Article

Assessment of Forest Above-Ground Biomass Estimation from PolInSAR in the Presence of Temporal Decorrelation

Nafiseh Ghasemi *, Valentyn Tolpekin and Alfred Stein

Faculty of Geo-Information Science and Earth Observation, University of Twente, 7500 AE Enschede, The Netherlands; v.a.tolpekin@utwente.nl (V.T.); a.stein@utwente.nl (A.S.)

* Correspondence: n.ghasemi@utwente.nl; Tel.: +31-534874444

Received: 16 April 2018; Accepted: 21 May 2018; Published: 24 May 2018

Abstract: In forestry studies, remote sensing has been widely used to monitor deforestation and estimate biomass, and it has contributed to forest carbon stock management. A major problem when estimating biomass from optical and SAR remote sensing images is the saturation effect. As a solution, PolInSAR offers a high coverage height map that can be transformed into a biomass map. Temporal decorrelation may affect the accuracy of PolInSAR and may also have an effect on the accuracy of the biomass estimates. In this study, we compared three different height estimation models: the Random-Volume-over-Ground (RVoG), Random-Motion-over-Ground (RMoG), and Random-Motion-over-Ground-Legendre (RMoGL) models. The RVoG model does not take into account the temporal decorrelation, while the other two compensate for temporal decorrelation but differ in structure function. The comparison was done on 214 field plots of the 10 m radius of the BioSAR2010 campaign. Different models relating PolInSAR height and biomass were developed by using polynomial, exponential, power series, and piece-wise linear regression. Different strategies for training and test subset selection were followed to obtain the best possible regression models. The study showed that the RMoGL model provided the most accurate biomass predictions. The relation between RMoGL height and biomass is well expressed by the exponential model with an average RMSE equal to 48 ton ha$^{-1}$ and $R^2$ value equal to 0.62. The relative errors for estimated biomass were equal to 46% for the RVoG model, to 37% for the RMoG, and to 30% for the RMoGL model. We concluded that taking the temporal decorrelation into account for estimating tree height has a significant effect on providing accurate biomass estimates.

Keywords: biomass; temporal decorrelation; PolInSAR height; accuracy; RVoG model; RMoG model; RMoGL model

1. Introduction

Forest biomass monitoring is important for several reasons: the critical role of forests on carbon stock and flux, quantification of the effect of deforestation on greenhouse gas emissions, and its use as an indicator of land cover change in UN-FAO's national statistics [1–3]. Remote sensing has been widely used in past decades to estimate biomass, monitor deforestation, and carbon stock management [4–6].

Remote sensing for estimating biomass can be divided into two-dimensional and three-dimensional methods. Two-dimensional methods use optical reflectance or SAR backscatter and relate it to biomass using regression [7–9]. Satellites such as LANDSAT and MODIS have been used for this purpose [10,11]. The main problem with these methods is that the signal is insensitive to the increase of biomass above a certain level, i.e., the saturation effect. Therefore, they can only be used for areas with low biomass such as savannas [9]. In the case of using SAR data, they can
be extended to medium-level biomass if L and P band or multi-temporal SAR data is used [12–15]. In contrast, biomass estimation using three-dimensional methods is more accurate since forest biomass is strongly related to the vertical structure of the trees [16–19]. Airborne and space-borne Lidar can provide an accurate height map of the forested areas that later can be converted into a biomass map [16,20,21]. In some studies, ground-based Lidar has been used to obtain biomass map as well [17,22]. Other means of acquiring a height map are photogrammetry, Interferometric SAR (InSAR) and Polarimetric Interferometric SAR (PolInSAR) [23,24]. PolInSAR records the reflected SAR signal from the same objects from two different points in two different times [25]. InSAR and PolInSAR give a height map of a lower accuracy than Lidar, but they cover large areas and can be used under all weather conditions [26,27]. InSAR can produce a tree height map in two ways, e.g., by subtracting a DEM from InSAR height or by using dual-wavelength InSAR i.e., TanDEM-X and E-SAR [19,28,29].

More advanced techniques such as TomoSAR and fusion of LiDAR and InSAR have also been used to estimate biomass [30–32]. These techniques provide more accurate results as compared to InSAR. The height estimation accuracy of PolInSAR and TomoSAR is directly affected by the presence of temporal decorrelation [33–35]. Temporal decorrelation is the change in SAR signal reflected from the objects changes in the position or attributes during the image acquisition time [36]. We may thus expect that the accuracy of biomass estimation improves after mitigating the temporal decorrelation. Recently, the Random-Motion-over-Ground (RMoG) model has been proposed to compensate the effect of temporal decorrelation on height estimation using PolInSAR. This model is based upon the RVoG model combined with modeling motion of trees in the vertical direction. It has shown promising results in improving height estimation accuracy up to 20% [33,37]. We have proposed the Random-Motion-over-Ground-Legendre (RMoGL) model based on the RMoG model, but with the Fourier–Legendre series as the structure function instead of the simple exponential function used in the RMoG model [38].

The main objective of this paper is to analyze the accuracy of biomass estimation after correcting for temporal decorrelation. To do so, the first step is to develop a model to estimate biomass from PolInSAR height maps resulting from the Random-Volume-over-Ground (RVoG), RMoG, and RMoGL models. We have selected the Remningstorp forest as a study area as it has been studied during BioSAR2010 campaign [39] as the test site. Both linear and non-linear models were developed to clarify the relation between PolInSAR height and biomass. The second step is to apply different strategies for train and test dataset selection to make sure that the chosen model is general and accurate enough [40]. At the final step, for evaluating the results both field data and Lidar data were used.

