



## Article

# A GNSS-IR Method for Retrieving Soil Moisture Content from Integrated Multi-Satellite Data That Accounts for the Impact of Vegetation Moisture Content

Jichao Lv <sup>1</sup>, Rui Zhang <sup>1,2,\*</sup>, Jinsheng Tu <sup>3</sup>, Mingjie Liao <sup>1</sup>, Jiatai Pang <sup>1</sup>, Bin Yu <sup>1</sup>, Kui Li <sup>1</sup>, Wei Xiang <sup>1</sup>, Yin Fu <sup>1</sup> and Guoxiang Liu <sup>1</sup>

- <sup>1</sup> Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China; lvjichao@my.swjtu.edu.cn (J.L.); lmj8407@my.swjtu.edu.cn (M.L.); Pjt\_Shawn@my.swjtu.edu.cn (J.P.); ybdouble@my.swjtu.edu.cn (B.Y.); likui@my.swjtu.edu.cn (K.L.); xiangwei@my.swjtu.edu.cn (W.X.); rsyinfu@my.swjtu.edu.cn (Y.F.); rsgxliu@swjtu.cn (G.L.)
- <sup>2</sup> State-Province Joint Engineering Laboratory of Spatial Information Technology for High-Speed Railway Safety, Southwest Jiaotong University, Chengdu 611756, China
- <sup>3</sup> College of Geographic Information and Tourism, Chuzhou University, Chuzhou 239099, China; tujinsheng@chzu.edu.cn
- \* Correspondence: zhangrui@swjtu.edu.cn

**Abstract:** There are two problems with using global navigation satellite system-interferometric reflectometry (GNSS-IR) to retrieve the soil moisture content (SMC) from single-satellite data: the difference between the reflection regions, and the difficulty in circumventing the impact of seasonal vegetation growth on reflected microwave signals. This study presents a multivariate adaptive regression spline (MARS) SMC retrieval model based on integrated multi-satellite data on the impact of the vegetation moisture content (VMC). The normalized microwave reflection index (NMRI) calculated with the multipath effect is mapped to the normalized difference vegetation index (NDVI) to estimate and eliminate the impact of VMC. A MARS model for retrieving the SMC from multi-satellite data is established based on the phase shift. To examine its reliability, the MARS model was compared with a multiple linear regression (MLR) model, a backpropagation neural network (BPNN) model, and a support vector regression (SVR) model in terms of the retrieval accuracy with time-series observation data collected at a typical station. The MARS model proposed in this study effectively retrieved the SMC, with a correlation coefficient ( $R^2$ ) of 0.916 and a root-mean-square error (RMSE) of  $0.021 \text{ cm}^3/\text{cm}^3$ . The elimination of the vegetation impact led to 3.7%, 13.9%, 11.7%, and 16.6% increases in  $R^2$  and 31.3%, 79.7%, 49.0%, and 90.5% decreases in the RMSE for the SMC retrieved by the MLR, BPNN, SVR, and MARS model, respectively. The results demonstrated the feasibility of correcting the vegetation changes based on the multipath effect and the reliability of the MARS model in retrieving the SMC.

**Keywords:** GNSS-IR; signal-to-noise ratio; soil moisture content retrieval; vegetation moisture content; MARS



**Citation:** Lv, J.; Zhang, R.; Tu, J.; Liao, M.; Pang, J.; Yu, B.; Li, K.; Xiang, W.; Fu, Y.; Liu, G. A GNSS-IR Method for Retrieving Soil Moisture Content from Integrated Multi-Satellite Data That Accounts for the Impact of Vegetation Moisture Content. *Remote Sens.* **2021**, *13*, 2442. <https://doi.org/10.3390/rs13132442>

Academic Editor: Damian Wierzbicki

Received: 25 May 2021

Accepted: 21 June 2021

Published: 22 June 2021

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



**Copyright:** © 2021 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/>).

## 1. Introduction

The soil moisture content (SMC) is an important index for terrestrial hydrologic circulation and research in fields such as agriculture, meteorology, and hydrology. Accurate real-time SMC is an important reference for agricultural irrigation, meteorological forecasting, and water resource recycling [1,2]. Global navigation satellite system-interferometric reflectometry (GNSS-IR) is a new microwave sensing technique that primarily takes advantage of the interference effect that is generated by direct and surface-reflected GNSS signals at the receiver, to retrieve surface parameters based on the characteristics of the interference signal. This technique is mainly employed to retrieve the SMC, snow depths, and vegetation parameters [3,4].

In recent years, researchers in China and other countries have achieved marked progress in the use of GNSS-IR to retrieve the SMC, made breakthroughs in areas such as the establishment of empirical models and the selection of optimum characteristic components, and determined a technical route for retrieving the SMC in single surface-cover areas [5–8]. Larson et al. proposed a normalized microwave reflection index and found a good correlation between normalized microwave reflection index (NMRI) and vegetation water content [9–11]. Chew et al. established a database for correcting the phase of the reflection signal based on the law of different vegetation perturbation reflection signals through a large number of simulation experiments, which further improved the accuracy of inversion of soil moisture [12,13]. Wan et al. established that the error of inversion of vegetation moisture content using this model was less than  $1 \text{ kg/m}^2$  [14]. Small et al. verified the effect of three different algorithms to weaken vegetation moisture content from bare soil, single vegetation, and multiple vegetation, respectively [15,16]. To address the impact of the surface vegetation moisture content (VMC) on SMC retrievals, Liang et al. have successively tested the performance of linear regression and backpropagation neural network (BPNN) [17] models in reducing the impact of the VMC. Most of the aforementioned soil moisture inversion algorithms are based on specific GNSS reference stations or the selection of specific satellites with a high inversion accuracy. That is, these models do not have generalized application and require the artificial selection of satellites with good quality data. Thus, there is an urgent need to develop models that can automatically select high-quality data for soil moisture inversion.

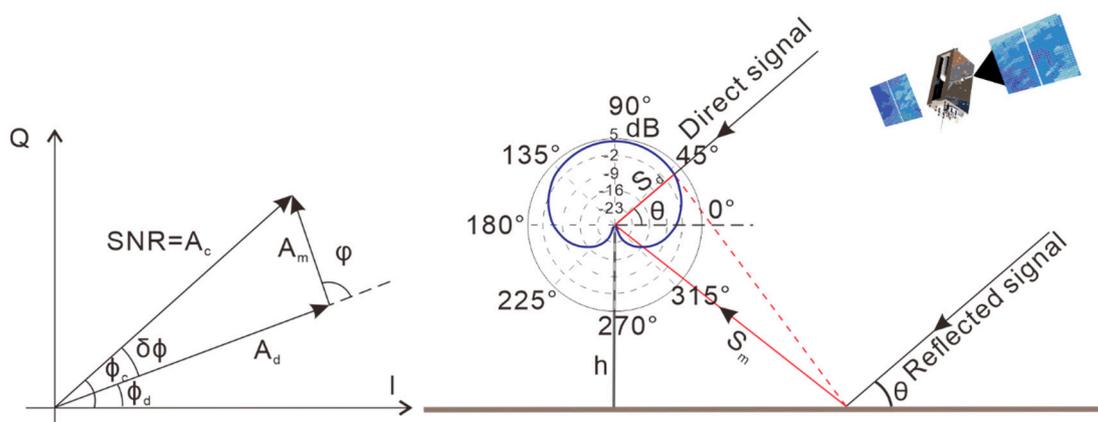
Overall, the current GNSS-IR SMC retrieval methods are mostly limited to the technical route for retrievals from single-satellite data [3–5]. There are relatively large errors and uncertainties in the available empirical or semi-empirical models, and most methods are applicable only to single experimental scenarios (e.g., bare soil) [18]. Considering that the advantages of the comprehensive use of multiple satellites in responding to GNSS-IR information from various angles are promising for reducing the impact of the composition of ground objects and VMC surrounding stations, recent studies have begun to establish combined least-squares (LS) and support vector regression (SVR) models and methods that jointly retrieve the SMC from multi-satellite GNSS-IR signals [19]. However, these relevant models are unable to account for the impact of seasonal vegetation growth on the reflected microwave signal.

Consequently, a novel GNSS-IR method to correct reflection signals for vegetation error was proposed in this study. And the attenuation effect of vegetation cover on the reflection signal was analyzed. The multipath signals obtained in the absence of measured VMC data were used to correct for the effect of vegetation seasonal changes on the reflection signal. A GNSS-IR soil moisture inversion model with multi-satellite data fusion was proposed based on the GNSS-IR information response from different perspectives of multiple satellites, thereby solving the problems of generalization and low automation of existing multi-satellite soil moisture inversion algorithms. The proposed method was implemented as follows: the correlation between the multipath information for the L1 carrier and the normalized difference vegetation index (NDVI) was modeled to estimate the phase shift induced by vegetation information and reduce the characteristic phase component of the signal reflected by vegetation. Then, a nonlinear regression model based on the multivariate adaptive regression spline (MARS) was established and used to retrieve the SMC from the multi-satellite data integrated using the GNSS-IR technique [20]. The SMC was retrieved using four models, namely, a multiple linear regression (MLR) model, a BPNN model, an SVR model, and the MARS model, from time-series observation data collected at a typical station. In addition, the four models were compared in terms of the retrieval accuracy to evaluate the feasibility and accuracy of the proposed model.

