

## Article

# A Comprehensive Comparison of Machine Learning and Feature Selection Methods for Maize Biomass Estimation Using Sentinel-1 SAR, Sentinel-2 Vegetation Indices, and Biophysical Variables

Chi Xu <sup>1</sup>, Yanling Ding <sup>1,\*</sup>, Xingming Zheng <sup>2</sup>, Yeqiao Wang <sup>3</sup>, Rui Zhang <sup>4</sup>, Hongyan Zhang <sup>1</sup>, Zewen Dai <sup>1</sup> and Qiaoyun Xie <sup>5</sup>

- <sup>1</sup> Key Laboratory of Geographical Processes and Ecological Security in Changbai Mountains, Ministry of Education, School of Geographical Sciences, Northeast Normal University, Changchun 130024, China  
<sup>2</sup> Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China  
<sup>3</sup> Department of Natural Resources Science, University of Rhode Island, Kingston, RI 02881, USA  
<sup>4</sup> North Automatic Control Technology Institute, Taiyuan 030000, China  
<sup>5</sup> Faculty of Science, University of Technology Sydney, Sydney, NSW 2007, Australia  
\* Correspondence: dingyl720@nenu.edu.cn; Tel.: +86-15943072686

**Abstract:** Rapid and accurate estimation of maize biomass is critical for predicting crop productivity. The launched Sentinel-1 (S-1) synthetic aperture radar (SAR) and Sentinel-2 (S-2) missions offer a new opportunity to map biomass. The selection of appropriate response variables is crucial for improving the accuracy of biomass estimation. We developed models from SAR polarization indices, vegetation indices (VIs), and biophysical variables (BPVs) based on gaussian process regression (GPR) and random forest (RF) with feature optimization to retrieve maize biomass in Changchun, Jilin province, Northeastern China. Three new predictors from each type of remote sensing data were proposed based on the correlations to biomass measured in June, July, and August 2018. The results showed that a predictor combined by vertical-horizontal polarization (VV), vertical-horizontal polarization (VH), and the difference of VH and VV (VH-VV) derived from S-1 images of June, July, and August, respectively, with GPR and RF, provided a more accurate estimation of biomass ( $R^2 = 0.81\text{--}0.83$ , RMSE = 0.40–0.41 kg/m<sup>2</sup>) than the models based on single SAR polarization indices or their combinations, or optimized features ( $R^2 = 0.04\text{--}0.39$ , RMSE = 0.84–1.08 kg/m<sup>2</sup>). Among the S-2 VIs, the GPR model using a combination of ratio vegetation index (RVI) of June, normalized different infrared index (NDII) of July, and normalized difference vegetation index (NDVI) of August achieved a result with  $R^2 = 0.83$  and RMSE = 0.39 kg/m<sup>2</sup>, much better than single VIs or their combination, or optimized features ( $R^2$  of 0.31–0.77, RMSE of 0.47–0.87 kg/m<sup>2</sup>). A BPV predictor, combined with leaf chlorophyll content (CAB) in June, canopy water content (CWC) in July, and fractional vegetation cover (FCOVER) in August, with RF, also yielded the highest accuracy ( $R^2 = 0.85$ , RMSE = 0.38 kg/m<sup>2</sup>) compared to that of single BPVs or their combinations, or optimized subset. Overall, the three combined predictors were found to be significant contributors to improving the estimation accuracy of biomass with GPR and RF methods. This study clearly sheds new insights on the application of S-1 and S-2 data on maize biomass modeling.

**Keywords:** maize biomass; Sentinel-1; Sentinel-2; polarization indices; vegetation indices; biophysical variables; gaussian processes regression; random forest; feature optimization



**Citation:** Xu, C.; Ding, Y.; Zheng, X.; Wang, Y.; Zhang, R.; Zhang, H.; Dai, Z.; Xie, Q. A Comprehensive Comparison of Machine Learning and Feature Selection Methods for Maize Biomass Estimation Using Sentinel-1 SAR, Sentinel-2 Vegetation Indices, and Biophysical Variables. *Remote Sens.* **2022**, *14*, 4083. <https://doi.org/10.3390/rs14164083>

Academic Editor: Giuseppe Modica

Received: 25 June 2022

Accepted: 18 August 2022

Published: 20 August 2022

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



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Maize is an important, globally cultivated food and energy crop. The availability of information about maize development and health during the growing season is essential in optimizing crop production. Above-ground biomass is a key biophysical metric for

monitoring crop growth status and health conditions, as well as predicting crop yield. Spatially continuous crop biomass information plays a prominent role in the strategies for managing fertilizer application [1,2], disease control [3], yield forecasting [4], and greenhouse gas emissions [5].

Traditionally, crop biomass is commonly collected using destructive sampling in field measurements, which is the most accurate approach for estimating maize biomass. However, direct monitoring of maize biomass is time-consuming, labor-intensive, and difficult to conduct across large regions. It has become a cost-effective method for estimating biomass in large regions based on the correlations between field-measured biomass and remote sensing data [2,6]. The selection of appropriate response variables and algorithms is critical to obtaining accurate biomass estimation [7,8].

Optical (i.e., Sentinel-2, S-2) and Synthetic Aperture Radar (i.e., Sentinel-1, S-1) remote sensing data from the European Commission's Copernicus program have been frequently used for biomass estimation [9–11]. Spectral information from optical data can efficiently reflect the development of plants. The amount of chlorophyll and the canopy structure of the maize crop are significantly correlated with the reflectance responses of vegetation in the red and near-infrared regions of the spectrum, respectively. Data acquired from these two regions has been widely used to create spectral transformations known as vegetation indices (VIs) [12]. Multiple types of VIs derived have been demonstrated to have a substantial relationship with biomass [13,14]. The normalized difference vegetation index (NDVI), the most extensively used index, is highly sensitive to low biomass [15]. Ratio vegetation index (RVI) and enhanced vegetation index (EVI) have been shown to be more correlated to high biomass [16]. VIs that incorporate red-edge bands also have a greater potential for accurately estimating biomass [17]. In addition to VIs, biophysical variables (BPVs) such as leaf area index (LAI) provide new capabilities for monitoring biomass. Castillo et al. [9] compared S-2 VIs and BPVs to estimate mangrove forest biomass and discovered that both LAI and fractional vegetation cover (FCOVER) outperformed NDVI.

SAR data shows an overwhelming potential for crop monitoring due to its capacity to obtain high-quality images in all weather conditions and penetrate the crop canopy to capture leaf and stem information [18]. The ability of the S-1 backscatter coefficient to estimate biomass has been demonstrated in recent investigations [19]. Wang et al. [11] employed vertical-horizontal polarization (VV) and vertical-horizontal polarization (VH) to estimate pasture biomass. Ndikumana et al. [20] showed that the correlation of S-1 signals to the forest biomass was high in VH polarization. In comparison with forest and grassland ecosystems, only limited attempts have been made to retrieve maize biomass using S-1 and S-2 datasets.

Machine learning (ML) algorithms such as gaussian process regression (GPR), support vector machine (SVM), random forest (RF), and artificial neural network (ANN), have been increasingly utilized to estimate biomass [21]. Among a variety of ML algorithms, GPR and RF algorithms have been regarded as one of the best methods for classification and regression because of their ability to capture complex non-linear relationships quickly, accurately, and automatically [22–24]. Alebele et al. [2] demonstrated the potential of integrating S-1 and S-2 data based on the GPR algorithm to retrieve rice biomass. Jachowski et al. [23] reported that the GPR model with VIs provided a more accurate result than linear regression in estimating the above-ground biomass of mangroves. Pandit et al. [25] applied S-2 VIs combined with an RF algorithm to estimate forest biomass. Forkuor et al. [10] made use of the integrated derivatives from S-1 and S-2 data with an RF algorithm to map biomass in West African dryland forests. Chen et al. [26] compared four models to estimate forest biomass in combination with S-1 and S-2 data and concluded that RF was superior to other models.

Feature selection plays a critical role in ML algorithms, which helps in removing irrelevant, redundant, and noisy features, avoiding significant loss of information, reducing computation requirements, and therefore improving the performance of ML [27–30]. Some ML algorithms include built-in feature selection methods [31]. Karlson et al. [32] used a

recursive feature elimination (RFE) method embedded in RF to assess its effect on the predictive performance of RF, and they found that variable selection improved the prediction of above-ground biomass. The kernel function of GPR  $\sigma$  is a dedicated parameter that controls the spread of the relationships for each predictor variable [33]. Verrelst et al. [34] determined the optimal number of predictor variables by iteratively removing the predictor variable with the lowest  $\sigma$  until only one variable remained for estimating LAI.

During different growth stages, VIs have different connections with crop biomass due to dynamic changes in canopy reflectance [35]. The dynamic changes may be affected by canopy biophysical properties such as the number of leaves per area, canopy biochemical features (chlorophylls and carotenoids), crop healthy conditions, and other factors [36]. Gnyp et al. [37] discovered that RVI had the strongest and weakest relationships with rice biomass at the early growth stage (tillering) and the later growth stage (booting), respectively, whereas it was less sensitive to biomass at the middle stage (elongation). Besides, some studies indicated that the growth stage of the crop also had an impact on the performance of radar backscatter coefficients and BPVs in estimating biomass [38,39]. Li et al. [40] indicated that the ratio of VV and VH had a strong correlation with rice biomass at the transplanting stage, while VH and horizontal transmit-horizontal (HH) showed the highest sensitivities to biomass at the tillering stage and the heading stage, respectively. Li et al. [41] found that LAI was highly related to maize biomass during the entire growing season, while leaf chlorophyll content (CAB) and FCOVER had the strongest correlations with maize biomass at the seedling stage, but they were not sensitive to biomass at the filling and tasseling stages, respectively. The responses of remote sensing derivatives to biomass vary in different growth stages. Thus, it is worth further exploration to improve the retrieval of biomass according to the sensitivity of each variable to biomass in different growth stages.

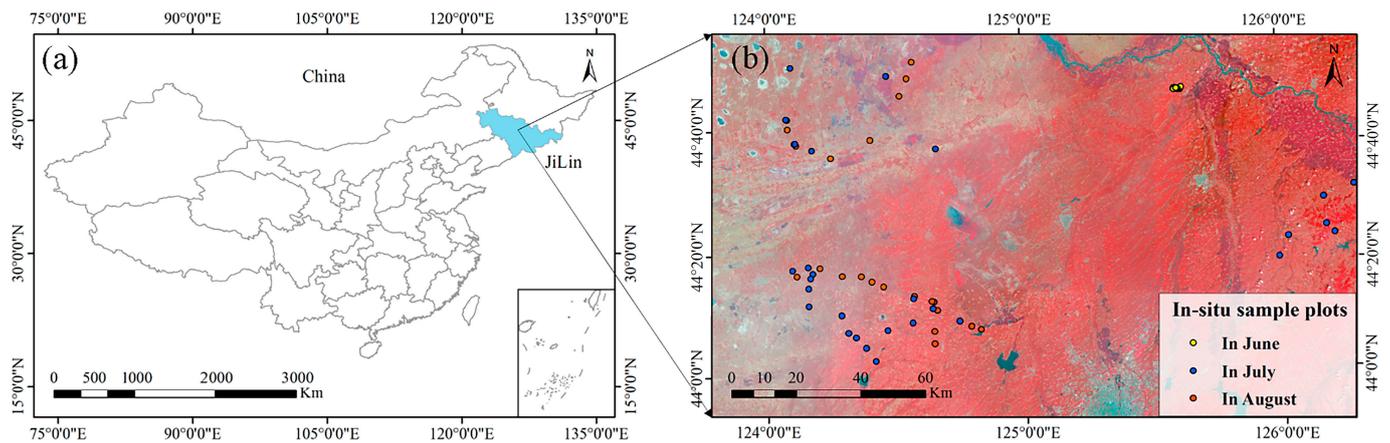
In this study, we evaluate the ability of multi-temporal S-1 and S-2 data to estimate maize biomass and explore improving the accuracy of biomass estimation. The specific objectives are the followings: (1) compare the accuracies of S-1 and S-2 derivatives, individually, to estimate maize biomass; (2) determine the optimal features optimized by GPR and RF for biomass modeling; (3) propose three combined features based on the correlations of S-1 and S-2 derivatives to biomass in different growth stages for improving biomass retrieval, respectively.

