Appendix

I. Software

The software used for data processing can be summarized as follows:

- PyCharm Python IDE 2019.2.1 with Miniconda3 4.3.31 and Python 3.6.7: Installed packages were (in this order): pandas 0.24.2, scipy 1.2.1, xlrd 1.2.0, xlwt 1.3.0, matplotlib 3.0.3, scikit-learn 0.20.3, scikit-image 0.15.0, cx-Oracle 7.0.0, py4j 0.10.8.1, statsmodels 0.9.0, open-cv 4.1.0.25, [GDAL 2.4.1, Fiona 1.8.6, pyproj 2.1.3, Rtree 0.8.3, Shapely 1.6.4 as Gohlke Wheelfiles], geopandas 0.5.0, pyshp 2.1.0, xgboost 0.90, psutil 5.6.3, patsy 0.5.1, seaborn 0.9.0, tensorflow 1.12, pillow 6.0.0 and Cython 0.29.12
- LAStools v. 4.4.2019 (Rapidlasso GmbH, 2019).
- Octave 5.1.0 with the Digital Forestry Toolbox by (Parkan, 2018).
- RStudio 1.2 user interface with R 3.6.1 (R Core Team, 2019): Installed packages were: boot, broom, dismo, dplyr, ggplot2, minpack.lm, nlstools, pander, RColorBrewer, reshape2 and xlsx
- QGIS 3.6

II. Supplementary methodical explanations

Allometry data: additional information (section 2.2)

From the EFM dataset 173’382 single-tree measurements were used. These were recorded at different plots in Switzerland over more than a century. For this study all measurements were used, without e.g. selection according to different regions or altitude levels. The TUM dataset comprised 39’624 single-tree measurements, 37’707 of which originated from long-term research plots in southern Germany. In 1’917 cases, park and city trees all over the world were sampled (Pretzsch et al., 2015). For the present study, the attributes DBH, tree species and crown radius of all 39’624 trees were used. The crown radius was either derived from average values of measurements in four orthogonal, horizontal directions (WSL dataset) or from root-mean-squared error values of measurements in eight cardinal points (N, NW, ..., NE; TUM dataset).

ALS data pre-processing: additional information (section 2.4)

The entire processing of the ALS data was carried out using the LAStools v. 4.4.2019 software (Rapidlasso GmbH, 2019). Names of LAStools tools are hereafter highlighted in typewriter font. With `lasclip`, the ALS data were cut to the search window of each sample-plot, and with `lasnoise` outliers were classified as background noise and excluded from further processing (class 7). Subsequently, ground points (class 2) were identified with `lasground` and the remaining points were classified with `lasclassify` (classes: unclassified 1; low vegetation 3; high vegetation 5; buildings 6). The calculation of a digital terrain model (DTM) with a resolution of 50 cm was done with `las2dem`. With the same tool and the pit-free algorithm, a digital surface model (DSM) and a normalized DSM (nDSM or CHM) with a resolution of 50 cm were created. Finally, the classified ALS point-cloud was normalized with `lasheight` and points with a negative normalized height were excluded.

Single-tree-identification (STI) methods: additional information (section 2.10)

The methods of Menk et al. (2017) and Kaartinen et al. (2012) both rely on filtering of the CHM prior to LM detection. Menk et al. (2017) used a] Gaussian filter with a raster-resolution-dependent standard deviation of $\sigma = 0$ (no smoothing) for resolutions (res) ≥ 0.5 m, $\sigma = 1$ for res ≥ 0.5 and < 1, and $\sigma = 2$ for res < 0.5 (method `LM1`). The following Gauss kernel (method `LM2`) was applied by (Kaartinen et al., 2012):

$$\frac{1}{28}\begin{bmatrix} 1 & 3 & 1 \\ 3 & 12 & 3 \\ 1 & 3 & 1 \end{bmatrix}$$

In another LM method, Parkan (2018) first calculated a CHM based on the ALS point-cloud, which was then smoothed with a $2.4\ x\ 2.4$ m Gaussian filter. Afterwards, single trees with a minimum height of 2 m and a minimum inter-tree distance of $0.28 - h^{0.59}$ m (Chen et al., 2006) were identified, where h is the height of the higher of the two tree candidate points that were read from the CHM (method `LM3`).

