

Article

# Considerations towards a Novel Approach for Integrating Angle-Count Sampling Data in Remote Sensing Based Forest Inventories

Melanie Kirchhoefer \* , Johannes Schumacher , Petra Adler and Gerald Kändler

Department of Biometrics, Forest Research Institute of Baden-Württemberg; Wonnhaldestr. 4, Freiburg D-79100, Germany; johannes.schumacher@forst.bwl.de (J.S.); petra.adler@forst.bwl.de (P.A.); gerald.kaendler@forst.bwl.de (G.K.)

\* Correspondence: melanie.kirchhoefer@forst.bwl.de; Tel.: +49-761-4018-285

Academic Editors: Christian Ginzler, Lars T. Waser and Timothy A. Martin

Received: 31 March 2017; Accepted: 3 July 2017; Published: 5 July 2017

**Abstract:** Integration of remote sensing (RS) data in forest inventories for enhancing plot-based forest variable prediction is a widely researched topic. Geometric consistency between forest inventory plots and areas for extraction of RS-based predictive metrics is considered a crucial factor for accurate modelling of forest variables. Achieving geometric consistency is particularly difficult with regard to angle-count sampling (ACS) plots, which have neither distinct shape nor distinct extent. This initial study considers a new approach for integrating ACS and RS data, where the concept of ACS is transferred to RS-based metrics extraction. By using the relationship between tree height and diameter at breast height (DBH), pixels of a RS-based canopy height model are extracted if their value suggests a DBH that would lead to inclusion in an angle-count sample at the given distance to the plot centre. Different variations of this approach are tested by modelling timber volume in national forest inventory plots in Germany. The results are compared to those achieved using fixed-radius plots. A root mean square error of approximately 42% is achieved by both the new and fixed-radius approaches. Therefore, the new approach is not yet considered sufficient for overcoming all difficulties concerning the integration of ACS plot and RS data. However, possibilities for improvement are discussed and will be the subject of further research.

**Keywords:** remote sensing; forest inventory; angle count sampling; timber volume modelling; aerial images; digital surface models; canopy height models

## 1. Introduction

The utilisation of remote sensing (RS) data—especially aerial images—for gaining information about forests has a long tradition [1], but for a long time, information extraction was laborious and expensive. Improved availability of digital image data and enhanced computational power, as well as advances in automatic image data processing, created new opportunities for fast and extensive information extraction. Dense image matching allows automatic generation of high-resolution digital surface models (DSM) from aerial stereo images, which enable the extraction of 3D data for forest stands. Another means of extracting this kind of information is the utilisation of airborne laser scanning (ALS), which directly measures 3D points for DSM generation. However, ALS data acquisition is more expensive than acquisition of aerial imagery. This reduces the availability and timeliness of ALS data compared to image data. Height information extracted from RS data sources was found to be well correlated with heights measured in the field e.g., [2]. This relationship was further used to model and predict many other forest variables, among them timber volume and biomass e.g., [3]. In forest inventories, this data can be used to improve the precision of forest variable estimates by combining

very high resolution RS data with forest inventory plot data in a modelling process. Deo et al. [4] noted that in plot-based approaches, accurate statistical modelling for forest inventories requires geometric consistency between the terrestrial inventory data and RS data. That is, location, size, and shape of plots for inventory data collection have to correspond to the geometries used for RS data extraction. The size and shape of a sample plot, and the distribution of sampled and non-sampled trees within this plot, affect the fit between plot-based forest variables and metrics derived from remotely sensed data. Therefore, the configuration of sample plots—that is, the methods for selecting sample trees—is a factor that should be considered when combining inventory plot data with RS data. This greatly influences the predictive power of RS data metrics.

Fixed-area plots are considered “ideal when working with remotely sensed data” [4], as geometric inconsistencies tend to be minimal. This is very different for angle-count sampling (ACS), which is a well-established method for forest inventory data acquisition that does not utilise fixed-area plots. The ACS method was developed by Bitterlich in 1948 [5]. Sample trees are selected with a probability proportional to their basal area, which is usually defined through the diameter at breast height (DBH), with a predefined basal area factor (BAF) used as weight [4–6]. Therefore, some trees might be close to the sampling point, but are not sampled, while other trees further away will be sampled, resulting in sample plots of unknown and varying shape and size. In fact, ACS has no actual plot size [7] and the trees sampled might be different from the trees sampled using a fixed-area plot with the same plot centre [8]. This leads to an increased geometric inconsistency between inventory and RS data, compared to fixed-area plots, impeding accurate statistical modelling in plot-based approaches [4].

In order to allow integration of ACS and RS data, the problem of geometric inconsistency needs to be attenuated. Finding the representative extent of ACS plots has been a topic of research for a long time. Matérn [7] aimed to find a plot size that allows ACS to be compared to fixed-area sampling with the same number and distribution of samples. He found that the plot area equals the ratio of the mean basal area to the BAF requiring some knowledge about diameter distribution.

More recently, the motivation behind finding a suitable ACS plot size is to enable the utilisation of ACS data together with RS data. So far in the literature, mostly ALS data was used with this prospect, but there are also instances of utilising optical image data. Maack et al. [9] examined the variation of ACS radii and chose a fixed-radius of 20 m as a compromise between average ACS radius and sensor (ALS and Landsat 7) resolution. Hollaus et al. [10] determined an appropriate fixed-radius sample plot size by testing five different radii, choosing the one leading to the highest accuracy in a predictive timber volume model. A similar approach was chosen by [11]. They selected the best suited plot size for timber volume determination by analysing the coefficient of determination and the standard deviation. Scrinzi et al. [12] used ACS data that were simulated using fixed-radius sampling data. They tested three approaches for approximating “variable pseudo areas” of ACS that are suitable for extraction of ALS data: using the fixed-radius plot area and two variations of the plot-size formula in [7]. They found that ACS was marginally better for model calibration than fixed-radius sampling. Deo et al. [4] optimised the plot area for ALS data extraction for timber volume modelling depending on different BAFs. They noted that the optimum plot area for ACS inventories is dependent on forest structure and the chosen BAF. Maltamo et al. [6] used truncated ACS plots with a BAF of 2 and a maximal plot radius of 9 m. Thus, trees of smaller diameters (e.g.,  $DBH \leq 25.5$  cm) are being sampled using the normal ACS approach. Larger trees (e.g.,  $DBH > 25.5$  cm), however, will be sampled comparable to fixed-radius sampling, because the maximal plot radius is the cut-off distance for sampling. Thus, for larger trees, the angle-count plot equals a fixed-radius plot [6]. They modelled several plot-based variables (timber volume, basal area, number of stems, basal area median diameter, and basal area median height) and achieved Root Mean Square Errors (RMSEs) comparable to modelling using the fixed-radius plot data. They noted that ACS without truncation would lead to less accurate results and complicate the selection of a plot size for ALS data extraction. Immitzer et al. [13] used multiple, concentric circles of radii from 5–20 m for extracting metrics from optical stereo satellite images (WorldView-2). The combined metrics from all circles were joined with ACS

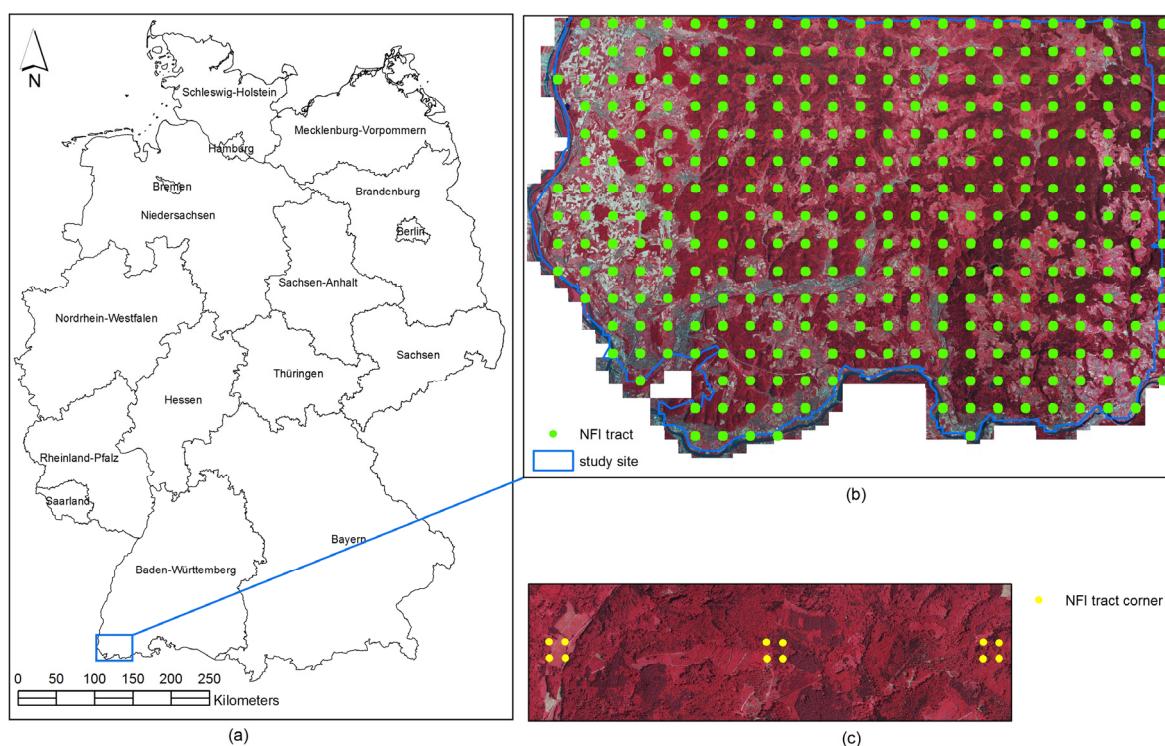
data in a Random Forest model for wall-to-wall mapping of growing stock. They also produced results using several specific circle radii separately and found that using multiple radii combined led to lower RMSEs. This is explained by single circles not entirely meeting the design of ACS plots, with sample trees always being excluded or non-sample trees being included [13].