This paper has been organized in the following sections: first, a short introduction of biomass estimation using remote sensing in general and SAR data specifically is given. Second, the dataset and study area is characterized. In the third section, explanation of the methodology is given. In the fourth section, the results are presented followed by the discussion and the conclusion sections.

2. Materials and Methods

A summary of the methodology used in this study is presented in Figure 1. Hereafter, the methodology is described in details.
Figure 1. Flowchart of the methodology used in this study.

2.1. Study Area

A complete report about the BioSAR2010 campaign and Remningstorp area has been published previously [39]. Here, a summary is provided that is relevant for this study. The Remningstorp forest (58°30′N, 13°40′E) is located in the southwestern part of Sweden and has an area of approximately 1200 ha. The forest type is hemi-boreal, which intermediates between boreal and temperate forest types [41]. Dominant forest species are Norway Spruce and Scots pine, and it is completed by a mixture of oak, birch, and aspen. The elevation range of this area is between 120 and 145 m above sea level.

2.2. Field and Lidar Data

Field observations used in this study have been collected resulting in 214 circular plots with a radius of 10 m [39]. The location of these field plots is shown in Figure 2.

Figure 2. Location of field plots on Remningstorp area image. Google earth V 6.0. Remningstorp, Sweden. The center coordinates are 58°30′N and 13°40′E, and the area is 1200 ha. Eye alt 3.58 km.
The area has been divided into several stands based upon the homogeneity of forest cover type. Field plots have been defined in such a way that they fall into one of these stands completely to minimize the effect of heterogeneity within the plots.

Measurements inside each plot according to the Heureka forestry system [42] include an H100 height, defined as the basal area weighted average of the 10 highest trees in each plot [18], Diameter at Breast Height (DBH), i.e., the diameter at 1.3 m above ground, and the dominant tree species. According to the recorded information, 60% of the field plots is covered by spruce, 30% covered by pine, and 10% dominated by oak, birch, and aspen (mixed deciduous) trees. Additionally, dry aboveground biomass, including stems, branches, bark, and needles, has been measured [39]. For all 10 m radius field plots, biomass was between 6 and 250 ton ha\(^{-1}\) with an average of 105 ton ha\(^{-1}\). The allometric equation used for measuring biomass is based upon height and DBH [11]. Allometric equations are the equations that relate biophysical parameters to biomass. At the plot level, the general form of this equation equals

\[
B = N \cdot \pi \cdot \left( \frac{1}{2} DBH \right)^2 \cdot H \cdot \rho \cdot f
\]

(1)

where \(B\) is the above-ground biomass, \(N\) is the number of trees per area unit, \(H\) is the tree height, \(\rho\) is the species-related wood density, and \(f\) is a form factor. These equations have been developed for different forests according to the FAO standard procedure [43]. Developing these equations requires excessive field work and the data are only valid for a 5-year period. After this period, the density and the factor \(f\) change, so the measurements are repeated every five years. In this paper, we used the previously developed allometric equations for the test site [39].

Lidar data of the study area were acquired with an average density of 69 returns m\(^{-2}\). The airborne Lidar data has been used for two main purposes. First for evaluating height estimation models i.e., the RVoG, RMoG, and RMoG\(_L\) models. Second, the biomass map produced from Lidar data [20,39,41] was used for the assessment of generated biomass maps from PolInSAR heights beside the field data. The predicted biomass map from Lidar height is presented in Figure 3.

![Figure 3](image-url)
Besides predicting the biomass map, we used the Lidar Digital Surface Model (DSM) with a cell size of \(0.5 \times 0.5\) m, as the reference height map. Since we have the canopy height map, the Lidar DSM is called the Canopy-Height-Model (CHM) hereafter. This map was derived from the differences between the first and the last returns of the Lidar pulses and is displayed in Figure 4. Both the predicted biomass map and the CHM have been used as the reference data sets.

![Figure 4](image_url)

**Figure 4.** Available CHM from Lidar of the Remningstorp area. Dark circles show the location of field plots.

The areas with zero height values on the Lidar CHM are masked on the predicted biomass map. The areas with taller trees on the Lidar CHM correspond to areas with the highest biomass values. The two maps show a clear positive correspondence between tree height and biomass.

### 2.3. PolInSAR Data

During the BioSAR 2010 campaign, 10 PolInSAR images were obtained. These images were collected with the ONERA SETHI airborne sensor, developed by the Office National d’Etudes et de Recherches Aérospatiales (ONERA) [39]. This device can acquire full Polarimetric Interferometric SAR images in both L and P bands.

Three pairs of P-band images, acquired specifically for PolInSAR analysis, were chosen for this study. Their spatial baseline is equal to 30 m, the heading angles equal 199°, 178°, and 270°, respectively, and the sensor height is approximately equal to 4000 m. The images have been delivered in SLC (Single Look Complex) format and have four polarization channels: \(HH\), \(HV\), \(VH\), and \(VV\) with \(HV = VH\) as the system is mono-static. The ground-range resolution of the InSAR data is \(0.5 \times 0.5\) m.

