water

MDPI

Article

Spatial Downscaling Methods of Soil Moisture Based on Multisource Remote Sensing Data and its Application

Shaodan Chen, Dunxian She *, Liping Zhang *, Mengyao Guo and Xin Liu

State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, No. 8 Donghu South Road, Wuhan 430072, China

* Correspondence: shedunxian@whu.edu.cn (D.S.); zhanglp@whu.edu.cn (L.Z.)

Received: 7 May 2019; Accepted: 4 July 2019; Published: 8 July 2019

Abstract: Soil moisture is an important indicator that is widely used in meteorology, hydrology, and agriculture. Two key problems must be addressed in the process of downscaling soil moisture: the selection of the downscaling method and the determination of the environmental variables, namely, the influencing factors of soil moisture. This study attempted to utilize machine learning and data mining algorithms to downscale the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) soil moisture data from 25 km to 1 km and compared the advantages and disadvantages of the random forest model and the Cubist algorithm to determine the more suitable soil moisture downscaling method for the middle and lower reaches of the Yangtze River Basin (MLRYRB). At present, either the normalized difference vegetation index (NDVI) or a digital elevation model (DEM) is selected as the environmental variable for the downscaling models. In contrast, variables, such as albedo and evapotranspiration, are infrequently applied; nevertheless, this study selected these two environmental variables, which have a considerable impact on soil moisture. Thus, the selected environmental variables in the downscaling process included the longitude, latitude, elevation, slope, NDVI, daytime and nighttime land surface temperature (LST_D and LST_N, respectively), albedo, evapotranspiration (ET), land cover (LC) type, and aspect. This study achieved downscaling on a 16-day timescale based on Moderate Resolution Imaging Spectroradiometer (MODIS) data. A comparison of the random forest model with the Cubist algorithm revealed that the R^2 of the random forest-based downscaling method is higher than that of the Cubist algorithm-based method by 0.0161; moreover, the root-mean-square error (RMSE) is reduced by 0.0006 and the mean absolute error (MAE) is reduced by 0.0014. Testing the accuracies of these two downscaling methods showed that the random forest model is more suitable than the Cubist algorithm for downscaling AMSR-E soil moisture data from 25 km to 1 km in the MLRYRB, which provides a theoretical basis for obtaining high spatial resolution soil moisture data.

Keywords: soil moisture; Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E); downscaling; random forest; Cubist

1. Introduction

Soil moisture is an important component of both the water cycle and the surface energy cycle; it is also an important indicator for reflecting land degradation and characterizing surface drought information [1–5]. Soil moisture is related to a number of factors, which include vegetation growth, crop growth, and food production, as well as important parameters in the fields of hydrology, climate research, agriculture, and ecology [6,7]. Consequently, soil moisture has been widely used in various environmental applications, such as hydrological modeling, land surface evapotranspiration simulation,

numerical weather prediction, and surface runoff prediction [8]. Furthermore, timely access to accurate soil moisture information can enable the spatial distribution of soil moisture to be mapped, which is of great significance for improving the efficiency of agricultural irrigation water use, agricultural production, sustainable water resource utilizing, and drought condition monitoring. Unfortunately, the traditional methods that are used to measure soil moisture, which include the drying and weighing method, negative pressure meter method, neutron meter detection method, indirect resistance method, and time domain reflection (TDR) method, suffer from numerous disadvantages, namely, they are limited by sparse sampling, they have poor representativeness and poor dynamics, and they are time consuming and laborious, while offering only a small monitoring range.

When compared with traditional soil moisture monitoring methods, in recent years remote sensing technology can provide a wide range of continuous monitoring, which are dynamic and inexpensive. The remote sensing inversion of soil moisture is mainly divided into optical remote sensing and microwave remote sensing methods [9]. The former is primarily based on the soil color and surface temperature, both of which are affected by soil moisture, and thus visible, near-infrared and thermal infrared remote sensing methods are used for soil moisture monitoring [10,11]. To date, many scholars have performed research on the optical remote sensing inversion of soil moisture. As a result, numerous soil moisture inversion models, such as the normalized difference vegetation index (NDVI), apparent thermal inertia (ATI), and temperature vegetation dryness index (TVDI), have been established [12,13]. However, the signals in the visible, near-infrared, and thermal infrared bands have weak penetration ability and they are easily affected by factors, such as atmospheric conditions and clouds. In contrast, microwaves boast a strong penetration ability, and the surface moisture content of soil can be monitored at any time of day [14]. To some extent, these advantages, namely, the acquisition of all-weather observations and the ability to penetrate clouds, fog, and snow, compensate for the shortcomings of optical remote sensing in soil moisture inversion and they introduce unique advantages into soil moisture monitoring.

Microwave remote sensing is divided into active and passive microwave remote sensing techniques on the basis of different energy sources. Active microwave remote sensors mainly employ synthetic aperture radar, while passive microwave remote sensors are known as microwave radiometers. Although active microwave remote sensing boasts high spatial and temporal resolutions, this approach is susceptible to the surface roughness, vegetation canopy structure, etc., and thus the inversion of soil moisture suffers from considerable uncertainty. The difference between active and passive microwave remote sensing is that microwave radiometers are less affected by these disturbances and they are sensitive to temperature, vegetation, and soil moisture. Because of these advantages, passive microwave remote sensing is widely utilized in hydrology, agriculture, and meteorology. With the development of microwave remote sensing technology, a series of microwave radiometer-based soil moisture products, such as the Scanning Multichannel Microwave Radiometer (SMMR) [15,16], Special Sensor Microwave Imager (SSM/I) [17–19], Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) [20–23], Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) [24–30], Soil Moisture and Ocean Salinity (SMOS) satellite [31,32], and Fengyun (FY) series of satellites launched by China, have emerged [30,33]. These passive microwave remote sensing-based soil moisture products provide large areas of soil moisture distributions, which constitute important basic data for global or regional research; however, 25–40 km is the spatial resolution of most soil moisture products, which makes it difficult for soil moisture data to satisfy the needs of applications, such as hydrological modeling. Therefore, passive microwave products must be downscaled to obtain high-resolution soil moisture data.

Researchers have established a variety of empirical equations that are based on the relationship between soil moisture and environmental variables. However, global linear regression assumes that the distributions of variables are spatially consistent before conducting statistical analysis, and thus local characteristics are ignored. Brunsdon et al. [34] proposed the concept of geographically weighted regression (GWR) based on the principles of local regression to improve the global regression model.

When compared with the global regression model, GWR mainly introduces geographic location information into the regression model [35,36]. However, due to the limitations of the GWR algorithm, it is impossible to effectively screen local environmental variables with the closest soil moisture based on the spatial distribution; as a result, GWR is not applicable for obtaining multiple factors to downscale the soil moisture. To date, following the emergence of machine learning algorithms, the random forest algorithm, which is an ensemble learning algorithm that was developed on the basis of decision trees, has been widely used in many fields, because it provides better ability in capturing the nonlinear relationships between variables. Shi et al. [37] downscaled TRMM products based on the random forest model and obtained precipitation data at a 1 km resolution over the Tibet Plateau. In addition, Ma et al. [38] introduced the Cubist data mining algorithm to downscale TRMM annual precipitation data across the Tibet Plateau from a resolution of 0.25° × 0.25° to 1 km × 1 km. The results demonstrated that the performance of the Cubist algorithm is better than that of the GWR model.

Although both algorithms are very effective for remote sensing precipitation products, the most widely used downscaling algorithm for passive microwave radiometer soil moisture measurements is still the empirical relationship with optical remote sensing images. Therefore, the purpose of this study is to use the random forest and Cubist models to downscale the AMSR-E soil moisture products in the study area to obtain higher-resolution soil moisture data. The specific objectives are as follows: (1) to apply the random forest and Cubist algorithms to downscale AMSR-E passive microwave products from 25 km to 1 km and (2) to compare the downscaling results of these two models to determine the most suitable downscaling algorithm for the study area.

2. Data and Methods

2.1. Study Area

The Yangtze River is one of the major rivers in China, maintaining a length of approximately 6300 km and a catchment area of approximately 1.8 million km^2. We choose the middle and lower reaches of the Yangtze River Basin (MLRYRB, see Figure 1), which lies between 25° and 35° N and between 106° and 122° E, as the study area. The Terra and Aqua combined Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) Version 6 data product adopts five different land cover classification schemes. This study adopts the second classification method of the Annual University of Maryland (UMD) classification, which shows the land cover types in the MLRYRB.

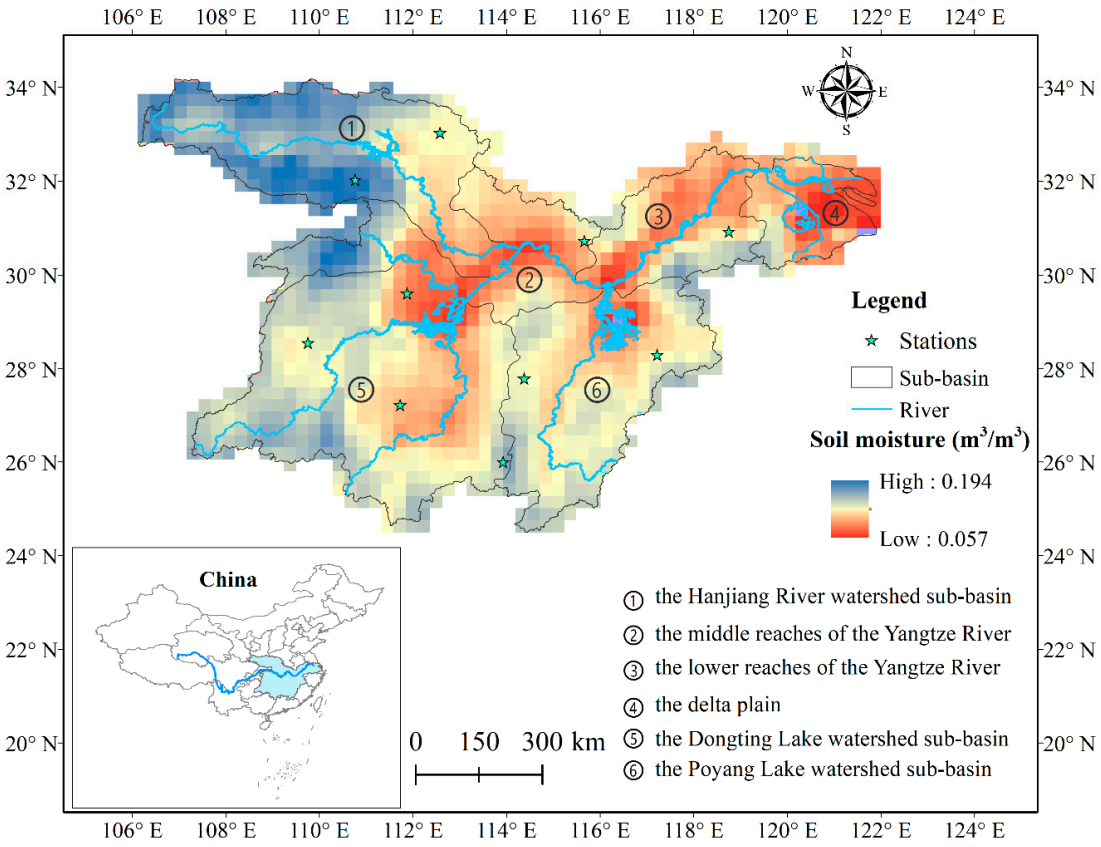


Figure 1. Annual average soil moisture distribution in the middle and lower reaches of the Yangtze River Basin (MLRYRB) based on Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E).

2.2. Data

The AMSR-E is a multichannel passive microwave sensor that was launched on NASA's Earth Observing System (EOS) Aqua satellite in May 2002, with daily ascending (13:30 equatorial local crossing time) and descending (01:30 equatorial local crossing time) overpasses [25,39]. In this study, we select the Level-3 land surface product of the AMSR-E (AE_Land3) onboard NASA's Aqua satellite with a spatial resolution of 25 km × 25 km (Figure 2). The MODIS products that were utilized in this study, including NDVI, daytime land surface temperature (LST_D), nighttime LST (LST_N), albedo, land cover (LC) type, and evapotranspiration (ET) products, are acquired from NASA. Table 1 summarizes the attribute information of the MODIS data set that was used in this study. In addition, the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) product was obtained by the International Scientific & Technical Data Mirror Site, Computer Network Information Center, Chinese Academy of Science, with a spatial resolution of 90 m (http://www.gscloud.cn). The slope and aspect are derived from the DEM data. Additionally, the observed soil moisture data (0–10 cm) in situ soil stations (as shown in Figure 1) used in this study were provided from the China Meteorological Data Sharing Service System (available at http://cdc.nmic.cn/home.do).