This study considers a novel approach for integration of ACS plot data and RS data for forest variable modelling. This approach aims to transfer the concept of ACS in the extraction of RS-based descriptive metrics in order to better meet the design of ACS plots. In this study, only 3D data (height) is extracted for modelling.

## 2. Materials and Methods

### 2.1. Study Site and Reference Data

The study site is located in the south-west of Germany in the federal state of Baden-Württemberg (Figure 1), covering about 120,000 ha of the southern parts of the Black Forest and the Upper Rhine Plain. This results in very diverse topographic conditions, with flat areas in the Rhine plain, and hills with partly steep slopes in the Black Forest mountain range. Altitudes range between 205 m and 1220 m above sea level. Hence, the forested areas within the study site—which cover approximately 64,150 ha—vary greatly in terms of species composition. The lower regions in the Rhine plain are dominated by broadleaves (mainly *Fagus sylvatica* (26%), *Quercus spec.* (8%), and *Fraxinus excelsior* (3%)), while the upper regions are dominated by conifers (mainly *Picea abies* (30%), *Abies alba* (19%), and *Pseudotsuga menziesii* (9%)). However, many stands are a mixture of both conifers and broadleaves, to a varying degree. The age class composition of stands varies to a similar degree.



**Figure 1.** Location and extent of study site: (a) Location of study site in Germany; (b) Extent of study site and distribution of national forest inventory (NFI) tracts superimposed on colour-infrared (CIR) orthophotos; (c) Details of NFI tract layout with four tract corners at each tract that establish NFI plot centres superimposed on CIR orthophotos.

The study site encompasses 552 woodland plots of the latest national forest inventory (NFI) in Germany, which was conducted 2011–2012. The sampling design of the NFI in Germany utilises a  $4\text{ km} \times 4\text{ km}$  grid aligned with the Gauss-Krüger coordinate system. In the federal state of Baden-Württemberg, however, the density of the grid is increased to  $2 \times 2\text{ km}$ . Each grid point defines the centre of one square inventory tract of  $150\text{ m} \times 150\text{ m}$ . The corner points of the tracts define the centres of the NFI sampling plots. The plots were geolocated by the field crews using the MxBox Global Navigation Satellite System (GNSS) device of GEOSat GmbH [14]. On each plot sample, trees are selected using ACS with a BAF of 4 and a minimum DBH of 7 cm. For each sample tree, azimuth and distance to the plot centre, as well as DBH and species is recorded, amongst other variables. Additionally, heights are measured for a subset of sample trees. The heights of the remaining sample trees are subsequently modelled.

## 2.2. Remote Sensing Data

This study utilised digital aerial stereo images acquired during the regular aerial surveys of the Baden-Württemberg land surveying authority (LGL). The aerial imagery is characterised by a forward overlap of 60%, side overlap of 30%, and a nominal ground resolution of 20 cm. The images were acquired by large size digital aerial matrix camera (UltraCam Eagle) in one survey flight in June 2013 with four spectral bands (red (R), green (G), blue (B), and near infrared (NIR), and a focal length of 79.8 mm. According to LGL, the orientation accuracy after aerotriangulation is 0.04 m in planimetric coordinates (x and y) and 0.14 m in height (z) on control points. The data delivery from LGL included the imagery and an Inpho project file that comprises the exterior and interior orientation, ready to be used for subsequent photogrammetric processing. In addition, a high quality digital terrain model (DTM) with 1 m resolution was also provided by LGL. The DTM was derived from ALS data collected between 2001 and 2004, with an approximate point density of 0.8 points per square meter. The nominal height accuracy of this ALS-DTM reported by LGL is 0.5 m or better [15].

A high resolution (0.4 m) DSM point cloud was generated from the stereo images in a dense image matching process using the software SURE of nFrames. Image matching is performed based on stereo image pairs and the results of multiple stereo pairs are subsequently fused [16]. The implemented matching algorithm is a modification of the SGM algorithm proposed by [17]. The software's default parameter set for aerial images of the given overlap was used. Following adaptions were made according to the experience of the research team with image matching for forest areas: triangulation will accept a maximum angle of  $99^\circ$  and 1 as minimum detections per cell. The interpolation for the extraction of a regular gridded point cloud was done with inverse distance weighted. A detailed description of the matching algorithm implemented in SURE can be found in [16,18].

After point cloud extraction, all points identified as outliers were eliminated from each cloud. These outliers were identified by considering the maximum tree height at the study site and the ALS-DTM available as reference. All points with elevations of 55 m and more above, and 1 m and more below the DTM, were regarded outliers and excluded from the point cloud. The upper threshold is based on the assumption that trees in this area are usually not taller than 55 m. The lower threshold allows for some height inaccuracy in the point cloud.

A standard process was applied to generate a canopy height model (CHM) from each filtered point cloud. First, a raster-DSM with spatial resolution of 1 m (equivalent to the resolution of the ALS-DTM) was derived from each point cloud using the software LAStools [19]. Each DSM pixel value was determined by the elevation of the highest point within the planimetric pixel extent ( $1\text{ m}^2$ ). In areas without matching points, the pixel values were calculated via Triangular Irregular Network (TIN) streaming [20]. The CHMs with 1 m resolution were obtained by subtracting the ALS-DTM from the photogrammetric DSMs.

### 2.3. Transferring ACS in Height Metric Extraction

Traditionally, plot-based metrics are extracted from raster data by selecting relevant pixels using a simple geometric shape—for example circle or rectangle—and in some cases, a pixel value (height) threshold. This is a suitable approach when the extraction geometry corresponds to the actual field sample plot geometry. However, ACS plots cannot be represented by such a simple geometry and pixels that do not correspond to a sample tree might be used for metric calculation. In addition, some sample trees might not be represented in the metrics if they are located outside the chosen geometric shape. In this initial study, a novel approach was devised and tested that seeks to extract only data that most likely corresponds to actual sample trees. The basic idea is to transfer the concept of sample tree selection in ACS. However, ACS requires some knowledge about the DBH of trees. This information is not available from airborne RS data and its derivatives such as CHMs. This is solved by utilising the relationship between DBH and tree height. So, for each pixel height a theoretical DBH can be calculated and compared to a minimum DBH, which depends on the horizontal distance between pixel location and plot centre. A single pixel does not represent a tree. However, a pixel of a certain height can be considered part of the crown of a tree that is an actual sample tree. This approach is expected to select at least the upper parts of sample tree crowns. For large trees (large basal area), more pixels will be selected than for smaller trees at the same distance from the plot centre. Nevertheless, all trees where at least some upper parts of the crown reach a certain height will be selected, with the amount of pixels selected being dependant on the tree size and its distance from the plot centre. Thus, the extraction of RS data mimics the sample tree selection process in ACS. This approach is detailed in the following sections. In this study, all data extraction, analysis, and modelling were implemented using the statistical software R version 3.3.1 [21].

