OPEN ACCESS

forests

Article

Large Area Mapping of Boreal Growing Stock Volume on an Annual and Multi-Temporal Level Using PALSAR L-Band Backscatter Mosaics

Sebastian Wilhelm [1,]*, **Christian Hüttich** [2], **Mikhail Korets** [3] **and Christiane Schmullius** [2]

[1] Earth Observation Services (EOS) Jena GmbH, Jena 07743, Germany

[2] Department of Earth Observation, Friedrich-Schiller-University, Jena 07743, Germany;
E-Mails: christian.hüttich@uni-jena.de (C.H.); c.schmullius@uni-jena.de (C.S.)

[3] V.N. Sukachev Institute of Forest, Siberian Branch of the Russian Academy of Sciences,
Krasnoyarsk 660036, Russia; E-Mail: mik@ksc.krasn.ru

* Author to whom correspondence should be addressed; E-Mail: wilhelm@eos-jena.com;
Tel.: +49-3641-9-48979; Fax: +49-3641-9-48882.

Abstract: The forests of the Russian Taiga can be described as an enormous biomass and carbon reservoir. Therefore, they are of utmost importance for the global carbon cycle. Large-area forest inventories in these mostly remote regions are associated with logistical problems and high financial efforts. Remotely-sensed data from satellite platforms may have the capability to provide such huge amounts of information. This study presents an application-oriented approach to derive aboveground growing stock volume (GSV) maps using the annual large-area L-band backscatter mosaics provided by the Japan Aerospace Exploration Agency (JAXA). Furthermore, a multi-temporal map has been created to improve GSV estimation accuracy. Based on information from Russian forest inventory data, the maps were generated using the machine learning algorithm, RandomForest. The results showed the high potential of this method for an operational, large-scale and high-resolution biomass estimation over boreal forests. An RMSE from 55.2 to 63.3 m³/ha could be obtained for the annual maps. Using the multi-temporal approach, the error could be slightly reduced to 54.4 m³/ha.

1. Introduction

Estimating large-scale forest biomass is a crucial issue for understanding global carbon cycling, as well as for monitoring global and regional changes in vegetation due to climate change effects or changes in land use and, therefore, also for local forest inventories [1]. The Russian land surface with its large forested areas, as well as peat and wetlands contains an enormous biomass reservoir. As 49% of Russia is covered with forest, its overall growing stock volume (GSV) plays an important role for Russia's carbon balance and, thus, for the global carbon cycle [2].

Due to the large-scale dimensions of the Russian Taiga, it is still very expensive in terms of costs and time to establish an area-covering monitoring system for the Siberian forest. Satellite imagery, especially radar data, can assist in achieving this task. These data provide a frequent observation method for monitoring GSV decrease caused by logging, clear cutting or forest fires and also for detecting forest regrowth from afforestation or forest succession processes, e.g., after fires or clear cuts [3,4].

In past and recent research, synthetic aperture radar (SAR) has been proven to be capable of estimating GSV over boreal and temperate forests [5–8]. Past investigations showed that the wavelength of L-band backscatter data (approximate 20 cm) is highly suitable for estimating GSV and changes in GSV over boreal forests. The attenuation of the forest canopy in L-band is less than for the shorter C- and X-bands. Thus, saturation takes place at a higher biomass level [9,10]. Though the retrieval of GSV from a single C-band measurement is generally poor, a multi-temporal combination of several GSV estimates can lead to adequate results [11].

Basically, the backscatter signal of short wavelengths, like X- and C-band, is mainly determined by scattering processes in the upper layers of the trees, such as small branches and upper leafs. For longer wavelengths, like L- and P-band, the scattering processes are also affected by the major components of trees, such as trunks and larger branches. P-band with its wavelength of approximately 100 cm shows the best capabilities for measuring GSV [12]. Since space-borne P-band data are not available at present or in the near future, L-band data are of high importance in the estimation of forest GSV, as is documented in several publications. Using JERS-1 L-band backscatter, Santoro *et al.* [13] obtained a relative RMSE of 25% when mapping a boreal forest test site in Sweden. An investigation of sub-tropical forest vegetation in southern China using polarimetric ALOS PALSAR data led to an RMSE of 28.58 t/ha with $r^2 = 0.9$ [14]. Based on the significant correlation between biomass and the L-band backscatter signal, several regression models have been published. Using empirical or semi-empirical linear regression models, correlation coefficients from 0.66 to 0.78 between measured and estimated GSV could be obtained [15–17].

As can be seen in several publications, there is a significant correlation between the aboveground biomass of woody covered areas and the backscatter signal of L-band SAR data. In general, the backscatter signal gains when aboveground biomass increases. This relationship may therefore be suitable for estimation and classification of GSV over forested areas [9,13,18,19].

This paper gives a combination of a methodological approach and a possible future application method for deriving GSV over large areas from operational annual backscatter mosaic products as provided by the Japan Aerospace Exploration Agency (JAXA). The aim of this investigation was to analyze how capable these mosaics are for large-scale, high-resolution mapping of boreal forests.

2. Data

2.1. SAR Data

The Japan Aerospace Exploration Agency (JAXA) launched the Advanced Land Observing Satellite (ALOS) on the 24 January 2006. It was placed in a polar and sun synchronous orbit at an altitude of approximately 700 km with a recurrence cycle of 46 days. Besides the Panchromatic Remote-sensing Instrument for Stereo Mapping (PRISM) and the Advanced Visible and Near-Infrared Radiometer Type 2 (AVNIR-2), it carried the Phased Array L-band Synthetic Aperture Radar (PALSAR), which recorded the imagery used in this study. Until its demise in April 2011, PALSAR operated in different modes: Fine Beam Single (FBS), Fine Beam Dual (FBD), Polarimetric (PLR), and ScanSAR. This study was carried out by using mosaics recorded in FBD mode, HH and HV polarization, where a single swath has a width of 70 km [20].

JAXA provides annual mosaics of these L-band SAR backscatter data delivered by the PALSAR sensor on ALOS at 25-m spatial resolution. These mosaics were provided through JAXA's Kyoto and Carbon Science initiative [21]. The mosaic algorithm is described in detail by Shimada and Ohtaki [22]. It includes SAR long strip processing, ortho-rectification, backscatter correction for slope effects and neighboring strip suppression. Since space-borne imagery has the benefit of observing the same location with the same spatial resolution in a repetitive way, these mosaics are highly suitable for monitoring the annual changes of the Earth's surface on a regional or even on a global scale. An advantage of these operational backscatter products is that they need no further processing. This imagery can be used directly for any further processing or mapping without having expert knowledge in synthetic aperture radar (SAR) raw data processing and geocoding. To facilitate layer stack applications, the images come also co-registered to each other.

