



# Article Enhancing Spatial Resolution of SMAP Soil Moisture Products through Spatial Downscaling over a Large Watershed: A Case Study for the Susquehanna River Basin in the Northeastern United States

Shimelis Asfaw Wakigari \* D and Robert Leconte

Département de Genie Civil et de Genie du Bâtiment, Université de Sherbrooke, Sherbrooke, QC J1K 2R1, Canada; robert.leconte@usherbrooke.ca

\* Correspondence: shimelis.asfaw.wakigari@usherbrooke.ca

Abstract: Soil moisture (SM) with a high spatial resolution plays a paramount role in many local and regional hydrological and agricultural applications. The advent of L-band passive microwave satellites allowed for it to be possible to measure near-surface SM at a global scale compared to in situ measurements. However, their use is often limited because of their coarse spatial resolution. Aiming to address this limitation, random forest (RF) models are adopted to downscale the SMAP level-3 (L3SMP, 36 km) and SMAP enhanced (L3SMP\_E, 9 km) SM to 1 km. A suite of predictors derived from the Sentinel-1 C-band SAR and MODIS is used in the downscaling process. The RF models are separately trained and verified at both spatial scales (i.e., 36 and 9 km) considering two experiments: (1) using predictors derived from the MODIS and Sentinel-1 along with other predictors such as elevation and brightness temperature and (2) using all predictors of the first experiment except for the Sentinel-1 predictors. Only dates when the Sentinel-1 images were available are considered for the comparison of the two experiments. The comparison of the results of the two experiments indicates that the removal of Sentinel-1 predictors from the second experiment only reduces the R value from 0.84 to 0.83 and from 0.91 to 0.86 for 36 and 9 km spatial scales, respectively. Among the predictors used in the downscaling, the brightness temperature in VV polarization is identified as the most important predictor, followed by NDVI, surface albedo and API. On the contrary, the Sentinel-1 predictors play a less important role with no marked contribution in enhancing the predictive accuracy of RF models. In general, the two experiments have limitation, such as a small sample size for the training of the RF model because of the scarcity of Sentinel-1 images (i.e., revisit time of 12 days). Therefore, based on this limitation, a third experiment is proposed, in which the Sentinel-1 predictors are not considered at all in the training of the RF models. The results of the third experiment show a good agreement between the downscaled L3SMP and L3SMP\_E SM, and in situ SM measurements at both spatial scales. In addition, the temporal availability of the downscaled SM increased. Moreover, the downscaled SM from both SMAP products presented greater spatial detail while preserving the spatial patterns found in their original products. The use of the two SMAP SM products as background fields for the downscaling process does not show marked differences. Overall, this study demonstrates encouraging results in the downscaling of SMAP SM products over humid climate with warm summers dominated by vegetation.

Keywords: soil moisture; downscaling; resolution; validation; Sentinel-1; MODIS; SMAP; random forest

# 1. Introduction

Soil moisture (SM), as a key hydrological variable, is of paramount importance in modulating terrestrial water cycle components such as precipitation by partitioning into surface runoff and infiltration; subsequently, affecting the streamflow and ground water recharge [1]. It also affects regional and global climate systems through its control of the



Citation: Wakigari, S.A.; Leconte, R. Enhancing Spatial Resolution of SMAP Soil Moisture Products through Spatial Downscaling over a Large Watershed: A Case Study for the Susquehanna River Basin in the Northeastern United States. *Remote Sens.* 2022, *14*, 776. https://doi.org/ 10.3390/rs14030776

Academic Editor: Gabriel Senay

Received: 26 December 2021 Accepted: 2 February 2022 Published: 7 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). interaction between land surface and atmosphere via evapotranspiration, thereby affecting air temperature and precipitation [2,3]. In addition, knowledge of SM is important in agriculture for crop growth [4,5]. Similarly, a large number of operational hydrological and agricultural applications, including flood forecasting [6,7], drought monitoring [8,9] and irrigation planning [10], are heavily dependent on SM. Hence, the proper characterization of the spatial and temporal variations of SM is imperative.

The spatio-temporal variability of SM is often ascribed to multifaceted interactions of numerous environmental factors, including soil properties, surface topography, vegetation and atmospheric forcing [11]. Traditionally, this variability is monitored using in situ measurements which offer an accurate estimate of SM at different depths with a high temporal resolution [12,13]. However, despite its accuracy, this method is costly, time consuming and, most importantly, suffers from low spatial representativeness (i.e., only in the order of a few centimeters). For these reasons, over the last few decades, a number of satellite remote sensing platforms have been launched to address the need for economically feasible global information on SM with a temporal frequency of a few days to a few weeks.

Among satellites, remote sensing, active and passive microwave remote sensing satellites have demonstrated the potential for measuring near surface soil moisture under all-weather conditions depending on the underlying land surface conditions [14,15]. Active microwave remote sensing has the capability to provide SM at a high spatial resolution (i.e., in the order of tens of meters) [16], but it is heavily affected by perturbing factors such as surface roughness and vegetation, thereby complicating SM retrieval [14]. Comparatively, passive microwave remote sensing has long been recognized as a prominent tool for the retrieval and monitoring of surface soil moisture, owing to its sensitivity to near surface soil moisture dynamics and the capability to penetrate to a certain extent through the vegetation canopy, but suffers from its low spatial resolution (i.e., in the order of tens of kilometers) [17,18].

Over the last few decades, several passive microwave satellites with different frequencies (e.g., L, C and X) have been launched, providing global coverage of surface SM information. Among the several frequencies, the L-band passive microwave satellites (e.g., Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active Passive (SMAP)) are the most promising and primarily devoted to global monitoring of surface soil moisture [19–21].

While the advent of L-band passive microwaves satellites has effectively addressed the difficulties associated with the continuous monitoring of near surface soil moisture over large areas, the spatial resolution of SM products derived from these satellites is in the order of several kilometers (i.e., 25-50 km), which impedes their utilization for a large variety of local and regional applications [22,23], such as agriculture and hydrology. Therefore, a detailed representation of sub-grid-scale spatial variability of SM within the coarse spatial resolution of passive microwave SM products would significantly benefit these applications.

In order to address the need for spatially detailed SM information, a growing number of studies have attempted to provide high-spatial-resolution SM maps through downscaling by leveraging the complementary strength of passive microwave and radar and/or optical/thermal observations [22–30]. In this context, downscaling takes advantage of high-spatial-resolution land surface variables derived from radar and/or optical/thermal sensors to disaggregate the coarse resolution of passive microwave SM products.

A variety of downscaling techniques has been proposed and tested, which can be roughly grouped into three categories based on the sources of high-resolution land surface variables used in downscaling [31,32]. These are (1) optical/thermal-based downscaling, (2) radar-based downscaling and (3) radiometer-based downscaling. In addition, there are also emerging techniques such as machine learning (ML) and data assimilation. Sabaghy et al. [32] summarized the performances of these downscaling techniques reported by many studies conducted in different regions of the world under varying climate and land surface conditions. Based on their comparison, optical/thermal-based downscaling was identified as the most widely applied technique to downscale passive microwave SM,

although with varying degrees of accuracy, whereas radar-based downscaling and ML models are considered the two most promising approaches for SM downscaling with a higher degree of accuracy. An extensive review of these techniques can be found in [31,32].

The optical/thermal-based downscaling generally relies on the triangular/trapezoidal feature space established between vegetation indexes and land surface temperature derived from optical/thermal satellites (e.g., MODIS) over heterogeneous land surfaces [33,34]. The premise of this technique is that the sensitivity of SM to land surface temperature varies under a wide range of land surface conditions [35]. The empirical polynomial fitting and semi-physical evaporation-based approaches are among the techniques that fall under this category. A large number of studies have applied these techniques to improve the spatial resolution of passive microwave observations of SM (e.g., [36–38])). For instance, Piles et al. [39] applied the polynomial function to downscale SMOS SM from 40 km to 10 and 1 km. Likewise, Djamai et al. [40] applied a semi-physical model called DISPATCH (DISaggregation based on Physical And Theoretical scale CHange) to downscale SMOS SM to 1 km over the Canadian Prairies.

Another alternative approach is the radar-based downscaling technique, which synergistically combines coarse-resolution passive microwave SM or brightness temperature with high-resolution synthetic aperture radar (SAR) observations to enhance the spatial detail of SM. A typical example of this technique is the SMAP baseline downscaling algorithm [41], which has been utilized by NASA on the SMAP mission to downscale the coarse-scale passive microwave brightness temperature (36 km) to an intermediate resolution (9 km) by making use of the high-resolution (3 km) SAR backscatter, which was then inverted to retrieve SM at a 9 km spatial resolution. In addition, there are several other radar-based downscaling techniques, including the SMAP optional method [42], the change detection method [43] and the Bayesian framework [44]. Despite the great potential of radar-based downscaling in improving the spatial resolution of SM, the lower revisit frequency (e.g., 6-12 days) and lack of a temporally collocated passive microwave radiometer and SAR measurements is still a big challenge after the failure of the SMAP radar. However, in order to continue the production of high spatial resolution SM by the SMAP mission, significant efforts have been undertaken to substitute the SMAP radar with an active radar from other satellites [45]. Among the existing radar satellites, the European Space Agency's (ESA) Sentinel-1A/1B satellites were considered suitable due to the similarity of their orbital configuration with that of SMAP [46,47].

Apart from the above downscaling methods, ML has recently received a great deal of attention for the downscaling of passive microwave SM by learning the relationship between SM and a set of land surface variables [22,48]. It has become increasingly popular mainly due to (1) its ability to handle the complex non-linear relationship between SM and a large set of land surface variables across a variety of heterogeneous surface conditions, (2) the rapid increase in computational power and public availability of remote sensing data and (3) the relatively good performance as compared to aforementioned downscaling techniques.

Numerous studies have reported successful applications of various ML techniques, e.g., artificial neural network (ANN), classification and regression trees (CART), gradientboosting decision tree (GBDT) and random forest (RF), in the downscaling of passive microwave SM products [22,23,49,50]. Among these techniques, a large number of studies have shown superiority of RF because of its potential to greatly reduce the problem of overfitting (e.g., [48,51,52]). For example, Zhao et al. [22] downscaled SMAP SM to 1 km over the Iberian Peninsula using RF, and found that the RF had the capability to capture the spatial variability of SM with great accuracy compared to the polynomial function fitting approach. Similarly, Liu et al. [52] compared six variants of ML methods and revealed that RF showed better performance than all others.

In general, the application of aforementioned downscaling techniques enhances the spatial resolution of passive microwave SM with varying degrees of accuracy. However, most of these techniques are predominantly applied to the arid and semi-arid climatic

regions, and their performances are often site-specific (e.g., [22,25,28,53]), requiring further investigation across diverse geographical and climatic regions with a range of land cover types.

In the present study, RF models are applied for the downscaling of two SMAP SM products (i.e., L3SMP (36 km) and L3SMP\_E (9 km)) to a 1 km spatial resolution over the warm summer humid climate of the Susquehanna watershed. Studies have exploited the potential strengths of a suite of high-resolution land surface variables and indices derived from optical/thermal remote sensing such as the land surface temperature (LST), normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), surface albedo and other auxiliary variables (e.g., precipitation, soil texture and surface topography) in the downscaling of passive microwave satellite SM using RF [22,25,54]. However, the use of synthetic aperture radar (SAR) observations, such as those provided by the Sentinel-1 C-band SAR, as predictors notably in RF-based downscaling is often limited [54]. In other words, although a number of other approaches were proposed to merge C-band SAR and passive microwave SM, the use of SAR information in RF-based downscaling is limited. As a result, it is necessary to clarify the contribution of SAR products (i.e., backscattering intensity and its textural information) along with other remotely-sensed products in the downscaling of SMAP SM products through comprehensive assessments.

