Annals of Forest Science (2019) 76: 33
https://doi.org/10.1007/s13595-019-0804-4

RESEARCH PAPER

Potential of remote sensing-based forest attribute models for harmonising large-scale forest inventories on regional level: a case study in Southwest Germany

Melanie Kirchhoefer[1] · Johannes Schumacher[1] · Petra Adler[1]

Received: 31 January 2018 / Accepted: 24 January 2019 / Published online: 2 April 2019

Abstract

• ***Key message*** **Image-based 3D information can provide metrics for forest attribute modelling that are robust within the study region. This enables transferability of predictive models to other data sets of the same region without loss of accuracy. As a result, aerial images can potentially provide auxiliary data for supporting different large-scale forest inventories within a geographic region.**

• ***Context*** Due to their high spatial coverage and acquisition frequencies, aerial imagery can provide auxiliary data to support large-scale forest inventories. If the methods are applicable to different data sets and forest inventory protocols, they will also facilitate harmonisation of forest inventory data.

• ***Aims*** This study aims to investigate the level of transferability of such image-based methods. This is crucial for their applicability across different large-scale forest inventories. The investigation focusses on one geographic region.

• ***Methods*** Three blocks of aerial images were used to generate models of forest canopies and extract 3D metrics. These were utilised for building timber volume models separately for each image block. The models were applied to the respective other blocks and the achieved accuracy of timber volume prediction was assessed. Additionally, 3D metric changes between blocks were also assessed.

• ***Results*** Some metrics were found more robust than others. Transferring models based on robust metrics achieved RMSE% between 38 and 45%, which is similar to the model accuracy.

• ***Conclusion*** This indicates transferability of models within the study region without loss of accuracy, and there is potential for further improvement of model accuracy. Therefore, forest attribute models based on remote sensing have potential to support harmonisation of large-scale forest inventories within the study region.

Handling Editor: Tuula Packalen and Klemens Schadauer

This article is part of the Topical collection on *Forest information for bioeconomy outlooks at European level*

Contributions of the co-authors Conceptualization: Melanie Kirchhoefer, Petra Adler
Methodology: Melanie Kirchhoefer, Johannes Schumacher, Petra Adler
Software: Melanie Kirchhoefer; Validation: Melanie Kirchhoefer, Johannes Schumacher
Formal analysis: Melanie Kirchhoefer, Joahnnes Schumacher
Investigation: Melanie Kirchhoefer, Johannes Schumacher
Resources: Petra Adler
Data curation: Melanie Kirchhoefer, Johannes Schumacher
Writing—original draft: Melanie Kirchhoefer
Writing—review and editing: Melanie Kirchhoefer, Johannes Schumacher
Visualisation: Melanie Kirchhoefer, Johannes Schumacher
Supervision: Petra Adler
Project administration: Petra Adler
Funding acquisition: Petra Adler

✉ Melanie Kirchhoefer
melanie.kirchhoefer@forst.bwl.de

[1] Department of Biometry and Informatics, Forest Research Institute (FVA) Baden-Württemberg, Freiburg i. Br, Germany

Keywords Aerial images · Canopy height model · Forest inventory · Timber volume · Model transferability

1 Introduction

Large-scale forest inventories are an important source of information for forest management and policy making. However, comparison and/or combination of forest inventory results that are based on different protocols requires suitable estimation procedures that account for inventories specificities and support harmonisation of target attributes at plot level (McRoberts et al. 2012). In recent years, increased efforts were made to develop approaches to forest inventory harmonisation (McRoberts et al. 2010). Within the European program "Cooperation in Science and Technology" (COST), actions E43 (COST Action E43 n.d., McRoberts et al. 2009; Tomppo and Schadauer 2012) and FP1001 (COST Action FP1001 2014) were concerned with harmonisation of National Forest Inventories (NFIs) on European level. Action FP1001 also considered the role of remote sensing in NFIs, including improving estimates of forest attributes. The work conducted within these COST Actions is continued in the EU Horizon 2020 project DIABOLO (Distributed, Integrated and Harmonised Forest Information for Bioeconomy Outlooks) (DIABOLO n.d.), which includes research towards the derivation of auxiliary data from aerial images for supporting NFI harmonisation across Europe.

The main benefit of utilising remote sensing methods in forest inventories is their ability to enable acquisition of detailed data over large areas in short-time intervals. This area is a widely researched field and there are several remote sensing technologies that have been the focus of projects concerned with supporting or enhancing forest inventories (White et al. 2016). In the majority of relevant literature, airborne laser scanning is utilised (e.g. McRoberts and Tomppo 2007; Fekety et al. 2015; McRoberts et al. 2015, Moser et al. 2016, Véga et al. 2016). Despite advantageous characteristics (White et al. 2016), laser scanning can currently not be considered suitable for harmonisation efforts at all scales, due to limited data availability. Digital aerial photogrammetry data on the other hand is commonly available with high spatial coverage and acquisition frequencies, aiding development of a universal approach for remote sensing data integration in large-scale forest inventory workflows.

Advances in dense image matching allow derivation of photogrammetric 3D point clouds from aerial imagery and, subsequently, provision of canopy height models (CHM) at very high spatial resolution (generally < 1 m), when a high-quality digital terrain model (DTM) is available. Much research has been conducted into the usability of aerial imagery for forest inventories and wall-to-wall forest attribute mapping (e.g. Ginzler and Hobi 2015; Rahlf 2017; Stepper et al. 2015; Straub and Stepper 2016). These approaches generally build relationships between forest inventory data and image-based data by fitting statistical models that are subsequently used for estimating forest attributes for areas where no inventory data is available. A majority of forest inventory attributes is depending on canopy height. When image-based data can represent the heights to a reasonable degree, as indicated by Leberl et al. (2010), comparable results between laser scanning and aerial imagery can be expected (White et al. 2016). In comparative research studies (e.g. Straub et al. 2013; Ullah et al. 2017; White et al. 2015), image-based data achieved accuracy levels slightly below laser scanning, but was considered practicable for forest management and planning to a certain degree.

Studies in this area of research are usually focussed on one study site where an image-based forest attribute model is fitted and validated on the same image data set (e.g. Rahlf et al. 2015; Stepper et al. 2015; Ullah et al. 2017). Aerial image data acquisition conditions (i.e. camera specifics, image overlap, illumination and geographical context) are known to influence image data quality and subsequently the quality of derived canopy height models (Haala 2014; Leberl et al. 2010; Remondino et al. 2014). Despite this, there are no standards or best practice guidelines concerning aerial image quality for 3D point cloud generation, especially for forestry applications (White et al. 2015). Image quality can vary between countries and even between aerial image data acquisition blocks within one country and in some instances only imagery of suboptimal quality might be available. This leads to questions concerning the robustness of extracted metrics and the transferability of predictive models, ultimately influencing the practicability of remote sensing data integration in forest inventory workflows. There are few research studies in the literature concerning the transferability of models, which was also noted in Stepper et al. (2017). Most of these studies used airborne laser scanning (e.g. Fekety et al. 2015; Zald et al. 2016). Stepper et al. (2017) used image-based canopy height models and investigated the transferability of predictive models to nearby forest areas aiming at forest attribute prediction for areas where no forest inventory data for reference is available. They tested three different Random Forest models (predicting the quadratic mean diameter of the 100 largest trees, the basal area-weighted mean height of the 100 trees per hectare with the largest diameters and the timber volume) and found that they could be transferred to nearby stands of similar structure without loss of accuracy as long as a sufficient amount of training data is used. Their test sites were covered by several blocks of aerial images that were all acquired using the same aerial camera (UltraCam-Xp), reducing the variation in aerial image data acquisition conditions.

The presented study is based on a data set from Southwest Germany and was conducted within the framework of the Horizon 2020 DIABOLO project. It focussed on investigating the level of transferability of timber volume models utilising

image-based canopy height model data. Here, transferability means applying predictive models that were fitted on one canopy height model data set to other canopy height model data sets for timber volume estimation. These canopy height model data sets are spatially different (i.e. not in the same location and not of the same extent) and are based on imagery acquired under different conditions. Within this scope, this study aims to answer the following research questions:

1. Are image-based 3D metrics robust with regard to data acquisition conditions?
2. What is the magnitude of difference when models are based on differing aerial image data?
3. Can image-based timber volume models be transferred without loss of accuracy?

Answering these questions will provide important information towards devising imagery-supported methods for large-scale forest inventory harmonisation.

2 Material and methods

2.1 Study site

The study site is located in the south-western part of the federal state of Baden-Württemberg, Germany, including the southern parts of the Upper Rhine Plain and the Black Forest (Fig. 1). It comprises an area of about 3860 km^2 with elevation ranging from 190 m in the Rhine plain to 1493 m (Feldberg) in the Black Forest. The terrain conditions range from flat to slopes of up to 61°. The forested areas within the study site—covering approximately 1860 km^2—vary greatly in terms of species composition and structure. The main tree species are *Picea abies* (L.) H. Karst. (48%), *Fagus sylvatica* L. (18%), *Abies alba* Mill. (12%), *Pseudotsuga menziesii* (Mirb.), Franco (5%), *Quercus* spec. (3%), *Fraxinus excelsior* L. (3%), *Acer* spec. (3%) and *Pinus sylvestris* L. (3%)). There are pure as well as mixed stands with regard to tree species and age class.

2.2 Remote sensing data

Aerial image data was acquired in several flight blocks as part of the regular aerial surveys of the Baden-Württemberg land surveying authority (LGL—Landesamt für Geoinformation und Landentwicklung) using large-sized digital aerial frame cameras. Every year, one third of the state area is covered in this way. According to LGL, the orientation accuracy of the image block is 0.04 m horizontally and 0.14 m vertically. Aerial image data with acquisition dates closest to the date of the latest German NFI (2011–2012) were acquired in three aerial surveys conducted in June 2012 (block A), June 2013 (block B) and July 2013 (block C) (Fig. 1). Image data was acquired using two different cameras, Vexcel's UltraCam-Xp and UltraCam Eagle (Vexcel n.d.). Data acquisition and camera specifications can be found in Table 1. The flight blocks of these aerial surveys overlap and some areas are covered by two or even three aerial surveys.