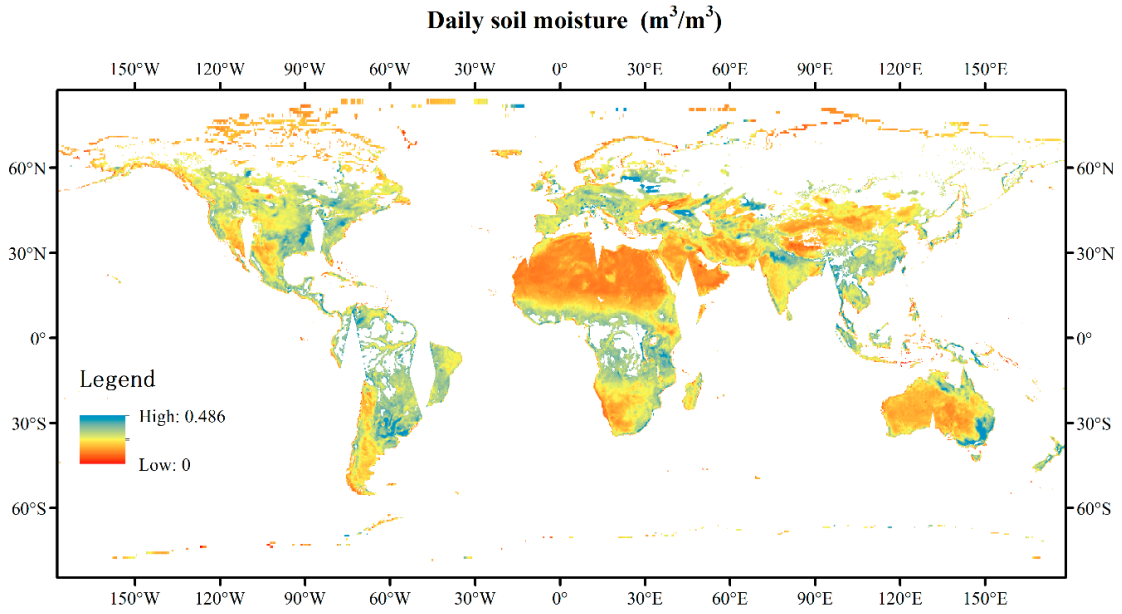


Figure 2. Global soil moisture distribution based on the AMSR-E (taking January 1, 2003, as an example).

Table 1. Attribute information of the Moderate Resolution Imaging Spectroradiometer (MODIS) data.

Scientific Data Sets (SDS)	Spatial Resolution	Scale Factor
MOD13A2 (MODIS/Terra Vegetation Indices 16-Day L3 Global 1 km SIN Grid V006)	1 km	0.0001
MOD11A2 (MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1 km SIN Grid V006)	1 km	0.02
MCD43A3 (MODIS/Terra+Aqua BRDF/Albedo Daily L3 Global-500 m V006)	500 m	0.001
MCD12Q1 (MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500 m SIN Grid V006)	500 m	N/A
MOD16A2 (MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500 m SIN Grid V006)	500 m	0.1

2.3. Methodology

We choose the rule-based machine learning approaches, including the random forest and Cubist algorithms, to downscale the AMSR-E soil moisture data from 25 km to 1 km. The random forest method, which is a popularly used machine-learning method, uses randomization when selecting the features at each node. The Cubist method is a spatial data mining algorithm that applies a divide-and-conquer strategy. The reader is referred to the literature for an introduction to these two machine learning algorithms [38,40–43]. The spatial downscaling method is based on the relationship between soil moisture and various environmental variables. Two key problems must be addressed to downscale the AMSR-E soil moisture data: one is the selection of the downscaling method, and the other is the determination of the surface variables of soil moisture. This study attempts to introduce machine learning algorithms into the downscaling model and compares the advantages and disadvantages of the random forest model with those of the Cubist algorithm to determine the most suitable soil moisture downscaling method for the MLRYRB. The NDVI, DEM, and surface temperature are the most commonly employed environmental variables. In contrast, there are few applications for variables, such as albedo and ET; nevertheless, this study considers these factors to have a greater impact on soil moisture. Therefore, the longitude, latitude, DEM, slope, aspect, NDVI, LST_D, LST_N, albedo, ET, and LC are the final environmental variables that were utilized in the downscaling process.

The basic idea of the downscaling method is to first establish the relationship between soil moisture and all of the environmental variables at a spatial resolution of 25 km and then apply the established

model to the environmental variables, which have a resolution of 1 km, to obtain spatially continuous soil moisture with a 1 km spatial resolution. The main steps of the downscaling process are as follows:

(1) First, the environmental variables during the 2003–2010 period are resampled to 25 km and 1 km, and the machine learning models, including the random forest and Cubist models, with a 16-day timescale, are established at a spatial resolution of 25 km:

$$f(SM_{25km}) = f(longitude_{25km}, latitude_{25km}, DEM_{25km}, slope_{25km}, aspect_{25km}, NDVI_{25km}, LST\_D_{25km}, LST\_N_{25km}, albedo_{25km}, ET_{25km}, LC_{25km}) \quad (1)$$

where SM_{25km} is the AMSR-E soil moisture data and $f(SM_{25km})$ is the downscaling model, namely, either the random forest model or the Cubist model.

(2) Subsequently, the AMSR-E soil moisture data are subtracted from the estimated soil moisture at a spatial resolution of 25 km to obtain a residual at 25 km, after which residual at 25 km model is resampled to 1 km;

(3) The established model is applied to the environmental variables at a spatial resolution of 1 km to obtain soil moisture data of 1 km before applying a residual correction:

$$f(SM_{1km}) = f(longitude_{1km}, latitude_{1km}, DEM_{1km}, slope_{1km}, aspect_{1km}, NDVI_{1km}, LST\_D_{1km}, LST\_N_{1km}, albedo_{1km}, ET_{1km}, LC_{1km}) \quad (2)$$

(4) Finally, the estimated value at 1 km is added to the residual at 1 km to obtain the 16-day soil moisture data with a spatial resolution of 1 km after a residual correction. Figure 3 provides the flowchart to illustrate the main steps of this procedure.

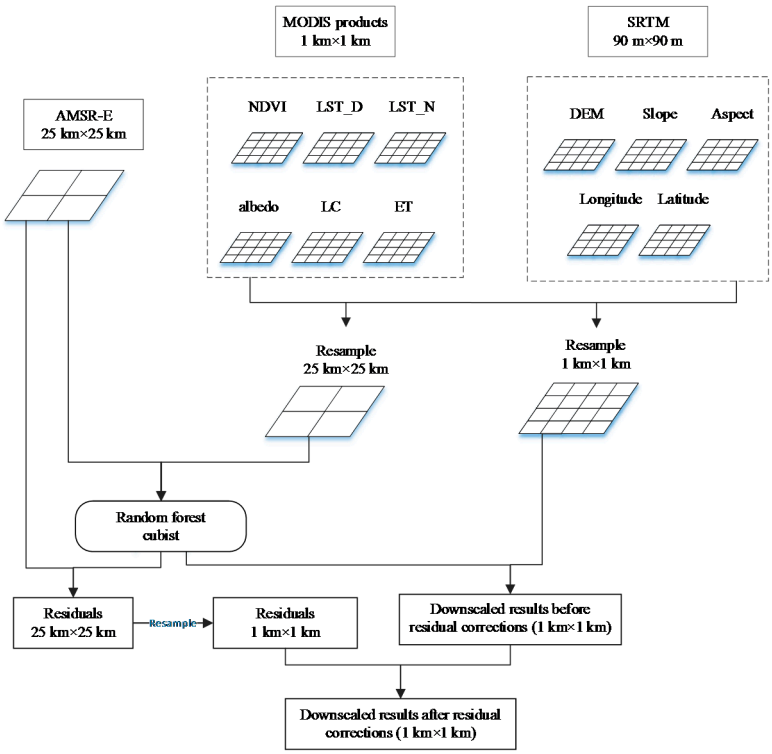


Figure 3. Flowchart of the downscaling process in this study.

3. Results and Discussion

3.1. Relationship between Soil Moisture and Environmental Variables

Most of the environmental variables, namely, the longitude, latitude, elevation, slope, NDVI, LST_D, LST_N, albedo, and ET, are numerical variables, whereas the LC and aspect are factor variables. Thus, Figure 4 shows the soil moisture values corresponding to different LC types, which demonstrates that the AMSR-E soil moisture values in different LC types in the MLRYRB exhibit the following trend: deciduous broadleaf forests > evergreen needleleaf forests > mixed forests > woody savannas > evergreen broadleaf forests > savannas > cropland/natural vegetation mosaics > croplands > grasslands > urban and built-up lands. Additionally, the AMSR-E soil moisture values of different forestland LC types are higher than those of the other LC types. Similarly, the distribution of soil moisture values that are based on the aspect from 2003 to 2010 is shown in Figure 5. As shown, in addition to the low AMSR-E value of the flat aspect (Flat = 0.1112), the AMSR-E values of the other eight geographical slopes do not substantially vary (the mean of each slope direction is given in parentheses): southeast (Southeast = 0.1296) > south (South = 0.1280) > northwest (Northwest = 0.1277) > west (West = 0.1269) > north (North = 0.1264) > southwest (Southwest = 0.1263) > northeast (Northeast = 0.1248) > east (East = 0.1234). Although the aspect has certain influence on the distribution of soil moisture, it is also closely related to the topography and surface vegetation cover type. Figure 6 shows the interannual variation of the AMSR-E soil moisture throughout the study area in 2003–2010. The shaded portion around the straight blue fitting line is the standard deviation of the linear fit. The soil moisture mean clearly exhibits a downward trend during the entire period of 2003–2010; in addition, the standard deviation in the middle of the study period is relatively small, whereas that in the beginning and end of the study period is relatively large.

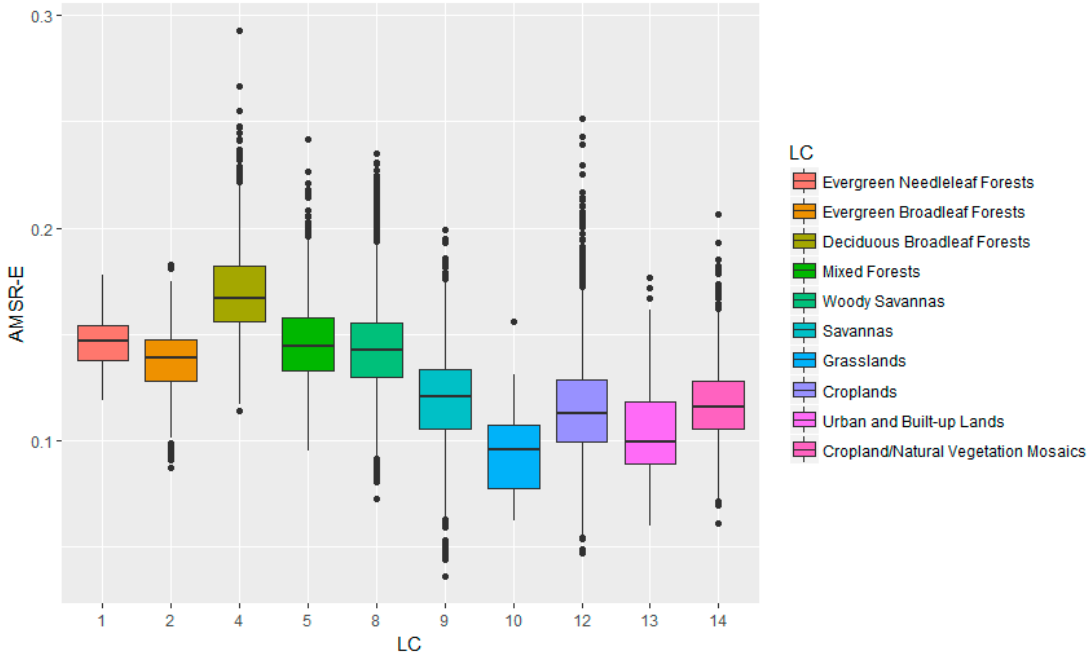


Figure 4. AMSR-E soil moisture values corresponding to different land cover (LC) types.

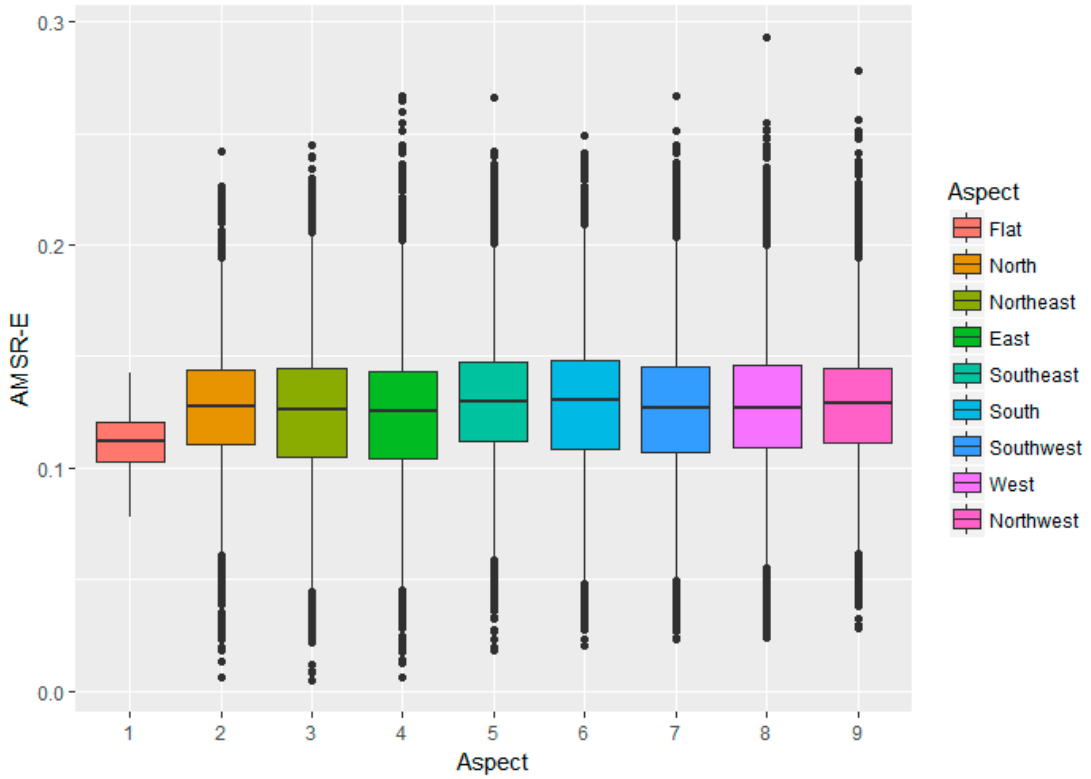


Figure 5. AMSR-E soil moisture values corresponding to different aspects.

