Modeling and Theoretical Analysis of GNSS-R Soil Moisture Retrieval Based on the Random Forest and Support Vector Machine Learning Approach

remote sensing

MDPI

Article

Modeling and Theoretical Analysis of GNSS-R Soil Moisture Retrieval Based on the Random Forest and Support Vector Machine Learning Approach

Yan Jia [1], Shuanggen Jin [2,3,*], Patrizia Savi [4], Qingyun Yan [2] and Wenmei Li [1]

1 Department of Surveying and Geoinformatics, Nanjing University of Posts and Telecommunications, Nanjing 210046, China; jiayan@njupt.edu.cn (Y.J.); liwm@njupt.edu.cn (W.L.)
2 School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; qy2543@mun.ca
3 Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China
4 Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy; patrizia.savi@polito.it
* Correspondence: sgjin@nuist.edu.cn; Tel.: +86-25-58235371

Received: 19 September 2020; Accepted: 6 November 2020; Published: 10 November 2020

Abstract: Global Navigation Satellite System-Reflectometry (GNSS-R) as a microwave remote sensing technique can retrieve the Earth's surface parameters using the GNSS reflected signal from the surface. These reflected signals convey the surface features and therefore can be utilized to detect certain physical properties of the reflecting surface such as soil moisture content (SMC). Up to now, a serial of electromagnetic models (e.g., bistatic radar and Fresnel equations, etc.) are employed and solved for SMC retrieval. However, due to the uncertainty of the physical characteristics of the sites, complexity, and nonlinearity of the inversion process, etc., it is still challenging to accurately retrieve the soil moisture. The popular machine learning (ML) methods are flexible and able to handle nonlinear problems. It can dig out and model the complex interactions between input and output and ultimately make good predictions. In this paper, two typical ML methods, specifically, random forest (RF) and support vector machine (SVM), are employed for SMC retrieval from GNSS-R data of self-designed experiments (in situ and airborne). A comprehensive simulated dataset involving different types of soil is constructed firstly to represent the complex interactions between the variables (reflectivity, elevation angle, dielectric constant, and SMC) for the requirement of training ML regression models. Correspondingly, the main task of soil moisture retrieval (regression) is addressed. Specifically, the post-processed data (reflectivity and elevation angle) from sensor acquisitions are used to make predictions by these two adopted ML methods and compared with the commonly used GNSS-R retrieval method (electromagnetic models). The results show that the RF outperforms the SVM method, and it is more suitable for handling the inversion problem. Moreover, the RF regression model built by the comprehensive dataset demonstrates satisfactory accuracy and strong universality, especially when the soil type is not uniform or unknown. Furthermore, the typical task of detecting water/soil (classification) is discussed. The ML algorithms demonstrate a high potential and efficiency in SMC retrieval from GNSS-R data.

Keywords: Global Navigation Satellite System-Reflectometry (GNSS-R); soil moisture retrieval; signal-to-noise ratio (SNR); random forest (RF); support vector machine (SVM)

1. Introduction

Soil moisture content (SMC) is an important determinant parameter of surface energy balance and plays an important role in the global water cycle. Existing ground-based experiments and satellite missions dedicated to SMC estimation commonly employ heavy and bulk passive or active sensors,

which limits partly the flexibility and mobility of SMC estimation [1]. With the development of the Global Navigation Satellite System (GNSS) [2], the GNSS reflected signals from objects were received and utilized. Its advantages—low cost, wide global coverage, a large amount of data storage and no need for a special radar transmitter—have made the GNSS-Reflectometry technique predominant. Moreover, it is also a powerful supplement to other traditional measurement methods, thus opening up a new field of research in microwave remote sensing.

The European Space Agency (ESA) proposed that the GPS L-band signal could be used as an ocean scatterometer in 1988. Then, the Passive Reflectometry and Interferometry System (PARIS) concept was first proposed in 1993 [3], using passive reflection and interference technology to carry out GPS L-band ocean remote sensing [4–8]. GNSS-R research work was further extended to the land surface [9–14]. In 2000, the GPS signal was reported for soil moisture retrieval by simulating GNSS-R signals as a function of soil moisture, including the use of tower-based GPS bistatic radar for sensing the seasonal polarization measurements [15]. NASA and the University of Colorado conducted the well-known Soil Moisture Experiment in 2002 (SMEX02) [16,17]. Later, many researchers carried out a large amount of research on GNSS-R soil moisture estimation models and methods [18,19]. Remote sensing laboratory of the Polytechnic University of Catalonia (UPC) utilized the Interference Pattern Technique (IPT) of GNSS direct and reflected signals to quantify the relationship between reflected signals and the SMC [20]. A SMIGOL reflectometer was specifically designed and developed with vertically polarized antennas, which can be used for soil moisture sensing [21]. Additionally, some ground-based and airborne polarimetric experiments were conducted to investigate the sensitivity of polarized components to SMC [22,23]. The soil moisture retrieval using GNSS-R signals was extended from cases of the bare surface [16–19] to the scenarios of ground covered by vegetation. For the latter scenarios, their applications have been expanded to measurements of vegetation height [20,22,24], moisture [25], and biomass [22,23,26].

In 2003, the UK Disaster Monitoring Constellation (UK-DMC) satellite with carrying GNSS-R equipment successfully obtained the data over varied land surfaces and observed the signal power fluctuations concerning different terrains [27]. After that, the TechDemoSat-1 (TDS-1) satellite was launched in 2014 and provided Delay–Doppler Map (DDM) data products, which opened a window for the GNSS-R onboard measurements [28,29] on soil moisture [30,31]. NASA has also launched the Cyclone GNSS (CYGNSS) constellation in December 2016 [32,33]. Some significant results have been found utilizing space-borne data for the soil moisture content (SMC) application [34–39]. It was also reported that the apparent reflectivity, the reflected signal-to-noise-ratio (SNR), and the polarimetric ratio (PR) were correlated with soil moisture well [40].

At present, the SMC retrieval using GNSS-R signals has not been established with an accurate analytical model to reveal the inherent rules and properties of SMC retrieval. Most of the efforts have been focused on quantifying the correlation between SMC and the amplitude of GNSS-R signals, which allow the estimation and monitoring of soil moisture trends. For example, the lately space-borne data-based SMC estimation can achieve accurate results, but it highly relies on the prior knowledge of SMC or the heavy-loaded ancillary data. More work is needed concerning the inversion process to improve GNSS-R SMC retrieval accuracy and usability. It should be noted that the main factors affecting the SMC analytical computation accuracy include the uncertainty physical characteristics of the soil, high complexity, and nonlinearity SMC retrieval process (e.g., electromagnetic and soil dielectric models), etc. Nonetheless, little evidence is available for how to resolve the interactions among these complex factors.

Machine learning (ML) algorithms have been growing in popularity in the applications of remote sensing, since they attempt to construct intrinsically nonlinear relationships between the input and output from data [41–44]. They can serve as a tool to uncover a function, especially when this function is too complicated to be formally expressed. As such, it is hypothesized here that ML methods could be used for the complex GNSS-R retrieval modeling and improving the estimation. Among the machine learning methods, support vector machines (SVMs) became popular in the last few years

for obtaining geo-/bio-physical parameters, such as soil moisture [45], wheat leaf rust [46], and sea ice [47]. SVMs have shown excellent capability in generalization and the resistance to noise with limited data [41]. A bagging ensemble algorithm, random forest (RF), has been widely used in remote sensing applications to obtain the land cover type [48], the boreal forest attributes [49], precipitation [50], vegetation water content [51], and metal concentration [52], since it is good at capturing nonlinear and complex relationships between inputs and predictors with good estimation results [50,51]. These two typical machine learning methods have great potential for interpreting remote sensing data in the fields of land and sea applications, because they are faster and require fewer training samples while exhibiting better prediction performance, compared to other learning methods [46–48,51]. Although SVMs and RF have been used in the past studies for soil moisture estimation, neither of them has been adopted for modeling and comparing with the GNSS-R SMC retrieval models.

Therefore, this study aims to investigate the feasibility of GNSS-R estimation (regression and classification) by using two typical ML algorithms with self-designed experiments (in situ and airborne) and establishes an optimization method for SMC retrieval. A simulated dataset involving different types of soil is constructed for training ML regression models. The performance of the two adopted ML methods and the GNSS-R retrieval method for SMC estimation are evaluated and compared. Additionally, the classifications of water and soils are discussed, and the predicted properties of the surfaces are presented by the classification function. This paper is organized as follows: Section 2 presents the theoretical background of the GNSS-R SMC retrieval and ML algorithms. Section 3 describes the methodology for training and modeling the GNSS-R inversion process. The experimental setup and the employed datasets are detailed in Section 4. Section 5 shows the regression results performed by ML and GNSS-R models with self-designed experimental data as well as some discussions. Finally, conclusions are given in Section 6.

2. Theoretical Background

2.1. Soil Moisture Retrieval Process from Bistatic GNSS-R

The GNSS-R system can be regarded as a bistatic radar system as shown in Figure 1, in which the satellite is the transmitter, and the receiver can be placed near the ground (in situ measurement) or on an aircraft for airborne experiments.

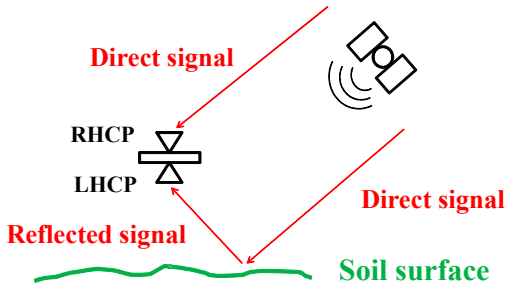

Figure 1. Bistatic GNSS-R receiving configuration.

GNSS-R aims to obtain the characteristics of the reflecting surface by analyzing the reflected signals or their difference from the direct signal. GNSS-R utilizes the L-band microwave signals that are immune to atmospheric attenuation and normally have a good penetration through vegetation [52]. As seen in Figure 1, the RHCP antenna receives the direct signal, and the LHCP antenna receives the reflected signal. The *SNR* peak power of the RHCP antenna is:

$$SNR_{peak}^{direct} = \frac{P^t G^t}{4\pi R_3^2} \frac{G^r \lambda^2 G_D}{4\pi P_N} \quad (1)$$

where P^t represents the satellite transmit power, G^t stands for the satellite gain, G^r and P_N are the antenna gain and noise power for the RHCP and the LHCP link, respectively. G_D is the processing gain due to the de-spread of the GPS C/A code, R_3 denotes the distance between the satellite and the receiver, and λ is the wavelength of the L1 band signal.