#### 2.3.1. Establishing Relationship between DBH and Tree Height

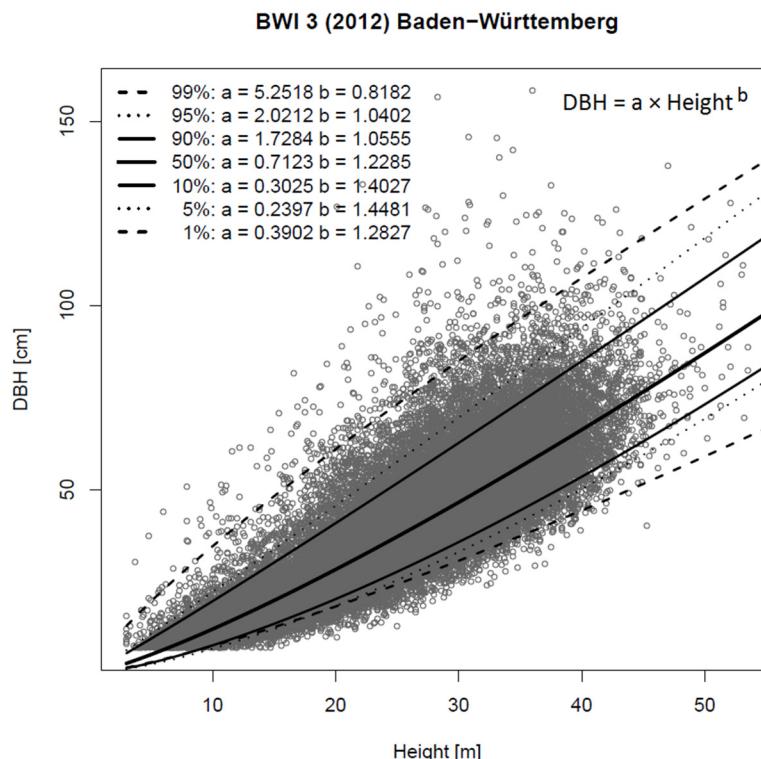
The relationship between DBH and tree height is a common and widely used factor in tree size determination [22]. However, this relationship is influenced by other factors such as tree species, stand characteristics, and site conditions [23]. In order to develop a practical data extraction approach that is applicable to a wide range of forest types and growing conditions, it is necessary to establish a generalised mathematical relationship between DBH and tree height. For this study, all tree heights and DBHs measured during the latest German NFI (BWI3) on plots located in Baden-Württemberg were included in a quantile regression analysis (Figure 2) using the interior point method for computing nonlinear quantile regression estimates as it is implemented in the function “nlrq” of the R-package “quantreg” [24]. The resulting relationship between DBH and tree height is denoted in Equation (1)

$$DBH = a \times (H_{NFI})^b, \quad (1)$$

where  $DBH$  is the diameter at breast height,  $H_{NFI}$  is the measured tree height and  $a$  and  $b$  are regression parameters that are dependent on the chosen quantile. The regression parameters were calculated for the 99%, 95%, 90%, 50%, 10%, 5%, and 1%-quantiles. However, no quantiles above 50% were chosen for further investigation, as this would assume relatively large diameters at a certain height. With quantiles above 50%, the DBH-height-relation would imply that more than 50% of all actually measured ACS trees have a DBH too small for inclusion in an angle-count sample. Furthermore, the 5%-quantile did not show any substantial difference to the 1%-quantile during analysis of results and was excluded in this publication in order to simplify the presentation of results. The parameters for the remaining quantiles of 50%, 10%, and 1%, which were used for further investigation, can be found in Table 1.

**Table 1.** Regression parameters  $a$  and  $b$  of quantile regression for 50%, 10% and 1%-quantiles.

Quantile	$a$	$b$
50%	0.7123	1.2285
10%	0.3025	1.4027
1%	0.3902	1.2827

**Figure 2.** Result of quantile regression analysis of DBH-height relationship using all NFI (BWI3) sample tree data for Baden-Württemberg.

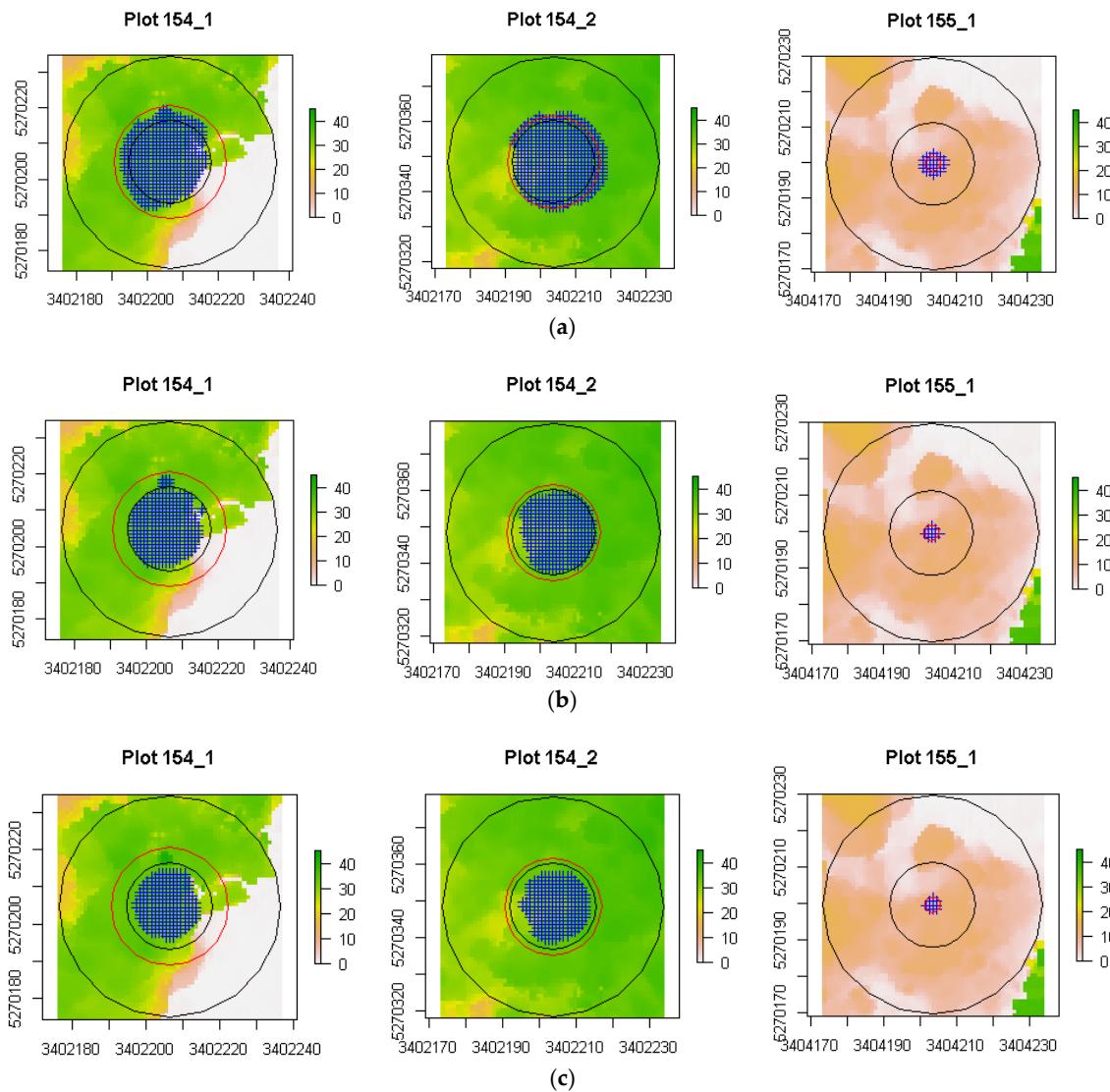
### 2.3.2. Data Extraction Algorithm

Pixel data at each NFI plot is extracted for each parameter set of the quantile regression separately using the following “transfer ACS” (tACS) approach. An initial search radius of 30 m around each plot centre was chosen, which is expected to cover all probable sample tree to plot centre distances in ACS with a BAF of 4. This decision was based on the fact that in the NFI data used in this study, no sample tree was more than 27 m distant from the plot centre. Choosing a 30 m radius allows extraction of all pixels belonging to a sample tree that is located at a distance of 27 m from the plot centre, with parts of its crown being even farther from the centre. For different data sets, one might consider using a different search radius. For each pixel where the centre coordinate falls within the search radius, the height value, as well as the coordinate were extracted and the distance to the plot centre was calculated. This distance was used for calculating a minimum DBH at each pixel by implementing Equation (2):

$$\min DBH_p = \frac{dist_p}{\sqrt{\frac{2500}{BAF}}}, \quad (2)$$

where  $\min DBH_p$  is the minimum DBH at pixel  $p$ ,  $dist_p$  is the distance between pixel  $p$  and the plot centre and BAF is the basal area factor, which in this study was set to 4. In cases where the thus calculated  $\min DBH$  was below 7 cm—which is the minimum required DBH for NFI field sampling in Germany—the calculated value was overridden and  $\min DBH$  set to 7 cm instead. Subsequently, a

theoretical DBH ( $tDBH$ ) was calculated for each pixel using Equation (1) and the respective parameters of Table 1. A pixel where  $tDBH$  is equal to or larger than  $minDBH$  is considered part of a tree that was probably sampled during NFI field data acquisition. Pixels where  $minDBH$  is not met are flagged as ‘not sampled’ and are eventually removed from the set of extracted pixels. All pixels that remain in the data set for each plot are used for metrics calculation (Section 2.4.1). Graphic representations of the results of tACS pixel data extraction for three exemplary plots can be found in Figure 3.



