*Article*

# Machine-Learning Applications for the Retrieval of Forest Biomass from Airborne P-Band SAR Data

**Emanuele Santi** [1,*] , **Simonetta Paloscia** [1] , **Simone Pettinato** [1] , **Giovanni Cuozzo** [2] , **Antonio Padovano** [2] , **Claudia Notarnicola** [2] **and Clement Albinet** [3]

[1]   Institute for Applied Physics of National Research Council (CNR-IFAC), 50019 Florence, Italy; s.paloscia@ifac.cnr.it (S.P.); s.pettinato@ifac.cnr.it (S.P.)

[2]   Institute for Earth Observation, EURAC Research, 39100 Bolzano, Italy; giovanni.cuozzo@eurac.edu (G.C.); antoniopadovano82@gmail.com (A.P.); claudia.notarnicola@eurac.edu (C.N.)

[3]   ESRIN ESA's Centre for Earth Observation, 00044 Frascati, Rome, Italy; clement.albinet@esa.int

*   Correspondence: e.santi@ifac.cnr.it

check for updates

**Abstract:** This study aimed at evaluating the potential of machine learning (ML) for estimating forest biomass from polarimetric Synthetic Aperture Radar (SAR) data. Retrieval algorithms based on two different machine-learning methods, namely Artificial Neural Networks (ANNs) and Supported Vector Regressions (SVRs), were implemented and validated using the airborne polarimetric SAR data derived from the AfriSAR, BioSAR, and TropiSAR campaigns. These datasets, composed of polarimetric airborne SAR data at P-band and corresponding biomass values from in situ and LiDAR measurements, were made available by the European Space Agency (ESA) in the framework of the Biomass Retrieval Algorithm Inter-Comparison Exercise (BRIX). The sensitivity of the SAR measurements at all polarizations to the target biomass was evaluated on the entire set of data from all the campaigns, and separately on the dataset of each campaign. Based on the results of the sensitivity analysis, the retrieval was attempted by implementing general algorithms, using the entire dataset, and specific algorithms, using data of each campaign. Algorithm inputs are the SAR data and the corresponding local incidence angles, and output is the estimated biomass. To allow the comparison, both ANN and SVR were trained using the same subset of data, composed of 50% of the available dataset, and validated on the remaining part of the dataset. The validation of the algorithms demonstrated that both machine-learning methods were able to estimate the forest biomass with comparable accuracies. In detail, the validation of the general ANN algorithm resulted in a correlation coefficient R = 0.88, RMSE = 60 t/ha, and negligible BIAS, while the specific ANN for data obtained R from 0.78 to 0.94 and RMSE between 15 and 50 t/ha, depending on the dataset. Similarly, the general SVR was able to estimate the target parameter with R = 0.84, RMSE = 69 t/ha, and BIAS negligible, while the specific algorithms obtained $0.22 \leq R \leq 0.92$ and $19 \leq RMSE \leq 70$ (t/ha). The study also pointed out that the computational cost is similar for both methods. In this respect, the training is the only time-demanding part, while applying the trained algorithm to the validation set or to any other dataset occurs in near real time. As a final step of the study, the ANN and SVR algorithms were applied to the available SAR images for obtaining biomass maps from the available SAR images.

**Keywords:** P-band SAR; forest biomass; artificial neural networks; supported vector regressions

## 1. Introduction

Forests act as one of the main terrestrial carbon sinks [1]. Monitoring forest changes and estimating forest biomass are therefore mandatory for several applications, including studies on global changes, natural disaster prevention, and management of forest resources. The possibility to observe forests from

satellite and/or aircraft instruments is thus very attractive. In this respect, Synthetic Aperture Radar (SAR) has proven to be a suitable instrument for forest investigations. The SAR capability of estimating several forest parameters, such as forest density, tree height, and forest biomass, was demonstrated in several studies (e.g., [2–11]). The SAR capability of observing forest parameters depends on the operating frequency, that is, at microwaves, the sensitivity of low frequencies (i.e., P- or L- band) to forest biomass was largely proven [12–14]. The sensitivity increases with the wavelength, that is, at L band, the relationship between the backscattering coefficient ($\sigma°$) and biomass is logarithmic, and a saturation is shown for biomass values higher than 100–150 t/ha [15,16]. Some research has also pointed out a decrease of the signal for biomasses above this threshold [17]. At P-band, previous experiments and model simulations pointed out that saturation occurs for biomass values higher than at least 300–350 t/ha (e.g., [15]). Research has also demonstrated that, depending on the penetration capability, P-band is better related to the trunk and stem biomass, whereas L-band is better related to the crown biomass [18,19]. At present, only few SAR data at P-band are available, mainly from airborne campaigns, and therefore the potential of this band has not yet been fully exploited. A significant improvement in the remote sensing of forests is therefore expected with the upcoming BIOMASS satellite mission of the European Space Agency (ESA), carrying onboard an SAR operating at P-band [20].

Waiting for BIOMASS, research mainly focused on L-band and on the few data available at P-band from airborne campaigns. The L-band SAR backscatter variability in tropical forests was assessed using ALOS-2/PALSAR-2 data collected on the entire tropical forest region from 2016 to 2018 [21]. The authors in [22] investigated the relationship between polarimetric SAR backscatter at L- and P-bands and forest biomass using data acquired within the BioSAR1 ESA campaign in southern Sweden, and they attempted the retrieval with regression methods. At L-band, the biomass was estimated with a root mean square error (RMSE) between 31% and 46%, whereas P-band exhibited better retrievals, with RMSE between 18% and 27%.

The retrieval of forest biomass in northeast Costa Rica was addressed by the authors in [23], using P- and L- band SAR data acquired by the AIRSAR NASA/JPL airborne SAR system in March 2004. By using algorithms based on polarized radar backscatter, they estimated forest biomass with RMSE = 22.6 t/ha for biomass < 300 t/ha at P-band and RMSE = 23.8 t/ha for biomass < 150 t/ha at L-band.

Tomographic applications have also been considered, thanks to their capability of separating the contributions of several layers of vegetation and soil surface. The potential of TomoSAR data at L-band to monitor temporal variations of forest structure was addressed by using simulated and experimental datasets in one study [24]. In another study [25], the application of tomography to SAOCOM L-band SAR data allowed the retrieval of biomass using a single polarization (HH) with a 26–30% RMSE and a small effect from the observation geometry and local topography.

The potential of machine-learning (ML) methods for addressing forest biomass retrieval is evaluated in this study. Depending on its ability to solve nonlinear problems, ML has been previously employed for the retrieval of surface parameters from remote sensing data, providing more accurate results than other methods (e.g., [26–28]).

In this framework, both Artificial Neural Networks (ANNs) and Supported Vector Regressions (SVRs) demonstrated their potential in addressing the retrieval of surface parameters, showing their peculiar capability of combining inputs from different sources for addressing the retrieval.

ANNs have been already applied to several remote sensing problems (e.g., [29–31]), demonstrating a good trade-off between retrieval accuracy and computational cost. Past studies have pointed out that the availability of an extensive reference dataset for training the ANN is the main constraint for obtaining satisfactory retrievals (e.g., [32]). However, only a few examples of ANN applications to the retrieval of forest biomass are available in the literature, and in a couple of these [33,34], ANN were applied to retrieve the forest biomass in Mediterranean areas by combining L- and C- band SAR data.

SVR is another method of machine learning that has gained an increasing interest in the remote sensing of geophysical variables [35,36], thanks to the flexibility and robustness to noise in the training

data. SVR was successfully applied to the retrieval of surface parameters by the authors in [37], where this method was also compared to ANN.

In this study, two algorithms for the retrieval of forest biomass, based on ANN and SVR, respectively, were implemented and validated using P- band polarimetric airborne SAR data acquired by the ESA during the BioSAR, AfriSAR, and TropiSAR campaigns [38–43]. Inputs of both algorithms are the polarimetric SAR backscattering, and output is the biomass. Considering the dependence of SAR measurements on the acquisition geometry [44–47], both algorithms have the local incidence angle (LIA) as additional input.

The paper is organized as follows. Section 2 gives an overview of the test areas and the datasets considered in this study. The ANN- and SVR-based retrieval algorithms are presented in Section 3. The sensitivity of SAR measurements to forest biomass is described in Section 4. The algorithm validation is presented in Section 5, the obtained results are discussed in Section 6 and conclusions and future work are presented in Section 7.