In this study, the reflected signal received by the antenna is considered to be dominated by the coherent reflections [16]. Thus, the reflected signal power of the LHCP antenna is:

$$SNR_{peak}^{reflect} = \frac{P^t G^t}{4\pi(R_1 + R_2)^2} \frac{G^r \lambda^2 G_D}{4\pi P_N} \Gamma. \tag{2}$$

In (2), R_1 is the distance between the satellite and the reflection point, and R_2 represents the distance between the reflection point and the receiver. The ratio of SNR_{peak}^{direct} to $SNR_{peak}^{reflect}$ can be written as:

$$\frac{SNR_{peak}^{reflect}}{SNR_{peak}^{direct}} = \frac{R_3^2}{(R_1 + R_2)^2} \Gamma \cdot C \tag{3}$$

where C is a calibration parameter summarizing the uncertainties of G^r and P_N. Γ is the power reflectivity that depends on the surface roughness [53,54]:

$$\Gamma = \left|\rho(\gamma)\right|^2 \chi(z) \tag{4}$$

where $\rho(\gamma)$ represents the Fresnel reflection coefficient of the reflecting surface, and γ denotes the elevation angle of the satellite. $\chi(z)$ is the probability density function for the surface height z. Under the assumption of a flat surface, the $\chi(z) = 1$.

The reflection coefficient $\rho(\gamma)$ is given by a linear combination of vertically and horizontally polarized components; therefore [55]:

$$\rho(\gamma) = \rho_{LR} = \frac{1}{2}\left|\rho_{VV} - \rho_{HH}\right| \tag{5}$$

where ρ_{VV} is the horizontal polarization reflection coefficient and ρ_{HH} is the vertical polarization reflection coefficient. More specifically [5]:

$$\rho_{VV} = \frac{\varepsilon \cdot \sin(\gamma) - \sqrt{\varepsilon - (\cos(\gamma))^2}}{\varepsilon \cdot \sin(\gamma) + \sqrt{\varepsilon - (\cos(\gamma))^2}} \tag{6}$$

$$\rho_{HH} = \frac{\sin(\gamma) - \sqrt{\varepsilon - (\cos(\gamma))^2}}{\sin(\gamma) + \sqrt{\varepsilon - (\cos(\gamma))^2}} \tag{7}$$

where ε is the complex permittivity of the reflecting surface. In the case of dry terrain or almost dry, the imaginary part of the permittivity can be neglected [56,57].

When the LH reflected signal and the RH direct signal are known, the real part of permittivity can be obtained from the combination of (3)–(7) with nearby water calibration [16]. Since the relationship between the dielectric constant of soil and soil moisture is given by the soil dielectric models [53,58], the SMC can be retrieved from the dielectric constant.

2.2. Support Vector Machines

The support vector machine (SVM) was established by Vapnik [59] on the basis of statistical learning theory. It is a typical machine learning algorithm, which was originally used for classification. Assuming the data sample set is denoted as $T = \{(x_i, y_i)\big|i = 1, 2, \ldots, l\}, x_i \in \mathfrak{R}^n, y_i \in \pm 1$, where $x_i \in \mathfrak{R}^n$

is the input vector and its components are features or attributes; $y_i \in \pm 1$ is the output value of corresponding x_i; l is the number of samples. SVM aims to find a classification hyperplane that maximizes the margin between different classes. The hyperplane is constructed as follows [59]:

$$w \cdot x + b = 0 \tag{8}$$

w is a weighting vector, x is an input vector, and b is the bias. A hyperplane that allows two dashed lines $\omega \cdot x + b = 1$ and $\omega \cdot x + b = -1$ to distinguish positive and negative samples was perfectly satisfied, and the maximum value of the distance between them is $\frac{2}{\|\omega\|}$ [59].

The optimization function can be expressed as follows [59]:

$$\begin{cases} min \frac{1}{2}\|w\|^2 \\ y_i(w \cdot x_i + b) \geq 1, \ i = 1, 2, \ldots\ldots, l \end{cases} \tag{9}$$

SVM is quite efficient and requires fewer samples [60]. Especially, SVM features have a kernel function that takes data as input and transforms it into the desired form [59]. These functions can be different types, for example, linear, nonlinear, polynomial, or radial basis function (RBF). Here, we adopted the RBF kernel function, since it has good generalization ability and demonstrated excellent performance [59]. Moreover, SVM is also a typical solution regarding the regression problem, maintaining all the main features that characterize the algorithm (maximal margin), which is known as the support vector regression (SVR). Similar to SVM, SVR can also estimate the nonlinear relationship between input vectors and corresponding predictors [61]. The core of the SVR is the iterative process of the sequential minimal optimization (SMO) algorithm [62].

2.3. Random Forest

Random forest (RF) is an integrated machine learning method proposed by Breiman [63], which uses bagging (bootstrap aggregation) and random split selection techniques to construct multiple decision trees and obtain final classification results by voting. Random forests can also be used for regression. An RF can analyze the complex interaction and even highly correlated variables. It has a fast learning speed and it is quite resistant to noisy data and the data with missing values [46–48,51].

The random forest is an integrated classifier consisting of a set of tree-structured classifiers $\{h(X, \vartheta_k), k = 1, 2, 3 \ldots, K\}$, simplified as $h_i(x)$, where $\{\vartheta_k\}$ is a random vector obeying independent and identical distribution, and K is the number of decision trees in the random forest. Under the given independent variable X, the optimal classification results will be determined by the majority vote from decision trees [63].

Building a random forest requires three steps: generating a training set (bootstrap sampling) for each decision tree, constructing each decision tree, and repeating the above two steps to generate a random forest. In order to construct k trees, we need to generate k random vectors ϑ_1, ϑ_2, ϑ_3 … ϑ_k. These random vectors ϑ_i are independent of each other and are equally distributed. The random vector ϑ_i is used to construct a collection of decision trees $h(x, \vartheta_i)$, and it is simplified as $h_i(x)$. When constructing a tree, a feature is selected from a subset of features and is used to grow each tree [63].

The prediction of the model is the average of the regression results for the k decision trees [63]:

$$H(x) = \frac{1}{k} \sum_{i=1}^{k} h_i(x). \tag{10}$$

When using bootstrap sampling, the unselected data is called out-of-bag (OOB) data. This part of the unselected OOB can be used to estimate the generalization error, classification strength, and correlation coefficient (CC) for the model of the ensembled decision trees; for each decision tree, OOB can be used to obtain an error estimate. The estimates of OOB error for all decision trees

in a random forest are averaged to evaluate the generalization error of the random forest model. More details about the implementation of RF can be found in e.g., [63].

3. Methodology

3.1. RF and SVMs Models for GNSS-R Soil Moisture Retrieval

In general, as demonstrated in Figures 1 and 2, the GNSS-R signals coming from direct and reflected links are received, and the collected raw data were post-processed respectively to obtain the correlation power and relevant navigation messages. Therefore, the soil reflectivity can be obtained by calculating the *SNR* of the received data collected from the reflected and direct signals. After that, as we have introduced in Section 2.1, the soil reflectivity is used to obtain the dielectric constants through the bistatic radar equations. Since the dielectric constants are strongly related to SMC, the relationship between soil dielectric constants and soil moisture is given by the soil dielectric models [53,58].

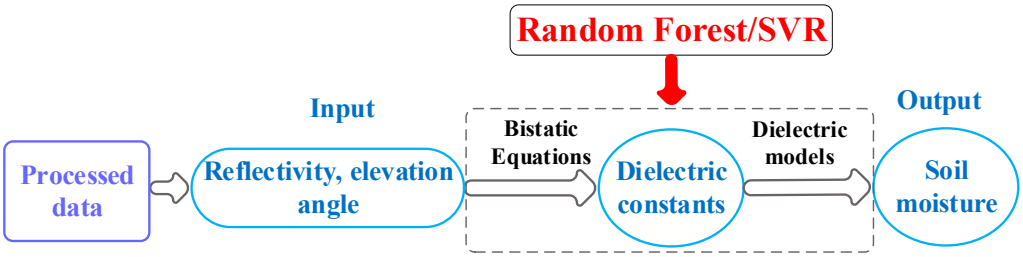

Figure 2. The flowchart of machine learning (ML) applied in Global Navigation Satellite System-Reflectometry (GNSS-R) soil moisture retrieval.

In fact, it has to be noted that the commonly used semi-empirical soil dielectric models [53,58] need the texture information (e.g., clay, sand, and silt proportions) of the soil. As shown in Figure 3, the SMC increases generally with dielectric constants. However, different soil types (identified with n) show an evident impact on SMC retrieval, which increases the difficulty and uncertainty in SMC retrieval when the texture of the soil is unknown or nonuniform. Moreover, operating field measurements for acquiring the soil texture in all test sites are practically impossible; therefore, most GNSS-R SMC measurements are conducted without knowing the information of the test site. On the other hand, the inversion process is quite complex and unable to be solved analytically. Thus, it is difficult to establish an accurate GNSS-R soil moisture model analytically due to the complex interaction of these parameters.

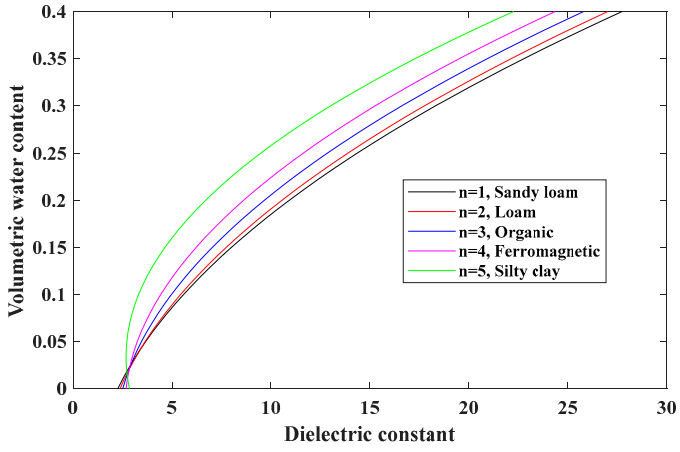

Figure 3. The dielectric constant versus volume of water content for five typical types of soil.

Hence, facing the above-mentioned challenges, here, the GNSS-R SMC retrieval is considered as a nonlinear regression problem and modeled by ML techniques (RF and SVMs), as shown in Figure 2. Input vectors are Γ, γ, and the SMC is the output to be predicted by ML methods. It is worth mentioning that during the GNSS-R experiment, the instability of the receiving equipment or other unexpected situations may cause missing data. ML methods are effective, flexible, and can maintain high accuracy prediction, even when a portion of data is lost [51], which is quite valuable for GNSS-R soil moisture retrieval. In this study, two ML algorithms of RF and SVM were applied for training the regression model and testing the performance of the proposed GNSS-R ML retrieval method.

3.2. Simulated GNSS-R Dataset for Training Regression Models

As noted previously, the regression problem is a typical task solved by ML methods. As such, in this study, we will use SVR and RF models to perform the SMC retrieval (regression) with data collected during self-designed in situ and airborne experiments. In principle, such learning techniques are based on building a regression model between the known SM values from a reference dataset (such as Soil Moisture Active Passive, SMAP, or ground-truth SMC networks) and the experiment observations, and then exploiting this model to perform future SMC estimations. However, as mentioned earlier, constructing an ML model may highly lie on the prior knowledge of SM or the heavy-loaded ancillary data. For particular regions with a self-designed experiment (airborne or in situ measurement), it is extremely difficult to obtain sufficient reference ground-truth data, satisfying numbers of samples for preferable ML models training. Therefore, in this paper, a comprehensive simulation dataset involving five types of soil was built firstly for training ML models. Next, selected real GNSS-R data from airborne and in situ measurements were processed and further tested to validate the prediction performance of the ML models.

