Fully Polarimetric L-Band Synthetic Aperture Radar for the Estimation of Tree Girth as a Representative of Stand Productivity in Rubber Plantations

Bambang H. Trisasongko 1,2,* , Dyah R. Panuju 1,2 , Amy L. Griffin 3 and David J. Paull 4

1 Department of Soil Science and Land Resources, Bogor Agricultural University, Jalan Meranti, Bogor 16680, Indonesia; panuju@apps.ipb.ac.id
2 Geospatial Information and Technologies for the Integrative and Intelligent Agriculture (GITIIA), Center for Regional System Analysis, Planning and Development (CRESTPENT/P4W), Bogor Agricultural University, Jalan Pajajaran, Bogor 16127, Indonesia
3 School of Science, RMIT University, Melbourne 3001, Australia; amy.griffin@rmit.edu.au
4 School of Science, UNSW Canberra, Campbell 2610, Australia; d.paull@unsw.edu.au
* Correspondence: trisasongko@apps.ipb.ac.id

Abstract: This article explores a potential exploitation of fully polarimetric radar data for the management of rubber plantations, specifically for predicting tree circumference as a crucial information need for sustainable plantation management. Conventional backscatter coefficients along with Eigen-based and model-based decomposition features served as the predictors in models of tree girth using ten regression approaches. The findings suggest that backscatter coefficients and Eigen-based decomposition features yielded lower accuracy than model-based decomposition features. Model-based decompositions, especially the Singh decomposition, provided the best accuracies when they were coupled with guided regularized random forests regression. This research demonstrates that L-band SAR data can provide an accurate estimation of rubber plantation tree girth, with an RMSE of about 8 cm.

Keywords: circumference; extreme gradient boosting; girth; polarimetric decomposition; rubber; regularized random forests

1. Introduction

Monitoring systems that can capture datasets systematically over wide areas are needed to investigate forested land covers. This is an area where spaceborne Earth observation platforms perform well. They permit frequent, repetitive monitoring of ever-changing woodlands at broad scales with remarkably diverse spatial resolutions, spanning from hundreds of meters to sub-meter pixel spacing. In fragmented areas, such as in populous regions of South East Asia, the detection and investigation of land cover alterations, especially in surrounding protected areas, is paramount for establishing usable land cover forecasts.

The dynamics of human activities in forested landscapes need special attention because of their undesirable environmental impacts. These include the development of plantations, especially monoculture plantations. Environmental conditions surrounding oil palm plantations, for instance, have been debated [1], and some question their environmental sustainability. Plantations are generally developed in areas experiencing tropical deforestation or degradation, which is well-known in South East Asia, Latin America, and, recently, in African regions. Large-scale plantations require rigorous, spatially aware systems capable of capturing and reporting these dynamics in remote areas.

While investigations of tropical rainforests using remote sensing remain important, monitoring schemes have also become increasingly crucial in plantations. Research attention has focused on pine [2,3], oil palm [4,5], and timber (acacia) plantations [6]. Although
they have not been as important as oil palm within the context of the local/regional economy and the environment, rubber plantations rank second in importance in Indonesia after oil palm. They occupy 3.6 million hectares countrywide (Statistik Perkebunan Indonesia Komoditas Karet/Indonesian Plantation Statistics for Rubber Commodities, released by the Ministry of Agriculture in 2016), a scale that requires a substantial amount of research.

Plantations require recurrent monitoring for at least two key reasons. The first is to safeguard the environment from their impacts. In Indonesia, in particular, many plantations are situated adjacent to or inside forested areas. In some cases, they share borders with conservation areas, including national parks. The potential of monitoring plantations employing spaceborne remote sensing data for this purpose has long been studied [7,8]. These studies generally conclude that Earth observation satellites are indispensable for safeguarding the environment. The second reason is related to the operation of plantations. This mission comprises the observation of diseases, nutrient uptake, and biomass estimation. Tree health has been investigated, including *Syrex noctilio* invasion in pine trees [9]. A survey of the literature indicates many studies have focused on pine plantations [10,11], while remote sensing studies that monitor tree health among other types of plantations are rarer. To improve precision fertilization, remote sensors have also been employed, including in oil palm plantations [12]. Greater research effort has been directed to estimating biomass, which may have been driven by the need to advance global biomass assessment by including Trees Outside Forest (TOF, [13]) areas. This terminology shares a similar meaning with Other Wooded Land (OWL), as introduced by Galidaki et al. [14]. There has been abundant study of pine tree biomass using various space-borne sensors [3,15]. In the tropics, including in Indonesia, studies estimating tree biomass have typically been implemented in oil palm and acacia plantations. The latter has been of increasing attention due to its rapid growth [6] and because of its consequent impacts on the surrounding environment.