## 2. Materials and Methods

### 2.1. GNSS-IR SMC Retrieval Principle

The signal-to-noise ratio (SNR), which is a measure of the signal quality of the antenna of a receiver, is primarily affected by the antenna gain, receiver noise, and multipath effect, the last of which has a particularly pronounced impact. When a satellite is at a low elevation angle  $\theta$ , the SNR is subject to a significant multipath effect, and the direct and reflected signals have approximately the same frequency and produce a relatively stable interference effect at the antenna. In addition, the interference signal oscillates periodically. Let us assume that reflection occurs only once. In Figure 1, I and Q are the in-phase and orthogonal space, respectively, of the vector signal within the carrier tracking loop of the receiver, and the signal received by the receiver is a vectorial superposition of the direct and reflected signals [21,22].



**Figure 1.** The figure on the left shows the geometric principle of signal reflection. The figure on the right is the principal diagram of the receiver antenna gain interferometry. The left half of the right figure is a diagram of the internal antenna gain (dB) of the receiver at a low satellite altitude angle and relatively low gains of the direct and reflected signals. Thus, the signal-to-noise ratio is small, and the oscillation of the reflected multipath signal depends on the distance difference in the travel distance of the direct and reflected signals to the antenna.

The SNR can be represented as:

$$SNR^2 = A_d^2(\theta) + A_m^2(\theta) + 2A_d(\theta)A_m(\theta) \cos \varphi, \tag{1}$$

where the SNR is the composite signal that is formed after interference;  $A_d$  and  $A_m$  are the amplitudes of the direct signal and the reflected signal, respectively;  $\theta$  is the elevation angle of the satellite;  $\varphi$  is the difference between the phases of the direct signal and the phases of the reflected signals; and  $\delta\phi$ ,  $\phi_c$ , and  $\phi_d$  are the phase error after one occurrence of reflection, the phase of the composite signal, and the phase of the direct signal, respectively.

When interference occurs once, the direct signal travels a shorter distance to reach the antenna than the reflected signal, as shown in Figure 1 [23]. The difference  $\Delta S$  between the distances that the direct and reflected signals travel to reach the antenna is expressed as [9–11]:

$$\Delta S = S_m - S_d = 2h \sin \theta, \tag{2}$$

where  $S_m$  is the distance from the reflected signal to the receiver antenna through the ground and  $S_d$  is the distance to the receiver antenna after the direct signal minus the same distance as the reflected signal. Let us assume that the GNSS signal is reflected once and that the satellite is at a low  $\theta$ . Thus, the reflected signal can be approximated from the horizontal reflecting surface, that is,  $\varphi$  is dictated by the distance  $h$  between the antenna of

the receiver and the reflecting surface as well as  $\theta$ . For a GNSS signal with a wavelength of  $\lambda$ ,  $\varphi$  can be derived using (3):

$$\varphi = \frac{2\pi}{\lambda} \Delta S = \frac{4\pi h}{\lambda} \sin \theta. \quad (3)$$

Thus, the rate of change in  $\varphi$  with time is expressed as:

$$\frac{d\varphi}{dt} = 4\pi \frac{h}{\lambda} \cos \theta \frac{d\theta}{dt}, \quad (4)$$

where  $t$  is time,  $h$  is the antenna height,  $\theta$  is the satellite altitude angle, and  $\lambda$  is the global positioning system (GPS) L2 carrier wavelength. According to (4), when  $h$  remains unchanged, as  $\theta$  gradually increases, there is a gradual decrease in the rate of change in the oscillation of the observed SNR. By letting  $\sin \theta = x$ , we can further simplify (4) as follows:

$$\frac{d\varphi}{dx} = 4\pi \frac{h}{\lambda}. \quad (5)$$

According to (5),  $\varphi$  changes linearly with the sine of  $\theta$ . Because the direct component of the interference signal is substantially greater than its reflected component, based on (1), the SNR is fitted to a low-order polynomial. The fitted result is approximated as the direct component. In addition, the reflected component is separated from the interference signal. The reflected signal can be represented by [14]:

$$\text{SNR}_m = A_m \cos\left(\frac{4\pi h}{\lambda} \sin \theta + \varphi\right), \quad (6)$$

where  $\text{SNR}_m$  is the residual SNR,  $A_m$  is the reflected signal amplitude, and  $\varphi$  is the reflected signal phase. The peak frequency  $f$  is obtained using the Lomb–Scargle spectral analysis method based on the sine of the low  $\theta$  and the value of the reflected signal. Next,  $f$  is converted to an equivalent  $h$  based on the following relation:  $h = 2f\lambda$ . Subsequently, the phase shift of the reflected signal is determined by fitting with (6). The phase shift of the reflected signal is mapped to the SMC.

## 2.2. Vegetation Error Correction Based on the Multipath Effect

A microwave signal that is reflected by surface vegetation contains a large amount of VMC information [24,25], which significantly affects the SMC retrieval accuracy of GNSS-IR. Therefore, it is necessary to consider the vegetation information in the retrieval of the SMC using GNSS-IR. Based on the GPS pseudorange and carrier phase, Larson and Small (2014) proposed a normalized microwave reflection index (NMRI) and mapped it to the NDVI [7,16]. The multipath error for the L1 carrier (MP1) can be represented by:

$$\text{MP1} = P_1 - \frac{f_1^2 + f_2^2}{f_1^2 - f_2^2} \lambda_1 \varphi_1 + \frac{2f_2^2}{f_1^2 - f_2^2} \lambda_2 \varphi_2, \quad (7)$$

where MP1 is the multipath error for the L1 carrier,  $P_1$  is the observed pseudorange of the L1 carrier,  $f_1$  and  $f_2$  are the frequencies of the L1 carrier and L2 carrier, respectively;  $\lambda_1$  and  $\lambda_2$  are the wavelengths of the L1 carrier and L2 carrier, respectively, and  $\varphi_1$  and  $\varphi_2$  are the observed phases of the L1 carrier and L2 carrier, respectively.

Based on MP1, which changes with the epoch, the root mean square (RMS) of the MP1 for each single satellite is calculated. Subsequently, the RMS of the MP1 of a single day is calculated by a weighted summation of the value observed by each satellite [26]. The NMRI can be represented by:

$$\text{NMRI} = \frac{\max(\text{RMS}_{\text{MP1}}) - \text{RMS}_{\text{MP1}}}{\max(\text{RMS})}, \quad (8)$$

where  $\max(\text{RMS}_{\text{MP1}})$  is the average value of the largest 5% of the  $\text{RMS}_{\text{MP1}}$  values in the annual time series and  $\text{RMS}_{\text{MP1}}$  is the RMS of the MP1 of a single day. For the same period as that of the NDVI values, the NMRI values are obtained by downsampling the calculated values of the NMRI. In addition, the mapping model is employed to calculate the NDVI of the experimental period.

The NDVI is a parameter that reflects the vegetation growth conditions and vegetation coverage. The vegetation surrounding a station is the primary factor that causes a shift in the phase of the reflected signal. In the absence of measured VMC data, the NDVI is an important factor that reflects the shift in the phase of the reflected signal. To ensure that the mathematical scale of the NDVI is consistent with that of the phase of the reflected signal, the value calculated by zeroing the median of the largest 15% of the NDVI values in the annual time series is employed to approximately replace the vegetation-induced phase shift  $\Delta\varphi_{\text{veg}}(t)$ . The corrected phase shift of the reflected signal can be represented by:

$$\varphi_r(t) = \varphi(t) - \Delta\varphi_{\text{veg}}(t), \quad (9)$$

where  $\varphi_r(t)$  is the corrected phase and  $\varphi(t)$  is the original phase of the SNR of the reflected signal. The SMC can be retrieved from  $\varphi_r(t)$ .

### 2.3. MARS Model

The SMC retrieved from single-satellite data is unable to sufficiently reflect the SMC surrounding a station. In addition, MLR hardly reflects the inverse accuracy. To address these problems, this study presents a MARS model that is capable of combining multi-satellite data to produce sufficient SMC information surrounding a station and improve the retrieval accuracy. The MARS method is a data analysis technique that was proposed by U.S. statistician Jerome Friedman in 1991 [27–29]; it has been extensively applied due to its high modeling efficiency, and high interpretability. This technique can be divided into three steps, namely, a forward stepwise procedure, a backward pruning procedure, and model selection. In the forward stepwise procedure, a basis function (BF) is automatically established based on the input data. Based on the BF, the data are divided into different spatial regions. Subsequently, a new BF is established by fitting a linear regression model in each region. An overfitted model is obtained when the forward stepwise procedure is completed. In the backward pruning procedure, the BFs that contribute insignificantly to the overfitted forward stepwise model are removed while ensuring their accuracy. An optimum combination of satellites is selected as a regression model.