For this study, four dual-pol scenes (HH, HV) have been used for four years from 2007 to 2010. The study area is located in the Russian Taiga in central Siberia. Vegetation in that region is dominated by coniferous forest. Radar imagery covers an area of about 569,000 km^2 that reaches from 92° E to 105° E and from 53° N to 60° N, approximately 870 km in longitude and 780 km in latitude, between Krasnoyarsk and Irkutsk (Figure 1).

2.2. Forest Inventory Data

The data used as training data were available as stand-wise polygon shapefiles from the forest inventory (FI), each containing several forest parameters, such as tree-height, stem volume, relative stocking and tree-species combination. A stand is the major FI unit in the boreal forests in Russia and is delineated by the homogeneity of forested areas. In some cases, a stand's boundaries can be set by aggregating several smaller areas with different forest properties, particularly tree species and stem volume. As the forest stand boundaries in Russia are set in a subjective way by human interpretation, often based on aerial photos, the delineations of the stands themselves can be seen as possible sources of error. Legally, the error in stem volume accuracy of Russian forest inventory should be between 12% and 20%. A detailed description of the Russian boreal FI is given in Santoro *et al.* [7]. The geo-location of the five FI sites used for training can be seen in Figure 1. Each of them consists of 814 to 2046 forest stands. The area covered by a single stand reaches from 1.85 up to 485.15 ha.

Stem volume is the major parameter for this investigation. It is recorded in steps of 5 up to 50 m³/ha and in steps of 10 for greater values, and it mainly reaches from 0 to 470 m³/ha with 520 m³/ha as an absolute peak value with an average 165 m³/ha among all stands. Besides less frequent aspen, larch and spruce, the dominant tree species in the test area are birch, pine and fir, as well as Siberian pine (or *Pinus sibirica*). Data used for validation cover different areas than the training data (Figure 1; see Section 4.3 for details).

Figure 1. Location of the training and validation sites within the area covered by the mosaics. Upper left corner: Study area covered by the mosaics in a global context.
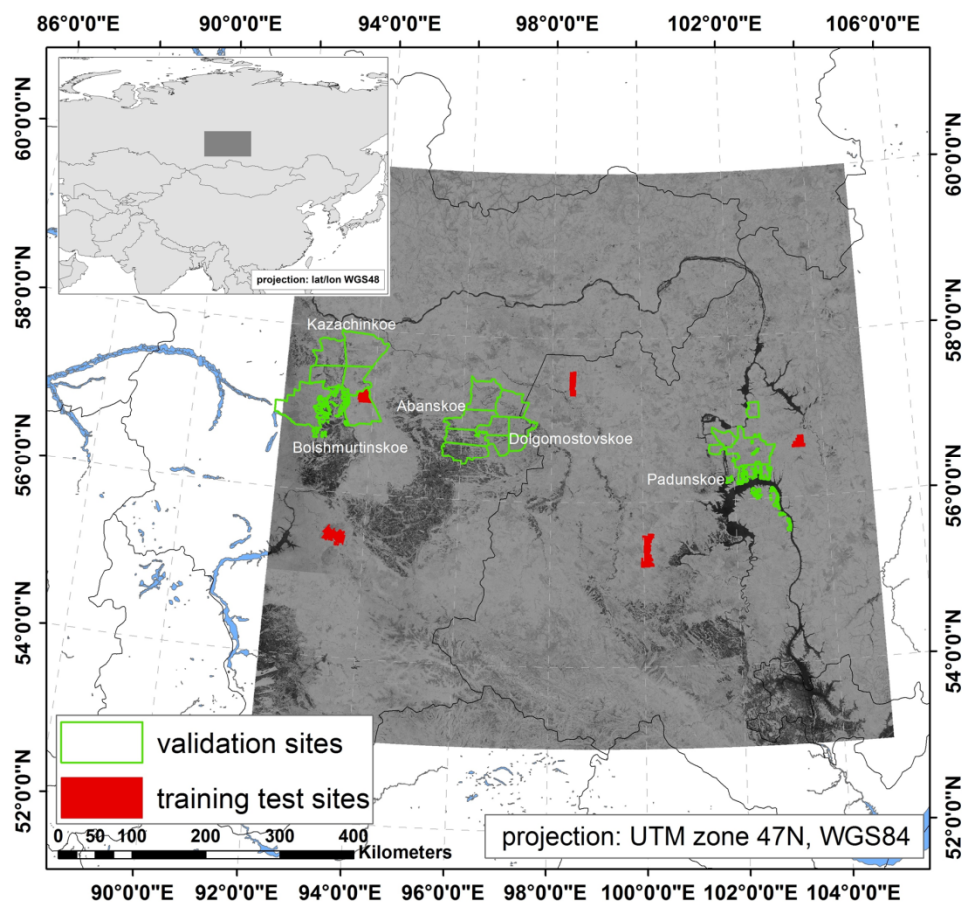
3. Methods

3.1. Random Forest Classifier

Within this study, we used the RandomForest (RF) classifier, which is a machine learning algorithm. It was originally implemented for classifying large statistical datasets. This property enables the RF classifier to deal with satellite data covering large areas [23]. Besides the common way of explaining a given variable as a function of one or more predictors (e.g., regression models), several other algorithms have been developed during recent decades. These machine learning methods have a large application field from the prediction to classification of large arrays of data [24–26]. For this investigation, GSV maps were derived by using a supervised random forest regression approach. RandomForest is a non-parametric machine learning algorithm. It uses random sampling and attribute selection to derive

the classification or, as in this study, regression of diverse multidimensional datasets. In general, RandomForest works by using an ensemble of many decision trees. This multiple tree-based regressions vote for the most likely class based on equally-weighted majority voting. Therefore, one third of the training dataset is excluded randomly for training each individual tree. This selection is the so-called out-of-bag (OOB) bootstrap sample, which is randomly permuted among the input features for each *tree*. Based on the other 2/3 of the input training data, the trees are grown, each to its maximum depth, by using the impurity Gini index [27], which is the result of the antagonism of the randomly permuted samples and features. These non-parametric regression tree methods have proven their capabilities in a wide range of ecological modeling studies [28–31], as well as in land cover mapping projects using optical remote sensing data [32,33]. In the case of SAR, some GSV estimations and height modeling has been performed [26,34,35].

RandomForest is available as an open source package for the statistic programming software, R [36], which has been used for processing the data for this investigation.