## 2. Experimental Sites and SAR Data

The entire dataset made available by the ESA in the framework of the Biomass Retrieval Algorithm Inter-Comparison Exercise (BRIX) [48] was considered for implementing, training, and validating both the ANN and SVR algorithms. BRIX is an initiative promoted by the ESA that is intended to intercompare different approaches for retrieving forest biomass from P- band fully polarimetric SAR sensors. The BRIX dataset was composed of the P-band polarimetric SAR data acquired during the airborne experiments carried out in preparation for BIOMASS. The calibrated and geocoded backscattering data from each campaign were delivered by the ESA through the "Testbed", as described in the BRIX documentation [48]. We referred to the support documentation of each campaign for the description of the SAR data processing and of the in situ biomass measurements [38–43]. The main dataset was derived from the joint NASA and ESA AfriSAR experiment, which was conducted in Gabon. The experiment was composed of two campaigns—the first was carried out in 2015 with the ONERA (Office National d'Études et de Recherches Aérospatiales) SETHI SAR system and the second one in 2016 with the NASA (National Aeronautics and Space Administration) Land Vegetation and Ice Sensor (LVIS) LiDAR, the NASA L-band UAVSAR, and the DLR (German Aerospace Center) F-SAR [38]. The area covered by the flights was composed of four test sites. The main site was located in Lopé National Park (0.5° S, 11.5° E), a UNESCO World Heritage site protected since 2007, in which high closed-canopy forests are merged with open savannas. SAR campaigns were carried out in three more sites, including the protected area of Mondah, close to Libreville (0.6° N, 9.6° E), the Mabounié mining site (0.7° S, 10.7° E), and the Rabi site (1.9° S, 10° E). In Mondah, integer patches of forest are mixed with others significantly disturbed by human activities. Mabounié is forested for the most part but with traces of ongoing building, while Rabi is an onshore oil-drilling site in which some areas were preserved. Ancillary LiDAR measurements were collected for generating biomass maps at 50 m to be used for validation. Direct biomass measurements were also carried out on a total of 12 plots of $50 \times 50$ m (33 subplots of $25 \times 25$ m) for calibration and validation purposes.

The other datasets were derived from the three BioSAR campaigns [39–41] and from the TropiSAR campaign [42,43]. The data included biomass values up to 500 t/ha, being representative of forest types that included boreal and equatorial forests. The datasets also included very low biomass values, clearly referring to non-forested areas. These data were kept in the training to make the algorithms capable of identifying the low vegetated areas as well. All the test areas had flat or gently undulated topography. For instance, the La Lopé area included a 600 m tall hill rising over a 200 m ground: thus, the effect of local slopes on the measured $\sigma^0$ was accounted for by including the LIA in the algorithm inputs.

By combining all the available data, a dataset was obtained composed of about 4500 sets of backscattering coefficients ($\sigma°$) at four polarizations (HH, HV, VH, and VV), as well as corresponding LIA and biomass values from LiDAR or in situ measurements.

The ANN and SVR algorithms were trained by dividing the entire BRIX dataset into two parts, or the subsets of each campaign in the case of specific algorithms. The training set was composed of 50% of the entire dataset, while the other 50% was used for validating the algorithms. The flowchart in Figure 1 shows how the training and validation sets were obtained. The interleaved sampling was also evaluated for dividing the dataset; however, negligible differences were obtained in the training and validation results. It should be mentioned that the sampling cannot be considered purely random in the case of the general algorithm, since the training and validation sets are obtained by combining together the specific sets randomly sampled, so the general training and validation sets include the 50% of data from each campaign.
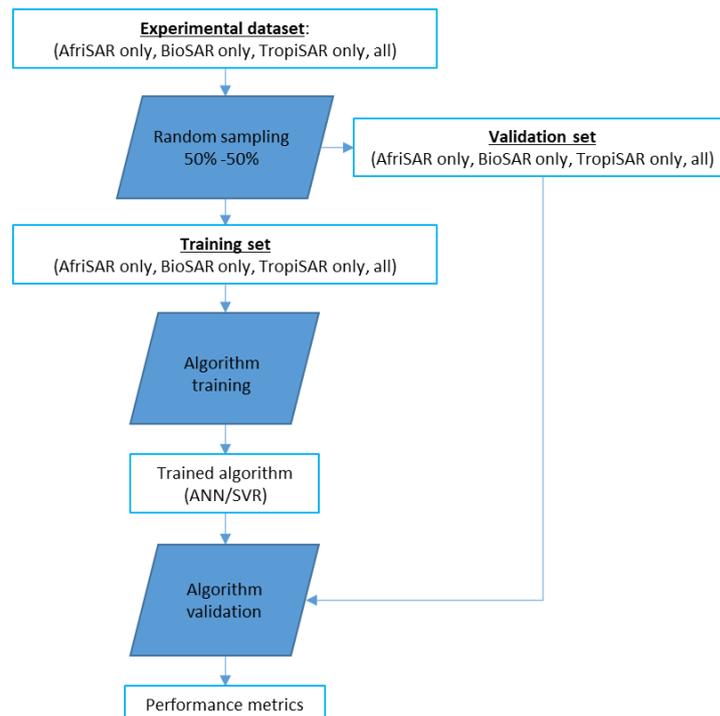


**Figure 1.** Generation of training and validation datasets.

## 3. Retrieval Methods

### 3.1. ANN

ANN can be regarded as a statistical method based on the minimum variance, which is able to approximate the existing relationship between the given input(s) and output(s) [49,50].

The core of the algorithm is based on the Feedforward Multi-layer Perceptron Artificial Neural Networks (MLP-ANNs) available in the MATLAB® Neural Networks toolbox. In MLP-ANN, each neuron is connected to all the other neurons of the previous and following layers. The connections are weighted and biased by coefficients that are iteratively adjusted during the training, thus modifying the strength of every connection. In each neuron, the inputs are weighted, biased, and added each other to produce an output value, called activation, which is then passed through the so-called transfer function and moved to the neurons of the following layer. In this study, transfer functions of linear, hyperbolic tangent sigmoid (tansig), and log-sigmoid (logsig) types were considered.

The weights and biases were adjusted by the Back Propagation (BP) learning rule, namely BP is a gradient descendent algorithm aimed at minimizing the Mean Square Error (MSE) between the ANN output and the desired value.

In the implemented algorithm, ANN inputs were the SAR backscattering coefficients (dB) at all polarizations (HH, HV, VH, VV) and the corresponding local incidence angle (LIA), and output was

the forest biomass (t/ha). It is worthy to mention that, according to the reciprocity property of passive targets for monostatic observations, no difference was found in using one or the other cross-polarized channel, and therefore the two cross-polarized channels were provided to the algorithms as mean average. Several combinations of inputs were evaluated, by implementing specific algorithms for each dataset (AfriSAR Onera, AfriSAR DLR, BioSAR, and TropiSAR), and a general algorithm which was trained and validated on the entire dataset.

The training set was further divided into 60%, 20%, and 20% by using random sampling [32]. The ANN training was carried out on the first subset, using BP for adjusting the ANN parameters. The other two subsets were used for a posteriori tests at each training iteration. The "early stopping" rule was applied for preventing overfitting, that is, training stops as soon as the errors on the three subsets are diverging.

The best dimensioning of ANN, in terms of number of neurons and hidden layers, is obtained by iteratively increasing the number of neurons and hidden layers from one hidden layer, with a number of neurons equal to the number of inputs, and up to two hidden layers, with a number of neurons each equal to three times the number of inputs. The training of each configuration is repeated 100 times for each transfer function, by resetting each time the initial ANN weights. At the end of the iterations, the results are compared and the ANN giving the highest correlation estimated vs. target is chosen as optimal. Such systematic optimization helps in preventing both underfitting and overfitting; underfitting occurs when the ANN architecture is too simplistic for the given problem, and overfitting happens when the ANN architecture is overly complex and it also fits the noise in the training set, causing large errors when applied to other datasets.

It should be noted that the training described here is carried out one time only; after training, the ANN is saved and loaded again to be applied to the validation set and, in this case, to any new datasets. Since the training is the only time-consuming step of the implementation, the algorithm can be applied to new data in near real time.

Moreover, the training can be updated every time new training sets are available, for improving the retrieval accuracy. This represents a unique feature of the ANN and, in general, of the data-driven approaches based on machine learning, in comparison to the conventional retrievals that are based on the inversion of electromagnetic forward models.

The validation, to which the results presented in Section 5 refer, was obtained by applying the saved ANN to the validation set, which was not involved in the training, to keep training and validation as independent as possible.

