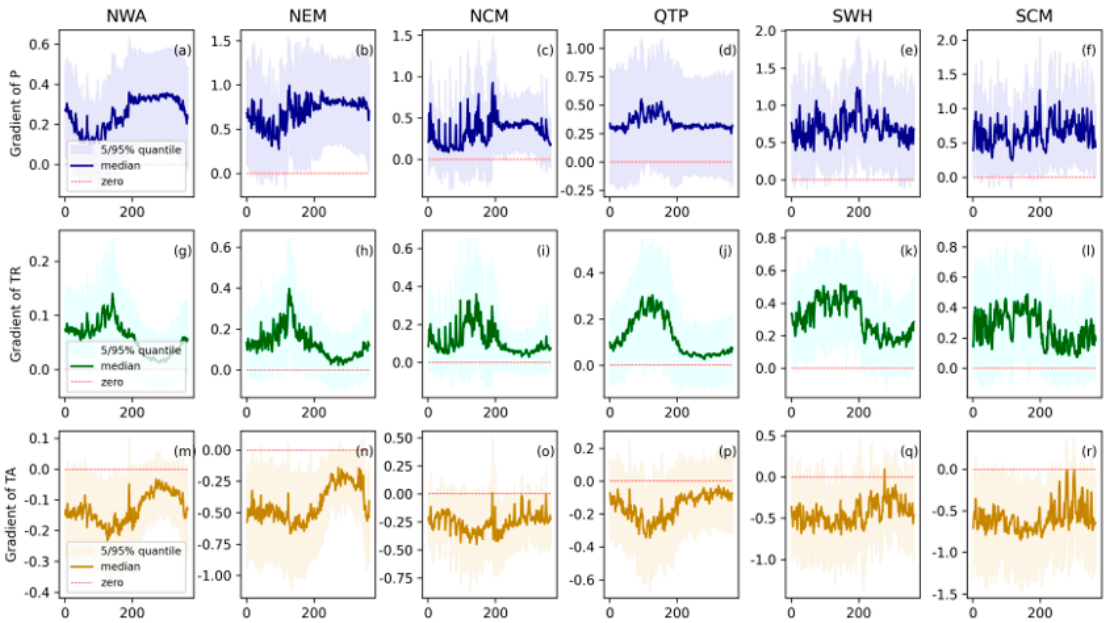


Figure 6. Temporal variation of dynamic variables' gradient across different regions. The figure shows the gradients of P (**a**–**f**), TR (**g**–**l**), and TA (**m**–**r**) for each day of the testing year in six regions (from the first column to the six column): NWA, NEM, NCM, QTP, SWH, and SCM. The rows represent the variables, and the columns represent the regions.