Rubber plantation monitoring has benefited from a wealth of Earth observation data, including both passive and active sensors. The long-term availability of multispectral data can assist with mapping and monitoring the dynamics of rubber estates. This has been widely proven in various environments [16–18]. While multispectral data are indispensable for rubber plantation monitoring, atmospheric disturbances such as cloud cover can reduce their value. Synthetic Aperture Radar (SAR) sensors provide an alternative data source. They could be used independently or in conjunction with multispectral data in the framework of data or information fusion [16,19]. Information extraction from a single SAR sensor benefits from the availability of multi-polarization SAR data. Experiments on plantation mapping have been conducted with promising results [20–22]. This work has been extended to estimating aboveground biomass to support the improvement of global biomass datasets, including a test case in Java [23]. Meanwhile, managing rubber plantations tends to focus more upon measuring tree girth or circumference [24] and their dynamics. Nonetheless, the rarity of related studies is clear and motivates the present study.

The primary goal of this article is to examine statistical learning models for estimating tree girth from SAR data as an indicator of tree productivity. We compared conventional regression methods to recently developed approaches such as random forests and support vector machines. Selected models retrieved using several fully polarimetric datasets were then evaluated to investigate the presence of negative estimates.

2. Materials and Methods

2.1. Test Site

We selected the Subang regency in West Java province, Indonesia, for this research (Figure 1). A field survey was conducted in Jalupang and Wanareja estates, which have been considered to be the best rubber estates in Indonesia. Both estates share similar soil characteristics, i.e., dominated by Inceptisols with some Ultisols, which suit the establishment of rubber plantations. The Köppen climate classification for both sites is Aw, which is a typical climate in the western part of Java. Therefore, this area is generally suitable for rubber plantations. The terrain in the research area is generally flat in the northern part of
the test site and undulating with gentle slopes in the south. The average altitude is about 100 m above mean sea level.

![Research site in Java, Indonesia.](image)

**Figure 1.** Research site in Java, Indonesia.

### 2.2. Datasets and Analysis

Fully polarimetric data acquired from Phased Array L-band SAR 2 (PALSAR-2) were used for this analysis. The data were provided by Japan Aerospace Exploration Agency (JAXA) (Tokyo, Japan) in Level 1.1 (Single Look Complex, SLC), imaged on 15 April 2015. The scene was in ascending mode; hence, the datasets were marginally impacted by Faraday rotation. To re-assure quality in our dataset, Faraday rotation was calculated using the Bickel and Bates technique [25] prior to the analysis of the data. Estimated rotations were approximately 3 degrees, which is below the proposed threshold of 5 degrees [26].

The availability of fully polarimetric SAR data in SLC format is crucial to deriving polarimetric features to improve the results of classification and regression. These features can be integrated with a more popular form of fully polarimetric SAR data, i.e., backscatter coefficients. This research employed linearly polarized backscatter coefficients, i.e., HH, HV, and VV polarizations, where H and V indicate horizontal and vertical linearly polarized waves, respectively. With PALSAR-2’s monostatic configuration, this study only considered HV polarization data.

A fully polarimetric SAR dataset includes signal measurements combining linearly horizontal and vertical wave propagation. A common representation of this is the Sinclair (or Scattering) matrix:

$$S = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix}$$

(1)

Each matrix component contains amplitude and phase measurements. To be able to further describe the associated phenomenon, $S$ is generally decomposed into distinctive, independent components using decomposition theorems. Applying the reciprocity theorem (i.e., $S_{hv} = S_{vh}$), the Pauli scattering vector $k_P$ can be established.

$$k_P = [ \begin{bmatrix} S_{hh} + S_{vv} & S_{hh} - S_{vv} & S_{hv} + S_{vh} \end{bmatrix} ]^T$$

(2)

Some theorems were developed using the coherence matrix, requiring a conversion from the scattering matrix. For this purpose, the Pauli scattering vector can be multiplied with its transposed conjugate, yielding a $3 \times 3$ coherence $[T]$ matrix.