## 3.2. SVR

The second machine-learning approach to the retrieval was based on the Supported Vector Regression (SVR) techniques. Similarly to ANN, past research has proven the SVR capabilities for remote sensing applications (e.g., [36,51]). SVRs were demonstrated capable of handling complex and nonlinear problems and of managing different kinds of inputs. While neural networks can handle nonlinear problems having only a large training dataset, SVRs can achieve high accuracies, even if few training data are available [51]. Actually, SVRs overcome this limitation because they are based on a geometrical concept [52]. The method uses a so-called kernel function, for mapping the m-dimensional input space of the original problem into a higher dimensional space, in which the function underlying the data can be linearly approximated.

While neural networks try to populate the feature space with as many data as possible considering all the available combinations of inputs and outputs for mapping the functions, SVRs aim at identifying the boundaries of the tolerance tube around the input data to map the function without the need of larger datasets [53].

The problem of retrieving the biomass from the P-band backscattering at different polarizations was set as follows:

$$y = f(x_1, x_2, \ldots x_m) + e \tag{1}$$

where $f$ is the desired function, $y$ is the biomass, $x_1$, $x_2$,... $x_m$ is the backscattering coefficient at different polarizations, and $e$ is the white noise. The y estimation is obtained by determining the mapping function $f'$ as close as possible to the true mapping $f$ for the given problem [37]. Given a set of N reference samples {$xi$, $yi$ |i=1, ... N}, the $\varepsilon$-insensitive SVR attempts to identify a smooth function $f'$ approximating $f$ while keeping at most $\varepsilon$ deviation from the targets $yi$ [54]. $f'$ is obtained by mapping the input domain at m-dimensions into a higher dimension feature space, in which the function flatness is increased. In the new space, $f'$ can be linearly approximated, according to the following equation:

$$f'(x) = w \cdot \Phi(x) + b \tag{2}$$

where $w$ represents the weights of the linear function, $\Phi$ is the mapping function that transforms the samples into the higher dimensional space, and $b$ is the bias.

The cost function combining the training error and the model complexity is minimized to obtain the optimal linear function in the transformed feature space [36]. The loss function $f'$ is $\varepsilon$-insensitive— $\varepsilon$ being the tolerance to errors, $f'$ ensures that losses smaller than $\varepsilon$ are neglected [53]. An example of a possible choice of the $\varepsilon$-insensitive loss function is shown in Figure 2. The second term of Equation (2) is computed as the Euclidean norm of the weight vector $w$. The latter is inversely related to the geometrical margins of the solution and therefore to the model complexity [36]. A regularization parameter $C$ is also introduced with the scope of adjusting the trade-off between the complexity (flatness) of the function $f'$ and the tolerance to the empirical errors. We referred to [37] for the detailed mathematical formulation.
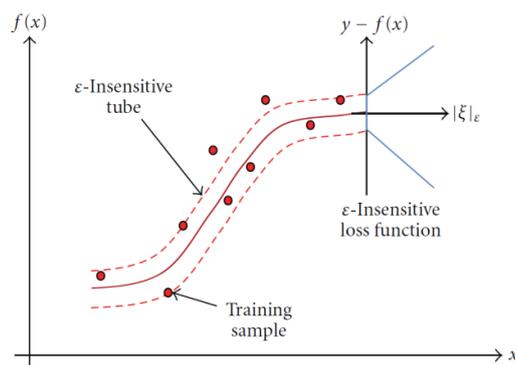


**Figure 2.** Example of a possible choice of the $\varepsilon$-insensitive loss function characterizing the Supported Vector Regression (SVR) learning approach (after [37]).

The training phase acts on reference samples composed by field data coupled with remote sensing data, to train the SVR algorithm and tune the kernel parameters C and $\varepsilon$. This process is called model selection [53]. During the training phase, the SVR algorithm uses a subset of the training dataset, called the test dataset, to tune its performances step by step. After this phase, the algorithm is applied to the validation set for evaluating the algorithm performances on independent data. At this point, the learning phase is over and the regressor can be used (operational estimation phase). The trained SVR is a good approximator of the mapping function between input data, which in this case are represented by HH, VV, HV, and VH P-band backscattering, and the target variable represented by the biomass. The separation among training, test, and validation follows the same approach proposed for ANN.

Similar to ANN, specific SVR training regressors were computed from each of the missions provided in the input dataset, while a general SVR training regressor was computed from the entire dataset. The input/output configuration is the same of ANN—algorithm inputs are the backscattering coefficients at all polarizations (HH, VV, HV, VH), while the output is the forest biomass.

## 4. Data Analysis

A sensitivity analysis was carried out to understand the relationship between σ° and forest biomass for all the experimental data available in BRIX.

The analysis of the BRIX dataset confirmed the sensitivity of σ° to the forest biomass; the direct sensitivity of σ° at each polarization to forest biomass for the AfriSAR dataset is shown in Figure 3. The plots show an increase of σ° when biomass increases up to 500 t/ha, although some saturation is evident for biomass values higher than 300–350 t/ha, as already pointed out by previous research (e.g., [15,16]).
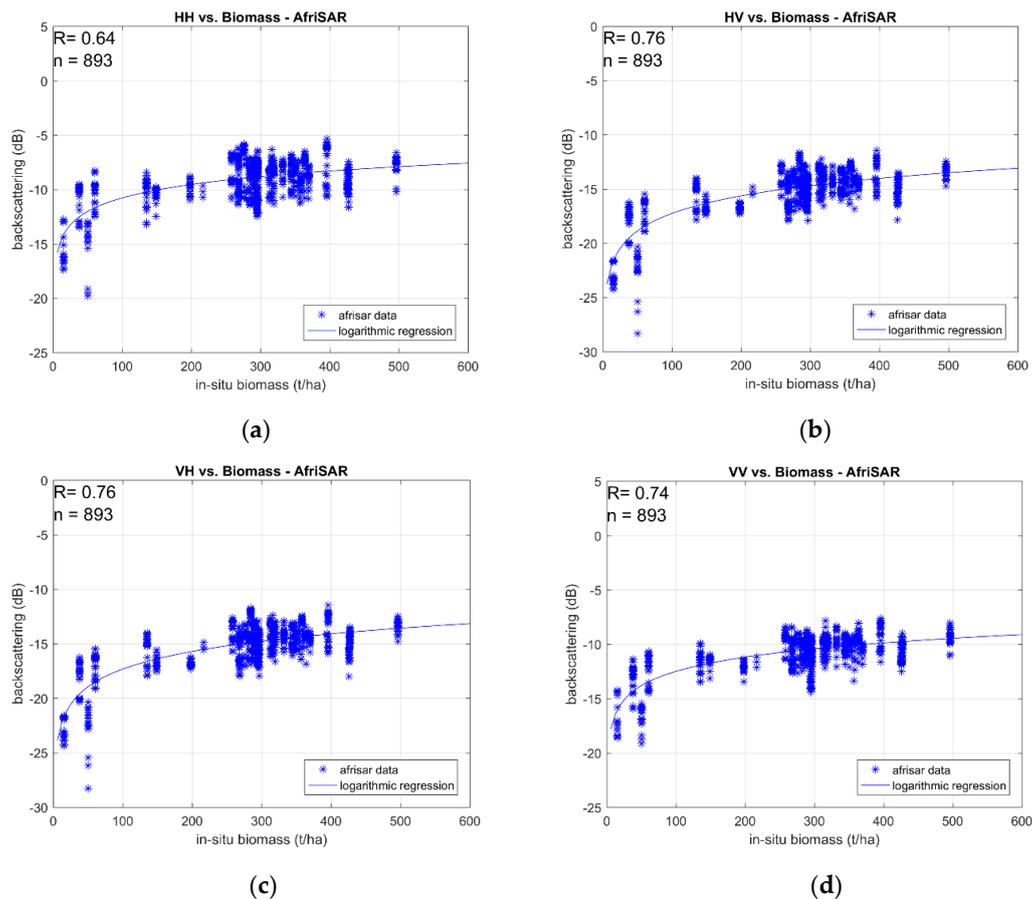


**Figure 3.** AfriSAR dataset: σ° at P-band vs. Biomass in (**a**) HH polarization, (**b**) HV polarization, (**c**) VH polarization, (**d**) VV polarization.

The backscattering however does not saturate completely, and some increase with biomass can also be observed beyond this threshold. All scatterplots exhibit vertical clustering of σ°; this depends on the σ° variability between the different SAR acquisitions in each test area that composed the dataset. Moreover, the same values of in situ biomass correspond to different subareas and, therefore, to different σ° values, depending on the spatial variability of forest structure and on the speckle [23]. This further increased the spread of data.