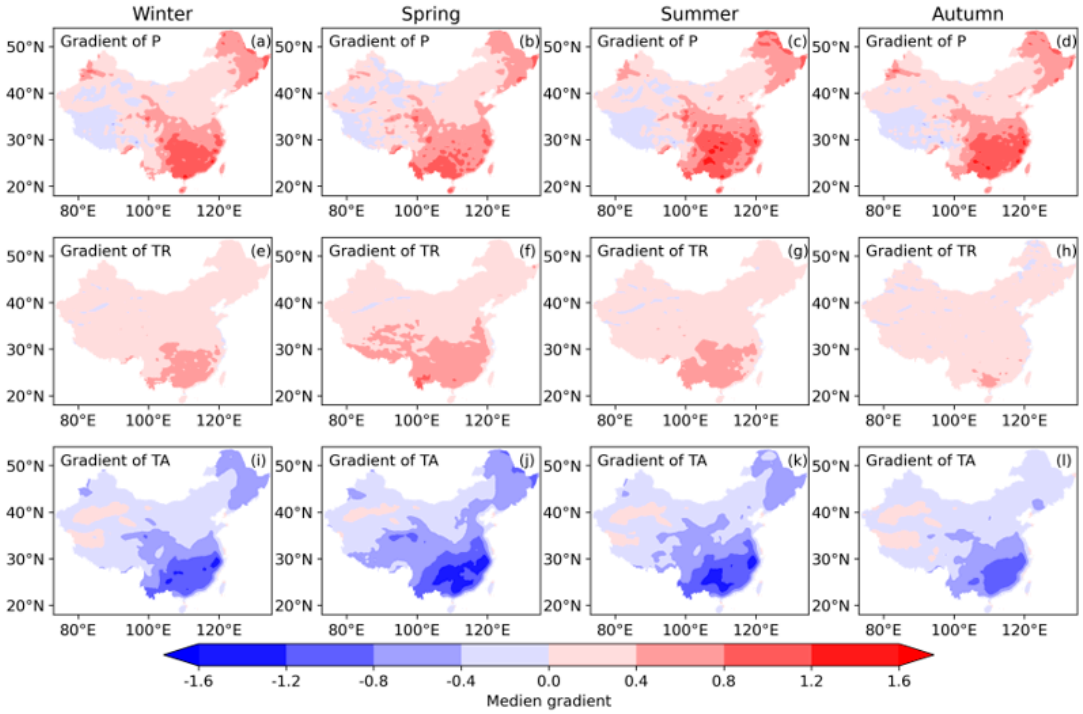


Figure 7. Spatial distribution of gradient medians of P (**a**–**d**), TR (**e**–**h**), and TA (**i**–**l**) for four seasons. Each row shows a different variable: P (**top**), TR (**middle**), and TA (**bottom**). Each column shows a different season: winter (**left**), spring (second from left), summer (third from left), and autumn (**right**). Winter is from 1 December 2017 to 28 February 2018; the other seasons are from 1 March to 31 May, 1 June to 31 August, and 1 September to 30 November, respectively.

4. Discussion

Our study demonstrated that DL is a promising technique for SM prediction in agricultural fields. We found that adding static variables to the Conv-LSTM model improved its performance and achieved an R^2 of 0.92 on the testing set, which was higher than the Conv-LSTM model without static variables. This result indicates that static variables such as elevation, soil properties, and land cover type, have significant effects on SM variation and should be considered in SM modeling. To verify the advantage of DL, we trained two random forest (RF) models as baselines: one that included only meteorological variables and one that included both meteorological and static variables (see supplementary material for details). Both RF models performed worse than the DL models in terms of R^2, MAE, and RMSE (Figure S1), indicating that DL can better capture the spatio-temporal features of SM dynamics than traditional machine learning methods. These findings are consistent with previous studies that applied DL for meteorological forecasting and intelligent agricultural management [50,51]. Our results have important implications for improving irrigation efficiency and water conservation in agriculture. By using DL models to predict SM at high spatial resolution, farmers can optimize their irrigation schedules based on the actual soil water status rather than empirical rules or fixed intervals. This can reduce water waste, enhance crop yield, and mitigate environmental impacts such as runoff, erosion, and nutrient leaching. Furthermore, our DL models can also provide valuable information for hydrological modeling, drought monitoring, climate change assessment, and land surface feedback analysis.

Although the DL is often criticized for its lack of model transparency in the agricultural community, XAI is capable of addressing the challenge of model transparency in DL. The conventional view suggests that there is an irreconcilable and unavoidable conflict between the model's predictive accuracy and the possibility of understanding its behaviors: this is known as the accuracy-interpretability trade-off [16]. However, we show that this trade-off can be overcome by using XAI techniques. We applied PI and SG as global and local interpretation methods, respectively, to explain the model's behavior and outputs. Using PI, we found that P was the most influential variable, followed by soil properties (BULK, CLAY, SILT) and meteorological variables (TR, TA). These results indicate that SM dynamics are largely governed by precipitation patterns and soil characteristics in China. We found that the gradients of predictors changed with their values and SM values using the SG method (Figure 5). For example, as P increased, the SM would maintain an upward trend fiercely especially where the soil was in a mild state (SM at 0.3~0.5 m^3/m^3). We also used the SG method to examine the temporal and spatial patterns of gradients for three important predictors: P, TR, and TA (Figures 6 and 7). We observed that these variables had different effects on SM prediction depending on the seasonality and latitude of regions. The seasonality of gradients was more explicit in the high-latitude regions. For example, P had a positive impact on SM prediction in most regions except for arid areas in northwest China and the Qinghai–Tibet plateau where snow cover was prevalent. Moreover, the effects of TR and TA were of seasonal characteristics, which were amplified in spring from the south to the north. This is a likely result of seasonal monsoon. These XAI tools helped us understand how the Conv-LSTM model captured the spatio-temporal patterns of SM and what factors influenced its predictions. This study demonstrated how XAI tools can enhance the credibility of applying DL techniques for environmental modeling by peeking insight into the black-box model, which is often criticized for its lack of interpretability. By doing so, the PI and SG were employed to offer the interpretability of the DL model. Specifically, the PI provided the feature importance of the DL model, and the SG identified the pixel attribution of the model.