The BFs in MARS are composed of truncated spline functions or the product of multiple spline functions. BFs can be defined as:

$$\left[ S_{km}(x_{v(k,m)} - t) \right]_+ = \begin{cases} x - t_{km} & (x > t_{km}) \\ 0 & (x \leq t_{km}) \end{cases}, \quad (10)$$

$$\left[ S_{km}(x_{v(k,m)} - t) \right]_+ = \begin{cases} t_{km} - x & (x < t_{km}) \\ 0 & (x \geq t_{km}) \end{cases}, \quad (11)$$

where  $S_m(x)$  is the  $m$ th spline function,  $t$  is the location of the node of the spline function, that is, all the observations of each input satellite as nodes,  $v(k, m)$  is an independent variable identifier, and  $t_{km}$  is the location of the identified node. Based on Equations (10) and (11), the MARS model can be defined as:

$$\hat{y} = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{k_m} \left[ S_{km}(x_{v(k,m)} - t_{km}) \right]_+, \quad (12)$$

where  $\hat{y}$  is the predicted value of the output variable of the model, which is the predicted value of soil moisture,  $a_0$  is a constant parameter,  $a_m$  is the coefficient of the  $m$ th BF, and  $M$

is the number of BFs, that is, the number of spline functions contained in the model. The  $a_m$  of a combination of BFs is determined by calculating the sum of squares of the LS residuals.

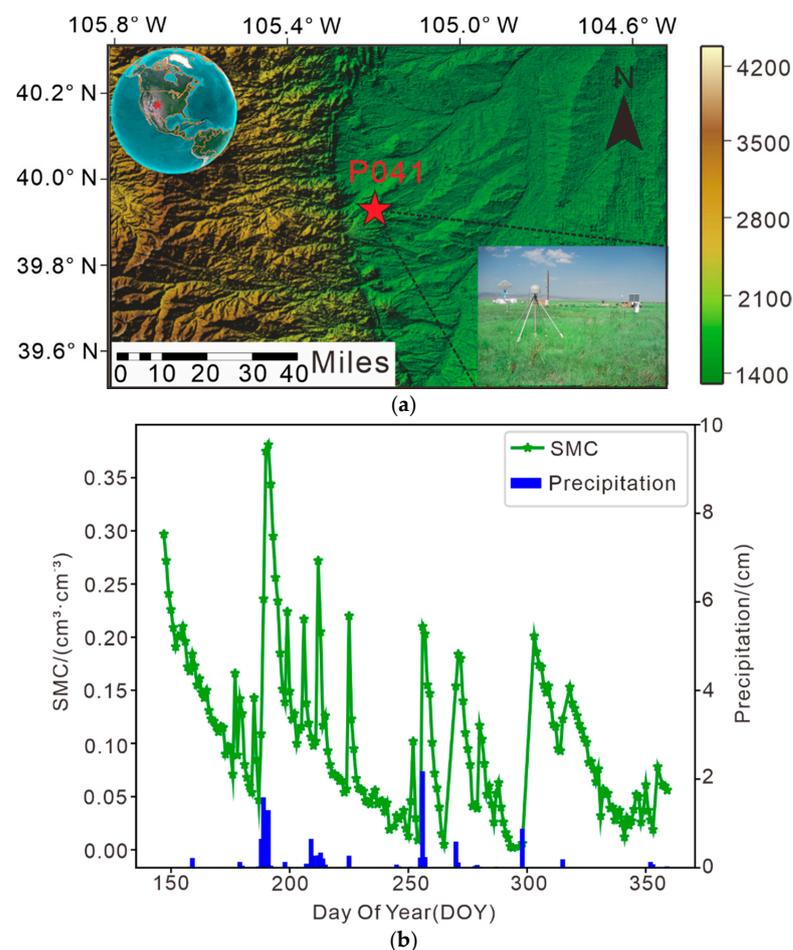
The BFs that contribute insignificantly to the overfitted MARS model are eliminated by backward pruning. Thus, an optimum combined BF model is obtained. An optimum MARS model is determined by generalized cross-validation (GCV). An optimum MARS model is obtained at the minimum GCV.

$$\text{GCV}(\lambda) = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{(1 - M(\lambda)/N)^2}, \quad (13)$$

where  $\lambda$  is the number of terms in the model,  $M(\lambda)$  is the number of effective parameters in the model, which is equal to the number of terms in the model, plus the number of parameters at the optimal node location.,  $N$  is the number of BFs, and  $\hat{y}_i$  is the optimum model value that is estimated in each step, which is the predicted value of soil moisture.

### 3. Data Sources

The GPS observation data and reference SMC data that are collected at the P041 station (Figure 2) of the U.S. Plate Boundary Observatory (PBO) between days of year (DOYs) 147 and 360 of 2012 were selected as the experimental data in this study. Located in the Colorado, the U.S., and positioned at an altitude of 1728.8 m, the P041 station (39.94949°N, 105.19427°W) is surrounded by flat terrain and is unobstructed by large obstacles. At this station, the reflected signal features are affected primarily by vegetation and precipitation.



**Figure 2.** (a) The digital elevation model diagram of the experimental site, (b) The soil moisture-rainfall diagram during the experimental period.

Figure 2 shows the observed SMC and precipitation data series for DOYs 147–360 of 2012, which are presented in a broken line graph and a histogram, respectively. As demonstrated in Figure 2, eight significant precipitation events occurred during the experimental period, with a maximum precipitation of 21.8 mm. There was a considerable increase in the SMC during the precipitation events, particularly on DOYs 188–191, 209–215, 255–257, 270, and 298. Continuous precipitation led to a significant nonlinear increase in the SMC. As precipitation decreased or stopped, there was a decrease in the SMC. Evidently, precipitation was the primary factor that caused sudden changes in the SMC. The precipitation at the P041 station during the experimental period was appropriate and suitable for SMC retrieval.

### 4. Experiment and Results

#### 4.1. Experimental Technical Scheme

Figure 3 shows the flow chart of the soil moisture inversion technique in this article. From the figure, it can be seen that the technical route of the article can be divided into three lines: (1) GNSS-IR soil moisture inversion data pre-processing, extracting the characteristic parameters of the reflection signal from the observation data acquired by the original GNSS receiver. (2) Establish a vegetation error correction model using GNSS multi-path data and MODIS NDVI data to correct the influence of vegetation interference on the reflected signal. (3) Inverse the soil moisture by establishing a multivariate adaptive regression model and do a comparative analysis with BPNN, SVRM, and MLR models.