Next, after correcting for the flat earth phase [44], the complex coherence was obtained according to [25]. The optimized coherence channels plus linear polarizations \((HH, HV, \text{ and } VV)\), circular polarization \((LL, LR, \text{ and } RR)\), and the Pauli basis polarization \((HH + VV \text{ and } HH − VV)\) channels [45] served as input for the RVoG, RMoG, and RMoGL models. These models are applied on each pair of P-band images separately and the final height maps are generated by mosaicking the resulting images.
2.4. Tree Height Estimation Using PolInSAR Data

Since the Polarimetric Interferometric SAR systems illuminates an area from two different positions at two different times, any change in the objects is represented in the form of signal decorrelation. Thus, the height of the trees, like other properties, can be retrieved by reversing the process of interferograms generation, e.g., quantifying the decorrelation sources. In the past decade, the models used for obtaining tree properties were only taking volumetric decorrelation into account. Volumetric decorrelation is the inconsistency of the signal caused by the vertical structure of the trees. In the presence of natural phenomena in particular wind, changes in the position of the scatterers cause temporal decorrelation. This should be taken into account when estimating tree height. The following models have been applied to estimate height without and with accounting for temporal decorrelation.

2.4.1. The RVoG Model

The RVoG model has been popular over the past decade for estimating forest height [46–48]. It relates the observed complex coherence to the height of the vegetation layer with the assumption that only volumetric decorrelation is present. It expresses the volumetric coherence as

\[
\gamma_v = e^{j\phi_g} \left[ \mu \left( \frac{p_1 (e^{h_v p_2} - 1)}{p_2 (e^{h_v p_1} - 1)} \right) + 1 \right]
\]  

(2)

where \( \phi_g \) is the ground phase, \( h_v \) is the vegetation height, and \( \mu \) is the ground-to-volume-ratio [47]. The parameter \( \mu \) is added to compensate for the surface scattering mechanism effects on the observed coherence. It can have any value between 0 and \( \infty \) [47]. The parameters \( p_1 \) and \( p_2 \) are defined as

\[
p_1 = \frac{2K_e}{\cos(\theta - \theta_s)}
\]  

(3)

\[
p_2 = p_1 + jk_z
\]  

(4)

where \( K_e \) is the mean wave extinction coefficient \((0 < K_e < 1)\), \( \theta \) is the average look angle \((0 < \theta < 90^\circ)\), \( \theta_s \) is the terrain slope angle \((0 \leq \theta_s \leq 90^\circ)\), and \( k_z \) is the vertical wave number \((0 < k_z < 1)\). The final form of complex coherence by the RVoG model \( \gamma_R \) is

\[
\gamma_R = e^{j\phi_g} \gamma_v + \frac{\mu}{1 + \mu}.
\]  

(5)

To solve Equation (5), we need to estimate \( \mu \), so we need a polarization channel as the volume and another channel as the ground scattering representation. Solving Equations (2)–(5) is done with a multi-stepwise procedure following [47].

As the RVoG model ignores temporal decorrelation, it provides biased results if the temporal decorrelation is high. This can happen due to wind or precipitation. To tackle this limitation, the RMoG model has been proposed [37], which models the induced motion caused by wind in the vertical direction.

2.4.2. The RMoG Model

The RMoG model assumes that the vegetation layer consists of randomly distributed vertical scattering objects located on a rough dielectric layer. The main difference with the RVoG model here is to assume that both layers have random movements along the vertical axis [33]. The complex coherence using the RMoG model \( \gamma_M \) is obtained as

\[
\gamma_M = \frac{\int_0^{h_v} \rho(z) \exp(jk_z z) \exp \left(-\frac{1}{2} \left( \frac{4\pi}{\lambda} \right)^2 \sigma_z^2(z) \right) dz}{\int_0^{h_v} \rho(z) dz}
\]  

(6)
where \( \sigma_v^2(z) \) is defined as
\[
\sigma_v^2(z) = \sigma_0^2 + (\sigma_0^2 - \sigma_0^2) \frac{z - h_r}{h_r}.
\]  
(7)

Here, \( h_r \) is the reference height, which is an arbitrary constant value [33], \( \lambda \) is the SAR system wavelength, and \( \sigma_0 \) and \( \sigma_v \) are the surface and canopy layer motion standard deviations, respectively. The function \( \rho(z) \) is the structure function, which defines the vertical structure of the vegetation layer. Similar to the RVoG model, the RMoG model assumes a Gaussian function to represent the structure of the trees. The procedure of solving Equation (6) is explained in details in [33].

Here, the primary assumption is that the maximum of the structure function occurs at the top of the canopy layer and the volume backscatter and from ground components are distinguishable. This assumption leads to biased height estimation especially in complex and multi-layer forests [37]. To improve the accuracy of the RMoG model, and inspired by [48], we have proposed the RMoG model [38].

2.4.3. The RMoG Model

The RMoG model is similar to the RMoG model, but instead of an exponential function, it uses a finite Fourier–Legendre series as the structure function when considering the motion of scatterers along the vertical axis [38]. Thus, the complex coherence of the RMoG model \( \gamma_{ML} \) equals
\[
\gamma_{ML} = e^{jh_v} \frac{1}{f_{-1}^1 f(z') \exp(1 + f(z')) \exp(i k_v z') \exp(-i \frac{1}{4} k^2 \sigma_v^2(z')) dz'}{f_{-1}^1 (1 + f(z')) dz'}.
\]
(8)

Here, \( z' = \frac{z}{h_v} - 1, k_v = \frac{k h_v}{2} \), and \( \sigma_v(z') = \sigma_v \frac{h_v(1 + z')}{2} \) [38]. The \( a_i \) are normalized Legendre coefficients, and the \( f_i, i = 1, 2, \ldots, n \) are Fourier functions of different orders [48] defined as
\[
f_i(z') = \sum_{i=0}^{n} a_i P_i(z')
\]
(9)

where the \( P_i(z') \), \( i = 1, 2, \ldots \) are Legendre polynomials [49]. If we assume that \( \sigma_v = 0 \) according to the short time interval between image acquisition times, the final form of complex coherence obtained by the RMoG model equals
\[
\gamma = \frac{\gamma_{ML}}{h + 1}.
\]
(10)