In addition to the dynamic variables, we also investigated the effects of static variables on SM prediction using XAI tools. As shown in Figure 4, static variables had an inevitable effect on the SM prediction in the global interpretation of PI. To further explore this phenomenon, we analyzed it with the SG method and found that the static variables often affected SM prediction in wet environments (Figure 5). The result was similar to the

streamflow prediction [23]. For example, in Figure 7h, the negative gradients of BULK were found where the soil was moist. We also examined their spatial distribution over time. The medians of the four seasons in the test year were shown in Figure S2. From the seasonal variability in space, we found that the effects of the static variables with lower importance (i.e., LAND and SAND) did not change much over time. However, the significant variables such as BULK, CLAY, and SILT changed more dramatically with time. Figure S3 shows the temporal gradient of static variables. The negative effects of BULK showed great seasonality in the six regions of China. The variation of CLAY and SILT's effects was not relatively significant and not different from the regions. Similar to the dynamic variables, the seasonality of static variables was more evident in the high-latitude regions and the effects were stronger in low-latitude regions. Thus, SG reflected both spatial and temporal information of static variables for SM prediction.

With the urgent demand for interpretable DL in agriculture, XAI becomes one of the research hotspots in recent years and an achievable approach to promote the DL advances. However, there is a lack of attention by the agricultural community on how to select feasible and effective XAI tools. Different XAI tools may provide different interpretations of the same DL model, which can affect the decision-making process. The feasibility of different XAI tools was discussed in the Earth system but did not propose an evaluation system for XAI [26]. In this paper, we applied six model-agnostic gradient-based methods to interpret a DL model for SM prediction. We compared these methods based on their ability to capture the relationship between predictors and SM. The details of these methods are presented in the Supplementary material. To illustrate our approach, we showed a random sample from the testing year in Figure S4 which include P, SM, and differential SM (diff SM). The interpretation obtained by six gradient-based methods is shown in Figure S5. We found that only the SG method could provide reasonable interpretations that corresponded to P and differential SM comprehensively. Hence, we argued that the SG might be the most suitable XAI technique for SM prediction using Conv-LSTM models in our case.

The main contribution of this study is that we presented an interpretation strategy of DL models for SM prediction, which is of high precision and considerable interpretability. Overall, in order to reply to the two scientific questions in Section 1, we summarized our main findings in this study: (a) The XAI techniques can offer the interpretability of complex DL models which provide comprehensive interpretations to answer why the DL would make the decision on the prediction. In this study, A spatio-temporal Conv-LSTM prediction model was proposed, which was interpreted both globally and locally; (b) The interpretation obtained by XAI in this study included feature importance from PI and pixel attribution from SG. We found that SG was a relatively appropriate method as it could properly reveal the spatial and temporal information hidden in the networks. Our study contributes to the advancement of explainable artificial intelligence in agriculture by providing a novel approach to enhance the interpretability and trustworthiness of DL models for SM prediction.