Parameters obtained from decomposition theorems were calculated using the PolSARpro software package from the European Space Agency (ESA). This research only considered incoherent decomposition methods, which have been the focus of contemporary research [27–30]. This research evaluated incoherent decomposition approaches, including the widely used Eigen-based or Cloude–Pottier [31] method. Decomposing the coherence matrix is done by using a unitary matrix, allowing the extraction of three eigenvalues.
Based on these eigenvalues, Cloude–Pottier proposed the first decomposition feature, the entropy \((H)\).

\[
H = \sum_{i=1}^{n} -P_i \log P_i; \quad P_i = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}
\]

(3)

Entropy measures the randomness of scattering objects, represented by a value ranging from zero to one. Zero entropy indicates a unique scatterer observable within the processed group of pixels. Random scatterers, for instance, returning waves from woody vegetation, would have an entropy value of one.

The secondary parameter, mean Alpha angle, signifies the type of scatterer, which is derived from the unitary matrix.

\[
\alpha = \sum_{i=1}^{n} P_i \alpha_i
\]

(4)

A value of zero designates an odd-bounce scattering process, including incoming waves from smooth surfaces. Dipole scattering, such as wave returns from woody vegetation and double-bounce scattering from human-made structures, are observed with alpha angle values around 45 and 90 degrees, respectively. Pure entropy is hardly observable. It is, therefore, crucial to assess an additional Cloude–Pottier parameter, i.e., the anisotropy, which is vital for investigating the contribution of secondary eigenvalues [31]. A small anisotropy value indicates a lesser importance of secondary scattering. Anisotropy is defined as:

\[
A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3}
\]

(5)

The second approach is model-based decomposition, which has been extensively used since its introduction by Freeman and Durden [32]. They suggest the significance of estimating single- and double-bounce as well as scattering mechanisms from vegetation for studying surface objects, hence:

\[
[T] = f_s[T_{\text{single}}] + f_d[T_{\text{double}}] + f_v[T_{\text{vol}}] + \ldots
\]

(6)

Further advancement is generally related to the modification of the model and introducing additional parameters. This includes the helix component proposed by Yamaguchi et al. [33], which served as the basis of further progress in polarimetric decomposition theorems, for instance, the four-component model [30] and the six-component model [29].

In total, six datasets were evaluated in this research, i.e., backscatter coefficients, Entropy-Alpha angle (Cloude–Pottier), three-component Freeman–Durden, four-component Yamaguchi, and four-component Singh decomposition features, and a mixture of all polarimetric features.

Similar to raw SAR data, datasets in SLC format are projected in radar geometry. To match with existing maps, conversion to Earth geometry was first done. To reduce the influence of terrain, this research implemented the Range-Doppler Terrain Correction using the SNAP software tools, employing Shuttle Radar Topography Mission (SRTM) one-arc-second data. The original data have a pixel size in the azimuth and range directions of about 3.22 m and 5.13 m, respectively. This spacing was increased to 10 m to diminish geolocation error during resampling. A 7 × 7 boxcar filter was used to minimize SAR speckle noise.

After a preliminary survey to obtain access to the plantation, field data collection was undertaken from December 2015 to January 2016 to assemble in situ data from plantation managers and ground measurements. Field data collection at the research site allowed elementary information to be gathered at the stand level, including age and latex yields. In situ data were measured at locations with equal spatial distribution of stand age. Ground data were collected from 2–30 years old stands. In this research, circular plots with a 15 m radius were used. Tree spacing was similar, with most stands using 7 m × 4 m spacing.
in a northerly direction. While individual tree height was also measured, in this article, only circumference is reported as stand productivity in rubber plantations is assessed primarily by tree circumference. Although this research used a monoculture plantation, stand statistics varied greatly within a stand. For instance, in a stand at full production, the girth ranged between 40 cm and 94 cm, with a mean of 66 cm. The deviation was lower in stands younger than about five years of age, an age before canopies become connected. This may be due to competition between trees for obtaining sunlight, as rubber trees prefer full light exposure.