The number of terms in \( f_i(z') \) depends upon the vegetation layer; for more complex vegetation layers, several terms are needed, whereas a single layer can be modeled by fewer terms. Note that, for estimating more than two Legendre coefficients, we need multi-baseline SAR data. It is argued in [48] that using the first and second terms of the series should be sufficient in most cases; the accuracy, however, depends upon the complexity of the vegetation layer. According to the relatively simple structure of the hemi-boreal forest and the availability of PolInSAR data, we fixed the number of terms to two. Therefore, parameters to be estimated are \( \phi_R, h_v, a_1, a_2, \) and \( \sigma_v(z') \). The way of solving Equation (10) is described in [38]. The solving method is similar to the RMoG model, but some extra steps should be taken in order to deal with more number of unknown parameters.

2.5. Biomass Estimation Using PolInSAR Data

The most important outcome of the RVoG, RMoG, and RMoG model is a height map of the area. Averaged, plot-wise height values are then extracted from these height maps. In this study, we applied a buffer zone of 10 m to reduce the border effect.

We used a robust regression model to retrieve biomass from PolInSAR height. The used algorithm is non-linear least square fitting with LAR (Least Absolute Residuals) algorithm for providing the prediction bounds within a 95% of confidence interval [50,51]. According to past studies, there are
different assumptions about the relation between biomass and PolInSAR height. The first and most common assumption is that the biomass can be obtained from PolInSAR height by a linear equation [9]. In other studies, this relation has been shown to be non-linear [41]. Therefore, to find the best model for estimating biomass, we started by a polynomial model to relate biomass to PolInSAR height. The general form of this model is

$$\ln(B) = \sum_{i=1}^{n+1} \alpha_i H^{n+1-i}$$  \hspace{1cm} (11)

where $n$ is the polynomial degree, $B$ is biomass, $H$ is the PolInSAR height obtained from the RVoG, RMoG, and RMoG_L models, and the $\alpha_i$ are coefficients to be estimated. The parameters $B$ and $H$ are standardized and thus are dimensionless. We determined the number of coefficients by trial and error, and the final form arrived at a value of $n = 3$. The polynomial model is

$$\ln(B) = a_1 H^3 + a_2 H^2 + a_3 H + a_4.$$  \hspace{1cm} (12)

We stopped to increase $n$ if $a_i \leq 0.005$ and consequently $a_i H^{n+1-i} \leq 0.15$ according to [52]. As is evident from Equation (12), the relation between biomass and PolInSAR is non-linear.

Another common assumption is that biomass increases proportionally to the increase of height. Therefore, as a further examination, the one-term and two-term exponential equations were tested. Furthermore, we tested whether the power series can describe the relation between two variables. The general form of the selected exponential model equals

$$\ln(B) = \beta_1 \left(1 - e^{-\beta_2 H}\right),$$  \hspace{1cm} (13)

and for the power series it equals

$$\ln(B) = \zeta_1 H^{\zeta_2}$$  \hspace{1cm} (14)

where the $\beta_i$ and $\zeta_i$, $i = 1, 2$ are the coefficients to be estimated.

The reliability of Equations (12)–(14) is based upon the assumption that the observed height values have a normal distribution of errors. In our case, according to [39,41], outliers occur in measured biomass values. Robust regression [53] was therefore applied to avoid the effect of outliers on estimated parameters. It assigns a weight value to each data point with weights being adjusted iteratively. In the first iteration, one value, in our case 0.5, is assigned to all data points, and by using normal least squares the coefficients are estimated. In the next iterations, the weights are re-estimated to place less emphasis on the points at a larger distance from model predictions. Next, the model coefficients are recalculated using weighted least squares. This procedure proceeds until the difference between each estimated coefficient in two different iterations is below a threshold, selected as 0.005 in this study [41].

2.5.1. Data Splitting

The selection of a training and a test data set directly affects the reliability of estimated parameters. Different strategies exist for solving the so-called “data splitting” problem [40,54]. It is important to choose a good subset of data for training that preserves the generality of the model while having high accuracy. These two aspects, however, are conflicting. For determining the best training subset, we applied two procedures. First, we selected 75%, 50%, and 25% of the data points randomly as the training set and carried out the regression. Second, the field plots were divided into three groups based upon measured biomass value as low (<120 ton ha$^{-1}$), medium (120–240 ton ha$^{-1}$), and high (>240 ton ha$^{-1}$). Half of the data points from each category was then used as training data and the other half as test data. These two approaches, called the random and stratified methods, respectively, were compared and the best method was selected. The modeling procedure was repeated for mean
diameter and basal area to understand how structural properties of trees are related to the height obtained from PolInSAR images.

2.6. Accuracy Assessment

To compare PolInSAR height with ground truth data, the first step is to convert the height maps from slant-range to ground-range. This was done using a precise DEM available in the campaign data and using the suggested method in [39]. We used both field and Lidar data as the reference height map. For evaluating the accuracy of three different height estimation models, the average height inside each plot was compared with the corresponding values on Lidar CHM and with the H100 value of each plot. Different metrics were used for evaluating the RVoG, RMoG, and RMoGL models. To test the distribution of measured and obtained biomass maps and for comparing them, the Kolmogorov–Smirnov (KS) test was applied [55]. A one-sample KS test examined the null hypothesis that the data distribution is a hypothesized distribution against the alternative that it does not follow such distribution. The test statistic is the maximum absolute difference between the empirical Cumulative Distribution Function (CDF) calculated from the dataset $\hat{F}(x)$ and the hypothesized CDF of $G(x)$, in our case the normal distribution:

$$D = \max(|\hat{F}(x) - G(x)|).$$ (15)

A two-sample KS test checks if the two datasets have similar distributions. The alternative is that the two tested datasets have different distributions.