A stem-detection-based STI for deciduous stands was performed according to the procedure proposed in the `Digital Forestry Toolbox`. The algorithm uses normalized leaf-off ALS last returns, which belong to class 4 or 5 (medium vegetation or high vegetation), to identify stems at positions of high ALS point density (method `Stems`).

Shannon entropy and GLCM: additional information (section 2.13)

The Shannon entropy is defined as $H = -\sum_k p_k * log_2(max(p_k, 1))$, where p_k is the probability that a pixel has the value k (Singh & Singh, 2008). The result depends on the underlying image data type (e.g. 8-bit integer with $k \in [0, 256)$ or 16-bit integer with $k \in [0, 65'536)$). In Bremgarten and Zurich 16-bit integer orthoimages were used.

GLCM contains information about the texture in an image. It has the four dimensions i, j, d and Θ. The matrix-values $p_{i,j,d,\Theta} \in [0, 1]$ give the probability that a pixel j at a distance d and an angle Θ occurs relative to the pixel value i (Soh & Tsatsoulis, 1999). In this study, d was defined as one pixel width (Bremgarten 25 cm; Zurich 10 cm) and Θ had a value of 0°. Thus, the dimensions of the GLCM were reduced to two (i, j). The 16-bit orthoimages were converted to 8-bit images prior to calculating the GLCM for each of the four orthoimage bands (RGB and NIR).

Removal of highly correlated features: additional information (section 2.14.1)

To get a subset of predictors that were not highly correlated, the following algorithm was repeated 1,000 times, counting how often each predictor occurred in these 1,000 runs. Starting with one predictor, further predictors were randomly added as long as their correlation with all of the previously chosen predictors was < 0.7. Finally, only those predictors which had been added more than 0.15 times as often as the one that was most frequently chosen in the 1,000 runs were kept in the subset. Since strongly correlated predictors could also occur in this subset, the algorithm was applied one more time to the subset with a correlation limit of 0.85 to ultimately exclude one of two predictors with a correlation > 0.85.

III. Supplementary tables

Table 5: This table shows the parameters a, b and c of the regression model equations, which were selected for each tree species (number of observations in the second column) to model the relationship between DBH and tree height (top) or DBH and crown size (bottom). σ denotes the standard deviation of the residuals. The $\mathbf{R^2}$ is only given for the linear models, since its interpretation is questionable for the *non*-linear models (Spiess & Neumeyer, 2010). The underlying model formulation, chosen by AIC comparison (see Table 6 and 7) is indicated in the column on the right.

DBH – Tree height

Tree species	n-Obs.	a	b	c	σ [m]	R^2	Model
Fagus sylvatica	36 381	5.695×10^{1}	2.452×10^{-2}	1.042	3.496	-	Logistic mod.
Quercus sp.	4 125	3.348×10^{1}	3.079×10^{-2}	1.127	2.763	-	Weibull
Acer sp.	4 447	7.970×10^{1}	-6.306×10^{1}	1.940×10^{1}	3.266	-	Exponential
Fraxinus excelsior	1 931	6.121×10^{1}	-3.673×10^{1}	1.349×10^{1}	3.142	-	Exponential
Tilia sp.	480	3.212×10^{1}	1.652×10^{-2}	1.250	2.510	-	Weibull
Prunus avium	785	9.102×10^{1}	-6.990×10^{1}	2.035×10^{1}	2.162	-	Exponential
Ulmus sp.	88	2.962×10^{2}	5.889	2.377×10^{-1}	3.057	-	Korf
Other broadleaf	21 368	3.313	7.984×10^{-1}	5.440×10^{-3}	3.362	0.806	Linear
Picea abies	57 592	4.923	1.001×10^{-2}	1.348	3.805	-	Logistic mod.
Abies alba	18 485	6.311×10^{1}	9.287	6.565×10^{-1}	3.335	-	Korf
Pinus sylvestris	19 078	2.673	7.749×10^{-1}	4.561×10^{-3}	3.770	0.656	Linear
Larix sp.	16 790	1.325	8.897×10^{-1}	5.301×10^{-3}	3.219	0.877	Linear
Other coniferous	31 456	2.289×10^{2}	7.392	3.378×10^{-1}	3.788	-	Korf