Thirty-three plots were measured and separated into two groups: 75% of total plots were exploited for model building (training), while the remainder were used for validation. We documented measurements in each plot by taking photographs in four directions. To better apprehend the canopy condition of rubber stands at various age classes, we also collected vertical photographs (Figure 2). Collected tree circumferences spanned from 11 cm to 100 cm, with the median girth at about 43 cm.

![Figure 2. Field documentation of vertical photographs (left column) and stand conditions (right column).](image-url)
Machine learning models were assessed using the aforementioned training data. In this research, Multivariate Adaptive Regression Splines (MARS), Cubist, and M5 were used as the benchmarks. Nawar et al. [34] found that pixel-wise soil salinity was best modeled through MARS. Similar findings were shown in studies of oil palm [35] or forest biomass [36]. It is important to note, however, that similarity in performance could exist among models derived from MARS and artificial neural networks (ANN) [37]. Cubist has also been utilized in diverse applications, including the estimation of soil organic carbon [38] and chlorophyll content [39]. For remotely sensed land surface temperature assessment, M5 was found to be more robust than artificial neural networks in the estimation of reference evapotranspiration [40].

This article also examined two popular methods, i.e., Support Vector Regression (SVR) and Random Forests Regression (RFR). The support vector machine technique [41] was initially used for classification and regression. The Radial Basis Function (RBF) kernel was assessed as it has the versatility to adapt to complex variables. Random Forests, proposed by Breiman [42], has been employed for estimating aboveground biomass. In addition, two variants of RF (regularized RF [43] and conditional inference RF [44]) were used to investigate their performance against the original RF model.

Less-examined models such as Bayesian Regularized Neural Networks (BRNN), Extreme Learning Machine (ELM), and Extreme Gradient Boosting (XGB) were included in the analysis to improve the understanding of their accuracy. BRNN has been employed in research on sub-pixel land cover [45], landslide detection and assessment [46], and surface temperature [47]. Studies involving ELM and its extensions have recently been growing, with numerous applications [48–51]. A variant of a tree-based model, the XGB, was proposed by Chen and Guestrin [52] to accommodate sparse data. In remote sensing, the performance of this approach has been investigated in drought [53], forest [54], and crop [55] studies.

Model performance was assessed using common measures, i.e., coefficient of determination ($R^2$) and the root mean square error (RMSE). Chosen models were then implemented for the whole image using the R Statistical Software using the “raster” and “caret” packages.

3. Results
3.1. Performance of Backscatter Coefficients

The performance of models using SAR backscatter coefficients to estimate tree girth is presented in Table 1. The PALSAR-2 backscatter coefficients dataset reliably delivered a high degree of accuracy in the validation dataset, with the mean $R^2$ of 0.71 and RMSE = 11.90 cm. As expected, the benchmark (MARS, Cubist, and M5) models did not produce the best accuracies, although they provided a more accurate model than did the more recently developed ELM approach. Backscatter coefficients accurately delivered the best model when combined with random forests or support vector machines, although it appears that the difference, in terms of RMSE, was not large.

It appears that the strong relationship was due to the wide span of HV backscatter, which assisted in better distinguishing various girth measurements. The finding is consistent with previous work, indicating that SAR applications for vegetation would likely benefit from cross-polarization [32]. In the case of SVM, its contribution was very significant, with an RMSE = 8.76 cm and $R^2 = 0.79$. A secondary contribution was due to VV polarization (RMSE = 9.29 cm; $R^2 = 0.87$). Neumann et al. [56] argued that the backscatter coefficients could estimate tree and canopy structure with negligible influence from ground-trunk interactions. This research, however, benefited from considering ground-trunk attenuation, especially in juvenile trees where canopy development was incomplete.

3.2. Accuracy Analysis of Decomposition Features

Employing Eigen-based decomposition features for model development was not particularly successful. The highest $R^2$ was about 0.70, with RMSE = 10.18 cm, obtained
using the Cubist algorithm. On average, the Cloude–Pottier entropy, mean alpha angle, and anisotropy datasets could only yield an RMSE of about 15.73 cm, with a mean $R^2$ of 0.31.