A 3D point cloud was derived from each aerial image block in an image matching process using the software SURE (nFrames 2018). These point clouds were further processed, and three canopy height models of a 1-m geometric resolution were produced applying the process described in Kirchhoefer et al. (2017).

A high-quality laser scanning-based terrain model with a 1-m resolution covering the study area was also available and was provided by LGL. The nominal height accuracy of the terrain model reported by LGL is 0.5 m or better (LGL n.d.). This terrain model was used to filter the image matching point clouds and to derive canopy height models.

2.3 Field inventory data

In this study, field data of the latest German NFI (2011–2012) (Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz n.d.) was used as reference data. The German NFI is conducted in a 10-year circle on permanent plots. The plots are organised in square 150 × 150 m inventory tracts with one plot at each tract corner. The tracts are located in the cross sections of a 4 × 4 km grid, which in Baden-Württemberg is densified to 2 × 2 km. The plots were located by the field crews using the MxBox Global Navigation Satellite System (GNSS) device of GEOSat GmbH (GEOsat n.d.-a). There is no data on the positional accuracy of the NFI plots, as this was not recorded when the plots were established. According to the specifications of the MxBox GNSS device, the nominal horizontal accuracy is 1.5 m or better (GEOSat n.d.-b). The height accuracy of the device is not specified, but it can be expected to be below the horizontal accuracy (Sass 2011). However, environmental and operating conditions (e.g. canopy cover, number and geometry of satellites, distance to reference stations) influence the effective positioning accuracy (GEOSat n.d.-b).

At each plot, sample trees are selected using angle-count sampling (ACS) with a basal area factor (BAF) of 4 m^2/ha and a minimum diameter at breast height (DBH) of 7 cm. For each sample, tree attributes such as the DBH, azimuth and distance to the plot centre and the tree species are recorded. Heights are measured for a subset of sample trees. The heights of the remaining trees are estimated using height-diameter curves based on Sloboda et al. (1993) (Riedel et al. 2017). For each tree, its representative volume per hectare was calculated. The aggregated volumes at each plot were used as reference data for timber volume modelling in this study.

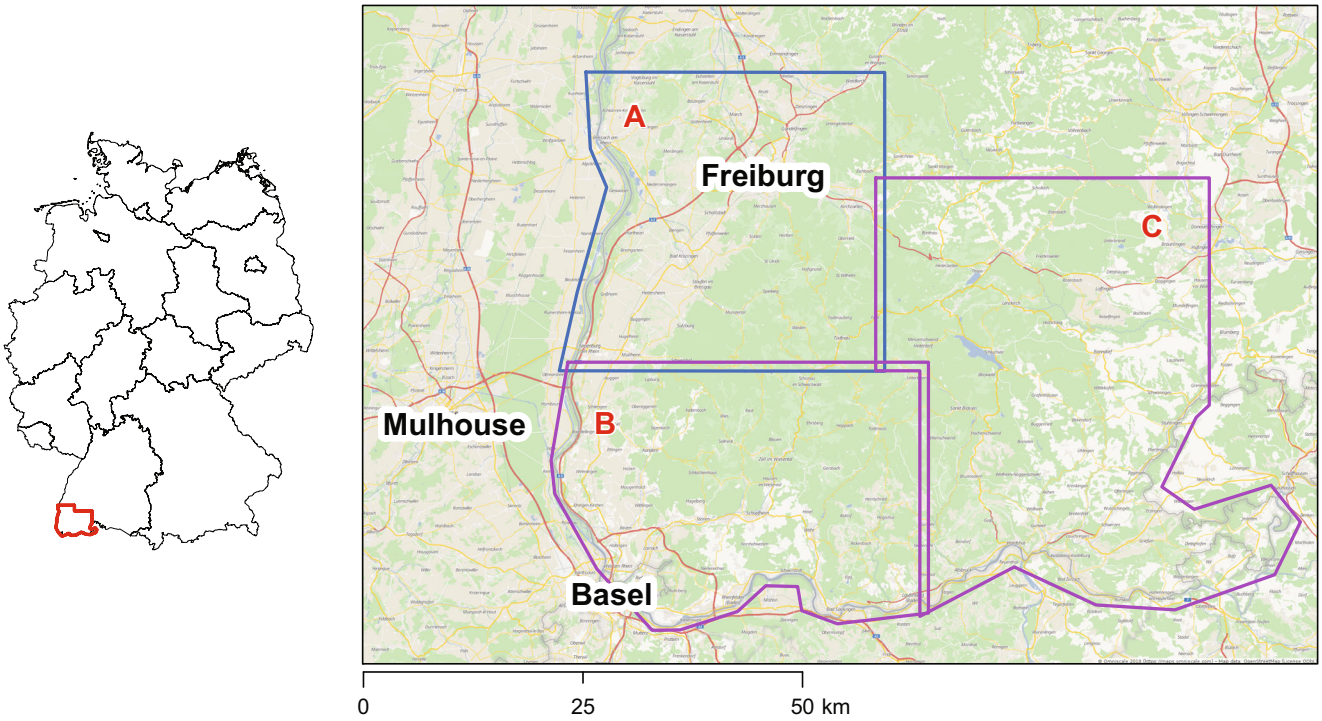

Fig. 1 Study site including block boundaries for aerial image data acquisition (background image © OpenStreetMap contributors)

A more detailed description of the German NFI can be found in Tomppo et al. (2010) and Riedel et al. (2017).

2.4 Metric extraction

In this study, only NFI plots located within a forested area were utilised, i.e. plots where observations as described in Section 2.3 are available. For extracting metrics from the canopy height model, a circular shape around each plot centre location was used. Determining a suitable radius for this circle is not straight forward, when the reference data is based on ACS. Different approaches (e.g. Deo et al. 2016; Immitzer et al. 2016; Kirchhoefer et al. 2017; Maack et al. 2015; Scrinzi et al. 2015) have been tested in order to find an optimal geometry. In this study, the median of the maximum measured distance between each NFI plot centre and corresponding sample trees was used as plot radius, resulting in canopy height model plot radii of 11.31 m (block A), 11.70 m (block B) and 10.75 m (block C), respectively. This approach achieved acceptable results in Kirchhoefer et al. (2017).

On these circular plots, only pixels with an elevation of 6 m and more were extracted from the canopy height model (Kirchhoefer et al. 2017) and used for 3D metrics calculation, including the following height statistics: minimum, maximum, mean, standard deviation, variance, coefficient of variation of height, percentiles (5, 10, 15, 20, 25, 50, 75, 80, 85, 90, 95, 99) as well as metrics related to canopy volume (inner volume, outer volume, Fig. 2) and the number of pixels used for metrics calculation (npix). These metrics are frequently found in publications concerned with predicting volume or biomass using inventory and remote sensing data (e.g. Stepper et al. 2015; Straub and Stepper 2016; Véga et al. 2016).

Table 1 Data acquisition and camera specifications

	Block A	Block B	Block C
Acquisition date	June 2012	June 2013	July 2013
GSD (cm)	20	20	20
Nominal overlap	60%/30%	60%/30%	60%/30%
Spectral resolution	R, G, B, NIR	R, G, B, NIR	R, G, B, NIR
Camera	UltraCam-Xp	UltraCam Eagle	UltraCam-Xp
Focal length (mm)	100.5	79.8	100.5
Sensor size (mm)	67.860 × 103.860	68.016 × 104.052	67.860 × 103.860
Pixel size (μm)	6.0 × 6.0	5.2 × 5.2	6.0 × 6.0

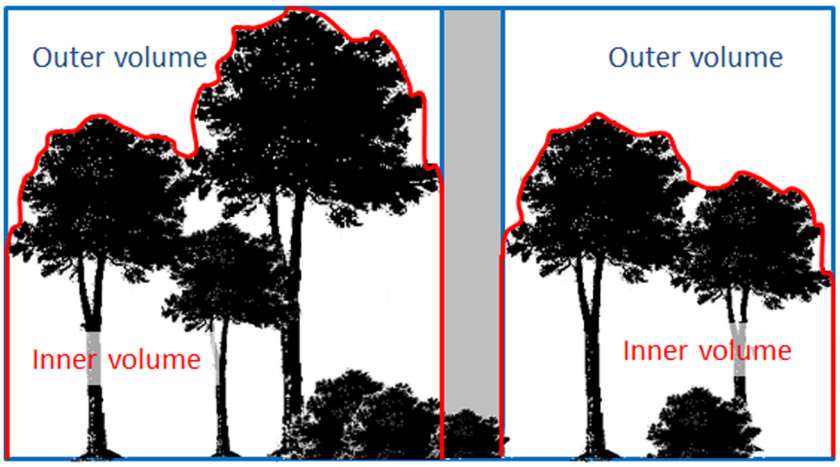

Fig. 2 Schematic description of Volin and Volout (from Kirchhoefer et al. 2017). The grey area, which represents space over vegetation of less than 6 m height, is excluded from any volume calculation

The inner volume (Volin) describes the 3D space that is probably occupied by trees, i.e. the space between the canopy height model and the height threshold of 6 m. It is calculated by Eq. (1):

$$\mathrm{Volin} = \sum \mathrm{CHM_{ep}} \times \mathrm{res}^2, \quad (1)$$

where $\mathrm{CHM_{ep}}$ is the height of the extracted pixels and res is the spatial resolution of the canopy height model.