DBH – Crown size

Tree species	n-Obs.	a	b	c	σ [m]	R^2	Model
Fagus sylvatica	36 381	1.312	6.824×10^{-2}	1.177×10^{-4}	8.852×10^{-1}	0.637	Linear
Quercus sp.	4 125	3.280×10^{1}	-1.777×10^{2}	4.074×10^{1}	7.426×10^{-1}	-	Exponential
Acer sp.	4 447	2.135×10^{4}	1.473×10^{-5}	6.967×10^{-1}	4.824×10^{-1}	-	Logistic mod.
Fraxinus excelsior	1 931	1.883×10^{-1}	1.085×10^{-1}	-1.612×10^{-4}	7.128×10^{-1}	0.803	Linear
Tilia sp.	480	1.319×10^{1}	-5.520×10^{1}	1.752×10^{1}	5.186×10^{-1}	-	Exponential
Prunus avium	785	1.370×10^{1}	-6.011×10^{1}	1.599×10^{1}	4.224×10^{-1}	-	Exponential
Ulmus sp.	88	1.448	8.045×10^{-2}	-9.470×10^{-6}	6.836×10^{-1}	0.745	Linear
Other broadleaf	21 368	3.008×10^{-1}	1.218×10^{-1}	-3.961×10^{-4}	7.061×10^{-1}	0.824	Linear
Picea abies	57 592	2.024×10^{1}	-2.143×10^{2}	6.832×10^{1}	4.730×10^{-1}	-	Exponential
Abies alba	18 485	9.646×10^{-1}	6.316×10^{-2}	-1.797×10^{-4}	6.270×10^{-1}	0.729	Linear
Pinus sylvestris	19 078	2.729×10^{1}	-2.125×10^{2}	5.781×10^{1}	4.623×10^{-1}	-	Exponential
Larix sp.	16 790	1.163×10^{1}	-9.742×10^{1}	3.534×10^{1}	4.958×10^{-1}	-	Exponential
Other coniferous	31 456	9.450×10^{1}	6.275	1.766×10^{-1}	5.783×10^{-1}	-	Korf

Table 6: Akaike's Information Criterion (AIC) of the DBH - tree height regression models. For each species the one model with the smallest AIC value was chosen for further modelling.

Model	*Fagus*	*Quercus*	*Acer*	*Fraxinus*	*Tilia*	*Prunus*	*Ulmus*
Linear	1.9515×10^5	2.1005×10^4	2.3178×10^4	9.9797×10^3	2.2512×10^3	3.4650×10^3	4.5476×10^2
Chapman	1.9434×10^5	2.0104×10^4	2.3218×10^4	9.9194×10^3	2.2539×10^3	3.4905×10^3	4.5231×10^2
Weibull	1.9433×10^5	2.0097×10^4	2.3220×10^4	9.9185×10^3	2.2508×10^3	3.4905×10^3	4.5217×10^2
Exponential	1.9439×10^5	2.0211×10^4	2.3151×10^4	9.9065×10^3	2.2560×10^3	3.4429×10^3	4.5312×10^2
Logistic mod.	1.9432×10^5	2.0158×10^4	2.3222×10^4	9.9176×10^3	2.2579×10^3	3.4905×10^3	4.5193×10^2
Korf	1.9433×10^5	2.0341×10^4	2.3231×10^4	9.9253×10^3	2.2710×10^3	3.4943×10^3	4.5136×10^2

Model	*Other coniferous*	*Picea*	*Abies*	*Pinus*	*Larix*	*Other broadleaf*
Linear	1.7346×10^5	3.1898×10^5	1.0013×10^5	1.0479×10^5	8.6915×10^4	1.1246×10^5
Chapman	1.7326×10^5	3.1747×10^5	9.7731×10^4	1.0495×10^5	8.6946×10^4	1.1249×10^5
Weibull	1.7327×10^5	3.1753×10^5	9.7803×10^4	1.0495×10^5	8.6918×10^4	1.1249×10^5
Exponential	1.7329×10^5	3.1739×10^5	9.7179×10^4	1.0484×10^5	8.7041×10^4	1.1252×10^5
Logistic mod.	1.7319×10^5	3.1737×10^5	9.7330×10^4	1.0499×10^5	8.7045×10^4	1.1260×10^5
Korf	1.7306×10^5	3.1759×10^5	9.6992×10^4	1.0505×10^5	8.7312×10^4	1.1280×10^5

Table 7: Akaike's Information Criterion (AIC) of the DBH – crown size regression models. For each species the one model with the smallest AIC value was chosen for further modelling.