Table 1. Accuracy of the backscatter coefficients dataset. Bolded values indicate the best performing models.

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Abbreviation</th>
<th>RMSE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multivariate Adaptive Regression Splines</td>
<td>MARS</td>
<td>10.72</td>
<td>0.65</td>
</tr>
<tr>
<td>Cubian</td>
<td>Cub</td>
<td>12.05</td>
<td>0.70</td>
</tr>
<tr>
<td>Weka-M5</td>
<td>M5</td>
<td>13.88</td>
<td>0.64</td>
</tr>
<tr>
<td>Bayesian Regularized Neural Networks</td>
<td>NNbr</td>
<td>14.56</td>
<td>0.65</td>
</tr>
<tr>
<td>Extreme Learning Machine</td>
<td>NNetm</td>
<td>15.61</td>
<td>0.51</td>
</tr>
<tr>
<td>Random Forests</td>
<td>RF</td>
<td>9.86</td>
<td>0.82</td>
</tr>
<tr>
<td>Regularized RF</td>
<td>RFg</td>
<td>9.48</td>
<td>0.81</td>
</tr>
<tr>
<td>Conditional Inference RF</td>
<td>RFc</td>
<td>12.75</td>
<td>0.75</td>
</tr>
<tr>
<td>Support Vector Machine with RBF kernel</td>
<td>SVMrb</td>
<td>8.82</td>
<td>0.77</td>
</tr>
<tr>
<td>Extreme Gradient Boosting</td>
<td>GBMxgb</td>
<td>11.29</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Model-based decomposition techniques have been increasingly developed as an alternative to Eigen-based approaches. Their main parameters, i.e., single-bounce, double-bounce, and volume scattering, serve as the main dataset for object characterization and analysis. Consistent with research findings given in Trisasongko et al. [20], this type of decomposition provides important variables for data analysis. Three decomposition techniques were compared, and the variation of outcomes is presented in Table 2.

Table 2. Accuracy of model-based decomposition features. Bolded values indicate the best performing models.

<table>
<thead>
<tr>
<th>Models</th>
<th>Freeman–Durden RMSE</th>
<th>R²</th>
<th>Yamaguchi RMSE</th>
<th>R²</th>
<th>Singh RMSE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARS</td>
<td>17.69</td>
<td>0.84</td>
<td>16.23</td>
<td>0.80</td>
<td>16.50</td>
<td>0.78</td>
</tr>
<tr>
<td>Cub</td>
<td>12.05</td>
<td>0.70</td>
<td>12.05</td>
<td>0.70</td>
<td>12.05</td>
<td>0.70</td>
</tr>
<tr>
<td>M5</td>
<td>13.88</td>
<td>0.64</td>
<td>13.88</td>
<td>0.64</td>
<td>13.88</td>
<td>0.64</td>
</tr>
<tr>
<td>NNbr</td>
<td>14.56</td>
<td>0.65</td>
<td>14.56</td>
<td>0.65</td>
<td>14.56</td>
<td>0.65</td>
</tr>
<tr>
<td>NNetm</td>
<td>18.27</td>
<td>0.74</td>
<td>17.13</td>
<td>0.79</td>
<td>18.08</td>
<td>0.82</td>
</tr>
<tr>
<td>RF</td>
<td>12.14</td>
<td>0.80</td>
<td>10.65</td>
<td>0.80</td>
<td>8.97</td>
<td>0.89</td>
</tr>
<tr>
<td>RFg</td>
<td>11.52</td>
<td>0.81</td>
<td>10.24</td>
<td>0.81</td>
<td>8.31</td>
<td>0.91</td>
</tr>
<tr>
<td>RFc</td>
<td>9.63</td>
<td>0.81</td>
<td>11.89</td>
<td>0.61</td>
<td>11.60</td>
<td>0.60</td>
</tr>
<tr>
<td>SVMrb</td>
<td>11.28</td>
<td>0.65</td>
<td>10.22</td>
<td>0.67</td>
<td>9.70</td>
<td>0.71</td>
</tr>
<tr>
<td>GBMxgb</td>
<td>9.50</td>
<td>0.81</td>
<td>18.40</td>
<td>0.66</td>
<td>10.13</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Table 2 indicates that, on average, model-based decomposition techniques yield sufficiently high accuracies ($R^2 = 0.74$, RMSE = 12.99 cm). As shown, the conventional Freeman–Durden model has a strong chance to obtain above-average accuracy. A similar situation occurs with Yamaguchi decomposition. The results also indicate that the Singh algorithm provides a much-improved estimation. Overall, the two most promising models were derived from Singh decomposition features. They were built using random forests and RFg, a new variant of random forests. The latter is more favorable since it provides an RMSE = 8.31 cm with $R^2$ about 0.91, which is slightly better than the conventional random forests model’s performance (RMSE = 8.97 cm; $R^2 = 0.89$). Both random forests models outperform support vector machines with a radial basis function kernel (RMSE = 9.70 cm; $R^2 = 0.71$).