3.2. GSV Modeling

Reference polygons were reduced in size by applying a 50-m buffer on each polygon. This is done to avoid mixed pixel information at clear cut edges or natural forest boundaries. The buffering also helps minimize possible location mismatches between the FI and the SAR data. Furthermore, buffering can lead to small fragments of polygons. Depending on noise or topographic effects, respectively, it may be possible that very small areas in the training data are not representative according to the relationship between stem volume and backscatter. Therefore, polygons that cover an area less than 10 pixels (approximately 6000 square meters) after buffering were excluded from the training data set. The five forest inventory shapefiles (Figure 1) have been combined to one single training shapefile containing all polygons (5021 in total). In order to reduce the influences of outdated FI or errors in SAR data (e.g., geometric and/or radiometric calibration errors or soil moisture and other weather effects), a threshold of two standard deviations was applied on the training data. Models have been generated on individual annual mosaics from 2007 to 2010 and on a four-year multi-temporal basis for all datasets available within this period. The number of trees was set to 500 for each single model run, which is usually the default value. The information for total stem volume on the forest stand level noted in the forest inventory polygons has been used for training the models. The training polygons were rasterized (transformed into pixels) based on the SAR data containing the information of stem volume in m^3/ha. All valid pixels with stem volume information within the rasterized training data were extracted. Due to the enormous amount of valid pixels, a random sample of 5% of these pixels has been taken for training the models, which is still a large number of 56,174 samples. A higher number of samples would have led to a tremendous processing time or caused allocation errors while generating the models. The models were derived by using a raster stack of a dual-pol imagery for each year and a raster stack of all 8 images (4 HV, 4 HH) for the multi-temporal approach as predictors. Applying the models on each corresponding raster stack, GSV maps were generated. As the prediction of the regression tree model on the SAR data led to float values, the output maps have been rounded to integer values for reasons of better operability.

4. Results

4.1. Random Forest Performance

Applying an RF regression and proving its performance for dealing with the high resolution and large area covering SAR data was a crucial issue during this investigation. Though the processing in R seems to be quite slow (especially when predicting the models on the SAR mosaics), RF in combination with the raster R-package gives a labor-saving approach for estimating GSV from large imagery. Note, that one single SAR mosaic can be almost 4 GB in size. As regression trees are grown automatically based on the features of the input data, there is no need to set training areas, and it is also not necessary to figure out the parameters of regression functions manually. This makes RF flexible and time-saving when dealing with large datasets. On the other hand, RF works as a black box classifier. Thus, it is difficult to retrace the origin of the decision tree structures, as well as to verify and adjust the random samples taken by RF by hand. Possible errors in the SAR data or the training data are also taken into account. Nevertheless, a large number of trees grown should reduce the overall error while the model is trained. As shown in Figure 2, a number of 200 to 300 trees seem to be adequate to reduce the error to its minimum. A larger number of trees would be more time consuming without a significant further reduction of the mean square error (MSE). While the regression trees were grown, the multi-temporal model led to the highest percent value of variance explained (% of variance explained, Table 1), which is defined as $1 - (MSE/variance(FI\ data))$.

Figure 2. Relationship between the number of trees grown and the reduction of the MSE during model generation in the case of the multi-temporal RF model.
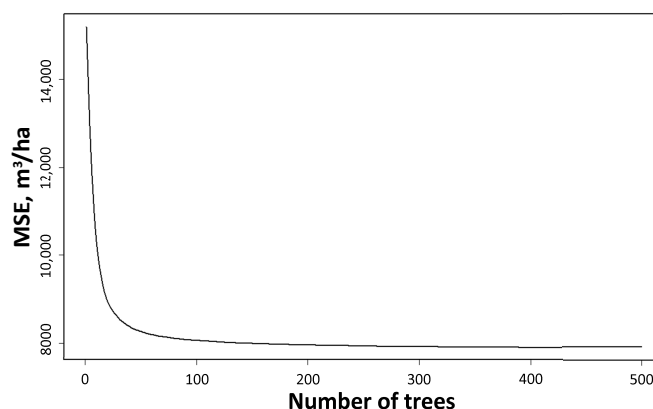

Table 1. Variance explained by the model.

Year	2007	2008	2009	2010	Multi-Temporal
Variance explained	29.83%	32.33%	29.92%	30.14%	46.63%

In addition to the main issue of this investigation, RF models have been generated using smaller subsets of the SAR imagery. For that purpose, the SAR data sets have been clipped to the smallest three of the five training datasets (Figure 1), and models have been generated for a different number of layers of these SAR data subsets. This was done to verify how RF performs under the condition of a varying depth of SAR imagery. One dataset consists of a dual-pol SAR scene. It was found that an

increasing number of datasets also increases the percent variance explained (% of variance explained). However, the increase of the variance explained is dependent on the number of datasets used, which seems to saturate when a larger number of SAR imagery is processed. Using two instead of one SAR datasets, the variance explained is increased between 9.29% and 13.13%, whereas an addition of one more datasets led to a further increase of approximately 3.3%, and using a fourth dataset increased the model accuracies further by only 1.92% to 2.73% (Figure 3).

Figure 3. The percentage of variance explained for three smaller subsets of SAR data under the consideration of an increasing number of datasets. Each dataset consisted of one dual-pol scene. Each color represents one test site.
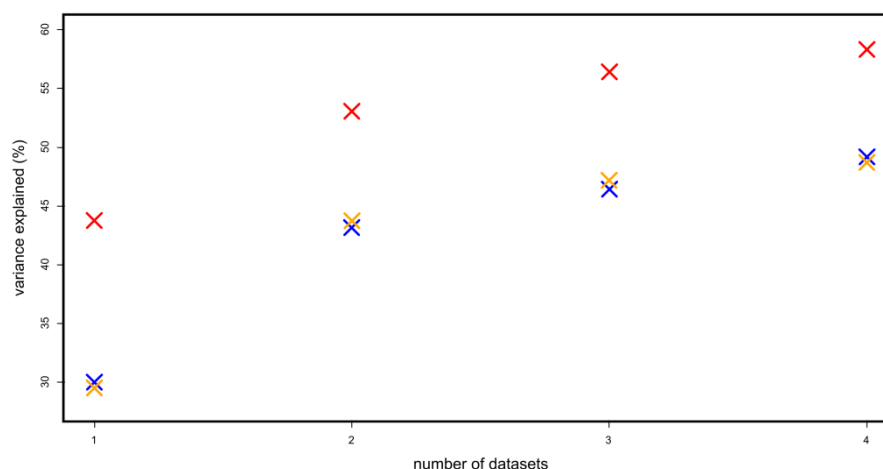
4.2. Mapping Results

Based on operational large-scale annual PALSAR mosaics, four annual and one multi-temporal GSV maps were derived with a spatial resolution of 25 m. Each of them contains RF regression results of GSV expressed in m^3/ha. Figure 4 gives an overview and a zoomed-in detailed view of the multi-temporal map. As can be seen in detail, forest structures are well identifiable, e.g., dense forest, sparse forest, clear-cuts or non-vegetated river beds and creeks. The range of the derived GSV for each map is shown in Table 2. It should be noted that these maximum values are absolute peaks within the generated maps. Figure 5 shows a comparison between the 2007 GSV map and the multi-temporal map. It can be seen that the multi-temporal map shows a less noisy structure than the one-year map (derived by only one dual-pol scene). Furthermore, the distinction of dense forest, sparse forest, non-forested areas and clear cuts is clearer in the multi-temporal mapping result. This is caused by the fact that the noise effects are reduced when applying RF on a higher amount of variables; in this case, multi-temporal recorded pixels.