## 2. Materials and Methods

### 2.1. Study Area and Field Data

The study area (43°5′N–45°15′N, 124°18′E–127°2′E) was situated in Changchun, Jilin province, Northeastern China (Figure 1). The region is situated in a temperate continental climate zone with four seasons, characterized by a hot and rainy summer and a cold and dry winter. The average annual mean air temperature and precipitation of the study area are 4.6 °C and 520 mm, respectively. The proportion of farmland in this area is over 90%. Maize is the main crop in this area, which is harvested once a year.

During the maize growing season, three field campaigns were conducted on 23 June, 20 and 22 July, and 9 and 10 August 2018 to collect maize above-ground dry biomass. Each sample point was selected in the center of a quadrat of 10 × 10 m based on the remote sensing image, and the location of each sample point was measured by GPS. According to the GPS recorded coordinates, the remote sensing data for each sample point was extracted using the “Extract Multi Values to Points” in the ArcGIS software. The unit area of single maize plant was calculated by maize column and row spacings. At each sample location, maize row spacings were measured three times and averaged, and column spacings were averaged by the distance of ten consecutive maize plants. In this study, the unit area maize varied from 0.07 m<sup>2</sup> to 0.32 m<sup>2</sup>.



**Figure 1.** Study area (a) and the location of sampling points (b). The image in display is false color composite of Sentinel-2 image acquired on 23 June 2018.

Within each sample point, three randomly selected plants were horizontally cut to the root. Each of three maize biomass samples was partitioned into stems, leaves, and fruits with each component processed separately and put into the bags. The fresh samples were dried in an oven at 70 °C for 72 h and weighed in dry conditions to obtain the dry weight of maize biomass samples. The total above-ground biomass of each sample point was calculated by dry weight and the unit area per maize plant. The biomass of the three maize plants was averaged and considered as the representative value of the dry biomass in this plot. In this study, the range of the measured dry biomass was between 0.02 and 4.24 kg/m<sup>2</sup>. Details on the number and location of 85 samples used in this study are given in Table 1 and Figure 1.

**Table 1.** Details of Sentinel-1, Sentinel-2 images, and field samples acquired for the study.

Sentinel-1 Acquisition Date	Product Type	Sentinel-2 Acquisition Date	Product Type	Field Acquisition Date	Sample Points
23 June 2018	GRD	23 June 2018	Level-1C	23 June 2018	30
22 July 2018	GRD	23 July 2018	Level-1C	20 July 2018, 22 July 2018	34
10 August 2018	GRD	2 August 2018	Level-1C	9 August 2018, 10 August 2018	21

## 2.2. Satellite Data Pre-Processing and Derived Variables

This study used data from S-1 and S-2 imagery of the European Space Agency acquired from ESA's Copernicus Open Access (<https://scihub.copernicus.eu/dhus/#/home>, accessed on 15 June 2021). The data acquisition time consistently matched with the dates of field campaigns. The information on Sentinel images used for the study is presented in Table 1.

The acquired S-1 C-band (5.405 GHz) data were collected in the interferometric wide swath (IW) mode with VV and VH dual polarizations, and in high-resolution Level-1 ground range detected (GRD) processing level with a pixel size of 10 m. The acquired SAR data was preprocessed by sentinel application platform (SNAP) software. The data processing steps consisted of orbit calibration, thermal noise removal, noise removal, image calibration, speckle filtering, and terrain correction [42]. The digital number (DN) of SAR images was transformed to radar intensity backscatter coefficient ( $\sigma^0$ ) using log scaling.

The acquired cloudless S-2 data was an orthorectified, top-of-atmosphere reflectance (Level-1C), with 13 spectral bands in the visible, near-infrared, and short-wave infrared regions. As the Level-1C product had been processed for radiometric and geometric corrections, S-2 images were only atmospherically corrected and converted to Level-2A

products using the Sen2cor atmospheric correction toolbox of SNAP software. In order to keep the spatial resolution consistent, the pre-processed S-2 images with 20 and 60 m resolution were resampled to 10 m spatial resolution by using the nearest neighbor method of SNAP software.

Three groups of predictors presented in Table 2 were extracted from Sentinel images. In total, we selected 23 predictors to test their performance in estimating biomass. The first group of predictors consisted of VH and VV polarization indices. In addition, the difference (VH – VV) and sum (VH + VV) of VH and VV were computed, which were considered as quotient products and used to estimate biomass [43]. We also calculated four different combinations of VH and VV, such as VH × VV and VH/(VH × VV). A total of eight SAR polarization indices were applied to estimate maize biomass.

The second group involved ten VIs computed from S-2 10 m multispectral bands, including six traditional VIs calculated from red and near-infrared (NIR) bands (e.g., NDVI and RVI) and four red-edge indices, which were normalized difference red-edge index (NDRE), red-edge simple ratio vegetation index (RERVI), red-edge chlorophyll index (CIre) and red-edge re-normalized difference vegetation index (RERDVI). These VIs are widely used to estimate vegetation parameters [44].

The last group included LAI, FCOVER, FAPAR, CAB, and canopy water content (CWC), which were calculated using the “Biophysical Processor” in the SNAP software. Previous studies have confirmed the performance of SNAP-derived biophysical variables is applicable for crop parameter retrieval [44–46]. Kamenova et al. [47] reported that the measured values and the SNAP-derived estimates for three BPVs (LAI, FCOVER, and FAPAR) were highly associated ( $R^2 > 0.89$ ). The principle of “Biophysical Processor” is to retrieve these parameters from Sentinel-2 instantaneous observations using neural network approach. This process consists mainly of the following three steps: (1) generating training database should be constituted of a representative set of Sentinel-2 top of canopy reflectance and observation geometry data obtained by the PROSPECT + SAIL radiative transfer model; (2) training the neural network architecture, the steps consisting of normalization of the input, network architecture, denormalization of the output; (3) generation of quality indicator [48].

**Table 2.** List of Sentinel-1 and Sentinel-2 predictors used for maize biomass modeling.

Indices	Variables	Definition	Reference
S-1 polarization indices	Vertical transmit-vertical channel	VV	—
	Vertical transmit-horizontal channel	VH	—
	SAR simple additive index	VH + VV	[43]
	SAR simple difference index	VH – VV	[43]
	SAR multiplication index	VH × VV	This paper
	SAR ratio index	VH/(VH × VV)	This paper
	SAR ratio index	(VH + VV)/(VH × VV)	This paper
	SAR square difference index	VH × VH – VV × VV	This paper
S-2 VIs	Normalized difference vegetation index (NDVI)	$(B8a - B4)/(B8a + B4)$	[49]
	Enhanced vegetation index (EVI)	$2.5 \times (B8a - B4)/(B8a + 6 \times B4 - 7.5 \times B2 + 1)$	[50]
	Ratio vegetation index (RVI)	$B8a/B4$	[51]
	Normalized difference infrared Index (NDII)	$(B11 - B4)/(B11 + B4)$	[52]
	Modified simple ratio (MSR)	$((B8a/B4) - 1)/\sqrt{(B8a/B4) + 1}$	[53]
	Soil adjusted vegetation index (SAVI)	$(1 + 0.5) \times (B8a - B4)/(B8a + B4 + 0.5)$	[54]

Table 2. Cont.

Indices	Variables	Definition	Reference
	Normalized difference red-edge Index (NDRE)	$(B8a - B5)/(B8a + B5)$	[55]
	Red-edge simple ratio vegetation Index (RERVI)	$B8a/B5$	[56]
	Red-edge chlorophyll index (CI <sub>re</sub> )	$B8a/B5 - 1$	[57]
	Red-edge re-normalized difference vegetation index (RERDVI)	$(B8a - B5)/\sqrt{B8a/B5}$	[58]
S-2 BPVs	Leaf area index	LAI	—
	Fractional vegetation cover	FCOVER	
	Fraction of absorbed photo-synthetically active radiation	FAPAR	
	Leaf chlorophyll content	CAB	
	Canopy water content	CWC	

Note: S-2 multispectral bands setting: B2 (blue, 490 nm), B4 (red, 665 nm), B5 (red-edge, 705 nm), B8a (near infrared, 865 nm), and B11 (shortwave infrared, 1610 nm).

### 2.3. Maize Biomass Modeling and Feature Selection

#### 2.3.1. Gaussian Process Regression and Feature Selection

Gaussian process regression (GPR) is a kernel-based machine learning algorithm. A Gaussian process assigns probability distribution over a set of possible functions that fit the input data and converts them into posterior probabilistic estimates [59]. A non-parametric Gaussian process model is specified as follows:

$$p(f(x)|\theta) \sim gp(0, k(x, x')) + I\sigma_y^2 \quad (1)$$

where  $x$  is the input predictors,  $k(x, x')$  is a kernel matrix to approximate covariance function, which can be implemented with a variety of functions and  $\sigma_y^2$  is Gaussian noise.

The model hyperparameters were automatically optimized using a “fitgp” function in Matlab R2019b. A general introduction to optimizing the hyperparameters of GPR algorithms can be found [60]. In this study, we used a squared exponential kernel. One of the advantages of GPR is that the predictive power of each predictor can be evaluated for the parameter of interest. The importance of input predictors can be interpreted by  $\sigma$ , which is a parameter of the covariance function of GPR as follows:  $k(x, x') = \exp\left(-\frac{\|x, x'\|^2}{2\sigma^2}\right)$ . High values of  $\sigma$  indicate that relations mostly extend along that predictor, hence, the lower the  $\sigma$ , the more relevant the predictor [33]. As such, the optimal number of input predictors was assessed by excluding the least important predictors according to the relevance of each variable to biomass. We used a stepwise elimination method to identify the optimal input combination in such a way to reduce the number of input variables, beginning at the variable with the highest  $\sigma$  and ending up with the combination that provided the lowest root mean squared error (RMSE).

#### 2.3.2. Random Forest and Feature Selection

The RF is an ensemble-learning algorithm that combines a large number of decision trees to improve prediction accuracy [61]. In random forest regression, each tree is built using a deterministic algorithm by selecting a random set of variables and a random sample from the training dataset [62]. In order to implement RF, the number of regression trees (ntree) and different predictors selected at each leaf node (mtry) need to be optimized [63]. In this study, the ntree values were tested from 50 to 200 at a step of 50, as well as mtry values were tested from 5 to 100. Two parameters were optimized using the grid search method in the “caret” package in R 4.1.2 software.

Feature selection with RF is achieved by the recursive feature elimination (RFE), which is a well-known wrapper-based feature-ranking method that searches within the space for the optimal subset provided by the caret package in R software [64,65]. The process iteratively calculates the importance of all variables added to the RF model and removes the least important variables until only one variable remained in the model. The importance of input predictors can be interpreted by mean decrease gini (IncNodePurity) and the root mean square error of (RMSE). IncNodePurity evaluates the quality of a split for every variable (node) of a tree by means of the gini index, a higher IncNodePurity value represents higher variable importance. RMSE is constructed by permuting the values of each variable of the test set, recording the prediction and comparing it with the unpermuted test set prediction of the variable [66]. In this study, the model was optimized by selecting best mtry and ntree, the smallest subset of variables with lowest RMSE was selected to predict biomass.

