

Article

Soil Moisture Retrievals by Combining Passive Microwave and Optical Data

Cheng Tong¹, Hongquan Wang^{1,*}, Ramata Magagi², Kalifa Goïta², Luyao Zhu¹, Mengying Yang¹ and Jinsong Deng¹

¹ College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310029, China; 21714124@zju.edu.cn (C.T.); luyao_zhu@zju.edu.cn (L.Z.); 21914160@zju.edu.cn (M.Y.); Jsong_deng@zju.edu.cn (J.D.)

² Centre d'Applications et de Recherche en Télédétection, Université de Sherbrooke, QC J1K 2R1, Canada; Ramata.Magagi@USherbrooke.ca (R.M.); Kalifa.Goita@USherbrooke.ca (K.G.)

* Correspondence: hongquan_wang@zju.edu.cn

Received: 16 August 2020; Accepted: 25 September 2020; Published: 28 September 2020



Abstract: This paper aims to retrieve the temporal dynamics of soil moisture from 2015 to 2019 over an agricultural site in Southeast Australia using the Soil Moisture Active Passive (SMAP) brightness temperature. To meet this objective, two machine learning approaches, Random Forest (RF), Support Vector Machine (SVM), as well as a statistical Ordinary Least Squares (OLS) model were established, with the auxiliary data including the 16-day composite MODIS NDVI (MOD13Q1) and Surface Temperature (ST). The entire data were divided into two parts corresponding to ascending (6:00 p.m. local time) and descending (6:00 a.m. local time) orbits of SMAP overpasses. Thus, the three models were trained using the descending data acquired during the five years (2015 to 2019), and validated using the ascending product of the same period. Consequently, three different temporal variations of the soil moisture were obtained based on the three models. To evaluate their accuracies, the retrieved soil moisture was compared against the SMAP level-2 soil moisture product, as well as to in-situ ground station data. The comparative results show that the soil moisture obtained using the OLS, RF and SVM algorithms are highly correlated to the SMAP level-2 product, with high coefficients of determination ($R^2_{OLS} = 0.981$, $R^2_{SVM} = 0.943$, $R^2_{RF} = 0.983$) and low RMSE ($RMSE_{OLS} = 0.016 \text{ cm}^3/\text{cm}^3$, $RMSE_{SVM} = 0.047 \text{ cm}^3/\text{cm}^3$, $RMSE_{RF} = 0.016 \text{ cm}^3/\text{cm}^3$). Meanwhile, the estimated soil moistures agree with in-situ station data across different years ($R^2_{OLS} = 0.376\sim 0.85$, $R^2_{SVM} = 0.376\sim 0.814$, $R^2_{RF} = 0.39\sim 0.854$; $RMSE_{OLS} = 0.049\sim 0.105 \text{ cm}^3/\text{cm}^3$, $RMSE_{SVM} = 0.073\sim 0.1 \text{ cm}^3/\text{cm}^3$, $RMSE_{RF} = 0.047\sim 0.102 \text{ cm}^3/\text{cm}^3$), but an overestimation issue is observed for high vegetation conditions. The RF algorithm outperformed the SVM and OLS, in terms of the agreement with the ground measurements. This study suggests an alternative soil moisture retrieval scheme, in complementary to the SMAP baseline algorithm, for a fast soil moisture retrieval.

Keywords: soil moisture; SMAP; random forest; support vector machine; ordinary least square regression

1. Introduction

Soil moisture (SM) influences water, energy and biogeochemical cycles [1], and is a key parameter in meteorological, hydrological, ecological and agricultural systems [2–6]. Traditionally, the soil moisture is measured over fields at some measuring points or ground stations. Although the accuracy of the measured values is high, the point-based soil moisture data cannot reflect the soil moisture spatial patterns at a large scale, due to the limited number of measurements. In addition, the point measurements of SM are time-consuming, labor-intensive and affected by extreme weather [7,8].

With the advantages of timely and objective observations, remote sensing technology was widely used to monitor surface information such as land use, land surface temperature and vegetation cover [9–12]. In comparison with optical remote sensing, microwave has the penetrating ability, allowing to observe the soil surface under moderate vegetation cover for all-weather and day/night conditions. Thus, it has become an effective method to estimate soil moisture [13–16].

Microwave remote sensing consists of active and passive options, and both can perform the tasks of soil moisture retrievals: (i) passive sensors measure the surface emissivity with high temporal resolution but at coarse spatial resolution; (ii) active sensor measures the backscattering power at high spatial resolution but it is limited by low temporal resolution, except the measurements provided by radar constellations such as the Sentinel-1 and Radarsat constellation missions [17,18]. Considering their individual advantages and limitations, the authors in [19–21] focused on synergistic studies between active-passive microwave observations for soil moisture estimation. For instance, Bai, et al. [22] used the synergy of Soil Moisture Active and Passive (SMAP) radar and radiometer observations at the same spatial scale into a discrete radiative transfer model, Tor Vergata (TVG) model, to gain insights into microwave scattering and emissions mechanisms over grasslands. The SMAP satellite launched in January 2015 consists of L-band (1.26 GHz) radar with resolution of 3 km and radiometer (1.41 GHz) with resolution of 36 km. Unfortunately, the radar sensor stopped working in July 2015. Since then, the SMAP passive radiometer operates alone, providing only passive brightness temperature at a coarse resolution [23,24]. SMAP mission provides a set of soil moisture products at different spatial scales through inversion of physical radiative transfer models [25]. The zeroth-order solution of the radiative transfer equation, known as τ - ω model was used to account for the vegetation effect on the brightness temperature [26,27]. In addition to brightness temperature, a number of ancillary data concerning the vegetation and soil characteristics such as effective soil temperature and vegetation water content were also required as inputs to generate the soil moisture products [28]. The Single Channel Algorithm (SCA) based on brightness temperature at V-polarization was considered as a baseline algorithm, and the Dual Channel Algorithm (DCA) was also proposed to achieve better retrieval performance. Compared to the SMAP SCA and DCA algorithms which used the NDVI climatology to account for vegetation contribution in the brightness temperature, Ebtehaj and Bras [29] proposed a multi-channel retrieval algorithm that considers the soil types and vegetation density as a priori information to constrain the temporal changes of vegetation characteristics. This algorithm allows soil moisture retrieval at higher spatial resolution than the original radiometer data.

With the developments of artificial intelligence in recent decades, machine learning methods such as Artificial Neural Network (ANN), Support Vector Machine (SVM) and Random Forest (RF) provide new ideas to retrieve soil moisture from satellite data [30,31]. Compared to traditional physical models [25,32], machine learning methods avoid complex physical relationships, although they lack the interpretability of the retrieval results due to their black-box nature. In the machine learning methods, the estimators are trained using one portion of the total data to optimize the nonlinear relationships between satellite observations and soil moisture, followed by the validation process performed using the other portion of the data.

To retrieve the soil moisture, Lu, et al. [33] used a recurrent autoregressive neural network algorithm with AMSR2 (Advanced Microwave Scanning Radiometer) and SMOS (Soil Moisture and Ocean Salinity) data, daily NDVI, Land Surface Temperature, precipitation and DEM as trained datasets. When compared with in-situ soil moisture measurements, the retrieved results show a higher correlation coefficient (R) and lower Root Mean Square Error (RMSE) than other satellite soil moisture products such as AMSR-E. Yao, et al. [34] used a back-propagation neural network (BPNN) method to derive global long-term soil moisture series from AMSR-E/AMSR2 brightness temperature. The retrieved results agree well with SMOS soil moisture products as well as the ground station measurements. This indicates that the BPNN method can capture the surface soil moisture in terms of absolute values and temporal variations. Qu, et al. [35] applied the RF model to AMSR-E/AMSR2 brightness temperature and auxiliary data such as latitude, longitude, DEM, Day of Year (DOY) and

land classification data, in order to estimate soil moisture from 2010 to 2015 in the Qinghai-Tibet plateau. During the unfrozen seasons, the retrieved soil moisture correlated well ($R = 0.75$, $RMSE = 0.06 \text{ m}^3/\text{m}^3$) with the in-situ soil moisture networks. In addition, the performance of the trained RF estimator was evaluated against the SMAP Single Channel Algorithm at V polarization (SCA-V), indicating a high reliability of the RF model. Kolassa, et al. [36] developed an ANN-based retrieval algorithm to estimate global surface soil moisture from SMAP brightness temperatures. Compared with ground validation data, the ANN retrievals have a significantly higher performance than the NASA Goddard Earth Observing System Model version 5 (GEOS-5) land modeling system. However, the accuracy of the ANN derived soil moisture is less than that of the SMAP Level-2 product, probably due to the inappropriate target soil moisture during the ANN training process. Senyurek, et al. [37] adopted a machine learning framework for the soil moisture retrieval using NASA's Cyclone GNSS observations. Three widely-used machine learning approaches, namely ANN, RF, and SVM were tested and validated. The results reveal that the machine learning algorithms particularly the RF can be applied in soil moisture monitoring over the agricultural areas. Furthermore, to obtain finer resolution soil moisture information, Park, et al. [38] used MODIS products to downscale the 25 km AMSR2 soil moisture products to 1km via statistical ordinary least squares and RF methods.

Previous studies on soil moisture retrievals were usually based on a single method such as physical models, empirical statistical models or machine learning approaches. However, each method has distinct advantages and limitations: physical models follow complex physical laws and have high universality, but involve many physical parameters which increase the complexity; empirical statistical model is based on fitting a large set of data, but with poor applicability; machine learning approaches have high retrieval accuracy, but need a large number of training samples. Therefore, it is thus necessary to compare their performances under different soil and vegetation conditions [31,39]. In addition, the data assimilation methods were also considered for soil moisture estimation [40], but they require many variables such as atmospheric parameters (e.g., temperature, humidity), soil parameters (e.g., temperature, texture, albedo), precipitation, wind speed, elevation and land surface process models [40,41]. Thus, the number of input parameters for data assimilation approaches was larger than the statistical or machine learning approaches. Within this context, the motivation of the current paper is to identify an alternative statistical or machine learning based soil moisture retrieval algorithm that is capable to provide soil moisture information when the SMAP physical algorithms fail due to the strict constraints in surface roughness and vegetation conditions. It provides the potential to fill the data gaps when the SMAP soil moisture products are not available. To realize this objective, our study proposes to retrieve the soil moisture from SMAP brightness temperature using two representative machine learning methods, SVM [42] and RF [43], as well as an Ordinary Least Square (OLS) algorithm. These three algorithms were selected, because they were widely used in the geophysical parameter retrievals from remote sensing data, and can provide reasonable retrieval accuracies [37,38]. Then, the retrieved soil moisture from the three methods was compared to the SMAP L2 soil moisture product obtained using physical radiative transfer models, and validated against the ground measured soil moisture.