Examining the results by a simple visual interpretation, the annual maps of 2007, 2008 and 2009 showed no significant difference in detail. In comparison to the other three annual maps, the mapping result of the annual 2010 dataset showed an obviously strongly underestimated GSV in the southwestern area of the mosaic. This effect may be caused by radiometric calibration errors when the mosaics were generated (Figure 6, see the Discussion for details).

Figure 4. Growing stock volume (GSV) mapping result of the multi-temporal approach highlighting a subset of a heterogeneous forested area, including dense forest, sparse forest and clear cuts (clearly identifiable by their unnatural geometric shape).
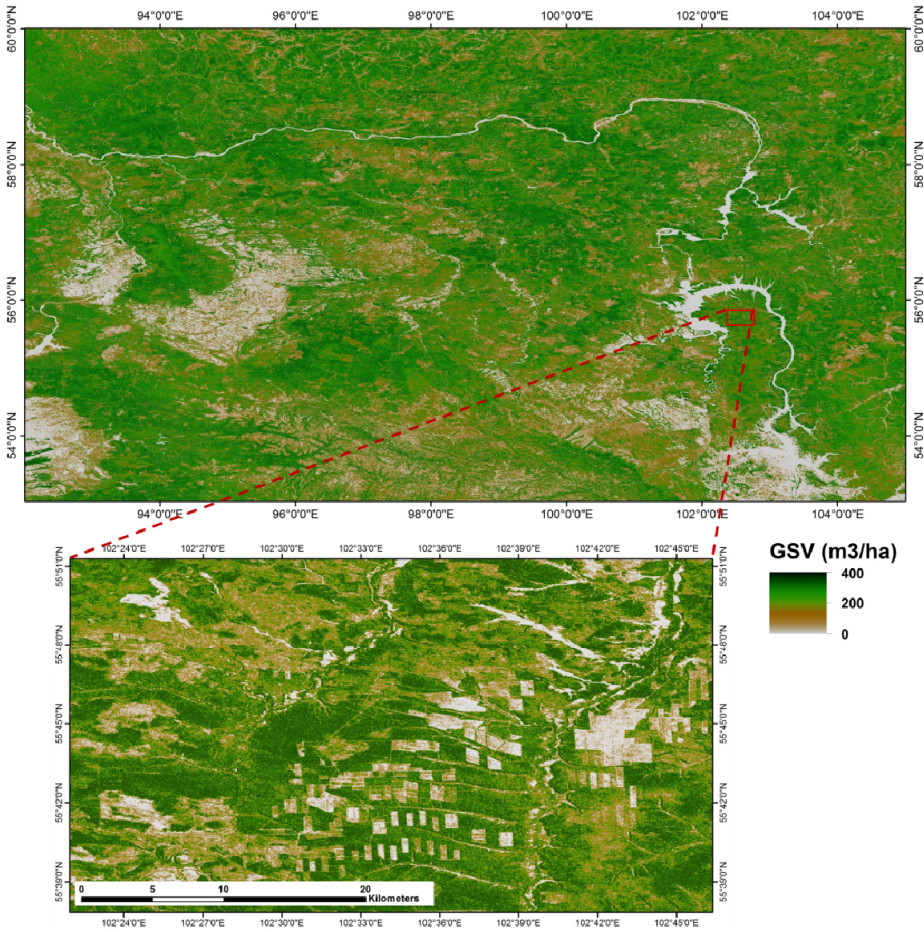

Table 2. Range of the derived GSV.

Year	2007	2008	2009	2010	Multi-Temporal
Range	0–412	0–401	0–408	0–421	0–409

Figure 5. Detailed comparison of the mapping results for the 2007 map (**a**) and the multi-temporal result (**b**).
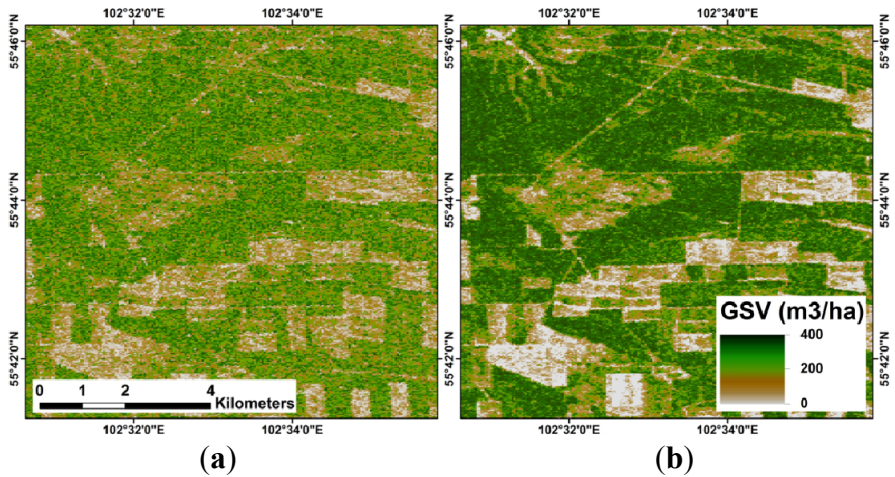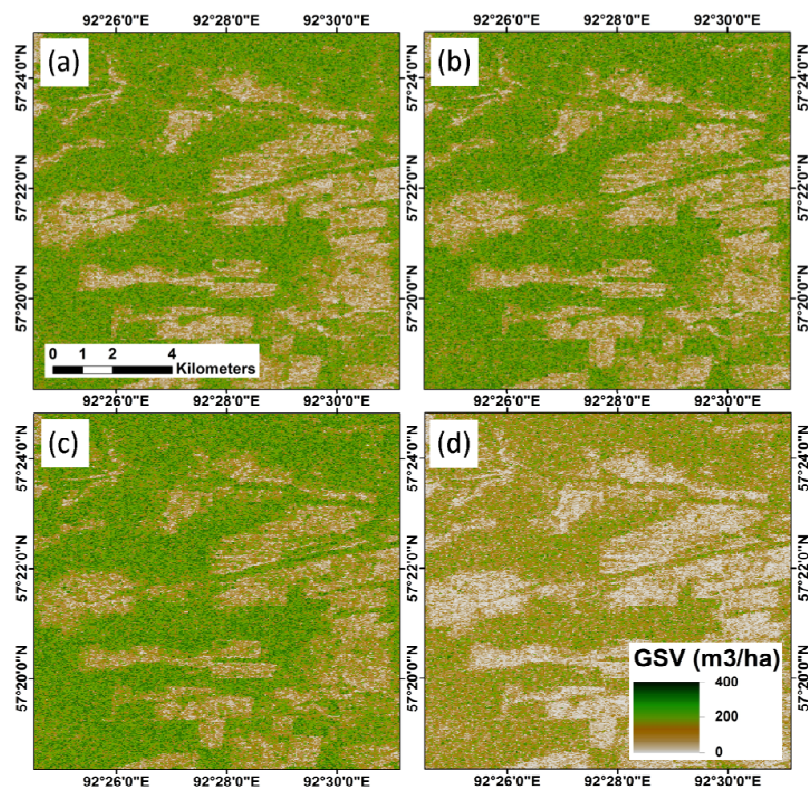

(**a**)   (**b**)

Figure 6. Detailed examples of the annual maps (southwestern area). (**a**) 2007; (**b**) 2008; (**c**) 2009; (**d**) 2010.