However, our study also has some limitations that need to be addressed in the future. First, we did not fully exploit the spatio-temporal information hidden in the Conv-LSTM, which could have a huge potential to improve the model's performance and behavior. We suggest that future studies should explore more effective ways to extract and utilize this information. Second, to date, DL models have been interpreted quantitatively and confirmed by human prior knowledge. The interpretation should be extremely understandable and verifiable [52,53]. We argue that future studies should develop more objective and quantitative metrics to assess the interpretability and physical consistency of the model. Third, we did not consider the effects of human activities (e.g., irrigation) on soil moisture, which could be significant in some regions or seasons. However, these effects are difficult to measure and incorporate into the model. We recommend that future studies should investigate how to account for these factors in soil moisture prediction and interpretation. Finally, this study only tested one type of deep learning model (Conv-LSTM), one dataset

(ERA5-land), and one region (China). Future research could test our framework on different models, data sources, regions, or variables to test the generalizability and applicability.

5. Conclusions

This study proposed an interpretable Conv-LSTM model for spatio-temporal soil moisture prediction in China. PI and SG were utilized to offer the global (feature importance) and local (pixel importance) interpretation, respectively. This allows us to extract the reasons behind the model's decision for the entire model and individual prediction. We found that our model achieved high accuracy and interpretability simultaneously. Our findings include several aspects. First, incorporating static variables (i.e., land cover, soil properties, and elevation) can enhance the model's accuracy. The Conv-LSTM with the static variables (R^2 = 0.92) performed better than the model without them (R^2 = 0.84). Second, the complexity of the DL model can be effectively interpreted by PI and SG methods. In terms of global interpretation, PI revealed that precipitation and soil properties were the most important factors for soil moisture prediction. In terms of local interpretation, the spatial and temporal gradients obtained by SG were reasonable and consistent with physical knowledge, demonstrating a certain rationality of spatial distribution and periodicity. The seasonality of variables was more evident in the high-latitude regions, but their effects were stronger in low-latitude regions. These interpretations enhanced the trustworthiness and transparency of the model for potential users and meet the requirements from a regulatory context. The interpretation obtained from SG also satisfied the criteria of model transparency, interpretation stability, and physical consistency. Overall, this study demonstrates the promising potential of using XAI to interpret DL models to overcome the classic trade-off between model accuracy and interpretability and provide a new perspective on applying more complex DL models for spatio-temporal agricultural element predictions.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agriculture13050971/s1, Figure S1: Density scatter plot of the observed SM versus predicted SM by: (a) random forest with only dynamic attributes (RF), (b) random forest with dynamic and static attributes (RF-S); Figure S2: Median spatial gradients of static variables for different seasons; Figure S3: Temporal variation of Static variables' gradient for each day of the testing year in six regions; Figure S4: A randomly selected example for soil moisture (SM) prediction: (a) precipitation, (b) predicted SM and (c) differential SM (diff SM); Figure S5: Interpretation of the example obtained by six gradient-based methods; Table S1: Overview of categories of interpretation methods; Table S2: Flowchart to help select a type of interpretation method based on the intended user; Table S3: Hyper-parameter settings for grid search in RF models. References [54–57] are cited in the supplementary materials.

Author Contributions: Conceptualization, W.S., F.H. and Y.Z. (Yongkun Zhang); methodology, W.S. and F.H.; software, F.H.; validation, F.H. and Y.Z.; formal analysis, F.H.; investigation, Q.L.; resources, L.L.; data curation, Y.Z. (Ye Zhang); writing—original draft preparation, F.H.; writing—review and editing, S.J. and W.S.; visualization, F.H.; supervision, W.S.; project administration, W.S.; funding acquisition, W.S. All authors have read and agreed to the published version of the manuscript. References [54–57] are cited in the supplementary materials.