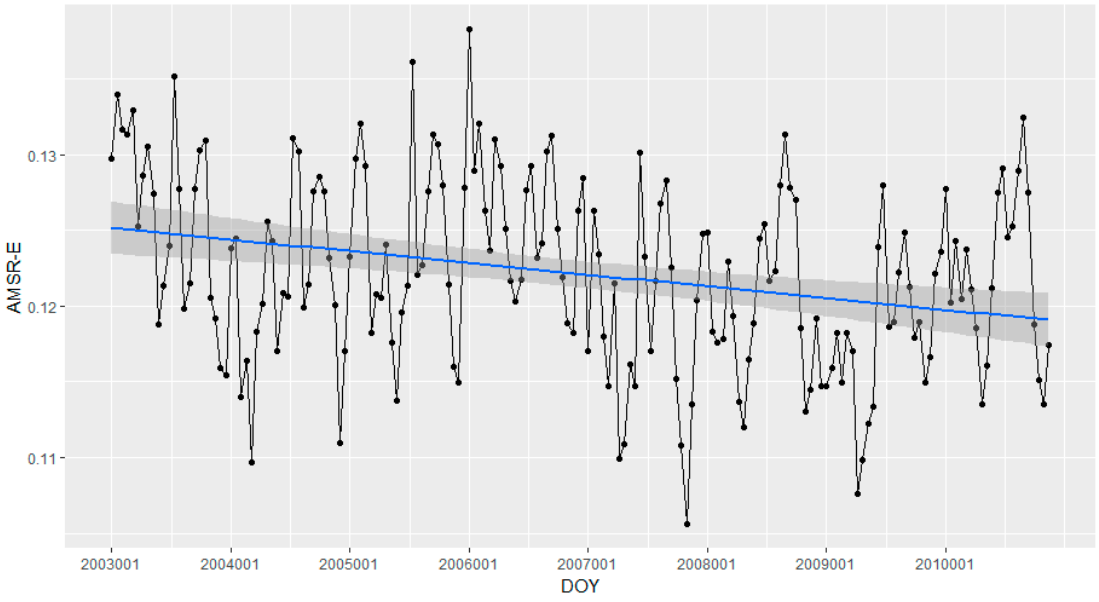


Figure 6. Interannual variation of the AMSR-E soil moisture in 2003–2010.

3.2. *Downscaling Results Based on the Random Forest Algorithm*

The environmental variables at 1 km spatial resolution are applied to estimate soil moisture at 1 km spatial resolution according to the soil moisture downscaling model that was constructed with the random forest at a spatial resolution of 25 km × 25 km (Figure 7b); then, the residuals at 25 km × 25 km are resampled to residuals at 1 km while using bilinear interpolation (Figure 7c). Next, the estimated values at 1 km are added to the residuals at 1 km to obtain 16-day soil moisture data with a spatial resolution of 1 km after applying a residual correction (Figure 7d). Figure 7 shows that the original AMSR-E soil moisture data are consistent with the trend of the estimated soil moisture data after

applying a residual correction, which indicates that the random forest-based downscaling method is well applicable to the study area. In addition, Figure 7 also shows that the soil moisture after the residual correction is more detailed and more representative of the spatial variation in the soil moisture values and is closer to the original AMSR-E soil moisture data distribution than the estimated results before the residual correction.

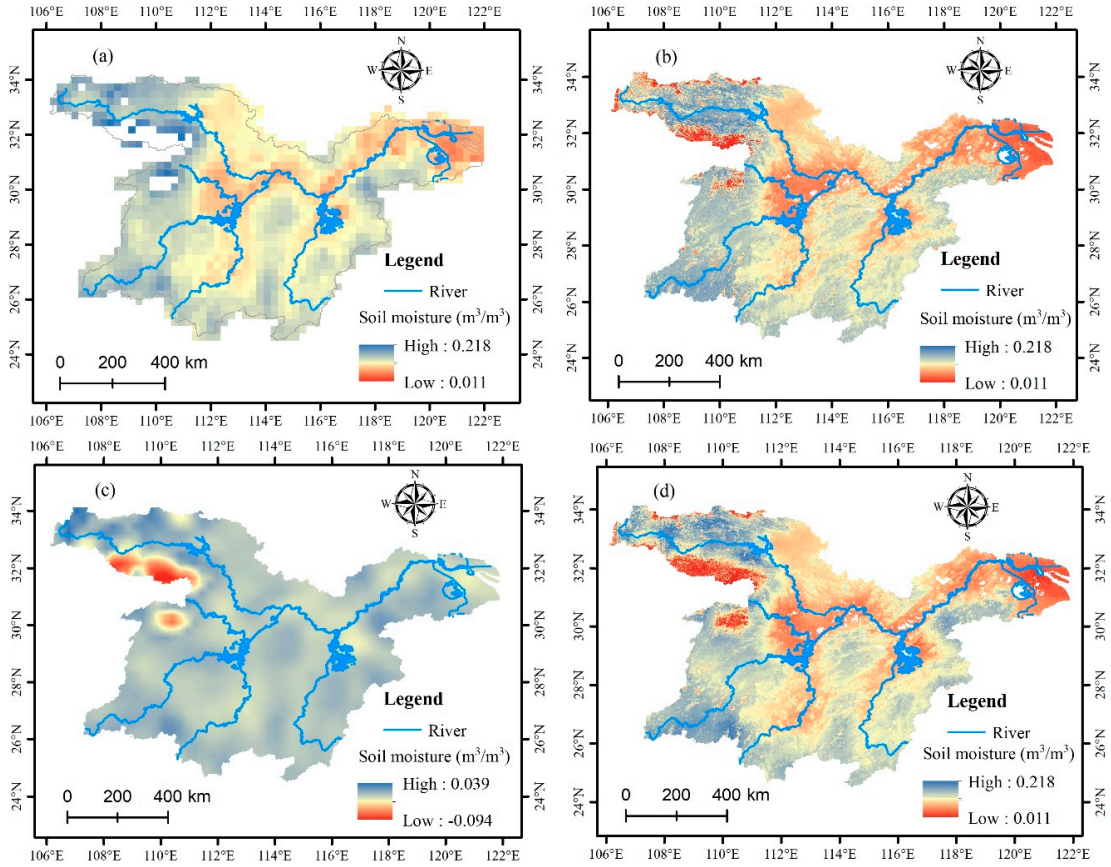
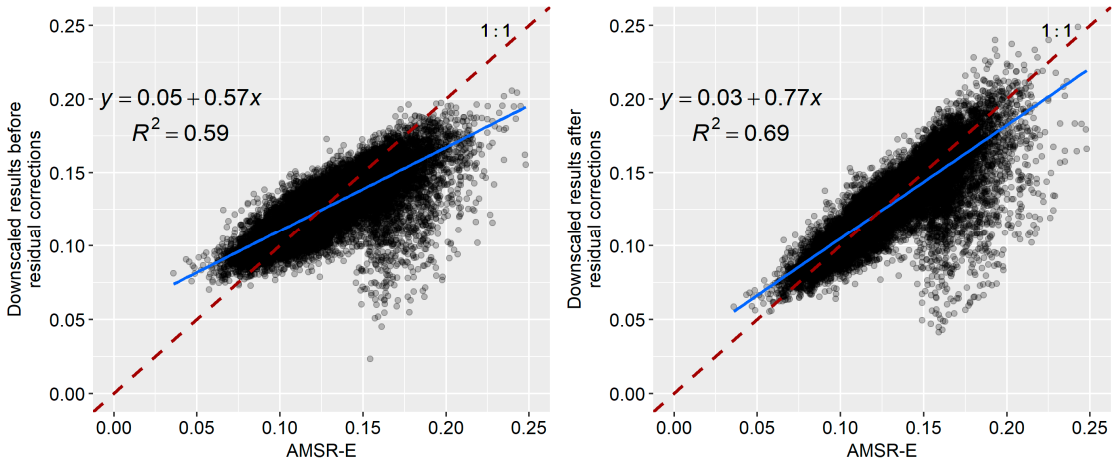
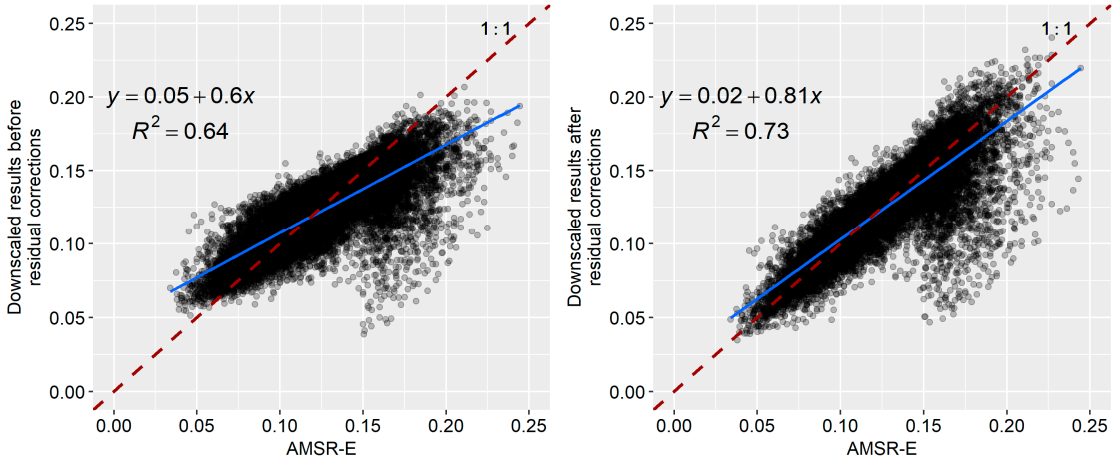


Figure 7. (**a**) AMSR-E soil moisture at a 25 km resolution; (**b**) downscaled results before the residual correction based on the random forest in the MLRYRB; (**c**) interpolated residuals using ordinary kriging; and, (**d**) downscaled results after the residual correction based on the random forest (taking DOY = 2003001 as an example).

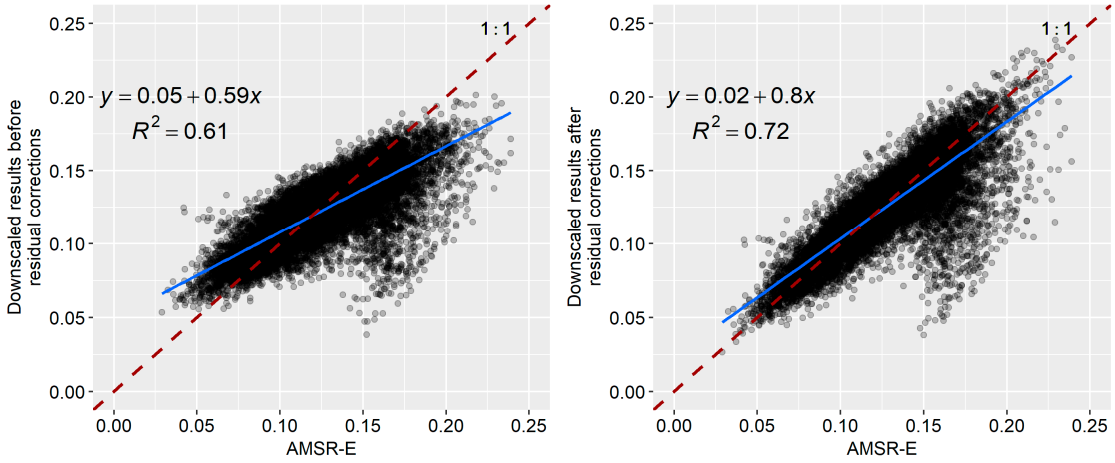
We compare the downscaled results with the original AMSR-E soil moisture to further validate the performance of the random forest-based downscaling model (Figure 8). The left panels in Figure 8 are scatterplots between the downscaling results before the residual correction and the original AMSR-E data, while the right panels are the scatterplots between the downscaling results after the residual correction and the original AMSR-E soil moisture. The transparency in Figure 8 is set according to the number of data points. A darker graph indicates that the density is higher and that the data points are very concentrated; the lighter the color is, the smaller the density, that is, the less the data are scattered. Evidently, the range of R^2 between the downscaling results before the residual correction and the original AMSR-E data is 0.55–0.64, while the range of R^2 after the residual correction is 0.68–0.74. The scatter plots on the left side of Figure 8 indicate that most of the data points are distributed above the 1:1 line, when the soil moisture is less than 0.12, and the data points are distributed below the 1:1 line when the soil moisture is greater than 0.12. In the scatter plots on the right side of Figure 8 corresponding to the downscaling results after the residual correction, the data points are evenly distributed on both sides of the 1:1 line. In general, the correlation between the AMSR-E soil moisture and the results of the random forest-based downscaling model are very good, which indicates that this downscaling model has good applicability in the MLRYRB.