3.3. Feature Stack

The limited number of variables captured in remote sensing data often inhibits achieving higher accuracy. This has frequently been overcome by adding derivative variables. For optical datasets, vegetation indices such as Normalized Difference Vegetation Index
(NDVI) are repeatedly reported to have a favorable impact [37]. In radar remote sensing, a similar idea was implemented, for instance, the Pope index for backscatter coefficients data [38]. With the era of polarimetric decomposition, combining backscatter coefficients and decomposition features could be seen as an analogue.

The outcome of stacking all polarimetric variables is presented in Table 3. The result suggests that extreme gradient boosting yielded the best accuracy with stacked datasets, providing further evidence of the strength of ensemble tree models. Although its $R^2$ is the greatest from all of the experiments, its difference to Singh decomposition features coupled with the global random forests method is clearly not large. Taking computation time into account as a weighing factor, the combined Singh features and global random forests is clearly the best choice because of the greater computation times associated with the extreme gradient boosting technique. This finding also suggests that stacking features does not always result in the best outcome.

Table 3. Accuracy of stacking all polarimetric variables. Bolded values indicate the best performing models.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>RMSE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARS</td>
<td>15.60</td>
<td>0.51</td>
</tr>
<tr>
<td>Cub</td>
<td>19.66</td>
<td>0.37</td>
</tr>
<tr>
<td>M5</td>
<td>12.04</td>
<td>0.73</td>
</tr>
<tr>
<td>NNbr</td>
<td>14.56</td>
<td>0.65</td>
</tr>
<tr>
<td>NNeim</td>
<td>20.66</td>
<td>0.17</td>
</tr>
<tr>
<td>RF</td>
<td>8.82</td>
<td>0.86</td>
</tr>
<tr>
<td>RFg</td>
<td>9.69</td>
<td>0.83</td>
</tr>
<tr>
<td>RFC</td>
<td>15.03</td>
<td>0.64</td>
</tr>
<tr>
<td>SVMrb</td>
<td>17.72</td>
<td>0.06</td>
</tr>
<tr>
<td>GBMxgb</td>
<td>8.58</td>
<td>0.92</td>
</tr>
</tbody>
</table>

3.4. Prediction

Based on the $R^2$, the most promising model used extreme gradient boosting with the combined dataset, i.e., 0.92 with RMSE = 8.58 cm. Nonetheless, further investigation indicates that a slightly lower $R^2$ (0.91) with an improved RMSE (8.31 cm) was presented by the combination of Singh decomposition features and the regularized random forests model. In terms of computing time, the latter is preferable. Faster computation is due to the smaller number of polarimetric variables used in regularized random forests modeling, i.e., only four variables in the Singh decomposition, instead of sixteen in the original feature stack. This research again indicated that combining variables does not always benefit data analysis, which is consistent with previous findings [20].