3. Results

A summary of the most important biophysical parameters measured within field plots is provided in Table 1 and a histogram in Figure 5. These values are reported for the whole dataset, whereas the regression analyses results are based on the selected training data as described in Section 2.5.1.

Table 1. Summary statistics of the biophysical parameters measured within plots with 10 m radius.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomass (ton ha$^{-1}$)</td>
<td>120.04</td>
<td>1.2–315.96</td>
</tr>
<tr>
<td>Basal area (m$^2$ ha$^{-1}$)</td>
<td>26.34</td>
<td>1.35–77.81</td>
</tr>
<tr>
<td>Mean diameter (m)</td>
<td>2.39</td>
<td>0.52–5.05</td>
</tr>
<tr>
<td>H100 (m)</td>
<td>28.24</td>
<td>13.77–34.77</td>
</tr>
</tbody>
</table>

Figure 5. Distribution of forest plot biomass for the 214 plots at the test site.
According to Table 1, the measured biomass varies from 1.2 to 315.96 ton ha\(^{-1}\). The mean value is 120 ton ha\(^{-1}\) and the median value is 105 ton ha\(^{-1}\), showing that the distribution of the field data deviates from the normal distribution. Other measured biophysical parameters i.e., basal area and mean diameter also have a wide range of variation.

The histogram is skewed and shows several peaks. This makes it difficult to fit a specific distribution function to biomass values. Most plots have a biomass around 100–130 ton ha\(^{-1}\), whereas the second peak occurs between 200 and 230 ton ha\(^{-1}\). This multi-modal histogram validates our choice to divide the data into different categories and to select a training data set for each category. The relation between measured biomass and biomass predicted by Lidar is presented in Figure 6. The relative error is approximately equal to 12.7%.

The plot shows a positive correlation between measured biomass and predicted biomass by Lidar height. Two points marked with circles are most likely measurement errors.

After processing SLC PolInSAR images and obtaining complex coherence [25], height maps using the RVoG, RMoG, and RMoG\(_L\) models were obtained. The height map generated by the RMoG\(_L\) model is shown in Figure 7.

![Figure 6. Relation between measured and predicted biomass using Lidar data with $R^2 = 0.60$ and RMSE = 0.68 ton ha\(^{-1}\). The points marked with dark circles are possibly the measurement errors.](image)

![Figure 7. Resulting height map from the RMoG\(_L\) model.](image)
Figure 7 shows a good agreement with the Lidar CHM of Figure 4. For shorter trees, we see that the RMoG\text{L} model overestimates tree heights, while for taller trees the height is underestimated. Plot-wise averaged height values from three different models were compared to the corresponding H100 values. The results showed that $R^2 = 0.43$ for the RVoG model, $R^2 = 0.47$ for the RMoG model, and $R^2 = 0.48$ for the RMoG\text{L} mode. The weak correlation between PolInSAR height and H100 is because H100 is measured based on the tallest trees inside each field plot while the PolInSAR height represents the average of canopy height. Similarly, height maps obtained from the RVoG, RMoG, and RMoG\text{L} models were compared with Lidar CHM, and results are shown in Figure 8. The Lidar and PolInSAR heights show a positive correlation. In the case of the RMoG model, the noise has decreased and the data points are closer to the least square line. We also observe that the PolInSAR height estimation error increases with increasing tree heights.

![Figure 7](image1)

![Figure 8](image2)

**Figure 8.** Relation (in red) between PolInSAR height resulting from (a) the RVoG model ($R^2 = 0.50$, RMSE = 0.67 m), (b) the RMoG model ($R^2 = 0.69$, RMSE = 0.60 m), and (c) the RMoG\text{L} model ($R^2 = 0.78$, RMSE = 0.55 m) with the corresponding averaged height values from Lidar. The black line presents 1:1 line.

Next, we selected training and test sets for performing regression analyses. The stratified sampling leads to more accurate biomass estimation. All results hereafter have been obtained using stratified sampling. The relation between PolInSAR height and biomass was examined using different models. First, we fitted a polynomial model to relate the PolInSAR height and biomass. The results are listed in Table 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\alpha_4$ (Constant Term)</th>
<th>RMSE (ton ha$^{-1}$)</th>
<th>Adjusted $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVoG</td>
<td>-0.028</td>
<td>-0.063</td>
<td>0.498</td>
<td>4.630</td>
<td>30.87</td>
<td>0.50</td>
</tr>
<tr>
<td>RMoG</td>
<td>-0.014</td>
<td>-0.083</td>
<td>0.498</td>
<td>3.660</td>
<td>30.80</td>
<td>0.62</td>
</tr>
<tr>
<td>RMoG\text{L}</td>
<td>0.005</td>
<td>-0.102</td>
<td>0.467</td>
<td>2.694</td>
<td>30.75</td>
<td>0.73</td>
</tr>
</tbody>
</table>
The highest $R^2$ and lowest RMSE values were obtained by the RMoG$_L$ model. Another important observation is that, if we set $H = 0$, then $B \neq 0$. The $B$ value shows the error of the fitted polynomial model for estimating biomass from PolInSAR height. We can force a no-intercept regression to estimate biomass, but the RMSE value then becomes high (0.75 ton ha$^{-1}$). Next, we correlated height obtained by PolInSAR and biomass via an exponential model and a power series. The results of fitting the exponential model are listed in Table 3.