The comprehensive simulation dataset was built and used for training and regression tests. This dataset is featured concerning five types of soils that correspond to the dielectric model as mentioned in Figure 3. The input vector consists of Γ (reflectivity) and γ (elevation angle). The output vector is SMC. The simulated dataset is built by the following input vectors concerning different soil types (n):

1. Γ, Reflectivity (from 0–0.8)
2. γ, Elevation angle (from 35 degrees to 85 degrees)

The designed range [55] of the input data for training aimed at covering the range of our acquired measured data. With the simulated input vectors and Equations (4)–(7) of GNSS-R, the SMC including different soil types can be calculated from the dielectric constant by using semi-empirical soil models [58], as illustrated in Section 2.1. Particularly, since the ML methods can build and reveal the nonlinear relationship between the input and output vectors, the regression model composed of different soil types is trained and used, which can increase the prediction accuracy when the soil type is unknown or uncertain. The overall simulated dataset having five different typical soil types is composed of 2000 points (Γ, γ, SMC), as shown in Figure 4.

3.3. Simulated GNSS-R Dataset for Training Classification Models

We further investigate the performance of solving both the classification and regression problems for the airborne data. Hence, the experimental airborne GNSS-R data used for the soil moisture content predictions are also tested for the classification task, and the satellites PRN4 and PRN32 are also considered. Similar to the procedure for our proposed SMC regression scheme, the simulation dataset is devised for training and building the RF and SVM prediction models, since the simulated data can provide sufficient samples and show a more accurate relationship between the input and output. For the classification task, we considered the dielectric constant and elevation angle as the input of the dataset, and the reflectivity (Γ) is the output that can be generated by considering the bistatic Equations (4)–(6) of GNSS-R, under the assumption of a flat surface. Generally, the dielectric

constant of soils does not exceed 25, so the simulated dataset was constructed by varying two input variables in the range:

1. ε, Dielectric constant, soils (from 1 to 25, with a step size of 1), water (78)
2. γ, Elevation angle (from 0 degrees to 90 degrees, with a step size of 3)

With the GNSS-R bistatic equations described in Section 2.1, the reflectivity (Γ) was obtained and the simulated 900 training samples were labeled with −1 (soils) and +1 (water), as presented in Figure 5. The label of soil/water is assigned based on the corresponding value of dielectric constant; specifically, that for water is 78, and for soil, it varies from 1 to 25. The simulated dataset is composed of (Γ, γ, labels).

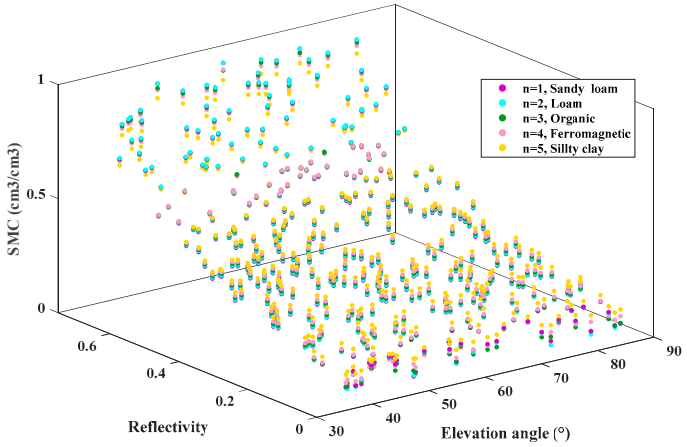

Figure 4. The simulated dataset for regression with different soil types (represented by different colors).

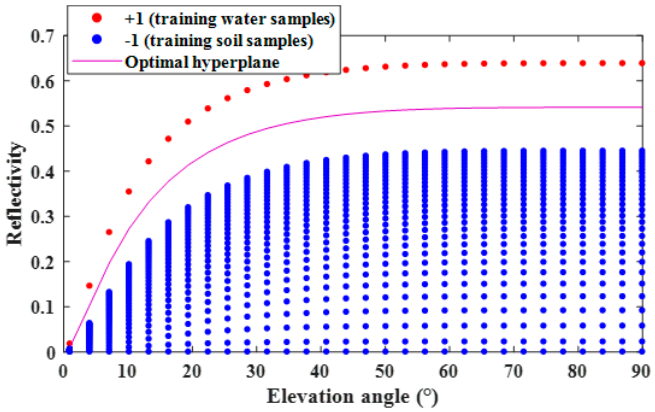

Figure 5. The simulated dataset for classification.

4. Experiments and Data

4.1. Airborne Experimental Data

To validate this work, we firstly consider data obtained from a low-altitude airborne experiment that was carried out by a P92 Digisky airplane over the Avigliana lake (45.099°N, 7.369°E) in Italy on the 11th of December 2014. The flight route and corresponding reflection points for different PRN satellites (PRN4 and PRN32) are shown in Figure 6, including an image of the experimental area from Google Earth.

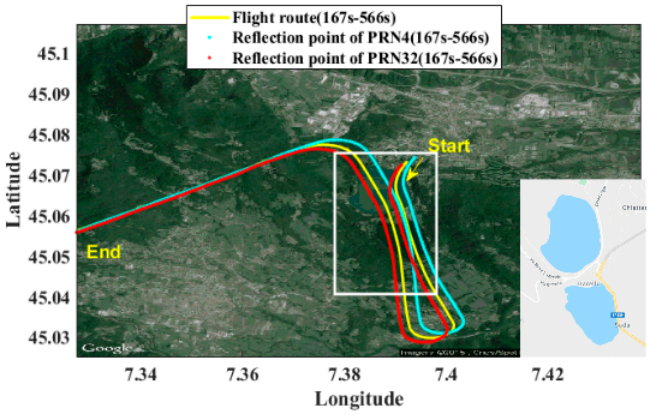

Figure 6. Flight route and corresponding reflection points (PRN4 and PRN32) on Avigliana lakes Piemonte, Italy, on Google Maps [64]. The presence of two lakes within the white box is illustrated on the right-bottom corner of the figure.

This flight experiment was mainly dedicated to investigating soil moisture retrieval from a large area. The type of terrain ranged from open water to terrain with small bushes to built-up areas [64]. It includes two lakes: the size of the northern lake (bigger) is approximately 1 km × 1.3 km, and the southern lake (smaller) is 700 m × 1.1 km. The area was selected for several reasons. First of all, in this area, the presence of two lakes can provide the reflections and the known dielectric constant for calibration. Second, the terrain slope variation can be neglected, and the terrain can be considered smooth [65]. Basically, the reflected signal power is composed of two parts: coherent and non-coherent power. The phase distribution of the coherent part is constant, while in the incoherent part, the phase is random and uniformly distributed over an interval of 2π [66]. If the surface can be considered smooth, the non-coherent component assumes very low values that can be ignored, and the total power received by the antenna can be approximated with the coherent part only [16,65,67].

Data are collected with a receiver working in a bistatic mode, as shown in Figure 1. The up-looking patch antenna is a traditional hemispherical GNSS L1 patch antenna mounted on top of the aircraft fuselage, and the down-looking antenna is a GNSS L1 antenna with LHCP polarization mounted on the bottom fuselage of the aircraft [65]. The antenna was enclosed in a 2-inch square radome (53 mm × 53 mm) and equipped by an Low Noise Amplifier (LNA) to provide 33 dB gain. The GNSS-R receiver [68] is fixed on a small aircraft, as shown in Figure 7 [65].

Figure 7. The receiver prototype [68] equipped on an aircraft.

The prototype used for the acquisition of the received power can measure both the direct and reflected GPS signals through two synchronized channels: one for the direct signal and the other for the reflected signal (see Figure 8). Two antennas are connected with two front ends, respectively. Each front-end is connected to the ODROID-X2 microprocessor board in the prototype, and two data streams are stored in the onboard memory for post-processing [64,65].

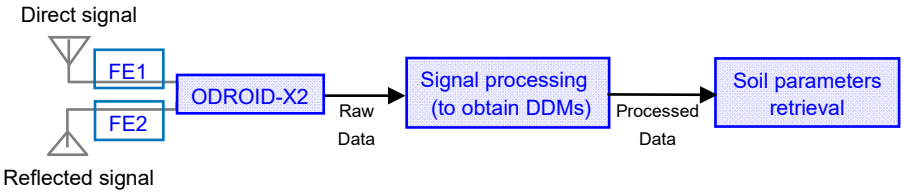

Figure 8. The scheme of the receiver prototype.

As shown in Figure 8, the received raw data are stored in an ODROID-X2 eMMC memory of the receiver prototype in order to be post-processed by an open-loop approach to obtain DDMs and the corresponding delay waveforms. Since a large amount of memory (GB/min) is required for storing the raw data, i.e., 1 s ≈ 1.6 GB data, the duration of the data collection is limited in the embedded Multi Media Card (eMMC) memory (64 GB) and external storage devices. To free more space for data storage, some of the data can be processed on board. Raw data are processed with software SOPRANO [69] and stored as much as hardware capability allowed. Especially, since the reflected GNSS signal is very weak, a combination of coherent and non-coherent integration algorithm was adopted in order to distinguish between the reflection peak and the noise [65]. The coherent integration (also known as signal correlation process) time we used is 1 ms depending on the length of GPS C/A code (1 ms). Several summations or averaging (called non-coherent integration) revealed the real signal shape and eliminated the fading noise effects. Comparing the delay waveforms (DW) performances, including average noise power and standard deviation of the noise, a final 500 ms non-coherent integration time is chosen to meet both the needs of system resolution and reliability to detect real signals [65,68].

4.2. In Situ Experimental Data

In this subsection, data obtained from several in situ measurements are introduced. The in situ data were collected from a serial of ground-based experiments in two bare and smooth sites with different SMC conditions (wet/dry) and terrain compositions. As shown in Figure 9, the first site is located in Grugliasco, Torino (45°03′58.5″N, 7°35′33.8″E), in the Dipartimento Inter-ateneo di Scienze Progetto e Politiche del Territorio (DIST) of Polito. The second site is located in Agliano (44°47′29.1″N, 8°15′19.8″E), which is an area of smooth hills mainly devoted to wine production. The in situ experiment campaign is summarized in Table 1.

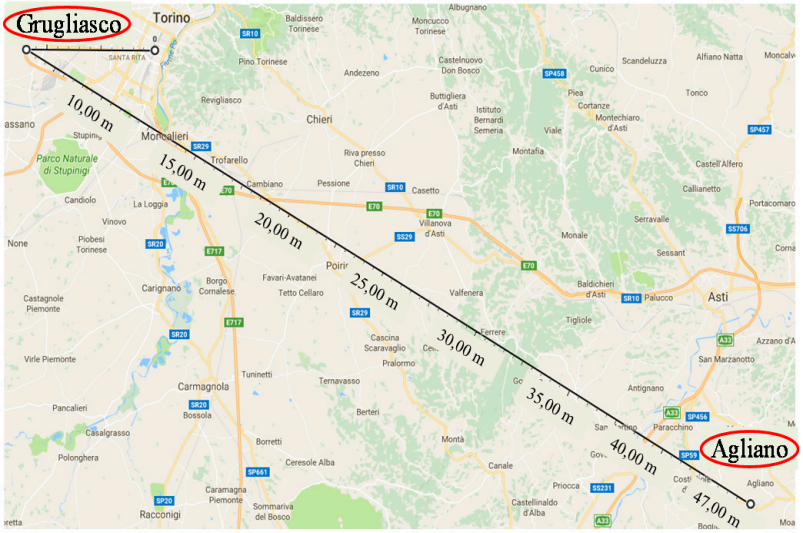

Figure 9. The location of Grugliasco and Agliano on Google Maps.