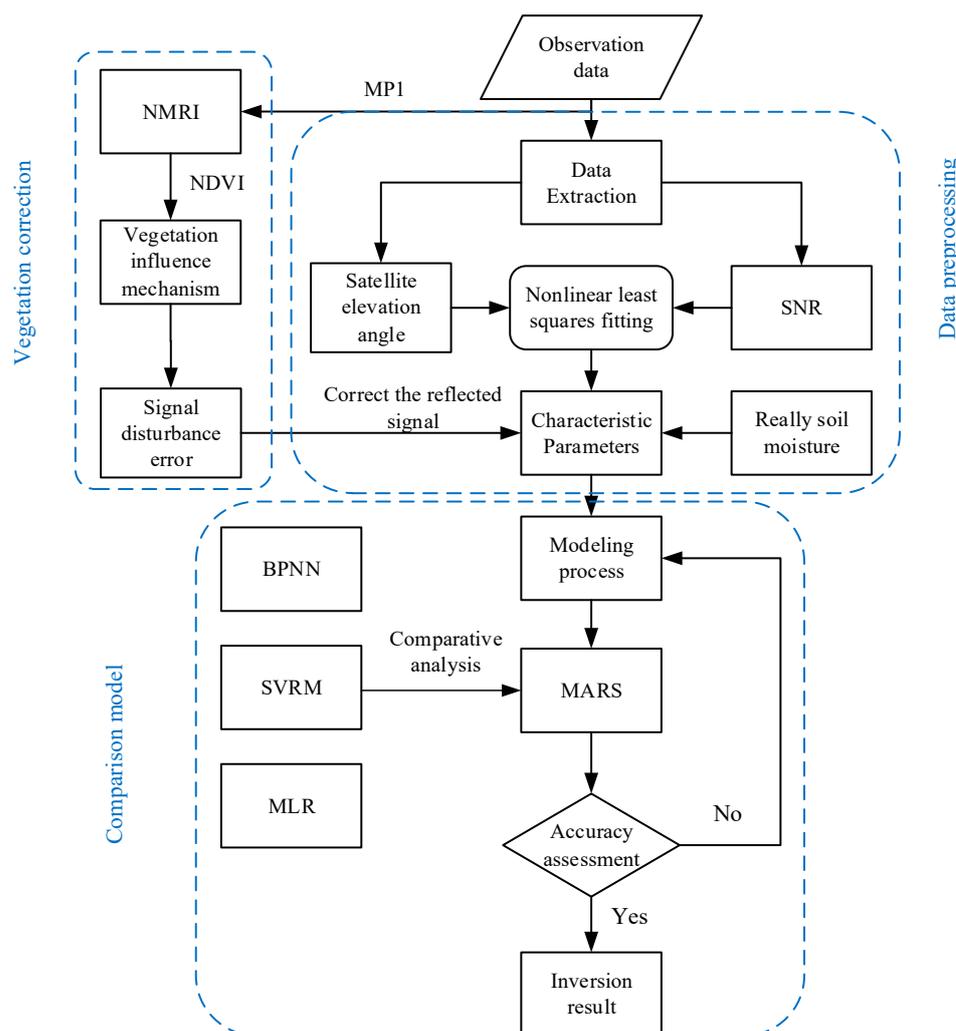


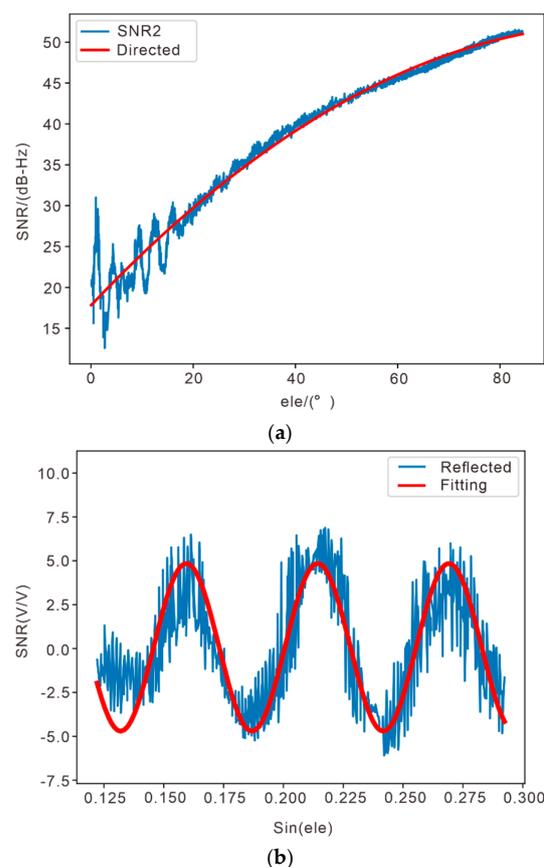
Figure 3. Technical process of soil moisture inversion.

Machine learning algorithms are used in current soil moisture inversion models to improve the accuracy of the inversion process. However, as the machine learning algorithm is a black box that contains the defects of the model itself, mathematical expressions for the soil moisture cannot be obtained, and the model parameters are difficult to modulate. The proposed MARS model for soil moisture inversion adaptively combines multiple satellite data and selects the best combination of satellites for soil moisture inversion, yielding highly accurate results and mathematical expressions.

#### 4.2. Reflected Signal Feature Parameter Extraction

GNSS receiver observation data are in the format of carrier phase and pseudorange, and GNSS-IR soil moisture inversion requires the use of satellite altitude angle and L2 carrier signal-to-noise ratio data. This needs to be calculated from the GNSS observation file and the navigation file by using the relevant equations to calculate the signal-to-noise ratio, satellite altitude angle, and other relevant data, respectively.

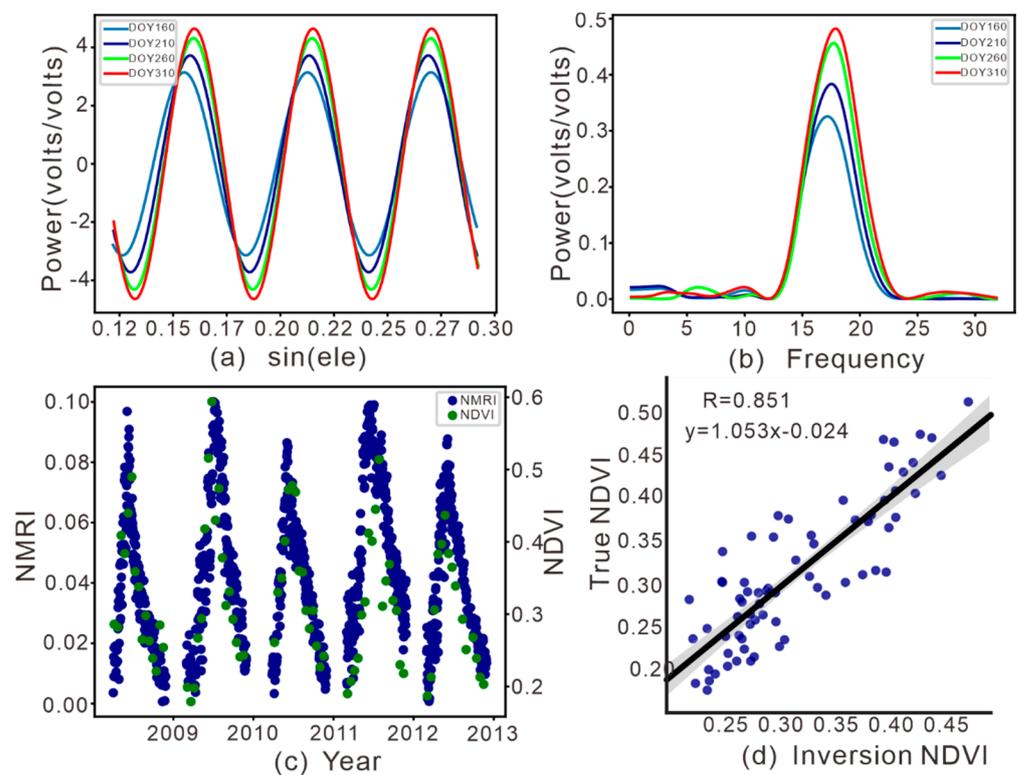
The upper panel of Figure 4 is a plot of the SNR versus the satellite altitude angle. The L2 carrier SNR is shown in blue, and the direct signal data are shown in red. At low satellite altitude angles, there is a severe multipath effect for the SNR, which exhibits periodic oscillations. As the satellite altitude angle gradually increases, there is considerable antenna gain, and the SNR stabilizes. To extract the reflected signal data from the GNSS SNR data, the direct signal is separated from the SNR by using Equation (1) and fitting the SNR data by a low-order polynomial. The lower part of Figure 4 shows the nonlinear least-squares cosine fit to the reflected signal. The reflected signal is a nonlinear least-squares cosine, which is fitted by Equation (6) to extract the reflected signal amplitude, phase, and frequency.



**Figure 4.** Reflected signal preprocessing. (a) L2 SNR data for one GPS satellite are shown in blue. The direct signal is represented by the smooth curve in red; (b) SNR data with the direct signal removed and converted to a linear scale.

### 4.3. Vegetation Impact Correction

When using GNSS-IR to retrieve the SMC, changes in the VMC yield corresponding an increase or decrease in the phase shift of the reflected signal. Therefore, it is necessary to correct the vegetation impact. Figure 5a shows the nonlinear LS cosine fit of the reflected signals on DOYs 160, 210, 260, and 310. As demonstrated in Figure 5a, the vigorous vegetation growth in the summer (DOY160) produced the smallest amplitude of the reflected signal of the DOYs that were compared, whereas the amplitude of the reflected signal in the winter (DOY310) was the largest. Figure 5b shows the Lomb–Scargle spectral analysis plots for the corresponding DOYs. As demonstrated in Figure 5b, as a result of the vegetation growth cycle, there was a corresponding increase or decrease in the main frequency on each DOY. To address the changes in the reflected signal caused by vegetation in different seasons, the previously established NMRI–NDVI correlation model was employed to correct the phase shift of the reflected signal.