Considering the abovementioned premises, the purpose of the present study is to: (1) assess the performance of the RF-ML technique in the downscaling of SMAP SM products over the warm summer humid climate of a large watershed with a variety of land cover types with a predominance of forests; (2) exploit the benefit of the high-resolution Sentinel-1 SAR backscattering intensity and its textural information along with MODIS products and other auxiliary variables in downscaling SMAP SM products; (3) identify the best set of predictor variables for downscaling for underlying surface and climate conditions of the study area; (4) compare the downscaled and the existing 1 km high-resolution SMAP/Sentinel-1 SM products against ground-based in situ SM and precipitation; (5) compare the relative merits of the SMAP level-3 (L3SMP) versus SMAP enhanced (L3SMP\_E) SM as background field in downscaling.

# 2. Materials and Methods

# 2.1. Study Area

The study area, Susquehanna River Basin, is located in the eastern part of the United States, with an approximate area of 71,000 km<sup>2</sup> (Figure 1). The basin experiences high flows during spring season due to combined effect of snowmelt and rainfall, while low flows occur in late summer and early fall. The elevation of the basin varies from approximately 52 m below mean sea level to 1444 m above mean sea level [55].

The watershed is characterized by a mild, subtemperate and humid climate marked by cold winters with snow events and warm to hot summers. The average annual precipitation ranges from approximately 800 to 1250 mm, while the mean annual temperature is roughly 6 and 12 °C in the northern and southern portion of the basin, respectively. The basin is mainly dominated by forests (55%), followed by agriculture (33%) [55].



Figure 1. Study area: (a) DEM and (b) land cover.

## 2.2. Description of Data

The dataset used in this study was derived from variety of sources, including: (1) remote sensing satellites, e.g., SMAP, MODIS, Sentinel-1; (2) models, e.g., PRISM, DEM; (3) in situ measurements (e.g., ISMN). The data collected during unfrozen period (i.e., May to October) of 2017 to 2020 were used to circumvent the poor quality of SMAP SM during the remaining months of the year because the soil was likely to be frozen near the surface during winter. A brief description of the main features of these data is presented as follows.

# 2.2.1. SMAP Soil Moisture and Brightness Temperature

The Soil Moisture Active Passive (SMAP) mission is the latest L-band satellite mission launched on 31 January 2015 by the National Aeronautics and Space administration (NASA). It was primarily dedicated to measuring near surface soil moisture and landscape freeze/thaw conditions below the ground surface (top ~5 cm). It is a sun-synchronized near-polar orbit satellite with local equator overpass times of 6:00AM (descending) and 6:00PM (ascending). SMAP incorporates both radar (active) and radiometer (passive) on the same platform, aiming to offer SM at range of spatial resolutions, including 3, 9 and 36 km. Nevertheless, due to SMAP radar failure that occurred on 7 July 2015, the generation of the 3 km (active) and 9 km (active–passive) SM products was halted. Consequently, in order to continue the generation of SM at aforementioned resolutions, the SMAP mission proposed: (1) to replace the 9 km active–passive SM with SMAP enhanced SM, which was retrieved from the inversion of interpolated coarse-resolution SMAP L3 brightness temperature using the Backus–Gilbert optimal interpolation technique; (2) to replace the original SMAP 3 km SM products with downscaled SMAP enhanced SM using Sentinel-1 SAR backscatter.

In this study, the SMAP level-3 (L3SMP) and SMAP enhanced (L3SMP\_E) SM and brightness temperature with a resolution of 36 km and 9 km, respectively, were used.

These products had temporal resolution of 2–3 days, posted on the 36 km and 9 km Equal-Area Scalable Earth (EASE) grid, respectively. In addition, 1 km high-spatial-resolution SMAP/Sentinel-1 (L2\_SM\_SP) images were collected to compare with the downscaled SMAP SM product in this study. L2\_SM\_SP had a temporal resolution of 12 days. These data can be downloaded freely from the NASA Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC) (https://nsidc.org/data/smap/ (accessed date 30 May 2019)).

# 2.2.2. Sentinel-1 Backscatter and Textural Features

Sentinel-1 is a constellation mission composed of two satellites (i.e., Sentinel-1A and Sentinel-1B), each carrying an imaging C-band (5.405 GHz) synthetic aperture radar (SAR) instrument in a near-polar sun-synchronous orbit to monitor the land and ocean [16]. It has four data acquisition modes: stripmap, interferometric wide (IW) swath, extra-wide (EW) swath and wave. The IW swath is the prime acquisition mode over land surfaces, which provides observations at spatial and temporal resolution of 10 m and 6–12 days, respectively, in both VV and VH polarizations.

In the present study, the Sentinel-1 images collected in IW mode during ascending overpass were used. The post-processing of the images, including thermal noise removal, speckle filtering, radiometric calibration, terrain correction and normalization, was performed using the Sentinel application platform (SNAP) tool. In addition, four less correlated textural features, namely, contrast, entropy, correlation and variance, were extracted from the Sentinel-1 images using the grey-level co-occurrence matrix (GLCM) provided by SNAP. These features were extracted with a moving window of  $5 \times 5$  and average orientation of all directions (0°, 45°, 90° and 135°). The Sentinel-1 imagery is openly available to users through ESA's Sentinel-1 internet data hub (https://scihub.esa.int/ (accessed date 25 June 2019)).

### 2.2.3. MODIS Products

The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor on board NASA's Earth Observing System (EOS) Terra and Aqua satellites is widely used for the monitoring and understanding of the spatiotemporal dynamics of land, ocean and atmosphere at the surface [56]. It operates on a solar synchronous polar orbit with local equator overpass times of around 10:30 a.m. (descending) and 10:30 p.m. (ascending) for Terra and 1:30 a.m. (descending) and 1:30 p.m. (ascending) for Aqua. It provides a suite of global products every 1–2 days with spatial resolution ranging from 250 m to 1000 m. In this study, the latest version (i.e., Collection 6) of MODIS land surface variables which had a strong connection with SM, including 1 km daily Terra and Aqua LST (MYD11A1, MOD11A1), 1 km 16-day vegetation index (MOD11A2) and 500 m daily surface albedo (MCD43A3), were retrieved from the NASA Land Processes Distributed Active Archive Center (LP DAAC) (https://lpdaac.usgs.gov (accessed date 28 June 2019)).

## 2.2.4. Surface Topography

Topographic attributes, including elevation, slope and aspect, were extracted from the digital elevation data obtained from the NASA Shuttle Radar Topography Mission (SRTM) digital elevation model at 30 m resolution. These data were downloaded from the Land Processes Distributed Active Archive center (https://lpdaac.usgs.gov/ (accessed date 1 July 2019)).

#### 2.2.5. Precipitation and Antecedent Precipitation Index (API)

The spatiotemporal variability of SM is highly affected by precipitation patterns. In addition, precipitation that occurred over preceding days has a considerable effect on SM (i.e., SM memory) of a given day. In this study, daily precipitation data obtained from Parameter–elevation Relationships on Independent Slopes Model (PRISM) [57] at a resolution of 4 km were used to calculate the modified antecedent precipitation index (modified

API), aiming to include the effect of precipitation of the past few days on downscaled SMAP SM of a given day. This data are available covering the entire conterminous United States (http://www.prism.oregonstate.edu/ (accessed date 3 July 2019)).

# 2.2.6. In Situ Soil Moisture

In situ SM measurements of the top 5 cm depth of the soil column and precipitation data were collected from the soil climate analysis network (SCAN) and the U.S. climate reference network (USCRN). There were about 4 SM probes which lay within the study area, including Avondale, Ithaca, Geneva and Rocksprings (Figure 1). The probes provided hourly values of volumetric SM. The dominant land cover types for Avondale and Geneva were mosaic herbaceous and grassland, respectively, whereas for Ithaca and Rocksprings it was crop. The average annual precipitation for the four stations varied from 850 to 1280 mm (i.e., Geneva (855 mm), Rockspring (1015 mm), Ithaca (1100 mm) and Avondale (1280 mm)), whereas average annual temperature ranged from 8 to 12 °C (i.e., Ithaca (8 °C), Geneva (9 °C), Rockspring (10 °C) and Avondale (12 °C)). The elevation of the probe locations varied from 122 to 374 m (i.e., Avondale (122 m), Geneva (221 m), Rockspring (372 m) and Ithaca (374 m)). These data were collected from the International Soil Moisture Network (ISMN) database (https://ismn.geo.tuwien.ac.at/en/ (accessed date 10 July 2019)).

The basic characteristics of dataset, such as spatial and temporal resolutions, used in this study were summarized in Table 1.

Datasets (or Product Name)	Variables/Indices	Temporal Resolution (Day)	Spatial resolution (km)
SMAP L3 SM	Soil moisture	1–3	36
SMAP enhanced	Brightness temperature	10	9
SMAP/Sentinel-1	Soil moisture	12	1/3
Sentinel-1 SAR	Sigma VV/VH, Contrast, Correlation, Variance, Entropy	12	10 m
MOD11A1 MYD11A1	LST	1	1
MOD13A2	NDVI	16	1
MCD43A3	Albedo	1	500 m
SRTM	Elevation, Slope, Aspect	-	90 m
PRISM	Precipitation, API	1	4
ISMN	Soil moisture, Precipitation	1	-

Table 1. Summary of data used for the downscaling process in this study.

#### 2.3. Data Pre-Processing

All the data used in this study were reprojected to the same coordinate system (i.e., WGS 84) and rigorous data quality control was applied to remove poor-quality pixels which might have been affected by the presence of clouds or aerosols. In addition, to solve the spatial resolution differences among the predictors and response variables, two methods were adopted: simple arithmetic averaging and the nearest-neighbor interpolation. The simple arithmetic averaging was used to aggregate the high-resolution predictors (e.g., LST, Sigma VV/VH) to coarse resolution, whereas the nearest-neighbor interpolation was adopted for resampling of coarse-resolution predictors (e.g., SMAP brightness temperature and precipitation) to fine resolution. The open water bodies in the study area were also masked out using MOD44W, which is a 250 m resolution static global water map provided by MODIS.

#### 2.4. Random Forest

In this study, random forest (RF) ML technique was adopted for downscaling of L3SMP (36 km) and L3SMP\_E SM (9 km) SM to 1 km. It has increasingly gained popularity within the remote sensing community to solve problems related to classification and regression [58]. It is a non-parametric tree-based model formed from an ensemble of decision trees [59].

RF has great capability to efficiently describe the complex non-linear relationship between SM and a suite of land surface variables as shown in Equation (1) [60].

$$SM = f(p_1, p_2, p_3 \dots p_n) + \varepsilon \tag{1}$$

where *SM* is soil moisture;  $\varepsilon$  is the model estimation error,  $p_1, p_2, p_3 \dots p_n$  represent predictors (i.e., NDVI, albedo, LST); *n* represents the total number of predictors.

RF is primarily based on the bootstrap aggregation (bagging), which is a statistical technique used for random resampling of the training dataset with a replacement in order to generate a number of subsets (i.e., bootstrap samples). In this work, about two-thirds of each bootstrap sample (i.e., in-bag (IB)) were used for independent training of each decision tree of the ensemble, whereas the remaining one-third (i.e., out-of-bag (OOB)) was used for verification of the model. In the end, the final predicted value was obtained through averaging of the prediction of each regressor.

RF has a few user-defined parameters to tune in order to increase its predictive power. These are the number of decision trees to grow, maximum number of features selected at each split and the maximum depth of the decision tree. In addition, another attractive feature of RF is its ability to provide the relative importance of predictors by using out-ofbag data. The scikit-learn Python library, which has several built-in functions, was used for implementation of RF algorithm [61].

In the present study, the construction of RF model was based on using all available data during the summer seasons of 2017 to 2020 over the Susquehanna watershed, as opposed to construction for each individual day. This was primarily because of the lack of sufficient training data if the model was constructed for each individual day for a small size of study area such as ours. For example, only 130 pixels of SMAP SM covered our study area. Thus, if the model was constructed based on each day, we would only have 130 data points of SMAP SM with their corresponding aggregated predictors for training of RF model at coarse spatial resolution, which is really very small. Thus, to circumvent the lack of sufficient training data, we used all available data during the study period for training of RF model. This agreed with the previous studies conducted in relatively small-sized study areas in different regions of the world [25,48,62].