Funding: This research was funded by the National Natural Science Foundation of China under Grants 41975122, U1811464, 42088101, 42105144, 4227515, and 42205149, the National Key R&D Program of China under Grant 2017YFA0604300, Guangdong Basic and Applied Basic Research Foundation 2021B0301030007, the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (311022006), and the Fundamental Research Funds for the Central Universities, Sun Yat-Sen University.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The ERA5-Land dataset is available at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land (last accessed on 6 March 2023). The Chinese soil properties dataset can be obtained at http://globalchange.bnu.edu.cn (last accessed on 6 March 2023). The land cover type data is available at http://edcwww.cr.usgs.gov/landdaac/glcc/glcc.html (last accessed on 6 March 2023). The Digital Elevation Model data can be obtained from the website of Multi-Error-Removed Improved-Terrain DEM (http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/, last accessed on 6 March 2023).

Acknowledgments: The authors thank the anonymous reviewers for providing such valuable comments.

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

References

1. Lesk, C.; Coffel, E.; Winter, J.; Ray, D.; Zscheischler, J.; Seneviratne, S.I.; Horton, R. Stronger temperature–moisture couplings exacerbate the impact of climate warming on global crop yields. *Nat. Food* **2021**, *2*, 683–691. [CrossRef]
2. Lesk, C.; Anderson, W.; Rigden, A.; Coast, O.; Jägermeyr, J.; McDermid, S.; Davis, K.F.; Konar, M. Compound heat and moisture extreme impacts on global crop yields under climate change. *Nat. Rev. Earth Environ.* **2022**, *3*, 872–889. [CrossRef]
3. Shen, C. A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resour. Res.* **2018**, *54*, 8558–8593. [CrossRef]
4. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Prabhat Deep learning and process understanding for data-driven earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef]
5. Kratzert, F.; Klotz, D.; Herrnegger, M.; Sampson, A.K.; Hochreiter, S.; Nearing, G.S. Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resour. Res.* **2019**, *55*, 11344–11354. [CrossRef]
6. Li, Q.; Zhu, Y.; Shangguan, W.; Wang, X.; Li, L.; Yu, F. An attention-aware LSTM model for soil moisture and soil temperature prediction. *Geoderma* **2022**, *409*, 115651. [CrossRef]
7. LeCun, Y.; Bengio, Y. Convolutional networks for images, speech, and time series. In *Handbook of Brain Theory & Neural Networks*; The MIT Press: London, UK, 1995; pp. 1–14.
8. Li, Q.; Wang, Z.; Shangguan, W.; Li, L.; Yao, Y.; Yu, F. Improved daily SMAP satellite soil moisture prediction over China using deep learning model with transfer learning. *J. Hydrol.* **2021**, *600*, 126698. [CrossRef]
9. Pan, B.; Hsu, K.; AghaKouchak, A.; Sorooshian, S. Improving precipitation estimation using convolutional neural network. *Water Resour. Res.* **2019**, *55*, 2301–2321. [CrossRef]
10. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.; Woo, W. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *arXiv* **2015**, arXiv:1506.04214.
11. ElSaadani, M.; Habib, E.; Abdelhameed, A.M.; Bayoumi, M. Assessment of a spatiotemporal deep learning approach for soil moisture prediction and filling the gaps in between soil moisture observations. *Front. Artif. Intell.* **2021**, *4*, 636234. [CrossRef]
12. Xiao, C.; Chen, N.; Hu, C.; Wang, K.; Xu, Z.; Cai, Y.; Xu, L.; Chen, Z.; Gong, J. A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. *Environ. Model. Softw.* **2019**, *120*, 104502. [CrossRef]
13. Chen, G.; Li, L.; Zhang, Z.; Li, S. Short-term wind speed forecasting with principle-subordinate predictor based on Conv-LSTM and improved BPNN. *IEEE Access* **2020**, *8*, 67955–67973. [CrossRef]
14. Sharma, G.; Singh, A.; Jain, S. A hybrid deep neural network approach to estimate reference evapotranspiration using limited climate data. *Neural Comput. Appl.* **2022**, *34*, 4013–4032. [CrossRef]
15. Murdoch, W.J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 22071–22080. [CrossRef] [PubMed]
16. Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.-Z. XAI—Explainable artificial intelligence. *Sci. Robot.* **2019**, *4*, eaay7120. [CrossRef]
17. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
18. Roscher, R.; Bohn, B.; Duarte, M.F.; Garcke, J. Explain it to me–facing remote sensing challenges in the bio- and geosciences with explainable machine learning. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *3*, 817–824. [CrossRef]
19. Nussberger, A.-M.; Luo, L.; Celis, L.E.; Crockett, M.J. Public attitudes value interpretability but prioritize accuracy in artificial intelligence. *Nat. Commun.* **2022**, *13*, 5821. [CrossRef]
20. Bergen, K.J.; Johnson, P.A.; de Hoop, M.V.; Beroza, G.C. Machine learning for data-driven discovery in solid earth geoscience. *Science* **2019**, *363*, eaau0323. [CrossRef]
21. Gagne II, D.J.; Haupt, S.E.; Nychka, D.W.; Thompson, G. Interpretable deep learning for spatial analysis of severe hailstorms. *Mon. Weather Rev.* **2019**, *147*, 2827–2845. [CrossRef]
22. Li, Z.; Wen, Y.; Schreier, M.; Behrangi, A.; Hong, Y.; Lambrigtsen, B. Advancing satellite precipitation retrievals with data driven approaches: Is black box model explainable? *Earth Space Sci.* **2021**, *8*, e2020EA001423. [CrossRef]
23. Althoff, D.; Rodrigues, L.N.; Silva, D.D. Addressing hydrological modeling in watersheds under land cover change with deep learning. *Adv. Water Resour.* **2021**, *154*, 103965. [CrossRef]