The outer volume (Volout) corresponds to the complement of the inner volume with respect to the bounding volume of the canopy height model. Along with the inner volume, it provides information about the complexity of the outer canopy structure. It is calculated by Eqs. (2) and (3):

$$\mathrm{maxVol_{CHM}} = \max\left(H_{ep}\right) \times \mathrm{res}^2 \times \mathrm{count(ep)} \quad (2)$$

$$\mathrm{Volout} = \mathrm{maxVol_{CHM}} - \mathrm{Volin}, \quad (3)$$

where $\mathrm{maxVol_{CHM}}$ is the volume defined by the maximum height of the canopy height model on the plot, H_{ep} are the heights of the extracted pixels, res is the spatial resolution and count(ep) the number of pixels extracted at the plot.

The terrain in the study site varied considerably (Section 2.1) and elevation influences forest structure as well as growth dynamics of trees. So, the mean elevation was included as a site-related metric. It was extracted from the digital terrain model described in Section 2.2 using a 30-m fixed radius circular plot.

2.5 Outlier removal

When combining terrestrial measurements with aerial image data, several factors influence data quality and combination possibilities. These include spatial co-registration, configuration of the sample plots (Deo et al. 2016), border effects and time lags between terrestrial measurements and remote sensing data (i.e. effect of thinning or harvesting activities). All these effects could be considered outliers and they have to be identified and removed from the data set before establishing statistical models for forest parameter estimation (Webster and Oliver 2007).

In this study, outlier plots were defined as plots with the maximum canopy height model elevation deviating by more than 25% from the maximum measured NFI tree height at the respective plot (Eq. (4)).

$$\mathrm{devH_{max}} = \left| 100 - \frac{100}{\max(H_{\mathrm{NFI}}) \times \max(H_{\mathrm{CHM}})} \right|, \quad (4)$$

where $\mathrm{devH_{max}}$ is the percentage of difference, $\max(H_{\mathrm{NFI}})$ is the maximum measured height on the NFI plot and $\max(H_{\mathrm{CHM}})$ is the maximum canopy height model height on this plot. The threshold of 25% was chosen based on results in a project conducted at the Forest Research Institute Baden-Württemberg that was concerned with supporting forest management inventories (not published).

In the outlier removal workflow, an individual radius was calculated for each plot. This was due to sample trees being selected using ACS, resulting in plots of undefined size and shape. The individual plot radius was calculated by the maximum measured distance between any tree of a plot and the plot centre plus a 2-m buffer, allowing for slightly inaccurate positioning of plot centres as well as for non-verticality of trees.

2.6 Volume model fitting

Finding the best suited model is not the focus of this study, but modelling timber volume provided means to investigate the robustness of metrics and transferability of models. For simplicity reasons, only linear models are considered in this investigation.

The metrics extracted are highly correlated. In order to reduce the number of metrics for modelling and to avoid co-linearity of the metrics, a correlation analysis was

conducted. Metrics with a correlation coefficient (Pearson) of minimum 80% were examined and the metric with the highest correlation with timber volume was kept. The correlation analysis was conducted separately for each aerial image block, and the results were combined in one initial set of metrics. This was used for fitting one timber volume model for each aerial image block separately. Model fitting started with an initial maximum model that was stepwise reduced using Akaike information criterion (AIC). Besides the initial set of metrics, interactions between metrics, squares of metrics, and weights were also considered in this process. Selection of the best suited model was based on R^2 and root mean square error percentage (RMSE%) (Eq. (5)) calculated in a 10-fold cross-validation.

$$\mathrm{RMSE\%} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{obs_i}-\mathrm{pred}_i)^2}}{\frac{1}{n}\sum_{i=1}^{n}\mathrm{obs}_i} \times 100, \quad (5)$$

where n is the number of plots with observations (i.e. NFI plots within forests), obs_i is the observation at plot i and pred_i is the prediction for plot i.

2.7 Assessment of metric robustness and model transferability

The assessment of metric robustness utilises the overlap of image blocks. In the overlap areas AB, AC and BC (Fig. 3) more than one set of extracted metrics is available for each NFI plot. Assessing robustness by investigating metrics differences is only feasible in overlap area BC, as the time span between image acquisition dates in the other overlap areas are considered too long (≥ 1 year, cf. Table 1).The effect of tree growth will significantly impact on the metrics and does not allow a direct comparison. In order to exclude this effect on the results of this study, metric comparisons were conducted in overlap area BC only.

In this investigation, the focus was on a subset of canopy height model metrics: the mean height of the canopy height model (meanCHM), npix, Volin, Volout, the 25th (p25) and the 95th (p95) height percentile. Differences in other metrics are expected to be of similar magnitude. The metric differences between blocks were calculated for each overlap area plot by subtracting the metrics derived in block B from metrics derived in block C. The investigation was conducted using scatterplots of metric differences as well as the normalised median absolute deviation (NMAD) (Höhle and Höhle 2009). The NMAD is proportional to the median of the absolute differences between single deviations and the median deviation and is calculated by Eq. (6):

$$\mathrm{NMAD} = 1.4826 \times \mathrm{median}\left(\left|\Delta\mathrm{metric}_j - m\Delta\mathrm{metric}\right|\right), \quad (6)$$

where 1.4826 is a constant scale factor for normally distributed data, $\Delta\mathrm{metric}_j$ is the individual metric difference at plot $j = 1 \ldots n$, n is the number of (non-outlier, forested) plots within the overlap area and $m\Delta\mathrm{metric}$ is the median of the metric differences.

During the assessment of metric robustness, it was discovered that some plots were identified as outliers in block B but not in block C, and vice versa. That let to further investigations into the cause for this by visually assessing canopy height model differences at these plots.

Assessment of model transferability was conducted by applying the models fitted in Section 2.6 to the canopy height model data of the respective other image blocks and predicting timber volume for each plot. The time span between data acquisition dates is not considered an obstacle when transferring predictive models from one image block to another. From the predicted and observed values, the RMSE% was calculated using Eq. (5).

The timber volume predictions for the NFI plots within overlap area BC were additionally subjected to a more in-depth investigation. The timber volumes predicted using fitted as well as transferred models were compared to each other by creating scatterplots and calculating the NMAD based on Eq. (6).

3 Results

3.1 Comparison of metrics in overlap areas

Figure 4 displays the metric differences in the BC overlap plots for meanCHM, npix, Volin, Volout, p25 and p95 while the NMADs are listed in Table 2. The medians of the metrics extracted from each image block within overlap area BC are also included in Table 2, supporting evaluation of the NMADs. The meanCHM differed up to 3 m with most plots differing by a value between 0 and − 1 m, which is reflected in the NMAD of 0.4 m. Similar results were achieved for p25 and p95. The medians of these metrics are of comparable magnitude and ranged from 19 to 28 m.

Volin and Volout show medians of larger magnitude with 8667.7 and 7660.1 m^3, respectively, for Volin and 3068.1 and 2280.5 m^3, respectively, for Volout. The absolute metric differences are of larger magnitude, too (up to − 2500–−

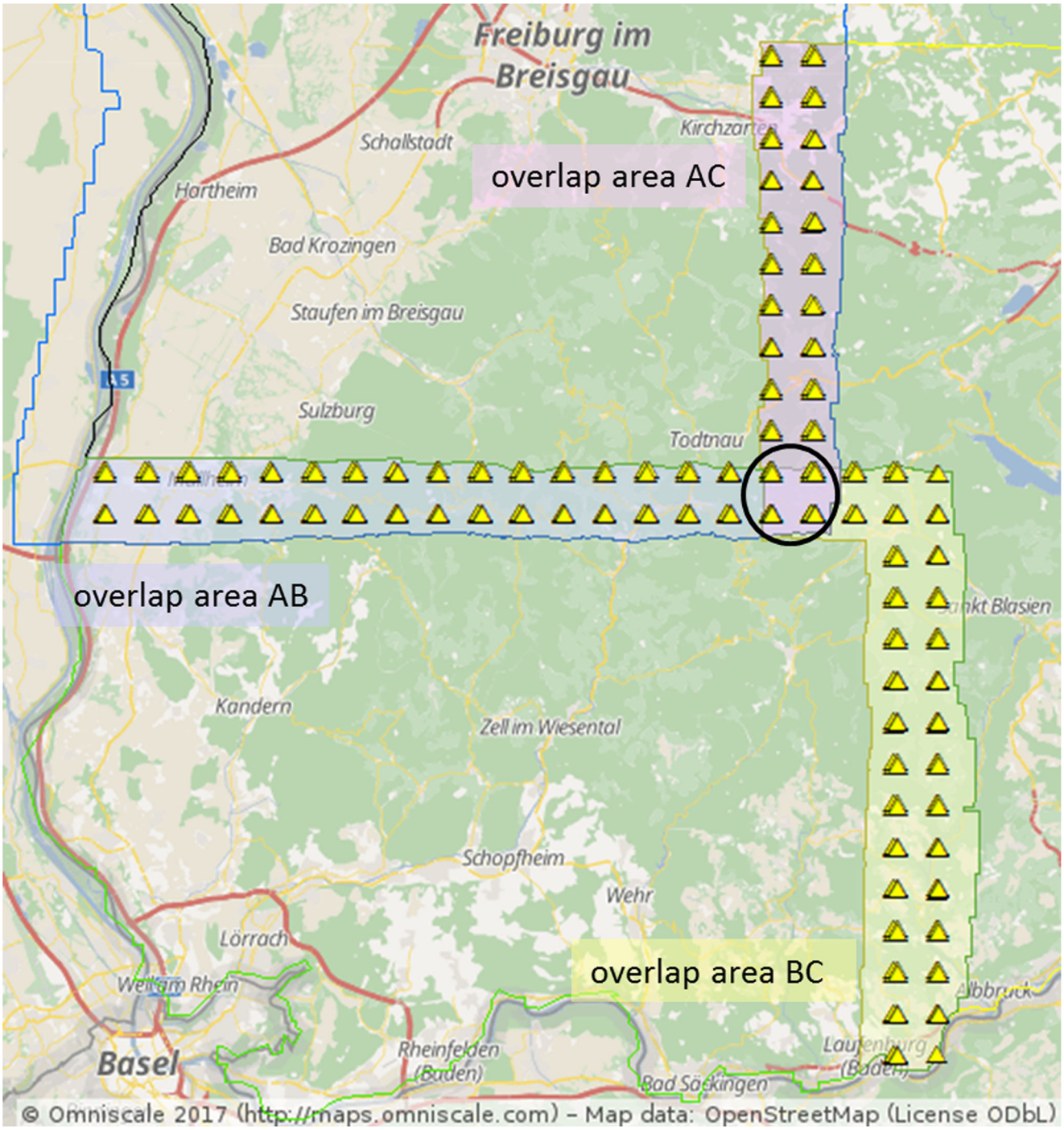

Fig. 3 Overlap areas in the study site. Triangles indicate the location of NFI plots (within and outside forests) (backgroound image © OpenStreetMap contributors)

4500 m^3, respectively) and vary to a larger extent (NMAD of 600.2 and 474.2 m^3, respectively).