In this study, the implementation of RF model for downscaling of SMAP SM entailed five general steps: (1) **aggregation:** involved aggregation of high-resolution land surface variables (i.e., predictors) to SMAP SM resolution using simple arithmetic averaging; (2) **splitting of dataset:** involved dividing of aggregated dataset into training and testing sets; (3) **developing the linking model**: this entailed training the RF using SMAP SM as response variable and aggregated land surface variables as predictors (Figure 2); (4) **model evaluation**: involved assessment of the model performance built in the third step using out-of-bag and testing datasets; (5) **model application**: this involved applying the developed model to high-resolution land surface variables to predict SM at 1 km. It is important to note here that the coarse resolution variables such as brightness temperature and precipitation were resampled to fine resolution during prediction of SM, since they were often not available at high spatial resolution.



Figure 2. Flow chart of downscaling procedure.

#### 2.5. Predictors Selection

In the present study, a total of 15 different predictors, including LST, NDVI, surface albedo, elevation, slope, aspect, antecedent precipitation index, brightness temperature, and Sentinel-1 co- and cross-polarized backscatter and textural information (i.e., contrast, entropy, variance and correlation), was used for downscaling of SMAP SM products using RF. The selection of these predictors was based on their strong association with SM variability.

For example, LST is strongly interlinked with SM through thermal inertia and evapotranspiration (ET), among others. Thermal inertia describes the resistance of soil to temperature change and it has positive relation with SM, which implies that wetter soil has higher thermal inertia which in turn leads to decrease in diurnal LST range [63], whereas, in the context of ET, the relationship between LST and SM varies depending on ET regimes: water- or energy-limited. In water-limited environments, increase in SM leads to decrease in LST, whereas in energy-limited environments, their relationship is not significant [64,65]. Similarly, brightness temperature is among the predictors which influence SM variability. It is dependent on emissivity and surface temperature of soil which, in turn, is determined by dielectric property of the soil. Interestingly, dielectric properties of the soil are primarily dependent on SM content, among many other factors [66]. On the other hand, precipitation is the main source of SM and controls its spatial and temporal variations.

Apart from above-mentioned predictors, Sentinel-1 co- and cross-polarized backscatter coefficients, which represent normalized radar return from targeted surface, have a direct positive relationship with SM, most notably over sparse to moderately vegetated areas [67]. It is worth to note that this relationship could be affected by perturbing factors such as vegetation and surface roughness. Similarly, topographic characteristics such as elevation, slope and aspect control the spatial distribution of SM. For example, flat terrain tends to be wetter than sloping hillside and vice versa [68].

Moreover, surface albedo is among the important variables that affect SM variability. It represents the proportion of incoming solar radiation reflected by the earth surface and plays a crucial role in the energy balance of land surface, thereby affecting near-surface climate. It can be affected by many factors, including SM, soil texture, elevation angle and vegetation cover. For example, in the context of SM, wet soils tend to have a darker

color than dry soils, which in turn reduces surface albedo, implying negative relationship between the two variables [69].

NDVI is also one of the predictors strongly related with SM. It reflects the greenness status of vegetation, which in turn depends on the availability of SM. This indicates a positive correlation between NDVI and SM. Similar to NDVI, Sentinel-1 image textural features are used to describe the spatial pattern of pixel values of an images and are often used for land cover classification. Here, as predictors, they were used to provide information about spatial variation of vegetation, which is heavily dependent on SM [70].

Overall, in this study, predictors such as LST, antecedent precipitation index and brightness temperature were used with the aim to preserve temporal dynamics of down-scaled SMAP SM, whereas static topographical information comprising elevation, slope and aspect were included to maintain the spatial patterns of downscaled SMAP SM. In addition, vegetation information, e.g., NDVI and Sentinel-1 textural information were used to be able to elucidate the impact of vegetation on the spatio-temporal patterns of downscaled SM. Sentinel-1 backscatter in VV polarization was included as additional predictor because of its high sensitivity to SM change (notably in sparse to moderate vegetation cover), while VH polarization was considered to account for the effect of vegetation dynamics on SM.

In this study, three experiments were implemented. The first two experiments (i.e., Exp1 and 2) were implemented to evaluate the contribution of predictors derived from Sentinel-1. In Exp1, Sentinel-1 predictors, along with other predictors, were used in the downscaling of SMAP SM products (i.e., L3SMP and L3SMP\_E), but in Exp2 they were removed and only the remaining predictors were used in the downscaling (see Table 2). For comparison purpose, both Exp1 and 2 had the same data length, which was determined by the availability of Sentinel-1 images (i.e., available every 12 days over the study area). On the other hand, Exp3 was proposed, in which the Sentinel-1 predictors were not considered. The exclusion of these predictors (Table 1) and SMAP SM as a response variable (i.e., available every 2 to 3 days). This implied that Exp3 could produces downscaled SM every 2 to 3 days as opposed to Exp1, which produced downscaled SM only every 12 days because of the temporal constraint imposed by the Sentinel-1 revisit time. Overall, it is worth to mention that the downscaling was implemented for spatio-temporally collocated predictors and response variables.

Experiments	Experiments Predictors		Remarks	
Exp1 Exp1 Exp1 Exp1 Exp1 Exp1 Exp1 Exp1		L3SMP L3SMP_E	Length of data was determined by availability of Sentinel-1 images Sentinel-1 variables were removed	
Exp2	LST, NDVI, Albedo, TB <sub>V</sub> , TBh, API, elevation, aspect, slope		from Exp2	
Exp3 LST, NDVI, Albedo, TB <sub>V</sub> , TB <sub>h</sub> , API, elevation, Aspect, Slope		L3SMP L3SMP_E	Relatively large sample size data for training of RF models	

Table 2. Combination of different predictors for different experiments.

\* Sentinel-1 predictors.

#### 3. Results

3.1. Random Forest Model Performance

In this study, two SMAP SM products, namely, the SMAP level-3 (L3SMP, 36 km) and SMAP enhanced (L3SMP\_E, 9 km), were used as background fields to downscale to a 1 km spatial resolution using RF models. A total of 84 images of SMAP SM which were spatio-temporally collocated with another 84 images of each predictor available for summer seasons of 2017 to 2020 was used in the RF model development (Exp1, Table 2). The number of available images was limited due to the low temporal availability of Sentinel-1

images. About two-thirds of these images were used to train RF models, whereas the remaining third was used for testing. Before applying RF models for prediction of SM at 1 km, their performances were evaluated at coarse spatial resolution (i.e., at 36 and 9 km) by constructing the models using the predictors aggregated at these resolutions. Figure 3a,b presents the scatter plots of the predicted SM as a function of observed SMAP SM at 36 and 9 km, respectively, using test data of Exp1. In addition, their respective performance metrics are also displayed in the figure.



**Figure 3.** RF models verification at (**a**) 36 km, (**b**) 9 km and their predictors importance scores at (**c**) 36 km and (**d**) 9 km.

As can be seen in Figure 3a,b, the predicted SM generally varied linearly with the observed SMAP SM; however, with a slight tendency to overestimate and underestimate low (0.2) and high (0.4) SM values, respectively, for both L3SMP and L3SMP\_E SM. In addition, the statistical performance metrics presented in the figure indicated a good agreement between the predicted and observed SMAP SM at both spatial scales. The RF model developed at the 9 km spatial scale performed better with R, ubRMSE and bias values of 0.91, 0.03 m<sup>3</sup>/m<sup>3</sup> and  $-0.001 \text{ m}^3/\text{m}^3$ , respectively, compared to the RF model built at 36 km with R, ubRMSE and bias values of 0.84, 0.04 m<sup>3</sup>/m<sup>3</sup> and  $-0.001 \text{ m}^3/\text{m}^3$ , respectively.

The evaluation of RF models at both spatial scales generally suggests their capability to predict SM at a coarse spatial resolution with acceptable accuracy. Therefore, these models were implemented for the prediction of SM at a 1 km spatial resolution, assuming that the RF model built at coarse spatial resolution was also valid for the prediction of SM at high spatial resolution using high-spatial-resolution predictors. In other words, it assumed that the developed RF models were spatial scale invariant.

#### 3.2. Relative Importance of Predictors

Whenever RF models are employed in the downscaling of geophysical variables, including SM, it is of interest to identify the relative importance of the predictors used. Relative importance is often computed by determining the percentage of the increase in the mean square error (MSE) when a given predictor is randomly permuted, while keeping the remaining predictors unaltered. Generally, higher values indicate the higher importance of a given predictor in enhancing the predictive power of the RF model. Figure 3c,d shows the relative importance of each predictor for both RF models. According to the figure, among the input predictors, the TBv, Albedo, NDVI, API, Tbh and slope were identified as the most important predictors for both models, while the predictors derived from Sentinel-1 had less importance as shown in Figure 3c.

#### 3.3. Validation of Downscaled Soil Moisture with In Situ Observations

Figure 4 shows the time series comparisons of the downscaled L3SMP SM, the most recent 1 km SMAP/Sentinel-1 existing SM product (L2\_SM\_SP) and the original L3SMP SM, along with in situ SM observations and precipitation collected from four selected sites (i.e., Avondale, Ithaca, Geneva and Rockspring) for the summer seasons (May to October) of 2017 to 2020. The downscaled L3SMP SM matched well with the temporal dynamics of the in situ SM across the four sites. Likewise, L2\_SM\_SP and the original L3SMP SM also captured well the temporal variability of in situ SM. Compared to L2\_SM\_SP and L3SMP SM, the downscaled L3SMP SM more closely corresponded with in situ SM observations, yet a slight systematic underestimation or overestimation of extreme SM values of all the products was observed; see, for example, the Geneva site. Moreover, all the three products, including the downscaled L3SMP, L2\_SM\_SP and the original L3SMP SM, responded quite well to the temporal change of rainfall events as can be seen in Figure 4, but in situ SM better correlated with the rainfall events, since both were point measurements located at the same site as opposed to all other products, which give an average SM over a large area (i.e.,  $1 \times 1$  km and  $36 \times 36$  km grid cell).

In a similar way, time series plots of the downscaled L3SMP\_E SM, the original L3SMP\_E SM and L2\_SM\_SP along with in situ SM observations and precipitation for the four in situ SM stations in the Susquehanna watershed are shown in Figure A1. It can be observed that both the downscaled L3SMP\_E and its original counterpart showed a good agreement with in situ SM observations. Comparing Figure A1 to Figure 4, it was found that both the downscaled L3SMP\_E SM had similar temporal behaviors in capturing the wetting and drying patterns of in situ SM in response to rainfall events.

#### 3.4. Quantitative Assessment of Downscaled Soil Moisture

Figure 5 presents the scatter plot of downscaled L3SMP, the original L3SMP and L2\_SM\_SP SM against in situ SM observations at the Avondale, Ithaca, Geneva and Rock-spring sites, whereas Table 3 shows their respective performance metrics. The downscaled L3SMP SM had a slightly higher correlation when compared to in situ SM observations than the original L3SMP SM and L2\_SM\_SP did, with R values in the range of 0.65 (i.e., at the Geneva site) to 0.76 (i.e., at the Ithaca site). Similarly, the ubRMSE values of the downscaled L3SMP SM were somewhat lower than that of L3SMP and L2\_SM\_SP SM for all the four sites, ranging from 0.047 to 0.064 m<sup>3</sup>/m<sup>3</sup>. In terms of bias, on the other hand, the downscaled L3SMP SM had a slightly higher bias particularly for the Ithaca and Geneva stations. Moreover, L2\_SM\_SP had less overall accuracy compared to both downscaled L3SMP and its corresponding original product (Table 3).