Table 1. Summary of the experimental campaign.

	Date	Soil Condition	Location
Before rain	27 January 2016	dry	Grugliasco
	4 February 2016	dry	Agliano
After rain	3 March 2016	wet	Grugliasco
	7 March 2016	wet	Agliano

The GNSS-R system used for in situ measurements was performed also in a bistatic GNSS-R configuration, as shown in Figure 1. It consists of two commercial front-ends connected to two antennas and PCs for data acquisition [64]. The raw data processing and calibration procedure were done the same way as for the airborne experiment. Therefore, the reflectivity and corresponding elevation angle can be collected also and would be tested by the proposed ML methods. Moreover, the reference ground-truth SMC was measured and recorded based on the time-domain reflectometry (TDR) technique [70]. A three-rod sensor Tektronix Metallic Cable Tester 1502 manufactured by Tektronix Inc., Beaverton, OR, USA was used in the measurements.

The measurements in dry conditions were done after a long drought, and the wet condition was determined after several rainfalls. The GNSS-R system and ground-truth rod sensor were both used to make measurements before and after rain in bare and smooth fields (Gruliasco/Agliano), as introduced before. The major axis of the first Fresnel zone for satellites in our geometrical condition (high elevation angle and a height of tripod of 1.5 m) is around 1 m. It was estimated for providing the coverage of the GNSS-R data for comparing the results with other kinds of measurements. In this measurement, this information is useful for indicating the location of the instrument probe to precisely evaluate the SMC. In both places (Grugliasco and Agliano), the portable sensor setup moved around in parallel to cover each estimated first Fresnel zones for obtaining the corresponding ground-truth SMC to the GNSS-R system.

5. Results and Analysis

5.1. In Situ Experiments

As we introduced before, the collected ground-based GNSS-R data are processed to obtain the calibrated reflectivity and the elevation angles. Each SNR time series (5 min) is averaged for obtaining the reflectivity. In each site, we obtained twelve groups of GNSS-R measurement data and the corresponding SMC measured by the portable rod sensor. It has to be noted that the measurements are intentionally selected before and after rain in bare and relatively smooth fields (the roughness of Agliano is slightly higher than Grugliasco). Moreover, the data with elevation angles that are smaller than 35 degrees were excluded for good signal reception. The obtained calibrated reflectivity (Γ) and the corresponding elevation angle are shown in Figure 10.

It is shown that in each site, the reflectivity obtained after rain (wet condition) is higher than before rain (dry condition), which corresponds to the theoretical knowledge that the GNSS-R reflectivity increases with SMC [1–3]. The standard deviations (SD) of reflectivity from each site are also shown in Figure 10. It indicates that the SD of reflectivity in Grugliasco is lower than the values obtained in Agliano, which is consistent with the fact that the roughness of Agliano is slightly higher than Grugliasco.

The ground-based GNSS-R data are considered as the testing set to demonstrate and validate the previously established model built by ML algorithms in the preceding section. With the data of reflectivity and elevation angles as an input, in Figure 11, the performance of predictions obtained from RF and SVR models is shown, which is accompanied by the derived GNSS-R SMC on one of the soil types (e.g., $n = 1$) that corresponds to the semi-empirical dielectric model [58] and the measured reference ground-truth SMC.

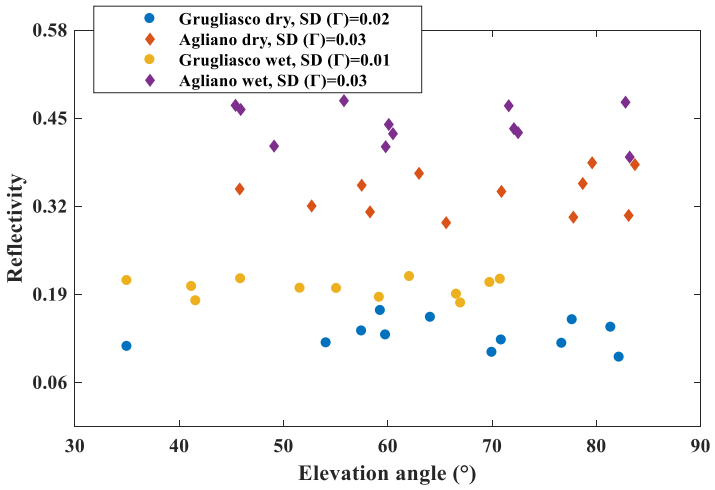

Figure 10. Power reflectivity and elevation angle for in situ measurements.

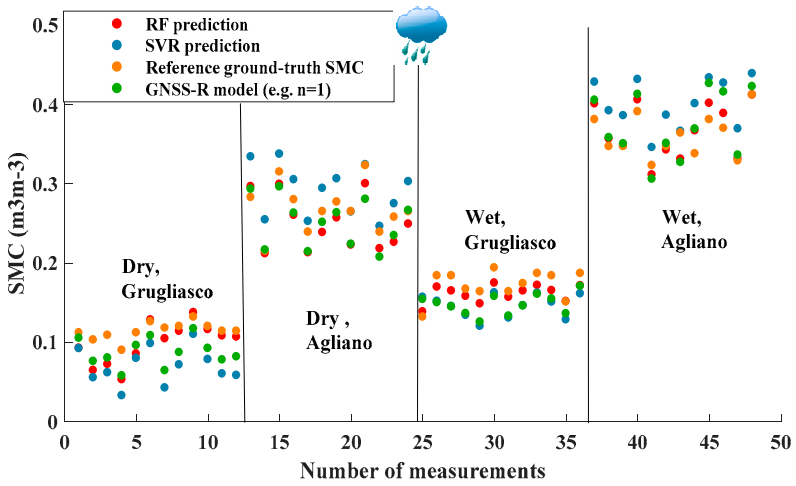

Figure 11. The soil moisture content (SMC) results are obtained from ML and GNSS-R models, compared with ground-truth measurements.

In Figure 11, the overall good estimations can be seen in these four campaigns. The SMC derived from the GNSS-R model (e.g., n = 1) is close to the reference ground-truth SMC. Meanwhile, the prediction results of RF and SVR are also all close to the GNSS-R model and reference ground-truth SMC, which show the good prediction ability of SMC by using ML models.

The results of SMC predictions in each campaign are summarized in Table 2, as well as the SMC obtained by using GNSS-R models under different soil types. Particularly, it demonstrates that the root mean square error (RMSE) obtained is higher in Aliagno than Grugliasco. This phenomenon can be explained by the fact that the GNSS-R models did not take into account the roughness effects; therefore, the higher roughness in Aliano leads to higher RMSE in SMC estimation. Moreover, compared to the two ML models, the SMC obtained from the RF model is much closer to that of the ground-truth and GNSS-R model. RF has a better prediction performance than SVR in GNSS-R SMC estimation, which will be validated also by the airborne experiment in the next subsection.

Compared to the SMC obtained from regression models, RF, SVR, and GNSS-R with different soil types (n), the RF model exhibits the best performance that is the most stable and accurate in all four campaigns. GNSS-R models show some good results, which can be observed only from certain campaigns or soil types (n). It is worth noting that the GNSS-R model relies on knowledge of soil type, while the RF model does not. Hence, when there is no available information about soil type, simply choosing one particular type of soil in the GNSS-R model to predict SMC is not a good

choice. Thus, the RF regression model is quite significant, especially when the soil type is unknown or nonuniform. As such, the flexible, efficient RF model with strong data mining ability becomes more undeniable.

Table 2. The performance matrix of soil moisture prediction by using ML and GNSS-R models with different soil types (n) for in situ measurement.

SMC (m^3/m^3)	Dry, Grugliasco		Dry, Agliano		Wet, Grugliasco		Wet, Agliano	
Ground-Truth	**0.11**		**0.28**		**0.16**		**0.36**	
	mean	**rmse**	**mean**	**rmse**	**mean**	**rmse**	**mean**	**rmse**
RF model	0.10	0.02	0.27	0.03	0.16	0.02	0.39	0.05
SVR model	0.07	0.05	0.32	0.06	0.16	0.03	0.45	0.10
GNSS-R (n = 1)	0.08	0.04	0.25	0.04	0.13	0.03	0.37	0.04
GNSS-R (n = 2)	0.08	0.04	0.26	0.04	0.13	0.03	0.38	0.04
GNSS-R (n = 3)	0.09	0.03	0.27	0.03	0.15	0.02	0.39	0.05
GNSS-R (n = 4)	0.10	0.02	0.29	0.03	0.16	0.01	0.40	0.06
GNSS-R (n = 5)	0.14	0.03	0.31	0.04	0.20	0.04	0.42	0.07
Average (GNSS-R)	0.10	0.02	0.27	0.03	0.15	0.01	0.39	0.05

To compare further the behavior of the regression models, Figure 12 illustrates the scatter plot, which compares the overall predicted and the ground-truth SMC. The results of the GNSS-R model shown in Figure 12 are the averages values (see Table 2) obtained from five soil types of GNSS-R models. From Figure 12, the consistency between predicted data (provided by RF, SVR ML models, and the GNSS-R model, respectively) and ground-truth data is observed.

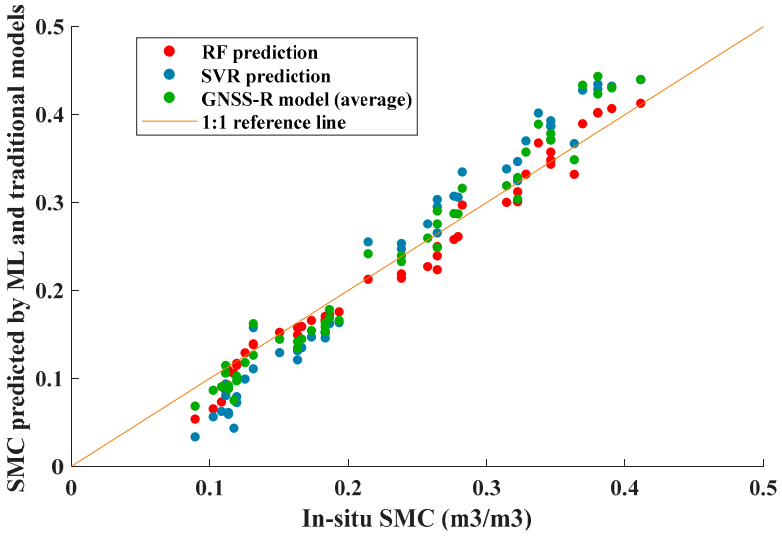

Figure 12. Density plots comparing SMC predicted by random forest (RF), support vector regression (SVR), the average of GNSS-R models, and reference ground-truth SMC measurement with 1:1 reference line.