2. Study Area and Datasets

2.1. Study Area

The study site is located in the western plains of the Murrumbidgee catchment near the town of Yanco, Australia. The Yanco hydrological monitoring network data available over this site will be used to validate our soil moisture retrieval algorithms. The Yanco network contains 37 soil moisture stations distributed over a $60 \times 60 \text{ km}$ area. The soil texture was mainly composed of 11% clay, 83% sand and 6% silt. However, as shown in Figure 1, we selected only 15 stations that are distributed within a single SMAP pixel of $36 \times 36 \text{ km}$. This study site belongs to a semi-arid agricultural and grazing area, and is dominated by annual crops including rice, corn, soybeans, wheat, barley, oats, and canola [39].

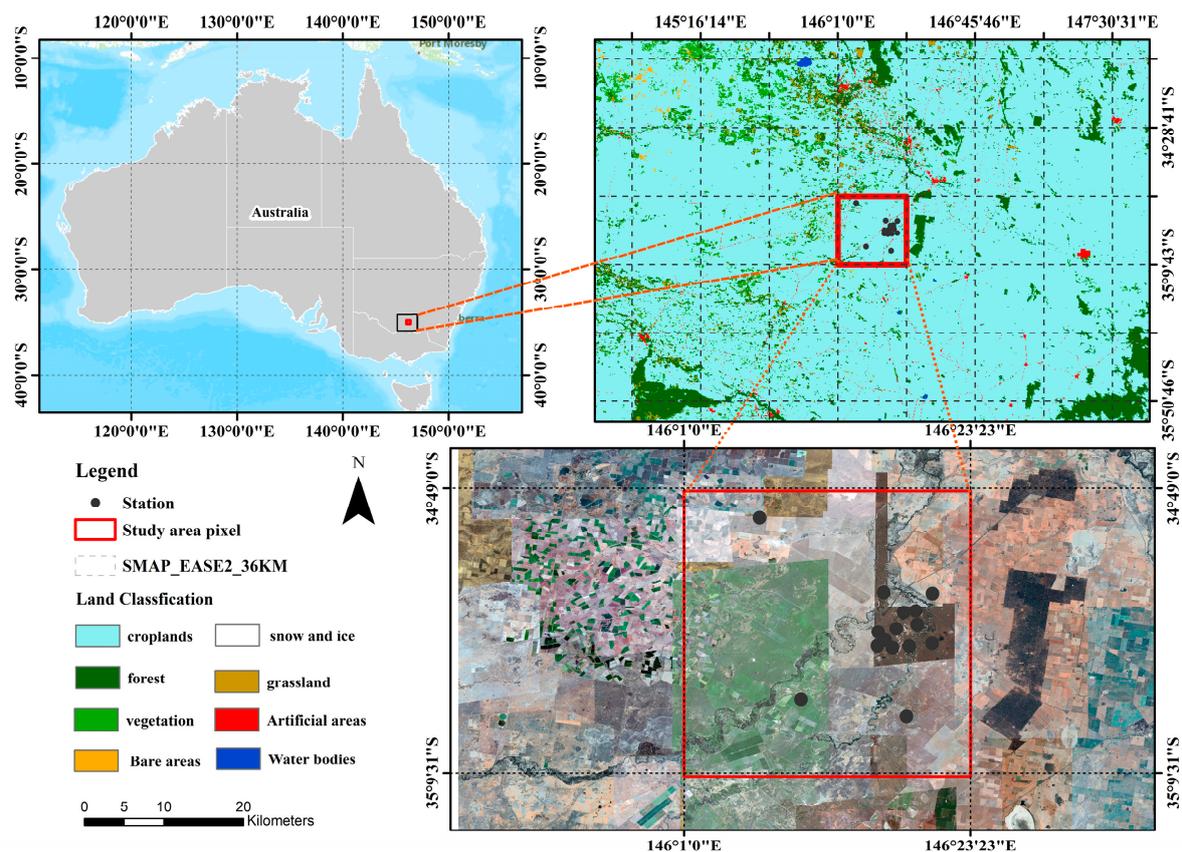


Figure 1. Study area and Yanco soil moisture ground stations.

2.2. Datasets

2.2.1. SMAP Data

The SMAP mission provides L-band microwave brightness temperature data and a series of application products. For instance, the SMAP Level-1C (L1C) data are the calibrated, geo-located and time-ordered brightness temperatures. The SMAP L2 products (version 6) include the global soil moisture and soil surface temperature at a spatial resolution of 36 km. These products were acquired from the National Snow and Ice Data Center (<http://nsidc.org/data/smap/smap-data.html>).

In Figure 2a,b, the SMAP L1C brightness temperature (descending orbits) at horizontal polarization TB_H and vertical polarization TB_V show seasonal and annual dynamics from 2015 to 2019. According to τ - ω model, the brightness temperature over vegetated soil was impacted by three emission processes: (i) vegetation direct upwelling emission; (ii) vegetation down-welling emission reflected by underlying soils and then attenuated by vegetation itself; (iii) soil emission attenuated by vegetation. Hence, the variation of brightness temperature is highly related to changes in vegetation and soil moisture.

However, compared to the radiative transfer model algorithms used to provide the SMAP soil moisture products, this study contributes to alternative machine learning algorithms to retrieve the soil moisture from the brightness temperature, surface temperature and vegetation feature such as the remotely sensed NDVI.

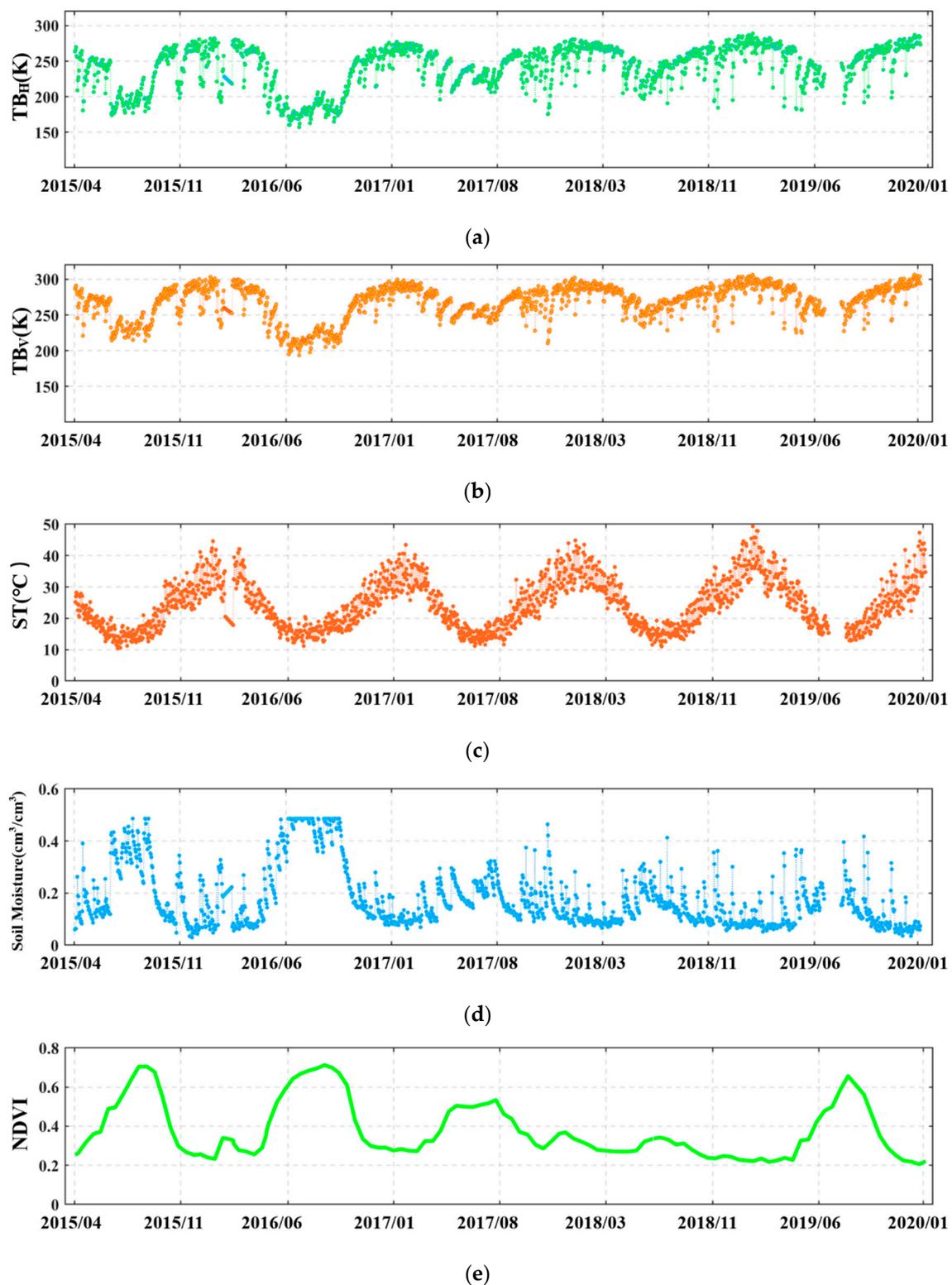


Figure 2. Temporal variations of (a) Soil Moisture Active Passive (SMAP) brightness temperature TB_H ; (b) TB_V ; (c) in situ surface temperature; (d) in situ soil moisture and; (e) MODIS NDVI composite.

2.2.2. In-Situ Soil Moisture and Surface Temperature

In-situ soil moisture, surface temperature and precipitation data were obtained from OzNet hydrological monitoring network (www.oznet.org.au) [44]. This network provides long time series of ground measurements at 20 min intervals from 2001 to the present, including the soil moisture at 0~5,

0~30, 30~60, and 60~80 cm depths, the surface temperature at 2.5 and 15 cm depths, and the precipitation measured using a rain gauge with a precision of 0.2 mm. Over each station, the soil moisture was measured using vertically installed 30 cm Campbell Scientific water content reflectometers, followed by a verification using the Time Domain Reflectometer (TDR) [44]. The accuracy of soil moisture in-situ measurement is about $0.03 \text{ m}^3/\text{m}^3$ across the stations. Given the penetration depth of the L-band microwave, this study used the in-situ soil moisture and surface temperature measurements at the top soil layer (0~5 cm and 0~2.5 cm, respectively) for the model evaluation and validation.

The SMAP satellite applied sun-synchronous high-inclination orbit, overpassing the test site at 6 a.m. and 6 p.m. for descending and ascending acquisitions, respectively. In order to match the SMAP acquisitions, the in-situ soil moisture and surface temperature data collected 3 h before and after the SMAP satellite overpasses were selected. However, the in-situ data are based on point measurements, reflecting the soil attribute for that point. Due to the spatial heterogeneity and scaling effects, it is still challenging to match the point-based in-situ measurements to the large-scale satellite product [45]. In this study, to lessen the spatial scale mismatch between the in-situ measurements and SMAP pixels, the average value of all the ground stations located within the selected pixel is calculated to represent the soil moisture for that pixel [46,47]. Furthermore, the current study used the in-situ soil temperature, since we assumed it was closer to the effective temperature adopted by the radiative transfer models. However, in case that the ground measured temperature is not available, other land surface temperature products simulated using land surface models such as the GLDAS Noah Model or the products retrieved from satellite data such as the MODIS can be considered to implement the proposed three algorithms.