Using the variable importance of RF, HV polarization could be identified as more important for the classification results than HH for all cases. Variable importance is a concept of RF, which estimates the importance of a variable (in this case, either HV or HH backscatter values). In general, this is done by permuting the out-of-bag data for that variable while the others are left unchanged. RF then calculates how much the error increases (see Section 3.1). The fact that HV usually shows a stronger correlation to forest biomass than HH has been proven in several former investigations, e.g., [37,38].

4.3. Validation

The generated GSV maps have been validated by the Sukachev Institute of Forest, Siberian Branch of the Russian Academy Of Sciences (SB RAS) located in Krasnoyarsk, Russia, Siberia. SAR-derived GSV estimates were compared to elementary forest inventory polygons (FIP), which contain information about land cover type, stand species composition, density, age, height, tree diameter and GSV per species in m^3/ha. The total GSV for all species in the FIP has been used for validating the SAR-based GSV. Statistical parameters (mean and standard deviation) were calculated for each FIP based on the SAR-GSV pixel values after the polygons have been rasterized. The comparison statistics were evaluated for four annual maps and the multi-temporal GSV map using per species and total averaging. The local validation sites were grouped into three areas (Figure 2, Table 3).

Table 3. Validation test sites.

Site number	Site Name (FMA Name)	Area, ha	FIP Number	Region
1	Kazachinskoe and Bolshemurtinskoe	903,573	47,962	Krasnoyarsk kray
2	Abanskoe and Dolgomostovskoe	698,479	41,241	Krasnoyarsk kray
3	Padunskoe	366,696	22,318	Irkutsk oblast
	Total for all test sites	1,968,748	111,521	

The overall statistics of the validation are gathered in Table 4. While the results of the first three maps are consistent in comparison to each other, the map of 2010 shows weaker results. In comparison to the other maps, the 2010 map contains a stronger underestimation of the GSV. The greatest underestimated GSV (Emin) of this map dropped to -285.8 m³/ha, which is 20.9 less than in the worst case of the other images. Furthermore, the mean difference between FIP and SAR pixel values (ME) shows the weakest result in the 2010 map (-14.6 m³/ha). The accuracy of the generated GSV maps was measured by the root mean square error (RMSE). In the case of the maps for the years 2007, 2008 and 2009, an RMSE between 55.2 and 57.9 m³/ha could be obtained, while the result of 2010 showed a higher error of 63.3 m³/ha. The multi-temporal approach, where we used the datasets of all four years, led to a lower RMSE of 54.4 m³/ha.

Table 4. Validation results, overall statistics of SAR and forest inventory (FI)-based GSV comparison (m³/ha).

Label	Characteristics	2007	2008	2009	2010	Multi-Temporal
Emin	Δ GSV$_{min}$	-259.1	-249.6	-264.9	-285.8	-247.9
Emax	Δ GSV$_{max}$	202.5	216.5	217.6	208	221.2
ME	Mean Δ GSV (SAR-FI)	-1.3	1.4	6.3	-14.6	3.7
SD	Δ GSV SD	55.3	55.2	57.6	61.6	54.3
RMSE	Root Mean Square Error	55.3	55.2	57.9	63.3	54.4

Figure 7 shows a comparison of the spatial distribution of the GSV differences between SAR-derived GSV and FI data for the annual mapping results of one validation test site. Though the first three images differ slightly from each other, they basically show a certain similarity with regions of GSV overestimation in the northeastern part of Kazachinskoe and in some areas of Bolshemurtinskoe, which are strongly dominated by fir and birch. In dense aspen- and birch-dominated forest stands, the GSV is often underestimated, as can be seen in the eastern part of Bolshemurtinskoe. Nevertheless, most areas show smaller GSV differences between -40 and $+40$ m³/ha, displayed in light-blue, light-green and yellow colors. For the 2010 mapping result, an area in the western part of Bolshemurtinskoe shows a strong GSV underestimation, which leads to an increase of the overall RMSE, as mentioned above. This effect is caused by radiometric calibration errors in this part of the backscatter mosaic, leading to lower dB values in this region (see the Discussion for details).

The species-dependent statistics show that areas where Siberian pine is the dominant species have the weakest accuracy, with an RMSE of 74.4 m³/ha. In these regions, the GSV is often also strongly underestimated. The best result could be obtained over larch-dominated areas showing an RMSE of 45.3 m³/ha (Table 5).

Figure 7. GSV difference between SAR-derived and FI GSV for the validation sites, Kazachinskoe and Bolshemurtinskoe. (**a**) 2007; (**b**) 2008; (**c**) 2009; (**d**) 2010.