For the metric npix, the magnitude of difference is comparatively large. The value derived in block B (median = 420) exceeds the value in block C (median = 356) by around 60 pixels/plot (Fig. 4). However, the NMAD of 7.4 pixels is comparatively small, indicating that the magnitude of differences is rather stable.

3.2 Differences in outlier detection

Within the overlap area BC five plots were identified as outliers in block B only, while one plot was identified as outlier in block C only. These plots were visually examined. One outlier plot in block B was due to an obvious co-registration error (Fig. 5), where the NFI plot centre and one sample tree fell on a forest road. This plot was not identified as an outlier in block C due to

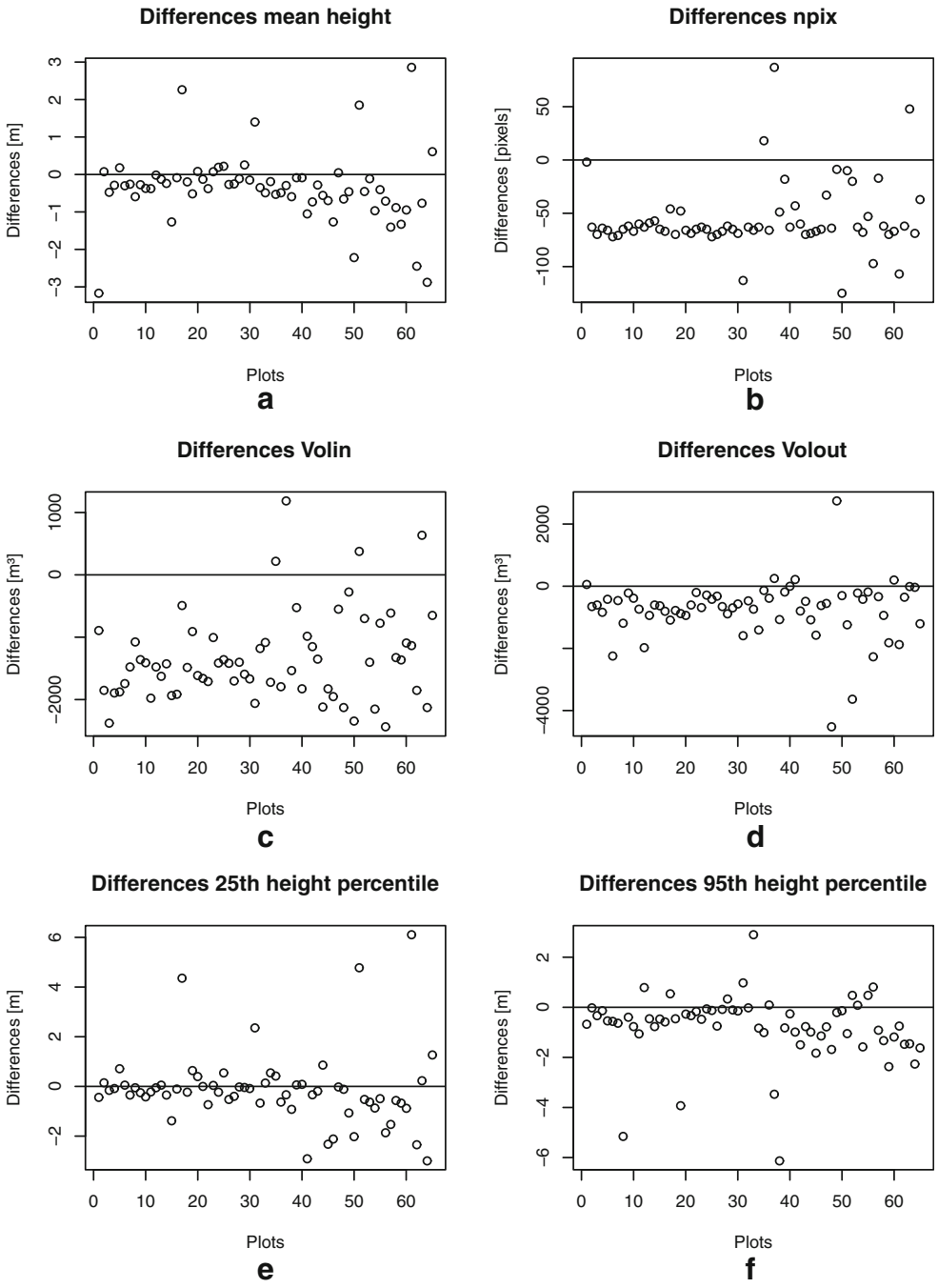

Fig. 4 Scatterplots of metric differences in overlap area BC for meanCHM a, npix (**b**), Volin (**c**), Volout (**d**), p25 (**e**) and p95 (**f**)

the location of one sample tree actually coinciding with a tree in the canopy height model of block C. This was sufficient for passing the automatic outlier removal. The outlier in block C that was not identified in block B can also be explained by an

Table 2 NMADs of metrics in plots of overlap area BC. The medians of the metrics extracted from image block B and block C, respectively, are also displayed, supporting evaluation of the NMADs

	meanCH (m)	npix (pixels)	Volin (m^3)	Volout (m^3)	p25 (m)	p95 (m)
NMAD	0.4	7.4	600.2	474.2	0.7	0.5
Median B	22.7	420	8667.7	3068.1	27.6	19.4
Median C	21.9	356	7660.1	2280.5	27.3	19.6

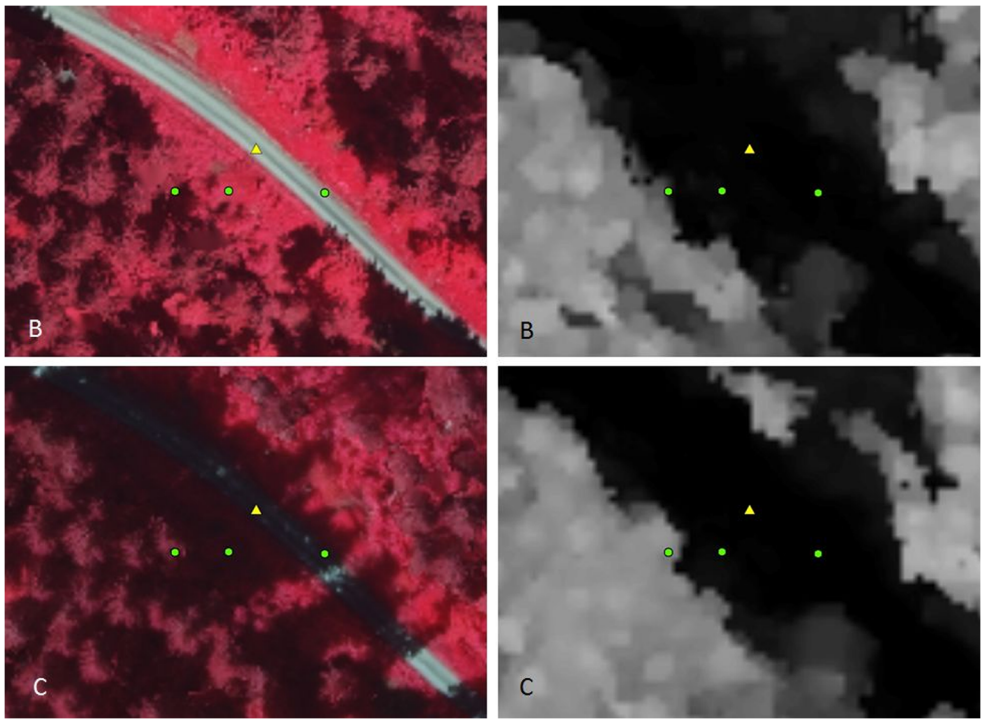

Fig. 5 Example of outlier due to co-registration error; the plot centre (triangle) and sample trees (circles) are superimposed on the true orthophotos (left) of blocks B and C as well the canopy height model (right)

error in co-registration. Here, differences in the canopy height model caused this plot not being identified as outlier in block B.

Errors in the canopy height model could be identified as the other source for the differences in outlier identification between block B and block C. Figure 6 exemplifies one plot where the errors in the canopy height model in block B let to this plot being identified as an outlier, while the canopy height model of block C in the same area was without obvious errors. Similar conditions were found for the two remaining outlier plots in block B that were non-outliers in block C.