Figure 2c,d shows the evolution of daily surface temperature and soil moisture from 2015 to 2019, respectively. The surface temperature shows significant annual and seasonal variation with the highest austral temperature in February and the lowest in August for each year. The study area belongs to a semi-arid agricultural and grazing area, where the increased temperature in austral summer (from December to the next year February) results in high evaporation, leading to decreasing soil moisture. Consequently, the soil moisture is lower from December to February for each year.

2.2.3. MODIS NDVI Composite

The MODIS-Terra MOD13Q1 is a level-3 grid data in a Sinusoidal projection mode at a 250 m spatial resolution and a 16-day temporal resolution. The NDVI products were generated by selecting the best available pixels from all the acquisitions during the 16-day period [48]. This paper employed the 16-day composite NDVI products to characterize the vegetation dynamics from 2015 to 2019. The MODIS data were re-projected to WGS84 using the MODIS Re-Projection Tool (MRT). To account for the difference in the spatial resolution between the SMAP and MODIS sensors, the MODIS NDVI data were aggregated into the SMAP pixel to represent the overall vegetation characteristics. Figure 2e shows the temporal variation of the aggregated NDVI for the studied pixel (Figure 1). As the study area is covered by agricultural crops, the NDVI behaviors are more induced by the phenological rhythms of the crop growths. In agreement with He, et al. [49], the NDVI reached the highest value in August of each year, revealing flourishing crop growth stages which include the flowering and anthesis. In contrast, the NDVI decreases for the ripe and the harvest periods.

3. Methodology

We developed a scheme to retrieve soil moisture from three different algorithms including Support Vector Machine (SVM), Random Forest (RF) and Ordinary Least Squares (OLS), as shown in Figure 3. Indeed, our study was motivated by the previous researches on the reconstruction of soil moisture time series. For instance, Qu, et al. [35] rebuilt a time series of soil moisture by applying the RF algorithm into the AMSR-E and AMSR2 brightness temperature. In the training process of the RF algorithm, the SMAP soil moisture was considered as the target output. In Yao, et al. [34], a global long time series of soil moisture was developed using a backpropagation artificial neural network.

They considered AMSR-E and AMSR2 brightness temperature as input variables, while the SMOS Level 3 soil moisture product as the target output. Following the above work, the current paper is to establish an alternative soil moisture retrieval algorithm when the SMAP physical algorithms fail due the roughness and vegetation conditions. In the retrieval models, the input features included SMAP brightness temperature (T_{BH} , T_{BV}), in situ Surface Temperature (ST) and MODIS NDVI product, and the output feature is the soil moisture from SMAP L2 product. By fitting the training data to the three proposed models, linear or non-linear relationships were developed between the selected input features and the SMAP soil moisture. After completing the training process, the established models were used to estimate the soil moisture from the testing data. Finally, the performances of the different models were compared against the ground station measurements and SMAP L2 product.

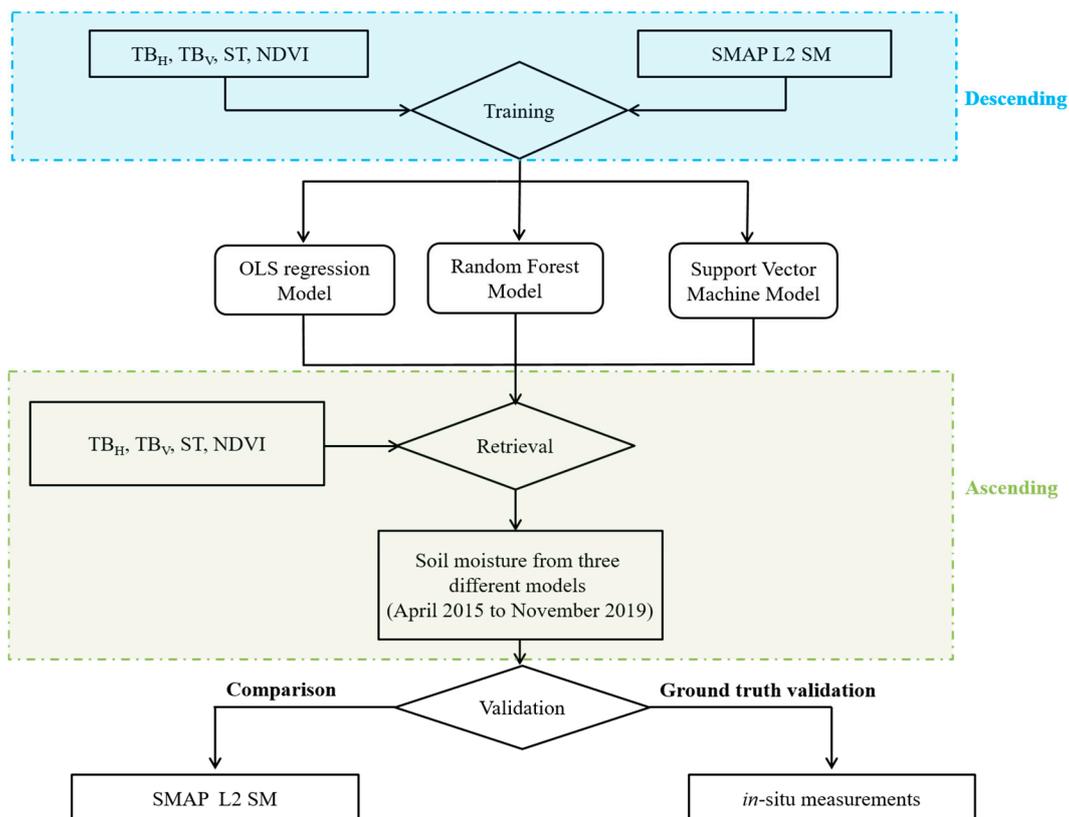


Figure 3. Diagram of soil moisture retrieval method.

3.1. Input Feature Selection Strategy

SMAP mission used the radiative transfer model τ - ω to retrieve the soil moisture from the brightness temperatures along with other ancillary data such as surface roughness, soil temperature, land cover and vegetation parameters. Thus, vegetation covers and to a lesser extent surface roughness are two key variables affecting the SMAP soil moisture retrieval algorithms. Previous studies show that the brightness temperature has a strong correlation with surface roughness, and also that the vegetation cover imposes a significant impact on surface emissivity [50]. For this reason, to select the ancillary input variables for our retrieval models which are based on the SMAP observations, the surface roughness is considered, at first. However, as our study is limited within a single SMAP pixel, the surface roughness is considered relatively stable during the multi-temporal radiometer observations. In addition, TB is less sensitive to surface roughness at coarse than high spatial resolution [51]. Consequently, the surface roughness acts like a constant in the proposed retrieval algorithms so that it isn't necessary to be included in the input features. As the brightness temperature is a product between soil effective

temperature and land surface microwave emissivity, the surface temperature is chosen as an input variable [25].

As for the vegetation characteristics, the vegetation water content (VWC) is usually used in the radiative transfer models. However, it is labor-consuming to directly measure the VWC. Previous studies reveal that NDVI can be transformed into VWC through empirical equations for different vegetation types [52,53]. As our study is based on the machine learning approaches, we assumed that the potential empirical relationships can be learned in the training process. Thus, we used the NDVI instead of VWC. As a vegetation descriptor, NDVI has the following advantages: (1) Over a large area such as the SMAP pixel, the NDVI is sensitive to abrupt accidents which changed external environment such as drought, flood and human activities. (2) It is associated with the overall vegetation coverage of the study area. (3) For long time series, the NDVI displays seasonal and inter-annual variations [54,55]. To develop the machine learning based soil moisture retrieval algorithms, daily NDVI was required. As the vegetation growth is a gradual process, the temporal NDVI data possesses significant seasonal patterns with respect to the vegetation phenology. Consequently, considering the temporal periodicity and stability of MODIS NDVI [56,57], a linear interpolation was applied to these 16-day composite products to provide daily NDVI, which can be consistent with the SMAP overpass. To derive the NDVI for a given day corresponding to the SMAP acquisition, two NDVI products before and after that day were used to realize the linear interpolation.

3.2. Model Development and Retrieval

With the selected observables, this study develops three retrieval models (OLS, SVM and RF) applied to the time series of SMAP data and MODIS products. In order to increase the dataset and to retrieve the soil moisture during the entire period 2015–2019, the training and validation processes make use of the descending and ascending SMAP measurements, respectively [58]. For soil moisture retrievals, some studies have shown the better quality of SMAP descending product at 6 a.m. than that of the ascending at 6 p.m. owing to more uniform soil/vegetation temperature and lower Faraday rotation in the morning [59,60]. Nevertheless, although the heterogeneity in soil and vegetation characteristics for the ascending orbit at 6 p.m. is complex, it was reported that the retrieved soil moisture products from ascending orbit are only slightly worse than those from the descending orbit [61,62]. In the current study, the five years SMAP descending product and the auxiliary data for the same period are selected as the training set, and the ascending data as the testing set to retrieve the 2015–2019 temporal soil moisture profile from the three algorithms. The retrieved soil moisture was then validated against the ground measurements. In addition, the SMAP L2 products that were used in training the algorithms, were also analyzed with respect to the in-situ measurements to identify the potential factors influencing the retrieval accuracy.

A. OLS Model

As a linear regression model, OLS aims to estimate the optimal coefficients for different observables to predict the soil moisture by minimizing the sum of squared errors [63]. The linearity between the four selected input variables (TB_H , TB_V , ST , $NDVI$) and the soil moisture was evaluated to determine the formula. For instance, Figure 4a shows the relationship between the SMAP brightness temperature and the soil moisture, indicating a high negative correlation ($R = -0.97$, $p < 0.001$). On the other hand, in Figure 4b, the surface temperature presents a moderate positive relationship ($R = 0.72$, $p < 0.001$) with the SMAP brightness temperature. These findings are in accordance with the formulation of linear regression for estimating soil moisture from $NDVI$, ST and other auxiliary data in Park, et al. [38]. Furthermore, the studies in [64,65] also reported that $NDVI$ and ST are strongly correlated with soil moisture and can be used to retrieve soil moisture.