(a) 2003

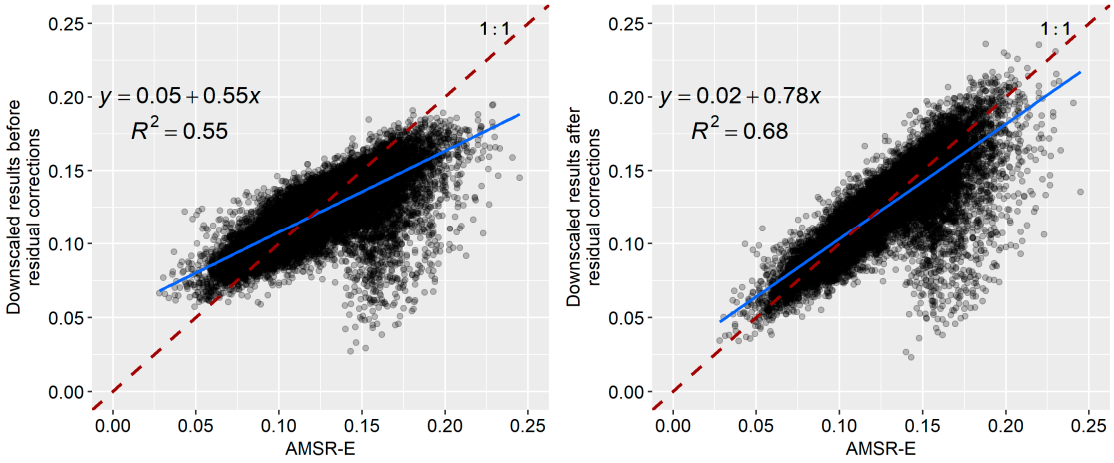


(b) 2004

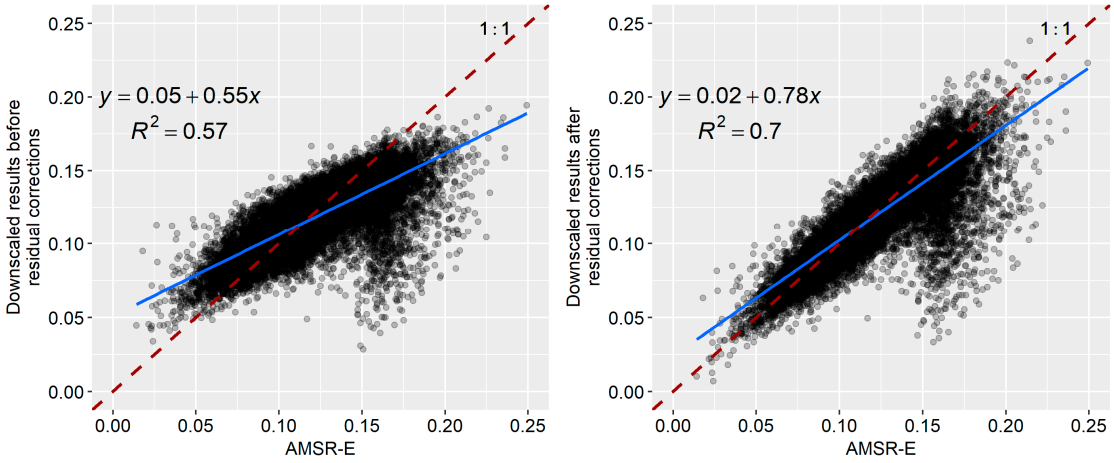


(c) 2005

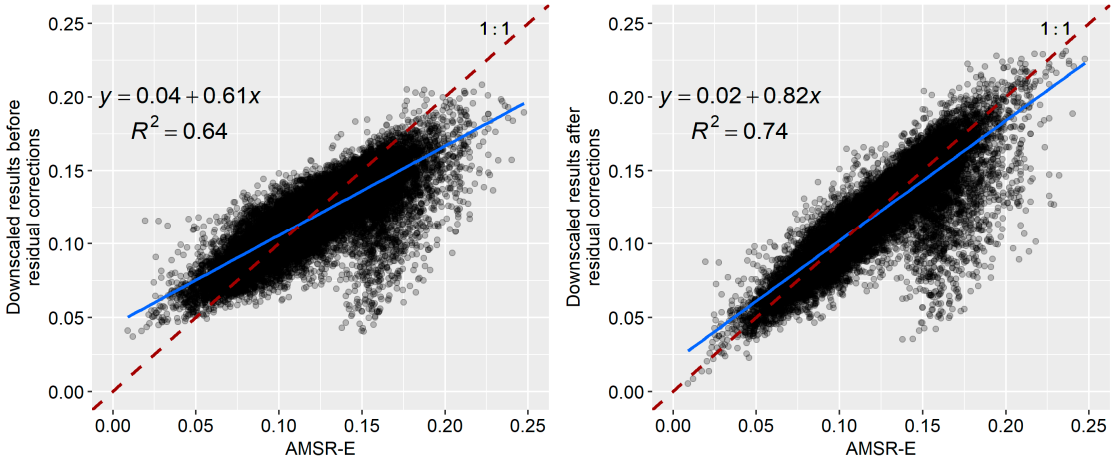
Figure 8. *Cont.*



(d) 2006

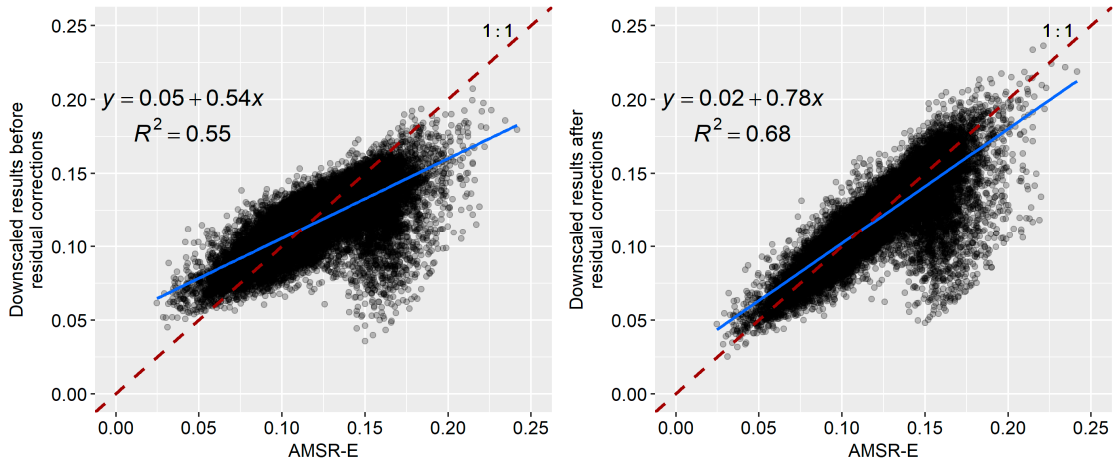


(e) 2007

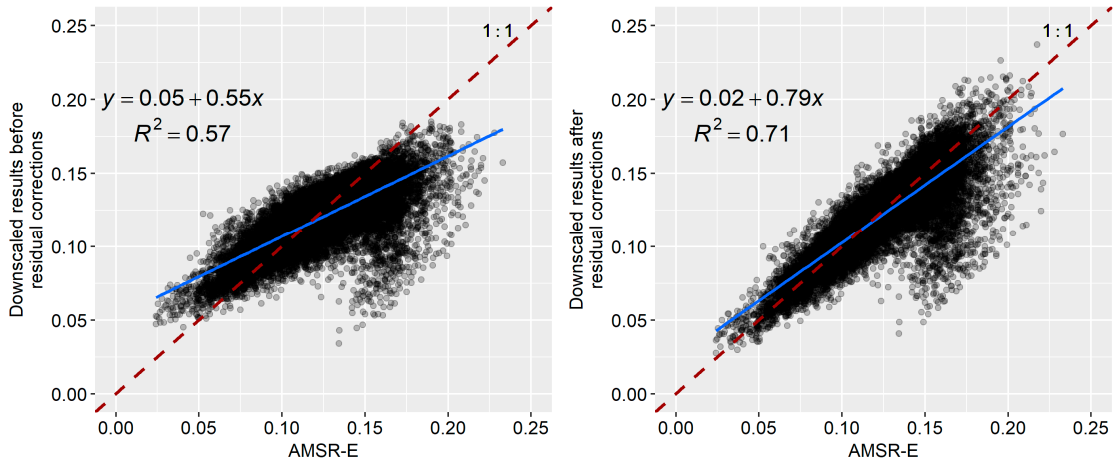


(f) 2008

Figure 8. *Cont.*



(g) 2009



(h) 2010

Figure 8. Scatter plots of the downscaled soil moisture results based on the random forest model and the original AMSR-E soil moisture from 2003 to 2010 (**left**: the downscaled results before the residual correction; **right**: the downscaled results after the residual correction).

3.3. Downscaling Results Based on the Cubist Model

This study selects the same environmental variables as those used for the random forest-based downscaling model for the Cubist downscaling algorithm: the longitude, latitude, elevation, slope, NDVI, LST_D, LST_N, albedo, ET, LC type, and aspect. However, the results show that, due to the linear relationship among longitude, latitude, and soil moisture, blocky features are too obvious in the downscaling results of the Cubist algorithm (see the red rectangles in Figure 9). In other words, the relationship for each rule between the soil moisture data and the latitude and longitude displays abrupt changes, which is obviously contrary to the continuity of soil moisture in space. Therefore, this study excludes two surface variables in the Cubist downscaling model, namely, longitude and latitude.

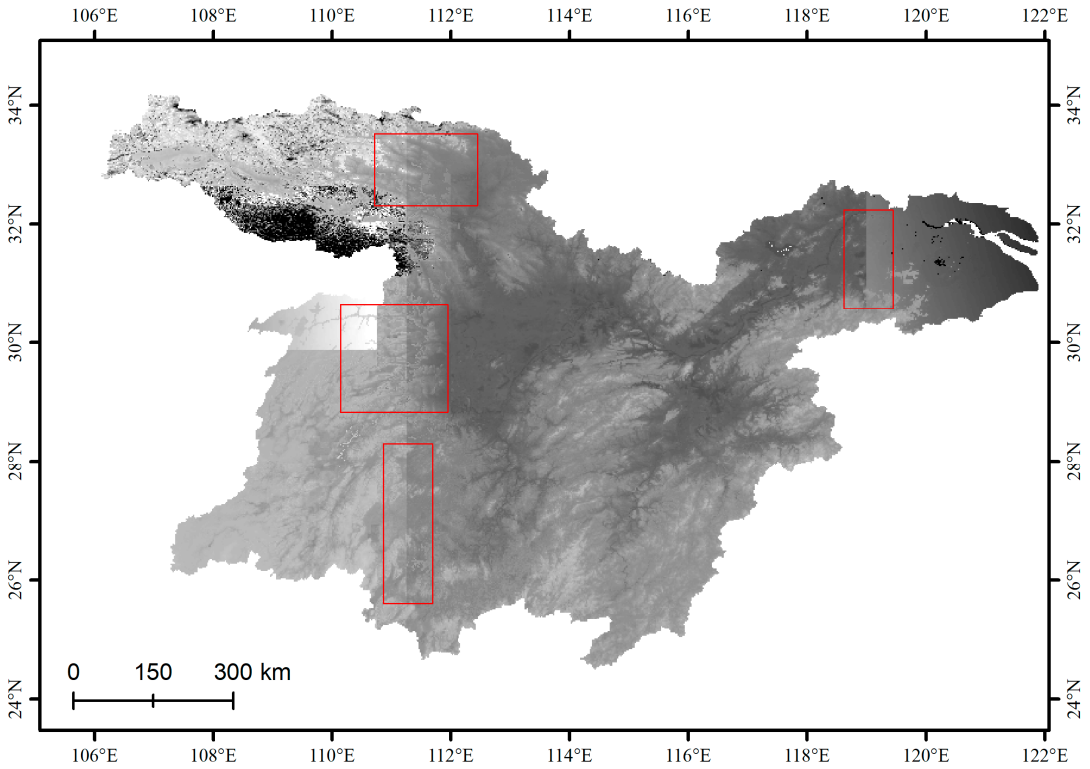


Figure 9. The downscaled results based on the Cubist algorithm (including longitude and latitude).

The downscaling process is carried out based on the Cubist model after removing the latitude and longitude surface variables. Table 2 lists the spatial regression relationship between the soil moisture and each surface variable (in the case of DOY = 2003001). 11 rules are generated after removing the longitude and latitude variables; however, not all of the variables participate in the downscaling model in each rule. This is one of the advantages of the Cubist algorithm, which automatically filters the optimal combination of variables that are required within the rules. To more intuitively ascertain whether each variable participates in the relationship of each rule, this study uses the R Lattice package to draw the regression coefficient graph of each rule (Figure 10). As the aspect and LC are factor variables, they only participate in the classification of each rule and do not participate in the regression calculations. Figure 10 shows that the albedo and ET surface variables participate in fewer rules; albedo participates in the model construction of the first and sixth rules, while the ET participates in the first, fifth, and tenth rules. These graphs can visualize the regression coefficient and intercept participating in each rule.