3.3 Transfer of models

Table 3 displays the best fitted models for each of the aerial image blocks, including the selected metrics, the weights applied, R^2, standard error, RMSE% and the number of NFI

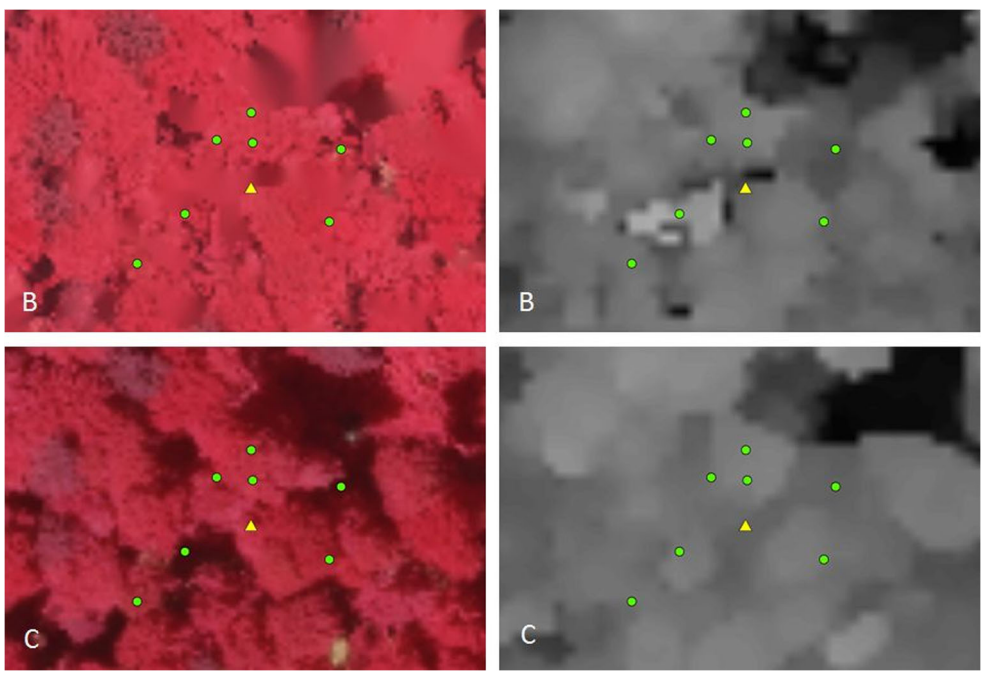

Fig. 6 Example of outlier due to error in canopy height model; the plot centre (triangle) and sample trees (circles) are superimposed on the true orthophotos (left) of blocks B and C as well the canopy height model (right)

Table 3 Best fitted linear models for each aerial image block

	Block A	Block B	Block C
Metrics selected (*p* values in parentheses)	meanCHM (0.10902) npix (0.10902) meanCHM*npix (0.00188) meanDTM (1.61e-13)	meanCHM (0.009459) meanCHM^2 (0.0027) npix (0.000442) meanCHM*npix (1.26e-5) meanDTM (5.01e-6)	Volin (< 2e-16) npix (< 2e-16) meanDTM (4.93e-9)
Weights	1/meanCHM	1/meanCHM	1/meanCHM
R^2	0.55	0.50	0.57
Standard error (m^3/ha)	32.37	39.56	37.88
RMSE%	38.6	42.1	37.9
Number of plots	464	517	765

plots used for model building. The models of blocks A and B are similar. In both cases, the metrics meanCHM, npix, the interaction between meanCHM and npix (meanCHM*npix) and meanDTM were selected. The only difference is the selection of meanCHM^2 in block B. The principal difference between the linear model fitted to block C and the other models is the replacement of meanCHM with Volin. However, they also have the metrics npix and meanDTM in common. The metrics npix and meanDTM were selected in all three models, indicating their importance for volume modelling. Applying a weight (1/meanCHM) was also beneficial in all three blocks.

All models achieved a similar R^2, ranging between 0.5 and 0.6, a standard error between 32 and 39.59 m^3/ha and RMSE% values between 37 and 43%. The scatterplots in Fig. 7 all show signs of overestimating lower timber volumes and underestimating higher timber volumes, but this seems to be less pronounced with the model fitted in block C.

Figure 8 shows the scatterplots of observed versus predicted volumes when transferring models to the respective other image blocks. When the models from block A or block B were used for volume prediction, the resulting scatterplots show an even stronger over- and underestimation effect as the respective scatterplots in Fig. 7. When transferring the model of block C, this effect disappears. The values are more evenly distributed around the 1:1 line, but show larger deviations from this line. This is also reflected in the RMSE% values in Table 4. Transferring the models fitted on block A or B results in RMSE% between 38 and 43%, which is comparable to model accuracy. The accuracy achieved when transferring model C is notably lower with an RMSE% of 47.9% in block A and 51.1% in block B.

3.4 Differences in predictions on plot level

Figure 9 displays the differences in predicted volume at plot level in overlap area BC for different combinations of predictive model and canopy height model data. The differences are smallest (− 99–25 m^3/ha) when model B is used for both blocks (Fig. 9c). Using fitted models only (Fig. 9a), the spread is a bit wider (− 100–110 m^3/ha), with most differences being positive. A similar magnitude of differences (− 135–45 m^3/ha) is achieved when only model C is used (Fig. 9b), but in this case, most differences are negative. When only transferred models were used for timber volume prediction (Fig. 9d), the magnitude of differences increases notably (− 327–10 m^3/ha). The different ranges of volume differences are also reflected in the NMAD calculated for each combination (Table 5). The lowest NMAD (30.7 m^3/ha) was achieved for the case where only model B was used for volume prediction (Fig. 9c) while

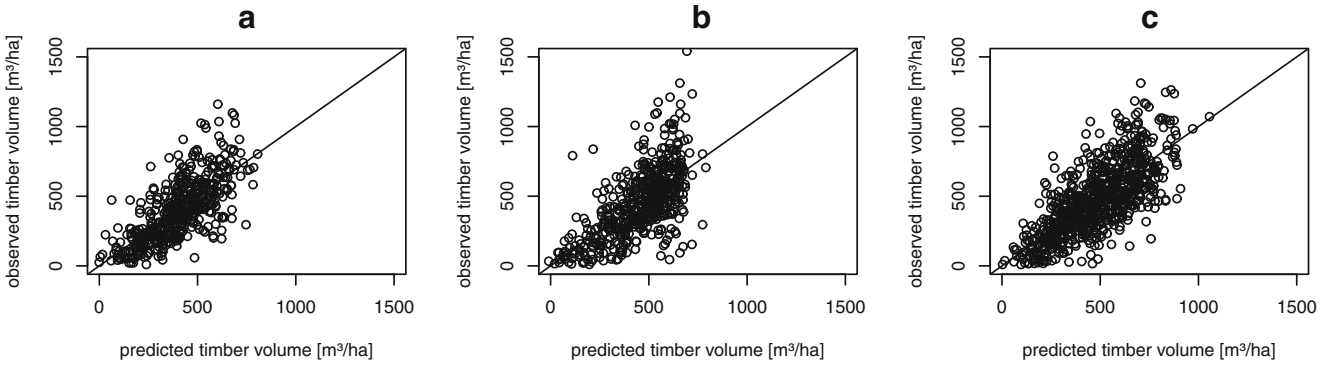

Fig. 7 Scatterplots of best fitted models

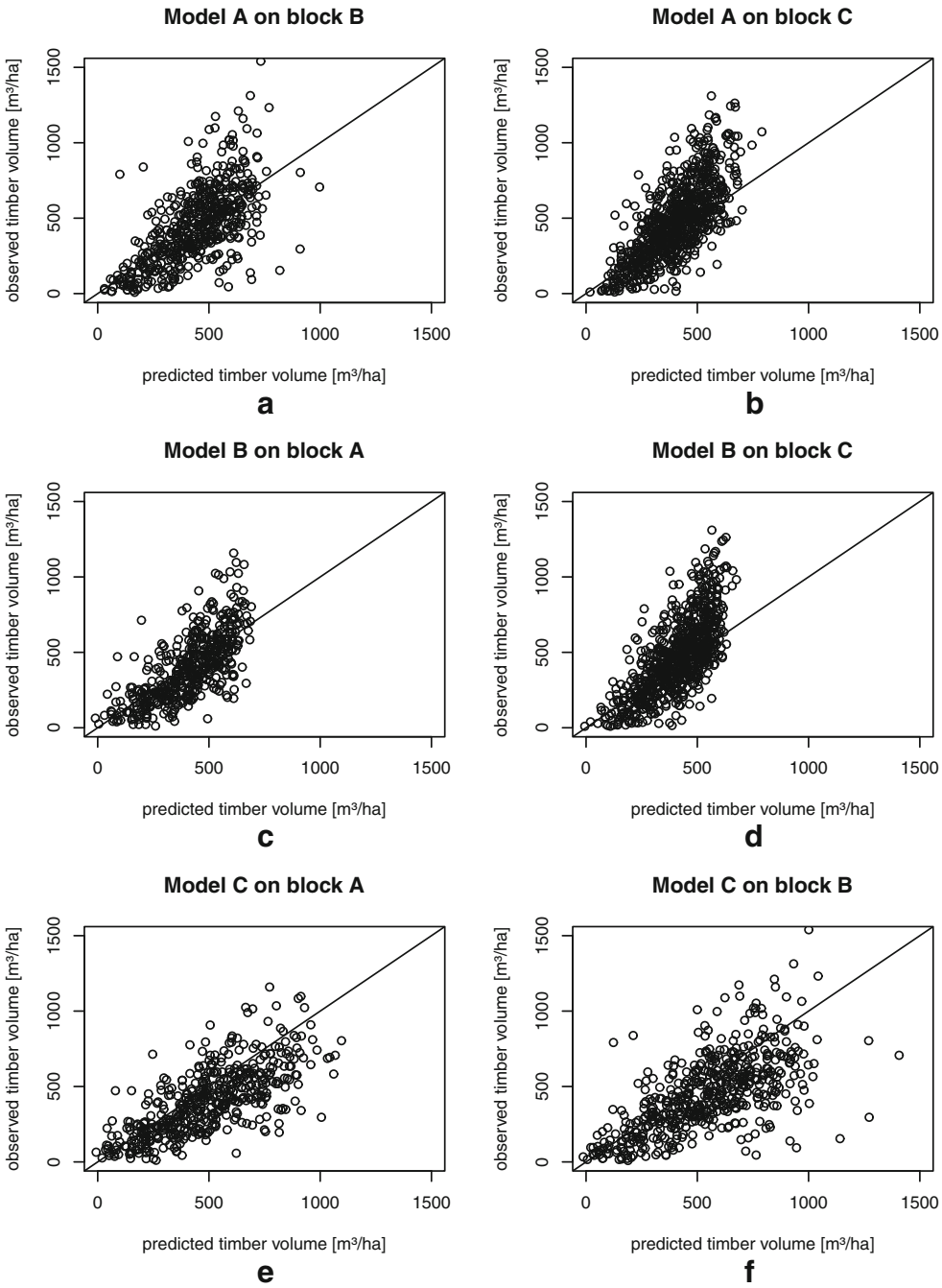

Fig. 8 Scatterplots of observed vs. predicted timber volume for all combinations of transferred models

the largest NMAD (83.5 m^3/ha) was achieved for the case where only transferred models were used (Fig. 9d).