### 2.3.3. Three New Predictors Proposed for Biomass Retrieval

Pearson's correlation coefficient (R) was calculated to assess the sensitivity of each variable in Table 2 to biomass measured in June, July, and August, respectively. The higher value represented stronger correlation between the variable and biomass. As for each type of remote sensing data, the new variables were proposed by combining the first three variables, which were the most correlated to biomass collected in the three months, respectively, and used to estimate biomass with GRP and RF.

GPR and RF methods were employed for biomass modeling. Firstly, the univariate models were developed based on single S-1 SAR polarization indices, S-2 VIs and BPVs to estimate biomass, and we evaluated the ability of single S-1 and S-2 data to estimate maize biomass. Then, we developed the models from SAR indices, VIs, and BPVs based on GPR and RF with feature optimization to improve the accuracy of biomass estimation. Finally, three integrated predictors were proposed to estimate biomass by combining the derivatives of S-1 and S-2, which were the most sensitive to biomass measured in June, July, and August of 2018, respectively. Three integrated predictors were combined with GPR and RF to explore the possibility of further improving biomass estimation. The overall methodological flowchart for estimating biomass from S-1 to S-2 images is presented in Figure 2.

### 2.3.4. Model Calibration and Validation

To explore the potential of different datasets to monitor maize biomass, eight SAR polarization indices, ten VIs, and five BPVs were used individually and integrally with GRP and RF. The performances of the above biomass estimation models were evaluated using the coefficient of determination ( $R^2$ ), RMSE, and the ratio of percent deviation (RPD), three statistical criteria for each algorithm were the average of 5-fold cross-validation repeated 50 times, which were calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - M)^2} \quad (2)$$

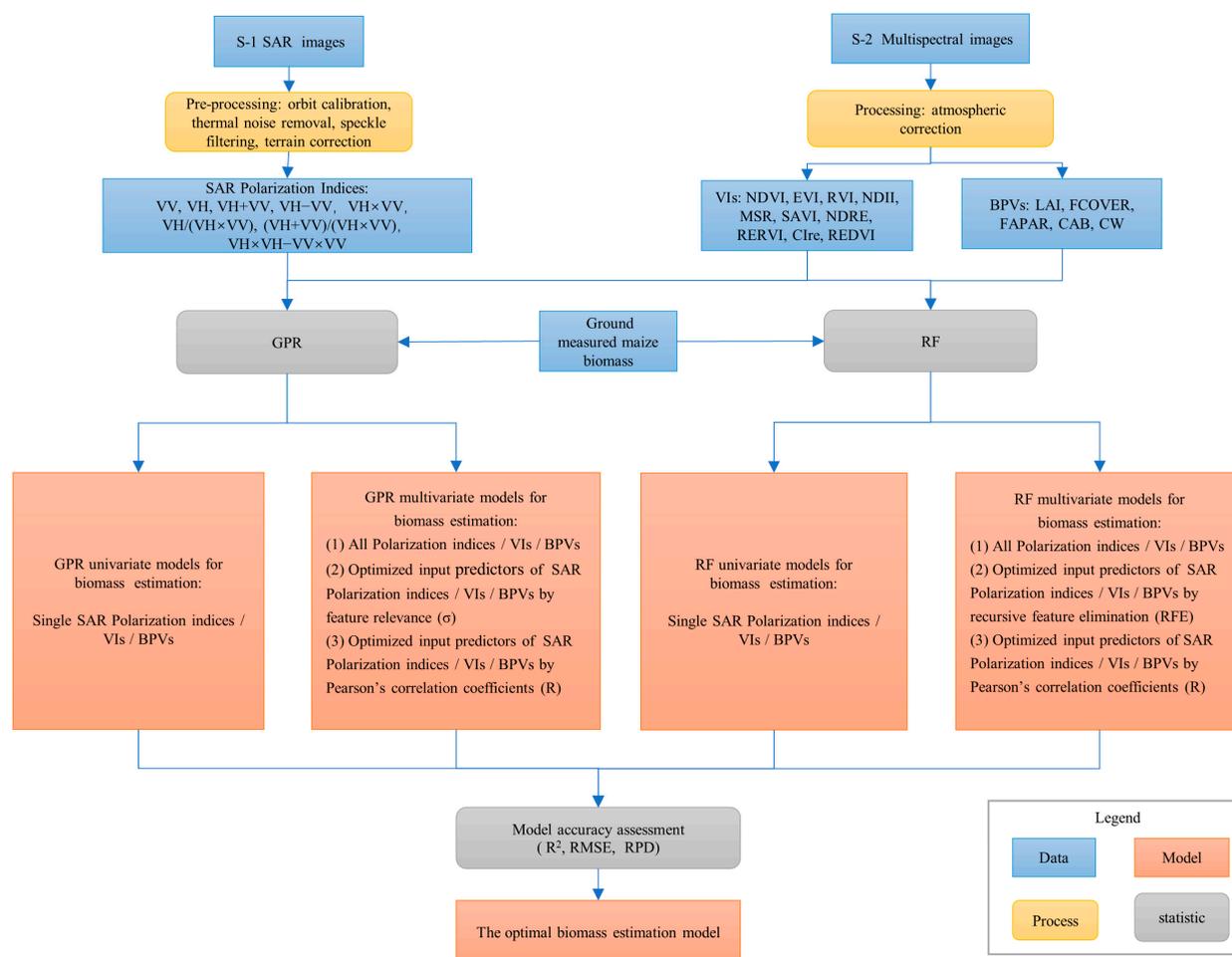
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (3)$$

$$\text{RPD} = \text{SD}/\text{RMSE} \quad (4)$$

where " $P$ " is the predicted value, " $O$ " is the observed value, " $M$ " is the mean of observed values. SD is the standard deviation of observed values. The quality of estimation is assessed based on RPD as follows: very poor (<1.0), poor (1.0–1.4), acceptable (1.4–1.8), good (1.8–2.0), very good (2.0–2.5), and excellent (>2.5) [67,68].

In order to avoid reliance on a single random split of the datasets, as well as to guarantee that all samples were used for both training and validation, we used a repeated

5-fold cross-validation procedure. All samples were randomly split into 5 equal-sized sub-datasets, and they were trained and tested 10 times. For each time, 4 sub-datasets were used iteratively for calibration and the remaining sub-dataset were used for validation. By repeating the training procedure 5 times, all observations were used for both calibration and validation, with each observation being used for validation only once.



**Figure 2.** Flowchart for maize biomass estimation from S-1 and S-2 data.

### 3. Results

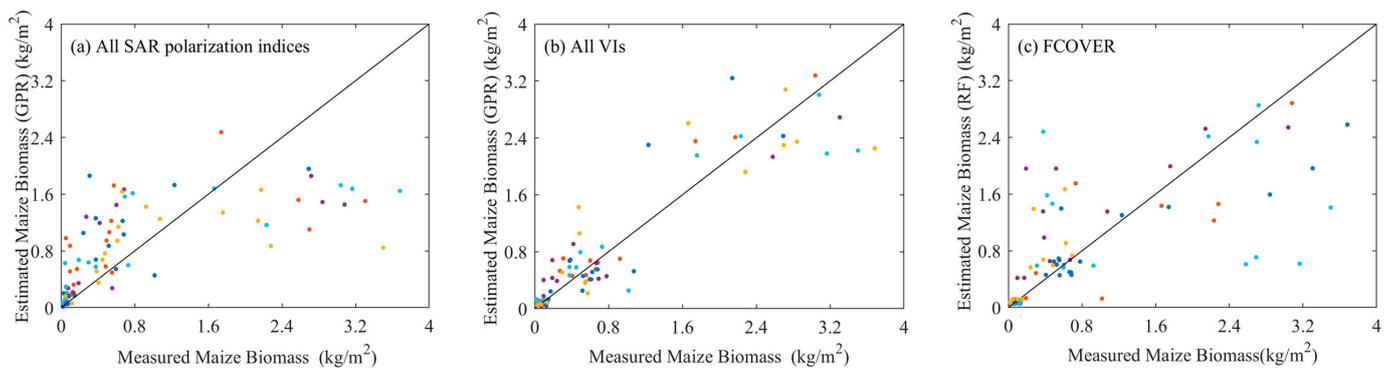
#### 3.1. Performance of Each S-1 SAR Polarization Indices, S-2 VIs, and BPVs on Estimating Maize Biomass with GPR and RF

The performance of SAR polarization indices on estimating biomass with GPR and RF is presented in Table 3. In terms of the GPR approach, among the eight indices, the highest accuracy was obtained by VH + VV ( $R^2 = 0.36$ ,  $RMSE = 0.82$  kg/m<sup>2</sup>,  $RPD = 1.30$ ). Testing of three derived indices of  $VH \times VV$ ,  $VH/(VH \times VV)$ , and  $(VH + VV)/(VH \times VV)$  in this study provided new information on the use of S-1 data for biomass modelling, which had a marginal advantage over VH and VV.  $VH - VV$  and  $VH \times VH - VV \times VV$  presented unreliable results with extremely low  $R^2$ . A slight improvement was achieved by using all eight polarization indices as input predictors ( $R^2 = 0.39$ ,  $RMSE = 0.84$  kg/m<sup>2</sup>). In the case of RF, VH + VV was also the best predictor, which yielded the highest accuracy ( $R^2 = 0.41$  and  $RMSE = 0.85$  kg/m<sup>2</sup>). However, the scatterplot of measured biomass and estimated biomass by all SAR polarization indices demonstrated that samples with biomass below 1.2 kg/m<sup>2</sup> were significantly overestimated and those higher than 1.2 kg/m<sup>2</sup> were underestimated (Figure 3a). Other polarization indices with RF performed similarly to

GPR. SAR polarization indices and their combinations with GPR and RF produced poor predictions of maize biomass ( $PRD < 1.4$ ).

**Table 3.** Performance of S-1 SAR polarization indices on estimating maize biomass based on GPR and RF.

Input Variables	GPR			RF		
	R <sup>2</sup>	RMSE (kg/m <sup>2</sup> )	RPD	R <sup>2</sup>	RMSE (kg/m <sup>2</sup> )	RPD
VH	0.32	0.87	1.23	0.30	0.96	1.17
VV	0.31	0.88	1.23	0.25	1.01	1.10
VH + VV	0.36	0.82	1.30	0.41	0.85	1.35
VH − VV	0.04	1.08	0.98	0.02	1.23	0.89
VH × VV	0.35	0.84	1.27	0.30	0.94	1.20
VH/(VH × VV)	0.31	0.89	1.21	0.28	1.01	1.11
(VH + VV)/(VH × VV)	0.34	0.85	1.25	0.26	1.02	1.09
VH × VH − VV × VV	0.20	0.99	1.06	0.28	0.99	1.13
All SAR	0.39	0.84	1.27	0.31	0.92	1.21



**Figure 3.** Measured vs. estimated biomass by the optimal models for each dataset (a) All SAR polarization indices with GPR, (b) all VIs with GPR, and (c) FCOVER with RF. The different colors indicated the 5-fold subsets.