Table 3. Coefficients of the fitted exponential model (within 95% confidence interval).

<table>
<thead>
<tr>
<th></th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>RMSE (ton ha$^{-1}$)</th>
<th>Adjusted $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVoG</td>
<td>5.15</td>
<td>0.13</td>
<td>30.87</td>
<td>0.54</td>
</tr>
<tr>
<td>RMoG</td>
<td>5.10</td>
<td>0.17</td>
<td>30.75</td>
<td>0.56</td>
</tr>
<tr>
<td>RMoG$_L$</td>
<td>5.10</td>
<td>0.18</td>
<td>30.61</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Similar to polynomial regression, the RMoG$_L$ model is the best model, with the lowest RMSE and highest $R^2$ value. Previous studies [52] suggest using an exponential model to estimate biomass from height. We observed that the RMSE is higher than when using the polynomial model, and $R^2$ is lower. In contrast with the third-degree polynomial, the exponential model passes through the origin, hence a zero biomass value for $H = 0$. In addition, the power series was tested to estimate biomass, and the results are presented in Table 4.

Table 4. Coefficients of the fitted power series (within 95% confidence interval).

<table>
<thead>
<tr>
<th></th>
<th>$\zeta_1$</th>
<th>$\zeta_2$</th>
<th>RMSE (ton ha$^{-1}$)</th>
<th>Adjusted $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVoG</td>
<td>2.068</td>
<td>0.277</td>
<td>40.01</td>
<td>0.56</td>
</tr>
<tr>
<td>RMoG</td>
<td>2.194</td>
<td>0.274</td>
<td>30.93</td>
<td>0.60</td>
</tr>
<tr>
<td>RMoG$_L$</td>
<td>2.224</td>
<td>0.276</td>
<td>30.87</td>
<td>0.65</td>
</tr>
</tbody>
</table>

We obtain higher RMSE values by using power series in comparison with the polynomial and exponential models. According to the variation of the data set and to provide further comparison, the piece-wise regression was also tested. In finding the break points of the dataset, the target is to find $y$ given the number of break points $x$, which minimizes the following function

$$p = |B - \text{interp}(H_b, B_b, H)|^2.$$  \hspace{1cm} (16)

Here, interp is the interpolation function, $H_b$ is the break point, and $B_b$ is the interpolation point for the $H_b$ [56]. The results of piece-wise linear regression relating biomass to the RMoG$_L$ height are listed in Table 5.

Table 5. The slope and intercept of piece-wise linear regression for the RMoG$_L$ model.

<table>
<thead>
<tr>
<th>Piece-Wise Regression</th>
<th>Slope</th>
<th>Intercept</th>
<th>Average RMSE (ton ha$^{-1}$)</th>
<th>Adjusted $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H &lt; 8$ m</td>
<td>1.73</td>
<td>0</td>
<td>30.91</td>
<td>0.90</td>
</tr>
<tr>
<td>$8$ m $\leq H$</td>
<td>0.04</td>
<td>3.20</td>
<td>30.64</td>
<td>0.62</td>
</tr>
</tbody>
</table>

The piece-wise regression showed good results as compared to the power-series and polynomial models listed in Tables 2 and 3. for $H > 8$ m, the slope of the line decreases significantly, showing the saturation effect. For both exponential and piece-wise regression models, the saturation effect happens when biomass reaches approximately 300 ton ha$^{-1}$.

The polynomial, exponential, power series, and piece-wise linear regression curves for the RMoG$_L$ model are shown in Figure 9.
Figure 9. Relation between logarithm of measured biomass and RMoG<sub>L</sub> height resulting from the (a) polynomial ($R^2 = 0.42$), (b) exponential ($R^2 = 0.67$) with prediction bounds within a 95% confidence interval calculated by the LAR method (c) power series models ($R^2 = 0.54$), and (d) piece-wise linear model ($R^2 = 0.60$). The biomass values on vertical axis represent $\ln(B)$.

We notice a saturation effect for $H > 8$ m in Figure 9b–d. Figure 8a does not show the saturation point clearly, which is in line with other studies [52].

According to Tables 2–5 and Figure 9, the most accurate model is the exponential model, while the piece-wise regression also showed good results. Since it is more straightforward to fit a single curve to the whole dataset, we selected the exponential to produce the biomass map. For accuracy assessment, the remaining 52 plots out of 214 field plots were used as the test data set. Different metrics were employed for this purpose [57,58]. These metrics included bias ($\mu$), co-variance ($\sigma$), root-mean-square-error ($RMSE$), and the coefficient of determination ($R^2$). The results are listed in Table 6.