The results of the sensitivity analysis conducted on the entire BRIX dataset is shown in Figure 4, where the backscattering at the four polarizations is represented as a function of the in situ biomass.

Different colors correspond to different campaigns (AfriSAR, BioSAR, TropiSAR), showing some separation of the data coming from the different campaigns. In particular, at least two different clusters can be identified, one including the AfriSAR and TropiSAR data and one including the BioSAR data. These clusters can be explained by considering the differences in the instrumental setups and calibrations of the different SAR instruments used in the campaigns [38–43].
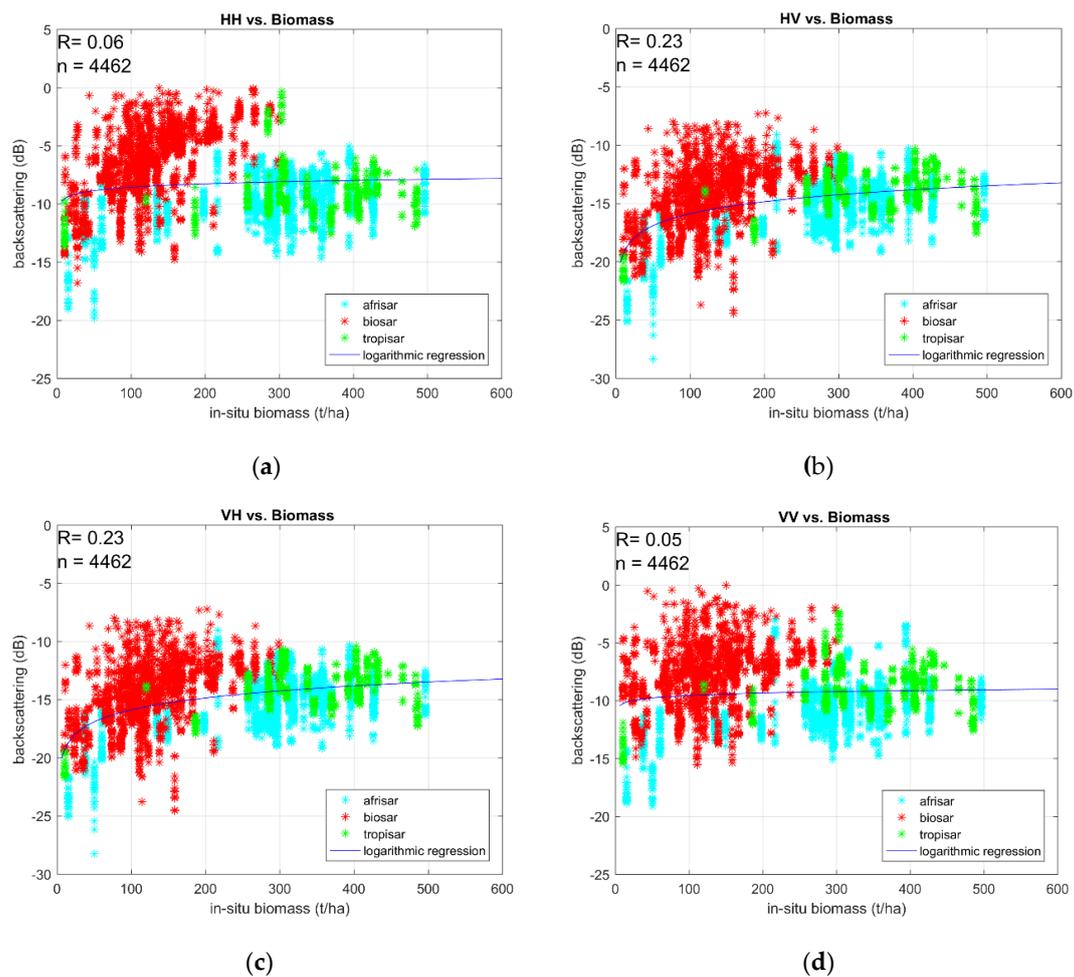
**Figure 4.** All data: σ° in (**a**) HH, (**b**) HV, (**c**) VH, and (**d**) VV vs. Biomass.

The analysis of correlation coefficients (R) reflects this behavior, with higher correlation when considering the AfriSAR dataset (R from 0.64 to 0.76) and low values when considering the entire dataset (R from 0.16 to 0.34). The R values for each campaign are summarized in Table 1.

**Table 1.** Correlation coefficients (R) between σ° at various polarizations and in situ biomass.

| Campaign | RVV | RVH | RHV | RHH |
|---|---|---|---|---|
| AfriSAR DLR | 0.74 | 0.76 | 0.76 | 0.64 |
| AfriSAR Onera | 0.51 | 0.64 | 0.65 | 0.53 |
| BioSAR 1 | 0.07 | 0.21 | 0.21 | 0.35 |
| BioSAR 2 | −0.04 | 0.04 | 0.03 | 0.05 |
| BioSAR 3 | 0.02 | −0.01 | 0 | 0.02 |
| TropiSAR | 0.5 | 0.69 | 0.68 | 0.27 |
| All | 0.05 | 0.23 | 0.23 | 0.06 |

Table 1 points out that the three BioSAR datasets have the worst correlation to in situ biomass. This can be attributed to the boreal forest type, which is characterized by more sparse trees and lower biomasses. Therefore, soil and undergrowing vegetation are expected to contribute to the total backscattering. The correlation computed by grouping the three sets slightly increases (0.17 ≤ R ≤ 0.24) but is still very low. Depending on the lower biomass range, σ° saturation was not observed in the BioSAR datasets.

The dependence of the relationship σ°–biomass on the site characteristics suggested that, along with a general algorithm able to apply to all datasets, specific algorithms for each dataset improve the retrieval performances.

## 5. Results

### 5.1. ANN

The ANN algorithm validation is summarized in the plots of Figure 5. Each scatterplot shows the predicted vs. in situ biomass values for the validation set of each campaign, obtained by applying the specific ANN to each given validation set, plus the result obtained by applying the general ANN to the whole validation set.
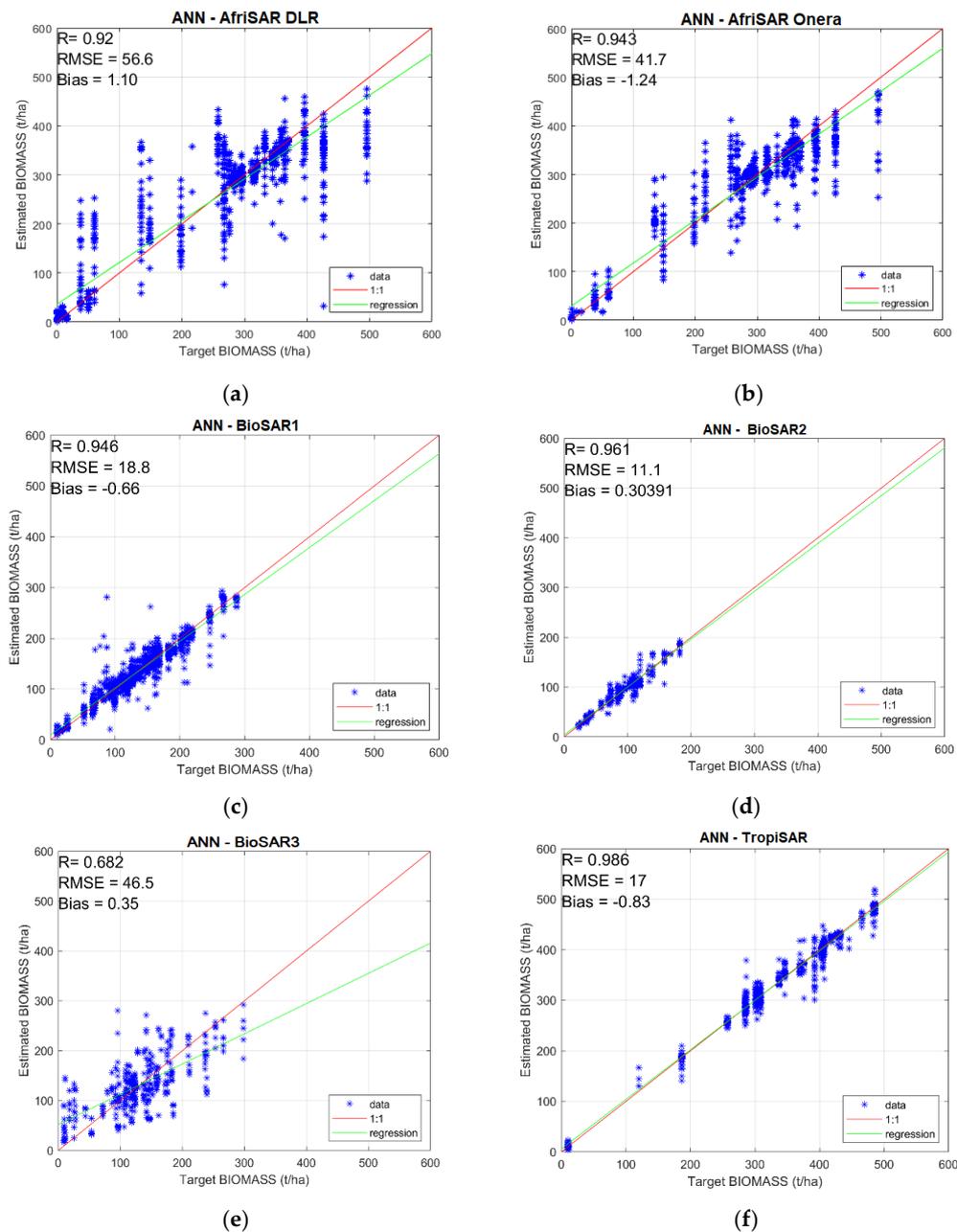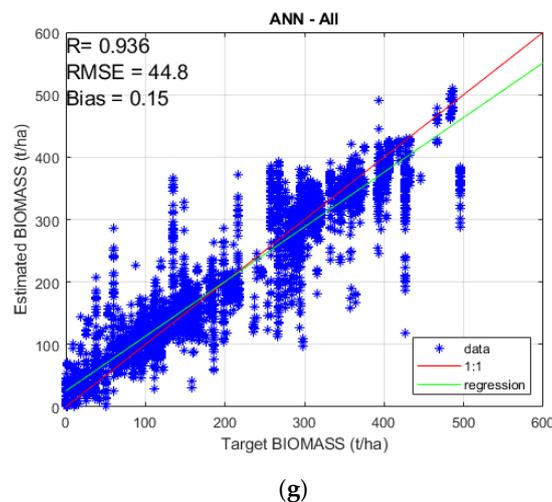