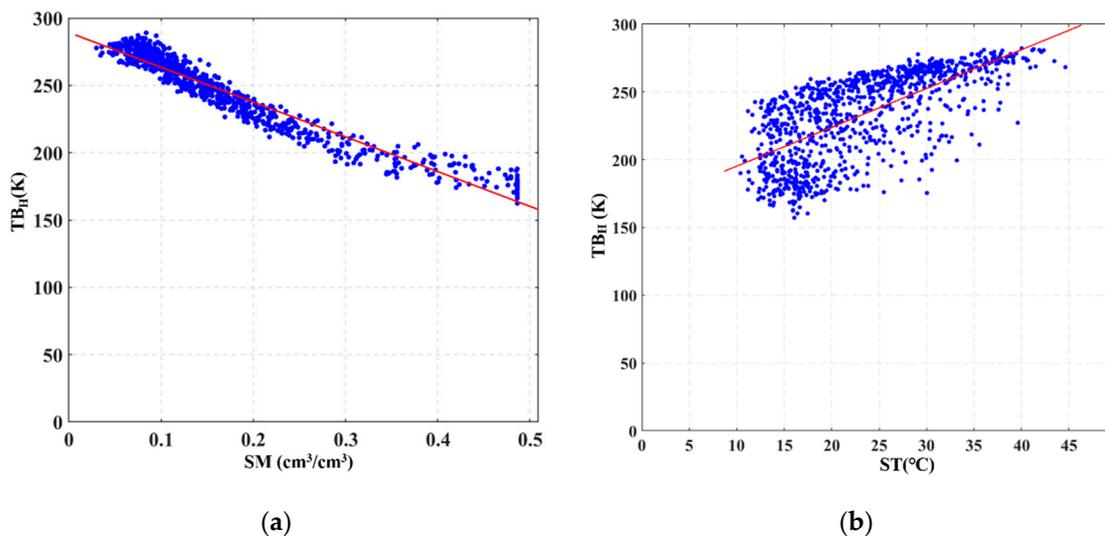


Figure 4. Relationship of brightness temperature with (a) soil moisture (b) land surface temperature.

Thus, a linear formulation for predicting soil moisture is assumed as:

$$mv = b + a_1NDVI + a_2ST + a_3TB_H + a_4TB_V \quad (1)$$

where a_1 , a_2 , a_3 , a_4 are the regression coefficients, and b represents an intercept. When the determined coefficient is positive, it means that the input variable positively influences the soil moisture values. Otherwise, the input variable negatively impacts the soil moisture values.

B. SVM Model

The non-parametric kernel-based SVM model is one of the most popular machine learning algorithms, and has been used in linear or nonlinear classification, regression and prediction [66]. The basic principle of SVM is to find optimal separating hyperplanes that divide training samples into different classes. Meanwhile, the SVM used kernel functions to transform data into a higher-dimensional space where the data can be separated via a hyperplane with maximum width. The accuracy of SVM depends on the selection of kernel functions, and the distance between the hyperplanes and training samples. Our study will evaluate the linear, polynomial and radial kernel functions used in the SVM algorithm to retrieve the soil moisture from the SMAP brightness temperature.

C. RF Model

The ensemble-learning RF algorithm has been widely applied in classification [67,68] and regression [43,69,70]. The RF consists of numerous decision trees, where each tree is built using a random subset of independent variables. Each tree casts a unit vote for the most popular class to classify the input vector [71]. Compared with other machine learning methods, RF has the advantages of high sensitivity, high precision and fast training speed. It reduced the issue of overfitting, and is especially suitable for processing high-dimensional data. In addition, RF model provides the relative importance of each variable, by measuring the increased mean square error (MSE) when that variable is changed [72]. The increased MSE represents the effect of a variable change on the prediction accuracy of the RF model: the larger the MSE value, the more important the corresponding input variable [38]. Thus, this algorithm is helpful for determining the important remote sensing variables for the soil moisture retrievals.

3.3. Statistic Metrics for Validation

To evaluate the accuracy of the developed soil moisture retrieval models, four statistical metrics including the correlation coefficient (R), root mean square error (RMSE), mean absolute percentage error (MAPE) and bias between the retrieved and measured soil moisture were computed [73]:

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum (x_i - y_i)^2}{N}} \quad (3)$$

$$MAPE = \frac{1}{N} \sum \left| \frac{x_i - y_i}{y_i} \right| \cdot 100\% \quad (4)$$

$$\text{bias} = \frac{1}{N} \sum (x_i - y_i) \quad (5)$$

where x_i , y_i represents the retrieved and referenced soil moisture for validation; \bar{x} and \bar{y} indicate their mean values, and N is the number of matching data for comparison.

4. Results and Discussion

This section discusses the training and testing retrieval results obtained from the proposed OLS, SVM and RF algorithms applied to the SMAP brightness temperature, MODIS NDVI data, and in situ soil temperature. The retrieved soil moisture was compared to SMAP L2 products, and validated using ground measurements.

4.1. Training Results

A. OLS Model

After fitting the OLS model to the training dataset, the optimal coefficients were determined by considering the minimum sum of errors. The resulting formula to retrieve soil moisture from 2015 to 2019 is given as:

$$mv = b + a_1NDVI + a_2ST + a_3TB_H + a_4TB_V \quad (6)$$

In accordance with Equation (6), the fitted positive coefficients for NDVI and ST indicate a positive relationship with the soil moisture, while the negative coefficients for TB_H and TB_V reveal a negative relationship, as expected. The R and RMSE of the formulation is 0.99 and 0.018 cm^3/cm^3 (p -value < 0.0001) respectively.

B. SVM Model

The SVM model uses different kernel functions for classification and regression. Table 1 compares the performances of the linear, polynomial and radial based functions in the SVM model. The linear kernel function achieves the best results with the highest correlation coefficient (0.93) and the lowest RMSE (0.047 cm^3/cm^3), and the corresponding cost value and gamma value are 1 and 0.25, respectively.

Table 1. Performances of different kernel functions in the Support Vector Machine (SVM) algorithm.

Kernel Function	Linear	Polynomial	Radial Basis
R ²	0.93	0.88	0.45
RMSE	0.047	0.047	0.11

C. RF Model

Once the training process is completed, the RF estimator with the importance rank of different input observables is obtained. Figure 5 shows the importance order of the four input variables in the RF algorithm. It indicates that the TB_H and TB_V play the most significant role in retrieving soil moisture, but the auxiliary NDVI and ST also have significant contributions.

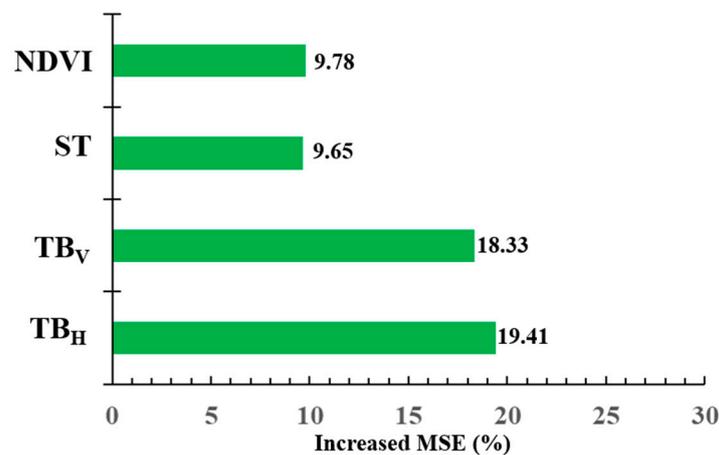


Figure 5. Relative importance of input variables for soil moisture retrieval using Random Forest (RF) algorithm.

4.2. Validation of Testing Results

The quality of the three retrieval algorithms is quantitatively evaluated by comparing the testing soil moisture results with the SMAP L2 soil moisture product and with the ground measurements from April 2015 to December 2019. The metrics used are R, RMSE, MAPE and bias values.

4.2.1. Comparison with SMAP L2 Product

Figure 6 shows linear regressions between the retrieved soil moisture from the three proposed algorithms and the SMAP L2 product. The red and black lines indicate the best-fitted curve and the nonbiased 1:1 line, respectively.

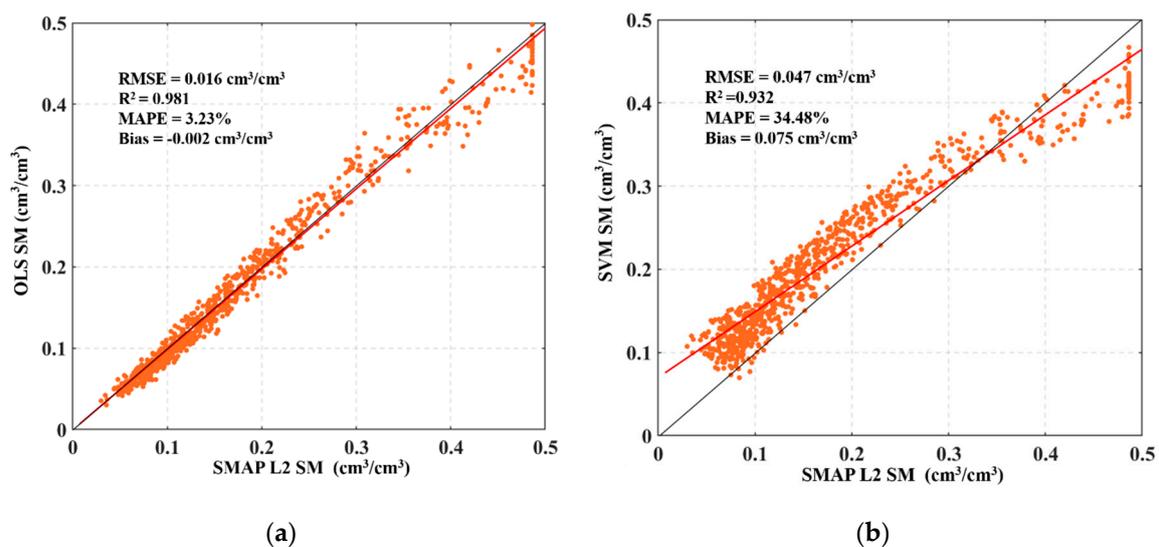


Figure 6. Cont.

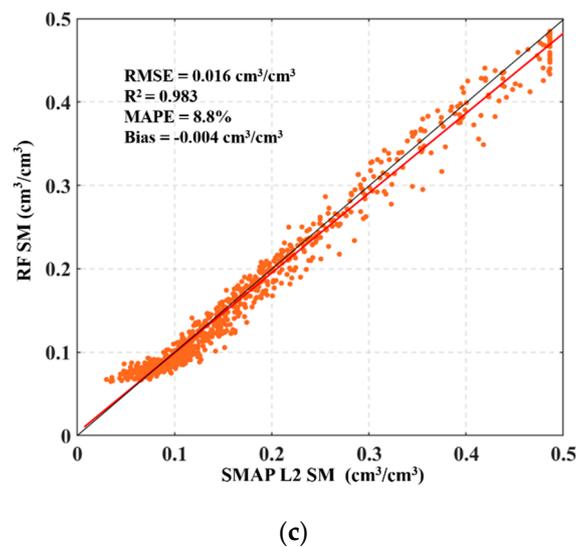


Figure 6. Comparison of the retrieved soil moisture using (a) Ordinary Least Squares (OLS) (b) SVM and (c) RF algorithms with the SMAP L2 soil moisture product.