Figure 3 depicts a comparison of estimated and reference (ground-measured) data, suggesting that the two selected models exhibit similar patterns. By considering computing time, a model developed using Singh features coupled with the RFg method was selected for developing a girth map. With the paucity of current literature on the performance of RFg in similar remote sensing data analyses, this finding was unable to be compared to other studies. It should be noted, however, that RFg was shown to be useful for the selection of key variables from hyperspectral data for detecting invasive weeds [59]. This technique was also invaluable for investigating growing stock volumes in temperate forests [60], thanks to its capability for handling multicollinearity. The results suggest that the accuracy of RFg was not much different from that achieved by the original RF model ($R^2 = 0.91$, RMSE = 8.97 cm). While a conditional inference RF model was shown to be useful in previous crop production research [61], this study showed that this was not the case in rubber tree girth estimation in a plantation. RFC only yielded an $R^2 = 0.60$ with RMSE = 11.60 cm, which is far below the accuracy of other RF models. The present study indicates the importance of tree-based models for regression problems, similar to conclusions drawn by previous research [62–64]. The final prediction based on RFg is presented in Figure 4. As shown, no negative estimation is identified, with the minimum estimated girth being about 10 cm.
A better choice among the two is yet to be conclusively identified since contributions of volume scattering component has a dominant influence on the final outcome, followed by double-bounce scattering. It is well understood that the tree canopy and its underlying structures (twigs and branches) relate strongly to the volume component [6], or in the case of backscatter coefficients, the HV polarization. Double-bounce scattering is likely to be prominent in the case of juvenile stands where the tree canopy is sparse, exposing ground conditions. In this situation, the dominant wave-tree interaction is double-bounce scattering.

While several variants of statistical models are available in the literature, this research indicates that RF and SVM remain the best options for classification or regression problems. A better choice among the two is yet to be conclusively identified since contributions of inputs (numbers and variable types, etc.) and the complexity of the problem (designated outputs) continue to be the keys to producing successful outcomes. Some previous research has favored SVM [65], while other studies found that RF provided more accurate results [66,67]. The variability of these findings suggests that large-scale comparisons should be made in the future, involving a variety of environments. From an implementation perspective, however, end-users benefit from having access to several statistical models. Libraries in python or the R programming language are readily available, despite minimal graphical user interface (GUI) development for end-users.

Analysis of the importance of different variables in this model indicates that the volume scattering component has a dominant influence on the final outcome, followed by double-bounce scattering. It is well understood that the tree canopy and its underlying structures (twigs and branches) relate strongly to the volume component [6], or in the case of backscatter coefficients, the HV polarization. Double-bounce scattering is likely to be prominent in the case of juvenile stands where the tree canopy is sparse, exposing ground conditions. In this situation, the dominant wave-tree interaction is dou-

Figure 3. Relationship between predicted and measured girth from the best models: Singh with the RFg model (left) and all variables combined with the XGB model (right). The blue line shows the relationship between predicted and measured tree girth.

Figure 4. Estimated circumferences using Singh decomposition features combined with the guided regularized random forests technique.

4. Discussion
ble bounce—a reflective wave from the ground, which subsequently interacts with tree stems. Adam et al. [68] proposed an interesting approach to investigate an early phase of maize infestation, i.e., using regularized random forests to select robust variables as inputs to random forests modeling. Future SAR research could adopt this approach to select robust decomposition theorems prior to data analysis, either for classification or regression problems.

This result agrees with previous results in Acacia plantations [6]. Nonetheless, we noted that Sai Bharadwaj et al. [69] experienced that polarimetric decomposition features derived from a model-based theorem were unable to yield a useful statistical model using the same sensor. This problem may have roots in the discrepancy between when field data measurements were made and when the remote sensing data were acquired. We should mention, however, that wave saturation should be further examined to allow a deeper understanding of saturation limits.

Rapid developments in SAR data processing and analysis in recent decades suggest a need for the expansion of current research related to pre-processing, transformation, and information extraction. Radar auto-focus, for instance, has been further developed through deep learning methodologies [70]. Polarimetric decompositions, applied to hybrid and fully polarimetric data, need to be further explored to allow better options for data analysis. Extracting information, as the primary goal of the whole process, has benefited from rapid progress in data science, in particular machine and deep learning.

5. Conclusions

Remote sensors dedicated to earth observation have long been employed to monitor the Earth’s remaining forest cover. Despite this, the conditions and the dynamics of woodlands, including plantations and mixed gardens, are the least studied land covers. Within the context of plantations, a larger amount of remote sensing research is needed to effectively contribute to monitoring the surrounding environment, managing production, and disease monitoring. This article presented the findings of an experiment that used fully polarimetric PALSAR-2 data for estimating rubber tree circumference. In the evaluation of rubber stands, tree girth is commonly used as an indicator for estimating site productivity during a planting rotation. The information can then be employed in tree biomass estimation at the end of the rotation.