The models based on published S-2 VIs yielded various retrieval results (Table 4). Generally, GPR and RF models with S-2 VIs produced better accuracy statistics than S-1 SAR polarization indices. Among the ten univariate models based on GPR, the highest retrieval accuracies were achieved by RVI and MSR with the same statistical values ( $R^2 = 0.65$ ,  $RMSE = 0.59$  kg/m<sup>2</sup>,  $RPD = 1.93$ ), followed by NDVI. Other VIs performed poorly, with low  $R^2$  values. Additionally, the retrieval accuracy was highly improved by using all VIs as input features, with an  $R^2$  of 0.77 and a  $RMSE$  of 0.47 kg/m<sup>2</sup>. The scatter plots show that the estimated versus measured values fall close to the 1:1 line (Figure 3b). However, this model had difficulty estimating higher biomass quantities ( $>1.2$  kg/m<sup>2</sup>). This can be explained by the fact that VIs, particularly those based on red and near-infrared bands, approach saturation level after a certain biomass density [69]. Compared with the univariate models, the combination of all VIs enriched the effective information and explained more variability for biomass estimation. Similar to RF models, RVI, MSR, and NDVI had the most significant contributions to predicting biomass, with  $R^2$  around 0.55 and  $RMSE$  at about 0.70 kg/m<sup>2</sup>, and the contributions of the other VIs were less significant. The accuracy was further improved by using all VIs as predictors, resulting in an  $R^2 = 0.73$  and  $RMSE = 0.53$  kg/m<sup>2</sup>. In comparison, the GPR models performed better than RF models for biomass estimation.

**Table 4.** Performance of S-2 VIs on estimating maize biomass based on GPR and RF.

Input Variables	GPR			RF		
	R <sup>2</sup>	RMSE (kg/m <sup>2</sup> )	RPD	R <sup>2</sup>	RMSE (kg/m <sup>2</sup> )	RPD
NDVI	0.64	0.60	1.86	0.54	0.71	1.66
EVI	0.41	0.81	1.32	0.34	0.93	1.20
RVI	0.65	0.59	1.93	0.55	0.71	1.65
NDII	0.35	0.85	1.26	0.27	0.99	1.13
MSR	0.65	0.59	1.92	0.56	0.70	1.70
SAVI	0.31	0.87	1.22	0.36	0.91	1.22
NDRE	0.40	0.79	1.33	0.37	0.85	1.31
RERVI	0.44	0.77	1.40	0.38	0.86	1.30
CIre	0.44	0.77	1.39	0.39	0.84	1.32
RERDVI	0.31	0.88	1.20	0.26	0.98	1.13
All VIs	0.77	0.47	2.42	0.73	0.53	2.28

Of the five BPVs investigated, FCOVER with GPR led to the most accurate estimates of biomass ( $R^2 = 0.44$ ,  $RMSE = 0.78$  kg/m<sup>2</sup>,  $RPD = 1.36$ ), followed by CWC and FAPAR with  $R^2$  around 0.35 (Table 5). CAB and LAI produced similar estimates of biomass. Compared to utilizing a single BPV as a predictor, the application of all five BPVs improved the performance of GPR with an  $R^2$  of 0.53 and a  $RMSE$  of 0.76 kg/m<sup>2</sup>. In terms of RF, the highest retrieval accuracy was also achieved by FCOVER, with  $R^2$  of 0.58 and a  $RMSE$  of 0.68 kg/m<sup>2</sup>, and the scatter plots in Figure 3c were similar to the scatterplot of measured biomass and estimated biomass by the GPR model of all SAR polarization indices (Figure 3a). It was noted that samples with biomass values below 1.2 kg/m<sup>2</sup> were overestimated and those higher than 1.2 kg/m<sup>2</sup> were underestimated. The other BPVs presented unstable estimates of biomass with low  $R^2$ . Compared to the RF model based on FCOVER, no improvement was achieved by the application of all BPVs with RF.

**Table 5.** Performance of S-2 vegetation biophysical variables on estimating maize biomass based on GPR and RF.

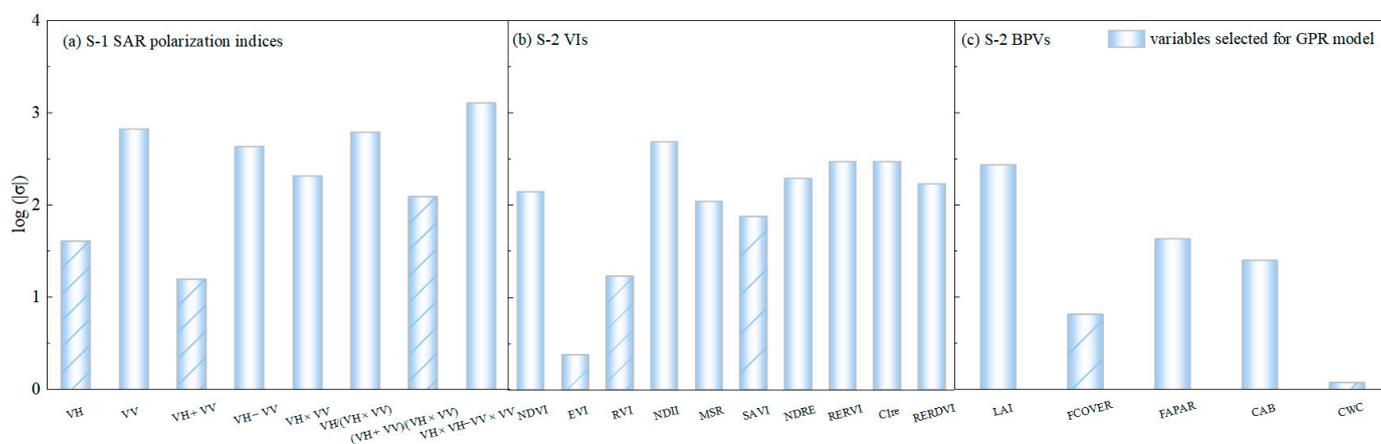
Input Variables	GPR			RF		
	R <sup>2</sup>	RMSE (kg/m <sup>2</sup> )	RPD	R <sup>2</sup>	RMSE (kg/m <sup>2</sup> )	RPD
LAI	0.34	0.86	1.24	0.23	1.02	1.09
FCOVER	0.44	0.78	1.36	0.58	0.68	1.70
FAPAR	0.36	0.86	1.23	0.29	0.96	1.16
CWC	0.38	0.82	1.29	0.33	0.91	1.21
CAB	0.32	0.87	1.22	0.17	1.08	1.02
All BPVs	0.53	0.76	1.50	0.46	0.77	1.45

### 3.2. Performance of GPR and RF on Estimating Maize Biomass with Feature Optimization

#### 3.2.1. Performance of GPR-Optimized by Feature Relevance

As noted earlier, one interesting feature of GPR is its ability to provide insight into the relevance of input predictors when developing the regression model. The  $\sigma$  in the kernel function of GPR is interpreted as the relevance of the predictor, which means the lower  $\sigma$ , the more relevant the predictor [70]. We calculated  $\sigma$  for each group of input predictors and illustrated it in Figure 4. The plot of  $\sigma$  for the eight SAR polarization indices (Figure 4a) showed that the most relevant indices were  $(VH + VV)/(VH \times VV)$ ,  $VH + VV$ , and  $VH$ . GPR models associated with these three indices have been proven powerful in estimating biomass (Table 3). Calculation of  $\sigma$  values for the ten VIs (Figure 4b) showed that EVI, RVI, and SAVI were more relevant to biomass than the other VIs. RVI with GPR has been shown to outperform other univariate models (Table 4). As for the five BPVs, CWC and FCOVER

yielded high correlations to biomass (Figure 4c). FCOVER as input also provided more accurate predictions of biomass than other BPVs (Table 5).



**Figure 4.** The logarithmic  $\sigma$  of GPR models with different groups of input predictors.

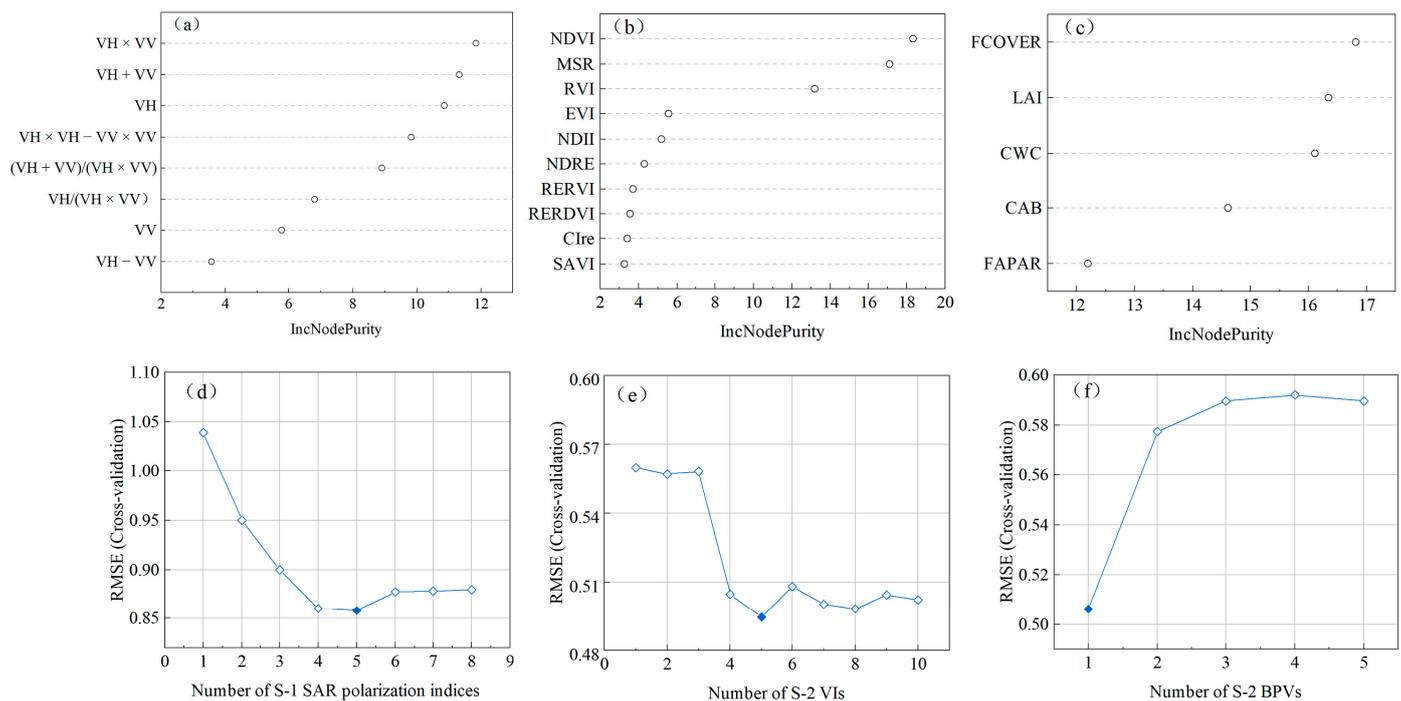
The slash-filled bars in Figure 4 were the final input variables for each group of data optimized by the stepwise elimination method. The accuracies retrieved by GPR models with the optimal input variables are listed in Table 6.  $(VH + VV)/(VH \times VV)$ ,  $VH + VV$ , and  $VH$  were selected as the most important input variables among the S-1 data, this selection did not yield more accurate estimates than all SAR polarization indices. The optimal variables among VIs were EVI, RVI, and SAVI. The GPR associated with three variables outperformed the models based on all VIs, enhancing  $R^2$  from 0.77 to 0.80 and RMSE from  $0.47 \text{ kg/m}^2$  to  $0.43 \text{ kg/m}^2$ . Compared to the original five BPVs, the RMSE was improved from  $0.76 \text{ kg/m}^2$  to  $0.69 \text{ kg/m}^2$  by using CWC and FCOVER as input variables. In general, the retrieval accuracy of GPR was improved by the stepwise elimination method based on the  $\sigma$ .