A. OLS Model

As can be seen in Figure 6a, the retrieved soil moisture from OLS model presents strong correlation ($R^2 = 0.981$) and low RMSE ($0.016 \text{ cm}^3/\text{cm}^3$) with the SMAP L2 soil moisture product. Thus, the developed empirical OLS formula seems promising to briefly describe the physical relationships between soil moisture and the four input observables. However, for the small study area corresponding to only one SMAP pixel (Figure 1), land use and vegetation cover types are relatively single and not fully representative. Consequently, the obtained simple linearly mathematical formula may only be applicable to the environments with similar soil and vegetation characteristics as our study site.

B. SVM Model

In Figure 6b, the SVM retrieved soil moisture was related to the SMAP L2 product with a high correlation of $R^2 = 0.932$ and RMSE of $0.047 \text{ cm}^3/\text{cm}^3$. However, an overestimation occurs for soil moisture less than $0.3 \text{ cm}^3/\text{cm}^3$; beyond this threshold value, the retrieved soil moisture is underestimated compared to SMAP L2 data. Consequently, the obtained MAPE (34.48%) and bias ($0.075 \text{ cm}^3/\text{cm}^3$) using the SVM algorithm are higher than the other two approaches. The performances of the SVM is therefore the worst among the three proposed algorithms in this study.

C. RF Model

Similarly, Figure 6c shows the comparison between the RF retrieved soil moisture and SMAP L2 product, indicating a correlation of $R^2 = 0.983$ and a RMSE of $0.016 \text{ cm}^3/\text{cm}^3$. The RF model is likely suitable to retrieve the soil moisture from the SMAP brightness temperature in an efficient and simple way. This is because RF comprises numerous decision trees and adopts the average values of multiple decision trees, so that the overfitting of single linear regression is relieved. This algorithm seems powerful in the practically near-real-time retrieval of soil moisture from the SMAP instantaneous observations [35].

4.2.2. Comparison with In-Situ Measurements

To further quantify the performances of the proposed three soil moisture retrieval algorithms, we compare the retrieved soil moisture to the in-situ data measured from the Yanco ground stations. For each algorithm, Figure 7 shows the temporal variation of the retrieved soil moisture with respect to the in-situ measurements and daily precipitation from April 2015 to November 2019. The statistical

metrics between the retrieved and measured soil moisture are presented in Figure 8. In addition, the SMAP L2 product was also analyzed in terms of the in-situ measurements to identify the potential factors influencing its retrieval accuracy.

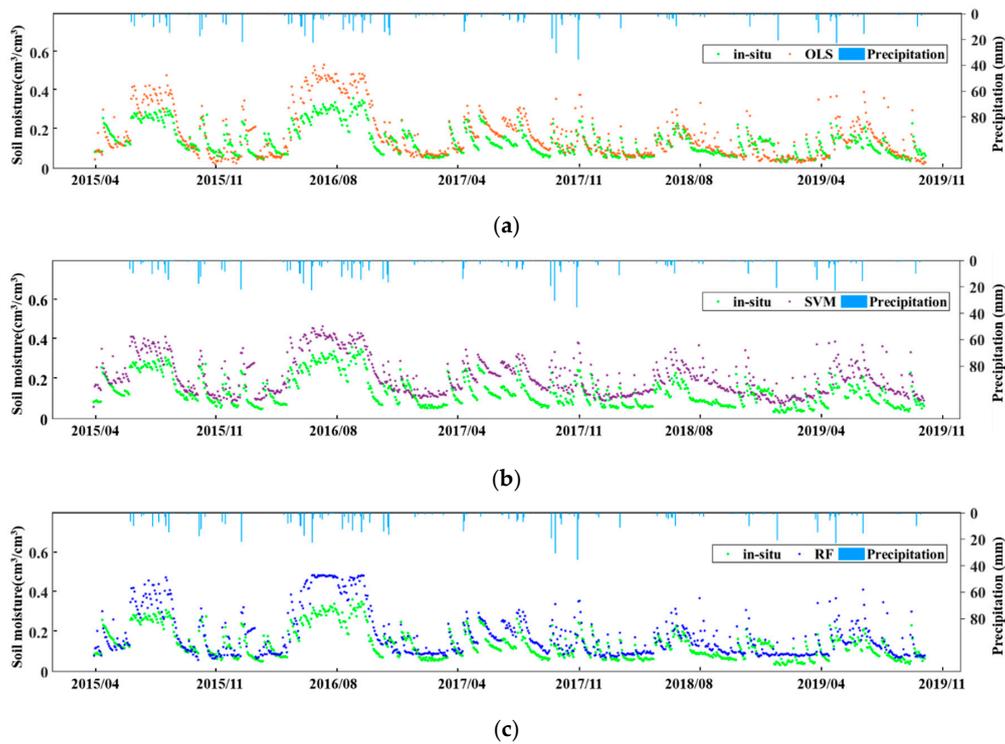


Figure 7. Temporal variation (from April 2015 to November in 2019) of the retrieved soil moisture using (a) OLS (b) SVM (c) RF along with the in-situ soil moisture and precipitation.

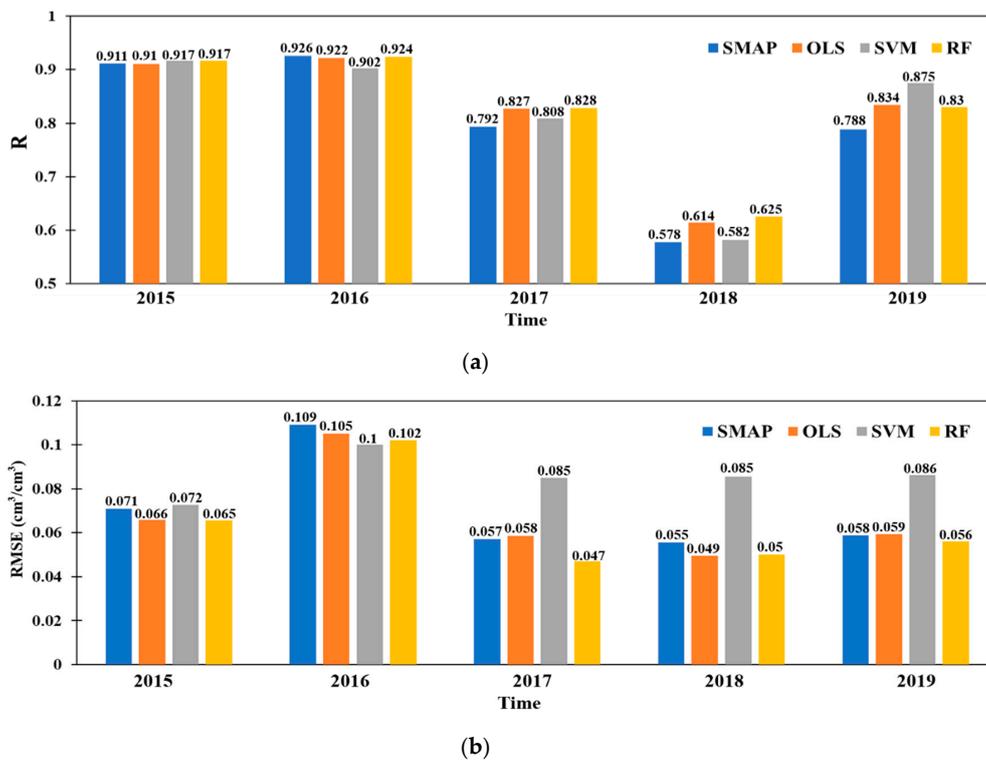


Figure 8. Annual evaluation of the retrieved soil moisture and SMAP L2 product using in-situ measurements from 2015 to 2019: (a) correlation R values; (b) RMSE values.

A. Performances of the Three Proposed Algorithms

In Figure 7, the retrieved soil moisture using the three developed models and the SMAP L2 SM product capture the temporal dynamic of in-situ measurements. However, the SVM algorithm (Figure 7a) overestimated the soil moisture during the entire period from 2015 to 2019. In contrast, for the RF (Figure 7b) and OLS (Figure 7c) derived soil moisture, an overestimation is mainly observed for specific times of the year. For example, from June to August of every year, the significant overestimation in the RF and OLS algorithms may be due to the high vegetation conditions as discussed in the next section. Furthermore, the retrieved soil moisture is related to the precipitation. The peak of soil moisture followed a significant rainfall amount but with a certain temporal lag.

Generally, the retrieved soil moisture correlates well with in-situ measurements, with encouraging metrics which are however weaker for 2018 (Figure 8). The validation against the in-situ data shows different temporal patterns. In 2015, the RF retrieved soil moisture presents the highest $R = 0.917$ and the lowest $RMSE = 0.065 \text{ cm}^3/\text{cm}^3$. In 2016, the SMAP L2 product outperforms the soil moisture retrievals from the proposed three algorithms. In 2017, RF retrieval obtains again the highest $R = 0.828$, while in 2018 OLS and RF retrieved soil moisture perform better than SVM as well as the SMAP L2 product. In 2019, SVM retrieved soil moisture provides the highest R , but with the highest $RMSE$, and RF results keep the lowest $RMSE$.

For the entire period from 2015 to 2019, Table 2 summarizes the correlation coefficient, $RMSE$ and Bias between the in-situ soil moisture measurements and the retrieved soil moisture using the three algorithms. The SMAP L2 soil moisture product was also included in the comparison with ground measurements. Compared to the SVM, the RF and OLS obtained retrieved soil moisture with a high resemblance to the SMAP L2 product.

Table 2. Comparison of model performances during the whole study period from 2015 to 2019.

	OLS	SVM	RF	SMAP
R	0.886	0.874	0.889	0.879
$RMSE (\text{cm}^3/\text{cm}^3)$	0.074	0.088	0.072	0.074
Bias (cm^3/cm^3)	0.04	0.074	0.044	0.042

B. Annual Performances of SMAP L2 Product

Figure 9 presents the temporal variation of the SMAP L2 soil moisture along with the NDVI variation from 2015 to 2019. For each year, during the period of high NDVI values, the SMAP algorithm overestimates the soil moisture. In 2015 (Figure 9a) and 2016 (Figure 9b), when the NDVI values were above 0.5 during the vegetation rapid growth season, a series of abnormal retrieved results appeared. In contrast, for the year 2017 (Figure 9a), the changes in NDVI are less pronounced and only a few NDVI values were greater than 0.5. Consequently, the issue of significant overestimation is not presented. In 2018, the NDVI fluctuated slightly with lower values than the past three years (2015~2017). Indeed, the drought of that year directly affected the normal growth of vegetation, resulting thus to low NDVI values. Meanwhile, the correlation ($R = 0.6$) between SMAP L2 product and in-situ soil moisture is the lowest compared with other years, due to the less variability in soil moisture as a result of the drought. In 2019, the drought ended, the NDVI value returned to normal; the performance of SMAP product also reached the former level (correlation coefficient R restored around 0.8). The study of Ma, et al. [74] has demonstrated the SMAP products perform well under moderate or dense vegetation conditions, whereas in the areas with sparse vegetation conditions, the SMAP products present relatively poor skills with lower time-series correlations. Our results are consistent with former research.

