MDPI

Article

Artificial Neural Network-Based Microwave Satellite Soil Moisture Reconstruction over the Qinghai–Tibet Plateau, China

Jie Wang [1,2,†] and **Duanyang Xu** [1,*,†]

1   Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; wangjie203@mails.ucas.ac.cn
2   Department of Environment and Resources, University of Chinese Academy of Sciences, Beijing 100049, China
*   Correspondence: xudy@igsnrr.ac.cn; Tel.: +86-10-64889795
†   These authors contributed equally to this work.

Abstract: Soil moisture is a key parameter for land-atmosphere interaction system; however, fewer existing spatial-temporally continuous and high-quality observation records impose great limitations on the application of soil moisture on long term climate change monitoring and predicting. Therefore, this study selected the Qinghai–Tibet Plateau (QTP) of China as research region, and explored the feasibility of using Artificial Neural Network (ANN) to reconstruct soil moisture product based on AMSR-2/AMSR-E brightness temperature and SMAP satellite data by introducing auxiliary variables, specifically considering the sensitivity of different combination of input variables, number of neurons in hidden layer, sample ratio, and precipitation threshold in model building. The results showed that the ANN model had the highest accuracy when all variables were used as inputs, it had a network containing 12 neurons in a hidden layer, it had a sample ratio 80%-10%-10% (training-validation-testing), and had a precipitation threshold of 8.75 mm, respectively. Furthermore, validation of the reconstructed soil moisture product (named ANN-SM) in other period were conducted by comparing with SMAP (April 2019 to July 2021) for all grid cells and in situ soil moisture sites (August 2010 to March 2015) of QTP, which achieved an ideal accuracy. In general, the proposed method is capable of rebuilding soil moisture products by adopting different satellite data and our soil moisture product is promising for serving the studies of long-term global and regional dynamics in water cycle and climate.

Keywords: soil moisture; artificial neural network; AMSR-2; SMAP; Qinghai–Tibet Plateau

1. Introduction

Soil moisture is a key factor in the climate-land surface coupling system. It not only plays a critical role in the hydrological process, vegetation and crop growth, and material and energy cycle of ecosystem [1–4], but it is also essential for understanding the dynamics of the earth system and predicting climate and land changes in the future [5]. The current rapid development of satellite remote sensing technology, including optical, radar, and microwave remote sensing, provides an effect way to achieve soil moisture observation at different scales [6–9]. Especially for passive microwave satellite remote sensing [10], due to the lessened influence of the atmosphere, deeper detection depth, and more direct physical relationship between remote sensing information and soil moisture [11–13], it has become the mainstream method to obtain soil moisture observation data at global and local scales, using C band (4–8 GHz) at higher frequencies and L band (0.5–1.5 GHz) at lower frequencies [14–16]. This includes Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) (2002–2011), Advanced Microwave Scanning Radiometer 2 (AMSR2) (2012–present), Soil Moisture Ocean Salinity (SMOS) (2010–present), and the Soil Moisture Active Passive (SMAP) (2015–present).

Due to the limitations of band coverage and sensor lifetime, it is difficult to acquire credible, high-precision, and long time series observations of soil moisture products, which is insufficient to satisfy the study of long-term climate change, terrestrial hydrology monitoring, and agriculture and ecosystem sustainable management. To solve this problem, many scholars have tried to fuse and reconstruct a variety of soil moisture data with different instrument characteristics (frequency, spatial resolution, temporal coverage, polarization, revisit time, etc.) to obtain the spatial-temporal consistent and long time series soil moisture products of passive microwave satellite. The cumulative distribution function (CDF) matching is the main soil moisture calibration method and widely used in merging multiple sources soil moisture data into long time series products [17–20]; for example, the soil moisture dataset from the European Space Agency's Climate Change Initiative (ESA CCI) combines various single-sensor active and passive microwave soil moisture products into three harmonized products using CDF matching [21]. On the other hand, the fusion algorithm based on CDF matching is relatively simple, which limits the fusion accuracy, especially in the regions with less observation data and complex land surface environments.

Machine learning algorithms are another way to acquire long time series soil moisture products. Compared to CDF matching, it is an exclusively data-driven technique with the capability to learn the functional dependencies and build the correlation among multiple variables without prior knowledge and assumptions about functional relationships [22–24]. It usually takes the most credible soil moisture products as the standard reference dataset and trains other data to obtain the non-linear function that can be applied to achieve more precise prediction. Previous studies have used machine learning algorithms, such as Neural Network (NN) [25], General Regression Neural Network (GRNN) [26], convolutional neural network (CNN) [27], Random Forest (RF) [28], etc., and adopted auxiliary data, including of the Normalized Difference Vegetation Index (NDVI) [29], Microwave Vegetation Index (MVI) [30], Land Surface temperature (LST) [31], Leaf Area Index (LAI) [32], Albedo, etc., to train multi-source soil moisture data to obtain longer homogeneous time series products. The Artificial Neural Network (ANN) is an effective approach to establish a nonlinear model and widely applied in microwave remote sensing soil moisture retrieval [24,33,34]. However, some problems are easily overlooked when using ANN for soil moisture reconstruction, such as the number of neurons in hidden layer of neural network and the different proportions of training samples, which have a vital impact on the model training results. In addition, rainfall may occur between the two sensors due to different crossing times, and the large precipitation may cause unreasonable regressions; thus, the precipitation data should be used to screen out ineffective dates [35]. These three factors were not considered sufficiently in previous studies, which limited the accuracy of using ANN for soil moisture calibration and also restricted the application of ANN in the reconstruction long time series soil moisture products to a certain extent. Furthermore, more auxiliary data related to soil moisture should be introduced in ANN to get a robust modeling, including vegetation status, land surface roughness, soil temperature, and soil moisture information from other sensors.

To overcome these problems, this study selected the Qinghai–Tibet Plateau (QTP) of China as a research region, and explored the feasibility of using ANN to reconstruct a long time series soil moisture product based on the AMSR-E/AMSR2 brightness temperature (due to its long time series since 2002 but relatively unreliable estimation in many regions [36–39]) and the SMAP soil moisture observation data (due to its more accurate and reliable soil moisture estimation but the short time span [40–44]), by aid of the GLDAS land datasets (soil moisture, soil temperature, and land surface temperature) and MODIS LAI. In addition, we designed a series of sensitivity analysis to analyze the factors that impact the model accuracy, including the different combination of input variables, the number of neurons in hidden layer, sample proportions, and precipitation thresholds. Then, the control variable method is used to determine the best combination of parameters that will be utilized for simulating long time series soil moisture product.

This paper is organized as follows: Section 2 presents the study area as well as the different data used in the process of ANN model training and validation. Section 3 provides the description of the ANN method, and procedure of sensitivity analysis involves the combination of input variables, selection of the number of neurons in hidden layer, sample ratio, and precipitation threshold. Then, Section 4 shows the results of sensitivity analysis, as well as the accuracy and advantages of ANN-SM in comparison to SMAP soil moisture products and in situ soil moisture sites from soil moisture ground measurement networks. Finally, Sections 5 and 6 provide discussions and conclusions.

2. Study Area and Data

2.1. Study Area

QTP is the largest plateau in China and the highest mountain plateau in the world. It is often referred to as the "Roof of the World" or the "water tower of Asia", and it is located between 26°–39°47′N and 73°19′–104°47′E (Figure 1). It stretches approximately 1500 km north to south and 2900 km east to west, with an area of 2.5 million km^2. The QTP is generally between 3000 and 5000 m above sea level, with an average elevation exceeding 4000 m and being surrounded by imposing mountain ranges. The mean annual temperature of QTP is usually from −2.2 °C to 0 °C, the mean annual sunshine duration is from 2730 h to 2915 h, the average annual potential evapotranspiration is about 940 mm, the mean annual precipitation is between 415 mm and 512 mm, and 73% of the precipitation occurs in summer and autumn. As a "sensitive area" of climate change in Asia and even the northern hemisphere, the QTP has a great impact on the Asian monsoon and the global atmospheric cycle [45]. Monitoring the long-term series of soil moisture observation records in the QTP plays an important role in response to regional or global climate change [46].
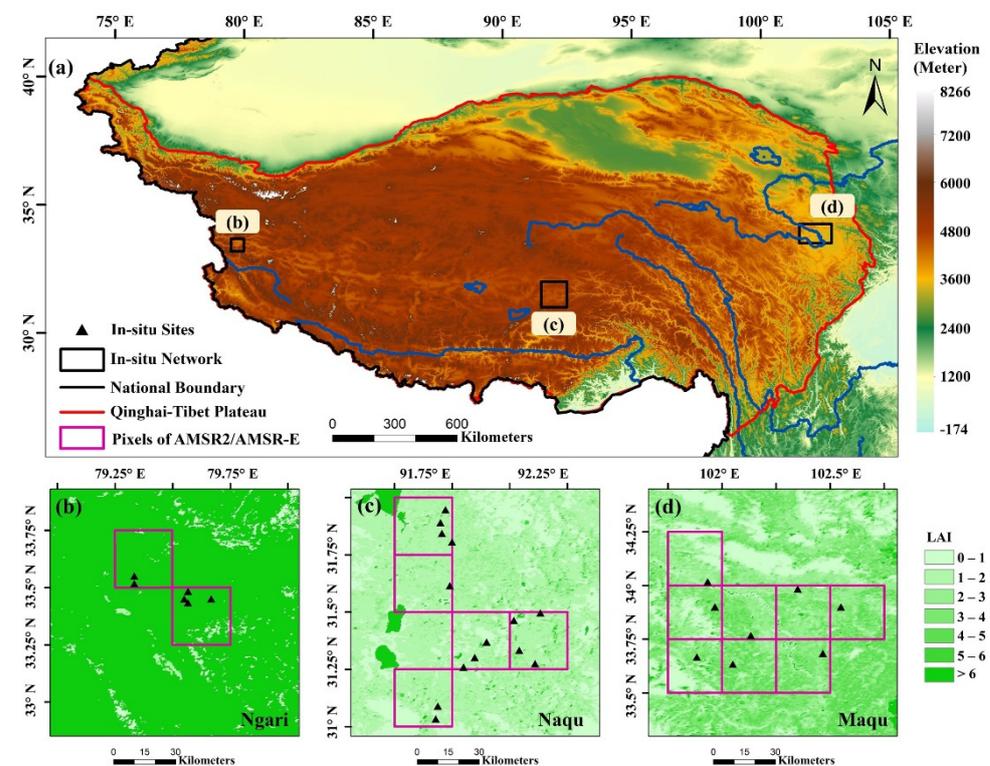

Figure 1. The location of Qinghai–Tibet Plateau (**a**) and distribution of in situ sites with LAI map. (**b**–**d**) corresponds to the Ngari network, the Naqu network, and the Maqu network, respectively.

2.2. Data

2.2.1. Brightness Temperature of AMSR-E/AMSR2

AMSR-E was launched on 4 May 2002, and stopped operations on 4 October 2011. It had a 13:30/01:30 equator-crossing orbit of ascending and descending orbits, with a one to two day repeat coverage. The AMSR-E sensor operated at six frequencies ranging from 6.925 GHz to 89.0 GHz (6.9 GHz, 10.7 GHz, 18.7 GHz, 23.8 GHz, 36.5 GHz, and 89.0 GHz). The first five frequencies of AMSR-E brightness temperature were selected due to the higher the frequency of the microwave, the shorter the corresponding wavelength, and the weaker the ability to penetrate the ground.