**Figure 3.** Example of selected pixels (blue crosses) in three different plots for each transfer angle-count sampling (tACS) data set. Background image is the CHM, the outer black circle the 30 m search radius, the red circle the terrestrially measured maximum distance between plot centre and sample trees ( $maxDist_{ind}$ ) and the inner black circle the median of maximum distances ( $maxDist_{med} = 11.7$  m) for reference. (a) tACS<sub>q50</sub>; (b) tACS<sub>q10</sub>; (c) tACS<sub>q1</sub>.

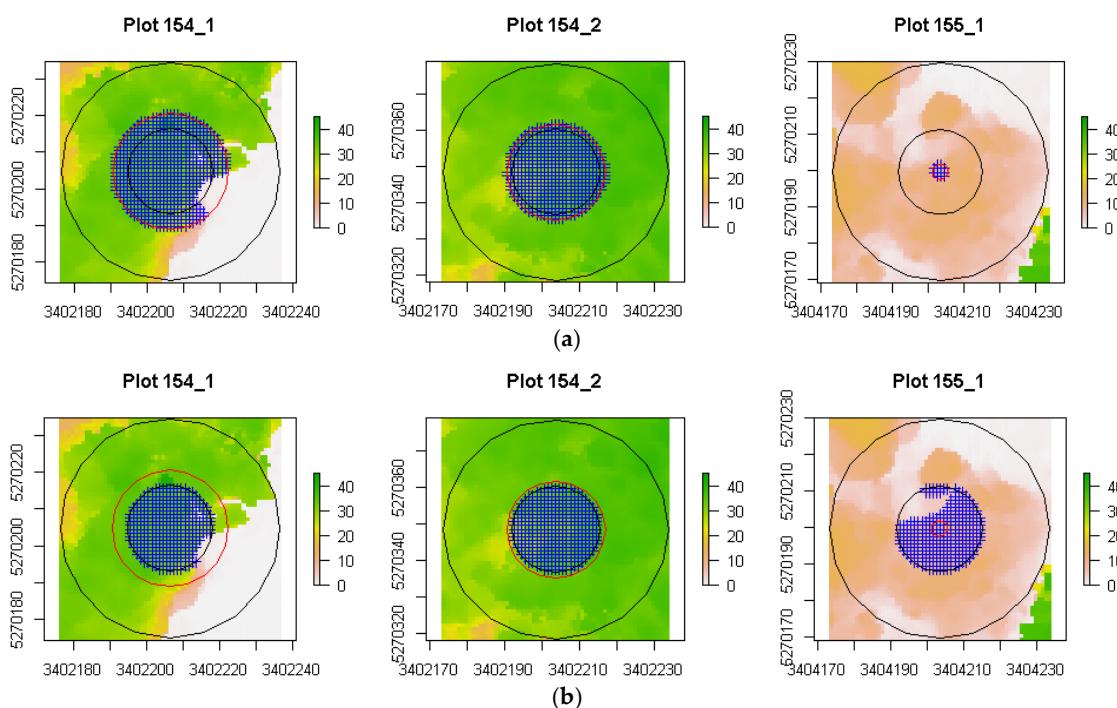
For comparison reasons, pixels were also extracted using two variations of the conventional method with fixed-radius plots. The radii used were derived in two different ways. In the first variation the maximum distance between plot centre and sample trees were derived for each plot individually ( $maxDist_{ind}$ ). This maximum distance was measured during the terrestrial inventory measurements at each plot and was used as the fixed radius for the respective plot centre. Of course, this is not a viable approach for actual forest inventory applications, but is expected to represent an ideal

plot size for reference. In the second variation the median of all maximum distances (of  $\text{maxDist}_{\text{ind}}$ ) was calculated and used as fixed radius for all plot centres ( $\text{maxDist}_{\text{med}}$ ). In this study the calculated radius for  $\text{maxDist}_{\text{med}}$  was 11.7 m. All pixels that are overlapped by a circle of the respective plot radius were extracted. In a second step all pixels with a height value below 6 m were removed from this set of pixels. Six metres is a threshold that resulted from the expert interviews of forest management planning at ForstBW—the Baden-Württemberg state office for forestry. It approximates the height of trees when they reach the minimum required DBH of 7 cm. This threshold was successfully applied in a project for improving forest management planning with RS methods at Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg (FVA), Germany. It helped to improve models for timber volume estimation based on enterprise inventory data with sampling on concentric plots. Graphic representations of the results of the fixed-radius data extraction approaches for three exemplary plots can be seen in Figure 4.

In total five variations of CHM data extraction approaches were applied, resulting in five data sets. Three variations use transfer ACS ( $\text{tACS}_{q50}$ ,  $\text{tACS}_{q10}$ ,  $\text{tACS}_{q1}$ ) and two variations use circular geometries for data extraction ( $\text{maxDist}_{\text{ind}}$ ,  $\text{maxDist}_{\text{med}}$ ). The variations are listed in Table 2.

**Table 2.** Data set resulting from the differing data extraction approaches and their variations.

Data Set	Geometry	Data Extraction
$\text{tACS}_{q50}$	complex	transfer ACS; 50%-quantile for DBH-height relationship
$\text{tACS}_{q10}$	complex	transfer ACS; 10%-quantile for DBH-height relationship
$\text{tACS}_{q1}$	complex	transfer ACS; 1%-quantile for DBH-height relationship
$\text{maxDist}_{\text{ind}}$	in general circular	plot radius = max. distance; min. height 6 m
$\text{maxDist}_{\text{med}}$	in general circular	plot radius = median max. distance; min. height 6 m



**Figure 4.** Example of selected pixels (blue crosses) in three different plots for each fixed-radius data set. Background image is the canopy height model (CHM), the outer black circle is the 30 m search radius from the tACS data sets, the red circle is the terrestrially measured maximum distance between plot centre and sample trees ( $\text{maxDist}_{\text{ind}}$ ), and the inner black circle is the median of maximum distances ( $\text{maxDist}_{\text{med}} = 11.7$  m) for reference: (a)  $\text{maxDist}_{\text{ind}}$ ; (b)  $\text{maxDist}_{\text{med}}$ .

### 2.3.3. Outlier Removal

NFI field data in this study was acquired between 2011 and 2012, while the aerial imagery was acquired in 2013. This time lag of 1–2 years can introduce a source of error due to tree growth and trees disappearing, which might be caused by forest operations or natural factors. Furthermore, insufficient spatial co-registration might lead to improper data being extracted. Plots that are affected by this can be considered outliers, and should be identified and removed from the data set before establishing statistical models for forest variable estimation [25]. In this study, outliers were removed by comparing the maximum measured tree height ( $\max(H_{NFI})$ ) at each plot with the maximum height in the CHM ( $\max(H_{CHM})$ ) of this plot. Plots where  $\max(H_{CHM})$  differed more than 25% from  $\max(H_{NFI})$  were considered outliers and removed from the data sets. The percentage of difference was calculated as follows (Equation (3)):

$$devH_{max} = \left| 100 - \frac{100}{\max(H_{NFI}) * \max(H_{CHM})} \right|, \quad (3)$$

where  $devH_{max}$  is the percentage of difference,  $\max(H_{NFI})$  is the maximum measured height on the NFI plot, and  $\max(H_{CHM})$  is the maximum CHM height on this plot. The plot area for CHM data extraction was defined by the maximum measured tree to plot centre distance of the respective plot plus a 2 m-buffer. The maximum distance is measured to the centre of the respective tree, which is probably one of the largest trees in the plot. The size of the buffer was chosen in order to allow for inaccurate positioning of plot centres and non-verticality of tree stems. With this approach, 34 plots were identified as outliers and removed from the data sets, reducing the number of plots for metric calculation to 518.

## 2.4. Descriptive Metrics and Modelling

Descriptive metrics were calculated and timber volume models were created separately for each data set in Table 2.