Similarly, the scatter plots of the downscaled L3SMP\_E, the original L3SMP\_E SM and L2\_SM\_SP versus in situ SM observations are presented in Figure A2. The downscaled L3SMP\_E SM had a slightly better accuracy with R, ubRMSE and bias values in ranges of 0.64 to 0.80, 0.044 to 0.059 m<sup>3</sup>/m<sup>3</sup> and -0.006 to 0.048 m<sup>3</sup>/m<sup>3</sup>, respectively, compared to its original version with R, ubRMSE and bias values in ranges of 0.59 to 0.69, 0.047 to 0.066 m<sup>3</sup>/m<sup>3</sup> and -0.031 to 0.038 m<sup>3</sup>/m<sup>3</sup>, respectively. On the other hand, L2\_SM\_SP SM had

less overall accuracy compared to both downscaled L3SMP\_E SM and its corresponding original product (see Table 3). As can be observed in Table 3, both downscaled L3SMP and L3SMP\_E had comparable performance metrics with only a slight outperformance for the downscaled L3SMP\_E.



**Figure 4.** Time series of downscaled L3SMP SM, original SMAP SM (L3SMP), L2\_SM\_SP, daily precipitation and in situ SM for (**a**) Avondale, (**b**) Ithaca, (**c**) Geneva and (**d**) Rockspring sites.



**Figure 5.** Scatter plot of the downscaled L3SMP, the original L3SMP and L2\_SM\_SP SM versus in situ SM observation for (**a**) Avondale, (**b**) Ithaca, (**c**) Geneva and (**d**) Rockspring sites.

<b>Fable 3.</b> Performance metrics of downscaled, the original and SMAP/Set	entinel-1 SMAP SM products.
--	-----------------------------

		SMAP L3 SM (L3SMP)		SMAP/Sentinel-1	SMAP enhanced SM (L3SMP_E)		SMAP/Sentinel-1
<b>Stations</b>	Metrics	L3SMP (36 km)	Down_ L3SMP (1 km)	L2_SM_SP (1 km)	L3SMP_E (9 km)	Down_L3SMP_E (1 km)	L2_SM_SP (1 km)
Avondale	R ubRMSE Bias	0.69 0.046 0.031	0.72 0.047 0.013	0.43 0.078 0.012	$0.67 \\ 0.047 \\ -0.031$	$0.73 \\ 0.044 \\ -0.006$	0.43 0.078 0.012
Ithaca	R ubRMSE Bias	0.71 0.056 0.014	0.76 0.054 0.071	0.75 0.055 0.019	0.69 0.056 0.038	0.80 0.05 0.048	0.75 0.055 0.019
Geneva	R ubRMSE Bias	0.57 0.068 0.005	0.65 0.064 0.037	$0.49 \\ 0.073 \\ -0.048$	0.59 0.063 0.0004	0.64 0.059 0.021	$0.49 \\ 0.073 \\ -0.048$
Rocksprings	R ubRMSE Bias	0.61 0.067 0.037	0.75 0.058 0.031	0.66 0.102 0.037	0.63 0.066 0.0123	0.73 0.058 -0.017	0.66 0.102 0.037

# 3.5. Evaluation of Spatial Pattern

A successfully downscaled SM should reproduce the spatial pattern of its corresponding coarse-resolution product, and its evaluation generally relies on a visual comparison of the downscaled SM with its respective native resolution. Figure 6a,b displays the spatial distribution of L3SMP and L3SMP\_E SM, whereas Figure 6c,d shows their corresponding downscaled SM at a 1 km spatial resolution using RF on 17th June 2018. From a visual comparison, it was evident that the downscaled L3SMP and L3SMP\_E SM by the RF model reproduced the spatial patterns of their corresponding native resolutions reasonably well. In addition, they provided spatially more detailed information of SM. The wet and dry regions in the native L3SMP or L3SMP\_E SM maps were clearly portrayed in their corresponding downscaled products and L2\_SM\_SP. As can be seen in the figure, the northern part was characterized by a drier SM, whereas the central–eastern part displayed wetter SM condition. The southern part represented a medium SM range. The L2\_SM\_SP also somewhat reproduced the spatial pattern of both L3SMP and L3SMP\_E SM, as can be observed in Figure 6e, but it was relatively less visually appealing compared to the downscaled L3SMP and L3SMP E SM.



**Figure 6.** Maps of (**a**) the original L3SMP SM, (**b**) the original L3SMP\_E SM, (**c**) downscaled L3SMP SM, (**d**) the downscaled L3SMP\_E SM and (**e**) L2\_SM\_SP on 17th June 2018.

# 3.6. Evaluation of the Value of Sentinel-1 Predictors in Downscaling

The results presented in previous sections were based on the downscaling of L3SMP and L3SMP\_E SM using the Sentinel-1 predictors together with other predictors (i.e., Exp 1, Table 2). Based on a visual inspection and the statistical metrics, the downscaled L3SMP and L3SMP\_E SM reproduced the temporal dynamics of in situ observation and the spatial pattern of their original counterparts reasonably well, as indicated in Figures 3–6.

However, in this section, the contribution of Sentinel-1 predictors in improving the predictive performance of RF models for each spatial scale (i.e., 36 and 9 km) was evaluated. Thus, four RF models were trained and tested separately. The first (RF\_1) and second (RF\_2) RF models were implemented for the downscaling of L3SMP and L3SMP\_E SM, respectively, using Sentinel-1 predictors along with other predictors (i.e., Exp 1, Table 2), while the third (RF\_3) and fourth (RF\_4) RF models were implemented using the same predictors as that of RF\_1 and RF\_2, but without Sentinel-1 predictors for the downscaling of L3SMP and L3SMP\_E SM, respectively (i.e., Exp 2, Table 2). It is also worth to mention that only both spatially and temporally overlapped Sentinel-1 and other predictors were considered in all four models for the purpose of the comparison. The result of this comparison is presented in Figure 7.

Figure 7 shows the scatterplots of predicted and observed SMAP SM for both Exp1 and 2 at both spatial scales for the four RF models mentioned above. The visual inspection clearly showed a good performance of the RF models, but with a slight tendency of overestimation and underestimation of minimum and maximum SM values, respectively, at both spatial scales. Likewise, the models attained higher statistical metrics for both experiments (see Table 4). Accordingly, at the spatial scale of 36 km, the R values of RF\_1 and RF\_3 were 0.84 and 0.83, respectively, while the ubRMSE and bias values were the same for both RF\_1 and RF\_3, which were 0.04 m<sup>3</sup>/m<sup>3</sup> and  $-0.001 \text{ m}^3/\text{m}^3$ , respectively, while the spatial scale of 9 km, the R values of RF\_2 and RF\_4 were 0.91 and 0.86, respectively, while the spatial scale of 9 km, the values of ubRMSE for RF\_2 and RF\_4 were 0.03 m<sup>3</sup>/m<sup>3</sup> and  $0.04 \text{ m}^3/\text{m}^3$ , respectively.



**Figure 7.** Comparison of RF model performances (**a**) with Sentinel-1 predictors at 36 km, (**b**) with Sentinel-1 predictors at 9 km, (**c**) without Sentinel-1 predictors at 36 km and (**d**) without Sentinel-1 predictors at 9 km.

Metrics	SMAP L3 (L	.3SMP, 36 km)	SMAP Enhanced (L3SMP_E, 9 km)		
	With Sentinel-1 Predictors	Without Sentinel-1 Predictors	With Sentinel-1 Predictors	Without Sentinel-1 Predictors	
R	0.84	0.83	0.91	0.86	
ubRMSE	0.04	0.04	0.03	0.04	
Bias	-0.001	-0.001	-0.001	-0.001	

**Table 4.** Performance metrics of RF at spatial scale of 36 and 9 km with and without Sentinel-1 predictors.

The comparison results of Exp1 and 2 showed that leaving out Sentinel-1 predictors slightly decreased the performance of RF models at both spatial scales (Figure 7 and Table 4). For instance, at the spatial scale of 36 km, the R value declined only by 0.01, while at the spatial scale of 9 km, it reduced by 0.05. Apart from R values, the ubRMSE and bias values remained the same for both experiments at 36 km. On the other hand, at the spatial scale of 9 km, the ubRMSE reduced by 0.01, while the bias remained the same. Overall, the models with Sentinel-1 (RF\_1 and RF\_2) slightly outperformed those without Sentinel-1 predictors (RF\_3 and RF\_4) at both spatial scales, but their difference in the overall accuracy was minimal. This indicated that the contribution of the Sentinel-1 predictors was marginal in improving the predictive accuracy RF models for the underlying surface and climatic conditions of our study area.

#### 3.7. Downscaling of SMAP SM Products without Inclusion of Sentinel-1 Predictors

By taking into account the lesser importance of the Sentinel-1 predictors for the condition of our study area as shown in Section 3.6, in the remaining section of the results, we primarily focused on the downscaling of L3SMP and L3SMP\_E SM using predictors of Exp3 (Table 2), namely, the TBv, TBh, Albedo, NDVI, API, slope, LST, DEM and aspect. It is important to note that, although Exp2 and 3 had the same predictors, the length of their data record was different. In the case of Exp3, the availability of data depended on the revisit time of the SMAP SM, which was every 2 to 3 days, whereas for Exp2, the length of the data record was limited by the availability of Sentinel-1 images (i.e., every 12 days) for the purpose of the comparison with Exp1. In other words, the RF models developed for Exp3 had substantially more data for training compared to Exp2, and, therefore, should have resulted in more robust models. In addition to having more data for training the RF models, other advantages of Exp3 relative to Exp2 were the increased temporal availability of the downscaled SM and a better spatial completeness. Therefore, based on this premise, new RF models were trained and tested based on Exp3 at both spatial scales.

Figure 8a,b displays the scatter plots of the observed versus predicted SMAP SM for two separately trained RF model at 36 and 9 km spatial scales, respectively, for the case of Exp3. A total of 325 images of SMAP SM, which were spatio-temporally collocated with other 325 images of each predictor listed in Table 2 for Exp3, was used. About two-thirds of these images were used to train the RF models, whereas the remaining third was used for testing. The RF model at the 36 km spatial scale showed good performance with R, ubRMSE and bias values of 0.91, 0.03 m<sup>3</sup>/m<sup>3</sup> and -0.0002 m<sup>3</sup>/m<sup>3</sup>, respectively (Table 5). Similarly, the RF model at the spatial scale of 9 km had R, ubRMSE and bias values of 0.87, 0.04 m<sup>3</sup>/m<sup>3</sup> and 0.0001 m<sup>3</sup>/m<sup>3</sup>, respectively (Table 5). In addition, Figure 8c,d presents the relative importance score of the predictors. Among the predictors, the TBv, NDVI, Albedo slope and API had the most influential roles at both spatial scales, with a high percentage of increase in MSE during their permutation. Comparing Tables 4 and 5, the performance of RF models of Exp3 increased compared to Exp2. In addition, the bias in extreme SM values was less pronounced in Exp3 compared to Exp1 and 2 (comparing Figures 7 and 8).



**Figure 8.** RF models verification at (**a**) 36 km, (**b**) 9 km and their predictors importance scores at (**c**) 36 km and (**d**) 9 km.

Metrics	L3SMP (36 km)	L3SMP_E (9 km)
R	0.91	0.87
ubRMSE	0.03	0.04
Bias	-0.0002	0.0001

Table 5. Performance of RF models for Exp3 at spatial scale 36 and 9 km.

# 3.7.1. Validation of Downscaled SM with In Situ Observations

To evaluate the temporal behavior of downscaled L3SMP and the original L3SMP SM, a time series comparison was performed with in situ SM and precipitation collected from four sites as shown in Figure 9. The downscaled SM showed a good temporal consistency with in situ SM observations across all sites. It also preserved the temporal dynamics of its corresponding original product (i.e., L3SMP). In addition, the downscaled L3SMP SM and its original counterpart exhibited similar wetting and drying patterns of in situ SM in response to the rainfall events. However, the lower SM values were not sufficiently captured at all sites, and, in a similar way, some underestimations of higher SM values were observed.