As AMSR-E's successor, AMSR2 was carried by the Global Change Observation Mission-W1 (GCOM-W1) satellite and launched on 3 July 2012, and it has the same ascending and descending times as AMSR-E. The AMSR2 has seven frequencies (6.9 GHz, 7.3 GHz, 10.7 GHz, 18.7 GHz, 23.8 GHz, 36.5 GHz, and 89.0 GHz) and we exclude the brightness temperature data of 36.5 GHZ and 89.0 GHZ, and the first five frequencies (F1–F5) at 6.9 to 23.8 GHZ were used as model input data.

The AMSR-E/AMSR2 brightness temperature products can be downloaded from the Globe Portal System (G-Portal), https://gportal.jaxa.jp/ (accessed on 14 December 2021).

2.2.2. The SMAP Soil Moisture

The SMAP mission is an orbiting observatory that measures the amount of water in the surface soil on Earth. It was launched in January 2015 and started operation in April 2015. Its ascending time is 6:00 pm (local time), and its descending time is 6:00 am (local time). The SMAP L3 descending soil moisture data were used in this study due to the temperature in the morning is more uniform than that at night. The SMAP L3 daily SM products are freely available on the National Snow and Ice Data Center (NSIDC), https://nsidc.org/data/ (accessed on 14 December 2021).

2.2.3. The Auxiliary Input Variables

The auxiliary variables included GLDAS datasets and LAI from MODIS. GLDAS is forced with a combination of model and observation data, providing ideal land surface variables and energy fields by using land surface models and data assimilation technologies, and GLDAS-2.1 Noah Land Surface Model L4 products with a temporal resolution of three hours and a spatial resolution of 0.25° were used in this study, including GLDAS L4 SM (GLDAS soil moisture product), GLDAS L4 LST (GLDAS surface temperature products), GLDAS L4 ST (GLDAS soil temperature products), and GLDAS L4 Prec (GLDAS precipitation products). These data are also available on the Goddard Earth Sciences Data and Information Services Center (GES DISC), at https://disc.gsfc.nasa.gov/ (accessed on 14 December 2021).

The LAI variable defines the number of equivalent layers of leaves relative to a unit of ground area. The Level-4 MODIS global Leaf Area Index (LAI) is an 8-day 500-m resolution product on a Sinusoidal grid, and these data can be downloaded at https://ladsweb.modaps.eosdis.nasa.gov/ (accessed on 14 December 2021).

2.2.4. The In Situ Soil Moisture Sites

Three soil moisture ground measurement networks provided by the National Tibetan Plateau Data Center (http://data.tpdc.ac.cn (accessed on 14 December 2021)) [47,48] are used in the article, namely, the Naqu network, the Maqu network, and the Ngari network [49]. The Naqu network is located in the central QTP with a cold, semi-humid climate, the Maqu network is located in the east of the QTP with a humid climate, and the Ngari network is located in the western of the QTP with an arid climate covering by bare land and grassland in most of the region [50,51]. A total of 28 sites are selected in these three networks, of which 6 are in the Ngari network and 14 are in the Naqu network and 8 are in the Maqu network. Some grids in the Naqu network contain multiple sites, and thus, the

average value of all sites in the same grid is calculated. The spatial distribution of sites is shown in Figure 1b–d.

2.2.5. The Other Soil Moisture Satellite Products

Three satellite soil moisture products were used for comparison with our rebuilt product, among which two soil moisture products (Level 3, JAXA, and LPRM) derived from AMSR-E/AMSR2 using the Land Parameter Retrieval Model (LPRM) and from the Japan Aerospace Exploration Agency (JAXA), respectively. The soil moisture products from JAXA can be downloaded from the G-Portal, and the soil moisture product from LPRM can be downloaded from the GES DISC. Another is the Soil Moisture and Ocean Salinity (SMOS) Level 3 soil moisture data. The SMOS satellite has an orbit with a 6:00 am/6:00 pm equator overpass time of ascending and descending orbits, providing global brightness temperature observations at the L-band from 2010 until the present. The SMOS ascending product (at 6:00 am local time) was selected in our study. The SMOS data are provided by Centre Aval de Traitement des Données (CATDS), at http://www.catds.fr/Products/Products-access (accessed on 14 December 2021). The detailed information of all data concerning the spatial resolution, temporal resolution, and transit time of each dataset are shown in Table 1.

Table 1. List of satellite and reanalysis datasets.

Name of Dataset	Spatial Resolution	Temporal Resolution	Transit Time (Local Time)
AMSR-2/AMSR-E L3 TB	0.25°	Half a day	1:30 a.m., 1:30 p.m.
SMAP L3	36 km	Half a day	6:00 a.m., 6:00 p.m.
GLDAS L4	0.25°	3 h	-
MODIS LAI	500 m	8 days	-
LPRM/JAXA SM	0.25°	Half a day	1:30 a.m., 1:30 p.m.
SMOS L3	35–50 km	Half a day	6:00 a.m., 6:00 p.m.

3. Methodology

3.1. ANN Processing Strategy for Soil Moisture

ANN is an important algorithm of machine learning, which can accurately approximate the complicated nonlinear relationship between different environmental variables with limited understanding of the physical process [40,52] and has a wide range of applications for remote sensing retrieval, multiple sources data fusion and reconstruction, downscaling, and so forth [53]. The structure of the feed-forward network includes the input layer, the hidden layer, and the output layer (Figure A1). A network is composed of a large number of nodes (or neurons) connected to each other. Each node represents an output function, which is called an activation function. Each connection between two nodes represents a weighted value of the connected signal. The main work of the neural network is to establish a model and determine the weights. The learning and training of a neural network requires a set of input data and output data. The neural network calculates the result after selecting the network model and passing and training functions, and corrects the weight according to the error between the actual output values and the expected output values. ANN in this study is utilized to train the input and output variables through using the Levenberg–Marquardt backpropagation algorithm. We use MATLAB to implement the ANN training and simulation.

To build the ANN model, nine input variables and one output variable were used. The input variables include AMSR-2 F1-F5, GLDAS SM, GLDAS LST, GLDAS ST, and MODIS LAI, and the output variable is SMAP SM. The training period of the ANN model is from April 2015 to March 2019. The descending SMAP SM (at 6:00 am local time) and descending AMSR-2 TB (at 1:30 am local time) were selected in this study. The SMAP SM (36 km resolution), MODIS LAI (500 m resolution), and SMOS (35–50 km resolution) were resampled to the 0.25° resolution using linear interpolation. The temporal resolution of GLDAS SM, GLDAS LST, and GLDAS ST was increased from 3 h to 1 h using nearest

interpolation and the values of these data at 2:00 am local time were selected. All variables were matched grid by grid (days for each variable does not contain the null value are selected) during the training period. Those grid cells with less than 50 matching pairs of input-output variables were removed. To keep the orders of magnitude consistent between variables, including the input and output variables, each variable was normalized to 0–1. After this simulation, the output variable was reversely normalized to obtain the ANN-SM value.

Performance of the ANN model is affected by many factors and we considered the combination of input variables, i.e., the number of neurons in hidden layer, sample ratios, and precipitation thresholds, as the three most important factors. Therefore, the sensitivity experiments on the three parameters were designed to get the optimal parameter settings using the variable-controlling approach. Then, we utilized the optimal parameters and adopted the input and output data over the training period (April 2015 to March 2019) to build the ANN model to obtain a nonlinear function, which can be applied to data prediction over the other test period. The flowchart of this process is shown in Figure 2.
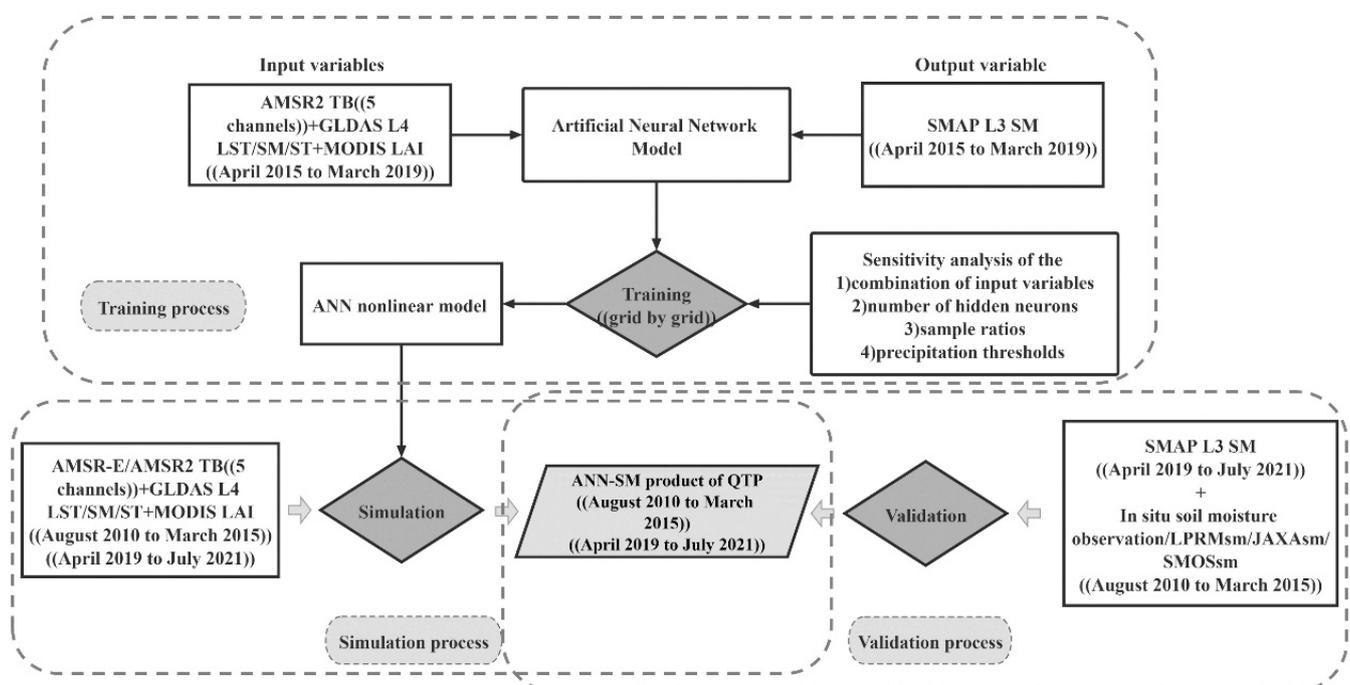

Figure 2. Flowchart of the training, simulation, and validation process of the artificial neural network soil moisture (ANN-SM).

3.2. Sensitivity Analysis

3.2.1. Combination of Input Variables

The input variables including the first five frequencies of AMSR2 brightness temperature (F1–F5), soil moisture, land surface temperature, soil temperature, and leaf area index were utilized. It is necessary to research the impact of the combination of different input variables on the ANN model accuracy. On the basis of F1–F5, another four input variables were successively added to test the sensitivity of these variables to the ANN model. Then, the variables that contribute most to ANN model were selected along with F1–F5 as inputs. The optimal combination of input variables can be found by these sensitivity experiments.