<table>
<thead>
<tr>
<th></th>
<th>$\mu$ (ton ha&lt;sup&gt;−1&lt;/sup&gt;)</th>
<th>$\sigma$ (ton ha&lt;sup&gt;−1&lt;/sup&gt;)</th>
<th>RMSE (ton ha&lt;sup&gt;−1&lt;/sup&gt;)</th>
<th>Adjusted $R^2$</th>
<th>Relative Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVoG</td>
<td>36.81</td>
<td>0.72</td>
<td>62.06</td>
<td>0.45</td>
<td>46</td>
</tr>
<tr>
<td>RMoG</td>
<td>27.12</td>
<td>1.97</td>
<td>55.44</td>
<td>0.74</td>
<td>37</td>
</tr>
<tr>
<td>RMoG&lt;sub&gt;L&lt;/sub&gt;</td>
<td>15.97</td>
<td>2.09</td>
<td>46.65</td>
<td>0.82</td>
<td>30</td>
</tr>
</tbody>
</table>

Biomass values of the test dataset were compared to corresponding predicted biomass by the Lidar data, and the results are plotted in Figure 10.
Figure 10. Relation between measured biomass by PolInSAR data and predicted biomass by Lidar data for the test dataset: (a) the RVoG model with $R^2 = 0.60$ and RMSE = 30.85 (ton ha$^{-1}$), (b) the RMoG model with $R^2 = 0.73$ and RMSE = 30.73 (ton ha$^{-1}$), and (c) the RMoGL model with $R^2 = 0.92$ and RMSE = 30.64 (ton ha$^{-1}$).

Furthermore, the relation between PolInSAR height and mean diameter on the one hand and basal area on the other have been examined to gain insight into how these biophysical parameters affect height measurements. The results are shown in Figure 11. The results showed a significantly better performance of the RMoGL model on test data.

Figure 11. The relationship (a) between the PolInSAR height from the RMoGL model and the basal area with $R^2 = 0.61$ and (b) between this height and the mean diameter with $R^2 = 0.81$ in red.
The results of applying two-dimensional KS-test between predicted biomass by PolInSAR and Lidar data are listed in Table 7, whereas the biomass map resulting from the RMoG\textsubscript{L} model is shown in Figure 12. The results showed a positive linear relation between mean diameter, basal area, and PolInSAR height, whereas the relation between mean diameter and basal area are stronger. This shows that inclusion of the mean diameter in obtaining biomass estimation models increases the accuracy of the biomass estimation.

Table 7. Parameters of the two-dimensional KS test between biomass predicted by Lidar and PolInSAR.

<table>
<thead>
<tr>
<th></th>
<th>h</th>
<th>p</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVoG</td>
<td>0</td>
<td>0.001</td>
<td>0.41</td>
</tr>
<tr>
<td>RMoG</td>
<td>0</td>
<td>0.000</td>
<td>0.31</td>
</tr>
<tr>
<td>RMoG\textsubscript{L}</td>
<td>0</td>
<td>0.000</td>
<td>0.21</td>
</tr>
</tbody>
</table>

The parameter $h$ equals 0 if the null hypothesis of having similar CDFs is not rejected and equals 1 if the CDFs are not similar at the 0.05 significance level. The parameter $D$ is the distance between two dataset distribution functions and shows the goodness of fit. From Table 7, we can see the $D$ decreased by 50% by applying the RMoG\textsubscript{L} model.

Figure 12. Resulting biomass map from the RMoG\textsubscript{L} model.

To provide a better comparison, the predicted biomass by PolInSAR height was also compared to the measured biomass of the field plots. The results are shown in Figure 13. The results shows a positive linear relation between PolInSAR predicted biomass and field biomass for all three height estimation models. Figure 13c shows better performance of the RMoG\textsubscript{L} model, while the other models show good correlation as well.
Figure 13. Relation between predicted biomass by PolInSAR data and measured biomass in field data: (a) the RVoG model with $R^2 = 0.55$ and RMSE = 30.97 (ton ha$^{-1}$), (b) the RMoG model with $R^2 = 0.69$ and RMSE = 30.76 (ton ha$^{-1}$), and (c) the RMoGL model with $R^2 = 0.74$ and RMSE = 30.69 (ton ha$^{-1}$). Red line is 1:1 line.

4. Discussion

In this paper, we investigated whether correcting for temporal decorrelation improves biomass estimation accuracy. From the BioSAR2010 campaign, we selected plots with a 10 m radius, in line with other studies [41]. Recorded values for these plots showed that biomass and other structural parameters had a broad range of varieties. This feature enabled us to examine how the PolInSAR height is related to different levels of biomass. Moreover, the effect of saturation was observed.

For evaluating three different height estimation models, the H100 and Lidar CHM were used as reference data. Neither Lidar nor PolInSAR height showed a significant correlation with the H100 parameter. The reason is that H100 represents the tallest trees, while the PolInSAR and Lidar height in each pixel represents the average of tree heights. In contrast, the Lidar and PolInSAR height values showed a strong relation. As we expected and as was shown previously, the relative error of height estimation decreased from 35% for the RVoG model to 18.6% for the RMoGL models.

Since there is no agreement in previous studies whether the correlation between biomass and tree height is linear or not, we examined different regression models. The outcome of the training sets should be a model that can predict the values for the test data set accurately and meanwhile is able to
predict new data. If the model is over-trained, the accuracy will be good but the model would not be able to produce good results with new data. These two aspects, accuracy and generality, is known as the bias and variance dilemma [54]. It becomes important when dealing with a high volume of data and when the developed model is being used in estimating biomass at the global scale. In this study, we tried two different strategies for data splitting to evaluate the effect of training set data selection on estimating biomass. The results showed a better performance of stratified sampling, but the computational time increased threefold. The best result is achieved by fitting an exponential model. The exponential curve and the piece-wise regression show a saturation effect for a biomass level above 300 ton ha\(^{-1}\). The saturation level increases from 100 ton ha\(^{-1}\) when using SAR backscatter [59] to 300 ton ha\(^{-1}\) by using PolInSAR height as the main predictor. In tropical forests, it is important to consider the saturation effect since biomass is mostly higher than 300 ton ha\(^{-1}\).