**Figure 9.** Time series of downscaled SM, the original SMAP (L3SMP), daily rainfall and in situ SM for (**a**) Avondale, (**b**) Ithaca, (**c**) Geneva and (**d**) Rockspring sites.

Similarly, a time series comparison of the downscaled L3SMP\_E and the original L3SMP\_E against in situ SM observations along with the precipitation was depicted in Figure A3. The downscaled L3SMP\_E SM showed a quite good agreement with the in situ SM observations across all sites. It also responded well to the temporal change of rainfall events. Comparing Figures 9 and A3, it could be clearly seen that both the downscaled L3SMP\_E and their corresponding original products generally followed a similar temporal pattern, and also reacted similarly to the rainfall events.

# 3.7.2. Qualitative Evaluation of Downscaled Soil Moisture

Figure 10 depicts the scatter plots of the downscaled L3SMP and their original counterpart against in situ SM observations collected from the four sites in the study area. The statistical performance metrics computed for each site are also displayed in Table 6. The downscaled L3SMP SM agreed quite well with the in situ measurements, with R values greater than 0.68 for all the sites. However, in terms of the ubRMSE, the Avondale and Ithaca sites had values which were quite closer to the SMAP mission accuracy requirement, which was  $0.04 \text{ m}^3/\text{m}^3$ , as can be observed in Table 6. In comparison to the original L3SMP SM, the downscaled L3SMP SM had a slightly higher R and lower ubRMSE values across

all the sites, while its bias was relatively higher. Moreover, from a visual inspection of the scatter plots, the model slightly overestimated lower SM values, while the higher SM values were somewhat underestimated.



**Figure 10.** Scatter plot of L3SMP and downscaled SM versus in situ SM observation for (**a**) Avondale, (**b**) Ithaca, (**c**) Geneva and (**d**) Rockspring sites.

Analogous to Figure 10, the scatter plots of the downscaled L3SMP\_E and the original L3SMP\_E SM against the in situ SM observations are presented in Figure A4. The downscaled L3SMP\_E SM showed a better performance with R, ubRMSE and bias values in the ranges of 0.74 to 0.85, 0.04 to 0.056 m<sup>3</sup>/m<sup>3</sup> and -0.0002 to 0.019 m<sup>3</sup>/m<sup>3</sup>, respectively, compared to its original counterpart (i.e., L3SMP\_E), with R, ubRMSE and bias values in the ranges of 0.58 to 0.83, 0.066 to 0.043 m<sup>3</sup>/m<sup>3</sup> and -0.0007 to 0.0392 m<sup>3</sup>/m<sup>3</sup>, respectively (Table 6). Among all sites, the downscaled L3SMP\_E and original L3SMP\_E SM performed better at the Avondale and Ithaca sites. The comparison of Figure 10 with Figure A4 indicated that the overall performance metrics of the downscaled L3SMP\_E SM were modestly higher than that of the downscaled L3SMP SM across all the four sites, although it was not significant (see Table 6). In addition, a similar tendency of a slight overestimation and underestimation of lower and higher SM values, respectively, was noticed at both spatial scales.

Stations	Metrics —	SMAP L3 (	L3SMP, 36 km)	SMAP Enhanced (L3SMP_E,9 km)		
		L3SMP	Down_L3SMP	L3SMP_E	Down_L3SMP_E	
	R	0.69	0.75	0.71	0.81	
Avondale	ubRMSE	0.049	0.044	0.048	0.043	
	Bias	-0.009	0.011	-0.02	-0.001	
Ithaca	R	0.76	0.83	0.71	0.82	
	ubRMSE	0.048	0.045	0.054	0.044	
	Bias	-0.001	0.003	0.005	-0.0045	
Geneva	R	0.49	0.68	0.51	0.70	
	ubRMSE	0.071	0.056	0.068	0.053	
	Bias	0.013	0.016	0.006	0.017	
Rocksprings	R	0.64	0.76	0.65	0.71	
	ubRMSE	0.066	0.057	0.066	0.06	
	Bias	0.025	0.042	0.018	0.043	

Table 6. Performance metrics of original SMAP and downscaled SM.

3.7.3. Visual Evaluations of Spatial Pattern of Downscaled Soil Moisture

The visual comparison between the downscaled SM and its corresponding original product allows one to comprehend whether the proposed model is reproducing the spatial pattern of the original product well. Figure 11a,b shows the maps of spatial variations of the downscaled L3SMP and L3SMP\_E SM, while their corresponding original equivalents are presented in Figure 11c,d, respectively. It can be inferred from the figure that the proposed downscaling method considerably improved the spatial details of the original SM, while maintaining their spatial pattern across the entire basin. It can be observed that a lower SM appeared in the northwest, whereas a higher SM occurred in the middle-eastern part of the basin. The southern part of the basin was mainly characterized by medium range SM values.

When comparing Figure 11c,d, similar spatial patterns of the downscaled SM maps were generated from the two background SM fields (i.e., L3SMP and L3SMP\_E). Both of them were characterized by higher and lower SM values in the southern and northwestern parts of the basin, respectively. However, a closer look at the maps of their spatial distribution revealed that the downscaled L3SMP\_E more closely matched its corresponding original version than the downscaled L3SMP SM did with its original counterpart. Finally, compared to the maps of SM produced from exp1 (Figure 6) on 17 June 2018, Exp3 provided a map of SM which a is spatially complete with less overestimations and underestimations of the lower and higher SM values, respectively (Figure 11).



**Figure 11.** Maps of (**a**) the L3SMP, (**b**) L3SMP\_E, (**c**) downscaled L3SMP and (**d**) downscaled L3SMP\_E on 17th June 2018.

# 4. Discussion

This study demonstrated the downscaling of two SMAP SM products, which were L3SMP (36 km) and L3SMP\_E (9 km), to a 1 km spatial resolution using an RF model by considering three different experiments. In the first experiment, the RF models were separately implemented at a resolution of L3SMP (i.e., 36 km) and L3SMP\_E (i.e., 9 km) using predictors derived from Sentinel-1 and MODIS, together with others, such as the brightness temperature, elevation, slope, aspect and antecedent precipitation index (Exp1, Table 2), while in the second experiment, the RF models were employed at both spatial scales by leaving out the Sentinel-1 predictors and focusing only on MODIS products, along with other predictors mentioned above (Exp2, Table 2). It is also important to note here that, for the purpose of the comparison of the two experiments (i.e., Exp1 and 2), only dates when the Sentinel-1 images were available were considered. Finally, the third experiment

(Exp3, Table 2) was proposed based on the limitations of the comparison results between Exp1 and 2 as described in Section 3.7.

## 4.1. RF Model Performance Evaluation

An adequate performance of the RF model at a coarse spatial resolution was required before its use for predicting SM at a high spatial resolution. Thus, in this study, the evaluation of RF models was carried out at two spatial scales: (1) a 36 km resolution for the SMAP L3SMP product and (2) a 9 km resolution for SMAP L3SMP\_E. In both contexts, the results showed the RF models to have a good capability in predicting SM at both spatial scales for both Exp1 and 2. In addition, the comparative evaluation of the performance measures between the RF models of Exp1 and 2 at both spatial scales indicated that the benefit of the Sentinel-1 predictors in enhancing the predictive accuracy of the RF models was minimal (Figure 7 and Table 4). This disagreed with the study conducted by Bai et al. [54], who found that the use of a Sentinel-1 predictor, namely, the backscatter in VV polarization (Sigma VV), markedly increased the predictive accuracy of the RF model. One reason for this inconsistency with results herein could be due to differences in the underlying land surface and climatic conditions of the study area. Their study area was mainly located in an arid area of northern China with low vegetation cover as opposed to the Susquehanna watershed, which is dominated by forests and agriculture, with a humid climate condition. The predominance of vegetation may lower and, in some cases, cancel out the sensitivity of the backscatter signal to SM [71].

As stated earlier, the results of the comparative evaluation indicated the lesser importance of Sentinel-1 predictors in improving the predictive accuracy RF models. For this reason, these predictors were further left out in order to (1) reduce the computational burden associated with the processing of their images, (2) reduce the computational cost associated with the RF model because of the decrease in the dimension of predictors after the removal of the Sentinel-1 predictors and (3) increase the training sample size (as the Sentinel images had a limited temporal and spatial coverage compared to SMAP and MODIS), resulting in more robust RF models. Therefore, the evaluation of the RF model was continued with the remaining predictors, which was denoted as Exp3 (Table 2), as mentioned in Section 3.7. The evaluation of the RF models based on Exp3 at both spatial scales showed higher performances (Figure 7). In addition, the underestimation and overestimation of higher and lower SM values were reduced compared to Exp1 and 2. This could have been due to the large sample size used for the training of the RF model, which resulted in a more robust model and, subsequently, led to better performances.

In general, the performance evaluation results indicated that the RF models developed at both spatial scales for all the experiments showed higher performance measures for the condition of our study area, which was a warm summer humid climate with a predominance of vegetation. This agreed quite well with the results reported by previous studies conducted in different regions of the world under varying climatic and land surface conditions using the RF model, e.g., the subtropical continental monsoon [72], semi-humid temperate monsoon [25], South Asian monsoon [23,62], and Mediterranean climate zone [73].

## 4.2. RF Model Predictors Importance

In this study, predictors derived from Sentinel-1 and MODIS, together with others, such as the brightness temperature, elevation, aspect, slope and antecedent precipitation index were used to downscale L3SMP and L3SMP\_E SM. Many studies on the spatial downscaling of satellite SM products have indicated that the use of numerous predictors with a strong connection to SM generally increased the predictive power of the RF model (e.g., [48,54,74]), which was also confirmed in this study. Nevertheless, the level of the contribution of each predictor in improving the model performance was not the same and varied from place to place and from spatial scale to spatial scale, at which the RF model was trained and verified.

The result of the analysis of the relative importance of predictors from Exp 1, at both spatial scales provided by the RF models, showed that the Sentinel-1 predictors, including Sigma VV, Sigma VH, the correlation, contrast, variance and entropy, had lower values of the percent of increase in MSE when each of them was randomly permuted (Figure 3c,d). This implied that the Sentinel-1 predictors played a less significant role in the predictive accuracy of the RF models over the Susquehanna watershed. This agreed with the study by Attarzadeh et al. [75], which reported a lesser importance of backscatter and its textural features, including the contrast, dissimilarity, energy and entropy, among others, for the retrieval of Sentinel-1 SM over vegetated areas in Kenya, Africa. However, in contrast to our results, Bai at el. [54] found that a Sentinel-1 predictor (i.e., Sigma VV) was the most important predictor among nine predictors (i.e., Sigma VV, LST, NDVI, EVI, NDWI, LAI, ALB, elevation and slope) used for the downscaling of L3SMP\_E SM over a semi-arid area of northern China, with minimal vegetation cover. As previously stated, the main reason for this inconsistency could have been due to a difference in the vegetative cover and climatic conditions of the study area.

On the other hand, among the predictors used, the vertically polarized SMAP brightness temperature (TBv) was identified as the most important predictor in all the experiments. This was because the brightness temperature carried the integrated information of the factors responsible for the variation of SM, such as vegetation opacity, soil temperature, soil texture and surface roughness, among others (e.g., [37,49,76]). This corresponded with the results of Hu et al. [25], who found the brightness temperature in VV polarization to be the most important among the predictors they used in the downscaling of SMAP SM over Inner Mongolia, northern China.