**Table 6.** Performance of the optimized GPR models on estimating maize biomass.

Optimized Input Predictors	GPR		
	$R^2$	RMSE ( $\text{kg/m}^2$ )	RPD
$(VH + VV)/(VH \times VV)$ , $VH + VV$ , $VH$	0.40	0.84	1.29
EVI, RVI, SAVI	0.80	0.43	2.68
CWC, FCOVER	0.57	0.69	1.62

### 3.2.2. Performance of RF-Optimized by RFE

The RF by combining an RFE based on predictor importance ranking was optimized. The importance of S-1 SAR polarization indices, S-2 VIs, and BPVs for maize biomass modeling is shown in Figure 5a–c. The effect of the number of variables on the RMSE for the biomass models is illustrated in Figure 5d–e. For S-1 SAR polarization indices, the top five important variables for biomass modeling were  $VH + VV$ ,  $VH \times VV$ ,  $(VH + VV)/(VH \times VV)$ ,  $VH$ , and  $VH \times VH-VV \times VV$  (Figure 5a) with the lowest RMSE (Figure 5d). As for S-2 VIs, a set of five variables including NDII, MSR, NDVI, RVI, and EVI showed the lowest RMSE for biomass prediction (Figure 5b,e). In terms of S-2 BPVs, the minimum RMSE was obtained by only using FCOVER (Figure 5c,f).



**Figure 5.** Random forest-based variable importance for (a) S-1 SAR polarization indices, (b) S-2 VIs, and (c) BPVs; selection of optimum number of variables based on the least RMSE for 5-fold cross-validation for (d) S-1 SAR polarization indices, (e) S-2 VIs, and (f) BPVs.

The performance of three groups of the optimized predictors for biomass estimation is listed in Table 7. The five RFE-optimized indices yielded similar accuracy ( $R^2 = 0.32$ ,  $RMSE = 0.91 \text{ kg/m}^2$ ) to the predictions by the RF model of SAR predictors individually and integrally (Table 3), failing to enhance the accuracy. Compared to the RF model based on all VIs, no improvement was achieved by using NDII, MSR, NDVI, RVI, and EVI. Among the five BPVs, the most important variable for biomass prediction was FCOVER, which resulted in  $R^2 = 0.58$  and  $RMSE = 0.68 \text{ kg/m}^2$ . Although RF with RFE did not enhance biomass prediction accuracy, it did minimize the number of input variables.

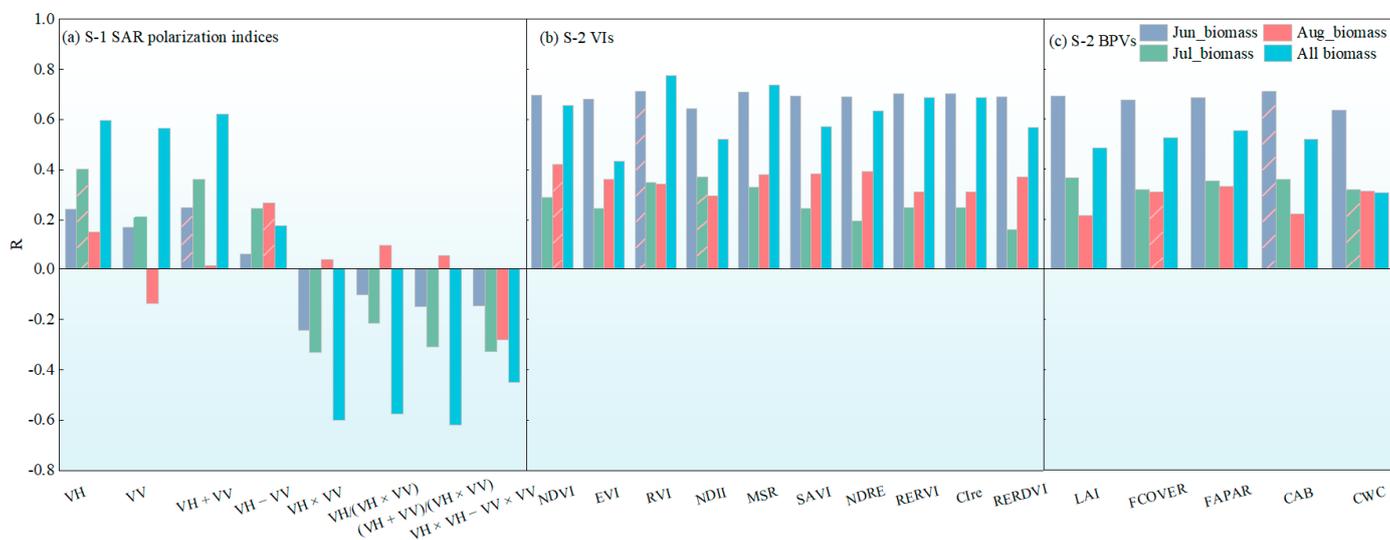
**Table 7.** Performance of the optimized RF models on estimating maize biomass.

Optimized Input Predictors	RF		
	$R^2$	RMSE ( $\text{kg/m}^2$ )	RPD
$VH + VV$ , $VH \times VV$ , $(VH + VV)/(VH \times VV)$ , $VH$ , $VH \times VH - VV \times VV$	0.32	0.91	1.24
NDII, MSR, NDVI, RVI, EVI	0.74	0.52	2.29
FCOVER	0.58	0.68	1.70

### 3.2.3. Performance of GPR and RF with New Features

The Pearson's correlation coefficients between derivatives of S-1 and S-2 and biomass measured in June, July, and August are presented in Figure 6. SAR polarization indices showed different degrees of correlation with biomass collected in different periods (Figure 6a). The  $VH + VV$  polarization yielded the highest correlation with biomass collected in June ( $R = 0.25$ ). The  $VH$  channel was more sensitive to biomass measured in July than the other indicators ( $R = 0.40$ ). The difference between  $VH$  and  $VV$  yielded the strongest correlation with August biomass ( $R = 0.27$ ). Therefore, the combination of the June  $VV + VH$  (Jun\_  $(VV + VH)$ ), derived from the S-1 image acquired on 23 June 2018), the July  $VH$  (Jul\_  $VH$ , derived from the S-1 image acquired on 22 July 2018), and the August

VH – VV (Aug\_(VH – VV)), derived from S-1 image acquired on 10 August 2018) can be a potential predictor to estimate maize biomass. Thus, Jun\_(VV + VH), Jul\_VH, and Aug\_(VH – VV) were used to build the retrieval model for biomass.



**Figure 6.** Pearson's correlation coefficients of different groups of input predictors and maize biomass measured in June, July and August 2018. The bars filled with diagonal lines were the most correlative to biomass of each month.

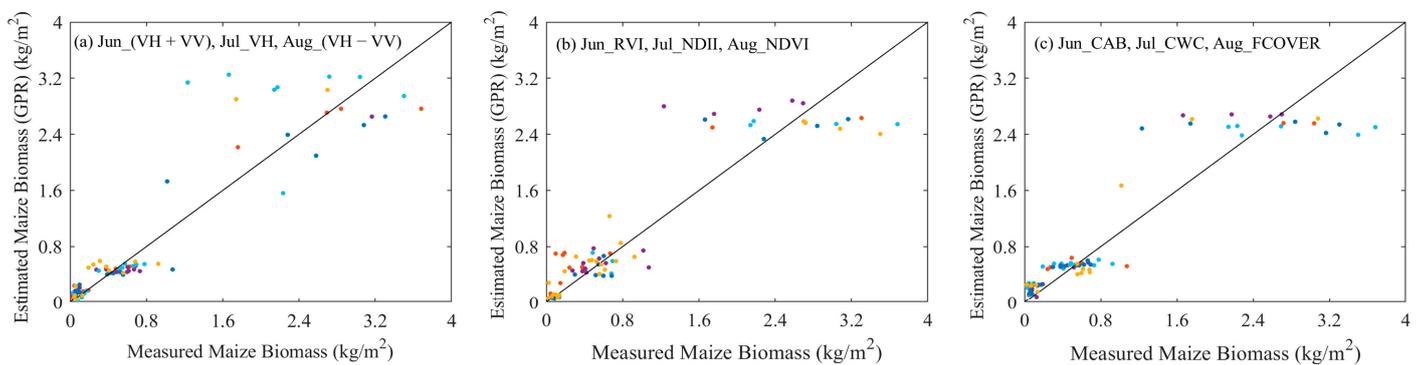
In terms of Sentinel-2 VIs, all VIs were significantly correlated to biomass measured in June ( $R > 0.60$ ), of which the highest was RVI ( $R = 0.71$ ) (Figure 6b). NDII was better correlated to July biomass ( $R = 0.35$ ) than the other VIs. NDVI yielded the strongest correlation with August biomass ( $R = 0.42$ ). RVI had the highest correlation ( $R = 0.78$ ) with all biomasses, followed by MSR ( $R = 0.74$ ). Further analysis was executed by using the June RVI, the July NDII, and the August NDVI as predictors to estimate maize biomass.

As for S-2 BPVs, all BPVs yielded high correlations to June biomass ( $R > 0.65$ ), of which the highest was CAB ( $R = 0.72$ ) (Figure 6c). CWC was more sensitive to July biomass ( $R = 0.33$ ). FCOVER had a stronger correlation with biomass collected in August ( $R = 0.35$ ). All variables were sensitive to biomass measured in the three periods except CWC. The combination of June CAB, July CWC, and August FCOVER was used for estimating biomass. Generally, S-2 VIs were more correlated to biomass than BPVs, which had marginal advantages over S-1 SAR derivatives.

The validation results of the above three types of combined features with GPR and RF for biomass estimation are presented in Table 8. In terms of GPR, the predictor of Jun\_(VV + VH), Jul\_VH, and Aug\_(VH – VV) significantly improved the retrieval accuracy of biomass and outperformed the GPR models with all SAR polarization indices and the optimized subset as input predictors, enhancing  $R^2$  from 0.40 to 0.81 and RMSE from  $0.84 \text{ kg/m}^2$  to  $0.43 \text{ kg/m}^2$ . The biomass was very good in estimation ( $\text{RPD} > 2.0$ ). Compared to the estimation results by all SAR polarization indices with GPR (Figure 3a), the phenomenon of underestimation and overestimation was obviously improved by using Jun\_(VH + VV), Jul\_VH, and Aug\_(VH – VV) (Figure 7a). It was suggested that the best prediction could be obtained by combining SAR polarization indices based on their sensitivities to biomass in different growing periods.

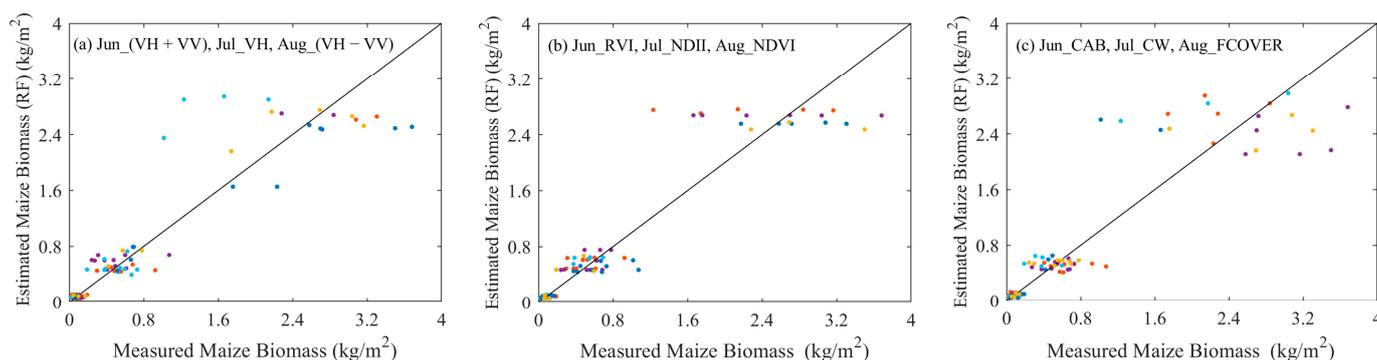
**Table 8.** Performance of GPR and RF on estimating maize biomass using combined features selected from correlation analysis.