(**a**) 2007
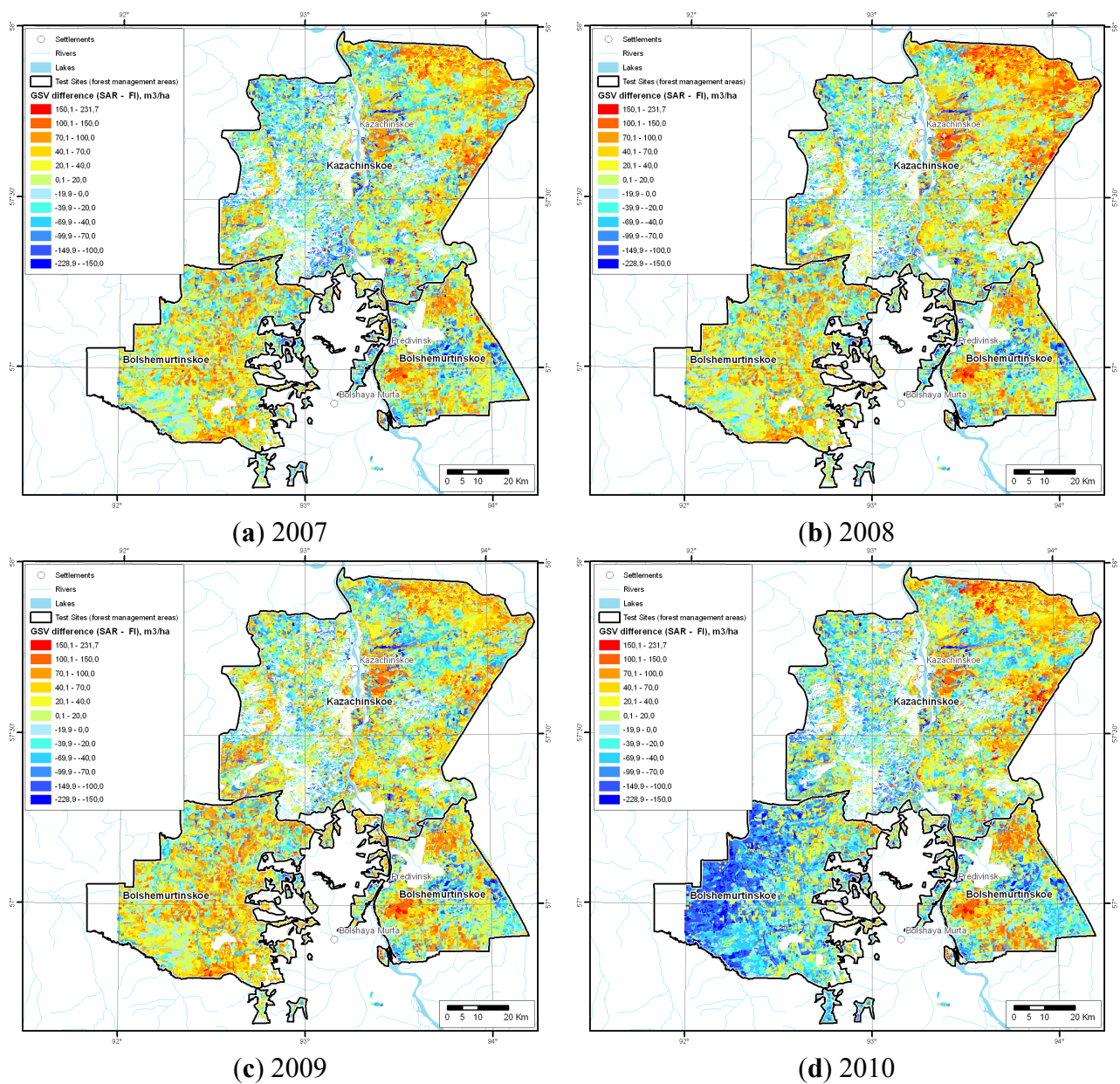
(**b**) 2008

(**c**) 2009

(**d**) 2010

Table 5. Validation results, species-dependent statistics of the SAR and FI-based GSV comparison (m^3/ha) for the multi-temporal approach.

Dominant tree species	Emin	Emax	ME	SD	RMSE
Aspen	−184.3	210.7	−5.9	56.3	56.6
Birch	−170.0	186.7	21.1	46.5	51.1
Fir	−270.0	221.2	29.2	72.0	72.1
Larch	−171.1	131.2	−3.4	45.2	45.3
Pine	−247.9	193.6	−16.9	58.0	60.4
Siberian Pine	−310.0	209.0	−39.1	63.3	74.4
Spruce	−260.0	219.5	9.9	51.4	52.4

5. Discussion

Considering the fact that this method represents a comparatively user-friendly and time-saving black box classifier for estimating GSV, the results of this study are basically of good quality when compared to other investigation results [13,39–41]. These studies used high-resolution L-band backscatter for GSV estimation over boreal and temperate forests, as well. They led to similar or even worse results, showing root mean square errors in GSV estimation between 36 and 102 m³/ha.

When GSV is estimated by using radar backscatter data, one main issue and well known problem is that saturation takes place at a certain point when the GSV increases. For L-band data over boreal and temperate forests, saturation levels of approximately 140 to 300 m³/ha can be expected [12,16,40]. Hence, the deviation between estimated and validation GSV becomes larger over areas covered by dense forests. This fact leads to an underestimation of the GSV over such regions and, thus, influences the overall error. As can be seen in Figure 7, there are areas in the eastern part of Bolshemurtinskoe that show a strongly underestimated GSV. Here, the ground truth GSV reaches from 300 up to at least 400 m³/ha, which leads to the abovementioned underestimation caused by signal saturation. Carreiras *et al.* [26] showed that in regions with a generally lower biomass, higher accuracies could be obtained using L-band backscatter, as saturation did not take place during their investigation in African savannas.

The quality and accuracy of the mapping results may also be affected by different moisture conditions at different image acquisition times. As the mosaics are created by individual images or strips, weather effects can lead to different moisture conditions within one mosaic in both, vegetation water content and soil moisture. This is likely to have an influence on the backscatter behavior of the surface [19,42,43]. Seasonal dynamics according to frozen/unfrozen conditions could also affect the characteristics of SAR data [44]. However, the mosaics consist of data recorded in summer from May to September. An influence of these dynamics is therefore not impossible, but unlikely. In the first instance, different weather and moisture conditions may be responsible for the weak mapping results in the western part of Bolshemurtinskoe based on the 2010 dataset. Radiometric calibration errors in this region could lead to lower dB values in comparison to the remaining parts of these mosaics. This can cause the strong GSV underestimation determined in that area. The 2010 data was also included in the multi-temporal approach, which showed slightly better results than any other annual GSV map. However, depending on the weak quality of some regions in the 2010 mosaic, it has to be expected that the 2010 dataset worsens the validation result of the multi-temporal map. However, excluding the 2010 dataset from the creation of a multi-temporal RF model, no significant improvement could be obtained.

Besides seasonal dynamics and moisture conditions, other factors can affect the SAR-based GSV estimation. For small GSV values, an uneven forest floor would increase the backscatter signal and lead to an overestimated GSV. Further influences can be the forest structure, the vegetation type or other biophysical parameters [13,45]. Considering tree species during this investigation, Siberian pine, pine and fir showed higher RMSE than other tree types, which reduces the overall accuracy.

6. Conclusions and Outlook

For this study, large-scale PALSAR backscatter mosaics with a spatial resolution of 25 m have been examined for their capabilities of mapping GSV over boreal forest in Siberia. This was done for

getting information about the potential of these annual mosaics to derive GSV operationally for forest and biomass monitoring. Another goal of this investigation was to increase the accuracy of the GSV estimation by using a multi-temporal approach. Based on FI data and using the backscatter mosaics as predictors, we used a machine learning algorithm, which is implemented in the R package RF, to predict the GSV for the whole area covered by the mosaics.