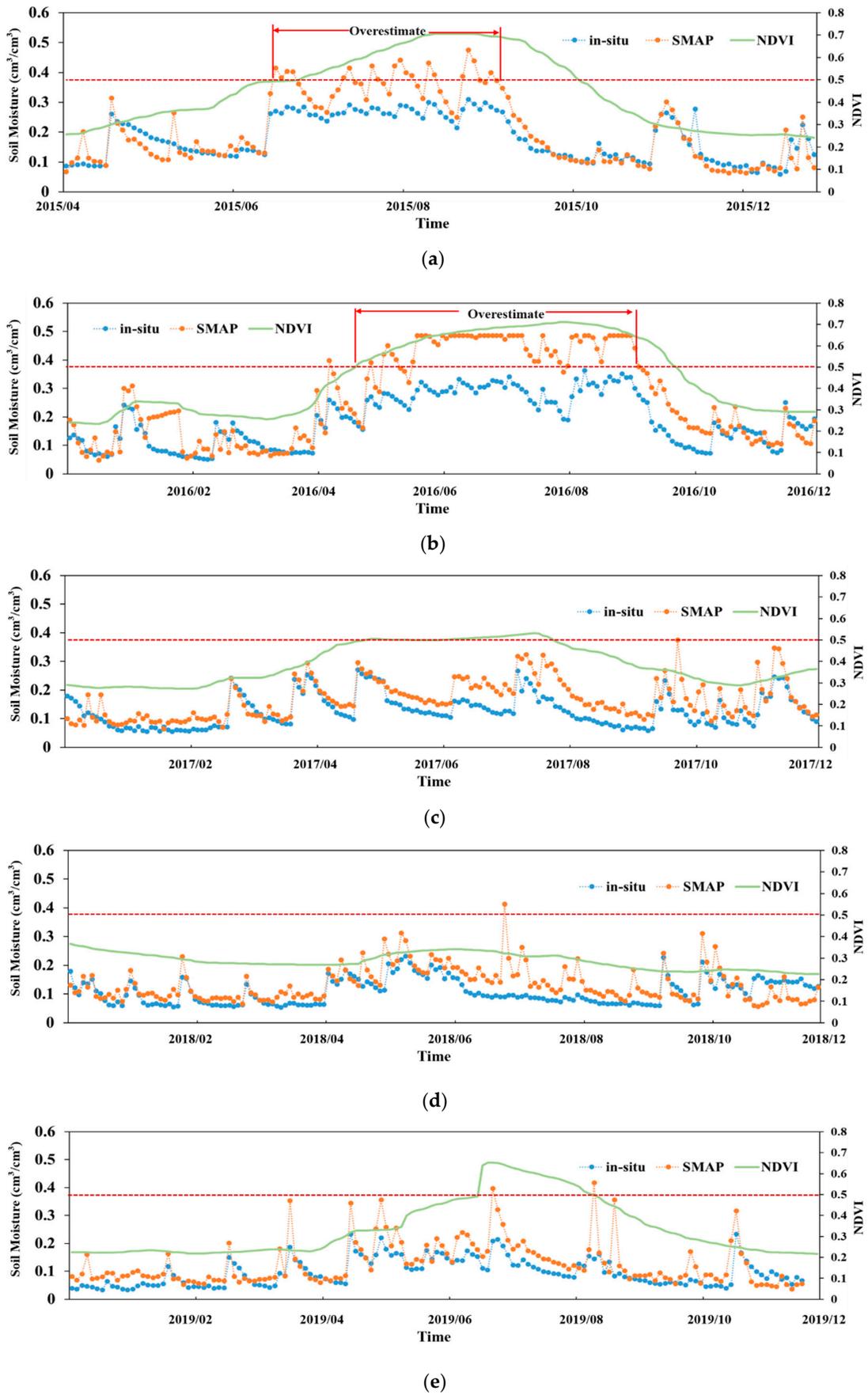


Figure 9. Temporal variation of SMAP L2 soil moisture along with in-situ measurements and NDVI in (a) 2015 (b) 2016 (c) 2017 (d) 2018 (e) 2019.

4.3. Discussion

In this study, we explored the added values of the combination of passive microwave and optical observations to retrieve soil moisture over the vegetated area. Soil moisture is not only related to these four variables, but it is also relevant to other variables such as elevation, soil texture, precipitation and land cover [16,75]. This study selects the most important physical variables similar to those used in the physical models, and obtains comparable results. Compared to the traditional physical models, the machine learning based retrieval algorithms avoid the complex formulation, and enhance the soil moisture retrieval efficiency.

Once completed, the validation process against ground measured data pointed out that the developed algorithms could provide soil moisture retrieval with consistent temporal behaviors. In recent research, Ma, et al. [74] assess several satellite soil moisture products including SMAP, SMOS, AMSR2 and ESA CCI, using global ground-based observations from dense and sparse networks. The results show that SMAP product has the capacity of capturing temporal trends of ground soil moisture. In our study, the soil moisture was retrieved using alternative machine learning algorithms instead of the radiative transfer models used in SMAP algorithms. The proposed machine learning algorithms provided retrieval results that are well correlated with the SMAP L2 product. Therefore, when the SMAP L2 soil moisture product is not available due to the limitation of the physical algorithms, the proposed machine learning approaches such as RF can cover the temporal gap using only the brightness temperature, surface temperature and NDVI as observables.

The validation of the results indicates that the retrieved soil moisture matched the temporal variation of SMAP soil moisture product, with a considerable overestimation that occurs around June to August due to the dense vegetation during this period. Indeed, when the NDVI reaches a certain high value, the flourished vegetation weakens the microwave penetration as well as the emitted signals from the soil layer, which in turn negatively impacts the accuracy of the soil moisture retrieval. On another hand, the empirical formula in Equation (6) indicated that vegetation had a positive correlation with soil moisture, and thus high NDVI value may be related to high soil moisture, justifying the observed overestimation. In further SMAP algorithm refinement, the issues of the overestimation caused by high NDVI are supposed to be taken into consideration [74].

For the SVM algorithm, we compared three kernel functions, and the linear one obtained the highest retrieval accuracy. From our knowledge, this is due to the high linearity between the SMAP brightness temperature and the soil moisture (Figure 4). In contrast, the radial basis function is more suitable for nonlinear and high dimensional relationships, and produced thus degraded performance than the linear kernel function. Furthermore, to train a reliable SVM estimator, a larger number of training data is required. However, in the current study, the number of samples is not sufficient. As for the RF model, the performance depends on the number and structure of decision trees. In the current study, the best performance was obtained with about 100 trees. Compared to SVM, the RF algorithm requires less number of training data [43], resulting thus a better performance than the SVM, given the limited number of the available training samples. Compared with the two machine-learning algorithms (SVM, RF), the empirical OLS model reflects the physical relationships between input observables and soil moisture, and provide a simple way to retrieve soil moisture. Surprisingly, the OLS obtained better performance than the SVM, which may also attribute to the significant linear relationship between TB and soil moisture. However, the regression coefficients need to be adjusted for any use with a different dataset or over different regions. Indeed, the machine learning methods can be also categorized as empirical statistical models, although they show limitations in analyzing the retrieval mechanisms due to the encapsulation in the black-box learning process. In the current study, the RF algorithm provides an overall better result than the SVM model which is impacted by a significant overestimation. This is because the RF algorithm involves the average of the estimates from multiple decision trees [35], which may decrease the variation in the predictions.

Furthermore, for a given area, the SMAP L2 soil moisture products are not always available for training the algorithms. In this case, two solutions were suggested. First, if the SMAP L2 products were

missing only for a short period, we can train the algorithms using the SMAP data beyond that missing time. Furthermore, the gap-filling algorithms [76] can be also considered to reconstruct the missing pixels in the SMAP baseline products, before training the proposed algorithms. Second, in the extreme condition that the SMAP L2 products were not available for almost all the observation time, we may consider other similar soil moisture products such as the European Space Agency (ESA) Climate Change Initiative (CCI) as the target output in the training process. According to Ma, et al. [74], the ESA CCI soil moisture is an appropriate complementarity to the SMAP L2 products across different climate conditions. However, additional bias may be introduced due to the uncertainty in the alternative soil moisture products, and the different configurations among diverse sensors.

5. Conclusions

This study proposed multiple models including Ordinary Least Squares, Random Forest and Support Vector Model to retrieve the temporal dynamics of soil moisture from SMAP brightness temperature, surface temperature and auxiliary MODIS data. The brightness temperature (T_{BH} , T_{BV}), NDVI and surface temperature were considered as input observables, while the SMAP L2 soil moisture product was considered as the output target variable. The linear or nonlinear relationships between the input and output features were obtained via the training process using a portion of the dataset. Then the obtained estimators were used to estimate the soil moisture from SMAP brightness temperature with the auxiliary vegetation feature provided by the NDVI of MODIS.

To evaluate the accuracy of the proposed three algorithms, we compared the retrieved soil moisture with the SMAP L2 product, based on the second portion of the dataset. The results indicate that the retrieved soil moisture agrees well with SMAP L2 soil moisture products. Both the RF and OLS retrieval results obtained high correlation coefficients ($R^2_{OLS} = 0.981$, $R^2_{RF} = 0.983$) and low RMSE ($RMSE_{OLS} = 0.016 \text{ m}^3/\text{m}^3$, $RMSE_{RF} = 0.016 \text{ m}^3/\text{m}^3$) with the SMAP L2 product. In contrast, the overall performance of the SVM was relatively weak.

The retrieved soil moisture was also compared to in-situ measurements. Good agreement is observed between the temporal profiles of the retrieved and in-situ soil moisture; an overestimation was also found for high vegetation conditions. Through a comprehensive comparison of the three methods, RF and OLS algorithms outperformed the SVM for soil moisture retrieval from the SMAP observations. However, the proposed algorithms were only evaluated over a limited area. In the perspective, the potential applications of the proposed algorithms along with the deep learning approaches [77,78] will also be investigated for soil moisture retrievals over other areas with diverse roughness and vegetation conditions, particularly at regional and global scales. Furthermore, other vegetation parameters such as LAI, vegetation water content and different kinds of vegetation index will be exploited to characterize the vegetation influence on soil moisture retrievals from brightness temperature.