The RMSE of biomass estimation decreased from 46% when using height resulting from the RVoG model to 30% for the RMoG\(_L\) model. As expected, there is a considerable improvement in estimating biomass after modeling temporal decorrelation. The results validated the hypothesis of the improvement of biomass estimation after correcting for temporal decorrelation.

Our results were compatible with previous studies for the same test site. For example, InSAR images at C-band were employed to estimate stem volume and achieved an RMSE value of 27% [60]. Similar studies using L and P-bands estimated biomass with RMSE values between 30–40% for L-band and between 20–30% for P-band [57,61]. Interestingly, recent research in Remningstorp area using TanDEM-X interferometric heights showed an RMSE value around 16.5% [41]. This improvement can be due to a larger number of baselines and a high number of field plots. Other studies in tropical [62], temperate [63,64], boreal, and hemi-boreal forests [65] showed an average error between 2 and 5 m with and average of 3.5 m in estimating tree heights using PolInSAR. This shows a PolInSAR height can be used to estimate biomass in different forest types.

The RMSE only, however, cannot be enough to show the model performance. Based on Table 6, the RVoG model has a significant bias, which, after correcting for temporal decorrelation, has decreased by about 50%. Moreover, the correlation increased from 0.45 for the RVoG model to 0.82 for the RMoG\(_L\) model. For the RMoG model, the results were close to the RMoG\(_L\) model. The difference is because of the more accurate structure function in the RMoG\(_L\) model.

The relation between the RMoG\(_L\) height and two other structural parameters, namely basal area and mean diameter, was also high. This is due to the capability of P-band in penetrating into the vegetation layer as well as the Fourier–Legendre series employed in the RMoG\(_L\) model to represent the structure of trees. This is an interesting finding since it shows the importance of taking different vegetation structures into account for estimating biomass. Involving other biophysical parameters for example the D\(_{BH}\) in estimating biomass can have a huge effect on improving accuracy.

The result of the one-dimensional KS test showed a different distribution of measured biomass and predicted biomass by all height estimation models and Lidar data. The two-dimensional KS test showed similar distribution functions for predicted values by Lidar and PolInSAR data. The main reason is that, at the plot level, biomass is measured for individual trees, while remotely sensed data measure the canopy of trees that are spread around the tree trunk.

There is a requirement for methods that can estimate biomass at broader scales. The method and dataset we used here are limited because we used one pair of PolInSAR images. By using multi-baseline tomographic PolInSAR images, we can add more terms to the structure function of the RMoG\(_L\) model. Therefore, reconstruction of the tree structure has a higher resolution. We expect that this leads to more accurate biomass maps, but with an increased computation time.

There are other biophysical parameters of trees i.e., diameter and basal area that affect the SAR signal. The ideal case would be to develop a model based on height, diameter, and the basal area of the trees. In the case of using PolInSAR data, the only observable parameter is tree height. One possibility is to use the field data together with PolInSAR height, but in doing so the developed model would
only be applicable to this area and we lose any generality. Thus, we chose to develop a model based on only PolInSAR height to maintain the general applicability of the model [41,52].

Our findings showed a strong relation between PolInSAR height and above-ground biomass. The accuracy of biomass estimation improved considerably after correcting for temporal decorrelation. This means that the accuracy of height estimation is directly related to the biomass estimation accuracy. This shows the importance of mitigating temporal decorrelation in estimating height using PolInSAR since such mitigation affects the further products of height maps such as biomass. At present, the P-band images, on which we tested our proposed model, are only available in airborne sensors. Although they will also be available on spaceborne sensors in the near future, it would be of great value to examine the effect of the correction of temporal decorrelation on biomass estimation accuracy using other wavelengths, especially the L-band. As a sequel to this study, we suggest developing a unique structure function for each tree type according to the availability of the forest classification map. This will lead to more accurate biomass maps [66].

5. Conclusions

This study showed that, after mitigation of temporal decorrelation, the accuracy of biomass estimation improved by approximately 10%. Therefore, for producing biomass maps of a desirable accuracy, the temporal decorrelation factor should be considered. The best model selected for estimating biomass from PolInSAR was the exponential curve. Furthermore, we showed, in forests with diverse levels of biomass, that it is a good approach to divide the area according to the different height classes and use different regression lines for each class. By using PolInSAR height, the saturation level increases from 150 to 300 ton ha$^{-1}$. This is important in tropical forests where biomass is high and saturation becomes an issue.

A weak relation was observed between remotely sensed height maps and H100 values. This is due to the sensitivity of remote sensing sensors to the average height of the tree canopy versus the H100, which represents the tallest trees inside each plot. Basal area and mean diameter also affect the SAR signals, and consequently PolInSAR height estimates.

The presented results demonstrate the capability of PolInSAR to obtain biomass maps. This is important since, in the near future, the satellite BIOMASS mission aims to produce biomass maps with wall-to-wall coverage of important forested areas.

Author Contributions: The study has been carried out by N.G. as a part of her PhD project with the help of V.T. as her supervisor. A.S. has been supervising, editing and correcting the results and the text of the paper.

Acknowledgments: The authors would like to thank the European Space Agency (ESA) for providing BioSAR2010 campaign data.

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

References


60. Askne, J.; Santoro, M. Experiences in boreal forest stem volume estimation from multitemporal C-band InSAR. In Recent Interferometry Applications in Topography and Astronomy; InTech: London, UK, 2012.

© 2018 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 (http://creativecommons.org/licenses/by/4.0/).