This study can be seen as an approach to derive GSV estimations over boreal forests in an operational way for large areas focusing on regional and (for possible future investigations) on global scale biomass estimates and biomass monitoring systems, since the mosaics are available annually. For processing the data, the RF package in combination with the Raster package for the statistical programming language, R, has been used. Though the raster processing in R, in comparison to commercial products, seems to be quite slow, the classification/regression results show good quality in comparison to other studies, as mentioned above. Furthermore, the performance of the open source software, R, was more than sufficient for the purposes of this study, especially considering that R comes free of charge. Applying the automated classification method provided by RF on large raster datasets, the results are consistent for each model run and, thus, comparable to each other, which is a fundamental requirement for monitoring changes in vegetation and land cover during possible future investigations (e.g., fires, clear cuts, regrowth, forest successions).

In addition, RF can easily be used for deriving multi-temporal GSV estimations, which probably show an increased accuracy.

The potential of large area-covering PALSAR dual-pol backscatter mosaics has also been verified by De Grandi *et al.* [46]. They presented a number of potential applications for a mosaic covering the whole African continent, highlighting the mapping of mangroves, plantations, secondary forest and the boundary between savannas and forest, as well as detecting changes in vegetation cover. Based on this potential of the mosaics, possible future investigations should consider further biophysical parameters besides GSV for mapping and monitoring boreal forest, e.g., forest structure, tree species, tree height.

It should be noted that a further possibility to improve forest-related work based on SAR data can be seen in interferometry, which refers to the measurements of the difference in phase observed when a target is measured at different times or from slightly different locations [47]. In that way, interferometric coherence can be derived, which is also sensitive to forest biomass and other biophysical parameters, as has been demonstrated in several publications [1,3,7,48,49].

A new generation of SAR satellite (ALOS-2) is expected to be launched in 2014 and will continue the L-band SAR observations of ALOS PALSAR. Data of ALOS-2 will come with a higher spatial resolution in single, dual and full polarimetry [50]. This will make them highly suitable for forest investigations and for continuing past work based on ALOS PALSAR data.

Acknowledgments

Space, Cross-cutting Activities, International Cooperation, Grant No. SPA.2010.3.2-01 EU-Russia Cooperation in Global Monitoring for Environment and Security (GMES).

Author Contributions

Sebastian Wilhelm was responsible for the research design and the editing coordination of the paper. Data preparation and analysis were partitioned as follows: forest inventory and validation (Mikhail Korets); ALOS-PALSAR GSV maps (Sebastian Wilhelm and Christian Hüttich). Christiane Schmullius provided significant input to the research design. All of the authors contributed in editing and reviewing the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Eriksson, L.E.B.; Santoro, M.; Wiesmann, A.; Schmullius, C.C. Multitemporal JERS repeat-pass coherence for growing-stock volume estimation of Siberian forest. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1561–1570.
2. Dolman, A.J.; Shvidenko, A.; Schepaschenko, D.; Ciais, P.; Tchebakova, N.; Chen, T.; van der Molen, M.K.; Marchesini, L.B.; Maximov, T.C.; Maksyutov, S.; *et al*. An estimate of the terrestrial carbon budget of Russia using inventory-based, eddy covariance and inversion methods. *Biogeosciences* **2012**, *9*, 5323–5340.
3. Thiel, C.J.; Thiel, C.; Schmullius, C.C. Operational Large-Area Forest Monitoring in Siberia Using ALOS PALSAR Summer Intensities and Winter Coherence. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3993–4000.
4. Hüttich, C.; Schmullius, C.C.; Thiel, C.J.; Pathe, C.; Bartalev, S.; Emelyanov, K.; Korets, M.; Shvidenko, A.; Schepaschenko, D. ZAPÁS: Assessment and monitoring of forest resources in the framework of the EU-Russia space dialogue. *Let's Embrace Space, Volume II—Space Research achievements under the 7th Framework Programme*; Directorate-General for Enterprise and Industry, European Commission: Luxembourg, 2012; pp. 164–171.
5. Beaudoin, A.; Toan, T.L.E.; Goze, S.; Nezry, E.; Lopes, A.; Mougin, E.; Hsu, C.C. Retrieval of forest biomass from SAR data. *Int. J. Remote Sens.* **1994**, *15*, 2777–2796.
6. Wang, Y.; Davis, F.W.; Melack, J.M.; Kasischke, E.S.; Christensen, N.L. The effects of changes in forest biomass on radar backscatter from tree canopies. *Int. J. Remote Sens.* **1995**, *16*, 503–513.
7. Santoro, M.; Shvidenko, A.; Mccallum, I.; Askne, J.; Schmullius, C. Properties of ERS-1/2 coherence in the Siberian boreal forest and implications for stem volume retrieval. *Remote Sens. Environ.* **2007**, *106*, 154–172.
8. Ni, W.; Sun, G.; Member, S.; Guo, Z.; Zhang, Z.; He, Y. Retrieval of forest biomass from ALOS PALSAR data using a lookup table method. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 875–886.

9. Pulliainen, J.T.; Kurvonen, L.; Hallikainen, M.T. Multitemporal behavior of L- and C-band SAR observations of boreal forests. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 927–937.

10. Rignot, E.; Way, J.; Williams, C.; Viereck, L. Radar estimates of aboveground biomass in boreal forests of interior Alaska. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*, 1117–1124.

11. Santoro, M.; Beer, C.; Cartus, O.; Schmullius, C.; Shvidenko, A.; McCallum, I.; Wegmüller, U.; Wiesmann, A. Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements. *Remote Sens. Environ.* **2011**, *115*, 490–507.

12. Dobson, M.; Ulaby, F.T.; LeToan, T.; Beaudoin, A.; Kasischke, E.S.; Christensen, N. Dependence of radar backscatter on coniferous forest biomass. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 412–415.

13. 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.

14. Sarker, M.L.R.; Nichol, J.; Ahmad, B.; Busu, I.; Rahman, A.A. Potential of texture measurements of two-date dual polarization PALSAR data for the improvement of forest biomass estimation. *ISPRS J. Photogramm. Remote Sens.* **2012**, *69*, 146–166.

15. Harrell, P.A.; Bourgeau-Chavez, L.L.; Kasischke, E.S.; French, N.H.F.; Christensen, N.L. Sensitivity of ERS-1 and JERS-1 radar data to biomass and stand structure in Alaskan boreal forest. *Remote Sens. Environ.* **1995**, *54*, 247–260.

16. Fransson, J.E.S.; Israelsson, H. Estimation of stem volume in boreal forests using ERS-1 C- and JERS-1 L- band SAR data. *Int. J. Remote Sens.* **1990**, *20*, 123–137.

17. Rauste, Y. Multi-temporal JERS SAR data in boreal forest biomass mapping. *Remote Sens. Environ.* **2005**, *97*, 263–275.

18. Luckman, A.; Baker, J.; Honza, M.; Lucas, R. Tropical forest biomass density estimation using JERS-1 SAR: Seasonal variation, confidence limits, and application to image mosaics. *Remote Sens. Environ.* **1998**, *63*, 126–139.