The performance matrix of SMC predictions acquired by using ML and also the GNSS-R models has been summarized in Table 3. Compared to the two ML models, the performance of RF is better than that of SVR. A correlation coefficient (CC) of $r = 0.92$ and an RMSE of 0.02 m^3/m^3 are obtained for RF. For the SVR algorithm, the correlation coefficient is $r = 0.82$ and the RMSE is 0.04 m^3/m^3. The SMC obtained from the average of the GNSS-R shows a correlation coefficient of $r = 0.80$ and an RMSE of 0.03 m^3/m^3. The RF ML prediction performed best and is slightly even better than the average of the

GNSS-R. The reason could be due to the sufficient training sample and the strong data mining ability of ML, which shows also the high potential of ML predictions in SMC estimations.

Table 3. The performance matrix of predicted SMC by ML and GNSS-R models.

In-Situ Meas.	RMSE (m^3/m^3)	CC
RF	0.02	0.92
SVR	0.04	0.82
GNSS-R (average)	0.03	0.80

5.2. Airborne Experiments

5.2.1. SMC Regression Predictions

In this subsection, the prediction performance of ML and GNSS-R models are also tested and validated by airborne experiments. Those that employed measured airborne GNSS-R data were received along some significant routes (PRN4 and PRN32), in which the elevation angles (see Table 4) were high enough for good signal reception. The specular points corresponding to these satellites fell on the lakes' surfaces, which enables us to calibrate the system.

Table 4. Azimuth, elevation angles for PRN32 and PRN4, at the 1st second of the route (11th December 2014).

PRN Number	Azimuth (°)	Elevation Angle (°)
4	49	76.6
32	222	80.1

Both direct and reflected signals were processed to obtain the signal-to-noise ratio, and a calibration process was performed through the over-water condition to determine the calibration constant c in (3). After obtaining the calibrated reflectivity as shown in Figure 13, the SMC was retrieved using the bistatic GNSS-R method, as described in Section 2.1, by combing (3)–(7) with a soil dielectric model [58]. Additionally, both the RF and SVMs methods were applied for the comparison of soil moisture retrieval.

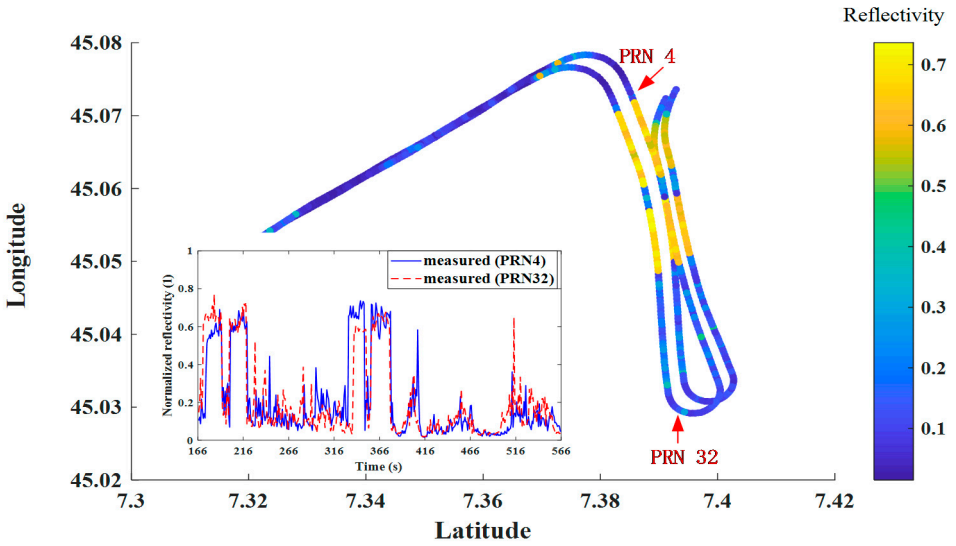

Figure 13. Normalized power reflectivity of PRN32 and PRN4.

After training and testing the proposed SVR- and RF-based regression models with the simulation data, predictions were made by inputting the measured GNSS-R data. The airborne experiment

results retrieved from GNSS-R are represented by the average of the model under different soil types, since the average values have been tested having a preferable result in the previous subsection. Here, the training data were randomly split into two subsets: a training set and a testing set in order to obtain two "unseen" datasets. The training set is a set of samples (1200) used for learning to create a model. The testing set is a set of examples (800) used only to assess the performance of the trained model. The performance of the test set of simulated data is shown in Figure 14,a and the prediction using the tested SVR model for measured data of route PRN32 and PRN4 is shown in Figure 14b.

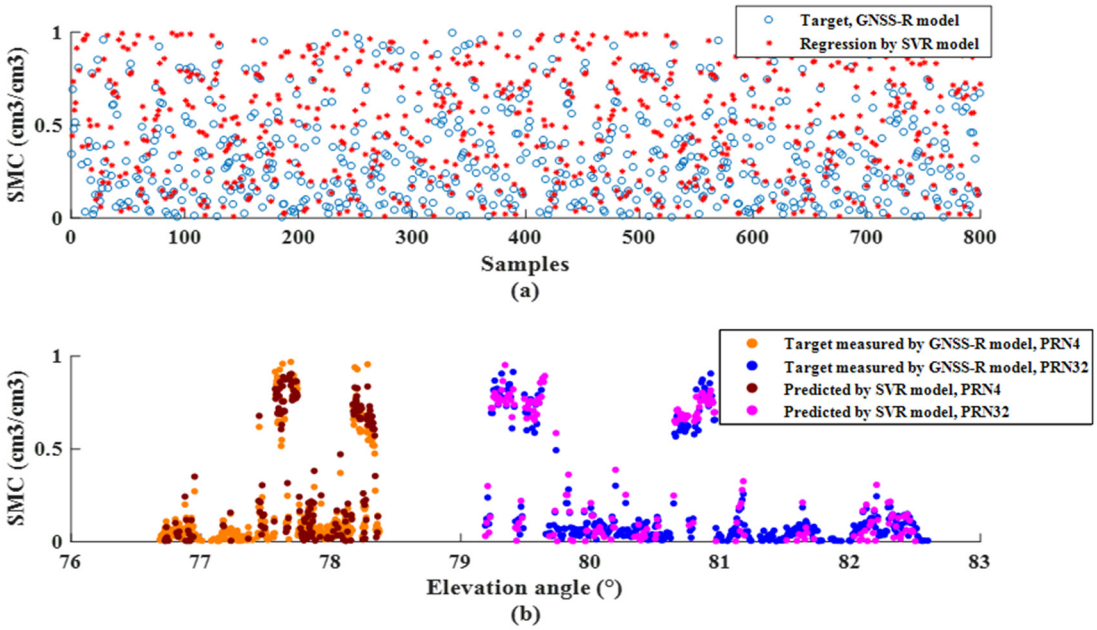

Figure 14. The SVR regression result of test data (**a**) and predictions (**b**) for the route of PRN32 and PRN4.

Figure 14a shows that the SVR regression model could obtain similar results with the target. The performance of regression is also observed in Figure 14b with inputting measured airborne data. The predicted SMC by using the SVR model is highly correlated with the results predicted by the GNSS-R model. In the first and the second periods of time for flying over the lake, the results are better than the others.

The density plot showing the comparison between SM predicted by ML and GNSS-R models for measured data is presented in Figure 15. From Figure 15, good consistency between SM predicted by the SVR model and SM retrieved by GNSS-R can be seen, especially for the densest data. Specifically, a correlation coefficient (CC) of $r = 0.98$ and an RMSE of 0.08 cm^3/cm^3 are obtained for PRN32 and PRN4. A similar performance achieved for both PRNs indicates the generalizability of the proposed method.

As was mentioned before, the RF prediction model was also built after the training and testing steps with the simulation data. Then, the GNSS-R acquisition data from the flight were used to perform the SMC predictions. In Figure 16, the performance of testing (Figure 16a) and the prediction (Figure 16b) of regression using RF for route PRN32 and PRN4 are shown.

Figure 16a shows that the built RF model has enhanced regression ability as compared with the SVR model shown in Figure 14a. The good regression performance can be seen also in the prediction for airborne measured data in Figure 16b. The prediction results are nearly the same as the target predicted by the GNSS-R model.

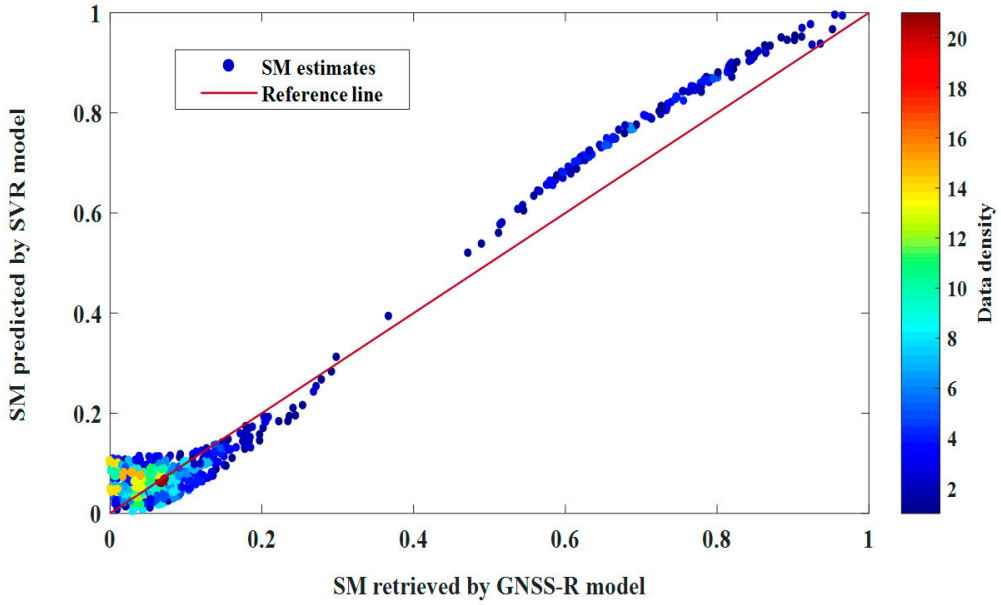

Figure 15. Density plots comparing SMC predicted by the SVR model and SM retrieved by GNSS-R with the 1:1 reference line.