3.2.2. Number of Hidden Neurons

The determination of the appropriate number of hidden neurons and the selection of the most appropriate activation functions in neurons of hidden and output layers required more attention. It is worth mentioning that a network with too few hidden neurons may

not be powerful enough for a given learning task, while with an excessive number of hidden neurons, ANNs "memorize" the training dataset and exhibit poor generalizing capabilities, which results in an unsatisfactory performance on new data. However, there is no universal, commonly accepted method to determine the number of neurons in the hidden layer. In this article, the suggested number of hidden layer neurons evaluated with the used rules of thumb was in the range of 2–12 nodes.

3.2.3. Sample Ratios

There are three types of samples in the training process of the ANN model, which are training samples, validation samples, and testing samples. Training samples are presented to the network during training, and the network is adjusted according to its error, validation samples are used to measure network generalization, and to halt training when generalization stops improving, and testing samples have no effect on training, and thus, provide an independent measure of network performance during and after training. The overall samples contain the three kinds of samples. It is obvious that model accuracy is greatly affected by different sample ratios. It is not difficult to understand that the ANN model with too many training samples will negatively squeeze the proportions of the other two samples, while with too few training samples it is not enough to simulate a reasonable nonlinear function, which results in relatively unsatisfactory accuracy of the model. Thus, the ratio of training samples with a step length of 10% from 40% to 80% is a reasonable interval. We utilized 13 different sample combinations that is in training-validation-testing form to train the input and output variables and they are 40%-15%-45%, 40%-30%-30%, 40%-45%-15%, 50%-15%-35%, 50%-20%-30%, 50%-25%-25%, 60%-10%-30%, 60%-20%-20%, 60%-30%-10%, 70%-10%-20%, 70%-15%-15%, 70%-20%-10%, and 80%-10%-10%, respectively. The optimal number of neurons in the hidden layer was used before the sensitivity analysis of the sample ratio.

3.2.4. Precipitation Thresholds

There is a time difference between AMSR-2 TB (at 1:30 am local time) and SMAP SM (at 6:00 am local time), and large rainfall within the time difference may cause unreasonable regressions theoretically, thereby interfering with the training accuracy of the ANN model. The precipitation data from GLDAS is selected for calculating the sum of precipitation within the time difference between AMSR-2 and SMAP. Specifically, the temporal resolution of GLDAS Prec were improved from 3 h to 1 h through the nearest neighbor interpolation method. The precipitation threshold refers to the different values of the sum of the precipitation from 1:00 am to 7:00 am of GLDAS Prec (with a temporal resolution of 1 h), and the unit of precipitation threshold is mm/7 h. Ineffective data filtering based on the precipitation threshold aims to eliminate precipitation events within the time difference of other input and output variables. A low precipitation threshold will result in a small amount of data for training, while a high precipitation cannot eliminate precipitation events effectively. Heavy rain (15–30 mm within 12 h) and torrential rain (>30 mm within 12 h) were selected as two standard precipitation thresholds according to the reference of GBT28592-2012, and they were set to 8.75 mm ($15 \times (7/12)$) and 17.5 mm ($30 \times (7/12)$), respectively. In addition, to explore the impact of different precipitation thresholds on training accuracy, the final precipitation thresholds were set to 8.75 mm, 10 mm, 11.25 mm, 12.5 mm, 13.75 mm, 15 mm, 16.25 mm, and 17.5 mm.

3.3. Validation Procedure

The validation procedure has two parts: one is the validation of the sensitivity analysis results, and the other one is the validation of the corrected soil moisture product (named ANN-SM).

First, the error of sensitivity analysis results is represented by the correlation coefficient (R) and the Root Mean Squared Error (RMSE). Specifically, the input and output variables of all grids in the QTP were trained using ANN model, and the mean value of R and RMSE

of all grids in the QTP were calculated as the final error. The error of the overall samples is used as the criterion to evaluate the accuracy sensitivity analysis results.

Then, in order to quantitatively evaluate the effectiveness of ANN-SM, we utilized R, RMSE, and bias (BIAS) to analyze the consistency between the ANN-SM and SMAP soil moisture during the test period (April 2019 to July 2021). Furthermore, to further evaluate the accuracy of the long time series ANN-SM product, the in situ sites from 15 grids in the soil moisture ground measurement networks are used to compare with the ANN-SM from August 2010 to March 2015. The time series scatterplots of ANN-SM and in situ soil moisture data were plotted to show the consistency between them, and the statistical indicators (R, RMSE, and BIAS) were calculated to represent the error of ANN-SM and in situ soil moisture. R, RMSE, and BIAS are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_i - y_i)^2} \qquad (1)$$

$$R = \frac{\sum_{t=1}^{N}(x_i - \overline{x})((y_i - \overline{y})}{\sqrt{\sum_{t=1}^{N}(x_i - \overline{x})^2(y_i - \overline{y})^2}} \qquad (2)$$

$$\text{BIAS} = \frac{1}{N}\sum_{t=1}^{N}(x_i - y_i) \qquad (3)$$

where $\overline{x}$ and $\overline{y}$ indicate the mean values, x_i is target value, and y_i is the predicted value.

4. Results

4.1. Sensitivity Analysis

Figure 3a presented the result of sensitivity experiment of combination of different input variables. The overall accuracy of ANN model improves as the number of input variables increases. The accuracy of ANN model is highest when the input variables are all variables, and the mean RMSE and R of the overall samples is 0.099 cm^3/cm^3 and 0.89, respectively. Compared with the other three variables, soil moisture derived from GLDAS improves the performance of ANN model the most when it is used as the input variables with F1–F5 initially. Then, on the basis of F1–F5 and soil moisture from GLDAS, the addition of soil temperature from GLDAS increases the model accuracy most obviously. Next, the land surface temperature from GLDAS slightly improves the model performance in comparison to the leaf area index from MODIS. Thus, soil moisture contributed the most to the ANN model, followed by soil temperature, land surface temperature, and leaf area index.

Figure 3b illustrated the result of sensitivity analysis of the number of neurons in the hidden layer. The best result is obtained for network containing 12 neurons in the hidden layer and its mean RMSE and R of overall samples is 0.099 cm^3/cm^3 and 0.89, respectively. It can be seen that the variation trend of accuracy in the overall samples increases when the number of hidden neurons increases. However, judging from the changes in the accuracy of the validation samples and testing samples from Figure 3, the accuracy of these two kinds of samples decreases as the number of hidden neurons increases. Furthermore, the accuracy of the overall samples changes little and mean RMSE and R stabilize near 0.1 cm^3/cm^3 and 0.89 when the number of neurons in the hidden layer is greater than 6.

According to Figure 3c, training samples, validation samples, and overall samples have all achieved acceptable training accuracy, while the accuracy of the testing samples is lower than the other three types of samples. From the overall accuracy of the proportion of all samples, as the proportion of validation samples and testing samples increases, the overall accuracy increases. Among these 13 training results, the highest accuracy is when training-validation-testing samples is 80%-10%-10%, where its mean of RMSE and R of the overall samples is 0.097 cm^3/cm^3 and 0.90, respectively. The overall accuracy is largely determined by the accuracy of the training samples when the proportion of

training samples is high. When the proportion of training samples is small, the proportion of test samples is large, and the accuracy of test samples is inherently low; thus, the overall accuracy will not be high, e.g., 40%-15%-35%. It is appropriate for the sample ratios to be 80%-10%-10% for training samples, validation samples, and testing samples, respectively.
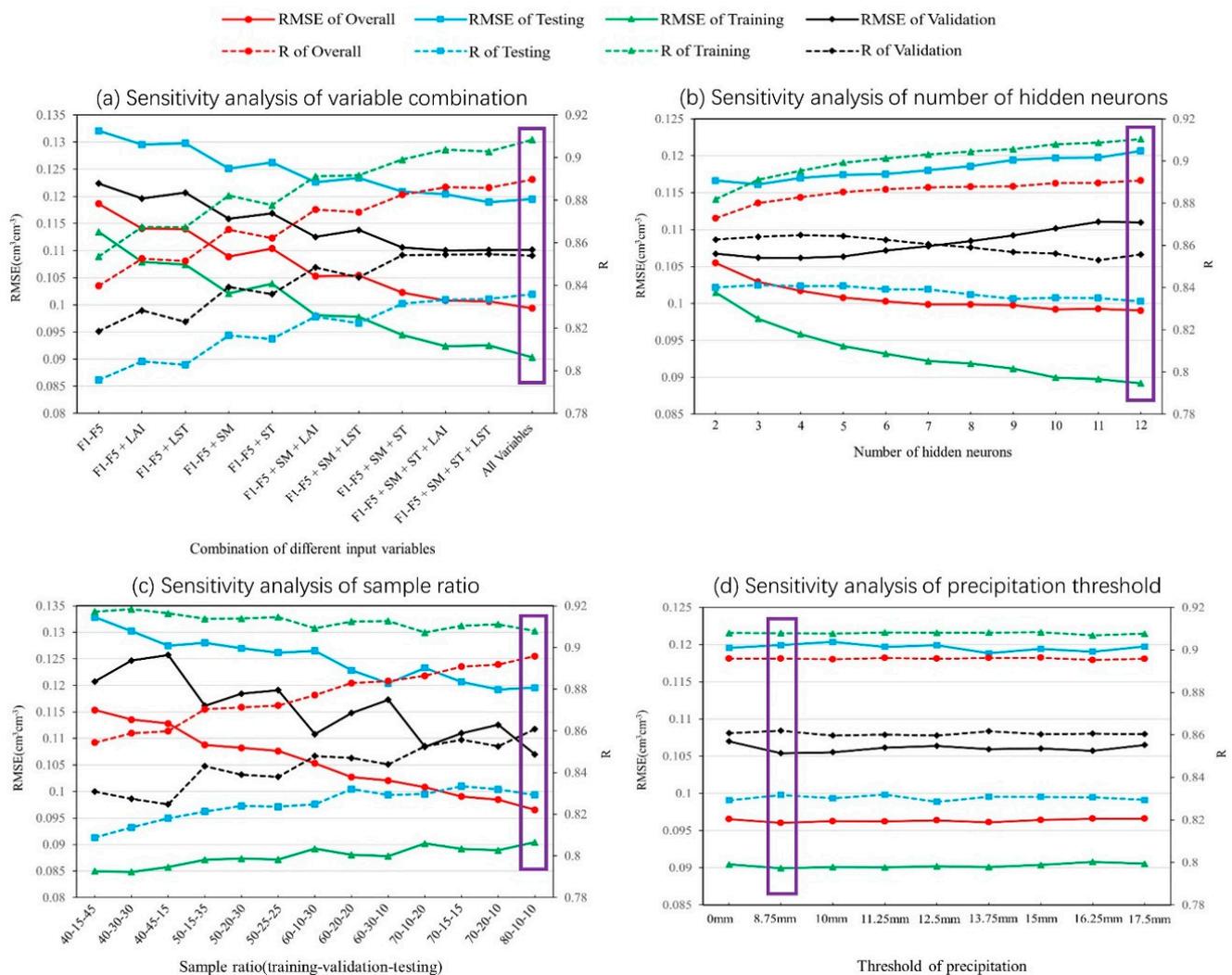

Figure 3. Mean of RMSE and R of all grid cells in the QTP for training, validation, testing, and overall samples in the context of sensitivity analysis in terms of (**a**) the combination of different input variables, (**b**) the number of hidden neurons, (**c**) different sample ratios, and (**d**) different value of precipitation thresholds. The purple box is the optimal parameter.