### 2.4.1. Metrics Calculation

For each plot in each data set, 20 plot-based descriptive metrics were calculated from the extracted CHM pixels. Those were 18 metrics related to height—5th ( $p5$ ), 10th ( $p10$ ), 15th ( $p15$ ), 20th ( $p20$ ), 25th ( $p25$ ), 50th ( $p50$ ), 75th ( $p75$ ), 80th ( $p80$ ), 85th ( $p85$ ), 90th ( $p90$ ), 95th ( $p95$ ), and 99th ( $p99$ ) percentiles, arithmetic mean ( $mean$ ), minimum ( $min$ ), maximum ( $max$ ), standard deviation ( $sd$ ), variance ( $var$ ), coefficient of variation ( $cv$ ), which are commonly used in other studies (e.g., [26]). Two metrics were related to volume—inner volume ( $volin$ ) and outer volume ( $volout$ ). The metric  $volin$  describes the volume below the pixels extracted from the CHM (Equation (4)), that is the 3D space that is probably occupied by trees [27].

$$volin = \sum CHM_{ep} \times res^2, \quad (4)$$

where  $CHM_{ep}$  are the extracted CHM pixels and  $res$  is the geometric resolution of the CHM. The  $volout$  metric is the volume between the outer canopy boundary and the maximum bounding box of the canopy [27] (Equations (5) and (6)):

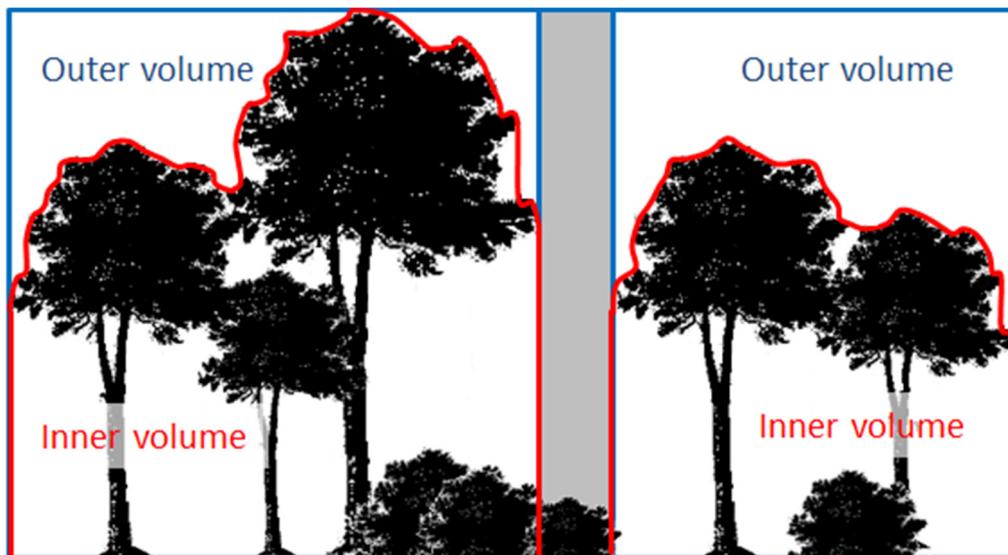
$$maxVol_{CHM} = max(H_{ep}) \times res^2 \times count(ep), \quad (5)$$

where  $maxVol_{CHM}$  is the volume defined by the maximum height in the CHM,  $H_{ep}$  are the heights of the extracted pixels,  $res$  is the geometric resolution of the CHM and  $count(ep)$  is the number of extracted pixels.

$$volout = maxVol_{CHM} - volin, \quad (6)$$

Both volume metrics provide information about the complexity of the outer canopy structure (Figure 5). Furthermore, the number of extracted pixels (*n<sub>pix</sub>*) were derived and included as metric.

One additional plot-based metric that was not derived from the CHM, but from the DTM, is the mean ground elevation (*meanDTM*). This mean was calculated based on all DTM data within a radius of 30 m around each plot centre. This metric is the same for all six CHM data extraction approaches.



**Figure 5.** Schematic description of inner volume (*volin*) and outer volume (*volout*). The grey area, which represents space over vegetation of less than 6 m height, is excluded from any volume calculation.

#### 2.4.2. Modelling

In total, 22 metrics were derived for timber volume modelling (Section 2.4.1). Some of these metrics were highly correlated and a correlation analysis was conducted in order to avoid collinearity. Metrics that were correlated (Pearson) by more than 80% to each other were carefully examined and the metric that was most correlated with timber volume was kept. This analysis was conducted for all data sets and the results were combined in one initial set of metrics that were used for timber volume modelling.

Linear models were fitted to each data set. Linear models were chosen for all data sets to be able to compare the effect of transferring the concept of ACS to RS-assisted timber volume modelling. Nonlinearity in the data was handled by adding some of the predictor metrics additionally as squared terms to the models. The initial maximum models were stepwise reduced based on Akaike information criterion (AIC) using forward and backward selection. Furthermore, square root transformation of the response and weighted least squares were tested during modelling. For applying weights, it was necessary to account for plots where no pixels were selected and all metrics were 0. This was achieved by adding a small absolute term (0.01) to the denominator. This value is much lower than the measurement accuracy of heights within aerial image derived CHMs.

#### 2.5. Validation

For each model, root mean square error percent (RMSE%) and *bias*% of model predictions were estimated for each fold in a 10-fold cross-validation as presented in Table 3. For each model, RMSE% and *bias*% values from each fold were averaged. Furthermore, standard deviation of *bias*% from each fold was calculated.

Several models with varying combinations of predictive metrics were tested for each data set and the model with lowest RMSE% was chosen as best model. These “best models” were used for comparing and assessing the performance of the different data extraction approaches.

These best performing models were used to predict timber volumes on all 518 sample plots. Model generation and volume prediction is consequently based on the same data set. This will lead to a prediction accuracy that is not realistic for practical wall-to-wall applications. However, these values are considered sufficient for comparison of the different data extraction approaches.

**Table 3.** Equations for calculating Root Mean Square Error (RMSE), RMSE%, and bias%. In each equation,  $\hat{y}_i$  and  $y_i$  represent predicted and observed timber volume for sample plot  $i$ , respectively, and  $n$  is the number of samples.

Metric	Equation
RMSE ( $m^3/\text{ha}$ )	$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
RMSE (%)	$100 \times \frac{\text{RMSE}}{\frac{1}{n} \sum_{i=1}^n (y_i)}$
Bias (%)	$100 \times \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{\sum_{i=1}^n (y_i)}$

### 3. Results

The initial, uncorrelated variables used for modelling for each data set were *mean*, *npix*, *cv*, *volout*, and *meanDTM*. The predictive metrics used for the best performing models for each data set are listed in Table 4. The number of extracted pixels (*npix*) and the mean ground elevation (*meanDTM*) were selected for all data sets, indicating high robustness of these metrics for modelling timber volume. The metrics *mean* and *cv*, and the interaction *mean* and *npix* (*mean*  $\times$  *npix*), were selected four of five times. Two metrics (*volout* and the squared *mean* (*mean*<sup>2</sup>)) were selected only once, each with one of the fixed-radius approaches. Table 4 also shows that including weights in the model only had a positive effect for the fixed-radius approaches.

**Table 4.** Predictive metrics selected in best performing model for each data set (*mean*<sup>2</sup> is the squared *mean* and *mean*  $\times$  *npix* is the interaction between metrics *mean* and *npix*).