Likewise, the NDVI and surface albedo appeared to exhibit a higher importance next to the brightness temperatures for both RF models developed for L3SMP and L3SMP\_E. This was because our study area was predominated by vegetation, which caused the NDVI to be more sensitive to vegetation dynamics and played an influential role in controlling the evapotranspiration, thereby regulating SM [77]. Similarly, the albedo also plays an important role in downscaling through its ability to control surface energy fluxes, thereby modulating surface SM [77]. In addition, the NDVI and albedo were used in the parameterization of the radiative transfer model (RTM) during SMAP SM retrieval, which, subsequently, impacted the downscaling process. For example, NDVI plays an important role through its control on vegetation optical depth, whereas albedo regulates the surface emissivity in the SMAP SM retrieval algorithm.

In addition, the API was among the most important predictors identified in this study, which was not surprising, because the spatio-temporal variation of SM was strongly influenced by precipitation from which API was calculated. It is worth nothing that, in this study, we considered API as a predictor instead of a precipitation in contrast to the studies by Abbaszadeh et al. [48] and Long et al. [78]. This was because the SM observation on a given day could be influenced by the precipitation of the past few days.

Lastly, satellite remote sensing, such as MODIS, provides a daily LST temperature at a 1km spatial resolution, yet its continuous spatio-temporal availability is highly affected by cloud contamination, which, subsequently, resulted in some missing or abnormal values. Indeed, this constrains its usability for several applications. For example, in the downscaling of satellite SM using ML techniques, such as RF as in this study, the deterioration of LST data quality due cloud cover reduced the training sample size. This was because of the exclusion of LST data, which were affected by the presence of clouds during the training of the RF model. However, in recent years, a number of studies have reported the successful reconstruction of cloud-affected LST pixels which, interestingly, would improve its usability in the downscaling of satellite SM, as well as many other applications [79,80]. In the current study, because of the weak link between SM and land surface temperature in a humid climate dominated by vegetation, such as the Susquehanna watershed, the LST showed less importance. However, other SMAP SM downscaling studies (e.g., [22,74]) in arid and semi-arid climates identified LST as an important variable in the downscaling process, because of the strong correlation between the LST and SM for such climates. Similarly, the DEM, aspect and slope were also identified as relatively less important in the downscaling process of the current study.

# 4.3. Validation of Downscaled Soil Moisture

The results of the validation of the downscaled L3SMP and L3SMP\_E SM showed a range performance for all the experiments depending on the spatial scale of SM used as background fields and the underlying surface and climate conditions of the in situ measurement sites. The performance measure of the downscaled SM indicated a slightly higher ubRMSE than the SMAP accuracy requirement of  $0.04 \text{ m}^3/\text{m}^3$  suggested in [81] across all four sites. This deviation could be partly explained by the disparity in scale between the downscaled SM and in situ SM observation, and partly by the predominance of vegetation in the study area as opposed to the conditions for which SMAP's accuracy of  $0.04 \text{ m}^3/\text{m}^3$  is asserted, which is in less vegetated areas with a vegetation water content (VWC) below 5 kg/m<sup>2</sup> [81]. It is also important to note that, although the in situ sites were specifically established in less vegetated areas such as crops and grasses, they were located in mixed pixels of SMAP SM, with a substantial fraction of forests, which in turn might have affected the validation accuracy.

Compared to Exp1 and 2, Exp3 provided relatively more temporally continuous downscaled SM at both spatial scales. This was because of the high temporal availability (i.e., 1 to 3 days) of predictors used for downscaling, which in turn presented an opportunity to have longer time series during validation with the in situ SM. In addition, a better spatial coverage was obtained when using Exp3 predictors, unless there was an effect of cloud cover and a scanning gap due to the SMAP radiometer. However, for Exp1, the temporal availability of downscaled SM was scarce because of the temporal/spatial constraints imposed by Sentinel-1 data, which might have affected the performance metrics calculation during validation with in situ SM. Besides metrics, the clear visualization of the temporal dynamics of the downscaled SM was problematic because of the discontinuity of time series of downscaled SM for the same reason mentioned above.

The study also compared the downscaled L3SMP and L3SMP\_E SM against the most recent SMAP/ Sentinel-1 (L2\_SM\_SP) 1km SM. Accordingly, the L2\_SM\_SP more or less exhibited a similar temporal variation as the downscaled SM and in situ SM observations. However, the performance metrics were relatively lower for L2\_SM\_SP compared to the downscaled SM (see Table 3). This agreed with the L2\_SM\_SP validation study by Das et al. [46] using an in situ SM observation from core validation sites. They found an inferior performance of L2\_SM\_SP, notably for the sites with crop cover, which were similar to the surface conditions of the Ithaca and Rocksprings sites in this study. This was partly explained by the strong effect of vegetation cover on the backscatter signal of Sentinel-1 SAR, which in turn affected the SM retrieval during the downscaling process.

This study also demonstrated that the RF models tended to provide biased downscaled SM extremes. Such a behavior was consistently exhibited in the models developed at both spatial scales for all the experiments. This was in agreement with previous studies. For instance, Long et al. [78] found an overestimation and underestimation of low and high values, respectively, of the downscaled CCI and CLDAS SM. Similarly, Hutengs and Vohland [60] found a tendency of an over and under estimation of low and high values of the downscaled LST. On the other hand, Zhao et al. [22] found a systematic dry bias in the downscaled SMAP SM. The potential reasons for these discrepancies include (1) an inadequate representation of low and high SM values in the data used for the training of the RF model, (2) the smoothing effect due to the aggregation of predictors used for the training of the RF model at a coarse spatial resolution and (3) the averaging of the predictions of the ensemble of RF trees, which was the main feature of RF to reduce bias and variance, and this in turn smoothed the final predicted SM.

The study also demonstrated that both the downscaled L3SMP and L3SMP\_E SM showed comparable results with respect to the temporal evolution and spatial patterns.

This was because the L3SMP\_E SM was basically derived from the oversampled L3SMP brightness temperature by the Backus–Gilbert interpolation method, indicating that they had similar characteristics. One possible advantage of L3SMP\_E over L3SMP was that L3SMP\_E had a resolution of 9 km, which needed a large number of pixels to cover the entire study area. This, in turn, helped to generate a large training sample size, which is important for RF modelling (and for machine learning models in general). For instance, about 2080 pixels of L3SMP\_E were required to cover the entire study area, while it was only 130 pixels for L3SMP. Thus, the training sample size of L3SMP\_E was about 16 times that of L3SMP, which helped to develop a more robust RF model, which made it more trustworthy for downscaling. In addition, this might also be the reason why the downscaled L3SMP\_E showed a slightly better performance than the downscaled L3SMP.

## 4.4. Advantages and Limitations of This Study

A key advantage of this study was its exploitation of combined optical (i.e., MODIS) and C-band SAR (i.e., Sentinel-1) data in the downscaling of L-band passive microwave SM (i.e., SMAP SM) using the RF model as opposed to most of the previous studies which solely based their research on optical data for the downscaling of satellite SM products. Another advantage was that even though the Sentinel-1-derived predictors did not add substantial information in the downscaling process of the present study because of the predominance of vegetation over the Susquehanna watershed, this study added noteworthy value to the growing interest in the use of SAR data. This is true in light of the increasing availability of SAR sensors with low frequencies, such as the existing C-band SAR, e.g., the Sentiniel-1 and RADARSAT constellation mission (RCM), and the upcoming SAR missions, e.g., the Earth Observing System Synthetic Aperture Radar (EOS SAR) with an S-band radar and the NASA-ISRO Synthetic Aperture Radar (NISAR) mission with an L-band satellite.

In addition, while most of the previous studies focused on the downscaling of either L3SMP SM [22,25,48] or L3SMP\_E SM [54] using the RF model, this study contributed to the limited study on the use of two SMAP SM products as background fields in downscaling, which notably helped to identify the most suitable SMAP SM products for the conditions of our study area. Moreover, this study attempted to discover whether the recent high-resolution SMAP/Sentinel-1 SM product could be potentially suitable compared to the downscaled SM. Furthermore, the downscaled SM had a great potential for local environmental applications such as flood forecasting and irrigation planning.

This study also had some limitations that are noteworthy to mention, which possibly offer the opportunity for further efforts to improve the spatial downscaling of coarse-resolution satellite SM products. First, our study area was predominantly forested, which appeared to affect the quality of the SMAP SM retrieval (i.e., L3SMP and L3SMP\_E) for some of the pixels that covered the study area. Thus, the use of these products as back-ground fields for downscaling may possibly lead to the propagation of errors through the downscaling algorithm. Recently, the NASA SMAP mission started the validation of SMAP SM over temperate forests (i.e., SMAP Validation Experiment 2019–2021 (SMAPVEX19-21)), which might provide the opportunity to revise the SMAP algorithm through a proper parameterization forest canopy in the future [82], which in turn would help in the retrieval of good-quality SMAP SM over vegetated pixels.

Second, the in situ SM measurements were sparse across the study area (i.e., only four stations were available, including Avondale, Ithaca, Geneva and Rockspring), which led to the validation of downscaled SM with a single in situ SM observation point within a given downscaled pixel at those four stations. However, it is known that in situ observations provide only a point measurement of SM in the immediate vicinity of the stations, which is in the order of a few centimeters, whereas the downscaled SM represents the average SM over an area of 1km. This indicated the scale disparity between the in situ SM observation and downscaled SM, implying that a single in situ observation within a downscaled pixel might not be able to sufficiently reflect the downscaled SM, which partly explains the observed systematic bias. Considering the spatial heterogeneity of SM within a pixel, many

satellite SM validation studies recommended to have multiple in situ observations within a pixels, which could then be upscaled during the validation of downscaled SM [83,84], but the present study lacked this.

Third, the spatial aggregation of high-resolution predictors, e.g., NDVI, Albedo, LST, had a smoothing effect on the extreme values, which led the training of the RF models with few extremes, as also stated by Zhao et al. [85]. Similarly, the resampling of coarse-resolution predictors, e.g., brightness temperature and precipitation, also had some smoothing effects on their extreme values, which in turn affected the prediction of the extreme SM values during downscaling. Yet, this was actually not special to our study, since the currently existing downscaling methods rely on the calibration of the downscaling methods at a coarse spatial resolution as the first step [22,25], which caused the aggregation of high-spatial-resolution predictors to be inevitable. In addition, given that some of the predictor variables were not available at a fine resolution, e.g., precipitation, resampling to a fine resolution during the prediction of SM at 1km was also unavoidable.

Fourth, although it was found that the RF is a powerful tool for the handling of the high-dimensional datasets used in this study, it tended to overestimate lower SM values and underestimate higher SM values. Several studies have also reported similar a behavior of RF in the spatial downscaling of SM [78], land surface temperature [60] and precipitation [86]. It seems likely that producing biased extreme values is an inherited behavior of the RF model. This is because the final prediction value is obtained by the averaging of the prediction of each tree, which could reduce the value range due to the smoothing effect of averaging [87,88]. Therefore, post processing, which entails the rescaling of the downscaled SM to in situ SM observations using approaches such as linear regression or cumulative distribution function (CDF) matching, could correct the observed bias of extreme SM values if a in situ measurements network was available.

Finally, despite the fact that the downscaling process presented spatially detailed SM information, their usability was recommended after the correction of bias in extreme values using the aforementioned approaches. The correction of bias in extreme values, notably higher values, is important for hydrological applications such as flood forecasting.

# 5. Conclusions

This study demonstrated the effectiveness of the RF models in the downscaling of two SMAP SM products (i.e., L3SMP (36 km) and L3SMP\_E (9 km)) over the Susquehanna watershed using a suite of predictors derived from different sources, including Sentinel-1 and MODIS. It appeared that the RF model developed at both spatial scales performed quite well for the underlying surface and climate condition of our study area. Among the predictors used in RF, the brightness temperature played the most important role in reinforcing the link between SMAP SM and predictors, followed by the NDVI, albedo and API, whereas the Sentinel-1 predictors only had a marginal impact on the predictive accuracy of the RF model.