Figure 11 shows the distributions of the main split nodes of the environmental variables in the 11 rules of the Cubist downscaling algorithm (taking DOY = 2003001 as an example). The x-axis is the range of each variable and it is normalized from 0 to 1, while the y-axis is the split node of the variable used in each rule. For example, if a rule's variable is less than a certain value, then the rule's line will be drawn blue; if a rule's variable is greater than a certain value, the rule's line will be drawn purple. Figure 11 shows that during the 16-day period corresponding to DOY = 2003001, the variables composing the main split nodes are the NDVI, slope, DEM, and ET. The DEM is a main split node in all 11 rules, whereas the slope is a main split node in every rule, except the fourth, sixth, and eighth rules, the ET plays the role of a main split node in four of the rules, and the NDVI is a main split node in the first and fifth rules. The regression relationship and split node information of each rule in the Cubist downscaling algorithm are fully reflected through Table 2 and the graphs in Figures 10 and 11.

Table 2. Spatial regression relationships between the AMSR-E soil moisture and environmental variables in the MLRYRB (in the case of DOY = 2003001) based on the Cubist algorithm. For the aspect, numbers 1 through 9 indicate Flat, North, Northeast, East, Southeast, South, Southwest, West, and Northwest.

Rule 1: if {DEM > 1062, Slope > 3.82698, Slope ≤ 8.79908, Aspect in {4, 5, 6, 7, 9}, NDVI > 0.42153, ET ≤ 14.8708} then
outcome = 0.0623238 + 0.0426 × Slope − 0.0217 × ET − 1.412 × albedo + 0.315 × NDVI
Rule 2: if {DEM > 1062, Slope > 2.77987, Slope ≤ 3.82698, LC in {4, 5}} then
outcome = −0.5929961 + 0.2089 × Slope
Rule 3: if {DEM > 646, Slope > 8.79908} then
outcome = $-0.6552099 - 1.79 \times 10^{-5} \times$ DEM + 0.0025 × LST_D
Rule 4: if {DEM ≤ 136} then
outcome = −0.3143938 + 0.0002028 × DEM + 0.0015 × LST_N + 0.0015 × Slope + 0.019 × NDVI
Rule 5: if {DEM > 1062, Slope > 3.82698, Aspect in {4, 5, 6, 7, 9}, NDVI ≤ 0.42153, ET ≤ 14.8708} then
outcome = 0.7520723 − 0.0417 × ET − 0.444 × NDVI + 0.0061 × Slope + 0.0002 × LST_D
Rule 6: if {DEM > 136, DEM ≤ 646} then
outcome = $-0.0459537 + 0.00228 \times$ LST_N $+ 3.57 \times 10^{-5} \times$ DEM − 0.0016 × LST_D + 0.0012 × Slope − 0.035 × albedo + 0.003 × NDVI
Rule 7: if {DEM > 646, Slope > 2.77987, Slope ≤ 3.82698, LC in {8, 9, 12}} then
outcome = $0.0276217 + 0.0255 \times$ Slope $+ 5.8 \times 10^{-6} \times$ DEM + 0.00016 × LST_N + 0.003 × NDVI
Rule 8: if {DEM > 646, DEM ≤ 1062} then
outcome = 0.1565938
Rule 9: if {DEM > 1062, Slope > 3.82698, Slope ≤ 8.79908, ET > 14.8708} then
outcome = $-0.0548823 + 0.0008 \times$ LST_N $+ 5.5 \times 10^{-6} \times$ DEM + 0.0003 × Slope + 0.003 × NDVI
Rule 10: if {DEM > 1062, Slope > 3.82698, Aspect in {3, 8}, ET ≤ 14.8708} then
outcome = 0.1484353 + 0.0081 × Slope − 0.0044 × ET + 0.00014 × LST_N
Rule 11: if {DEM > 646, Slope ≤ 2.77987} then
outcome = 0.7492527 − 0.0021 × LST_D $+ 8 \times 10^{-7} \times$ DEM

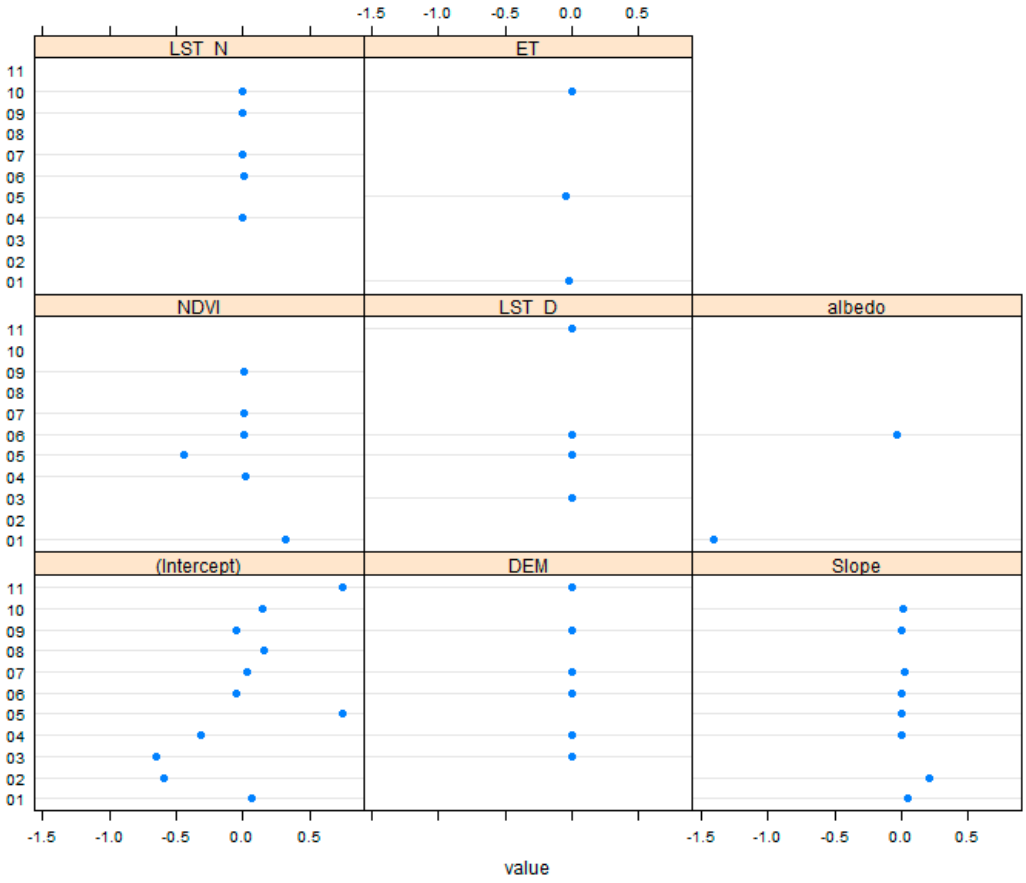


Figure 10. Regression coefficient of each relationship in the Cubist downscaling algorithm (taking DOY = 2003001 as an example).

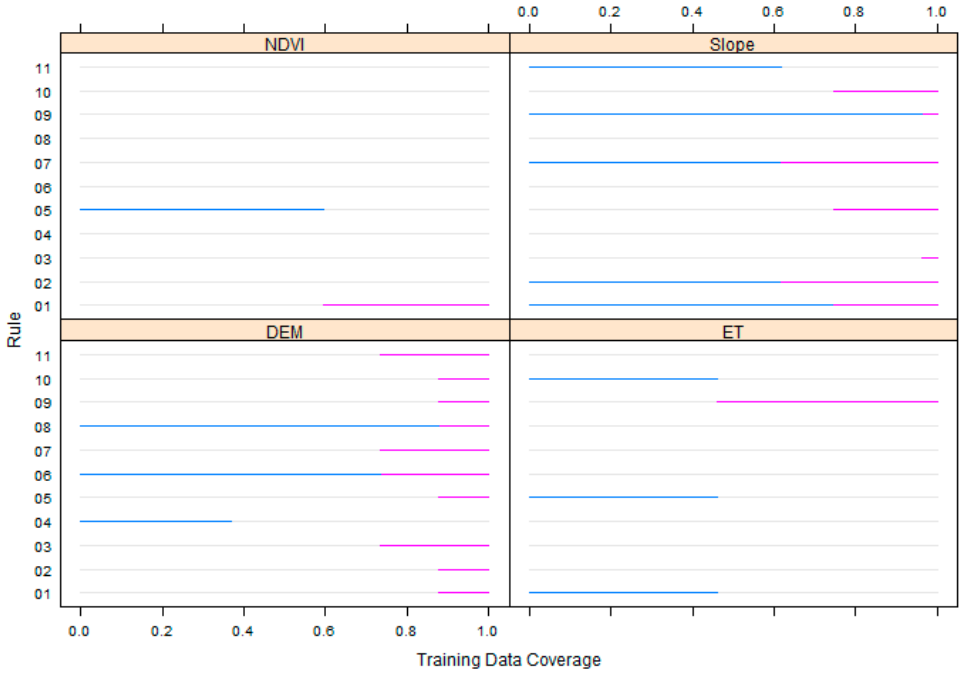


Figure 11. Information of the main split node in each relationship in the Cubist downscaling algorithm (taking DOY = 2003001 as an example).

The environmental variables with a 1 km spatial resolution are applied to estimate soil moisture at a 1 km spatial resolution according to the Cubist-based soil moisture downscaling model at a spatial resolution of 25 km × 25 km (Figure 12b); then, the residuals at 25 km × 25 km are resampled to residuals at 1 km using bilinear interpolation (Figure 12c). Next, the estimated values at 1 km are added to the residuals at 1 km to obtain 16-day soil moisture data with a spatial resolution of 1 km after applying a residual correction. Figure 12 demonstrates that the AMSR-E passive microwave soil moisture data are consistent with the distribution of the residual-corrected estimated values, which indicates that the Cubist-based downscaling method also shows good performance in the MLRYRB.

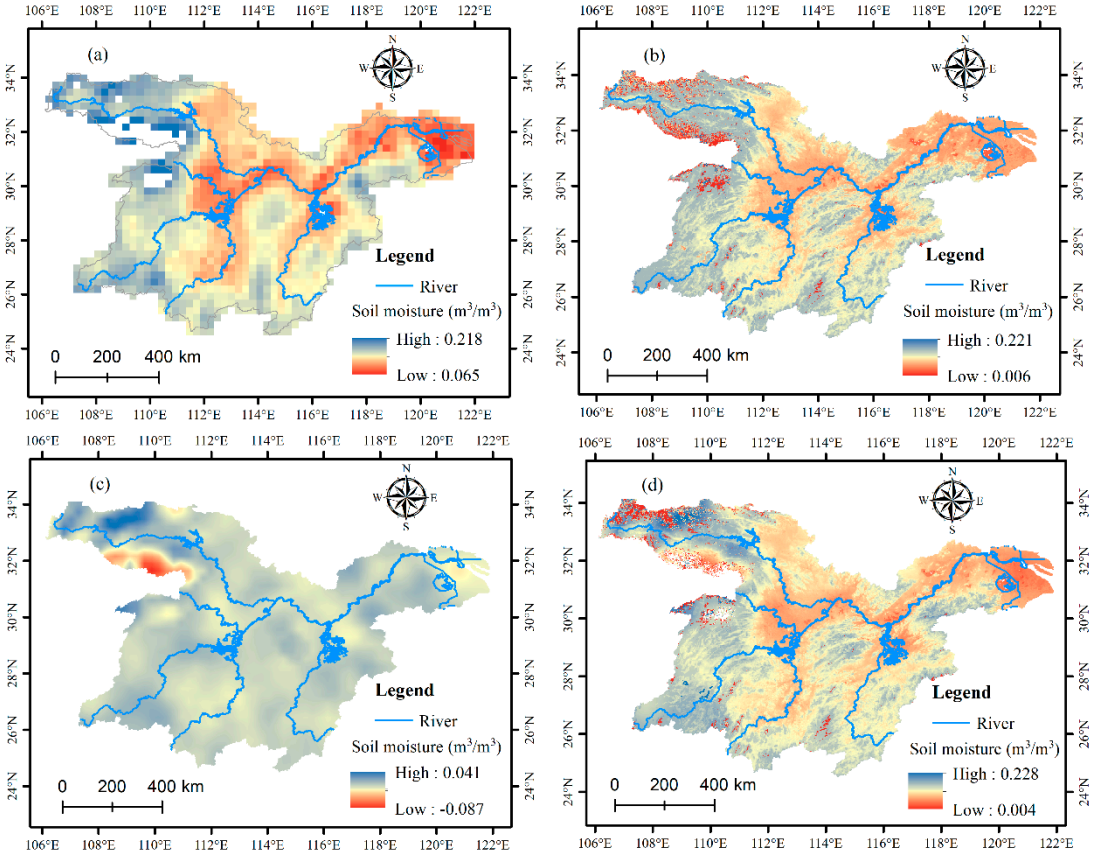
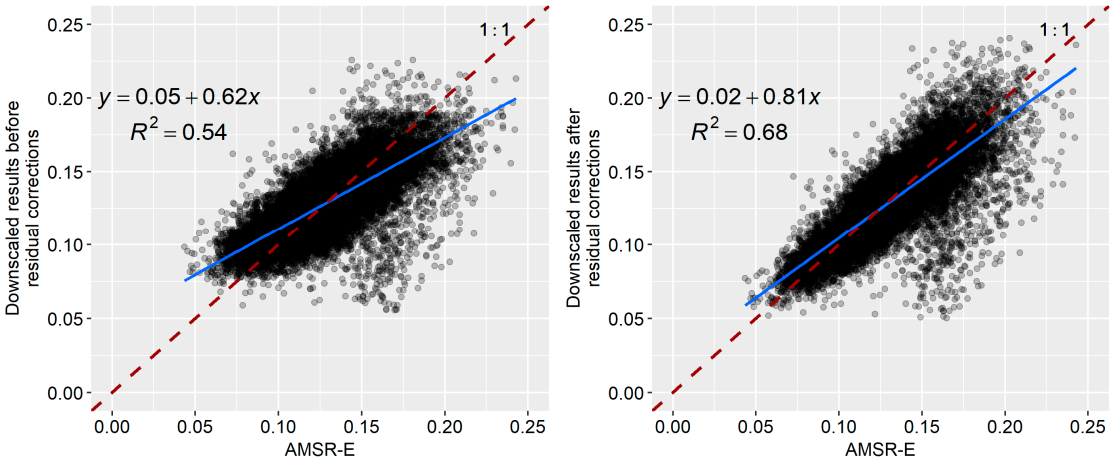