(**a**)

(**b**)

(**c**)

(**d**)

(**e**)

(**f**)

**Figure 5.** *Cont.*

(**g**)

**Figure 5.** ANN predicted vs. target biomass for (**a**) AfriSAR DLR, (**b**) AfriSAR Onera, (**c**) BioSAR 1, (**d**) BioSAR 2, (**e**) BioSAR 3, (**f**) TropiSAR, and (**g**) all (general algorithm).

The correlation coefficient between estimated and target biomass ranged from R = 0.69 to R = 0.98, while the RMSE was between 14 t/ha and 58 t/ha (Figure 5a–f). The result obtained by the specific ANN for the two AfriSAR missions was $0.92 \leq R \leq 0.94$ with $42 \leq RMSE \leq 57$ t/ha. Some underestimation of the highest biomass values can be identified in the scatterplot, possibly attributed to the saturation exhibited by the SAR signal for biomasses higher than 350 t/ha (Figure 5a,b). The ANN for TropiSAR dataset obtained better results (R = 0.98 and RMSE = 17 t/ha—Figure 5f); in that case, the underestimation of higher values was not evident. Among the specific ANN for BioSAR missions, the first and the second obtained similar results in terms of both R and RMSE (Figure 5c,d), whereas worse results were obtained for the third one (Figure 5e).

Finally, the validation result for the general ANN is reported in Figure 5g). Looking at the $\sigma^\circ$ correlation to the forest biomass listed in Figure 4, this result is quite surprising, since the correlation between biomass estimated by the ANN and reference values was R = 0.93, while the $\sigma^\circ$ at various polarizations and target biomass on the entire dataset was $R \leq 0.23$. The *p*-value was < 0.05 for the general and all the specific algorithms.

*5.2. SVR*

The learning process of the proposed algorithm was performed for any of the six missions and a further analysis was performed for the learning process by using the complete dataset. Then a trained SVR regressor was obtained for each dataset.

The performances of each SVR regressor were computed using the validation dataset and using as metrics the correlation and the RMSE between true biomass and estimated biomass (Figure 6a–g), where the green line represents the regression line.

From the scatterplots, we can see that the correlation coefficient between estimated and target biomass ranges from R = 0.65 to R = 0.93, while the RMSE is between 15 t/ha and 65 t/ha.

Despite a lower sensitivity of the backscattering coefficient with high biomass values, the SVR regression appears to predict quite well the higher biomasses. On the other hand, the higher sensitivity of the backscattering coefficient with low biomass values does not correspond to the lower ability of the SVR regressor to forecast low values of biomass.
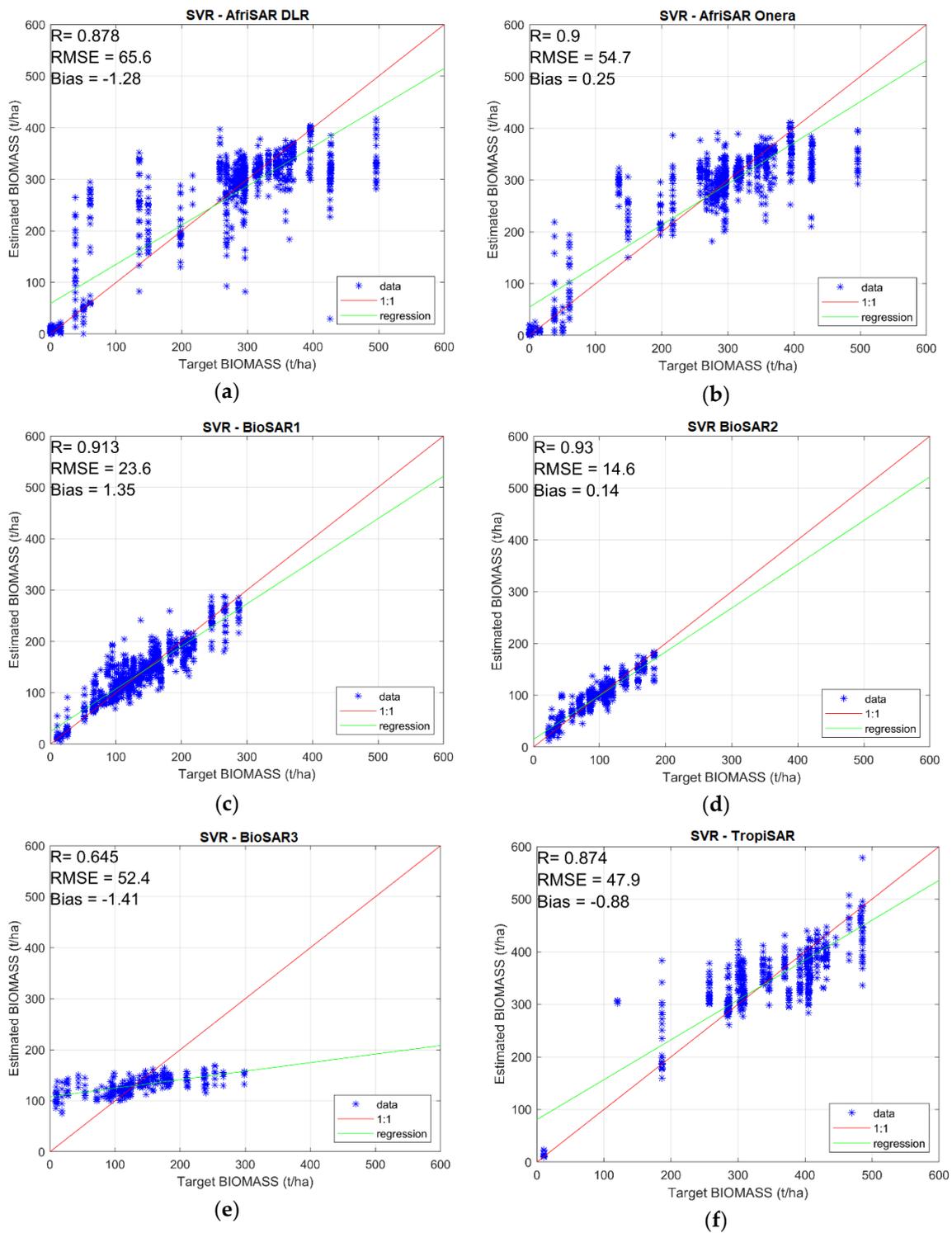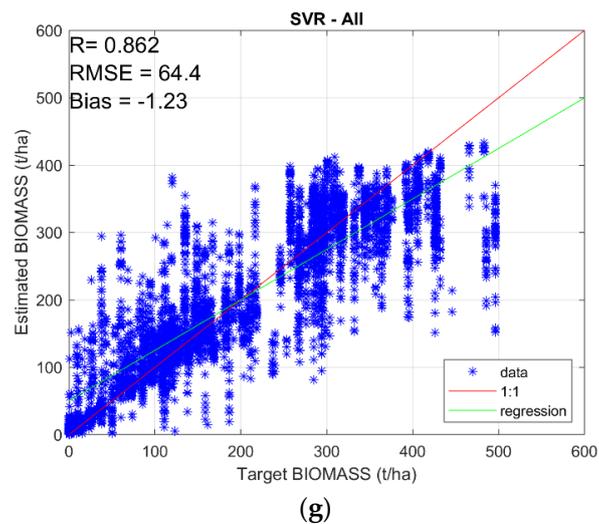
(**a**)