The validation results indicated that the time series of the downscaled L3SMP and L3SMP\_E SM closely agreed with in situ SM observations. Similarly, the statistical performance measures indicated that the downscaled SM had a better overall performance than their original counterparts. The results also suggested that the downscaled L3SMP and L3SMP\_E SM reproduced the spatial pattern of their original counterparts sufficiently well, while providing detailed spatial information. Nevertheless, a slightly higher ubRMSE was obtained compared to the retrieval accuracy requirement of the SMAP mission (ubRMSE =  $0.040 \text{ m}^3/\text{m}^3$ ). In addition, the downscaled extreme SM values were slightly biased for both spatial scales.

The downscaled L3SMP and L3SMP\_E SM using Sentinel-1 predictors, along with others (i.e., Exp1, Table 2), only had a low temporal availability because of the long revisit time of Sentinel-1 satellites. In addition, due to the narrow swath width of Sentinel-1 compared to that of SMAP, the downscaled SM only partially covered the study area, thereby reducing its spatial completeness. In contrast, the downscaled L3SMP and L3SMP\_E SM

using other predictors rather than that of Sentinel-1 (i.e., Exp3, Table 2) improved the temporal availability of the downscaled SM and also expanded the spatial completeness of the downscaled SM apart from during the presence of clouds and the SMAP scanning gap.

The potential of L3SMP and L3SMP\_E SM was also explored to identify the best performing product for the condition of the study area. A comparison between the down-scaled L3SMP and L3SMP\_E SM and between their corresponding original version showed comparable results, but with a slightly better performance by L3SMP\_E SM.

In future studies, one could consider the use of recently launched RADARSAT constellation mission (RCM) data, which have the temporal resolution of a few days (i.e., 4 days) compared to Sentinel-1 SAR for the downscaling of satellite SM over varieties of land surfaces and climatic conditions. In addition, the upcoming Earth Observing System Synthetic Aperture Radar (EOS SAR) with an S-band radar and NASA-ISRO Synthetic Aperture Radar (NISAR) mission with an L-band satellite would offer a new opportunity, especially for the better representation of SM in forested regions, since they have the capability to penetrate through the forest canopy in contrast to C-band SAR. Another possible new perspective can be improving the RF algorithm and/or bias correction after downscaling. As noted previously, as long as RF is used for downscaling SM, there is a possibility of predicting biased SM extremes due to its inherent behavior. Thus, improving the algorithm of the RF model in order to accommodate all ranges of SM may be important, otherwise, bias correction before the use of the downscaled SM for subsequent application such as flood forecasting is crucial. Moreover, one could also explore a suite of satellite SM products, e.g., SMAP, SMOS, and AMSR-E (to name a few), as background fields for downscaling in order to identify the best-performing downscaled SM product for the condition of the selected study area. Furthermore, future studies should also consider the potential of cascade downscaling (e.g., from 36 km to 9 km and then from 9 km to 1km) compared to direct downscaling (i.e., directly from 36 km to 1km). Finally, revising the SM retrieval algorithms as well requires further investigation, notably over forested regions, in order to hamper the propagation of poor-quality retrieval to the downscaled SM.

**Author Contributions:** S.A.W. and R.L. contributed to the study conception and design; S.A.W. performed the data analysis; S.A.W. prepared the manuscript (i.e., original draft) and edited; R.L. reviewed and edited the manuscript and supervised the study. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by industrial research partners, including Hydro-Quebec, Brook-field and the city of Sherbrooke.

Data Availability Statement: Not applicable.

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

## Appendix A

In this study, downscaling was implemented to two SMAP SM products (L3SMP and L3SMP\_E). Some of the results of the downscaling of L3SMP\_E were presented here in the appendix.

Validation of downscaled L3SMP\_E based on Exp1.



**Figure A1.** Time series of downscaled L3S L3SMP\_E SM, L3SMP\_E, L2\_SM\_SP, daily rainfall and in situ SM for (**a**) Avondale, (**b**) Ithaca, (**c**) Geneva and (**d**) Rockspring sites.



**Figure A2.** Scatter plot of L3SMP\_E, L2\_SM\_SP and downscaled SM versus in situ soil moisture observation for (**a**) Avondale, (**b**) Ithaca, (**c**) Geneva and (**d**) Rockspring sites.

Validation of downscaled L3SMP\_E based on Exp3.



**Figure A3.** Time series of the downscaled L3SMP\_E SM, L3SMP\_E SM, L2\_SM\_SP, daily rainfall and in situ observation for (**a**) Avondale, (**b**) Ithaca, (**c**) Geneva and (**d**) Rockspring sites.



**Figure A4.** Scatter plot of L3SMP\_E, and downscaled SM versus in situ SM observation for (a) Avondale, (b) Ithaca, (c) Geneva and (d) Rockspring sites.

# References

- 1. Tuttle, S.E.; Salvucci, G.D. A new approach for validating satellite estimates of soil moisture using large-scale precipitation: Comparing AMSR-E products. *Remote Sens. Environ.* **2014**, 142, 207–222. [CrossRef]
- 2. Entekhabi, D. Recent advances in land-atmosphere interaction research. Rev. Geophys. 1995, 33, 995–1003. [CrossRef]
- 3. 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]
- Vereecken, H.; Huisman, J.A.; Bogena, H.; Vanderborght, J.; Vrugt, J.A.; Hopmans, J.W. On the value of soil moisture measurements in vadose zone hydrology: A review. *Water Resour. Res.* 2008, 44, 6829. [CrossRef]
- Lakhankar, T.; Krakauer, N.; Khanbilvardi, R. Applications of microwave remote sensing of soil moisture for agricultural applications. *Int. J. Terrasp. Sci. Eng.* 2009, 2, 81–91.
- 6. Wanders, N.; Karssenberg, D.; De Roo, A.; De Jong, S.M.; Bierkens, M.F.P. The suitability of remotely sensed soil moisture for improving operational flood forecasting. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 2343–2357. [CrossRef]
- 7. Massari, C.; Camici, S.; Ciabatta, L.; Brocca, L. Exploiting satellite-based surface soil moisture for flood forecasting in the Mediterranean area: State update versus rainfall correction. *Remote Sens.* **2018**, *10*, 292. [CrossRef]
- Park, S.; Im, J.; Park, S.; Rhee, J. Drought monitoring using high resolution soil moisture through multi-sensor satellite data fusion over the Korean peninsula. *Agric. For. Meteorol.* 2017, 237–238, 257–269. [CrossRef]

- 9. Velpuri, N.M.; Senay, G.B.; Morisette, J.T. Evaluating New SMAP Soil Moisture for Drought Monitoring in the Rangelands of the US High Plains. *Rangelands* 2015, *38*, 183–190. [CrossRef]
- 10. El Sharif, H.; Wang, J.; Georgakakos, A.P. Modeling regional crop yield and irrigation demand using SMAP type of soil moisture data. *J. Hydrometeorol.* **2015**, *16*, 904–916. [CrossRef]
- Crow, W.T.; Berg, A.A.; Cosh, M.H.; Loew, A.; Mohanty, B.P.; Panciera, R.; De Rosnay, P.; Ryu, D.; Walker, J.P. Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Rev. Geophys.* 2012, 50. [CrossRef]
- 12. Dorigo, W.A.; Wagner, W.; Hohensinn, R.; Hahn, S.; Paulik, C.; Xaver, A.; Gruber, A.; Drusch, M.; Mecklenburg, S.; Van Oevelen, P.; et al. The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 1675–1698. [CrossRef]
- 13. Dobriyal, P.; Qureshi, A.; Badola, R.; Hussain, S.A. A review of the methods available for estimating soil moisture and its implications for water resource management. *J. Hydrol.* **2012**, *458–459*, 110–117. [CrossRef]
- 14. Petropoulos, G.P.; Ireland, G.; Barrett, B. Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Phys. Chem. Earth* **2015**, *83–84*, 36–56.
- 15. Schmugge, T.; O'neill, P.E.; Wang, J.R. Passive Microwave Soil Moisture Research. *IEEE Trans. Geosci. Remote Sens.* **1986**, *GE*-24, 12–22. [CrossRef]
- 16. Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.Ö.; Floury, N.; Brown, M.; et al. GMES Sentinel-1 mission. *Remote Sens. Environ.* **2012**, *120*, 9–24. [CrossRef]
- 17. Njoku, E.G.; Entekhabi, D. Passive microwave remote sensing of soil moisture. J. Hydrol. 1996, 184, 101–129. [CrossRef]
- 18. Jackson, T.J.; Schmugge, T.J. Passive Microwave Remote Sensing System for Soil Moisture: Some Supporting Research. *IEEE Trans. Geosci. Remote Sens.* **1989**, 27, 225–235. [CrossRef]
- 19. Kerr, Y.H.; Waldteufel, P.; Wigneron, J.P.; Martinuzzi, J.M.; Font, J.; Berger, M. Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geosci. Remote Sens.* 2001, *39*, 1729–1735. [CrossRef]
- Narayan, U.; Lakshmi, V.; Jackson, T.J. High-resolution change estimation of soil moisture using L-band radiometer and radar observations made during the SMEX02 experiments. *IEEE Trans. Geosci. Remote Sens.* 2006, 44, 1545–1554. [CrossRef]
- 21. 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]
- Zhao, W.; Sánchez, 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]
- 23. Wei, Z.; Meng, Y.; Zhang, W.; Peng, J.; Meng, L. Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau. *Remote Sens. Environ.* **2019**, 225, 30–44. [CrossRef]
- 24. Fan, S.D.; Hu, Y.M.; Wang, L.; Liu, Z.H.; Shi, Z.; Wu, W.B.; Pan, Y.C.; Wang, G.X.; Zhu, A.X.; Li, B. Improving spatial soil moisture representation through the integration of SMAP and PROBA-V products. *Sustainability* **2018**, *10*, 3459. [CrossRef]
- 25. Hu, F.; Wei, Z.; Zhang, W.; Dorjee, D.; Meng, L. A spatial downscaling method for SMAP soil moisture through visible and shortwave-infrared remote sensing data. *J. Hydrol.* **2020**, *590*, 125360. [CrossRef]
- 26. Rudiger, C.; Su, C.H.; Ryu, D.; Wagner, W. Disaggregation of Low-Resolution L-Band Radiometry Using C-Band Radar Data. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1425–1429. [CrossRef]
- Sánchez-Ruiz, S.; Piles, M.; Sánchez, N.; Martínez-Fernández, J.; Vall-llossera, M.; Camps, A. Combining SMOS with visible and near/shortwave/thermal infrared satellite data for high resolution soil moisture estimates. *J. Hydrol.* 2014, 516, 273–283. [CrossRef]
- 28. Senanayake, I.P.; Yeo, I.Y.; Walker, J.P.; Willgoose, G.R. Estimating catchment scale soil moisture at a high spatial resolution: Integrating remote sensing and machine learning. *Sci. Total Environ.* **2021**, *776*, 145924. [CrossRef]
- Wu, X.; Walker, J.P.; Rüdiger, C.; Panciera, R.; Monerris, A.; Das, N.N. Towards Medium-Resolution Brightness Temperature Retrieval from Active and Passive Microwave. In Proceedings of the 19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011.
- Zhan, X.; Fang, L.; Liu, J.; Hain, C.; Yin, J.; Schull, M.; Cosh, M.; Wen, J.; Lakhankar, T.; Yang, K.; et al. Fusing Microwave and Optical Satellite Observations for High Resolution Soil Moisture Data Products. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Forth Worth, TX, USA, 23–28 July 2017; Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2017.
- 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]
- Sabaghy, S.; Walker, J.P.; Renzullo, L.J.; Jackson, T.J. Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. *Remote Sens. Environ.* 2018, 209, 551–580. [CrossRef]
- 33. Carlson, T. An overview of the "triangle method" for estimating surface evapotranspiration and soil moisture from satellite imagery. *Sensors* **2007**, *7*, 1612–1629. [CrossRef]
- 34. Moran, M.S.; Clarke, T.R.; Inoue, Y.; Vidal, A. Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote Sens. Environ.* **1994**, *49*, 246–263. [CrossRef]
- 35. Peng, J.; Niesel, J.; Loew, A. Evaluation of soil moisture downscaling using a simple thermal-based proxy—The REMEDHUS network (Spain) example. *Hydrol. Earth Syst. Sci.* 2015, *19*, 4765–4782. [CrossRef]