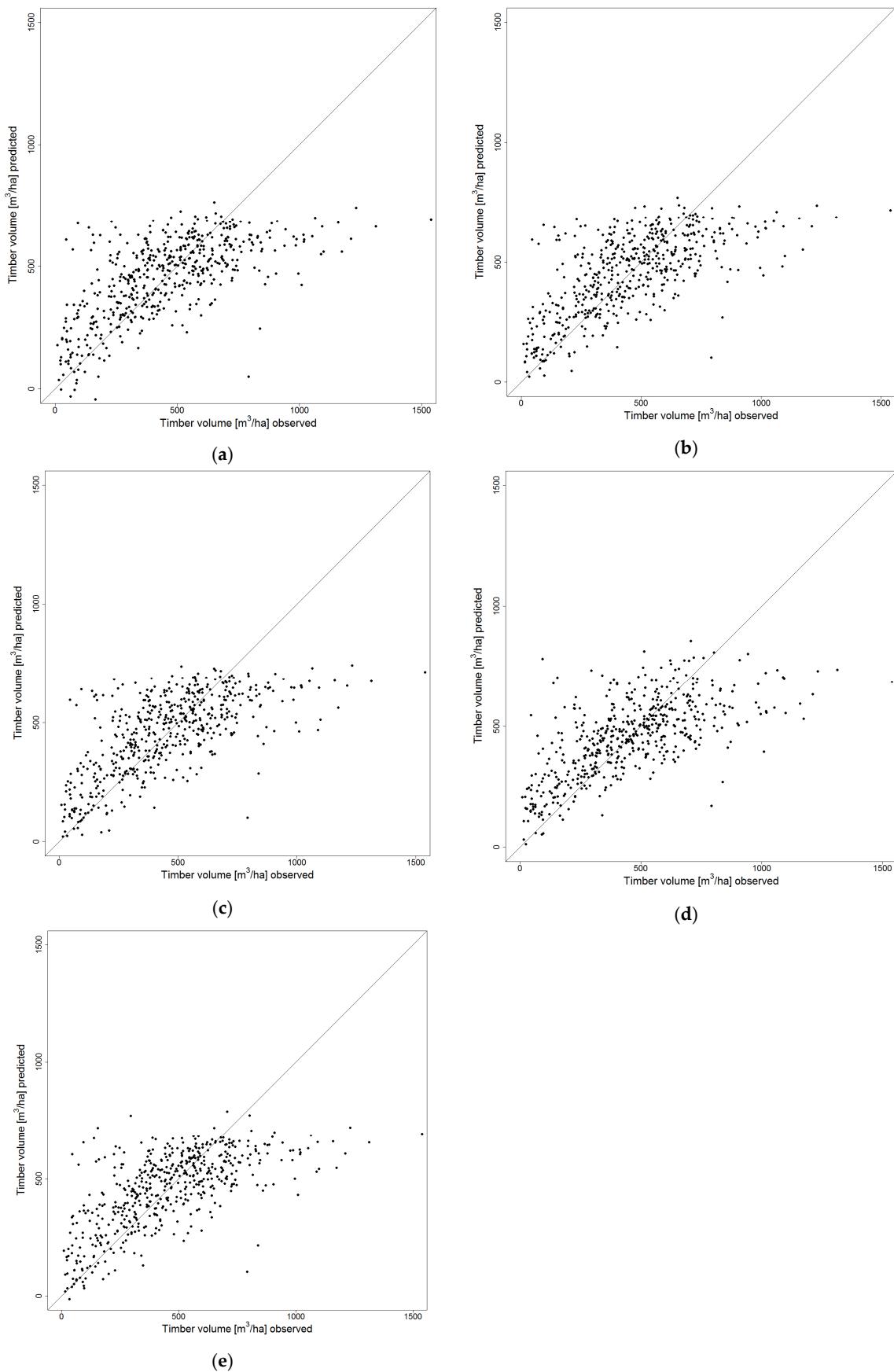
Data set	<i>mean</i>	<i>mean</i> <sup>2</sup>	<i>npix</i>	<i>mean</i> $\times$ <i>npix</i>	<i>cv</i>	<i>meanDTM</i>	<i>volout</i>	weights
tACS <sub>q50</sub>	x	-	x	x	x	x	-	-
tACS <sub>q10</sub>	x	-	x	x	x	x	-	-
tACS <sub>q1</sub>	-	-	x	x	x	x	-	-
maxDist <sub>ind</sub>	x	-	x	-	x	x	x	$1/\text{mean}^1$
maxDist <sub>med</sub>	x	x	x	x	-	x	-	$1/\text{mean}^1$

<sup>1</sup> An absolute term of 0.01 was added to *mean* in order to account for plots where no pixels were selected and *mean* = 0.

Table 5 displays the RMSE% achieved in a 10-fold cross-validation for all data sets. The values range between 41.57–42.43%, with both the lowest (tACS<sub>q1</sub>) and the highest (tACS<sub>q50</sub>) values achieved with the transferred ACS approach. The difference is less than one percentage point. Bias% varied around zero between negative and positive values for all data sets (Table 5), reflecting underestimation in high volume ranges and overestimation in low volume ranges, respectively (Figure 6).

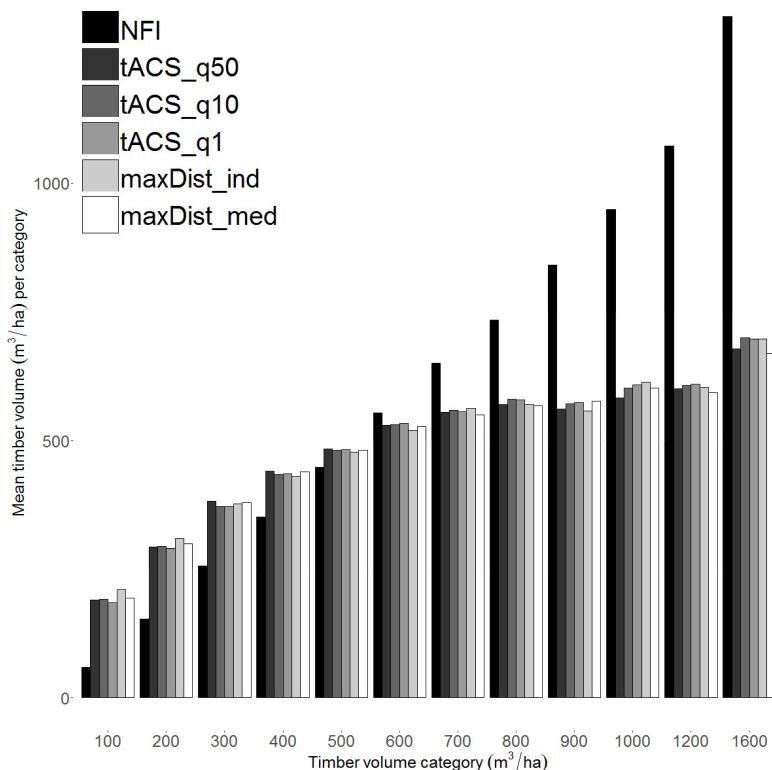
**Table 5.** Mean of Root Mean Square Error percent (RMSE%) and mean and standard deviation (sd) of bias% for each data set achieved in 10-fold cross-validation.

Data Set	RMSE%	Bias% Mean (sd)
tACS <sub>q50</sub>	42.43	0.0 (4.55)
tACS <sub>q10</sub>	41.73	0.02 (4.94)
tACS <sub>q1</sub>	41.57	0.04 (4.69)
maxDist <sub>ind</sub>	42.04	0.14 (4.36)
maxDist <sub>med</sub>	42.23	0.13 (5.13)



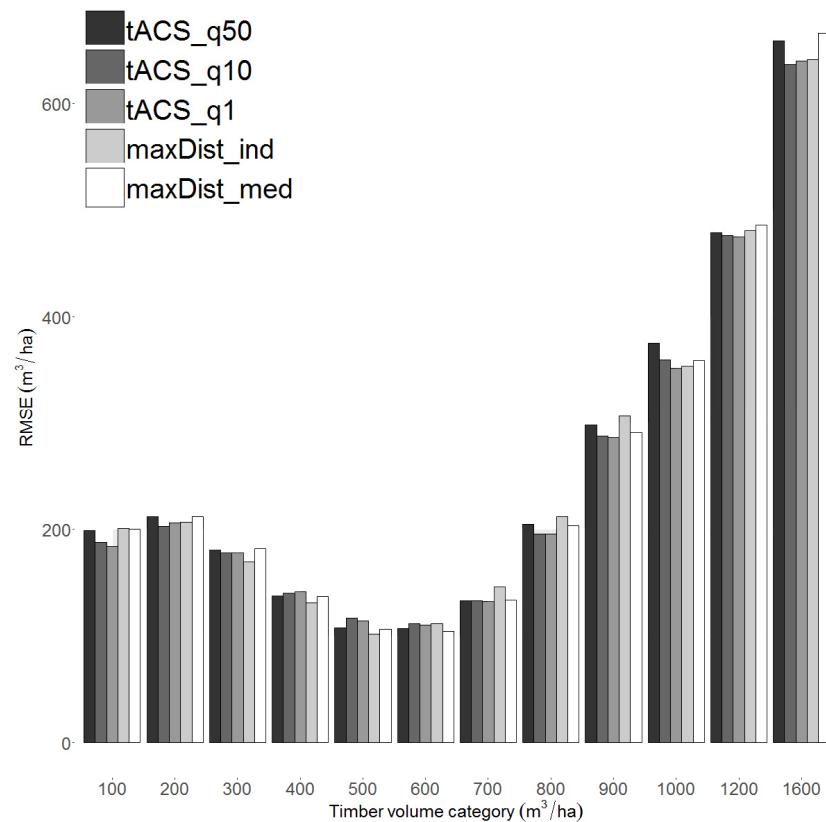
**Figure 6.** Scatterplots of observed and predicted timber volumes at each plot for all data sets: (a) tACS<sub>q50</sub>; (b) tACS<sub>q10</sub>; (c) tACS<sub>q1</sub>; (d) maxDist<sub>ind</sub>; (e) maxDist<sub>med</sub>.