24. Huang, F.; Shangguan, W.; Li, Q.; Li, L.; Zhang, Y. Beyond prediction: An integrated post–hoc approach to interpret complex model in hydrometeorology. *SSRN J.* **2022**, 59. [CrossRef]
25. Dikshit, A.; Pradhan, B. Interpretable and explainable AI (XAI) model for spatial drought prediction. *Sci. Total Environ.* **2021**, *801*, 149797. [CrossRef]
26. McGovern, A.; Lagerquist, R.; John Gagne, D.; Jergensen, G.E.; Elmore, K.L.; Homeyer, C.R.; Smith, T. Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 2175–2199. [CrossRef]
27. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed. Available online: https://christophm.github.io/interpretable-ml-book/ (accessed on 6 March 2023).
28. Gevaert, C.M. Explainable AI for Earth observation: A review including societal and regulatory perspectives. *Int. J. Appl. Earth Obs.* **2022**, *112*, 102869. [CrossRef]
29. Fisher, A.; Rudin, C.; Dominici, F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 177.
30. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
31. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA, 4–9 December 2017; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 4768–4777.
32. Simonyan, K.; Vedaldi, A.; Zisserman, A. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv* **2014**, arXiv:1312.6034.
33. Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic attribution for deep networks. *arXiv* **2017**, arXiv:1703.01365.
34. Ancona, M.; Ceolini, E.; Öztireli, C.; Gross, M. Towards better understanding of gradient-based attribution methods for deep neural networks. *arXiv* **2018**, arXiv:1711.06104.
35. Hooker, S.; Erhan, D.; Kindermans, P.-J.; Kim, B. A benchmark for interpretability methods in deep neural networks. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Curran Associates Inc.: Red Hook, NY, USA, 2019; pp. 9737–9748.
36. Seo, J.; Choe, J.; Koo, J.; Jeon, S.; Kim, B.; Jeon, T. Noise-adding methods of saliency map as series of higher order partial derivative. *arXiv* **2018**, arXiv:1806.03000.
37. Smilkov, D.; Thorat, N.; Kim, B.; Viégas, F.; Wattenberg, M. SmoothGrad: Removing noise by adding noise. *arXiv* **2017**, arXiv:1706.03825.
38. Piao, S.; Ciais, P.; Huang, Y.; Shen, Z.; Peng, S.; Li, J.; Zhou, L.; Liu, H.; Ma, Y.; Ding, Y.; et al. The impacts of climate change on water resources and agriculture in China. *Nature* **2010**, *467*, 43–51. [CrossRef] [PubMed]
39. Meng, X.; Mao, K.; Meng, F.; Shi, J.; Zeng, J.; Shen, X.; Cui, Y.; Jiang, L.; Guo, Z. A fine-resolution soil moisture dataset for China in 2002–2018. *Earth Syst. Sci. Data* **2021**, *13*, 3239–3261. [CrossRef]
40. Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [CrossRef]
41. Beck, H.E.; Pan, M.; Miralles, D.G.; Reichle, R.H.; Dorigo, W.A.; Hahn, S.; Sheffield, J.; Karthikeyan, L.; Balsamo, G.; Parinussa, R.M.; et al. Evaluation of 18 satellite- and model-based soil moisture products using in situ measurements from 826 sensors. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 17–40. [CrossRef]
42. Shangguan, W.; Dai, Y.; Liu, B.; Zhu, A.; Duan, Q.; Wu, L.; Ji, D.; Ye, A.; Yuan, H.; Zhang, Q.; et al. A China data set of soil properties for land surface modeling. *J. Adv. Model. Earth Syst.* **2013**, *5*, 212–224. [CrossRef]
43. Loveland, T.R.; Reed, B.C.; Brown, J.F.; Ohlen, D.O.; Zhu, Z.; Yang, L.; Merchant, J.W. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sens.* **2000**, *21*, 1303–1330. [CrossRef]
44. Balenović, I. Quality Assessment of high density digital surface model over different land cover classes. *Period. Biol.* **2016**, *117*, 459–470. [CrossRef]
45. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
46. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37 (ICML'15), Lille, France, 6–11 July 2015; pp. 448–456.
47. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
48. O'Malley, T.; Bursztein, E.; Long, J.; Chollet, F.; Jin, H.; Invernizzi, L.; Gabriel, d.M.; Fu, Y.; Hahn, A.; Mullenbach, J.; et al. KerasTuner. Available online: https://github.com/keras-team/keras-tuner (accessed on 6 March 2023).
49. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2017**, arXiv:1412.6980.
50. Zhang, Z.; Chen, X.; Pan, Z.; Zhao, P.; Zhang, J.; Jiang, K.; Wang, J.; Han, G.; Song, Y.; Huang, N.; et al. Quantitative estimation of the effects of soil moisture on temperature using a soil water and heat coupling model. *Agriculture* **2022**, *12*, 1371. [CrossRef]
51. Wang, G.; Zhuang, L.; Mo, L.; Yi, X.; Wu, P.; Wu, X. BAG: A linear-nonlinear hybrid time series prediction model for soil moisture. *Agriculture* **2023**, *13*, 379. [CrossRef]