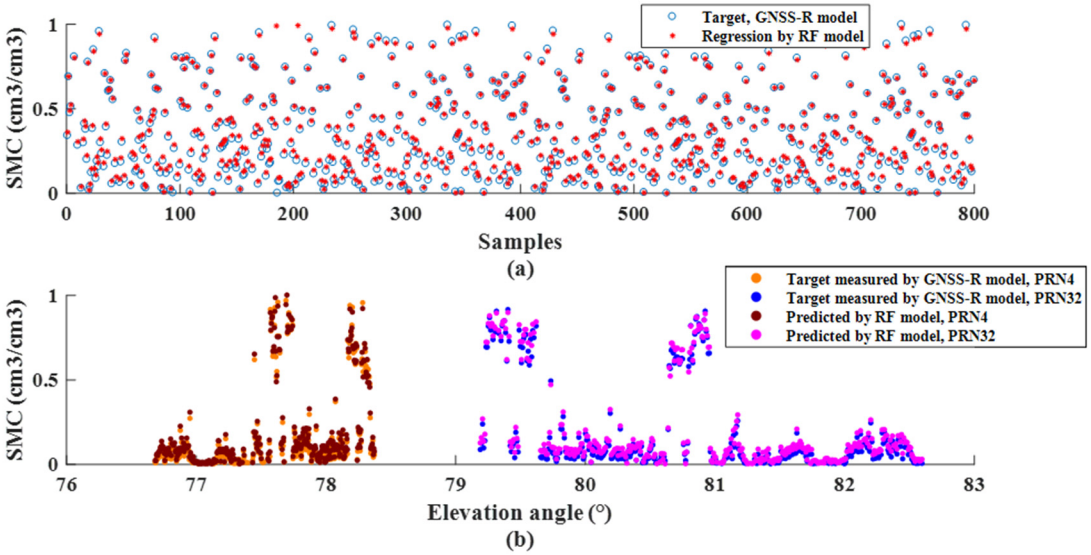

Figure 16. The random forest (RF) regression result of test data (**a**) and predictions (**b**) for the route of PRN32 and PRN4.

The density plot is shown for comparing the predicted SMC by using RF and GNSS-R models as in Figure 17. From Figure 17, good consistency between SM predicted by the RF model and SM estimates by GNSS-R can be seen for the whole dataset. The performance is better than the result obtained from SVR (Figure 15). A correlation coefficient of $r = 0.99$ and an RMSE of 0.02 cm^3/cm^3 are obtained for PRN32 and PRN4. It is observed obviously that the prediction accuracy of RF outperformed SVR and with good generalizability.

The performance matrix of SMC predictions by using RF and SVR with measured PRN4 and PRN32 is summarized in Table 5. We concluded that compared with the SVR algorithm, the prediction performance of RF is better. It is evidenced by its higher correlation coefficient and lower root mean square error, which are also observed in the previous in situ measurement.

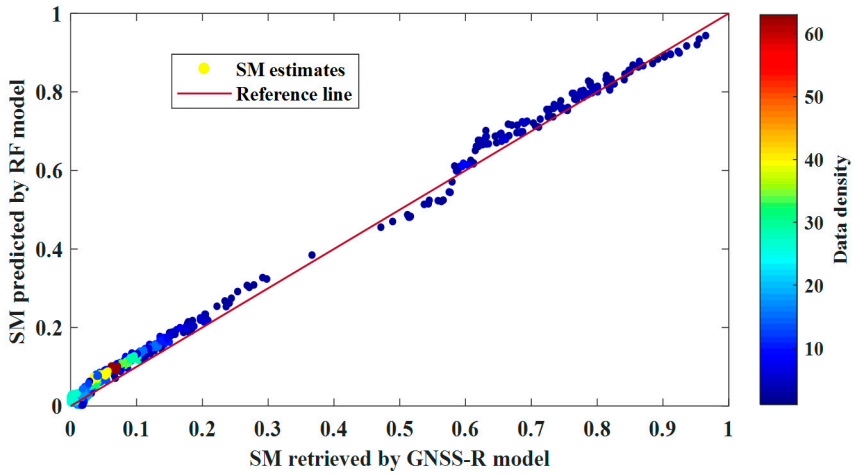

Figure 17. Density plots comparing SM predicted by RF model and SM retrieved by GNSS-R measurement with the 1:1 reference line.

Table 5. The performance matrix of SMC by using RF and SVR for PRN32 and PRN4.

PRN32 and 4	RMSE (cm^3/cm^3)	CC
RF	0.02	0.99
SVR	0.08	0.98

5.2.2. Open Water Classification

The objective of SVM is to find a plane that has the maximum margin to separate the two classes of data points. Many possible hyperplanes could be chosen. With the simulated training set, we built the SVM learning model. In Figure 5, we show the adopted optimal hyperplane (RBF kernel function) that distinctly classifies the data point to achieve the water/soil classification. Then, the processed measured data (Γ, γ) were taken to do the classification. As shown in Figure 18, the results obtained from the data of two satellites (PRN4 and PRN32) are classified into water and soil.

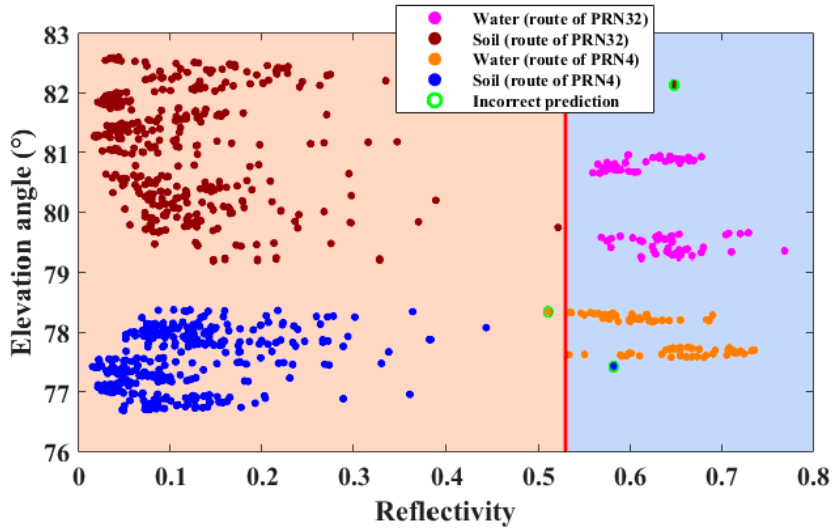

Figure 18. The classification (water/soil) result of support vector machine (SVM) for the route of PRN32 and PRN4.

In this figure, some data points with high reflectivity (oranges and pink points) stand for the presence of lakes in the measurement. Based on the obtained results, the spatial resolution is found to

be about 20 m. In an ideal case, the reflectivity of water should be 0.63. Due to some random factors, e.g., the wave of the water surface, floating plants and microorganisms, etc., the reflectivity is not constant. The measured reflectivity of the water surface ranged between 0.53 and 0.76.

Considering the characteristics of the SVM method, it is anticipated that the trained SVM model would find a hyperplane between the maximum of the soils and the minimal of the water samples. When the elevation angle is around 80° , the maximum of the soil reflectivity is 0.44, and the minimal of the water samples is 0.63 in the training samples, as shown in Figure 5 (here, the reflectivity is an average, and slight variation was made depending on the satellite elevation angle). In Figure 18, the optimal hyperplane (red line) in prediction results shows that the reflectivity higher than 0.54 is judged as water; otherwise, it is considered to be soil. This is consistent with the trained SVM model and the theoretical background of [1–3].

In this case, it can be observed that a majority of data points could be clearly distinguished to be water and soil. Notably, the transitions between soil and water including the soil contents between the two lakes are also distinct, except for three outliers (green circles). The prediction accuracy of both PRN4 and PRN32 is 99.5% and 99.75% respectively, as shown in Figure 18. The support vector machine algorithm can determine the water/soil regions in the figure. Furthermore, the performance of the prediction results is also dependent on the set of training samples. It means that in the training step, the range of the dielectric constant and elevation angles for training samples would be estimated and selected as close as possible to the area of interest, which has similar behavior with the testing samples, in order to train a model with better prediction performance.

The RF algorithm was applied in the simulated dataset to make a comparison with the SVM method. The processed airborne experimental dataset is also used for testing the performance of the classification task. As it has been mentioned in the SVM method, the measured data points and the classification results are shown in Figure 19. Four periods of flight over lakes were distinguished with a spatial resolution of around 20 m. The prediction accuracy of PRN4 and PRN32 is both with 99.75% as illustrated in Figure 19. In this case, the two reflection routes (PRN4 and PRN32) show different classification accuracy, as compared to the 99.5% and 99.75% obtained by applying SVM. The RF shows a similar performance with the SVM algorithm.

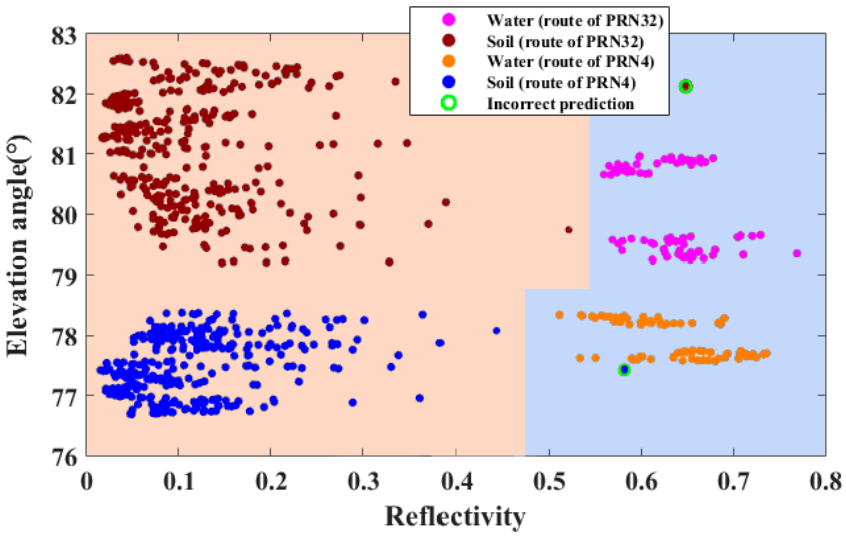

Figure 19. The classification (water/soil) result of RF for the route of PRN32 and PRN4.

6. Discussions

The major focus on GNSS-R soil moisture currently is to build ML models with ongoing knowledge of SMC. However, the comparison of ML models and GNSS-R SMC retrieval using physically-based models is rarely presented. The motivation and the aim of this paper are to build SMC prediction

models using ML, replacing traditional GNSS-R forwarding modeling methods to predict soil moisture from GNSS-R observations, especially in the most of the cases, where the distribution of soil texture is nonuniform or unknown. The study demonstrated that the RF is stable and performs well in all fields with different soil textures. Notably, the ML model does not rely on soil type, while the GNSS-R model does. This distinct advantage is quite useful and significant. The proposed RF model can be used as an alternative to GNSS-R SMC retrieval, which could be applied in various fields and applications in an easy and practical way.

The in situ and airborne GNSS-R experiments are investigated in detail. The technical approaches and the observational data of the experiments are rarely presented in the state of the art, which is regrettable, since field data experiments are very significant and can be a good tool for discovering and studying the inherent GNSS-R problems. Moreover, many researchers are considering assembling their equipment and will be interested in conducting GNSS-R experiments, especially for the airborne platform. In this study, we would like to generalize the finding from the ground to the airborne platform. Despite the lack of the reference ground-truth data for the airborne experiments, the data of input vectors are collected from the real surface and participate in the testing stage in order to show and test the availability of these established ML models and traditional GNSS-R.