Input Variables	GPR			RF		
	R <sup>2</sup>	RMSE (kg/m <sup>2</sup> )	RPD	R <sup>2</sup>	RMSE (kg/m <sup>2</sup> )	RPD
Jun_(VH + VV), Jul_VH, Aug_(VH - VV)	0.81	0.41	2.85	0.83	0.40	2.80
Jun_RVI, Jul_NDII, Aug_NDVI	0.83	0.39	2.93	0.82	0.43	2.69
Jun_CAB, Jul_CWC, Aug_FCOVER	0.82	0.40	2.73	0.85	0.38	2.97

**Figure 7.** Measured vs. estimated biomass by GPR with combined predictors. The different colors indicated the 5-fold subsets.

In comparison with the optimized GPR model based on EVI, RVI, and SAVI, the integrated predictor of Jun\_RVI, Jul\_NDII, and Aug\_NDVI with GPR improved R<sup>2</sup> from 0.80 to 0.83 and RMSE from 0.43 to 0.39 kg/m<sup>2</sup>. Point-by-point analysis showed a good linear relationship between the modeled biomass and sampled biomass (Figure 7b). The combined predictor of Jun\_CAB, Jul\_CWC, and Aug\_FCOVER with GPR achieved a result with R<sup>2</sup> = 0.82 and RMSE = 0.40 kg/m<sup>2</sup>, much better than the GPR model based on CAB and FCOVER (R<sup>2</sup> = 0.57, RMSE = 0.69 kg/m<sup>2</sup>). The scatter plots show that the estimated versus measured values fall close to the 1:1 line (Figure 7c). The results indicated that the biomass estimation models incorporating these three predictors were more robust and reliable and improved the accuracy of the biomass estimation.

To understand whether these three predictors improved the biomass estimation with another regression method, we also analyzed the performance of three predictors with RF for comparison. The RF models associated with the integrated SAR predictor and the combined VI predictor achieved similar results, with R<sup>2</sup> at about 0.82 and RMSE around 0.43 kg/m<sup>2</sup>. The Jun\_CAB, Jul\_CWC, and Aug\_FCOVER outperformed the other two predictors, with an R<sup>2</sup> of 0.85 and an RMSE of 0.38 kg/m<sup>2</sup>. The results showed that these three predictors with RF exceeded the univariate and optimized subset with RF models (Tables 3 and 7). Similar to the results in Figure 7, the biomass estimated by these three predictors with RF lies close to the 1:1 line (Figure 8a–c). These findings demonstrate that the combination of remote sensing derivatives according to their sensitivities to biomass in different growing periods is promising to provide sufficient information for biomass estimation.



**Figure 8.** Measured vs. estimated biomass by RF with combined predictors. The different colors indicated the 5-fold subsets.

#### 4. Discussion

##### 4.1. Proficiency of Single S-1 SAR Polarization Indices, S-2 VIs and BPVs for Maize Biomass Modelling

In this study, different types of SAR and optical indices were identified and regressed with ground-sampled biomass. This study demonstrated that compared with the models based solely on S-1 SAR polarization indices ( $R^2 = 0.04$ – $0.36$ ), VIs and BPVs models ( $R^2 = 0.31$ – $0.65$ ) showed better performance in predicting maize biomass. The findings were consistent with the study of Alebele et al., who discovered that SAR polarization indices did not yield more accurate estimates than VIs and BPVs [2]. Similar findings were reported in some studies that compared S-1 and S-2 data for biomass modeling [60]. One possible explanation is that the single-date SAR predictors had obvious limitations in monitoring seasonal variations of crops. The combination of multi-date can enhance the sensitivity of SAR indices to biomass and improve the accuracy of biomass. Results of Jun\_(VH + VV), Jul\_VH, and Aug\_(VH – VV) provided more accurate estimates of biomass than single SAR polarization indices. Castillo et al. [9] also found that multi-date S-1 images had a better correlation than single-date images in retrieving mangrove forests' biomass. A second possible reason for the performance of S-1 data is that SAR polarization indices are influenced by the height of the crop. As maize height continuously increased, the short wavelength of S-1 (C-band) exhibited limited ability to penetrate deeply into the maize canopy to capture structural information. In order to analyze the effect of crop height on S-1 performance, new integration methods were accomplished by combining the optimal SAR predictor (VH + VV) with height. The performance of VH + VV was significantly improved by the inclusion of height (Table 9). Results demonstrate that the effect of crop height on SAR response is real.

**Table 9.** Performance of GPR and RF on estimating maize biomass using VH + VV and combined maize height.

Input Variables	GPR			RF		
	$R^2$	RMSE ( $\text{kg/m}^2$ )	RPD	$R^2$	RMSE ( $\text{kg/m}^2$ )	RPD
VH + VV	0.36	0.82	1.30	0.41	0.85	1.35
VH + VV, height	0.59	0.65	1.68	0.59	0.65	1.74

With respect to S-2 VIs, GPR and RF models involving MSR and RVI, respectively, obtained better biomass predictions, followed by NDVI. The results supported the findings that VIs calculated by simple combinations of visible and near-infrared bands were sensitive to maize biomass [71,72]. However, the other selected optical indices obtained poor estimations ( $RPD < 1.4$ ). The main reason is that these VIs are highly sensitive to low

biomass; however, maize biomass in August is no longer low, and the saturation of some VIs occurs at high biomass. The red-edge indices such as Clre also did not show any significant improvements in biomass estimation over MSR and RVI, which was consistent with findings from Jin et al. [73], who indicated that red-edge indices had no significant effect on biomass estimation. Furthermore, S-2 BPVs provided insights into the importance of biomass modeling. Baloloy et al. [74] found CAB was more correlated to mangrove biomass than LAI and FCOVER, while Chen et al. [26] found a stronger relationship between LAI and forest biomass than CAB, FAPAR, and FCOVER. In this study, FCOVER with an RF model achieved better accuracy ( $R^2 = 0.58$ ,  $RMSE = 0.70 \text{ kg/m}^2$ ) than univariate models based on SAR polarization indices. It was confirmed that BPVs could provide a reliable prediction of maize biomass.

#### 4.2. Efficiency of Feature Selection Methods

The findings of this study revealed that for three types of remote-sensing datasets, GPR and RF exhibited their own characteristic in estimating maize biomass with feature optimization. The inputting variables had a direct impact on the performance of GPR and RF.  $(VH + VV)/(VH \times VV)$ ,  $VH + VV$ , and  $VH$  were identified as important predictors among the eight SAR polarization indices by utilizing the GPR feature optimization method. These three predictors were also highly correlated to the whole biomass (Figure 6). The use of these three indices as inputs to GPR had an equivalent ability to estimate biomass in comparison with the use of all polarization indices. With respect to S-2, the most valuable VI was easily identified. RVI was found to be the most important variable, followed by SAVI and EVI, which were known to reduce soil background disturbance [75,76]. EVI also keeps sensitivity to dense vegetation [77]. The optimized GPR involving these three VIs yielded accurate estimates of biomass with an  $R^2$  of 0.80 and an RMSE of  $0.43 \text{ kg/m}^2$ , which was better than the GPR models with all VIs and RVI as inputs, respectively. Among the five BPVs, CWC and FCOVER had the most significant contributions to estimating biomass. The GPR model with the two BPVs outperformed the GPR involving all BPVs. Actually, GPR automatically provides physical insight into the ranking of input variables based on their relevance, making it more capable of selecting optimal input variables and establishing estimation relationships.

The RFE method embedded in RF has limited capability to gather useful information for improving maize biomass estimation in this study. Compared to utilizing all predictors of each type of dataset as input predictors, the optimal subsets of polarization indices, VIs, and BPVs selected by RFE were all ineffective in improving the accuracy of biomass estimation. One possible explanation is that the ability of RFE may be influenced by the number of predictors. Previous studies on biomass estimation found that RFE was widely used to find the optimal subset of features from a large number of different types of variable combinations [32,78]. The prediction from the RF-RFE model is based on the average value of each tree generated by samples [67]. If the dataset contains limited sample units, they may be consistently underrepresented in the tree construction and RF-RFE may therefore result in variance in biomass estimation.

#### 4.3. Optimal Features Based on the Response of Remote Sensing Indicators to Biomass in Different Growth Periods

Few studies explored the responses of S-1 derivatives to maize biomass dynamics at different growth periods. During the growth periods, the contribution of the crop to microwave response is variable due to changes in plant structure, total biomass, canopy water content, and so on [79]. In the early phase of growth, maize plants are shorter in height and have a loose canopy. The scattering signal from the maize canopy is relatively limited and heavily influenced by the soil background [10,80]. Most maize plants reach their maximum height and canopy density during the middle development stage, and the scattering mainly comes from the canopy [81]. As maize matures, biomass components continue to increase due to fruit development. The radar signal is influenced not only by

the leaves and stems of maize but also by the fruits [12]. As shown by Pearson's correlation values in Figure 6, eight SAR derivatives showed different sensitivities to biomass measured at the three periods. The new S-1 predictor, Jun\_(VH + VV), Jul\_VH, and Aug\_(VH – VV) was proposed because these three SAR predictors had the strongest correlations with maize biomass measured in June, July, and August, respectively. This combination with GRP and RF both improved the retrieval accuracy of biomass, being the best model among the models associated with SAR information. The combination of multi-month SAR data can reflect the time-series variation of radar response, which is more conducive to monitoring crop dynamics.

Seasonal changes in canopy structures, biochemical traits, and soil background can significantly modify the spectral response of the canopy [82,83]. In different stages of crop development, the relationships between biomass and VIs are often significantly different [36,84]. The new S-2 predictor was proposed by RVI, NDII, and NDVI for June, July, and August, respectively. The results revealed that Jun\_RVI, Jul\_NDII, and Aug\_NDVI greatly improved the retrieval accuracies of GPR and RF. The GPR model with this integrated predictor achieved a higher accuracy with an  $R^2$  of 0.83 and an RMSE of 0.39 kg/m<sup>2</sup>, which was better than the GPR optimized by feature relevance. This can be explained by the fact that the integrated VI predictor can minimize the impact of the spectral response of the canopy and reduce the problem of saturation under high biomass conditions.

Not only did SAR polarization indices and VIs show different degrees of correlations to biomass, but five BPVs also responded differently to biomass dynamics. Similar findings have been reported by Li et al. [41], who reported that the relationships between BPVs and maize biomass changed at different growth periods. As a result, the same strategy was applied to S-2 BPVs, and the new predictor combined by Jun\_CAB, Jul\_CWC, and Aug\_FCOVER was proposed. Although the optimal univariate model based on FCOVER and the model associated with all BPVs obtained unreliable results, Jun\_CAB, Jul\_CWC, and Aug\_FCOVER with RF improved  $R^2$  to 0.85 and RMSE to 0.38 kg/m<sup>2</sup>. This predictor outperformed GPR and RF models with other features. In all, the results suggest that there is great potential for the retrieval of biomass by combining remote sensing predictors according to their responses to biomass dynamics.