Figure 6 shows the scatterplots of observed and predicted timber volume for the best performing models of all data sets using all plots. In all instances, there is a distinct roof-like limit for predictions at about  $700 \text{ m}^3$  solid volume including bark per ha. This is least pronounced with the maxDist<sub>ind</sub> data set (Figure 6d). For high volumes, this results in severe underestimation. For some low volume plots, substantial overestimation can also be observed. This was already indicated in the values for bias% and is also reflected in Figure 7, which shows mean volume prediction for observed volume classes (class width:  $100 \text{ m}^3/\text{ha}$ ) for each data set. For these predictions, the same plots were used as for model generation. The mean timber volume of the NFI is also depicted, showing that the lower volume classes ( $100, 200, 300 \text{ m}^3/\text{ha}$ ) are overestimated in all data sets, while the upper volume classes ( $600\text{--}1600 \text{ m}^3/\text{ha}$ ) are underestimated. There were no significant differences between the predictions of the models ( $p = 1$ ).



**Figure 7.** Mean volume prediction for observed volume classes (categories) for each data set utilising all plots, as well as the mean volume terrestrially measured at all NFI plots.

Figure 8 depicts the absolute RMSE achieved with each data set in different volume classes (observed values). These values were calculated using all plots. The RMSE in the low volume classes ( $100\text{--}400 \text{ m}^3/\text{ha}$ ) reaches values up to  $200 \text{ m}^3/\text{ha}$ . It is lowest for the medium volume classes ( $500$  and  $600 \text{ m}^3/\text{ha}$ ) and rapidly increases from approximately  $150 \text{ m}^3/\text{ha}$  in the  $700 \text{ m}^3/\text{ha}$  class, to  $650 \text{ m}^3/\text{ha}$  in the  $1600 \text{ m}^3/\text{ha}$  class. There is no noteworthy difference between the different data sets.

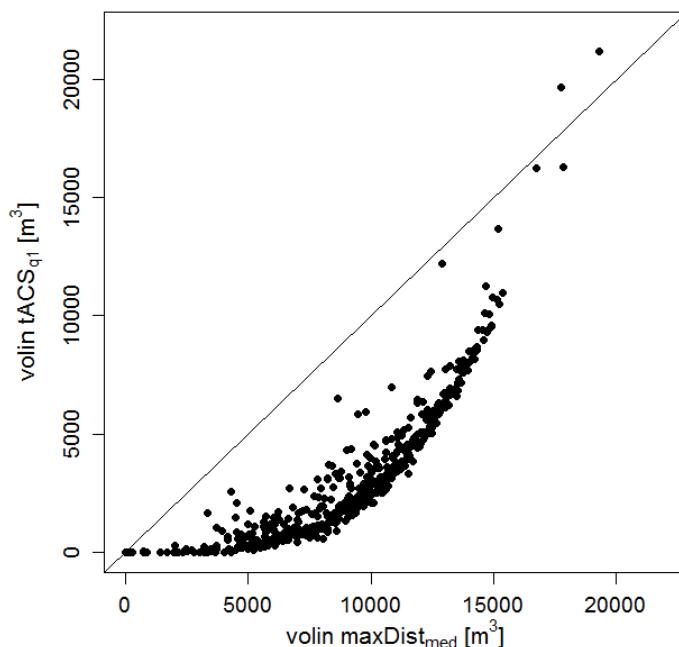


**Figure 8.** Root Mean Square Error (RMSE) achieved with each data set in different volume classes (categories) (observed values) utilising all plots.

#### 4. Discussion

This study was an initial investigation into a novel approach for integrating ACS and RS data for enhancing forest inventories. This new approach transfers the concept of ACS to the RS data extraction process for calculation of predictive metrics, aiming to better represent the plot design of ACS. Figure 3 shows that transferred ACS actually extracts data in irregular plot shapes and, especially in the cases of  $t\text{ACS}_{q50}$  and  $t\text{ACS}_{q10}$ , meets the maximum measured distance of the terrestrial ACS data acquisition at the respective plot. The implementable fixed-radius approach ( $\text{maxDist}_{\text{med}}$ ) extracts data using a circle where the radius is usually smaller or larger than the maximum measured distance (Figure 4b), resulting in data extraction areas that do not correspond to the actual area from which sample trees were selected. Deo et al. [4] suspected geometric inconsistency between RS and ACS field data as a source for decreased modelling precision, especially for old stands. It was therefore expected to improve timber volume modelling using the transferred ACS approach. However, the current implementation of this approach did not result in an improvement compared to the conventional fixed-radius plot approach. In fact, the achieved RMSE% after 10-fold cross-validation (Table 5) differed by less than one percentage point and demonstrates that the transferred ACS and fixed-radius approaches led to predictions of similar quality. One explanation could be the missing information about understory sample trees in image-based RS data. This might influence both data extraction approaches in the same way and could indicate that tACS might actually be more successful when applied to ALS data, where understory information can be retrieved. However, further tests restricted to single layer stands will be necessary in order to verify this. Furthermore, comparing the height metrics of the different approaches directly reveals that the values of the respective plots vary notably. The most obvious difference occurs in the metric *volin*. Figure 9 exemplarily shows the relation between *volin* from fixed-radius plots ( $\text{maxDist}_{\text{med}}$ ) to *volin* based on one variation of the tACS approach ( $t\text{ACS}_{1q}$ ). The  $t\text{ACS}_{1q}$  *volin* metric

increases exponentially compared to the  $\text{maxDist}_{\text{med}}$  *volin* metric. This relationship was discovered during the assessment of results in this study and it could not be sufficiently exploited in the volume estimation, yet. The form of the exponential relationship suggests that it might be helpful in adjusting for the over- and underestimation of timber volume (Figure 6). This requires further investigation into the usability of the information contained in the relationship in Figure 9.



**Figure 9.** Relation between metric *volin* from *maxDist<sub>med</sub>* and *volin* from *tACS<sub>q1</sub>*.

With the current linear model used, all different approaches show a typical pattern of over- and underestimation of timber volume. This leads to two different interpretations.

1. The variation of volume within the ACS data cannot be explained by the height metrics used for modelling.
2. The reference data and the explaining data do not match regarding their position and content.

This pattern of over and underestimation was also observed in other studies where either photogrammetrically derived CHMs (e.g., [28]) or ALS-based CHMs (e.g., [9]) were used for forest timber volume modelling. Maack et al. [9] suspected a unevenly distribution of samples with regard to timber volume, with most samples being in the mid-range between 200 and 400 m<sup>3</sup>/ha. They also remark that the relationship between growth in height and growth in diameter is not linear and changes depending on the age of the trees.

Other influences such as positional accuracy of plot location, bias in the NFI data due to non-detection [29,30], or spatial and species specific variety of height-volume relation, could significantly influence the model accuracy. This could be the reason for all different methods leading to similar trends in overestimation of lower volume classes and underestimation of higher volume classes. Unfortunately, when the NFI plots were established, no assessment of positional accuracy was carried out. Due to the general uncertainties of GNSS positioning under canopy, even the GNSS device's nominal accuracy is not a reliable indicator for the positional accuracy of NFI plots. Furthermore, the unknown shape and size of ACS plots hinders any reliable manual or automatic co-registration of ACS and RS data. This prevents a thorough assessment of the influence of the positional accuracy of terrestrial measurements on the model accuracy.

The highest potential for improvement is expected in adapting the DBH-height relationship utilised for CHM data extraction (Section 2.3.1). The DBH-height relationship was established using

all NFI data for the entire state of Baden-Württemberg, comprising data from very different regions in terms of site quality and prevailing forest types. Adapting this relationship to the geographic region of the study site by using data from this region only might improve the selection of relevant pixels of the CHM. Further adaption of the DBH-height relationship should consider the major differences between conifers and broadleaf trees. Under more homogeneous conditions—for example, predominantly coniferous forests—overall model accuracy is expected to improve, as edge effects and growing differences between tree species are reduced. For heterogeneous forests, separate functions of DBH-height relationship for conifers and broadleaves is expected to cause a similar effect and notably improve data extraction. Data could be extracted based on the dominant, plot-based tree type: conifers, broadleaves, and mixed. However, this requires a high-quality tree type classification. This classification can theoretically be achieved using the spectral information of the aerial images. Generally, including suitable spectral metrics in the model should improve prediction accuracy. However, aerial images suffer from changing illumination conditions between, as well as within, images [31,32]. Both affect the reliability and robustness of spectral metrics derived from aerial images, which makes them not well suited for forest variable prediction, unless the effect of changing illumination can be compensated.

Another possibility for improvement could be to use transferred ACS not for extracting single pixels from the CHM, but for defining an ideal plot radius for any plot position. The basic idea would be to use the DBH-height relationship to identify the pixel that is farthest away from the plot centre, while still indicating a tree that is theoretically an ACS sample tree. The distance between this pixel and the plot centre then defines the respective plot radius. This approach would be comparable to the approach using the individual measured maximum distance ( $\text{maxDist}_{\text{ind}}$ ), but would actually be usable in practical applications.