Figure 3d depicts the mean values of errors from different precipitation thresholds. Comparing the model accuracy of different precipitation threshold, the correlation between the level of the precipitation threshold and the accuracy of model training was weak. It is not difficult to find that the lowest average RMSE of overall samples and the highest average R of the overall samples could be obtained by a small margin (mean RMSE of the overall samples is 0.096 cm^3/cm^3 and mean R of the overall samples is about 0.90) when the precipitation threshold is 8.75 mm. On the one hand, some ineffective data that negatively affect the accuracy can be eliminated, and on the other hand, most of the effective data can be retained, that is, the number of valid days is moderate when the precipitation threshold of 8.75 mm.

4.2. Comparative Analysis

4.2.1. Comparative Analysis of ANN-SM and SMAP Soil Moisture

Figure 4a showed the correlation coefficient between ANN-SM and SMAP soil moisture for each grid cell of the QTP. It can be seen that the corrected soil moisture values (ANN-SM) correlate well with the SMAP soil moisture for most of the grids of the QTP, with the mean value of R = 0.72. The distribution areas with higher correlation coefficient value (R > 0.8) are concentrated in the central and southern parts of the QTP, while the areas with low R values are mainly distributed in the northwest and central-eastern parts of the QTP. The land type in the northwest is dominated by barren land and sparsely vegetated land, and this area is normally covered with ice and snow. The leaf area index variable that participated in ANN simulation is more like a constant than the soil moisture variable with a larger variation trend due to the generally low vegetation coverage and low interannual variability of vegetation coverage. Thus, the simulation effect in this area was no better than other areas. The spatial distribution of the RMSE value for each grid of the QTP is plotted in Figure 4b, and the average value of RMSE = 0.034 cm^3/cm^3. The spatial distribution of the RMSE value was consistent well with the spatial distribution of the correlation coefficient. Figure 4c showed the spatial distribution of the value of BIAS, with the mean absolute value of BIAS = 0.0122 cm^3/cm^3. The BIAS value of the most grids was around 0, indicating the error between ANN-SM and SMAP soil moisture is small. The results of the above parameters (R, RMSE, and BIAS) indicated the feasibility of correcting soil moisture products based on the ANN algorithm.

4.2.2. Comparative Analysis of ANN-SM and In Situ Sites

The time series scatterplots of the six in situ grids in three soil moisture ground measurement networks were drawn in Figure 5. The grid cells shown in Figure 5 are representative grids for their respective network, and the time series scatterplots of the other nine grids are plotted in Figure A2. The purple hollow circles represent the value of ANN-SM and the black solid line represent the value of the in situ soil moisture. Missing values of AMSR-E/AMSR2 brightness temperature data between October 2011 and June 2012 were not shown on the scatterplots. As can be seen from Figures 5 and A2, the trend of ANN-SM was highly consistent with that of in situ soil moisture, and the seasonal and interannual variation of the in situ soil moisture was generally captured by ANN-SM. Moreover, to further clarify the advantages of ANN algorithm and our ANN-SM product, we validated the performance of ANN-SM by comparing the simulated soil moisture output with the satellite standard soil moisture products of AMSR-E/AMSR2 from JAXA and LPRM and SMOS, over the above six in situ grids. Results are also shown in Figures 5 and A2. ANN-SM seemed to be more consistent with in situ soil moisture in comparison to LPRM, JAXA, and SMOS soil moisture data. Specifically, soil moisture from LPRM is significantly higher than the in situ measured value, while soil moisture from JAXA and SMOS underestimated the in situ measured value in most grids. Therefore, the performance of ANN-SM is better than that of AMSR-E/AMSR2 soil moisture from JAXA and LPRM and SMOS.
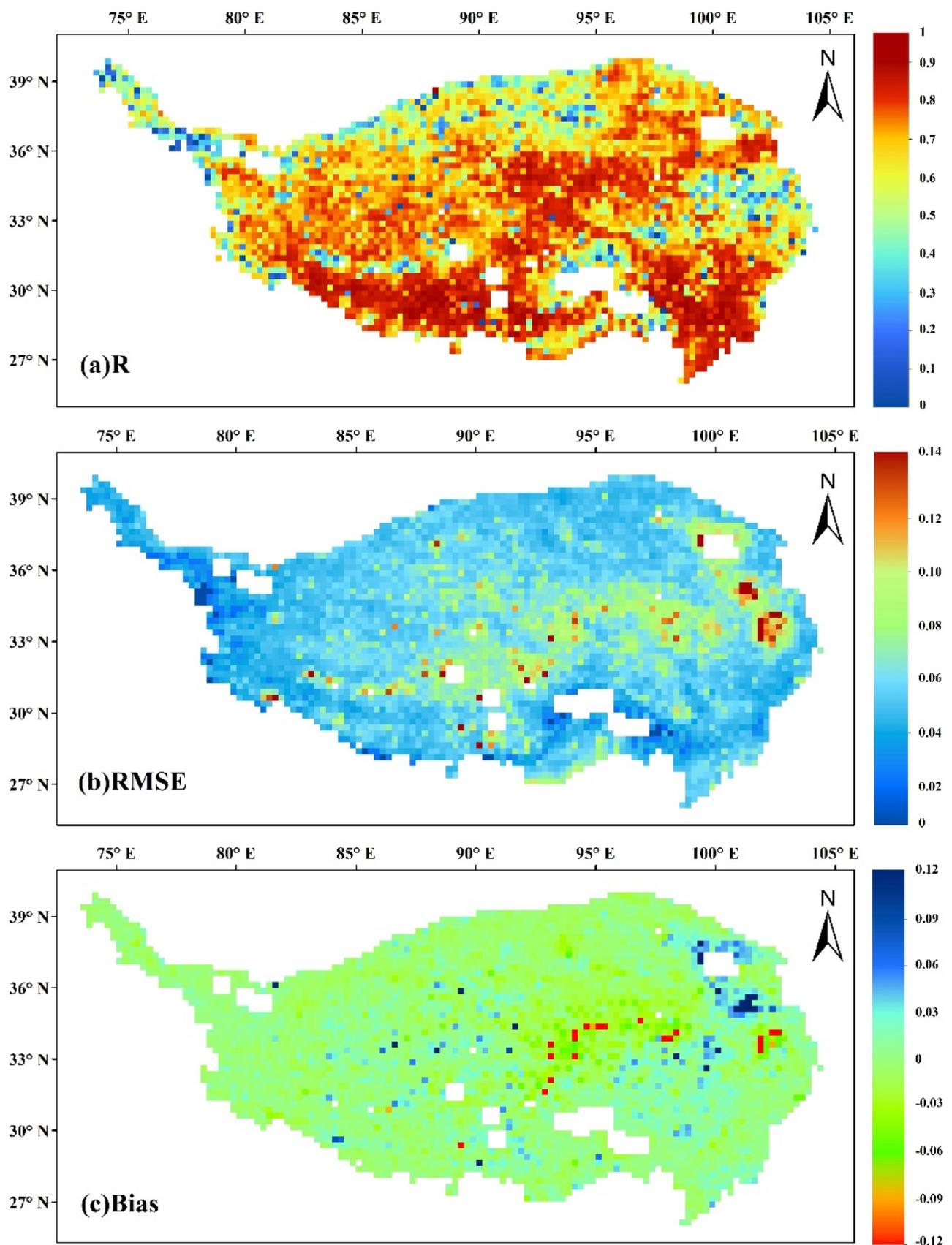
Figure 4. Error distribution maps of ANN-SM vs. SMAP soil moisture during the test period (April 2019 to March 2021) in terms of (**a**) R, (**b**) RMSE, and (**c**) BIAS.
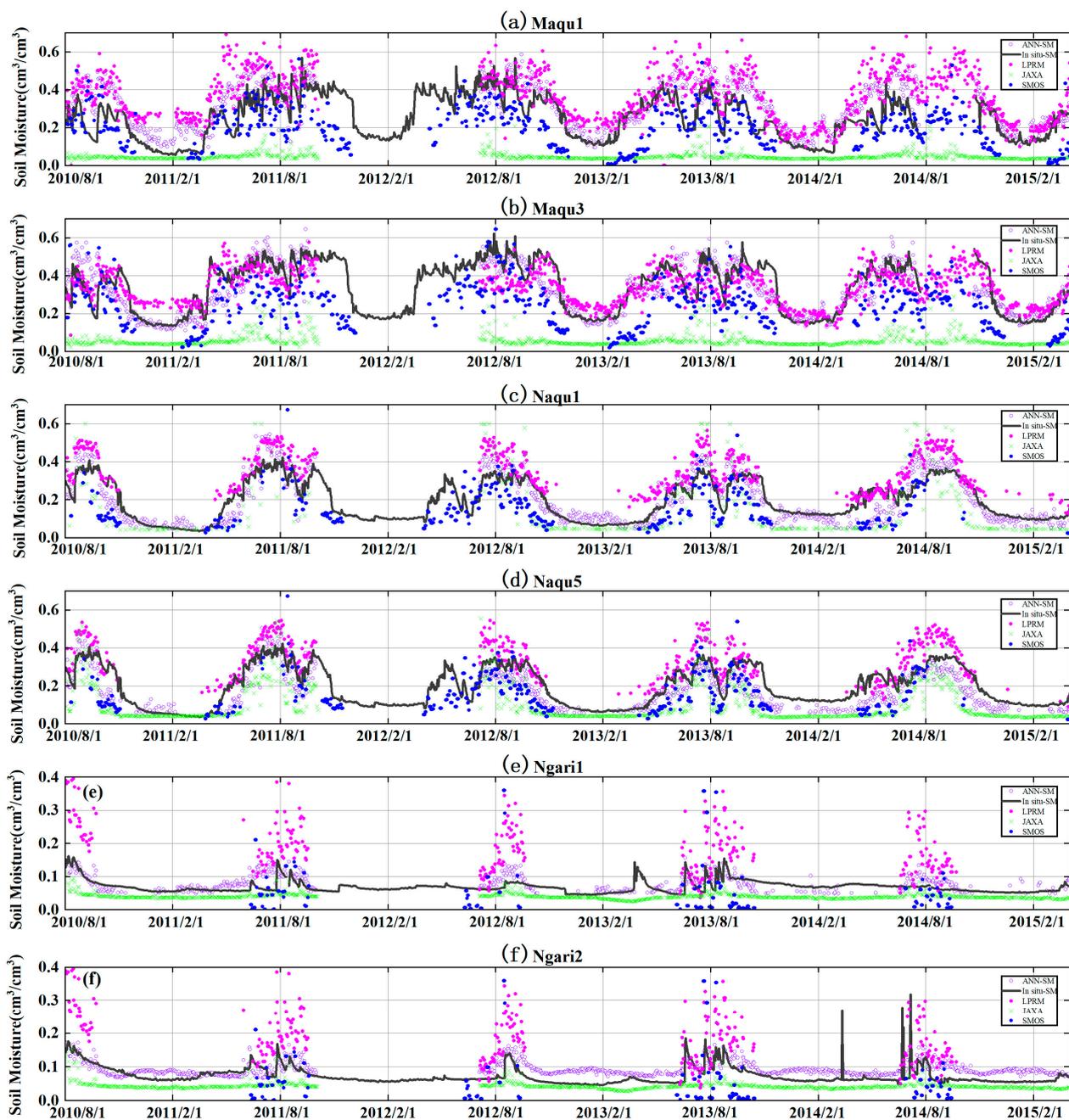
Figure 5. Temporal series scatterplots for ANN-SM (purple hollow circles), LPRMsm (purple dots), JAXAsm (green cross), and SMOSsm (blue dots) vs. in situ soil moisture (black solid line) during a time period (August 2010 to March 2015) of (**a**) Maqu1 grid, (**b**) Maqu3 grid, (**c**) Naqu1 grid, (**d**) Naqu5 grid, (**e**) Ngari1 grid, and (**f**) Ngari2 grid.