Model	*Fagus*	*Quercus*	*Acer*	*Fraxinus*	*Tilia*	*Prunus*	*Ulmus*
Linear	8.8517×10^{-1}	7.5588×10^{-1}	4.8411×10^{-1}	7.1285×10^{-1}	5.1873×10^{-1}	4.2425×10^{-1}	6.8362×10^{-1}
Chapman	9.0215×10^{-1}	7.5726×10^{-1}	4.8240×10^{-1}	7.1313×10^{-1}	5.1910×10^{-1}	4.2544×10^{-1}	7.0289×10^{-1}
Weibull	9.0215×10^{-1}	7.5722×10^{-1}	4.8240×10^{-1}	7.1313×10^{-1}	5.1913×10^{-1}	4.2547×10^{-1}	7.0289×10^{-1}
Exponential	8.8568×10^{-1}	7.4256×10^{-1}	4.8548×10^{-1}	7.1385×10^{-1}	5.1857×10^{-1}	4.2238×10^{-1}	6.8400×10^{-1}
Logistic mod.	9.0215×10^{-1}	7.5724×10^{-1}	4.8240×10^{-1}	7.1313×10^{-1}	5.1915×10^{-1}	4.2549×10^{-1}	7.0289×10^{-1}
Korf	9.0368×10^{-1}	7.5853×10^{-1}	4.8270×10^{-1}	7.1353×10^{-1}	5.1933×10^{-1}	4.2583×10^{-1}	7.0456×10^{-1}

Model	*Other coniferous*	*Picea*	*Abies*	*Pinus*	*Larix*	*Other broadleaf*
Linear	5.8148×10^{-1}	4.7337×10^{-1}	6.2700×10^{-1}	4.6256×10^{-1}	4.9605×10^{-1}	7.0609×10^{-1}
Chapman	5.7861×10^{-1}	4.8214×10^{-1}	6.3474×10^{-1}	4.6761×10^{-1}	4.9914×10^{-1}	7.0683×10^{-1}
Weibull	5.7850×10^{-1}	4.8214×10^{-1}	6.3474×10^{-1}	4.6761×10^{-1}	4.9914×10^{-1}	7.0688×10^{-1}
Exponential	5.7884×10^{-1}	4.7304×10^{-1}	6.2764×10^{-1}	4.6227×10^{-1}	4.9580×10^{-1}	7.0676×10^{-1}
Logistic mod.	5.7839×10^{-1}	4.8214×10^{-1}	6.3474×10^{-1}	4.6760×10^{-1}	4.9914×10^{-1}	7.0693×10^{-1}
Korf	5.7829×10^{-1}	4.8320×10^{-1}	6.3544×10^{-1}	4.6846×10^{-1}	4.9949×10^{-1}	7.0727×10^{-1}