In principle, such ML techniques are based on building a regression model between the known SM values from a reference dataset (such as SMAP, or ground-truth SM networks) and the experiment observations, then exploiting this model to perform future SMC estimations. As many samples as possible are needed to achieve the accuracy and stability of a model. In this study, as mentioned earlier, to obtain a batch of reference ground-truth data at every single observational point is almost impossible. So, we built a machine learning algorithm model through the simulation dataset to satisfy the requirement for training the ML models. In future work, it will conduct the proposed ML methods for a larger area with sufficient ground-based reference SMC to generalize the findings (e.g., International Soil Moisture Network or the others).

Another possible future work could be investigating the proposed ML and GNSS-R models with representative soil. The acquisition of knowledge about the site is complex. The GNSS-R model may achieve good results, since the GNSS-R model contains the details of the parameters that better represent the physical components of the site. While, in this case, apart from the accuracy of the GNSS-R, it will give rise to an issue of the significance of existence for building ML models, since the soil composition is already known. Moreover, as we have mentioned, it is not practical, since the soil texture is unknown in most of the GNSS-R experiments, where the ML has demonstrated its efficiency and simplicity in this case.

The distinct advantage of the machine learning algorithms is that they can dig out intrinsically the rules from the dataset. The SM retrieval process itself possesses high complexity and nonlinearity. Here, the ML and traditional SM retrievals are compared. The ML models captured the nonlinear dependencies of the GNSS-R observables (e.g., reflectivity) and the output SMC values directly without intermediate variables. The highly efficient modeling ability and strong data mining capability make it perform well in SM retrieval. Especially for the GNSS-R experiment, the soil texture is commonly not available or nonuniform. The results obtained from this study show the significant advantages of ML methods. The RF model does not rely on knowledge of soil type, while the GNSS-R does. Hence, the RF model could be a very stable and efficient solution employed in different fields even with different scales data of GNSS-R. Moreover, from the perspective of ML algorithms, different ML algorithms are good at handling different data relationships. This paper also shows that the RF has better prediction ability than SVM in solving the SMC estimation problems, which is also one of the significance achievements of our paper.

7. Conclusions

In this study, two ML methods, i.e., SVMs and RF, are adopted for GNSS-R SMC retrieval. Regression results obtained from airborne and in situ data are presented and compared with the

traditional GNSS-R retrieval method. Furthermore, the results obtained from the in situ experiments of two sites using ML models are also validated by the reference ground-truth SMC sensor, respectively. Overall, good predictions are obtained, and the parameters of the performance metrics of applied SVMs and RF with different experiments are analyzed. Particularly, the RF shows the best prediction performance, compared with the SVR model and GNSS-R model under different soil types, which exhibits its high data mining and efficient ability, especially when the soil type is unknown or nonuniform. It is worth noting that the GNSS-R model relies on knowledge of soil type, while the RF model does not. Its good performance with a higher correlation coefficient and a smaller root mean square error is quite noticeable both in the airborne and in situ experiments. In addition, it is also apparent that GNSS-R observations are well suited for open water classification. It is feasible to judge the nature of the reflective surface such as water or soil from the two dependent input variables—reflectivity and elevation angles, which indicates the high potential of ML models.

The study shows the prospects of using ML to represent a complex process that is difficult to model using analytical approaches. The ML methods can help reveal the complex interactions and also make a good prediction, especially since in most of the cases the soil type is unknown or nonuniform. Therefore, regarding the GNSS-R SMC retrieval complexities and challenges, the regression techniques by ML can be practical for the GNSS-R SMC retrieval problem instead of a pure explicit solution of the physical model. This study shows its feasibility by the fact that it can minimize unpredictable influences and help improve the accuracy of soil moisture retrieval. New experiments would be deployed, and the proposed ML techniques will be further validated. Despite a flat surface, validation with SM experiments under a scattering dominated scene is meaningful and will be carried out in the future. They can be used as an alternative to the complex and data-intensive retrieval process and could be applicable in various situations.

Author Contributions: Y.J. proposed the original idea, performed the experiments, and organized the paper. S.J. provided suggestions to improve the whole framework and revised the manuscript. P.S. organized the measurements and revised the manuscript. Q.Y. and W.L. helped with the improvement of the algorithms. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 42001375, 42001362, by the Natural Science Foundation of Jiangsu Province under Grant BK20180765, BK20191384, in part by the Nanjing Technology Innovation Foundation for Selected Overseas Scientists under Grant RK032YZZ18003, 'Research on Advanced Land Surface Detection System using GNSS-R', in part by the Scientific Research Fund of Nanjing University of Posts and Telecommunications (NUPTSF) under Grant 217152, 219066, Shanghai Leading Talent Project (Grant No. E056061) and Strategic Priority Research Program Project of the Chinese Academy of Sciences (Grant No. XDA23040100).

Acknowledgments: The authors would thank the Istituto Superiore Mario Boella (ISMB) for the realization of the prototype, the NASA group of Politecnico di Torino and Digisky s.r.l. (Turin, Italy) for the flight campaigns performed with the Tecnam P92 aircraft during the SMAT-F2 project.

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

References

1. Wigneron, J.P.; Calvet, J.C.; Pellarin, T.; Van De Griend, A.A.; Berger, M.; Ferrazzoli, P. Retrieving near-surface soil moisture from microwave radiometric observations: Current status and future plans. *Remote Sens. Environ.* **2003**, *85*, 489–506. [CrossRef]
2. Jin, S.G.; Su, K. PPP models and performances from single- to quad-frequency BDS observations. *Satell. Navig.* **2020**, *1*, 16. [CrossRef]
3. Martin-Neira, M. A Passive Reflectometry and Interferometry System (PARIS): Application to ocean altimetry. *ESA J.* **1993**, *17*, 331–355.
4. Garrison, J.L.; Katzberg, S.J.; Hill, M.I. Effect of sea roughness on bistatically scattered range coded signals from the Global Positioning System. *Geophys. Res. Lett.* **1998**, *25*, 2257–2260. [CrossRef]
5. Zavorotny, V.U.; Voronovich, A.G. Scattering of GPS signals from the ocean with wind remote sensing application. *IEEE Trans. Geoence Remote Sens.* **2000**, *38*, 951–964. [CrossRef]

6. Lowe, S.; Kroger, P.; Franklin, G.; Labrecque, J.; Lerma, J.; Lough, M.; Marcin, M.; Muellerschoen, R.; Spitzmesser, D.; Young, L. A delay/Doppler-mapping receiver system for GPS-reflection remote sensing. *IEEE Trans. Geoence Remote Sens.* **2002**, *40*, 1150–1163. [CrossRef]
7. Marchan-Hernandez, J.F.; Rodriguez-Alvarez, N.; Camps, A.; Bosch-Lluis, X.; Ramos-Perez, I.; Valencia-Domènech, E. Correction of the Sea State Impact in the L-Band Brightness Temperature by Means of Delay-Doppler Maps of Global Navigation Satellite Signals Reflected Over the Sea Surface. *IEEE Trans. Geoence Remote Sens.* **2008**, *46*, 2914–2923. [CrossRef]
8. Jin, S.G.; Feng, G.; Gleason, S. Remote sensing using GNSS signals: Current status and future directions. *Adv. Space Res.* **2011**, *47*, 1645–1653. [CrossRef]
9. Larson, K.M.; Small, E.E.; Gutmann, E.D.; Bilich, A.L.; Braun, J.J.; Zavorotny, V.U. Use of GPS receivers as a soil moisture network for water cycle studies. *Geophys. Res. Lett.* **2008**, *35*, 24–28. [CrossRef]
10. Jin, S.G.; Qian, X.; Kutoglu, H. Snow Depth Variations Estimated from GPS-Reflectometry: A Case Study in Alaska from L2P SNR Data. *Remote Sens.* **2016**, *8*, 63. [CrossRef]
11. Li, X.; Zheng, K.; Li, X.; Liu, G.; Ge, M.; Wickert, J.; Schuh, H. Real-time capturing of seismic waveforms using high-rate BDS, GPS and GLONASS observations: The 2017 Mw 6.5 Jiuzhaigou earthquake in China. *GPS Solut.* **2019**, *23*, 17. [CrossRef]
12. Edokossi, K.; Calabia, A.; Jin, S.; Molina, I. GNSS in Reflectometry and Remote Sensing of Soil Moisture: A Review of Measurement Techniques, Methods, and Applications. *Remote Sens.* **2020**, *12*, 614. [CrossRef]
13. Camps, A.; Park, H.; Castellví, J.; Corbera, J.; Ascaso, E. Single-Pass Soil Moisture Retrievals Using GNSS-R: Lessons Learned. *Remote Sens.* **2020**, *12*, 2064. [CrossRef]
14. Wei, C.C.; Hsu, C.C. Extreme Gradient Boosting Model for Rain Retrieval using Radar Reflectivity from Various Elevation Angles. *Remote Sens.* **2020**, *12*, 2203. [CrossRef]
15. Masters, D.; Zavorotny, V.; Katzberg, S.; Emery, W. GPS signal scattering from land for moisture content determination. In Proceedings of the IEEE International Geoscience & Remote Sensing Symposium, Honolulu, HI, USA, 24–28 July 2000.
16. Masters, D.; Axelrad, P.; Katzberg, S. Initial results of land-reflected GPS bistatic radar measurements in SMEX02. *Remote Sens. Environ.* **2004**, *92*, 507–520. [CrossRef]
17. Katzberg, S.J.; Torres, O.; Grant, M.S.; Masters, D. Utilizing calibrated GPS reflected signals to estimate soil reflectivity and dielectric constant: Results from SMEX02. *Remote Sens. Environ.* **2006**, *100*, 17–28. [CrossRef]
18. Zavorotny, V.U.; Larson, K.M.; Braun, J.J.; Small, E.E.; Gutmann, E.D.; Bilich, A. A Physical Model for GPS Multipath Caused by Land Reflections: Toward Bare Soil Moisture Retrievals. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *3*, 100–110. [CrossRef]
19. Larson, K.M.; Braun, J.J.; Small, E.E.; Zavorotny, V.U.; Gutmann, E.D.; Bilich, A. GPS Multipath and Its Relation to Near-Surface Soil Moisture Content. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 91–99. [CrossRef]
20. Rodriguez-Alvarez, N.; Camps, A.; Vall-Llossera, M.; Bosch-Lluis, X.; Monerris, A.; Ramos-Perez, I.; Valencia, E.; Marchan-Hernandez, J.F.; Martinez-Fernandez, J.; Baroncini-Turricchia, G.; et al. Land Geophysical Parameters Retrieval Using the Interference Pattern GNSS-R Technique. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 71–84. [CrossRef]
21. Rodriguez-Alvarez, N.; Bosch-Lluis, X.; Camps, A.; Vall-Llossera, M.; Valencia, E.; Marchan-Hernandez, J.F.; Ramos-Perez, I. Soil moisture retrieval using GNSS-R techniques: Experimental results over a bare soil field. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3616–3624. [CrossRef]
22. Egido, A.; Caparrini, M.; Ruffini, G.; Paloscia, S.; Santi, E.; Guerriero, L.; Pierdicca, N.; Floury, N. Global Navigation Satellite Systems Reflectometry as a Remote Sensing Tool for Agriculture. *Remote Sens.* **2012**, *4*, 2356–2372. [CrossRef]
23. Carreno-Luengo, H.; Amèzaga, A.; Vidal, D.; Olivé, R.; Munoz, J.F.; Camps, A. First polarimetric GNSS-R measurements from a stratospheric flight over boreal forests. *Remote Sens.* **2015**, *7*, 13120–13138. [CrossRef]
24. Small, E.E.; Larson, K.M.; Braun, J.J. Sensing vegetation growth with reflected GPS signals. *Geophys. Res. Lett.* **2010**, *37*. [CrossRef]