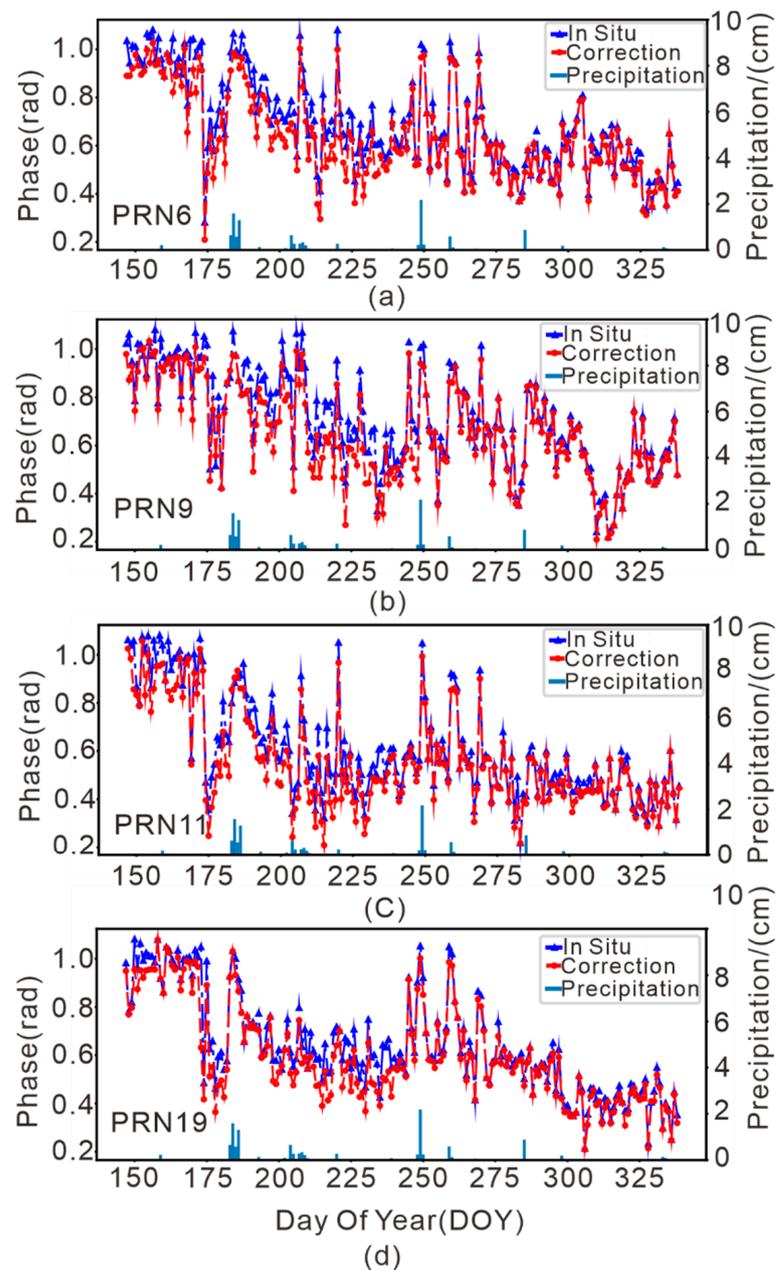


**Figure 5.** (a) The nonlinear least square cosine fitting diagram with the typical characteristics of four days during the experimental period. (b) The upper right figure is the L-S spectrum analysis diagram at the corresponding time; (c) The NMRI-NDVI diagram; (d) The inverse NDVI linear regression diagram.

Figure 5c shows the distribution of single-day NMRI values and 16-day NDVI values in the period from 2008 to 2012. The NMRI changed periodically with vegetation in the time series and its fluctuations tended to be consistent within the time domain. In addition, the peak and valley values of the NMRI matched relatively well. The NMRI values from the same period as that of the NDVI values were obtained by downsampling. A simple linear regression model between the NMRI and the NDVI was established. As demonstrated in Figure 5d, the correlation coefficient  $R$  and the RMS error (RMSE) of the linear regression inversion model were approximately 0.851 and  $0.043 \text{ cm}^3/\text{cm}^3$ , respectively. This finding suggests a relatively strong linear relationship between the NMRI and the NDVI.

Based on the phase shift estimated using the NDVI, the phase shift of the reflected signal was corrected. Due to the limited length of this article, the corrected phase shifts of only the four satellites with pseudorandom noise numbers (PRN) 6, 9, 11, and 19 are

presented, as shown in Figure 6 (in each plot, the blue line and red line show the uncorrected original phase and the corrected phase shift, respectively). As demonstrated in Figure 6, the phase shift of the reflected signal was primarily corrected for DOYs 150–250. The corrected phase shift fluctuated less than the original phase. In addition, the phase shift of the reflected signal fluctuated to a relatively large extent several times during the experimental period, which was related to the sharp increase in the SMC due to continuous precipitation.



**Figure 6.** (a) PRN 6 reflection signal delay phase correction diagrams. (b) PRN 9 reflection signal delay phase correction diagrams. (c) PRN 11 reflection signal delay phase correction diagrams. (d) PRN 19 reflection signal delay phase correction diagrams.

As demonstrated in Figure 6, the response to the SMC varied among the satellites. This response was primarily caused by the differences among the geometric motion trajectories relative to the GPS antennas during the observation period and the performance of the satellites as well as the multipath surface environment. Therefore, the direct retrieval of the SMC from single-satellite data involves relatively large uncertainties, and it is difficult

to employ a method to treat the outliers in single-satellite data. In this study, the MARS model was used to treat the values observed by multiple satellites as input and to establish BFs for all the observed values by forward fitting. Moreover, the BFs that contribute insignificantly to SMC retrievals were eliminated by the GCV. Thus, the automatic selection of a combination of satellites with the highest SMC retrieval accuracy by the MARS model was achieved.

#### 4.4. Soil Moisture Inversion Results

MARS is a model that is specifically utilized to process high-dimensional data. Thus, the observation data collected by 17 effective satellites during the observation period were used as input, and the combination of satellites in the optimum model was treated as output. Based on these data, the MARS method was employed to establish an SMC retrieval model. The maximum number of interactive BFs was set to 1, and the number of maximum BF values was set to 40. An optimum SMC retrieval model was obtained at a minimum GCV of 0.0008, as shown in Equation (14). Table 1 summarizes all the BFs in Equation (14).

$$y = 0.1766 + 0.034BF_1 + 0.054BF_2 - 0.507BF_3 - 0.073BF_4 - 0.061BF_5 - 0.079BF_6 - 0.471BF_7 + 0.082BF_8 + 0.092BF_9 + 0.801BF_{10} - 0.302BF_{11} \quad (14)$$

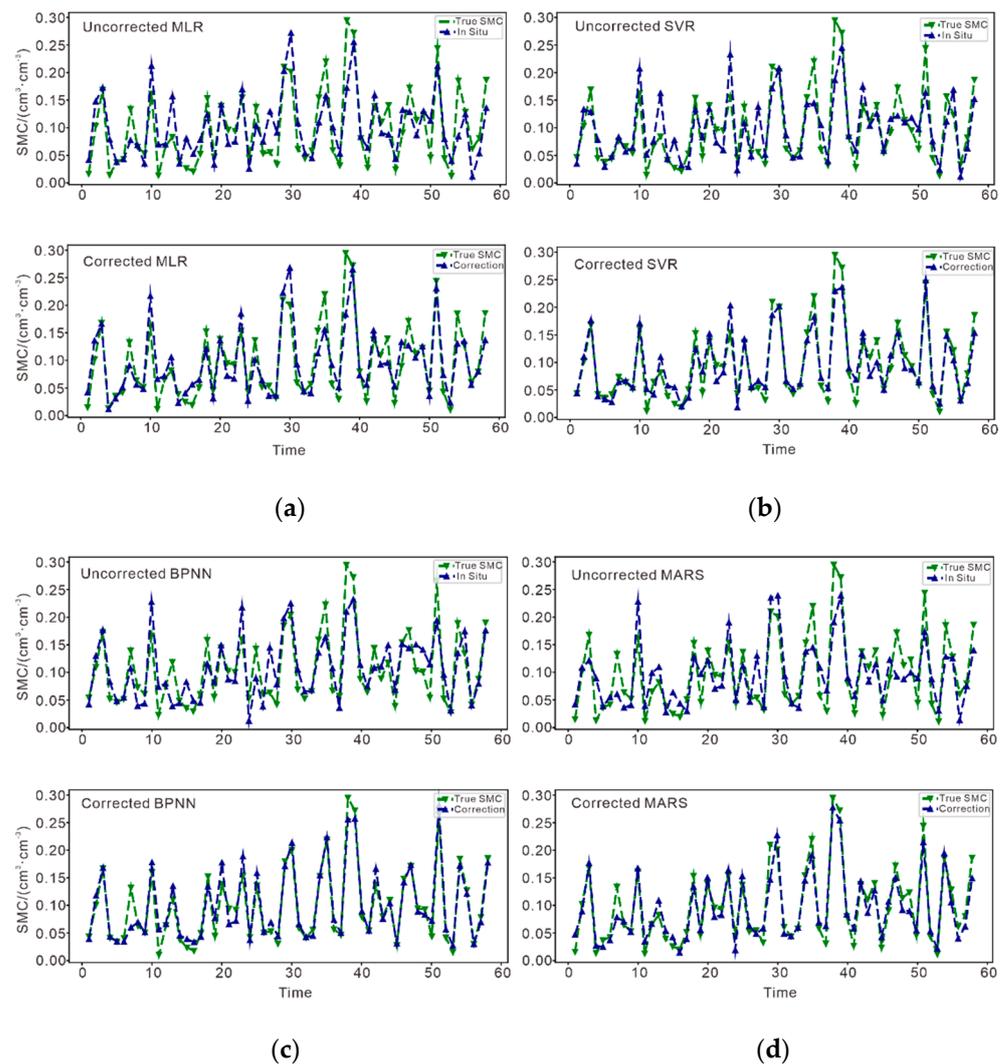
**Table 1.** BFs in MARS.