Author Contributions: Conceptualization, H.W. and C.T.; methodology, H.W.; software, C.T., L.Z. and M.Y.; validation, H.W., R.M., K.G. and J.D.; formal analysis, C.T.; investigation, C.T., L.Z. and M.Y.; resources, H.W.; writing—original draft preparation, C.T.; writing—review and editing, H.W., R.M., K.G. and J.D.; supervision, H.W., R.M., K.G. and J.D.; funding acquisition, H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (No. 41801232) and the Fundamental Research Funds for the Central Universities (No. 2018QNA6011).

Acknowledgments: The authors thank the OzNet hydrological monitoring network for providing the soil moisture station data, and the anonymous reviewers for improving the paper.

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

References

1. Seneviratne, S.; Corti, T.; Davin, E.; Hirschi, M.; Jaeger, E.; Lehner, I.; Orlowsky, B.; Teuling, A. Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Sci. Rev.* **2010**, *99*, 125–161. [[CrossRef](#)]
2. Zveryaev, I.; Arkhipkin, A.V. Leading modes of interannual soil moisture variability in European Russia and their relation to regional climate during the summer season. *Clim. Dyn.* **2019**, *53*, 3007–3022. [[CrossRef](#)]
3. Albertson, J.; Parlange, M. Natural integration of scalar fluxes from complex terrain. *Adv. Water Resour.* **1999**, *23*, 239–252. [[CrossRef](#)]
4. Koster, R.D.; Dirmeyer, P.; Guo, Z.; Bonan, G.; Chan, E.; Cox, P.; Gordon, C.T.; Kanae, S.; Kowalczyk, E.; Lawrence, D.; et al. Regions of Strong Coupling between Soil Moisture and Precipitation. *Science* **2004**, *305*, 1138–1140. [[CrossRef](#)] [[PubMed](#)]
5. Fang, B.; Lakshmi, V. Soil moisture at watershed scale: Remote sensing techniques. *J. Hydrol.* **2014**, *516*, 258–272. [[CrossRef](#)]
6. Zhang, L.; Meng, Q.; Yao, S.; Wang, Q.; Zeng, J.; Zhao, S.; Ma, J. Soil Moisture Retrieval from the Chinese GF-3 Satellite and Optical Data over Agricultural Fields. *Sensors* **2018**, *18*, 2675. [[CrossRef](#)]
7. Shin, Y.; Mohanty, B. Development of a deterministic downscaling algorithm for remote sensing soil moisture footprint using soil and vegetation classifications. *Water Resour. Res.* **2013**, *49*, 6208–6228. [[CrossRef](#)]
8. Owe, M.; de Jeu, R.; Walker, J. A Methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 1643–1654. [[CrossRef](#)]
9. Deng, J.; Wang, K.; Hong, Y.; Qi, J. Spatio-temporal dynamics and evolution of land use change and landscape pattern in response to rapid urbanization. *Landsc. Urban Plan.* **2009**, *92*, 187–198. [[CrossRef](#)]
10. Phan, T.-N.; Kappas, M.; Thoi Nguyen, K.; Tran, T.; Vinh Tran, Q.; Rafiei Emam, A. Evaluation of MODIS land surface temperature products for daily air surface temperature estimation in northwest Vietnam. *Int. J. Remote Sens.* **2019**, *40*, 5544–5562. [[CrossRef](#)]
11. Carlson, T.; Ripley, D.A. On the Relation between NDVI, Fractional Vegetation Cover, and Leaf Area Index. *Remote Sens. Environ.* **1997**, *62*, 241–252. [[CrossRef](#)]
12. Zhuo, L.; Dai, Q.; Han, D.; Chen, N.; Zhao, B.; Berti, M. Evaluation of Remotely Sensed Soil Moisture for Landslide Hazard Assessment. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 162–173. [[CrossRef](#)]
13. Sabaghy, S.; Walker, J.; Renzullo, L.; Jackson, T.J. Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. *Remote Sens. Environ.* **2018**, *209*, 551–580. [[CrossRef](#)]
14. Mohanty, B.; Cosh, M.; Lakshmi, V.; Montzka, C. Soil Moisture Remote Sensing: State-of-the-Science. *Vadose Zone J.* **2017**, *16*, 1–9. [[CrossRef](#)]
15. Zhang, X.; Chen, B.; Zhao, H.; Li, T.; Chen, Q. Physical-based soil moisture retrieval method over bare agricultural areas by means of multi-sensor SAR data. *Int. J. Remote Sens.* **2018**, *39*, 3870–3890. [[CrossRef](#)]
16. Mohamed, E.S.; Ali, A.; El-Shirbeny, M.; Abutaleb, K.; Shaddad, S.M. Mapping soil moisture and their correlation with crop pattern using remotely sensed data in arid region. *Egypt. J. Remote Sens. Space Sci.* **2019**. [[CrossRef](#)]
17. Montzka, C.; Rötzer, K.; Bogena, H.; Sanchez, N.; Vereecken, H. A new soil moisture downscaling approach for SMAP, SMOS, and ASCAT by predicting sub-grid variability. *Remote Sens.* **2018**, *10*, 427. [[CrossRef](#)]
18. Amazirh, A.; Merlin, O.; Er-Raki, S.; Gao, Q.; Rivalland, V.; Malbeteau, Y.; Khabba, S.; Escorihuela, M.J. Retrieving surface soil moisture at high spatio-temporal resolution from a synergy between Sentinel-1 radar and Landsat thermal data: A study case over bare soil. *Remote Sens. Environ.* **2018**, *211*, 321–337. [[CrossRef](#)]
19. Du, Y.; Ulaby, F.T.; Craig Dobson, M. Sensitivity to soil moisture by active and passive microwave sensors. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 105–114. [[CrossRef](#)]
20. Kolassa, J.; Reichle, R.H.; Draper, C. Merging active and passive microwave observations in soil moisture data assimilation. *Remote Sens. Environ.* **2017**, *191*, 117–130. [[CrossRef](#)]
21. Kim, H.; Parinussa, R.; Konings, A.G.; Wagner, W.; Choi, M.; Zohaib, M. Global-scale assessment and combination of SMAP with ASCAT (active) and AMSR2 (passive) soil moisture products. *Remote Sens. Environ.* **2018**, *204*, 260–275. [[CrossRef](#)]
22. Bai, X.; Zeng, J.; Chen, K.; Li, Z.; Zeng, Y.; Wen, J.; Wang, X.; Dong, X.; Su, Z. Parameter Optimization of a Discrete Scattering Model by Integration of Global Sensitivity Analysis Using SMAP Active and Passive Observations. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1084–1099. [[CrossRef](#)]

23. Entekhabi, D.; Njoku, E.G.; Neill, P.E.O.; Kellogg, K.H.; Crow, W.T.; Edelstein, W.N.; Entin, J.K.; Goodman, S.D.; Jackson, T.J.; Johnson, J.; et al. The Soil Moisture Active Passive (SMAP) Mission. *Proc. IEEE* **2010**, *98*, 704–716. [[CrossRef](#)]
24. Das, N.N.; Entekhabi, D.; Njoku, E.G. An Algorithm for Merging SMAP Radiometer and Radar Data for High-Resolution Soil-Moisture Retrieval. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1504–1512. [[CrossRef](#)]
25. 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](#)]
26. Zheng, D.; Wang, X.; Van Der Velde, R.; Ferrazzoli, P.; Wen, J.; Wang, Z.; Schwank, M.; Colliander, A.; Bindlish, R.; Su, B. Impact of surface roughness, vegetation opacity and soil permittivity on L-band microwave emission and soil moisture retrieval in the third pole environment. *Remote Sens. Environ.* **2018**, *209*, 633–647. [[CrossRef](#)]
27. Ebrahimi-Khusfi, M.; Alavipanah, S.K.; Hamzeh, S.; Amiraslani, F.; Neysani Samany, N.; Wigneron, J.-P. Comparison of soil moisture retrieval algorithms based on the synergy between SMAP and SMOS-IC. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *67*, 148–160. [[CrossRef](#)]
28. O'Neill, P.E.; Chan, S.; Njoku, E.G.; Jackson, T.; Bindlish, R. *SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 6*; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2019.
29. Ebtehaj, A.; Bras, R.L. A physically constrained inversion for high-resolution passive microwave retrieval of soil moisture and vegetation water content in L-band. *Remote Sens. Environ.* **2019**, *233*, 111346. [[CrossRef](#)]
30. Ali, I.; Greifeneder, F.; Stamenkovic, J.; Neumann, M.; Notarnicola, C. Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. *Remote Sens.* **2015**, *7*, 16398–16421. [[CrossRef](#)]
31. Ahmad, S.; Kalra, A.; Stephen, H. Estimating soil moisture using remote sensing data: A machine learning approach. *Adv. Water Resour.* **2010**, *33*, 69–80. [[CrossRef](#)]
32. Xing, M.; He, B.; Ni, X.; Wang, J.; An, G.; Shang, J.; Huang, X. Retrieving Surface Soil Moisture over Wheat and Soybean Fields during Growing Season Using Modified Water Cloud Model from Radarsat-2 SAR Data. *Remote Sens.* **2019**, *11*, 1956. [[CrossRef](#)]
33. Lu, Z.; Chai, L.; Ye, Q.; Zhang, T. Reconstruction of time-series soil moisture from AMSR2 and SMOS data by using recurrent nonlinear autoregressive neural networks. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015.
34. 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](#)]
35. 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](#)]
36. Kolassa, J.; Reichle, R.H.; Liu, Q.; Alemohammad, H.; Gentine, P.; Aida, K.; Asanuma, J.; Bircher, S.; Caldwell, T.; Colliander, A.; et al. Estimating surface soil moisture from SMAP observations using a Neural Network technique. *Remote Sens. Environ.* **2017**, *204*, 43–59. [[CrossRef](#)]
37. Senyurek, V.; Lei, F.; Boyd, D.; Kurum, M.; Gurbuz, A.C.; Moorhead, R. Machine Learning-Based CYGNSS Soil Moisture Estimates over ISMN sites in CONUS. *Remote Sens.* **2020**, *12*, 1168. [[CrossRef](#)]
38. Park, S.; Im, J.; Park, S.; Rhee, J. AMSR2 soil moisture downscaling using multisensor products through machine learning approach. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015.
39. Gruber, A.; Paloscia, S.; Santi, E.; Notarnicola, C.; Pasolli, L.; Smolander, T.; Pulliainen, J.; Mittelbach, H.; Dorigo, W.; Wagner, W. Performance inter-comparison of soil moisture retrieval models for the MetOp-A ASCAT instrument. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, QC, Canada, 13–18 July 2014.
40. Galantowicz, J.F.; Entekhabi, D.; Njoku, E.G. Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1860–1870. [[CrossRef](#)]
41. Dumedah, G.; Walker, J. Evaluation of Model Parameter Convergence when Using Data Assimilation for Soil Moisture Estimation. *J. Hydrometeorol.* **2014**, *15*, 359–375. [[CrossRef](#)]