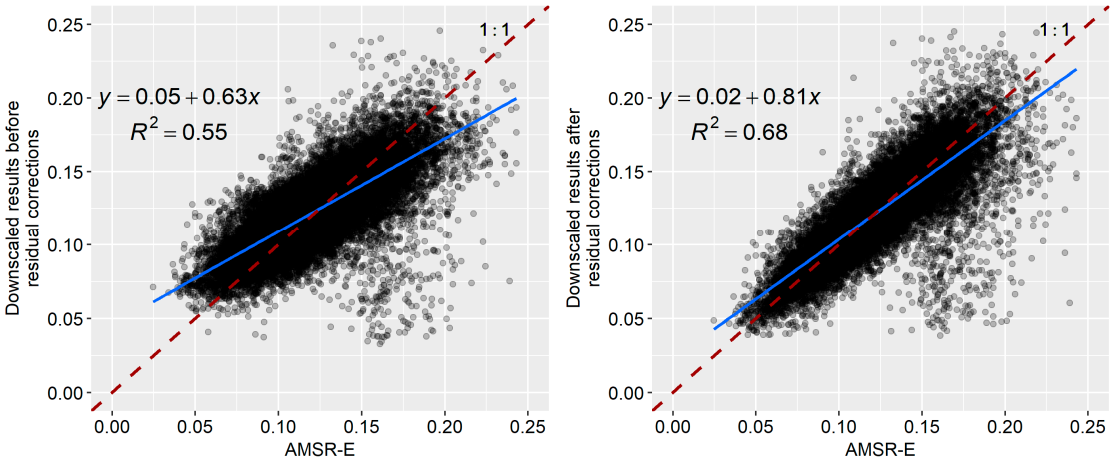
Figure 12. (**a**) AMSR-E soil moisture at a 25 km resolution; (**b**) downscaled results before the residual correction based on the Cubist model in the MLRYRB; (**c**) interpolated residuals using ordinary kriging; and, (**d**) downscaled results after the residual correction based on the Cubist model (taking DOY = 2003001 as an example).

We compare the original AMSR-E soil moisture data with the downscaling results to further validate the effectiveness of the Cubist-based downscaling algorithm (including the downscaling results before and after the residual correction). Again, we set the transparency in Figure 13 according to the data density: the darker areas correspond to a higher data density, while lighter areas indicate fewer data points. The graphs on the left in Figure 13 are the downscaling results of the Cubist algorithm before the residual correction, and the graphs on the right are the soil moisture downscaling results after the residual correction. Figure 13 shows that the R^2 values of the downscaled data before the residual correction vary from 0.50 to 0.56, while the R^2 values of the downscaled data based on the random forest model before the residual correction range from 0.55–0.64. The correlations of the random forest-based results are higher than those of the Cubist-downscaling results. In addition, the distributions of data points in the scatter plots of the Cubist model are similar to those of the random forest model. Most of the downscaled data before the residual correction are higher than the original AMSR-E soil moisture values when the soil moisture is less than 0.12; when the soil moisture

is greater than 0.12, most of the original AMSR-E soil moisture values are larger than the downscaled soil moisture values. Moreover, the results are not evenly distributed on both sides of the 1:1 line; that is, when compared with the random forest algorithm, the data points that are based on the Cubist algorithm are more dispersed. The R^2 values between the residual-corrected downscaling results of the Cubist algorithm and the AMSR-E soil moisture values range from 0.68 to 0.71, which is lower than the range of R^2 values corresponding to the random forest downscaling algorithm (0.68 to 0.74). The distribution of data points is more concentrated after the residual correction than the downscaling results before the residual correction, which indicates that a residual correction can significantly improve the downscaling results of soil moisture data.