(**b**)

(**c**)

(**d**)

(**e**)

(**f**)

**Figure 6.** *Cont.*

**(g)**

**Figure 6.** True vs. estimated biomass using SVR: (**a**) AfriSAR DLR, (**b**) AfriSAR Onera, (**c**) BioSAR 1, (**d**) BioSAR 2, (**e**) BioSAR 3, (**f**) TropiSAR, and (**g**) All (general algorithm).

The SVR regressor obtained for AfriSAR DLR dataset shows R = 0.88 and RMSE = 65 t/ha. For biomass values greater than 200, the SVR regressor shows rather good performances with RMSE = 47 t/ha, whereas for values lower than 200 it shows an overestimation in the estimated biomass (Figure 6a). The result obtained for AfriSAR Onera shows R = 0.9 and RMSE = 55 t/ha with some bias that however does not hamper the accuracy (Figure 6b).

The results obtained for the BioSAR 1 and 2 datasets are quite good, as demonstrated by R = 0.92 and 0.93 and RMSE = 24 t/ha and 15 t/ha, respectively (Figure 6c,d). As for the ANN algorithm, the accuracy was lower in the case of the BioSAR 3 dataset (RMSE = 53 t/ha and R = 0.65). Moreover, a consistent underestimation of the target biomass can be observed (Figure 6e). For the TropiSAR dataset, SVR obtained R = 0.87 and RMSE = 48 t/ha, with some overestimation of the biomasses in the range 200–300 t/ha (Figure 6f).

The general SVR regressor for the complete dataset obtained R = 0.86 and RMSE = 64 t/ha (Figure 6g). Similar to ANN, the condition $p$-value < 0.05 was verified in all cases.

### 5.3. Biomass Maps

After validation, the ANN and SVR algorithms were applied to the available SAR images for generating biomass maps of the entire area covered by the SAR acquisition. Figure 7 shows, as examples, three biomass maps, namely two generated by ANN and one by SVR using the AfriSAR DLR dataset. The maps are generated pixel by pixel from the input SAR data. Along with each map, the corresponding validation scatterplot using the in situ data available is shown. The areas other than savannah are covered by very dense forest and therefore the majority of points in the scatterplot is concentrated around very low and very high values. The correlation coefficient is $0.88 \leq R \leq 0.95$ and both machine-learning approaches behave similarly, by slightly underestimating the highest values of biomass. The qualitative inspection of the maps shows that non-forested areas, mainly composed of meadows and grassland, are identified and the local patterns of biomass are correctly reproduced.
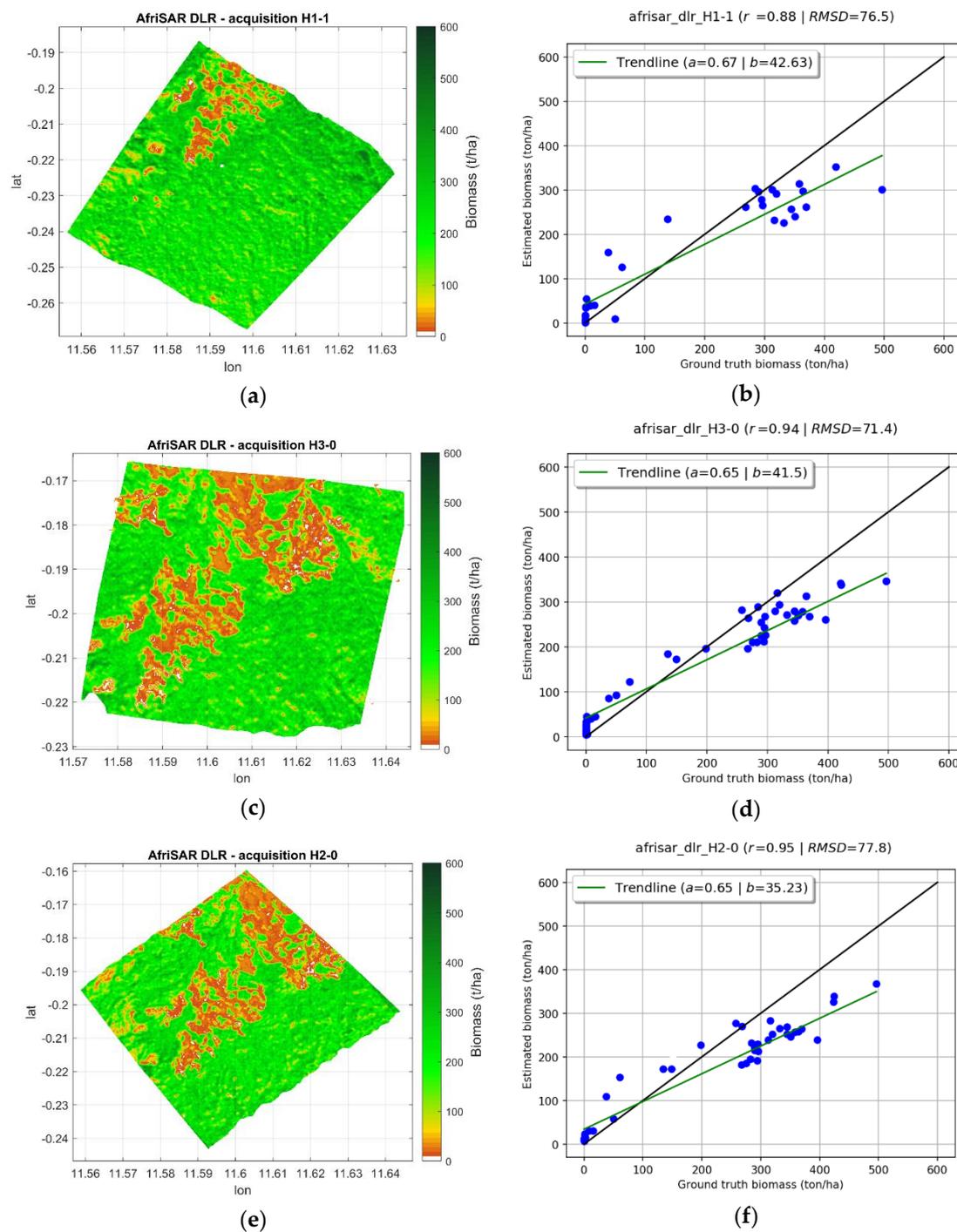
**Figure 7.** Examples of biomass maps estimated from AfriSAR DLR SAR images by using (**a**,**e**) the ANN algorithm and (**c**) the SVR algorithm. In (**b**,**d**,**f**) are shown the corresponding validation scatterplots.

## 6. Discussion

The comparison of the results obtained by ANN and SVR (Figures 5 and 6) did not indicate significant differences in the retrieval performances between the two methods. Both ANN and SVR exhibited similar accuracies, being able to estimate the target biomass with a slight underestimation of the values higher than 350 t/ha. Such underestimation can be attributed to the saturation of SAR data for biomass values higher than this threshold (see Figures 3 and 4), as already pointed out by past research [15]. It should be noted that both algorithms require only SAR data as input, without the need

of any ancillary information. The results of the general algorithms could be affected by the different instrumental setups that generated the clusters of data in Figure 4. Better retrievals could therefore be expected if data are collected by the same instrument, thus overcoming any intercalibration issue.

The computational cost of both ANN and SVR was also similar. The training was the only time-consuming step—for both algorithms, training each configuration using the BRIX dataset took a few minutes on a recent machine with an INTEL I7 6 Core processor, while applying the trained algorithm to the validation set occurred in near real time.