## 5. Conclusions

This paper focused on the estimation of maize biomass based on GRP and RF methods from S-1 SAR polarization indices, S-2 VIs, and BPVs. Three new predictors were proposed based on the responses of these remote sensing derivatives to biomass measured in different periods. The results showed that neither GPR nor RF with sole or total SAR polarization indices or the optimized subset achieved reliable estimation of biomass. The best-performing SAR indicator was Jun\_(VV + VH), Jul\_VH, and Aug\_(VH – VV), obtaining an accuracy of  $R^2$  of 0.83 and RMSE of 0.40 kg/m<sup>2</sup> with GPR. The total VIs and the optimized features with GRP and RF both obtained higher accuracy than SAR polarization information, but the accuracy was further improved by using Jun\_RVI, Jul\_NDII, and Aug\_NDVI as predictors ( $R^2 = 0.83$ , RMSE = 0.39 kg/m<sup>2</sup>, RPD = 2.93). Moreover, the integrated predictor of Jun\_CAB, Jul\_CWC, and Aug\_FCOVER delivered excellent accuracies with RF ( $R^2 = 0.85$ , RMSE = 0.38 kg/m<sup>2</sup>, RPD = 2.97), much better than single or total BPVs, or optimized subsets. Compared to conventional remote sensing derivatives, the three integrated predictors reduced the overestimation of low biomass and underestimation of high biomass, significantly improving biomass retrieval accuracy. Overall, this study provided a reference for using S-1 and S-2 to estimate maize biomass.

**Author Contributions:** Conceptualization, Y.D.; methodology, C.X., Y.W. and Y.D.; software, C.X.; validation, Z.D.; formal analysis, R.Z.; investigation, X.Z.; resources, X.Z.; data curation, C.X., X.Z. and R.Z.; writing—original draft preparation, C.X.; writing—review and editing, Y.D. and Q.X.; visualization, C.X.; supervision, Y.D. and H.Z.; project administration, Y.D.; funding acquisition, Y.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences, grant number XDA28070500, Land Observation Satellite Supporting Platform of National Civil Space Infrastructure Project, grant number CASPLOS-CCSI, the Fundamental Research Funds for the Central Universities, grant number 2412020FZ004, and the Science and Technology Project of the Department of Education, Jilin Province, grant number JJKH20221163K.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors are grateful to the European Space Agency for open data policy. The authors wish to thank the five anonymous reviewers and the academic editor for their constructive comments, which improved the manuscript.

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

## References

- Scharf, P.C.; Lory, J.A. Calibrating corn color from aerial photographs to predict sidedress nitrogen need. *Agron. J.* **2002**, *94*, 397–404. [[CrossRef](#)]
- Alebele, Y.; Zhang, X.; Wang, W.; Yang, G.; Yao, X.; Zheng, H.; Zhu, Y.; Cao, W.; Cheng, T. Estimation of Canopy Biomass Components in Paddy Rice from Combined Optical and SAR Data Using Multi-Target Gaussian Regressor Stacking. *Remote Sens.* **2020**, *12*, 2564. [[CrossRef](#)]
- Mahlein, A.-K.; Oerke, E.-C.; Steiner, U.; Dehne, H.-W. Recent advances in sensing plant diseases for precision crop protection. *Eur. J. Plant Pathol.* **2012**, *133*, 197–209. [[CrossRef](#)]
- Padilla, F.L.M.; Maas, S.J.; González-Dugo, M.P.; Mansilla, F.; Rajan, N.; Gavilán, P.; Domínguez, J. Monitoring regional wheat yield in Southern Spain using the GRAMI model and satellite imagery. *Field Crops Res.* **2012**, *130*, 145–154. [[CrossRef](#)]
- Marshall, M.; Thenkabail, P. Developing in situ Non-Destructive Estimates of Crop Biomass to Address Issues of Scale in Remote Sensing. *Remote Sens.* **2015**, *7*, 808–835. [[CrossRef](#)]
- Shoko, C.; Mutanga, O.; Dube, T. Progress in the remote sensing of C3 and C4 grass species aboveground biomass over time and space. *ISPRS J. Photogramm. Remote Sens.* **2016**, *120*, 13–24. [[CrossRef](#)]
- Jiang, F.; Kutia, M.; Ma, K.; Chen, S.; Long, J.; Sun, H. Estimating the aboveground biomass of coniferous forest in Northeast China using spectral variables, land surface temperature and soil moisture. *Sci. Total Environ.* **2021**, *785*, 147335. [[CrossRef](#)]
- Zhao, Q.; Yu, S.; Zhao, F.; Tian, L.; Zhao, Z. Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. *For. Ecol. Manag.* **2019**, *434*, 224–234. [[CrossRef](#)]
- Castillo, J.A.A.; Apan, A.A.; Maraseni, T.N.; Salmo, S.G. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *134*, 70–85. [[CrossRef](#)]
- Forkuor, G.; Benewinde Zoungrana, J.-B.; Dimobe, K.; Ouattara, B.; Vadrevu, K.P.; Tondoh, J.E. Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets—A case study. *Remote Sens. Environ.* **2020**, *236*, 111496. [[CrossRef](#)]
- Wang, J.; Xiao, X.; Bajgain, R.; Starks, P.; Steiner, J.; Doughty, R.B.; Chang, Q. Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. *ISPRS J. Photogramm. Remote Sens.* **2019**, *154*, 189–201. [[CrossRef](#)]
- Gao, S.; Niu, Z.; Huang, N.; Hou, X. Estimating the Leaf Area Index, height and biomass of maize using HJ-1 and RADARSAT-2. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *24*, 1–8. [[CrossRef](#)]
- Nandy, S.; Srinet, R.; Padalia, H. Mapping Forest Height and Aboveground Biomass by Integrating ICESat-2, Sentinel-1 and Sentinel-2 Data Using Random Forest Algorithm in Northwest Himalayan Foothills of India. *Geophys. Res. Lett.* **2021**, *48*, e2021GL093799. [[CrossRef](#)]
- Sibanda, M.; Mutanga, O.; Rouget, M. Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. *ISPRS J. Photogramm. Remote Sens.* **2015**, *110*, 55–65. [[CrossRef](#)]
- Lu, D. The potential and challenge of remote sensing-based biomass estimation. *Int. J. Remote Sens.* **2007**, *27*, 1297–1328. [[CrossRef](#)]
- Vuorinne, I.; Heiskanen, J.; Pellikka, P.K.E. Assessing Leaf Biomass of Agave sisalana Using Sentinel-2 Vegetation Indices. *Remote Sens.* **2021**, *13*, 233. [[CrossRef](#)]
- Kanke, Y.; Tubaña, B.; Dalen, M.; Harrell, D. Evaluation of red and red-edge reflectance-based vegetation indices for rice biomass and grain yield prediction models in paddy fields. *Precis. Agric.* **2016**, *17*, 507–530. [[CrossRef](#)]
- Chao, Z.; Liu, N.; Zhang, P.; Ying, T.; Song, K. Estimation methods developing with remote sensing information for energy crop biomass: A comparative review. *Biomass Bioenergy* **2019**, *122*, 414–425. [[CrossRef](#)]
- Ghasemloo, N.; Matkan, A.A.; Alimohammadi, A.; Aghighi, H.; Mirbagheri, B. Estimating the Agricultural Farm Soil Moisture Using Spectral Indices of Landsat 8, and Sentinel-1, and Artificial Neural Networks. *JGSA* **2022**, *6*, 19. [[CrossRef](#)]
- Ndikumana, E.; Ho Tong Minh, D.; Dang Nguyen, H.; Baghdadi, N.; Courault, D.; Hossard, L.; El Moussawi, I. Estimation of Rice Height and Biomass Using Multitemporal SAR Sentinel-1 for Camargue, Southern France. *Remote Sens.* **2018**, *10*, 1394. [[CrossRef](#)]
- Du, P.; Bai, X.; Tan, K.; Xue, Z.; Samat, A.; Xia, J.; Li, E.; Su, H.; Liu, W. Advances of Four Machine Learning Methods for Spatial Data Handling: A Review. *JGSA* **2020**, *4*, 13. [[CrossRef](#)]

22. Torre-Tojal, L.; Bastarrika, A.; Boyano, A.; Lopez-Guede, J.M.; Graña, M. Above-ground biomass estimation from LiDAR data using random forest algorithms. *J. Comput. Sci.* **2022**, *58*, 101517. [[CrossRef](#)]
23. Jachowski, N.R.A.; Quak, M.S.Y.; Friess, D.A.; Duangnamon, D.; Webb, E.L.; Ziegler, A.D. Mangrove biomass estimation in Southwest Thailand using machine learning. *Appl. Geogr.* **2013**, *45*, 311–321. [[CrossRef](#)]
24. López-Serrano, P.M.; Cárdenas Domínguez, J.L.; Corral-Rivas, J.J.; Jiménez, E.; López-Sánchez, C.A.; Vega-Nieva, D.J. Modeling of Aboveground Biomass with Landsat 8 OLI and Machine Learning in Temperate Forests. *Forests* **2019**, *11*, 11. [[CrossRef](#)]
25. Pandit, S.; Tsuyuki, S.; Dube, T. Estimating Above-Ground Biomass in Sub-Tropical Buffer Zone Community Forests, Nepal, Using Sentinel 2 Data. *Remote Sens.* **2018**, *10*, 601. [[CrossRef](#)]
26. Chen, L.; Ren, C.; Zhang, B.; Wang, Z.; Xi, Y. Estimation of Forest Above-Ground Biomass by Geographically Weighted Regression and Machine Learning with Sentinel Imagery. *Forests* **2018**, *9*, 582. [[CrossRef](#)]
27. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature Selection. *ACM Comput. Surv.* **2018**, *50*, 1–45. [[CrossRef](#)]
28. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [[CrossRef](#)]
29. Brede, B.; Verrelst, J.; Gastellu-Etchegorry, J.-P.; Clevers, J.G.P.W.; Goudzwaard, L.; den Ouden, J.; Verbesselt, J.; Herold, M. Assessment of Workflow Feature Selection on Forest LAI Prediction with Sentinel-2A MSI, Landsat 7 ETM+ and Landsat 8 OLI. *Remote Sens.* **2020**, *12*, 915. [[CrossRef](#)]
30. Luo, M.; Wang, Y.; Xie, Y.; Zhou, L.; Qiao, J.; Qiu, S.; Sun, Y. Combination of Feature Selection and CatBoost for Prediction: The First Application to the Estimation of Aboveground Biomass. *Forests* **2021**, *12*, 216. [[CrossRef](#)]
31. Li, B.; Xu, X.; Zhang, L.; Han, J.; Bian, C.; Li, G.; Liu, J.; Jin, L. Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 161–172. [[CrossRef](#)]
32. Karlson, M.; Ostwald, M.; Reese, H.; Sanou, J.; Tankoano, B.; Mattsson, E. Mapping Tree Canopy Cover and Aboveground Biomass in Sudano-Sahelian Woodlands Using Landsat 8 and Random Forest. *Remote Sens.* **2015**, *7*, 10017–10041. [[CrossRef](#)]
33. Verrelst, J.; Alonso, L.; Camps-Valls, G.; Delegido, J.; Moreno, J. Retrieval of Vegetation Biophysical Parameters Using Gaussian Process Techniques. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1832–1843. [[CrossRef](#)]
34. Verrelst, J.; Rivera, J.P.; Gitelson, A.; Delegido, J.; Moreno, J.; Camps-Valls, G. Spectral band selection for vegetation properties retrieval using Gaussian processes regression. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 554–567. [[CrossRef](#)]
35. Son, N.T.; Chen, C.F.; Chen, C.R.; Minh, V.Q.; Trung, N.H. A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation. *Agric. For. Meteorol.* **2014**, *197*, 52–64. [[CrossRef](#)]
36. Huang, X.; Ziniti, B.; Torbick, N.; Ducey, M. Assessment of Forest above Ground Biomass Estimation Using Multi-Temporal C-band Sentinel-1 and Polarimetric L-band PALSAR-2 Data. *Remote Sens.* **2018**, *10*, 1424. [[CrossRef](#)]
37. Gnyp, M.L.; Miao, Y.; Yuan, F.; Ustin, S.L.; Yu, K.; Yao, Y.; Huang, S.; Bareth, G. Hyperspectral canopy sensing of paddy rice aboveground biomass at different growth stages. *Field Crops Res.* **2014**, *155*, 42–55. [[CrossRef](#)]
38. Mandal, D.; Kumar, V.; Ratha, D.; Dey, S.; Bhattacharya, A.; Lopez-Sanchez, J.M.; McNairn, H.; Rao, Y.S. Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data. *Remote Sens. Environ.* **2020**, *247*, 111954. [[CrossRef](#)]
39. Mansaray, L.R.; Zhang, K.; Kanu, A.S. Dry biomass estimation of paddy rice with Sentinel-1A satellite data using machine learning regression algorithms. *Comput. Electron. Agric.* **2020**, *176*, 105674. [[CrossRef](#)]
40. Li, S.; Ni, P.; Cui, G.; He, P.; Liu, H.; Li, L.; Liang, Z. Estimation of rice biophysical parameters using multitemporal RADARSAT-2 images. *IOP Conf. Ser. Earth Environ. Sci.* **2016**, *34*, 012019. [[CrossRef](#)]
41. Li, W.; Niu, Z.; Huang, N.; Wang, C.; Gao, S.; Wu, C. Airborne LiDAR technique for estimating biomass components of maize: A case study in Zhangye City, Northwest China. *Ecol. Indic.* **2015**, *57*, 486–496. [[CrossRef](#)]
42. Filippini, F. Sentinel-1 GRD Preprocessing Workflow. *Proceedings* **2019**, *18*, 11. [[CrossRef](#)]
43. Urban, M.; Truckenbrodt, J.; Rizzo, M.; Puletti, N.; Papale, D.; Mattioli, W.; Corona, P.; Balling, J.; Laurin, G.V. Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data. *J. Appl. Remote Sens.* **2018**, *12*, 016008. [[CrossRef](#)]
44. Xie, Q.; Dash, J.; Huete, A.; Jiang, A.; Yin, G.; Ding, Y.; Peng, D.; Hall, C.C.; Brown, L.; Shi, Y.; et al. Retrieval of crop biophysical parameters from Sentinel-2 remote sensing imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *80*, 187–195. [[CrossRef](#)]
45. Djamai, N.; Fernandes, R.; Weiss, M.; McNairn, H.; Goita, K. Validation of the Sentinel Simplified Level 2 Product Prototype Processor (SL2P) for mapping cropland biophysical variables using Sentinel-2/MSI and Landsat-8/OLI data. *Remote Sens. Environ.* **2019**, *225*, 416–430. [[CrossRef](#)]
46. Estévez, J.; Salinero-Delgado, M.; Berger, K.; Pipia, L.; Rivera-Cacedo, J.P.; Wocher, M.; Reyes-Muñoz, P.; Tagliabue, G.; Boschetti, M.; Verrelst, J. Gaussian processes retrieval of crop traits in Google Earth Engine based on Sentinel-2 top-of-atmosphere data. *Remote Sens. Environ.* **2022**, *273*, 112958. [[CrossRef](#)]
47. Kamenova, I.; Dimitrov, P. Evaluation of Sentinel-2 vegetation indices for prediction of LAI, fAPAR and fCover of winter wheat in Bulgaria. *Eur. J. Remote Sens.* **2020**, *54*, 89–108. [[CrossRef](#)]
48. Weiss, M.; Baret, F.; Jay, S. S2ToolBox Level 2 Products: LAI, FAPAR, FCOVER. 2021. Available online: <https://hal.inrae.fr/hal-03584016> (accessed on 23 April 2022).
49. Rouse, J.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ.* **1974**, *351*, 309.