Table 4 RMSE% results for the volume predictions using transferred models

RMSE%		Prediction		
		A	B	C
Models	A		42.5	42.4
	B	38.4		42.5
	C	47.9	51.1	

4 Discussion

4.1 Metric robustness

The results in Section 3.1 show that the metrics related directly to canopy height are sufficiently robust with comparatively small absolute differences and a small NMAD. Compared to these the NMADs of the canopy volume related metrics—Volin and Volout—seem to indicate notably reduced robustness with respect to changing data acquisition conditions. However, Volin and Volout are influenced by both, the canopy

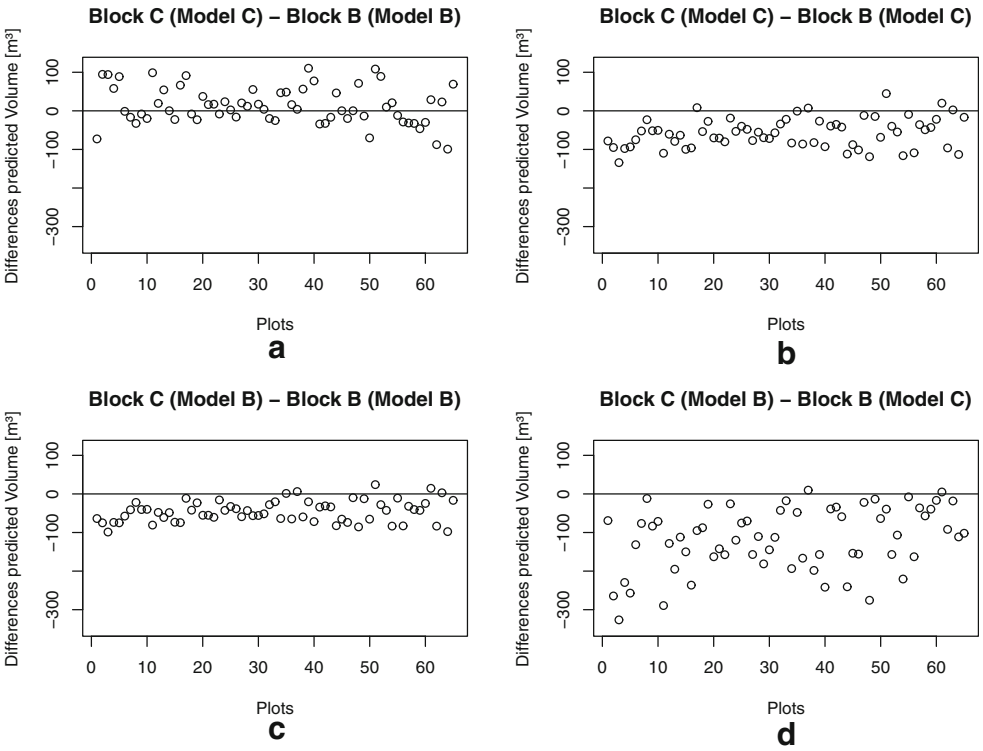

Fig. 9 Differences in predicted volume in overlap area BC for all combinations of models and canopy height models

height and npix. Additionally, the metrics values themselves are of larger magnitude compared to the height-related metrics and npix. These conditions could explain the greater variation and comparatively large differences (Table 2). Relating the NMADs to the respective medians of the extracted metrics (Table 2) normalises the results. The NMADs of the height-related metrics range between 1.7 and 2.6% of their respective medians in blocks B and C. The NMAD of Volin corresponds to 6.9–7.4% of its respective medians while these values increase for Volout to 15.5–20.8%. This indicates that Volin is less robust than the height-related metrics but could still be considered applicable for timber volume modelling. However, based on the presented results, Volout cannot be considered robust with respect to changing data acquisition conditions.

The metric npix showed comparatively large magnitude of differences (about 60 pixels), but the variation in differences is rather stable and corresponds to 1.8–2.1% of the npix median in block B and block C, respectively (Table 2). This indicates some fixed offset, which is caused by differing plot radii for canopy height model metrics extraction. In block B a radius of 11.70 m was used and in block C a radius of 10.75 m, resulting in plot areas of 430.05 and 363.05 m^2, respectively. The area difference is 67 m^2, which is equivalent to 67 pixels of size 1 m^2. Therefore, npix can be considered a robust metric, if plot sizes do not vary.

Table 5 NMAD of volume predictions in overlap area BC using different combinations of predictive models and canopy height models

NMAD (m^3)		Block B	
		Model B	Model C
Block C	Model B	30.7	83.5
	Model C	43.1	41.8

Here, the variation of plot size is due to the approach for determining a plot radius that suits ACS reference data, and this issue could easily be solved by adjusting plot radii. On large scales this could prove more difficult. Geometric consistency between field data and metric extraction plots is crucial for accurate forest attribute modelling (Deo et al. 2016) while plot sizes usually vary between forest inventories. Normalising metrics to a fixed plot size (e.g. plot with radius = 12 m) could provide a solution. The effect of this on the metrics npix, Volin and Volout in overlap are BC can be seen in Fig. 10 and Table 6. The fixed offset for npix disappears and the NMAD decreased slightly from 7.4 to 6.8 pixels (corresponding to 1.5% of the medians in both blocks) (Table 6), showing an increased robustness. The scatter for Volin decreased (Fig. 10b), which is also reflected in the NMAD decreasing from 600.2 (Table 2) to 290.8 m^3 (Table 6). After relating this to the Volin metric medians (3.2 and 3.0%, respectively; based on values in Table 6), a notable increase in robustness towards a level comparable to the height-related metrics and npix is indicated. The NMAD for Volout, on the other hand, increased from 474.2 (Table 2) to 626.3 m^3 (Table 6). Also, in relation to the Volout medians, the NMAD did increase with values corresponding to 19.4% (before 15.5%) of the median in block B and 22.0% (before

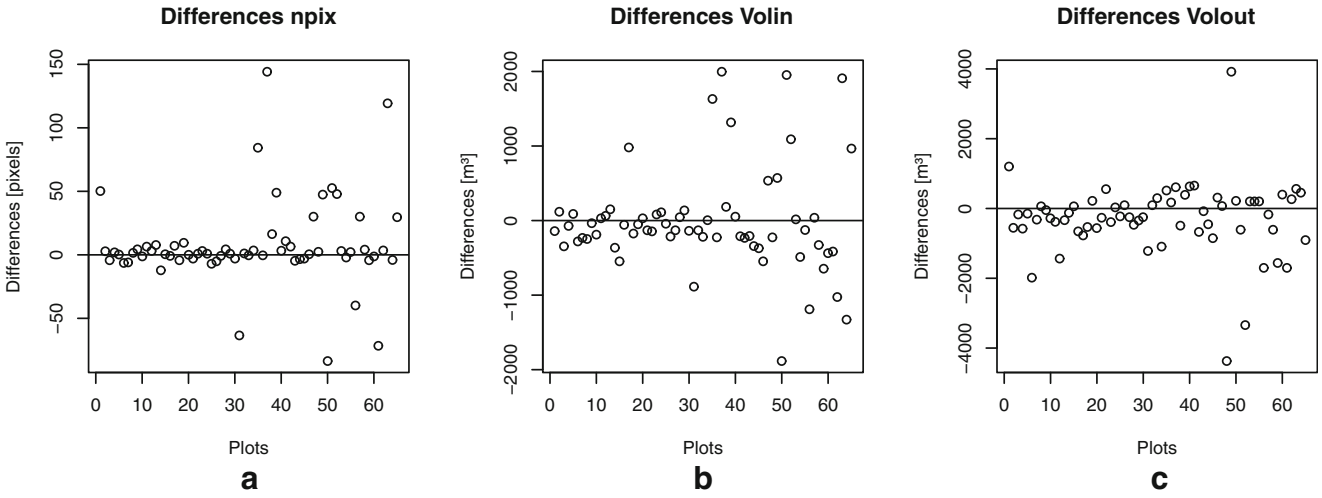

Fig. 10 Scatterplots of metric differences for npix a, Volin (**b**) and Volout (**c**) after normalisation (using a plot radius of 12 m)

20.8%) of the median in block C. Even when normalising Volout by plot size, the robustness cannot be increased to an acceptable level. The height-related metrics (meanCHM, p25, p95) are stable as they already imply a level of normalisation.

The differences between the canopy height models of block B and block C as well as the issue of inconsistent outlier removal could be explained by varying image acquisition conditions. Aerial image data in block B was acquired using UltraCam Eagle digital aerial camera while block C was acquired using UltraCam-Xp (Table 1). The shorter focal length and slightly larger sensor of the UltraCam Eagle resulted in wider aperture angles compared to the UltraCam-Xp (see, Table 1). For the UltraCam Eagle, these are 46.16° along track and 66.20° across track and for the UltraCam-Xp 37.32° along track and 54.66° across track. Wider angles cause a less favourable viewing geometry for image matching, resulting in more occlusions and mismatched points. The comparatively small image overlap (60% endlap, 30% sidelap) provides no means for compensation by redundancy. This affects the accuracy of the canopy height model in general (Leberl et al. 2010), and it is not surprising that all inconsistencies in the outlier removal due to errors in the canopy height model were found in block B. These errors can severely affect model building and forest attribute prediction based on canopy height model data. It is recommended to utilise aerial image data acquired with a narrower aperture angle (e.g. focal length of aerial camera > 100 mm) and to increase the image overlap (e.g. 80% endlap and 50% sidelap) in order to enhance image matching quality and subsequently quality of 3D metrics.