52. Krenn, M.; Pollice, R.; Guo, S.Y.; Aldeghi, M.; Cervera-Lierta, A.; Friederich, P.; dos Passos Gomes, G.; Häse, F.; Jinich, A.; Nigam, A.; et al. On Scientific Understanding with Artificial Intelligence. *Nat. Rev. Phys.* **2022**, *4*, 761–769. [CrossRef]
53. Schwartz, M.D. Should Artificial Intelligence Be Interpretable to Humans? *Nat. Rev. Phys.* **2022**, *4*, 741–742. [CrossRef]
54. Pan, J.; Shangguan, W.; Li, L.; Yuan, H.; Zhang, S.; Lu, X.; Wei, N.; Dai, Y. Using data-driven methods to explore the predictability of surface soil moisture with FLUXNET site data. *Hydrol. Process.* **2019**, *33*, 2978–2996. [CrossRef]
55. Belgiu, M.; Drăguţ, L. Random Forest in remote sensing: A review of applications and future directions. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2016**, *114*, 24–31. [CrossRef]
56. Bisong, E. More Supervised Machine Learning Techniques with Scikit-learn. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform*; Apress: Berkeley, CA, USA, 2019. [CrossRef]
57. Shrikumar, A.; Greenside, P.; Kundaje, A. Learning Important Features Through Propagating Activation Differences. In *International Conference on Machine Learning*; PMLR: Westminster, CA, USA, 2017.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.