## 5. Conclusions

The described approach of tACS is a first step in adapting the information from airborne derived height information to terrestrial ACS. Despite the similar results of the tested methods, one advantage of the new approach is that it does not require preliminary investigations for finding a best suited plot radius for ACS plots, as was the case in [9–11], or extracting large amounts of data, as in [13]. Furthermore, the results presented here are based on some initial considerations towards enhancing integration of ACS field data and RS data. After all, the results did not worsen and there is still potential for further improvement of the approach.

The presented approach of tACS will be further developed, aiming to enhance modelling results. An investigation into whether all actual sample trees of the terrestrial ACS data acquisition were covered by extracted pixels is expected to support identification of the most promising alterations to the data extraction algorithm and modelling approach. Furthermore, a test regarding temporal and spatial robustness of the tACS metrics and models is planned to be carried out. Another interesting point for future research could be the comparison between the tACS approach and other approaches that seek to combine ACS and RS data, such as truncated ACS e.g., [6] and finding suitable fixed plot radii e.g., [10,11,13].

**Acknowledgments:** This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 633464 (DIABOLO). We acknowledge the provision of the remote sensing data by the land surveying authority of Baden-Württemberg (LGL).

**Author Contributions:** M.K. and J.S. conceived, designed and performed the experiments; M.K., J.S. and P.A. analysed the data; G.K. contributed materials; M.K., J.S. and P.A. wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. Spurr, S.H. *Aerial Photographs in Forestry*; Ronald Press: New York, NY, USA, 1948; p. 340.
2. Durrieu, S.; Véga, C.; Bouvier, M.; Gosselin, F.; Renaud, J.-P.; Saint-André, L. Optical remote sensing of tree and stand heights. In *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing*; CRC Press: Boca Raton, FL, USA, 2015; pp. 449–485.
3. Ginzler, C.; Hobi, M.L. Countrywide stereo-image matching for updating digital surface models in the framework of the Swiss national forest inventory. *Remote Sens.* **2015**, *7*, 4343–4370. [CrossRef]
4. Deo, R.K.; Froese, R.E.; Falkowski, M.J.; Hudak, A.T. Optimizing variable radius plot size and lidar resolution to model standing volume in conifer forests. *Can. J. Remote Sens.* **2016**, *42*, 428–442. [CrossRef]
5. Zöhrer, F. Zur Theorie der Winkelzählprobe für die Forstinventur. *Forstwiss. Cent.* **1973**, *92*, 53–68. [CrossRef]
6. Maltamo, M.; Korhonen, K.; Packalén, P.; Mehtätalo, L.; Suvanto, A. Testing the usability of truncated angle count sample plots as ground truth in airborne laser scanning-based forest inventories. *Forestry* **2007**, *80*, 73–81. [CrossRef]
7. Matérn, B. Wie groß ist die “Relaskopfläche”? *Allg. Forstztg.* **1969**, *79*, 21–22.
8. Hasenauer, H.; Eastaugh, C.S. Assessing forest production using terrestrial monitoring data. *Int. J. For. Res.* **2012**, *2012*. [CrossRef]
9. Maack, J.; Lingenfelder, M.; Weinacker, H.; Koch, B. Modelling the standing timber volume of Baden-Württemberg—A large-scale approach using a fusion of landsat, airborne lidar and national forest inventory data. *Int. J. Appl. Earth Obs. Geoinform.* **2016**, *49*, 107–116. [CrossRef]
10. Hollaus, M.; Wagner, W.; Maier, B.; Schadauer, K. Airborne laser scanning of forest stem volume in a mountainous environment. *Sensors* **2007**, *7*, 1559–1577. [CrossRef]
11. Hollaus, M.; Dorigo, W.; Wagner, W.; Schadauer, K.; Höfle, B.; Maier, B. Operational wide-area stem volume estimation based on airborne laser scanning and national forest inventory data. *Int. J. Remote Sens.* **2009**, *30*, 5159–5175. [CrossRef]
12. Scrinzi, G.; Clementel, F.; Floris, A. Angle count sampling reliability as ground truth for area-based lidar applications in forest inventories. *Can. J. For. Res.* **2015**, *45*, 506–514. [CrossRef]
13. Immitzer, M.; Stepper, C.; Böck, S.; Straub, C.; Atzberger, C. Use of worldview-2 stereo imagery and national forest inventory data for wall-to-wall mapping of growing stock. *For. Ecol. Manag.* **2016**, *359*, 232–246. [CrossRef]
14. GEOsat Mxbox. Available online: <http://www.geosat.de/en/products/system-geometer-mx/geobox.html> (accessed on 29 May 2017).
15. LGL Das digitale Geländemodell (DGM) von Baden-Württemberg. Available online: [http://www.lv-bw.de/lvshop2/start\\_ns.asp?openkey=&keyinfo=dgm&os=Win32&mapw=400](http://www.lv-bw.de/lvshop2/start_ns.asp?openkey=&keyinfo=dgm&os=Win32&mapw=400) (accessed on 10 March 2017).
16. Wenzel, K.; Rothermel, M.; Haala, N.; Fritsch, D. SURE—the ifp software for dense image matching. *Photogramm. Week* **2013**, *27*, 59–70.
17. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [CrossRef] [PubMed]
18. Rothermel, M.; Wenzel, K.; Fritsch, D.; Haala, N. SURE: Photogrammetric Surface Reconstruction from Imagery. Available online: [http://www.ifp.uni-stuttgart.de/publications/2012/Rothermel\\_etal\\_lc3d.pdf](http://www.ifp.uni-stuttgart.de/publications/2012/Rothermel_etal_lc3d.pdf) (accessed on 4 July 2017).
19. Rapidlasso GmbH | Fast Tools to Catch Reality. Available online: <https://rapidlasso.com/> (accessed on 4 July 2017).
20. Isenburg, M.; Liu, Y.; Shewchuk, J.; Snoeyink, J.; Thirion, T. Generating raster dem from mass points via tin streaming. In *International Conference on Geographic Information Science*; Springer: Berlin, Germany, 2006; pp. 186–198.
21. The-R-Foundation R: The r project for statistical computing. Available online: <https://www.r-project.org/> (accessed on 23 May 2017).
22. Sumida, A.; Miaura, T.; Torii, H. Relationships of tree height and diameter at breast height revisited: Analyses of stem growth using 20-year data of an even-aged chamaecyparis obtusa stand. *Tree Physiol.* **2013**, *33*, 106–118. [CrossRef] [PubMed]
23. Prieditis, G.; Mits, I.; Arhipova, I. Allometric models for predicting tree diameter at breast height. *Energy Syst. Sustain.* **2012**, *4*, 105–110.

24. Koenker, R.; Ng, P.; Portnoy, S. Quantile smoothing splines. *Biometrika* **1994**, *81*, 673–680. [[CrossRef](#)]
25. Webster, R.; Oliver, M.A. *Geostatistics for Environmental Scientists*, 2nd ed.; John Wiley & Sons: Chichester, UK, 2007.
26. Stepper, C.; Straub, C.; Immitzter, M.; Pretzsch, H. Using canopy heights from digital aerial photogrammetry to enable spatial transfer of forest attribute models: A case study in central Europe. *Scand. J. For. Res.* **2016**. [[CrossRef](#)]
27. Véga, C.; Renaud, J.-P.; Durrieu, S.; Bouvier, M. On the interest of penetration depth, canopy area and volume metrics to improve lidar-based models of forest parameters. *Remote Sens. Environ.* **2016**, *175*, 32–42. [[CrossRef](#)]
28. Stepper, C.; Straub, C.; Pretzsch, H. Using semi-global matching point clouds to estimate growing stock at the plot and stand levels: Application for a broadleaf-dominated forest in central Europe. *Can. J. For. Res.* **2014**, *45*, 111–123. [[CrossRef](#)]
29. Ritter, T.; Nothdurft, A.; Saborowski, J. Correcting the nondetection bias of angle count sampling. *Can. J. For. Res.* **2013**, *43*, 344–354. [[CrossRef](#)]
30. Seidel, D.; Ammer, C. Efficient measurements of basal area in short rotation forests based on terrestrial laser scanning under special consideration of shadowing. *Iforest-Biogeosciences For.* **2014**, *7*, 227. [[CrossRef](#)]
31. Movic, A.; Beinat, A.; Sandri, T. Land use classification from vhr aerial images using invariant colour components and texture. *ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B7*, 311–317. [[CrossRef](#)]
32. Korpela, I.; Heikkilä, V.; Honkavaara, E.; Rohrbach, F.; Tokola, T. Variation and directional anisotropy of reflectance at the crown scale—Implications for tree species classification in digital aerial images. *Remote Sens. Environ.* **2011**, *115*, 2062–2074. [[CrossRef](#)]



© 2017 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/>).