Table 6 NMAD of metrics in overlap area plots after normalisation using a plot radius of 12 m. The medians of the metrics extracted from image block B and block C, respectively, are also displayed, supporting evaluation of the NMADs

	npix (pixels)	Volin (m^3)	Volout (m^3)
NMAD	6.8	290.8	626.3
Median B	441.8	9117.9	3227.5
Median C	443.6	9545.1	2841.7

The outlier plot due to co-registration problems depicted in Fig. 5 is an example for the issues concerning plot positioning. Despite using high-end GNSS-devices (e.g. see, Section 2.3), site conditions might reduce the achievable positional accuracy considerably. This demonstrates the importance of a reliable outlier removal process. However, the results of this study revealed that the currently used method of outlier removal needs to be improved. The obvious outlier due to co-registration errors in Fig. 5 was missed. The current approach comparing maximum heights could be improved by including other indicators, such as a comparison between meanCHM and the mean of the NFI heights.

4.2 Transferability of predictive models

Three types of metrics appear to be of importance for volume modelling: (1) one metric related to either canopy height or volume, (2) one metric related to the canopy structure (e.g. npix) and (3) one site-related metric (e.g. meanDTM). In some cases, the interaction of these metrics is significant. Table 3 displays the *p* values of the selected metrics and their interactions, respectively. The importance of meanDTM is certainly due to the large elevation changes within the study area. In a flat area, elevation could be expected to be insignificant, and other site-related metrics might be more useful.

Despite similar standard errors, R^2 and RMSE% values of the fitted models, models A and B differed from model C, most notably by the selection of Volin as prediction variable in block C. The scatterplots in Fig. 7 also show a mixture of over- and underestimation, which is less pronounced in block C. Also, in block C, the scatter was greater. When transferring the models, this difference

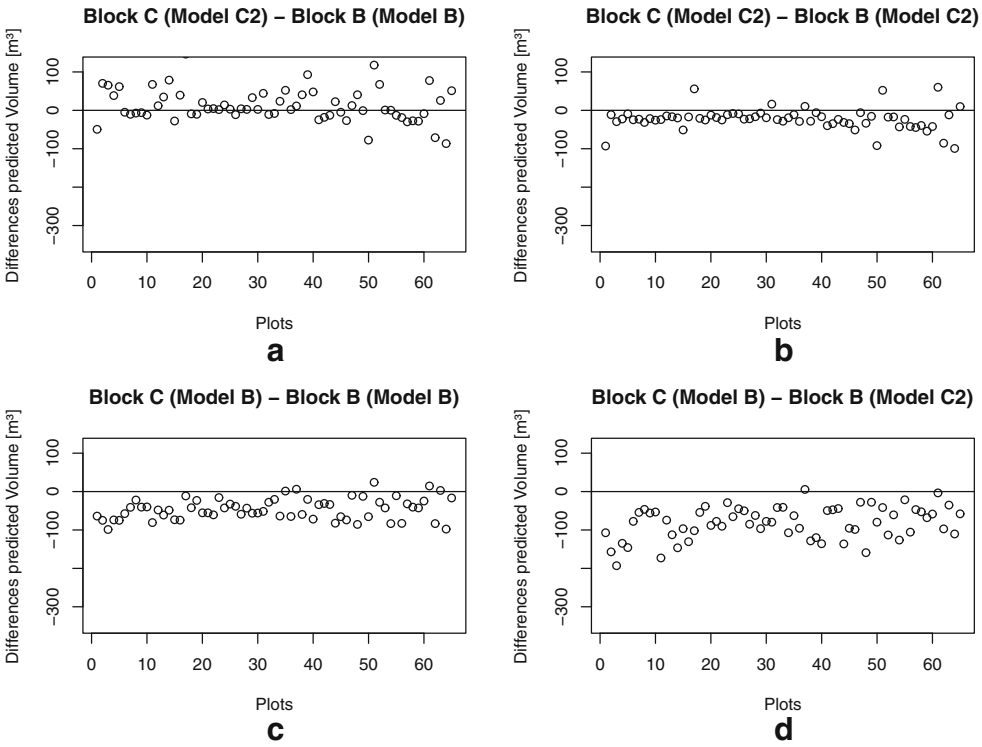

Fig. 11 Using refitted model C2: differences in predicted volume in overlap area BC for all combinations of models and canopy height models

became even more pronounced, with the accuracy achieved with the transferred model C being notably lower (Table 4). This could be explained by the lower robustness of Volin compared to the height-related metrics discussed in Section 4.1, as the models were fitted using metrics without normalisation based on plot size. In the model fitting process, Volin performed only slightly better than meanCHM. When refitting model C using meanCHM instead of Volin (model C2), RMSE% notably improves to 43.4% in block A and 44.9% in block B. The effect of using model C2 instead of model C on the plot level is displayed in Fig. 11 and Table 7. The ranges of differences in predicted volume in Fig. 11b and d were reduced considerably compared to those displayed in Fig. 9b and d, respectively. This is also reflected in reduced NMADs for all variations (Table 7). The differences between the two blocks are smallest when only model C2 is used. Alternatively, refitting model C with normalised Volin and npix (model C3) and transferring this model to the other blocks further improved the RMSE% to 38.6% in block A and 42.7% in block B. For that purpose, normalisation of Volin and npix was also conducted for these blocks. On plot level in overlap area BC, the differences in predictions were reduced compared to those achieved with model C (Fig. 12), but to a smaller magnitude as model C2 (Table 7). However, using model C3 for prediction in block C and model B for prediction in block B resulted in the same NMAD as when using model C. This demonstrates that the improved robustness of Volin due to normalisation by plot size has a positive effect on the transferability of the model in the study region, similar to the effect of using the robust metric meanCHM.

Table 7 NMAD of volume predictions in overlap area BC using the refitted models C2 and C3. (Results using model C in parentheses)

	NMAD (m^3)	Block B	
		Model B	Refitted model
Block C	Model B	–	C2: 44.6 C3: 51.0 (C: 83.5)
	Refitted model	C2: 31.6 C3: 43.1 (C: 43.1)	C2: 15.0 C3: 17.7 (C: 41.8)

The RMSE% achieved when transferring models (Table 4) based on robust metrics (models A, B, C2 and C3) are comparable to the RMSE% of the fitted models themselves (Table 3), i.e. models for volume prediction could be transferred between sites within the study region without loss of accuracy. The accuracy achieved is also comparable to values reported for other studies (e.g. Straub and Stepper 2016; Straub et al. 2013). There is potential for further improvement. Inclusion of tree species is expected to notably improve the accuracy of volume prediction. This information can theoretically be extracted from remote sensing data. However, so far no method for reliable tree species classification over large

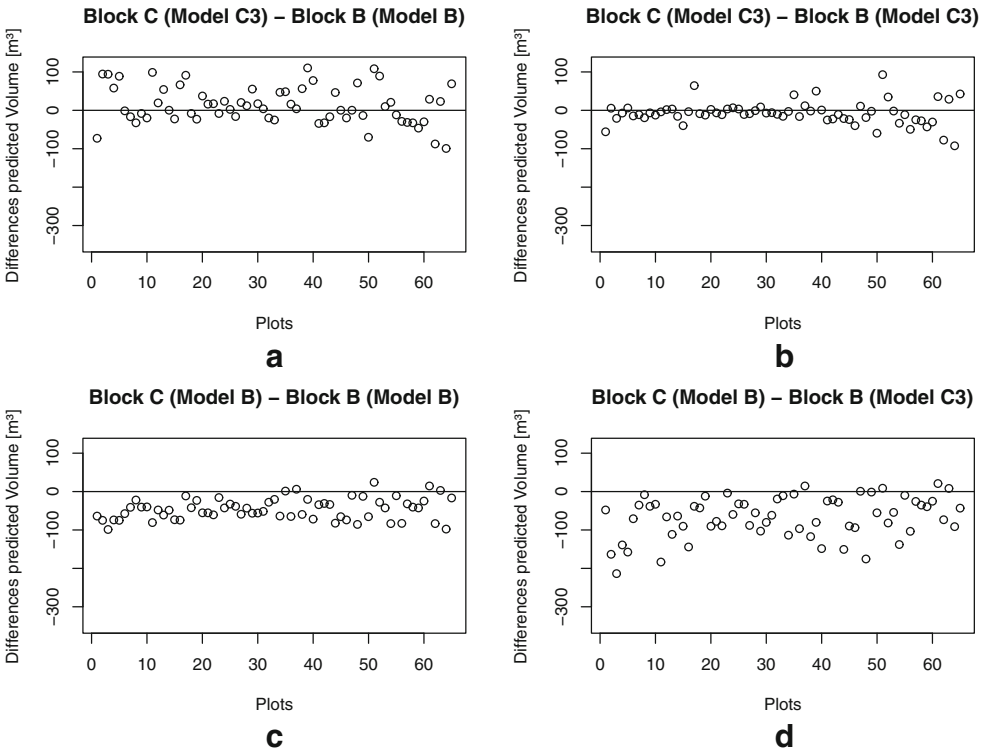

Fig. 12 Using refitted model C3: differences in predicted volume in overlap area BC for all combinations of models and canopy height models

areas is available. The recently broad availability of high resolution, multispectral satellite data by the Copernicus programme and related research efforts might soon provide a solution.

This study did not seek finding the best suited model for predicting timber volume based on canopy height model data, but investigating robustness of canopy height model metrics and general transferability of predictive models in the study region. That is why less complex linear models were used. Using a different type of models might be better suited for timber volume prediction, but this kind of investigation was not within the scope of this study. This might also help solving the issue over- and underestimation observed here and in other studies (Maack et al. 2015; Stepper et al. 2015).