19. Lucas, R.; Armston, J.; Fairfax, R.; Fensham, R.; Accad, A.; Carreiras, J.; Kelley, J.; Bunting, P.; Clewley, D.; Bray, S.; *et al.* An evaluation of the ALOS PALSAR L-band backscatter—Above ground biomass relationship Queensland, Australia: Impacts of surface moisture condition and vegetation structure. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 576–593.

20. Rosenqvist, A.; Shimada, M.; Ito, N.; Watanabe, M. ALOS PALSAR: A pathfinder mission for global-scale monitoring of the environment. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3307–3316.

21. Rosenqvist, Å.; Ogawa, T.; Shimada, M.; Igarashi, T. Initiating the ALOS Kyoto & carbon initiative. In Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium (IGARSS), Sydney, NSW, Austrilia, 9–13 July 2001; pp. 546–548.

22. Shimada, M.; Ohtaki, T. Generating large-scale high-quality SAR mosaic datasets: Application to PALSAR data for global monitoring. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 637–656.

23. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

24. Carreiras, J.M.B.; Pereira, J.M.C.; Shimabukuro, Y.E. Land-cover mapping in the Brazilian Amazon using SPOT-4 vegetation data and machine learning classification methods. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 897–910.

25. Moisen, G.G.; Freeman, E.A.; Blackard, J.A.; Frescino, T.S.; Zimmermann, N.E.; Edwards, T.C. Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecol. Model.* **2006**, *199*, 176–187.

26. Carreiras, J.M.B.; Melo, J.B.; Vasconcelos, M.J. Estimating the above-ground biomass in Miombo Savanna woodlands (Mozambique, East Africa) using L-band synthetic aperture radar data. *Remote Sens.* **2013**, *5*, 1524–1548.

27. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Wadsworth: Pacific Grove, CA, USA, 1984; p. 368.

28. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **2006**, *9*, 181–199.

29. Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792.

30. Hüttich, C.; Herold, M.; Strohbach, B.J.; Dech, S. Integrating in-situ, Landsat, and MODIS data for mapping in Southern African savannas: Experiences of LCCS-based land-cover mapping in the Kalahari in Namibia. *Environ. Monit. Assess.* **2011**, *176*, 531–547.

31. Carreiras, J.M.B.; Pereira, J.M.C.; Campagnolo, M.L.; Shimabukuro, Y.E. Assessing the extent of agriculture/pasture and secondary succession forest in the Brazilian Legal Amazon using SPOT VEGETATION data. *Remote Sens. Environ.* **2006**, *101*, 283–298.

32. Houghton, R.A.; Butman, D.; Bunn, A.G.; Krankina, O.N.; Schlesinger, P.; Stone, T.A. Mapping Russian forest biomass with data from satellites and forest inventories. *Environ. Res. Lett.* **2007**, *2*, doi:10.1088/1748-9326/2/4/045032.

33. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300.

34. Simard, M.; Pinto, N.; Fisher, J.B.; Baccini, A. Mapping forest canopy height globally with spaceborne lidar. *J. Geophys. Res.* **2011**, *116*, 1–12.

35. Cartus, O.; Kellndorfer, J.; Rombach, M.; Walker, W. mapping canopy height and growing stock volume using Airborne Lidar, ALOS PALSAR and Landsat ETM$^+$. *Remote Sens.* **2012**, *4*, 3320–3345.

36. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22,.

37. Rauste, Y.; Hame, T.; Pulliainen, J.; Heiska, K.; Hallikainen, M. Radar-based forest biomass estimation. *Int. J. Remote Sens.* **1994**, *15*, 2797–2808.

38. Le Toan, T.; Beaudoin, A.; Riom, J.; Guyon, D. Relating forest biomass to SAR data. *IEEE Trans. Geosci. Remote Sens.* **1994**, *30*, 403–411.

39. 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.

40. Cartus, O.; Santoro, M.; Kellndorfer, J. Mapping forest aboveground biomass in the Northeastern United States with ALOS PALSAR dual-polarization L-band. *Remote Sens. Environ.* **2012**, *124*, 466–478.

41. Peregon, A.; Yamagata, Y. The use of ALOS/PALSAR backscatter to estimate above-ground forest biomass: A case study in Western Siberia. *Remote Sens. Environ.* **2013**, *137*, 139–146.

42. Harrell, P.A.; Kasischke, E.S.; Bourgeau-chavez, L.L.; Haney, E.M.; Christensen, N.L, Jr. Evaluation of approaches to estimating aboveground biomass in southern pine forests using SIR-C data. *Remote Sens. Environ.* **1997**, *59*, 223–233.

43. Kasischke, E.S.; Tanase, M.A.; Bourgeau-Chavez, L.L.; Borr, M. Soil moisture limitations on monitoring boreal forest regrowth using spaceborne L-band SAR data. *Remote Sens.Environ.* **2011**, *115*, 227–232.

44. Santoro, M.; Askne, J.; Eriksson, L.; Schmullius, C.; Wiesmann, A. Seasonal dynamics and stem volume retrieval in boreal forests using JERS-1 backscatter. *Proc. SPIE* **2002**, *41*, 231–242,.

45. Suzuki, R.; Kim, Y.; Ishii, R. Sensitivity of the backscatter intensity of ALOS/PALSAR to the above-ground biomass and other biophysical parameters of boreal forest in Alaska. *Polar Sci.* **2013**, *7*, 100–112.

46. De Grandi, G.D.; Bouvet, A.; Lucas, R.M.; Shimada, M.; Monaco, S.; Rosenqvist, A. The K&C PALSAR mosaic of the African continent: Processing issues and first thematic results. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3593–3610.

47. Patenaude, G.; Milne, R.; Dawson, T.P. Synthesis of remote sensing approaches for forest carbon estimation: reporting to the Kyoto Protocol. *Environ. Sci. Policy* **2005**, *8*, 161–178.

48. Watanabe, M.; Shimada, M. Relation between coherence, forest biomass, and L-band σ^0. In Proceedings of the IEEE International Conference on International Geoscience and Remote Sensing Symposium (IGARSS), Denver, CO, USA, 31 July–4 August 2006; pp. 1651–1654.

49. Gaveau, D.L.A.; Balzter, H.; Plummer, S. Forest woody biomass classification with satellite-based radar coherence over 900 000 km^2 in Central Siberia. *Forest Ecol. Management*, **2003**, *174*, 65–75.

50. Okada, Y.; Hamasaki, T.; Tsuji, M.; Iwamato, M.; Hariu, K.; Kankaku, Y.; Suzuki, S.; Osawa, Y. Hardware performance of L-band SAR system onboard ALOS-2. In Proceedings of the 2001 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Vancouver, BC, Canada, 24–29 July 2011; pp. 894–897.