To better understand the training accuracy of each grid, the scatter plots of all grids are drawn in Figure 6, and the corresponding R, RMSE and BIAS values of each grid are shown in it. As can be seen from the scatterplots in Figure 6, most of the points over these three networks are located near the 1:1 line, while there are still some points far away from the 1:1 line, which negatively affect the comparative results. Due to the influence of the spatial scale effect, the in situ soil moisture as a ground verification point cannot fully reflect the soil moisture of the ANN-SM pixel, and the more valid sites in the ANN-SM pixel, the more accurately the true soil moisture value can be represented. On the whole, the accuracy of comparative analysis between ANN-SM and in situ soil moisture for each grid was acceptable, with the average value of R = 0.74, the average value of RMSE = 0.106 cm^3/cm^3, and the average absolute value of BIAS = 0.0610 cm^3/cm^3.
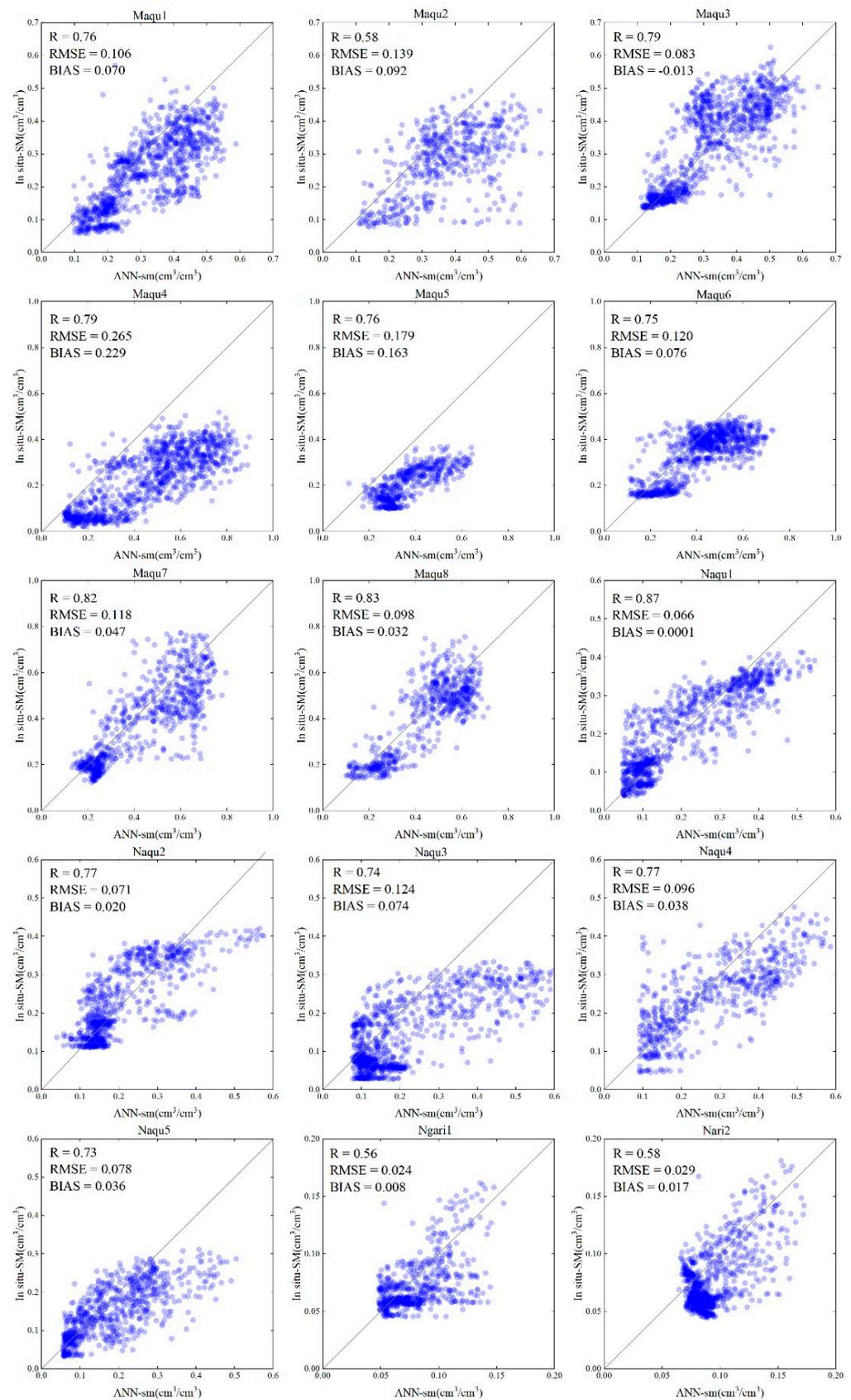
Figure 6. The scatterplots for ANN-SM (x axis) vs. in situ soil moisture (y axis) during a time period (August 2010 to March 2015) of all grids.

5. Discussion

5.1. Optimal Parameters from Sensitivity Analysis

As an important machine learning method, ANN showed good application potential in the fusion and reconstruction of multi-source soil moisture data, while the model was also greatly affected by the choice of parameters. This study focused on the sensitivity analysis of the combination of the input variables, number of neurons in the hidden layer, sample ratios, and precipitation thresholds, finding that the different parameters have a significant impact on the training results of the ANN model. For example, in regard to the combination of different input variables, multi-variable input is beneficial to improve the accuracy of an ANN model compared with the input of only the first five frequencies of AMSR2 brightness temperature. In addition, in terms of the number of neurons in hidden layer, a rule of thumb was used to set the number of neurons ranging from 2 to 12. A tendency to reduce RMSE and increase R of the overall samples can be seen when the number of neurons in hidden layer is increased, as found by other researchers [54–57]. However, the variation trend in RMSE of validation and testing samples decreases with the increase of the number of neurons, which had no obvious effect on the overall samples that tremendously influenced by the training samples rather than validation and testing samples. Furthermore, after using a number of neurons equal to 6, RMSE values of the overall samples stabilized near 0.1 cm^3/cm^3, which indicated that all networks with a number of neurons greater than 6 have relatively equal potential. This study adopted the number of neurons = 12 for all networks due to highest training accuracy can be achieved. In addition, with respect to the sample ratios, 13 different combinations of sample ratios were used in the study, finding that the ANN model has the highest accuracy when the training-validation-testing samples are 80%-10%-10%. Few studies explored the impact of different sample ratio combinations on the ANN model accuracy. Instead, the default sample ratio (70% was used to train the network, and the remaining 30% was utilized to validate and test the network) is usually used for model training [25,58]. However, ANN model performance varied greatly due to different sample ratios. Generally, the accuracy of the overall samples is largely determined by the training samples' performance. Although the default sample ratio made the proportion among training, validation, and testing samples look reasonable, further increasing the ratio of training samples (such as 80% training samples) indeed resulted in higher ANN model accuracy. However, the ratio of 80% training samples may be close to the limit value, because too many training samples will squeeze the proportion of other samples, which led to insufficient validation and testing for the ANN model. Finally, with respect to precipitation thresholds, nine different precipitation threshold values were utilized in the sensitivity analysis. It was found that different precipitation thresholds had little effect on the accuracy of the ANN model due to few ineffective dates of each grid cell in QTP that were screened out by the precipitation thresholds. It should be mentioned that large precipitation thresholds (heavy rain and torrential rain) were used in this study to preserve more days of input and output variables. It can be seen from Figure A3 that most regions of QTP are mountain climate areas with scarce precipitation, and the number of ineffective days of grid cells in this area is only within 10 days. Only a small part of southern Tibet valley and the small eastern part of the QTP with abundant rainfall, which belongs to the subtropical monsoon climate or the tropical monsoon climate that are affected by the southwest monsoon in the Indian Ocean. The number of ineffective dates of grid cells in that area is more than 20 days. Therefore, different precipitation thresholds had little influence on the accuracy of the ANN model, especially after averaging the accuracy of all grid cells of QTP. However, it is necessary to eliminate ineffective days by precipitation, which is beneficial to improve model performance [35]. The precipitation threshold = 8.75 mm was adopted in this study due to its best model performance.

5.2. Selection of Auxiliary Variables

The selection of auxiliary variables for ANN model training also needs to be considered. Selecting the appropriate auxiliary variables can improve the accuracy of the ANN model. For machine learning algorithms applied to multi-source soil moisture data fusion and reconstruction, previous studies mainly selected hydrological variables, geospatial variables, and vegetation variables as the auxiliary variables, such as soil moisture, evapotranspiration, rainfall, temperature, digital elevation model (DEM), slope, aspect, latitude, longitude, and NDVI, MVI, LAI, vegetation classification, Albedo, etc. The environmental information about vegetation, soil temperature, and surface temperature are the most closely related to soil moisture. Therefore, we choose the GLDAS datasets (including soil moisture, soil temperature, and land surface temperature) and MODIS LAI as the auxiliary input variables in order to improve the accuracy of ANN model training. Considering the best combination of different parameters and the use of auxiliary variables, ANN model achieved acceptable accuracy with the average RMSE = 0.096 cm^3/cm^3 and the mean R of overall samples = 0.90 for all grid cells of QTP. Then, we used the ANN model with highest accuracy to simulate long-term series soil moisture and obtained the ANN-SM soil moisture product of QTP. To quantitatively evaluate the effectiveness of ANN-SM, the in situ soil moisture observations from the soil moisture ground measurement networks were used to compare with the ANN-SM from August 2010 to March 2015, with mean R = 0.74, mean RMSE = 0.106 cm^3/cm^3, and the mean absolute BIAS = 0.0610 cm^3/cm^3 for all grids. Finally, to further clarify the advantages of our ANN-SM product, satellite standard soil moisture products of AMSR-E/AMSR2 from JAXA and LPRM and SMOS were used for comparative verification. The results indicated that ANN-SM is more consistent with in situ soil moisture observation.

5.3. Comparison of Different Reconstruction Products

Many scholars have conducted long time series soil moisture products reconstruction in the QTP. For example, the RFSM soil moisture product of QTP that based on random forest method had high accuracy by comparing RFSM with the in situ soil moisture observation (R = 0.75, RMSE = 0.06 m^3/m^3, and bias = -0.03 m^3/m^3) [28]; a new fused soil moisture product in QTP using GRNN to train essential climate variables (ECV) and Fengyun (FY) SM had acceptable accuracy by comparing it with original ECV and FY SM (R = 0.809, RMSE = 0.081 cm^3/cm^3, and bias = 0.050 cm^3/cm^3) [31]; and a global soil moisture product (NNsm) utilizing NN to train AMSR-E/AMSR-2 and SMOS obtained high-accuracy NNsm data by comparing the data with the in situ soil moisture (R = 0.52, RMSE = 0.084 m^3/m^3 and Bias = -0.002 m^3/m^3 for the all land grid cells of global) [59]. Compared with the previous soil moisture products mentioned above, the R and BIAS of the ANNSM in this study were acceptable, but the RMSE was relatively low. One possible reason is the uncertainty that remained in the input and output data involved in the ANN training and the validation error in the spatial scale mismatch between the in situ soil moisture sites and ANN-SM.