IV. Supplementary figures

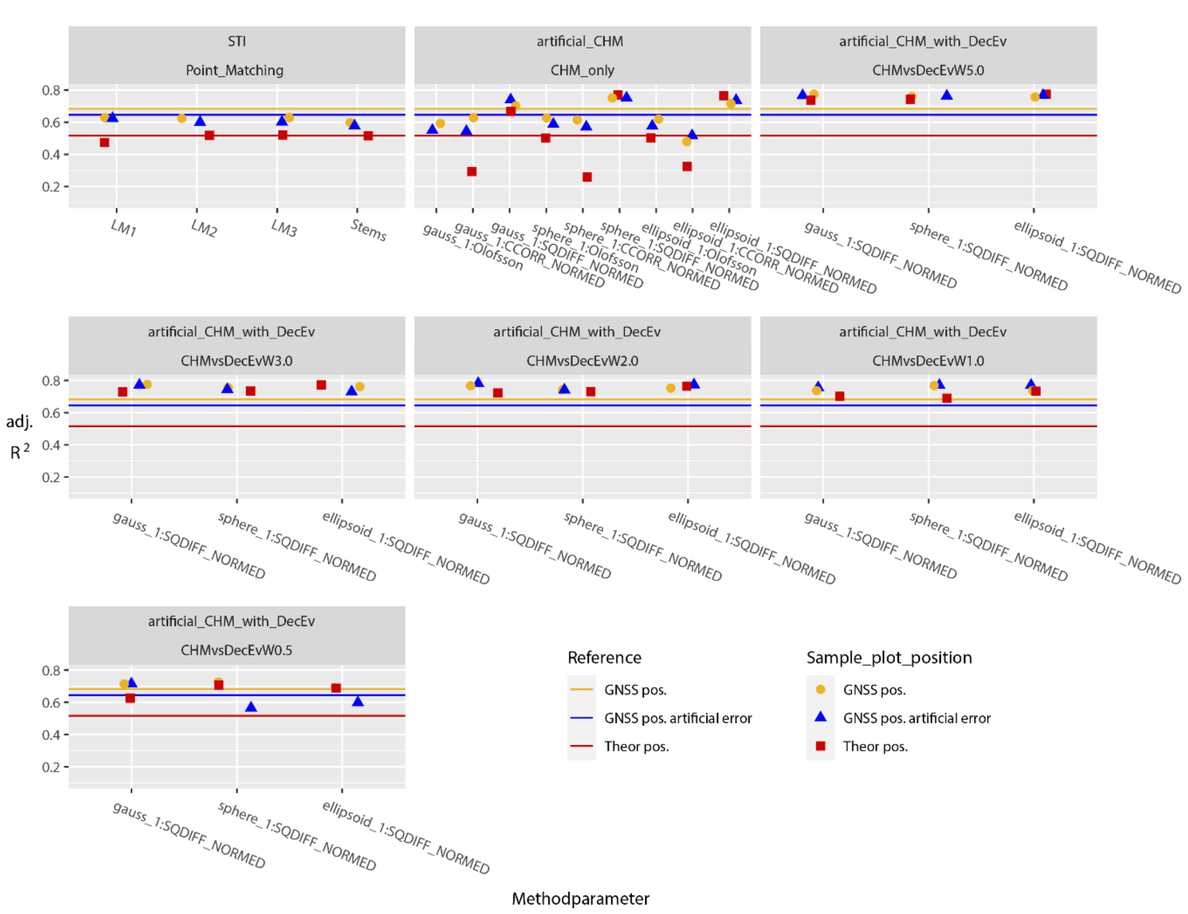

Figure 16: Comparison of co-registration methods (horizontal axis) based on the leave-one-out cross-validated adjusted R^2 (vertical axis) of the timber volume in Bremgarten. The values shown are the average values of three OLS regression model calibration runs, including predictor variable selection and random removal of strongly correlated predictors. The colour indicates the different co-registration starting positions. Horizontal lines show the reference values without co-registration. To minimize overlapping, all points were slightly shifted horizontally.