25. Rodriguez-Alvarez, N.; Bosch-Lluis, X.; Camps, A.; Ramos-Perez, I.; Valencia, E.; Park, H.; Vall-Llossera, M. Vegetation water content estimation using GNSS measurements. *IEEE Geosci. Remote Sens. Lett.* **2011**, *9*, 282–286. [CrossRef]
26. Ferrazzoli, P.; Guerriero, L.; Pierdicca, N.; Rahmoune, R. Forest biomass monitoring with GNSS-R: Theoretical simulations. *Adv. Space Res.* **2011**, *47*, 1823–1832. [CrossRef]
27. Gleason, S. Detecting bistatically reflected GPS signals from low earth orbit over land surfaces. In Proceedings of the 2006 IEEE International Symposium on Geoscience and Remote Sensing, Denver, CO, USA, 31 July–4 August 2006; pp. 3086–3089.
28. Li, W.; Cardellach, E.; Fabra, F.; Rius, A.; Ribó, S.; Martín-Neira, M. First spaceborne phase altimetry over sea ice using TechDemoSat-1 GNSS-R signals. *Geophys. Res. Lett.* **2017**, *44*, 8369–8376. [CrossRef]
29. Foti, G.; Gommenginger, C.; Jales, P.; Unwin, M.; Shaw, A.; Robertson, C.; Rosello, J. Spaceborne GNSS reflectometry for ocean winds: First results from the UK TechDemoSat-1 mission. *Geophys. Res. Lett.* **2015**, *42*, 5435–5441. [CrossRef]
30. Chew, C.; Shah, R.; Zuffada, C.; Hajj, G.; Masters, D.; Mannucci, A.J. Demonstrating soil moisture remote sensing with observations from the UK TechDemoSat-1 satellite mission. *Geophys. Res. Lett.* **2016**, *43*, 3317–3324. [CrossRef]
31. Camps, A.; Park, H.; Portal, G.; Rossato, L. Sensitivity of TDS-1 GNSS-R reflectivity to soil moisture: Global and regional differences and impact of different spatial scales. *Remote Sens.* **2018**, *10*, 1856. [CrossRef]
32. Carreno-Luengo, H.; Luzi, G.; Crosetto, M. Sensitivity of CyGNSS bistatic reflectivity and SMAP microwave radiometry brightness temperature to geophysical parameters over land surfaces. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *12*, 107–122. [CrossRef]
33. Ruf, C.S.; Gleason, S.; McKague, D.S. Assessment of CYGNSS wind speed retrieval uncertainty. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *12*, 87–97. [CrossRef]
34. Chew, C.C.; Small, E.E. Soil moisture sensing using spaceborne GNSS reflections: Comparison of CYGNSS reflectivity to SMAP soil moisture. *Geophys. Res. Lett.* **2018**, *45*, 4049–4057. [CrossRef]
35. Clarizia, M.P.; Pierdicca, N.; Costantini, F.; Floury, N. Analysis of CYGNSS data for soil moisture retrieval. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2227–2235. [CrossRef]
36. Yan, Q.; Huang, W.; Jin, S.; Jia, Y. Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data. *Remote Sens. Environ.* **2020**, *247*, 111944. [CrossRef]
37. Al-Khaldi, M.M.; Johnson, J.T.; O'Brien, A.J.; Balenzano, A.; Mattia, F. Time-series retrieval of soil moisture using CYGNSS. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4322–4331. [CrossRef]
38. Calabia, A.; Molina, I.; Jin, S. Soil Moisture Content from GNSS Reflectometry Using Dielectric Permittivity from Fresnel Reflection Coefficients. *Remote Sens.* **2020**, *12*, 122. [CrossRef]
39. Wu, X.; Dong, Z.; Jin, S.; He, Y.; Song, Y.; Ma, W.; Yang, L. First Measurement of Soil Freeze/Thaw Cycles in the Tibetan Plateau Using CYGNSS GNSS-R Data. *Remote Sens.* **2020**, *12*, 2361. [CrossRef]
40. Zribi, M.; Motte, E.; Baghdadi, N.; Baup, F.; Dayau, S.; Fanise, P.; Guyon, D.; Huc, M.; Wigneron, J.P. Potential Applications of GNSS-R Observations over Agricultural Areas: Results from the GLORI Airborne Campaign. *Remote Sens.* **2018**, *10*, 1245. [CrossRef]
41. Ali, I.; Greifeneder, F.; Stamenkovic, J.; Neumann, M.; Notarnicola, C. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sens.* **2015**, *7*, 16398–16421. [CrossRef]
42. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
43. Wang, J.; Yuan, Q.; Shen, H.; Liu, T.; Li, T.; Yue, L.; Shi, X.; Zhang, L. Estimating snow depth by combining satellite data and ground-based observations over Alaska: A deep learning approach. *J. Hydrol.* **2020**, *585*, 124828. [CrossRef]
44. Jia, Y.; Jin, S.; Savi, P.; Gao, Y.; Tang, J.; Chen, Y.; Li, W. GNSS-R soil moisture retrieval based on a XGboost machine learning aided method: Performance and validation. *Remote Sens.* **2019**, *11*, 1655. [CrossRef]
45. Ahmad, S.; Kalra, A.; Stephen, H. Estimating soil moisture using remote sensing data: A machine learning approach. *Adv. Water Resour.* **2010**, *33*, 69–80. [CrossRef]

46. Ashourloo, D.; Aghighi, H.; Matkan, A.A.; Mobasheri, M.R.; Rad, A.M. An investigation into machine learning regression techniques for the leaf rust disease detection using hyperspectral measurement. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4344–4351. [CrossRef]
47. Yan, Q.; Huang, W. Detecting sea ice from TechDemoSat-1 data using support vector machines with feature selection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 1409–1416. [CrossRef]
48. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300. [CrossRef]
49. Liu, J.; Hyyppä, J.; Yu, X.; Jaakkola, A.; Kukko, A.; Kaartinen, H.; Zhu, L.; Liang, X.; Wang, Y.; Hyyppä, H. A novel GNSS technique for predicting boreal forest attributes at low cost. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4855–4867. [CrossRef]
50. Min, M.; Bai, C.; Guo, J.; Sun, F.; Liu, C.; Wang, F.; Xu, H.; Tang, S.; Li, B.; Di, D.; et al. Estimating summertime precipitation from Himawari-8 and global forecast system based on machine learning. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 2557–2570. [CrossRef]
51. Yuan, Q.; Li, S.; Yue, L.; Li, T.; Shen, H.; Zhang, L. Monitoring the Variation of Vegetation Water Content with Machine Learning Methods: Point–Surface Fusion of MODIS Products and GNSS-IR Observations. *Remote Sens.* **2019**, *11*, 1440. [CrossRef]
52. Tan, K.; Ma, W.; Wu, F.; Du, Q. Random forest–based estimation of heavy metal concentration in agricultural soils with hyperspectral sensor data. *Environ. Monit. Assess.* **2019**, *191*, 446. [CrossRef]
53. Wang, J.R.; Schmugge, T.J. An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Trans. Geosci. Remote Sens.* **1980**, 288–295. [CrossRef]
54. Beckmann, P.; Spizzichino, A. *The Scattering of Electromagnetic Waves from Rough Surfaces*; Artech House: Norwood, MA, USA, 1987.
55. Stutzman, W.L. *Polarization in Electromagnetic Systems*; Artech House: Norwood, MA, USA, 2018.
56. Behari, J. *Microwave Dielectric Behaviour of Wet Soils*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2006.
57. Hong, S.; Shin, I. A physically-based inversion algorithm for retrieving soil moisture in passive microwave remote sensing. *J. Hydrol.* **2011**, *405*, 24–30. [CrossRef]
58. Hallikainen, M.T.; Ulaby, F.T.; Dobson, M.C.; El-Rayes, M.A.; Wu, L.K. Microwave dielectric behavior of wet soil-part 1: Empirical models and experimental observations. *IEEE Trans. Geosci. Remote Sens.* **1985**, 25–34. [CrossRef]
59. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
60. Moser, G.; Serpico, S.B. Automatic parameter optimization for support vector regression for land and sea surface temperature estimation from remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 909–921. [CrossRef]
61. Okujeni, A.; Van der Linden, S.; Jakimow, B.; Rabe, A.; Verrelst, J.; Hostert, P. A comparison of advanced regression algorithms for quantifying urban land cover. *Remote Sens.* **2014**, *6*, 6324–6346. [CrossRef]
62. Chang, C.C.; Lin, C.J. Training v-support vector regression: Theory and algorithms. *Neural Comput.* **2002**, *14*, 1959–1977. [CrossRef]
63. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
64. Pei, Y.; Notarpietro, R.; De Mattia, S.; Savi, P.; Dovis, F.; Pini, M. Remote sensing of soil based on a compact and fully software GNSS-R receiver. In Proceedings of the 26th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS + 2013), Nashville, TN, USA, 16–20 September 2013; pp. 56–61.
65. Pei, Y.; Notarpietro, R.; Savi, P.; Cucca, M.; Dovis, F. A fully software Global Navigation Satellite System reflectometry (GNSS-R) receiver for soil monitoring. *Int. J. Remote Sens.* **2014**, *35*, 2378–2391. [CrossRef]
66. Ulaby, F.T. *Microwave Remote Sensing Active and Passive-Volume III: From Theory to Applications*; Artech House: London, UK, 1986.
67. Jia, Y.; Savi, P. Sensing soil moisture and vegetation using GNSS-R polarimetric measurement. *Adv. Space Res.* **2017**, *59*, 858–869. [CrossRef]

68. Istituto SuperioreMario Boella (ISMB), Torino, Italy. Available online: http://www.ismb.it/ (accessed on 9 November 2020).
69. Falletti, E.; Margaria, D.; Nicola, M.; Povero, G.; Gamba, M.T. N-FUELS and SOPRANO: Educational tools for simulation, analysis and processing of satellite navigation signals. In Proceedings of the 2013 IEEE Frontiers in Education Conference (FIE), Oklahoma City, OK, USA, 23–26 October 2013; pp. 303–308.
70. Savi, P.; Maio, I.A.; Ferraris, S. The role of probe attenuation in the time-domain reflectometry characterization of dielectrics. *Electromagnetics* **2010**, *30*, 554–564. [CrossRef]