5.4. Limitations of the Method

The reconstructed soil moisture product ANN-SM in this study indicated that ANN is a feasible and effective approach for rebuilding long time soil moisture products and this soil moisture product is promising for application to research that requires long time series soil moisture observation records. However, it is worth mentioning that the accuracy of ANN-SM soil moisture product still has room for improvement. The ANN models could be further modified by improving the following limitations: (1) in our research, the data involved in the ANN model had limitations due to the inconsistent soil depth in different datasets (SMAP is 0–5 cm, AMSR-2/AMSR-E is 0–1 cm, and GLDAS is 0–10 cm), and more auxiliary variables, for example, vegetation index, soil types, and other conditioning factors, could be involved for studying the impact of different auxiliary variables on ANN model accuracy; (2) the training process of ANN models had unavoidable systematic uncertainty

that was influenced by the random initialization of free parameters (weights and biases), and different results were acquired in each training session; hence, it is recommended to train the same network several times to obtain more consistent results with higher precision (most of the grid cells in QTP were only trained once in this study); (3) the validation of ANN-SM still remained unclear due to the spatial scale mismatch between point-scale in situ soil moisture measurements and pixel-scale ANN-SM product; strategies that can upscale in situ soil moisture observations to a coarse resolution should be considered in the future studies.

6. Conclusions

In this study, a machine learning method, ANN, has been used to investigate its potential in rebuilding a long time series soil moisture product. Specifically, ANN was utilized to reconstruct a new soil moisture product over the QTP based on AMSR-2/AMSR-E brightness temperature and SMAP satellite by introducing the soil moisture, land surface temperature, and soil temperature of GLDAS and vegetation information of MODIS LAI as auxiliary variables, and considering the impact from the different combination of input variables, the number of neurons in hidden layer, sample ratios, and precipitation thresholds. In response to these research problems, a series of sensitivity experiments were designed to investigate the overall ANN performance. It is concluded that the highest ANN model accuracy can be obtained when all variables were used as inputs and network containing 12 neurons in the hidden layer, when the sample ratio is 80%-10%-10% (training samples-validation samples-testing samples), and when the precipitation threshold is 8.75 mm. Validation of the model in other period shows that ANN-SM is highly consistent with the SMAP soil moisture over the test period (April 2019 to July 2021), with mean $R = 0.72$, and mean RMSE = 0.034 cm^3/cm^3, and mean absolute BIAS = 0.0122 cm^3/cm^3. The ANN-SM is also consistent with the in situ soil moisture over a time period (August 2010 to March 2015), with mean $R = 0.74$, mean RMSE = 0.106 cm^3/cm^3, and the mean absolute BIAS = 0.0610 cm^3/cm^3 for all grids. Our study demonstrated that ANN is a feasible and effective approach for rebuilding spatial-temporal consistent long time series soil moisture products.

Author Contributions: Conceptualization, D.X. and J.W.; data curation, J.W.; formal analysis, D.X. and J.W.; methodology, J.W. and D.X.; writing—original draft preparation, J.W. and D.X.; supervision, D.X.; funding acquisition, D.X. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The AMSR-E/AMSR2 brightness temperature products can be downloaded at https://gportal.jaxa.jp/ (accessed on 14 December 2021). The SMAP L3 daily SM products are freely available at https://nsidc.org/data/ (accessed on 14 December 2021). The GLDAS datasets are available at https://disc.gsfc.nasa.gov/ (accessed on 14 December 2021). The Level-4 MODIS LAI can be downloaded at https://ladsweb.modaps.eosdis.nasa.gov/ (accessed on 14 December 2021). The soil moisture ground measurement networks data can be download at http://data.tpdc.ac.cn (accessed on 14 December 2021). The SMOS data are available at http://www.catds.fr/Products/Products-access (accessed on 14 December 2021). The soil moisture products from JAXA can be downloaded at https://gportal.jaxa.jp/ (accessed on 14 December 2021). The soil moisture products from LPRM can be downloaded at https://disc.gsfc.nasa.gov/ (accessed on 14 December 2021).

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

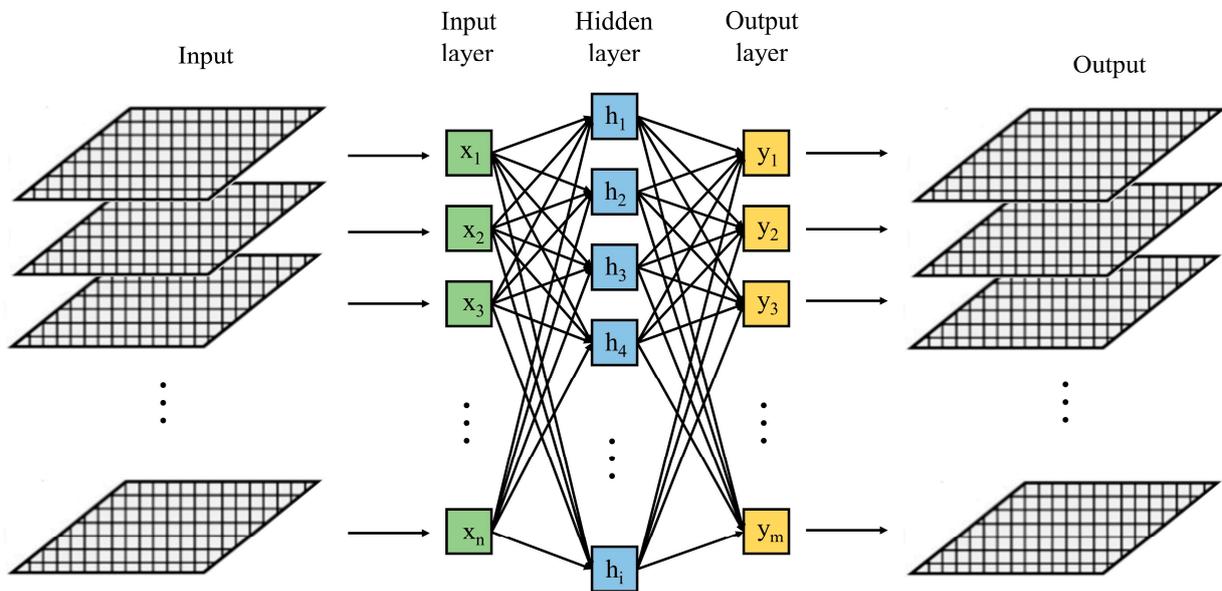**Appendix A**

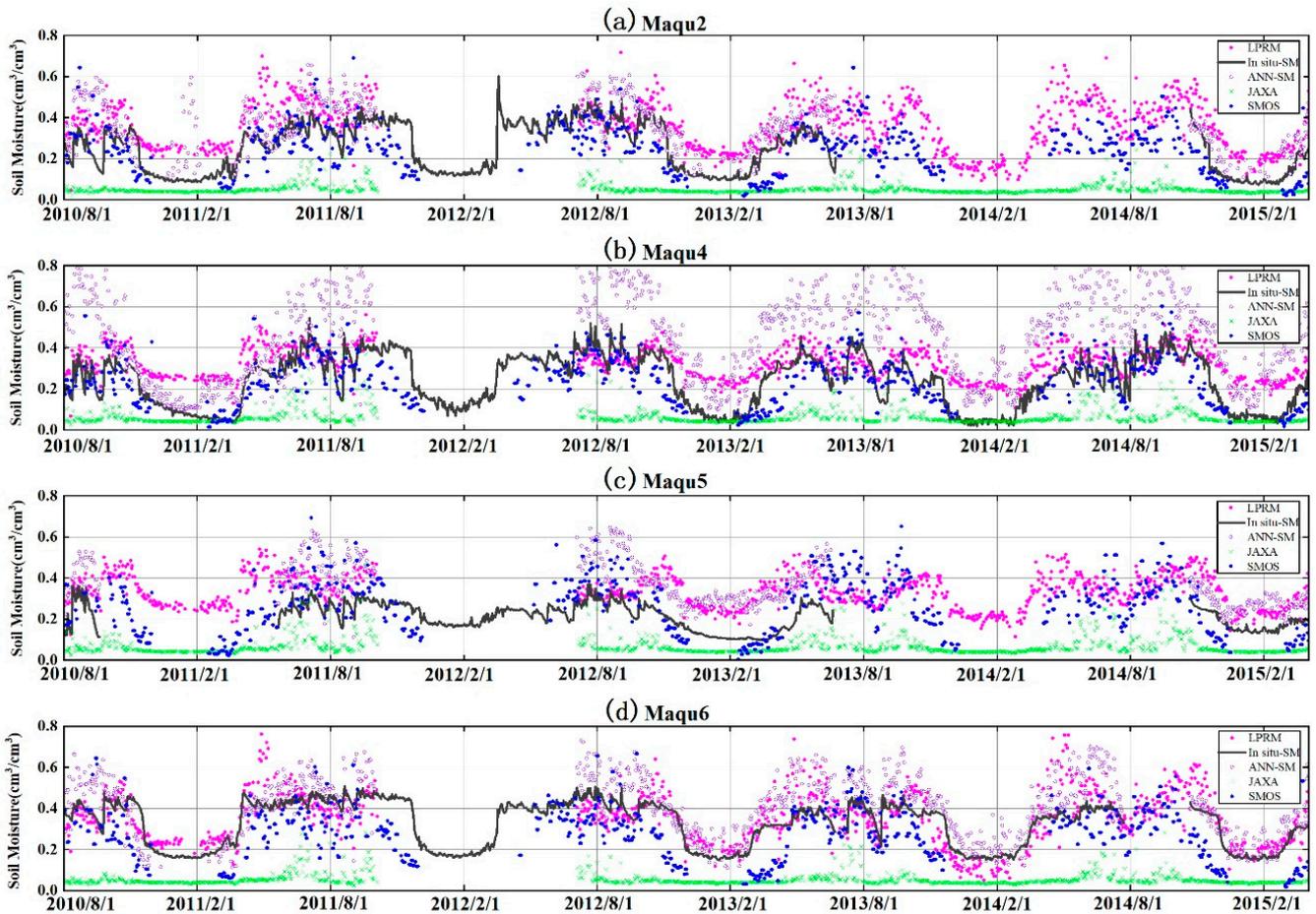**Figure A1.** Flowchart showing the training process of the ANN method.
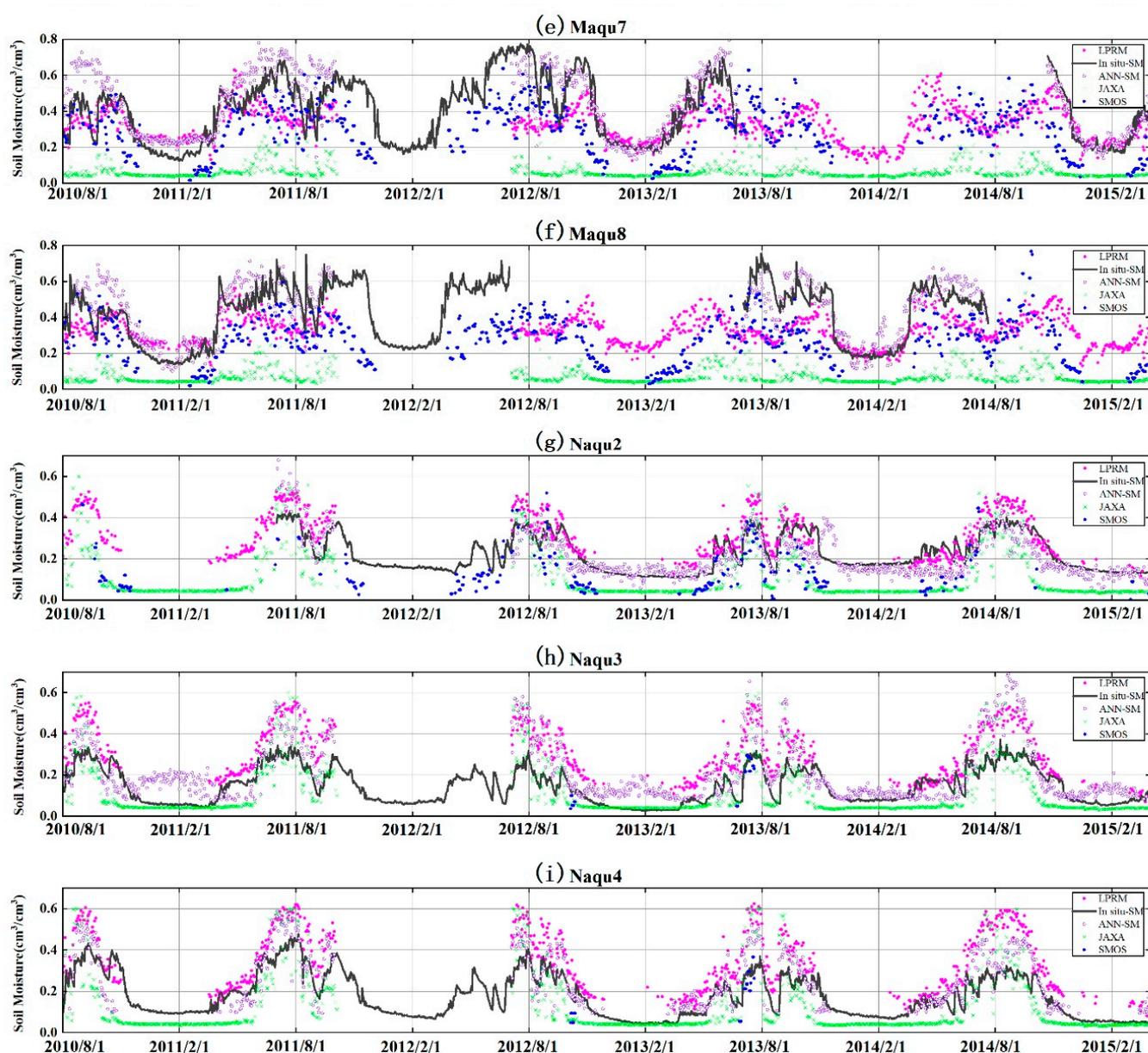