BF No.	BF	Variable	Coefficient
BF1	X16	PRN16	0.034
BF2	Max(0, 0.065-X9)	PRN9	0.054
BF3	Max(0, 0.682-X9)	PRN9	-0.507
BF4	Max(0, 0.742-X4)	PRN4	-0.073
BF5	Max(0, 0.920-X5)	PRN5	-0.061
BF6	Max(0, 0.995-X14)	PRN14	-0.079
BF7	Max(0, X14-0.995)	PRN14	-0.471
BF8	Max(0, X15-0.195)	PRN15	-0.082
BF9	Max(0, X17-0.304)	PRN17	0.092
BF10	Max(0, X9-1.029)	PRN9	0.801
BF11	Max(0, X9-0.682)	PRN9	-0.302

By the stepwise selection of variables using the “forward” and “backward” algorithms of MARS, an optimum combination of satellites (i.e., those with PRN numbers 4, 5, 9, 14, 15, 16, and 17) for SMC retrievals was determined.

To examine the feasibility and efficacy of the MARS models and considering that machine learning algorithms solve high-dimensional nonlinear problems with self-learning and self-adaptive capabilities, four schemes (1–4) were formulated for comparative analysis. Specifically, schemes 1–4 involved the multi-satellite integration-based MARS estimation model, a multi-satellite-based MLR estimation model, a multi-satellite-based BPNN model [30], and a multi-satellite-based SVR estimation model [31]. To reduce the modeling errors, the method described previously was employed to standardize the phase shift of the reflected signal. In addition, the feasibility of vegetation impact correction in each of the four SMC retrieval schemes was investigated. Moreover, 70% of the data collected at the P041 station on DOYs 147–360 of 2012 were randomly selected to establish an SMC retrieval model, and the remaining 30% of the data were applied to examine the reliability and accuracy of the model.

Figure 7 shows the validation of the vegetation-uncorrected and vegetation-corrected SMC values that were retrieved using the four models for the experimental period (blue and green lines signify the retrieved SMC values and measured SMC values, respectively).



**Figure 7.** (a) MLR invert the soil moisture comparison map for the two conditions of uncorrected phase and corrected phase. (b) SVR invert the soil moisture comparison map for the two conditions of uncorrected phase and corrected phase. (c) BPNN invert the soil moisture comparison map for the two conditions of uncorrected phase and corrected phase. (d) MARS invert the soil moisture comparison map for the two conditions of uncorrected phase and corrected phase.

As demonstrated in Figure 7, the SMC retrieved using the MLR model was, to a certain extent, consistent with the measured SMC in terms of the variation trend, but its curve did not satisfactorily coincide with that of the measured SMC. Compared to the retrieval accuracy of the other three models, the SMC retrieval accuracy of the MLR model needs to be improved. While the BPNN, and SVR models were able to retrieve the changes relatively satisfactorily in SMC, they failed to give a specific explicit expression. The MARS model was able to retrieve the variation trend of the SMC more satisfactorily, and its estimation error was more stable. This model effectively improved the low SMC retrieval accuracy associated with the MLR model. Moreover, the SMC that was retrieved using each of the four models without vegetation correction exhibited certain fluctuations. Furthermore, the SMC retrieved from the corrected phase shift using each of the four models was considerably more accurate than that retrieved from the original phase. This finding further demonstrates the feasibility of correcting vegetation changes when using GNSS-IR to retrieve SMC.

## 5. Discussion

### 5.1. Correction of Vegetation Error Term Analysis

Considering Figures 5 and 6 together shows that the calculated NMRI for station P041 is correlated with the NDVI. The effect of seasonal vegetation changes on the reflection signal over the investigated time period has the same periodicity as the vegetation growth season. During the summer season, relatively luxuriant vegetation has a strong effect on the reflected signal, resulting in a decrease in the amplitude, phase, and frequency of the reflected signal. Figure 6 shows that the vegetation correction is most noticeable in the summer. By comparison, relatively sparse vegetation in the winter has a relatively small effect on the reflected signal, and the phase correction of the reflected signal is not discernible in the figure.

Figure 8 verifies the accuracy and reliability of the vegetation correction. The correlation plot of the results obtained using the four models validates the proposed MARS model and the vegetation error correction. The results show that the correlation of all four models improves by 10% by correcting the vegetation error. This result demonstrates that the NMRI calculated using the GNSS multipath signal can be used to effectively correct the phase error from the seasonal vegetation changes.

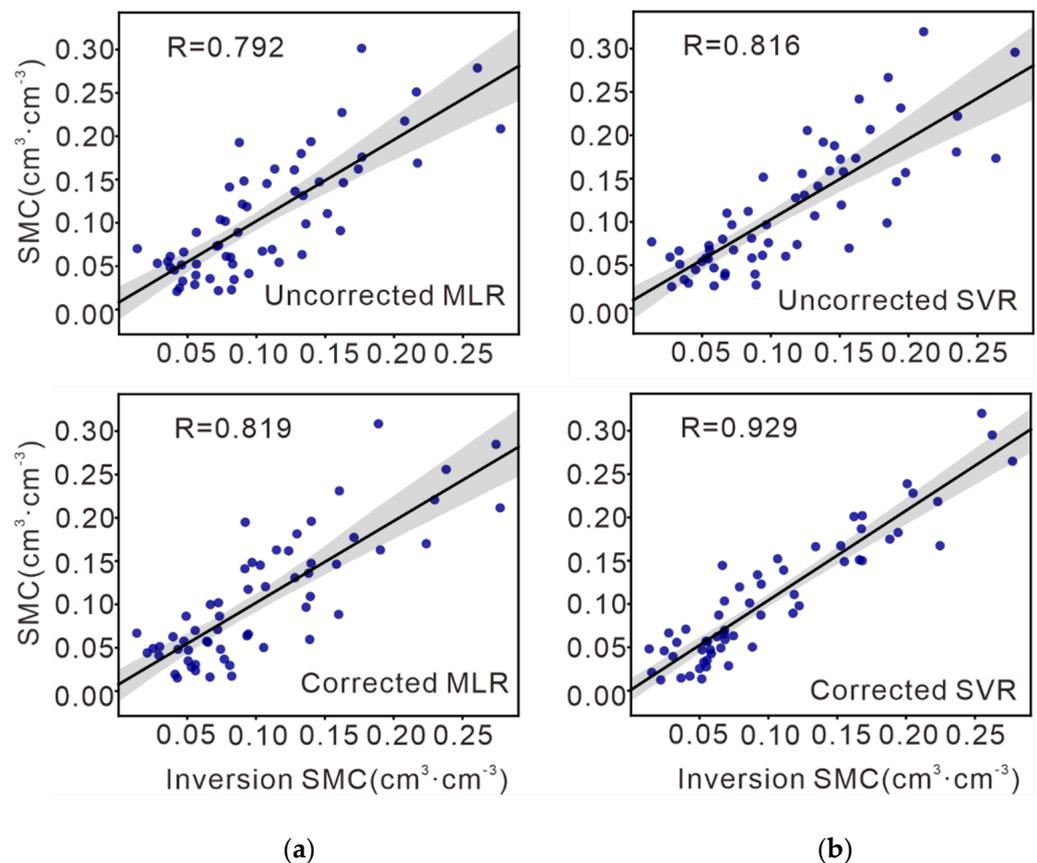
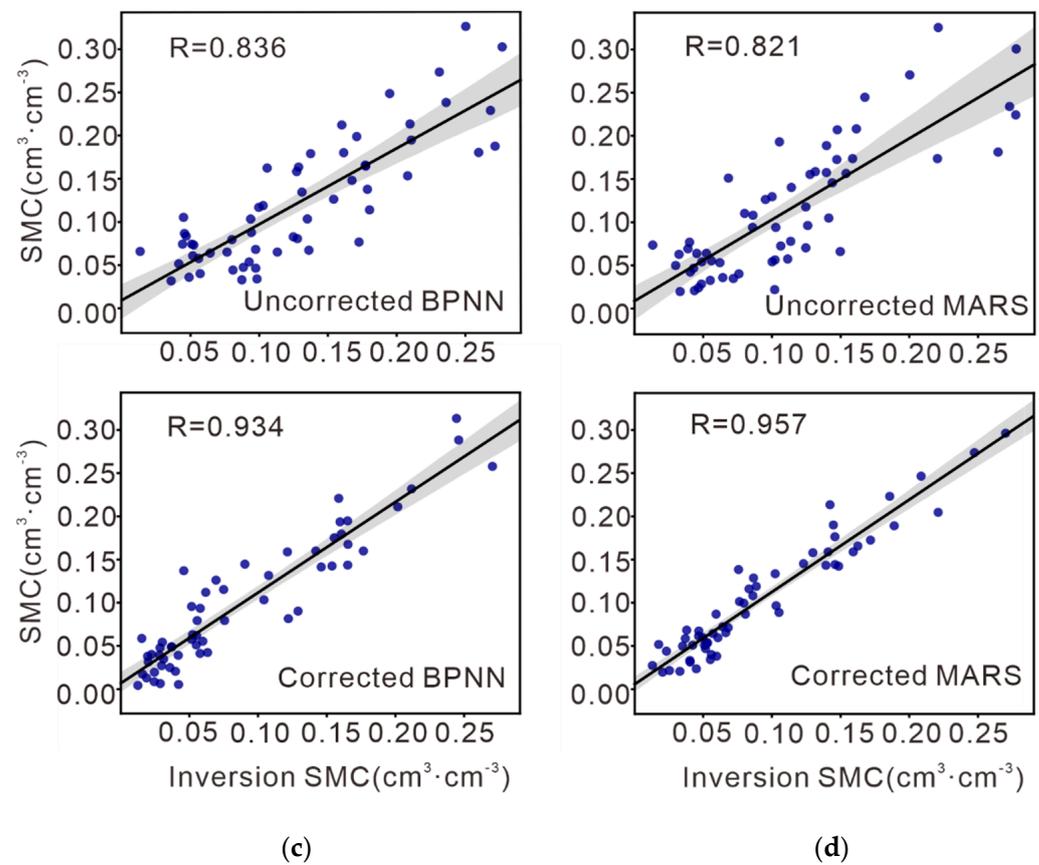


Figure 8. Cont.