- Chauhan, N.S.; Miller, S.; Ardanuy, P. Spaceborne soil moisture estimation at high resolution: A microwave-optical/IR synergistic approach. Int. J. Remote Sens. 2003, 24, 4599–4622. [CrossRef]
- Choi, M.; Hur, Y. A microwave-optical/infrared disaggregation for improving spatial representation of soil moisture using AMSR-E and MODIS products. *Remote Sens. Environ.* 2012, 124, 259–269. [CrossRef]
- Merlin, O.; Rüdiger, C.; Al Bitar, A.; Richaume, P.; Walker, J.P.; Kerr, Y.H. Disaggregation of SMOS soil moisture in Southeastern Australia. *IEEE Trans. Geosci. Remote Sens.* 2012, 50, 1556–1571. [CrossRef]
- 39. Piles, M.; Camps, A.; Vall-Llossera, M.; Corbella, I.; Panciera, R.; Rudiger, C.; Kerr, Y.H.; Walker, J. Downscaling SMOS-derived soil moisture using MODIS visible/infrared data. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3156–3166. [CrossRef]
- 40. Djamai, N.; Magagi, R.; Goita, K.; Merlin, O.; Kerr, Y.; Walker, A. Disaggregation of SMOS soil moisture over the Canadian Prairies. *Remote Sens. Environ.* 2015, 170, 255–268. [CrossRef]
- Das, N.N.; Entekhabi, D.; Njoku, E.G.; Shi, J.J.C.; Johnson, J.T.; Colliander, A. Tests of the SMAP combined radar and radiometer algorithm using airborne field campaign observations and simulated data. *IEEE Trans. Geosci. Remote Sens.* 2014, 52, 2018–2028. [CrossRef]
- 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]
- Piles, M.; Entekhabi, D.; Camps, A. A change detection algorithm for retrieving high-resolution soil moisture from SMAP radar and radiometer observations. *IEEE Trans. Geosci. Remote Sens.* 2009, 47, 4125–4131. [CrossRef]
- 44. Zhan, X.; Houser, P.R.; Walker, J.P.; Crow, W.T. A method for retrieving high-resolution surface soil moisture from hydros L-band radiometer and radar observations. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1534–1544. [CrossRef]
- 45. Das, N.N.; Entekhabi, D.; Dunbar, R.S.; Njoku, E.G.; Yueh, S.H. Uncertainty Estimates in the SMAP Combined Active-Passive Downscaled Brightness Temperature. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 640–650. [CrossRef]
- Das, N.N.; Entekhabi, D.; Dunbar, R.S.; Chaubell, M.J.; Colliander, A.; Yueh, S.; Jagdhuber, T.; Chen, F.; Crow, W.; O'Neill, P.E.; et al. The SMAP and Copernicus Sentinel 1A/B microwave active-passive high resolution surface soil moisture product. *Remote Sens. Environ.* 2019, 233, 111380. [CrossRef]
- 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]
- Abbaszadeh, P.; Moradkhani, H.; Zhan, X. Downscaling SMAP Radiometer Soil Moisture Over the CONUS Using an Ensemble Learning Method. *Water Resour. Res.* 2019, 55, 324–344. [CrossRef]
- 49. Liu, Y.; Yang, Y.; Jing, W.; Yue, X. Comparison of different machine learning approaches for monthly satellite-based soil moisture downscaling over Northeast China. *Remote Sens.* **2018**, *10*, 31. [CrossRef]
- Lv, A.; Zhang, Z.; Zhu, H. A neural-network based spatial resolution downscaling method for soil moisture: Case study of qinghai province. *Remote Sens.* 2021, 13, 1583. [CrossRef]
- 51. 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]
- 52. Liu, Y.; Jing, W.; Wang, Q.; Xia, X. Generating high-resolution daily soil moisture by using spatial downscaling techniques: A comparison of six machine learning algorithms. *Adv. Water Resour.* **2020**, *141*, 103601. [CrossRef]
- Yin, J.; Zhan, X.; Liu, J.; Moradkhani, H.; Fang, L.; Walker, J.P. Near-real-time one-kilometre Soil Moisture Active Passive soil moisture data product. *Hydrol. Process.* 2020, 34, 4083–4096. [CrossRef]
- 54. Bai, J.; Cui, Q.; Zhang, W.; Meng, L. An approach for downscaling SMAP soil moisture by combining Sentinel-1 SAR and MODIS data. *Remote Sens.* 2019, *11*, 2736. [CrossRef]
- 55. DePhilip, M.; Moberg, T. *Ecosystem Flow Recommendations for the Susquehanna River Basin;* Report to the Susquehanna River Basin Comission and U.S. Army Corps of Engineers; The Nature Conservancy: Harrisburg, PA, USA, 2010; p. 191.
- Justice, C.O.; Townshend, J.R.G.; Vermote, E.F.; Masuoka, E.; Wolfe, R.E.; Saleous, N.; Roy, D.P.; Morisette, J.T. An overview of MODIS Land data processing and product status. *Remote Sens. Environ.* 2002, 83, 3–15. [CrossRef]
- 57. Di Luzio, M.; Johnson, G.L.; Daly, C.; Eischeid, J.K.; Arnold, J.G. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. J. Appl. Meteorol. Climatol. 2008, 47, 475–497. [CrossRef]
- Belgiu, M.; Dragut, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 2016, 114, 24–31. [CrossRef]
- 59. Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
- 60. Hutengs, C.; Vohland, M. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sens. Environ.* **2016**, *178*, 127–141. [CrossRef]
- 61. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- 62. Chen, Q.; Miao, F.; Xu, Z.X.; Wang, H.; Yang, L.; Tang, Z. Downscaling of Remote Sensing Soil Moisture Products Based on TVDI in Complex Terrain Areas. In Proceedings of the 2019 International Conference on Meteorology Observations (ICMO), Chengdu, China, 28–31 December 2019; Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2019.

- 63. Pablos, M.; Gonzalez-Zamora, A.; Sanchez, N.; Martinez-Fernandez, J.; Portal, G.; Vall-Llossera, M. Surface Soil Moisture Estimation from MODIS Apparent Thermal Inertia: A Comparison with SMOS and SMAP Soil Moisture Products. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2018.
- 64. Sun, H.; Zhou, B.; Liu, H. Spatial evaluation of soil moisture (SM), land surface temperature (LST), and LST-Derived SM indexes dynamics during SMAPVEX12. *Sensors* **2019**, *19*, 1247. [CrossRef]
- Hamed Alemohammad, S.; 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]
- 66. Ye, N.; Walker, J.P.; Guerschman, J.; Ryu, D.; Gurney, R.J. Standing water effect on soil moisture retrieval from L-band passive microwave observations. *Remote Sens. Environ.* **2015**, *169*, 232–242. [CrossRef]
- Hajj, M.E.; Baghdadi, N.; Zribi, M.; Bazzi, H. Synergic use of Sentinel-1 and Sentinel-2 images for operational soil moisture mapping at high spatial resolution over agricultural areas. *Remote Sens.* 2017, *9*, 1292. [CrossRef]
- 68. Van Meerveld, H.J.; McDonnell, J.J. On the interrelations between topography, soil depth, soil moisture, transpiration rates and species distribution at the hillslope scale. *Adv. Water Resour.* **2006**, *29*, 293–310. [CrossRef]
- 69. Guan, X.; Huang, J.; Guo, N.; Bi, J.; Wang, G. Variability of soil moisture and its relationship with surface albedo and soil thermal parameters over the Loess Plateau. *Adv. Atmos. Sci.* **2009**, *26*, 692–700. [CrossRef]
- Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* 1983, SMC-3, 610–621. [CrossRef]
- Li, J.; Wang, S.; Gunn, G.; Joosse, P.; Russell, H.A.J. A model for downscaling SMOS soil moisture using Sentinel-1 SAR data. *Int. J. Appl. Earth Obs. Geoinf.* 2018, 72, 109–121. [CrossRef]
- Abowarda, A.S.; Bai, L.; Zhang, C.; Long, D.; Li, X.; Huang, Q.; Sun, Z. Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale. *Remote Sens. Environ.* 2021, 255, 112301. [CrossRef]
- Liu, Y.; Xia, X.; Yao, L.; Jing, W.; Zhou, C.; Huang, W.; Li, Y.; Yang, J. Downscaling Satellite Retrieved Soil Moisture Using Regression Tree-Based Machine Learning Algorithms Over Southwest France. *Earth Space Sci.* 2020, 7, e2020EA001267. [CrossRef]
- 74. Sun, H.; Cui, Y. Evaluating downscaling factors of microwave satellite soil moisture based on machine learning method. *Remote Sens.* **2021**, *13*, 133. [CrossRef]
- Attarzadeh, R.; Amini, J.; Notarnicola, C.; Greifeneder, F. Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at plot scale. *Remote Sens.* 2018, 10, 1285. [CrossRef]
- Neelam, M.; Mohanty, B.P. On the radiative transfer model for soil moisture across space, time and hydro-climates. *Remote Sens.* 2020, 12, 2645. [CrossRef]
- 77. Soliman, A.; Heck, R.J.; Brenning, A.; Brown, R.; Miller, S. Remote sensing of soil moisture in vineyards using airborne and ground-based thermal inertia data. *Remote Sens.* 2013, *5*, 3729–3748. [CrossRef]
- Long, D.; Bai, L.; Yan, L.; Zhang, C.; Yang, W.; Lei, H.; Quan, J.; Meng, X.; Shi, C. Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution. *Remote Sens. Environ.* 2019, 233, 111364. [CrossRef]
- 79. Zhao, W.; Duan, S.B. Reconstruction of daytime land surface temperatures under cloud-covered conditions using integrated MODIS/Terra land products and MSG geostationary satellite data. *Remote Sens. Environ.* **2020**, 247, 111931. [CrossRef]
- Zeng, C.; Long, D.; Shen, H.; Wu, P.; Cui, Y.; Hong, Y. A two-step framework for reconstructing remotely sensed land surface temperatures contaminated by cloud. *ISPRS J. Photogramm. Remote Sens.* 2018, 141, 30–45. [CrossRef]
- 81. Entekhabi, D.; Njoku, E.; O'Neill, P.E.; 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 Mission (SMAP). *Proc. IEEE* **2010**, *98*, 704–716. [CrossRef]
- Colliander, A.; Cosh, M.H.; Kelly, V.R.; Kraatz, S.; Bourgeau-Chavez, L.; Siqueira, P.; Roy, A.; Konings, A.G.; Holtzman, N.; Misra, S.; et al. SMAP Detects Soil Moisture Under Temperate Forest Canopies. *Geophys. Res. Lett.* 2020, 47, e2020GL089697. [CrossRef]
- 83. 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]
- Yee, M.S.; Walker, J.P.; Monerris, A.; Rüdiger, C.; Jackson, T.J. On the identification of representative in situ soil moisture monitoring stations for the validation of SMAP soil moisture products in Australia. *J. Hydrol.* 2016, 537, 367–381. [CrossRef]
- 85. Zhao, W.; Li, A. A downscaling method for improving the spatial resolution of AMSR-E derived soil moisture product based on MSG-SEVIRI data. *Remote Sens.* 2013, *5*, 6790–6811. [CrossRef]
- 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]
- 87. Wolfensberger, D.; Gabella, M.; Boscacci, M.; Germann, U.; Berne, A. RainForest: A random forest algorithm for quantitative precipitation estimation over Swizerland. *Atmos. Meas. Tech. Discuss.* **2020**, *14*, 1–35. [CrossRef]
- 88. Zhang, G.; Lu, Y. Bias-corrected random forests in regression. J. Appl. Stat. 2012, 39, 151–160. [CrossRef]