Figure A2. *Cont.*

Figure A2. Temporal series scatterplots for ANN-SM (purple hollow circles), LPRMsm (purple dots), JAXAsm (green cross), and SMOSsm (blue dots) vs. in situ soil moisture (black solid line) during a time period (August 2010 to March 2015) of (**a**) Maqu2 grid, (**b**) Maqu4 grid, (**c**) Maqu5 grid, (**d**) Maqu6 grid, (**e**) Maqu7 grid, (**f**) Maqu8 grid, (**g**) Naqu2 grid, (**h**) Naqu3 grid, and (**i**) Naqu4 grid.
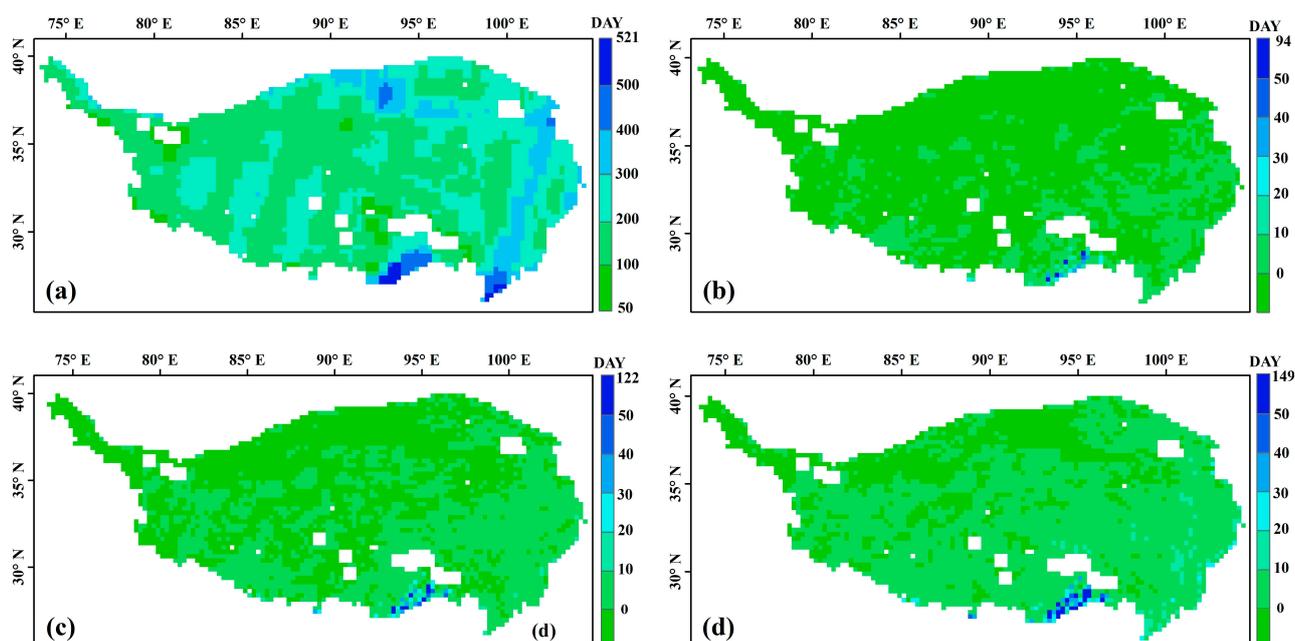
Figure A3. The number of valid days retained after data removal through different precipitation thresholds. The number of days reserved when the precipitation threshold is 0 mm (**a**), and the number of further reduced days when the precipitation threshold is (**b**) 17.5 mm, (**c**) 12.5 mm, and (**d**) 8.75 mm.

References

1. Dobriyal, P.; Qureshi, A.; Badola, R.; Hussain, S.A. A review of the methods available for estimating soil moisture and its implications for water resource management. *J. Hydrol.* **2012**, *458–459*, 110–117. [CrossRef]
2. Gomez-Casanovas, N.; Matamala, R.; Cook, D.R.; Gonzalez-Meler, M.A. Net ecosystem exchange modifies the relationship between the autotrophic and heterotrophic components of soil respiration with abiotic factors in prairie grasslands. *Glob. Chang. Biol.* **2012**, *18*, 2532–2545. [CrossRef]
3. Tao, F.; Yokozawa, M.; Hayashi, Y.; Lin, E. Changes in agricultural water demands and soil moisture in China over the last half-century and their effects on agricultural production. *Agric. Forest Meteorol.* **2003**, *118*, 251–261. [CrossRef]
4. Zhuo, W.; Huang, J.; Li, L.; Zhang, X.; Ma, H.; Gao, X.; Huang, H.; Xu, B.; Xiao, X. Assimilating Soil Moisture Retrieved from Sentinel-1 and Sentinel-2 Data into WOFOST Model to Improve Winter Wheat Yield Estimation. *Remote Sens.* **2019**, *11*, 1618. [CrossRef]
5. Seneviratne, S.I.; Corti, T.; Davin, E.L.; Hirschi, M.; Jaeger, E.B.; Lehner, I.; Orlowsky, B.; Teuling, A.J. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Sci. Rev.* **2010**, *99*, 125–161. [CrossRef]
6. Brocca, L.; Crow, W.T.; Ciabatta, L.; Massari, C.; de Rosnay, P.; Enenkel, M.; Hahn, S.; Amarnath, G.; Camici, S.; Tarpanelli, A.; et al. A Review of the Applications of ASCAT Soil Moisture Products. *IEEE J.-STARS* **2017**, *10*, 2285–2306. [CrossRef]
7. Karthikeyan, L.; Pan, M.; Wanders, N.; Kumar, D.N.; Wood, E.F. Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms. *Adv. Water Resour.* **2017**, *109*, 106–120. [CrossRef]
8. Laachrate, H.; Fadil, A.; Ghafiri, A. Soil Moisture Retrieval Using Microwave Remote Sensing: Review of Techniques and Applications. In *Geospatial Technology: Application in Water Resources Management*; Jarar Oulidi, H., Fadil, A., Semane, N., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 31–50.
9. Peng, J.; Albergel, C.; Balenzano, A.; Brocca, L.; Cartus, O.; Cosh, M.H.; Crow, W.T.; Dabrowska-Zielinska, K.; Dadson, S.; Davidson, M.W.J.; et al. A roadmap for high-resolution satellite soil moisture applications—Confronting product characteristics with user requirements. *Remote Sens. Environ.* **2021**, *252*, 112162. [CrossRef]
10. Edokossi, K.; Calabia, A.; Jin, S.; Molina, I. GNSS-Reflectometry and Remote Sensing of Soil Moisture: A Review of Measurement Techniques, Methods, and Applications. *Remote Sens.* **2020**, *12*, 614. [CrossRef]
11. Njoku, E.; Chan, S. Vegetation and surface roughness effects on AMSR-E land observations. *Remote Sens. Environ.* **2006**, *100*, 190–199. [CrossRef]
12. Reul, N.; Grodsky, S.A.; Arias, M.; Boutin, J.; Catany, R.; Chapron, B.; D'Amico, F.; Dinnat, E.; Donlon, C.; Fore, A.; et al. Sea surface salinity estimates from spaceborne L-band radiometers: An overview of the first decade of observation (2010–2019). *Remote Sens. Environ.* **2020**, *242*, 11769. [CrossRef]
13. Zhao, T. Recent advances of L-band application in the passive microwave remote sensing of soil moisture and its prospects. *Prog. Geogr.* **2018**, *37*, 198–213.