50. Liu, H.Q.; Huete, A. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 457–465. [[CrossRef](#)]
51. Jordan, C.F. Derivation of leaf-area index from quality of light on the forest floor. *Ecology* **1969**, *50*, 663–666. [[CrossRef](#)]
52. Hunt, E.R., Jr.; Rock, B.N. Detection of changes in leaf water content using near-and middle-infrared reflectances. *Remote Sens. Environ.* **1989**, *30*, 43–54. [[CrossRef](#)]
53. Chen, J.M. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Can. J. Remote Sens.* **1996**, *22*, 229–242. [[CrossRef](#)]
54. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [[CrossRef](#)]
55. Gitelson, A.; Merzlyak, M.N. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *J. Photochem. Photobiol. B Biol.* **1994**, *22*, 247–252. [[CrossRef](#)]
56. Carter, G.A. Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *Int. J. Remote Sens.* **1994**, *15*, 697–703. [[CrossRef](#)]
57. Gitelson, A.A. Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* **2005**, *32*. [[CrossRef](#)]
58. Cao, Q.; Miao, Y.; Shen, J.; Yu, W.; Yuan, F.; Cheng, S.; Huang, S.; Wang, H.; Yang, W.; Liu, F. Improving in-season estimation of rice yield potential and responsiveness to topdressing nitrogen application with Crop Circle active crop canopy sensor. *Precis. Agric.* **2015**, *17*, 136–154. [[CrossRef](#)]
59. Rasmussen, C.E.; Nickisch, H. Gaussian processes for machine learning (GPML) toolbox. *J. Mach. Learn. Res.* **2010**, *11*, 3011–3015.
60. Belda, S.; Pipia, L.; Morcillo-Pallarés, P.; Verrelst, J. Optimizing Gaussian Process Regression for Image Time Series Gap-Filling and Crop Monitoring. *Agronomy* **2020**, *10*, 618. [[CrossRef](#)]
61. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
62. Mutanga, O.; Adam, E.; Cho, M.A. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 399–406. [[CrossRef](#)]
63. Dewi, C.; Chen, R.-C. Random forest and support vector machine on features selection for regression analysis. *Int. J. Innov. Comput. Inf. Control* **2019**, *15*, 2027–2037. [[CrossRef](#)]
64. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Software* **2008**, *28*, 1–26. [[CrossRef](#)]
65. Brungard, C.W.; Boettinger, J.L.; Duniway, M.C.; Wills, S.A.; Edwards, T.C. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma* **2015**, *239–240*, 68–83. [[CrossRef](#)]
66. Fathima, A.S.; Sheriff, L.A.K. Exploring support vector machines and random forests for the prognostic study of an arboviral disease. *Int. J. Comput. Appl. Technol.* **2012**, *57*, 1–5.
67. Viscarra Rossel, R.A.; McGlynn, R.N.; McBratney, A.B. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. *Geoderma* **2006**, *137*, 70–82. [[CrossRef](#)]
68. Dong, T.; Liu, J.; Qian, B.; He, L.; Liu, J.; Wang, R.; Jing, Q.; Champagne, C.; McNairn, H.; Powers, J.; et al. Estimating crop biomass using leaf area index derived from Landsat 8 and Sentinel-2 data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 236–250. [[CrossRef](#)]
69. Mutanga, O.; Skidmore, A.K. Hyperspectral band depth analysis for a better estimation of grass biomass (*Cenchrus ciliaris*) measured under controlled laboratory conditions. *Int. J. Appl. Earth Obs. Geoinf.* **2004**, *5*, 87–96. [[CrossRef](#)]
70. Verrelst, J.; Rivera, J.P.; Veroustraete, F.; Muñoz-Marí, J.; Clevers, J.G.P.W.; Camps-Valls, G.; Moreno, J. Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods—A comparison. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 260–272. [[CrossRef](#)]
71. Zhang, Y.; Xia, C.; Zhang, X.; Cheng, X.; Feng, G.; Wang, Y.; Gao, Q. Estimating the maize biomass by crop height and narrowband vegetation indices derived from UAV-based hyperspectral images. *Ecol. Indic.* **2021**, *129*, 107985. [[CrossRef](#)]
72. Wang, C.; Nie, S.; Xi, X.; Luo, S.; Sun, X. Estimating the Biomass of Maize with Hyperspectral and LiDAR Data. *Remote Sens.* **2016**, *9*, 11. [[CrossRef](#)]
73. Jin, X.; Li, Z.; Feng, H.; Ren, Z.; Li, S. Deep neural network algorithm for estimating maize biomass based on simulated Sentinel 2A vegetation indices and leaf area index. *Crop J.* **2020**, *8*, 87–97. [[CrossRef](#)]
74. Baloloy, A.B.; Blanco, A.C.; Candido, C.G.; Argamosa, R.J.L.; Dimalag, J.B.L.C.; Dimapilis, L.L.C.; Paringit, E.C. Estimation of Mangrove Forest Aboveground Biomass Using Multispectral Bands, Vegetation Indices and Biophysical Variables Derived from Optical Satellite Imageries: Rapideye, Planetscope and Sentinel-2. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2018**, *IV-3*, 29–36. [[CrossRef](#)]
75. Mathew, A.; Sreekumar, S.; Khandelwal, S.; Kaul, N.; Kumar, R. Prediction of surface temperatures for the assessment of urban heat island effect over Ahmedabad city using linear time series model. *Energy Build.* **2016**, *128*, 605–616. [[CrossRef](#)]
76. Ren, H.; Zhou, G.; Zhang, F. Using negative soil adjustment factor in soil-adjusted vegetation index (SAVI) for aboveground living biomass estimation in arid grasslands. *Remote Sens. Environ.* **2018**, *209*, 439–445. [[CrossRef](#)]
77. Clark, M.L.; Roberts, D.A.; Ewel, J.J.; Clark, D.B. Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sens. Environ.* **2011**, *115*, 2931–2942. [[CrossRef](#)]
78. Zhang, J.; Qiu, X.; Wu, Y.; Zhu, Y.; Cao, Q.; Liu, X.; Cao, W. Combining texture, color, and vegetation indices from fixed-wing UAS imagery to estimate wheat growth parameters using multivariate regression methods. *Comput. Electron. Agric.* **2021**, *185*, 106138. [[CrossRef](#)]

79. Wiseman, G.; McNairn, H.; Homayouni, S.; Shang, J. RADARSAT-2 Polarimetric SAR Response to Crop Biomass for Agricultural Production Monitoring. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4461–4471. [[CrossRef](#)]
80. Wang, Y.; Fang, S.; Zhao, L.; Huang, X.; Jiang, X. Parcel-based summer maize mapping and phenology estimation combined using Sentinel-2 and time series Sentinel-1 data. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102720. [[CrossRef](#)]
81. Khabbazan, S.; Vermunt, P.; Steele-Dunne, S.; Ratering Arntz, L.; Marinetti, C.; van der Valk, D.; Iannini, L.; Molijn, R.; Westerdijk, K.; van der Sande, C. Crop Monitoring Using Sentinel-1 Data: A Case Study from The Netherlands. *Remote Sens.* **2019**, *11*, 1887. [[CrossRef](#)]
82. Qiao, L.; Gao, D.; Zhang, J.; Li, M.; Sun, H.; Ma, J. Dynamic Influence Elimination and Chlorophyll Content Diagnosis of Maize Using UAV Spectral Imagery. *Remote Sens.* **2020**, *12*, 2650. [[CrossRef](#)]
83. Peng, Y.; Nguy-Robertson, A.; Arkebauer, T.; Gitelson, A. Assessment of Canopy Chlorophyll Content Retrieval in Maize and Soybean: Implications of Hysteresis on the Development of Generic Algorithms. *Remote Sens.* **2017**, *9*, 226. [[CrossRef](#)]
84. Babar, M.A.; Reynolds, M.P.; van Ginkel, M.; Klatt, A.R.; Raun, W.R.; Stone, M.L. Spectral Reflectance to Estimate Genetic Variation for In-Season Biomass, Leaf Chlorophyll, and Canopy Temperature in Wheat. *Crop Sci.* **2006**, *46*, 1046–1057. [[CrossRef](#)]