This research suggests that conventional data analysis methods such as MARS did not sufficiently manage complex problems with conventional backscatter coefficients data. With this type of dataset, the best models were those derived from SVM and RF models, yielding $R^2$ of between 0.77 and 0.82, with RMSE ranging around 9–10 cm. Options to use polarimetric decomposition indicate that a model-based approach outperformed Cloude–Pottier features. Nonetheless, model-based decomposition features behaved variably, with the RMSEs ranging between 8 and 18 cm. A greater likelihood of obtaining a suitable outcome was achieved by using Singh’s theorem, while Freeman–Durden and Yamaguchi decomposition theorems provided mean RMSEs above 10 cm. Combining datasets was somewhat beneficial, although it comes with computational complexity. While some combinations produced RMSE values far larger than those from model-based decomposition (even backscatter coefficients), good accuracy was observed when combined datasets were analyzed using extreme gradient boosting. This leads to the selection of two potential models, the first derived from Singh features coupled with guided regularized RF ($R^2 = 0.91$; RMSE = 8.31 cm), and the second from combined features processed with extreme gradient boosting ($R^2 = 0.92$; RMSE = 8.58 cm). Neither model yielded a negative estimation when applied to the full image scene.

When computing time is considered, the Singh model is preferable due to the limited number of variables it uses. By implementing zonal statistics over an estimated girth map, one could conclude that young rubber stands in the test site were generally in a normal condition. Some blocks in the northern part of the region, however, remained below the normal growth stage. In contrast, stands labeled as an excellent growth stage were situated
in the southern parts of the region. The reason for this is yet to be determined and therefore serves as the basis for future investigations.

**Author Contributions:** Conceptualization, B.H.T.; methodology, B.H.T.; software, B.H.T. and D.R.P.; validation, B.H.T. and D.R.P.; formal analysis, B.H.T. and D.R.P.; investigation, B.H.T. and D.R.P.; resources, B.H.T., D.R.P., A.L.G., and D.J.P.; data curation, B.H.T.; writing—original draft preparation, B.H.T.; writing—review and editing, B.H.T., D.R.P., A.L.G., and D.J.P.; visualization, B.H.T.; supervision, A.L.G. and D.J.P.; project administration, B.H.T.; funding acquisition, B.H.T. and D.R.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** Parts of this research were supported by Australia Awards, UIPA Scholarship, and JAXA.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Restrictions apply to the availability of these data. Synthetic Aperture Radar data were obtained from Japan Aerospace Exploration Agency (JAXA) through RA6-3004 research project and are commercially available from PASCO Corporation (https://alos-pasco.com/en/alos-2/spec/, accessed on 16 March 2022).

**Acknowledgments:** The authors would like to thank the reviewers and editors for their comments.

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

**References**

15. Dube, T.; Mutanga, O. Quantifying the variability and allocation patterns of aboveground carbon stocks across plantation forest types, structural attributes and age in sub-tropical coastal region of KwaZulu Natal, South Africa using remote sensing. *Appl. Geogr.* 2015, 64, 55–65. [CrossRef]


49. Feng, Z.; Huang, G.; Chi, D. Classification of the complex agricultural planting structure with a semi-supervised extreme learning machine framework. Remote Sens. 2020, 12, 3708. [CrossRef]


59. Mureriwa, N.; Adam, E.; Sahu, A.; Tesfamichael, S. Examining the spectral separability of prosopis glandulosa from co-existent species using field spectral measurement and guided regularized random forest. Remote Sens. 2016, 8, 144. [CrossRef]


64. Adab, H.; Morbidelli, R.; Saltalippi, C.; Moradian, M.; Ghalhari, G.A.F. Machine learning to estimate surface soil moisture from remote sensing data. Water 2020, 12, 3223. [CrossRef]


68. Adam, E.; Deng, H.; Odindi, J.; Abdel-Rahman, E.M.; Mutanga, O. Detecting the early stage of phaeosphaeria leaf spot infestations in maize crop using in situ hyperspectral data and guided regularized random forest algorithm. J. Spectrosc. 2017, 2017, 6961387. [CrossRef]