The validation results were in line with other studies (e.g., [22,23,25]), although differences in the test areas and datasets make the direct comparison difficult. For instance, the RMSE obtained by both ANN and SVR in the validation using BIOSAR1 data is in the same range reported in [22]. Similar conclusions can be drawn from the comparison between the ANN and SVR results obtained on equatorial forests (AfriSAR and TropiSAR datasets) and the results at P-band presented in [23]. It should be remarked that, in this case, both algorithms were also able to retrieve biomass beyond the 300 t/ha threshold indicated in [23], although with a slight underestimation of the higher values. Both algorithms can manage nonlinear relationships and, therefore, they are able to exploit the residual sensitivity of backscattering to biomass higher than 300 t/ha shown in Figure 3.

Concerning the disadvantages of these methods, the algorithm exportability to other areas should be verified before claiming a general validity. Indeed, depending on the experiment-driven training, the obtained results could be site-dependent and they could change significantly if applying the algorithms to other test areas. Previous studies indeed reported that retrieval errors of ML methods could be large if the test data are not properly represented in the training (e.g., [32]). However, updating the training with new data to enable the algorithms working on other areas is quite straightforward and it can be achieved without modifying the algorithm structure. Another possibility that will be investigated in the pursuance of this study is to train the algorithms by merging the experimental datasets with data simulated by electromagnetic models, such as the Water Cloud Model [54,55], for a wider range of forest conditions. Such a strategy should allow overcoming the site dependency of experiment-driven training, by obtaining more general algorithms, which can also retrieve the forest biomass with satisfactory accuracy in areas other than the ones considered in the training.

## 7. Conclusions and Future Work

In this study, two algorithms based on machine learning, namely ANN and SVR, were implemented, trained, and validated to estimate forest biomass from P- band airborne SAR data.

Both ANN and SVR exhibited similar retrieval performances, displaying a general capability of retrieving the target biomass with a slight underestimation of the values higher than 350 t/ha. Such underestimation can be attributed to some saturation exhibited by the SAR signal for the highest biomass values.

The characteristics of the available dataset suggested implementing general algorithms trained and tested on the entire dataset and specific algorithms for each test area.

The validation of the general algorithms resulted in R > 0.85 for both SVR and ANN, with RMSE ≃ 60–70 t/ha and bias negligible, while the validation of the specific algorithms resulted in R from 0.65 to 0.98 and RMSE between 11 and 65 t/ha, depending on the dataset and on the algorithm.

This investigation demonstrated the capability of machine-learning techniques for the remote sensing of forest biomass by using SAR. In this respect, ANN and SVR can be substantially considered equivalent in both retrieval accuracy and computational cost, since the investigation did not point out any aspect in which one of the two methods outperformed the other.

In the pursuance of this study, we plan to merge the experimental dataset with data simulated by electromagnetic forward models for training the algorithms. This strategy should overcome the limits of experiment-driven training, by filling the gaps in the training set and enabling the application to other areas.

## References

1.　Waring, H.R.; Running, S.W. *Forest Ecosystems. Analysis at Multiples Scales*, 3rd ed.; Academic Press: San Diego, CA, USA, 2007.

2.　Le Toan, T.; Laur, H.; Mougin, E. Multitemporal and dualpolarization observations of agricultural vegetation covers by X-band SAR images. *IEEE Trans. Geosci. Remote Sens.* **1989**, *27*, 709–718. [CrossRef]

3.　Ulaby, F.T.; Dobson, M.C. *Handbook of Radar Statistics for Terrain*; Artech House: Norwood, MA, USA, 1989.

4.　Wang, Y.; Kasischke, E.S.; Melack, J.M.; Davis, F.W.; Christensen, N.L. The effects of changes in loblolly pine biomass and soil moisture on ERS-1 SAR backscatter. *Remote Sens. Environ.* **1994**, *49*, 25–31. [CrossRef]

5.　Wang, Y.; Hess, L.H.; Filoso, S.; Melack, J.M. Understanding the radar backscattering from flooded and nonflooded Amazonian forests: Results from canopy backscatter modeling. *Remote Sens. Environ.* **1995**, *54*, 324–332. [CrossRef]

6.　Ranson, K.J.; Sun, G. Mapping biomass of a northern forest using multifrequency SAR data. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*, 388–396. [CrossRef]

7.　Paloscia, S.; Macelloni, G.; Pampaloni, P.; Sigismondi, S. The Potential of L- and C-Band SAR in Estimating Vegetation Biomass: The ERS-1 and JERS-1 Experiments. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 4. [CrossRef]

8.　Kasischke, E.S.; Melack, J.M.; Dobson, M.C. The use of imaging radars for ecological applications—A review. *Remote Sens. Environ.* **1997**, *59*, 141–156. [CrossRef]

9.　Ackermann, N.; Thiel, C.; Borgeaud, M.; Schmullius, C. Cosmo-SkyMed backscatter intensity and interferometric coherence signatures over Germany's low mountain range forested areas. In Proceedings of the Geoscience and Remote Sensing Symposium (IGARSS2012), Munich, Germany, 22–27 July 2012. [CrossRef]

10.　Kaasalainen, S.; Holopainen, M.; Karjalainen, M.; Vastaranta, M.; Kankare, V.; Karila, K.; Osmanoglu, B. Combining LiDAR and Synthetic Aperture Radar Data to Estimate Forest Biomass: Status and Prospects. *Forests* **2015**, *6*, 252. [CrossRef]

11.　Baghdadi, N.; Le Maire, G.; Bailly, J.-S.; Osé, K.; Nouvellon, Y.; Zribi, M.; Lemos, C.; Hakamada, R. Evaluation of ALOS/PALSAR L-Band Data for the Estimation of *Eucalyptus* Plantations Aboveground Biomass in Brazil. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2015. [CrossRef]

12.　Dobson, M.C.; Ulaby, F.T.; Le Toan, T.; Beaudoin, A.; Kasischke, E. Dependence of radar backscatter on coniferous forest biomass. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 412–415. [CrossRef]

13.　Le Toan, T.; Beaudoin, A.; Riom, J.; Guyon, D. Relating forest biomass to SAR data. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 403–411. [CrossRef]

14.　Le Toan, T.; Quegan, S.; Woodward, F.I.; Lomas, M.R.; Delbart, N. Relating radar remote sensing of biomass to modelling of forest carbon budgets. *Clim. Chang.* **2004**, *67*, 379–402. [CrossRef]

15.　Ferrazzoli, P.; Paloscia, S.; Pampaloni, P.; Schiavon, G.; Sigismondi, S.; Solimini, D. The potential of multifrequency polarimetric SAR in assessing agricultural and arboreous biomass. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 5–179. [CrossRef]

16.　Joshi, N.; Mitchard, E.T.A.; Brolly, M.; Schumacher, J.; Fernández-Landa, A.; Johannsen, V.K.; Marchamalo, M.; Fensholt, R. Understanding 'saturation' of radar signals over forests. *Sci. Rep.* **2017**, *7*, 3505. [CrossRef] [PubMed]

17.　Mermoz, S.; Réjou-Méchain, M.; Villard, L.; le Toan, T.; Rossi, V.; Gourlet-Fleury, S. Decrease of L-band SAR backscatter with biomass of dense forests. *Remote Sens. Environ.* **2015**, *159*, 307–317. [CrossRef]

18. Saatchi, S.S.; Moghaddam, M. Estimation of crown and stem water content and biomass of boreal forest using polarimetric SAR imagery. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 697–709. [CrossRef]

19. Moghaddam, M.; Saatchi, S.; Cuenca, R.H. Estimating subcanopy soil moisture with radar. *J. Geophys. Res.* **2000**, *105*, 14899. [CrossRef]

20. Le Toan, T.; Quegan, S.; Davidson, M.; Balzter, H.; Paillou, P.; Papathanassiou, K.; Plummer, S.; Saatchi, S.; Shugart, H.; Ulander, L. The BIOMASS Mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sens. Environ.* **2011**, *115*, 2850–2860. [CrossRef]

21. Christian, N.; Watanabe, M.; Hayashi, M.; Ogawa, T.; Shimada, M. Mapping the spatial-temporal variability of tropical forests by ALOS-2 L-band SAR big data analysis. *Remote Sens. Environ.* **2019**, *233*, 111372. [CrossRef]

22. Sandberg, G.; Ulander, L.M.H.; Fransson, J.E.S.; Holmgren, J.; Le Toan, T. L- and P-band backscatter intensity for biomass retrieval in hemiboreal forest. *Remote Sens. Environ.* **2011**, *115*, 2874–2886. [CrossRef]