42. Pasolli, L.; Notarnicola, C.; Bertoldi, G.; Bruzzone, L.; Remelgado, R.; Greifeneder, F.; Niedrist, G.; Chiesa, S.D.; Tappeiner, U.; Zebisch, M. Estimation of Soil Moisture in Mountain Areas Using SVR Technique Applied to Multiscale Active Radar Images at C-Band. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 262–283. [[CrossRef](#)]
43. Wang, H.; Magagi, R.; Goita, K.; Trudel, M.; McNairn, H.; Powers, J. Crop phenology retrieval via polarimetric SAR decomposition and Random Forest algorithm. *Remote Sens. Environ.* **2019**, *231*, 111234. [[CrossRef](#)]
44. Smith, A.; Walker, J.; Western, A.; Young, R.; Ellett, K.; Pipunic, R.; Grayson, R.; Siriwardena, L.; Chiew, F.; Richter, H. The Murrumbidgee Soil Moisture Monitoring Network data set. *Water Resour. Res.* **2012**, *48*, 7701. [[CrossRef](#)]
45. Wu, X.; Xiao, Q.; Wen, J.; You, D.; Hueni, A. Advances in quantitative remote sensing product validation: Overview and current status. *Earth-Sci. Rev.* **2019**, *196*, 102875. [[CrossRef](#)]
46. Jackson, T.J.; Cosh, M.; Bindlish, R.; Starks, P.; Bosch, D.; Seyfried, M.; Goodrich, D.; Susan Moran, M.; Du, J. Validation of Advanced Microwave Scanning Radiometer Soil Moisture Products. *IEEE Trans. Geosci. Remote Sens.* **2011**, *48*, 4256–4272. [[CrossRef](#)]
47. Zeng, J.; Li, Z.; Quan, C.; Bi, H.; Qiu, J.; Zou, P. Evaluation of remotely sensed and reanalysis soil moisture products over the Tibetan Plateau using in-situ observations. *Remote Sens. Environ.* **2015**, *163*, 91–110. [[CrossRef](#)]
48. Didan, K.; Munoz, A.B.; Solano, R.; Huete, A. *MODIS Vegetation Index User's Guide (MOD13 Series)*; Vegetation Index and Phenology Lab, University of Arizona: Tucson, AZ, USA, 2015.
49. He, L.; Hong, Y.; Wu, X.; Ye, N.; Walker, J.P.; Chen, X. Investigation of SMAP Active–Passive Downscaling Algorithms Using Combined Sentinel-1 SAR and SMAP Radiometer Data. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4906–4918. [[CrossRef](#)]
50. Schmugge, T.; Neill, P.E.O.; Wang, J.R. Passive Microwave Soil Moisture Research. *IEEE Trans. Geosci. Remote Sens.* **1986**, *GE-24*, 12–22. [[CrossRef](#)]
51. Gherboudj, I.; Magagi, R.; Goita, K.; Berg, A.; Toth, B.; Walker, A. Validation of SMOS Data over Agricultural and Boreal Forest Areas in Canada. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1623–1635. [[CrossRef](#)]
52. Cheng, Y.-B.; Zarco-Tejada, P.J.; Riaño, D.; Rueda, C.A.; Ustin, S.L. Estimating vegetation water content with hyperspectral data for different canopy scenarios: Relationships between AVIRIS and MODIS indexes. *Remote Sens. Environ.* **2006**, *105*, 354–366. [[CrossRef](#)]
53. Gao, Y.; Walker, J.P.; Allahmoradi, M.; Monerris, A.; Ryu, D.; Jackson, T.J. Optical Sensing of Vegetation Water Content: A Synthesis Study. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 1456–1464. [[CrossRef](#)]
54. Jin, Z.; Xu, B. A novel compound smoother—RMMEH to reconstruct MODIS NDVI time series. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 942–946. [[CrossRef](#)]
55. Beck, P.S.A.; Atzberger, C.; Høgda, K.A.; Johansen, B.; Skidmore, A.K. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote Sens. Environ.* **2006**, *100*, 321–334. [[CrossRef](#)]
56. Xu, L.; Li, B.; Yuan, Y.; Gao, X.; Zhang, T. A Temporal-Spatial Iteration Method to Reconstruct NDVI Time Series Datasets. *Remote Sens.* **2015**, *7*, 8906–8924. [[CrossRef](#)]
57. Hu, Z.; Zhengwei, Y.; Liping, D.; Lin, L.; Haihong, Z. Crop phenology date estimation based on NDVI derived from the reconstructed MODIS daily surface reflectance data. In Proceedings of the 17th International Conference on Geoinformatics, Fairfax, VA, USA, 12–14 August 2009; pp. 1–6.
58. Chan, S.K.; Bindlish, R.; Neill, P.E.O.; Njoku, E.; Jackson, T.; Colliander, A.; Chen, F.; Burgin, M.; Dunbar, S.; Piepmeier, J.; et al. Assessment of the SMAP Passive Soil Moisture Product. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4994–5007. [[CrossRef](#)]
59. de Jeu, R.; Wagner, W.; Holmes, T.; Dolman, H.; van de Giesen, N.; Friesen, J. Global Soil Moisture Patterns Observed by Space Borne Microwave Radiometers and Scatterometers. *Surv. Geophys.* **2008**, *29*, 399–420. [[CrossRef](#)]
60. Burgin, M.; Colliander, A.; Njoku, E.; Chan, S.; Cabot, F.; Kerr, Y.; Bindlish, R.; Jackson, T.; Entekhabi, D.; Yueh, S. A Comparative Study of the SMAP Passive Soil Moisture Product with Existing Satellite-Based Soil Moisture Products. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2959–2971. [[CrossRef](#)] [[PubMed](#)]
61. Jackson, T.; O'Neill, P.; Chan, S.; Bindlish, R.; Colliander, A.; Chen, F.; Burgin, M.; Dunbar, S.; Piepmeier, J.; Cosh, M.; et al. *Calibration and Validation for the L2/3_SM_P Version 4 and L2/3_SM_P_E Version 1 Data Products*; Jet Propulsion Laboratory: Pasadena, CA, USA, 2016.

62. Jackson, T.; O'Neill, P.; Chan, S.; Bindlish, R.; Colliander, A.; Chen, F.; Burgin, M.; Dunbar, S.; Piepmeier, J.; Cosh, M.; et al. *Calibration and Validation for the L2/3_SM_P Version 5 and L2/3_SM_P_E Version 2 Data Products*; Jet Propulsion Laboratory: Pasadena, CA, USA, 2018.
63. Ouadhi, K.; Ourbih-Tari, M. Monte Carlo simulation of ordinary least squares estimator through linear regression adaptive refined descriptive sampling algorithm. *Commun. Stat. Theory Methods* **2018**, *48*, 865–875. [[CrossRef](#)]
64. Park, J.-Y.; Ahn, S.-R.; Hwang, S.J.; Jang, C.; Park, G.-A.; Kim, S. Evaluation of MODIS NDVI and LST for indicating soil moisture of forest areas based on SWAT modeling. *Paddy Water Environ.* **2014**, *12*, 77–88. [[CrossRef](#)]
65. Engstrom, R.; Hope, A.; Kwon, H.; Stow, D. The Relationship between Soil Moisture and NDVI near Barrow, Alaska. *Phys. Geogr.* **2013**, *29*, 38–53. [[CrossRef](#)]
66. Cao, X.; Chen, J.; Imura, H.; Higashi, O. A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data. *Remote Sens. Environ.* **2009**, *113*, 2205–2209. [[CrossRef](#)]
67. Zhang, Y.; Li, Q.; Huang, H.; Wu, W.; Du, X.; Wang, H. The Combined Use of Remote Sensing and Social Sensing Data in Fine-Grained Urban Land Use Mapping: A Case Study in Beijing, China. *Remote Sens.* **2017**, *9*, 865. [[CrossRef](#)]
68. Rodriguez-Galiano, V.; Chica-Olmo, M.; Abarca-Hernández, F.; Atkinson, P.; Chockalingam, J. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens. Environ.* **2012**, *121*, 93–107. [[CrossRef](#)]
69. Esteban, J.; McRoberts, R.; Fernández-Landa, A.; Tomé, J.; Næsset, E. Estimating Forest Volume and Biomass and Their Changes Using Random Forests and Remotely Sensed Data. *Remote Sens.* **2019**, *11*, 1944. [[CrossRef](#)]
70. Wang, Y.; Qi, Y.; Chen, Y.; Xie, F. Prediction of soil organic matter based on multi-resolution remote sensing data and random forest algorithm. *Acta Pedol. Sin.* **2016**, *53*, 342–354. [[CrossRef](#)]
71. Río, S.; López, V.; Benítez, J.; Herrera, F. On the use of MapReduce for Imbalanced Big Data using Random Forest. *Inf. Sci.* **2014**, *285*, 112–137. [[CrossRef](#)]
72. Chehreh Chelgani, S.; Matin, S.S.; Makaremi, S. Modeling of free swelling index based on variable importance measurements of parent coal properties by random forest method. *Measurement* **2016**, *94*, 416–422. [[CrossRef](#)]
73. Colliander, A.; Jackson, T.J.; Bindlish, R.; Chan, S.; Das, N.; Kim, S.B.; Cosh, M.H.; Dunbar, R.S.; Dang, L.; Pashaian, L.; et al. Validation of SMAP surface soil moisture products with core validation sites. *Remote Sens. Environ.* **2017**, *191*, 215–231. [[CrossRef](#)]
74. Ma, H.; Zeng, J.; Chen, N.; Zhang, X.; Cosh, M.H.; Wang, W. Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations. *Remote Sens. Environ.* **2019**, *231*, 111215. [[CrossRef](#)]
75. Hassan, A.M.; Belal, A.A.; Hassan, M.A.; Farag, F.M.; Mohamed, E.S. Potential of thermal remote sensing techniques in monitoring waterlogged area based on surface soil moisture retrieval. *J. Afr. Earth Sci.* **2019**, *155*, 64–74. [[CrossRef](#)]
76. Mao, H.; Kathuria, D.; Duffield, N.; Mohanty, B.P. Gap Filling of High-Resolution Soil Moisture for SMAP/Sentinel-1: A Two-Layer Machine Learning-Based Framework. *Water Resour. Res.* **2019**, *55*, 6986–7009. [[CrossRef](#)]
77. Chi, J.; Kim, H.-C.; Lee, S.; Crawford, M.M. Deep learning based retrieval algorithm for Arctic sea ice concentration from AMSR2 passive microwave and MODIS optical data. *Remote Sens. Environ.* **2019**, *231*, 111204. [[CrossRef](#)]
78. Fang, K.; Pan, M.; Shen, C. The Value of SMAP for Long-Term Soil Moisture Estimation with the Help of Deep Learning. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2221–2233. [[CrossRef](#)]