**Figure 8.** (a) Linear regression analysis of the estimation results of MLR with the uncorrected and corrected phases and the reference value of soil moisture. (b) Linear regression analysis of the estimation results of SVR with the uncorrected and corrected phase and the reference value of soil moisture. (c) Linear regression analysis of the estimation results of MLR with the uncorrected and corrected phases and the reference value of soil moisture. (d) Linear regression analysis of the estimation results of MLR with the uncorrected and corrected phases and the reference value of soil moisture.

### 5.2. Soil Moisture Inversion Correlation Analysis

In order to further evaluate the performance of each scheme comprehensively and verify the reliability and generalizability of the MARS model proposed in this paper, the correlation coefficient  $R$ , root mean square error (RMSE), and mean absolute error (MAE) are used for accuracy evaluation. Figure 8 shows the  $R$  values for the four models with the uncorrected and corrected phase shifts.

Table 2 shows that the indexes for the machine learning algorithms were relatively similar to those for the MARS model.

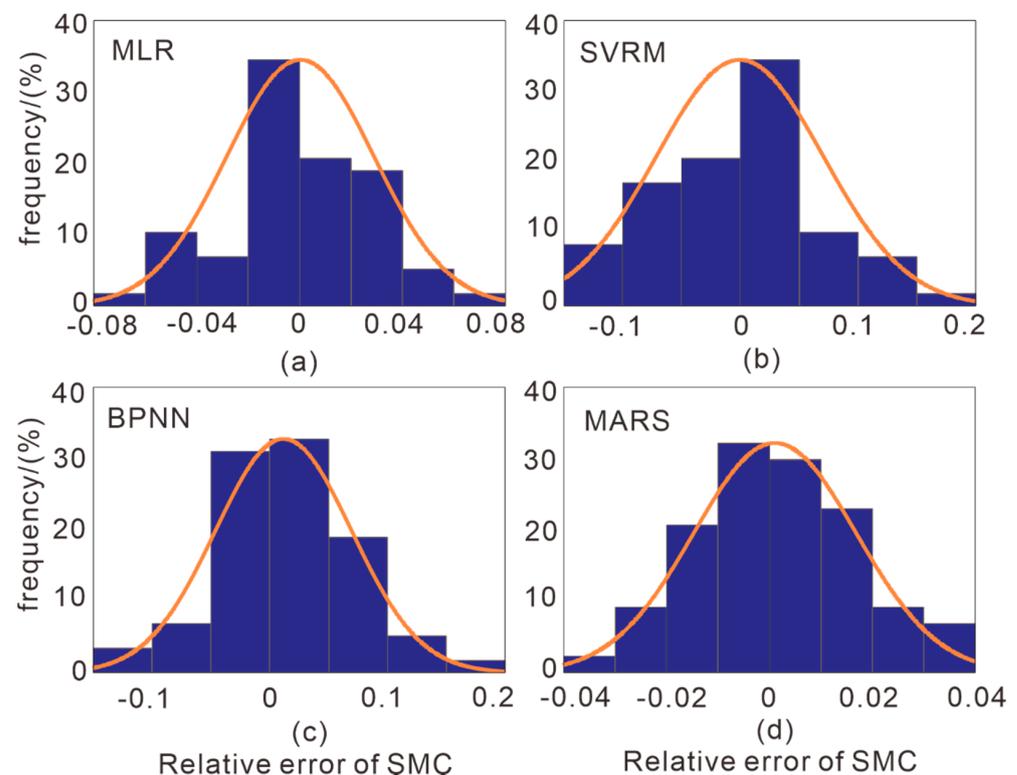
**Table 2.** Statistics of the SMC estimation accuracy of each model.

	Before Vegetation Correction				After Vegetation Correction			
	MLR	SVR	BPNN	MARS	MLR	SVR	BPNN	MARS
R	0.792	0.816	0.836	0.821	0.819	0.929	0.934	0.957
RMSE	0.126	0.115	0.076	0.040	0.096	0.064	0.051	0.021
MAE	0.097	0.086	0.034	0.032	0.060	0.045	0.032	0.017

The  $R$  values for the SVR, BPNN, and MARS models were 0.816, 0.836, and 0.821 before vegetation correction and 0.929, 0.934, and 0.957 after vegetation correction, respectively. This finding suggests that the accuracy was improved by at least 11.7% after vegetation

correction. After vegetation correction, the accuracy of the MARS model was 16.9% higher than that of each of the other three models, and its RMSE and MAE were  $0.021 \text{ cm}^3/\text{cm}^3$  and  $0.017 \text{ cm}^3/\text{cm}^3$ , respectively, which were the smallest of the four models. A comprehensive comparative analysis based on Figures 7 and 8 reveals the following results: The MARS estimation model based on integrated multi-satellite data ensured local error stability during the estimation process and obtained an optimum combination of satellites for SMC retrievals by eliminating overfitted BFs during the modeling process by GCV. Compared to conventional methods, the MARS model was capable of more effectively inhibiting the gross errors caused by single-satellite data.

We analyzed the soil moisture inversion error, calculated the absolute soil moisture inversion error (the difference between the inversion error and the true value, as shown in Figure 9) and analyzed the interval distribution pattern.



**Figure 9.** (a) Statistical results of the relative error of MLR. (b) Statistical results of the relative error of SVRM. (c) Statistical results of the relative error of BPNN. (d) Statistical results of the relative error of MARS.

Figure 9 is a statistical histogram of the proportion of the absolute error for the soil moisture inversion obtained using the four models. There is a large proportion of small absolute errors in the inverse soil moisture. The absolute error distribution for the MARS model lies within the range  $-0.04 \sim 0.04$ , which is narrower than that of the other three models and conforms to a normal distribution overall. These results further illustrate the advantages resulting from the high accuracy of the MARS model for soil moisture inversion.

As the MARS estimation model is based on integrated multi-satellite data, the advantages of multiple satellites are combined to obtain GNSS-IR information from various angles. The selection of an optimal combination of satellites for SMC retrieval produces complementary information from the phase shifts of the different satellites. The MARS model was found to adapt satisfactorily to the phase shift from seasonal vegetation growth. The MARS model obtained by GCV was not overfitted, resulting in satisfactory performance. The MARS model exhibited superior performance to the conventional MLR, SVR, and BPNN models under the same conditions.

## 6. Conclusions

Long-term accurate monitoring of the SMC, which is an important index that measures global water circulation, is exceedingly important and has significant application prospects. Based on the current limitations in GNSS-IR SMC research in areas such as data utilization and application scenarios, a MARS SMC retrieval model based on integrated multi-satellite data, which accounts for the impact of VMC, was established in this study by combining the technical approaches of vegetation impact correction and multi-satellite data integration by using GNSS-IR multi-path and signal-to-noise ratio data. The following conclusions were derived from the experimental analysis:

- (1) The NMRI that was generated based on MP1 exhibits notable periodicity and is strongly linearly correlated with the NDVI. The zeroed NDVI can adequately correct the phase shift of the reflected signal caused by vegetation.
- (2) The MARS algorithm fully realized the advantages of multi-satellite data integration in retrieving SMC and effectively addressed that the SMC estimated from single-satellite data cannot sufficiently reflect actual surface conditions. In addition, GCV was conducive to eliminating the satellites that significantly interfere with SMC retrievals and determining the combination of satellites with the highest SMC retrieval accuracy.
- (3) Compared to the SVR and BPNN models, the MARS model could obtain a combined multi-satellite expression when it was used to retrieve the SMC and had excellent generalization capability. In addition, the MARS algorithm could fully exploit its capabilities to ensure fast modeling, a relatively stable fitting process, and a relatively stable estimation error.