14. Sabaghy, S.; Walker, J.P.; Renzullo, L.J.; Jackson, T.J. Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. *Remote Sens. Environ.* **2018**, *209*, 551–580. [CrossRef]

15. Wigneron, J.-P.; Li, X.; Frappart, F.; Fan, L.; Al-Yaari, A.; De Lannoy, G.; Liu, X.; Wang, M.; Le Masson, E.; Moisy, C. SMOS-IC data record of soil moisture and L-VOD: Historical development, applications and perspectives. *Remote Sens. Environ.* **2021**, *254*, 112238. [CrossRef]

16. Wigneron, J.P.; Jackson, T.J.; O'Neill, P.; De Lannoy, G.; de Rosnay, P.; Walker, J.P.; Ferrazzoli, P.; Mironov, V.; Bircher, S.; Grant, J.P.; et al. Modelling the passive microwave signature from land surfaces: A review of recent results and application to the L-band SMOS SMAP soil moisture retrieval algorithms. *Remote Sens. Environ.* **2017**, *192*, 238–262. [CrossRef]

17. Chanda, K.; Maity, R.; Sharma, A.; Mehrotra, R. Spatiotemporal variation of long-term drought propensity through reliability-resilience-vulnerability based Drought Management Index. *Water Resour. Res.* **2014**, *50*, 7662–7676. [CrossRef]

18. Liu, Y.Y.; van Dijk, A.I.J.M.; de Jeu, R.A.M.; Holmes, T.R.H. An analysis of spatiotemporal variations of soil and vegetation moisture from a 29-year satellite-derived data set over mainland Australia. *Water Resour. Res.* **2009**, *45*, W07405. [CrossRef]

19. Moesinger, L.; Dorigo, W.; de Jeu, R.; van der Schalie, R.; Scanlon, T.; Teubner, I.; Forkel, M. The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA). *Earth Syst. Sci. Data* **2020**, *12*, 177–196. [CrossRef]

20. Reichle, R.H. Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett.* **2004**, *31*, L19501. [CrossRef]

21. Gruber, A.; Scanlon, T.; van der Schalie, R.; Wagner, W.; Dorigo, W. Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth Syst. Sci. Data* **2019**, *11*, 717–739. [CrossRef]

22. Salcedo-Sanz, S.; Ghamisi, P.; Piles, M.; Werner, M.; Cuadra, L.; Moreno-Martínez, A.; Izquierdo-Verdiguier, E.; Muñoz-Marí, J.; Mosavi, A.; Camps-Valls, G. Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources. *Inform. Fusion.* **2020**, *63*, 256–272. [CrossRef]

23. Sharma, A.; Jain, A.; Gupta, P.; Chowdary, V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access.* **2021**, *9*, 4843–4873. [CrossRef]

24. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [CrossRef]

25. Yao, P.; Lu, H.; Shi, J.; Zhao, T.; Yang, K.; Cosh, M.H.; Gianotti, D.J.S.; Entekhabi, D. A long term global daily soil moisture dataset derived from AMSR-E and AMSR2 (2002–2019). *Sci. Data* **2021**, *8*, 143. [CrossRef]

26. Cui, Y.; Zeng, C.; Zhou, J.; Xie, H.; Wan, W.; Hu, L.; Xiong, W.; Chen, X.; Fan, W.; Hong, Y. A spatio-temporal continuous soil moisture dataset over the Tibet Plateau from 2002 to 2015. *Sci. Data* **2019**, *6*, 247. [CrossRef] [PubMed]

27. Zhang, Q.; Yuan, Q.; Li, J.; Wang, Y.; Sun, F.; Zhang, L. Generating seamless global daily AMSR2 soil moisture (SGD-SM) long-term products for the years 2013–2019. *Earth Syst. Sci. Data* **2021**, *13*, 1385–1401. [CrossRef]

28. Qu, Y.; Zhu, Z.; Chai, L.; Liu, S.; Montzka, C.; Liu, J.; Yang, X.; Lu, Z.; Jin, R.; Li, X.; et al. Rebuilding a Microwave Soil Moisture Product Using Random Forest Adopting AMSR-E/AMSR2 Brightness Temperature and SMAP over the Qinghai–Tibet Plateau, China. *Remote Sens.* **2019**, *11*, 683. [CrossRef]

29. Cui, Y.; Yang, X.; Chen, X.; Fan, W.; Zeng, C.; Xiong, W.; Hong, Y. A two-step fusion framework for quality improvement of a remotely sensed soil moisture product: A case study for the ECV product over the Tibetan Plateau. *J. Hydrol.* **2020**, *587*, 124993. [CrossRef]

30. Yao, P.P.; Lu, H.; Yue, S.Y.; Yang, F.; Lyu, H.B.; Yang, K.; McColl, K.A.; Gianotti, D.; Entekhabi, D. Estimating Surface Soil Moisture from AMSR2 TB with Artificial Neural Network Method and SMAP Products. In Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2019), Yokohama, Japan, 28 July–2 August 2019; pp. 6998–7001.

31. Cui, Y.; Zeng, C.; Chen, X.; Fan, W.; Liu, H.; Liu, Y.; Xiong, W.; Sun, C.; Luo, Z. A New Fusion Algorithm for Simultaneously Improving Spatio-Temporal Continuity and Quality of Remotely Sensed Soil Moisture Over the Tibetan Plateau. *IEEE J.-STARS* **2021**, *14*, 83–91. [CrossRef]

32. Chen, M.; Willgoose, G.R.; Saco, P.M. Investigating the impact of leaf area index temporal variability on soil moisture predictions using remote sensing vegetation data. *J. Hydrol.* **2015**, *522*, 274–284. [CrossRef]

33. Hall, F.G.; Townshend, J.R.; Engman, E.T. status of remote-sensing algorithms for estimation of land-surface state parameters. *Remote Sens. Environ.* **1995**, *51*, 138–156. [CrossRef]

34. Santi, E.; Paloscia, S.; Pettinato, S.; Brocca, L.; Ciabatta, L.; Entekhabi, D. Integration of microwave data from SMAP and AMSR2 for soil moisture monitoring in Italy. *Remote Sens. Environ.* **2018**, *212*, 21–30. [CrossRef]

35. Leroux, D.J.; Kerr, Y.H.; Wood, E.F.; Sahoo, A.K.; Bindlish, R.; Jackson, T.J. An Approach to Constructing a Homogeneous Time Series of Soil Moisture Using SMOS. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 393–405. [CrossRef]

36. Chen, Y.; Yang, K.; Qin, J.; Cui, Q.; Lu, H.; La, Z.; Han, M.; Tang, W. Evaluation of SMAP, SMOS, and AMSR2 soil moisture retrievals against observations from two networks on the Tibetan Plateau. *J. Geophys. Res. Atmos.* **2017**, *122*, 5780–5792. [CrossRef]

37. Chen, Y.; Yang, K.; Qin, J.; Zhao, L.; Tang, W.; Han, M. Evaluation of AMSR-E retrievals and GLDAS simulations against observations of a soil moisture network on the central Tibetan Plateau. *J. Geophys. Res. Atmos.* **2013**, *118*, 4466–4475. [CrossRef]

38. Kumar, S.V.; Dirmeyer, P.A.; Peters-Lidard, C.D.; Bindlish, R.; Bolten, J. Information theoretic evaluation of satellite soil moisture retrievals. *Remote Sens. Environ.* **2018**, *204*, 392–400. [CrossRef]

39. Wang, Z.; Che, T.; Zhao, T.; Dai, L.; Li, X.; Wigneron, J.-P. Evaluation of SMAP, SMOS, and AMSR2 Soil Moisture Products Based on Distributed Ground Observation Network in Cold and Arid Regions of China. *IEEE J.-STARS* **2021**, *14*, 8955–8970. [CrossRef]

40. Bengio, Y.; Courville, A.; Vincent, P. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [CrossRef] [PubMed]

41. Liu, J.; Chai, L.; Lu, Z.; Liu, S.; Qu, Y.; Geng, D.; Song, Y.; Guan, Y.; Guo, Z.; Wang, J.; et al. Evaluation of SMAP, SMOS-IC, FY3B, JAXA, and LPRM Soil Moisture Products over the Qinghai–Tibet Plateau and Its Surrounding Areas. *Remote Sens.* **2019**, *11*, 792. [CrossRef]

42. Pan, M.; Cai, X.; Chaney, N.W.; Entekhabi, D.; Wood, E.F. An initial assessment of SMAP soil moisture retrievals using high-resolution model simulations and in situ observations. *Geophys. Res. Lett.* **2016**, *43*, 9662–9668. [CrossRef]

43. Wang, J.; Jiang, L.; Cui, H.; Wang, G.; Yang, J.; Liu, X.; Su, X. Evaluation and analysis of SMAP, AMSR2 and MEaSUREs freeze/thaw products in China. *Remote Sens. Environ.* **2020**, *242*, 111734. [CrossRef]

44. Zhang, X.; Zhang, T.; Zhou, P.; Shao, Y.; Gao, S. Validation Analysis of SMAP and AMSR2 Soil Moisture Products over the United States Using Ground-Based Measurements. *Remote Sens.* **2017**, *9*, 104. [CrossRef]

45. Sun, H.; Zheng, D.; Yao, T.; Zhang, Y. Protection and Construction of the National Ecological Security Shelter Zone on Tibetan Plateau. *Acta Geogr. Sin.* **2012**, *67*, 3–12.

46. Van der Velde, R.; Su, Z.; van Oevelen, P.; Wen, J.; Ma, Y.; Salama, M.S. Soil moisture mapping over the central part of the Tibetan Plateau using a series of ASAR WS images. *Remote Sens. Environ.* **2012**, *120*, 175–187. [CrossRef]

47. Yang, K.; Qin, J.; Zhao, L.; Chen, Y.; Tang, W.; Han, M.; Chen, Z.; Lv, N.; Ding, B.; Wu, H.; et al. A Multiscale Soil Moisture and Freeze–Thaw Monitoring Network on the Third Pole. *Bull. Am. Meteorol. Soc.* **2013**, *94*, 1907–1916. [CrossRef]

48. Yang, K.; Su, B. *Time-Lapse Observation Dataset of Soil Temperature and Humidity on the Tibetan Plateau (2008–2016)*; National Tibetan Plateau Data Center: Beijing, China, 2019.

49. Su, Z.; Wen, J.; Dente, L.; van der Velde, R.; Wang, L.; Ma, Y.; Yang, K.; Hu, Z. The Tibetan Plateau observatory of plateau scale soil moisture and soil temperature (Tibet-Obs) for quantifying uncertainties in coarse resolution satellite and model products. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 2303–2316. [CrossRef]

50. Dente, L.; Vekerdy, Z.; Wen, J.; Su, Z. Maqu network for validation of satellite-derived soil moisture products. *Int. J. Appl. Earth Obs.* **2012**, *17*, 55–65. [CrossRef]

51. Su, Z.; de Rosnay, P.; Wen, J.; Wang, L.; Zeng, Y. Evaluation of ECMWF's soil moisture analyses using observations on the Tibetan Plateau. *J. Geophys. Res. Atmos.* **2013**, *118*, 5304–5318. [CrossRef]

52. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

53. 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]

54. Aladag, C.H.; Kayabasi, A.; Gokceoglu, C. Estimation of pressuremeter modulus and limit pressure of clayey soils by various artificial neural network models. *Neural Comput. Appl.* **2012**, *23*, 333–339. [CrossRef]

55. Arsoy, S.; Ozgur, M.; Keskin, E.; Yilmaz, C. Enhancing TDR based water content measurements by ANN in sandy soils. *Geoderma* **2013**, *195–196*, 133–144. [CrossRef]

56. Collins, W.; Tissot, P. An artificial neural network model to predict thunderstorms within 400 km2 South Texas domains. *Meteorol. Appl.* **2015**, *22*, 650–665. [CrossRef]

57. Zanetti, S.S.; Cecílio, R.A.; Alves, E.G.; Silva, V.H.; Sousa, E.F. Estimation of the moisture content of tropical soils using colour images and artificial neural networks. *Catena* **2015**, *135*, 100–106. [CrossRef]

58. Garajeh, M.K.; Malakyar, F.; Weng, Q.; Feizizadeh, B.; Blaschke, T.; Lakes, T. An automated deep learning convolutional neural network algorithm applied for soil salinity distribution mapping in Lake Urmia, Iran. *Sci. Total Environ.* **2021**, *778*, 146253. [CrossRef] [PubMed]

59. Yao, P.; Shi, J.; Zhao, T.; Lu, H.; Al-Yaari, A. Rebuilding Long Time Series Global Soil Moisture Products Using the Neural Network Adopting the Microwave Vegetation Index. *Remote Sens.* **2017**, *9*, 35. [CrossRef]