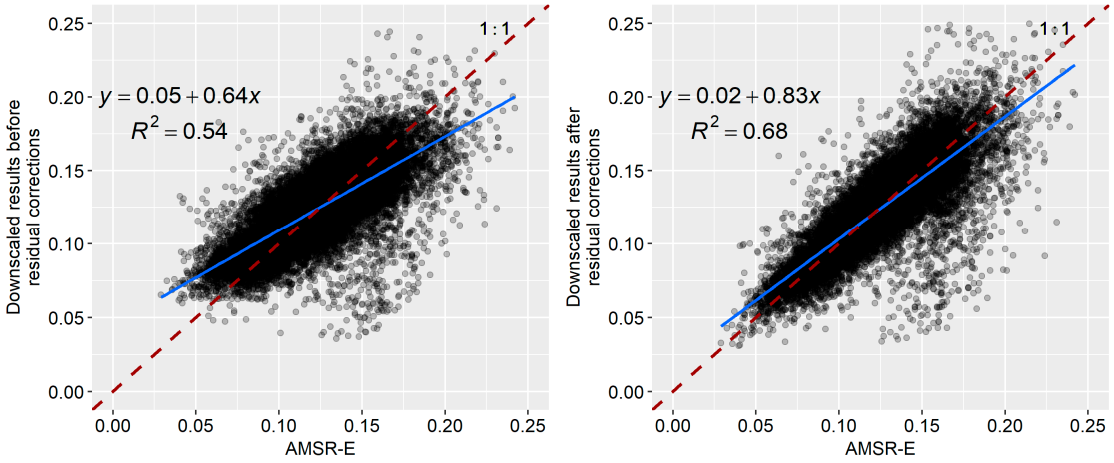


(a) 2003

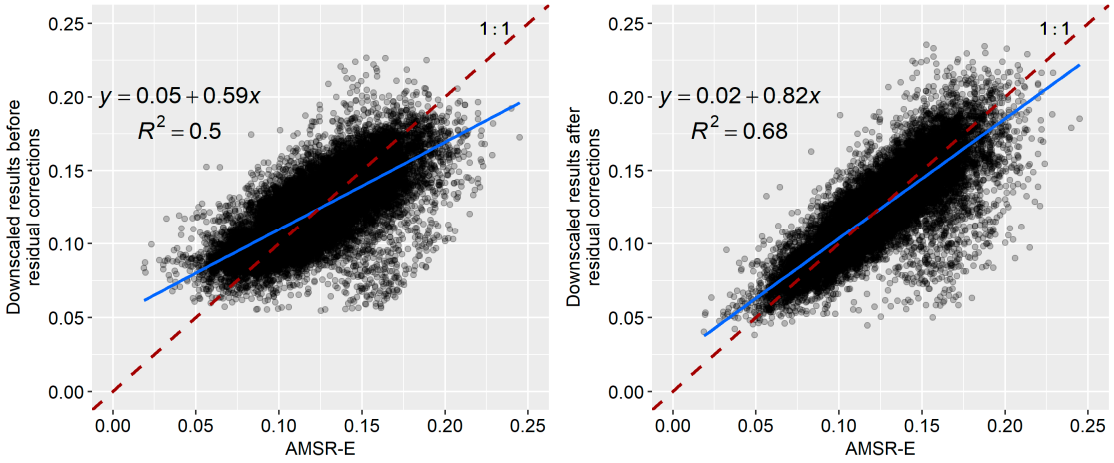


(b) 2004

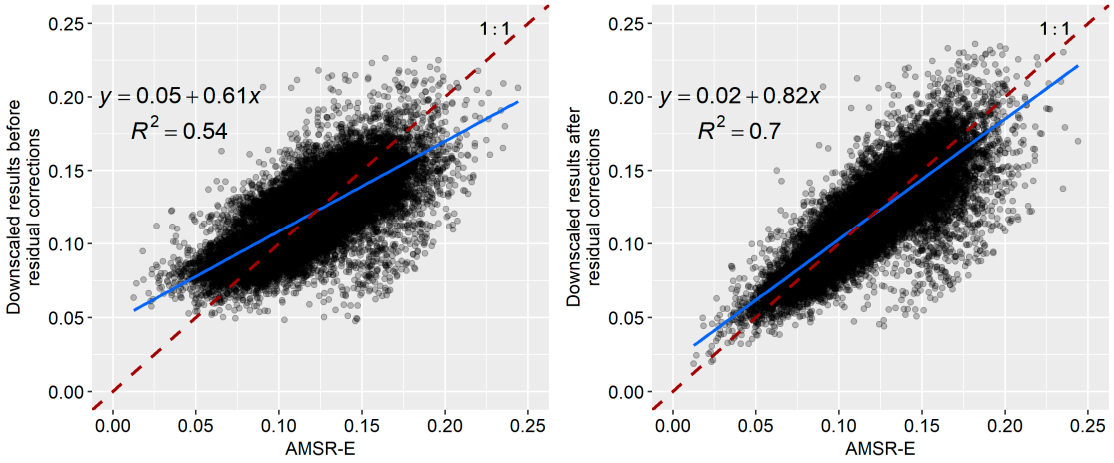
Figure 13. *Cont.*



(c) 2005

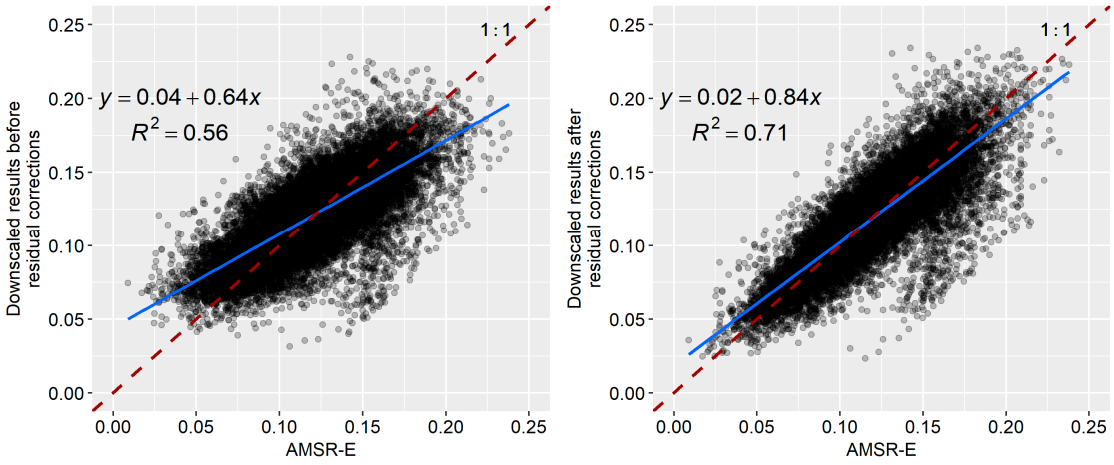


(d) 2006

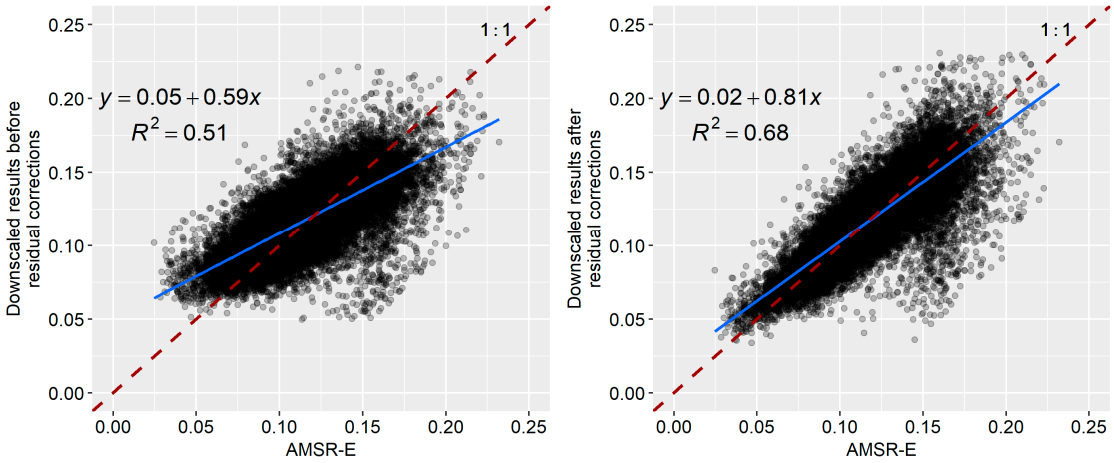


(e) 2007

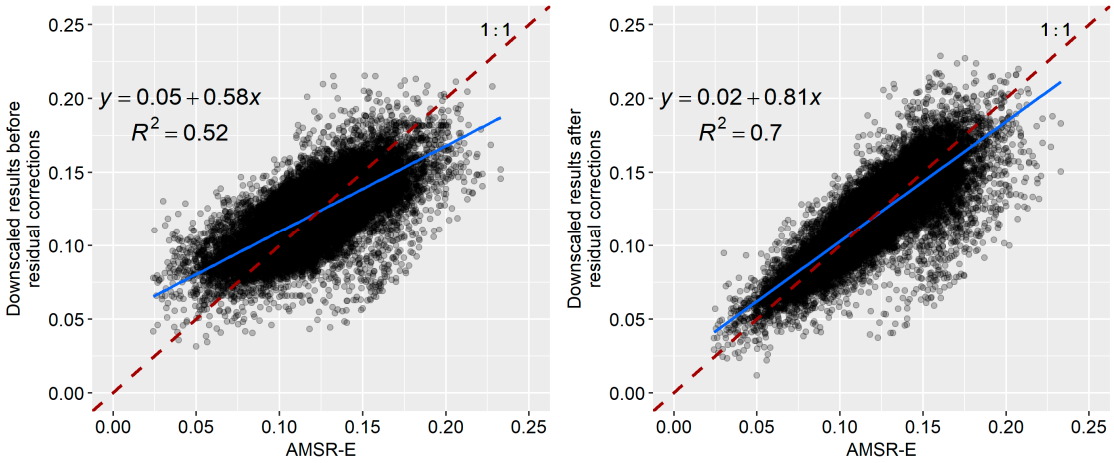
Figure 13. *Cont.*



(f) 2008



(g) 2009



(h) 2010

Figure 13. Scatter plots of the downscaled soil moisture results based on the Cubist model and the original AMSR-E soil moisture from 2003 to 2010 (**left**: the downscaled results before the residual correction; **right**: the downscaled results after the residual correction).

3.4. Comparative Analysis of the Downscaling Results Based on the Random Forest and Cubist Algorithms

Table 3 shows the accuracy verification between the downscaled results and the observed in situ soil moisture. The results showed that the mean R^2, root-mean-square error (RMSE), and mean absolute error (MAE) values between the original AMSR-E data and in situ soil moisture were 0.6018, 0.0131 m^3/m^3, and 0.0113 m^3/m^3, respectively. Furthermore, the results showed that the mean R^2, RMSE, and MAE values between the downscaled results based on the random forest and the in situ soil moisture was 0.7819, 0.0090 m^3/m^3, and 0.0076 m^3/m^3, which was better than that of the downscaled results that were based on Cubist (R^2 = 0.6722, RMSE = 0.0128 m^3/ m^3, and MAE = 0.0111 m^3/m^3). The two developed downscaling methods in this study can improve not only the spatial resolution of the remote sensing AMSR-E data, but also the accuracy of the data.

Table 3. Comparison between the downscaling results based on the two methods and in situ soil moisture.

Station Name	AMSR-E			Random Forest			Cubist		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
Nanyang	0.6836	0.0066	0.0053	0.8545	0.0064	0.0045	0.7331	0.0077	0.0059
Fangxian	0.4002	0.0161	0.0120	0.5024	0.0126	0.0104	0.4673	0.0132	0.0104
Lixian	0.8234	0.0125	0.0107	0.8757	0.0074	0.0061	0.7390	0.0088	0.0062
Guzhang	0.7278	0.0085	0.0070	0.8632	0.0059	0.0047	0.8368	0.0052	0.0039
Shaodong	0.5923	0.0151	0.0134	0.8569	0.0070	0.0056	0.6687	0.0176	0.0160
Yichun	0.7286	0.0120	0.0102	0.8375	0.0061	0.0047	0.8145	0.0055	0.0043
Guidong	0.5469	0.0084	0.0075	0.6758	0.0109	0.0096	0.6202	0.0203	0.0195
Yingshan	0.4692	0.0242	0.0228	0.7510	0.0179	0.0172	0.5008	0.0265	0.0256
Xuancheng	0.4979	0.0116	0.0095	0.7959	0.0077	0.0059	0.6706	0.0095	0.0076
Guixi	0.5484	0.0162	0.0144	0.8063	0.0084	0.0071	0.6708	0.0134	0.0120
Mean values	0.6018	0.0131	0.0113	0.7819	0.0090	0.0076	0.6722	0.0128	0.0111

This study compares the random forest and Cubist algorithms to determine the most suitable method for downscaling soil moisture in the MLRYRB. Figure 14 illustrates the downscaled results of the two downscaling methods. Similarly, we set the transparency according to the density of the points: the darker the plot, the more data points there are, while the lighter areas indicate fewer data points. Evidently, the correlation between the downscaling results based on the random forest method and the AMSR-E soil moisture (R^2 = 0.71) is better than that between the results that are based on the Cubist algorithm and the AMSR-E soil moisture (R^2 = 0.70). The data points in Figure 14a are more densely concentrated around the 1:1 line than are those in Figure 14b, while the data points of the Cubist algorithm (Figure 14b) are more dispersed; however, in both panels, there are some outliers below the 1:1 line, although the number of points is small. This is related to the quality of the MODIS images, because clouds cover some images, though other reasons can also lead to noise.

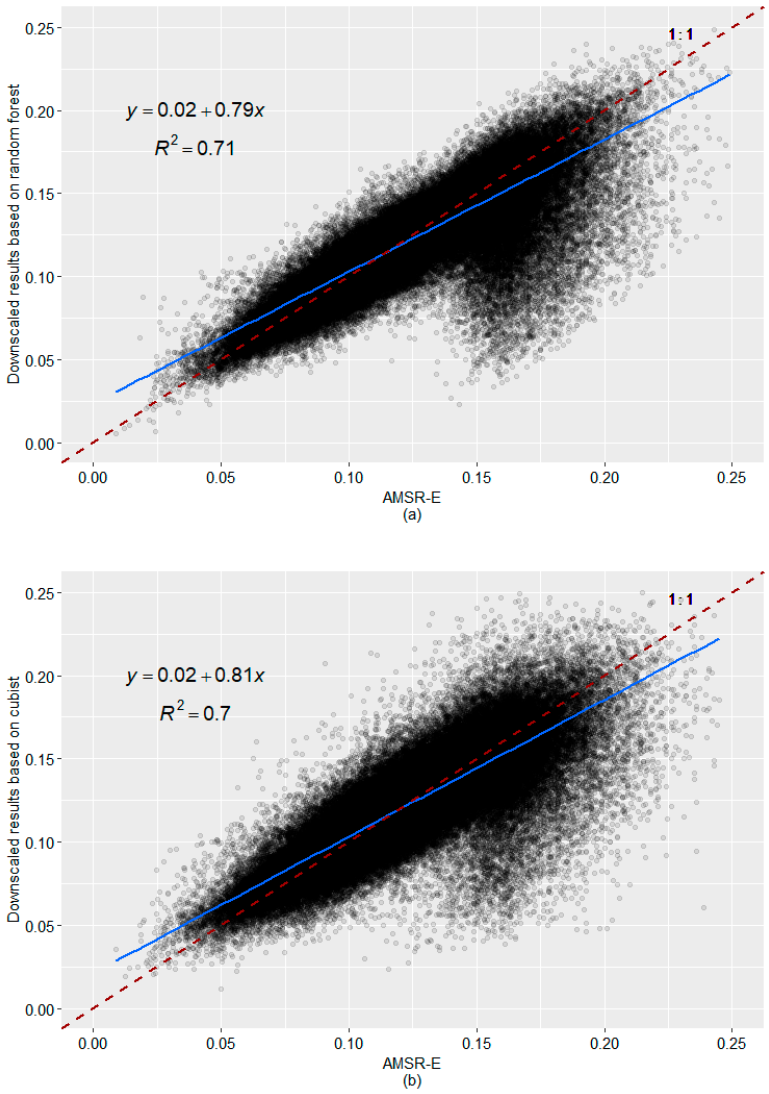


Figure 14. Comparative analysis of the two downscaling methods. (**a**) Scatter plot between the random forest-based downscaled results and the AMSR-E soil moisture; and, (**b**) scatter plot between the Cubist-based downscaled results and the AMSR-E soil moisture.

To further analyze the performance of the two methods, their accuracies were tested while using three evaluation indicators, namely, R^2, the root-mean-square error (RMSE), and the mean absolute error (MAE); Table 4 shows the results. The results show that the R^2, RMSE, and MAE values between the soil moisture data downscaled by the random forest method and the original AMSR-E soil moisture are 0.7045, 0.0155, and 0.0096, respectively, while those between the downscaling results that are based on the Cubist algorithm and the AMSR-E soil moisture are 0.6884, 0.0162, and 0.0010, respectively. The R^2 values that are based on the random forest downscaling method are 0.011, which is higher than those based on the Cubist algorithm, while the RMSE is reduced by 0.0006, and the MAE is reduced by 0.0014. The accuracies of the two downscaling methods are verified through a box plot to more intuitively compare the performance of the two downscaling methods (Figure 15). Figure 15 illustrates that the R^2, RMSE, and MAE of the soil moisture data obtained by the random forest-based downscaling method are better than those that were obtained by the Cubist algorithm, which indicated that the former model is a more suitable AMSR-E soil moisture downscaling method than the latter in the MLRYRB. This finding also provides a theoretical basis for obtaining high spatial resolution soil moisture data.

Table 4. Accuracy validation between the downscaling results based on the two methods and the AMSR-E soil moisture.

Year	Random Forest			Cubist		
	R^2	RMSE	MAE	R^2	RMSE	MAE
2003	0.6866	0.0146	0.0088	0.6759	0.0150	0.0100
2004	0.7311	0.0153	0.0094	0.6783	0.0172	0.0113
2005	0.7155	0.0151	0.0093	0.6841	0.0164	0.0109
2006	0.6780	0.0162	0.0098	0.6808	0.0161	0.0110
2007	0.6965	0.0162	0.0100	0.6950	0.0163	0.0114
2008	0.7381	0.0156	0.0097	0.7115	0.0166	0.0114
2009	0.6840	0.0161	0.0099	0.6783	0.0163	0.0112
2010	0.7066	0.0150	0.0095	0.7036	0.0152	0.0106
Mean values	0.7045	0.0155	0.0096	0.6884	0.0162	0.0110

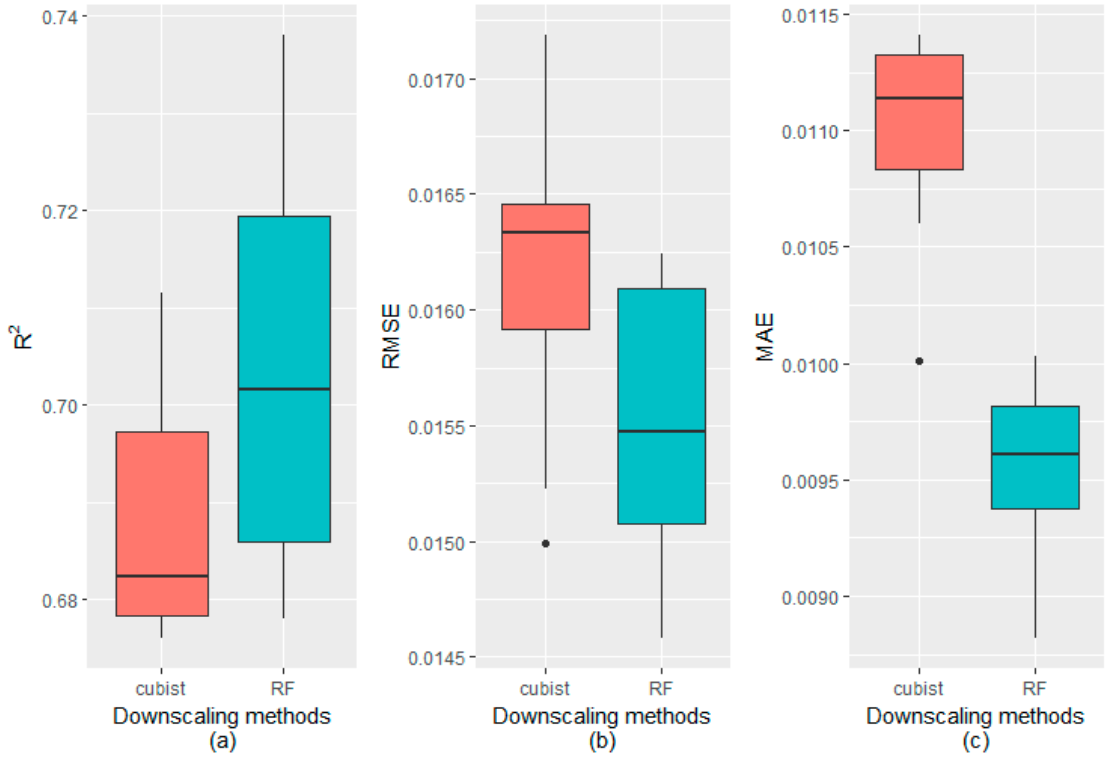


Figure 15. Boxplots based on the random forest model and Cubist algorithm to verify the accuracy. (**a**) R^2; (**b**) root-mean-square error (RMSE); and, (**c**) mean absolute error (MAE).