It is feasible and effective to use the MARS model to retrieve the SMC from integrated multi-satellite data. This method is, to a certain extent, better than the available models and SMC retrieval methods in reliability and stability. However, the GNSS-IR SMC retrieval process is affected by terrain conditions and soil roughness, and the physical mechanism of the interaction between the reflected signal and vegetation remains unclear. This topic warrants further investigation.

**Author Contributions:** Conceptualization, R.Z. and G.L.; methodology, J.L.; formal analysis, J.T.; data curation, M.L., B.Y., K.L. and J.P.; writing—original draft preparation, J.L.; writing—review and editing, Y.F. and W.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was jointly funded by the National Natural Science Foundation of China (Grant No. 41771402); the National Key R&D Program of China (Grant No. 2017YFB0502700); and the Sichuan Science and Technology Program (No. 2018JY0664, 2019ZDZX0042, 2020JDTD0003, 2020YJ0322).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The GNSS site data were provided under the PBO Observation Program of the United States, and the measured soil moisture data were obtained from <http://xenon.colorado.edu/portal> (accessed on 22 June 2021).

**Acknowledgments:** We thank the PBO Observation Program of the United States for providing the GNSS data, and the University of Colorado for providing data for soil moisture comparison analysis.

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

## References

1. Koster, R.D.; Dirmeyer, P.A.; Guo, Z.; Bonan, G.; Chan, E.; Cox, P.; Gordon, C.T.; Kanae, S.; Kowalczyk, E.; Lawrence, D.; et al. Regions of Strong Coupling between Soil Moisture and Precipitation. *Science* **2004**, *305*, 1138–1140. [[CrossRef](#)] [[PubMed](#)]
2. Jackson, T.J.; Schmugge, J.; Engman, E.T. Remote sensing applications to hydrology: Soil moisture. *Int. Assoc. Sci. Hydrol.* **1996**, *41*, 517–530. [[CrossRef](#)]
3. Larson, K.M.; Small, E.E. GPS ground networks for water cycle sensing. *IEEE Geosci. Remote Sens. Sym.* **2014**, 3822–3825.

4. Larson, K.M. GPS interferometric reflectometry: Applications to surface soil moisture, snow depth, and vegetation water content in the western United States. *Wiley Interdiscip. Rev. Water*. **2016**, *3*, 775–787.
5. Zavorotny, V.U.; Larson, K.M.; Braun, J.J.; Small, E.E.; Gutmann, E.D.; Bilich, A.L. A Physical Model for GPS Multipath Caused by Land Reflections: Toward Bare Soil Moisture Retrievals. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2010**, *3*, 100–110. [[CrossRef](#)]
6. Larson, K.M.; Small, E.E. Normalized Microwave Reflection Index: A Vegetation Measurement Derived from GPS Networks. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2014**, *7*, 1501–1511. [[CrossRef](#)]
7. Chew, C.C.; Small, E.E.; Larson, K.M.; Zavorotny, V.U. Effects of Near-Surface Soil Moisture on GPS SNR Data: Development of a Retrieval Algorithm for Soil Moisture. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 537–543. [[CrossRef](#)]
8. Roussel, N.; Darrozes, J.; Ha, C.; Boniface, K.; Frappart, F.; Ramillien, G. Multi-scale volumetric soil moisture detection from GNSS SNR data: Ground-based and airborne applications. *MetroAeroSpace* **2016**, 573–578.
9. Larson, K.M.; Small, E.E.; Gutmann, E.; Bilich, A.; Axelrad, P.; Braun, J. Using GPS Multipath to Measure Soil Moisture Fluctuations: Initial Results. *GPS Solut.* **2008**, *12*, 173–177. [[CrossRef](#)]
10. 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*, 24–35. [[CrossRef](#)]
11. Larson, K.M.; Braun, J.J.; Small, E.E.; Zavorotny, V.U.; Gutmann, V.U.; Bilich, A.L. GPS Multipath and Its Relation to Near-Surface Soil Moisture Content. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2010**, *3*, 91–99. [[CrossRef](#)]
12. Chew, C.C.; Small, E.E.; Larson, K.M.; Zavorotny, V.U. Vegetation Sensing Using GPS-Interferometric Reflectometry: Theoretical Effects of Canopy Parameters on Signal-to-Noise Ratio Data. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2755–2764. [[CrossRef](#)]
13. Chew, C.C.; Small, E.E.; Larson, K.M. An algorithm for soil moisture estimation using GPS-interferometric reflectometry for bare and vegetated soil. *GPS Solut.* **2016**, *20*, 525–537. [[CrossRef](#)]
14. Wan, W.; Larson, K.M.; Small, E.E.; Chew, C.C.; Braun, J.J. Using geodetic GPS receivers to measure vegetation water content. *GPS Solut.* **2015**, *19*, 237–248. [[CrossRef](#)]
15. Small, E.E.; Larson, K.M.; Braun, J.J. Sensing vegetation growth with reflected GPS signals. *Geophys. Res. Lett.* **2010**, *37*, L12401. [[CrossRef](#)]
16. Small, E.E.; Larson, K.M.; Smith, W.K. Normalized Microwave Reflection Index: Validation of Vegetation Water Content Estimates from Montana Grasslands. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2014**, *7*, 1512–1521. [[CrossRef](#)]
17. Liang, Y.; Ren, C.; Wang, H.; Huang, Y.; Zheng, Z. Research on soil moisture inversion method based on GA-BP neural network model. *Int. J. Remote Sens.* **2019**, *40*, 2087–2103. [[CrossRef](#)]
18. Small, E.E.; Larson, K.M.; Chew, C.C.; Dong, J.; Ochsner, T.E. Validation of GPS-IR Soil Moisture Retrievals: Comparison of Different Algorithms to Remove Vegetation Effects. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2016**, *9*, 4759–4770. [[CrossRef](#)]
19. Ren, C.; Liang, Y.; Lu, X.; Yan, H. Research on the soil moisture sliding estimation method using the LS-SVM based on multi-satellite fusion. *Int. J. Remote Sens.* **2019**, *40*, 2104–2119. [[CrossRef](#)]
20. Lewis, P.A.W.; Stevens, J.G. Nonlinear Modelling of Time Series Using Multivariate Adaptive Regression Splines (MARS). *J. Am. Stat. Assoc.* **1991**, *86*, 864–877. [[CrossRef](#)]
21. Roggenbuck, O.; Reinking, J.; Lambertus, T. Determination of Significant Wave Heights Using Damping Coefficients of Attenuated GNSS SNR Data from Static and Kinematic Observations. *Remote Sens.* **2019**, *11*, 409. [[CrossRef](#)]
22. Bilich, A.; Larson, K.M. Mapping the GPS multipath environment using the signal-to-noise ratio (SNR). *Radio Sci.* **2007**, *42*, RS6003. [[CrossRef](#)]
23. Han, M.; Zhu, Y.; Yang, D.; Hong, X.; Song, S. A Semi-Empirical SNR Model for Soil Moisture Retrieval Using GNSS SNR Data. *Remote Sens.* **2018**, *10*, 280. [[CrossRef](#)]
24. 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](#)]
25. Camps, A.; Alonso-Arroyo, A.; Park, H.; Onrubia, R.; Pascual, D.; Querol, J. L-Band Vegetation Optical Depth Estimation Using Transmitted GNSS Signals: Application to GNSS-Reflectometry and Positioning. *Remote Sens.* **2020**, *12*, 2352. [[CrossRef](#)]
26. Evans, S.G.; Small, E.E.; Larson, K.M. Comparison of vegetation phenology in the western USA determined from reflected GPS microwave signals and NDVI. *Int. J. Remote Sens.* **2014**, *35*, 2996–3017. [[CrossRef](#)]
27. Leathwick, J.R.; Elith, J.; Hastie, T. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species. *Ecol. Model.* **2006**, *199*, 188–196. [[CrossRef](#)]
28. Jakub, S.; David, I. A Generalized Estimating Equation Approach to Multivariate Adaptive Regression Splines. *J. Comput. Graph. Stat.* **2018**, *27*, 245–253.
29. Hamidi, O.; Tapak, L.; Abbasi, H.; Maryanaji, Z. Application of random forest time series, support vector regression and multivariate adaptive regression splines models in prediction of snowfall (a case study of Alvand in the middle Zagros, Iran). *Theor Appl Climatol.* **2018**, *134*, 769–776. [[CrossRef](#)]
30. Jin, W.; Li, Z.J.; Wei, L.S.; Zhen, H. The Improvements of BP Neural Network Learning Algorithm. In Proceedings of the IEEE 5th International Conference on Signal Processing Proceedings, Rotorua, New Zealand, 21–25 August 2000; Volume 3, pp. 1647–1649.
31. Chen, P.H.; Lin, C.J.; Scholkopf, B. A tutorial on v-support vector machines. *Appl. Stoch. Models. Bus. Ind.* **2005**, *21*, 111–136. [[CrossRef](#)]