23. Saatchi, S.; Marlier, M.; Chazdon, R.L.; Clark, D.B.; Russell, A.E. Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass. *Remote Sens. Environ.* **2011**, *115*, 2836–2849. [CrossRef]

24. Cazcarra-Bes, V.; Tello-Alonso, M.; Fischer, R.; Heym, M.; Papathanassiou, K. Monitoring of Forest Structure Dynamics by Means of L-Band SAR Tomography. *Remote Sens.* **2017**, *9*, 1229. [CrossRef]

25. Blomberg, E.; Ferro-Famil, L.; Soja, M.J.; Ulander, L.M.H.; Tebaldini, S. Forest Biomass Retrieval from L-Band SAR Using Tomographic Ground Backscatter Removal. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1030–1034. [CrossRef]

26. Notarnicola, C. A Bayesian Change Detection Approach for Retrieval of Soil Moisture Variations under Different Roughness Conditions. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 414–418. [CrossRef]

27. Pasolli, L.; Notarnicola, C.; Bruzzone, L. Estimating soil moisture with the support vector regression technique. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 1080–1084. [CrossRef]

28. Pierdicca, N.; Pulvirenti, L.; Pace, G. A Prototype Software Package to Retrieve Soil Moisture from Sentinel-1 Data by Using a Bayesian Multitemporal Algorithm. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 153–163. [CrossRef]

29. Del Frate, F.; Ferrazzoli, P.; Schiavon, G. Retrieving soil moisture and agricultural variables by microwave radiometry using neural networks. *Remote Sens. Environ.* **2003**, *84*, 174–183. [CrossRef]

30. Paloscia, S.; Pampaloni, P.; Pettinato, S.; Santi, E. A comparison of algorithms for retrieving soil moisture from ENVISAT/ASAR images. *IEEE Tran. Geosci. Remote Sens.* **2008**, *46*, 3274–3284. [CrossRef]

31. Paloscia, S.; Pettinato, S.; Santi, E.; Notarnicola, C.; Pasolli, L.; Reppucci, A. Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. *Remote Sens. Environ.* **2013**, *134*, 234–248. [CrossRef]

32. Santi, E.; Paloscia, S.; Pettinato, S.; Fontanelli, G. Application of artificial neural networks for the soil moisture retrieval from active and passive microwave spaceborne sensors. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *48*, 61–73. [CrossRef]

33. Santi, E.; Paloscia, S.; Pettinato, S.; Chirici, G.; Mura, M.; Maselli, F. Application of Neural Networks for the retrieval of forest woody volume from SAR multifrequency data at L and C bands. *Eur. J. Remote Sens.* **2015**, *48*, 673–687. [CrossRef]

34. Santi, E.; Paloscia, S.; Pettinato, S.; Fontanelli, G.; Mura, M.; Zolli, C.; Maselli, F.; Chiesi, M.; Bottai, L.; Chirici, G. The potential of multifrequency SAR images for estimating forest biomass in Mediterranean areas. *Remote Sens. Environ.* **2017**, *200*, 63–73. [CrossRef]

35. Evgeniou, T.; Pontil, M.; Poggio, T. Regularization Networks and Support Vector Machines. *Adv. Comput. Math.* **2000**, *13*, 1–50. [CrossRef]

36. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]

37. Pasolli, L.; Notarnicola, C.; Bruzzone, L.; Bertoldi, G.; Della Chiesa, S.; Hell, V.; Niedrist, G.; Tappeiner, U.; Zebisch, M.; Del Frate, F.; et al. Estimation of Soil Moisture in an Alpine Catchment with RADARSAT2 Images. Hindawi Publishing Corporation. *Appl. Environ. Soil Sci.* **2011**. [CrossRef]

38.  Fatoyinbo, L.; Pinto, N.; Hofton, M.; Simard, M.; Blair, B.; Saatchi, S.; Lou, Y.; Dubayah, R.; Hensley, S.; Armston, J.; et al. The 2016 NASA AfriSAR campaign: Airborne SAR and Lidar measurements of tropical forest structure and biomass in support of future satellite missions. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 4286–4287. [CrossRef]

39.  Hajnsek, I.; Scheiber, R.; Lee, S.; Ulander, L.; Gustavsson, A.; Tebaldini, S.; Monte Guarnieri, A. BioSAR 2007. Technical Assistance for the Development of Airborne SAR and Geophysical Measurements during the BioSAR 2007 Experiment: Final Report without Synthesis. Ph.D. Thesis, European Space Agency, Paris, France, 2008.

40.  Hajnsek, I.; Scheiber, R.; Keller, M.; Horn, R.; Lee, S.; Ulander, L.; Gustavsson, A.; Sandberg, G.; le Toan, T.; Tebaldini, S.; et al. BioSAR 2008. Technical assistance for the development of airborne SAR and geophysical measurements during the BioSAR 2008 Experiment DRAFT Final Report-BIOSAR Campaign. Ph.D. Thesis, European Space Agency, Paris, France, 2009.

41.  Ulander, L.M.; Gustavsson, A.; Dubois-Fernandez, P.; Dupuis, X.; Fransson, J.E.; Holmgren, J.; Wallerman, J.; Eriksson, L.; Sandberg, G.; Soja, M. Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 25–29 July 2011.

42.  Dubois-Fernandez, P.; Oriot, H.; Coulombeix, C.; Cantalloube, H.; du Plessis, O.R.; Le Toan, T.; Daniel, S.; Chave, J.; Blanc, L.; Davidson, M. TropiSAR a SAR data acquisition campaign in French Guiana. In Proceedings of the 8th European Conference on Synthetic Aperture Radar, Aachen, Germany, 7–10 June 2010; pp. 1074–1077.

43.  Dubois-Fernandez, P.; Le Toan, T.; Chave, J.; Blanc, L.; Daniel, S.; Oriot, H.; Arnaubec, A.; Réjou-Méchain, M.; Villard, L.; Lasne, Y.; et al. *Technical Assistance for the Development of Airborne SAR and Geophysical Measurements during the TropiSAR 2009 Experiment: Final Report*; TROPISAR-Final Report, ESA CONTRACT N° 22446/09/NL/CT; European Space Agency: Paris, France, February 2011.

44.  Mladenova, I.E.; Jackson, T.J.; Bindlish, R.; Hensley, S. Incidence Angle Normalization of Radar Backscatter Data. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1791–1804. [CrossRef]

45.  Villard, L.; le Toan, T. Relating P-Band SAR Intensity to Biomass for Tropical Dense Forests in Hilly Terrain: $\gamma^0$or$t^0$. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 214–223. [CrossRef]

46.  van Zyl, J.J. The effect of topography on radar scattering from vegetated areas *IEEE Trans. Geosci. Remote Sens.* **1993**, *31*, 153–160. [CrossRef]

47.  Schlund, M.; Davidson, M.W.J. Aboveground Forest Biomass Estimation Combining L- and P-Band SAR Acquisitions. *Remote Sens.* **2018**, *10*, 1151. [CrossRef]

48.  Available online: https://earth.esa.int/web/sppa/meetings-workshops/hosted-and-co-sponsored-meetings/brix (accessed on 26 February 2020).

49.  Hornik, K. Multilayer feed forward network are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]

50.  Linden, A.; Kinderman, J. Inversion of multi-layer nets. *Proc. Int. Joint Conf. Neural Netw.* **1989**, *2*, 425–443.

51.  Bruzzone, L.; Melgani, F. Robust multiple estimator system for the analysis of biophysical parameters from remotely sensed data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 159–174. [CrossRef]

52.  Pasolli, L.; Notarnicola, C.; Bertoldi, G.; Bruzzone, L.; Remelgado, R.; Greifeneder, F.; Niedrist, G.; Della Chiesa Tappeiner, U.; Zebisch, M. Estimation of Soil Moisture in Mountain Areas Using SVR Technique Applied to Multiscale Active Radar Images at C-Band. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**. [CrossRef]

53.  Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.

54.  Santoro, M.; Eriksson, L.; Askne, J.; Schmullius, C. Assessment of stand-wise stem volume retrieval in boreal forest from JERS-1 L-band SAR backscatter. *Int. J. Remote Sens.* **2006**, *27*, 3425–3454. [CrossRef]

55.  Attema, E.P.W.; Ulaby, F.T. Vegetation modeled as a water cloud. *Radio Sci.* **1978**, *13*, 357–364. [CrossRef]