Recent research suggests that timber volume can be modelled using remote sensing-based metrics independent of producing a canopy height model first (Giannetti et al. 2018). However, the mentioned study was conducted on a comparatively small site only and applicability to other sites needs to be proven. Additionally, there still would be the need to identify robust metrics for predictive models that can be transferred to different data sets.

5 Conclusion

The purpose of this study was to investigate the robustness of 3D metrics and the transferability of models between different canopy height model data sets in Southwest Germany. Metrics directly related to canopy height (e.g. meanCHM, height percentiles) were found most robust. For metrics influenced by the plot size such as the number of pixels used for metric calculation (npix) and the inner canopy volume (Volin), normalisation based on plot size is necessary in order to achieve robust metrics. It was also shown that the quality of aerial image-based 3D metrics is highly dependent on the data acquisition conditions. Different camera specifications were the reason for notable differences in canopy height model quality.

Predictive models based on robust metrics were transferred to other data sets in the study region without loss of accuracy. This shows potential for supporting harmonising of large-scale forest inventories in this region. Further research should test the applicability of these findings in different geographic regions as well as transferability of models between these regions. At the same time, the effect of using forest inventory data based on different sampling protocols as reference should also be investigated in order to find methods for supporting cross-regional forest inventory harmononisation.

Besides that, further improvements in prediction accuracy are necessary and could be achieved by including metrics related to tree type information.

Acknowledgements We acknowledge the provision of the remote sensing data by the land surveying authority of Baden-Württemberg (LGL).

Funding This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 633464 (DIABOLO).

Data availability The data that support the findings of this study are available from the land surveying authority of Baden-Württemberg (LGL) and Thünen-Institut, respectively, but restrictions apply to the availability of these data, which were used under licence for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of LGL and Thünen-Institut, respectively.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz (n.d.) BMEL - Federal Forest Inventory:Home. https://www.bundeswaldinventur.de/en/. Accessed 29 Oct 2018

COST Action E43 (n.d.) Harmonisation of National Forest Inventories in Europe: techniques for common reporting. https://www.cost.eu/actions/E43#tabs|Name:overview. Accessed 29 Oct 2018

COST Action FP1001 (2014) Improving data and information on the potential supply of wood resources—a European Approach from multisource National Forest Inventories. https://sites.google.com/site/costactionfp1001/. Accessed 26 Oct 2018

Deo RK, Froese RE, Falkowski MJ, Hudak AT (2016) Optimizing variable radius plot size and LiDAR resolution to model standing volume in conifer forests. Can J Remote Sens 42:428–442

DIABOLO (n.d.) Distributed, Integrated and Harmonised Forest Information for Bioeconomy Outlooks. http://diabolo-project.eu/. Accessed 26 Oct 2018

Fekety PA, Falkowski MJ, Hudak AT (2015) Temporal transferability of LiDAR-based imputation of forest inventory attributes. Can J For Res 45:422–435

GEOsat (n.d.-a) Mxbox. http://www.geosat.de/index.php/en/products/system-geometer-mx/geobox.html. Accessed 30 Oct 2018

GEOsat (n.d.-b) DGNSS-Sensor Mxbox hybrid – Technische Daten. http://geosat.de/images/stories/pdf/technik_MXbox_hybrid_28channel_201101.pdf. Accessed 30 Oct 2018

Giannetti F, Chirici G, Gobakken T, Næsset E, Travaglini D, Puliti S (2018) A new approach with DTM-independent metrics for forest growing stock prediction using UAV photogrammetric data. Remote Sens Environ 213:95–205

Ginzler C, Hobi ML (2015) Countrywide stereo-image matching for updating digital surface models in the framework of the Swiss National Forest Inventory. Remote Sens 7(4):4343–4370

Haala N (2014) Dense image matching final report. EuroSDR Official Publication, pp 115–145

Höhle J, Höhle M (2009) Accuracy assessment of digital elevation models by means of robust statistical methods. ISPRS J Photogramm Remote Sens 64:398–406

Immitzer M, Stepper C, Böck S, Straub C, Atzberger C (2016) Use of WorldView-2 stereo imagery and National Forest Inventory data for wall-to-wall mapping of growing stock. For Ecol Manag 359:232–246

Kirchhoefer M, Schumacher J, Adler P, Kändler G (2017) Considerations towards a novel approach for integrating angle-count sampling data in remote sensing based forest inventories. Forests 8:18

Leberl F, Irschara A, Pock T, Meixner P, Gruber M, Scholz S, Wiechert A (2010) Point clouds: Lidar versus 3D vision. Photogramm Eng Remote Sens 76:1123–1134

LGL (n.d.) Das digitale Geländemodell (DGM) von Baden-Württemberg. https://shop.lgl-bw.de/lvshop2/ProduktInfo/menuecontrol/frameSetProInfo.asp?keyInfo=dgm. Accessed 22 Jan 2018

Maack J, Kattenborn T, Fassnacht FE, Enßle F, Hernández J, Corvalán P, Koch B (2015) Modeling forest biomass using Very-High-Resolution data - combining textural, spectral and photogrammetric predictors derived from spaceborne stereo images. Eur J Remote Sens 48

McRoberts RE, Tomppo EO (2007) Remote sensing support for national forest inventories. Remote Sens Environ 110:412–419

McRoberts RE, Tomppo EO, Schadauer K, Vidal C, Ståhl G, Chirici G, Lanz A, Cienciala E, Winter S, Smith WB (2009) Harmonizing national forest inventories. J For 107(4):179–187

McRoberts RE, Tomppo EO, Næsset E (2010) Advances and emerging issues in national forest inventories. Scand J For Res 25:368–381

McRoberts RE, Tomppo EO, Schadauer K, Ståhl G (2012) Harmonizing national forest inventories. For Sci 58(3):189–190

McRoberts RE, Næsset E, Gobakken T, Bollandsås OM (2015) Indirect and direct estimation of forest biomass change using forest inventory and airborne laser scanning data. Remote Sens Environ 164:36–42

Moser P, Vibrans AC, McRoberts RE, Næsset E, Gobakken T, Chirici G, Mura M, Marchetti M (2016) Methods for variable selection in LiDAR-assisted forest inventories. For Int J For Res 90(1):112–124

nFrames (2018) SURE aerial. http://www.nframes.com/products/sure-aerial/. Accessed 22 Jan 2018

Rahlf J (2017) Forest resource mapping using 3D remote sensing: combining national forest inventory data and digital aerial photogrammetry. Norwegian University of Life Sciences

Rahlf J, Breidenbach J, Solberg S, Astrup R (2015) Forest parameter prediction using an image-based point cloud: a comparison of semi-ITC with ABA. Forests 6:4059–4071

Remondino F, Grazia Spera M, Nocerino E, Menna F, Nex F (2014) State of the art in high density image matching. Photogramm Rec 29:144–166

Riedel T, Hennig P, Kroiher F, Polley H, Schmitz F, Schwitzgebel F (2017) Die dritte Bundeswaldinventur (BWI2012). Inventur und Auswertmethoden:124 pages

Sass J (2011) GNSS Accuracy – specifications versus reality. ftp://ftp.ashtech.com/Spectra-precision/General%20Info/GNSS%20Accuracy%20-%20Specification%20versus%20Reality.pdf. Accessed 30 Oct 2018

Scrinzi G, Clementel F, Floris A (2015) Angle count sampling reliability as ground truth for area-based LiDAR applications in forest inventories. NRC Research Press

Sloboda J, Gaffrey D, Matsumura N (1993) Regionale und lokale Systeme von Höhenkurven gleichartiger Waldbestände. Allg Forst Jagdzeitung 164(12):225–228

Stepper C, Straub C, Pretzsch H (2015) 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 45:111–123

Stepper C, Straub C, Immitzer M, Pretzsch H (2017) 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 32:748–761

Straub C, Stepper C (2016) Using digital aerial photogrammetry and the random forest approach to model forest inventory attributes in Beech-and Spruce-dominated Central European Forests. Photogramm Fernerkund Geoinf 2016:109–123

Straub C, Stepper C, Seitz R, Waser LT (2013) Potential of UltraCamX stereo images for estimating timber volume and basal area at the plot level in mixed European forests. Can J For Res 43:731–741

Tomppo EO, Schadauer K (2012) Harmonization of National Forest Inventories in Europe: advances under COST Action 43. For Sci 58(3):191–200

Tomppo E, Gschwantner T, Lawrence M, McRoberts RE (eds) (2010) National forest inventories. Pathways for Common Reporting. Springer, Heidelberg

Ullah S, Dees M, Datta P, Adler P, Koch B (2017) Comparing airborne laser scanning, and image-based point clouds by semi-global matching and enhanced automatic terrain extraction to estimate forest timber volume. Forests 8:215

Véga C, Renaud J-P, Durrieu S, Bouvier M (2016) On the interest of penetration depth, canopy area and volume metrics to improve Lidar-based models of forest parameters. Remote Sens Environ 175:32–42

Vexcel (n.d.) Vexcel imaging – home of the UltraCam. https://www.vexcel-imaging.com/. Accessed 29 Oct 2018

Webster R, Oliver MA (2007) Geostatistics for environmental scientists (statistics in practice), 2nd edn. John Wiley & Sons, Chichester

White JC, Stepper C, Tompalski P, Coops NC, Wulder MA (2015) Comparing ALS and image-based point cloud metrics and modelled forest inventory attributes in a complex coastal forest environment. Forests 2016:6

White JC, Coops NC, Wulder MA, Vastaranta M, Hilker T, Tompalski P (2016) Remote sensing technologies for enhancing forest inventories: a review. Can J Remote Sens 42:619–641

Zald HS, Wulder MA, White JC, Hilker T, Hermosilla T, Hobart GW, Coops NC (2016) Integrating Landsat pixel composites and change metrics with lidar plots to predictively map forest structure and aboveground biomass in Saskatchewan, Canada. Remote Sens Environ 176:188–201