4. Conclusions

In this study, we choose the random forest model and Cubist algorithm to downscale AMSR-E soil moisture from 25 km to 1 km in the MLRYRB; for this task, the longitude, latitude, elevation, slope, NDVI, LST_D, LST_N, albedo, ET, LC, and aspect are selected as the environmental variables. Moreover, the random forest model and Cubist algorithm are compared and analyzed. The main conclusions can be summarized, as follows:

(1) Based on the random forest model, we downscale the AMSR-E soil moisture from 25 km to 1 km in the MLRYRB. The results show that the random forest downscaling method is strongly applicable in the MLRYRB, and the downscaled results after a residual correction have more details and they are more representative of the spatial distribution of soil moisture.

(2) The R^2 between the downscaling results that are based on the Cubist downscaling algorithm after the residual correction and the original AMSR-E soil moisture values range from 0.68 to 0.71,

which is lower than the range of the R^2 between those that are based on the random forest downscaling algorithm and original AMSR-E soil moisture data (0.68 to 0.74).

(3) A comparison between the random forest model and the Cubist algorithm reveals that the R^2 of the random forest-based downscaling method is higher than that of the Cubist algorithm-based downscaling method by 0.0161; moreover, the RMSE is reduced by 0.0006 and the MAE is reduced by 0.0014. Furthermore, testing the accuracies of the two downscaling methods reveals that the random forest model is a more suitable method than the Cubist algorithm for downscaling AMSR-E soil moisture data from 25 km to 1 km in the MLRYRB, and this finding provides a theoretical basis for obtaining high spatial resolution soil moisture data.

Author Contributions: For Conceptualization, S.C. and L.Z.; methodology, S.C. and L.Z.; software, S.C.; validation, D.S., M.G. and X.L.; formal analysis, S.C.; investigation, S.C.; resources, S.C. and D.S.; data curation, S.C.; writing—original draft preparation, S.C.; writing—review and editing, L.Z. and D.S.; visualization, S.C.; supervision, D.S.; project administration, D.S.; funding acquisition, L.Z.

Funding: This study was supported by the National Key Research and Development Program of China (NO.2017YFA0603704).

Acknowledgments: Thanks to the Earth Observing System Data Gateway of the United States' National Aeronautics and Space Administration (NASA) (https://earthdata.nasa.gov/) for offering the MODIS data. The meteorological data used in this study were collected from the National Climatic Centre of China Meteorological Administration (available on http://cdc.nmic.cn/home.do), which was highly appreciated. We are very grateful to the editor and anonymous reviewers.

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

References

1. 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]
2. Merlin, O.; Walker, J.; Chehbouni, A.; Kerr, Y. Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency. *Remote Sens. Environ.* **2008**, *112*, 3935–3946. [CrossRef]
3. Alemohammad, S.H.; Kolassa, J.; Prigent, C.; Aires, F.; Gentine, P. Global downscaling of remotely sensed soil moisture using neural networks. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 5341–5356. [CrossRef]
4. Berg, A.; Lintner, B.R.; Findell, K.L.; Malyshev, S.; Loikith, P.C.; Gentine, P. Impact of Soil Moisture–Atmosphere Interactions on Surface Temperature Distribution. *J. Clim.* **2014**, *27*, 7976–7993. [CrossRef]
5. Pablos, M.; Martínez-Fernández, J.; Sánchez, N.; González-Zamora, Á. Temporal and Spatial Comparison of Agricultural Drought Indices from Moderate Resolution Satellite Soil Moisture Data over Northwest Spain. *Remote Sens.* **2017**, *9*, 1168. [CrossRef]
6. Kang, Y.H.; Khan, S.; Ma, X.Y. Climate change impacts on crop yield, crop water productivity and food security—A review. *Prog. Nat. Sci.* **2009**, *19*, 1665–1674. [CrossRef]
7. Peng, J.; Loew, A.; Merlin, O.; Verhoest, N.E.C. A review of spatial downscaling of satellite remotely sensed soil moisture. *Rev. Geophys.* **2017**, *55*, 341–366. [CrossRef]
8. Brocca, L.; Hasenauer, S.; Lacava, T.; Melone, F.; Moramarco, T.; Wagner, W.; Dorigo, W.; Matgen, P.; Martínez-Fernández, J.; Llorens, P.; et al. Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. *Remote Sens. Environ.* **2011**, *115*, 3390–3408. [CrossRef]
9. Van doninck, J.; Peters, J.; De Baets, B.; De Clercq, E.M.; Ducheyne, E.; Verhoest, N.E.C. The potential of multitemporal Aqua and Terra MODIS apparent thermal inertia as a soil moisture indicator. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 934–941. [CrossRef]
10. Verstraeten, W.W.; Veroustraete, F.; Feyen, J. Assessment of Evapotranspiration and Soil Moisture Content Across Different Scales of Observation. *Sensors* **2008**, *8*, 70–117. [CrossRef]
11. Kalma, J.D.; McVicar, T.R.; McCabe, M.F. Estimating Land Surface Evaporation: A Review of Methods Using Remotely Sensed Surface Temperature Data. *Surv. Geophys.* **2008**, *29*, 421–469. [CrossRef]

12. Mitra, D.S.; Majumdar, T.J. Thermal inertia mapping over the Brahmaputra basin, India using NOAA-AVHRR data and its possible geological applications. *Int. J. Remote Sens.* **2004**, *25*, 3245–3260. [CrossRef]
13. Sandholt, I.; Rasmussen, K.; Andersen, J. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sens. Environ.* **2002**, *79*, 213–224. [CrossRef]
14. Zhao, W.; Sanchez, N.; Lu, H.; Li, A. A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. *J. Hydrol.* **2018**, *563*, 1009–1024. [CrossRef]
15. Reichle, R.H.; Koster, R.D.; Liu, P.; Mahanama, S.P.P.; Njoku, E.G.; Owe, M. Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *J. Geophys. Res. Atmos.* **2007**, *112*, 112. [CrossRef]
16. Guha, A.; Lakshmi, V. Use of the scanning multichannel microwave radiometer (SMMR) to retrieve soil moisture and surface temperature over the central United States. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1482–1494. [CrossRef]
17. Yan, F.H.; Jin, Y.Q. Monitoring flooding of the Huaihe River, China, in summer 2003 using characteristic indices derived from special sensor microwave/imager (SSM/I) multitemporal observations. *Can. J. Remote Sens.* **2004**, *30*, 764–768. [CrossRef]
18. De Ridder, K. Surface soil moisture monitoring over Europe using Special Sensor Microwave/Imager (SSM/I) imagery. *J. Geophys. Res. Atmos.* **2003**, *108*, 4422. [CrossRef]
19. De Ridder, K. Quantitative estimation of skin soil moisture with the Special Sensor Microwave/Imager. *Bound. Layer Meteorol.* **2000**, *96*, 421–432. [CrossRef]
20. Wen, J.; Su, Z.; Ma, Y. Determination of land surface temperature and soil moisture from Tropical Rainfall Measuring Mission/Microwave Imager remote sensing data. *J. Geophys. Res. Atmos.* **2003**, *108*, 4038–4050. [CrossRef]
21. Seyyedi, H.; Anagnostou, E.N.; Kirstetter, P.E.; Maggioni, V.; Hong, Y.; Gourley, J.J. Incorporating Surface Soil Moisture Information in Error Modeling of TRMM Passive Microwave Rainfall. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6226–6240. [CrossRef]
22. Lee, K. A combined passive/active microwave remote sensing approach for surface variable retrieval using Tropical Rainfall Measuring Mission observations. *Remote Sens. Environ.* **2004**, *92*, 112–125. [CrossRef]
23. Jackson, T.; Hsu, A. Soil moisture and TRMM microwave imager relationships in the Southern Great Plains 1999 (SGP99) experiment. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 1632–1642. [CrossRef]
24. Wu, S.L.; Zhu, X.X. USING AMSR-E land product to monitor the drought process in China. In Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium, Honolulu, HI, USA, 25–30 July 2010; pp. 3894–3897. [CrossRef]
25. Sridhar, V.; Jaksa, W.; Fang, B.; Lakshmi, V.; Hubbard, K.G.; Jin, X. Evaluating Bias-Corrected AMSR-E Soil Moisture using in situ Observations and Model Estimates. *Vadose Zone J.* **2013**, *12*, 12. [CrossRef]
26. Ray, R.L.; Jacobs, J.M. Relationships among remotely sensed soil moisture, precipitation and landslide events. *Nat. Hazards* **2007**, *43*, 211–222. [CrossRef]
27. Mikai, H.; Arai, Y.; Mutoh, T.; Imaoka, K.; Shibata, A. A comparison of in situ precipitation with soil moisture retrieved from AMSR-E. In Proceedings of the 2005 IEEE International Geoscience and Remote Sensing Symposium, Seoul, Korea, 29–29 July 2005; pp. 3460–3461.
28. Choi, M.; Hur, Y.M.; Kim, H.; Kim, T.W.; Ahn, J.H. Passive microwave soil moisture evaluations by ground based measurements in Korea. In *100 Years Isprs Advancing Remote Sensing Science*; Wagner, W., Szekely, B., Eds.; ISPRS: Vienna, Austri, 2010; Volume 38, pp. 132–136.
29. Chakraborty, A.; Sai, M.S.; Murthy, C.; Roy, P.; Behera, G. Assessment of area favourable for crop sowing using AMSR-E derived Soil Moisture Index (AMSR-E SMI). *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 537–547. [CrossRef]
30. Al-Jassar, H.K.; Rao, K.S. Monitoring of soil moisture over the Kuwait desert using remote sensing techniques. *Int. J. Remote Sens.* **2010**, *31*, 4373–4385. [CrossRef]
31. Usowicz, B.; Marczewski, W.; Usowicz, J.B.; Lukowski, M.I.; Lipiec, J. Comparison of Surface Soil Moisture from SMOS Satellite and Ground Measurements. *Int. Agrophys.* **2014**, *28*, 359–369. [CrossRef]
32. Kornelsen, K.C.; Coulibaly, P. Design of an optimum soil moisture monitoring network using SMOS. In Proceedings of the 2014 IEEE International Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014.

33. Park, S.; Park, S.; Im, J.; Rhee, J.; Shin, J.; Park, J.D. Downscaling GLDAS Soil Moisture Data in East Asia through Fusion of Multi-Sensors by Optimizing Modified Regression Trees. *Water* **2017**, *9*, 332. [CrossRef]
34. Brunsdon, C.; Fotheringham, A.S.; Charlton, M.E. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geogr. Anal.* **1996**, *28*, 281–298. [CrossRef]
35. Zhan, C.; Han, J.; Hu, S.; Liu, L.; Dong, Y. Spatial Downscaling of GPM Annual and Monthly Precipitation Using Regression-Based Algorithms in a Mountainous Area. *Adv. Meteorol.* **2018**, *2018*, 1–13. [CrossRef]
36. Xu, S.; Wu, C.; Wang, L.; Gonsamo, A.; Shen, Y.; Niu, Z. A new satellite-based monthly precipitation downscaling algorithm with non-stationary relationship between precipitation and land surface characteristics. *Remote Sens. Environ.* **2015**, *162*, 119–140. [CrossRef]
37. Shi, Y.; Song, L.; Xia, Z.; Lin, Y.; Myneni, R.B.; Choi, S.; Wang, L.; Ni, X.; Lao, C.; Yang, F. Mapping Annual Precipitation across Mainland China in the Period 2001–2010 from TRMM3B43 Product Using Spatial Downscaling Approach. *Remote Sens.* **2015**, *7*, 5849–5878. [CrossRef]
38. Ma, Z.; Shi, Z.; Zhou, Y.; Xu, J.; Yu, W.; Yang, Y. A spatial data mining algorithm for downscaling TMPA 3B43 V7 data over the Qinghai–Tibet Plateau with the effects of systematic anomalies removed. *Remote Sens. Environ.* **2017**, *200*, 378–395. [CrossRef]
39. Draper, C.S.; Walker, J.P.; Steinle, P.J.; De Jeu, R.A.; Holmes, T.R. An evaluation of AMSR–E derived soil moisture over Australia. *Remote Sens. Environ.* **2009**, *113*, 703–710. [CrossRef]
40. Im, J.; Park, S.; Rhee, J.; Baik, J.; Choi, M. Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. *Environ. Earth Sci.* **2016**, *75*, 1120. [CrossRef]
41. Jing, W.; Yang, Y.; Yue, X.; Zhao, X. A Comparison of Different Regression Algorithms for Downscaling Monthly Satellite-Based Precipitation over North China. *Remote Sens.* **2016**, *8*, 835. [CrossRef]
42. Ma, Z.; Zhou, Y.; Hu, B.; Liang, Z.; Shi, Z. Downscaling annual precipitation with TMPA and land surface characteristics in China. *Int. J. Clim.* **2017**, *37*, 5107–5119. [CrossRef]
43. Lee, Y.J.; Park, C.; Lee, M.L. Identification of a Contaminant Source Location in a River System Using Random Forest Models. *Water* **2018**, *10*, 391. [CrossRef]