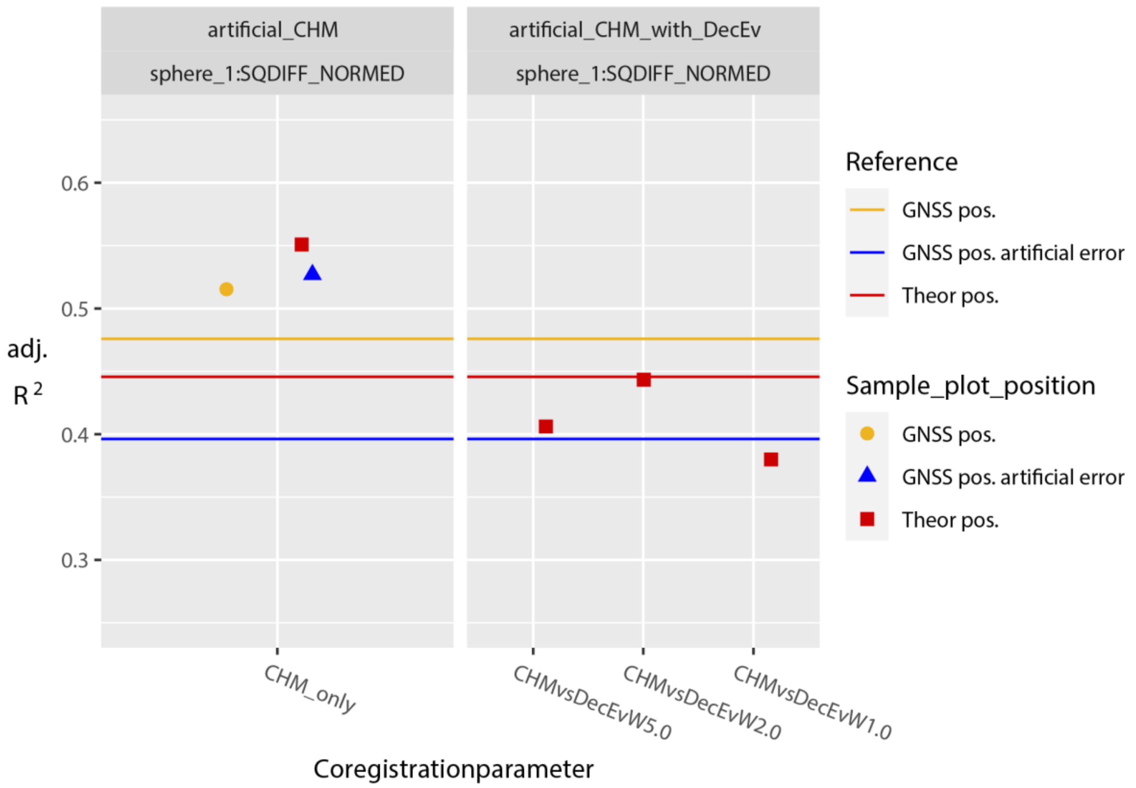

Figure 17: Application of the co-registration method that performed best in Bremgarten to the Zurich dataset, with and without taking the DecEv raster into account. The figure shows the leave-one-out cross-validated adjusted R^2 (vertical axis) of the OLS modelled BA in Zurich. The reference is given by the non-co-registered sample-plots (horizontal lines). All values are averages from three independent calibration runs, including predictor variable selection and random removal of strongly correlated predictors. The colour denotes the respective co-registration starting position. To minimize overlapping, all points were slightly shifted horizontally.

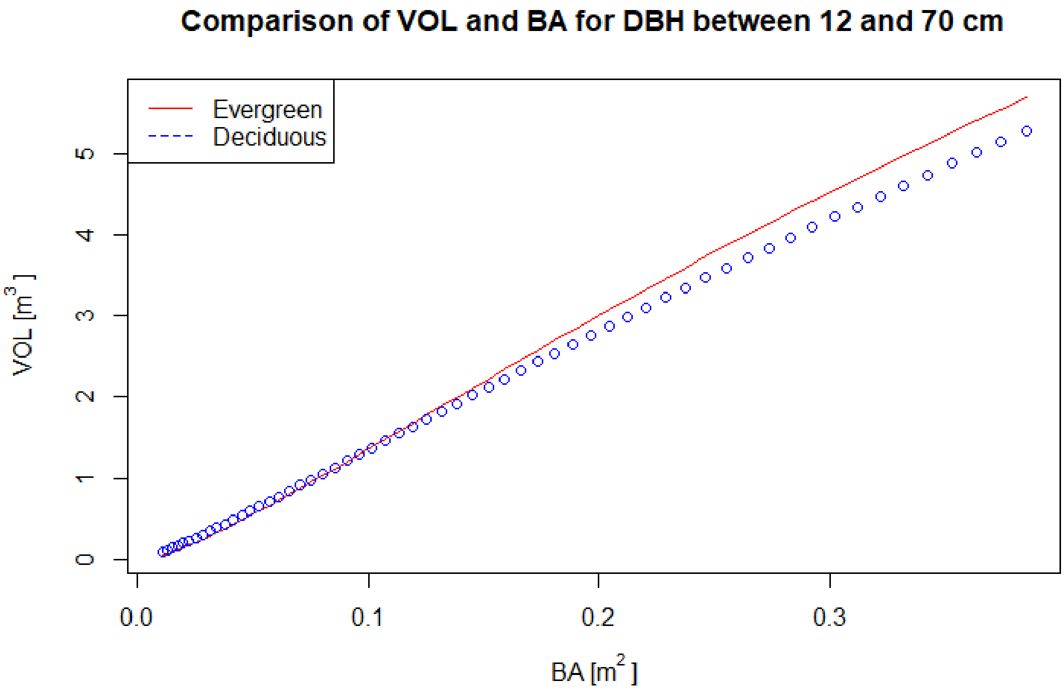

Figure 18: Comparison of the Basal Area (BA) and Volume (VOL) for DBH values of single trees between 12 and 70 cm. The volume was calculated according to the tariff functions, that were applied in Bremgarten. The resulting curves are almost linear.